



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO  
UNIDADE ACADÊMICA DE SERRA TALHADA  
CURSO DE BACHARELADO EM CIÊNCIAS BIOLÓGICAS

**INFERÊNCIA FILOGENÉTICA DE APOIDEA  
(HYMENOPTERA) A PARTIR DA ANÁLISE DO  
GRAU DE HOMOLOGIA DO CITOCROMO C POR  
FERRAMENTAS DE BIOINFORMÁTICA**

MONOGRAFIA

ERICLES CHARLES DA SILVA MELO

SERRA TALHADA – PE

2022

ERICLES CHARLES DA SILVA MELO

**INFERÊNCIA FILOGENÉTICA DE APOIDEA  
(HYMENOPTERA) A PARTIR DA ANÁLISE DO GRAU DE  
HOMOLOGIA DO CITOCROMO C POR FERRAMENTAS  
DE BIOINFORMÁTICA**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Ciências Biológicas da Universidade Federal Rural de Pernambuco / Unidade Acadêmica de Serra Talhada (UFRPE/UAST), como exigência para obtenção do título de Bacharel em Ciências Biológicas.

Orientador: Prof. Dr. Diego de Souza Buarque

SERRA TALHADA – PE

2022

Dados Internacionais de Catalogação na Publicação  
Universidade Federal Rural de Pernambuco  
Sistema Integrado de Bibliotecas  
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

---

M528i MELO, ERICLES CHARLES DA SILVA  
INFERÊNCIA FILOGENÉTICA DE APOIDEA (HYMENOPTERA) A PARTIR DA ANÁLISE DO GRAU  
DE HOMOLOGIA DO CITOCROMO C POR FERRAMENTAS DE BIOINFORMÁTICA / ERICLES CHARLES  
DA SILVA MELO. - 2022.  
50 f. : il.

Orientador: Diego de Souza .  
Inclui referências e anexo(s).

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco,  
Bacharelado em Ciências Biológicas, Serra Talhada, 2022.

1. Citocromo c. 2. Apoidea. 3. bioinformática. 4. filogenia. I. , Diego de Souza, orient. II. Título

CDD 574

---

**INFERÊNCIA FILOGENÉTICA DE APOIDEA (HYMENOPTERA) A  
PARTIR DA ANÁLISE DO GRAU DE HOMOLOGIA DO CITOCROMO  
C POR FERRAMENTAS DE BIOINFORMÁTICA**

Monografia submetida à coordenação do Curso de Bacharelado em Ciências Biológicas da Universidade Federal Rural de Pernambuco/Unidade Acadêmica de Serra Talhada (UFRPE/UAST) como requisito obrigatório para obtenção do grau de bacharel em Ciências Biológicas.

Data da defesa: 01/06/2022.

**BANCA EXAMINADORA**

---

Prof. Dr. Diego de Souza Buarque (Presidente/Orientador)  
UFRPE/UAST

---

Prof. Dr. Rogério de Aquino Saraiva (UFCA)  
(2° TITULAR)

---

Prof. Dr. Airton Torres Carvalho (UFERSA)  
(3° TITULAR)

Serra Talhada – PE  
2022

*Dedico esse trabalho à ciência brasileira, que permanece de pé mesmo com os ataques deste atual governo.*

## Agradecimentos

Agradeço,

Grandemente ao meu orientador de pesquisa Airton Torres Carvalho por me apresentar e despertar a paixão pelas abelhas, pelo conhecimento compartilhado, pelos incentivos durante os 04 anos de convivência, por abrir as portas de sua casa, pelos ensinamentos e por toda a paciência e compreensão.

Ao meu orientador de monografia Rogério de Aquino Saraiva, por me auxiliar em um dos momentos mais questionáveis de minha vida acadêmica, por sua paciência como pessoa e como professor, por suas contribuições diretas e indiretas em minha vida.

À Universidade Federal Rural de Pernambuco/Unidade Acadêmica de Serra Talhada pela oportunidade de viver um curso tão maravilhoso, apesar dos percalços ainda assim agradeço profundamente;

Ao GEEP (Grupo de Estudos em Ecologia e Polinização) por toda a ajuda e toda a amizade durante esse tempo, e que possamos continuar trabalhando juntos por muito mais tempo;

A *SoEnglish* e ao seu CEO Jeffeson Sobral, pela sua amizade, pela bolsa concedida no curso, que tem sido tão importante e tão útil na minha construção profissional.

Especialmente à minha esposa Erika Vanessa por estar sempre presente, por ser meu barco, minha âncora e meu porto seguro, espero que o tempo nos permita estar juntos por muitos longos anos. Te amo!

Ao meu amigo Murilo Inácio por ter me acompanhado nesses 19 anos de amizade, pelos conselhos, pela ajuda prestada em todos os momentos em que eu precisei. Agradeço também a sua mãe, Lucy Alves, por me aconselhar e guiar em situações difíceis.

À minha amiga, Maria Beatriz, por estar presente em meus momentos de choro, de tristeza, mas também em meus momentos de felicidade genuína em que contraio meu zigomático superior. Espero que o tempo te guarde em um potinho e me permita ser seu amigo para o resto de minha vida.

Ao meu quarteto fantástico, Allysson Vinicius, Eduarda Ferreira, Lyvia Albino e eu, por formamos o melhor grupo de fofocas, pelos melhores momentos de nossa amizade, que o tempo seja generoso o bastante para nos conceder muito mais.

A todos os meus familiares que acreditaram em mim e no meu potencial nos momentos em que fui falho, em especial a minha mãe, Patrícia Moreira, e minha tia, Conceição Moreira, pelo empenho em me ajudar na vida e me apoiar nesse sonho.

Ao Pontos das Cópias e funcionários, especialmente ao seu CEO Wandercleidson Gomes, pela oportunidade dada de trabalharmos lado a lado durante esses 05 anos, pela amizade que construímos, por ter sido o pai que sempre desejei, e tornou-se real. Agradeço profundamente.

A minha turma (2017.1) que esteve comigo durante toda a graduação, e onde fiz amigos que com certeza irei levar para vida toda. Pelos "rolês", pelas preocupações, pelos amparos, pelos copos de nevada, pelos momentos felizes que passamos juntos. Não esquecerei vocês de forma alguma.

Não poderia deixar de agradecer a comunidade científica por continuar lutando em meio aos ataques desse governo atual, que continua de pé contra o negacionismo, e que continuemos lutando para que mais pessoas como eu, de cidades do interior e do campo, tenham acesso ao ensino superior público, inclusivo e de qualidade.

E por fim agradeço a todos que de alguma forma estiveram presentes durante essa minha jornada, meu muito obrigado!

"Nada na Biologia faz sentido exceto à luz da Evolução"  
(Theodosius Dobzhansky)

## RESUMO

A superfamília Apoidea (Hymenoptera) abriga cerca de 30.000 espécies registradas e é constituída por vários subgrupos de abelhas apoides e vespas, com tipos de hábitos diferentes desde o cleptoparitismo até a eussocialidade. As espécies dessa subfamília podem ser identificadas pela divisão corporal (cabeça, mesosoma e metassoma) e pela presença ou ausência de pelos sobre o corpo. Devido à grande dificuldade em estabelecer marcadores moleculares que possam inferir a filogenia deste grupo, objetivamos nesse trabalho verificar se o citocromo c, uma proteína essencial ao metabolismo oxidativo e produção de energia, altamente conservada nas espécies, pode cumprir satisfatoriamente com esse papel. O trabalho consistiu em duas abordagens: 1) a busca das sequências primárias do citocromo c de organismos da superfamília Apoidea disponíveis em banco de dados biológicos (NCBI Protein), para posterior alinhamento múltiplo e obtenção de uma árvore filogenética através do programa *MAFFT*; 2) comparação com as filogenias presentes na literatura. Foram obtidas sequências FASTA de 15 espécies, todas contendo 108 resíduos de aminoácidos. O cladograma obtido a partir do alinhamento proposto mostra que as tribos Bombini e Apini parecem formar um grupo-irmão. O estudo também mostrou que o ancestral em comum que origina a tribo Euglossini, também origina as tribos Bombini, Apini e Meliponini, assim demonstrando que Apini e Euglossini podem ser grupos parafiléticos. Apesar das poucas sequências de citocromo c de Apoidea disponíveis no NCBI Protein, foi possível observar que o cladograma parece andar de encontro com algumas propostas da literatura, sugerindo que o citocromo c parece ser promissor nesse propósito. No entanto, este estudo não propõe de imediato uma nova classificação, pois seria necessário analisar um número maior de sequências primárias de diferentes espécies de Apoidea. Portanto, são necessários mais estudos em variados grupos dentre os himenópteros para que seja possível utilizar do citocromo c como um possível marcador.

**Palavras-chave:** Citocromo c; Apoidea; bioinformática; filogenia.

## ABSTRACT

The Apoidea superfamily (Hymenoptera) is home to about 30,000 recorded species and consists of several subgroups of apoid bees and wasps, with different types of habits from kleptoparasit to eusociality. The species of this subfamily can be identified by body division (head, mesosoma and metasoma) and by the presence or absence of hair on the body. Due to the great difficulty in establishing molecular markers that can infer the phylogeny of this group, we aimed in this work to verify if cytochrome c, a protein essential to oxidative metabolism and energy production, highly conserved in species, can satisfactorily fulfill this role. The work consisted of two approaches: 1) the search for the primary sequences of cytochrome c of organisms of the Apoidea superfamily available in biological databases (NCBI Protein), for subsequent multiple alignment and obtaining a phylogenetic tree through the MAFFT program; 2) comparison with the phylogenies present in the literature. FASTA sequences were obtained from 15 species, all containing 108 amino acid residues. The cladogram obtained from the proposed alignment shows that the Bombini and Apini tribes seem to form a sister group. The study also showed that the common ancestor that gives rise to the Euglossini tribe also gives rise to the Bombini, Apini and Meliponini tribes, thus demonstrating that Apini and Euglossini can be paraphyletic groups. Despite the few Apoidea cytochrome c sequences available in the NCBI Protein, it was possible to observe that the cladogram seems to go against some proposals in the literature, suggesting that cytochrome c seems to be promising for this purpose. However, this study does not immediately propose a new classification, as it would be necessary to analyze a greater number of primary sequences from different Apoidea species. Therefore, more studies are needed in different groups among the Hymenoptera so that it is possible to use cytochrome c as a possible marker.

