



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

INSTITUTO DE BIOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM BIODIVERSIDADE E BIOLOGIA

EVOLUTIVA – MESTRADO

MARCOS DE OLIVEIRA RAFAEL DA CRUZ

**ANÁLISE FILOGENÉTICA DA FAMÍLIA CRICETIDAE (RODENTIA:
MUROIDEA) A PARTIR DO GENOMA MITOCONDRIAL**

Rio de Janeiro – RJ

Julho de 2019

MARCOS DE OLIVEIRA RAFAEL DA CRUZ

**ANÁLISE FILOGENÉTICA DA FAMÍLIA CRICETIDAE (RODENTIA:
MUROIDEA) A PARTIR DO GENOMA MITOCONDRIAL.**

**Dissertação de mestrado apresentada no
Instituto de Biologia da Universidade
Federal do Rio de Janeiro como requisito
para obtenção do título de mestrado em
Biodiversidade e Biologia Evolutiva.**

ORIENTADOR: MARCELO WEKSLER

BANCA EXAMINADORA

Dra. Beatriz Mello Carvalho (Doutora em Genética) - UFRJ

Dr. João Alves de Oliveira (Doutor em Zoologia) - UFRJ

Dr. William Corrêa Tavares (Doutor em Genética) – UFRJ

Agradecimentos

Agradeço aos meus pais Rogério e Eliane por sempre apoiarem os meus estudos. Mesmo em períodos de muita dificuldade financeira vocês nunca me negaram um livro. Vocês são uma das razões pelo qual eu amo livros. Desde pequeno vocês me ensinaram que é possível viajar e expandir meu universo através da leitura. Além disso, sinto-me grato por vocês me darem a liberdade de seguir meu próprio caminho mesmo não concordando com alguns dos meus pontos de vista sobre quais escolhas tomar e como pensar sobre a vida. A liberdade é um dos aspectos que eu mais prezo na vida e vocês sempre me fizeram sentir livre a optar por fazer aquilo que eu julgava ser o melhor para mim. Essa liberdade e apoio nos estudos são essências para eu trilhar o meu caminho. Muito obrigado por vocês estarem ao meu lado me oferecendo apoio emocional em todos os momentos. Eu amo muito vocês.

Agradeço ao meu orientador Marcelo que me inspira a tentar me tornar um pesquisador melhor e a perseverar quando quase tudo não acontece como planejado. Você me passa a impressão que o mais importante é perseverar na busca de soluções e a manter o foco nessa busca. Essa atitude sua me deixa confortável a perseverar mesmo quando cometo alguns equívocos pelo caminho. Em poucas palavras, você me inspira a tentar sem ficar com muito medo de errar. Obrigado pela confiança que você deposita em mim e por me ajudar a manter ou recuperar a calma quando surgem obstáculos pelo caminho. Apesar dos vários projetos que você lidera você sempre encontra um tempo para me orientar nessa jornada acadêmica, e sou muito grato por isso também. Você foi essencial para eu conseguir finalizar esse trabalho e alguns outros trabalhos que eu me dediquei nos últimos anos. Sinto-me privilegiado em trabalhar com um cientista tão

acima da média como você. Em resumo, você é o melhor orientador que eu já tive e eu sentirei muito a sua falta quando não pudermos trabalhar juntos rotineiramente.

Agradeço ao nosso colaborador Francisco por ceder os servidores que me permitiram montar os genomas mitocondriais que eu utilizei nesse projeto. Também sou muito grato ao Francisco por me ajudar a aprender montar/anotar mitogenomas que eu utilizei nesse projeto, a navegar pela linha de comando do LINUX e me inspirar a desenvolver mais scripts no Python. Gostei muito de ter passado alguns momentos com você me auxiliando a escrever scripts. Obrigado por ter estado disposto a me orientar em um momento complicado da minha vida acadêmica. Você foi essencial para eu conseguir desenvolver esse projeto.

Agradeço aos meus amigxs de laboratório Aldo, Ana Lazar, Ana Pantaleão, Augusto, Carol, Cintia, Filipe, Giulia, João, Karina, Maíra, Mariana, Patrícia, Paula, Rafael, Roberto, Samantha e Sérgio, por sempre oferecerem um apoio moral com palavras de incentivo e por estarem dispostos a me ajudar a realizar algumas tarefas. Nesses últimos anos vocês foram muito importantes para eu conseguir recuperar o equilíbrio emocional em alguns momentos. Eu passei mais tempo com vocês do que com qualquer outra pessoa durante o período de mestrado, conseqüentemente vocês tiveram um impacto significativo e muito positivo na minha jornada acadêmica. Obrigado também por organizarem as confraternizações para comemorar meus aniversários. O ambiente de trabalho fica muito mais agradável e leve com a presença de vocês e com essas festinhas ocasionais. Eu me sinto muito bem acolhido por vocês todos. Eu sou uma pessoa muito introspectiva, calada e desligada do que está acontecendo a minha volta na maior parte do tempo, e por isso às vezes eu fico com a sensação que eu não sou muito recíproco a todo apoio que vocês me oferecem, mas saibam que eu gosto muito de vocês e fico extremamente feliz quando posso ajudar de

alguma maneira. Eu sentirei muito a falta de todos vocês quando a vida nos colocar em caminhos diferentes.

Também agradeço o apoio dos meus amigxs Ana, Gabriel, Deise, Larissa e Matheus que eu conheci no laboratório do Francisco e que frequentei bastante durante um período. Apesar de não nos encontrarmos há um bom tempo devido a correria da vida acadêmica, eu tenho boas lembranças dos momentos que passamos juntos tentando aprender navegar pela linha de comando do LINUX para montar/anotar mitogenomas. Eu já sinto muito a falta de vocês. Vocês são ótimas pessoas que me deram um apoio fundamental para eu continuar seguindo na vida acadêmica em um momento complicado da minha jornada.

Agradeço meus amigxs da faculdade Gabriella, Hanna, Julia França, Julia Niemeyer, Lisiane, Maria Rachel Tiago e Victor. Estamos juntos há 10 anos; hoje em dia um pouco mais distantes do que juntos devido às circunstâncias da vida. Apesar de eu estar um pouco mais ausente nos últimos anos, o mais importante é que ainda nos encontramos quando as circunstâncias permitem. Vocês já estiveram comigo em vários momentos de felicidade e dificuldade e eu sempre soube que eu poderia contar com vocês para me ajudar mesmo em períodos que estávamos um pouco mais distantes. Todos os momentos que eu passei com vocês na faculdade e fora dela me ajudaram indiretamente ou diretamente a trilhar o caminho que sigo hoje. Eu me sinto um sortudo de ter encontrado vocês nesta vida. Eu já sinto a falta de vocês.

Finalmente gostaria de agradecer aos professores que aceitaram a participar da banca avaliadora e pelo apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior CAPES através da bolsa de mestrado.

Resumo

A família Cricetidae (Rodentia: Muroidea) é a segunda mais diversa entre os mamíferos, com mais de 700 espécies distribuídas em todos continentes do mundo exceto Oceania e Antártida. A família é composta por cinco subfamílias, Arvicolinae, Cricetinae, Neotominae, Sigmodontinae e Tylomyinae, e trabalhos anteriores que buscaram inferir as relações evolutivas entre estes grupos de Muroidea apresentaram resultados contrastantes e em geral com baixo grau de suporte em termos de bootstrap e probabilidade posterior. O objetivo central do presente estudo é inferir as relações filogenéticas entre as subfamílias de Cricetidae baseando-se em genomas mitocondriais completos. As análises filogenéticas empregaram métodos de máxima verossimilhança e inferência bayesiana e contaram com 57 mitogenomas (aproximadamente 16,000 pares de base), incluindo oito representantes da família Cricetidae cujos genomas mitocondriais foram sequenciados neste estudo, 43 espécies de Cricetidae cujos genomas mitocondriais foram retirados do Genbank, e seis grupos externos (Dipodoidea e Muroidea não Cricetidae). As análises filogenéticas foram baseadas em conjuntos de dados com diferentes concatenações de classes de dados (genes codificadores de proteínas, genes codificadores de rRNA, genes codificadores de tRNA, D-loop e aminoácidos) e diferentes esquemas de partições (partição única, por posição do códon, por gene e posição do códon). Também foram realizadas análises de máxima verossimilhança fixadas (*constrained*) às 105 topologias bifurcadas que representavam as hipóteses de relacionamento entre as subfamílias de Cricetidae. As análises baseadas em diferentes esquemas de partição e alinhamentos com diferentes classes de dados produziram quatro topologias subfamiliares com diferentes graus de suporte, medidos

por bootstrap e probabilidade posterior. Contudo, todas as análises recuperaram consistentemente Sigmodontinae como grupo irmão de Tylomyinae, e Arvicolinae como grupo irmão de Cricetinae, com médio ou alto grau de suporte. O posicionamento filogenético de Neotominae foi conflitante nas análises de máxima verossimilhança e inferência bayesiana, mas a maior parte das análises, incluindo aquela fixada, sugere que Neotominae forma o grupo irmão do clado contendo as outras subfamílias, Arvicolinae, Cricetinae, Sigmodontinae e Tylomyinae. Análises adicionais com maior abrangência taxonômica e combinando estes dados com genes nucleares e morfologia provavelmente esclarecerão melhor essas questões filogenéticas.

Palavras chave: sequenciamento de nova geração, Illumina, genômica.

Abstract

The family Cricetidae (Rodentia: Muroidea) is the second most diverse family within mammals with more than 700 species distributed throughout the world except Oceania and Antarctica. The family is comprised of 5 subfamilies: Arvicolinae, Cricetinae, Neotominae, Sigmodontinae and Tylomyinae. Previous studies that inferred phylogenetic relationships among Muroidea groups including Cricetidae subfamilies, generally found contrasting results with low bootstrap and posterior probabilities support. The main objective of the present study is to infer phylogenetic relationships among Cricetidae subfamilies based on the complete mitochondrial genome. Phylogenetic analyses were based on maximum likelihood and bayesian inference including 57 mitogenomes (approximately 16,000 base pairs) represented by 8 new

mitogenomes of Cricetidae species sequenced in this study, 43 mitogenomes of Cricetidae species downloaded from GenBank and 6 outgroup representatives (Dipodoidea and Muroidea excluding Cricetidae). Phylogenetic analyses were based on different datasets that included protein coding sequences, rRNA coding sequences, tRNA coding sequences, D-loop control region, amino acid sequences; analyses were also based on different partition schemes (single partition, partitioned by codon position, partitioned by gene and codon position). Furthermore, phylogenetic analyses constrained to 105 bifurcated topologies that represented evolutionary relationship hypotheses among Cricetidae subfamilies were also performed. Analyses based on different partition schemes and datasets resulted in four subfamilial topologies with different bootstrap and posterior probability support. Nonetheless, most of the analyses consistently recovered Sigmodontinae as sister group to Tylomyinae and Arvicolinae as sister group to Cricetinae forming clades with medium or high support. The phylogenetic position of Neotominae was different among some maximum likelihood and bayesian analyses, but most of them, including the constrained analyses, recovered Neotominae as sister group to the clade comprised by Tylomyinae, Sigmodontinae, Arvicolinae and Cricetinae. Further studies including a broader taxonomic sampling combined with nuclear, mitochondrial genes and morphological data will probably bring more resolution to phylogenetic conundrum.

Keywords: next generation sequencing, Illumina, genomics.

Lista de figuras

- Figura 1:** Mapa do genoma mitocondrial da espécie *Oligoryzomys stramineus*. As regiões em vermelho, amarelo e rosa se referem a genes codificadores de rRNA, genes codificadores de proteínas e genes codificadores de tRNA, respectivamente. 52
- Figura 2:** Mapa do genoma mitocondrial da espécie *Irenomys tarsalis*. As regiões em vermelho, amarelo e rosa se referem a genes codificadores de rRNA, genes codificadores de proteínas e genes codificadores de tRNA, respectivamente. 53
- Figura 3:** Mapa do genoma mitocondrial da espécie *Delomys dorsalis*. As regiões em vermelho, amarelo e rosa se referem a genes codificadores de rRNA, genes codificadores de proteínas e genes codificadores de tRNA, respectivamente. 53
- Figura 4:** Mapa do genoma mitocondrial da espécie *Tylomys nudicaudus*. As regiões em vermelho, amarelo e rosa se referem a genes codificadores de rRNA, genes codificadores de proteínas e genes codificadores de tRNA, respectivamente. 54
- Figura 5:** Mapa do genoma mitocondrial da espécie *Sigmodon alstoni*. As regiões em vermelho, amarelo e rosa se referem a genes codificadores de rRNA, genes codificadores de proteínas e genes codificadores de tRNA, respectivamente. 55
- Figura 6:** Mapa do genoma mitocondrial da espécie *Scotinomys teguina*. As regiões em vermelho, amarelo e rosa se referem a genes codificadores de rRNA, genes codificadores de proteínas e genes codificadores de tRNA, respectivamente. 56
- Figura 7:** Mapa do genoma mitocondrial da espécie *Rhipidomys mastacalis*. As regiões em vermelho, amarelo e rosa se referem a genes codificadores de rRNA, genes codificadores de proteínas e genes codificadores de tRNA, respectivamente. 56
- Figura 8:** Mapa do genoma mitocondrial da espécie *Nyctomys sumichrasti*. As regiões em vermelho, amarelo e rosa se referem a genes codificadores de rRNA, genes codificadores de proteínas e genes codificadores de tRNA, respectivamente. 57
- Figura 9:** Árvores com topologias bifurcadas que representam as 105 hipóteses de relacionamento entre as subfamílias de Cricetidae. O identificador numérico para cada topologia se encontra acima das árvores. 59
- Figura 10:** Árvore de máxima verossimilhança baseada no alinhamento de rRNA-CDS-tRNA particionado por gene. Os números se referem ao suporte do bootstrap e os números em azul representam o suporte do bootstrap para os principais clados dentro de Cricetidae (ver Tabela 6). 60

Figura 11: Árvore bayesiana inferida baseando-se no alinhamento de rRNA-CDS-tRNA particionado por gene. Os números se referem à probabilidade posterior e os números em azul representam a probabilidade posterior para os principais clados dentro de Cricetidae (ver Tabela 6). 61

Figura 12: Gráfico para os valores de bootstrap do clado B encontrados nas análises filogenéticas baseadas em uma determinada classe de dados que estão representadas pelas cores rosa, azul, vermelha, verde e amarela na legenda. Os valores do bootstrap estão no eixo Y enquanto que os conjuntos (classe de dados + esquemas de partição) utilizados nas análises estão no eixo X. No eixo X, PU, PC, PG e PGC se referem à partição única, particionamento por posição do códon, particionamento por gene e particionamento por gene e posição do códon, respectivamente..... 62

Figura 13: Gráfico para os valores de bootstrap do clado C encontrados nas análises filogenéticas baseadas em uma determinada classe de dados que estão representadas pelas cores rosa, azul, vermelha, verde e amarela na legenda. Os valores do bootstrap estão no eixo Y enquanto que os conjuntos (classe de dados + esquemas de partição) utilizados nas análises estão no eixo X. No eixo X, PU, PC, PG e PGC se referem à partição única, particionamento por posição do códon, particionamento por gene e particionamento por gene e posição do códon, respectivamente..... 62

Figura 14: Gráfico para os valores de bootstrap do clado D encontrados nas análises filogenéticas baseadas em uma determinada classe de dados que estão representadas pelas cores rosa, azul, vermelha, verde e amarela na legenda. Os valores do bootstrap estão no eixo Y enquanto que os conjuntos (classe de dados + esquemas de partição) utilizados nas análises estão no eixo X. No eixo X, PU, PC, PG e PGC se referem à partição única, particionamento por posição do códon, particionamento por gene e particionamento por gene e posição do códon, respectivamente..... 63

Figura 15: Gráfico para os valores de probabilidade posterior do clado B encontrados nas análises filogenéticas baseadas em uma determinada classe de dados que estão representadas pelas cores rosa, azul, vermelha, verde e amarela na legenda. Os valores do bootstrap estão no eixo Y enquanto que os conjuntos (classe de dados + esquemas de partição) utilizados nas análises estão no eixo X. No eixo X, PU, PC, PG e PGC se referem à partição única, particionamento por posição do códon, particionamento por gene e particionamento por gene e posição do códon, respectivamente. 64

Figura 16: Gráfico para os valores de probabilidade posterior do clado C encontrados nas análises filogenéticas baseadas em uma determinada classe de dados que estão

representadas pelas cores rosa, azul, vermelha, verde e amarela na legenda. Os valores do bootstrap estão no eixo Y enquanto que os conjuntos (classe de dados + esquemas de partição) utilizados nas análises estão no eixo X. No eixo X, PU, PC, PG e PGC se referem à partição única, particionamento por posição do códon, particionamento por gene e particionamento por gene e posição do códon, respectivamente. 64

Figura 17: Gráfico para os valores de probabilidade posterior do clado D encontrados nas análises filogenéticas baseadas em uma determinada classe de dados que estão representadas pelas cores rosa, azul, vermelha, verde e amarela na legenda. Os valores do bootstrap estão no eixo Y enquanto que os conjuntos (classe de dados + esquemas de partição) utilizados nas análises estão no eixo X. No eixo X, PU, PC, PG e PGC se referem à partição única, particionamento por posição do códon, particionamento por gene e particionamento por gene e posição do códon, respectivamente. 65

Lista de tabelas

Tabela 1: Informações sobre os mitogenomas dos espécimes que foram utilizados no presente trabalho..... 66

Tabela 2: Características de cada conjunto em que as análises filogenéticas foram baseadas com os seus respectivos modelos de substituição. Número* se refere à quantidade de vezes em que um modelo de substituição foi sugerido pelo PartitionFinder2 para ser utilizado nas partições de um determinado esquema de partição. 71

Tabela 3: Valores de verossimilhança, AIC e delta para cada uma das 105 topologias bifurcadas que representam o relacionamento entre as subfamílias de Cricetidae..... 72

Tabela 4: resultados do processo de montagem dos mitogenomas. 78

Tabela 5: Informações sobre o genoma mitocondrial das espécies cujo os genomas foram sequenciados no presente trabalho. Os números representam o comprimento em pares de bases (bp) dos genes mitocondriais de cada espécie. Entre parênteses estão os códons de início/códon de terminação dos CDS. Taa representa códon de terminação incompleto em que a cauda poli A é adicionada posteriormente durante o processamento do mRNA..... 79

Tabela 6: Suporte do bootstrap e probabilidade posterior para os principais clados da família Cricetidae. 81

Lista de apêndices

Apêndice 1 – Script na linguagem bash para realizar análises de verossimilhança no RAxML com uma das 105 topologias subfamiliares fixadas a cada iteração do loop. . 83	83
Apêndice 2 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 1. 84	84
Apêndice 3 – Árvore bayesiana inferida baseando-se no conjunto 1. 85	85
Apêndice 4 – Árvore de máxima verossimilhança inferida baseando-se no conjunto 2. 86	86
Apêndice 5 - Árvore bayesiana inferida baseando-se no conjunto 2..... 87	87
Apêndice 6 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 3. 88	88
Apêndice 7 - Árvore bayesiana inferida baseando-se no conjunto 3..... 89	89
Apêndice 8 – Árvore de máxima verossimilhança inferida baseando-se no conjunto 4. 90	90
Apêndice 9 - Árvore bayesiana inferida baseando-se no conjunto 4..... 91	91
Apêndice 10 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 5. 92	92
Apêndice 11 - Árvore bayesiana inferida baseando-se no conjunto 5..... 93	93
Apêndice 12 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 6. 94	94
Apêndice 13 - Árvore bayesiana inferida baseando-se no conjunto 6..... 95	95
Apêndice 14 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 7. 96	96
Apêndice 15 - Árvore bayesiana inferida baseando-se no conjunto 7..... 97	97
Apêndice 16 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 8. 98	98
Apêndice 17 - Árvore bayesiana inferida baseando-se no conjunto 8..... 99	99
Apêndice 18 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 9. 100	100
Apêndice 19 - Árvore bayesiana inferida baseando-se no conjunto 9..... 101	101

Apêndice 20 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 10.	102
Apêndice 21 - Árvore bayesiana inferida baseando-se no conjunto 10.....	103
Apêndice 22 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 12.	104
Apêndice 23 - Árvore bayesiana inferida baseando-se no conjunto 12.....	105
Apêndice 24 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 13.	106
Apêndice 25 - Árvore bayesiana inferida baseando-se no conjunto 13.....	107
Apêndice 26 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 14.	108
Apêndice 27 - Árvore bayesiana inferida baseando-se no conjunto 14.....	109
Apêndice 28 - Árvore de máxima verossimilhança baseada em CYTB.....	110
Apêndice 29 - Árvore de máxima verossimilhança baseada em ND1.....	111
Apêndice 30 - Árvore de máxima verossimilhança baseada em ND2.....	112
Apêndice 31 - Árvore de máxima verossimilhança baseada em COX1.....	113
Apêndice 32 - Árvore de máxima verossimilhança baseada em COX2.....	114
Apêndice 33 - Árvore de máxima verossimilhança baseada em ATP8.....	115
Apêndice 34 - Árvore de máxima verossimilhança baseada em ATP6.....	116
Apêndice 35 - Árvore de máxima verossimilhança baseada em COX3.....	117
Apêndice 36 - Árvore de máxima verossimilhança baseada em ND3.....	118
Apêndice 37 - Árvore de máxima verossimilhança baseada em ND4L.....	119
Apêndice 38 - Árvore de máxima verossimilhança baseada em ND4.....	120
Apêndice 39 - Árvore de máxima verossimilhança baseada em ND5.....	121
Apêndice 40 - Árvore de máxima verossimilhança baseada em ND6.....	122

SUMÁRIO

Agradecimentos	iii
Resumo	vi
Abstract.....	vii
Lista de figuras	ix
Lista de tabelas	xi
Lista de apêndices.....	xii
1. Introdução.....	1
1.1 Família Cricetidae: composição e relações filogenéticas	1
1.2 Genoma mitocondrial	5
1.3 Análise filogenética de verossimilhança	8
1.4 Análise filogenética bayesiana	11
1.5 Hipótese	12
2. Objetivo geral	13
2.1 Objetivos específicos.....	13
3. Materiais e métodos.....	13
3.1 Amostragem taxonômica.....	14
3.2 Sequenciamento, montagem e anotação dos mitogenomas.....	14
3.3 Análises filogenéticas, alinhamentos e esquemas de partição.....	16
3.5 Análises filogenéticas com topologias fixadas.....	22
4. Resultados.....	25
4.1 Montagem dos mitogenomas e anotações	25
4.2 Análises filogenéticas baseadas em genes concatenados e aminoácidos.	27
4.2.1 Análises com partição única.	28
4.2.2 Análises particionados por posição do códon.....	30
4.2.3 Análises particionadas por gene.	31
4.2.4 Análises particionadas por gene e posição do códon.	31

4.3 Análises baseadas em genes individuais.	32
4.4 Análise filogenética com topologias fixadas	36
5. Discussão	37
5.1 Análises filogenéticas	37
5.2 Impacto das partições e classes de dados nas árvores filogenéticas.....	40
6. Conclusões.....	44
Referências	45
FIGURAS	52
TABELAS	66
APÊNDICES	83

1. Introdução

1.1 Família Cricetidae: composição e relações filogenéticas

A família Cricetidae pertence à ordem Rodentia que é o grupo mais diverso dentro da classe Mammalia, com aproximadamente 2.360 espécies conhecidas, representando 42% dos mamíferos (Musser and Carleton, 2005). Cricetidae está incluída na subordem Myomorpha que é subdividida nas superfamílias Muroidea e Dipodoidea, sendo que Cricetidae pertence à superfamília Muroidea (Musser and Carleton, 2005). A superfamília Muroidea é atualmente subdividida nas famílias Platacanthomyidae, Spalacidae, Calomyscidae, Nesomyidae, Cricetidae e Muridae, sendo as subfamílias Arvicolinae, Cricetinae, Neotominae, Sigmodontinae e Tylomyinae alocadas dentro de Cricetidae (Jansa and Weksler, 2004; Musser and Carleton, 2005). As espécies da subfamília Cricetidae estão distribuídas pelas regiões Neotropical (Sigmodontinae e Tylomyinae), Paleoártica (Cricetinae), Holoártica (Arvicolinae) e Neoártica (Neotominae) (Musser and Carleton, 2005; Nowak, 1999). Sigmodontinae é a subfamília que apresenta a maior diversidade entre os roedores da região Neotropical (Musser and Carleton, 2005; Patton et al., 2015).

Contudo, a família Cricetidae nem sempre foi reconhecida como um grupo válido entre os roedores murídeos. Um dos primeiros estudos a tentar classificar roedores murídeos a partir de caracteres cranianos e dentários reconhecia apenas as famílias Lophiomyidae, Spalacidae e Muridae, sendo esta última composta pelas subfamílias Sminthinae, Hydromyinae, Platacanthomyinae, Gerbilinae, Phloeomyinae, Dendromyinae, Cricetinae, Murinae, Arvicolinae e Siphneinae (Alston, 1876). Em um estudo que reavaliou a classificação taxonômica proposta por Alston (1876) baseando-se em caracteres morfológicos

de exemplares ainda não disponíveis na época de Alston (1876), Thomas (1896) separou os roedores murídeos nas famílias Muridae, Spalacidae e realocou a família Lophiomyidae a nível de subfamília (Lophiomyinae) dentro Muridae. Além disso, Thomas (1896) reconheceu a família Muridae sendo composta pelas subfamílias Hydromyinae, Rhynchomyinae, Phloeomyinae, Gerbillinae, Otomyinae, Dendromyinae, Murinae, Lophiomyinae, Sigmodontinae, Neotominae, Microtinae e Siphneinae.

. Em 1899, Tullberg (1899) separou os roedores murídeos nas famílias Spalacidae, Nesomyidae, Cricetidae, Lophiomyidae, Arvicolidae, Hesperomyidae (Sigmodontinae), Muridae, Gerbilidae. Depois de alguns anos, Miller and Gidley (1918) propuseram a superfamília Muroidea sendo composta pelas famílias Muscardinidae, Cricetidae, Platacanthomyidae, Rhizomyidae, Spalacidae e Muridae. Segundo Miller and Gidley (1918) a família Cricetidae era composta pelas subfamílias Cricetinae (incluindo sigmodontíneos, cricetíneos e nesomyíneos), Gerbillinae, Microtinae e Lophiomyinae.

Ellerman (1941) reconheceu a superfamília Muroidea subdividindo-a nas famílias Muscardinidae, Lophiomyidae, Spalacidae, Rhizomyidae e alocou os roedores cricetídeos na família Muridae. Em contrapartida, (Simpson, 1945) subdividiu a superfamília Muroidea nas famílias Spalacidae, Rhizomyidae, Muridae e Cricetidae, sendo esta última composta pelas subfamílias Cricetinae, Nesomyinae, Lophiomyinae, Microtinae e Gerbillinae.

Chaline et al. (1977) subdividiu a superfamília Muroidea nas famílias Nesomyidae, Rhizomyidae, Gerbilidae, Arvicolidae, Dendromuridae, Cricetomyidae, Muridae e Cricetidae, sendo esta última composta pelas subfamílias Cricetodontinae, Hesperomyinae, Cricetinae, Spalacinae, Myospalacinae, Lophiomyinae e Platacanthomyinae.

Reig (1980) reconhecia a superfamília Muroidea como sendo composta pelas famílias Muridae, Arvicolidae, Gerbilidae, Spalacidae, Myospalacidae, Lophiomyidae e alocou dentro de Cricetidae as subfamílias Cricetodontinae, Melissiodontinae, Platacanthomyinae,

Cricetinae, Neotominae, Sigmodontinae, Nesomyinae, Dendromurinae, Petromiscinae e Cricetomyinae.

Atualmente, a família Cricetidae é reconhecida como um grupo monofilético segundo alguns estudos que realizaram análises filogenéticas com dados moleculares (Jansa et al., 2009; Jansa and Weksler, 2004; Leite et al., 2014; Stepan et al., 2004; Stepan and Schenk, 2017). Contudo, as relações filogenéticas entre as subfamílias de Cricetidae ainda não estão bem estabelecidas quando se compara os resultados desses estudos. Nesses trabalhos, a maioria dos principais clados dentro de Cricetidae apresenta baixo a médio suporte do bootstrap e a maior parte deles encontra árvores com topologias conflitantes para o relacionamento entre as subfamílias de Cricetidae.

No trabalho de Jansa and Weksler (2004) em que os autores inferiram as relações filogenéticas entre as subfamílias da superfamília Muroidea baseando-se no primeiro éxon do gene IRBP (~ 1.200 pares de bases (pb)), a árvore de máxima verossimilhança recuperou Cricetinae como grupo irmão do clado formado por Tylomyinae, Neotominae, Arvicolinae e Sigmodontinae, o qual apresentou 70% de suporte do bootstrap. Nesse mesmo trabalho, Arvicolinae foi recuperado como grupo irmão de Sigmodontinae, formando um clado com 72% de suporte do bootstrap, e as relações entre Tylomyinae, Neotominae e o clado formado por Arvicolinae e Sigmodontinae não foram bem resolvidas.

No trabalho de Stepan et al. (2004), em que os autores inferiram as relações filogenéticas entre membros da família Muroidea baseando-se nos genes nucleares GHR, BRCA1, RAG1 e c-myc concatenados 6.400 pb, a subfamília Arvicolinae foi recuperada como grupo irmão de Cricetinae, formando um clado que teve 89% de suporte do bootstrap. Esse clado constituído por Arvicolinae e Cricetinae foi recuperado como grupo irmão do clado formado por Neotominae, Tylomyinae e Sigmodontinae o qual apresentou suporte do

bootstrap igual a 51%. Além disso, Neotominae foi grupo irmão do clado constituído por Tylomyinae e Sigmodontinae que teve 56% de suporte do bootstrap.

No trabalho de Jansa et al. (2009) os autores inferiram as relações filogenéticas entre roedores murídeos baseando-se nos genes nucleares IRBP e GHR que apresentavam 1.257 bp e 954 bp, respectivamente. Nesse trabalho, a árvore de verossimilhança baseada em IRBP recuperou Cricetinae como grupo irmão do clado formado por Tylomyinae, Neotominae, Arvicolinae e Sigmodontinae o qual apresentou suporte do bootstrap superior a 75%. Ademais, a subfamília Tylomyinae saiu como grupo irmão do clado formado por Neotominae, Arvicolinae e Sigmodontinae o qual apresentou bootstrap inferior a 50%, sendo que Sigmodontinae foi grupo irmão do clado constituído por Arvicolinae e Neotominae o qual apresentou suporte do bootstrap inferior a 50%. Em contrapartida, a árvore de verossimilhança baseada no gene GHR recuperou Neotominae como grupo irmão da subfamília Arvicolinae formando um clado com suporte do bootstrap inferior a 50% que saiu como grupo irmão do clado constituído por Tylomyinae, Cricetinae e Sigmodontinae que também teve suporte do bootstrap inferior a 50%. Nessa árvore Tylomyinae foi recuperado como grupo irmão do clado formado por Cricetinae e Sigmodontinae que teve suporte do bootstrap inferior a 50%. Finalmente, a árvore de verossimilhança baseada em IRBP e GHR concatenados resultou em Cricetinae como grupo irmão do clado formado por Tylomyinae, Arvicolinae, Neotominae e Sigmodontinae que teve suporte do bootstrap inferior a 50%. Nessa mesma árvore, a subfamília Tylomyinae saiu como grupo irmão do clado formado por Arvicolinae, Neotominae e Sigmodontinae o qual apresentou suporte do bootstrap inferior a 50%, e a subfamília Arvicolinae foi recuperada como grupo irmão do clado formado por Neotominae e Sigmodontinae que teve suporte do bootstrap inferior a 50%.

No trabalho realizado por Leite et al. (2014) os autores estimaram o tempo de divergência entre linhagens da subfamília Sigmodontinae e também inferiram o posicionamento

filogenético dessa subfamília em relação às outras quatro subfamílias de Cricetidae usando os genes CYTB e IRBP que concatenados apresentaram no total 2,376 pb. Nesse estudo, a subfamília Neotominae foi recuperada como grupo irmão do clado formado por Arvicolinae, Cricetinae, Tylomyinae e Sigmodontinae o qual apresentou 71% de suporte do bootstrap, e Arvicolinae saiu como grupo irmão do clado formado por Cricetinae, Tylomyinae e Sigmodontinae o qual apresentou 45% de suporte do bootstrap. Nessa mesma árvore, a subfamília Tylomyinae foi recuperada como grupo irmão de Sigmodontinae formando um clado com 74% de suporte do bootstrap.

No trabalho realizado por Stepan and Schenk (2017) os autores inferiram relações filogenéticas entre roedores murídeos baseando-se nos genes nucleares BRCA1, GHR, Rbp3, RAG1 e Acp5 e no gene mitocondrial CYTB que concatenados somavam em média 3,070 pb por espécie. Na árvore de verossimilhança gerada nesse trabalho, as subfamílias Neotominae, Sigmodontinae e Tylomyinae formaram um clado com 80% de suporte do bootstrap e Sigmodontinae saiu como grupo irmão de Tylomyinae formando um subclado com 85% de suporte do bootstrap. Nessa mesma árvore, o clado formado por Neotominae, Sigmodontinae, Tylomyinae saiu como grupo irmão do clado formado por Arvicolinae e Cricetinae o qual apresentou 96% de suporte do bootstrap.

1.2 Genoma mitocondrial

A mitocôndria, com o auxílio do genoma mitocondrial, promove a apoptose, envelhecimento celular, e seu mau funcionamento pode causar doenças (Graeber and Müller, 1998; Kroemer et al., 1998; Wei, 1998). Ademais, o genoma mitocondrial codifica várias proteínas que são necessárias para a síntese de adenosina trifosfato (ATP) que é a principal fonte de energia da célula (Ballard and Whitlock, 2004). O genoma mitocondrial de animais é

composto por duas fitas de DNA circulares, geralmente apresenta 13 genes codificadores de proteínas (CDS), dois genes codificadores de rRNA (rRNA), 22 genes codificadores de tRNA (tRNA) e uma região não codificadora conhecida como D-loop que está associada ao processo de replicação do mitogenoma (Bibb et al., 1981; Boore, 1999). Todos esses genes mais a região D-loop somam no total aproximadamente 16,000 pb (Boore, 1999).

O genoma mitocondrial de animais é haploide, não sofre recombinação, é herdado maternamente, e apresenta taxa de mutação mais alta que o genoma nuclear porque o seu sistema de reparo de mutação é menos sofisticado em comparação aquele utilizado pelo genoma nuclear (Ballard and Whitlock, 2004; Brown and Kraus, 1979). Devido ao fato do genoma mitocondrial não sofrer recombinação, todos os genes mitocondriais estão em desequilíbrio de ligação, e, portanto, forças evolutivas que afetam diretamente uma região do mitogenoma poderão influenciar indiretamente todas as outras regiões do mitogenoma (Ballard and Whitlock, 2004; Hartl and Clark, 2007). A ausência de recombinação possibilita que todos os genes mitocondriais tenham a mesma história evolutiva em termos de ancestralidade (Cann et al., 1987; Hartl and Clark, 2007; Zink and Barrowclough, 2008). Portanto, pode-se esperar que todos os genes mitocondriais recuperarão as mesmas relações evolutivas entre os táxons em uma árvore filogenética (Hartl and Clark, 2007; McVay and Carstens, 2013). Contudo, se existirem incongruências entre as relações evolutivas nas árvores filogenéticas reconstruídas a partir de diferentes genes mitocondriais, essas incongruências podem ser resultado de enviesamento dos métodos de reconstrução filogenética (McVay and Carstens, 2013).

Os genes mitocondriais apresentam restrições seletivas tendo em vista que eles desempenham importantes funções para o bom funcionamento da célula. Portanto, muitos sítios dos genes mitocondriais não estão sob evolução neutra (Matioli and Fernandes, 2012; Ridley, 2013). Entretanto, a terceira posição do códon tende a ser degenerada devido ao

pareamento oscilante do anticódon de tRNA nessa posição, e por isso muitos sítios da terceira posição do códon estão sujeitos a uma alta taxa de substituição sinônima que não terá um grande impacto na função do gene (Pesole et al., 1999). Essa alta taxa de substituição na terceira posição pode saturar o sinal filogenético nesse sítio, o que aumenta a probabilidade de ocorrência de homoplasias entre as sequências, podendo ocasionar eventos de atração dos ramos longos (Bergsten, 2005).

