

**Universidade Federal de Pernambuco
Centro de Ciências Biológicas
Programa de Pós-Graduação em Genética**

Patrícia Keytth Lins Rocha

**Qual região do DNA mitocondrial reflete a história
evolutiva da ordem Lepidoptera?**

**Recife
2013**

Patrícia Keytth Lins Rocha

**Qual região do DNA mitocondrial reflete a história
evolutiva da ordem Lepidoptera?**

Dissertação apresentada ao Programa de Pós-Graduação em Genética da Universidade Federal de Pernambuco como parte dos requisitos exigidos para obtenção do título de Mestre em Genética.

Orientador: Marcos Antônio de Moraes Júnior

Coorientador: Valdir de Queiroz Balbino

Recife

2013

Catálogo na fonte
Elaine Barroso
CRB 1728

Rocha, Patrícia Keytth Lins

Qual região do DNA mitocondrial reflete a história evolutiva da ordem Lepidoptera?/ Patrícia Keytth Lins Rocha– Recife: O Autor, 2013.

90 folhas : il., fig., tab.

Orientador: Marcos Antônio de Moraes Júnior

Coorientador: Valdir de Queiroz Balbino

Dissertação (mestrado) – Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Genética, 2013.

Inclui bibliografia e apêndices

1. Filogenia 2. Lepidoptera 3. DNA I. Moraes Júnior, Marcos Antônio de (orientador) II. Balbino, Valdir de Queiroz (coorientador) IV. Título

576.88

CDD (22.ed.)

UFPE/CCB- 2013- 305

Patrícia Keytth Lins Rocha

**Qual região do DNA mitocondrial reflete a história evolutiva da
ordem Lepidoptera?**

Aprovado em ____/____/____

Banca Examinadora:

Dr. Valdir de Queiroz Balbino
Universidade Federal de Pernambuco

Dr. Marcos André Cavalcanti Bezerra
Universidade Federal de Pernambuco

Dr. Herbert Álvaro Abreu de Siqueira
Universidade Federal Rural de Pernambuco

Recife
2013

Aos meus pais, Rivaldo Lins Rocha e Maria Virgínia de Jesus Lins Rocha, que sempre me apoiaram e me incentivaram a lutar pelo que almejo. Divido com eles a concretização de mais um sonho.

Agradecimentos

Agradeço a Deus por tudo que ele colocou em minha vida e por nunca ter me abandonado.

Ao Prof. Dr. Marcos Morais pela oportunidade, orientação, confiança e por ser um exemplo de pesquisador.

Ao Prof. Dr. Valdir Balbino pela coorientação, pelos ensinamentos de bioinformática, por indicar os melhores caminhos e por ser exemplo profissional.

Ao Dr. José Roberto Tavares pela sua dedicação, por compartilhar suas experiências de pesquisador e pelas discussões produtivas.

Aos meus pais por serem exemplos de luta e aos meus irmãos pelo companheirismo. À toda minha família que me apoia e me incentiva.

Ao meu noivo Marcos Martins, que me incentivou a buscar o mestrado na UFPE e esteve sempre presente durante este tempo. Com palavras de consolo e estímulos, mostrou-se compreensivo e companheiro em todos os momentos.

À minha amiga Lidianne Gomes, que foi o meu apoio em Recife, meu conforto e minha família.

Ao Prof. Dr. Herbert Siqueira por ceder gentilmente as larvas de *Diatraea flavipennela* para a pesquisa.

A técnica do laboratório de Genômica e Expressão Gênica, Heidi Lacerda, e a Dra. Theresa Liberal, por fornecerem os cromatogramas.

Aos membros do Núcleo de Engenharia Metabólica pelo companheirismo, ensinamentos e momentos de alegria.

Aos membros do Laboratório de Bioinformática e Biologia Evolutiva por me auxiliarem nas análises de bioinformática, pelo acolhimento e bom humor. Em especial a Marco dos Santos e Marcus Batista, que tiveram uma contribuição maior neste trabalho.

A todos os professores do Programa de Pós-Graduação em Genética da UFPE que contribuíram para a minha formação acadêmica.

Aos colegas de disciplinas por ter compartilhado momentos alegres e difíceis e se mostrado um verdadeiro grupo.

A FACEPE pela concessão da bolsa.

“Disse a flor para o pequeno príncipe: é preciso que eu suporte duas ou três larvas se quiser conhecer as borboletas.”

(Antoine de Saint-Exupéry)

Resumo

A ordem Lepidoptera apresenta uma gama de espécimes de importância econômica, algumas são agentes polinizadores e outras são pragas. Vários estudos têm sido realizados com base na morfologia com base em dados moleculares para elucidar a evolução da ordem. O DNA mitocondrial é muito utilizado por proporcionar boa resolução filogenética. Com base em genes mitocondriais informativos, nós propusemos um conjunto de dados que pode ser utilizado em análise filogenética de Lepidoptera obtendo a mesma robustez que a análise com mtDNAs completos. Para isso, as sequências dos mitogenomas de Lepidoptera foram recuperadas no banco de dados do NCBI. Foi identificada a ordem gênica das sequências utilizando o programa MAUVE. As regiões de interesse em *D. flavipennella* foram sequenciadas para testar a eficiência dos marcadores moleculares em sequência nova. Foi realizada análise de entropia, teste de sinal filogenético e de saturação para verificar características de bons marcadores moleculares e foram realizadas análises filogenéticas nos programas PhyML e MrBayes. Foram realizados também testes com e sem a terceira posição dos códons para verificar a influência da terceira posição nas análises filogenéticas de Lepidoptera. As regiões estudadas foram concatenadas para aumentar os valores de confiança das árvores. Verificamos que com a concatenação dos genes COI, ATP6, COIII, ND3, ND5, CYTB, ND1 e 16S foi possível obter resultados com robustez semelhante a dos mitogenomas completos.

Palavras-chave: Concatenação de genes; Filogenia; Lepidoptera; Marcadores Moleculares; Mitogenoma.

Abstract

Lepidoptera specimens present a range of economically important. They are pollinators, pests and harm human health. Several studies based on morphology and molecular data to elucidate the evolution of the order exists. The mtDNA is widely used for providing good phylogenetic resolution. Based on information mitochondrial genes, we propose a set of data used in phylogenetic analysis to obtaining the same lepidopteran robustness of complete mtDNAs analysis. For this, we recover the sequences of lepidopteran mitogenomes in the NCBI database. We identified the order of gene sequences using the program MAUVE. We sequenced the region of interest in *D. flavipennella* mitogenome in order to test the efficiency of the new molecular markers in a new sequence. We analyzed the entropy, the phylogenetic signal and the saturation of the sequences to verify characteristics of good molecular markers. We performed phylogenetic analysis using PhyML and MrBayes. We also performed tests with and without the 3rd position of codons to verify the influence of the 3rd position in the phylogenetic analysis of lepidopteran. Finally, we concatenate the regions studied to increase the confidence values of the trees. We found that 4.6 kb was sufficient to group all families of Lepidoptera with strong support.

Keywords: Concatenation of genes; Lepidoptera; Mitogenome; Molecular Markers; Phylogeny.

Lista de Ilustrações

Figura 1 - Filogenia dos Lepidoptera calibrada por ocorrências fósseis (círculos)	3
Figura 2 - Diversidade da ordem Lepidóptera	4
Figura 3 - Mapa circular do genoma mitocondrial de <i>Diatraea saccharalis</i>	5
Figura 4 - Representação esquemática dos arranjos de genes mitocondriais	6
Quadro 1 – Espécies de Lepidoptera utilizadas neste trabalho	13
Quadro 2 - Tipos de análises realizadas neste trabalho	20
Quadro 3 - Testes realizados neste trabalho	21
Figura 5 - Representação linear dos mtDNAs de Lepidoptera	26
Figura 6 - Representação linear do mitogenoma de <i>D. saccharalis</i> e produto dos pares de <i>primers</i> utilizados para <i>D. flavipennella</i>	29
Figura 7 - Gel dos produtos amplificados dos pares de <i>primers</i> utilizados	30
Figura 8 - Análise de entropia dos genes	32
Figura 9 - Análise de entropia das regiões delimitadas pelos pares de <i>primers</i>	33
Figura 10 - Teste de saturação dos genes	39
Figura 11 - Teste de saturação dos pares de <i>primers</i>	41
Figura 12 - Árvore filogenética de Lepidoptera a partir do gene COI	43
Figura 13 - Árvore filogenética de Lepidoptera a partir do gene COIII	44
Figura 14 - Árvore filogenética de Lepidoptera a partir do 16S	45

Figura 15 - Árvore filogenética de Lepidoptera a partir da região delimitada pelo par de <i>primers</i> LABBE-Dip1	46
Figura 16 - Relações filogenéticas inferidas para Lepidoptera baseado na concatenação dos genes por análise bayesiana e máxima verossimilhança, respectivamente	48
Figura 17 - Relações filogenéticas inferidas para Lepidoptera baseado na concatenação dos pares de <i>primers</i> por análise bayesiana e máxima verossimilhança, respectivamente	49

Lista de Tabelas

Tabela 1 - Pares de <i>primers</i> utilizados para <i>Diatraea flavipennella</i>	17
Tabela 2 - Modelo evolutivo dos testes	23
Tabela 3 - Informações sobre as regiões de <i>Diatraea flavipennella</i> sequenciadas	31
Tabela 4 - Análise dos sítios dos genes de Lepidoptera	34
Tabela 5 - Análise dos sítios das regiões delimitadas pelos pares de <i>primers</i>	35
Tabela 6 - Sinal filogenético dos genes	36
Tabela 7 - Sinal filogenético das regiões delimitadas pelos pares de <i>primers</i>	36

Lista de Abreviaturas, Siglas e Símbolos

Item	Definição
A	Adenina / Alanina
A+T%	Porcentagem de Adenina e Timina
ATP	Adenosina Trifosfato
ATP6	ATP F0 Sintase Subunidade 6
ATP8	ATP F0 Sintase Subunidade 8
BIC	<i>Bayesian Information Criterion</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
C	Citosina / Cisteína
COI	Subunidade 1 da citocromo c oxidase
COII	Subunidade 2 da citocromo c oxidase
COIII	Subunidade 3 da citocromo c oxidase
CCB	Centro de Ciências Biológicas
CYTB	Citocromo B
D	Ácido Aspártico
DAMBE	<i>Data Analysis in Molecular Biology and Evolution</i>
DNA	Ácido desoxirribonucléico
mtDNA	DNA Mitocondrial
dNTP	Desorribonucleotídeo trifosfatado
FACEPE	Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco
E	Ácido glutâmico
F	Fenilalanina

G	Guanina / Glicina
GCPs	Genes Codificadores de Proteínas
H	Histidina
I	Isoleucina
IDT	<i>Integrated DNA Technologies</i>
IUPAC-IUB	<i>International Union of Pure and Applied Chemistry - International Union of Biochemistry</i>
IS	Espaço intergênico
K	Lisina
Kb	Kilobases
L	Leucina
M	Metionina
Ma	Milhões de anos
MCMC	<i>Markov Chain Monte Carlo</i>
MgCl₂	Cloreto de Magnésio
Min	Minuto
mM	Milimolar
N	Nucleotídeo/ Asparagina
NCBI	<i>National Center of Biotechnology Information</i>
ND1	Subunidade 1 da NADH Desidrogenase
ND2	Subunidade 2 da NADH Desidrogenase
ND3	Subunidade 3 da NADH Desidrogenase
ND4	Subunidade 4 da NADH Desidrogenase

ND4L	Subunidade 4L da NADH Desidrogenase
ND5	Subunidade 5 da NADH Desidrogenase
ND6	Subunidade 6 da NADH Desidrogenase
ng	Nanogramas
NNI	Nearest Neighbor Interchange
P	Prolina
PCR	Reação em cadeia da polimerase
PLANALSUCAR	Programa Nacional de Melhoramento da Cana-de-Açúcar
Q	Glutamina
R	Arginina
RNA	Ácido ribonucléico
rRNA	RNA ribossomal
S	Serina
SAL	Sítios alinhados
SC	Sítios conservados
Seg	Segundo
SING	<i>Singletons</i>
SPi	Sítios informativos para parcimônia
SPR	Subtree <i>Pruning and Regrafting</i>
SV	Sítios variáveis
T	Timina / Treonina
Ti	Taxa de transição
TM	Temperatura de <i>melting</i>
tRNA	RNA transportador
tRNA^{Gln}	tRNA – Glutamina

tRNA^{Gly}	tRNA – Glicina
tRNA^{Ile}	tRNA – Isoleucina
tRNA^{Leu}	tRNA – Leucina
tRNA^{Met}	tRNA – Metionina
tRNA^{Ser}	tRNA – Serina
Tv	Taxa de transversão
U	Uracila
UFPE	Universidade Federal de Pernambuco
UFRPE	Universidade Federal Rural de Pernambuco
U\$	Dólar americano
v.	Versão
V	Valina
W	Triptofano
Y	Tirosina
12S	RNA ribossomal 12S
16S	RNA ribossomal 16S
%	Porcentagem
μL	Microlitro
μM	Micromolar
°C	Graus Celsius
=	Igual
≥	Maior igual a

Sumário

Resumo	viii
Abstract	ix
Lista de ilustrações	x
Lista de tabelas	xii
Lista de abreviaturas, siglas e símbolos	xiii
1. Introdução	1
2. Revisão da Literatura	2
2.1 Ordem Lepidoptera	2
2.2 DNA mitocondrial de Lepidoptera	4
2.3 Filogenia	7
2.4 Marcadores Moleculares	10
3. Objetivos	12
3.1 Geral	12
3.2 Específicos	12
4. Material e Métodos	13
4.1 Recuperação das sequências	13
4.2 Análise de sintenia	14
4.3 Sequenciamento das regiões filogeneticamente	15

informativas no DNA mitocondrial de <i>Diatraea flavipennella</i>	
4.3.1 Material Biológico	16
4.3.2 Extração de DNA e amplificação	17
4.3.3 Clonagem e sequenciamento	18
4.3.4 Análises das sequências	18
4.4 Análise das regiões filogeneticamente informativas para Lepidoptera	19
4.4.1 Análise de entropia	19
4.4.2 Análises da variação nucleotídica na sequências analisadas	20
4.4.3 Sinal filogenético	22
4.4.4 Teste de saturação	22
4.4.5 Análise filogenética	22
4.5 Concatenação	23
5. Resultados	25
5.1 Análise de sintenia	25
5.2 Sequenciamento das regiões filogeneticamente informativas no mitogenoma de <i>Diatraea flavipennella</i>	29
5.3 Análise das regiões informativas para a filogenia de Lepidoptera	31
5.3.1 Análise de entropia	31

5.3.2 Análises da variação nucleotídica na sequências analisadas	34
5.3.3 Sinal filogenético	35
5.3.4 Teste de saturação	37
5.3.5 Árvores filogenéticas	41
5.4 Concatenação	47
6. Discussão	49
7. Conclusões	59
8. Referências Bibliográficas	60
9. Apêndices	65
10. Curriculum vitae (Lattes)	86

1. Introdução

Lepidoptera é o segundo maior grupo da classe Insecta. Apresenta grande importância econômica por incluir espécies de pragas e agentes polinizadores. Além disso, o grupo tem sido amplamente estudado como sistema-modelo em várias pesquisas científicas.

A evolução de Lepidoptera foi intensamente estudada com base na morfologia e mais recentemente com base em dados de DNA. A sequência do DNA mitocondrial completa é cada vez mais utilizada como o marcador de escolha por apresentar forte sinal filogenético e por proporcionar melhor resolução filogenética e precisão em relação a marcadores tradicionais.

Visando a rapidez e a economia, estudos de concatenação de genes mitocondriais evolutivamente informativos tornaram-se uma opção em análises filogenéticas. Neste trabalho, nos propusemos a identificar região(ões) dos mitogenomas de Lepidoptera capazes de reproduzir resultados semelhantes aos obtidos pelas análises com as sequências completas dos mtDNAs e otimizar as análises de grandes conjuntos de dados ou de Lepidoptera com sequências mitogenômicas incompletas.

2. Revisão da Literatura

2.1 Ordem Lepidoptera

Lepidoptera é formada por mais de 150.000 espécies de mariposas e borboletas. São organismos holometábolos, cujas asas, corpo e apêndices são cobertos por escamas pigmentadas ou cerdas em forma de pelo (Freitas e Marini-Filho, 2011).

As primeiras espécies de Lepidoptera surgiram no período Jurássico-Cretáceo (antes da deriva continental, portanto) e se diversificaram no período Cenozoico, principalmente nas regiões Neotropicais. A expansão do grupo relaciona-se com o início da diversidade das angiospermas, que abrange o intervalo de 130 a 95 milhões de anos (Ma) (Figura 1). Um grupo de Lepidoptera (Glossata) desenvolveu um sistema bucal sugador que permitiu a alimentação a partir do néctar das flores. A coevolução entre angiospermas e Lepidoptera foi fundamental para a diversidade da ordem (Labandeira *et al.*, 1994).

A classificação atual de Lepidoptera divide o grupo em quatro subordens: Zeugloptera, Aglossata, Heterobathmiina e Glossata (Figura 2). Essa última é a mais numerosa e apresenta distribuição cosmopolita (Buzzi, 2005). É composta por espécies com espirotromba e de grande importância ecológica, econômica e social. Seus adultos são agentes polinizadores fundamentais que permitem o equilíbrio dinâmico de ecossistemas e são utilizados em pesquisas ecológicas como indicadores no monitoramento da diversidade biológica. As larvas das famílias Pyralidae, Pieridae e Papilionidae são pragas agrícolas de culturas potencialmente importantes para o ser humano. Existem ainda espécies que causam prejuízos à saúde humana, como as da família Saturniidae. As larvas de

Lonomia spp. possuem cerdas urticantes que ao serem inoculadas pelo homem podem ocasionar desde simples lesões até mesmo o óbito (Oliveira-Júnior *et al.*, 2007).

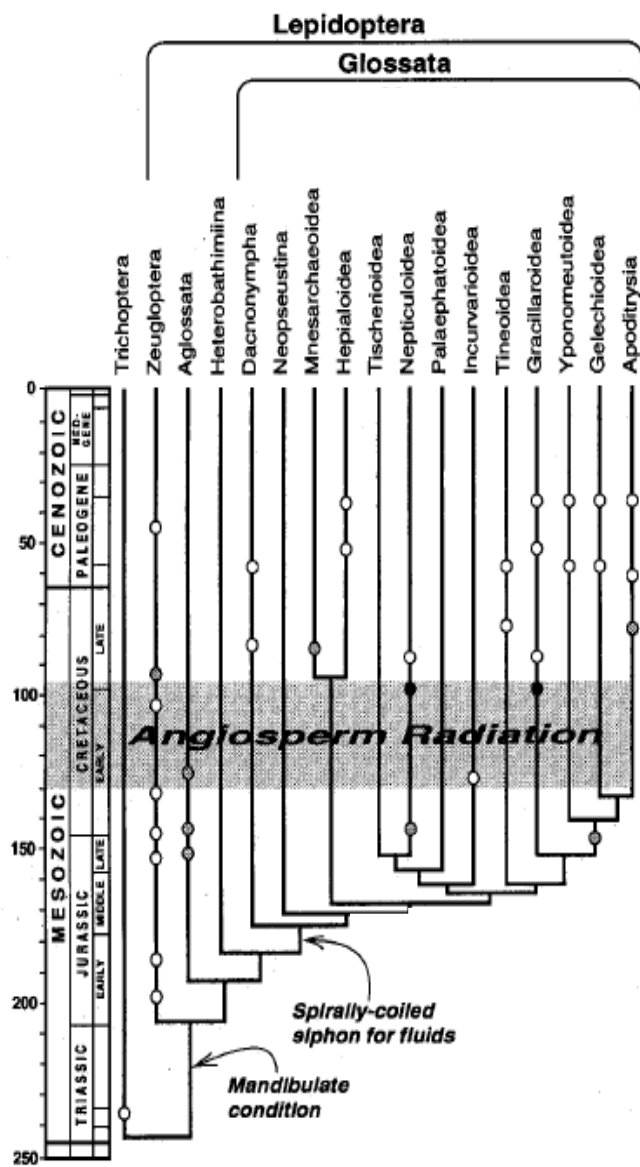


Figura 1 - Filogenia dos Lepidoptera calibrada por ocorrências fósseis (círculos). Os círculos abertos são atribuições confiáveis, os sombreados são possíveis e os escuros são do trabalho de Labandeira *et al.* (1994). O tempo geológico sombreado representa o tempo da radiação das angiospermas. Fonte: Labandeira *et al.*, 1994.

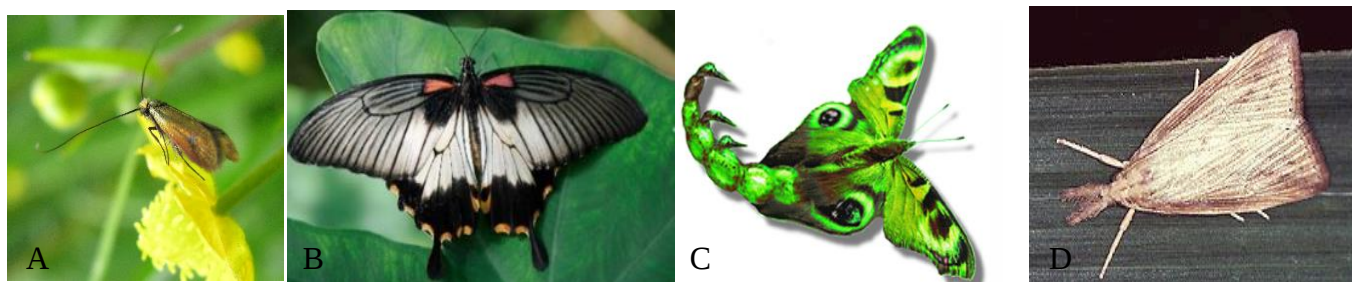


Figura 2 - Diversidade da ordem Lepidoptera. Representantes das quatro subordens: Zeugloptera (A), Aglossata (B), eterobathmiina (C) e Glossata (D). Fonte: Ondřej Zicha, Annina Merk, Lambert, William White, respectivamente.

2.2 DNA mitocondrial de Lepidoptera

Diversos estudos evolutivos de animais, incluindo genética populacional e filogenética utilizam o DNA mitocondrial (mtDNA) (Zhai *et al.*, 2012; Silva-Brandão *et al.*, 2011; Bravo, 2008). O mtDNA é uma molécula circular, rica em adenina e timina (A+T%), de tamanho entre 14 kb a 20 kb e apresenta densidade gênica elevada (Wolstenholme *et al.*, 1992). Os mitogenomas de Lepidoptera seguem a estrutura dos mtDNAs dos metazoários, que é composta por: 13 genes que codificam subunidades proteicas envolvidas na fosforilação oxidativa (três subunidades do citocromo c oxidase: COI, COII e COIII, sete subunidades da NADH desidrogenase: ND1-ND6 e ND4L, citocromo b [CYTB] e duas subunidades da ATP F0 sintase: ATP6 e ATP8), duas subunidades do ribossomo mitocondrial, 22 tRNAs necessários para a tradução das proteínas codificadas pelo mtDNA e uma região não codificante rica em A+T%, como pode ser visualizado na Figura 4 (Li *et al.*, 2010).

Os primeiros mitogenomas de Lepidoptera sequenciados foram da mariposa da seda *Bombyx mori* (Yukuhiro *et al.*, 2002) e da broca de milho

Ostrinia spp. (Coates *et al.*, 2005). Desde então, pelo menos outros 53 mitogenomas foram sequenciados (Zhai *et al.*, 2012). Todos pertencem a espécies da subordem Glossata e estão distribuídos em 13 famílias: Hepialidae (2); Papilionidae (5), Hesperidae (1), Nymphalidae (13), Lycaenidae (3), Pieridae (3), Crambidae (5), Pyralidae (1), Bombycidae (2), Sphingidae (1), Saturniidae (6), Noctuoidea (3), Erebiidae (2), Notodontidae (2), Geometridae (1), Tortricidae (4), Lyonetiidae (1).



Figura 3 - Mapa circular do genoma mitocondrial de *Diatraea saccharalis*. tRNAs são denotados com uma única letra do aminoácido correspondente de acordo com a IUPAC-IUB. Nomes de genes estão entre as duas fitas. F1-F10 são os fragmentos amplificados pelos pares de *primers* utilizados no estudo de Li *et al* (2010). Fonte: Li *et al.*, 2010

A família Hepialidae, infraordem Exoporia apresenta duas espécies com mitogenomas sequenciados: *Thitarodes renzhiensis* e *Ahamus yunnanensis*. As outras 12 famílias pertencem a infraordem Heteroneura. A ordem gênica dos

mtDNAs dos Exoporia é semelhante a do ancestral comum dos artrópodes e diferente da dos Heteroneura. Os mtDNAs das espécies da infraordem Heteroneura apresentam uma translocação do tRNA^{Met} na posição 5' *upstream* do tRNA^{Ile} em relação ao ancestral comum dos artrópodes, como observado na Figura 4 (Coates *et al.*, 2005; Salvato *et al.*, 2008; Li, 2010). Devido ao fato da translocação não ser observada nos Exoporia, acredita-se que o arranjo gênico em Lepidoptera evoluiu após a separação da linhagem dos artrópodes e pode ter se mantido nas infraordens mais antigas de Lepidoptera (Cao *et al.*, 2012). O compartilhamento da ordem gênica em um táxon é uma evidência importante a ser considerada nos estudos de reconstruções filogenéticas (Boore *et al.*, 1999).

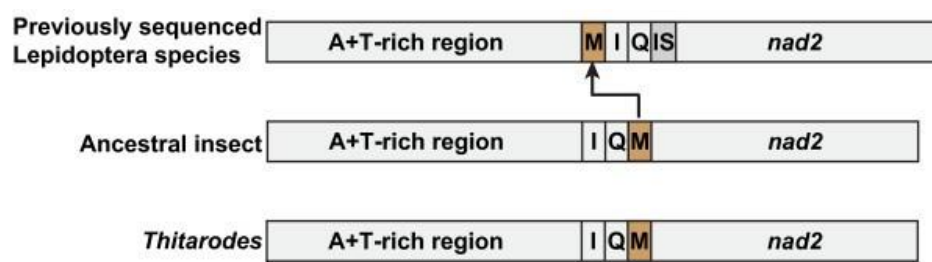


Figura 4 - Representação esquemática dos arranjos de genes mitocondriais. *Thitarodes sp.* (infraordem Exoporia) tem a ordem gênica ancestral. Espécies previamente sequenciadas de Lepidoptera (infraordem Heteroneura) têm o tRNA^{Met} a montante do tRNA^{Ile}, tRNA^{Gln}, e um espaço intergênico (IS). (Fonte: Cao *et al.*, 2012).

Os menores mitogenomas de Lepidoptera pertencem a *Aporia crataegi* (Pieridae), *Pieris melete* (Pieridae) e *Fabriciana nerippe* (Nymphalidae), que apresentam 15.140 pb e o maior mitogenoma pertence a *Thitarodes renzhiensis*

(Hepialidae) com 16.173 pb. A variação corresponde a pouco mais de 1.000 pb e refere-se principalmente a alterações na região rica em A+T% (Cao *et al.*, 2012).

O elevado conteúdo A+T% de Lepidoptera está diretamente relacionado ao viés do uso de códons. As regiões codificantes utilizam com frequência os códons degenerados NNA e NNU (Gibson *et al.*, 2005). A preferência pelo códon NNA pode ser explicada pela abundância do ribonucleotídeo ATP na matriz celular, onde o mRNA é transcrito (Behura e Serverson, 2013). As regiões codificantes raramente usam o códon degenerado NNG para não comprometer a vida-útil do mRNA, uma vez que a guanina é a base menos estável e pode ser facilmente convertida a xantina devido ao estresse oxidativo na mitocôndria. (Gibson *et al.*, 2005).

Existem algumas discrepâncias nos limites dos genes mitocondriais de Lepidoptera. O gene COI apresenta ambiguidade no início da tradução. Em algumas espécies o códon de iniciação é CGA (arginina) (Kim *et al.* 2010); em outras são tetranucleotídeos, como TTAG (Li *et al.*, 2010), ou até mesmo hexanucleotídeos, como TTTTAG (Liu *et al.*, 2012). Outra discrepância refere-se à terminação de determinados genes. A maioria dos genes mitocondriais termina com o códon TAA. Contudo, em algumas espécies, os genes ND3, ND4, ND5 e principalmente COI e COII apresentam códon de terminação incompleto com um único nucleotídeo T. Os códons de parada incompletos originam códons de terminação funcionais após o mecanismo de poliadenilação, que completa o códon truncado (TAA) (Negrisolo *et al.*, 2004).

2.3 Filogenia

Estudos de filogenia molecular são importantes na elucidação do processo evolutivo em todos os níveis de classificação dos organismos. Alguns aspectos devem ser considerados em análises filogenéticas, como: tamanho da amostra; origem gênica; alinhamento das sequências; modelo evolutivo e método filogenético.

Uma amostra com poucos representantes do grupo taxonômico pode não apresentar informações moleculares suficientes para construir uma árvore que condiz com a filogenia correta do grupo. É indicado realizar a análise com o maior número de grupos taxonômicos possível para reduzir a probabilidade da filogenia ser determinada por substituições raras (Phillips, 2000).

Em relação à origem gênica, utilizam-se os genes homólogos ortólogos para a construção de filogenia de espécies, uma vez que possuem a mesma origem e passaram a ter histórias evolutivas distintas. Os genes parálogos (originários de duplicação), os xenólogos (originários de transferência lateral) e os plerólogos (originários de conversão gênica) não são indicados para reconstruções filogenéticas, mas são utilizados na identificação da origem dos genes e na compreensão da evolução dos genomas (Phillips, 2000).

O alinhamento das sequências de DNA é utilizado para identificar as posições homólogas. Regiões homólogas significam nucleotídeos que ocupam a mesma posição, pois eles supostamente têm um ancestral comum, porém não são necessariamente iguais (Silva, 2008).