**Keywords:** Cytochrome c; Apoidea; bioinformatics; phylogeny.

## LISTA DE FIGURAS

<b>Figura 1-</b> Estrutura da mitocôndria .....	18
<b>Figura 2-</b> Estrutura molecular do citocromo c do <i>Equus caballus</i> , obtida no Protein Data Bank .....	19
<b>Figura 3-</b> Orientação do corpo hipotético de um Hymenoptera.....	23
<b>Figura 4-</b> Principais divisões morfológicas de uma abelha hipotética .....	23
<b>Figura 5-</b> Aminoácidos do sítio de ligação do citocromo c de <i>Equus caballus</i> (Código PDB 1AKK) obtidos a partir do programa LIGPLOT .....	27
<b>Figura 6-</b> Alinhamento múltiplo das sequências FASTA do citocromo c de <i>Apis mellifera</i> e <i>Equus caballus</i> (estrutura tridimensional de código PDB 1AKK) obtido a partir do Clustal Omega.....	28
<b>Figura 7-</b> Alinhamento múltiplo das sequências FASTA do citocromo c de diferentes organismos do clado Apoidea determinado a partir do Clustal Omega .....	29
<b>Figura 8-</b> Alinhamento múltiplo das sequências de citocromo c de organismos do clado Apoidea no Jalview (multiple sequence alignment editor and analysis workbench) .....	30
<b>Figura 9-</b> . Dendrograma/árvore filogenética do MAFFT (KATO, ROZEWICKI, YAMADA 2019).....	31
<b>Figura 10-</b> (A) Representação do resultado obtido pelo trabalho do BOSSERT ET AL. 2019; (B) Representação do resultado obtido pelo MAFFT .....	32
<b>Figura 11-</b> Representação da árvore com respectivas famílias .....	33
<b>Figura 12-</b> Introdução no MAFFT – Inserido sequências.....	43
<b>Figura 13-</b> Introdução no MAFFT – Submetendo sequências.....	44
<b>Figura 14-</b> Introdução no MAFFT – Obtenção do alinhamento.....	45
<b>Figura 15-</b> Introdução no MAFFT – Obtenção da árvore.....	46
<b>Figura 16-</b> Introdução no MAFFT – Configuração da árvore.....	47
<b>Figura 17-</b> Introdução no MAFFT – Visualização da árvore.....	48
<b>Figura 18-</b> Introdução no MAFFT – Reroot da árvore.....	49

## LISTA DE TABELAS

<b>Tabela 1-</b> Classificação de Apoidea por Michener (2007) .....	21
<b>Tabela 2-</b> Organismos do clado Apoidea utilizados para obtenção da sequência FASTA do citocromo c no NCBI Protein.....	26
<b>Tabela 3-</b> Resíduos de aminoácidos do citocromo c envolvidos nas interações intermoleculares com o grupo heme em Apoidea.....	29

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	14
<b>2. REFERÊNCIAL TEÓRICO</b> .....	16
2.1 Filogenia e Relógio Molecular.....	17
2.2 Bioinformática.....	17
2.3 O Citocromo C.....	18
2.4 Hymenoptera: Apoidea.....	20
2.4.1 Visão sobre a classificação de Apoidea.....	20
2.4.2 Características morfológicas gerais dos Apoidea.....	22
<b>3. MATERIAIS E MÉTODOS</b> .....	24
3.1 Obtenção da sequência de citocromo C de interesse.....	24
3.2 Obtenção do alinhamento das sequências de citocromo C de interesse.....	24
3.3 Obtenção do dendrograma/árvore filogenética.....	25
<b>4. RESULTADOS E DISCUSSÃO</b> .....	26
4.1. Sequência de citocromo C.....	26
4.2. Análise comparativa estrutural do citocromo c em Apoidea.....	26
4.3. Dendrograma/Árvore filogenética.....	31
<b>5. CONSIDERAÇÕES FINAIS</b> .....	33
<b>6. REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	34
<b>ANEXO</b> .....	40

## 1. INTRODUÇÃO

**Apoidea** é uma superfamília que abriga cerca de 30.000 espécies registradas, constituída por vários subgrupos de abelhas apoídes e vespas, com tipos de hábitos diferentes que figuram desde o cleptoparismo até a eussocialidade (SANN et al., 2018). A monofilia de Apoidea é suportada por uma série de caracteres morfológicos, incluindo (i) a forma arredondada lobo pronotal bem separado da tégula, (ii) a extensão ventral do pronoto para cercar ou quase circundar o tórax atrás da coxa anterior, e (iii) aumento do metapostnoto (Michener 2007). Estudos acerca desses indivíduos são de grande importância para conservação do meio ambiente, delinear parâmetros ecológicos e compreender os serviços ambientais. As abelhas, por exemplo, são importantes no equilíbrio ecológico, sendo responsáveis pela polinização de cerca de 70% das plantas floríferas, além de proverem produtos como cera e mel de considerável valor econômico (Ollerton et al 2011).

A bioinformática tem contribuído de maneira significativa para o armazenamento e análise de dados biológicos moleculares, melhorando significativamente o entendimento de várias questões em biologia. Como exemplo mais prático, podemos citar a elucidação do genoma humano, divulgado em 2001, com milhões de bases nitrogenadas, no qual uma infinidade de dados foi gerada, se fazendo necessário que houvesse uma boa ferramenta de armazenamento e organização de dados. Nesse sentido, a bioinformática surge como um suporte organizacional e de armazenamento, diminuindo os riscos de perda. Muito além disso, a bioinformática utiliza esses dados para auxiliar em estudos filogenéticos, tornando possível através do mesmo sanar dúvidas acerca das classificações atuais da biodiversidade que carecem de estudos mais aprofundados, em nível molecular (LEAL, 2004).

O citocromo c é uma molécula fortemente conservada, e preserva sua função, encontrada na célula, precisamente na mitocôndria, participando da cadeia de transporte de elétrons, refletindo o caráter de origem evolutiva nos indivíduos (LEHNINGER et al, 2014). O caráter evolutivo significa mudança ao longo das gerações, em todos os níveis, desde sequências de DNA até a morfologia externa e o comportamento social, e que se podem ser modificadas a partir dos seus ancestrais durante a evolução (RIDLEY, 2006), sendo assim

com a baixa taxa de modificação do citocromo c é possível através desta informação fazer um comparativo de várias sequências de proteínas de determinados organismos, possibilitando o alinhamento, e, portanto, compartilhamento evolutivo, bem como obter uma árvore filogenética.

Em artigo publicado na *Science* em 1978, Robert Schwartz e Margareth Dayhoff apresentaram árvores filogenéticas mostrando as relações de parentesco entre vários organismos diferentes (procariotos e eucariotos) obtida a partir da análise do alinhamento múltiplo das sequências de algumas proteínas, incluindo o citocromo c (SCHARWITZ; DAYHOFF, 1978). Essa árvore apresenta forte correlação com as ideias evolutivas já propostas por outras abordagens metodológicas, validando a ideia de que ferramentas de bioinformática podem ser úteis para um melhor entendimento destas questões.

Devido à grande dificuldade em estabelecer marcadores moleculares que possam sanar as relações filogenéticas de Apoidea, o objetivo foi verificar se o citocromo c pode ser um possível marcador molecular para este propósito. Diante disso, buscamos obter sequências primárias do citocromo c das espécies pertencentes ao clado Apoidea que estão disponíveis no banco de dados, para assim fazer um comparativo com as informações cladísticas presentes na literatura atual.

## 2 REFERENCIAL TEÓRICO

### 2.1 Filogenia e Relógio Molecular

Um dos primeiros estudiosos a criar um sistema que estabelece um critério de organizar a diversidade biológica em categorias foi Carolus Linnaeus (ou Carl Linné). Linnaeus criou uma série de categorias taxonômicas que segundo Amorim (1994) foram: Espécie, Ordem, Classe e Reino. Porém, a visão de Linnaeus era essencialista tipológica, ou seja, procuravam características morfológicas que fossem compartilhados por uma ou mais espécies e a partir da definição de um *tipus*, ou característica perfeita. Variações dessas características eram então agrupadas formando amplos grupos de organismos, criando uma classificação hierarquizada (MOREIRA, 2009). Séculos após a classificação de Linnaeus, surgiu a Teoria da Evolução de Charles Darwin. Essa Teoria tinha como hipótese a ideia de que todos os indivíduos descendiam de um ancestral comum, isso fez com que vários taxonomistas utilizassem o sistema de Linnaeus, adaptando-o com as relações filogenéticas. Foi através das análises genéticas de RNA ribossômico 16S, que Carl Woese em 1977 dividiu os seres vivos fundamentalmente em três domínios: Bacteria, Archaea e Eukarya (LESK, 2014).

Na década de 1950, com a criação da eletroforese em amido, que consiste na migração de micromoléculas sob influência de um campo elétrico, muitos estudos acerca das proteínas surgiram. Alguns anos depois, surgiu a primeira sequência completa da proteína insulina, através desse sequenciamento foi possível comparar a insulina de diversos indivíduos (boi, ovelha, porco, cavalo e baleia) e perceber que haviam substituições em aminoácidos em diferentes partes da proteína, entretanto parecia não ter efeito em sua atividade. Assim a Era das Proteínas teve seu auge na década de 1960, quando já havia o conhecimento de várias sequências de proteínas e sua análise proveu dados moleculares capazes de inferir sobre a evolução dos grupos biológicos. Através desses avanços na ciência moderna foi-se capaz de propor hipóteses de relações evolutivas entre ordens, classes, filos, reinos baseando-se em filogenia molecular. As ideias revolucionárias de Willi Hennig (1950), foram essenciais para o desenvolvimento do pensamento filogenético (cladísticas), especialmente depois da tradução para o inglês de seu livro em 1966. A proposição fundamental

da ideia de Henning é de que uma vez surgido, um caráter permanece em todos os seus descendentes, o que permite inferir o grau de parentesco evolutivo entre espécies. Ele definiu os termos até hoje utilizados como apomorfia, plesiomorfia, monofilia, parafilia, etc, e propôs que a evolução era parcimoniosa, propondo então um método simples e eficaz para construção de árvores filogenéticas (AMORIN,2002).

Muitas proteínas são “fosseis vivos”, em que suas estruturas foram conservadas pelo processo evolutivo ao longo dos bilhões de anos, evoluindo de sequências ancestrais comuns por um grande número de pequenas mudanças (SCHWARTZ; DAYHOFF, 1978). Para estimar as distâncias dessas mudanças, Emile Zuckerkandl e Linus Pauling (1962) criaram a Teoria do Relógio Molecular, que leva em consideração o número de substituições moleculares medidas através das sequências de proteínas ou DNA supondo que a taxa de substituições é constante, ou seja, quanto mais similares geneticamente, menor o tempo de separação entre as espécies (SOLÉ-CAVA; SILVA; LÔBO-HAJDU, 2010).

Essa teoria foi baseada em análises de moléculas de hemoglobina, em que as taxas evolutivas eram aproximadamente constantes, o que foi corroborado por Margoliash (1963) que fez testes com os aminoácidos do citocromo c de vários organismos, percebendo que este fenômeno ocorria nesta proteína (ARAUJO, 2022). Contudo, alguns autores se posicionam contra essa ideia pois dizem que é biologicamente improvável que as taxas de substituição permaneçam constantes em linhagens que são filogeneticamente distantes (RIDLEY, 2006).

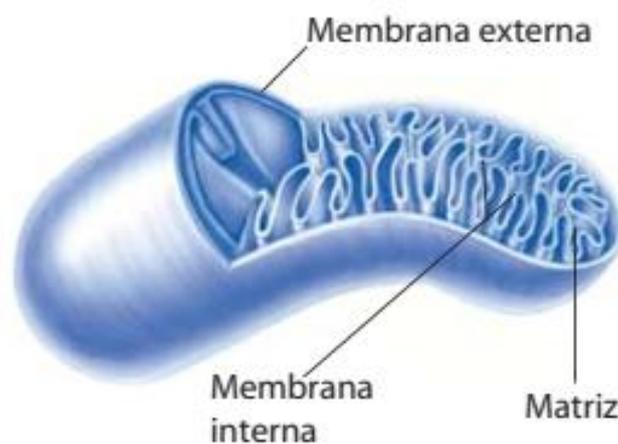
## **2.2 Bioinformática**

Pode-se dizer que a bioinformática teve seu começo quando em 1950, James Watson e Francis Crick publicaram na Revista *Nature* o famoso trabalho sobre a estrutura em hélice do DNA. Outro ponto histórico que muito contribuiu foram os trabalhos de Linus Pauling e Robert Corey no início dos anos 1950, bem como o trabalho de Gopalasamudram N. Ramachandran no início dos anos 1960, que forneceram as bases para a compreensão da estrutura tridimensional das proteínas (VERLI,2014).