Os primeiros genomas não virais a serem sequenciados foram os genomas mitocondriais do rato e da espécie humana (Anderson et al., 1981; Bibb et al., 1981). Alguns anos depois, os genomas mitocondriais completos de outros animais, algas e fungos também foram sequenciados (Anderson et al., 1982; Boore, 1999; Cummings et al., 1990; Gray, 1982; Vahrenholz et al., 1993). O sequenciamento de genomas mitocondriais completos alavancou linhas de pesquisas voltadas a genômica comparativa, estratégias de replicação, padrões de herança do mitogenoma, processos de transcrição e tradução (Bridge et al., 1992; Clayton, 1982; Gray and Boer, 1988; Nosek and Tomaska, 2003). Durante a décadas de 80, 90 até o início da virada do século os genomas mitocondriais eram sequenciados baseando-se em variações do método Sanger de sequenciamento que é capaz de sequenciar fragmentos de DNA com 700 pb ou mais (Metzker, 2010; Schuster, 2008; Smith, 2016). Contudo, o sequenciamento de genomas mitocondriais aumentou drasticamente com o advento do sequenciamento de nova geração (Smith, 2016). Métodos de sequenciamento de nova geração como pirosequenciamento 454, Illumina e Ion Torrent são capazes de sequenciar muito mais fragmentos de DNA em comparação aos métodos Sanger de sequenciamento (Shendure et al., 2017). Enquanto os métodos de Sanger sequenciam um único fragmento de DNA com aproximadamente 700 pb ou mais de cada vez, os métodos de sequenciamento de nova geração sequenciam milhares de fragmentos de DNA menores que 500 pb em paralelo, permitindo que os genomas sejam montados com uma maior profundidade com vários

fragmentos validando um determinado sítio (Schuster, 2008; Shendure and Ji, 2008; Sims et al., 2014). Além disso, uma das principais características dos métodos de sequenciamento de nova geração é que eles são capazes de sequenciar genomas inteiros de maneira mais veloz e a um custo menor em comparação ao do sequenciamento baseado no método Sanger (Shendure et al., 2017). Devido a esta facilidade, ocorreu uma explosão de trabalhos que visavam sequenciar genomas mitocondriais e descrever a sua estrutura e características (Smith, 2016). Esses trabalhos geralmente explicitam os métodos utilizados para sequenciar, montar e anotar os mitogenomas; também descrevem a estrutura e características dos mitogenomas, como por exemplo, número de pares de bases, ordem dos genes, quantidade de cada tipo de gene (genes codificadores de proteína, genes codificadores de tRNA, genes codificadores de rRNA), descrevem as estruturas das moléculas de tRNA, explicitam quais são os códons de início e de terminação, e calculam a porcentagem de conteúdo AT ou GC (Chen and Zhao, 2009; Chen et al., 2013; Dietrich and Brune, 2016; Gan et al., 2016; He et al., 2013). Atualmente, existe um esforço maior para realizar trabalhos que além de descrever métodos de sequenciamento, montagem, anotação, estrutura e características dos mitogenomas, também visam estudar evolução molecular, inferir filogenias e estudar o metabolismo da mitocôndria (da Cruz et al., 2019; Silva et al., 2018; Tavares and Seuánez, 2018; Vasta et al., 2009; Vieira and Prosdocimi, 2019).

1.3 Análise filogenética de verossimilhança

Dentro do contexto de análises filogenéticas, a verossimilhança calcula a probabilidade do alinhamento de sequências de DNA dado um modelo de substituição, uma topologia de árvore e o tamanho de ramos da árvore (Matioli and Fernandes, 2012). Em outras palavras, a verossimilhança realiza uma busca pela topologia, tamanhos de ramos e valores dos

parâmetros do modelo de substituição que maximizam a probabilidade de que o alinhamento de DNA tenha evoluído seguindo um padrão determinado pelo modelo evolutivo, isto é, a árvore que melhor se ajusta a evolução das sequências. (Felsenstein, 2004; Yang et al., 1995).

Na análise filogenética baseada em máxima verossimilhança, uma árvore inicial é utilizada para calcular a verossimilhança dessa árvore em relação a cada um dos sítios, do alinhamento. A verossimilhança final da árvore é o resultado do somatório (Σ) do log do valor de verossimilhança da árvore em relação a cada um dos sítios que pode ser representado pela seguinte fórmula:

$$l = \log\{L(\theta)\} = \sum_{h=1}^n \log\{f(x_h | \theta)\}$$

Aonde n é a quantidade de sítios no alinhamento, h representa um determinado sítio do alinhamento, Θ representa os valores dos parâmetros e l representa o valor de verossimilhança final (Yang, 2014). O método de máxima verossimilhança calcula os valores de Θ que maximizam o valor de verossimilhança da árvore (Felsenstein, 2004). Durante a busca da árvore de máxima verossimilhança, a topologia e o comprimento dos ramos da árvore inicial são alterados, e um novo cálculo de verossimilhança é realizado para essa árvore modificada. Se o valor de verossimilhança da árvore modificada for maior que a valor de verossimilhança da árvore inicial, a árvore modificada é aceita como o novo estado a partir do qual a análise continuará (Matioli and Fernandes, 2012). Esse processo se repetirá até o algoritmo de busca por árvores não ser mais capaz de encontrar árvores que apresentam valor de verossimilhança superior em comparação à árvore anterior (Felsenstein, 2004; Matioli and Fernandes, 2012).

Em análises baseadas em sequências de nucleotídeos, a verossimilhança de uma árvore em relação a cada um dos sítios é calculada a partir da probabilidade de que um determinado sítio que apresenta uma base nucleotídica (A, C, G ou T) tenha sido A, C, G ou T no mesmo sítio da sequência de DNA ancestral (Felsenstein, 2004). A probabilidade de que um sítio tenha apresentado outro nucleotídeo ou o mesmo nucleotídeo na sequência ancestral é definida por um determinado modelo de substituição (Yang, 2014). Por exemplo, o modelo de substituição mais simples conhecido como Jukes & Cantor 1969 (Jukes and Cantor, 1969) considera que os seis tipos de substituição ($A \leftrightarrow C$; $A \leftrightarrow G$; $A \leftrightarrow T$; $C \leftrightarrow T$; $C \leftrightarrow G$; $G \leftrightarrow T$) apresentam a mesma probabilidade de ocorrência e os quatro nucleotídeos têm a mesma frequência. Em contrapartida, o modelo de substituição conhecido como *General Time-Reversible (GTR) model* (Tavaré, 1986), que também é o modelo mais complexo entre os modelos da família tempo reversível, considera que cada um dos seis tipos de substituição apresenta probabilidade de ocorrência diferente e os quatro nucleotídeos têm frequências diferentes. Os modelos de substituição mencionados anteriormente pertencem à família de modelos tempo reversível em que uma mudança de um nucleotídeo i para um nucleotídeo j apresenta a mesma probabilidade de mudança de um nucleotídeo j para um nucleotídeo i (Yang, 2014).

Análises filogenéticas de verossimilhança geralmente aplicam uma técnica conhecida como *bootstrap* para estimar o grau de suporte dos clados da árvore (Felsenstein, 2004). O *bootstrap* é uma técnica de reamostragem com reposição que tem como objetivo principal inferir a variabilidade de uma distribuição desconhecida de onde os dados foram amostrados. Em outras palavras, o *bootstrap* infere a variabilidade de uma distribuição da qual desconhecemos qual seria a variabilidade real (Felsenstein, 1985). Dentro do contexto de análises filogenéticas, o *bootstrap* reamostra com reposição os sítios do alinhamento das sequências utilizadas para inferir as árvores filogenéticas e reordena esses sítios em cada

evento de reamostragem, sendo que os alinhamentos com os sítios reamostrados apresentam o mesmo comprimento da sequência original. Esses alinhamentos com sítios reamostrados são utilizados para inferir árvores filogenéticas durante o processo do bootstrap (Felsenstein, 1985). Os clados presentes na árvore inferida com o alinhamento original e que se repetem com maior frequência no conjunto de árvores inferidas a partir dos alinhamentos com sítios reamostrados recebem os maiores suportes do bootstrap (Felsenstein, 2004).

1.4 Análise filogenética bayesiana

Uma das principais características da análise bayesiana é a probabilidade *a priori* que é definida antes de uma determinada análise ser realizada. A probabilidade *a priori* pode ser definida baseando-se em um conhecimento prévio sobre um determinado parâmetro, nesse caso, o prior é objetivo; caso não exista nenhuma informação prévia para definir o prior, também existe a opção de se estabelecer um prior subjetivo baseando-se naquilo que o pesquisador considera ser uma distribuição adequada para esse prior (Yang, 2014).

Análises filogenéticas baseadas na estatística bayesiana calculam a probabilidade posterior da árvore e dos parâmetros dado um alinhamento de nucleotídeos ou aminoácidos (Yang, 2014). A fórmula para calcular a probabilidade posterior de uma árvore filogenética é

$$P(T, \theta | D) = \frac{P(T, \theta)P(D|T, \theta)}{P(D)}$$

Nessa fórmula, $P(T, \theta | D)$ representa a probabilidade posterior da árvore T e parâmetros θ dado um alinhamento D; $P(T, \theta)$ é a probabilidade *a priori* para árvore T e parâmetros θ estabelecida antes de qualquer análise; $P(D|T, \theta)$ é a função de verossimilhança que busca a

árvore T e valores dos parâmetros Θ que maximizam a probabilidade do alinhamento D ; $P(D)$ representa a probabilidade do alinhamento D (Yang and Rannala, 2012).

A probabilidade $P(D)$ envolve integrar todos os valores possíveis dos parâmetros e todas as árvores possíveis, o que seria computacionalmente inviável (Yang and Rannala, 2012). Análises filogenéticas bayesianas são computacionalmente viáveis com o auxílio de um algoritmo conhecido como *Monte Carlo Markov Chain* (MCMC) que amostra da distribuição de probabilidade (Yang, 2014). O algoritmo MCMC considera que a probabilidade de um estado futuro depende apenas da probabilidade do estado presente e não depende da probabilidade de estados passados (Felsenstein, 2004). Quando aplicado a análises filogenéticas, a cadeia de Markov (MCMC) começa com uma árvore inicial que tem probabilidade posterior p . Logo em seguida, alguns dos parâmetros associados a essa árvore são alterados (e.g. tamanho dos ramos da árvore) gerando outra árvore com probabilidade posterior p' . Se $p' > p$, a árvore com probabilidade p' é aceita como o novo estado atual sobre o qual o algoritmo se baseará para continuar a análise. Se $p' < p$, calcula-se a razão $\frac{p'}{p}$, e árvore com probabilidade p' será aceita como novo estado apenas se $\frac{p'}{p}$ for maior que um número aleatório entre 0 e 1 que é sorteado pelo algoritmo (Matioli and Fernandes, 2012).

1.5 Hipótese

Steppan and Schenk (2017) inferiram a árvore filogenética com a mais extensa amostragem - 904 espécies de roedores murídeos - que também resultou nos maiores suportes de bootstrap para os principais clados dentro Cricetidae até o momento. Portanto, a topologia subfamiliar de Cricetidae inferida por Steppan and Schenk (2017) será considerada a hipótese inicial para verificar se os resultados do presente trabalho corroboram ou não essa topologia. Para isso, o genoma mitocondrial completo – com aproximadamente 16000 pb - de espécies representantes de cada uma das subfamílias de Cricetidae serão utilizados em análises

filogenéticas de máxima verossimilhança e inferência bayesiana utilizando diferentes esquemas de partição para contornar possíveis problemas de saturação do sinal filogenético e para verificar o impacto dessas diferentes partições na topologia e suporte das árvores. Além disso, também serão realizadas análises filogenéticas de verossimilhança restringidas a cada uma das 105 topologias subfamiliares bifurcadas e enraizadas que representam as relações entre as subfamílias de Cricetidae com o objetivo de verificar qual topologia se ajusta consideravelmente melhor ao alinhamento dos genomas mitocôndrias incluídos no presente trabalho.

2. Objetivo geral

- Inferir as relações filogenéticas entre as subfamílias de Cricetidae a partir de dados mitogenômicos

2.1 Objetivos específicos

- Verificar o efeito de diferentes conjuntos de classes dados (CDS, rRNA-CDS, rRNA-CDS-tRNA, rRNA-CDS-tRNA-D-loop, aminoácidos) e esquemas de partição nas inferências filogenéticas
- Verificar se existe alguma topologia que representa as relações entre as subfamílias de Cricetidae que se ajusta consideravelmente melhor aos dados em comparação a outras topologias.

3. Materiais e métodos

3.1 Amostragem taxonômica

Foram analisados os genomas mitocondriais de 51 espécies de Cricetidae, sendo oito novos mitogenomas e 43 obtidos no GenBank (Tabela 1). A amostragem buscou abranger o máximo possível da diversidade da família Cricetidae, incluindo espécies representantes de todas as cinco subfamílias e todos os gêneros disponíveis no Genbank. Decidiu-se priorizar o sequenciamento de mitogenomas das espécies das subfamílias Sigmodontinae e Tylomyinae, que estavam subrepresentadas no GenBank. Desta forma, foram sequenciados cinco novos mitogenomas de espécies da subfamília Sigmodontinae (*Delomys dorsalis*, *Irenomys tarsalis*, *Oligoryzomys stramineus*, *Rhipidomys mastacalis* e *Sigmodon alstoni*), 2 novos mitogenomas de espécies da subfamília Tylomyinae (*Tylomys nudicaudus* e *Nyctomys sumichrasti*) e 1 novo mitogenoma da espécie *Scotinomys teguina* pertencente a subfamília Neotominae (Figuras 1 a 8). O mitogenoma de *Oligoryzomys stramineus* (Figura 1) já foi reportado (da Cruz et al., 2019). O grupo externo foi composto por sete espécies das superfamílias Muroidea e Dipodoidea consideradas gradualmente distantes filogeneticamente do grupo interno (Cricetidae) segundo filogenias anteriores (Jansa and Weksler, 2004; Steppan and Schenk, 2017). *Mus musculus*, *Psammomys obesus* e *Acomys cahirinus* pertencem à família Muridae que é grupo irmão de Cricetidae, enquanto que *Rhizomys sinensis* e *Spalax judaei* pertencem à família Spalacidae que é grupo irmão do clado composto por Calomyscidae, Nesomyidae, Muridae e Cricetidae (Steppan et al., 2004). Finalmente, *Sicista concolor* pertence à superfamília Dipodoidea que é grupo irmão da superfamília Muroidea (Steppan and Schenk, 2017), e foi utilizada no enraizamento da árvore.

3.2 Sequenciamento, montagem e anotação dos mitogenomas

Sequenciamento: O DNA foi extraído de oito espécimes de roedores da família Cricetidae depositados no Museu Nacional da Universidade Federal do Rio de Janeiro com a técnica de extração salina (Miller et al., 1988) utilizando-se o *Kit Promega Mouse Tail*. O sequenciamento do genoma parcial das oito espécies de Cricetidae foi realizado no sequenciador Illumina HiSeq 2500 do INCA (Instituto Nacional do Câncer, Rio de Janeiro, Brasil) utilizando-se o Kit Nextera, que permitiu multiplexar amostras de DNA de espécies diferentes em um único poço da placa de vidro do sequenciador Illumina. O sequenciamento do genoma parcial gera uma porcentagem de fragmentos de DNA (*DNA reads*) suficiente para ser possível montar o genoma mitocondrial completo (Smith, 2016).

Montagem: Os fragmentos de DNA sequenciados pelo Illumina foram utilizados durante o processo de montagem do genoma mitocondrial completo. Após o sequenciamento, a informação das sequências de fragmentos de DNA foi armazenada em dois arquivos fastq. Um dos arquivos continham reads que foram sequenciadas no sentido *forward* e outro arquivo fastq continha as reads que foram sequenciadas no sentido *reverse*. Esses arquivos fastq serviram como entrada para o programa de montagem conhecido como MIRA (Chevreux et al., 1999). Primeiro, tentava-se realizar a montagem *de novo* usando o MIRA; quando a montagem *de novo* não era capaz de montar completamente o genoma mitocondrial, realizava-se, no próprio MIRA, o mapeamento das *reads* em um genoma mitocondrial de referência de uma espécie evolutivamente próxima aquela cujo o mitogenoma pretendia-se montar. Quando o processo de mapeamento do MIRA também não era capaz de montar completamente o genoma mitocondrial, utilizava-se o genoma mitocondrial montado incompletamente pelo MIRA como referência para o processo de mapeamento no programa MITObim (Hahn et al., 2013) que utiliza o MIRA iterativamente. O MITObim realiza várias rodadas de mapeamento usando como referência a montagem do MIRA. Quando a primeira rodada de mapeamento realizada pelo MITObim não é capaz de montar completamente o

genoma mitocondrial, ele utiliza como referência a última montagem incompleta para mapear as reads na próxima rodada de mapeamento. A montagem com o MIRA e MITOBIM foram executadas na linha de comando do LINUX utilizando a configuração default.

Anotação: O processo de anotação dos mitogenomas foi realizado primeiramente com o auxílio do script *genechecker.py* (Schomaker-Bastos and Prosdocimi, 2018), que realiza uma anotação automática baseada em genomas de referência. Posteriormente, os mitogenomas passaram pela plataforma de anotação automática MITOS WebServer, com o objetivo de aumentar a acurácia da anotação dos tRNA e rRNA. Após o processo de anotação automática, realizou-se a anotação manual com o auxílio dos programas Artemis (Carver et al., 2012) e BLAST (Camacho et al., 2009) para confirmação dos início e fim dos genes codificadores de proteínas.

Como a espécie *Mus musculus* é um organismo modelo que apresenta um genoma bem anotado e disponível no GenBank, o genoma mitocondrial dessa espécie serviu como referência para refazer algumas anotações dos 48 mitogenomas que foram obtidos do GenBank e também serviu para conferir a anotação manual dos 8 mitogenomas que foram sequenciados nesse projeto. Para isso, o programa Geneious v.11.1 (Kearse et al., 2012) foi utilizado para comparar a anotação dos outros 56 mitogenomas incluídos nas análises com o mitogenoma de *Mus musculus*. As anotações dos genes que não começavam e terminavam em códons de início e término, respectivamente, eram alteradas para se adequarem melhor ao padrão de anotação do *Mus musculus*.

3.3 Análises filogenéticas, alinhamentos e esquemas de partição

As sequências foram alinhadas com o algoritmo MUSCLE v.3.7 (Edgar, 2004) na plataforma CIPRES (Miller et al., 2010). As relações filogenéticas foram inferidas com os

métodos de máxima verossimilhança e bayesianos com o modelo de substituição GTR + G + I escolhido pelo jModelTest v.2.1.10 (Posada, 2008) através da plataforma CIPRES para análises baseadas em todos os genes mitocondriais concatenados com partição única. As árvores de máxima verossimilhança foram inferidas no RAxML v.8.2.10 (Stamatakis, 2014) com 1000 pseudo réplicas e o algoritmo rapid bootstrap no CIPRES (Miller et al., 2010) nas análises com partição única. No caso das análises particionadas, o RAxML foi executado com a função -q, que realizou o particionamento baseando-se em um arquivo que continha as partições desejadas. As árvores bayesianas foram inferidas no MrBayes v.3.2.6 (Ronquist et al., 2012) na plataforma CIPRES com duas análises (MCMC) independentes que apresentavam comprimento de 20,000,000, 20,000,000, 25,000,000 e 45,000,000 com amostragem a cada 20,000, 20,000, 25,000 e 45,000 de gerações, nas análises com partição única, particionadas por posição do códon, particionadas por gene e particionadas por gene e posição do códon, respectivamente. Cada uma dessas análises apresentavam três cadeias quentes e uma fria e as 100 primeiras gerações eram descartadas. O tamanho efetivo da amostragem (*effective sampling size* - ESS) foi verificado no Tracer v.1.7 (Rambaut et al., 2018). Todos os parâmetros apresentaram valores de ESS superiores a 200.

Além de inferir as relações filogenéticas entre as subfamílias de Cricetidae, as análises filogenéticas foram realizadas para verificar o impacto das diferentes classes de dados na topologia e nos valores de suporte para os nós das árvores. Para isso, foram realizadas análises baseadas em alinhamentos com sequências de DNA codificadoras de proteína (CDS), análises baseadas em genes codificadores de rRNA concatenados aos CDS (rRNA-CDS), análises baseadas rRNA, CDS e genes codificadores de tRNA concatenados (rRNA-CDS-tRNA) e análises baseadas em todos genes mitocondriais concatenados mais a região D-Loop (rRNA-CDS-tRNA-D-loop). Foi possível criar esses quatro tipos de alinhamento com o auxílio do programa Geneious v.7.1 (Kearse et al., 2012), que permitiu a extração das regiões

codificadores de rRNA, codificadoras de proteínas, codificadoras de tRNA e região D-Loop a partir dos genomas mitocôndrias completos das 57 espécies incluídas nesse estudo. Essa extração permitiu que os dois rRNA, 13 genes codificadores de proteínas e 21 tRNA e região D-Loop fossem separados e alinhados individualmente no próprio Geneious usando o algoritmo MUSCLE. Após o alinhamento individual dos genes, as sequências de cada gene foram exportadas do Geneious em arquivos no formato fasta. Os alinhamentos que apresentavam alguns sítios com *gaps* em muitas sequências, passaram pelo programa trimAl (Capella-Gutierrez et al., 2009) usando o parâmetro *-gt 0.90* para eliminar os sítios em que mais de 10% das sequências apresentavam gaps. As sequências dos genes em arquivos fasta foram concatenados com o comando *seqkit concat* (Shen et al., 2016) através da linha de comando do LINUX. Esse processo de concatenação gerou quatro arquivos, sendo que cada arquivo armazenou as sequências de DNA de apenas um dos quatro tipos de classe de dados (rRNA, CDS, tRNA, D-loop). Posteriormente, as sequências de DNA armazenadas no arquivo de rRNA foram concatenadas às sequências do arquivo que armazenava as sequências CDS, gerando um arquivo contendo às sequências de rRNA e CDS concatenados (rRNA-CDS). Ademais, as sequências do arquivo contendo as sequências de rRNA e CDS foram concatenadas as sequências do arquivo contendo tRNA, gerando um arquivo em que as sequências de rRNA, CDS e tRNA estavam concatenadas (rRNA-CDS-tRNA). Finalmente, o alinhamento de rRNA-CDS-tRNA foi concatenado à região D-loop (rRNA-CDS-tRNA-D-Loop). Os resultados das análises baseadas em cada um desses quatro tipos alinhamentos foram comparados para verificar o impacto dessas classes de dados na topologia e suporte da árvore.

O alinhamento do genoma mitocondrial foi realizado na plataforma CIPRES utilizando o algoritmo MUSCLE com os parâmetros default. Esses quatro tipos de alinhamento também foram utilizadas em análises filogenéticas que além de inferirem as

relações evolutivas entre as subfamílias de Cricetidae, também serviram para verificar o impacto dos esquemas de partição na topologia e no suporte das árvores. As análises particionadas foram baseadas nos seguintes três tipos de partições: particionamento por posição do códon, particionamento por gene e particionamento por gene e posição do códon. Essas partições foram estabelecidas no arquivo de configuração do Partitionfinder2 antes de executar o programa. Além disso, no arquivo de configuração do Partitionfinder2 foi estabelecido que (1) o tamanho dos ramos da árvore de verossimilhança que serviu como base para a escolha do modelo de substituição seria o mesmo para cada partição (opção *branchlengths = linked*); (2) os modelos de substituição candidatos seriam aqueles disponíveis no MrBayes (opção *models = mrbayes*); (3) que o critério para a escolha do modelo de substituição mais adequado seria o BIC porque ele penaliza mais fortemente a adição de parâmetros em comparação ao AIC (opção *model_selection = BIC*); (4) que o algoritmo de busca do melhor esquema de partição seria o *greedy* (opção *search = greedy*). Posteriormente, esse arquivo de configuração foi enviado para a plataforma CIPRES aonde o Partitionfinder2 foi executado para encontrar os modelos de substituição mais adequados a um determinado esquema de partição que também foi escolhido pelo programa. Com o objetivo de verificar o impacto dos esquemas de partição na topologia e suporte das árvores, os resultados das análises baseadas em um determinado tipo de alinhamento particionado por posição do códon, particionado por gene e particionado por gene e posição do códon foram comparados.

As análises filogenéticas de verossimilhança e bayesianas foram baseadas em 14 conjuntos que compreendiam um tipo de alinhamento (CDS, rRNA-CDS, rRNA-CDS-tRNA, rRNA-CDS-tRNA-D-loop, aminoácidos) sem partição ou com um determinado tipo de partição somando 28 análises filogenéticas no total. Os 14 conjuntos serão descritos nos próximos parágrafos. No caso das análises particionadas, as partições descritas nos próximos

parágrafos se referem às partições pré-definidas no arquivo de configuração do Partitionfinder2 antes de executar esse programa (ver Tabela 2 para informações sobre os modelos de substituição sugeridos pelo Partitionfinder2 para cada partição após execução desse programa).

Conjunto 1: esse conjunto é composto por um alinhamento de todos os genes codificadores de proteínas (CDS) concatenados com partição única.

Conjunto 2: composto por um alinhamento de genes codificadores de rRNA e CDS concatenados (rRNA-CDS) com partição única.

Conjunto 3: composto por um alinhamento de genes codificadores de rRNA, CDS e genes codificadores de tRNA concatenados (rRNA-CDS-tRNA) com partição única.

Conjunto 4: composto por um alinhamento de rRNA, CDS, tRNA e a região D-loop concatenados (rRNA-CDS-tRNA-D-loop) com partição única.

Conjunto 5: composto por um alinhamento de aminoácidos com partição única.

Conjunto 6: composto por um alinhamento de CDS particionado por posição do códon. Nesse conjunto a primeira, segunda e terceira posição dos códons dos CDS estão em partições separadas.

Conjunto 7: composto por um alinhamento de rRNA-CDS particionado por posição do códon. Nesse conjunto os genes codificadores de rRNA 12S e 16S estão em uma mesma partição, enquanto que a primeira, segunda e terceira posição dos códons dos CDS estão em partições separadas.

Conjunto 8: composto por um alinhamento de rRNA-CDS-tRNA particionado por posição do códon. Nesse conjunto os dois genes codificadores de rRNA estão em uma mesma

partição, todos os 21 genes codificadores de tRNA estão em uma mesma partição separada das partições dos genes de RNA e CDS, e a primeira, segunda e terceira posição dos códons dos CDS estão em partições separadas.

Conjunto 9: composto por um alinhamento de CDS particionado por gene. Nesse conjunto cada um dos 13 genes codificadores de proteínas (CDS) foi alocado em partições separadas.

Conjunto 10: composto por um alinhamento de rRNA-CDS particionado por gene. Nesse conjunto os genes de rRNA foram alocados em uma mesma partição, enquanto que cada um dos CDS foi alocado em partições separadas.

Conjunto 11: composto por um alinhamento de rRNA-CDS-tRNA particionado por gene. Nesse conjunto os genes de rRNA foram alocados em uma mesma partição, todos os genes de tRNA foram alocados em uma mesma partição separada das partições de rRNA e CDS, e cada um dos CDS foi alocado em partições separadas.

Conjunto 12: composto por um alinhamento de CDS particionado por gene e posição do códon. Nesse conjunto todos os 13 CDS foram alocados em partições separadas e cada uma dessas partições foram subparticionadas na primeira, segunda e terceira posição do códon.

Conjunto 13: composto por um alinhamento de rRNA-CDS particionado por gene e posição do códon. Nesse conjunto os dois genes de rRNA foram alocados em uma mesma partição, enquanto que todos os 13 CDS foram alocados em partições separadas, sendo que cada um dessas partições foi subparticionadas em primeira, segunda e terceira posição do códon.

Conjunto 14: composto por um alinhamento de rRNA-CDS-tRNA particionado por gene e posição do códon. Nesse conjunto, os dois genes de rRNA foram alocados em uma mesma partição, os 21 genes de tRNA foram alocados na mesma partição separada das partições de rRNA, e todos os 13 CDS foram alocados em partições separadas, sendo que cada uma dessas partições foi subparticionadas em primeira, segunda e terceira posição do códon.

3.5 Análises filogenéticas com topologias fixadas

Análises de verossimilhança para todas as 105 topologias subfamiliares bifurcadas possíveis que representavam as relações entre as subfamílias de Cricetidae foram realizadas no RAxML com o objetivo de verificar qual topologia explica consideravelmente melhor os dados usando o critério de Akaike. O resultado dessa análise indicaria se existe ou não uma topologia subfamiliar que se ajusta consideravelmente melhor ao alinhamento dos genomas mitocondriais das espécies incluídas no presente trabalho.

Antes de realizar as análises no RAxML, todas as topologias subfamiliares bifurcadas que representavam as relações entre as subfamílias de Cricetidae foram desenhadas no formato newick usando o comando `-- newick-tools --utree_all filename.nwk` <https://github.com/xflouris/newick-tools> que é capaz de desenhar todas as topologias bifurcadas para um conjunto de táxons terminais cujos nomes são escritos em um arquivo no formato newick. Nesse primeiro passo, os táxons terminais foram as cinco subfamílias de Cricetidae mais o grupo externo e.g. (Outgroup, Tylomyinae, Sigmodontinae, Neotominae, Cricetinae e Arvicolinae). Esse comando armazena cada uma das topologias no formato newick em arquivos separados e nomeia os arquivos em ordem alfabética. Por exemplo, a primeira topologia é armazenada no arquivo com nome xaa, a segunda topologia é

armazenada no arquivo com nome xba, a terceira topologia no xca e assim sucessivamente. Portanto, foi tomada a decisão de dar uma legenda numérica para cada uma das topologias seguindo a ordem alfabética dos nomes do arquivo. Logo, a topologia no arquivo xaa se refere à topologia subfamiliar #1, a topologia no arquivo xba se refere à topologia subfamiliar #2 e assim sucessivamente (ver Figura 9). Ao final da execução do comando *newick-tools --utree_all* filename.nwk foram gerados 105 arquivos com nomes que sempre começavam com a letra x. Assim foi possível usar o comando *cat. x* > unrooted_trees.newick* para armazenar as 105 topologias em um único arquivo. Depois de armazenar todas as 105 topologias subfamiliares bifurcadas no arquivo *unrooted_trees.newick*, o comando *newick-tools --root Outgroup --unrooted_trees.newick --output_file cricetidae_topologies.txt* foi executado para enraizar as 105 topologias no grupo externo, o que gerou o arquivo *cricetidae_topologies.txt* que armazenou todas as 105 topologias subfamiliares bifurcadas e enraizadas para os 6 táxons terminais. Logo em seguida, o arquivo *cricetidae_topologies.txt* foi importado para o PAUP v.4 (Swofford, 2002) juntamente com outro arquivo fasta que continha o alinhamento dos mitogenomas das espécies analisadas no presente trabalho. No PAUP foi criado um subconjunto para cada subfamília e o grupo externo. Esses subconjuntos foram preenchidos com o cabeçalho (gênero_espécie_identificador) das sequências dos mitogenomas armazenados no arquivo fasta. Desta maneira, cada nome de espécie presente no cabeçalho do arquivo fasta foi alocado no subconjunto referente à sua respectiva subfamília. Assim foi possível realizar no PAUP, uma análise filogenética para cada uma das 105 topologias subfamiliares com todas as sequências que estavam presentes no arquivo de alinhamento fasta. Desta maneira, o PAUP gerou 105 árvores no formato newick representando todas as topologias subfamiliares bifurcadas de Cricetidae com os terminais da árvore sendo as sequências dos mitogenomas das espécies de Cricetidae presentes no alinhamento utilizado nesse estudo.

Depois de todas as etapas descritas no parágrafo anterior, foi possível executar no RAxML análises de verossimilhança fixadas as 105 topologias subfamiliares bifurcadas baseando-se no alinhamento dos mitogenomas das espécies incluídas neste trabalho. O objetivo desta análise foi calcular o valor de verossimilhança para cada uma das 105 topologias bifurcadas que representavam as relações entre as subfamílias de Cricetidae para verificar se existia uma topologia que se ajustava consideravelmente melhor ao alinhamento dos mitogenomas. Em outras palavras, essa seria a topologia que explica consideravelmente melhor a evolução das sequências dos mitogenomas em comparação as outras 104 topologias. Como existiam 105 topologias a serem analisadas baseando-se em um alinhamento, foi escrito um script na linguagem *bash* (Apêndice 1) para automatizar o processo de executar o RAxML com todas 105 topologias baseando-se no alinhamento dos mitogenomas das espécies incluídas neste trabalho. O script realizava um *loop* para executar o RAxML com uma das 105 topologias a cada iteração do loop contra o alinhamento dos mitogenomas. Desta forma foi possível calcular o valor de verossimilhança de cada uma das 105 topologias subfamiliares bifurcadas.

Através dos valores de verossimilhança para cada uma das topologias foi possível usar o critério de Akaike (AIC) para saber qual topologia se ajustava consideravelmente melhor ao alinhamento dos mitogenomas das espécies incluídas no presente estudo. Para saber qual topologia se ajusta consideravelmente melhor ao alinhamento dos mitogenomas, calculou-se o valor de AIC usando o valor de verossimilhança de cada uma das 105 topologias através da fórmula:

$$AIC = -2L + 2K$$

Nessa fórmula, L representa o valor de verossimilhança de uma determinada topologia e K representa o número de parâmetros que no caso foi 121 (111 parâmetros relacionados ao tamanho de ramos mais 10 parâmetros do modelo de substituição GTR + G + I). Com o valor de AIC para cada topologia (ver Tabela 3), foi possível calcular a diferença entre o valor de AIC de cada uma das 104 topologias e o menor valor de AIC da topologia subfamiliar que melhor se ajustava ao alinhamento dos mitogenomas. O valor de delta serviu como parâmetro para saber se a topologia subfamiliar com o menor valor de AIC se ajustava consideravelmente melhor ao alinhamento em comparação a outras topologias subfamiliares. A topologia com menor valor de AIC se ajustava consideravelmente melhor ao alinhamento em comparação a outra topologia quando o valor de delta entre elas era superior a quatro (Burnham and Anderson, 2002).

4. Resultados

4.1 Montagem dos mitogenomas e anotações

Os mitogenomas das oito espécies de Cricetidae foram montados com aproximadamente 99,796 a 219,506 fragmentos de DNA resultando em uma profundidade média entre 509x a 1,254x dependendo da espécie (Tabela 4). A montagem do mitogenoma da espécie *Nyctomys sumichrasti* foi aquela que utilizou a maior quantidade de reads resultando também na maior profundidade média em comparação à montagem dos mitogenomas das outras espécies. A montagem do mitogenoma da espécie *Sigmodon alstoni* foi aquela que utilizou a menor quantidade de fragmentos de DNA resultando na menor profundidade média em comparação a montagem dos mitogenomas das outras espécies. *Irenomys tarsalis* foi a espécie o qual apresentou o mitogenoma de maior comprimento com

16,921 bp enquanto que *Oligoryzomys stramineus* apresentou o menor mitogenoma com 16,314 bp (Tabela 4).

Os genomas mitocondriais das oito espécies de Cricetidae sequenciadas no presente trabalho apresentavam dois rRNA, 13 genes codificadores de proteínas (CDS) e 21 tRNA na mesma ordem dos genes mitocondriais presentes no genoma mitocondrial de espécies de roedores sequenciados em outros trabalhos (Kim and Park, 2012). Esses mitogenomas apresentavam tamanho entre 16,314 bp – 16,921 bp. Entre esses mitogenomas, o maior CDS foi o ND5, com comprimento que variava entre 1,809 bp – 1,827 bp dependendo da espécie, enquanto que o menor CDS foi o ATP8 com comprimento de 204 bp. Durante o processo de anotação do genoma mitocondrial das espécies cujo genoma foi sequenciado nesse projeto, pode-se perceber que os códons de início para os CDS eram GTG, ATT, ATG, ATA, ATC enquanto que os códons de terminação eram TAA, T++ - representação para o códon de terminação incompleto em que a cauda poli A é adicionada pela enzima poli A polimerase na extremidade 3' do mRNA durante o processamento dessa molécula (Tomecki et al., 2004) - e TAG. O códon de início mais frequente foi o ATG com 71 casos, enquanto que o menos frequente foi ATC com quatro casos. O códon de terminação mais frequente foi o TAA com 51 casos, enquanto que o TAG ocorreu apenas sete vezes entre as sequências deste trabalho (Tabela 5). Todos os CDS estavam no sentido *forward*, exceto o ND6, que estava no sentido *reverse*, o que também é um padrão comum em genomas mitocondriais de vertebrados em geral. A espécie cujo mitogenoma apresentou a maior porcentagem (42%) de conteúdo GC foi *Irenomys tarsalis*, enquanto que o mitogenoma da espécie *Oligoryzomys stramineus* apresentou a menor porcentagem (32,9%) de conteúdo GC.

4.2 Análises filogenéticas baseadas em genes concatenados e aminoácidos.

Foram realizadas 14 análises de verossimilhança (MV) e 14 análises de inferência bayesiana (IB) baseadas em diferentes conjuntos (Tabela 6). Todas as análises filogenéticas baseadas em genes mitocondriais concatenados recuperaram a família Cricetidae como grupo irmão da família Muridae e todas as subfamílias de Cricetidae formaram grupos monofiléticos com alto suporte do bootstrap e probabilidade posterior (Figuras 10 e 11, Apêndices 2 a 27). A maioria dos resultados indica que as árvores com a topologia subfamiliar #32 são provavelmente aquelas que melhor explicam a evolução das sequências dos mitogenomas e que também apresentam a maior probabilidade de representarem as reais relações entre as subfamílias de Cricetidae segundo o alinhamento utilizado nas análises. Nessa topologia, a subfamília Neotominae foi recuperada como grupo irmão do clado formado pelas subfamílias Tylomyinae, Sigmodontinae, Cricetinae e Arvicolinae (clado B). Dentre deste clado, dois subgrupos são recuperados: Cricetinae e Arvicolinae (clado C) e Tylomyinae e Sigmodontinae (Clado D). Entre as 14 análises de máxima verossimilhança e inferência bayesiana, o clado D apresentou os maiores suportes do bootstrap e probabilidade posterior. O clado D não foi recuperado apenas nas análises bayesianas baseadas no alinhamento de aminoácidos (Tabela 6). Ademais, a topologia subfamiliar #32 apresentou o maior valor de verossimilhança (-345646) e o menor valor de AIC (691537) em comparação as outras 104 topologias enraizadas e bifurcadas que representavam o relacionamento entre as cinco subfamílias de Cricetidae (Tabela 3).

Entre as 14 análises de verossimilhança, a topologia subfamiliar #32 foi recuperada em 11 análises, enquanto que nas análises bayesianas essa topologia foi recuperada em sete das 14 análises (Tabela 6). Além disso, a árvore com a topologia subfamiliar #32 foi recuperada em algumas análises de MV e IB baseadas em um mesmo conjunto, havendo

cinco casos em que esses dois métodos convergiram na inferência dessa topologia subfamiliar. Entre as quatro topologias subfamiliares recuperadas nas análises filogenéticas, a topologia subfamiliar #32 foi a única a ser recuperada nas análises de MV e IB baseadas no mesmo conjunto (Tabela 6). As análises baseadas nos conjuntos 1 (CDS com partição única), 3 (rRNA-CDS-tRNA com partição única), 11 (rRNA-CDS-tRNA particionada por gene), 12 (CDS particionado por gene e posição do códon), 14 (rRNA-CDS-tRNA particionado por gene e posição do códon) foram aquelas em que a verossimilhança e a inferência bayesiana recuperaram a mesma topologia subfamiliar #32. Entre as análises de MV e IB que resultaram na mesma topologia subfamiliar, aquela baseada no conjunto 11 resultou nos maiores suportes do bootstrap e probabilidade posterior para os principais clados dentro de Cricetidae (Tabela 6). Os detalhes dos resultados de todas as análises filogenéticas serão apresentados em subseções reservadas a cada um dos conjuntos com o objetivo de expor esses resultados de maneira organizada.

