

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA



Marco Antonio Tonus Marinho

**“Análises genético-evolutivas em espécies da família Calliphoridae
(Diptera: Brachycera: Calyptratae)”**

Este exemplar corresponde à redação final da tese defendida pelo(a) candidato (a) <u>Marco Antonio Tonus Marinho</u>

e aprovada pela Comissão Julgadora.

Tese apresentada ao Instituto de Biologia para obtenção do Título de Doutor em Genética e Biologia Molecular, na área de Genética Animal e Evolução.

Orientadora: Prof^a. Dr^a. Ana Maria Lima de Azeredo-Espin

Co-Orientador: Prof. Dr. Nilson Ivo Tonin Zanchin

Campinas, 2012

FICHA CATALOGRÁFICA ELABORADA POR
ROBERTA CRISTINA DAL' EVEDOVE TARTAROTTI – CRB8/7430
BIBLIOTECA DO INSTITUTO DE BIOLOGIA - UNICAMP

M338a Marinho, Marco Antonio Tonus, 1984-
Análises genético-evolutivas em espécies da família
Calliphoridae (Diptera: Brachycera: Calyptratae) / Marco
Antonio Tonus Marinho. – Campinas, SP: [s.n.], 2012.

Orientador: Ana Maria Lima de Azeredo-Espin.
Coorientador: Nilson Ivo Tonin Zanchin.
Tese (doutorado) – Universidade Estadual de
Campinas, Instituto de Biologia.

1. Oestroidea. 2. Mosca-varejeira. 3. Filogenia –
Aspectos moleculares. 4. DNA espaçador ribossômico.
5. Estrutura secundária de RNA. I. Azeredo-Espin, Ana
Maria Lima de, 1955-. II. Zanchin, Nilson Ivo Tonin. III.
Universidade Estadual de Campinas. Instituto de
Biologia. IV. Título.

Informações para Biblioteca Digital

Título em Inglês: Genetic and evolutionary analyses in species of the Calliphoridae family (Diptera: Brachycera: Calyptratae)

Palavras-chave em Inglês:

Oestroidea

Blowflies

Phylogeny – Molecular aspects

Ribosomal spacer DNA

RNA secondary structure

Área de concentração: Genética Animal e Evolução

Titulação: Doutor em Genética e Biologia Molecular

Banca examinadora:

Ana Maria Lima de Azeredo Espin [Orientador]

Louis Bernard Klaczko

André Victor Lucci Freitas

Dalton de Souza Amorim

Cláudia Augusta de Moraes Russo

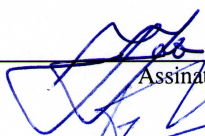
Data da defesa: 15-02-2012

Programa de Pós Graduação: Genética e Biologia Molecular

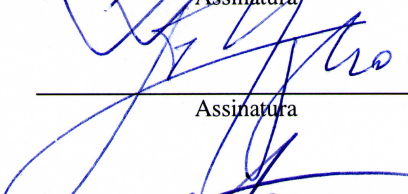
Campinas, 15 de fevereiro de 2012

BANCA EXAMINADORA

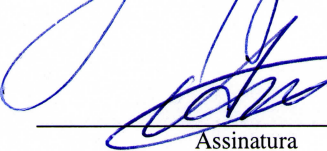
Prof(a). Dra. Ana Maria Lima de Azeredo-Espin
(Orientadora)


Assinatura

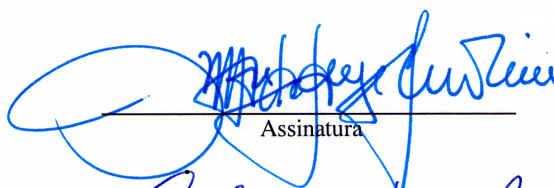
Prof(a). Dra. Louis Bernard Klaczko


Assinatura

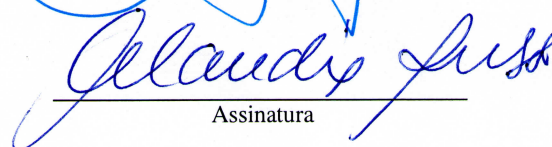
Prof(a). Dr(a) . André Victor Lucci Freitas


Assinatura

Prof(a). Dr(a) . Dalton de Souza Amorim


Assinatura

Prof(a). Dr(a) . Cláudia Augusta de Moraes Russo


Assinatura

Prof(a). Dr(a) . Karina Lucas da Silva Brandão

Assinatura

Prof(a). Dr(a) . Jörg Kobarg

Assinatura

Prof(a). Dr(a) . Arício Xavier Linhares

Assinatura

"Vi Veri Veniversum Vivus Vici"

("Pelo poder da verdade, eu, enquanto vivo, conquistei o universo")

V de Vingança (Alan Moore & David Lloyd).

"Omnia mutantur, nihil interit"

("Tudo muda, nada perece")

Metamorfoses (Ovídio)

Aos meus pais Antonio e Maria Bernadete,
minha irmã Patrícia e minha sobrinha Alice.