As regiões evolutivamente muito distantes (mesmo quando corretamente alinhadas) podem apresentar um alto grau de saturação mutacional. A saturação de substituições diminui a informação filogenética contida nas sequências e prejudica as análises filogenéticas que envolvem grupos antigos. Quando as

sequências apresentam alto grau de saturação, a similaridade entre as sequências depende inteiramente da semelhança nas frequências dos nucleotídeos, que muitas vezes não reflete as verdadeiras relações filogenéticas (Xia e Lemey, 2009).

A escolha de um modelo teórico de evolução que reflita o processo evolutivo dos dados moleculares também caracteriza um passo importante. As mudanças nas sequências de nucleotídeos podem ser justificadas por um modelo evolutivo que trabalha com suposições. Quando se infere uma filogenia, estima-se a história evolutiva baseada nas alterações das características herdadas que refletem na sequência de DNA (Vinuesa *et al.*, 2005).

Os métodos filogenéticos são algoritmos de reconstrução de filogenia, os quais podem ser divididos em três grupos: distância, métodos baseados em caracteres e probabilísticos. Os métodos probabilísticos são dois: máxima verossimilhança e inferência bayesiana (Tavares, 2005; Simon *et al.*, 1994).

O método da máxima verossimilhança busca a hipótese filogenética com maior probabilidade de explicar a distribuição dos estados de caráter nos táxons atuais com base em um modelo evolutivo previamente definido e estado de caracteres observados nos terminais da árvore filogenética (Rogers e Swofford, 1999). Por esse método probabilístico *a priori*, diversas hipóteses são testadas pela multiplicação das suas probabilidades para cada sítio (Rogers e Swofford, 1999; Holder e Lewis, 2003; Felsenstein, 2004). Para avaliar o quanto os dados sustentam cada clado da melhor topologia obtida, utiliza-se o método *bootstrap*. Esse algoritmo permite avaliar individualmente o quanto as partes da topologia são suportadas pelo conjunto de dados (Holder e Lewis, 2003).

A análise bayesiana é baseada no conceito de probabilidade *a posteriori*, calculado por meio do teorema de Bayes. O teorema combina a probabilidade *a priori* da topologia, com sua probabilidade de explicar a distribuição dos estados de caráter nos táxons atuais com base em um modelo evolutivo. A escolha da melhor topologia implica em calcular a probabilidade *posteriori* para todas as hipóteses, assim como sua integração com todas as possíveis combinações de tamanho de ramos e valores dos parâmetros dos modelos de substituição de nucleotídeos (Huelsenbeck *et al.*, 2001; Holder e Lewis 2003; Felsenstein, 2004). A aproximação das probabilidades posteriores das topologias é realizada através da Cadeia Markov de Monte Carlo (MCMC) (Felsenstein, 2004).

2.4 Marcadores Moleculares

Muitos genes mitocondriais são utilizados como marcadores moleculares em estudos de filogenia molecular e na classificação de espécies. Os marcadores moleculares são locus gênicos variáveis utilizados para inferir padrões de diversidade espaço-temporal dos organismos (Bleidorn *et al.*, 2003). Em muitos casos, as técnicas moleculares auxilia na identificação de espécies que não conseguem ser diferenciadas a partir de dados morfológicos (Köhler *et al.*, 2006).

O genoma mitocondrial é muito utilizado como marcador molecular tanto para estudos evolutivos entre táxons distantemente relacionados como em estudos filo-biogeográficos de espécies estreitamente relacionadas. O mtDNA oferece vantagens sobre outros marcadores em análises filogenéticas devido as suas características, como: cópia única; presença de polimorfismos frequentes; e herança uniparental (materna). Essas características são vantajosas por

permitirem a reconstrução de linhas evolutivas sem que haja interferência de parálogos, dos efeitos da hereditariedade biparental e da inerente recombinação existente no DNA nuclear (Pakendorf e Stoneking, 2005).

Em Lepidoptera os genes mitocondriais mais utilizados como marcadores moleculares em estudos evolutivos são: ND3, ATP6, COI, COIII, CYTB, ND1, ND5 e 16S (Duchêne *et al.*, 2011; Hong *et al.*, 2008; Salvato *et al.*, 2008; Kim *et al.*, 2010; Ohshima *et al.*, 2010; Li *et al.*, 2005; Strutzenberger *et al.*, 2012; Jeratthitikul *et al.*, 2013; Ashfag *et al.*, 2013; Hong *et al.*, 2008; Tachi, 2013; Albre *et al.*, 2008; Katto *et al.*, 2005; Nazari *et al.*, 2007; Meng *et al.*, 2008).

Visando os vários estudos com marcadores moleculares em Lepidoptera, este trabalho originou-se com a finalidade de sugerir uma forma alternativa da análise evolutiva dos Lepidoptera sem a necessidade de sequenciar mitogenomas completos. Esta análise deverá resultar na identificação de regiões do mitogenoma que reflitam fidedignamente a filogenia do grupo.

3. Objetivos

3.1 Geral

Comparar o grau de adequação de diferentes regiões do genoma mitocondrial utilizadas em estudos filogenéticos de Lepidoptera, determinando aquelas que melhor refletem a sua história evolutiva.

3.2 Específicos

- Inferir as regiões mais informativas dos mitogenomas de Lepidoptera para a filogenia do grupo;
- Verificar a eficácia da inferência no mitogenoma de *Diatraea flavipennella* em relação a dos outros Lepidoptera em função do sequenciamento das regiões marcadoras;
- Analisar a influência da terceira posição dos códons das regiões informativas na filogenia de Lepidoptera.

4. Material e Métodos

4.1 Recuperação das sequências

Foram recuperadas as sequências nucleotídicas e as anotações de 55 mitogenomas de Lepidoptera (Quadro 1) disponíveis no banco de dados do Centro Nacional de Informação de Biotecnologia (NCBI). Recuperou-se também a sequência nucleotídica de *Drosophila melanogaster* (Classe Insecta, Filo Artropoda), utilizada como grupo externo nas análises filogenéticas.

Quadro 1 – Espécies de Lepidoptera utilizadas neste trabalho

Infraordem	Clado	Superfamília	Família	Espécie
Exoporia	-	Hepialoidea	Hepialidae	<i>Thitarodes renzhiensis</i> <i>Ahamus yunnanensis</i>
Heteroneura	Obtectomera	Papilionoidea	Papilionidae	<i>Papilio maraho</i> <i>Papilio bianor</i> <i>Papilio machaon</i> <i>Teinopalpus aureus</i> <i>Parnassius bremeri</i>
			Hesperiidae	<i>Ochlodes veneta</i>
			Nymphalidae	<i>Issoria lathonia</i> <i>Melitaea cinxia</i> <i>Athymia sulpitia</i> <i>Euploea mulciber</i> <i>Libythea celtis</i> <i>Fabriciana nerippe</i> <i>Kallima inachus</i> <i>Apatura ilia</i> <i>Apatura metis</i> <i>Argynnis hyperbius</i> <i>Calinaga davidis</i> <i>Hipparchia autonoe</i> <i>Acraea issoria</i>
			Lycaenidae	<i>Celastrina hersilia</i> <i>Protantigius superans</i> <i>Spindasis takanonis</i>
			Pieridae	<i>Pieris rapae</i> <i>Pieris melete</i>

		Pyraloidea		<i>Aporia crataegi</i>
			Crambidae	<i>Cnaphalocrocis medinalis</i> <i>Chilo suppressalis</i> <i>Diatraea saccharalis</i> <i>Ostrinia nubilalis</i> <i>Ostrinia furnacalis</i>
			Pyralidae	<i>Corcyra cephalonica</i>
	Macroheterocera	Bombycoidea	Bombycidae	<i>Bombyx mori</i> <i>Bombyx mandarina</i>
			Sphingidae	<i>Manduca sexta</i>
			Saturniidae	<i>Actias selene</i> <i>Samia cynthia</i> <i>Antheraea yamamai</i> <i>Antheraea pernyi</i> <i>Eriogyna pyretorum</i> <i>Saturnia boisduvalli</i>
		Noctuoidea	Noctuoidea	<i>Sesamia inferens</i> <i>Helicoverpa armigera</i> <i>Spodoptera exigua</i>
			Erebidae	<i>Lymantria dispar</i> <i>Hyphantria cunea</i>
			Notodontidae	<i>Phalera flavescent</i> <i>Ochrogaster lunifer</i>
		Geometroidea	Geometridae	<i>Phthonandria atrilineata</i>
	Apoditrysia	Tortricoidea	Tortricidae	<i>Acleris fimbriana</i> <i>Grapholita molesta</i> <i>Spilonota lechriaspis</i> <i>Adoxophyes honmai</i>
	Ditrysia	Yponomeutoidea	Lyonetiidae	<i>Leucoptera malifoliella</i>

Fonte: van Nieukerken *et al.*, 2011

4.2 Análise de sintenia

A partir das sequências genômicas recuperadas, foi realizada uma análise de sintenia para avaliar a existência de alterações na ordem gênica de Lepidoptera. Foram selecionadas prioritariamente as regiões analisadas correspondentes aos genes: COI, ATP6, COIII, CYTB, ND3, ND5, ND1 e 16S, por serem os mais utilizados em análises filogenéticas de Lepidoptera. É importante

conhecer a ordem gênica dos mitogenomas, uma vez que alterações podem comprometer o alinhamento múltiplo das sequências e consequentemente a análise filogenética.

A análise de sintenia foi realizada utilizando os parâmetros *default* do programa MAUVE (Darling *et al.*, 2004). O programa realiza o alinhamento de regiões ortólogas entre os genomas e permite a identificação da ordem gênica. As sequências de Lepidoptera que apresentaram alterações na ordem gênica tiveram suas posições manualmente alteradas de acordo com àquela característica de Lepidoptera para permitir o alinhamento com os demais mitogenomas de Lepidoptera.

4.3 Sequenciamento das regiões filogeneticamente informativas no DNA mitocondrial de *Diatraea flavipennella*

Diatraea flavipennella (Fabricius, 1974) pertence à família Crambidae, subordem Glossata e é popularmente conhecida como broca amarela da cana-de-açúcar. A espécie se destaca no Brasil por causar prejuízos econômicos ao atacar as plantações de cana-de-açúcar dos estados nordestinos, do Rio de Janeiro e de Minas Gerais (Mendonça *et al.*, 1996).

As regiões marcadoras foram sequenciadas a partir do mtDNA de *Diatraea flavipennella* para verificar a eficiência da metodologia aplicada para analisar filogeneticamente espécies que não apresentam seus mitogenomas completos sequenciados.

4.3.1 Material Biológico

As larvas de *D. flavipennella* foram cultivadas no Laboratório de Interação Insetos-Tóxicos da Universidade Federal Rural de Pernambuco (UFRPE) sob a supervisão do Prof. Dr. Herbert Siqueira e gentilmente cedidas para este trabalho. As larvas foram mantidas a temperatura de -20°C até o momento em que seriam processadas.

4.3.2 Extração de DNA e amplificação

Extraíu-se o mtDNA das larvas por meio do método fenol-clorofórmio descrito por Sambrook *et al.* (1989). Para amplificar as regiões de interesse, utilizaram-se cinco pares de *primers* degenerados para o filo Arthropoda (LABBE-Dip 1 - LABBE-Dip 6 [Santos, 2012]; CYTB [Hodgkinson *et al.*, 2002]) e dois desenvolvidos especificamente para este trabalho (Tabela 1).

Para desenhar os dois pares de *primers*, foram realizados alinhamentos múltiplos com as sequências recuperadas dos mitogenomas de Lepidoptera, um para o gene ND5 e outro para o gene 16S. Os alinhamentos foram realizados com o programa MUSCLE (Edgar, 2004), inserido no pacote do MEGA 5.1 (Tamura *et al.*, 2011), utilizando os parâmetros *default* e o código genético mitocondrial de invertebrados. Cada alinhamento apresentou cerca de 2.000 pb para permitir a cobertura total dos genes de interesse.

Foram selecionadas regiões conservadas de cada alinhamento que seriam possíveis candidatos à definição dos pares de *primers*. Utilizaram-se os programas Primer3Plus (Untergasser *et al.*, 2007) e OligoAnalyzer v. 3.1 da *Integrated DNA Technologies* (IDT) para selecionar as regiões que apresentavam

valores de temperatura de *melting* (TM) elevados e valores de formações de grampos e de dímeros baixos. O *software* OligoAnalyzer v. 3.1 está disponível em <http://www.idtdna.com/analyzer/applications/oligoanalyzer/>. As duas regiões que apresentaram as melhores características foram selecionadas para a definição dos pares de *primers* Df-ND5 (para o gene ND5) e Df-Leu (para o gene 16S).

Tabela 1 - Pares de *primers* utilizados para *Diatraea flavipennella*.

Primer	Sequências	TM (°C)	Posição (pb)	Tamanho (pb)
LABBE-Dip1	5 ' TCAACTAAYCATAARGATATTGG 3 '	46,1 – 50	1.442 -	1.098
	5 ' CAACWACATAATAWGTATCGTG 3 '	46,3 - 47,3	2.540	
LABBE-Dip2	5 ' AGTTGATTAGCTACWTTWCAYGG 3 '	50,4 – 53	2.381-	361
	5 ' TCTGAGTAWCGTCGAGGTATTCC 3 '	55,3 - 55,7	2.742	
LABBE-Dip5	5 ' TTYCCTTATATTTTACAAGAAC 3 '	44,9 - 46,7	4.337-	801
	5 ' GGAATTTGRAAGGGATTAAA 3 '	45,8 - 47,9	5.138	
LABBE-Dip6	5 ' CTATATAYCAATGATGACGAGA 3 '	46,7 - 48,9	4.863-	1.060
	5 ' TGRTTTCATTCATGRTATAATCC 3 '	46,4 - 50,7	5.885	
Df-ND5*	5 ' CCTTGACCTCTATAAATCTCTC 3 '	49,6	6.570-	1.008
	5 ' CTTTTAGACCTTGATTACCAGC 3 '	51,2	7.578	
CYT B	5 'CAY ATT CAA CCW GAA TGA TA 3 '	46,6 – 47,6	11.458-	529
	5 ' GGC AYW TTG CCT CGA WTT CGW TAT GA 3 '	57,1 – 59,6	11.986	
Df-Leu*	5 ' AATCATTACATTTTTCTGCC 3 '	46,5	12.868-	1.049
	5 ' ATTTTATATTTTAAGGGATAAGC 3 '	44,7	13.917	

* Pares de *primers* construídos neste trabalho

As reações de PCR foram definidas para um volume final de 25 µL, utilizando: 50 ng/µL de DNA; 1 µM de cada *primer*; 0,5 µM de dNTP; 2,5 µL de

tampão x10; 1,5 mM de MgCl₂; e 1,5 U de Taq Platinum Invitrogen. As condições de amplificação utilizadas foram: desnaturação inicial por 4 min a 94°C; seguida por 35 ciclos de: desnaturação a 94 °C por 1 min, anelamento de 42 °C - 50 °C por 1 min, e alongação a 72 °C por 1 min e 30 seg; a extensão final de 7 min.

Os produtos amplificados foram visualizados em gel de agarose a 1% e purificados seguindo as instruções do fabricante (Promega).

4.3.3 Clonagem e sequenciamento

Os *amplicons* foram clonados no pGEM-T, utilizando o kit pGEM®-T Vector Systems, seguindo as recomendações do fabricante (Promega).

O sequenciamento dos *amplicons* foi realizado por meio do sequenciador ABI 3500 (Applied Biosystems) da Unidade Multiusuária de Sequenciamento do Centro de Ciências Biológicas (CCB) da UFPE utilizando o kit BigDye® Terminator v3.1 Cycle Sequencing (Applied Biosystems). O sequenciamento foi realizado ao menos quatro vezes para cada *amplicon*.

4.3.4 Análises das sequências

A análise dos cromatogramas foi realizada por meio do pacote de programas Staden (Staden *et al.*, 2000). O pacote contém funcionalidades (e.g. Trev, Pregap4 e Gap4), que permitem as análises dos cromatogramas advindos de sequenciadores automáticos. Os *contigs* foram formados considerando apenas as sequências que apresentavam bases com valores de PHRED iguais ou superiores a 30. As sequências primárias de nucleotídeos foram comparadas com as sequências do GenBank utilizando a Ferramenta de Busca de Alinhamento Local Básico (BLAST) (Altschul *et al.*, 1990).

4.4 Análise das regiões filogeneticamente informativas para Lepidoptera

Nesta etapa foram consideradas apenas as sequências obtidas a partir do sequenciamento de *D. flavipennella* e os seguimentos correspondentes dos 55 mitogenomas de Lepidoptera recuperados.

As análises foram divididas em dois grupos: (1) região delimitada pelos genes amplificados; (2) região delimitada pelos pares de *primers*. No grupo 1 foram consideradas apenas a informação contida nos genes individualmente e, uma vez que todos os genes já foram utilizados em análises filogenéticas de Lepidoptera, esperou-se identificar o que apresentava características de um bom marcador molecular para a ordem Lepidoptera. Já no grupo 2 foram consideradas as regiões amplificadas pelos pares de *primers* como um todo, analisando as sequências que apresentavam mais de um gene e tRNAs.

Foi realizado um alinhamento múltiplo para cada região no programa MUSCLE, incorporado no MEGA v. 5.1, utilizando o código genético mitocondrial de invertebrados. O alinhamento das regiões não codificadoras de proteínas foi realizado com as sequências nucleotídicas; enquanto que para as regiões codificantes, as sequências nucleotídicas foram traduzidas para sequência de aminoácidos, realizado o alinhamento e a sequência transformada novamente para nucleotídeos.

4.4.1 Análise de entropia

Para cada alinhamento foi realizado uma análise de entropia utilizando o programa Análise de Dados em Biologia Molecular e Evolução - DAMBE (Xia e Xie, 2001), que permite a obtenção, organização, manipulação, alinhamento,

análise das sequências de dados moleculares e pode ser utilizado para o cálculo das distâncias genéticas (Xia & Xie, 2001). A metodologia de análise de entropia aplicada foi proposta por Batista e colaboradores (2011) com a finalidade de se retirar possíveis regiões que interferissem no sinal filogenético. O resultado da análise de entropia do DAMBE é expresso em gráfico, cujos valores de entropia determinam o grau de conservação de uma dada região. Quanto menor for o valor de entropia, maior a probabilidade das sequências serem homólogas. Partes das sequências que apresentaram valores acima do ponto de corte (entropia $\geq 1,0$) foram retiradas da sequência inicial observando a fase de leitura.

4.4.2 Análises da variação nucleotídica na sequências analisadas

Dentro de cada grupo foram analisadas as situações: com a terceira posição de códons (A) e exclusão da terceira posição de códons (B). O objetivo de se estudar as duas situações foi verificar a influência da terceira posição nas análises filogenéticas de Lepidoptera baseado em dados moleculares mitocondriais, uma vez que vários artigos discutem a influência dessa posição nas análises filogenéticas de outros grupos. Para melhor entendimento dos tipos de análises realizadas neste trabalho, pode-se visualizar o Quadro 2, abaixo.

Quadro 2 - Tipos de análises realizadas neste trabalho

Análise	A	B
1	GCPs e 16S + terceira posição dos códons	GCPs e 16S – terceira posição dos códons
2	<i>Amplicons</i> + terceira posição dos códons	<i>Amplicons</i> – terceira posição dos códons

GCPs – genes codificadores de proteínas

O grupo 1 foi dividido em 15 testes, sendo dois para os GCPs estudados: COI, ATP6, COIII, ND3, ND5, ND1, CYTB e um para o 16S. O grupo 2 foi dividido em 11 testes, referente aos pares de *primers* da Tabela 1: LABBE-Dip1, LABBE-Dip2, LABBE-Dip5, LABBE-Dip6, CYT B, Df-Leu. Apenas o par de *primers* Df-ND5 não foi considerado no grupo 2, uma vez que amplifica apenas o gene ND5 e por isso, já havia sido analisado no grupo 1. O par de *primer* Df-Leu foi excluído da análise B por corresponder a região ribossomal 16S. Todos os 26 testes estão descritos no Quadro 3.

Quadro 3 – Testes realizados neste trabalho

GCPs e 16S + terceira posição dos códons	GCPs e 16S - terceira posição dos códons	<i>Amplicons</i> + terceira posição dos códons	<i>Amplicons</i> - terceira posição dos códons
1.A.COI	1.B.COI	2.A. LABBE-Dip1	2.B. LABBE-Dip1
1.A. ATP6	1.B. ATP6	2.A. LABBE-Dip2	2.B. LABBE-Dip2
1.A. COIII	1.B. COIII	2.A. LABBE-Dip5	2.B. LABBE-Dip5
1.A. ND3	1.B. ND3	2.A. LABBE-Dip6	2.B. LABBE-Dip6
1.A. ND5	1.B. ND5	2.A. CYT B	2.B. CYT B
1.A. ND1	1.B. ND1	2.A. Df-16S	
1.A. CYTB	1.B. CYTB		
1.A. 16S			

Calculou-se no programa MEGA v. 5.1, a partir de cada alinhamento, o número total de sítios alinhados, sítios variáveis, sítios parcimoniosos informativos e *singletons*. Nesta análise pode-se identificar o grau de variação entre as sequências dos mitogenomas de Lepidoptera.

4.4.3 Sinal filogenético

O sinal filogenético de cada teste foi analisado no programa TreePuzzle v. 5.2 (Schmidt *et al.*, 2002), utilizando *likelihood mapping*, que analisa os grupos de quatro sequências aleatórias chamadas de quartetos. O somatório dos quartetos 1, 2 e 3 representa o valor do sinal filogenético. A escala do sinal filogenético foi determinado empiricamente como: regular para valores > 70; bom para valores entre 70 e 90; e ótimo para valores >90.

4.4.4 Teste de saturação

O nível de saturação das sequências foi verificado por meio do programa DAMBE. O objetivo desta etapa foi analisar a qualidade dos dados de cada alinhamento. O nível de saturação das sequências foi avaliado através da análise gráfica do número transições e transversões (ti/tv) versus a distância genética corrigida segundo o modelo de Felsenstein 84.

4.4.5 Análise filogenética

O modelo de substituição de nucleotídeos para cada teste foi selecionado por meio do programa jModelTest v. 3.7 (Posada e Crandall, 1998), utilizando a estrutura de seleção de modelos Bayesian Information Criterion (BIC), mostrado na Tabela 2.

Duas análises filogenéticas foram realizadas para cada teste: uma baseada em máxima verossimilhança, com o programa PhyML (Guindon *et al.*, 2010) e outra em inferência bayesiana, com o software MrBayes (Huelsenbeck e Ronquist, 2001). No PhyML foi utilizado o método que une características do *Nearest Neighbor Interchange* (NNI) e *Subtree Pruning and Regrafting* (SPR) para

estimar a topologia da árvore. Uma árvore de BioNJ (algoritmo desenvolvido por Gascuel [1997] a partir do *neighbor-joining*) foi utilizada como árvore inicial e os táxons foram adicionados aleatoriamente. A robustez da árvore foi verificada pelo teste de confiança *bootstrap* utilizando-se 1.000 pseudo-réplicas. No MrBayes a topologia final foi estimada com 5.000.000 de interações descartando os primeiros 1.250.000 ciclos. Foram considerados os valores de agrupamento acima de 50 para PhyML e acima de 0,9 para MrBayes. Utilizou-se *D. melanogaster* como grupo externo para ambas as análises. As árvores filogenéticas foram visualizadas no programa FigTree v1.4.0, disponibilizado online no site: <http://tree.bio.ed.ac.uk/software/>.

Tabela 2 – Modelo evolutivo dos testes

Teste	Modelo evolutivo
1.A.COI / 1.B.COI	GTR+I+G
1.A.ATP6 / 1.B.ATP6 / 1.A.COIII / 2.B.COIII	TVM+I+G
1.A.ND3 / 1.B.COI	TIM2+I+G
1.A.ND5 / 1.B.ATP6	TVM+G
1.A.CYTB	TrN+I+G
1.A.ND1	HKY+G
1.A.16S	HKY+I+G
2.A.LABBE-Dip1 / 2.B.LABBE-Dip1 / 1.B.COIII	TIM1+I+G
2.A.LABBE-Dip2	TPM3uf+I+G
2.A.LABBE-Dip5 / 2.B.LABBE-Dip5	TVM+I+G
2.A.LABBE-Dip6	GTR+I+G
2.A.CYT B	TrN+I+G
2.A.Df-16S	TPM2uf+I+G

4.5 Concatenação

Cada um dos grupos tiveram suas sequências concatenadas. No grupo 1 concatenaram-se todos os GCPs e 16S; enquanto que no grupo 2 concatenaram-

se todas as regiões delimitadas pelos pares de *primers*. Para unir cada grupo foi necessário acoplar as sequências dos alinhamentos no programa MEGA 5. No grupo 2, retiraram-se as sequências duplicadas nas regiões de sobreposição.

O modelo evolutivo para os dois alinhamentos foi GTR+I+G, calculado no programa J-ModelTest. Foram geradas duas árvores filogenéticas para cada concatenação, sendo uma no programa PhyML e a outra no MrBayes.

5. Resultados

5.1 Análise de sintenia

A análise de sintenia dos 55 mitogenomas de Lepidoptera mostrou que as regiões de interesse apresentaram em geral a ordem gênica conservada. Contudo, foi identificada uma translocação de fita no mitogenoma de *L. malifoliella* e outra no de *A. selene*. Em *L. malifoliella* a translocação ocorreu com o gene 16S e em *A. selene* com o gene tRNA^{Leu}. Esta última região é considerada importante por anteceder o 16S e foi utilizada na construção do par de *primer* Df-Leu.

A sequência de 16S de *L. malifoliella* e do tRNA^{Leu} de *A. selene* foram posicionados na mesma orientação que as demais sequências de Lepidoptera e permitiu o alinhamento das sequências. O resultado da análise de sintenia de Lepidoptera encontra-se na Figura 5.