4.2.1 Análises com partição única.

Conjunto 1: As análises de máxima verossimilhança (MV) e inferência bayesiana (IB) baseadas em alinhamento de CDS com partição única resultaram na mesma topologia subfamiliar #32. Os clados B, C e D apresentaram suporte do bootstrap igual 46%, 68% e 98% (Apêndice 2) e probabilidade posterior igual a 0,92, 0,99 e 1,00, respectivamente (Apêndice 3).

Conjunto 2: as análises MV e IB baseadas no alinhamento de rRNA-CDS com partição única resultaram nas topologias subfamiliares #29 e #32, respectivamente. Na árvore o qual apresentou a topologia #29, o clado formado por Sigmodontinae e Tylomyinae (clado

D) saiu como grupo irmão do clado formado por Neotominae, Cricetinae e Arvicolinae (clado E), sendo Neotominae grupo irmão do subclado formado por Cricetinae e Arvicolinae (clado C) (Apêndice 4). Na árvore MV de topologia subfamiliar #29 os cladogramas D, C e E apresentaram suporte do bootstrap igual 77%, 96% e 45%, enquanto que a árvore IB de topologia subfamiliar #32 apresentou os cladogramas B, C e D com probabilidade posterior igual a 0,70, 1 e 1, respectivamente (Apêndice 5).

Conjunto 3: as análises MV e IB baseadas no alinhamento de rRNA-CDS-tRNA com partição única resultaram na topologia subfamiliar #32. Nas árvores MV e IB, os cladogramas B, C e D apresentaram suporte do bootstrap igual a 54%, 82% e 97% (Apêndice 6) e probabilidade posterior igual a 0,97, 1 e 1 (Apêndice 7), respectivamente.

Conjunto 4: as análises MV e IB baseadas no alinhamento de rRNA-CDS-tRNA-D-loop com partição única resultaram nas topologias subfamiliares #32 e #29, respectivamente. A árvore MV recuperou os cladogramas B, C e D apresentando suporte do bootstrap igual a 48%, 82% e 95% (Apêndice 8), enquanto que a árvore IB recuperou os cladogramas C, D e E com probabilidade posterior igual a 1, 1 e 0,56 (Apêndice 9), respectivamente.

Conjunto 5: a análise MV baseada no alinhamento de aminoácidos com partição única resultou na topologia subfamiliar #29, enquanto que a análise IB resultou na topologia subfamiliar #64. Na árvore MV com topologia subfamiliar #29, os cladogramas C, D e E apresentaram suporte do bootstrap igual 62%, 52% e 43% (Apêndice 10), respectivamente. Em contrapartida, a árvore IB com topologia subfamiliar #64 resultou em Tylomyinae como grupo irmão do clado formado por Sigmodontinae, Neotominae, Cricetinae e Arvicolinae o qual apresentou probabilidade posterior igual a 0,99 e a subfamília Sigmodontinae foi recuperada como grupo irmão do clado formado por Neotominae, Cricetinae e Arvicolinae o qual apresentou probabilidade posterior igual a 1. Essa árvore também recuperou o clado C

(Arvicolinae, Cricetinae) o qual apresentou probabilidade posterior igual a 0,56 (Apêndice 11).

4.2.2 Análises particionados por posição do códon.

Conjunto 6: as análises MV e IB baseada no alinhamento de CDS particionado por posição do códon resultaram na topologia subfamiliar #32 e #26, respectivamente. A árvore MV recuperou os clados B, C e D com valores de bootstrap igual 46%, 64% e 98% (Apêndice 12), respectivamente. A árvore IB com topologia subfamiliar #26 recuperou a subfamília Cricetinae como grupo irmão do clado formado por Tylomyinae, Sigmodontinae, Neotominae e Arvicolinae (clado F), e o clado D (Tylomyinae, Sigmodontinae) saiu como grupo irmão do clado formado por Neotominae e Arvicolinae (clado G). Nessa árvore, os clados D, F e G apresentaram probabilidade posterior igual a 1, 0,54 e 0,54 (Apêndice 13), respectivamente.

Conjunto 7: as análises MV e IB baseadas no alinhamento de rRNA-CDS com CDS particionado por posição do códon resultaram na topologia subfamiliar #32 e #26, respectivamente. Nas árvores MV e IB, os clados B, C e D apresentaram bootstrap igual 43%, 75% e 96% (Apêndice 14) e os clados D, F e G apresentaram probabilidade posterior igual 1, 0,54 e 0,55 (Apêndice 15), respectivamente.

Conjunto 8: a árvore MV baseada no alinhamento de rRNA-CDS-tRNA particionada por posição do códon resultou na topologia subfamiliar #32 com os clados B, C e D apresentando 58%, 86% e 97% de suporte do bootstrap (Apêndice 16), respectivamente. Em contrapartida, a árvore IB com topologia subfamiliar #29 recuperou os clados C, D e E com probabilidade posterior igual a 0,79, 1 e 0,65 (Apêndice 17), respectivamente.

4.2.3 Análises particionadas por gene.

Conjunto 9: a análise MV baseada no alinhamento de CDS particionado por gene resultou na topologia subfamiliar #32, enquanto que a análise IB resultou na topologia subfamiliar #26. A árvore com topologia subfamiliar #32 recuperou os clados B, C e D apresentando 54%, 46% e 97% de suporte do bootstrap (Apêndice 18), respectivamente. Em contrapartida, a árvore com topologia subfamiliar #26 resultou nos clados D, F e G com probabilidade posterior igual a 1, 0,94 e 0,94 (Apêndice 19), respectivamente.

Conjunto 10: as análises MV e IB baseadas no alinhamento de rRNA-CDS particionado por gene resultaram na topologia subfamiliar #32 e #26, respectivamente. A árvore MV recuperou os clados B, C e D com suporte do bootstrap igual a 44%, 67% e 91% (Apêndice 20), enquanto que a árvore IB recuperou os clados D, F e G com probabilidade posterior igual a 1,00, 0,51 e 0,51 (Apêndice 21), respectivamente.

Conjunto 11: as análises MV e IB baseada no alinhamento rRNA-CDS-tRNA particionado por gene resultaram na topologia subfamiliar #32 (Figuras 10 e 11). As árvores MV e IB recuperaram os clados B, C e D com suporte do bootstrap igual a 56%, 80% e 96% e probabilidade posterior igual a 0,99, 0,99 e 1, respectivamente.

4.2.4 Análises particionadas por gene e posição do códon.

Conjunto 12: a análise de MV baseada no alinhamento de CDS particionado por gene e posição do códon resultou na topologia subfamiliar #32. As análises MV e IB recuperaram os clados B, C e D com suporte do bootstrap igual 43%, 53% e 98% (Apêndice 22) e probabilidade posterior igual a 0,60, 0,61 e 1 (Apêndice 23), respectivamente.

Conjunto 13: as análises MV e IB baseadas no alinhamento de rRNA-CDS particionado por gene e posição do códon resultaram nas topologias subfamiliares #29 e #32 respectivamente. A árvore MV recuperou os clados C, D e E com suporte do bootstrap igual a 48%, 83% e 42% (Apêndice 24), enquanto que a análise IB recuperou os clados B, C e D com probabilidade posterior igual a 0,56, 0,59 e 1 (Apêndice 25), respectivamente.

Conjunto 14: As análises MV e IB baseadas no alinhamento de rRNA-CDS-tRNA particionado por gene e posição do códon resultaram na topologia subfamiliar #32. As árvores MV e IB recuperaram os clados B, C e D com suporte do bootstrap igual a 49%, 74% e 93% (Apêndice 26) e probabilidade posterior igual a 0,97, 0,98 e 1 (Apêndice 27), respectivamente.

4.3 Análises baseadas em genes individuais.

Várias análises filogenéticas baseadas em genes individuais não recuperaram a família Cricetidae como grupo monofilético. Integrantes do grupo externo foram recuperados dentro do grupo interno nas análises filogenéticas baseadas nos genes ND1, COX1, COX2, ATP8, ATP6 e ND3. Ademais, as subfamílias de Cricetidae não formaram grupos monofiléticos em algumas das análises baseadas em genes individuais. A subfamília Arvicolinae não formou grupo monofilético nas análises baseadas em ATP8 e ND3; a subfamília Cricetinae não foi recuperada como grupo monofiléticos nas análises baseadas em COX2, COX3 e ND6; a subfamília Neotominae não saiu como grupo monofilético na análise baseada em COX3; a subfamília Sigmodontinae não foi recuperada como grupo monofilético nas análises baseadas em ATP8 e COX3; a subfamília Tylomyinae não saiu como grupo monofilético nas análises baseadas nos genes ND1, ND2, COX1, ATP8, COX3, ND4L e ND6. As análises filogenéticas baseadas em CYTB, ND4 e ND5 foram às únicas que recuperaram a família

Cricetidae e todas as suas subfamílias como grupos monofiléticos, mas a maioria dos principais clados dentro dessa família apresentou baixo suporte do bootstrap. Os seguintes parágrafos descrevem os detalhes dos resultados de cada uma das análises baseadas em genes individuais.

A análise 14 MV baseada no alinhamento de CYTB resultou em uma árvore o qual apresentou a topologia subfamiliar #32. Nessa árvore, os clados B, C e D apresentaram suporte do bootstrap igual a 73%, 66% e 54% (Apêndice 28), respectivamente.

A análise MV baseada no alinhamento de ND1 resultou em uma árvore (Apêndice 29) em que *Mus musculus*, *Psammomys obesus* e *Acomys cahirinus* foram recuperadas dentro do grupo com integrantes da família Cricetidae. Nessa árvore a subfamília Tylomyinae não foi recuperada como grupo monofilético, sendo a espécie *Tylomys nudicaudus* irmã da subfamília Arvicolinae e *Nyctomys sumichrasti* espécie irmã da subfamília Cricetinae.

A análise MV baseada no alinhamento de ND2 resultou em uma árvore (Apêndice 30) em que a subfamília Tylomyinae não foi recuperada como grupo monofilético sendo a espécie *Tylomys nudicaudus* irmã do grupo formado por *Nyctomys sumichrasti* e espécies da subfamília Sigmodontinae. Além disso, Cricetinae saiu como grupo irmão do clado formado por Neotominae e Arvicolinae o qual apresentou suporte do bootstrap igual a 43%.

A análise MV baseada no alinhamento de COX1 resultou em uma árvore (Apêndice 31) em que as espécies *Acomys cahirinus*, *Mus musculus* e *Psammomys obesus* foram recuperadas dentro do grupo formado por integrantes da família Cricetidae. Além disso, a subfamília Tylomyinae não formou um grupo monofilético, sendo a espécie *Nyctomys sumichrasti* irmã do grupo formado *Tylomys nudicaudus*, *Acomys cahirinus*, *Mus musculus*, *Psammomys obesus* e os de mais integrantes da família Cricetidae. Essa árvore também recuperou Sigmodontinae

como grupo irmão do clado formado por Neotominae e Arvicolinae o qual apresentou 19% de suporte do bootstrap.

A análise de verossimilhança baseada no alinhamento do gene COX2 resultou em uma árvore (Apêndice 32) em que *Acomys cahirinus* foi recuperado dentro do grupo formado por integrantes da família Cricetidae. Ademais, a subfamília Cricetinae não formou um grupo monofilético, sendo *Cricetulus kamensis* recuperado como espécie irmã da subfamília Sigmodontinae constituindo um clado com 1% de suporte do bootstrap. Além disso, Tylomyinae saiu como grupo irmão da subfamília Arvicolinae formando um clado com 11% de suporte do bootstrap.

A análise MV baseada no alinhamento do gene ATP8 resultou em uma árvore (Apêndice 33) em que *Rhizomys sinensis* e *Nannospalax judaei* foram recuperados dentro do grupo formado por integrantes da família Cricetidae. Nessa árvore, Cricetinae e Neotominae foram as únicas subfamílias de Cricetidae recuperadas como grupos monofiléticos.

A análise MV baseada no gene ATP6 resultou em uma árvore (Apêndice 34) em que *Psammomys obesus*, *Mus musculus* e *Acomys cahirinus* foram recuperados dentro do grupo formado por alguns integrantes da família Cricetidae. Ademais, Sigmodontinae saiu como grupo irmão do clado formado *Psammomys obesus*, *Mus musculus*, *Acomys cahirinus*, Tylomyinae, Neotominae e Arvicolinae o qual apresentou 55% de suporte do bootstrap. Além disso, Arvicolinae saiu como grupo irmão da subfamília Neotominae formando um clado com 69% de suporte do bootstrap.

A análise MV baseada no alinhamento do gene COX3 (Apêndice 35) em que *Psammomys obesus* e *Nannospalax judaei* foram recuperados no grupo formado por integrantes da família Cricetidae. Além disso, Tylomyiane, Sigmodontinae, Cricetinae e Neotominae não formaram grupos monofiléticos.

A análise MV baseada no alinhamento do gene ND3, resultou em uma árvore (Apêndice 36) em que *Mus musculus*, *Psammomys obesus*, *Nannospalax judaei* e *Rhizomys sinensis* foram recuperados no mesmo grupo em que os integrantes da família Cricetidae estavam incluídos. Além disso, o posicionamento de Neotominae em relação a outros grupos presentes na árvore não ficou bem resolvido e a subfamília Sigmodontinae não formou um grupo monofilético.

A análise MV baseada no alinhamento do gene ND4L resultou em uma árvore (Apêndice 37) em que a subfamília Sigmodontinae não formou um grupo monofilético sendo *Oligoryzomys stramineus* espécie irmã do grupo formado por todos os outros integrantes da família Cricetidae incluídos nas análises desse trabalho. Além disso, a subfamília Tylomyinae não foi recuperada como grupo monofilético.

A análise MV baseada no alinhamento do gene ND4 resultou em uma árvore (Apêndice 38) em que a subfamília Cricetinae saiu como grupo irmão do clado formado por Tylomyinae, Sigmodontinae, Neotominae e Arvicolinae o qual apresentou 55% de suporte do bootstrap. Além disso, o clado formado por Sigmodontinae e Tylomyinae apresentou 32% de suporte do bootstrap e foi recuperado como grupo irmão do clado formado por Neotominae e Arvicolinae que teve 25% de suporte do bootstrap.

A análise MV baseada no alinhamento do gene ND5 resultou em uma árvore (Apêndice 39) em que a subfamília Arvicolinae saiu como grupo irmão do clado formado por Sigmodontinae, Tylomyinae, Cricetinae e Neotominae o qual apresentou 31% de suporte do bootstrap. Além disso, Sigmodontinae saiu como grupo irmão do clado formado por Tylomyinae, Cricetinae e Neotominae o qual apresentou 33% de suporte do bootstrap. Ademais, a subfamília Neotominae foi recuperada como grupo irmão do clado formado por Tylomyinae e Cricetinae o qual apresentou 24% de suporte do bootstrap.

A análise MV baseada no alinhamento do gene ND6 resultou em uma árvore (Apêndice 40) em que as subfamílias Cricetinae e Tylomyinae não formaram grupos monofiléticos. A espécie *Nyctomys sumichrasti* foi recuperada como irmã da subfamília Sigmodontinae, e a espécie *Tylomys nudicaudus* foi recuperada como irmã da subfamília Arvicolinae.

4.4 Análise filogenética com topologias fixadas

A topologia subfamiliar #32 apresentou o maior valor de verossimilhança (-345646) e o menor valor de AIC (691537) em comparação as outras 104 topologias enraizadas e bifurcadas que representavam o relacionamento entre as 5 subfamílias de Cricetidae. Em contra partida, a topologia subfamiliar #29 apresentou o segundo maior valor de verossimilhança (-345651) e o segundo menor valor de AIC (691547). A diferença do valor de AIC da topologia 32 para o valor de AIC da topologia subfamiliar #29 foi de 9,53 (Tabela 3).

As topologias subfamiliares #33, #27 e #26 apresentaram o terceiro, quarto e quinto menores valores de AIC, respectivamente. A diferença do valor de AIC da topologia subfamiliar #32 para os valores de AIC das topologias subfamiliares #33, #27 e #26 foi 41,29, 45,58 e 46,46, respectivamente. A topologia subfamiliar #62 apresentou o menor valor de verossimilhança (-345739) e o maior valor de AIC (691723) entre as 104 outras topologias enraizadas e bifurcadas. A diferença do valor de AIC da topologia 32 para o valor de AIC da topologia subfamiliar #62 foi 185,35.

5. Discussão

5.1 Análises filogenéticas

Neste trabalho foram analisados os genomas mitocondriais de 51 espécies pertencentes às cinco subfamílias de Cricetidae. As análises filogenéticas recuperaram com maior consistência a árvore com a topologia subfamiliar #32, o qual apresentou os maiores suportes do bootstrap e probabilidade posterior para os principais clados dentro de Cricetidae e que também obteve o menor valor de AIC na análise com topologias subfamiliares fixadas. O valor delta calculado através da diferença do valor de AIC da topologia subfamiliar #32 para o valor de AIC de cada uma das 104 outras topologias subfamiliares foi superior a 9 em todos os casos. Valores de delta superiores a quatro indicam que o modelo com o menor valor de AIC se ajusta consideravelmente melhor aos dados em comparação aos outros modelos sendo comparados a ele (Burnham and Anderson, 2002). Na topologia subfamiliar #32, o clado B recebe alto suporte apenas nas análises bayesianas baseadas nos conjuntos 3 e 11, enquanto que os clados C e D receberam médio a alto suporte na maioria das análises (Tabela 6). Os principais clados de Cricetidae encontrados no presente estudo serão discutidos nos próximos parágrafos dentro do contexto de outros trabalhos que também recuperam esses grupos ou grupos alternativos.

No presente estudo a subfamília Cricetinae saiu como grupo irmão de Arvicolinae formando o clado C em várias análises filogenéticas. O clado C foi recuperado em todas as análises MV com baixo a médio suporte do bootstrap e com alta probabilidade posterior em cinco análises bayesianas. Em contrapartida, no trabalho de Jansa et al. (2009) a árvore de verossimilhança baseada em IRBP recuperou Cricetinae como grupo irmão do clado F, formado por Tylomyinae, Neotominae, Arvicolinae e Sigmodontinae (clado F no presente

trabalho) o qual apresentou suporte do bootstrap superior a 75%. No presente trabalho, nenhuma análise de MV recuperou Cricetinae como grupo irmão do clado F, enquanto que quatro análises de IB recuperaram Cricetinae como grupo irmão do clado F, o qual apresentou baixa ($pp < 0,95$) probabilidade posterior que variou entre 0,51 a 0,94. Os resultados do presente trabalho indicam que Cricetinae provavelmente não é grupo irmão do clado F, tendo em vista que esse clado não foi recuperado em nenhuma das análises de MV e apresentou baixa ($pp < 0,90$) probabilidade posterior nas análises IB.

O trabalho de Leite et al. (2014) encontrou a subfamília Cricetinae como irmã do grupo constituído por Tylomyinae e Sigmodontinae com essas três subfamílias formando um clado mais abrangente (Cricetinae, (Tylomyinae, Sigmodontinae)) com 45% de suporte do bootstrap. Em contrapartida, no presente estudo todas as análises de MV recuperaram Cricetinae como grupo irmão de Arvicolinae (clado C) com baixo a médio suporte do bootstrap entre 56% e 86%, enquanto que seis análises IB recuperaram o clado C com alta probabilidade posterior e quatro análises IB recuperaram o clado C com baixa probabilidade posterior. Além disso, Steppan and Schenk (2017) também recuperaram a subfamília Cricetinae como grupo irmão de Arvicolinae formando um clado com alto suporte do bootstrap igual a 96%. Os resultados do presente trabalho mais o resultado de Steppan and Schenk (2017) sugerem que Cricetinae é realmente grupo irmão de Arvicolinae.

A subfamília Sigmodontinae foi recuperada como grupo irmão de Arvicolinae formando um clado com médio suporte do bootstrap igual 72% no trabalho de Jansa and Weksler (2004). Em outro trabalho, Jansa et al. (2009) recuperaram Sigmodontinae como grupo irmão de Neotominae formando um clado com baixo suporte do bootstrap ($bs \leq 50\%$) na árvore baseada no gene IRBP e na árvore baseada nos genes IRBP e GHR concatenados. Nesse mesmo trabalho, Jansa et al. (2009) recuperaram a subfamília Sigmodontinae como grupo irmão de Cricetinae com baixo suporte do bootstrap na árvore baseada no gene GHR.

Em contrapartida, Leite et al. (2014) recuperaram Tylomyinae como grupo irmão de Sigmodontinae formando um clado com médio suporte do bootstrap igual a 74%. Além disso, Steppan and Schenk (2017) também encontraram o clado formado por Tylomyinae e Sigmodontinae, com médio suporte do bootstrap igual a 85%. No trabalho de Steppan and Schenk (2017) os autores advogam que o relacionamento entre Sigmodontinae, Tylomyinae e Neotominae ainda não está bem definido com o posicionamento dessas três subfamílias variando bastante entre análises baseadas em diferentes genes, e que provavelmente muito mais genes serão necessários para melhorar a resolução do posicionamento filogenético dessas subfamílias. No presente trabalho, entre as 14 análises de MV, 12 análises recuperam o clado formado por Sigmodontinae e Tylomyinae com alto suporte do bootstrap superior a 90%, a análise de MV baseada em rRNA-CDS recuperou esse clado com 83% de suporte do bootstrap, e análise baseada em alinhamento de aminoácidos recuperou o clado D com 52% de suporte do bootstrap. Ademais, todas as análises de IB baseadas em genes mitocondriais concatenados recuperaram o clado D com probabilidade posterior igual a 1, sendo que a análise de IB baseada em alinhamento de aminoácidos foi a única a não recuperar esse clado. Todos esses resultados indicam fortemente que Tylomyinae é grupo irmão de Sigmodontinae.

Neste trabalho o clado B é constituído pelas subfamílias Tylomyinae, Sigmodontinae, Cricetinae e Arvicolinae, tendo como grupo irmão a subfamília Neotominae. O clado B foi recuperado em 11 das 14 análises de máxima verossimilhança, com baixo suporte do bootstrap ($bs < 70\%$) que variou entre 43% a 58%, enquanto que nas análises de IB esse clado foi recuperado com baixa probabilidade posterior ($pp < 0,90$) em três análises, com média probabilidade posterior ($0,90 \leq pp < 0,95$) em um análise e com alta probabilidade posterior ($pp \geq 0,95$) em três análises. Dentre os trabalhos anteriores que inferiram relações filogenéticas entre roedores da superfamília Muroidea, o clado B foi recuperado apenas no trabalho de Jansa et al. (2006) e no trabalho de Leite et al. (2014) em que esse clado foi

recuperado com médio suporte do bootstrap (71%). Apesar do clado B não ser recuperado em todas as análises, ele recebe alto suporte nas análises de inferência bayesiana baseadas em rRNA-CDS-tRNA com partição única, particionada por gene e particionada por gene e posição do códon (Tabela 6).

No presente trabalho a subfamília Neotominae saiu como grupo irmão do clado B na topologia subfamiliar #32, enquanto que Neotominae saiu como grupo irmão de Arvicolinae formando o clado G na topologia subfamiliar #26. Jansa et al. (2009) encontraram a subfamília Neotominae como grupo irmão de Arvicolinae formando um clado com suporte do bootstrap inferior a 50%. No presente trabalho, nenhuma das 14 análises de MV recuperou Neotominae como grupo irmão de Arvicolinae (clado G), enquanto que cinco análises IB recuperaram o clado G com baixa probabilidade posterior ($pp < 0,95$). Em contrapartida, Steppan and Schenk (2017) recuperaram a subfamília Neotominae como grupo irmão do clado formado por Tylomyinae e Sigmodontinae com essas três subfamílias formando um clado mais abrangente (Neotominae, (Sigmodontinae, Tylomyinae)) o qual apresentou 80% de suporte do bootstrap. Contudo, nenhuma das 14 análises MV e IB do presente trabalho recuperaram Neotominae como grupo irmão do clado formado por Tylomyinae e Sigmodontinae. Portanto, esses resultados contrastantes recuperados por Jansa et al. (2009), Steppan and Schenk (2017) e pelo presente trabalho, indicam que o posicionamento filogenético de Neotominae dentro da família Cricetidae ainda não está bem definido.

5.2 Impacto das partições e classes de dados nas árvores filogenéticas.

No trabalho de Li et al. (2008) que buscou analisar o impacto de esquemas de partição nas análises filogenéticas, os autores demonstraram evidências de que diferentes esquemas de partição podem resultar em variação no suporte para os nós das árvores. Neste trabalho, pode-

se perceber que as variações mais acentuadas dos valores de suporte ocorrem em nós que tendem apresentar baixo suporte de probabilidade posterior, enquanto que os nós que tendem a obter alto suporte não apresentam brusca variação nos valores de probabilidade posterior entre as análises. Um padrão semelhante a esse foi encontrado no presente estudo. Dentre os nós das árvores referentes aos principais clados de Cricetidae, os clados B e C que tendem a recuperar médio a baixo suporte apresentam maior variação nos valores de suporte em comparação ao clado D que recebe alto suporte em grande parte das análises (Figuras 12, 13, 14, 15, 16 e 17).

Os diferentes esquemas de partição afetaram o suporte do bootstrap do clado B com alguns valores do bootstrap variando até 11% entre as análises. A análise que resultou no maior suporte do bootstrap (54%) para o clado B foi aquela particionada por gene, enquanto que a análise que resultou no menor suporte do bootstrap (43%) para esse clado foi aquela baseada em um alinhamento de CDS particionado por gene e posição do códon (Figura 12). Além disso, as análises baseadas em rRNA-CDS-tRNA com diferentes esquemas de partição também resultaram em uma variação nos valores do bootstrap para o clado B, com a análise particionada por posição do códon apresentando o maior suporte do bootstrap (58%) e aquela particionada por gene e posição do códon apresentando o menor suporte (49%) (Figura 12).

Os diferentes esquemas de partição impactaram mais fortemente os valores do bootstrap do clado C. As análises baseadas em CDS que resultaram em maior suporte do bootstrap (68%) para o clado C foram aquelas com partição única, enquanto que a análise baseada em CDS que resultou no menor suporte do bootstrap (46%) foi aquela particionada por gene (Figura 13). As análises baseadas em rRNA-CDS com diferentes esquemas de partição foram aquelas que resultaram em maior variação para o suporte do bootstrap do clado C apresentado valor máximo de 77% e valor mínimo de 48% nas análises com partição única e particionada por gene e posição do códon, respectivamente (Figura 13). Finalmente, as

análises baseadas em rRNA-CDS-tRNA apresentaram suporte máximo do bootstrap de 86% e suporte mínimo de 74% para o clado C nas análises particionadas por posição códon e particionada por gene e posição do códon, respectivamente (Figura 13). O suporte do bootstrap (Figura 14) e probabilidade posterior (Figura 17) para o clado D variaram pouco entre as análises baseadas em alinhamentos de CDS, rRNA-CDS e rRNA-CDS-tRNA com diferentes esquemas de partição.

Estudos baseados em simulações e dados empíricos demonstram evidências de que particionar os dados por gene ou por gene com cada gene subparticionadas por posição do códon pode aumentar significativamente o suporte para os nós das árvores (Brandley et al., 2005; Brown and Lemmon, 2007; Castoe et al., 2004; Caterino et al., 2001; Pupko et al., 2002). No presente trabalho não foi possível identificar tal padrão nos resultados de várias análises. Nas análises baseadas em CDS com partição única o clado C obteve 68% de suporte do bootstrap enquanto que as análises baseadas nessa mesma classe de dados particionadas por gene ou por gene e posição do códon resultaram em 46% e 53% de suporte do bootstrap, respectivamente (Figura 13). Além disso, as análises baseadas em rRNA-CDS com partição única recebeu 77% de suporte do bootstrap para o clado C, enquanto que as análises baseadas nessa mesma classe de dados particionada por gene ou por gene e posição do códon obtiveram suporte do bootstrap igual 67% e 48%, respectivamente (Figura 13). Ademais, as análises baseadas em rRNA-CDS-tRNA com partição única recebeu 82% de suporte do bootstrap para o clado C, enquanto que as análises baseadas nessa classe de dados particionados por gene ou por gene e posição do códon resultaram em suporte do bootstrap igual 80% e 74%, respectivamente. A probabilidade posterior para o clado C também diminuiu bastante em análises com esquemas de partição mais complexos. Todas as análises com partição única resultaram em probabilidade posterior bem mais alta que o suporte de probabilidade posterior encontrado nas análises particionadas por gene e posição do códon

(Figura 16). Esses resultados indicam que o suporte para o clado C das análises do presente estudo tende a diminuir com esquemas de partição mais complexos com um número maior de parâmetros. Contudo, a análise de verossimilhança que recuperou os maiores valores de bootstrap para os principais clados dentro de Cricetidae foi aquela particionada por posição do códon (Figura 10), enquanto que análise bayesiana que recuperou os maiores valores de probabilidade posterior para os principais clados de Cricetidae foi aquela particionada por gene (Figura 11).

Os resultados de alguns trabalhos (Li et al., 2008; Rota and Wahlberg, 2012) também indicam que diferentes esquemas de partição e número de parâmetros podem afetar a topologia das árvores, sendo que as análises bayesianas são mais sensíveis aos diferentes esquemas de partição em comparação as análises de máxima verossimilhança. No presente trabalho pode-se perceber que os diferentes esquemas de partição influenciaram pouco a topologia subfamiliar das árvores de máxima verossimilhança, enquanto que as análises bayesianas baseadas em diferentes esquemas de partição resultaram em maior variação em suas topologias subfamiliares (Tabela 6). Dentre as análises de máxima verossimilhança do presente estudo, aquelas baseadas em rRNA-CDS com diferentes esquemas de partição foram as únicas a resultarem em diferentes topologias subfamiliares (Tabela 6). As análises baseadas no alinhamento de rRNA-CDS com partição única e particionado por gene e posição do códon resultaram na topologia subfamiliar #29, enquanto que as análises de MV baseadas nesse mesmo alinhamento particionado por posição do códon e particionado por gene resultaram na topologia subfamiliar #32 (Tabela 6). As 14 análises IB baseadas em alinhamentos com partição única ou com algum tipo de partição variaram entre as topologias subfamiliares #26, #29, #32 e #64, que foram recuperadas cinco vezes, uma vez, seis vezes e uma vez, respectivamente (Tabela 6).

As cinco análises MV que apresentavam diferentes classes de dados (CDS, rRNA-CDS, rRNA-CDS-tRNA, rRNA-CDS-tRNA-D-loop) no alinhamento com partição única recuperaram a topologia subfamiliar #32 três vezes, enquanto que a topologia subfamiliar #29 foi recuperada duas vezes. As cinco análises IB que apresentavam diferentes classes de dados com partição única resultaram nas topologias subfamiliares #29, #32 e #64, sendo que a topologia subfamiliar #32 foi recuperada três vezes (Tabela 5). Contudo, a topologia das árvores recuperadas no presente estudo pode ter apresentado maior variação a níveis abaixo de subfamília. Essa variação topológica pode ser analisada com mais acurácia através de métodos (Robinson and Foulds, 1981) que calculam distâncias topológicas entre árvores filogenéticas.

Alguns estudos indicam que análises baseadas em alinhamentos com vasta quantidade de caracteres tendem a recuperar árvores bem resolvidas e com alto suporte (Baurain et al., 2007; Comas et al., 2007; McMahon and Sanderson, 2006; Philippe et al., 2005; Rokas et al., 2003a; Rokas et al., 2005; Rokas et al., 2003b). No presente trabalho, as análises baseadas em uma maior quantidade de classes de dados tendiam a apresentar maior suporte do bootstrap e probabilidade posterior para os clados B e C (Figuras 12, 13, 15 e 16). Contudo, adição da região D-loop não aumentou o suporte do bootstrap do clado C (Figura 13) e diminuiu o suporte do bootstrap clado B (Figura 12). As análises filogenéticas inferidas a partir de sequências de aminoácidos, que também tinham a menor quantidade de caracteres em comparação a outras análises, recuperaram os menores valores de suporte para os clados B e C (Figuras 12, 13 e 16).

6. Conclusões

Os resultados do presente trabalho corroboram a hipótese de que Sigmodontinae é grupo irmão de Tylomyinae e que Arvicolinae é grupo irmão de Cricetinae. A subfamília

Neotominae saiu como grupo irmão do clado B que recebeu baixo suporte do bootstrap em todas as análises de máxima verossimilhança que recuperaram esse clado e duas análises bayesianas, recuperaram esse clado com alta probabilidade posterior. Portanto, um trabalho com uma amostragem mais extensa e baseando-se em marcadores nucleares, genoma mitocondrial e caracteres morfológicos, provavelmente resultará em suporte mais robustos para o clado B, podendo melhor esclarecer o posicionamento filogenético da subfamília Neotominae dentro de Cricetidae. Além disso, os diferentes esquemas de partição e classes de dados afetaram pouco a topologia das árvores apresentando um impacto maior no suporte para os principais clados dentro de Cricetidae.

Referências

- Akaike, H., 1973. Information theory as an extension of the maximum likelihood principle. In B. N. Petrov, and F. Csaki, (eds). *Second International Symposium on Information Theory.*, Akademiai Kiado, Budapest. .
- Alston, E.R., 1876. On the classification of the order Glires. *Proc Zool Soc Lond* 44, 61-98.
- Anderson, S., Bankier, A.T., Barrel, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R., Young, I.G., 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290, 457-465.
- Anderson, S., de Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F., Young, I.G., 1982. Complete sequence of bovine mitochondrial DNA conserved features of the mammalian mitochondrial genome. *J Mol Biol* 156, 683-717.
- Ballard, J.W.O., Whitlock, M.C., 2004. The incomplete natural history of mitochondria. *Mol Ecol* 13, 729-744.
- Baurain, D., Brinkmann, H., Philippe, H., 2007. Lack of resolution in the animal phylogeny: closely spaced cladogeneses or undetected systematic errors? *Mol Biol Evol* 24, 6-9.
- Bergsten, J., 2005. A review of long-branch attraction. *Cladistics* 21, 163-193.
- Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W., Clayton, D.A., 1981. Sequence and gene organization of mouse mitochondrial DNA. *Cell* 26, 167-180.
- Boore, J.L., 1999. Animal mitochondrial genomes. *Nucleic Acids Res* 27, 1767-1780.

- Brandley, M.C., Schmitz, A., Reeder, T.W., 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Syst Biol* 54, 373-390.
- Bridge, D., Cunningham, C.W., Schierwater, B., Desalle, R., Buss, L.W., 1992. Class-level relationships in the phylum Cnidaria: Evidence from mitochondrial genome structure. *Proc Natl Acad Sci USA* 89, 8750-8753.
- Brown, J.M., Lemmon, A.R., 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Syst Biol* 56, 643-655.
- Brown, T., Kraus, M., 1979. Origin of the tribosphenic molar and metatherian and eutherian dental formula In: Lilegraven, J., Kielan-Jaworowska, Z., Clements, W. (Eds.), *Mesozoic Mammals*. University of California Press, Berkeley.
- Burnham, K.P., Anderson, D.R., 2002 *Model selection and multimodel inference*. Springer Science + Business Media., New York.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC bioinformatics* 10, 421.
- Cann, L.R., Stoneking, M., Wilson, A.C., 1987. Mitochondrial DNA and human evolution. *Nature* 325, 31-36.
- Capella-Gutierrez, S., Silla-Martinez, J.M., Gabaldon, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972-1973.
- Carver, T., Harris, S.R., Berriman, M., Parkhill, J., McQuillan, J.A., 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28, 464-469.
- Castoe, T.A., Doan, T.M., Parkinson, C.L., 2004. Data partitions and complex models in Bayesian analysis: the phylogeny of Gymnophthalmid lizards. *Syst Biol* 53, 448-469.
- Caterino, M.S., Reed, R.D., Kuo, M.M., Sperling, F.A., 2001. A partitioned likelihood analysis of swallowtail butterfly phylogeny (Lepidoptera:Papilionidae). *Syst Biol* 50, 106-127.
- Chaline, J., Mein, P., Petter, F., 1977. Les grandes lignes d'une classification évolutive des Muroidea. *Mammalia* 41, 245-252.
- Chen, N., Zhao, S., 2009. New progress in snake mitochondrial gene rearrangement. *Mitochondrial DNA* 20, 69-71.
- Chen, Y.C., Liu, T., Yu, C.H., Chiang, T.Y., Hwang, C.C., 2013. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS One* 8, 1-20.
- Clayton, D.A., 1982. Replication of animal mitochondrial DNA. *Cell* 28, 693-705.
- Comas, I., Moya, A., González-Candelas, F., 2007. From Phylogenetics to Phylogenomics: The Evolutionary Relationships of InsectEndosymbiotic γ -Proteobacteria as a Test Case. *Syst Biol* 56, 1-16.