Na década de 1970, quanto mais sequências de DNA e proteínas estavam disponíveis, cresceu o interesse em desenvolver programas de computador para analisar essas sequências de várias maneiras. A *Nucleic Acids Research* publicou duas edições especiais em 1982 e 1984 dedicadas ao uso de computadores para análise de sequências, incluindo programas para computadores *mainframe* acompanhado o desenvolvimento do processamento dos microcomputadores. Pouco depois, o *Genetics Computer Group* (GCG) foi fundado na Universidade de Wisconsin por J. Devereux, oferecendo um conjunto de programas para análise que rodavam em um computador VAX (MOUNT, 2004).

### 2.3 O citocromo c

A mitocôndria é um dos componentes celulares mais importantes, uma vez que é responsável pelo metabolismo energético oxidativo, ou seja, pela respiração celular. Possui uma membrana interna e outra externa, há também uma parte aquosa em sua estrutura interna na qual denomina-se matriz mitocondrial (LAURENCE et al, 2013).

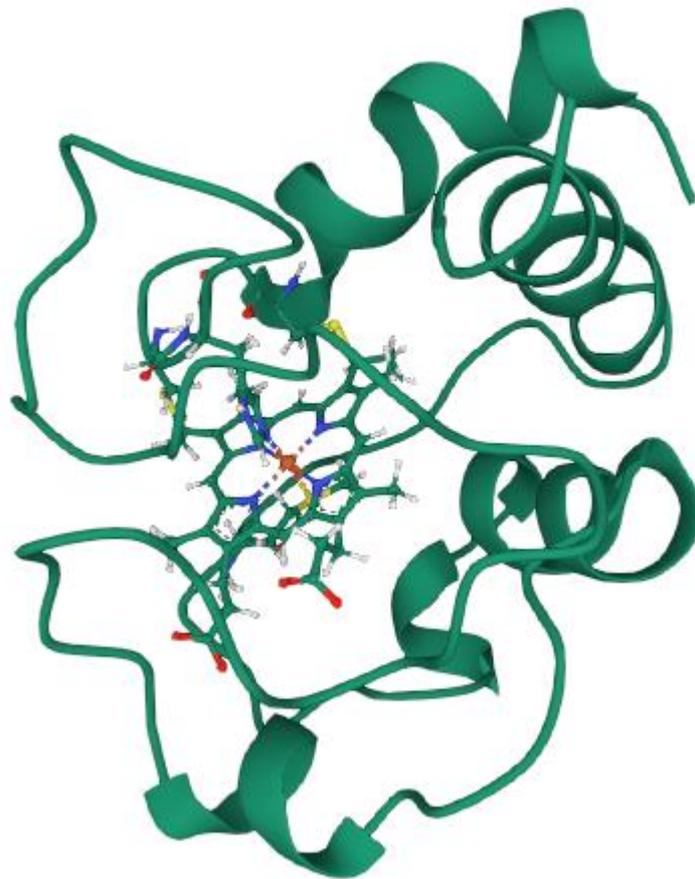


**Figura 1.** Estrutura da mitocôndria.

**Fonte:** LAURENCE, Moran et al. 2013

Na composição proteica da mitocôndria tem o que é comumente conhecido como citocromos. Os citocromos são coenzimas proteicas classificadas em diferentes tipos em função de sua cadeia polipeptídica e que garante sua propriedade distinta, classificados de maneira geral em citocromo *a*, *b* e *c*, esses apresentam um grupo ferro associado à porfirina em sua estrutura, geralmente

nominados de grupo heme (SILVEIRA, 2012). O citocromo c está associado à superfície da membrana interna mitocondrial, e é uma proteína produto de um gene nuclear, pequena e altamente conservada que apresenta massa molecular de 12,4 kDa, e é responsável por desempenhar o transporte de elétrons (AUGUSTO, 2011).



**Figura 2.** Estrutura molecular do citocromo c do *Equus caballus*, obtida no Protein Data Bank (Código 1AKK). **Fonte:** BANCI, Lucia et al. 2022.

O citocromo c participa ativamente da cadeia transportadora de elétrons, contendo algumas cópias da proteína, são elas: citocromo c oxidorreductase, também chamado complexo do citocromo *bc1*, que catalisa a oxidação de moléculas de ubiquinol (QH<sub>2</sub>) na membrana e a redução de uma molécula móvel e hidrossolúvel de citocromo c na superfície externa, servindo como carreador

móvel, levando elétrons para o próximo complexo da cadeia. Como o citocromo é uma proteína transportadora de elétrons em diversas espécies, a conservação da estrutura da proteína é um reflexo de sua interação com o grupo heme presente nela e sua função conservada em quase todos os organismos. Sendo assim a estrutura do citocromo c de todas as espécies são muito semelhantes e as sequências de aminoácidos da cadeia polipeptídica são bem conservadas (LAURENCE et al,2013). Alguns termos na leitura das sequencias proteicas se culminaram, são elas: *regiões conservadas* que são aminoácidos idênticos das sequências analisadas; *regiões conservativas* aminoácidos que foram substituídos, porém não perderam sua função, e *regiões seminconservativas* que são aminoácidos substituídos e podem ou não mudar a configuração da função proteica.

A filogenia baseada na análise de homologia do citocromo c ajuda a entender como se originou a mitocôndria e como ela evoluiu. Como o citocromo c é codificado no núcleo, mas funciona na mitocôndria, essa relação é explicada na teoria simbiótica pela transferência de informação genética, o que inclui o gene do citocromo c, da bactéria invasora aeróbia para o hospedeiro protoeucarioto durante o desenvolvimento de seu relacionamento (SCHWARTZ; DAYHOFF, 1978).

## **2.4 Hymenoptera: Apoidea**

### **2.4.1 Visão sobre a classificação de Apoidea**

Embora visualmente falando abelhas apóides e vespas esfecóides sejam grupos com diferenças, eles compartilham relações evolutivas muito próximas. Pode-se dizer que abelhas são vespas, cujo indivíduos, em vez de caçar predatoriamente, coletam pólen e néctar nas flores para alimentarem suas proles (SILVEIRA; MELO; ALMEIDA, 2002). Em uma hipótese de Sann et al, 2018, a família Ammoplanidae possivelmente representa a linhagem irmã existente das abelhas. As espécies de Ammoplanidae são conhecidas por caçar tripes, dos quais alguns se agregam em flores e se alimentam de pólen. A biologia específica de Ammoplanidae como predadores indica como a transição de uma vida predatória para uma vida de coleta de pólen estilo poderia ter ocorrido na evolução das abelhas.

As abelhas e vespas esfecóides são classificadas como uma superfamília há muito tempo. Em seu livro *An introduction to entomology*, publicado em 1925, John Henry Comstock combinou esses dois grandes grupos em uma superfamília conhecida como Sphecoidea. Alguns autores no passado separavam esses indivíduos em duas superfamílias Sphecoidea e Apoidea (MICHENER, 2007). Em trabalhos de Brothers e Carpenter (1993) defendem colocar as vespas esfecóides e as abelhas na mesma superfamília, tornando-os um grupo monofilético. O que segundo Michener (2007) levou autores a reconhecer apenas duas famílias principais no clado Apoidea, sendo eles: Sphecidae e Apidae. Entretanto, esses são grupos muito diversos, como é proposto por Michener (2007) a superfamília é subdividida em:

**Tabela 1.** Classificação de Apoidea por Michener (2007).

Fonte: MICHENER, C. D. 2007.

<p><b>Família Ampulicidae</b></p> <p><b>Família Crabronidae</b></p> <p><b>Família Sphecidae</b></p>	<p><b>Vespas esfecóides</b></p>
<p><b>Família Andrenidae</b></p> <p><b>Família Apidae</b></p> <p><b>Família Colletidae</b></p> <p><b>Família Halictidae</b></p> <p><b>Família Megachilidae</b></p> <p><b>Família Melittidae</b></p> <p><b>Família Stenotritidae</b></p>	<p><b>Abelhas</b></p>

Melo (1999) reconheceu, além das abelhas, os grupos de vespas em Apoidea: Heterogynaidae, Ampulicidae, Sphecidae *s.str.* e Crabronidae. Em publicação recente no Catálogo de Sphecidae sensu lato (= Apoidea excluindo Apidae), é proposto por Pulawski (2021) que esse grupo deve ser dividido em: Heterogynaidae, Ampulicidae, Sphecidae *s.str.* e Crabronidae.

Em estudos de Debevec et al, 2012 a superfamília Apoidea inclui as abelhas (Anthophila), bem como quatro famílias de vespas predadoras (Heterogynaidae, Ampulicidae, Sphecidae e Crabronidae) coletivamente referidas como as vespas “esfecoides” ou “apoides”. O autor diz que os estudos mais citados sugerem que as abelhas são irmãs da família de vespas Crabronidae, entretanto algumas hipóteses foram propostas com base em dados

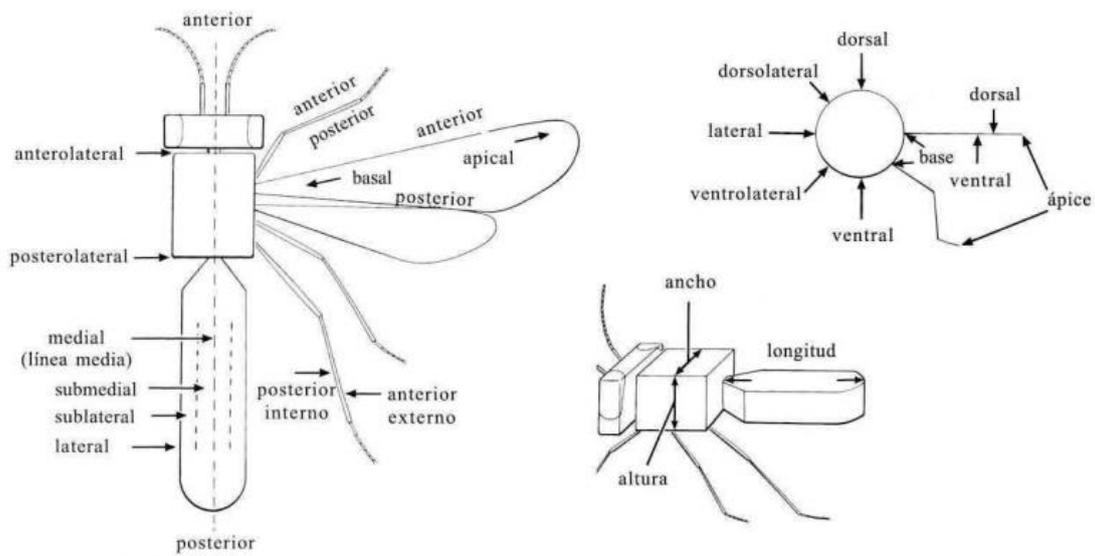
morfológicos e moleculares. Os resultados do estudo de Debevec et al, 2012 mostram que os possíveis grupos irmãos das abelhas incluem Philanthinae Pemphredoninae ou Philanthinae + Pemphredoninae (Crabronidae). Além de trazer um suporte adicional para a hipótese de que as abelhas surgem de Crabronidae, em vez de serem irmãs de Crabronidae.

O Danforth, 2013 fala que a compreensão da filogenia desse grupo melhorou nos últimos anos através de novos dados, principalmente dados de sequência de nucleotídeos e novos métodos, principalmente métodos baseados em modelos de reconstrução de filogenia ajudaram resolver a colocação de abelhas dentro da superfamília Apoidea.