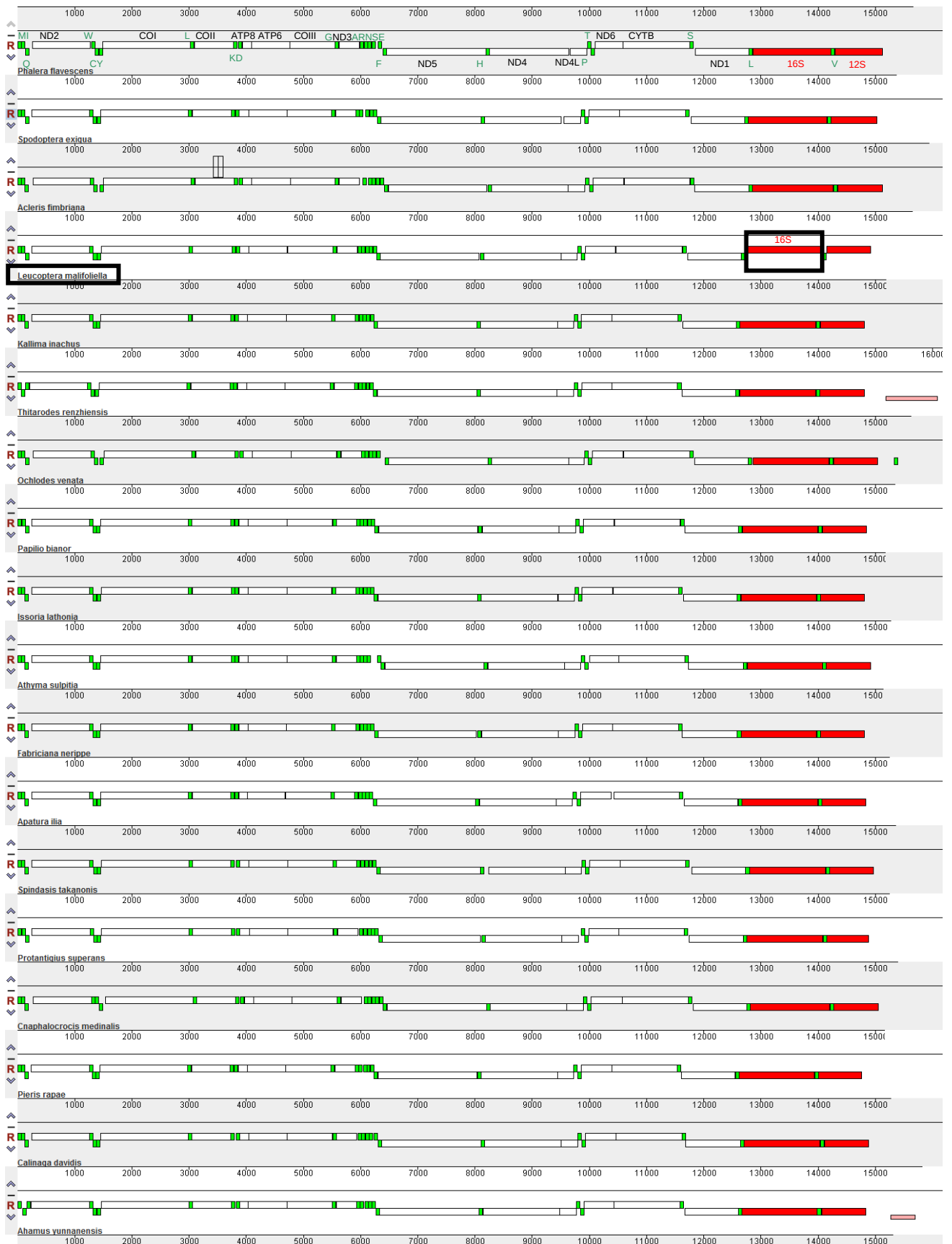
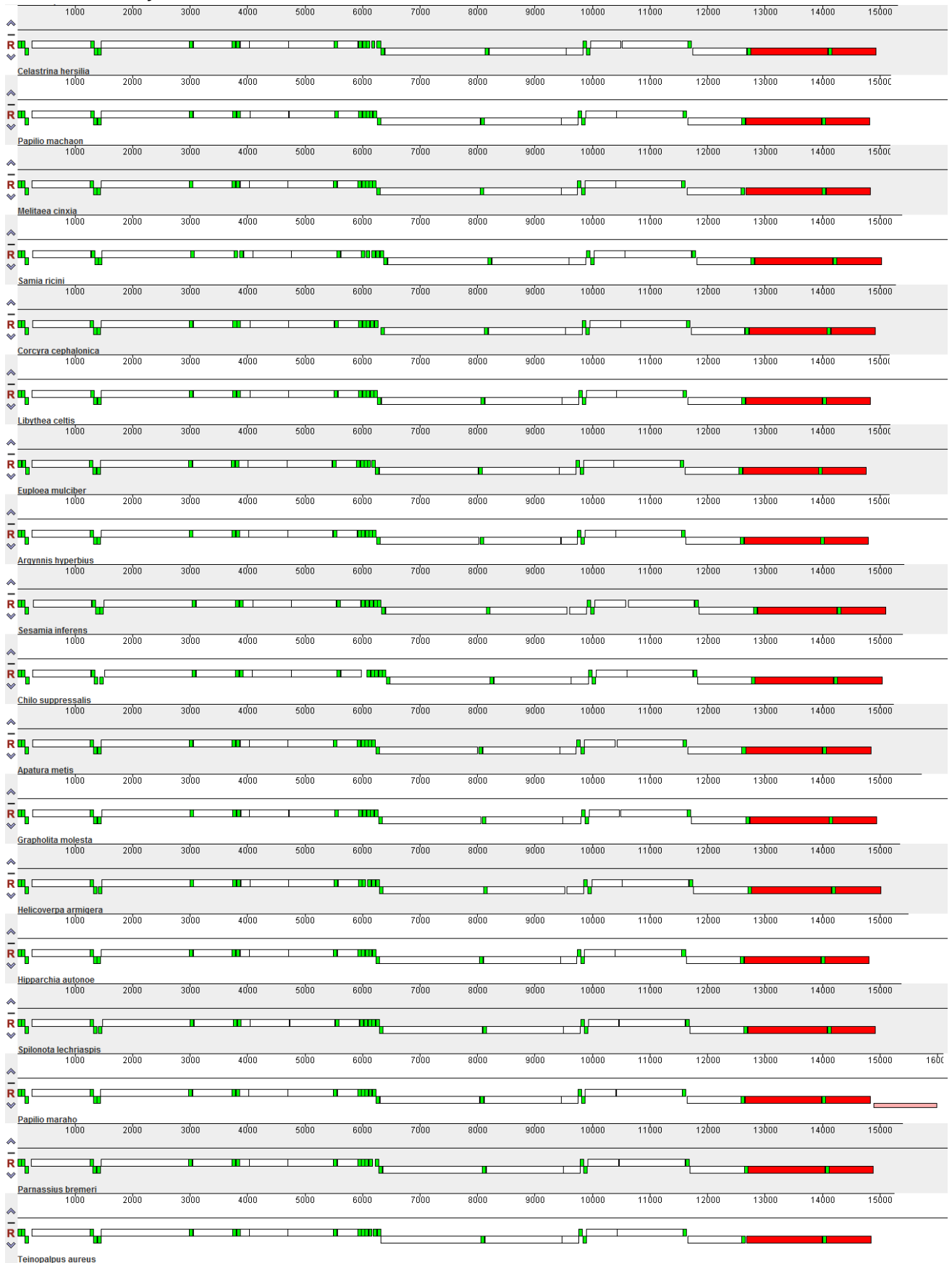


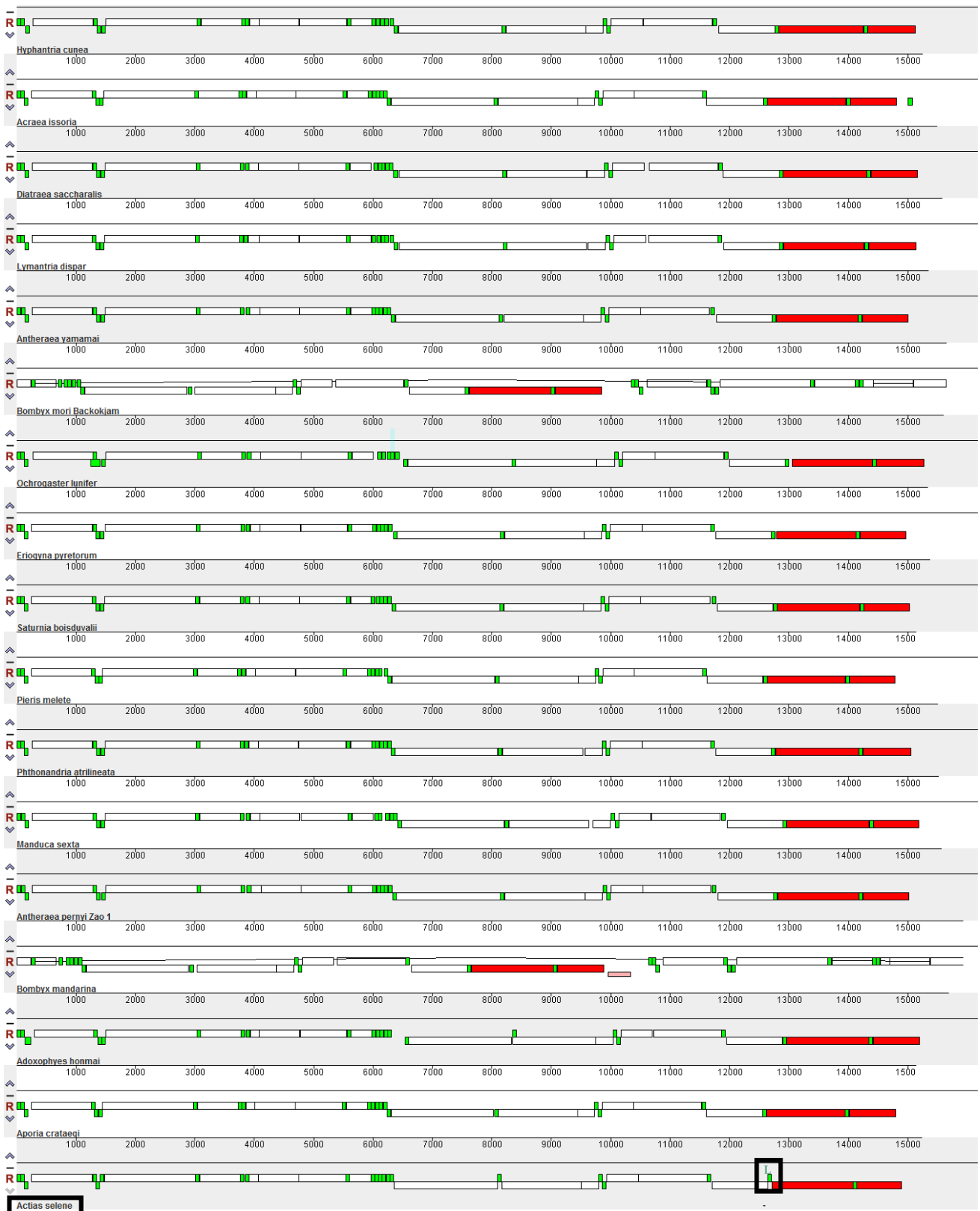
Figura 5 - Representação linear dos mtDNAs de Lepidoptera. Os genes estão representados nas fitas pela cor branca, os tRNAs em verde e os rRNAs em vermelho. As alterações na ordem gênica estão destacadas por quadrados pretos. (Fonte: MAUVE).

continuação...



continua

continuação...



5.2 Sequenciamento das regiões filogeneticamente informativas no mitogenoma de *Diatraea flavipennella*

As regiões amplificadas do mtDNA de *D. flavipennella* correspondem aos genes COI, ATP6, COIII, tRNA^{Gly}, ND3, ND5, CYTB, tRNA^{Ser}, ND1, tRNA^{Leu} e 16S de *D. saccharalis*, como podem ser visualizadas na Figura 6.

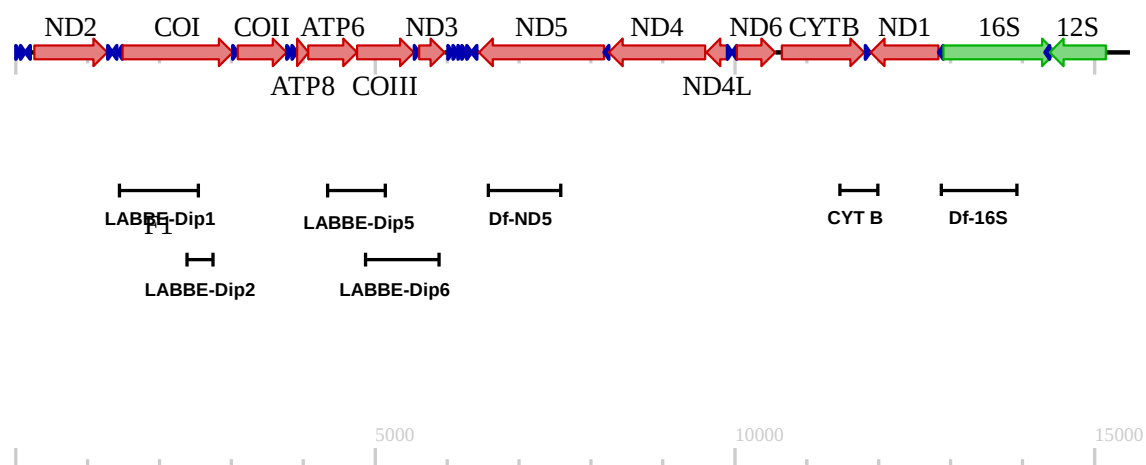


Figura 6 – Representação linear do mitogenoma de *D. saccharalis* e produto dos pares de primers utilizados para *D. flavipennella*. As direções de síntese estão demonstradas pelas setas; as vermelhas representam as regiões codificadoras de proteínas; as azuis, os tRNAs; as verdes, os rRNAs. Os produtos dos pares de primers estão representados por linhas pretas.

Foram obtidos 4.604 pb a partir dos sete pares de primers utilizados, o que equivale a 30% do mitogenoma de *D. saccharalis*. Os amplicons correspondentes aos segmentos do mtDNA de *D. flavipennella* estão mostrados na Figura 7.

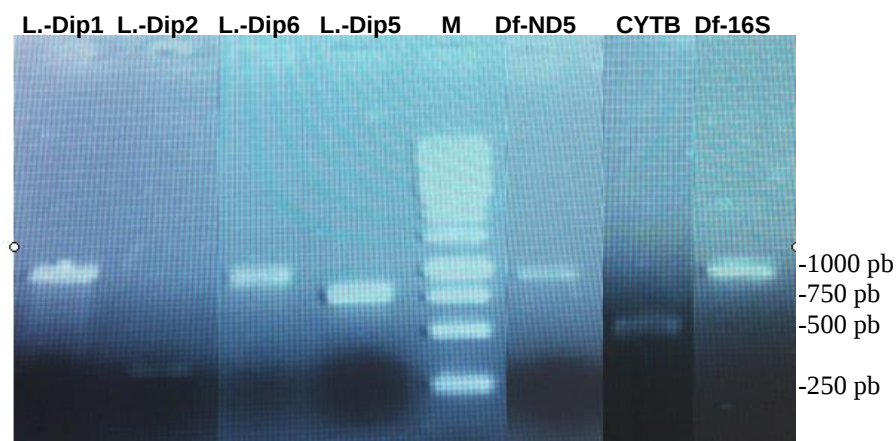


Figura 7 – Gel dos produtos amplificados dos pares de *primers* utilizados. M representa o marcador. As fotos originais encontram-se no APÊNDICE A.

Devido às sobreposições entre os pares de *primers* foram formados dois *contigs*. O primeiro apresentou 1.302 pb e foi formado pelos *amplicons* LABBE-Dip1 e LABBE-Dip2 que em conjunto correspondem a região central do gene COI. O segundo *contig* foi formado por LABBE-Dip5 e LABBE-Dip6 (1.378 pb) que continha os genes completos de COIII, do tRNA^{Gly} e regiões dos genes ATP6 e NAD3, nesta ordem.

Os índices de similaridade das regiões sequenciadas de *D. flavipennella* detectadas pelo BLAST variaram entre 88% a 100% em relação a *D. saccharalis*. A cobertura da sequência de nucleotídeos dos *amplicons* variou de 10% a 100%, (Tabela 3).

Tabela 3 - Informações sobre as regiões de *Diatraea flavipennella* sequenciadas

Regiões sequenciadas	Par de <i>primers</i>	Tamanho (pb)	Cobertura da região sequenciada (%)	Similaridade com <i>D. saccharalis</i> (%)
COI	LABBE-Dip1+	1.301	85	91
	LABBE-Dip2			
ATP6	LABBE-Dip5	277	41	91
COIII	LABBE-Dip5+	789	100	90
	LABBE-Dip6			
tRNA ^{Gly}	LABBE-Dip6	67	100	96
ND3	LABBE-Dip6	244	70	89
ND5	Df-ND5	409	24	88
CYTB	CYT B	339	34	88
tRNA ^{Ser}	CYT B	66	100	100
ND1	CYT B	96	10	91
tRNA ^{Leu}	Df-16S	28	42	100
16S	Df-16S	988	70	89

5.3 Análise das regiões informativas para a filogenia de Lepidoptera

5.3.1 Análise de entropia

Foram gerados 14 gráficos de análise de entropia, sendo oito correspondentes aos GCPs e mais o 16S (Figura 8) e seis correspondentes aos *amplicons* (Figura 9). Os genes COIII, ND3 e 16S apresentaram regiões que ultrapassaram o ponto de corte (entropia $\geq 1,0$). Apenas a sequência da região delimitada por Df-16S, que corresponde ao segmento tRNA^{Leu} e 16S, apresentou regiões acima deste ponto de corte. Essas regiões foram identificadas pela análise de entropia e retiradas do alinhamento múltiplo com a finalidade de evitar perturbações nas análises filogenéticas.

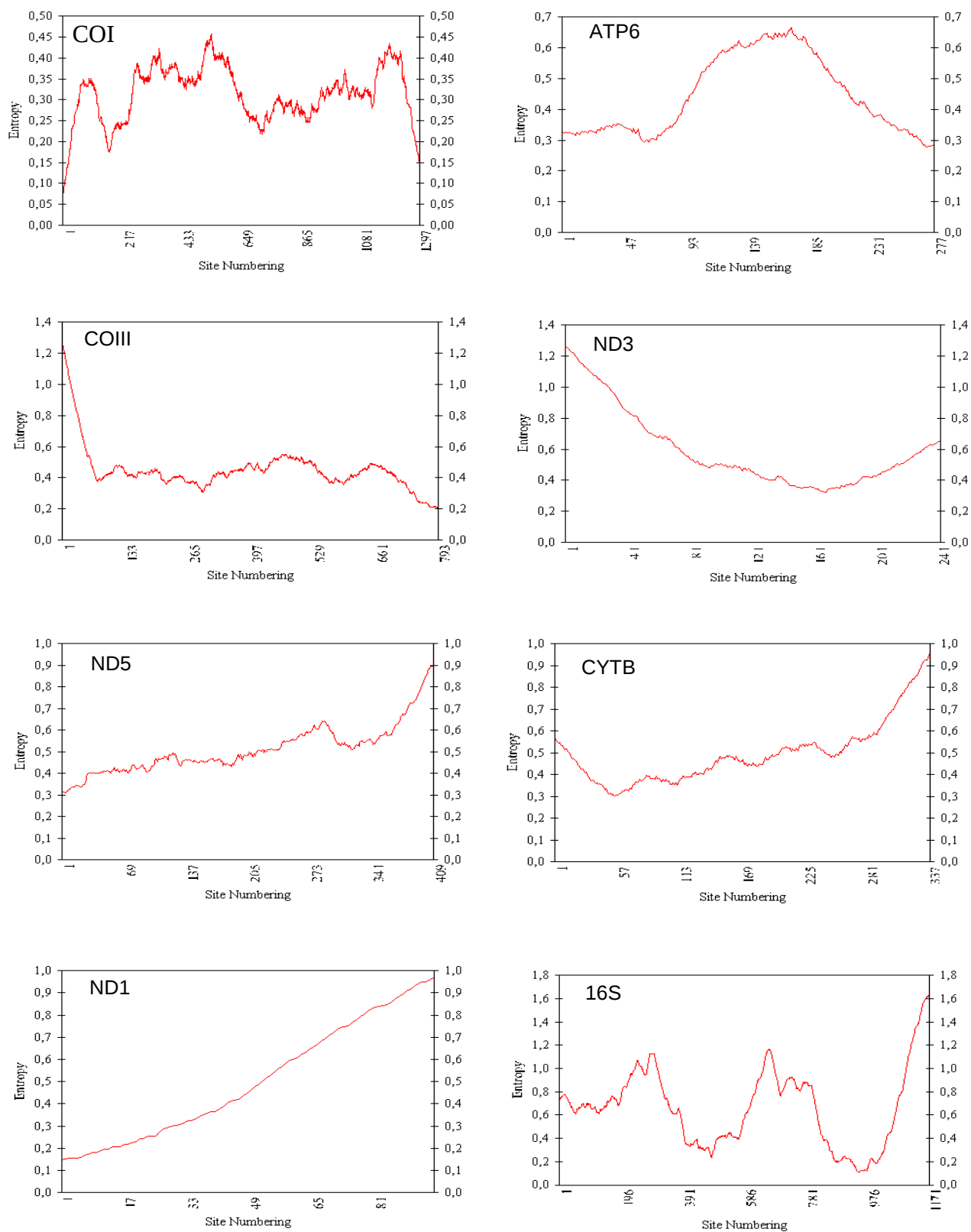


Figura 8 - Análise de entropia dos genes. O eixo das abcissas representa a posição na região analisada e o eixo das ordenadas o valor de entropia calculado.

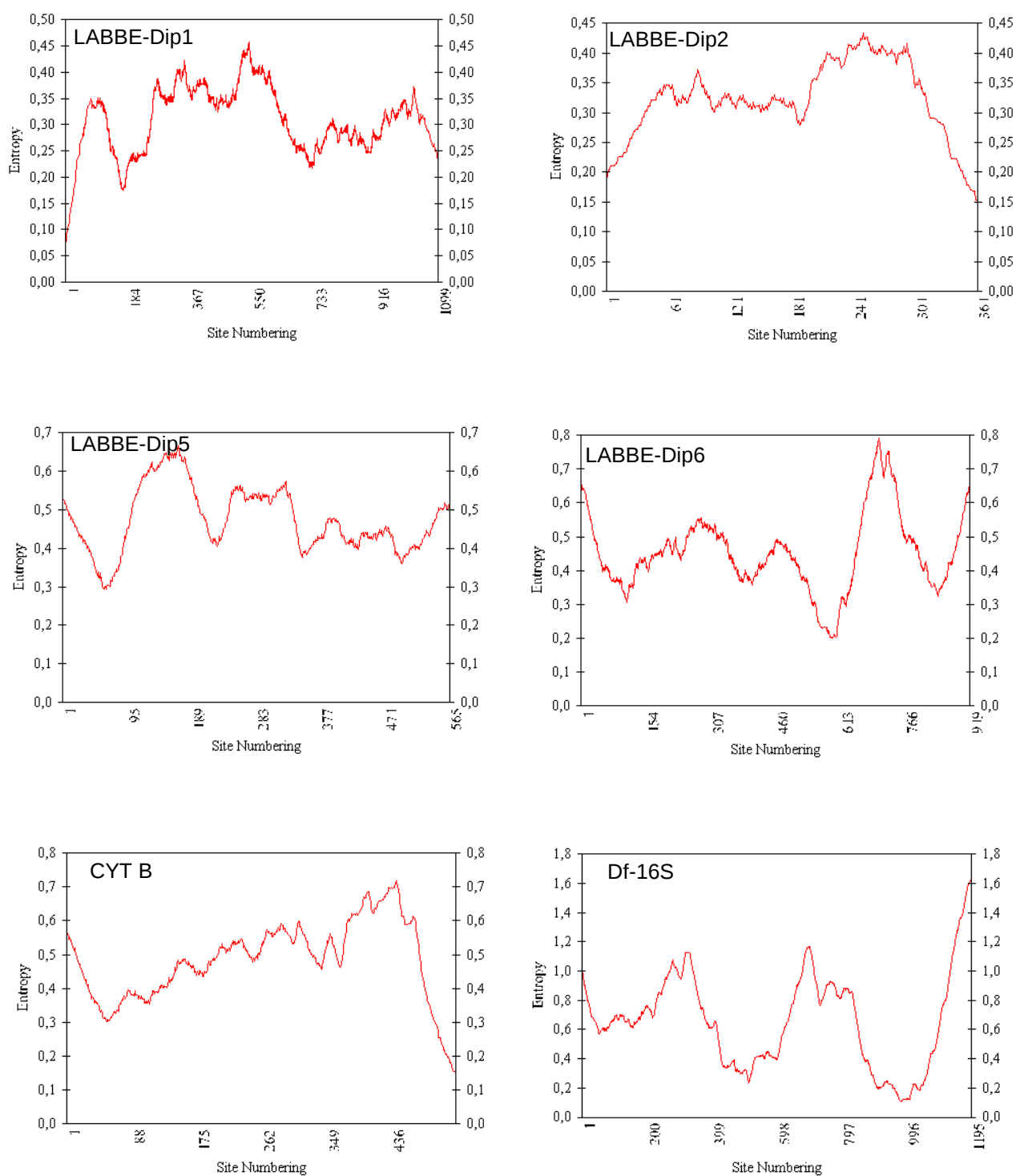


Figura 9 - Análise de entropia das regiões delimitadas pelos pares de *primers*. O eixo das abcissas representa a posição na região analisada e o eixo das ordenadas o valor de entropia calculado.

5.3.2 Análises da variação nucleotídica na sequências analisadas

Os ensaios de alinhamento revelaram a alta taxa de variação dos genes mitocondriais em Lepidoptera. O gene que apresentou a maior proporção de sítios conservados em relação ao número de sítios alinhados foi COI (52%) e o que apresentou o menor grau de conservação foi ND5 (25%). Quando excluída a terceira posição, o número de sítios conservados aumentou já o número de sítios parcimoniosos informativos diminuiu (Tabela 4).

Tabela 4 - Análise dos sítios dos genes de Lepidoptera

Teste	SAL (pb)	SC (%)	SV (%)	SPI* (%)	SING* (%)
1.A.COI	1.301	52	48	82	18
1.B.COI	868	74	26	68	32
1.A.ATP6	277	40	60	86	14
1.B.ATP6	184	57	43	82	18
1.A.COIII	774	41	59	85	15
1.B.COIII	516	59	41	74	26
1.A.ND3	213	38	62	86	14
1.B.ND3	142	54	46	77	23
1.A.ND5	409	25	75	85	15
1.B.ND5	273	35	65	80	20
1.A.CYTB	339	39	61	82	18
1.B.CYTB	226	54	46	73	27
1.A.ND1	96	35	65	81	19
1.B.ND1	64	48	52	70	30
1.A.16S	908	40	60	72	28

SAL: Sítios alinhados; SC: Sítios conservados; SV: Sítios variáveis; SPI: Sítios informativos para parcimônia; SING: *Singletons*. * Valores calculados em relação ao número de SV.

A região delimitada pelos pares de *primers* mais conservada foi LABBE-Dip1 (52%), que corresponde à região central do gene COI. Entretanto, a região mais variada (LABBE-Dip2) também correspondeu a região terminal 3' do gene

COI (Tabela 5). Em relação à exclusão da terceira posição, o mesmo perfil observado para os testes do grupo 1 foi observado para o grupo 2 (Tabela 5). O número de sítios conservados e de *singletons* foi maior nos testes da situação B que os da situação A.

Tabela 5 - Análise dos sítios das regiões delimitadas pelos pares de *primers*

Par de <i>primers</i>	SAL (pb)	SC (%)	SV (%)	SPi [*] (%)	SING [*] (%)
2.A.LABBE-Dip1	1.098	52	48	82	18
2.B.LABBE-Dip1	732	74	26	65	35
2.A.LABBE-Dip2	363	20	80	82	18
2.B.LABBE-Dip2	242	72	28	71	29
2.A.LABBE-Dip5	567	38	62	83	17
2.B.LABBE-Dip5	378	54	46	75	25
2.A.LABBE-Dip6	920	42	58	85	15
2.B.LABBE-Dip6	564	59	41	77	23
2.A.CYT B	522	40	60	82	18
2.B.CYT B	290	53	47	72	28
2.A.Df-Leu	937	40	60	72	28

SAL: Sítios alinhados; SC: Sítios conservados; SV: Sítios variáveis; SPi: Sítios informativos para parcimônia; SING: *Singletons*. Valores calculados em relação ao número de SV.

5.3.3 Sinal filogenético

Todos os genes apresentaram sinal filogenético bom ou ótimo. Na situação A, o gene que apresentou o menor sinal foi ND1 com 72,9. Os genes que apresentaram sinal filogenético bom foram: ATP6, ND3, ND5, CYTB e os que apresentaram sinal filogenético ótimo foram: COI e 16S.

Pode-se perceber que na maioria dos testes da situação B, o sinal filogenético diminuiu em relação aos testes da situação A. Para o gene ND1, o

sinal filogenético na situação B foi o único considerado baixo (67,7), como mostrado (Tabela 6).

Tabela 6 – Sinal filogenético dos genes

Grupo 1	Sinal Filogenético A	Sinal Filogenético B
COI	96,3	95,9
ATP6	86,6	85
COIII	94,5	93,5
ND3	85,1	82,5
ND5	88	87,9
CYTB	81,5	88,6
ND1	72,9	67,7
16S	95,6	-

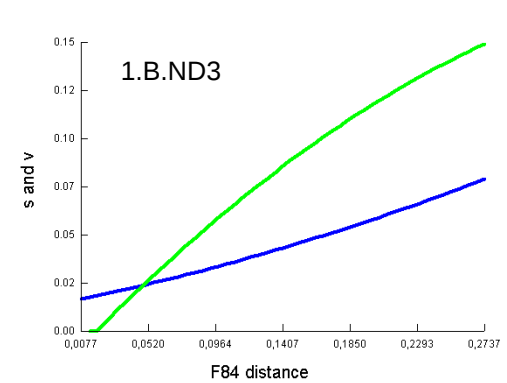
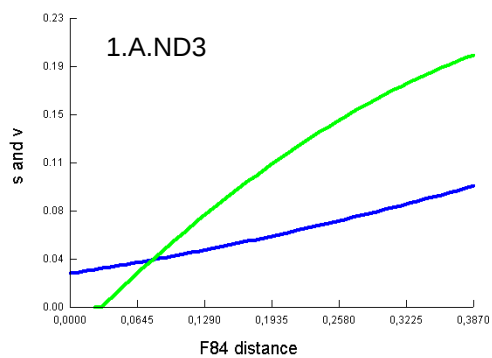
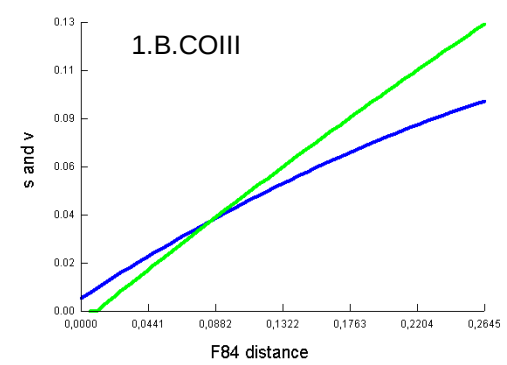
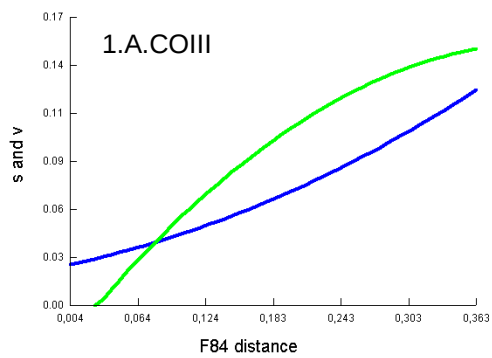
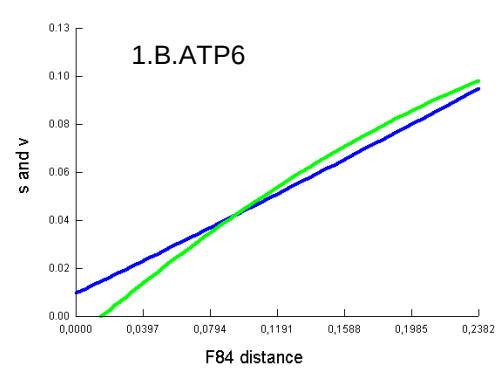
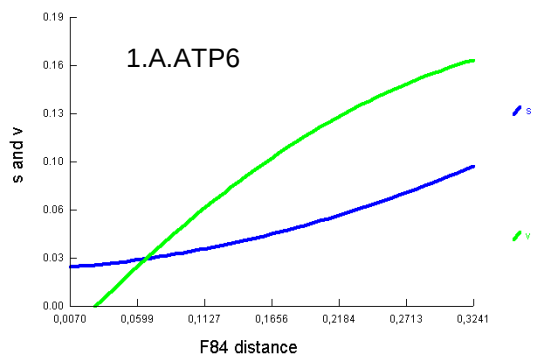
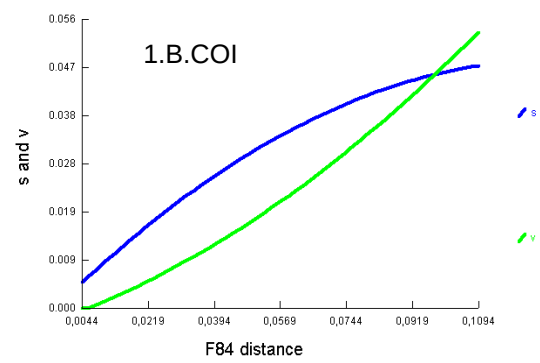
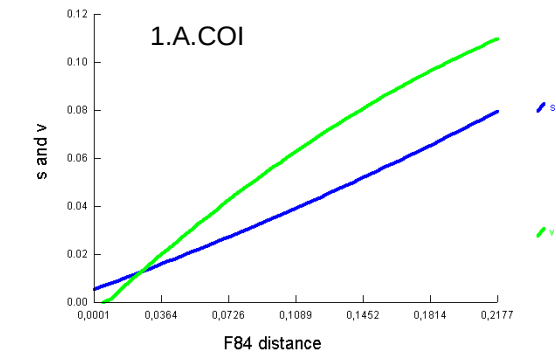
Todos os testes do grupo 2 apresentaram sinal filogenético ótimo. As sequências de nucleotídeos das regiões delimitadas pelos pares de *primers* que continham mais de um gene apresentaram sinal filogenético maior que a dos genes individualizados (Tabela 7). Quando retirada a terceira posição dos genes, o sinal filogenético diminuiu, porém, continuou ótimo, com exceção de LABBE-Dip2 que passou a ser bom.