- Cummings, D.J., McNally, K.L., Domenico, J.M., Matsuura, E.T., 1990. The complete DNA sequence of the mitochondrial genome of *Podospira anserina*. *Curr Genet* 17, 375-402.
- da Cruz, M.O.R., Weksler, M., Bonvicino, C.R., Bezerra, A.M.R., Prosdocimi, F., Furtado, C., Geise, L., Catzeflis, F., de Thoisy, B., de Oliveira, L.F.B., Silva, C., de Oliveira, J.A., 2019. DNA barcoding of the rodent genus *Oligoryzomys* (Cricetidae: Sigmodontinae): mitogenomic-anchored database and identification of nuclear mitochondrial translocations (Numts). *Mitochondrial DNA A DNA Mapp Seq Anal*, in press.
- Dietrich, C., Brune, A., 2016. The complete mitogenomes of six higher termite species reconstructed from metagenomic datasets (*Cornitermes* sp., *Cubitermes ugandensis*, *Microcerotermes parvus*, *Nasutitermes corniger*, *Neocapritermes taracua*, and *Termes hospes*). *Mitochondrial DNA A DNA Mapp Seq Anal* 27, 3903-3904.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32, 1792-1797.
- Ellerman, J.R., 1941. The families and genera of living rodents, vol. II. British Museum (Natural History), London.
- Felsenstein, J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39, 783-791.
- Felsenstein, J., 2004. *Inferring Phylogenies*. Sinauer Associates., Sunderland, MA.
- Gan, H.M., Tan, M.H., Austin, C.M., 2016. The complete mitogenome of the red claw crayfish *Cherax quadricarinatus* (Von Martens, 1868) (Crustacea: Decapoda: Parastacidae). *Mitochondrial DNA A DNA Mapp Seq Anal* 27, 385-386.
- Graeber, M.B., Müller, U., 1998. Recent developments in the molecular genetics of mitochondrial disorders. *J Neurol Sci* 153, 251-263.
- Gray, M.W., 1982. Mitochondrial genome diversity and the evolution of mitochondrial DNA. *Can. J. Biochem* 60, 157-171.
- Gray, M.W., Boer, P.H., 1988. Organization and expression of algal (*Chlamydomonas reinhardtii*) mitochondrial DNA. *Philos Trans R Soc Lon B Biol Sci* 319, 135-147.
- Hahn, C., Bachmann, L., Chevreur, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. *Nucleic Acids Res* 41, e129.
- Hartl, D.L., Clark, A.G., 2007. *Principles of population genetics*. Sinauer Associates Inc., Sunderland, Massachusetts.
- He, B., Lai, T., Peng, Z., Wang, X., Pan, L., 2013. Complete mitogenome of the Areolate grouper *Epinephelus areolatus* (Serranidae, Epinephelinae). *Mitochondrial DNA A DNA Mapp Seq Anal* 24, 498-500.

- Jansa, S.A., Barker, F.K., Heaney, L.R., 2006. The pattern and timing of diversification of Philippine endemic rodents: Evidence from mitochondrial and nuclear gene sequences. *Syst Biol* 55, 73-88.
- Jansa, S.A., Giarla, T.C., Lim, B.K., 2009. The phylogenetic position of the rodent genus *Typhlomys* and the geographic origin of Muroidea. *J Mammal* 90, 1083-1094.
- Jansa, S.A., Weksler, M., 2004. Phylogeny of muroid rodents: Relationships within and among major lineages as determined by IRBP gene sequences. *Mol Phylogenet Evol* 31, 256-276.
- Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. Mammalian protein metabolism, III. New York: Academic Press., New York.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647-1649.
- Kim, H.R., Park, Y.C., 2012. The complete mitochondrial genome of the striped field mouse, *Apodemus agrarius* (Rodentia, Murinae) from Korea. *Mitochondrial DNA A DNA Mapp Seq Anal.* 23, 145-147.
- Kroemer, G., Dallaporta, B., Resche-Rigon, M., 1998. The mitochondrial death/life regulator in apoptosis and necrosis. *Annu Rev Physiol* 619-642.
- Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol Biol Evol.* 34, 772-773.
- Leite, R.N., Kolokotronis, S.O., Almeida, F.C., Werneck, F.P., Rogers, D.S., Weksler, M., 2014. In the wake of invasion: tracing the historical biogeography of the South American cricetid radiation (Rodentia, Sigmodontinae). *PLoS One* 9, e100687.
- Li, C., Lu, G., Orti, G., 2008. Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. *Syst Biol* 57, 519-539.
- Matioli, S.R., Fernandes, F.M.C., 2012. *Biologia molecular e evolução*. Holos, Ribeirão Preto, São Paulo.
- McMahon, M.M., Sanderson, M.J., 2006. Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. *Syst Biol* 55, 818-836.
- McVay, J.D., Carstens, B.C., 2013. Phylogenetic Model Choice: Justifying a Species Tree or Concatenation Analysis. *J Phylogenetics Evol Biol.* 1, 1-114.
- Metzker, M.L., 2010. Sequencing technologies - the next generation. *Nature reviews. Genetics* 11, 31-46.
- Miller, G.S., Gidley, J.W., 1918. Synopsis of supergeneric groups of rodents. *J Wash Acad Sci* 8, 431-448.

- Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Enviroments Workshops, New Orleans, LA.
- Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 15, 1.
- Musser, G.G., Carleton, M.D., 2005. Order Rodentia. In: Wilson, D.E., Reeder, D.M. (Eds.), *Mammal Species of the World*, 3rd ed. Johns Hopkins University Press, Baltimore, pp. 894-1531.
- Nosek, J., Tomaska, L., 2003. Mitochondrial genome diversity: evolution of the molecular architecture and replication strategy. *Curr Genet* 44, 73-84.
- Nowak, R., 1999. *Walker's Mammals of the World*, vol. II. Baltimore and London: The Johns Hopkins University Press.
- Patton, J.L., Pardinás, U.F., D'Elía, G., 2015. *Mammals of South America*. The University of Chicago Press, Chicago and London.
- Pesole, G., Gissi, C., De Chirico, A., Saccone, C., 1999. Nucleotide substitution rate of mammalian mitochondrial genomes. *J Mol Evol* 48, 427-434.
- Philippe, H., Lartillot, N., Brinkmann, H., 2005. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol Biol Evol* 22, 1246-1253.
- Posada, D., 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25, 1253-1256.
- Pupko, T., Huchon, D., Cao, Y., Okada, N., Hasegawa, M., 2002. Combining Multiple Data Sets in a Likelihood Analysis: Which Models are the Best? *Mol Biol Evol* 19, 2294-2307.
- Rambaut, A., Suchard, M.A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7 *Syst Biol.* 67, 901-904.
- Ridley, M., 2013. *Evolução ARTMED*, Porto Alegre, Rio Grande do Sul.
- Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. *Math Biosci* 53, 131-147.
- Rokas, A., King, N., Finnerty, J., Carroll, S.B., 2003a. Conflicting phylogenetic signals at the base of the metazoan tree. *Evol Dev* 5, 346-359.
- Rokas, A., Krüger, D., Carroll, S.B., 2005. Animal evolution and the molecular signature of radiations compressed in time. *Science* 310, 1933-1938.
- Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003b. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798-804.

- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61, 539-542.
- Rota, J., Wahlberg, N., 2012. Exploration of data partitioning in an eight-gene data set: phylogeny of metalmark moths (Lepidoptera, Choreutidae). *Zool Scr* 41, 536-546.
- Russo, C.A.M., Takezaki, N., Nei, M., 1995. Molecular phylogeny and divergence times of drosophilid species. *Evol* 12, 391-404.
- Schomaker-Bastos, A., Prosdocimi, F., 2018. mitoMaker: a pipeline for automatic assembly and annotation of animal mitochondria using raw NGS data. *Preprints* 1, 1-10.
- Schuster, S.C., 2008. Next-generation sequencing transforms today's biology. *Nat Methods* 5, 16-18.
- Shen, W., Le, S., Li, Y., Hu, F., 2016. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One* 11, 1-10.
- Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A., Waterston, R.H., 2017. DNA sequencing at 40: past, present and future. *Nature* 550, 345-353.
- Shendure, J., Ji, H., 2008. Next-generation DNA sequencing. *Nat Biotechnol* 26, 1135-1145.
- Silva, S., Camino, L.P., Aguilera, A., 2018. Human mitochondrial degradosome prevents harmful mitochondrial R loops and mitochondrial genome instability. *Proc Natl Acad Sci USA* 115, 11024-11029.
- Simpson, G.G., 1945. The principles of classification and a classification of mammals. *Bull Am Mus Nat Hist* 85, 1-350.
- Sims, D., Sudbery, I., Iltott, N.E., Heger, A., Ponting, C.P., 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nature reviews. Genetics* 15, 121-132.
- Smith, D.R., 2016. The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Brief Funct Genomics* 15, 47-54.
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312-1313.
- Steppan, S.J., Adkins, R.M., Anderson, J., 2004. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. *Syst Biol* 53, 533-553.
- Steppan, S.J., Schenk, J.J., 2017. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. *PLoS One* 12, 1-31.
- Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

- Tavares, W.C., Seuanéz, H.N. Changes in selection intensity on the mitogenome of subterranean and fossorial rodents respective to aboveground species. *Mamm Genome* 29, 353-363.
- Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences. Am Math Soc* 17, 57-86.
- Thomas, O., 1896. On the Genera of Rodents: an Attempt to Bring up to Date the Current Arrangement of the Order. *Proc Zool Soc Lond* 1896, 1012-1028.
- Tomecki, R., Dmochowska, A., Gewartowski, K., Dziembowski, A., Stepień, P.P., 2004. Identification of a novel human nuclear-encoded mitochondrial poly(A) polymerase. *Nucleic Acids Res* 32, 6001-6014.
- Tullberg, T., 1899. Ueber das System der Nagethiere; eine phylogenetische Studie. *Nova Acta regiae Soc Sci Upsal* 3, 1-514.
- Vahrenholz, C., Riemen, G., Pratje, E., Dujon, B., Michaelis, G., 1993. Mitochondrial DNA of *Chlamydomonas reinhardtii*: the structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication. *Curren Genet* 24, 241-247.
- Vasta, V., B Ng, S., Turner, E.H., Shendure, J., Hahn, S.H., 2009. Next generation sequence analysis for mitochondrial disorders. *Genome Med* 1, 1-100.
- Vieira, G.A., Prosdocimi, F., 2019. Accessible molecular phylogenomics at no cost: obtaining 14 new mitogenomes for the ant subfamily Pseudomyrmecinae from public data. *PeerJ* 7, 1-26.
- Wei, Y.H., 1998. Oxidative stress and mitochondrial DNA mutations in human aging. *SEBM* 217, 53-63.
- Yang, Z., 2014. *Molecular evolution: A statistical approach*. Oxford University Press., Oxford.
- Yang, Z., Goldman, N., Friday, A., 1995. Maximum Likelihood Trees from DNA Sequences: A Peculiar Statistical Estimation Problem. *Syst Biol* 44, 384-399.
- Yang, Z., Rannala, B., 2012. *Molecular phylogenetics: principles and practice*. Nature 13, 303-314.
- Zink, R.M., Barrowclough, G.F., 2008. Mitochondrial DNA under siege in avian phylogeography. *Mol Ecol* 17, 2107-2121.

FIGURAS

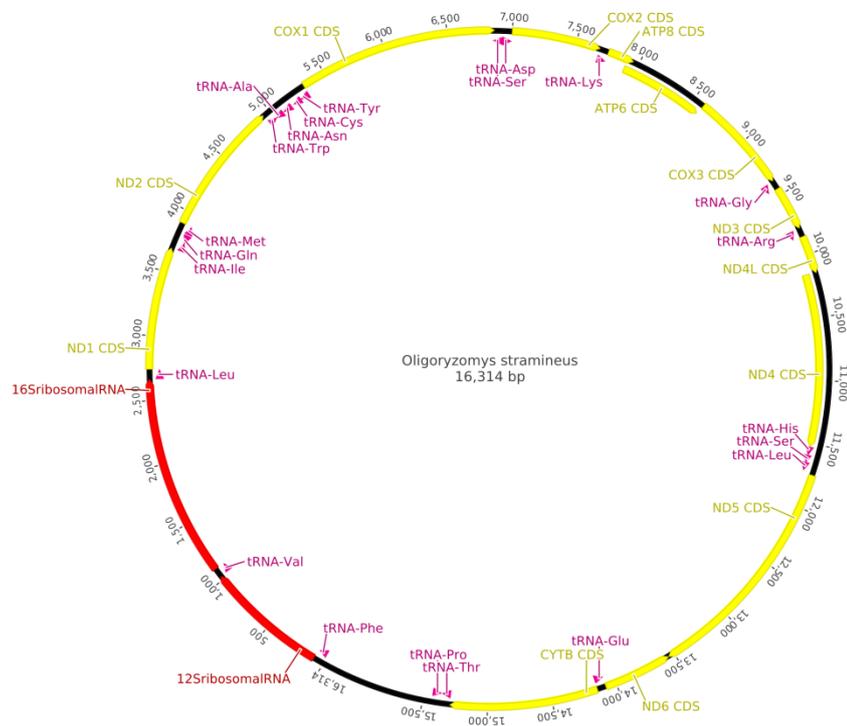


Figura 1: Mapa do genoma mitocondrial da espécie *Oligoryzomys stramineus*. As regiões em vermelho, amarelo e rosa se referem a genes codificadores de rRNA, genes codificadores de proteínas e genes codificadores de tRNA, respectivamente.

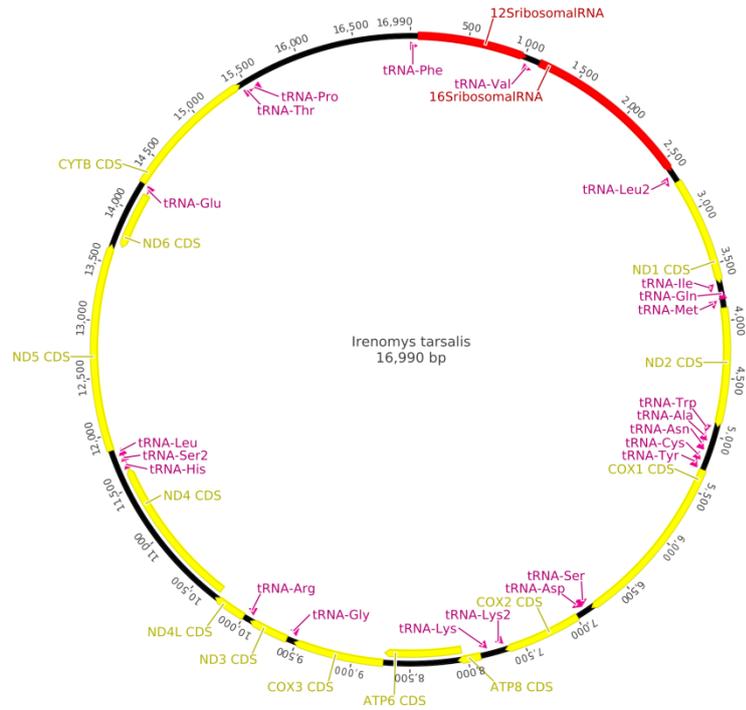


Figura 2: Mapa do genoma mitocondrial da espécie *Irenomys tarsalis*. As regiões em vermelho, amarelo e rosa se referem a genes codificadores de rRNA, genes codificadores de proteínas e genes codificadores de tRNA, respectivamente.

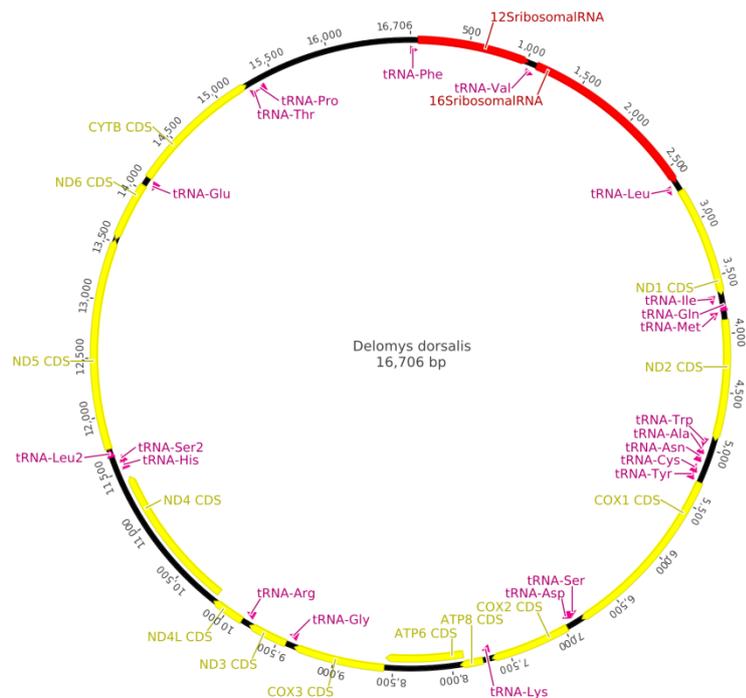


Figura 3: Mapa do genoma mitocondrial da espécie *Delomys dorsalis*. As regiões em vermelho, amarelo e rosa se referem a genes codificadores de rRNA, genes codificadores de proteínas e genes codificadores de tRNA, respectivamente.

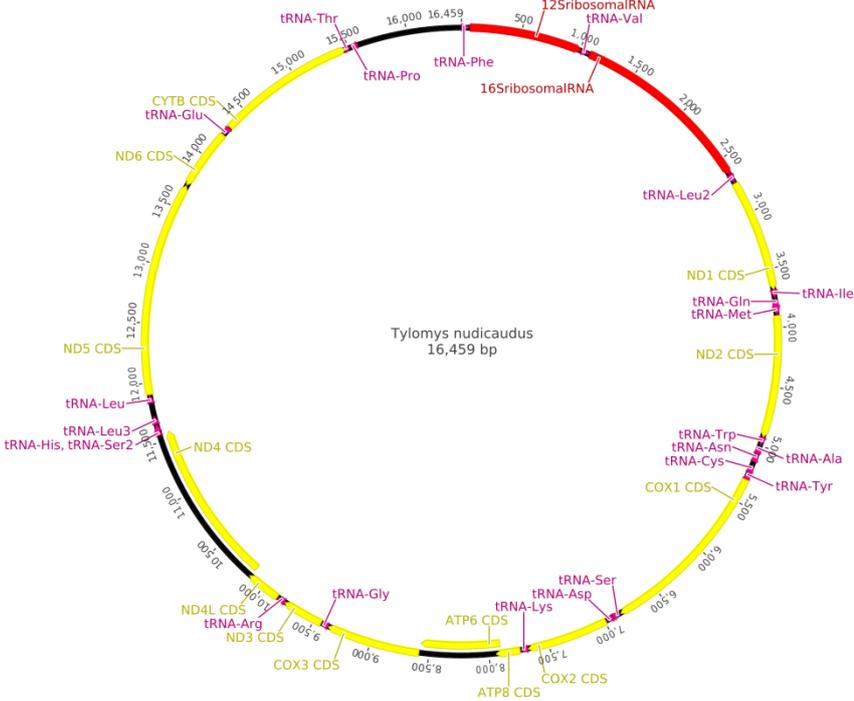


Figura 4: Mapa do genoma mitocondrial da espécie *Tylomys nudicaudus*. As regiões em vermelho, amarelo e rosa se referem a genes codificadores de rRNA, genes codificadores de proteínas e genes codificadores de tRNA, respectivamente.

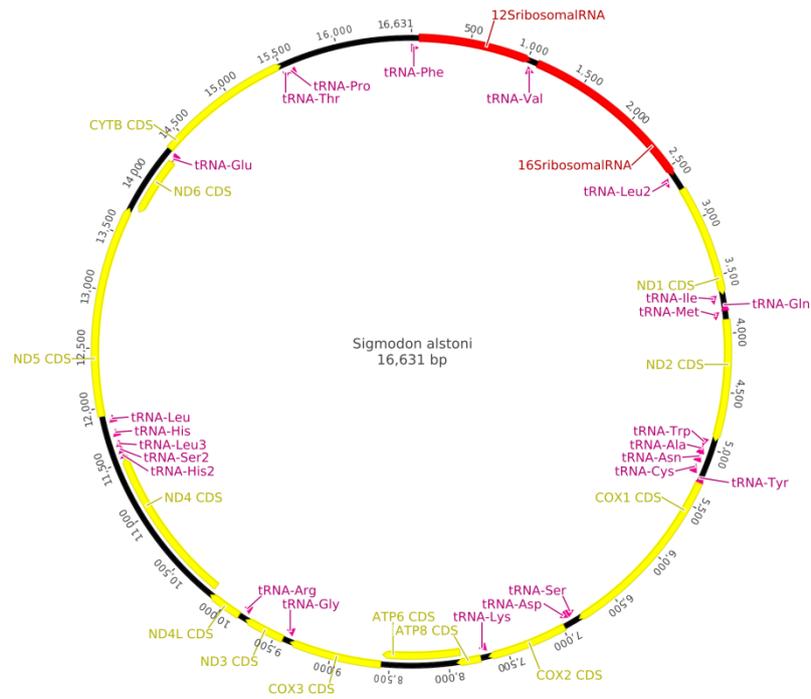


Figura 5: Mapa do genoma mitocondrial da espécie *Sigmodon alstoni*. As regiões em vermelho, amarelo e rosa se referem a genes codificadores de rRNA, genes codificadores de proteínas e genes codificadores de tRNA, respectivamente.

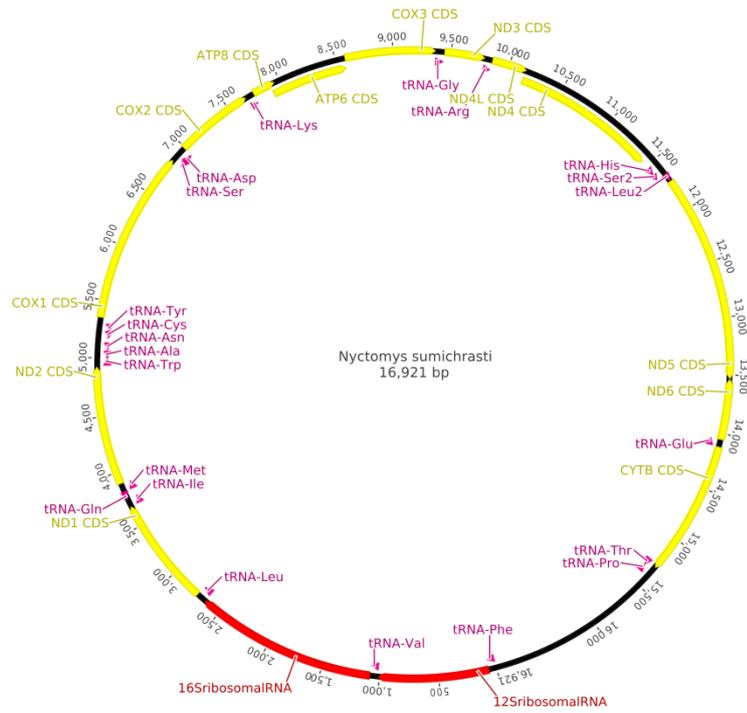


Figura 8: Mapa do genoma mitocondrial da espécie *Nyctomys sumichrasti*. As regiões em vermelho, amarelo e rosa se referem a genes codificadores de rRNA, genes codificadores de proteínas e genes codificadores de tRNA, respectivamente.





Figura 9: Árvores com topologias bifurcadas que representam as 105 hipóteses de relacionamento entre as subfamílias de Cricetidae. O identificador numérico para cada topologia se encontra acima das árvores.

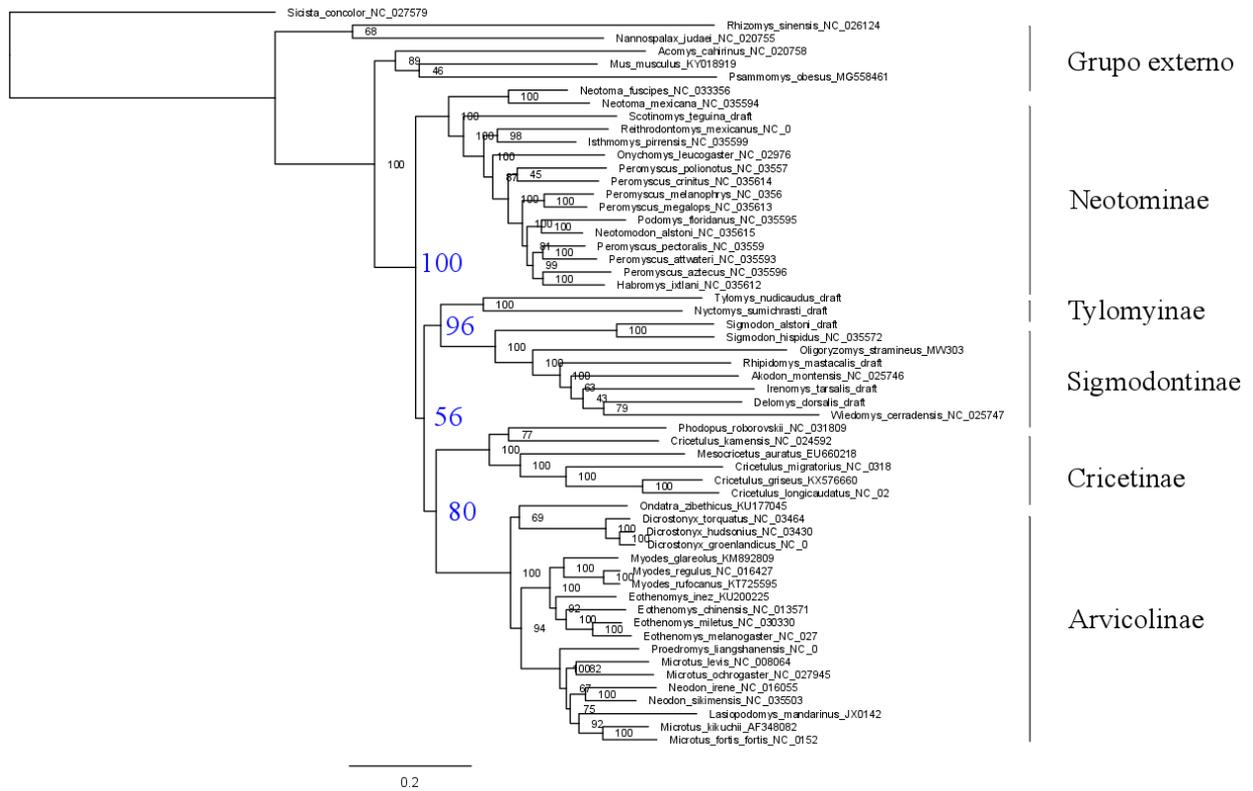


Figura 10: Árvore de máxima verossimilhança baseada no alinhamento de rRNA-CDS-tRNA particionado por gene. Os números se referem ao suporte do bootstrap e os números em azul representam o suporte do bootstrap para os principais clados dentro de Cricetidae (ver Tabela 6).

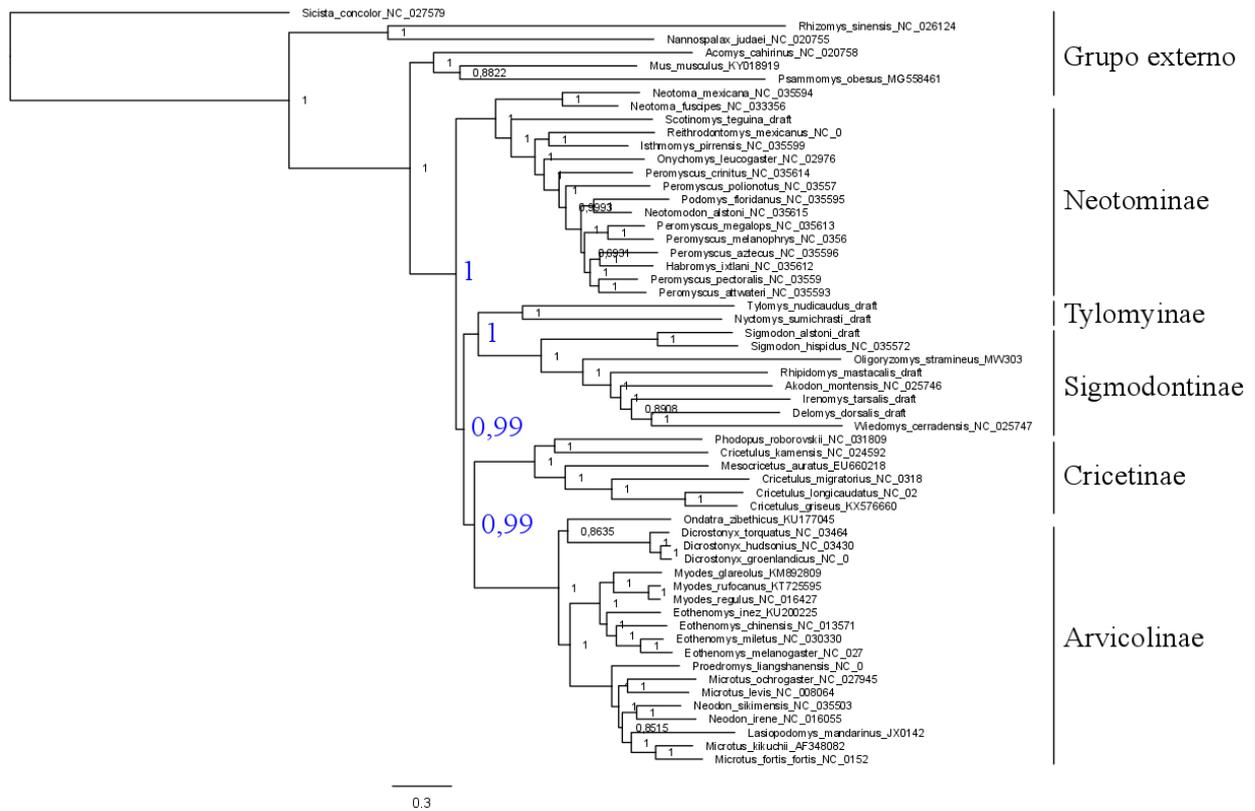


Figura 11: Árvore bayesiana inferida baseando-se no alinhamento de rRNA-CDS-tRNA particionado por gene. Os números se referem à probabilidade posterior e os números em azul representam a probabilidade posterior para os principais clados dentro de Cricetidae (ver Tabela 6).

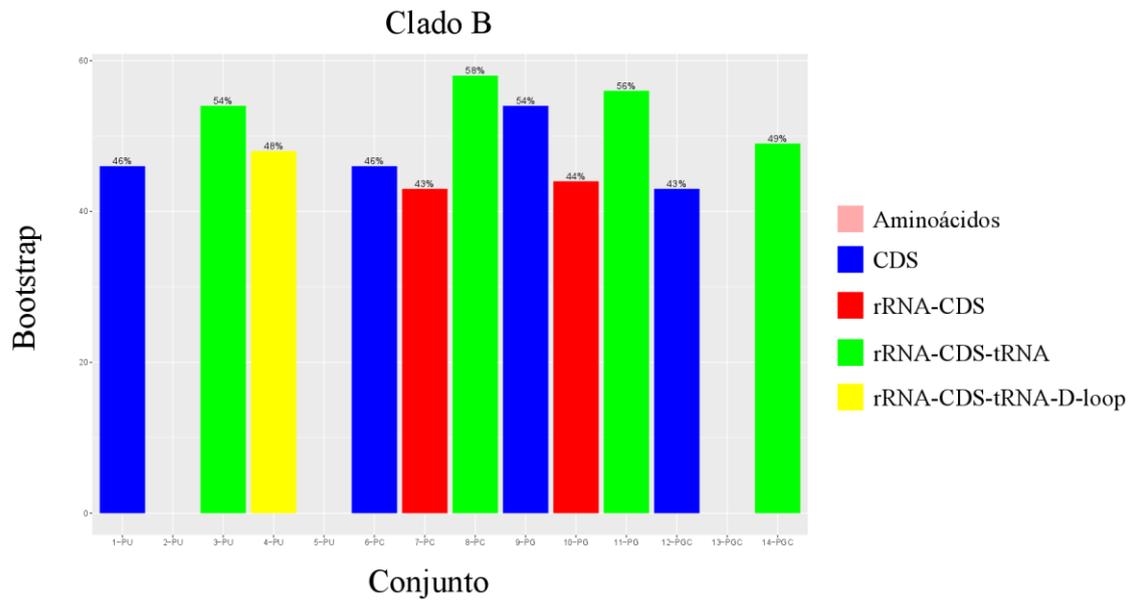


Figura 12: Gráfico para os valores de bootstrap do clado B encontrados nas análises filogenéticas baseadas em uma determinada classe de dados que estão representadas pelas cores rosa, azul, vermelha, verde e amarela na legenda. Os valores do bootstrap estão no eixo Y enquanto que os conjuntos (classe de dados + esquemas de partição) utilizados nas análises estão no eixo X. No eixo X, PU, PC, PG e PGC se referem à partição única, particionamento por posição do códon, particionamento por gene e particionamento por gene e posição do códon, respectivamente.

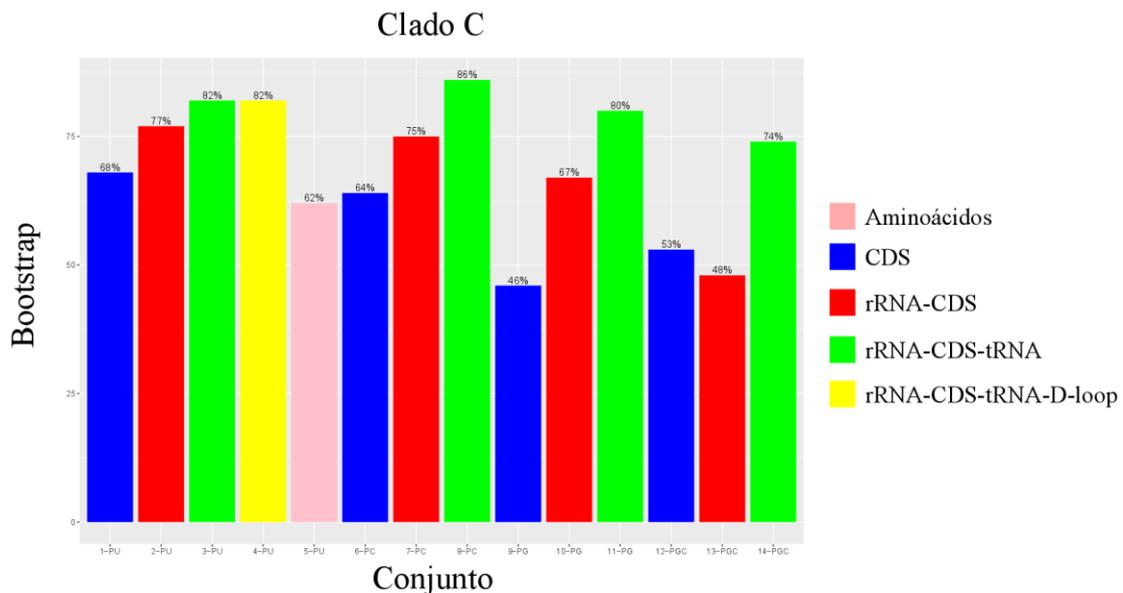


Figura 13: Gráfico para os valores de bootstrap do clado C encontrados nas análises filogenéticas baseadas em uma determinada classe de dados que estão representadas pelas cores rosa, azul, vermelha, verde e amarela na legenda. Os valores do bootstrap estão no eixo Y enquanto que os conjuntos (classe de dados + esquemas de partição) utilizados nas análises

estão no eixo X. No eixo X, PU, PC, PG e PGC se referem à partição única, particionamento por posição do códon, particionamento por gene e particionamento por gene e posição do códon, respectivamente.

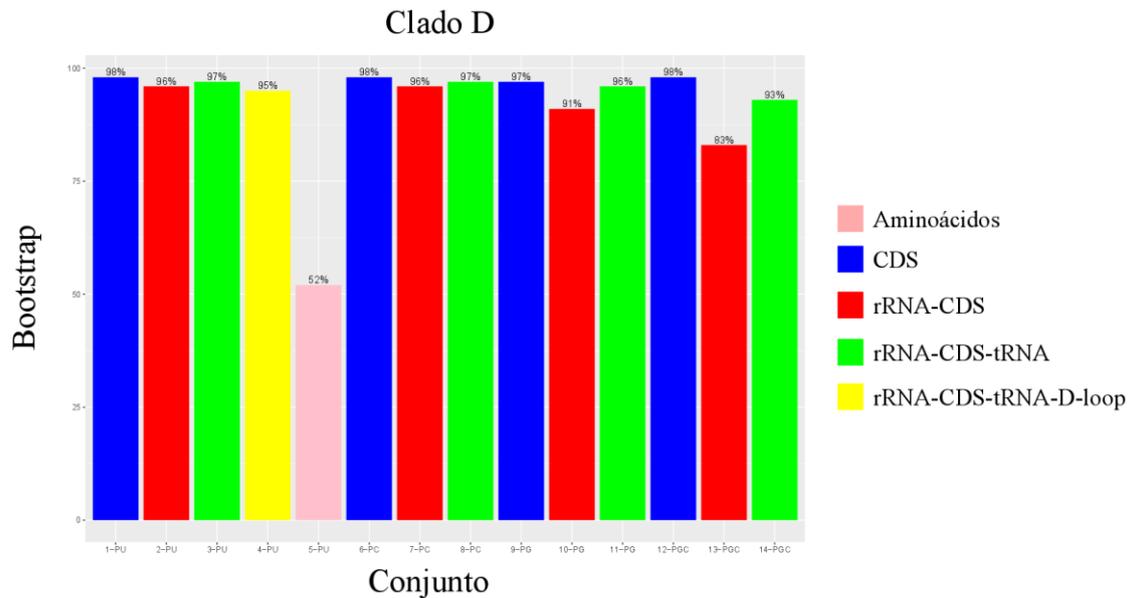


Figura 14: Gráfico para os valores de bootstrap do clado D encontrados nas análises filogenéticas baseadas em uma determinada classe de dados que estão representadas pelas cores rosa, azul, vermelha, verde e amarela na legenda. Os valores do bootstrap estão no eixo Y enquanto que os conjuntos (classe de dados + esquemas de partição) utilizados nas análises estão no eixo X. No eixo X, PU, PC, PG e PGC se referem à partição única, particionamento por posição do códon, particionamento por gene e particionamento por gene e posição do códon, respectivamente.