#### **2.4.2 Características morfológicas gerais de Apoidea**

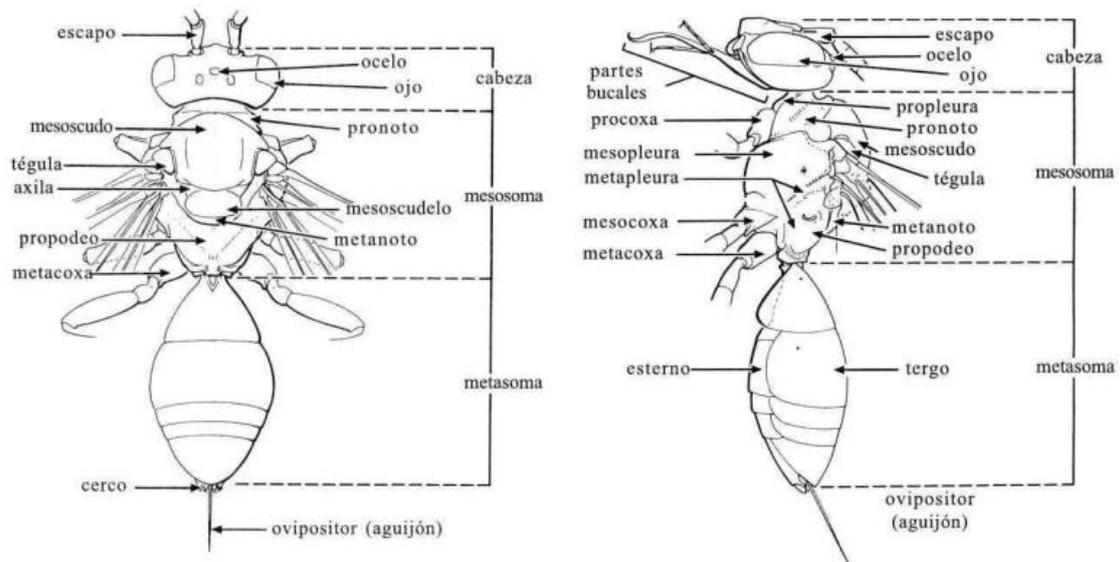
O corpo de Apoidea, assim como os corpos de outros insetos, é formado por unidades denominadas metâmeros. Esses segmentos são divididos em três categorias chamadas de tagmas, são elas: cabeça, mesosoma e metassoma (Fig. 3 e 4). Esse táxon possui algumas características compartilhadas, tais sinapomorfias garantem que esses indivíduos sejam agrupados, são elas: Pronoto com ângulo posterolateral reduzido acima do lóbulo espiracular; Pronoto com um par de carenas laterais em ângulo oblicuo; Metapostnoto expandido posteriormente para formar o triângulo propodeal;

Muitas abelhas têm um par de linhas impressas no propódeo, começando em seus dorsolaterais anteriores e estendendo-se para baixo e posteromedial, quase atingindo a fossa propodeal, uma depressão mediana da face pósteroinferior. Essas linhas, juntamente com a margem dorsal anterior do propódeo, completam o triângulo propodeal. Este triângulo é conhecido como metapostnoto em morfologia (MICHENER apud BROTHERS, 1976).



**Figura 3.** Orientação do corpo hipotético de um Hymenoptera.

Fonte: FERNÁNDEZ, F.; SHARKEY, M. J. 2006.



**Figura 4.** Principais divisões morfológicas de uma abelha hipotética.

Fonte: FERNÁNDEZ, F.; SHARKEY, M. J. 2006.



### 3. MATERIAL E MÉTODOS

#### 3.1. Obtenção da sequência de citocromo c de interesse

A busca pelas sequências de citocromo c foi realizada no banco de dados do *National Center for Biotechnology Information* (NCBI) <<https://www.ncbi.nlm.nih.gov/>>. Como critério de exclusão, foram buscados organismos com percentual de identidade igual ou superior a 86,24% comparado à sequência FASTA do citocromo c de *Apis mellifera* ou *Melipona quadrifasciata*, a partir da ferramenta *Basic Local Algorithm Search Tool* (BLAST) <<https://blast.ncbi.nlm.nih.gov/Blast.cgi>> que fazem parte do clado Apoidea. Optou-se por analisar apenas os organismos que apresentavam todos os 108 resíduos de aminoácidos.

#### 3.2 Obtenção do alinhamento das sequências de citocromo C de interesse no *Clustal Omega*

As sequências obtidas foram submetidas ao site *Clustal Omega* (<https://www.ebi.ac.uk/Tools/msa/clustalo/>), que também é um software de alinhamento de múltiplas sequências. Na *homepage*, após a codificação das sequências, essas sequências são colocadas em uma caixa de texto. Para conseguir fazer o alinhamento foi escolhido nas configurações avançadas a estratégia de análise automática, isto fez com que o software reconhecesse qual melhor estratégia para as sequências. O alinhamento apresentou as características: Regiões conservadas, conservativas e semiconservativas, representadas respectivamente pelos símbolos: (\*); (:); (.). (Ver fig. 10).

Para verificar se o sítio de ligação do grupo prostético heme é altamente conservado em Apoidea, realizamos o alinhamento das sequências FASTA do citocromo c de *Apis mellifera* e da estrutura tridimensional de *Equus caballus* depositada no Protein Data Bank (PDB), de código 1AKK (essa estrutura foi identificada como a mais próxima de *Apis mellifera*, de acordo com o BLAST). A partir do programa LIGPLOT implementado no PDBSum <<http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl>>, foi

possível identificar os aminoácidos que fazem parte do sítio de ligação do grupo prostético heme em *Equus caballus*, para posterior análise do grau de homologia nas demais sequências alinhadas de Apoidea.

### **3.3 Obtenção de dendrograma/árvore filogenética no MAFFT**

O software online MAFFT<sup>1</sup> (*Multiple Alignment Program for amino acid or nucleotide sequences*) foi utilizado para a análise filogenética. Após a aquisição do alinhamento das sequências, para gerar o dendrograma/árvore filogenética, foi obtido pela opção: *Phylogenetic Tree*. em configurações avançadas o método escolhido foi o *Neighbor Joining* (NJ) que se baseia no princípio de evolução mínima a cada união de nós vizinhos (SOUZA, 2010) (*Ver fig. 14*). Na visualização, foi utilizado a configuração *Reroot* (*Ver fig. 16*) na espécie considerada mais distante, assim o software pôde reconfigurar a árvore de acordo com a distância por meio do relógio molecular. Todas as configurações padrão do software MAFFT para a determinação da árvore filogenética foram adotadas neste estudo.

---

<sup>1</sup> <https://mafft.cbrc.jp/alignment/server/>

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Sequências de citocromo C

Tendo como base os organismos com percentual de identidade igual ou superior a 86,24% comparado à sequência FASTA do citocromo c de *Apis mellifera* ou *Melipona quadrifasciata*, foram reunidas 15 espécies, prevalecendo aquelas com resíduos de 108 aminoácidos (Tabela 2).

**Tabela 2.** Organismos do clado Apoidea utilizados para obtenção da sequência FASTA do citocromo c no NCBI Protein. Fonte: DADOS DO AUTOR, 2022.

Organismo	Número de acesso (NCBI Protein)	Superfamília/Família*
<i>Ampulex compressa</i>	KAG7208637.1	Apoidea/Ampulicidae
<i>Apis dorsata</i>	XP_006612971.1	Apoidea/Anthophila/Apidae
<i>Apis florea</i>	XP_003692897.1	Apoidea/Anthophila/Apidae
<i>Apis laboriosa</i>	XP_043791397.1	Apoidea/Anthophila/Apidae
<i>Apis mellifera</i>	NP_001170961.1	Apoidea/Anthophila/Apidae
<i>Bombus terrestris</i>	XP_003394648.1	Apoidea/Anthophila/Apidae
<i>Ceratina calcarata</i>	XP_017890006.1	Apoidea/Anthophila/Apidae
<i>Colletes gigas</i>	XP_043260830.1	Apoidea/Anthophila/Colletidae
<i>Eufriesea mexicana</i>	XP_017759263.1	Apoidea/Anthophila/Apidae
<i>Frieseomelitta varia</i>	XP_043513165.1	Apoidea/Anthophila/Apidae
<i>Megachile rotundata</i>	XP_003701289.1	Apoidea/Anthophila/Megachilidae
<i>Melipona quadrifasciata</i>	KOX69899.1	Apoidea/Anthophila/Apidae
<i>Nomia melanderi</i>	XP_031827659.1	Apoidea/Anthophila/Halictidae
<i>Osmia bicornis bicornis</i>	XP_029042037.1	Apoidea/Anthophila/Megachilidae
<i>Osmia lignaria</i>	XP_034187101.1	Apoidea/Anthophila/Megachilidae

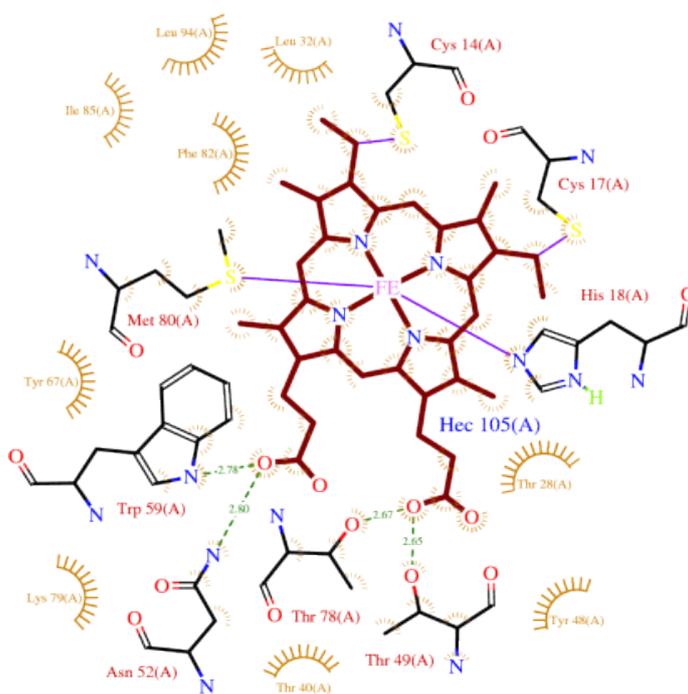
Como resultado da tabela obtida é possível perceber que, através do código FASTA que reúne os organismos mais próximo com base na porcentagem comparativa. No anexo A são apresentadas as sequências FASTA usadas para este estudo.

### 4.2 Análise comparativa estrutural do citocromo c em Apoidea

O resultado do alinhamento múltiplo obtido pelo Clustal Omega demonstra que o citocromo c em Apoidea é altamente conservado, uma vez que se que se

notou que 82 aminoácidos presentes nessas sequências são conservados (75,92%), 13 aminoácidos são conservativos (12,04%) e 13 aminoácidos são semiconservativos (12,04%). Esse dado corrobora com o trabalho de Margoliash (1963) (apud ARAUJO, 2022) quando fez alinhamento do citocromo c de vários indivíduos de diferentes espécies e observou alto grau de homologia.

Na figura 5 temos os aminoácidos do sítio de ligação do grupo prostético heme na estrutura tridimensional do citocromo c de *Equus caballus* (código PDB 1AKK) obtido a partir do programa LIGPLOT. A partir desses dados, identificamos no alinhamento com a sequência FASTA do citocromo c de *Apis mellifera* todos os aminoácidos que também interagem com a molécula de heme, os quais estão destacados na figura 5. A partir dessa análise, foi possível verificar que dos 17 aminoácidos do sítio de ligação, 15 são conservados (88,24%), 1 é conservativo (Ile por Leu, 5,88%) e 1 é semiconservativo (Thr por Val, 5,88%). Essas pequenas mudanças não parecem alterar a função da proteína, uma vez que são resíduos de aminoácidos envolvidos em interações de Van der Waals com o grupo prostético (para manter a estabilidade do complexo) e não estão envolvidos diretamente com as reações de oxirredução ou com o transporte de elétrons. Esses aminoácidos reforçam que a interação com o grupo heme permaneceu conservada mesmo após várias mudanças ao longo da evolução.