Tabela 7 – Sinal filogenético das regiões delimitadas pelos pares de *primers*

Grupo 2	Sinal Filogenético A	Sinal Filogenético B
LABBE-Dip1	96,2	93,4
LABBE-Dip2	91	87,8
LABBE-Dip5	92,9	92,2
LABBE-Dip6	95,2	95,9
CYT B	93,2	92,5
Df-Leu	95,6	-

5.3.4 Teste de saturação

Os resultados do teste de saturação mostraram que as sequências nucleotídicas apresentaram muitas mutações em relação à distância genética. Quando comparadas as situações A e B, apenas os genes COI, ATP6 e COIII e as regiões delimitadas pelos pares de *primers* LABBE-Dip1 e LABBE-Dip5 mostraram alterações consideradas. A saturação dos citados diminuiu quando foi retirada a terceira posição. Os gráficos de cada teste podem ser visualizados nas Figuras 10 (para os GCPs e 16S) e 11 (para as regiões delimitadas pelos pares de *primers*).



Continua

Continuação...

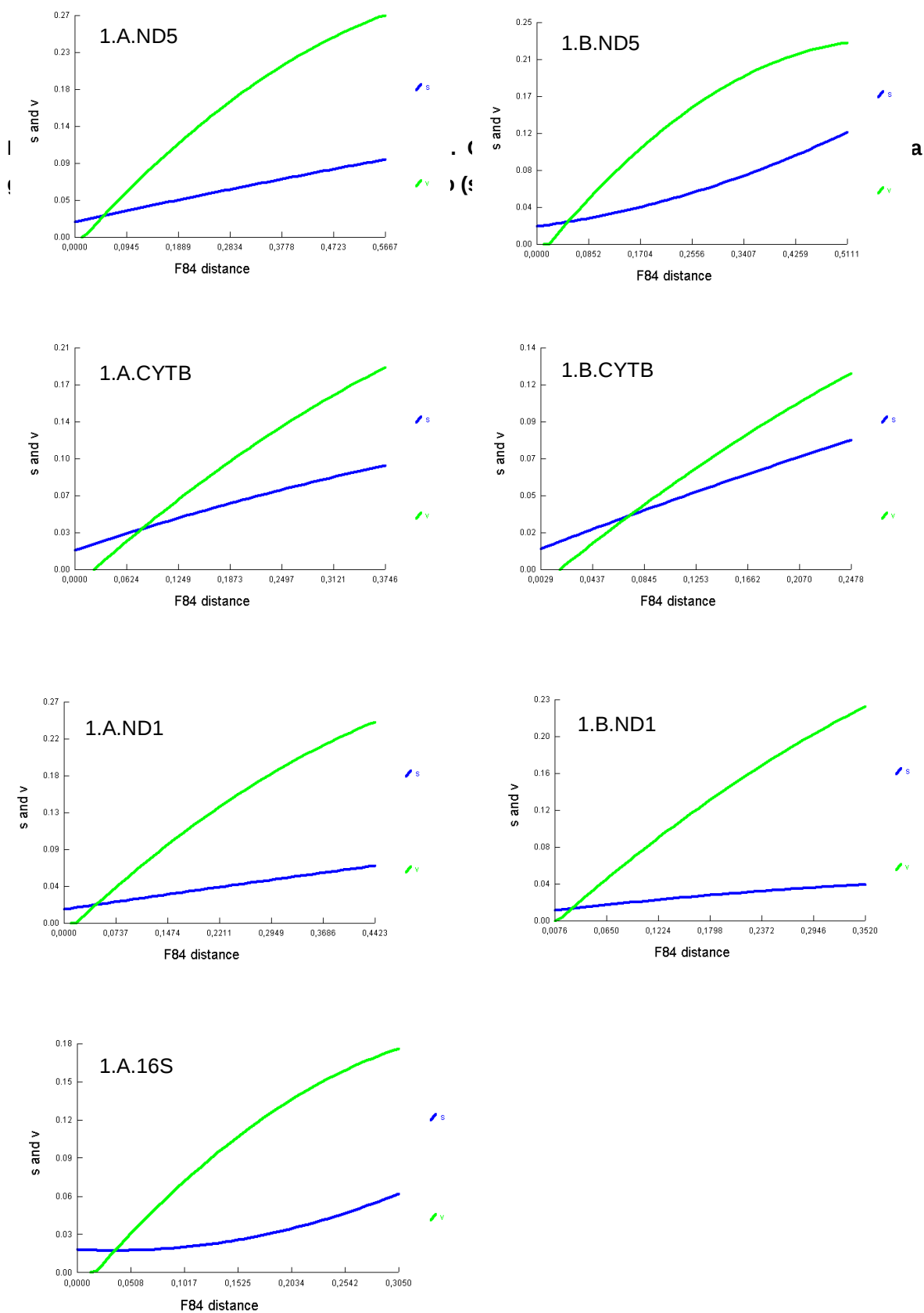
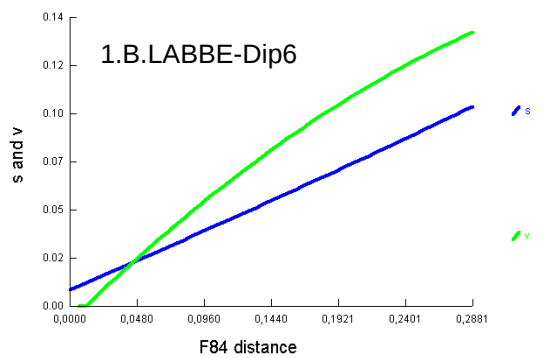
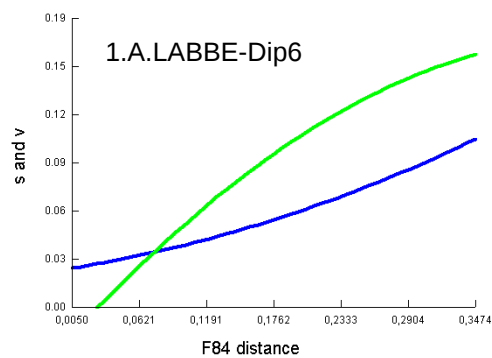
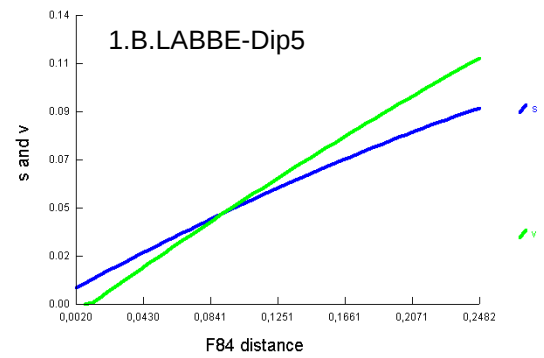
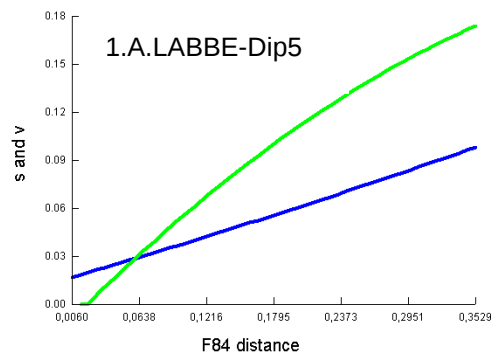
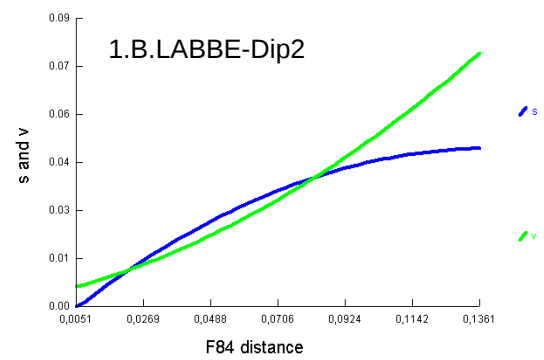
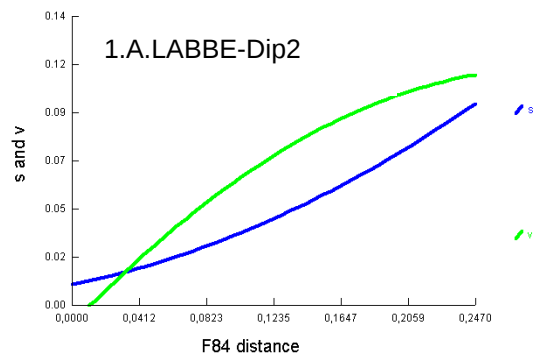
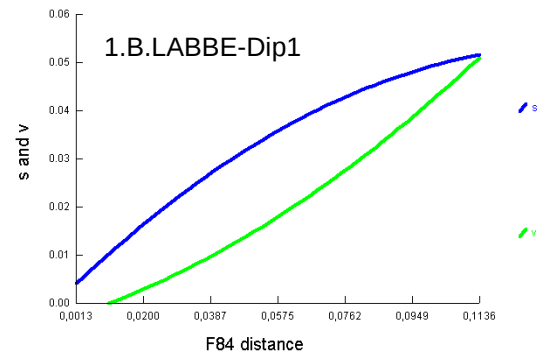
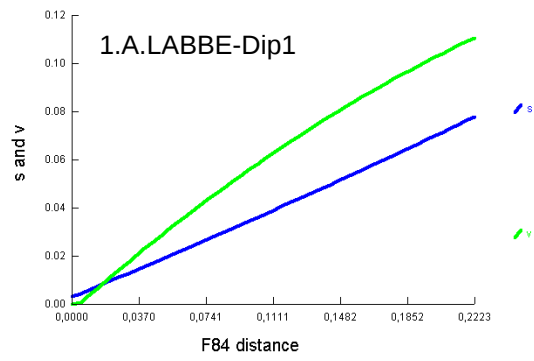


Figura 10 - Teste de saturação dos genes. O eixo das abcissas representa a distância genética e das ordenadas a taxa de transição (s) e transversão (v).



Continua

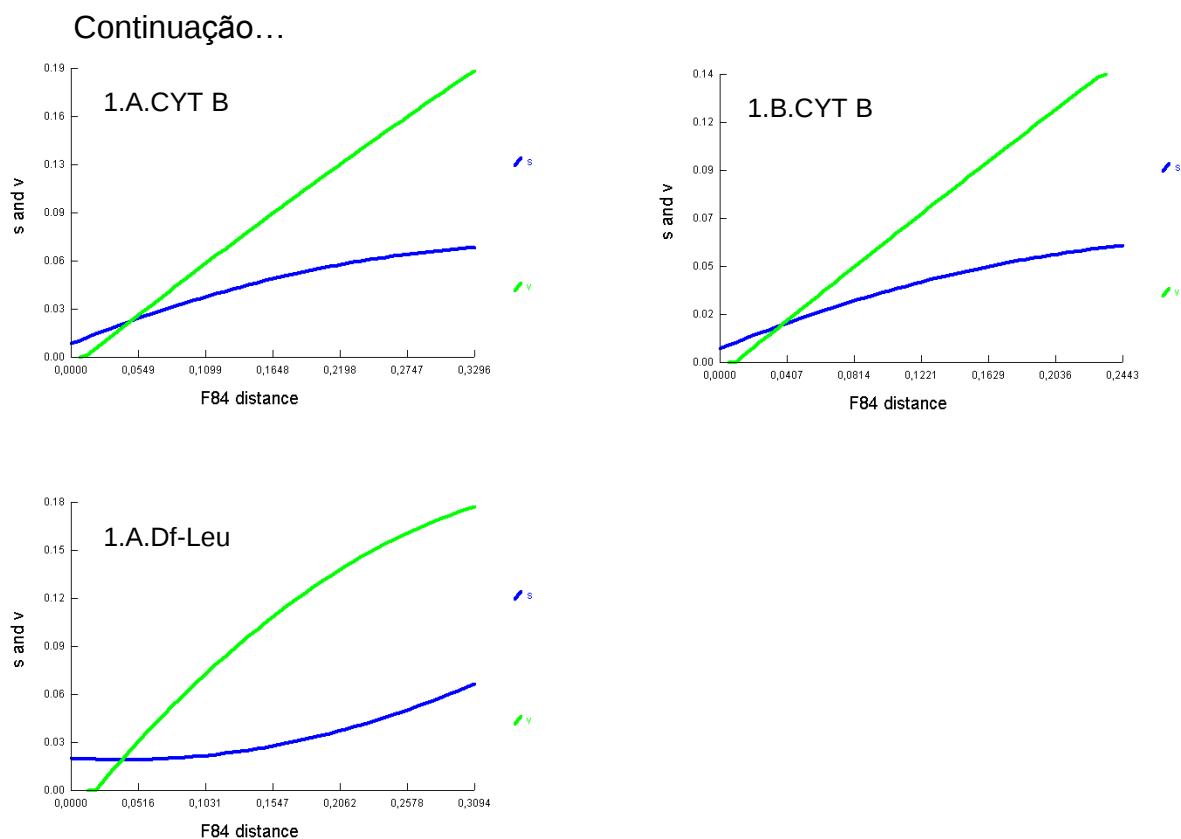


Figura 11 - Teste de saturação dos pares de *primers*. O eixo das abcissas representa a distância genética e das ordenadas a taxa de transição (s) e transversão (v).

5.3.5 Árvores filogenéticas

As árvores filogenéticas foram construídas com base nas análises prévias de sinal filogenético e de saturação. Não foram geradas árvores filogenéticas para os genes ND3, ND5, CYTB e ND1 e para as regiões delimitadas pelos pares de *primers* LABBE-Dip2, LABBE-Dip6 e CYT B na situação B. Os testes de saturação da situação B não apresentaram melhoras em relação aos testes da situação A e o sinal filogenético da situação A foi maior que o da situação B. Dessa forma, foram geradas apenas duas árvores filogenéticas (sendo uma para

análise de máxima parcimônia e outra para análise bayesiana) para a situação A para os genes ND3, ND5, CYTB e ND1 e para as regiões delimitadas pelos pares de *primers* LABBE-Dip2, LABBE-Dip6 e CYT B. Por outro lado, para os genes COI, ATP6 e COIII e para as regiões delimitadas pelos pares de *primers* LABBE-Dip1 e LABBE-Dip5 foram geradas duas árvores para a situação A e duas para a situação B. Além disso, foram geradas duas árvores filogenéticas para o gene 16S e para a região delimitada pelo par de *primer* Df-Leu, que não estavam inclusos nas análises da situação B. No total foram geradas 38 árvores filogenéticas.

Em todas as árvores filogenéticas, a família Hepialoidea foi agrupada, seja pela análise de máxima verossimilhança ou pela bayesiana. A família Tortricoidea foi determinada na maioria dos testes, com exceção dos genes ND1, CYTB e da região delimitada pelo par de *primers* LABBE-Dip2. Nem todos os testes foram satisfatórios nas análises filogenéticas, já que os nós mais externos não conseguiram ser determinados nem pela análise bayesiana e nem pela máxima verossimilhança. A família Nymphalidae não conseguiu ser agrupada em nenhuma das análises. Essas informações podem ser conferidas visualizando as árvores geradas por todos os testes no APÊNDICE B.

Os genes que apresentaram as árvores filogenéticas com maiores valores de confiabilidade e mais compatíveis com a filogenia de Lepidoptera relatada na literatura foram COI, COIII e 16S. Os resultados mostraram-se melhores para os testes da situação A do que da B, e estão descritos abaixo.

O teste 1.A.COI conseguiu agrupar as famílias: Crambidae, Tortricidae, Papilionidae, Notodontidae, Saturniidae, Bombycidae, Noctuidae, Pieridae, Lycaenidae e Hepialidae (Figura 12). Apesar de agrupar a família Erebidae, os

valores se mostraram abaixo do valor de corte para as análises (0,82 para a bayesiana e 46 para máxima verossimilhança). Apenas a família Nymphalidae não foi agrupada.

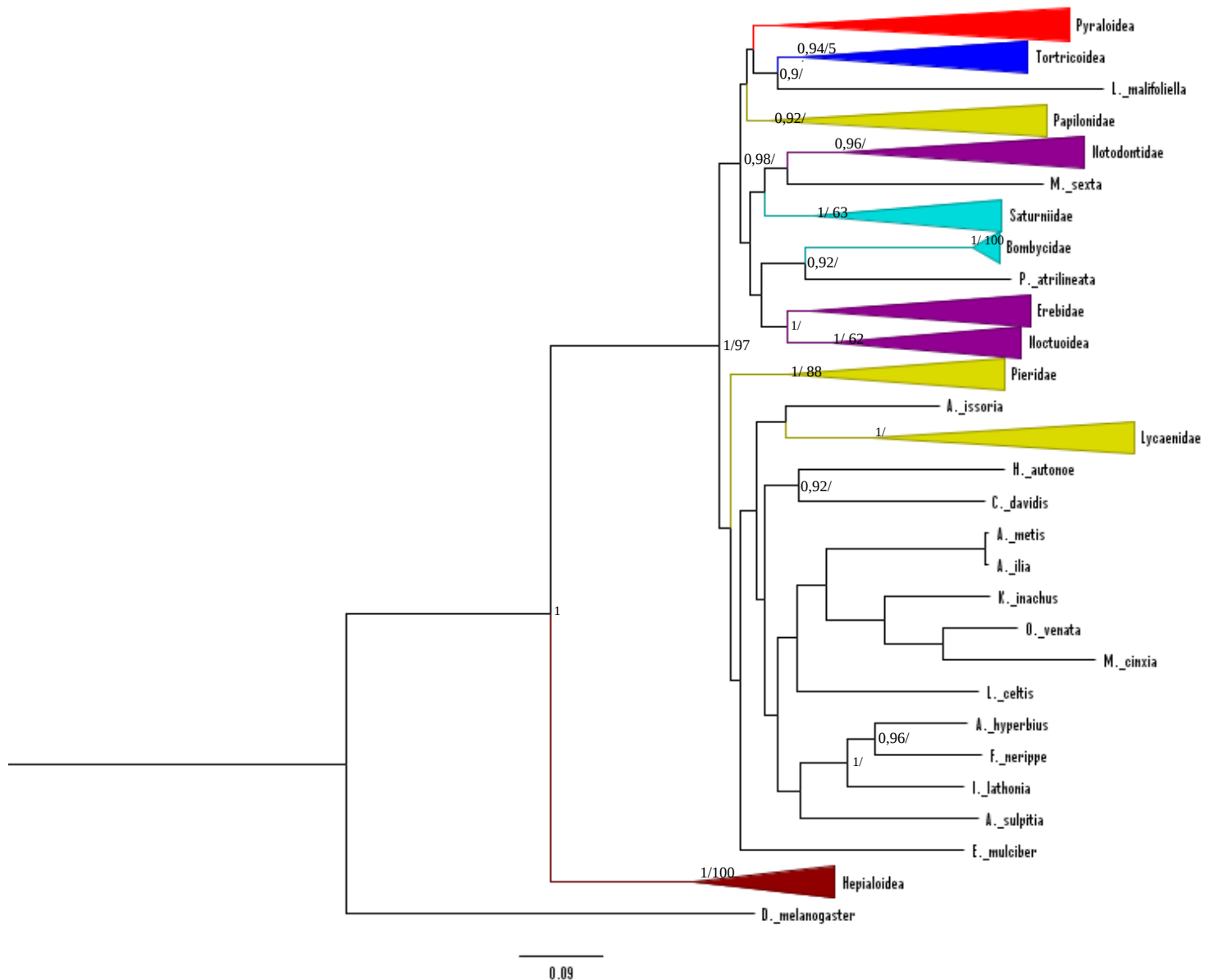


Figura 12 - Árvore filogenética de Lepidoptera a partir do gene COI. Os números representam os valores gerados pela análise bayesiana e máxima verossimilhança, respectivamente.

O teste 1.A.COIII agrupou as famílias: Tortricidae, Saturniidae, Bombycidae, Noctuoidae, Pieridae, Lycaenidae e Hepialidae (Figura 13). O teste

1.B.COIII agrupou Crambidae pela análise bayesiana, porém, não agrupou Tortricidae, pois Hepialidae se agrupou aos membros dessa família.

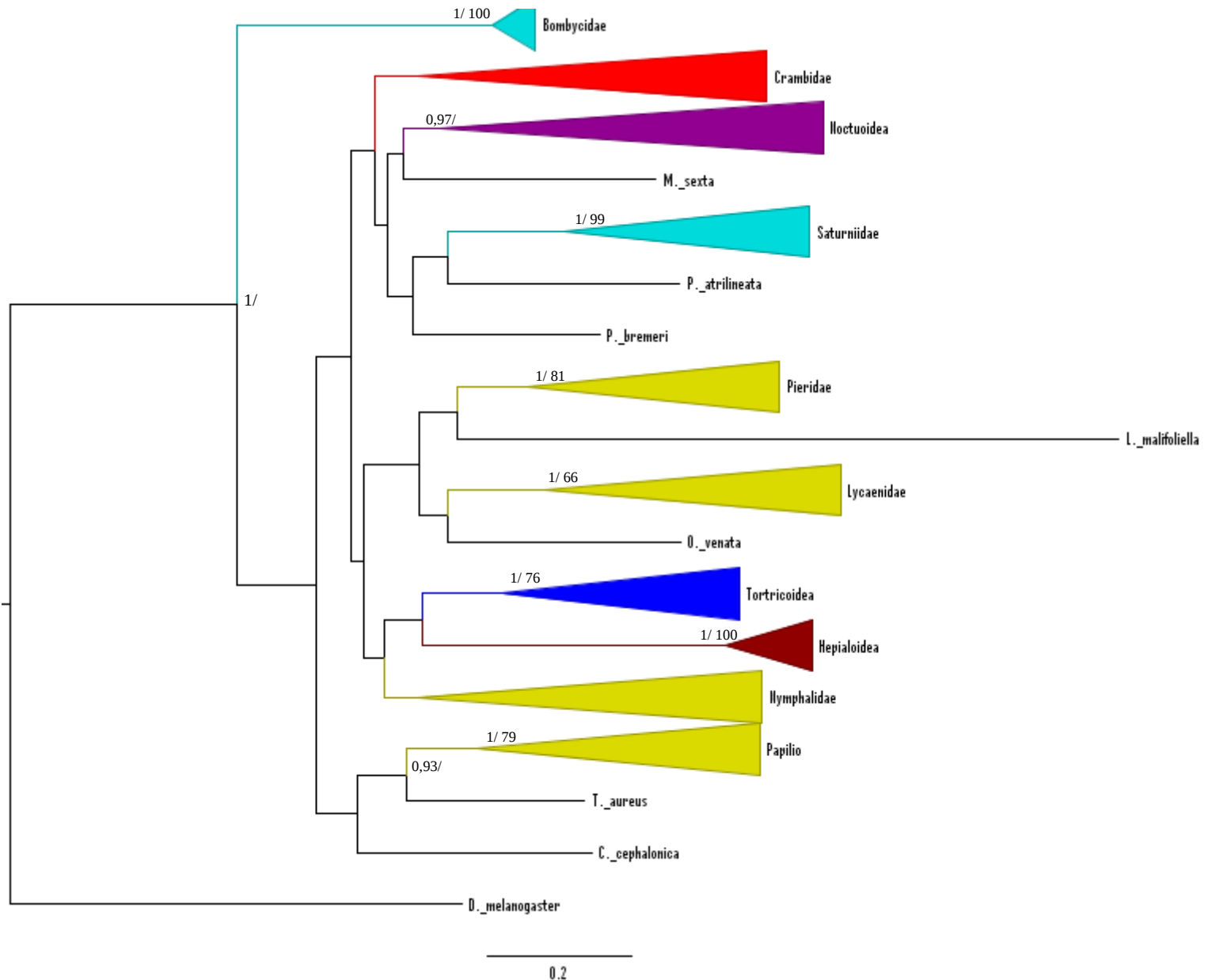


Figura 13 - Árvore filogenética de Lepidoptera a partir do gene COIII. Os números representam os valores gerados pela análise bayesiana e máxima verossimilhança, respectivamente.

O teste 1.A.16S foi capaz de agrupar Crambidae, Tortricidae, Saturniidae, Bombycidae, Noctuoidea, Lycaenidae e Hepialidae (Figura 14). Formou ainda a

superfamília Noctuoidea, formada pelas famílias Notodontidae, Noctuoidea e Erebidae e a superfamília Bombycoidea, formada por Bombycidae e Noctuoidea.

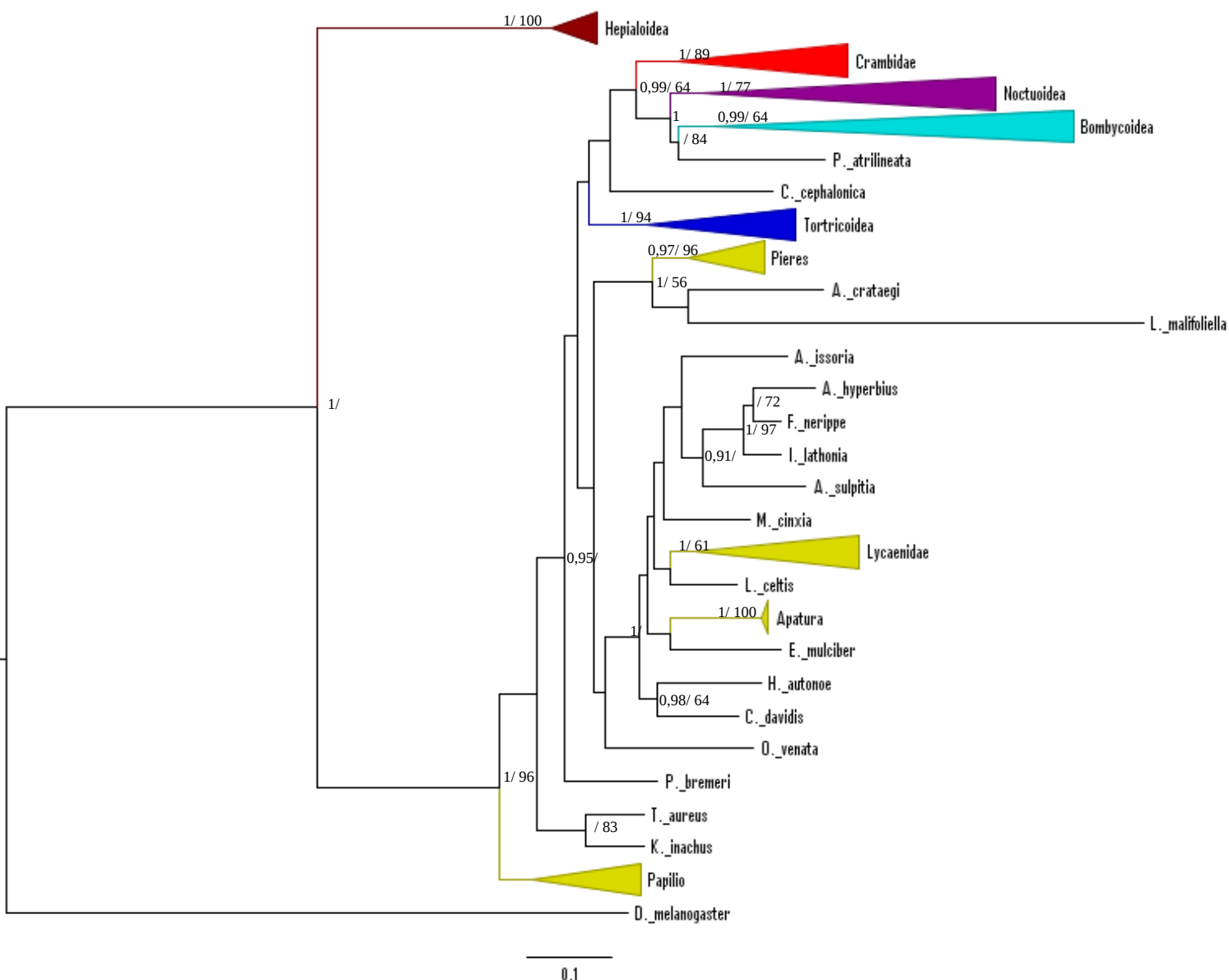


Figura 14 - Árvore filogenética de Lepidoptera a partir do 16S. Os números representam os valores gerados pela análise bayesiana e máxima verossimilhança, respectivamente.

A região delimitada pelo par de *primers* LABBE-Dip1 foi a que apresentou a árvore filogenética mais consistente, agrupando as famílias: Crambidae, Tortricidae, Notodontidae, Erebidae, Saturniidae, Bombycidae, Noctuoidea,

Pieridae, Lycaenidae e Hepialidae (Figura 15). Além disso, também agrupou a superfamília Pyraloidea pela análise bayesiana.