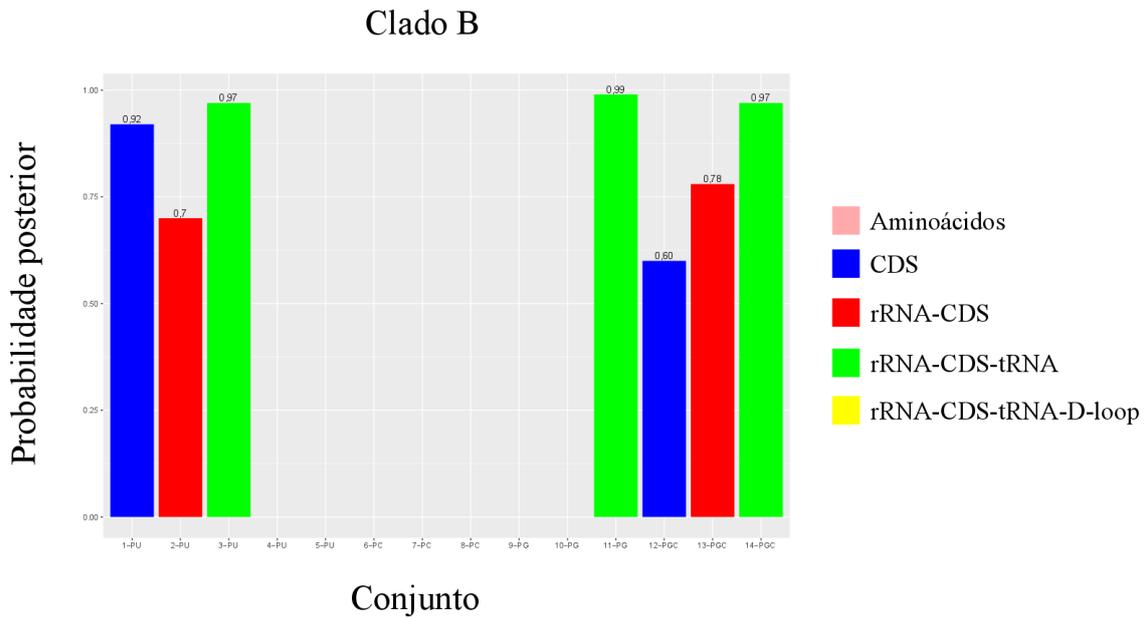


Figura 15: Gráfico para os valores de probabilidade posterior do clado B encontrados nas análises filogenéticas baseadas em uma determinada classe de dados que estão representadas pelas cores rosa, azul, vermelha, verde e amarela na legenda. Os valores do bootstrap estão no eixo Y enquanto que os conjuntos (classe de dados + esquemas de partição) utilizados nas análises estão no eixo X. No eixo X, PU, PC, PG e PGC se referem à partição única, particionamento por posição do códon, particionamento por gene e particionamento por gene e posição do códon, respectivamente.

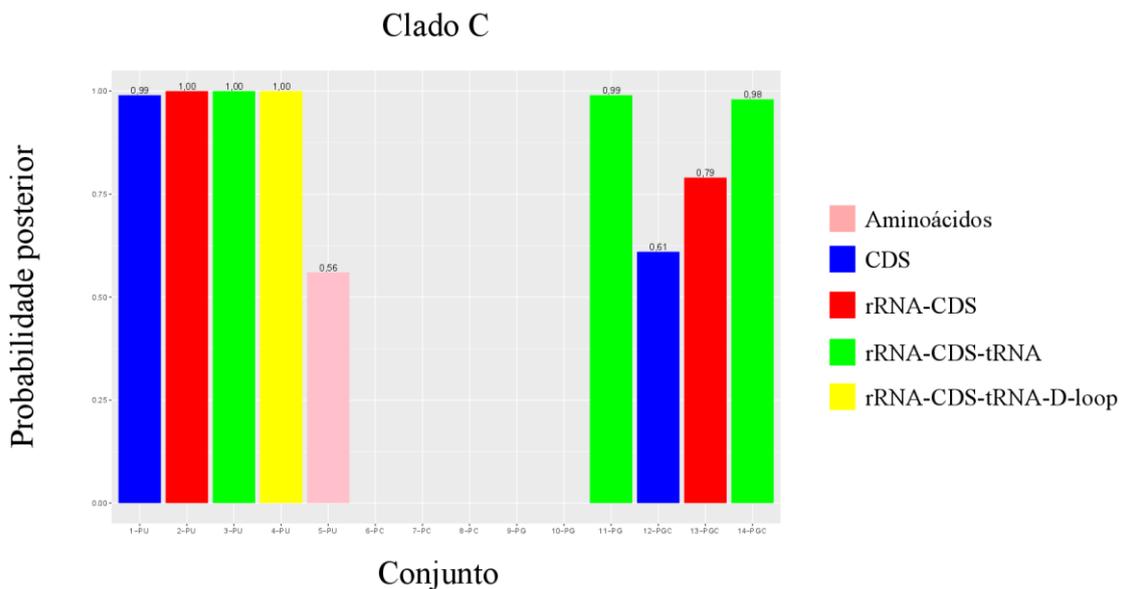


Figura 16: Gráfico para os valores de probabilidade posterior do clado C encontrados nas análises filogenéticas baseadas em uma determinada classe de dados que estão representadas pelas cores rosa, azul, vermelha, verde e amarela na legenda. Os valores do bootstrap estão no eixo Y enquanto que os conjuntos (classe de dados + esquemas de partição) utilizados nas

análises estão no eixo X. No eixo X, PU, PC, PG e PGC se referem à partição única, particionamento por posição do códon, particionamento por gene e particionamento por gene e posição do códon, respectivamente.

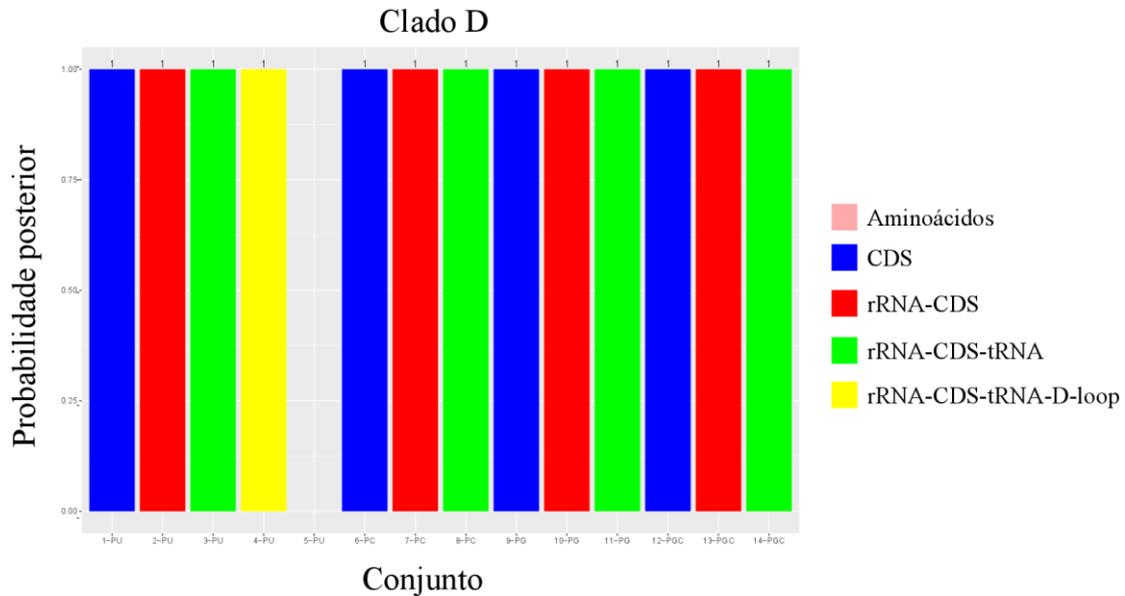


Figura 17: Gráfico para os valores de probabilidade posterior do clado D encontrados nas análises filogenéticas baseadas em uma determinada classe de dados que estão representadas pelas cores rosa, azul, vermelha, verde e amarela na legenda. Os valores do bootstrap estão no eixo Y enquanto que os conjuntos (classe de dados + esquemas de partição) utilizados nas análises estão no eixo X. No eixo X, PU, PC, PG e PGC se referem à partição única, particionamento por posição do códon, particionamento por gene e particionamento por gene e posição do códon, respectivamente.

TABELAS

Tabela 1: Informações sobre os mitogenomas dos espécimes que foram utilizados no presente trabalho.

ID	Família	Subfamília	Gênero	Espécie	Fonte
NC035615.1	Cricetidae	Neotominae	Neotomodon	<i>Neotomodon alstoni</i>	GenBank
NC035593.1	Cricetidae	Neotominae	Peromyscus	<i>Peromyscus attwateri</i>	GenBank
NC035598.1	Cricetidae	Neotominae	Peromyscus	<i>Peromyscus pectoralis</i>	GenBank
NC035612.1	Cricetidae	Neotominae	Habromys	<i>Habromys ixtlani</i>	GenBank
NC035611.1	Cricetidae	Neotominae	Peromyscus	<i>Peromyscus melanophrys</i>	GenBank
NC035613.1	Cricetidae	Neotominae	Peromyscus	<i>Peromyscus megalops</i>	GenBank
NC035596.1	Cricetidae	Neotominae	Peromyscus	<i>Peromyscus aztecus</i>	GenBank
NC035595.1	Cricetidae	Neotominae	Podomys	<i>Podomys floridanus</i>	GenBank
NC035614.1	Cricetidae	Neotominae	Peromyscus	<i>Peromyscus crinitus</i>	GenBank

NC035571.1	Cricetidae	Neotominae	Peromyscus	<i>Peromyscus polionotus</i>	GenBank
NC035599.1	Cricetidae	Neotominae	Isthmomyss	<i>Isthmomyss pirrensis</i>	GenBank
NC029760.1	Cricetidae	Neotominae	Onychomys	<i>Onychomys leucogaster</i>	GenBank
NC035597.1	Cricetidae	Neotominae	Reithrodontomys	<i>Reithrodontomys mexicanus</i>	GenBank
NC035594.1	Cricetidae	Neotominae	Neotoma	<i>Neotoma mexicana</i>	GenBank
NC033356.1	Cricetidae	Neotominae	Neotoma	<i>Neotoma fuscipes</i>	GenBank
MVZ148724	Cricetidae	Neotominae	Scotinomys	<i>Scotinomys teguina</i>	Presente estudo
NC034646.1	Cricetidae	Arvicolinae	Dicrostonyx	<i>Dicrostonyx torquatus</i>	GenBank
NC034307.1	Cricetidae	Arvicolinae	Dicrostonyx	<i>Dicrostonyx hudsonius</i>	GenBank
NC034313.1	Cricetidae	Arvicolinae	Dicrostonyx	<i>Dicrostonyx groenlandicus</i>	GenBank
KU177045.1	Cricetidae	Arvicolinae	Ondatra	<i>Ondatra zibethicus</i>	GenBank
KM892809.1	Cricetidae	Arvicolinae	Myodes	<i>Myodes glareolus</i>	GenBank
KT725595.1	Cricetidae	Arvicolinae	Myodes	<i>Myodes rufocanus</i>	GenBank
NC016427.1	Cricetidae	Arvicolinae	Myodes	<i>Myodes regulus</i>	GenBank

KU200225.1	Cricetidae	Arvicolinae	Eothenomys	<i>Eothenomys inez</i>	GenBank
NC013571.1	Cricetidae	Arvicolinae	Eothenomys	<i>Eothenomys chinensis</i>	GenBank
NC030330.1	Cricetidae	Arvicolinae	Eothenomys	<i>Eothenomys miletus</i>	GenBank
NC027418.1	Cricetidae	Arvicolinae	Eothenomys	<i>Eothenomys melanogaster</i>	GenBank
JX014233.1	Cricetidae	Arvicolinae	Lasiopodomys	<i>Lasiopodomys mandarinus</i>	GenBank
NC013563.1	Cricetidae	Arvicolinae	Proedromys	<i>Proedromys liangshanensis</i>	GenBank
NC027945.1	Cricetidae	Arvicolinae	Microtus	<i>Microtus ochrogaster</i>	GenBank
NC008064.1	Cricetidae	Arvicolinae	Microtus	<i>Microtus levis</i>	GenBank
NC015241.1	Cricetidae	Arvicolinae	Microtus	<i>Microtus fortis</i>	GenBank
AF348082.1	Cricetidae	Arvicolinae	Microtus	<i>Microtus kikuchii</i>	GenBank
NC035503.1	Cricetidae	Arvicolinae	Neodon	<i>Neodon sikimensis</i>	GenBank
NC016055.1	Cricetidae	Arvicolinae	Neodon	<i>Neodon irene</i>	GenBank
NC024592.1	Cricetidae	Cricetinae	Cricetulus	<i>Cricetulus kamensis</i>	GenBank
NC031809.1	Cricetidae	Cricetinae	Phodopus	<i>Phodopus roborovskii</i>	GenBank

EU660218.1	Cricetidae	Cricetinae	Mesocricetus	<i>Mesocricetus auratus</i>	GenBank
NC031802.1	Cricetidae	Cricetinae	Cricetulus	<i>Cricetulus migratorius</i>	GenBank
KX576660.1	Cricetidae	Cricetinae	Cricetulus	<i>Cricetulus griseus</i>	GenBank
NC025330.1	Cricetidae	Cricetinae	Cricetulus	<i>Cricetulus longicaudatus</i>	GenBank
F35709	Cricetidae	Tylomyinae	Nyctomys	<i>Nyctomys sumichrasti</i>	Presente estudo
F35721	Cricetidae	Tylomyinae	Tylomys	<i>Tylomys nudicaudus</i>	Presente estudo
NC025747.1	Cricetidae	Sigmodontinae	Wiedomys	<i>Wiedomys cerradensis</i>	GenBank
NC025746.1	Cricetidae	Sigmodontinae	Akodon	<i>Akodon montensis</i>	GenBank
MNLM4489	Cricetidae	Sigmodontinae	Delomys	<i>Delomys dorsalis</i>	Presente estudo
MNLM4468	Cricetidae	Sigmodontinae	Rhipidomys	<i>Rhipidomys mastacalis</i>	Presente estudo
MVZ184955	Cricetidae	Sigmodontinae	Irenomys	<i>Irenomys tarsalis</i>	Presente estudo
MN84348	Cricetidae	Sigmodontinae	Oligoryzomys	<i>Oligoryzomys stramineus</i>	Presente estudo
NC035572.1	Cricetidae	Sigmodontinae	Sigmodon	<i>Sigmodon hispidus</i>	GenBank
RSV1210	Cricetidae	Sigmodontinae	Sigmodon	<i>Sigmodon alstoni</i>	Presente estudo

NC020758.1	Muridae	Deomyinae	Acomys	<i>Acomys cahirinus</i>	Presente estudo
MG558461.1	Muridae	Gerbillinae	Psammomys	<i>Psammomys obesus</i>	Presente estudo
NC020755.1	Spalacidae	Spalacinae	Spalax	<i>Spalax judaei</i>	Presente estudo
NC026124.1	Spalacidae	Spalacinae	Rhizomys	<i>Rhizomys sinensis</i>	Presente estudo
NC027579.1	Dipodoidea	Sicistinae	Sicista	<i>Sicista concolor</i>	Presente estudo
KY018919.1	Muridae	Murinae	Mus	<i>Mus musculus</i>	Presente estudo

Tabela 2: Características de cada conjunto em que as análises filogenéticas foram baseadas com os seus respectivos modelos de substituição. Número* se refere à quantidade de vezes em que um modelo de substituição foi sugerido pelo PartitionFinder2 para ser utilizado nas partições de um determinado esquema de partição.

Conjunto	Alinhamento	Número de sítios	Particionamento	Número de partições	Modelo	Número de parâmetros
1	CDS	11,359	Partição única	1	1*(GTR + G + I)	121
2	rRNA-CDS	14,777	Partição única	1	1*(GTR + G + I)	121
3	rRNA-CDS-tRNA	14,777	Partição única	1	1*(GTR + G + I)	121
4	rRNA-CDS-tRNA-D-loop	16,785	Partição única	1	1*(GTR + G + I)	121
5	Aminoacids	3,677	Partição única	1	1*(mtMAM)	131
6	CDS	11,359	Por posição do códon	3	3*(GTR + G + I)	143
7	rRNA-CDS	14,777	Por posição do códon rRNA	4	4*(GTR + G + I)	154
8	rRNA-CDS-tRNA	16,113	Por posição do códon rRNA tRNA	5	5*(GTR + G + I)	165
9	CDS	11,359	Por gene	5	5*(GTR + G + I)	165
10	rRNA-CDS	14,777	Por gene	6	6*(GTR + G + I)	176
11	rRNA-CDS-tRNA	16,113	Por gene	7	7*(GTR + G + I)	187

12	CDS	11,359	Por gene e posição do códon	14	11*(GTR + G + I) 1*(SYM + G + I) 1*(HKY + G + I) 1*(HKY + G)	252
13	rRNA-CDS	14,777	Por gene e posição do códon	15	12*(GTR + G + I) 1*(SYM + G + I) 1*(HKY + G + I) 1*(HKY + G)	263
14	rRNA-CDS-tRNA	16,113	Por gene e posição do códon	16	13*(GTR + G + I) 1*(SYM + G + I) 1*(HKY + G + I) 1*(HKY + G)	274

Tabela 3: Valores de verossimilhança, AIC e delta para cada uma das 105 topologias bifurcadas que representam o relacionamento entre as subfamílias de Cricetidae.

Topologia	Legenda	Verossimilhança	AIC	Delta	Referência
32	xbf	-345646,9007	691537,8015		Presente trabalho
29	xbc	-345651,6662	691547,3324	9,530918	Presente trabalho
33	xbg	-345667,5485	691579,097	41,29555	Steppan et al, 2017
27	xba	-345669,6914	691583,3828	45,58127	
26	xaz	-345670,1348	691584,2696	46,4681	Presente trabalho
34	xbh	-345674,6808	691593,3617	55,56018	
22	xav	-345676,4275	691596,8551	59,05358	
99	xdu	-345677,1913	691598,3826	60,58115	
4	xad	-345677,5184	691599,0367	61,23523	

64	xcl	-345679,3762	691602,7524	64,95091	Presente trabalho
5	xae	-345683,6198	691611,2396	73,43809	
35	xbi	-345684,5494	691613,0987	75,29724	
28	xbb	-345685,4682	691614,9365	77,13496	
25	xay	-345687,5846	691619,1691	81,36762	
65	xcm	-345687,6427	691619,2853	81,48382	
102	xdx	-345688,5172	691621,0344	83,23291	
101	xdw	-345688,6957	691621,3914	83,5899	
7	xag	-345689,5108	691623,0215	85,22001	
6	xaf	-345691,2749	691626,5498	88,74832	
103	xdy	-345691,5443	691627,0886	89,28707	
1	xaa	-345692,2961	691628,5923	90,79076	
66	xcn	-345693,8158	691631,6315	93,83004	
68	xcp	-345694,009	691632,0181	94,21657	
37	xbk	-345698,751	691641,502	103,7005	
51	xby	-345699,3286	691642,6572	104,8557	
104	xdz	-345699,3443	691642,6885	104,8871	
48	xbv	-345700,8497	691645,6993	107,8978	
45	xbs	-345701,1109	691646,2217	108,4202	
100	xdv	-345702,0156	691648,0311	110,2296	

31	xbe	-345702,016	691648,0319	110,2304
69	xcq	-345702,8392	691649,6784	111,8769
67	xco	-345703,2071	691650,4142	112,6127
44	xbr	-345704,3689	691652,7378	114,9363
36	xbj	-345705,4392	691654,8784	117,0769
38	xbl	-345706,3918	691656,7836	118,9821
90	xdl	-345706,5513	691657,1025	119,3011
43	xbq	-345706,7866	691657,5732	119,7717
73	xcu	-345707,1276	691658,2552	120,4537
80	xdb	-345709,2728	691662,5456	124,7441
47	xbu	-345709,4721	691662,9442	125,1427
71	xcs	-345709,8266	691663,6532	125,8517
105	xea	-345709,9281	691663,8562	126,0547
30	xbd	-345709,945	691663,8901	126,0886
94	xdp	-345710,0736	691664,1472	126,3457
96	xdr	-345710,7287	691665,4574	127,6559
41	xbo	-345711,8496	691667,6992	129,8977
52	xbz	-345711,8923	691667,7845	129,983
58	xcf	-345711,9527	691667,9054	130,1039
61	xcj	-345712,2824	691668,5649	130,7634

85	xdg	-345712,6018	691669,2037	131,4022
70	xcr	-345713,1173	691670,2347	132,4332
95	xdq	-345713,2072	691670,4145	132,613
89	xdk	-345713,302	691670,604	132,8025
55	xcc	-345715,0263	691674,0525	136,251
49	xbw	-345715,0713	691674,1426	136,3411
11	xak	-345715,5166	691675,0331	137,2317
18	xar	-345715,87	691675,74	137,9385
82	xdd	-345716,3536	691676,7072	138,9057
75	xcw	-345716,3764	691676,7528	138,9513
42	xbp	-345716,7113	691677,4225	139,621
12	xal	-345716,7311	691677,4621	139,6606
54	xcb	-345717,2826	691678,5651	140,7636
50	xbx	-345717,7639	691679,5278	141,7263
19	xas	-345718,3583	691680,7166	142,9151
60	xch	-345718,6013	691681,2025	143,401
40	xbn	-345719,0419	691682,0837	144,2822
59	xcg	-345719,3702	691682,7404	144,9389
91	xdm	-345719,619	691683,2379	145,4364
72	xct	-345721,4387	691686,8774	149,0759

56	xcd	-345721,8016	691687,6032	149,8017
86	xdh	-345721,9184	691687,8368	150,0353
77	xcy	-345722,3768	691688,7536	150,9521
10	xaj	-345722,6145	691689,2291	151,4276
24	xax	-345723,4834	691690,9668	153,1653
46	xbt	-345724,1102	691692,2204	154,4189
97	xds	-345724,2158	691692,4315	154,63
93	xdo	-345724,389	691692,7781	154,9766
88	xdj	-345724,914	691693,828	156,0265
74	xcv	-345725,5628	691695,1256	157,3242
87	xdi	-345725,8193	691695,6387	157,8372
20	xat	-345726,3398	691696,6795	158,878
76	xcx	-345726,4204	691696,8407	159,0392
78	xcz	-345726,6678	691697,3355	159,534
92	xdn	-345727,8503	691699,7006	161,8991
81	xdc	-345728,2999	691700,5999	162,7984
79	xda	-345728,4137	691700,8273	163,0258
3	xac	-345728,4304	691700,8609	163,0594
17	xaq	-345728,6822	691701,3645	163,563
9	xai	-345728,7569	691701,5137	163,7122

15	xao	-345729,0203	691702,0406	164,2391
84	xdf	-345729,1334	691702,2668	164,4653
23	xaw	-345730,5291	691705,0582	167,2567
53	xca	-345732,0764	691708,1527	170,3512
39	xbm	-345732,099	691708,1979	170,3964
16	xap	-345732,2376	691708,4752	170,6737
57	xce	-345732,8212	691709,6425	171,841
8	xah	-345733,4357	691710,8714	173,0699
98	xdt	-345733,8043	691711,6086	173,8071
83	xde	-345734,575	691713,15	175,3485
21	xau	-345734,6851	691713,3702	175,5687
14	xan	-345735,8627	691715,7253	177,9238
2	xab	-345735,9492	691715,8984	178,0969
63	xck	-345736,6189	691717,2378	179,4363
13	xam	-345739,4728	691722,9456	185,1441
62	xcj	-345739,579	691723,1581	185,3566

Tabela 4: resultados do processo de montagem dos mitogenomas.

Amostra	Reads (total)	Reads (montagem)	Profundidade média	Conteúdo GC	Tamanho dos mitogenomas
<i>Oligoryzomys stramineus</i>	25,328,022	99,796	509x	32,9%	16,314 bp
<i>Delomys dorsalis</i>	20,977,704	78,194	449x	37,9%	16,706 bp
<i>Rhipidomys mastacalis</i>	15,520,970	101,901	592x	37,7%	16,824 bp
<i>Scotinomys teguina</i>	14,039,650	27,542	142x	36,5%	16,472 bp
<i>Sigmodon alstoni</i>	12,500,852	22,146	119x	41,2%	16,565 bp
<i>Irenomys tarsalis</i>	12,518,672	27,419	141x	42,6%	16,924 bp
<i>Tylomys nudicaudus</i>	18,671,406	44,781	224x	41,1%	16,459 bp
<i>Nyctomys sumichrasti</i>	14,754,414	219,506	1,254x	37,0%	16,921 bp

Tabela 5: Informações sobre o genoma mitocondrial das espécies cujo os genomas foram sequenciados no presente trabalho. Os números representam o comprimento em pares de bases (bp) dos genes mitocondriais de cada espécie. Entre parênteses estão os códons de início/códon de terminação dos CDS. Taa representa códon de terminação incompleto em que a cauda poli A é adicionada posteriormente durante o processamento do mRNA.

	<i>O. stramineus</i>	<i>T. nudicaudus</i>	<i>S. alstoni</i>	<i>S. teguina</i>	<i>R. mastacalis</i>	<i>N. sumichrasti</i>	<i>I. tarsalis</i>	<i>D. dorsalis</i>
tRNA-Phe	70	69	67	68	65	66	66	66
12S rRNA	956	954	954	957	958	953	955	953
tRNA-Val	67	66	67	67	68	70	68	67
16S rRNA	1569	1563	1555	1558	1578	1564	1551	1556
tRNA-Leu	74	75	75	75	75	75	75	75
ND1	955 (GTG/Taa)	955 (GTG/Taa)	955 (GTG/Taa)	955 (GTG/Taa)	955 (GTG/Taa)	955 (GTG/Taa)	955 (GTG/Taa)	955 (GTG/Taa)
tRNA-Ile	68	69	69	69	69	67	68	68
tRNA-Gln	70	73	74	72	78	71	78	74
tRNA-Met	69	69	69	69	69	69	68	68
ND2	1036 (ATT/Taa)	1038 (ATA/TAA)	1036 (ATT/Taa)	1038 (ATC/TAG)	1036 (ATA/Taa)	1038 (ATA/TAA)	1036 (ATT/Taa)	1036 (ATT/Taa)
tRNA-Trp	68	67	66	67	68	67	67	70

tRNA-Ala	69	69	68	69	69	69	68	68
tRNA-Asn	70	70	71	71	69	70	69	71
tRNA-Cys	66	67	67	66	67	67	66	67
tRNA-Tyr	67	66	67	66	67	65	68	67
COX1	1545 (ATG/TAA)							
tRNA-Ser	69	69	69	69	69	69	69	69
tRNA-Asp	68	68	69	68	70	69	68	68
COX2	684 (ATG/TAA)							
tRNA-Lys	66	64	65	64	69	65	212	68
ATP8	204 (ATG/TAA)							
ATP6	681 (ATG/TAA)							
COX3	784 (ATG/Taa)							
tRNA-Gly	68	68	68	70	69	69	69	70
ND3	348 (ATA/TAA)	348 (ATT/TAG)	348 (ATA/TAG)	348 (ATA/TAA)	348 (ATA/TAA)	348 (ATT/TAA)	348 (ATA/TAG)	348 (ATA/TAA)
tRNA-Arg	66	66	69	71	68	65	67	67
ND4L	297 (ATG/TAA)	297 (ATT/TAA)	297 (ATG/TAA)	297 (ATG/TAA)				
ND4	1375 (ATG/Taa)	1378 (ATG/Taa)	1378 (ATG/Taa)	1378 (ATG/Taa)	1375 (ATG/Taa)	1376 (ATG/Taa)	1375 (ATG/Taa)	1375 (ATG/Taa)
tRNA-His	70	68	286	67	69	65	68	68
tRNA-Ser	59	243	59	61	59	60	59	60

tRNA-Leu	70	70	70	70	71	67	70	70
ND5	1827 (ATA/TAA)	1812 (ATC/TAA)	1812 (ATT/TAG)	1812 (ATT/TAA)	1809 (ATC/TAA)	1812 (ATT/TAG)	1812 (ATT/TAG)	1812 (ATC/TAA)
ND6	527 (ATG/Taa)	524 (ATG/Taa)	523 (ATG/Taa)	524 (ATG/Taa)	524 (ATG/Taa)	523 (ATG/Taa)	524 (ATG/Taa)	524 (ATG/Taa)
tRNA-Glu	69	69	69	69	69	76	69	69
CYTb	1144 (ATG/Taa)	1144 (ATG/Taa)	1144 (ATG/Taa)	1144 (ATG/Taa)	1143 (ATG/TAA)	1141 (ATG/Taa)	1144 (ATG/Taa)	1144 (ATG/Taa)
tRNA-Thr	67	67	66	67	118	67	111	69
tRNA-Pro	65	65	66	67	70	58	67	65
Comprimento total	16314	16459	16631	16539	16824	16921	16990	16706

Tabela 6: Suporte do bootstrap e probabilidade posterior para os principais clados da família Cricetidae.

Conjunto	Alinhamento	Topologia (MV)	Clado	Clado	Clado	Clado	Clado	Clado	Topologia (IB)	Clado	Clado	Clado	Clado	Clado	Clado
			B	C	D	E	F	G		B	C	D	E	F	G
1	CDS	32	46%	68%	98%	*	*	*	32	0,92	0,99	1,00	*	*	*
2	rRNA-CDS	29	*	77%	96%	45%	*	*	32	0,7	1,00	1,00	*	*	*

3	rRNA-CDS-tRNA	32	54%	82%	97%	*	*	*	32	0,97	1,00	1,00	*	*	*
4	rRNA-CDS-tRNA-D-loop	32	48%	82%	95%	*	*	*	29	*	1,00	1,00	0,56	*	*
5	Aminoacids	29	*	62%	52%	43%	*	*	64	*	0,56	*	*	*	*
6	CDS	32	46%	64%	98%	*	*	*	26	*	*	1,00	*	0,54	0,54
7	rRNA-CDS	32	43%	75%	96%	*	*	*	26	*	*	1,00	*	0,54	0,55
8	rRNA-CDS-tRNA	32	58%	86%	97%	*	*	*	26	*	*	1,00	*	0,66	0,79
9	CDS	32	54%	46%	97%	*	*	*	26	*	*	1,00	*	0,94	0,94
10	rRNA-CDS	32	44%	67%	91%	*	*	*	26	*	*	1,00	*	0,51	0,51
11	rRNA-CDS-tRNA	32	56%	80%	96%	*	*	*	32	0,99	0,99	1,00	*	*	*
12	CDS	32	43%	53%	98%	*	*	*	32	0,60	0,61	1,00	*	*	*
13	rRNA-CDS	29	*	48%	83%	42%	*	*	32	0,78	0,79	1,00	*	*	*
14	rRNA-CDS-tRNA	32	49%	74%	93%	*	*	*	32	0,97	0,98	1,00	*	0,93	0,93

APÊNDICES

Apêndice 1 – Script na linguagem bash para realizar análises de verossimilhança no RAxML com uma das 105 topologias subfamiliares fixadas a cada iteração do loop.

```
#!/bin/bash
```

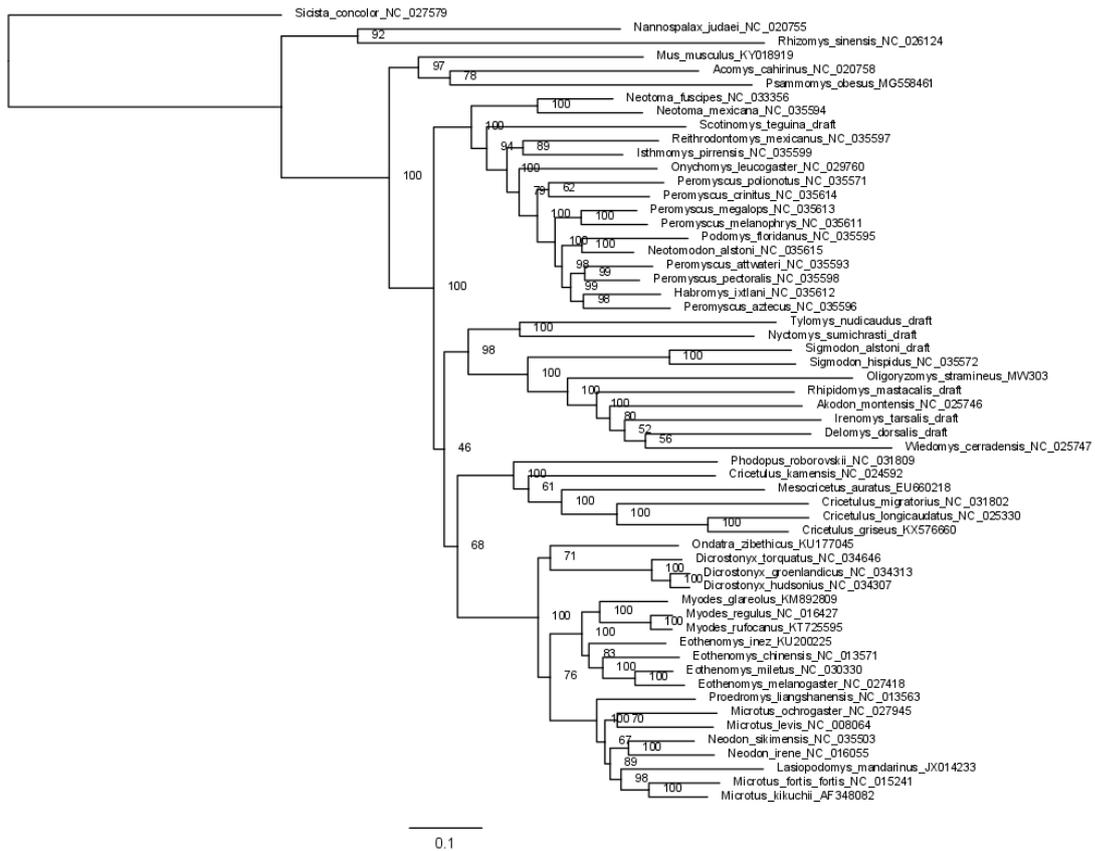
```
# A letra x era a inicial dos nomes de todos os 105 arquivos que continham uma das 105 topologias subfamiliares.
```

```
for i in $(ls x*); do
```

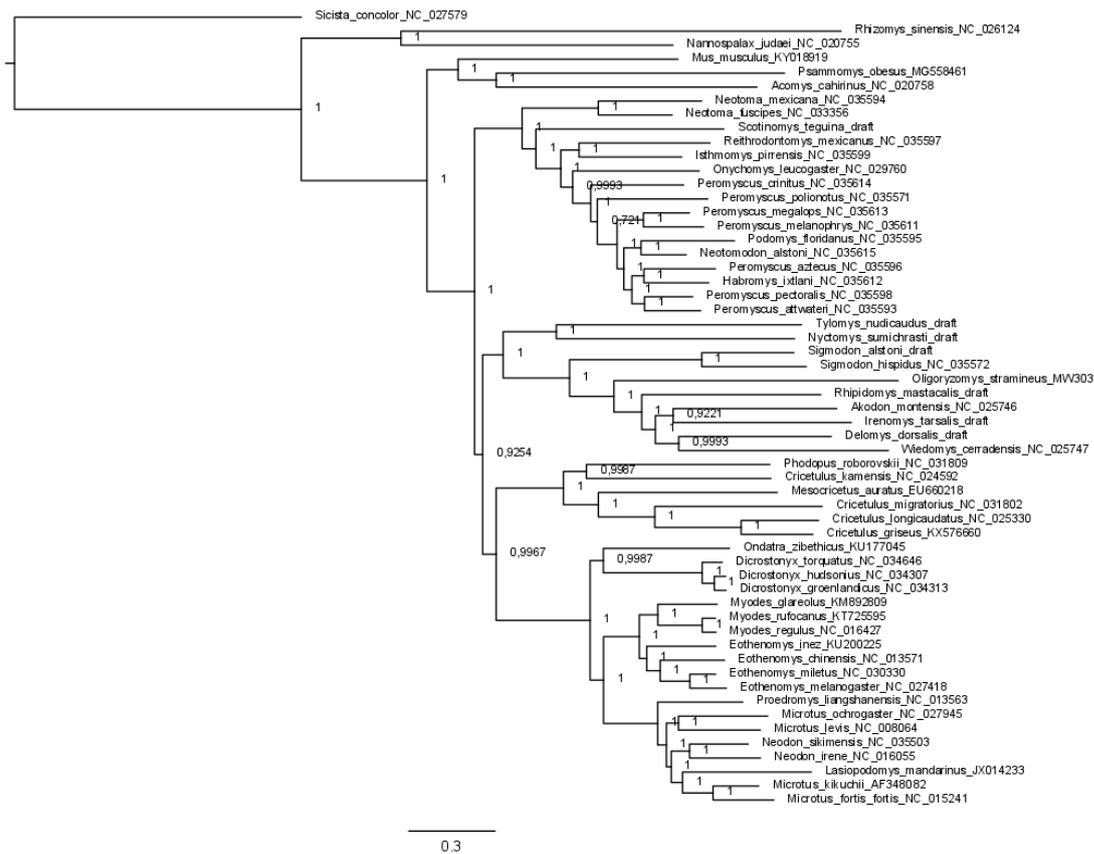
```
raxmlHPC -m GTRGAMMA -n out_$(i) -s arquivo_de_alinhamento.fasta -g $(i) -p 31321
```

```
done
```

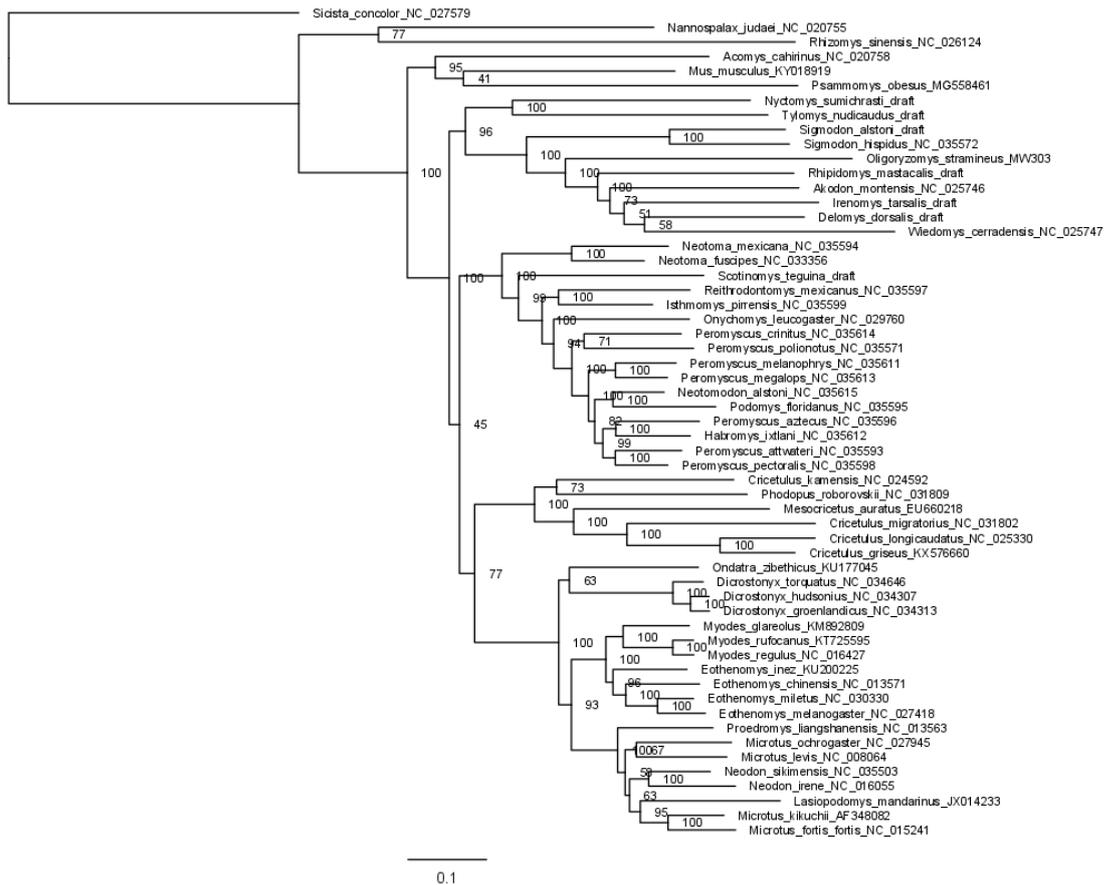
Apêndice 2 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 1.