**Figura 5.** Aminoácidos do sítio de ligação do citocromo c de *Equus caballus* (Código PDB 1AKK) obtidos a partir do programa LIGPLOT.

CLUSTAL O(1.2.4) multiple sequence alignment

```

Equus_caballus  ----GDVEKGGKIFVQKCAQCHTVEKGGKHKGPNHGLFGRKGGQAPGFTTDANKNK  55
Apis_mellifera  MGIPAGDPEKGGKIFVQKCAQCHTIESGGKHKGPNYGVYGRKGGQAPGYSITDANKGK  60
                ** *****:*.*****.***:*.:*****:*****.*

Equus_caballus  GITWKEETLMELENPKKYIPGTKMIAGKKKTEREDYIAYLKKATNE  104
Apis_mellifera  GITWNKETLFELENPKKYIPGTKMVAGKKPQERADYIAYIEQASK-  108
                ****:***.*****:***:** ** *****:*.:

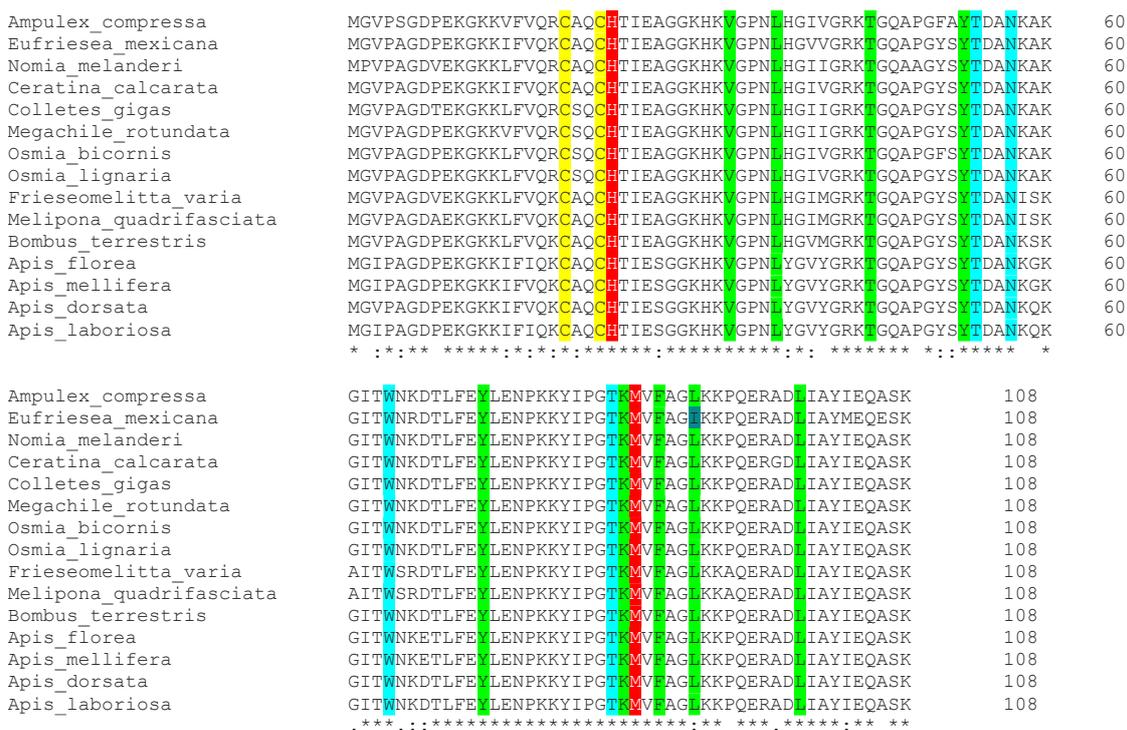
```

**Figura 6.** Alinhamento múltiplo das sequências FASTA do citocromo c de *Apis mellifera* e *Equus caballus* (estrutura tridimensional de código PDB 1AKK) obtido a partir do Clustal Omega. Os resíduos de aminoácidos destacados estão envolvidos nas interações intermoleculares com o grupo prostético heme: Em verde, resíduos de aminoácidos envolvidos em interações hidrofóbicas ou forças de Van der Waals; em azul, resíduos de aminoácidos envolvidos em ligações de hidrogênio; em amarelo, cisteínas ligadas ao grupo heme e em cinza, resíduos envolvidos em interação com o íon ferro (Fe). (\*): resíduos conservados; (:): resíduos conservativos e (.) resíduos semiconservativos.

**Fonte:** Dados do autor.

Considerando o clado Apoidea, 16 aminoácidos são idênticos e conservados (figura 6 e Tabela 3), com exceção da Leu90 que em *E. mexicana* aparece uma substituição por Ile90. No entanto, essa substituição não acarreta em alterações funcionais significativas, uma vez que esses dois aminoácidos possuem cadeia lateral hidrofóbica e ramificada, mantendo a estabilidade do heme no citocromo c. Ressalta-se também que todos os resíduos de aminoácidos envolvidos em ligações de hidrogênio e em interações com o cofator ferro (Fe) são conservados em todas as espécies analisadas (Figura 6 e tabela 3).

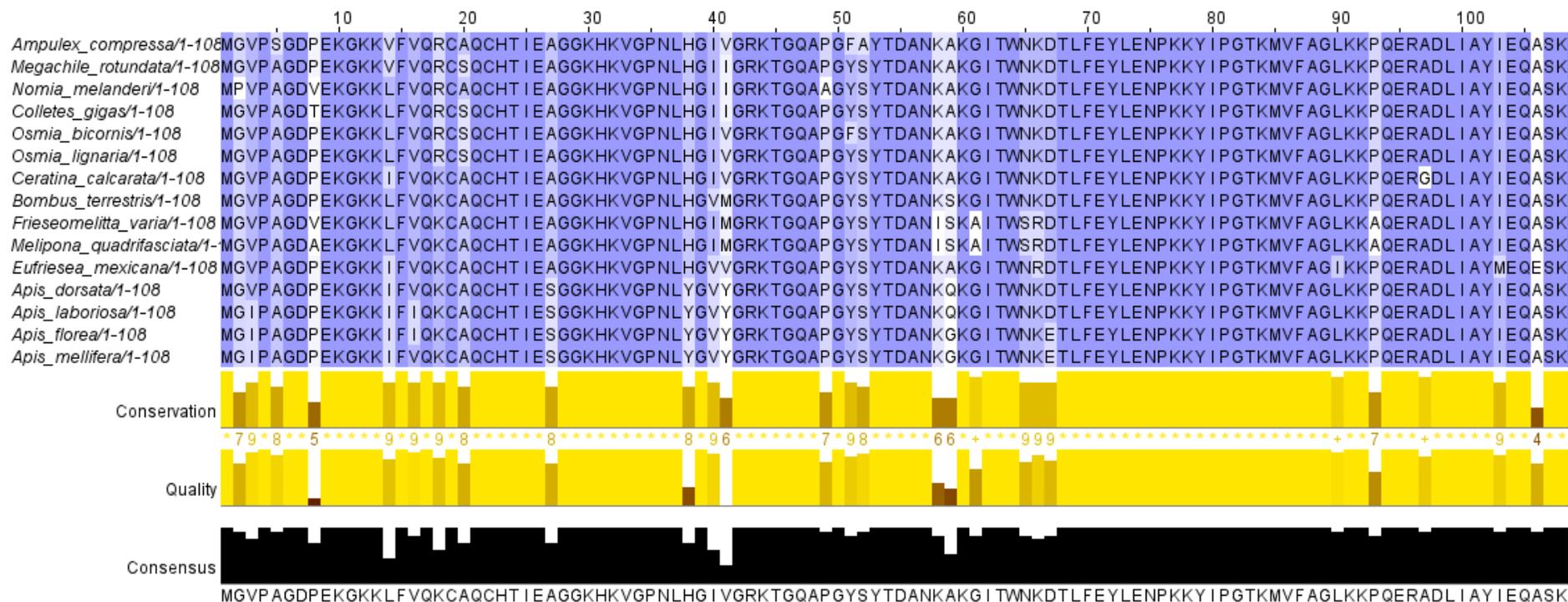
CLUSTAL O(1.2.4) multiple sequence alignment



**Figura 7.** Alinhamento múltiplo das sequências FASTA do citocromo c de diferentes organismos do clado Apoidea determinado a partir do Clustal Omega. Os resíduos de aminoácidos destacados estão envolvidos nas interações intermoleculares com o grupo prostético heme: Em verde, resíduos de aminoácidos envolvidos em interações hidrofóbicas ou forças de Van der Waals; em azul, resíduos de aminoácidos envolvidos em ligações de hidrogênio; em amarelo, cisteínas ligadas ao grupo heme e em vermelho, resíduos envolvidos em interação com o íon ferro (Fe). (\*) resíduos conservados; (.) resíduos conservativos e (.) resíduos semiconservativos. **Fonte:** Dados do autor, 2022.

**Tabela 3.** Resíduos de aminoácidos do citocromo c envolvidos nas interações intermoleculares com o grupo heme em Apoidea. O número após o aminoácido indica a posição deste na sequência primária.

Interações intermoleculares com o grupo heme	Aminoácidos (em Apoidea)
Ligação de hidrogênio	Thr54, Asn57, Trp64, Thr83
Interação com íon Fe	His23; Met85
Cisteínas ligadas ao grupo heme	Cys19 e Cys22
Interações hidrofóbicas ou de Van der Waals	Val33, Leu37, Thr45, Tyr53, Tyr72, Lys84, Phe87, Leu90 (Ile90), Leu99.

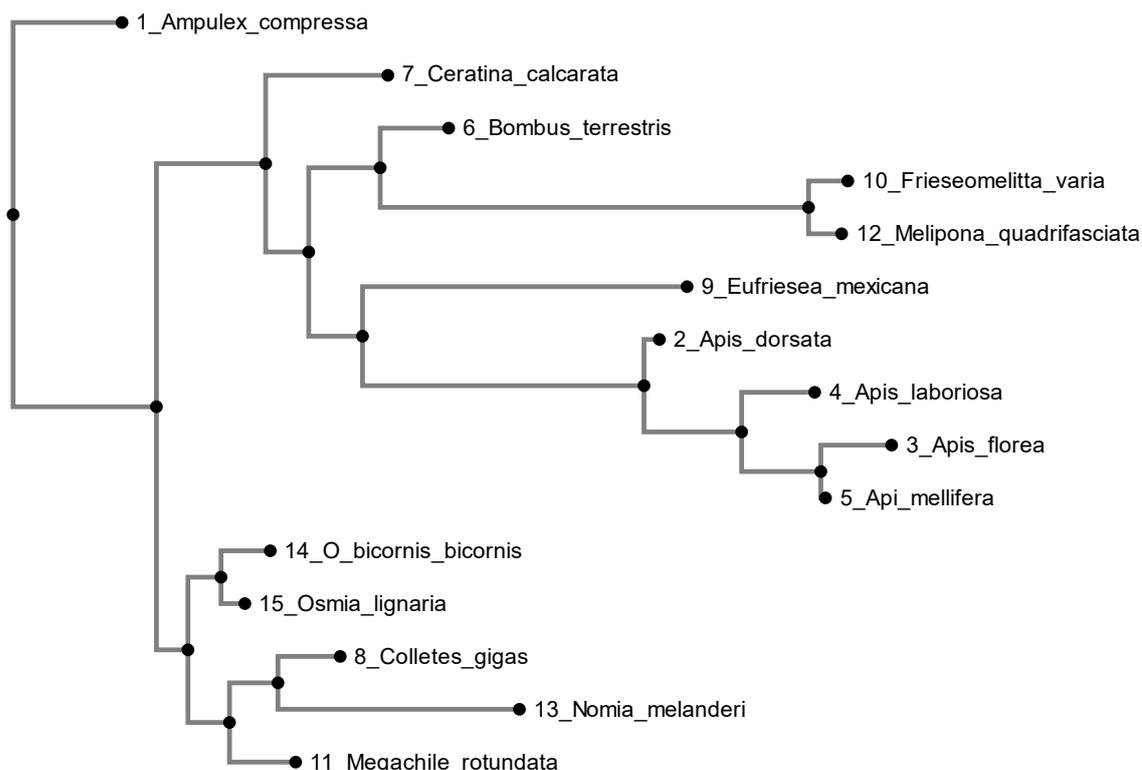


**Figura 8.** Alinhamento múltiplo das sequências de citocromo c de organismos do clado Apoidea no Jalview (multiple sequence alignment editor and analysis workbench) . Em azul escuro, são destacados os resíduos conservados.