Agradecimentos

Gostaria de agradecer, primeiramente e acima de tudo, à minha família, em especial a meus pais, Antonio e Maria Bernadete, que, desde cedo, sacrificaram muito de suas vidas em prol da minha formação, tanto intelectual quanto pessoal, e a quem hoje devo, se não tudo, muito do que sou; e a minha irmã Patrícia e meu cunhado Denys, que sempre me apoiaram muito e que, há cerca de quase dois anos, proporcionaram a minha família uma de suas maiores alegrias: o nascimento de minha sobrinha Alice.

Agradeço também à Prof^a Ana Maria Lima de Azeredo-Espin, que há sete anos me aceitou como aluno no Laboratório de Genética e Evolução Animal (CBMEG / UNICAMP) e que, desde então, tem orientado e supervisionado todo o crescimento científico e acadêmico que alcancei.

Agradeço aos colegas de laboratório, atuais e passados, que por muitos anos foram, mais do que companheiros de trabalho, amigos: Ana Cláudia, Ana Carolina (que me acompanhou durante toda a iniciação científica e a quem devo muito do que sei hoje), Tatiana, Marcos Túlio, Gustavo (Johnny), Luciane, Karina (e Marcelo e Miguel) Brandão, Aline, Priscila (amiga dentro e fora do lab, para toda a vida), Joan, Bárbara, Pedro, Salete, Alessandra, Daniel, Alberto, Ana Paula, Luana, Gabriel e Giselle. Um agradecimento especial à Rosângela, que mais do que a técnica do laboratório, foi também a pessoa que mais aguentou meus estresses ao longo desses anos. Um agradecimento mais do que especial aos amigos de todas as horas, Pablo, Mariana e Renato, os quais compartilharam comigo todas as alegrias e decepções do período do doutorado e muito mais além dele.

Agradeço também a todos os demais amigos que fiz ao longo do meu tempo de permanência no CBMEG, sobretudo Eduardo Kiota, pelos conhecimentos químicos e risadas, e Renato Vicentini, pelos conhecimentos computacionais e cervejas.

Agradeço também ao Prof^o Nilson Ivo Tonin Zanchin, que aceitou ser meu co-orientador e que me recebeu, por um curto porém bastante significativo período de tempo, no então Laboratório de Biologia Molecular do LNLS. Agradeço também a todos os companheiros de laboratório e amigos que fiz por lá: Patrícia (que me acompanhou mais diretamente no período de tempo que permaneci no laboratório, me ensinando muitas coisas) (e Miúdo), Beatriz (e Jorge), Juliana, Thais, Sandra, Dani, Larissa, Thiago, Carlos, Melissa, Elaine, Tereza e Wanderley.

Agradeço a todos os meus amigos, muitos de longa data e outros que fiz durante a vida acadêmica, os quais, por precaução, prefiro não nomear individualmente por medo de esquecer alguém. A todos vocês que sempre estiveram ao meu lado, fisicamente ou, por algumas vezes, apenas em pensamento, compartilhando todos os momentos bons e ruins da minha vida: muito obrigado! Cabe aqui um agradecimento particular a todos os ex-moradores da Casa do Edge, que me acolheram por três anos de minha vida, e que me proporcionaram, além de alguns estresses gerados pela convivência (inevitáveis, infelizmente), muitos momentos excepcionais....

Agradeço aos professores Louis Bernard Klaczko, André Victor Lucci Freitas, Cláudia Augusta de Moraes Russo e Dalton de Souza Amorim por terem aceitado o convite para minha banca de tese e também pelas muitas e valorosas contribuições que tenho certeza serão dadas a este trabalho. Ao Prof^o Louis, um agradecimento especial por ter me mostrado, ao longo das disciplinas de graduação e pós que cursei, o lado deslumbrante e fantástico da biologia.

Um agradecimento também a todos aqueles que contribuíram para esse trabalho, enviando espécimes e os identificando: Alexandre Ururahy Rodrigues, Cláudio José Barros de Carvalho, Silvio Nihei, Dalton de Souza Amorim, Maria Cristina Espósito, Arício X. Linhares, Martin H. Villet, Bernard Greenberg, Michelle L. Harvey, Carole Ames, Rod Mahon, Bernard Chauvet, Élvia E. S. Vianna e Jeffrey D. Wells. Um agradecimento também ao aluno de iniciação científica Daniel F. Paulo, por ceder parte das sequências da região 16S produzidas por ele, as quais foram incluídas nas análises descritas no capítulo III.

Agradeço à FAPESP pelo financiamento ao longo desses anos, tanto na iniciação científica (Processo nº 05/50100-5) quanto no doutorado (Processo nº 06/61217-3), o qual foi essencial para o desenvolvimento desta tese, e também ao CNPq e CAPES, pelo financiamento de diversos projetos e bolsas do laboratório. Agradeço ainda, por fim, à UNICAMP e seus professores, a quem devo minha formação acadêmica e científica.