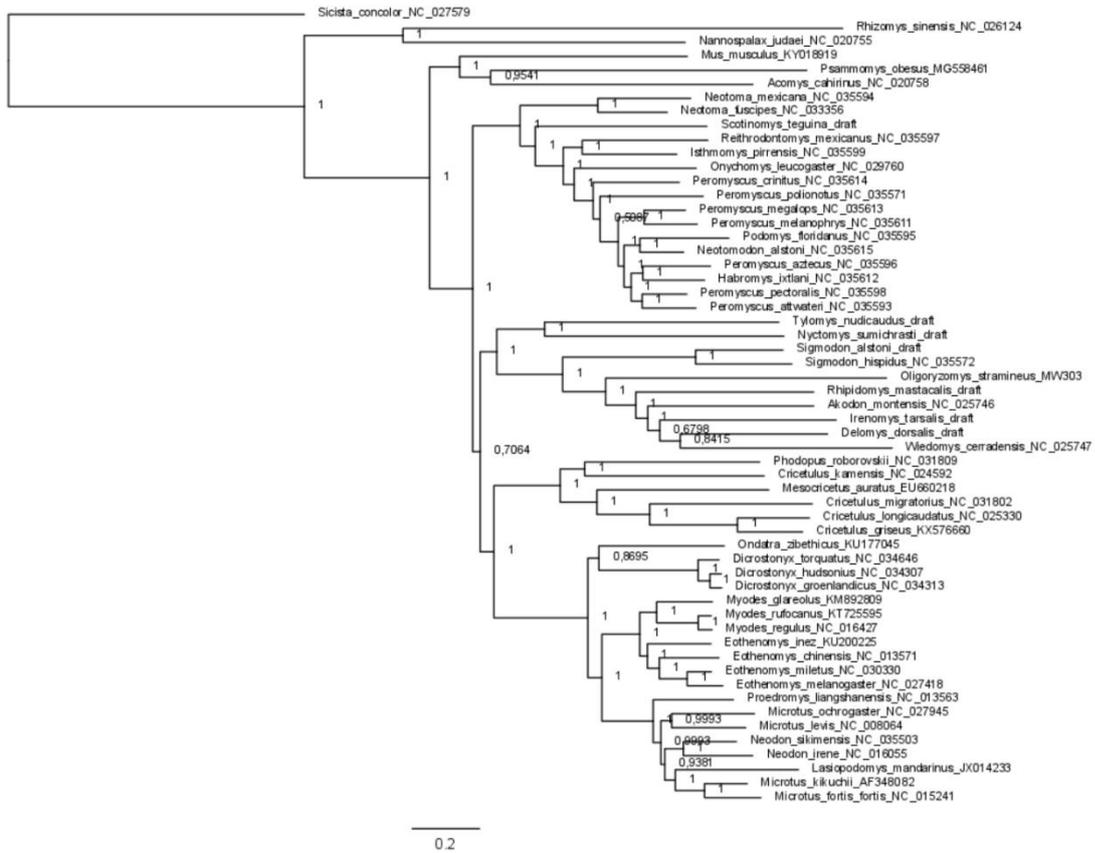
Apêndice 3 – Árvore bayesiana inferida baseando-se no conjunto 1.



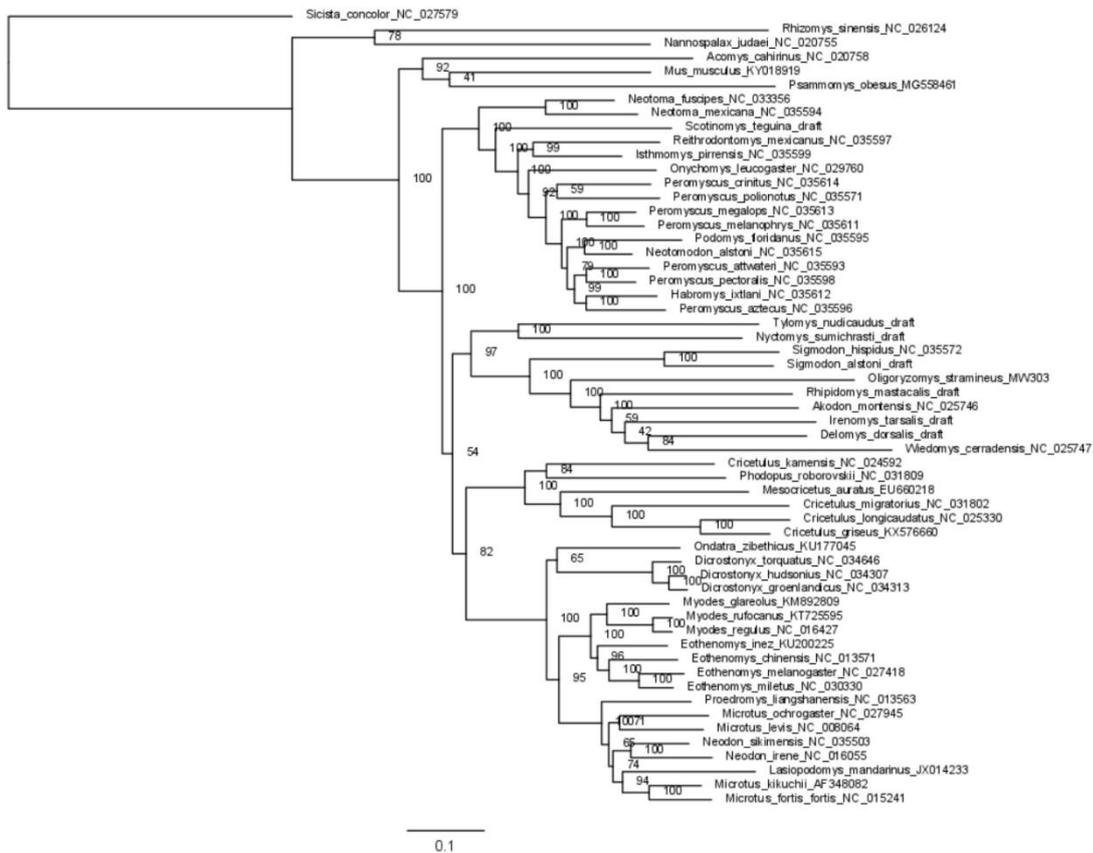
Apêndice 4 – Árvore de máxima verossimilhança inferida baseando-se no conjunto 2.



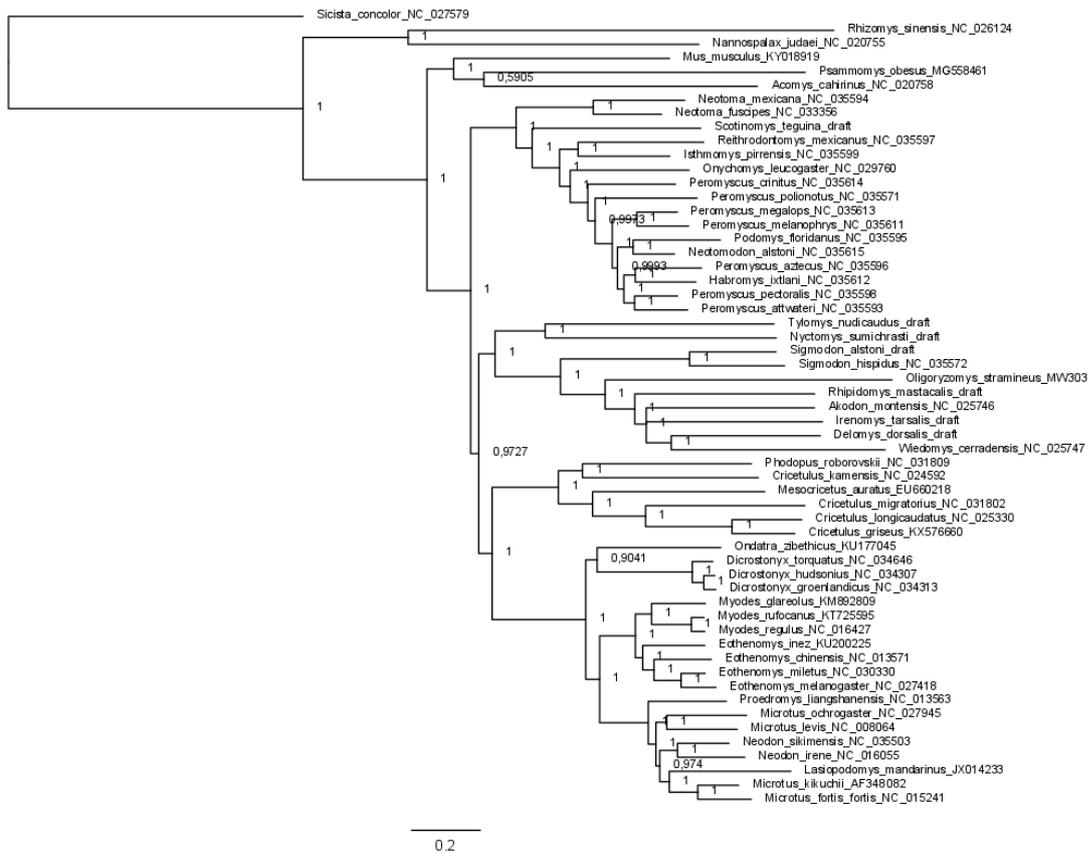
Apêndice 5 - Árvore bayesiana inferida baseando-se no conjunto 2.



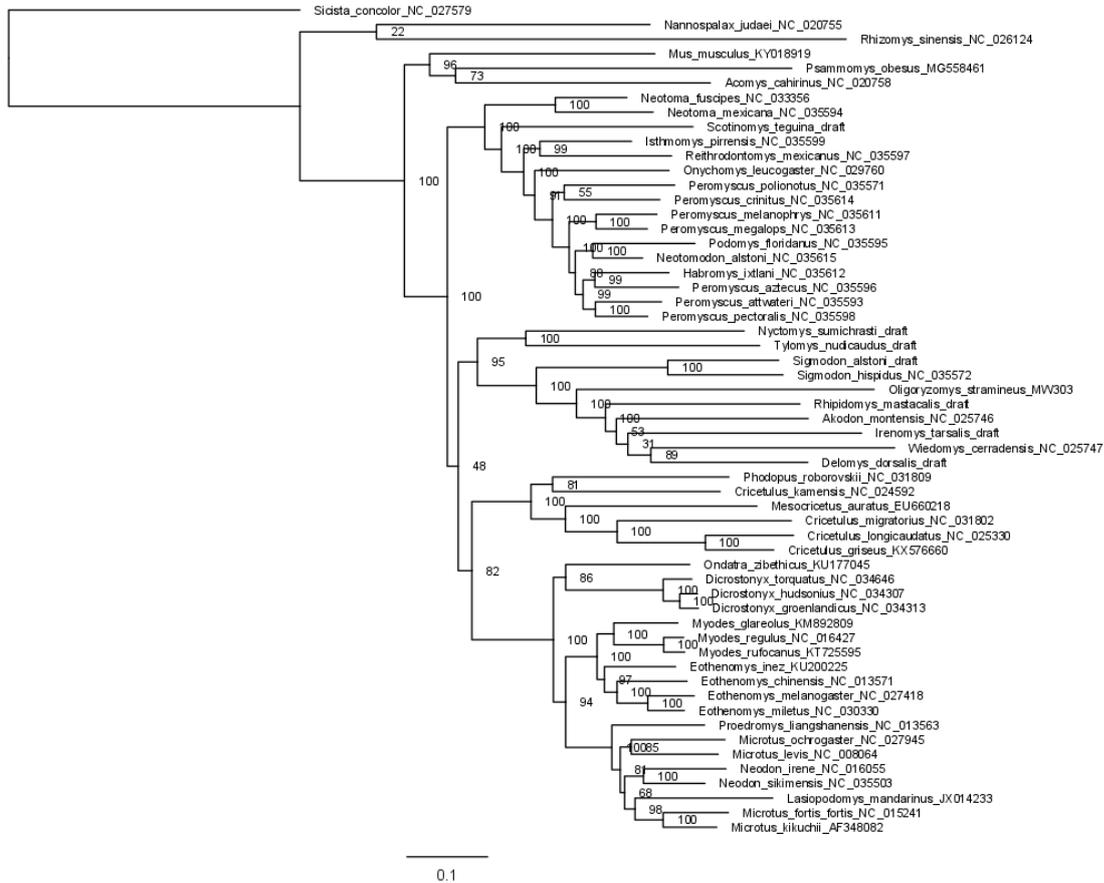
Apêndice 6 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 3.



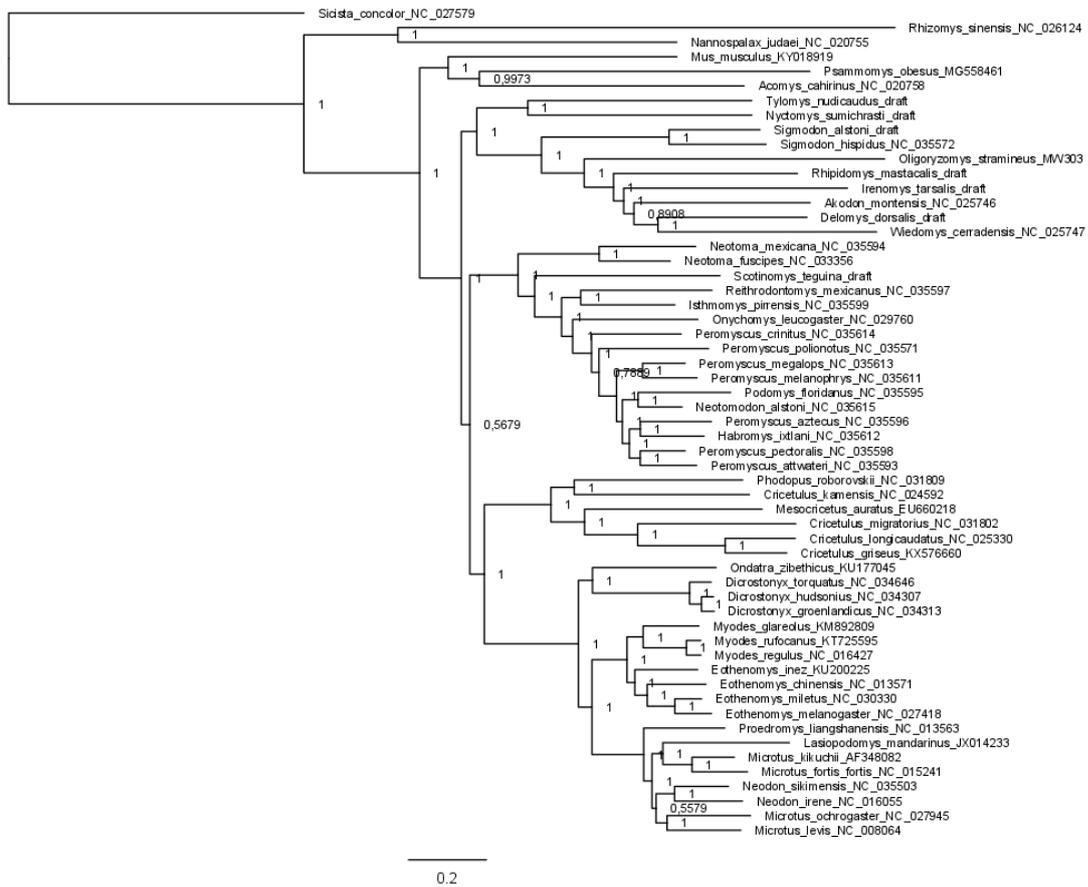
Apêndice 7 - Árvore bayesiana inferida baseando-se no conjunto 3.



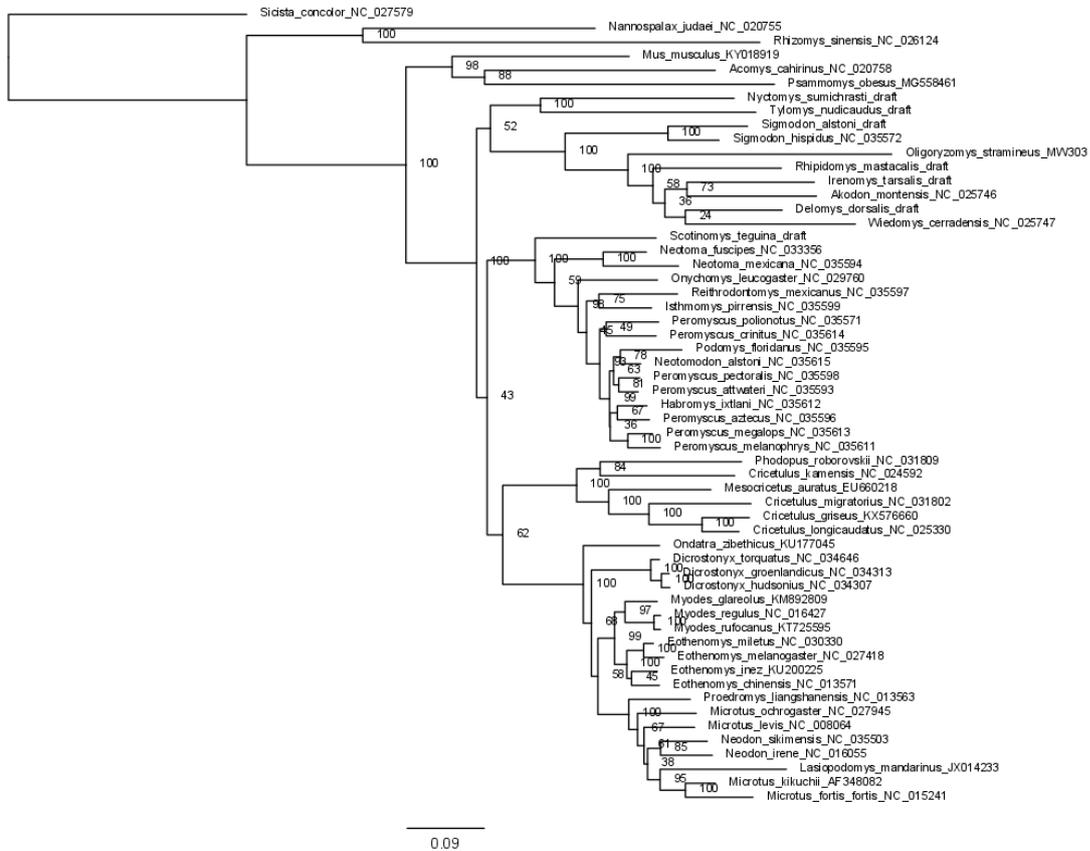
Apêndice 8 – Árvore de máxima verossimilhança inferida baseando-se no conjunto 4.



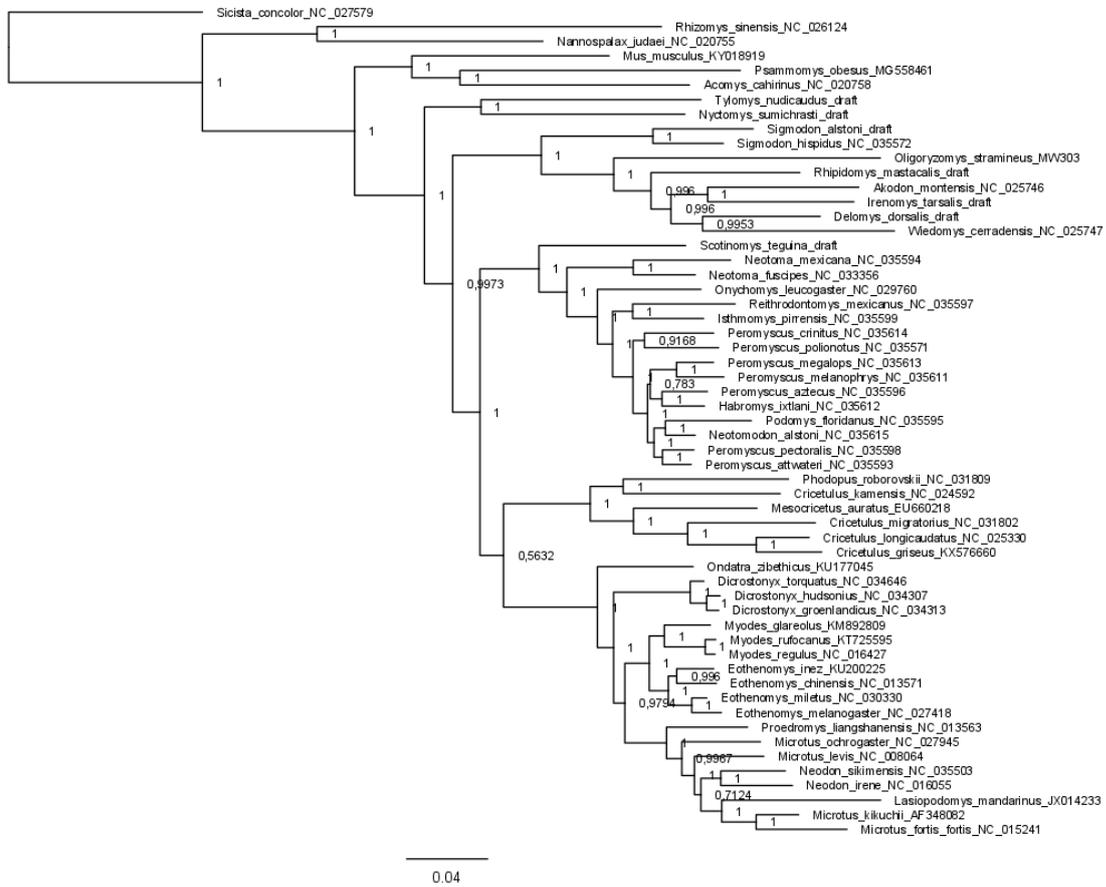
Apêndice 9 - Árvore bayesiana inferida baseando-se no conjunto 4.



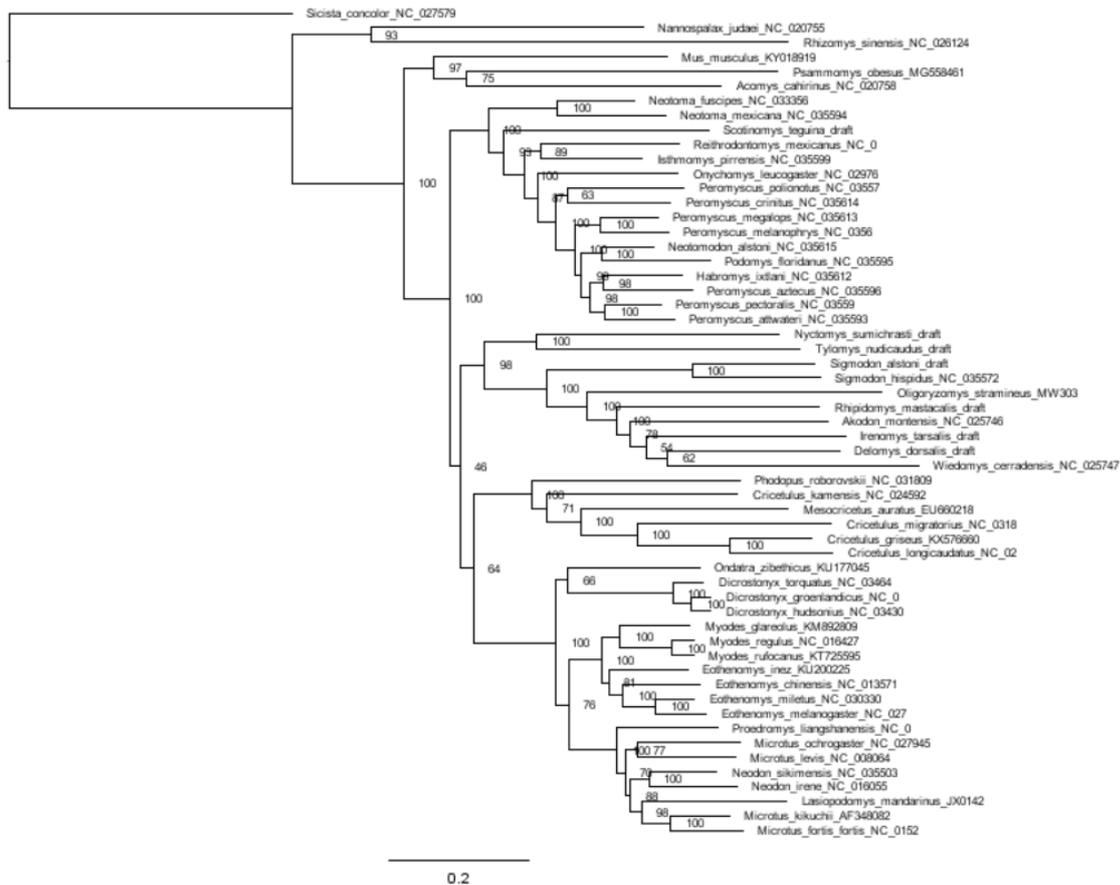
Apêndice 10 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 5.



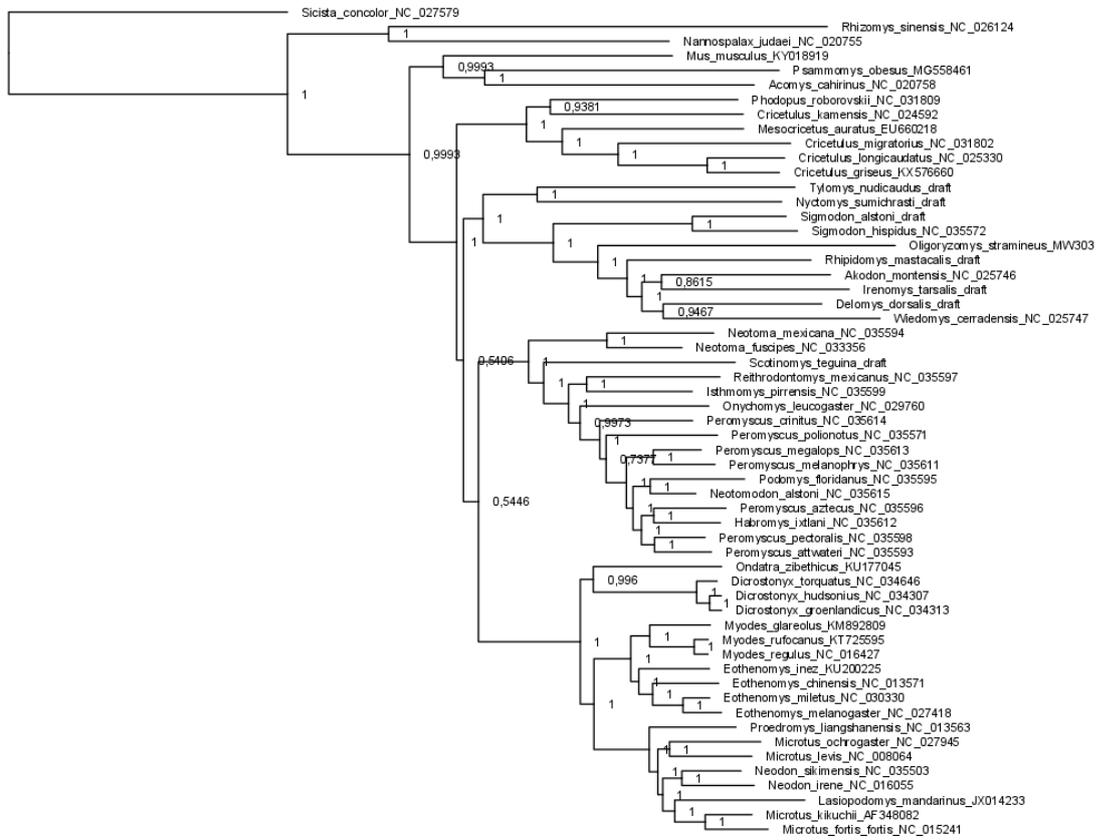
Apêndice 11 - Árvore bayesiana inferida baseando-se no conjunto 5.



Apêndice 12 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 6.

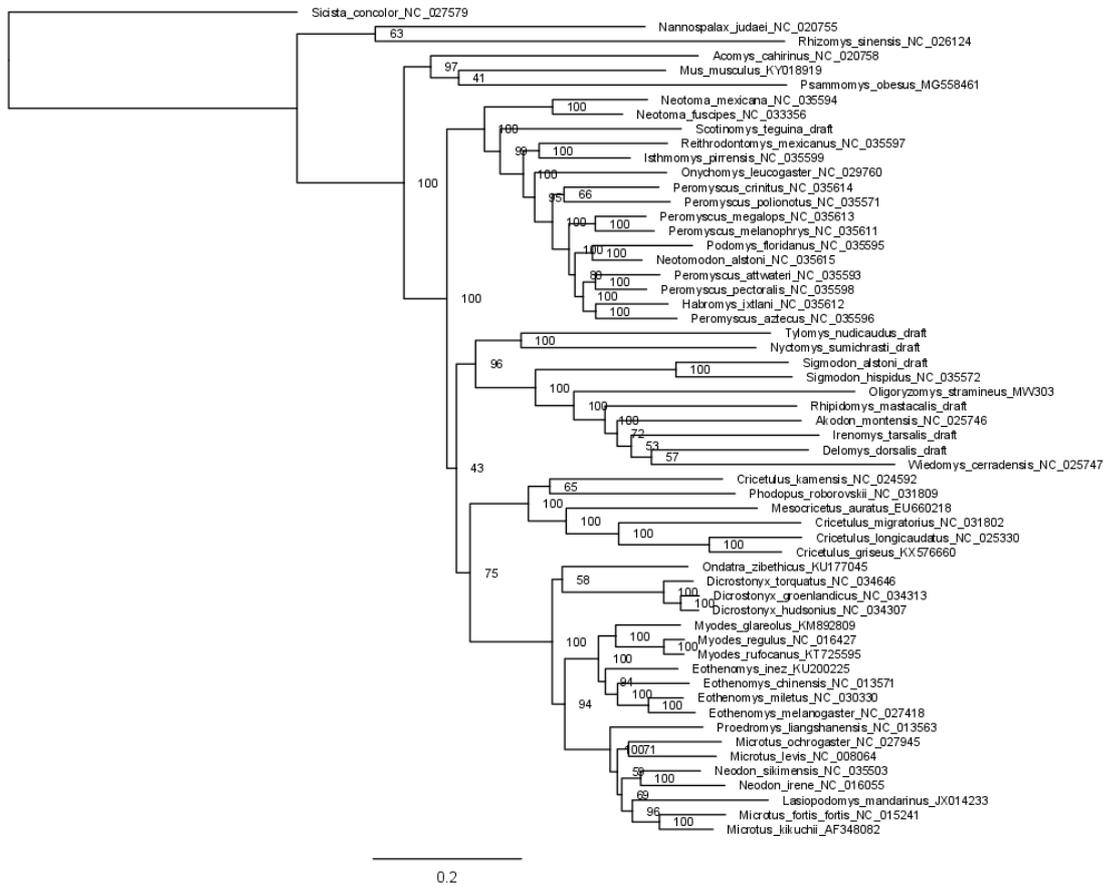


Apêndice 13 - Árvore bayesiana inferida baseando-se no conjunto 6.

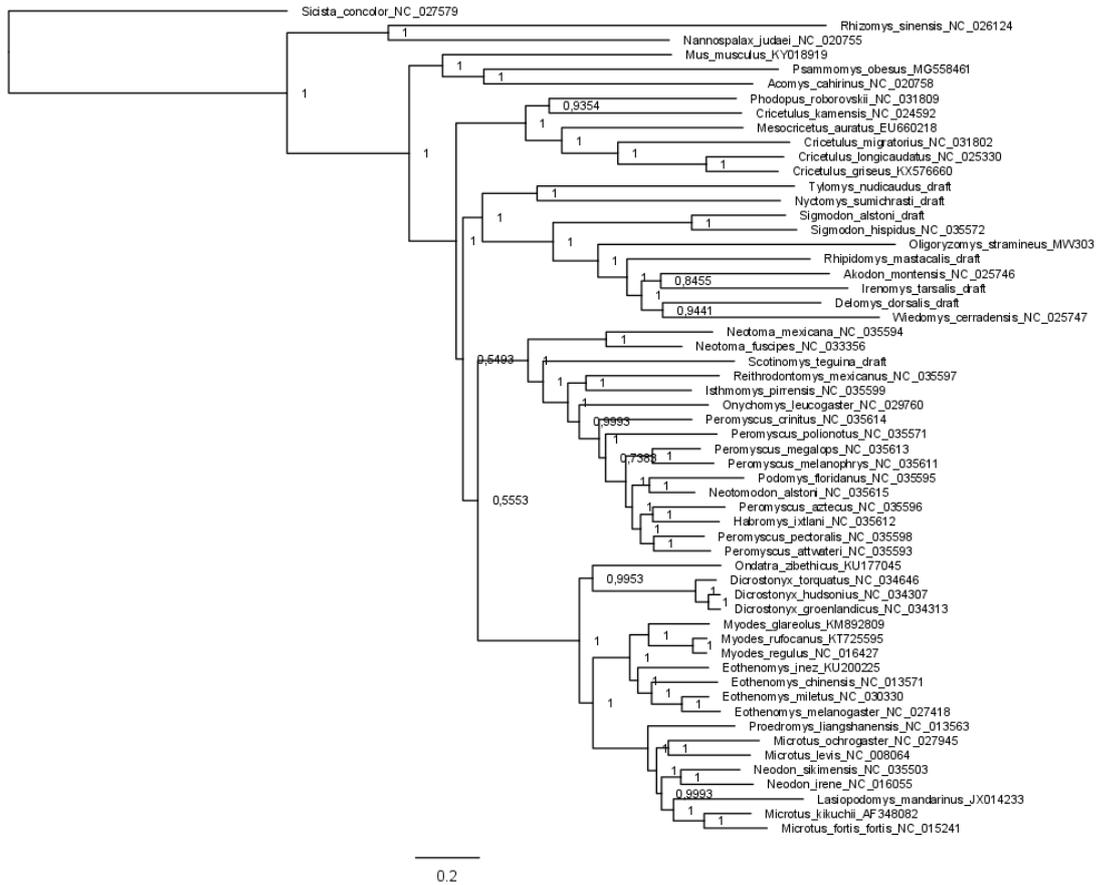


0.2

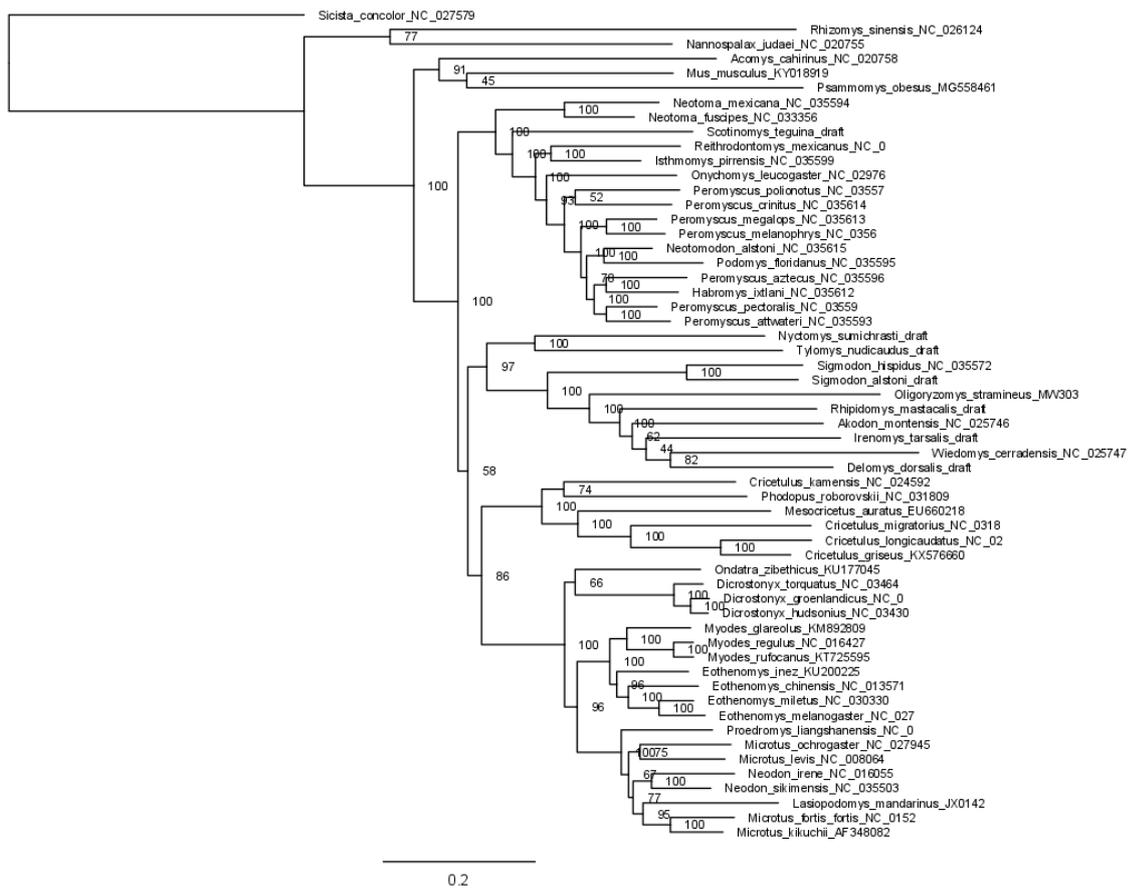
Apêndice 14 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 7.