Fonte: Dados do autor, 2022.

### 4.3 Dendrograma/ Árvore filogenética

Na figura 9 é apresentada a árvore filogenética obtida no programa MAFFT. Para facilitar a análise dessa árvore, foi escolhido o grupo mais basal como grupo-externo (Ampulicidae) representado pela espécie *Ampulex compressa*, assim como constatado no trabalho de Bransteeter et al. 2017, Ampulicidae pode ser considerado um grupo-irmão dos Anthophila (Abelhas e vespas esfecóides).



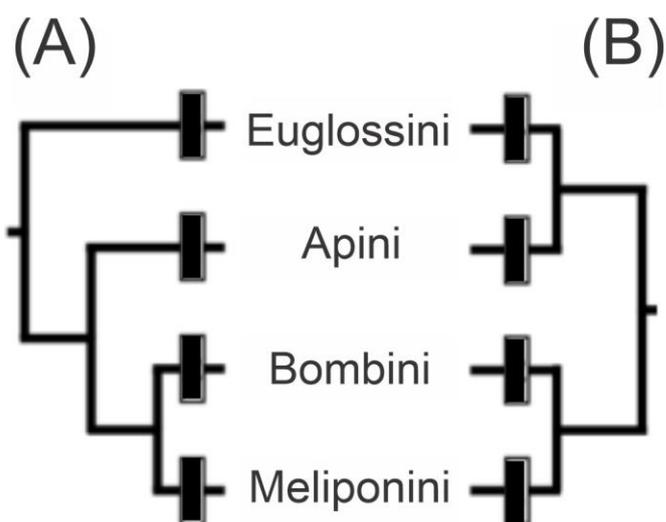
**Figura 9.** Dendrograma/árvore filogenética do MAFFT. (KATOH, ROZEWICKI, YAMADA 2019)

Fonte: Dados do autor, 2022.

Na árvore apresentada, os táxons Bombini e Meliponini aparecem como grupo-irmãos, Apini e Euglossini estão restritas em outro ramo também como grupo-irmãos. Esses quatro grupos são grupos-irmãos de Xylocopini representado pela espécie *Ceratina calcarata*, formando assim um grupo monofilético. Na outra extremidade dos ramos, as famílias Megachilidae, Colletidae e Halictidae dividem o mesmo clado como um grupo monofilético, tendo Colletidae como grupo-irmão de Halictidae, e estes sendo mais próximos

de Megachilidae. Em outro ramo outras espécies de Megachilidae estão dispostas.

Embasado em estudos atuais de sequências ultraconservadas de Bossert e colaboradores (2019), o clado das abelhas corbiculadas que inclui as tribos Bombini, Apini, Meliponini e Euglossini formam um grupo monofilético, em que Bombini, Meliponini e Apini são grupo-irmãs, e esses grupos tem relações mais próximas de Euglossini. Entretanto no presente resultado, apenas Bombini e Meliponini são grupo-irmãs, Apini e Euglossini estão restritas em outro ramo. Ambas hipóteses de que Bombini e Meliponini são grupo-irmãs, Apini e Euglossini demonstram o mesmo comportamento na árvore, pode ser corroborada pelo estudo de Porto (2015), explicando que essas ideias são comumente encontradas ao analisar dados moleculares. Essas tribos são representadas neste presente trabalho pelas espécies: *Bombus terrestris* (Bombini); *Apis dorsata*, *Apis florea*, *Apis laboriosa*, *Apis mellifera* (Apini); e *Melipona quadrifasciata*, *Frieseomelitta varia* (Meliponini); *Eufriesea mexicana* (Euglossini).



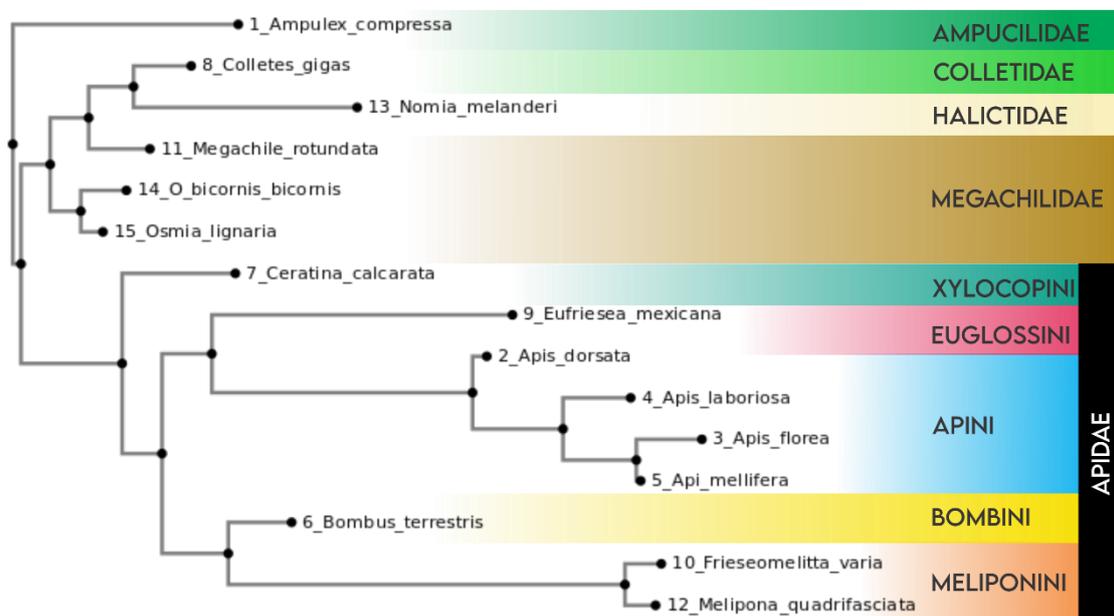
**Figura 10.** (A) Representação do resultado obtido pelo trabalho do BOSSERT *ET AL.* 2019; (B) Representação do resultado obtido pelo MAFFT .

**Fonte:** Figura adaptada de Almeida & Porto (2014).

A hipótese (B) é recuperada por grandes conjuntos de dados filogenômicos disponíveis no estudo recente de Porto e colaboradores (2021) e Noll e

colaboradores (2021), obtendo o agrupamento Bombini restringido como grupo irmão de Meliponini, mas todas as outras relações permaneceram inalteradas.

Segundo Debevec et al. 2012, e mais recentemente Bransteeter et al. 2017, *Osmia lignaria*, *Osmia bicornis bicornis* e *Megachile rotundata*, fazem parte da mesma família (Megachilidae), *Colletes gigas* (Colletidae) e *Nomia melanderi* (Halictidae). No resultado essa classificação encontra-se *O. lignaria* e *O. bicornis bicornis* restritas em um ramo, *C. gigas* e *N. melanderi* restritas em outro ramo como grupo irmão de *M. rotundata*. Essa inconsistência pode estar relacionada à quantidade de dados de sequência primária de citocromo c de espécies dessas famílias insuficientes para essa análise.



**Figura 11.** Representação da árvore com as respectivas famílias.

**Fonte:** Autor

## 5 CONSIDERAÇÕES FINAIS

Destaca-se que as classificações são esquemas de organização e os critérios que são utilizados para classificação dependem da finalidade para a qual se propõe a sua classificação (SILVEIRA; MELO; ALMEIDA, 2002). Muitas dúvidas ainda permeiam nas relações filogenéticas da superfamília **Apoidea**, pois este grupo necessita de outros métodos para uma análise mais completa. O Danforth, 2013 cita que a compreensão da filogenia desse grupo melhorou nos últimos anos através de novos dados, principalmente dados de sequência de nucleotídeos e novos métodos, principalmente métodos baseados em modelos de reconstrução de filogenia ajudaram resolver a colocação de abelhas dentro da superfamília Apoidea. Entretanto ainda é escasso os dados de sequências primárias do citocromo c desses himenópteros nos bancos de dados, logo conferindo um desafio dizer se essa proteína pode ser considerada um marcador molecular que possa definir as relações filogenéticas desse grupo (FAIRCLOTH et al, 2014). Contudo, essa primeira inferência filogenética parece sugerir o citocromo c como sendo um bom marcador para estudos filogenéticos.

Portanto é possível a utilização de citocromo c como marcador no clado das abelhas corbiculadas, sendo possível o entendimento de algumas relações internas, contudo tratando-se de outros grupos que fazem parte da superfamília Apoidea, essa proteína pode não ser essencialmente a indicada para analisar relações filogenéticas. É preciso também ressaltar a importância de outras abordagens metodológicas como importantes para compreender a filogenia do grupo, como é o caso dos dados morfológicos.

O número pequeno de indivíduos dentro de nossa análise também dificultou que o programa entendesse qual o grupo mais distante e mais recente, o que foi possível corrigir isto a partir da função *reroot*. Contudo, o citocromo c ainda é um “fóssil vivo” pois carrega na sua codificação as pequenas substituições que aconteceram ao longo dos anos, ajudando a entender a evolução da mitocôndria (SCHWARTZ; DAYHOFF, 1978). As ferramentas de bioinformática demonstraram grande importância na análise facilitando-a e organizando com rapidez, apesar das limitações. São necessários mais estudos em variados grupos dentre os

himenópteros para se obter as sequências de citocromo c de muitas espécies, uma vez que este é um dos clados mais ricos e diversos ainda vivos na terra.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

- ARAUJO, Edilson Divino de. **Avanços em biologia evolutiva**. Disponível em: [https://cesad.ufs.br/ORBI/public/uploadCatalogo/08594402092013Evoluc\\_ao\\_Aula\\_8.pdf](https://cesad.ufs.br/ORBI/public/uploadCatalogo/08594402092013Evoluc_ao_Aula_8.pdf). Acesso em: 12 maio 2022.
- AMORIM, D. de S. **Elementos básicos de sistemática filogenética**. São Paulo: Sociedade Brasileira de Entomologia, 1994.
- AMORIM, Dalton de Souza. **Fundamentos de sistemática filogenética**. In: Fundamentos de sistemática filogenética. 2002. p. 154-154.
- BANCI, Lucia et al. **Solution Structure of Oxidized Horse Heart Cytochrome c**. 1997. Disponível em: <https://www.rcsb.org/structure/1akk>. Acesso em: 12 maio 2022
- BOSSERT, Silas; MURRAY, Elizabeth A.; ALMEIDA, Eduardo A.B.; BRADY, Seán G.; BLAIMER, Bonnie B.; DANFORTH, Bryan N.. **Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of Apidae**. Molecular Phylogenetics And Evolution, [S.L.], v. 130, p. 121-131, jan. 2019. Elsevier BV. <http://dx.doi.org/10.1016/j.ympev.2018.10.012>.
- BRANSTETTER, Michael G.; DANFORTH, Bryan N.; PITTS, James P.; FAIRCLOTH, Brant C.; WARD, Philip S.; BUFFINGTON, Matthew L.; GATES, Michael W.; KULA, Robert R.; BRADY, Seán G.. **Phylogenomic Insights into the Evolution of Stinging Wasps and the Origins of Ants and Bees**. Current Biology, [S.L.], v. 27, n. 7, p. 1019-1025, abr. 2017. Elsevier BV. <http://dx.doi.org/10.1016/j.cub.2017.03.027>.
- CAMERON, S.A., MARDULYN, P., 2001. **Multiple molecular data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera: Apinae)**. Syst. Biol. 50, 192–214
- EMBL-EBI (Hinxton). **Clustal Omega**. Disponível em: <https://www.ebi.ac.uk/Tools/msa/clustalo/>. Acesso em: 22 abr. 2022.
- FAIRCLOTH, Brant C.; BRANSTETTER, Michael G.; WHITE, Noor D.; BRADY, Seán G.. **Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among**

**Hymenoptera**. Molecular Ecology Resources, [S.L.], v. 15, n. 3, p. 489-501, 29 set. 2014. Wiley. <http://dx.doi.org/10.1111/1755-0998.12328>.