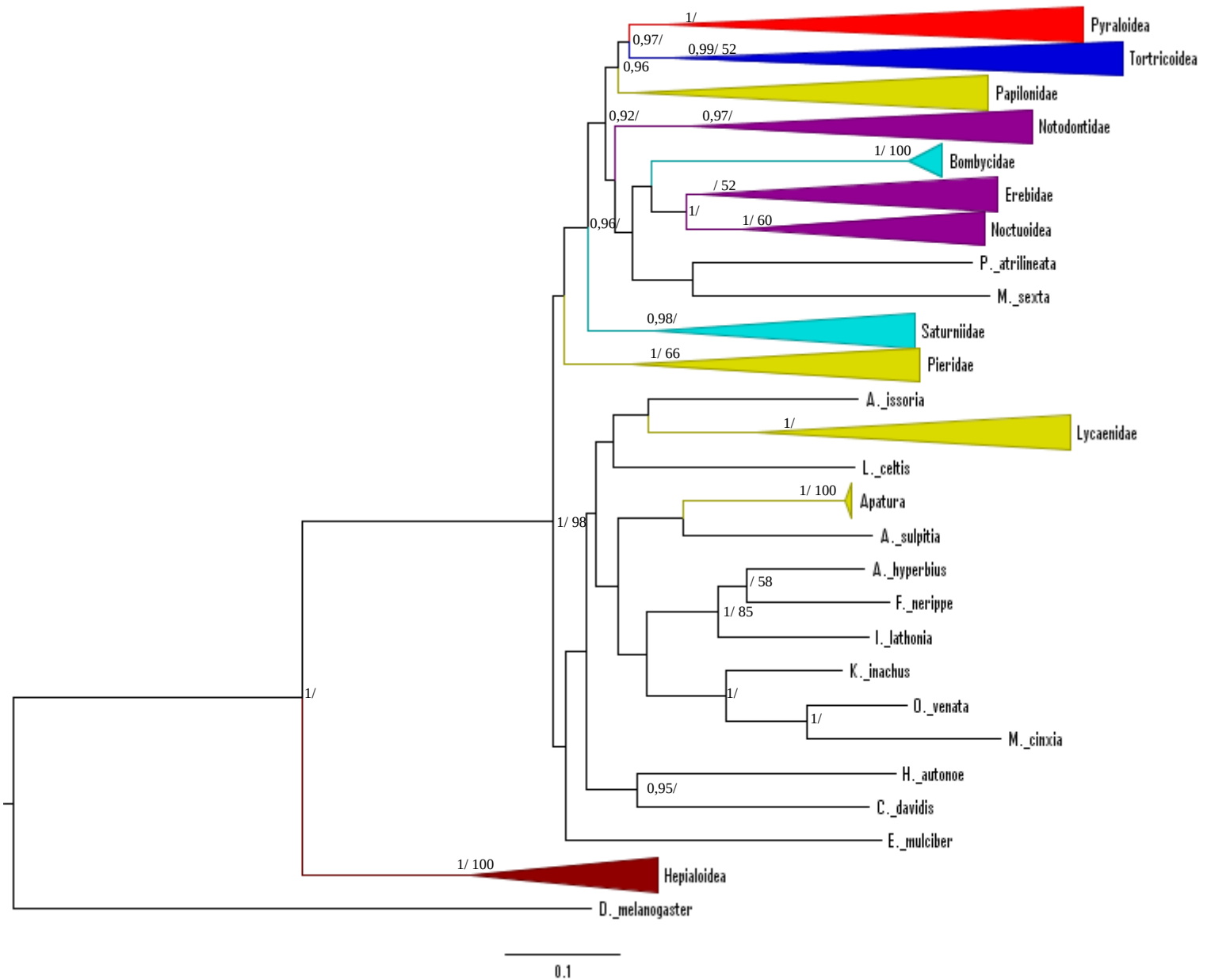


Figura 15 - Árvore filogenética de Lepidoptera a partir da região delimitada pelo par de primers LABBE-Dip1. Os números representam os valores gerados pela análise bayesiana e máxima verossimilhança, respectivamente.

5.4 Concatenação

Foram geradas duas árvores filogenéticas para os genes concatenados e mais duas para as regiões delimitadas pelos pares de *primers* concatenados (Figuras 16 e 17). As análises filigenéticas concatenadas mostraram que é possível agrupar as espécies de Lepidoptera tanto a nível de família quanto a nível de superfamília com apenas 30% da sequência dos mitogenomas.

Diatraea flavipennella foi agrupada na família Crambidae, juntamente com: *C. medinalis*, *O. nubilalis*, *O. furnacalis*, *C. suppressalis* e *D. saccharalis*. Foram formadas as famílias Notodontidae, Noctuoidea, Erebidae, Bombycidae, Saturnidae, Papilionidae, Nymphalidae, Pieridae, Lycanidae, Tortricidae, Hepialoidea. As famílias Notodontidae, Noctuoidea e Erebidae formaram a superfamília Noctuoidea; as famílias Bombycidae, Saturnidae, juntamente com a Sphingidae formou a superfamília Bombycoidea. Essas duas superfamílias mais a superfamília Geometroidea formaram o clado Macroheterocera. As famílias Papilionidae, Nymphalidae, Pieridae, Lycanidae e Hesperidae formam a superfamília Papilionoidea. Das superfamílias estudadas, apenas a Pyraloidea não foi formada, pois não foi possível agrupar *C. cephalonica* com os representantes da família Crambidae.

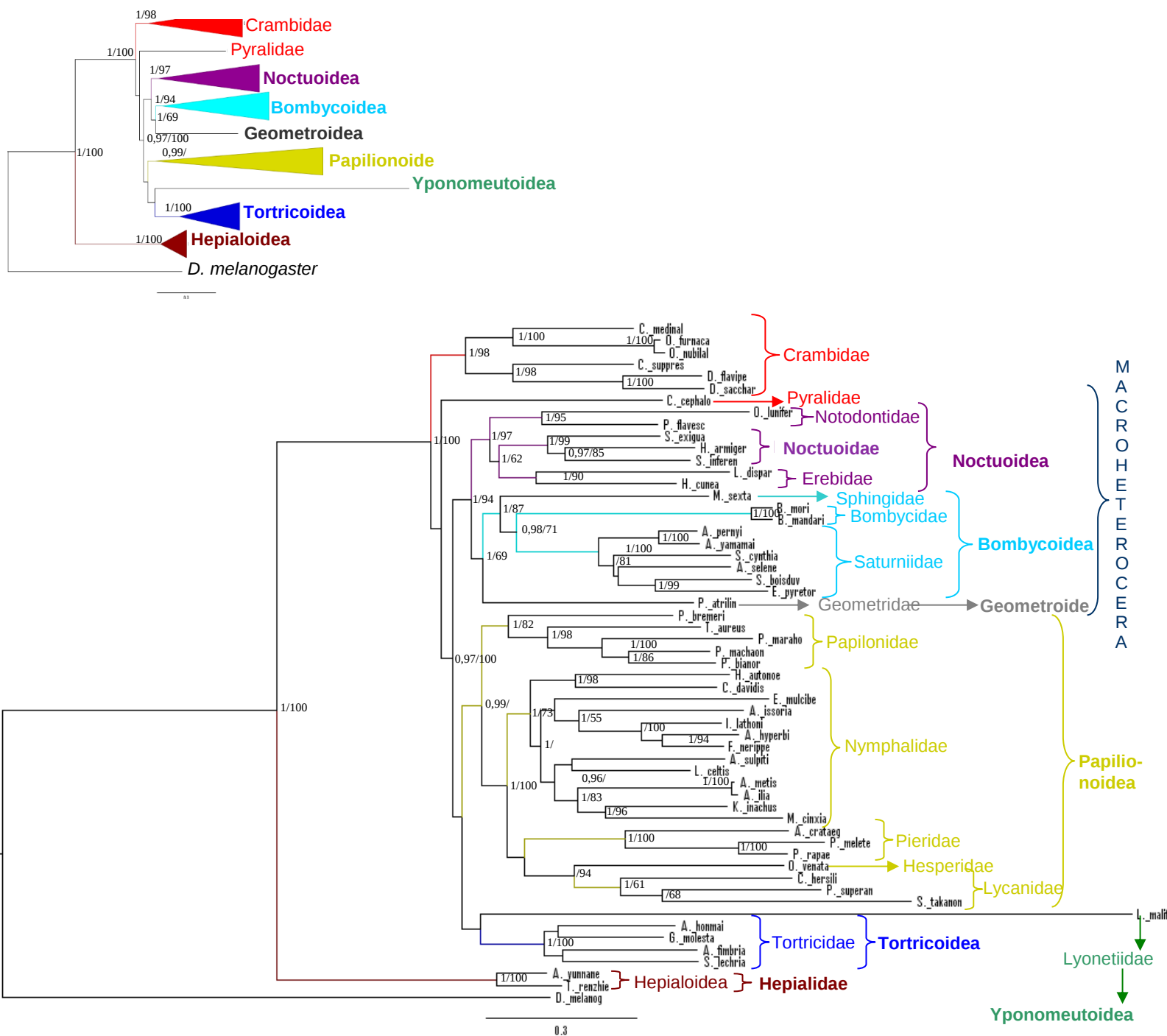


Figura 16 - Relações filogenéticas inferidas para Lepidoptera baseado na concatenação dos genes por análise bayesiana e máxima verossimilhança, respectivamente. *D. melanogaster* foi utilizado como grupo extremo.

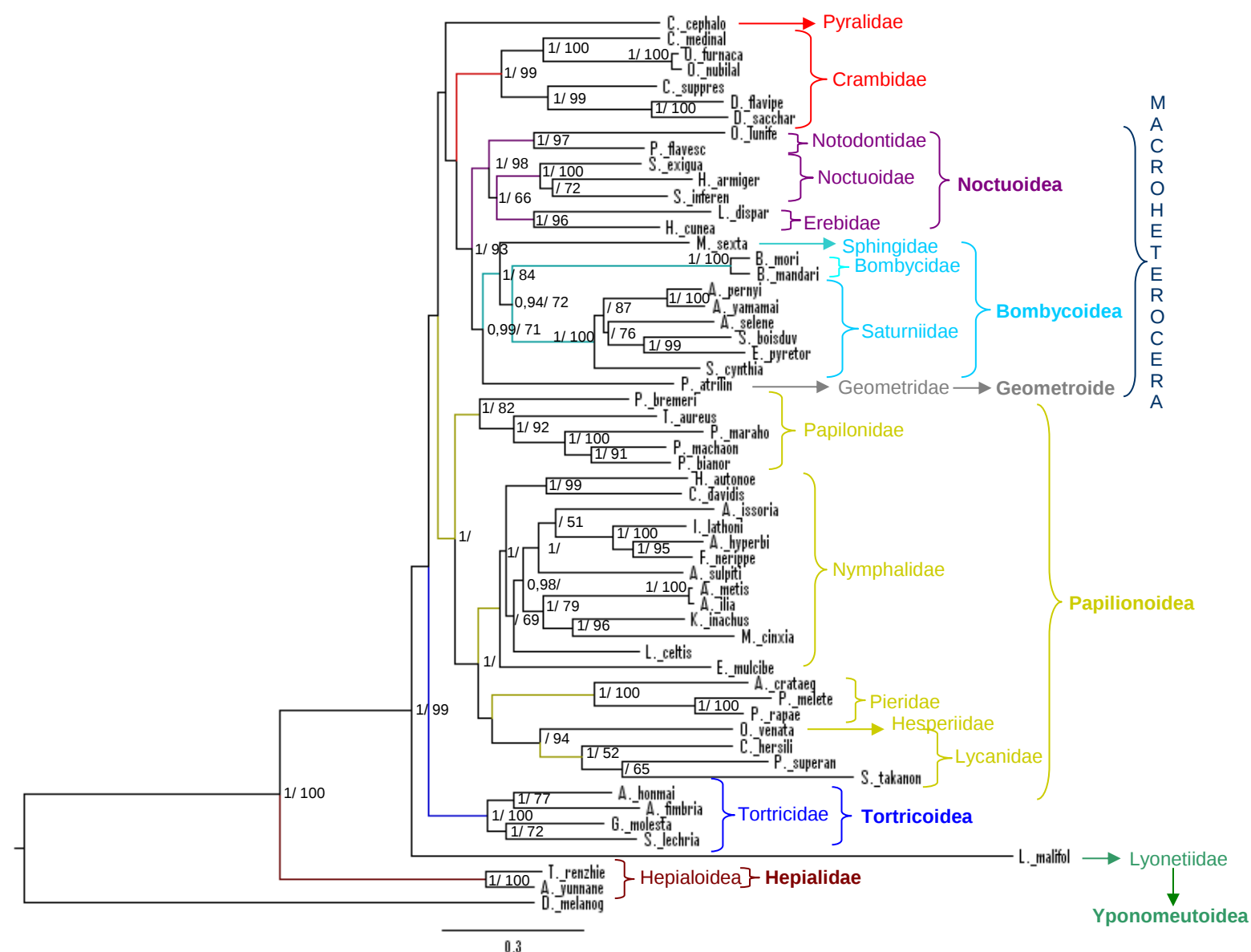
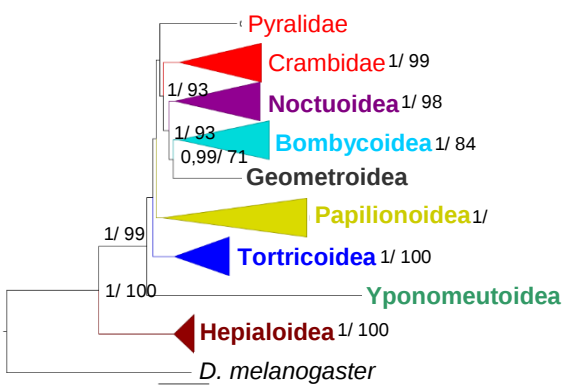


Figura 17 - Relações filogenéticas inferidas para Lepidoptera baseado na concatenação dos pares de *primers* por análise bayesiana e máxima verossimilhança, respectivamente. *D. melanogaster* foi utilizado como grupo extremo.

6. Discussão

Vários trabalhos têm sido realizados baseados na morfologia e em dados moleculares da ordem Lepidoptera com a finalidade de elucidar sua evolução (Nieukerken *et al.*, 2011; Mutanen *et al.*, 2010). O DNA mitocondrial é muito utilizado em estudos evolutivos por apresentar elevado sinal filogenético que possibilita resolver robustamente os relacionamentos entre os táxons. Neste trabalho, foram analisadas regiões informativas dos mitogenomas de Lepidoptera já utilizadas em análises filogenéticas da ordem. Os resultados mostraram que a concatenação dessas regiões produziu perfil filogenético com a robustez semelhante das análises das sequências completas dos mtDNAs. Com apenas 30% das sequências dos mitogenomas foi possível obter a filogenia de Lepidoptera, com redução de tempo computacional e de custos de bancada. Para comprovar a eficiência dos marcadores moleculares analisados para estudos filogenéticos da ordem Lepidoptera foram sequenciadas pela primeira vez regiões do mitogenoma de *D. flavipennella*. Esta espécie foi escolhida devido a sua importância econômica para o país, pois é uma das principais pragas da cana-de-açúcar no Nordeste brasileiro.

Análises de sintenia foram realizadas com as sequências disponibilizadas no GenBank para se determinar o grau de ordenação gênica nos mitogenomas escolhidos (Bhutkar, 2006). A partir disso foi possível verificar que a ordem gênica de Lepidoptera é bastante conservada, apesar de apresentar algumas alterações. Duas alterações na ordem gênica relevantes para este trabalho foram a translocação do tRNA^{Leu} de *A. selene* e a do 16S de *L. malifoliella*. Ambas foram consideradas eventos evolutivos independentes por não terem sido identificados previamente em nenhum outro lepidóptero.

Tanto o gene de tRNA de *A. selene* quanto o de 16S de *L. malifoliella* sofreram translocações de uma das fitas de DNA para a outra fita, como a maioria dos eventos de translocações nos mitogenomas de Lepidoptera. Os rearranjos envolvendo alteração de fitas podem ser explicados pelo mecanismo de inversão. A recombinação em um genoma circular pode conduzir a excisão de um pequeno círculo do genoma circular ou a inversão de uma região do genoma, dependendo da maneira como as fitas de DNA são reconectadas (Dowton e Campbell, 2001).

Os tRNAs sofrem mais eventos de translocação dentro dos mitogenomas que as regiões codificadoras de proteínas e que o rRNAs. Eles são frequentemente encontrados em fragmentos de genomas mitocondriais rearranjados por apresentarem capacidade de formar estruturas do tipo laço. Quando a estrutura é formada, parte do genoma pode ser “transportada” para uma nova região do genoma (Higgs *et al.*, 2006).

Apesar de serem as regiões que sofrem mais translocações, os tRNAs são as regiões mais conservadas nos mitogenomas. As regiões codificadoras de proteínas, ao contrário, sofrem poucas translocações e são as mais variáveis. Dentre as regiões codificadoras, COI é o gene mais conservado tanto em Lepidoptera quanto em outros organismos. Os dados obtidos pelo sequenciamento de *D. flavipennella* também suportam essa observação. Quando comparadas as suas sequências com *D. saccharalis*, as regiões mais conservadas foram os tRNAs, seguidos pelos genes COI, ATP6 e ND1. COI manteve-se como gene mais conservado também na comparação das sequências de *D. flavipennella* com os demais Lepidoptera. Hua *et al.* (2008) observou que os genes COI e ATP6 eram uns dos mais conservados em Hemiptera (classe Insecta) e Santos (2012) teve a mesma percepção quando analisou o filo

Arthropoda. O gene ND5 foi o mais variado tanto na comparação direta com *D. saccharalis*, quanto na comparação com os demais Lepidoptera, sugerindo que devido à elevada taxa de substituição, esse gene pode ser mais adequado para análises de evolução rápida como filogenética de populações.

Com a exclusão da terceira posição, o valor da conservação aumentou tanto nos genes quanto nas regiões delimitadas pelos pares de *primers*, enquanto que o sinal filogenético diminuiu. Esse fato mostra a importância da terceira posição nas análises filogenéticas envolvendo genes mitocondriais de Lepidoptera. A exclusão da terceira posição em estudos evolutivos tornou-se comum devido à região ser a mais mutável. O elevado número de mutações pode “camuflar” o sinal filogenético e muitas vezes é melhor realizar a análise filogenética sem a terceira posição. Contudo, em alguns estudos evolutivos envolvendo dados moleculares mitogenômicos identificou-se que a exclusão da terceira posição gerava árvores com incongruências. Nem sempre a exclusão da terceira posição torna a análise favorável e por isso, aconselha-se realizar estudos de terceira posição em diferentes táxons (Regier *et al.*, 2013; Regier *et al.*, 2009).

Se o sinal filogenético de um conjunto de dados for fraco, a inclusão de pequenas regiões sem informação filogenética pode eliminar o verdadeiro sinal. Nestes casos, a utilização de métodos que ajudam a discernir e remover as regiões pouco informativas é recomendada (Rota-Stabelli e Telford, 2008). Com a finalidade de evitar que artefatos interferissem nas análises filogenéticas, além do teste do sinal filogenético foram realizadas ainda análise de entropia e teste de saturação.

A análise de entropia identificou regiões com alto grau de variação que perturbariam ainda mais o sinal filogenético. Esta análise permitiu a retirada de regiões muito variáveis nos genes COIII, ND3 e 16S e na região do par de *primer* Df-Leu. Santos (2012) também empregou esta metodologia com mitogenomas de Arthropoda e Batista e colaboradores (2011) em genomas de Pappilomavírus. Eles obtiveram árvores filogenéticas com níveis mais altos de confiança quando utilizaram a entropia para a seleção de regiões filogeneticamente informativas.

Apesar da exclusão de regiões na análise de entropia, o teste de saturação nos mostrou que as sequências ainda estavam saturadas. Percebemos que nem todas as sequências melhoraram quando retirada a terceira posição, ou seja, a saturação das sequências encontrava-se também nas outras posições. As sequências que obtivemos melhoras com a exclusão da terceira posição foram COI, ATP6, COIII e as regiões delimitadas pelos pares de *primers* LABBE-Dip-1 e LABBE-Dip-5. Perceberam-se que as árvores filogenéticas geradas com a terceira posição (situação A) para esses testes mostraram-se mais robustas em relação às árvores que excluíram a terceira posição (situação B).

Apesar das sequências estarem saturadas, o sinal filogenético foi considerado no mínimo regular (>70) em todas as sequências. A maioria foi classificada como bom e COI e 16S apresentaram sinal filogenético ótimo. COIII apresentou o terceiro maior valor de sinal filogenético e esses três genes foram os que apresentaram as melhores árvores filogenéticas. O maior sinal filogenético entre as regiões delimitadas pelos pares de *primers* foi LABBE-Dip1 e também foi o que apresentou a melhor árvore filogenética dentre os pares de *primers*. Ainda assim, as árvores geradas por esses testes apresentaram incongruências principalmente em relação às superfamílias Papilionoidea e Pyraloidea.

O sinal filogenético está diretamente relacionado as mudanças sinônimas e não-sinônimas das sequências. Ambas as mudanças fornecem sinal filogenético valioso para os Lepidoptera, contudo estes sinais são mais informativos em diferentes níveis filogenéticos. Para divergências relativamente recentes de Lepidoptera, como dentro de famílias a mudança sinônima (que se acumula mais rapidamente) fornece uma abundância de informações filogenéticas úteis enquanto que mudanças não-sinônimas (que evolui de forma mais lenta) é mais escassa. Para resolver problemas filogenéticos mais distantes, a mudança sinônima é quase que indecifrável devido a várias substituições que se sobrepõem e pode mesmo tornar-se misinformativa devido à heterogeneidade composicional. Essa heterogeneidade pode introduzir um viés analítico que distorce o sinal filogenético da evolução da sequência primária, e pode até mesmo resultar em nós incorretos nas árvores (Cho *et al.*, 2011).

Em termos de tempos de divergência, relações de Lepidoptera de nível superior provavelmente apresentam uma situação intermédia, na qual tanto as alterações não-sinônimas e sinônimas são potencialmente úteis. Estudos voltados para relações dentro de superfamílias, como Bombycoidea, Tortricoidea e Pyraloidea demonstraram que as análises envolvendo mudanças sinônimas e não-sinônimas juntas proporcionam uma análise mais robusta (Regier *et al.*, 2013).

Os problemas de incongruência de árvore em relação à superfamília Papilionoidea foram resolvidos ao se concatenar todos os genes e todas as regiões delimitadas pelos pares de *primers*. Os valores de confiança do agrupamento das famílias e das superfamílias foram aumentados nas análises concatenadas em relação ao das análises individuais. A concatenação dos genes

é uma opção para evitar a perda de dados e mesclar a história evolutiva dos genes para se aproximar mais com a realidade evolutiva do grupo. Os genes individuais são susceptíveis a variar por causa das diferentes taxas de substituição e as árvores resultantes de diferentes genes resultam em uma variabilidade de topologias. Um número mínimo de genes podem ser usado para reproduzir resultados mitogenômicos e otimizar as análises de grandes conjuntos de dados ou de Lepidoptera com sequências mitogenômicas incompletas (Duchêne *et al.*, 2011).

Outros autores já haviam proposto a utilização de genes concatenados para análises filogenéticas e obtiveram resultados semelhantes às análises geradas a partir de mtDNAs completos em outros grupos. Regiões dos genes COI, CYTB, ND3 e ATP6 foram concatenadas para descrever a história evolutiva dos golfinhos (Delphinidae) e os genes ND1, COI e ND4 para o gênero das orcas (*Orcinus*) (Dulchêne *et al.*, 2011). Nós utilizamos cinco dos seis genes utilizados por Dulchêne mais o gene COIII e a região 16S, que individualmente mostraram árvores filogenéticas mais robustas e elevado sinal filogenético. Estudos de concatenação de genes também foram realizados na ordem Lepidoptera. Cho *et al.* (2011) utilizou cinco genes nucleares para Lepidoptera totalizando 6,6 kb. Nós utilizamos 4,6 kb com oito diferentes genes mitocondriais informativos e obtivemos resultados semelhantes.

Todos os 55 Lepidoptera estudados, inclusive *D. flavipennella*, pertenciam à subordem Glossata. Dentro de Glossata foram estudados dois representantes da infraordem Exoporia (*T. renzhiensis* e *A. yunnanensis*) e 53 da infraordem Heteroneura. Este último estava representado por quatro clados: Obtectomera, Macroheterocera, Apodritrysia e Ditrysia. Obtectomera é formado por todas as

borboletas (Papilionoidea) e por Pyraloidea (superfamília de *D. flavipennella*). Macroheterocera é formada por três superfamílias: Bombycoidea, Noctuoidea e Geometroidea. Apoditrysia é representada apenas pela superfamília Tortricoidea e Ditrysia apenas pela superfamília Yponomeutoidea (*L. malifoliella*) (van Nieukerken *et al.*, 2011).

Em todos os 38 testes realizados neste trabalho, foi possível agrupar a infraordem Exoporia como grupo mais antigo de Lepidoptera. Na análise de sintenia também identificamos que *T. renzhiensis* e de *A. yunnanensis* apresentavam o tRNA^{Leu} em posição diferente dos demais Lepidoptera. Eles apresentam o tRNA^{Leu} na mesma posição que o ancestral comum dos artrópodes. Dessa forma, acredita-se que a translocação do tRNA^{Leu} nos demais Lepidoptera tenha ocorrido posterior a divisão da infraordem Exoporia e se manteve nos mitogenomas dos decendentes de Heteroneura (Cao *et al.*, 2012).

L. malifoliella se agrupou em diferentes locais dependendo da região que estava sendo analisada. Em algumas análises se posicionou próximo a superfamília Tortricoidea (1.A.COI); em outras próximo a família Pieridae (1.A.COIII); e até mesmo entre os membros das famílias impedindo a monofilia dessas (1.A.16S). Um dos motivos para esse tipo de incongruência é a escassez de números de táxons (Kristensen & Skalski 1998). Quanto mais sequências disponíveis para realizar as análises filogenéticas, maior o número de dados a respeito da história evolutiva é obtida e consequentemente, melhor será a análise realizada.

Dos testes individuais, apenas 1.A.COI agrupou o clado Macroheterocera. Com a concatenação de todos os genes e das regiões delimitadas pelos pares de *pirmers* foi possível resolver esse problema observado nos demais testes

individuais. Contudo, em nenhuma análise, incluindo nas concatenadas foi possível agupar o clado Obtectomera, pois os mitogenomas de Macroheterocera muitas vezes se agrupavam com mitogenomas de Obtectomera. Alguns estudos evolutivos desafiam a monofilia do clado Obtectomera (Regier *et al.*, 2013; Regier *et al.*, 2009). O fato se deve principalmente devido a superfamília Pyraloide se agrupar ao clado Macroheterocera ao invés da superfamília das borboletas (Papilionidae). Alguns dados morfológicos contribuem com o agrupamento de Pyraloidea com Macroheterocera, como a presença de órgãos timpânicos bilaterais no tórax ou no abdomen que detectam o ultra-som e são utilizados principalmente para evitar a predação por morcegos. Essas "orelhas" podem ou não ser homólogas dentro do grupo 'Macroheterocera + Pyraloidea', porém, ocorrem apenas esporadicamente em outros Lepidoptera (Heikkilä *et al.*, 2012).

A posição exata das borboletas varia consideravelmente entre os conjuntos de dados, mas elas são sempre separadas do núcleo Macrolepidoptera + Pyraloidea. Apesar de ser um dos grupos mais conhecidos e mais bem estudado de insetos, a filogenia das borboletas verdadeiras (Papilionoidea) tem sido difícil de ser elucidada. As inter-relações das famílias têm sido contraditórias e os resultados de estudos recentes ainda não conseguiram resolver este impasse (Heikkilä *et al.*, 2012; Regier *et al.*, 2009; Wahlberg *et al.*, 2005). A incapacidade dos dados moleculares e morfológicos de resolverem as relações filogenéticas é muitas vezes atribuída à diversificação rápida e antiga das linhagens. A diversificação geográfica e temporal de insetos neotropicais ainda é mal compreendida por causa das complexas mudanças nas condições geológicas e climáticas que ocorreram durante o Cenozóico (Heikkilä *et al.*, 2012). Apesar desses impasses em relação ao grupo das borboletas, nós sugerimos o

agrupamento (Papilionidae + (+Nymphalidae + (Pieridae + (Hesperidae+ Lycaenidae)))) que foi fortemente suportado tanto pela concatenação dos genes, quanto pela concatenação das regiões delimita dos pares de *primers*.

Sendo assim, com apenas 4,6 Kb do mitogenoma de Lepidoptera foi possível obter resultados semelhantes as análises filogenéticas da ordem Lepidoptera descritas. Foi possível também corroborar com os estudos que defendem a parafilia do clado Obtectomera e o agrupamento de 'Macroheterocera + Pyraloidea'. Assim como propor o agrupamento das famílias das borboletas (Papilionidae + (+Nymphalidae + (Pieridae + (Hesperidae+ Lycaenidae)))).

7. Conclusão

Nenhuma região individualmente foi suficiente para elucidar a história evolutiva da ordem Lepidoptera, porém a concatenação dos genes COI, ATP6, COIII, ND3, ND5, CYTB, ND1 e 16S resultou na obtenção de padrões filogenéticos similares aqueles apresentados pelos estudos com mtDNAs completos evidenciando a eficiência da metodologia empregada.