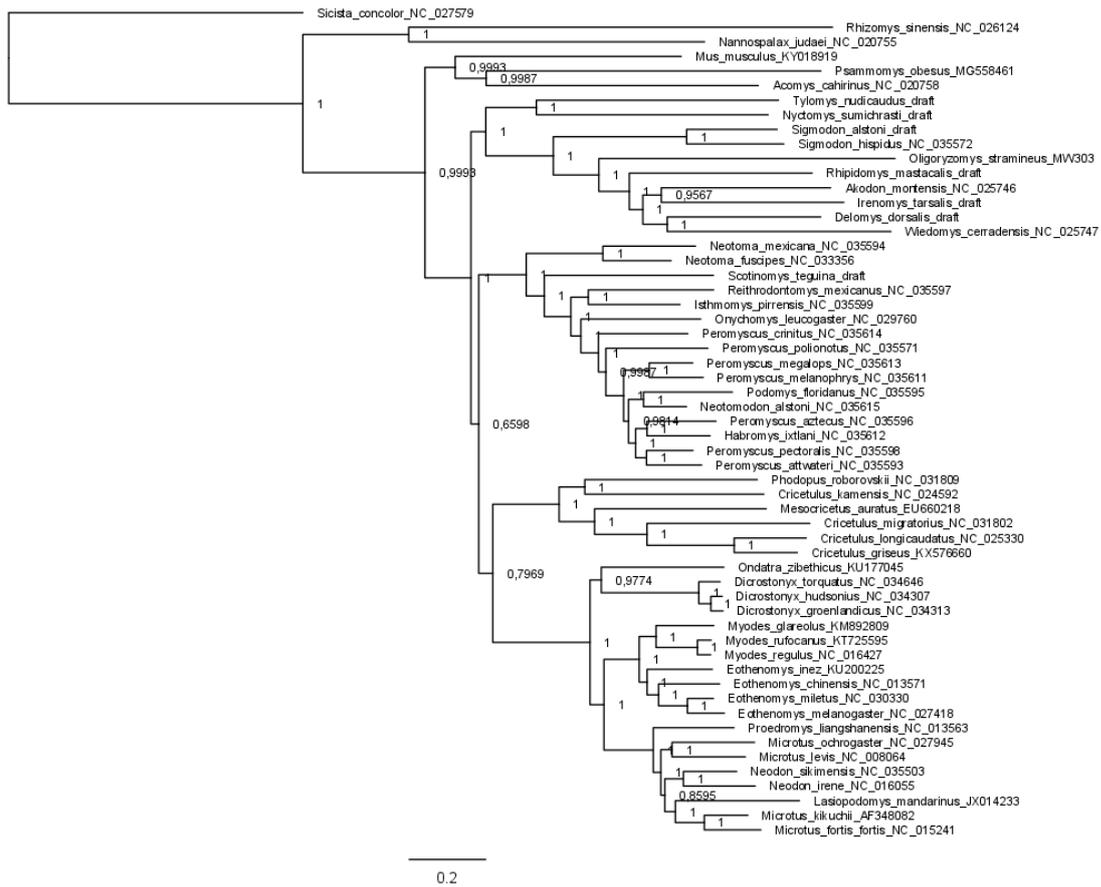
Apêndice 15 - Árvore bayesiana inferida baseando-se no conjunto 7.



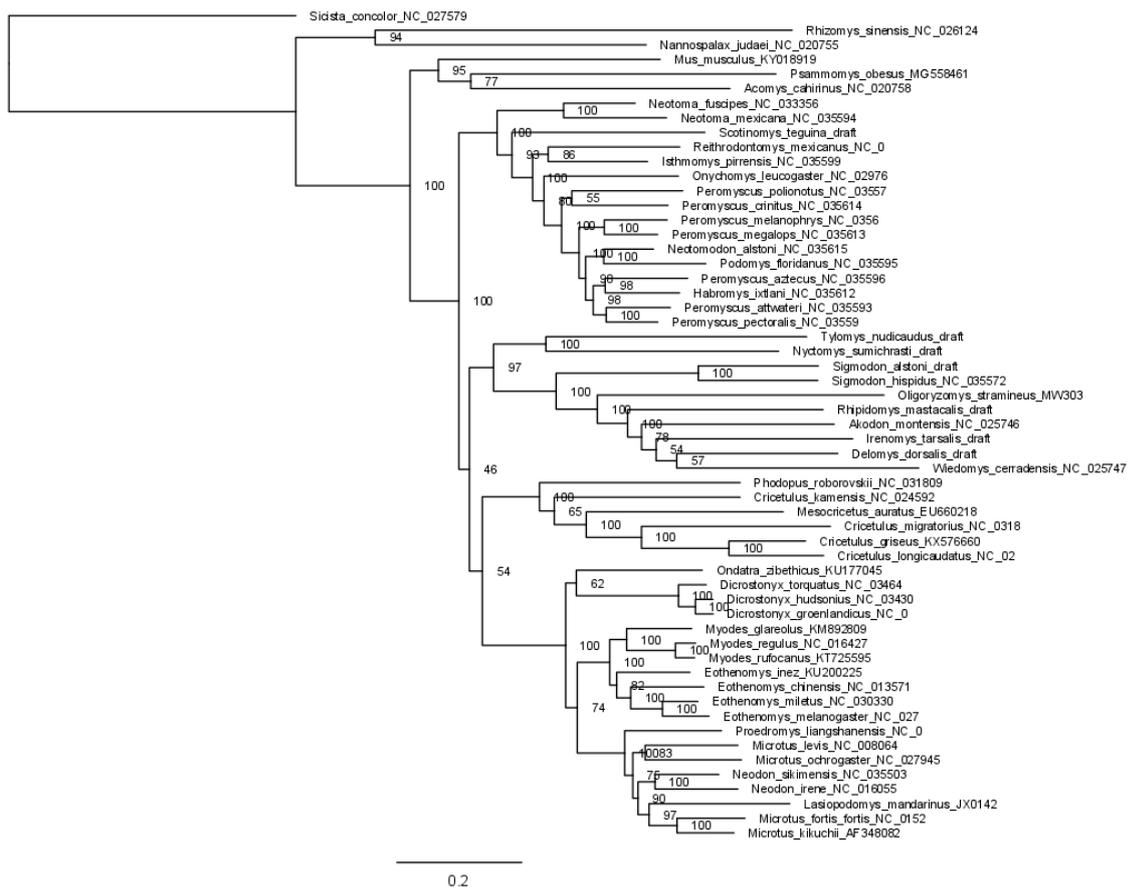
Apêndice 16 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 8.



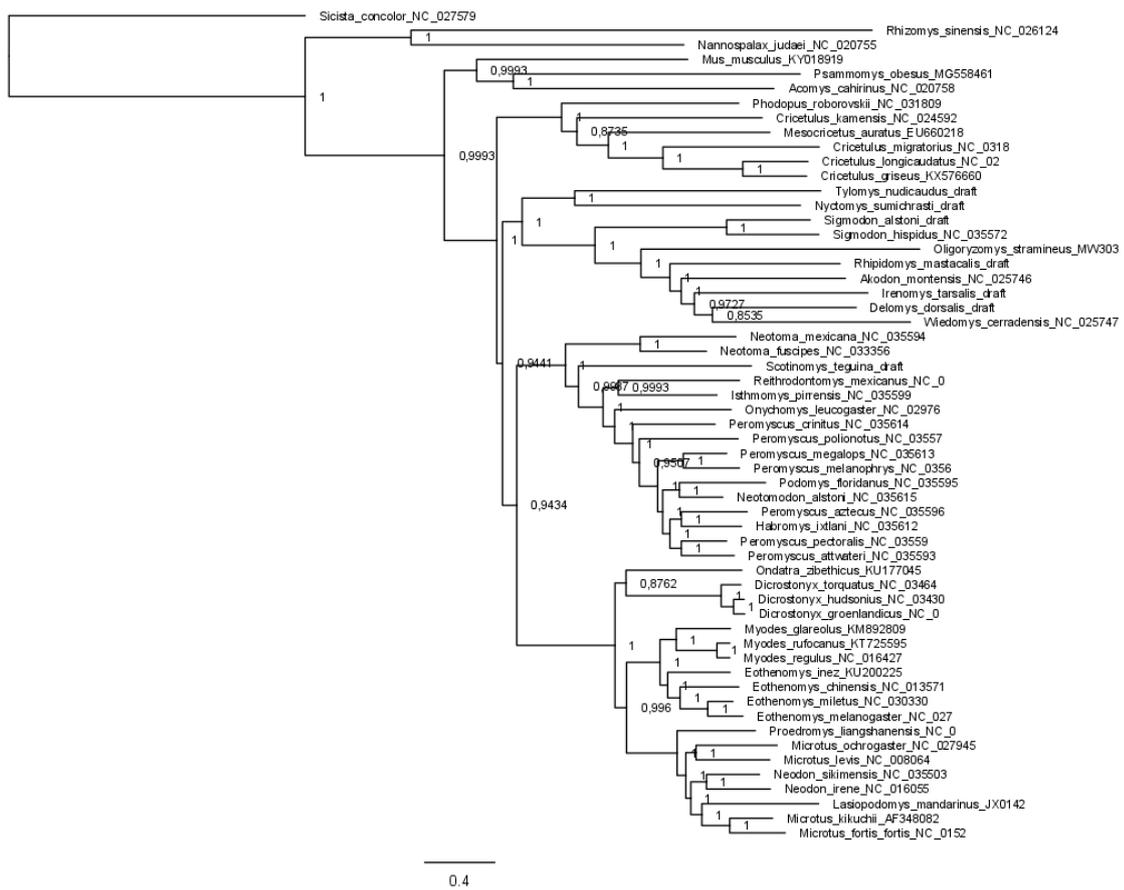
Apêndice 17 - Árvore bayesiana inferida baseando-se no conjunto 8.



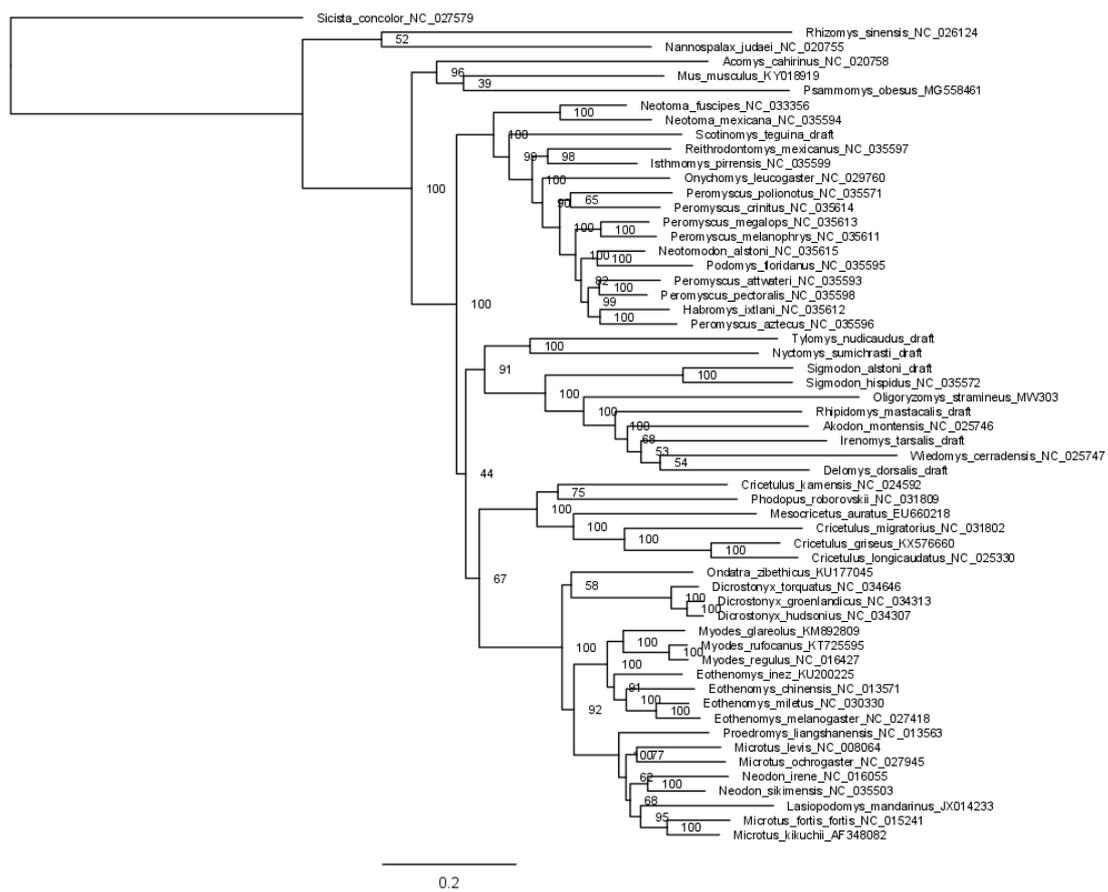
Apêndice 18 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 9.



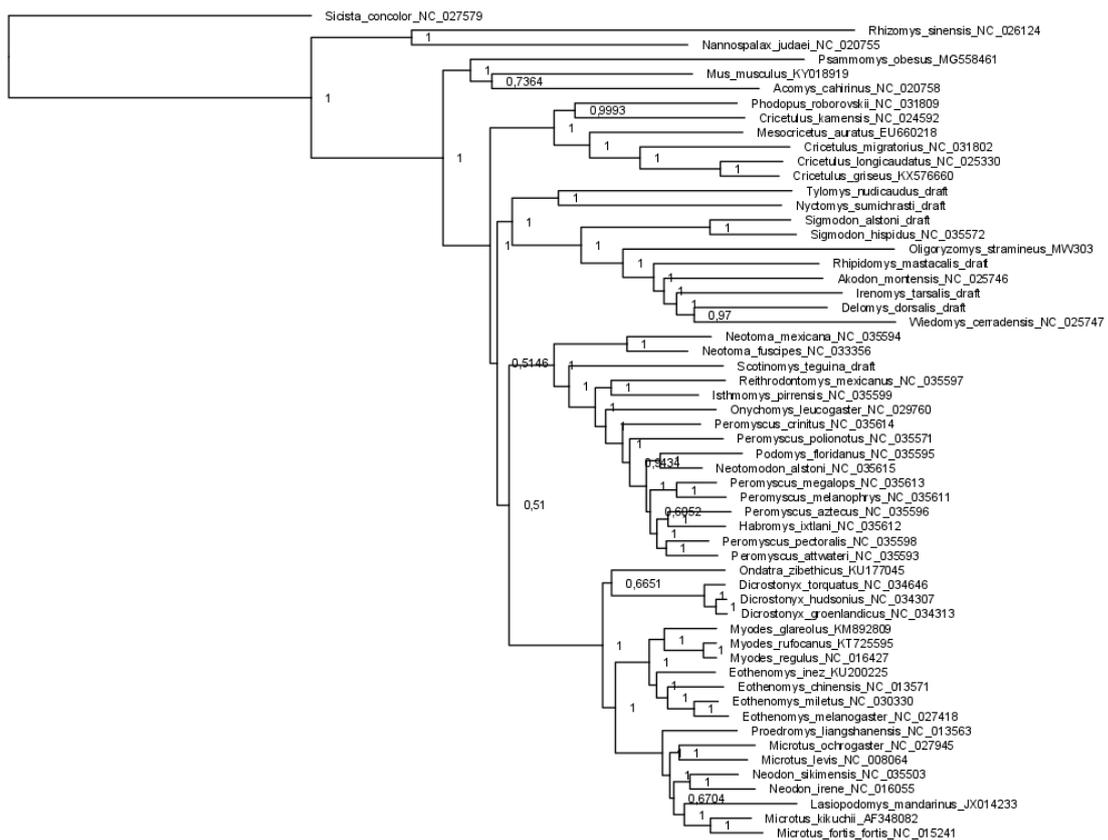
Apêndice 19 - Árvore bayesiana inferida baseando-se no conjunto 9.



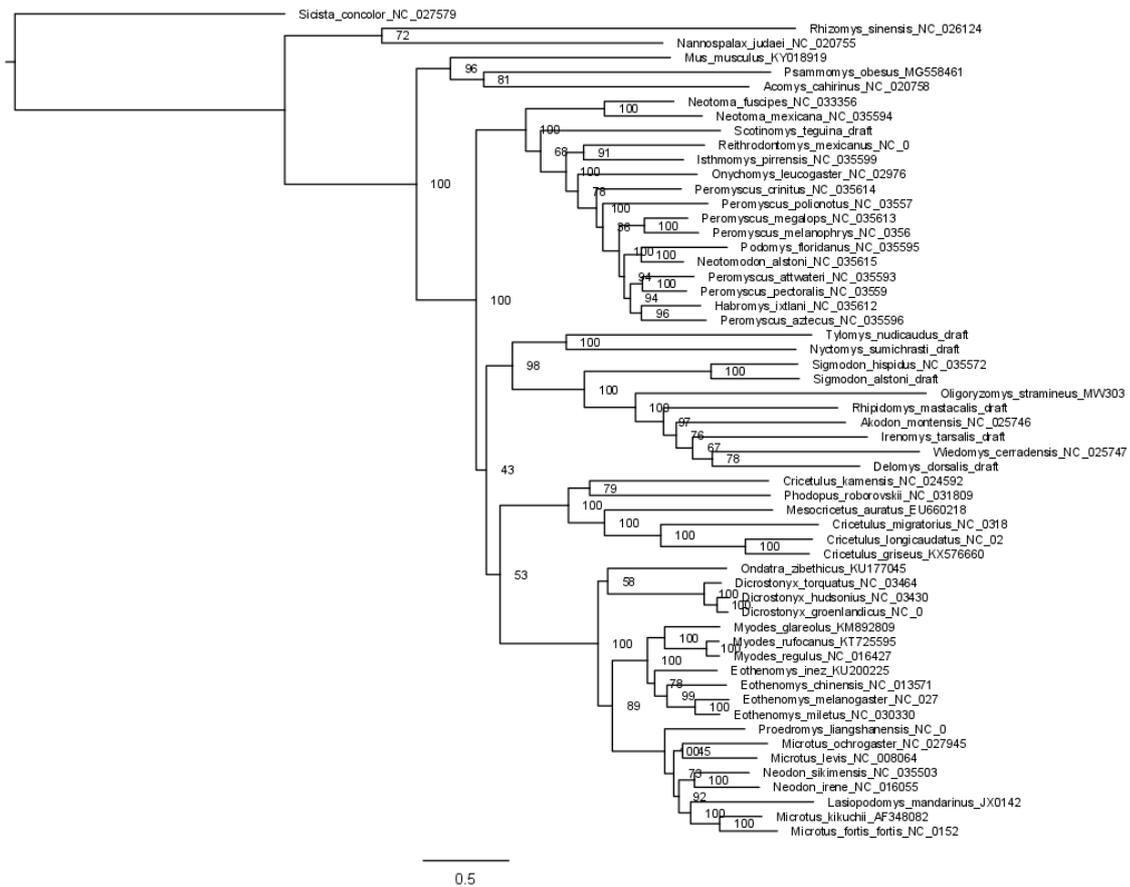
Apêndice 20 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 10.



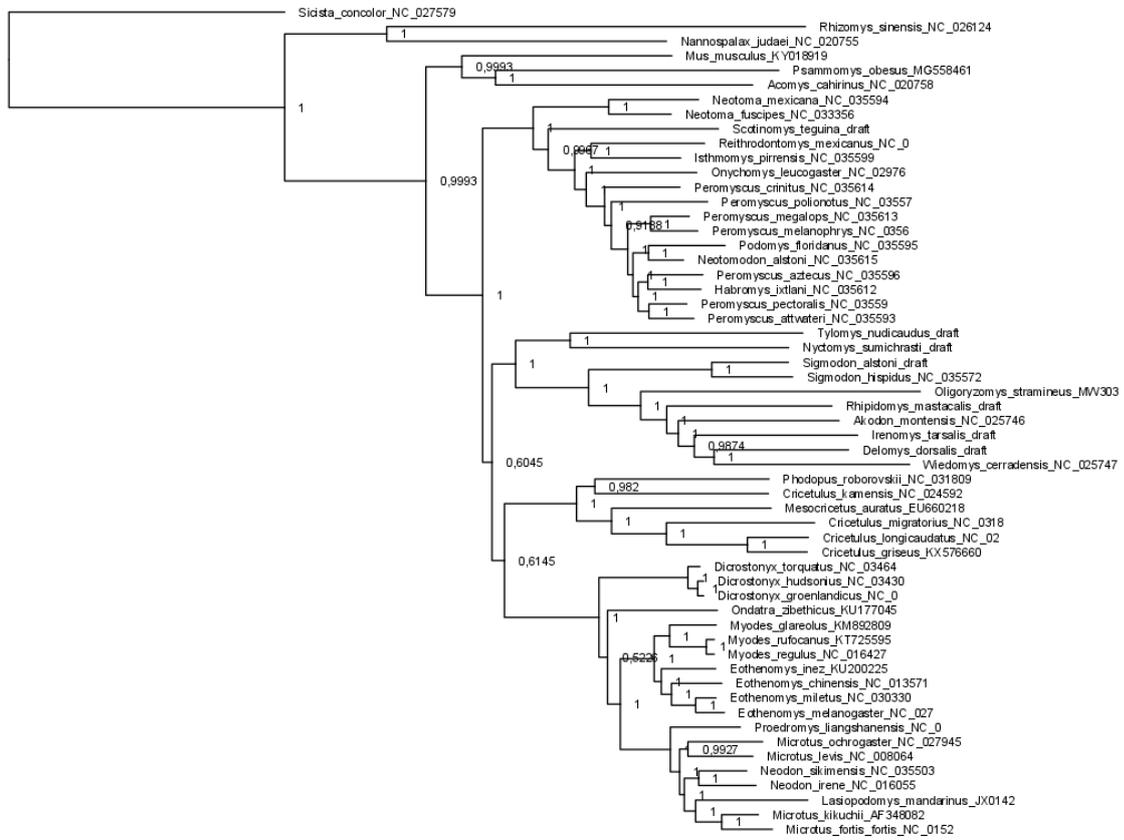
Apêndice 21 - Árvore bayesiana inferida baseando-se no conjunto 10.



Apêndice 22 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 12.

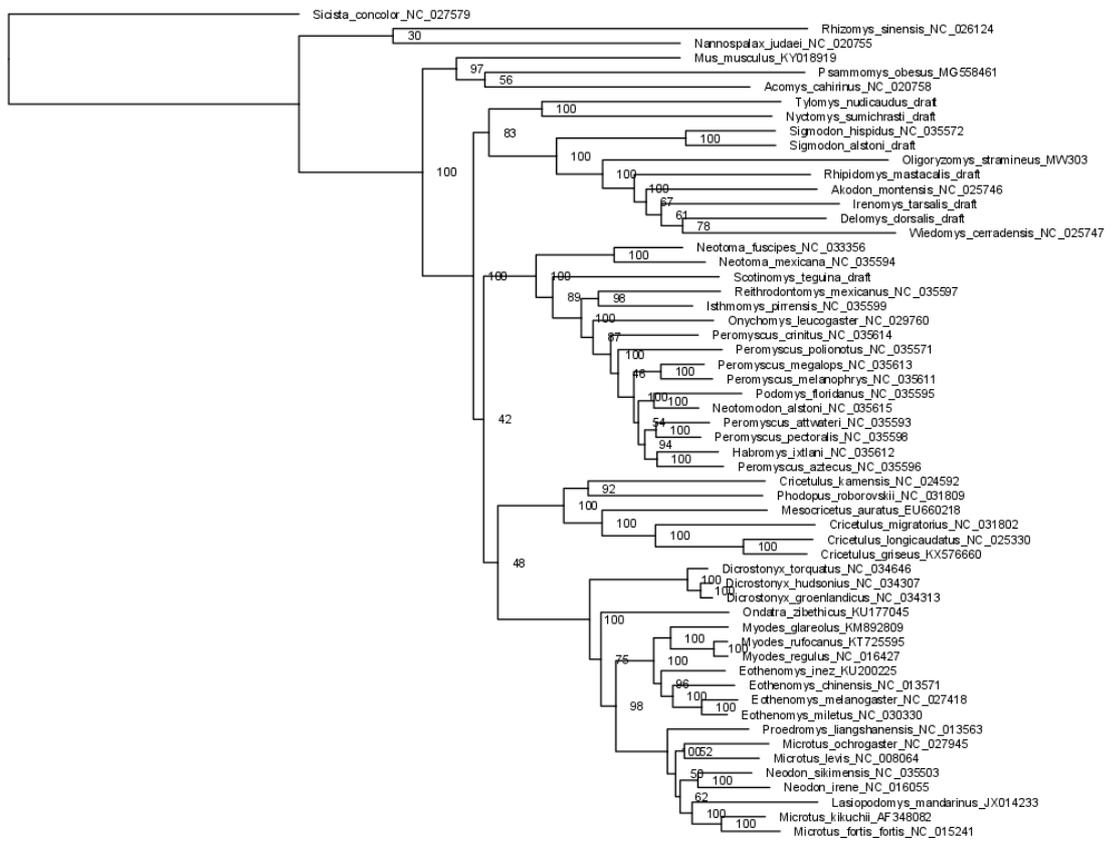


Apêndice 23 - Árvore bayesiana inferida baseando-se no conjunto 12.



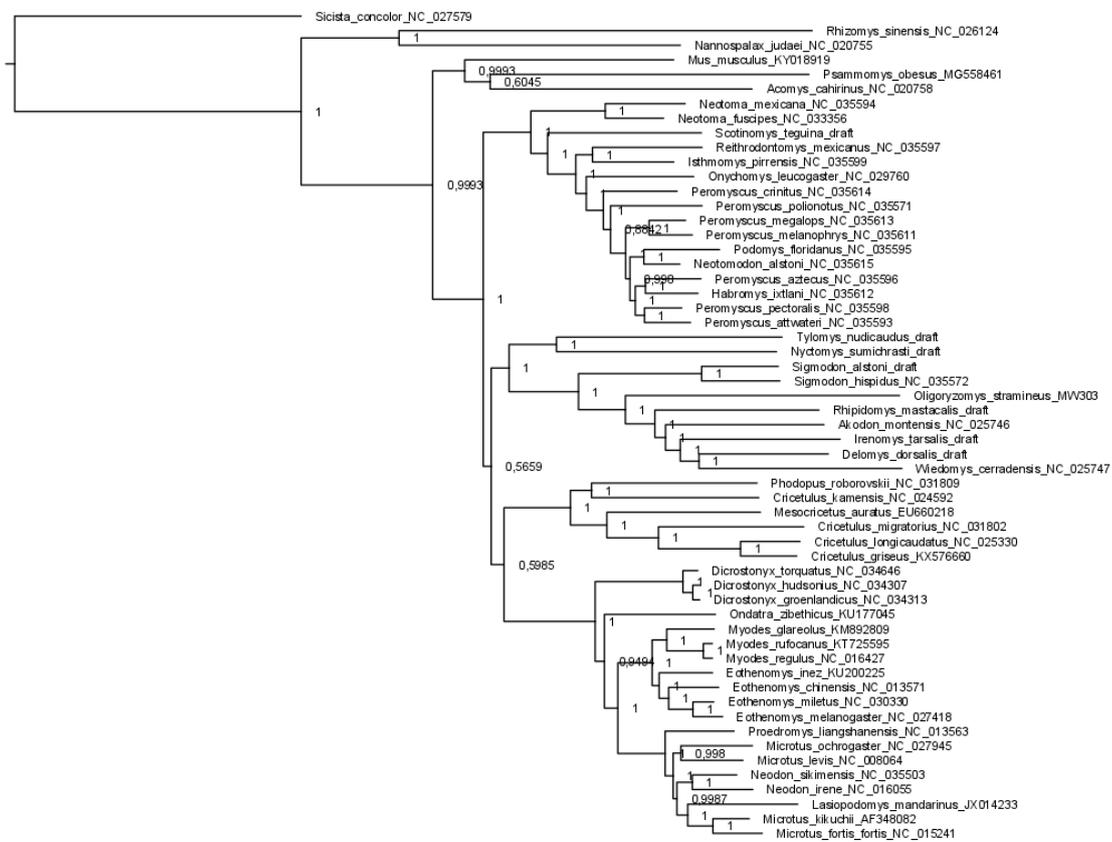
0.5

Apêndice 24 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 13.

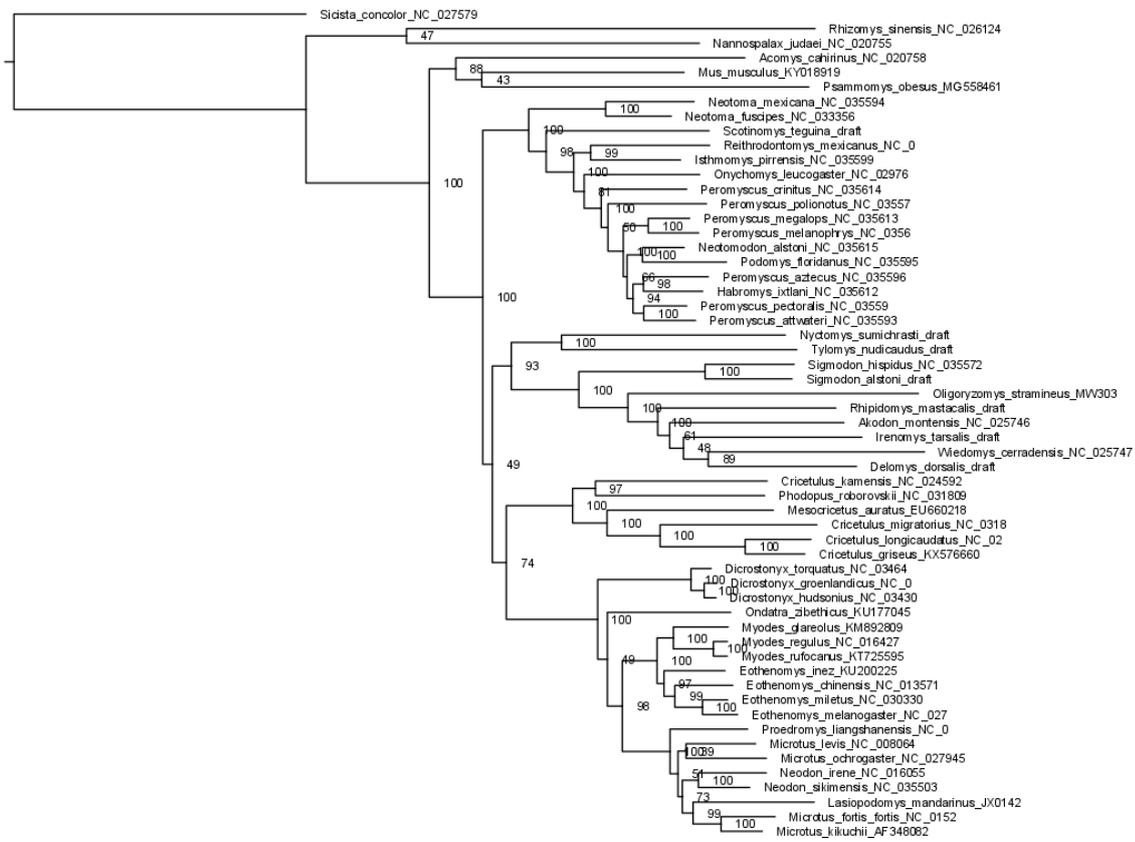


0.5

Apêndice 25 - Árvore bayesiana inferida baseando-se no conjunto 13.

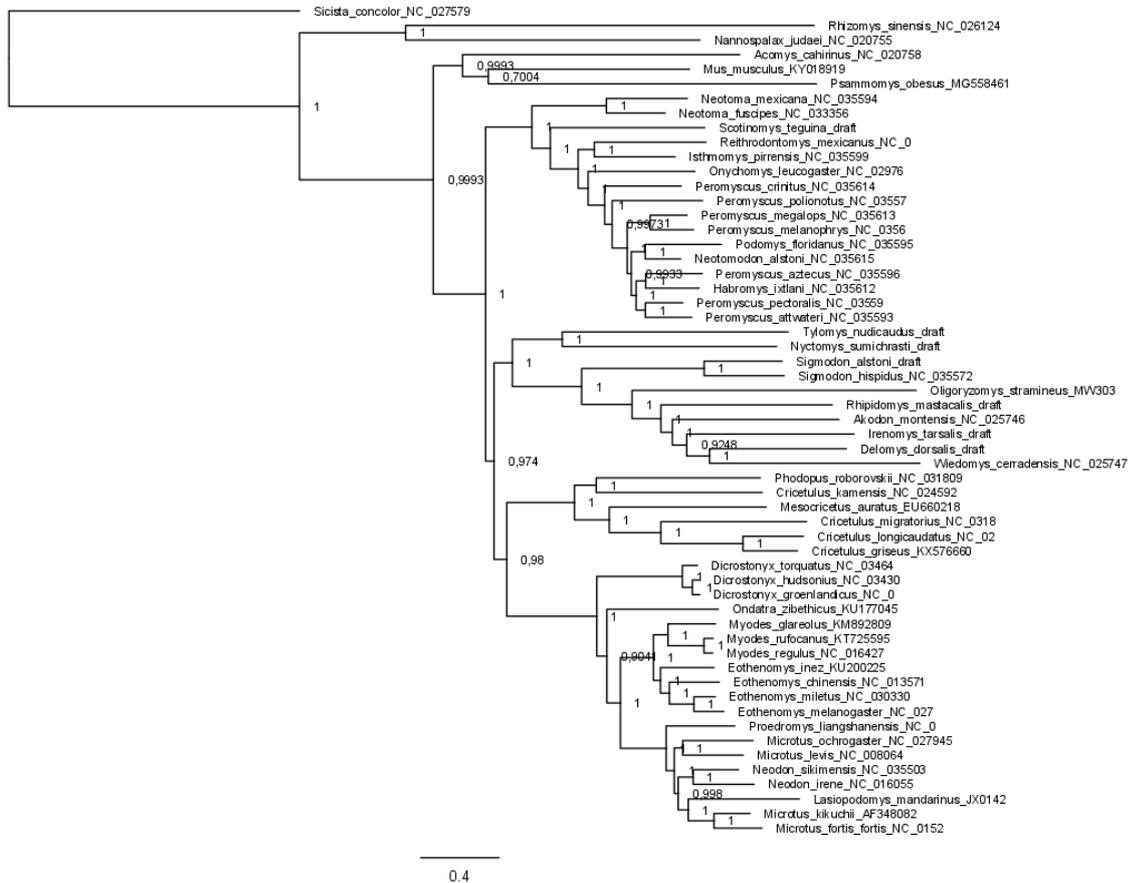


Apêndice 26 - Árvore de máxima verossimilhança inferida baseando-se no conjunto 14.