FERNÁNDEZ, F.; SHARKEY, M. J. **Introducción a los Hymenoptera de la Region Neotropical**. p. 920, 2006.

DANFORTH, Bryan N.; CARDINAL, Sophie; PRAZ, Christophe; ALMEIDA, Eduardo A.B.; MICHEZ, Denis. **The Impact of Molecular Data on Our Understanding of Bee Phylogeny and Evolution**. Annual Review Of Entomology, [S.L.], v. 58, n. 1, p. 57-78, 7 jan. 2013. Annual Reviews. <http://dx.doi.org/10.1146/annurev-ento-120811-153633>.

DEBEVEC, A.H., CARDINAL, S., DANFORTH, B.N., 2012. **Identifying the sister group to the bees: a molecular phylogeny of Aculeata with an emphasis on the superfamily Apoidea**. Zoologica Scripta 41 (5), 527e535. <https://doi.org/10.1126/sciadv.1501080>

KERR, W. E.; CARVALHO, G. A.; SILVA, A. C. D. **Aspectos pouco mencionados da biodiversidade amazônica**. p. 22, 2001.

LAURENCE A. MORAN, H. Robert Horton, K. Gray Scrimgeour e Marc D.Perry. **Bioquímica**, 5ª edição. Pearson 836 ISBN 9788581431260 . EBOOK. P-22, 2013.

LEAL, J. **Estudio e implementación de algunas herramientas bioinformáticas para su posible integración en un sistema que permita la cara**. p. 132, 2004.

LEHNINGER, T. M., NELSON, D. L. & COX, M. M. **Princípios de Bioquímica**. 6ª Edição, 2014. Ed. Artmed.

LESK, A. M. **Introduction to bioinformatics**. Fourth edition ed. Oxford, United Kingdom: Oxford University Press, 2014.

MAFFT version 7. Disponível em: <https://mafft.cbrc.jp/alignment/server/spool/phyloio.220423123732212.html>. Acesso em: 22 abr. 2022.

MICHENER, C. D. **The bees of the world**. 2nd ed ed. Baltimore: Johns Hopkins University Press, 2007.

- MILICCHIO, Vitor Augusto. **Determinação dos valores de pKa dos grupos prototrópicos de citocromo c identificado pela reatividade com dietilpirocarbonato em função de pH.** 2011. 77 f. Dissertação (Mestrado) - Curso de Biotecnologia Aplicada À Saúde, Universidade Mogi das Cruzes, Mogi das Cruzes, 2011.
- MOREIRA, Alexandre Paulo Teixeira. **Zoologia de Invertebrados I.** Florianópolis: Universidade Federal de Santa Catarina. Biologia/Ead, 2009. 213 p.
- MOUNT, David W.. **Bioinformatics: sequence and genome analysis.** 2. ed. Nova York: Cold Spring Harbor Laboratory Press, 2004. 692 p.
- NOLL, F.B., Justino, C.E.L., Almeida, E.A.B. et al. **Would wax glands help us to understand the relationships among corbiculate bees?** Insect. Soc. 68, 191–197 (2021). <https://doi.org/10.1007/s00040-021-00813-9>
- OLLERTON, Jeff; WINFREE, Rachael; TARRANT, Sam. **How many flowering plants are pollinated by animals?** Oikos, v. 120, n. 3, p. 321-326, 2011.
- PORTO, Diego Sasso. **Phylogenetic relationships of corbiculate bees (Hymenoptera: Apidae: Apini).** 2015. 231 f. Dissertação (Mestrado) - Curso de Programa de Pós-Graduação em Entomologia, Ffclrp - Departamento de Biologia, Universidade de São Paulo, Ribeirão Preto - Sp, 2015. Cap. 1.
- PORTO, Diego S.; VILHELMSSEN, Lars; ALMEIDA, Eduardo A. B.. **Comparative morphology of the mandibles and head structures of corbiculate bees (Hymenoptera: Apidae).** Systematic Entomology, [S.L.], v. 41, n. 2, p. 339-368, 27 out. 2015. Wiley. <http://dx.doi.org/10.1111/syen.12156>.
- PORTO, Diego Sasso; ALMEIDA, Eduardo A. B. **Corbiculate Bees (Hymenoptera: Apidae): Exploring the Limits of Morphological Data to Solve a Hard Phylogenetic Problem: apidae).** Insect Systematics And Diversity, [S.L.], v. 5, n. 3, p. 1-40, 1 maio 2021. Oxford University Press (OUP). <http://dx.doi.org/10.1093/isd/ixab008>.
- PULAWSKI, Wojciech J.. **FAMILY GROUP NAMES AND CLASSIFICATION and taxa excluded from Sphecidae sensu lato.** San Francisco: California Academy Of Sciences, 2021. 14 p.

- RIDLEY, M. 2006. **Evolução**. 3a. ed. Porto Alegre: ArtMed Editora. 752p.
- SANN, M. et al. **Phylogenomic analysis of Apoidea sheds new light on the sister group of bees**. BMC Evolutionary Biology, v. 18, n. 1, p. 71, dez. 2018.
- SOLÉ-CAVA, Antonio; SILVA, Edson Pereira da; LÔBO-HAJDU, Gisele. **Evolução**. 3. ed. Rio de Janeiro-Rj: Fundação Cecierj, 2010. 271 p.
- SCHWARTZ, R. M.; DAYHOFF, M. O. **Origins of Prokaryotes, Eukaryotes, Mitochondria, and Chloroplasts: A perspective is derived from protein and nucleic acid sequence data**. Science, v. 199, n. 4327, p. 395–403, 27 jan. 1978.
- SILVEIRA, F. A.; MELO, G. A. R.; ALMEIDA, E. A. B. **Abelhas brasileiras: sistemática e identificação**. 1. ed ed. Belo Horizonte: Silveira, 2002.
- SILVEIRA, L. T. DA. **Estudo eletroquímico da interação do citocromo c com líquidos iônicos em diferentes substratos para aplicação em biossensores**. Doutorado em Química—São Paulo: Universidade de São Paulo, 12 set. 2012.
- SOUZA, Maria Angélica Lopes de. **Alinhamento Múltiplo Progressivo de Sequências de Proteínas**. 2010. 146 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Unicamp, Campinas, 2010.
- VERLI, Hugo. **Bioinformática: da biologia à flexibilidade molecular**. Sociedade Brasileira de Bioquímica e Biologia Molecular. 1. ed. - São Paulo: SBBq, 2014. p. 191, 282, (2014).

# ANEXOS

## Anexo A

>KAG7208637.1 hypothetical protein KM043\_014844 [Ampulex compressa]

MGVPSGDPEKGKKVQVQRCQAQCHTIEAGGKHKVGNLHGIVGRKTGQAPGF  
AYTDANKAKGITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQ  
ASK

>XP\_006612971.1 cytochrome c [Apis dorsata]

MGVPAGDPEKGKKIFVQKCAQCHTIESGGKHKVGNLYGVYGRKTGQAPGY  
SYTDANKQKGITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQ  
ASK

>XP\_003692897.1 cytochrome c [Apis florea]

MGIPAGDPEKGKKIFIQKCAQCHTIESGGKHKVGNLYGVYGRKTGQAPGYS  
YTDANKGKGITWNKETLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQA  
SK

>XP\_043791397.1 cytochrome c [Apis laboriosa]

MGIPAGDPEKGKKIFIQKCAQCHTIESGGKHKVGNLYGVYGRKTGQAPGYS  
YTDANKQKGITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQA  
SK

>NP\_001170961.1 cytochrome c [Apis mellifera]

MGIPAGDPEKGKKIFVQKCAQCHTIESGGKHKVGNLYGVYGRKTGQAPGYS  
YTDANKGKGITWNKETLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQA  
SK

>XP\_003394648.1 cytochrome c [Bombus terrestris]

MGVPAGDPEKGKKLFVQKCAQCHTIEAGGKHKVGNLHGVMGRKTGQAPG  
YSYTDANKSKGITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIE  
QASK

>XP\_017890006.1 cytochrome c [Ceratina calcarata]

MGVPAGDPEKGKKIFVQKCAQCHTIEAGGKHKVGNLHGIVGRKTGQAPGYS  
YTDANKAKGITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQA  
SK

>XP\_043260830.1 cytochrome c [Colletes gigas]

MGVPAGDTEKGKKLFVQRCSQCHTIEAGGKHKVGNLHGIIGRKTGQAPGYS  
YTDANKAKGITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQA  
SK

>XP\_017759263.1 PREDICTED: cytochrome c [Eufriesea mexicana]

MGVPAGDPEKGKKIFVQKCAQCHTIEAGGKHKVGP NLHG VVGRKTGQAPGY  
SYTDANKAKGITWNRDTLFEYLENPKKYIPGTKMVFAGIKKPQERADLIAYMEQ  
ESK

>XP\_043513165.1 cytochrome c [Frieeseomelitta varia]

MGVPAGDVEKGKKLFVQKCAQCHTIEAGGKHKVGP NLHG IMGRKTGQAPGY  
SYTDANISKAITWSRDTLFEYLENPKKYIPGTKMVFAGLKKAQERADLIAYIEQA  
SK

>XP\_003701289.1 PREDICTED: cytochrome c [Megachile rotundata]

MGVPAGDPEKGKKV FVQRCSQCHTIEAGGKHKVGP NLHG IIGRKTGQAPGYS  
YTDANKAKGITW NKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQA  
SK

>KOX69899.1 Cytochrome c [Melipona quadrifasciata]

MGVPAGDAEKGKKLFVQKCAQCHTIEAGGKHKVGP NLHG IMGRKTGQAPGY  
SYTDANISKAITWSRDTLFEYLENPKKYIPGTKMVFAGLKKAQERADLIAYIEQA  
SK

>XP\_031827659.1 cytochrome c-2 [Nomia melanderi]

MPVPAGDVEKGKKLFVQRCAQCHTIEAGGKHKVGP NLHG IIGRKTGQAAGYS  
YTDANKAKGITW NKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQA  
SK

>XP\_029042037.1 cytochrome c [Osmia bicornis bicornis]

MGVPAGDPEKGKKLFVQRCSQCHTIEAGGKHKVGP NLHG IVGRKTGQAPGF  
SYTDANKAKGITW NKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQ  
ASK

>XP\_034187101.1 cytochrome c [Osmia lignaria]

MGVPAGDPEKGKKLFVQRCSQCHTIEAGGKHKVGP NLHG IVGRKTGQAPGY  
SYTDANKAKGITW NKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQ  
ASK

## MAFFT version 7

Multiple alignment program for amino acid or nucleotide sequences



[Download version](#)

[Mac OS X](#)

[Windows](#)

[Linux](#)

[Source](#)

[Online version](#)

[Alignment](#)

[mafft --add](#)

[Merge](#)

[Phylogeny](#)

[Rough tree](#)

[Merits / limitations](#)

[Algorithms](#)

[Tips](#)

[Benchmarks](#)

[Feedback](#)

[Follow](#)

Hardware was upgraded, Jan 16, 2022. There should be no change in user interface. If you notice any unexpected changes, then [please let us know](#).