8. Referências Bibliográficas

- Albre J, Gers C, Legal L. (2008) Molecular phylogeny of the *Erebia tyndarus* (Lepidoptera, Rhopalocera, Nymphalidae, Satyrinae) species group combining CoxII and ND5 mitochondrial genes: a case study of a recent radiation. *Molecular Phylogenetic Evolution* 47(1):196-210
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology* 215: 403-410.
- Batista MV, Ferreira TA, Freitas AC, Balbino VQ (2011) An entropy-based approach for the identification of phylogenetically informative genomic regions of Papillomavirus. *Infection, Genetics and Evolution* 11: 2026-2033.
- Behura S e Severson DW (2013) Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biological Reviews* 88: 49–61.
- Bhutkar A, Russo S, Smith T F, Gelbart W M (2006) Techniques for multi-genome synteny analysis to overcome assembly limitations. *Genome Information* 7(2):152-61.
- Bleidorn C, Vogt L, Bartolomaeu S T (2003). New insights into polychaete phylogeny (Annelida) inferred from 18S rDNA sequences. *Molecular Phylogenetics and Evolution* 29(2):279-288.
- Boore JL (1999). Animal mitochondrial genomes *Nucleic Acids Research* 27:1767–1780.
- Bravo JP, Felipes J, Zanatta DB, Silva JL, Fernandez MA (2008) Sequence and analysis of the mitochondrial DNA control region in the sugarcane borer *Diatraea saccharalis* (Lepidoptera: Crambidae). *Brazilian Archives of Biology and Technology* 51: 671-677.
- Buzzi Z.J (2005). *Entomologia didática*. 4.ed. Curitiba: Ed. da UFPR. 347p
- Cao YQ, Ma C, Chen JY, Yang DR (2012) The complete mitochondrial genomes of two ghost moths, *Thitarodes renzhiensis* and *Thitarodes yunnanensis*: the ancestral gene arrangement in Lepidoptera. *BioMed Central Genomics* 13:276-296.
- Cho S, Zwick A, Regier JC, Mitter C, Cummings MP, Yao J, Du Z, Zhao H, Kawahara AY, Weller S, Davis Dr, Baixeras J, Brown JW, Parr C. (2011) Can Deliberately Incomplete Gene Sample Augmentation Improve a Phylogeny Estimate for the Advanced Moths and Butterflies (Hexapoda: Lepidoptera)? *Systematic Biology* 60(6):782–796.
- Coates BS, Sumerford DV, Hellmich RL, Lewis LC (2005) Partial mitochondrial genome sequences of *Ostrinia nubilalis* and *Ostrinia furnacalis*. *International Journal of Biologics Sciences* 1: 13–18.
- Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14(7): 1394-403.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5): 1792-97.
- EMBRAPA. Área de conhecimento: Cana-de-açúcar. Disponível em: <http://www.agencia.cnptia.embrapa.br/gestor/cana-de-acucar/Abertura.html> Acesso em ago de 2012.

- Felsenstein J (2004) The importance of proper model assumption in bayesian phylogenetics. *Systematic Biology* 53(2):265-77.
- Freitas AVL, Marini-Filho OJ (2011) Plano de Ação Nacional para Conservação de Lepidoptera Ameaçados de Extinção. Instituto Chico Mendes de Conservação da Biodiversidade, ICMBIO, Brasília. 124pp.
- Freitas MRT, Silva EL, Mendonça CES, Fonseca APP, Mendonça AL, Santos JS, Nascimento RR, Sant'ana AEG (2007) The biology of *Diatraea flavipennella* (Lepidoptera: Crambidae) reared under laboratory conditions. *Florida Entomologist* 90: 309-313.
- Gascuel O. (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution* 14:685-695.
- Gibson A, Gowri-Shankar V, Higgs PG, Rattray M (2005) A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. *Molecular Biology Evolution* 22(2):251–264.
- Guindon S, Dufayard J F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* 59(3):307-21.
- Heikkilä M, Kaila L, Mutanen M, Peña C, Wahlberg N (2012) Cretaceous origin and repeated tertiary diversification of the redefined butterflies. *Proceedings Biological Sciences* 279(1731):1093-9.
- Holder M, Lewis P O (2003) Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics*. 4(4):275-84.
- Hong M Y, Lee E M, Jo Y H, Park H C, Kim S R, Hwang J S, Jin B R, Kang P D, Kim K G, Han Y S, Kim I (2008) Complete nucleotide sequence and organization of the mitogenome of the silk moth *Caligula boisduvalii* (Lepidoptera: Saturniidae) and comparison with other lepidopteran insects. *Gene* 413(1-2):49-57.
- Hua J, Li M, Dong P, Cui Y, Xie Q, Bu W (2008) Comparative and phylogenomic studies on the mitochondrial genomes of Pentatomomorpha (Insecta: Hemiptera: Heteroptera). *BioMed Central Genomics*. 17(9):610.
- Huelsenbeck JPF, Ronquist R, Nielsen JP, Bollback (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294: 2310-2314.
- Katoh T, Chichvarkhin A, Yagi T, Omoto K (2005) Phylogeny and evolution of butterflies of the genus *Parnassius*: inferences from mitochondrial 16S and ND1 sequences. *Zoologie Science* 22(3):343-51.
- Kim MII, Wan X, Kim MJ, Jeong HC, Ahn NH, Kim KG, Han YS, Kim I (2010) Phylogenetic Relationships of True Butterflies (Lepidoptera: Papilionoidea) Inferred from COI, 16S rRNA and EF-1 α Sequences. *Molecules and Cells* 30: 409-425.
- Köhler M, Leitner B, Blaschitz M, Michel R, Aspöck H, Walochnik J, (2006) ITS1 sequence variabilities correlate with 18S rDNA sequence types in the genus *Acanthamoeba* (Protozoa:Amoebozoa). *Parasitol Research* 98: 86–93.
- Labandeira CC, Dilcheri DL, Davis DR, Wagner DL (1994) Ninety-seven million years of angiosperm-insect association: Paleobiological insights into the meaning of coevolution. *Proceedings of the Academy of Natural Sciences* 91: 12278-12282.
- Leavitt JR, Hiatt KD, Whiting MF, Song H (2013) Searching for the optimal data partitioning strategy in mitochondrial phylogenomics: A phylogeny of

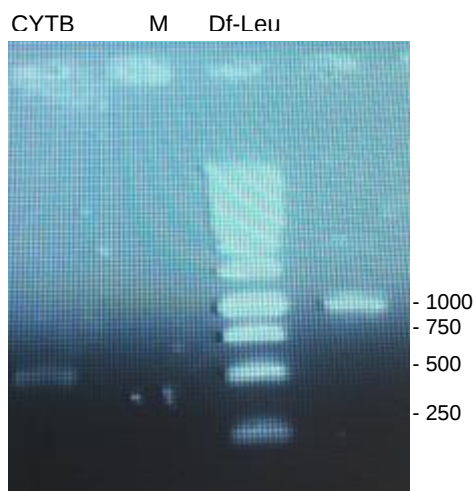
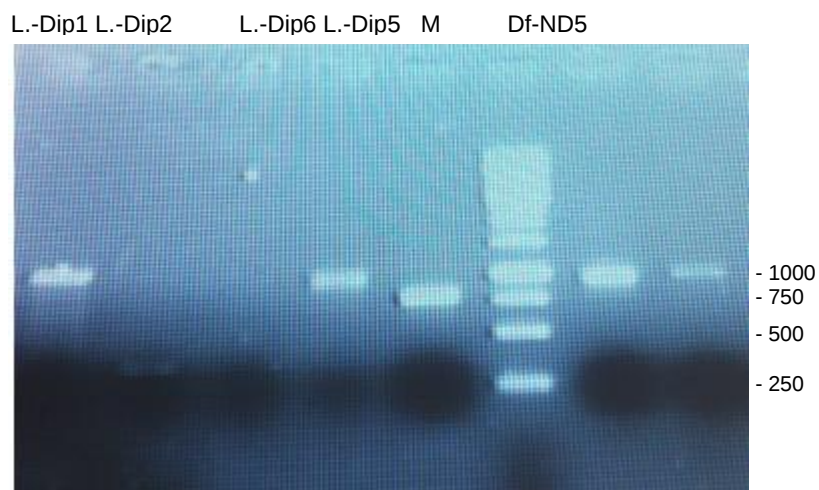
- Acridoidea (Insecta: Orthoptera: Caelifera) as a case study. *Molecular Phylogenetics and Evolution* 67: 494–508.
- Li W, Zhang X, Fan Z, Yue B, Huang F, King E, Ran J (2010) Structural characteristics and phylogenetic analysis of the mitochondrial genome of the sugarcane borer *Diatraea saccharalis* (Lepidoptera: Crambidae). *DNA and Cell Biology* 00: 1-6.
- Liu QN, Zhu BJ, Dai LS, Wei GQ, Liu CL (2012) The complete mitochondrial genome of the wild silkworm moth, *Actias selene*. *Gene* 505: 291–299.
- Long WH, Hensley SD (1972) Insect pests of sugar cane. *Annual Review of Entomology* 149 -176.
- Macedo N, Botelho PSM (1988) Controle integrado da broca da cana-de-açúcar *Diatraea saccharalis* (Fabricius 1794) (Lepidoptera Pyralidae). *Brasil Açucareiro* 162: 2-11.
- Mendonça AF (1996) Pragas da cana-de-açúcar. Maceió: insetos & Cia 200p.
- Meng X F, Shi M, Chen X X (2008) Population genetic structure of *Chilo suppressalis* (Walker) (Lepidoptera: Crambidae): strong subdivision in China inferred from microsatellite markers and mtDNA gene sequences. *Molecular Ecology* 17(12):2880-97.
- Mutanen M, Wahlberg N, Kaila L (2010) Comprehensive gene and taxon coverage elucidates radiation patterns in moths and butterflies. *Proceedings of the Royal Society Biological Sciences* Published online.
- Nazari V, Zakharov EV, Sperling FA (2007) Phylogeny, historical biogeography, and taxonomic ranking of Parnassiinae (Lepidoptera, Papilionidae) based on morphology and seven genes. *Molecular Phylogenetic Evolution* 42(1):131-56.
- Oliveira-Júnior JB, Silva CG, Marchiori CH (2007) Himenópteros parasitóides capturados em imaturos de *Lonomia* sp.(Lepidoptera: Saturniidae) oriundas de plantas de cafeeiro. *Revista Trópica – Ciências Agrárias e Biológicas* 1: 4.
- Ohshima I, Tanikawa-Dodo Y, Saigusa T, Nishiyama T, Kitani M, Hasebe M, Mohri H. (2010) Phylogeny, biogeography, and host-plant association in the subfamily Apaturinae (Insecta: Lepidoptera: Nymphalidae) inferred from eight nuclear and seven mitochondrial genes. *Molecular Phylogenetic Evolution* 57(3):1026-36.
- Pakendorf B, Stoneking M (2005) Mitochondrial DNA and Human Evolution. *Annual Review of Genomics and Human Genetics* 6: 165-183.
- PLANALSUCAR (1982) Guia: Principais pragas da cana-de-açúcar no Brasil. Piracicaba 28p.
- Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
- Regier J C, Mitter C, Zwick A, Bazinet A L, Cummings M P, Kawahara A Y, Sohn J C, Zwickl D J, Cho S, Davis D R, Baixeras J, Brown J, Parr C, Weller S, Lees D C, Mitter K T (2013) A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). *PLoS One* 8(3):e58568.
- Regier J C, Zwick A, Cummings M P, Kawahara A Y, Cho S, Weller S, Roe A, Baixeras J, Brown J W, Parr C, Davis D R, Epstein M, Hallwachs W, Hausmann A, Janzen D H, Kitching I J, Solis M A, Yen S H, Bazinet A L, Mitter C (2009) *BMC Evolutionary Biology* 9:280.

- Rogers J S, Swofford D L (1999) Multiple Local Maxima for Likelihoods of Phylogenetic Trees: A Simulation Study. *Molecular Biology and Evolution*. 16(8):1079–1085.
- Salvato P, Simonato M, Battisti A, Negrisoló E (2008) The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). *BioMed Central Genomics* 9:331-346.
- Sambrook J, Fritsch E F, Maniatis T (1989) *Molecular cloning: a laboratory manual*. 2. ed., 3. vol., Cold Spring Harbor Laboratory Press, New York, 253p.
- Santos M (2012). *Reconstrução Filogenética do Filo Arthropoda Baseada no Genoma Mitocondrial*. Dissertação (Mestrado em Genética) – Universidade de Federal de Pernambuco, Recife, 2012.
- Schmidt H A, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*. 18: 502-504.
- Silva-Brandão KL, Lyra ML, [...], Omoto C (2011) Exploitation of mitochondrial *nad6* as a complementary marker for studying population variability in Lepidoptera. *Genetics and Molecular Biology* 34(4):719-725.
- Silva U B (2008) *Reconstrução Filogenética de microorganismos a partir da Via de Reparo por Excisão de Nucleotídeo*. Dissertação (Mestrado em Genética e Biologia Molecular) – Universidade de Federal do Rio Grande do Norte, Natal, 2008.
- Simon C, Frati F, Becknbach A, Crespi B, Liu H, Flook P (1994) Evolution weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* 87: 651–701.
- Staden R, Beal KF, Bonfield JK (2000) The Staden package 1998. *Methods Molecular Biology* 132: 115–130.
- Tachi T (2013) Molecular phylogeny and host use evolution of the genus *Exorista* Meigen (Diptera: Tachinidae). *Molecular Phylogenetic Evolution*. 66(1): 401-11.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood evolutionary distance and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731-2739.
- Tavares E S (2005). *Relações filogenéticas, biogeografia histórica e evolução da organização de genes mitocondriais dos psitacídeos neotropicais (tribo Arini: Psittacidae: Psittaciformes)*. Tese (Doutorado em Ciências) – Universidade de São Paulo, São Paulo, 2005.
- Untergasser A, Veen HN, Rao X, Bisseling T, Guerts R, Leunissen JAM (2007) Primer3Plus an enhanced web interface to Primer3. *Nucleic Acids Res* 35: 71-74.
- van Nieukerken *et al.* (2011). Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. Ed. *Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness*.
- Wahlberg N, Braby M F, Brower A V, de Jong R, Lee M M, Nylin S, Pierce N E, Sperling F A, Vila R, Warren A D, Zakharov E (2005) Synergistic effects of combining morphological and molecular data in resolving the phylogeny of butterflies and skippers. *Proceedings Biological Sciences* 272(1572):1577-86.

- Wolstenholme DR (1992) Genetic novelties in mitochondrial genomes of multicellular animals. *Current Opinion in Genetics & Development* 2: 918-925
- Xia X, Lemey P (2009) Assessing substitution saturation with DAMBE. Pp. 615-630. Ed. *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny*. 2nd edition. Cambridge University Press
- Xia X, Xie Z (2001) DAMBE: data analysis and molecular biology and evolution. *Journal of Heredity* 92: 371-373.
- Xu W, Jameson D, Tang B, Higgs P G (2006) The relationship between the rate of molecular evolution and the rate of genome rearrangement in animal mitochondrial genomes *Journal Molecular Evolution* 63(3):375-92.
- Yukuhiro K, Sezutsu H, Itoh M, Shimizu K, Banno Y (2002) Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silkmoth, *Bombyx mandarina*, and its close relative, the domesticated silkmoth, *Bombyx mori*. *Molecular Biology Evolution* 19:1385–1389.
- Zhai B P, Du Y Z, Chai H N (2012) Characterization of the Complete Mitochondrial Genomes of *Cnaphalocrocis medinalis* and *Chilo suppressalis* (Lepidoptera: Pyralidae). *International Journal of Biological Sciences* 8: 561–579.

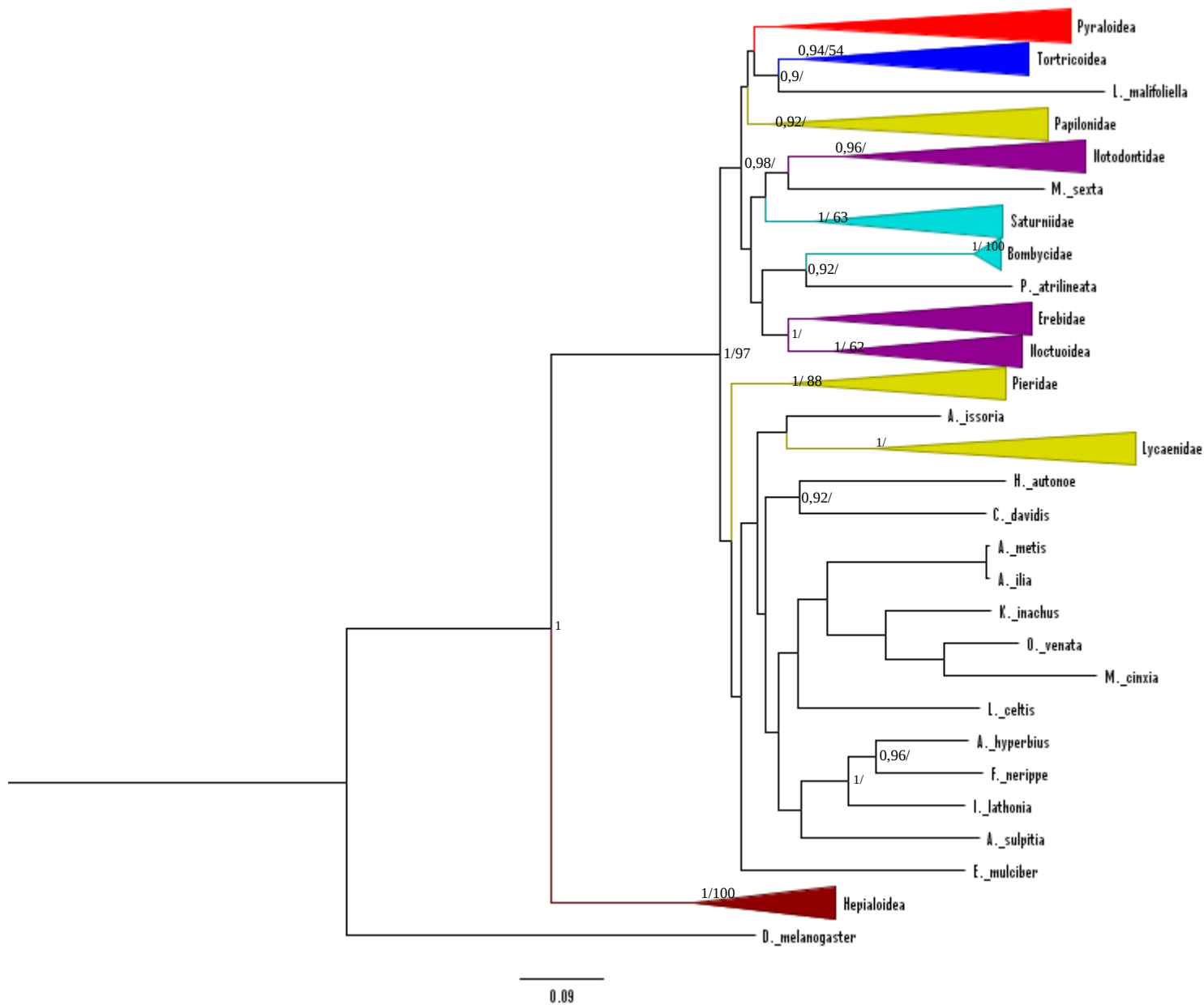
9. Apêndices

APÊNDICE A - Géis originais dos produtos amplificados dos pares de *primers* utilizados. M representa o marcador e L. é a abreviatura de LABBE

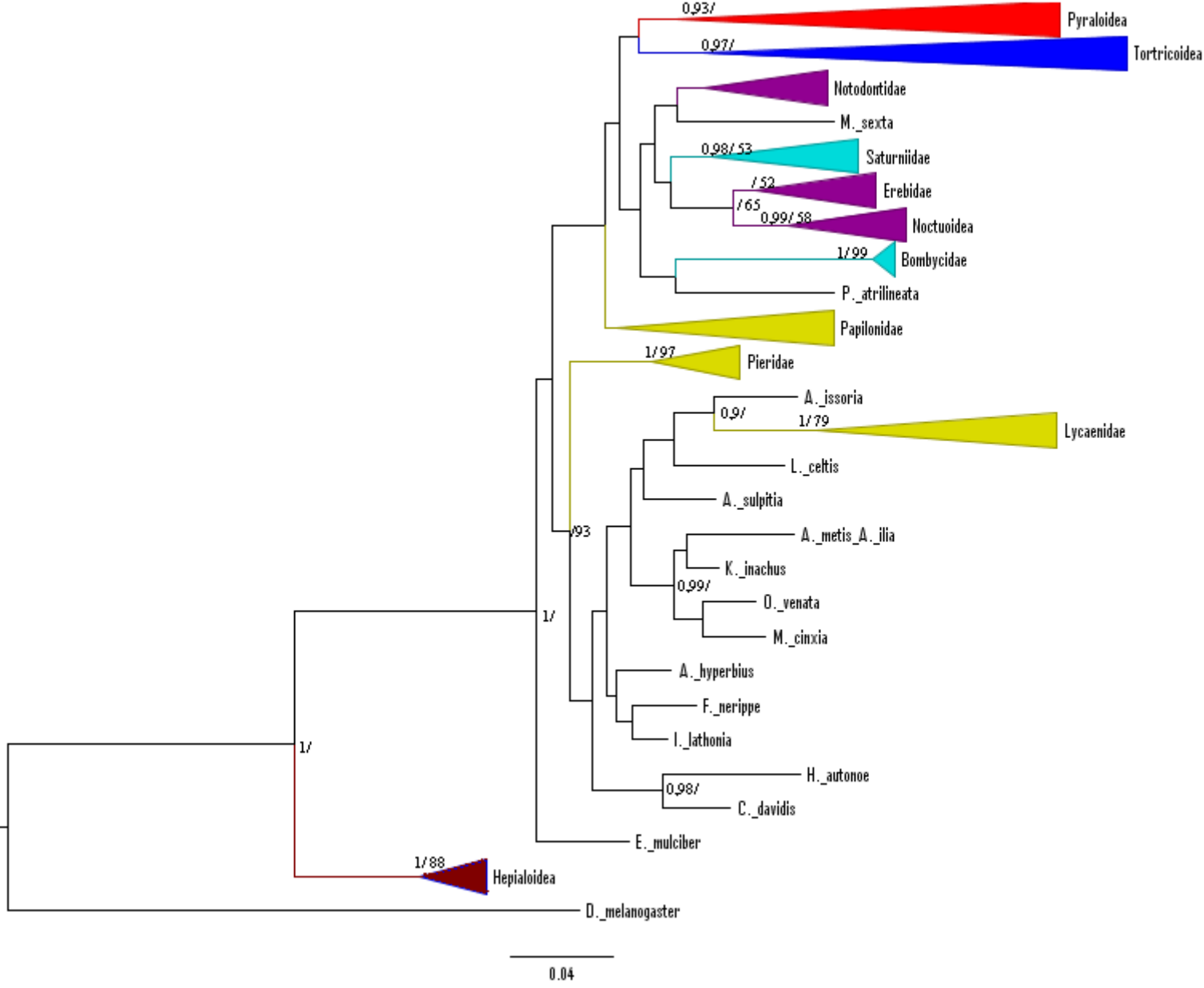


APÊNDICE B – Árvores filogenética de Lepidoptera a partir dos testes gerados neste trabalho. Os números representam os valores gerados pela análise bayesiana e máxima verossimilhança, respectivamente.