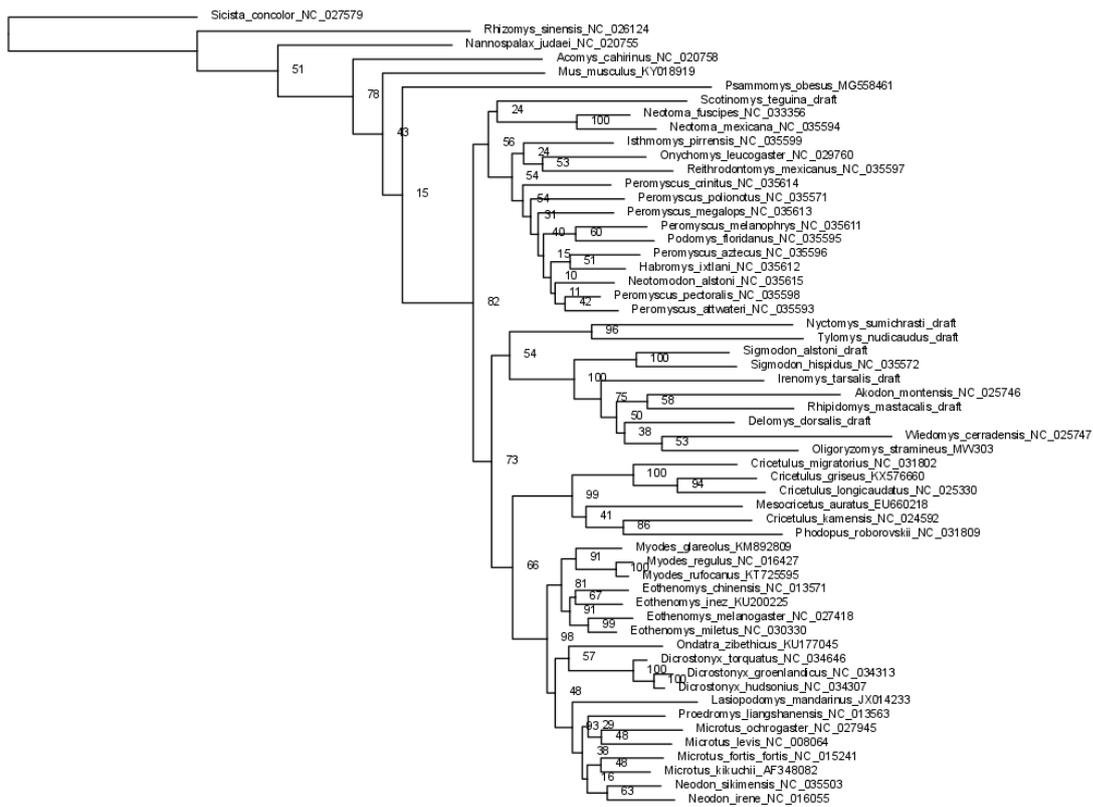


0.4

Apêndice 27 - Árvore bayesiana inferida baseando-se no conjunto 14.

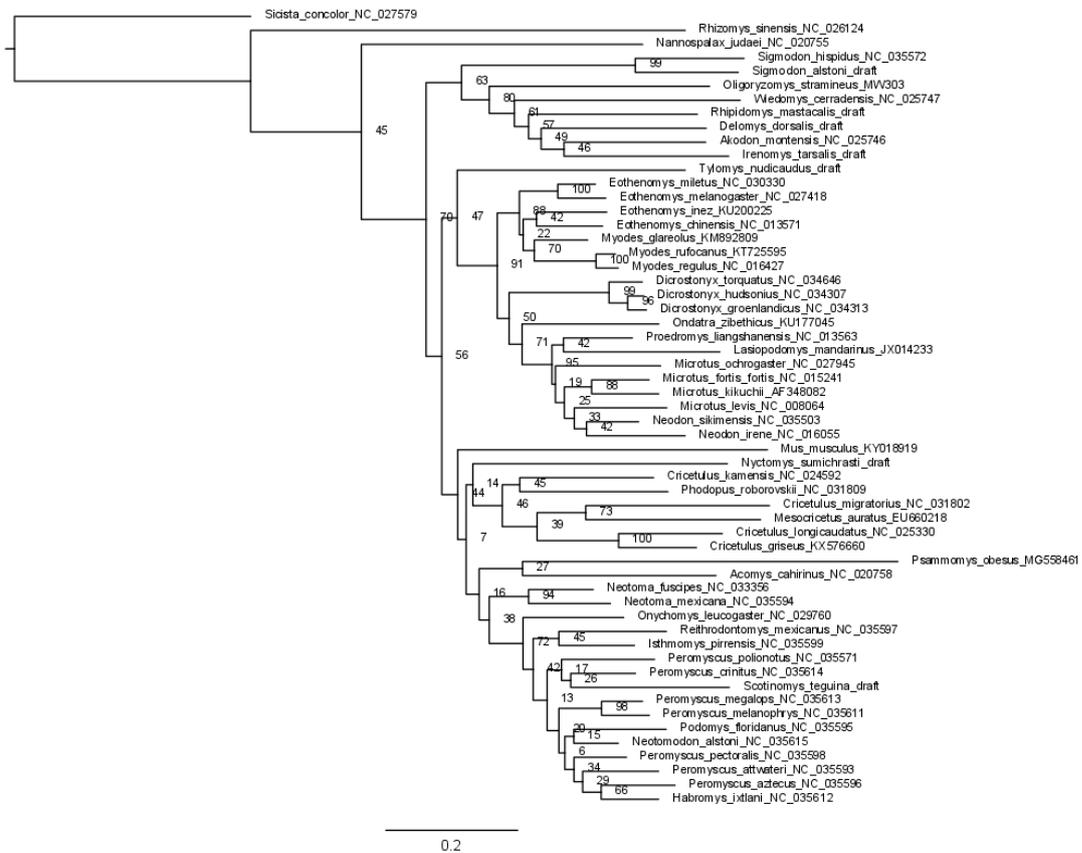


Apêndice 28 - Árvore de máxima verossimilhança baseada em CYTB.

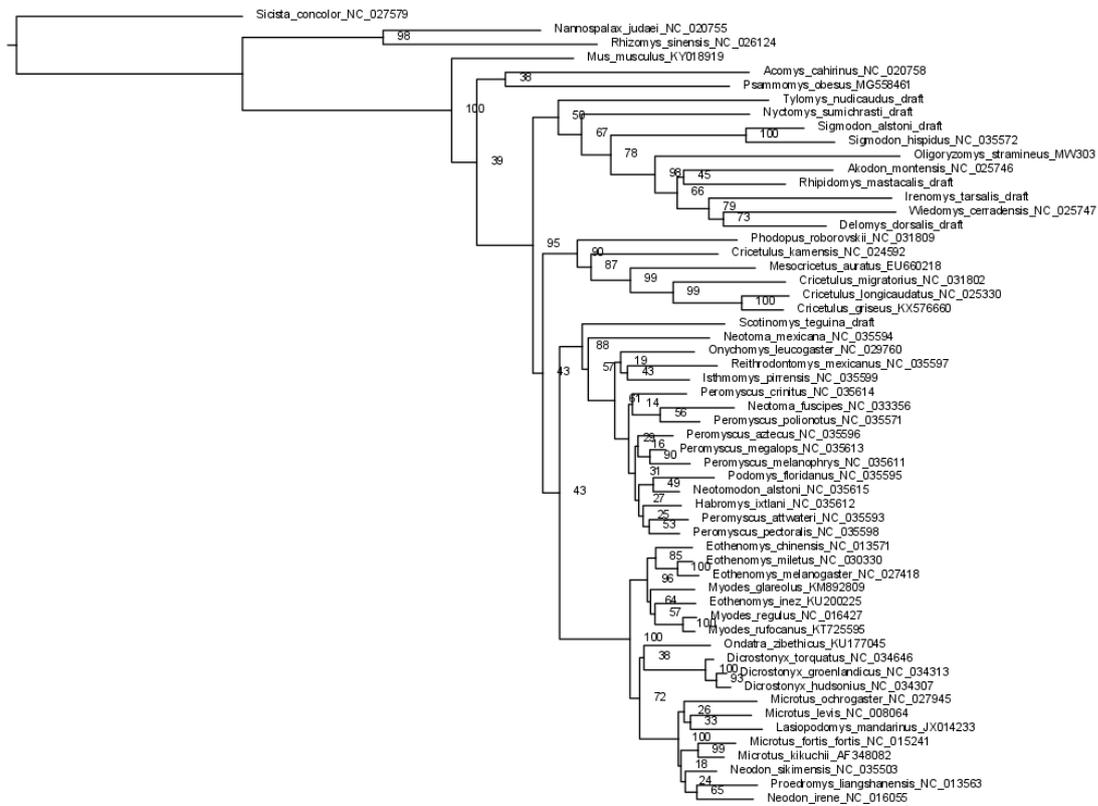


0.1

Apêndice 29 - Árvore de máxima verossimilhança baseada em ND1.

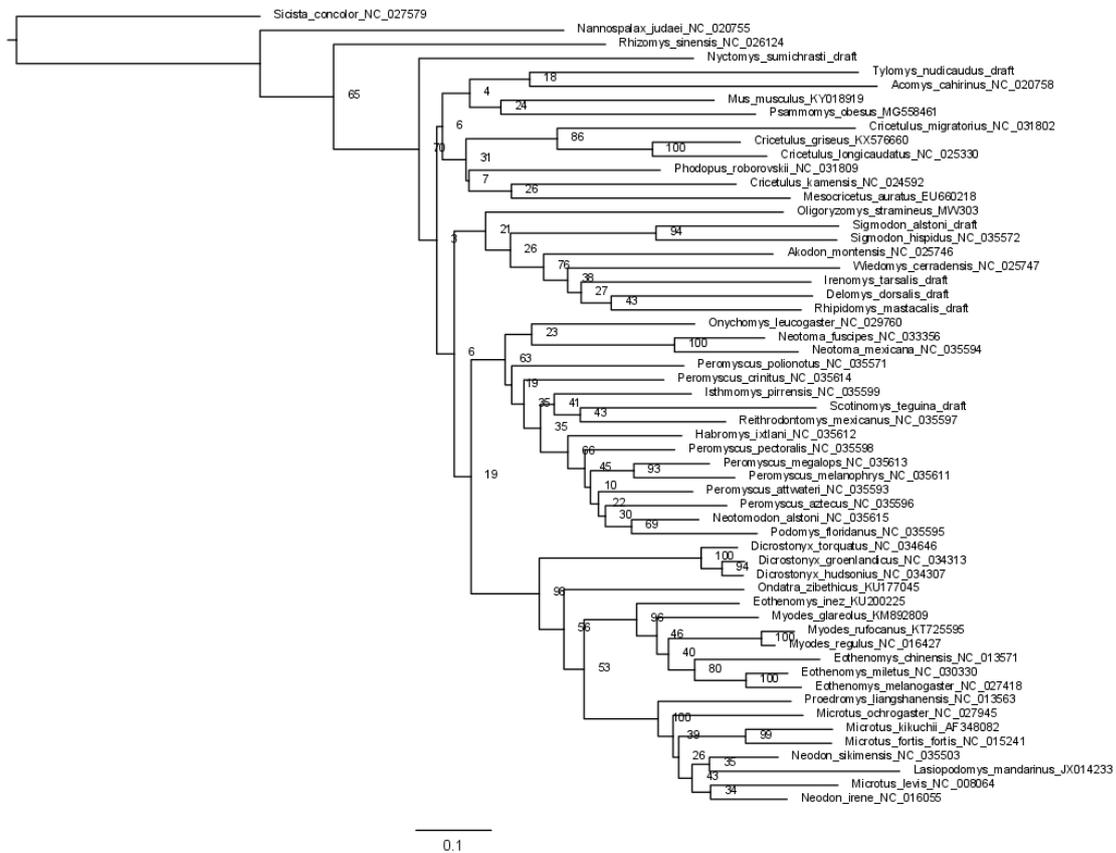


Apêndice 30 - Árvore de máxima verossimilhança baseada em ND2.

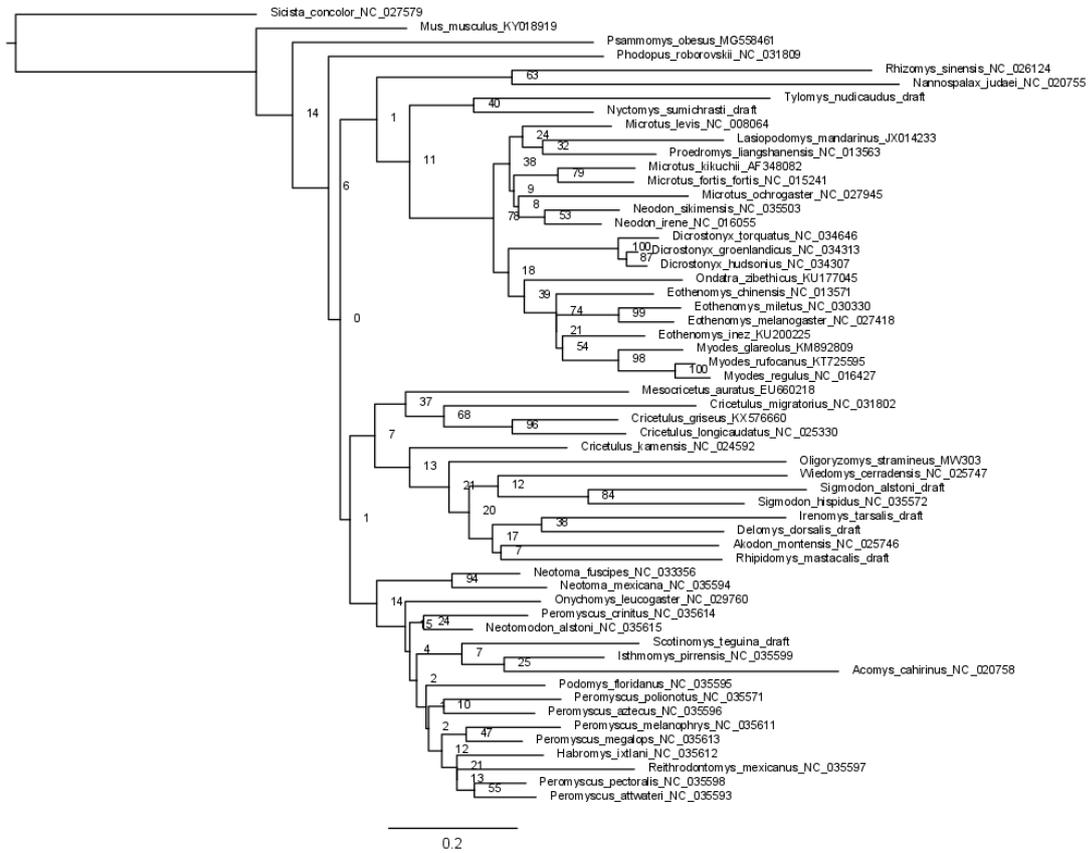


0.2

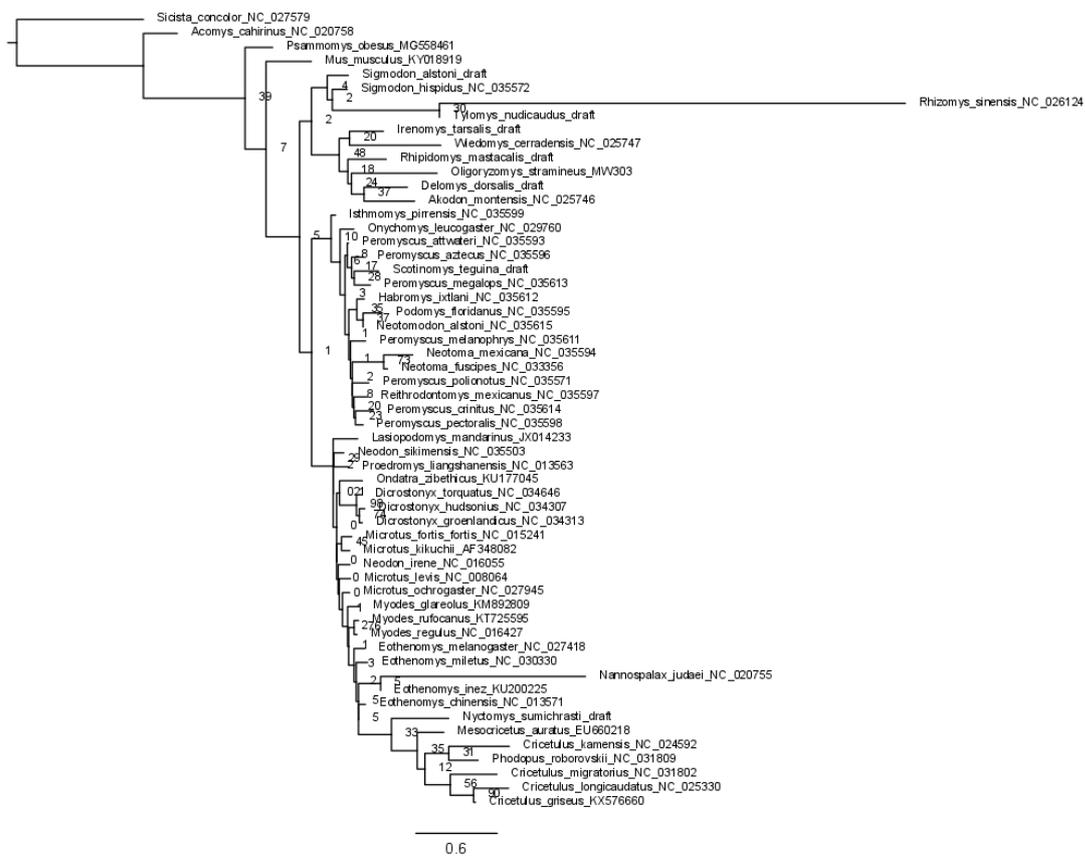
Apêndice 31 - Árvore de máxima verossimilhança baseada em COX1.



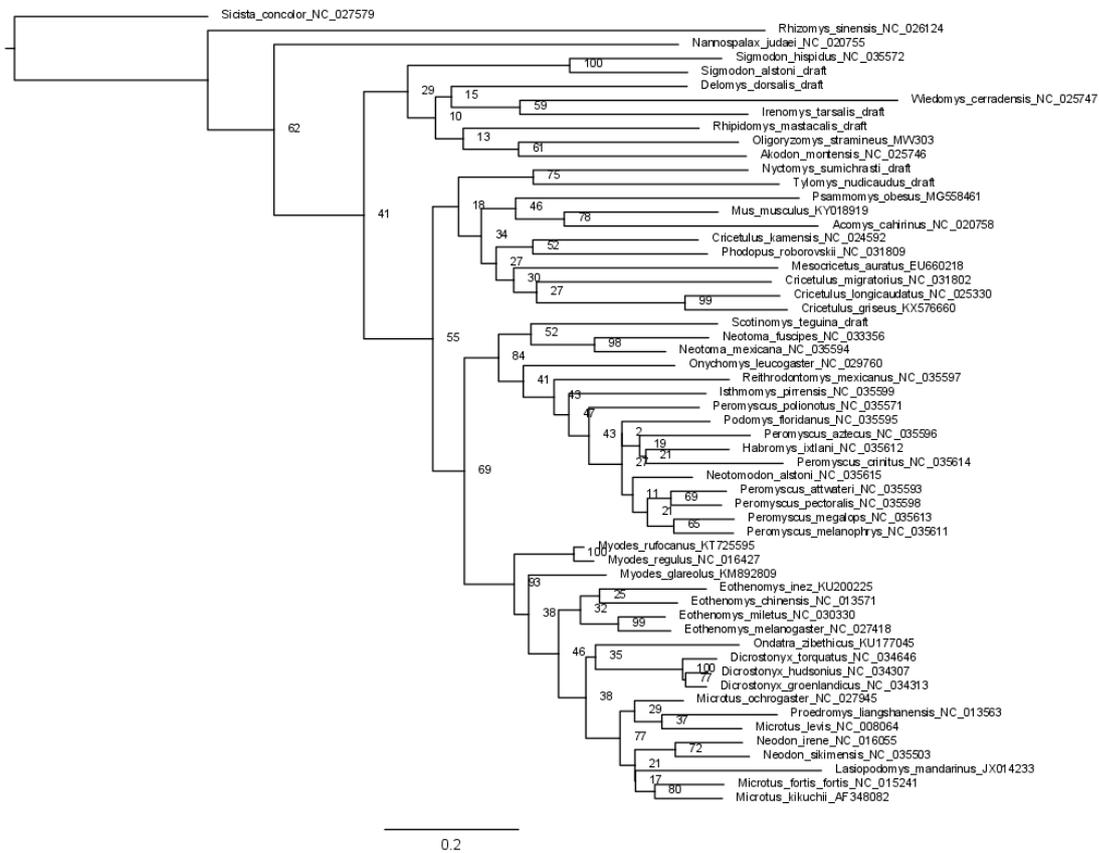
Apêndice 32 - Árvore de máxima verossimilhança baseada em COX2.



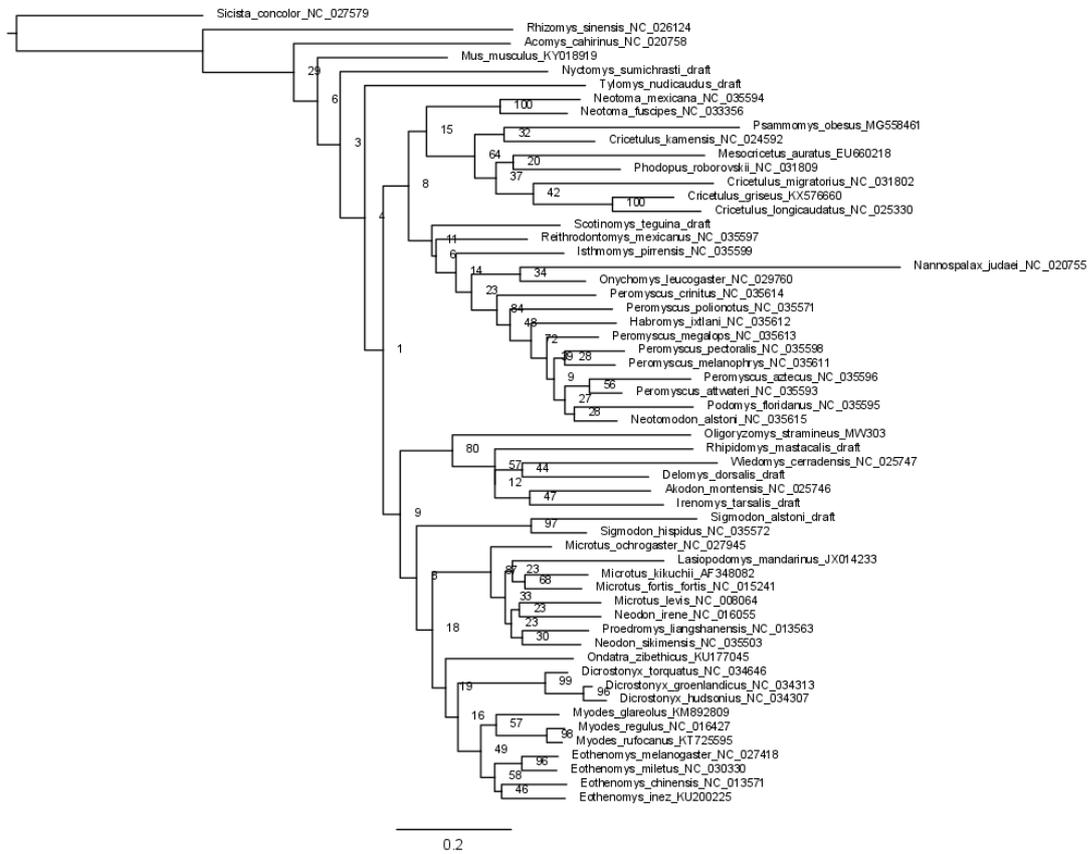
Apêndice 33 - Árvore de máxima verossimilhança baseada em ATP8.



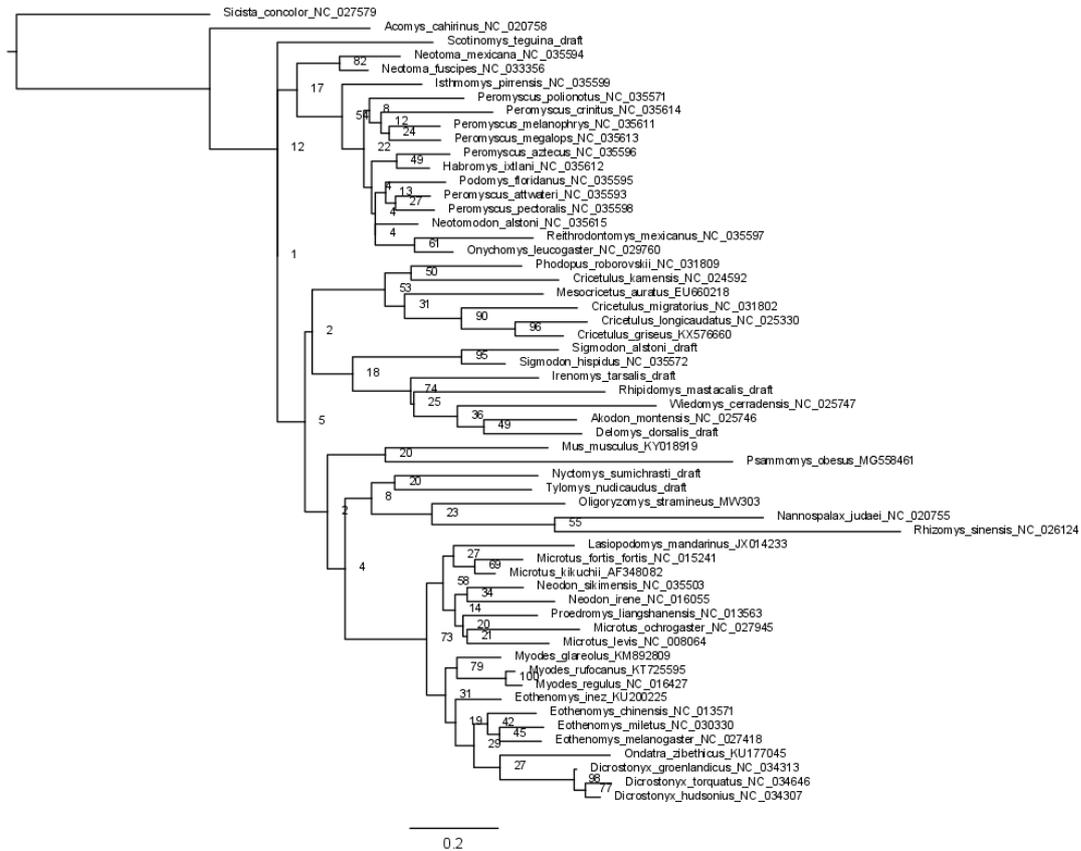
Apêndice 34 - Árvore de máxima verossimilhança baseada em ATP6.



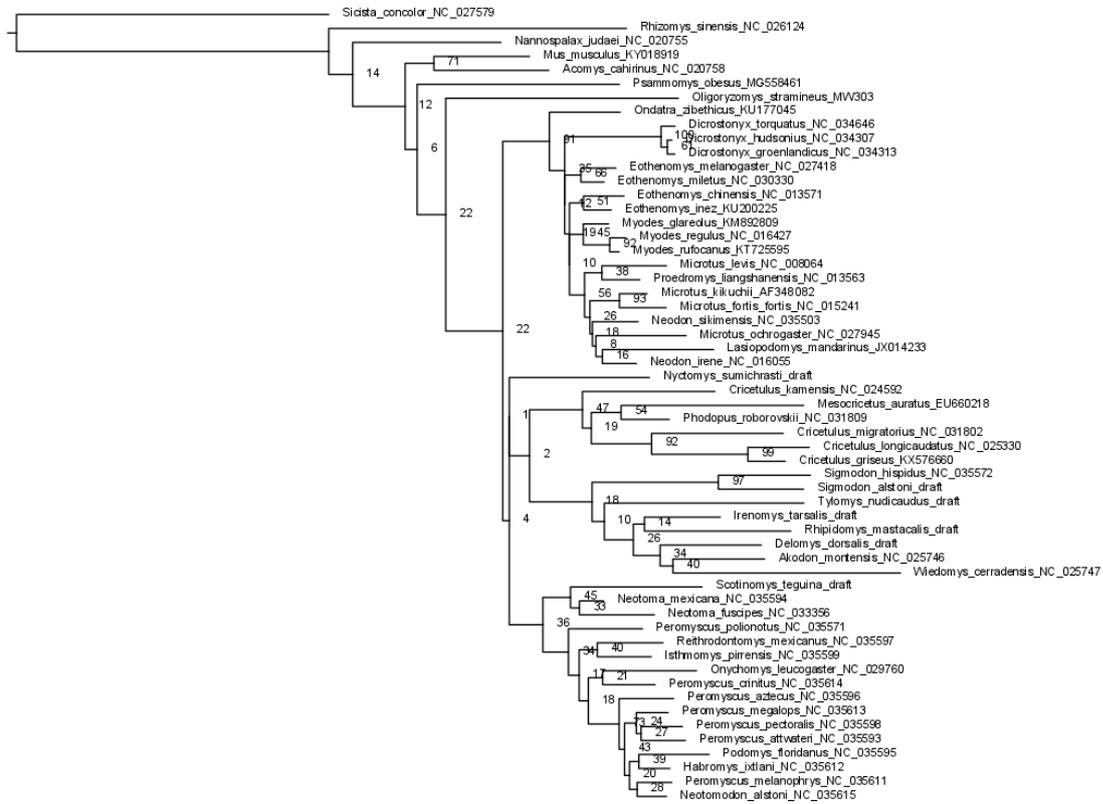
Apêndice 35 - Árvore de máxima verossimilhança baseada em COX3.



Apêndice 36 - Árvore de máxima verossimilhança baseada em ND3.

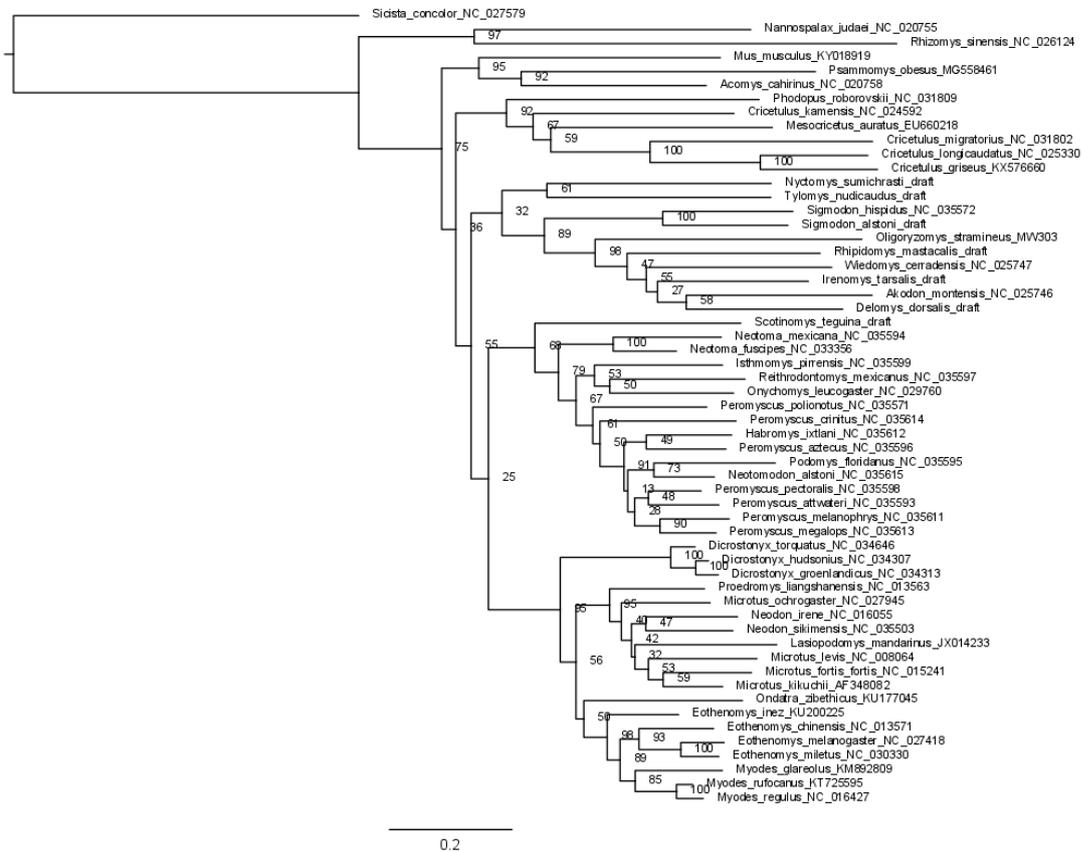


Apêndice 37 - Árvore de máxima verossimilhança baseada em ND4L.

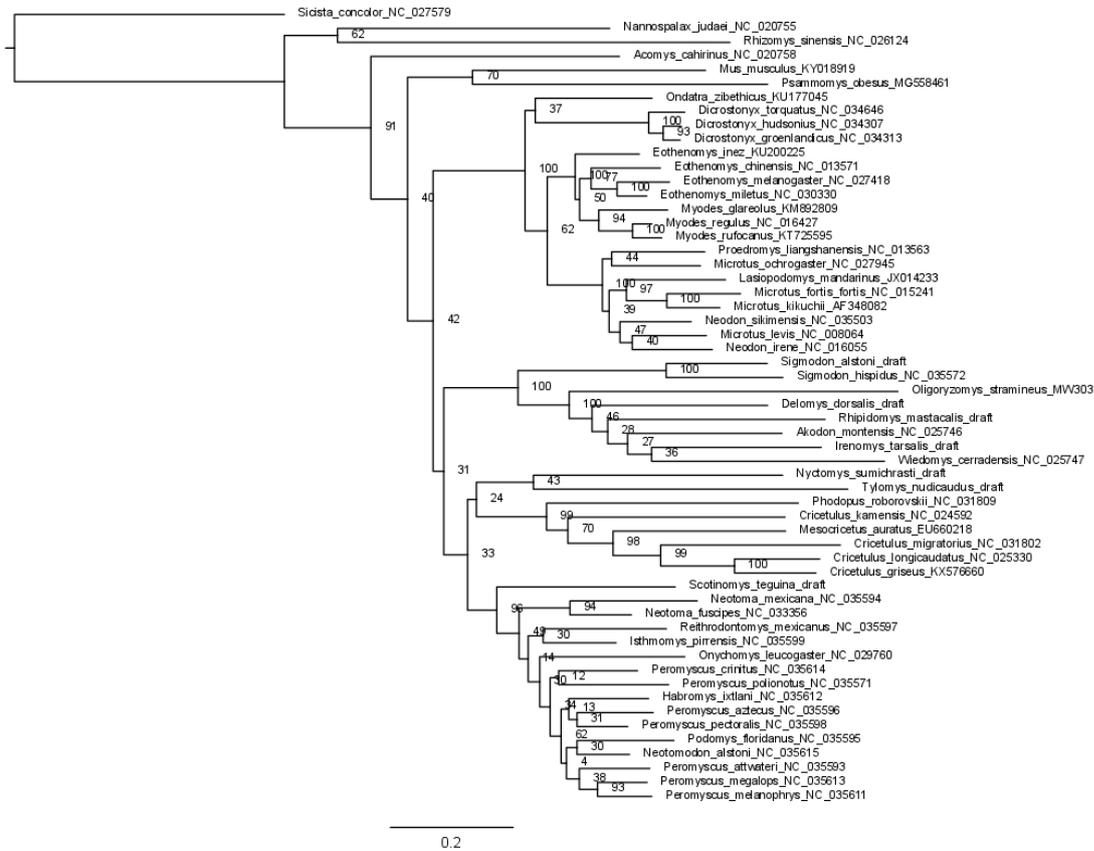


0.3

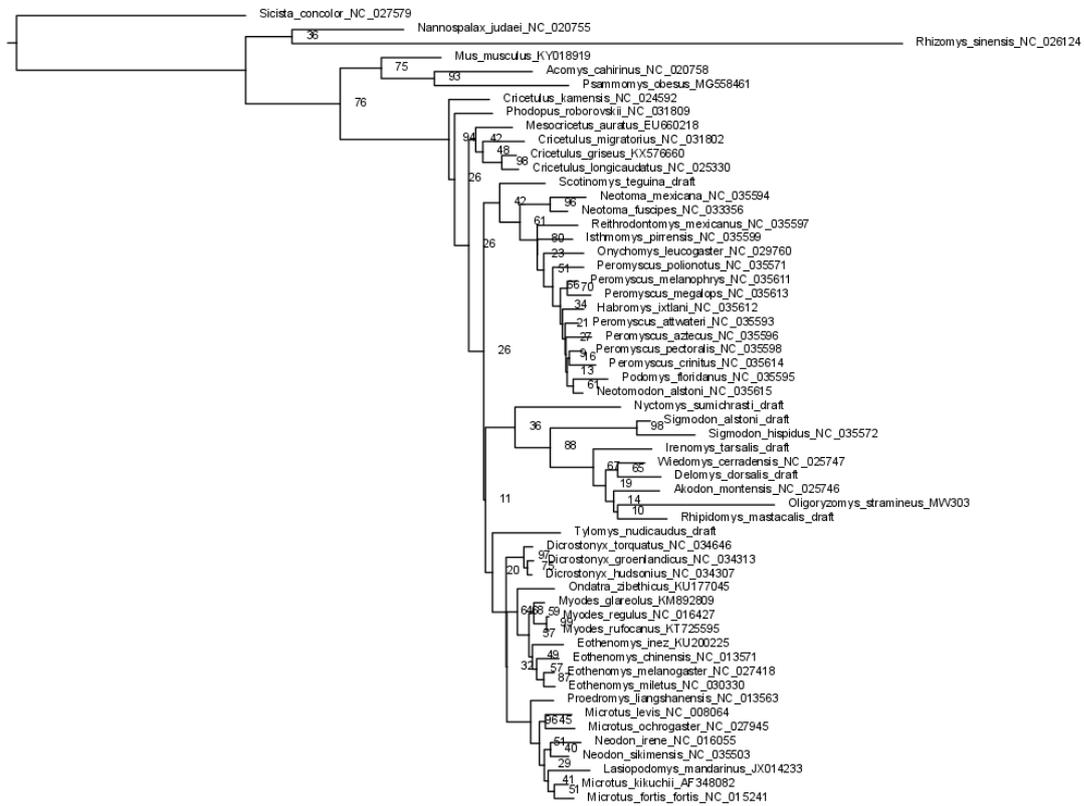
Apêndice 38 - Árvore de máxima verossimilhança baseada em ND4.



Apêndice 39 - Árvore de máxima verossimilhança baseada em ND5.



Apêndice 40 - Árvore de máxima verossimilhança baseada em ND6.



0.6