To avoid overload, try [a light-weight option](#), for MSA of full-length SARS-CoV-2 genomes (2020/Apr).

For a large number of short sequences, try [an experimental service](#).

[Experimental service for aligning raw reads \(2019/Aug\)](#)

Multiple sequence alignment and NJ / UPGMA phylogeny



**Figura 12.** Introdução no MAFFT – Inserindo sequências.

Fonte: MAFFT 2022.

## MAFFT version 7

Multiple alignment program for amino acid or nucleotide sequences



### [Download version](#)

[Mac OS X](#)

[Windows](#)

[Linux](#)

[Source](#)

### Online version

[Alignment](#)

[mafft --add](#)

[Merge](#)

[Phylogeny](#)

[Rough tree](#)

### [Merits / limitations](#)

[Algorithms](#)

[Tips](#)

[Benchmarks](#)

[Feedback](#)

[Follow](#)

#### UPPERCASE / lowercase:

- Same as input
- Amino acid → UPPERCASE / Nucleotide → lowercase

#### Direction of nucleotide sequences: [Help](#)

- Same as input
- Adjust direction according to the first sequence (accurate enough for most cases)
- Adjust direction according to the first sequence (only for highly divergent data; **extremely slow**)

#### Output order:

- Same as input
- Aligned

#### Job name (optional; used as output file name and subject of emails):

(basic Latin alphabet, number and space only)

#### Notify when finished (optional; recommended when submitting large data):

Email address:

**Figura 13.** Introdução no *MAFFT* – Submetendo sequências.

Fonte: MAFFT 2022.

MAFFT-L-INS-i Result

```

CLUSTAL format alignment by MAFFT (v7.505)

Ampulex      MGVPSPGDPEKGGKVFVQRCQAQCHTIEAGGKHKVGNLHGIVGRKGTQAPGFAYTDANKAK
Colletes     MGVPAGDTEKGGKLFVQRCSQCHTIEAGGKHKVGNLHGIIIGRKTGQAPGYSYTDANKAK
Megachile    MGVPAGDPEKGGKVFVQRCQAQCHTIEAGGKHKVGNLHGIIIGRKTGQAPGYSYTDANKAK
O            MGVPAGDPEKGGKLFVQRCSQCHTIEAGGKHKVGNLHGIVGRKGTQAPGFYSYTDANKAK
Osmia        MGVPAGDPEKGGKLFVQRCSQCHTIEAGGKHKVGNLHGIVGRKGTQAPGYSYTDANKAK
Ceratina     MGVPAGDPEKGGKIFVQKCAQCHTIEAGGKHKVGNLHGIVGRKGTQAPGYSYTDANKAK
Bombus       MGVPAGDPEKGGKLFVQKCAQCHTIEAGGKHKVGNLHGVMGRKGTQAPGYSYTDANKSK
Eufriesea    MGVPAGDPEKGGKIFVQKCAQCHTIEAGGKHKVGNLHGIVGRKGTQAPGYSYTDANKAK
Apis         MGVPAGDPEKGGKIFVQKCAQCHTIEAGGKHKVGNLYGVYGRKGTQAPGYSYTDANKQK
Apis         MGIPAGDPEKGGKIFIQKCAQCHTIEAGGKHKVGNLYGVYGRKGTQAPGYSYTDANKQK
Apis         MGIPAGDPEKGGKIFIQKCAQCHTIEAGGKHKVGNLYGVYGRKGTQAPGYSYTDANKGK
Api          MGIPAGDPEKGGKIFVQKCAQCHTIEAGGKHKVGNLYGVYGRKGTQAPGYSYTDANKGK
Nomia        MPVPAGDVEKGGKLFVQRCQAQCHTIEAGGKHKVGNLHGIIIGRKTQAAGYSYTDANKAK
Frieseomelitta MGVPAGDVEKGGKLFVQKCAQCHTIEAGGKHKVGNLHGIMGRKGTQAPGYSYTDANISK
Melipona     MGVPAGDAEKGGKLFVQKCAQCHTIEAGGKHKVGNLHGIMGRKGTQAPGYSYTDANISK
* :*:** *****:~*~*~*****:*****~*~*~***** *

Ampulex      GITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQASK
Colletes     GITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQASK
Megachile    GITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQASK
O            GITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQASK
Osmia        GITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQASK
Ceratina     GITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQASK
Bombus       GITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQASK
Eufriesea    GITWNRDTLFEYLENPKKYIPGTKMVFAGIKKPQERADLIAYMEQESK
Apis         GITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQASK
Apis         GITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQASK
Apis         GITWNKETLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQASK
Api          GITWNKETLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQASK
Nomia        GITWNKDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQASK
Frieseomelitta AITWSRDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQASK
Melipona     AITWSRDTLFEYLENPKKYIPGTKMVFAGLKKPQERADLIAYIEQASK
.***.~*~*~*****:***.***.*****~*~*~***** *

```

Figura 14. Introdução no MAFFT – Obtenção do alinhamento.

Fonte: MAFFT 2022.

[Clustal format](#) | [Fasta format](#) | [MAFFT result](#) | [View](#) | [Tree](#) | [Refine dataset](#) | [Return to home](#)

View

Reformat to GCG, PHYLIP, MSF, NEXUS, uppercase/lowercase, etc. with Readseq

GUIDANCE2 computes the residue-wise confidence scores and extracts well-aligned residues.

Refine dataset

Phylogenetic tree

**Figura 15.** Introdução no *MAFFT* – Obtenção da árvore.

Fonte: MAFFT 2022.

NJ or UPGMA tree ( $\beta$ )

sequences,  total sites,  gap-free sites,  [conserved sites](#)

Settings

Method:

- [NJ – Conserved sites \(108 AAs\)](#)
- [NJ](#) – All of gap-free sites (108 AAs)
- Average linkage (UPGMA) – alignment scores (for up to 50,000 sequences)
- Minimum linkage – alignment scores (for up to 50,000 sequences)
- Memory-saving tree – alignment scores (for larger data)

Substitution model (valid when NJ is selected):

- [JTT](#)
- [WAG Alpha](#)
- Poisson
- Raw difference

Heterogeneity among sites (valid when JTT or WAG is selected):

- Ignore ( $\alpha = \infty$ )
- Estimate
- Specify:  $\alpha =$   (0.10 – 5.0)

Bootstrap (valid for NJ):

- On
- Number of resampling:  (5 – 1000)  
(The number of sequences must be <1000 for Poisson model, or <100 for other models.)

**Figura 16.** Introdução no *MAFFT* – Configuração da árvore.

Fonte: MAFFT 2022.

[Clustal format](#) | [Fasta format](#) | [MAFFT result](#) | [View](#) | [Tree](#) | [Refine dataset](#) | [Return to home](#)

## Result (Phylo.io 1.0.0)

---

Phylo.io runs on any modern browser.

[View tree on Phylo.io](#)

[Refine dataset on tree](#)

**Figura 17.** Introdução no *MAFFT* – Visualização da árvore.

Fonte: MAFFT 2022.

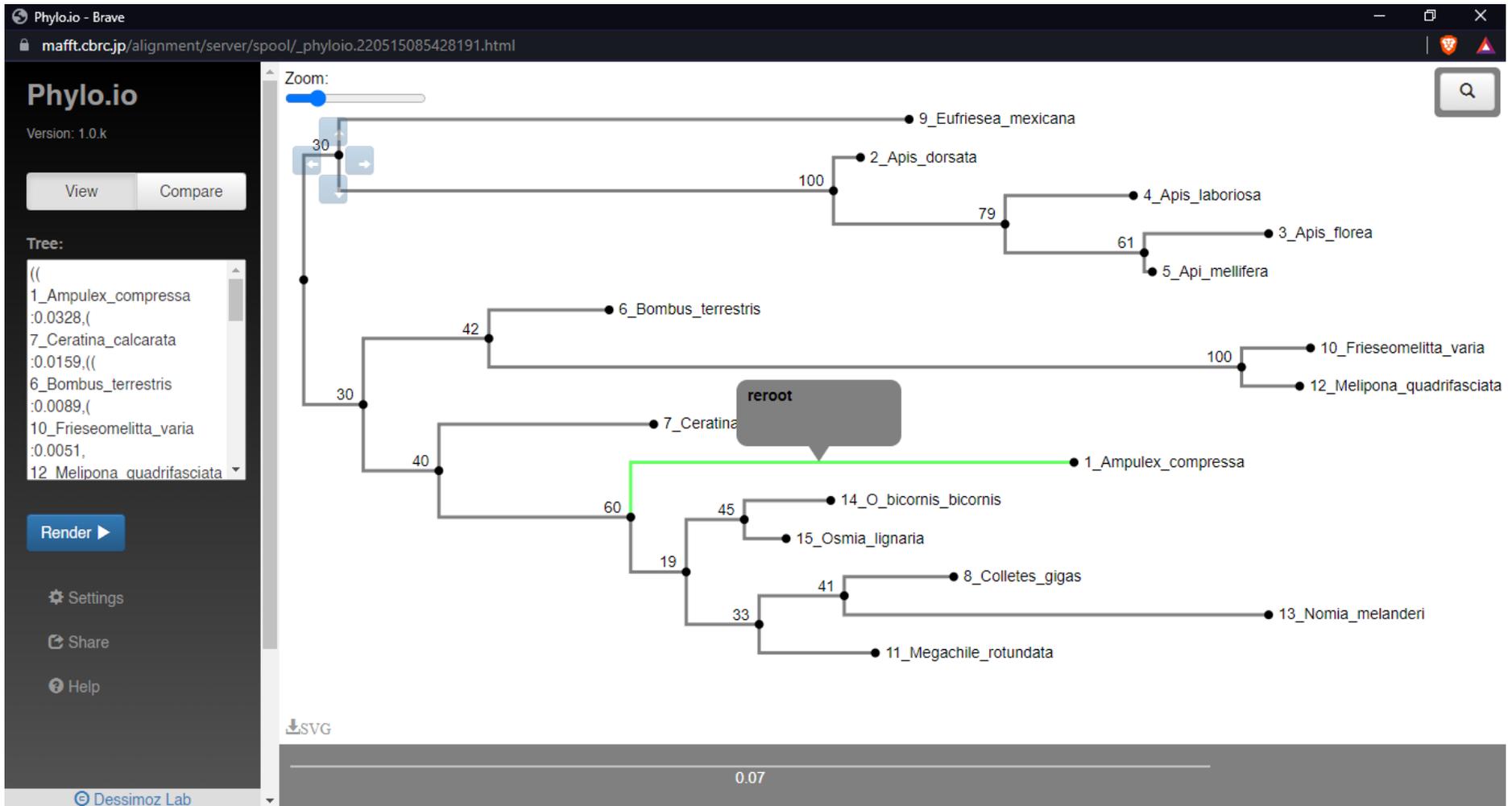


Figura 18. Introdução no MAFFT – Reroot da árvore.

Fonte: MAFFT 2022

Essa página foi deixada em branco propositalmente