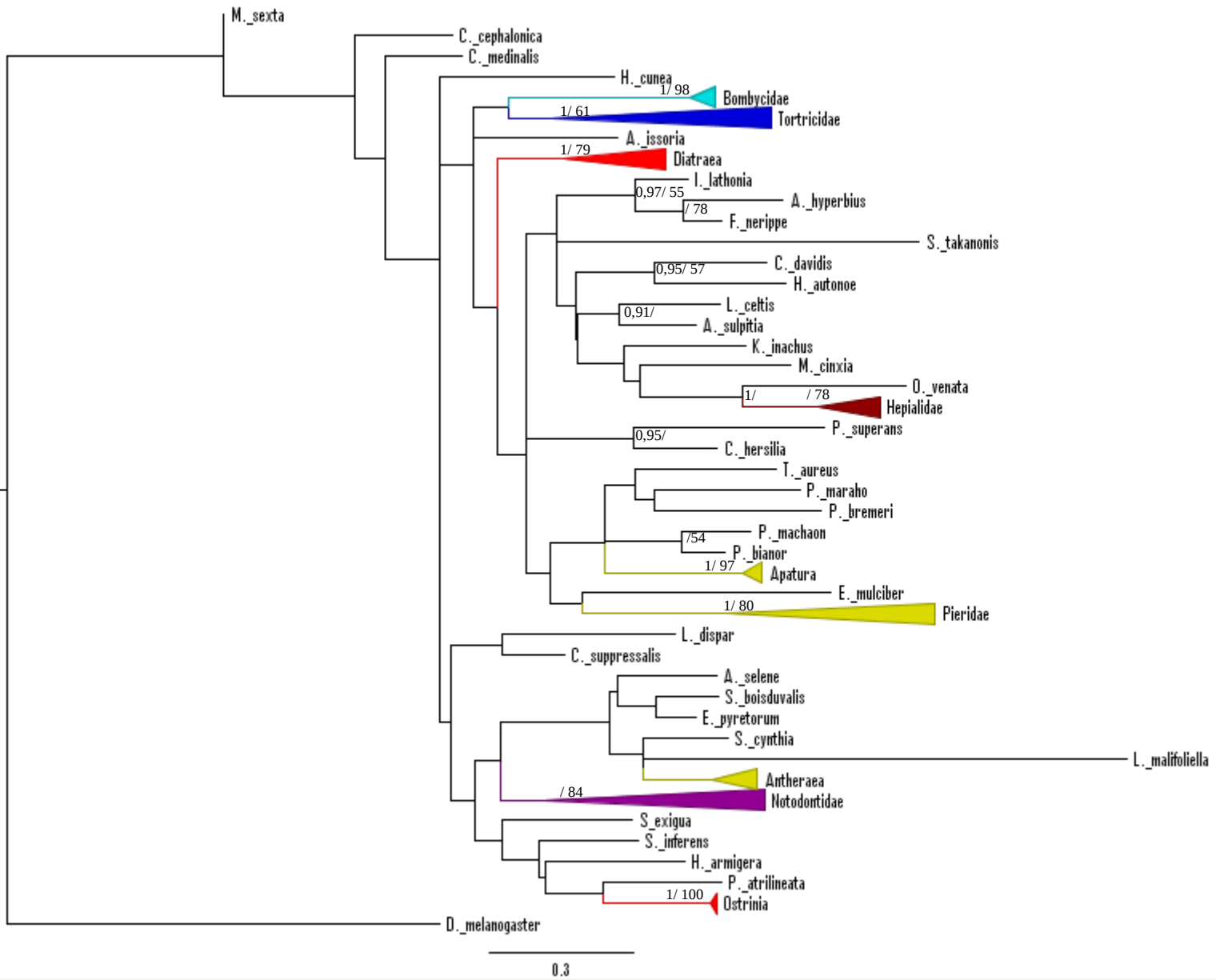
1.A.COI



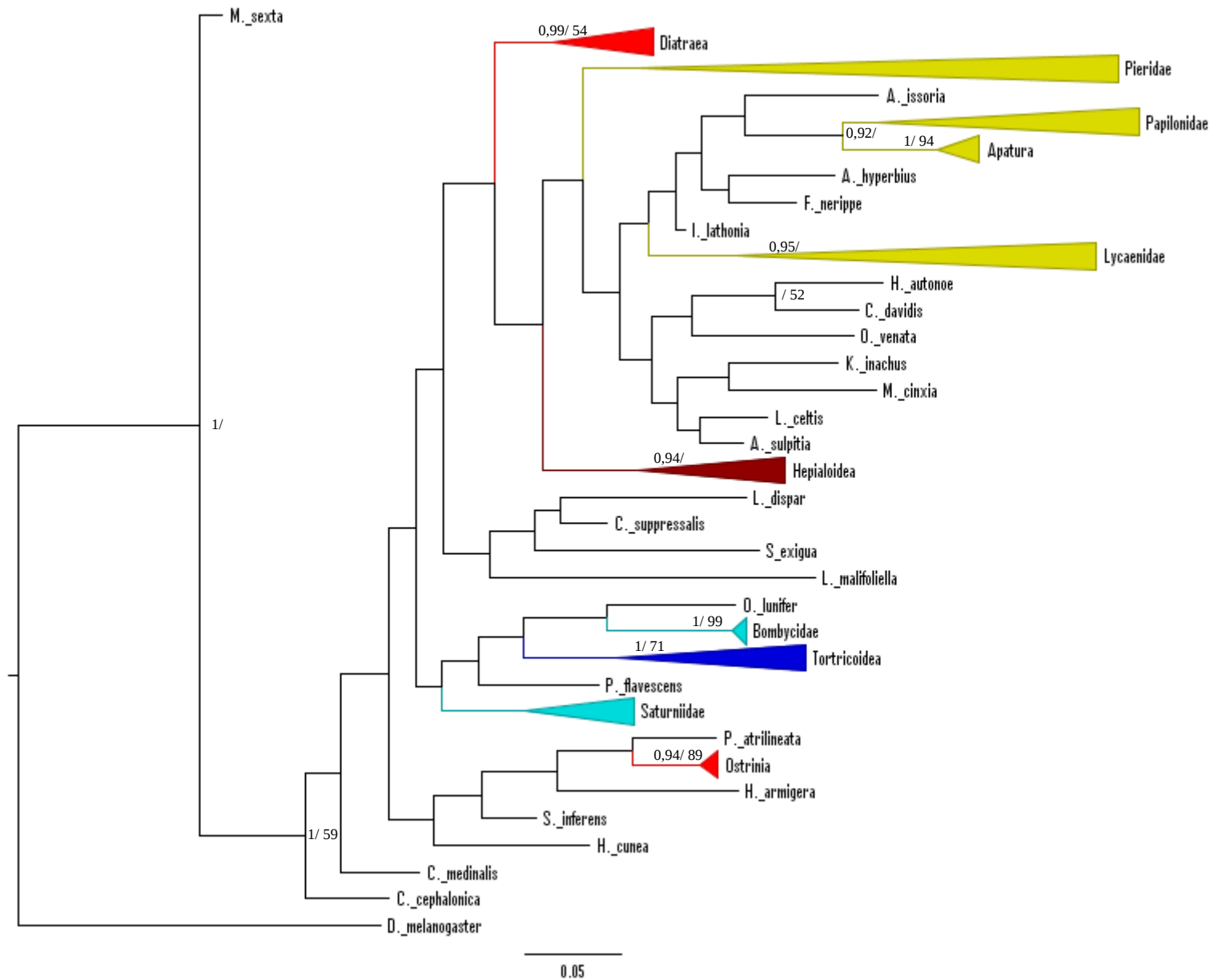
1.B.COI



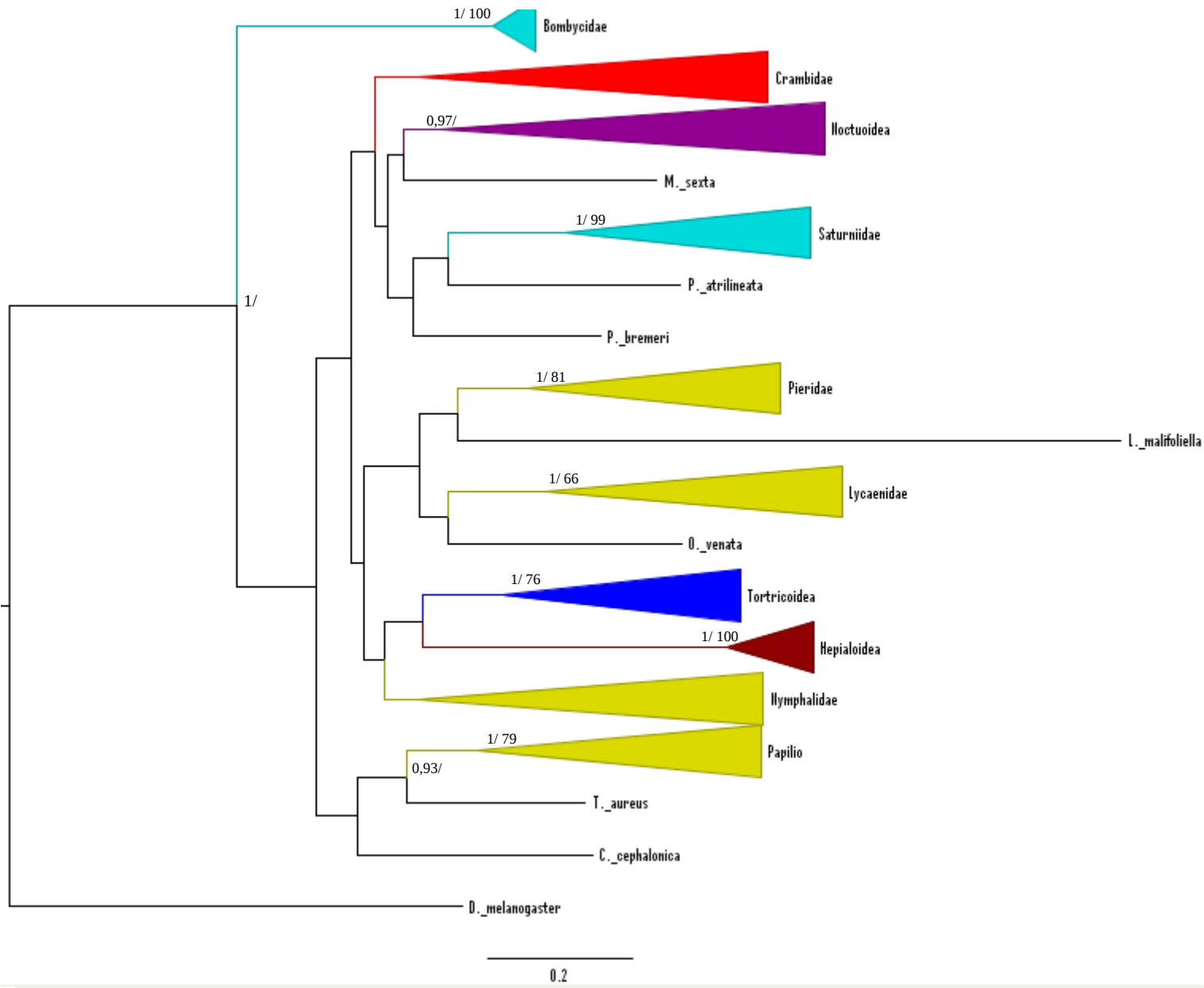
1.A.ATP6

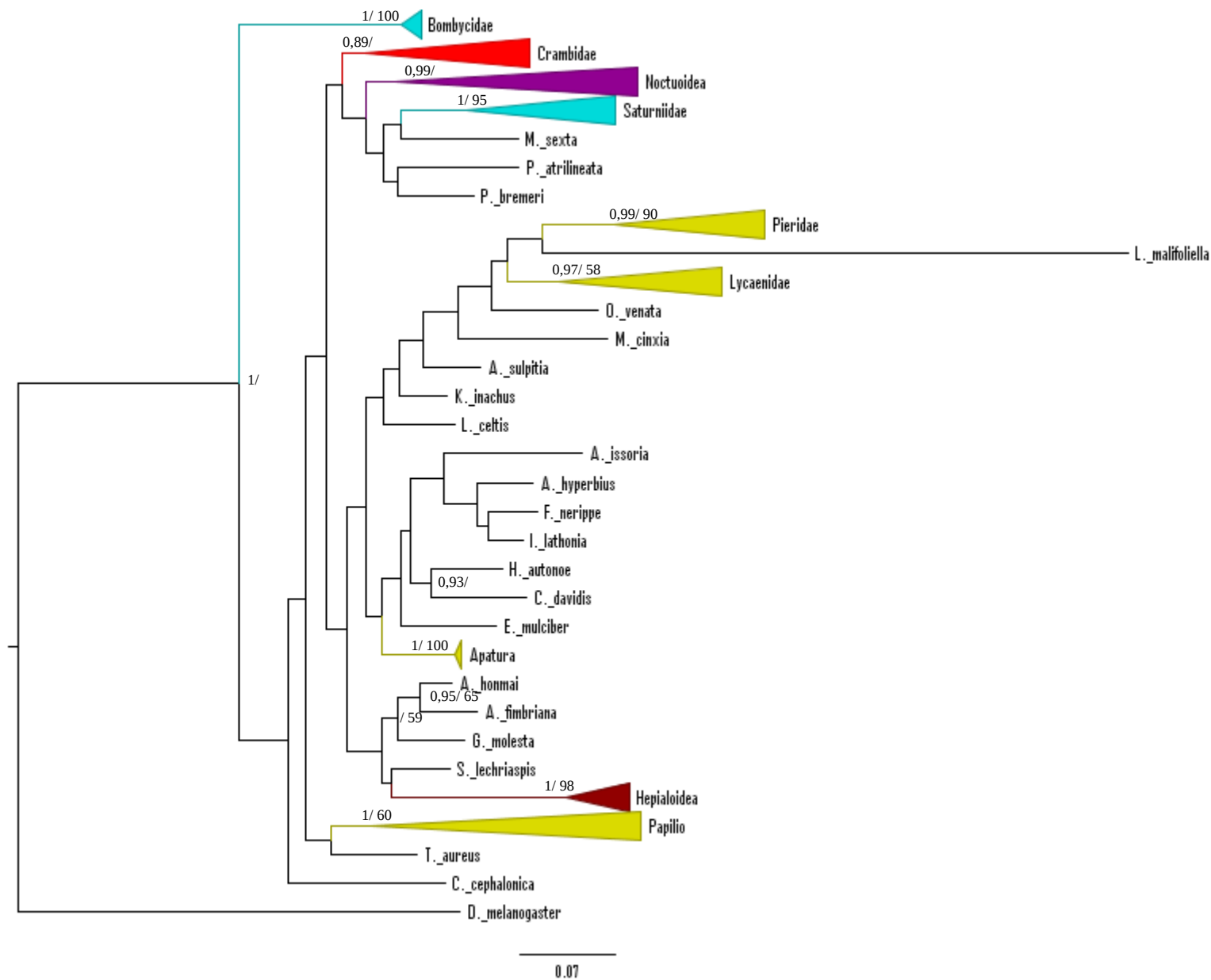


1.B.ATP6

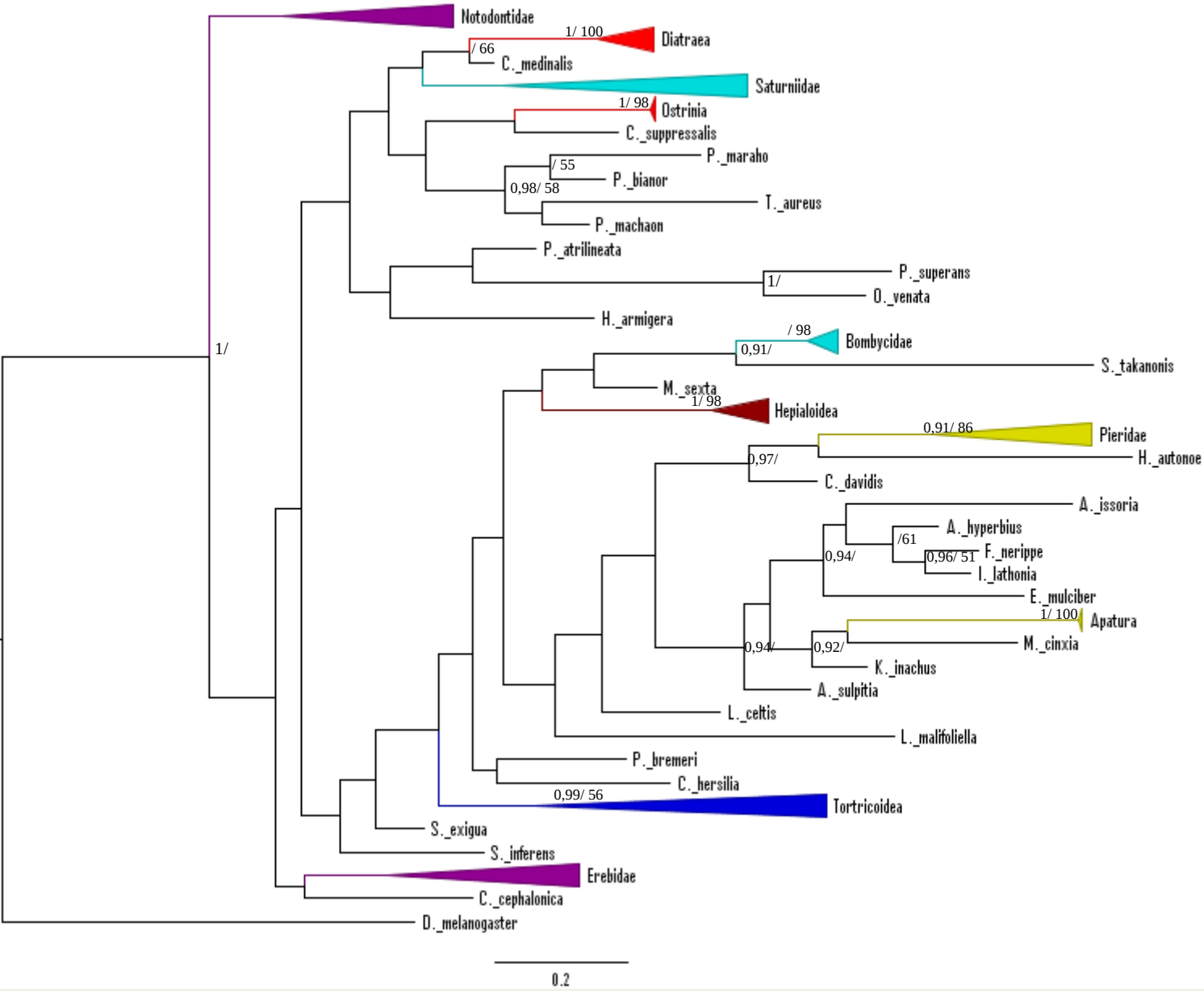


1.A.COIII



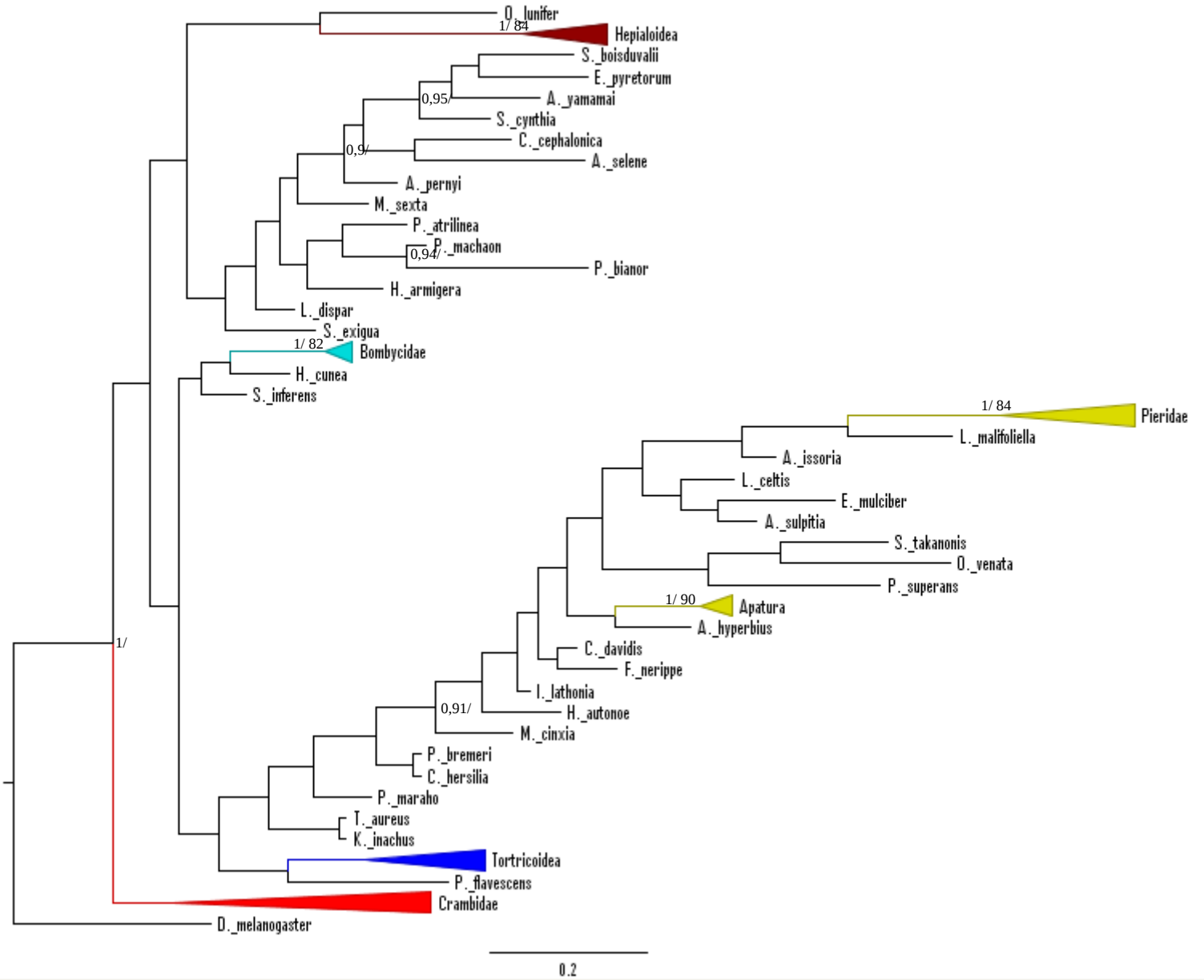


1.A. ND3

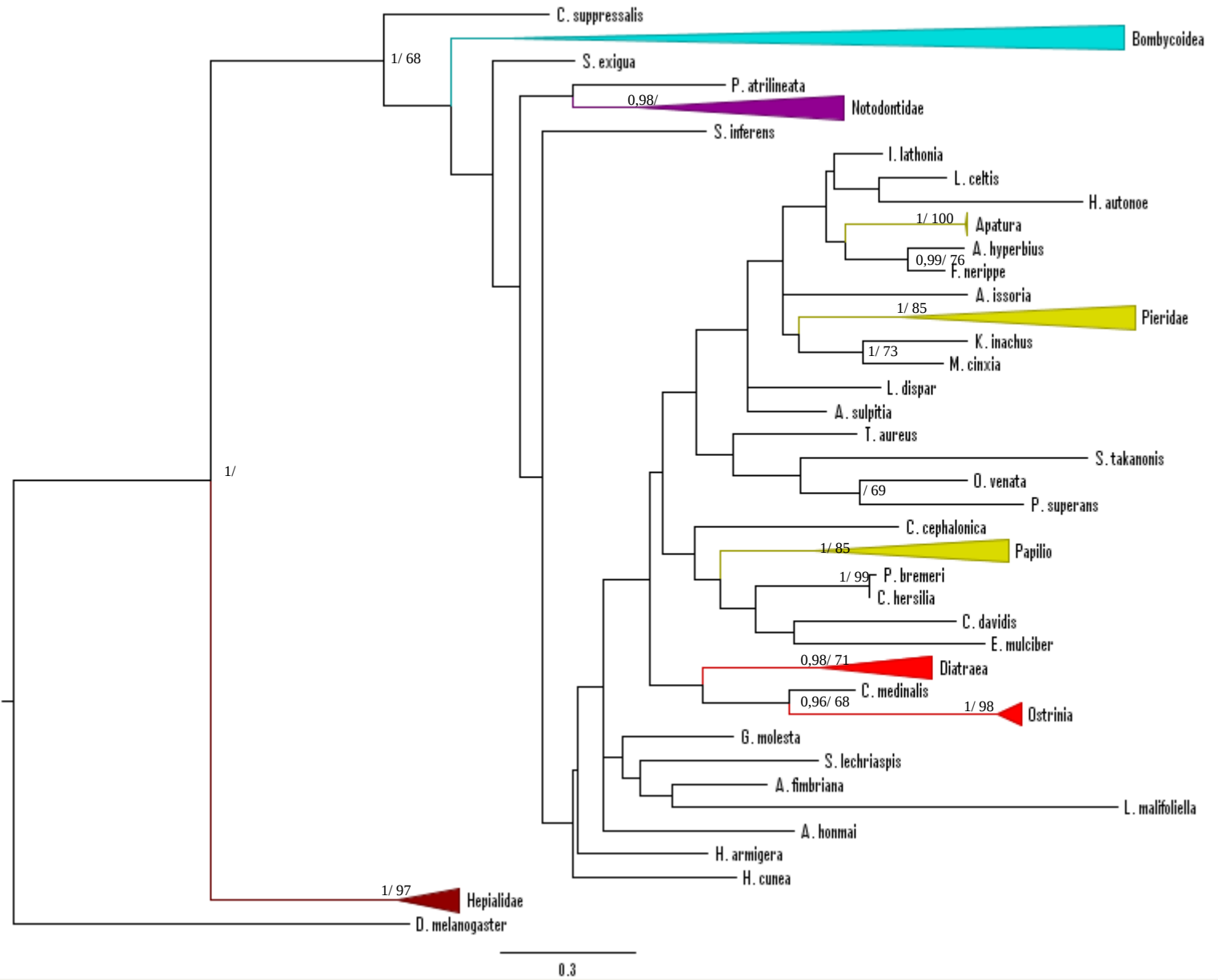


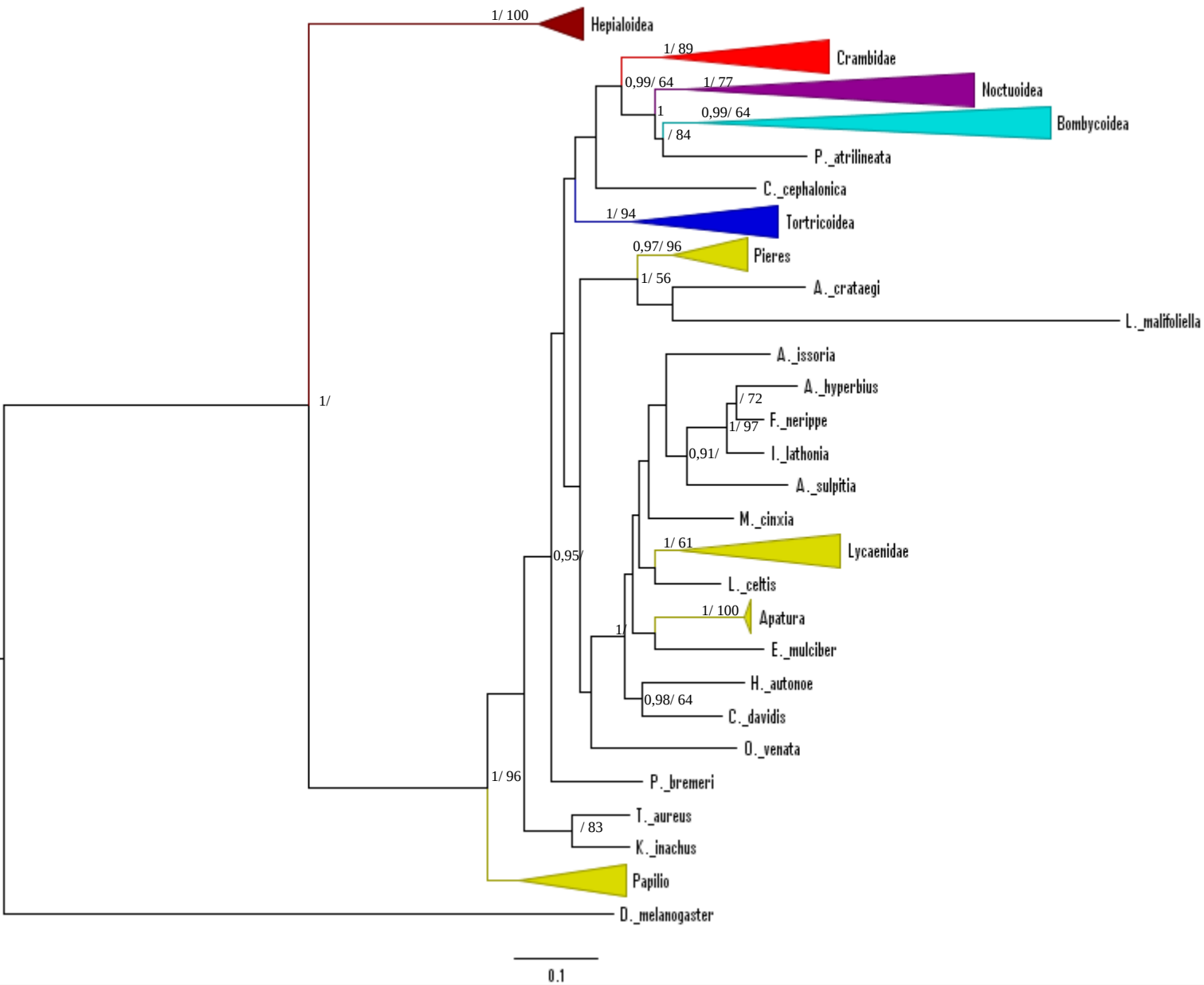


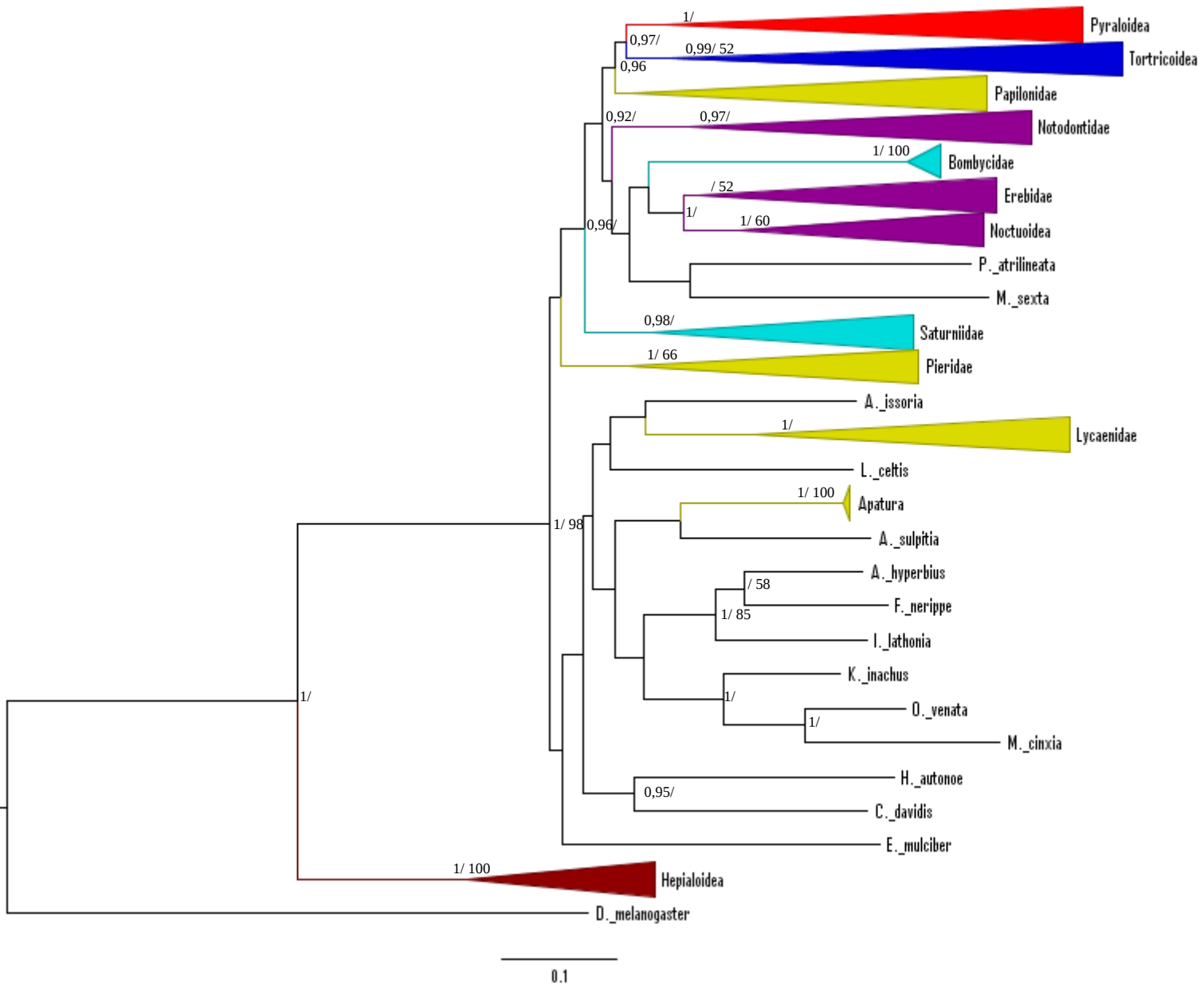
1.A.ND1

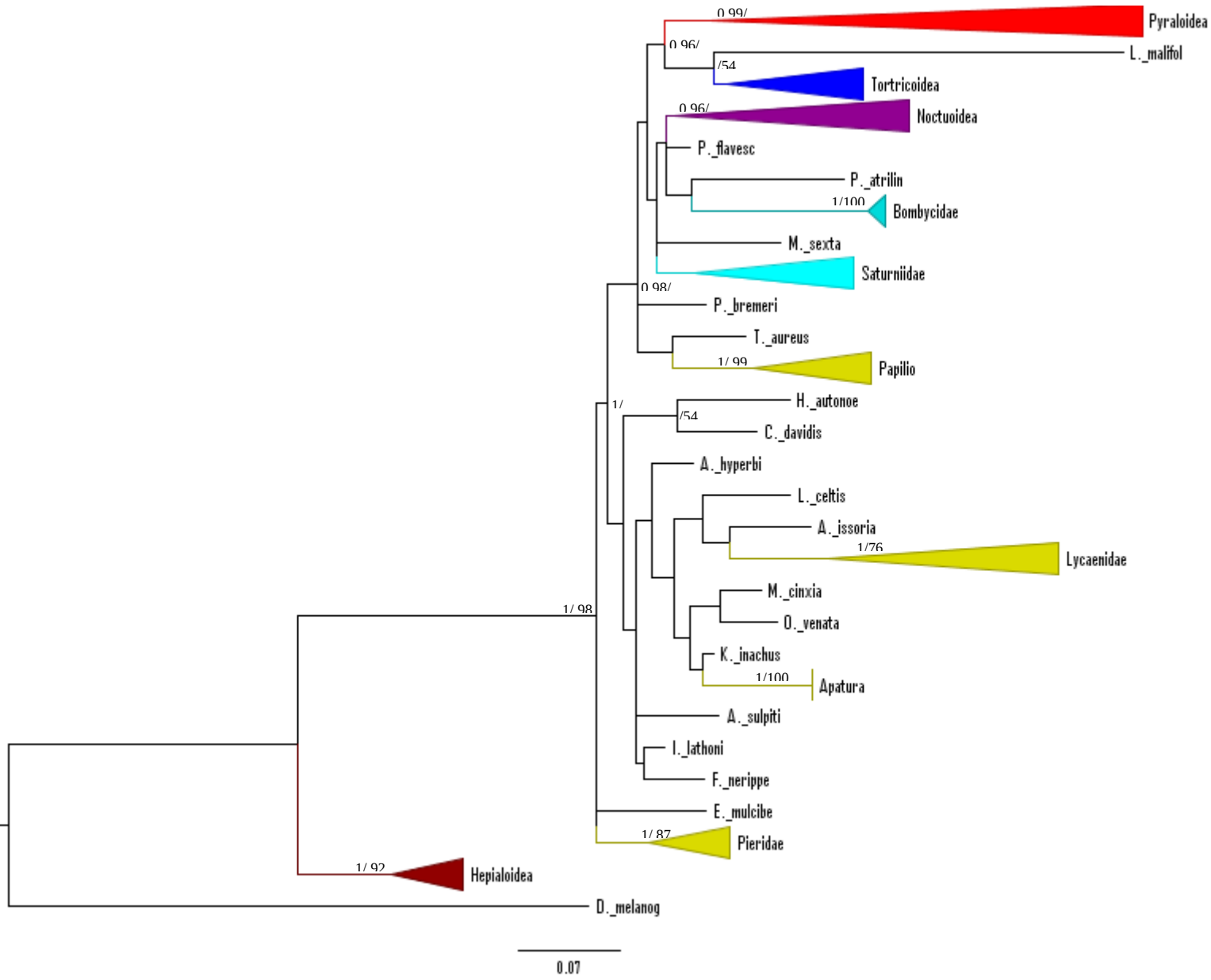


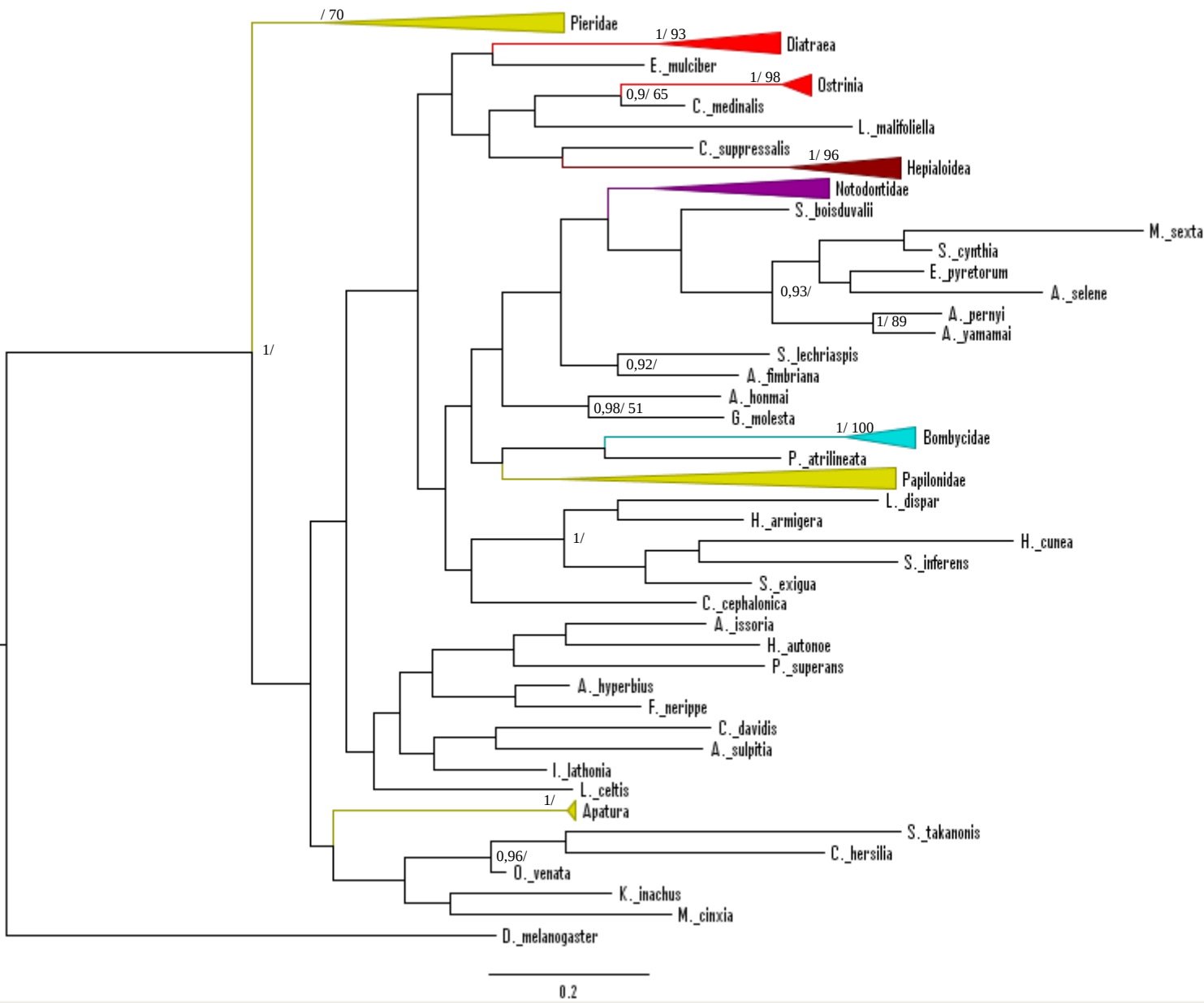
1.A.CYTB

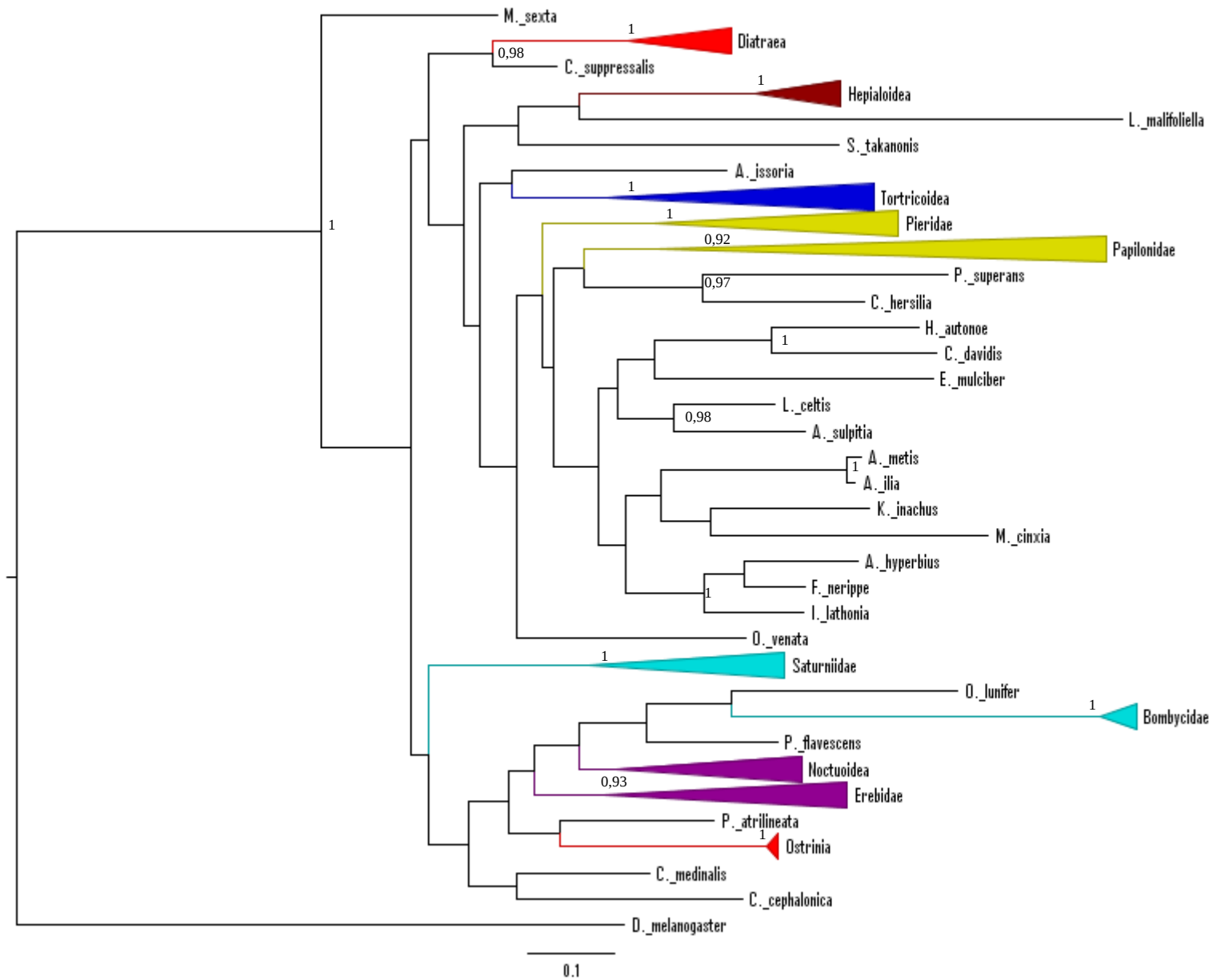


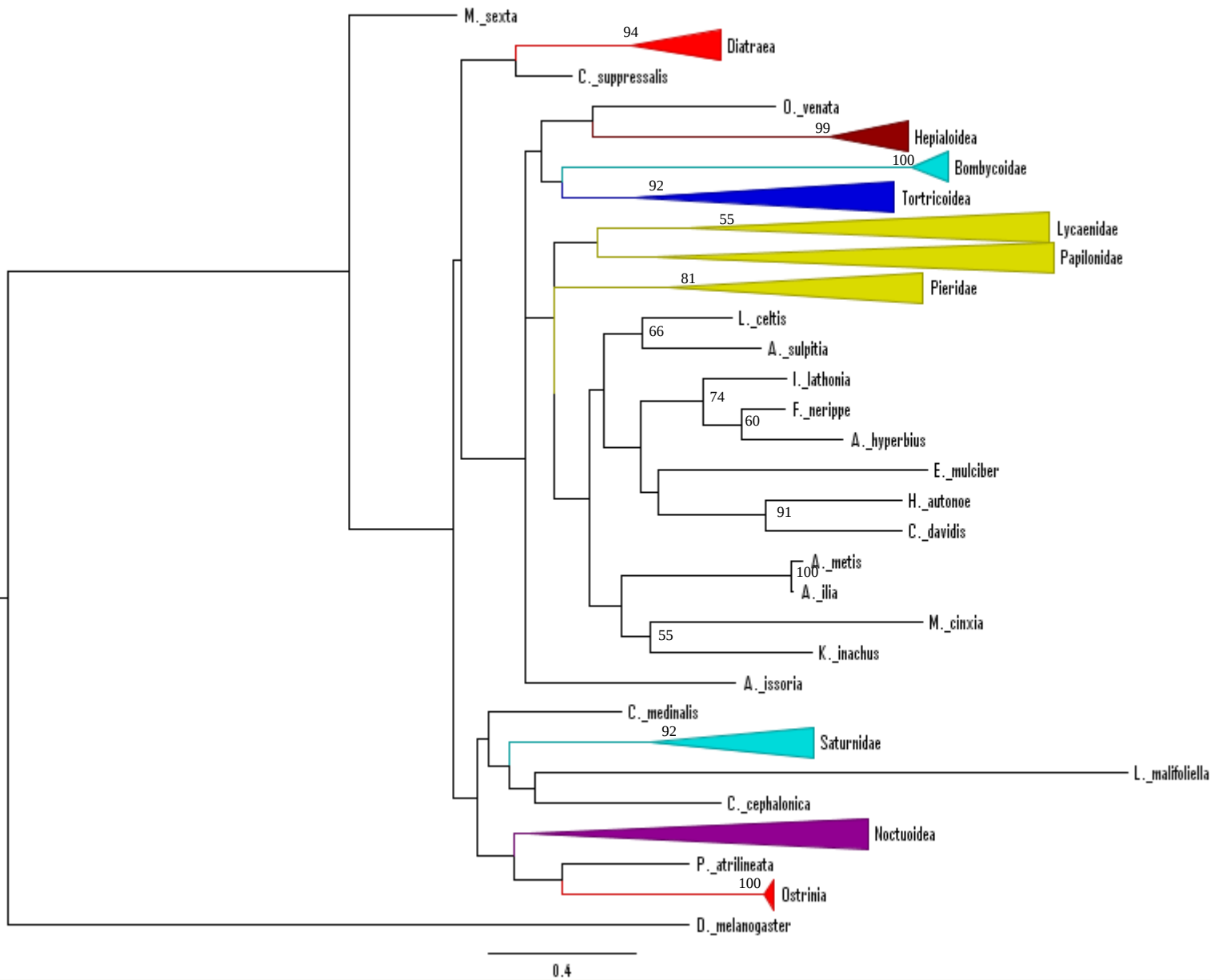


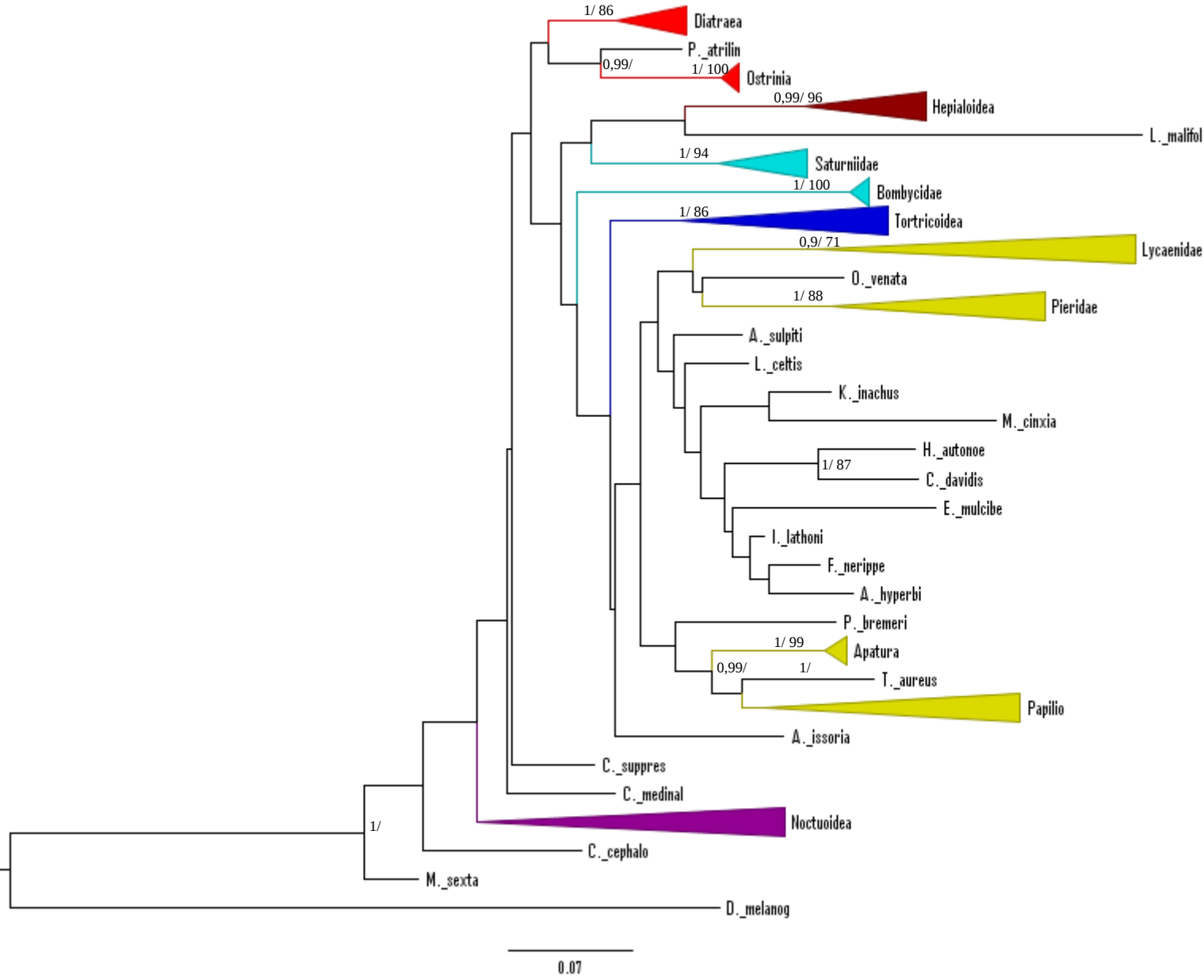


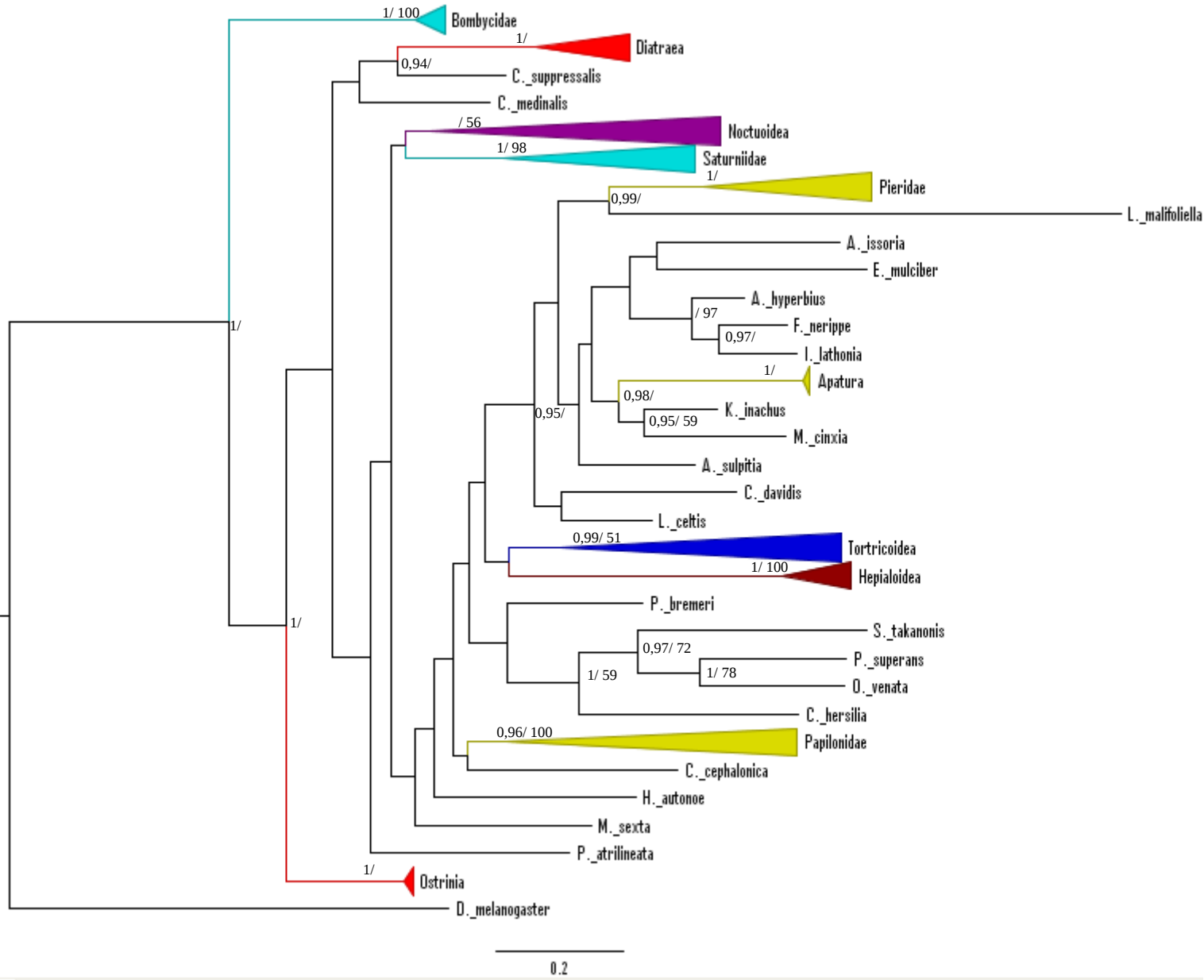




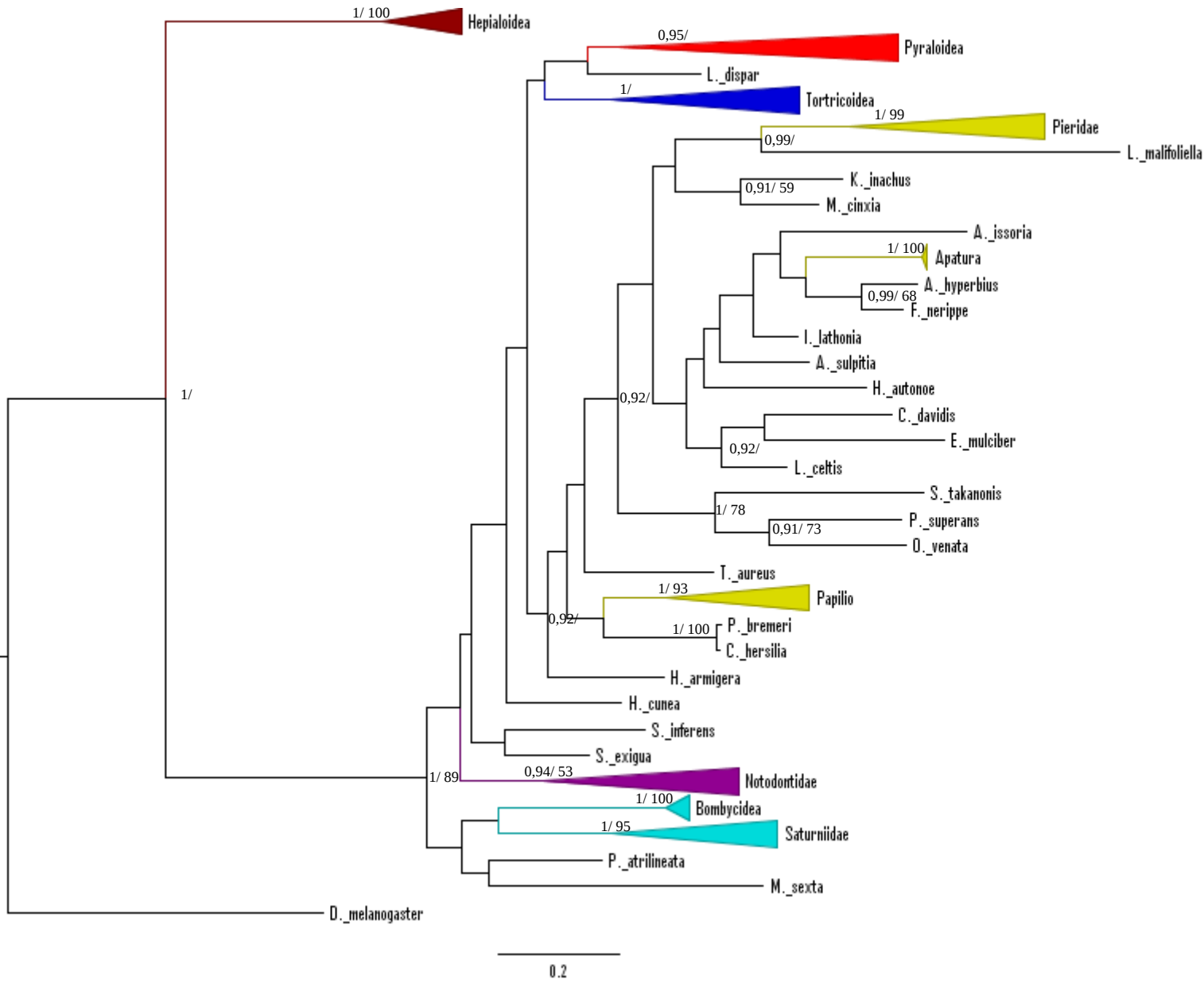








2.A.CYTB





10. Curriculum vitae (Lattes)

Patrícia Keytth Lins Rocha

Endereço para acessar este CV: <http://lattes.cnpq.br/6774917103336502>

Última atualização do currículo em 22/02/2013

Endereço eletrônico: patykeytth@gmail.com

Formação acadêmica/titulação

2011 - Mestrado em Genética. Universidade Federal de Pernambuco, UFPE, Recife - Brasil. Título: Mitogenômica comparativa de Lepidoptera: um modelo com o sequenciamento de *Diatraea flavipennella* (Lepidoptera: Crambidae).

Orientador: Marcos Antônio de Moraes Júnior. Co-orientador: Valdir de Queiroz Balbino. Bolsista do(a): Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco.

Formação Complementar

2012 - Extensão universitária em IV Curso de Bioinformática: Análise de dados molec. (Carga horária: 75h). Universidade Federal de Pernambuco.

2012 - Bioinformática: Recursos web aplicados ao estudo d. (Carga horária: 4h). Universidade Federal do Vale do São Francisco.

2012 - Programa genes como ferramenta para análise de div. (Carga horária: 4h). Universidade Federal do Vale do São Francisco.

2012 - Sequenciamento de DNA de nova Geração e Metagenô. (Carga horária: 3h). SBG

Atuação Profissional

1. Universidade Federal de Pernambuco - UFPE

Vínculo institucional 2011 – Atual. Vínculo: Discente de Mestrado.

2. Universidade Estadual da Paraíba - UEPB

Vínculo institucional 2012 – Atual. Vínculo: Servidor Público, Enquadramento

Funcional: Técnica de Laboratório em Genética.

Áreas de atuação:

1. Genética

Projetos de pesquisa

2011 - Atual

Mitogenômica comparativa de Lepidoptera: um modelo com o sequenciamento de *Diatraea flavipennella* (Lepidoptera: Crambidae)

Situação: Em andamento; Natureza: Pesquisa.

Integrantes: Patrícia Keyth Lins Rocha / José Roberto tavares - Integrante /

Hebert Álvaro Abreu de Siqueira - Integrante / Marcos Antonio de Moraes Júnior -

Coordenador / Valdir de Queiroz Balbino - Integrante / Marco Antonio Oliveira dos Santos - Integrante.

Resumos publicados em anais de congressos

1. ROCHA, P. K. L. ; TAVARES, J. R. ; SIQUEIRA, H. A. A. ; MORAIS, M. A. ;

BALBINO, V. Q. . Mitogenoma da broca pequena da cana-de-açúcar. In: XIX

Encontro de Genética do Nordeste, 2012, Juazeiro - BA, Petrolina - PE. Resumos do XIX Encontro de Genética do Nordeste, 2012.

2. ROCHA, P. K. L. ; SANTOS, M. A. O. ; FIGUEREDO JUNIOR, C. A. S. ; TAVARES, J. R. ; MORAIS, M. A. ; BALBINO, V. Q. . Sequencing of mitochondrial DNA of *Chrysomya megacephala*. In: I Simpósio Brasileiro de Identificação Molecular de espécies, 2012, Foz do Iguaçu. Anais eletrônico, 2012.

3. ROCHA, P. K. L. ; SANTOS, M. A. O. ; SILVA, L. G. ; ARAGAO, N. C. ; FIGUEREDO JUNIOR, C. A. S. ; COSTA JUNIOR, C. R. L. ; TAVARES, J. R. ; MORAIS, M. A. ; BALBINO, V. Q. . Mitogenome of the Main Vector of American Visceral Leishmaniose. In: 58º Congresso Brasileiro de Genética., 2012, Foz do Iguaçu. 58º Congresso Brasileiro de Genética, 2012.

4. SILVA, L. G. ; ROCHA, P. K. L. ; FIGUEREDO JUNIOR, C. A. S. ; ARAGAO, N. C. ; FREITAS, M. T. S. ; SANTOS, M. A. O. ; TAVARES, J. R. ; MORAIS, M. A. ; BALBINO, V. Q. . Sequencing Mitochondrial Genome of *Lutzomyia umbratilis*, vector of Cutaneous Leishmaniasis. In: 58º Congresso Brasileiro de Genética., 2012, Foz do Iguaçu. 58º Congresso Brasileiro de Genética, 2012.

5. ARAGAO, N. C. ; ROCHA, P. K. L. ; SANTOS, M. A. O. ; SILVA, L. G. ; MARCONDES, C. B. ; BALBINO, V. Q. . Primeira Parte do Genoma Mitocondrial de *Psorophora ferox* (Diptera: Culicidae). In: XXIV Congresso Brasileiro de Entomologia, 2012, Curitiba. XXIV Congresso Brasileiro de Entomologia, 2012.

6. ROCHA, P. K. L. ; TAVARES, J. R. ; SANTOS, M. A. O. ; SIQUEIRA, H. A. A. ; MORAIS, M. A. ; BALBINO, V. Q. . Mitogenoma de *Diatraea flavipennella*. In: II Jornada de pços-Graduação de Genética da UFPE, 2012, Recife. II Jornada de pços-Graduação de Genética da UFPE, 2012.

Apresentações de Trabalho

1. ROCHA, P. K. L. ; TAVARES, J. R. ; SIQUEIRA, H. A. A. ; MORAIS, M. A. ; BALBINO, V. Q. . Mitogenoma da broca pequena da cana-de-açúcar.. 2012. (Apresentação de Trabalho/Outra).

2. ROCHA, P. K. L. ; SANTOS, M. A. O. ; FIGUEREDO JUNIOR, C. A. S. ; MORAIS, M. A. ; BALBINO, V. Q. . Sequencing of mitochondrial DNA of *Chrysomya megacephala*. 2012. (Apresentação de Trabalho/Simpósio).

3. ROCHA, P. K. L. ; SANTOS, M. A. O. ; SILVA, L. G. ; ARAGAO, N. C. ; COSTA JUNIOR, C. R. L. ; TAVARES, J. R. ; MORAIS, M. A. ; BALBINO, V. Q. Mitogenome of the Main Vector of American Visceral Leishmaniasis. 2012. (Apresentação de Trabalho/Congresso).

4. ROCHA, P. K. L. ; TAVARES, J. R. ; SIQUEIRA, H. A. A. ; MORAIS, M. A. ; BALBINO, V. Q. . Mitogenoma de *Diatraea flavipennella*. 2012. (Apresentação de Trabalho/Outra).

Participação em eventos, congressos, exposições e feiras

- 1.** XIX Encontro de Genética do Nordeste. Mitogenoma da broca pequena da cana-de-açúcar. 2012. (Congresso).
- 2.** 58 Congresso Brasileiro de Genética. Mitogenome of the Main Vector of American Visceral Leishmaniose. 2012. (Congresso).
- 3.** I Simpósio Brasileiro de Identificação molecular de espécies. Sequencing of mitochondrial DNA of *Chrysomya megacephala*. 2012. (Simpósio).
- 4.** II Jornada de Pós-Graduação em Genética da UFPE. Mitogenoma de *Diatraea flavipennella*. 2012. (Outra).