Índice

Índice	pág. x
Lista de Tabelas e Figuras	pág. xi
Lista de Abreviações Utilizadas na Tese.....	pág. xvi
Estrutura e Apresentação da Tese	pág. xvii
 Resumo	 pág. 1
Abstract	pág. 3
Introdução Geral	pág. 5

Capítulo I

Resumo	pág. 15
Introdução	pág. 17
Materiais e Métodos	pág. 20
Resultados	pág. 33
Discussão	pág. 53
Conclusões	pág. 60
Referências	pág. 61
Anexo A	pág. 69
Anexo B	pág. 79

Capítulo II

Resumo	pág. 85
Introdução	pág. 87
Materiais e Métodos	pág. 91
Resultados	pág. 99
Discussão	pág. 109
Conclusões	pág. 114
Referências	pág. 115
Anexo A	pág. 119

Capítulo III

Resumo	pág. 149
Introdução	pág. 151
Materiais e Métodos	pág. 153
Resultados	pág. 168
Discussão	pág. 193
Conclusões	pág. 204
Referências	pág. 206
 Conclusões Gerais	 pág. 211
Referências	pág. 215

Lista de Tabelas e Figuras

Tabelas

Capítulo I - Tabela 1. Lista das espécies de Calliphoridae utilizadas nas análises filogenéticas - pág. 20

Capítulo I - Tabela 2. Lista dos espécimes de *C. hominivorax*, e suas respectivas regiões de origem, utilizados na avaliação da variabilidade intra-específica do ITS2 - pág. 24

Capítulo I - Tabela 3. Resumo de todas as combinações de modelos de substituição utilizados nas análises de inferência Bayesiana - pág. 30

Capítulo I - Tabela 4. Razão entre as taxas de transição e transversão (R); conteúdo nucleotídico (A+T%) e distância p não-corrigida (média) de cada domínio da estrutura do ITS2 para todas as espécies de Calliphoridae analisadas - pág. 34

Capítulo I - Tabela 5. Resultados dos testes de saturação de substituição - pág. 36

Capítulo I - Tabela 6. Resumo das CBCs e hemi-CBCs encontradas nos alinhamentos de pares de espécies congêneras em Calliphoridae - pág. 44

Capítulo I - Tabela 7. Resumo dos resultados encontrados na comparação entre as árvores inferidas usando a região ITS2 e as relações mais recorrentemente encontradas em publicações prévias - pág. 50

Capítulo I - Tabela 8. Resultado das análises de comparação de modelos baseadas em máxima-verossimilhança - pág. 51

Capítulo I - Tabela 9. Resultados das comparações por "Bayes factor" para as análises realizadas - pág. 52

Capítulo I - Tabela A1. Número total de pares de bases encontrados nas estruturas secundárias do ITS2 para cada uma das espécies analisadas - pág. 70

Capítulo I - Tabela A2. Número total de pares de bases encontrados nas áreas de interface entre regiões em simples e dupla-fita (junções, "bojos", alças internas e terminais). Apenas um par de bases na vizinhança dessas regiões foi considerado - pág. 71

Capítulo I - Tabela A3. Número total de pares de bases encontrados nas áreas de interface entre regiões em simples e dupla-fita (junções, "bojos", alças internas e terminais). Dois pares de bases na vizinhança dessas regiões foi considerado - pág. 73

Capítulo I - Tabela A4. Número total de pares de bases encontrados nas áreas de interface entre regiões em simples e dupla-fita (apenas "bojos" e alças internas). Apenas um par de bases na vizinhança dessas regiões foi considerado - pág. 75

Capítulo I - Tabela A5. Número total de pares de bases encontrados nas áreas de interface entre regiões em simples e dupla-fita (apenas "bojos" e alças internas). Dois pares de bases na vizinhança dessas regiões foi considerado - pág. 77

Capítulo II - Tabela 1. Tamanhos esperados (com base no sequenciamento do DNA molde clonado), observados e considerados no mapeamento dos fragmentos gerados a partir da digestão dos RNAs transcritos da região do ITS2 nas espécies *C. hominivorax*, *M. domestica* e *G. morsitans* - pág. 104

Capítulo III - Tabela 1. Lista das espécies utilizadas nas análises de filogenia molecular - pág. 155

Capítulo III - Tabela 2. Resumo de todas as estratégias de particionamento empregadas nas análises de inferência Bayesiana - pág. 165

Capítulo III - Tabela 3. Tamanho total do fragmento amplificado via PCR e sequenciado (em pares de bases) e conteúdo nucleotídico (A+T %) de cada uma das quatro regiões gênicas utilizadas nas análises filogenéticas - pág. 169

Capítulo III - Tabela 4. Resultados dos testes de sinal filogenético - pág. 176

Capítulo III - Tabela 5. Resultado das análises de comparação de modelos baseadas em máxima-verossimilhança - pág. 184

Capítulo III - Tabela 6. Resultado das comparações de modelo usando métodos Bayesianos - pág. 185

Capítulo III - Tabela 7. Suporte médio (média aritmética), comprimento total da árvore (TL) e alterações topológicas (distâncias simétricas) para as diferentes estratégias de particionamento utilizadas nas análises Bayesianas - pág. 186

Capítulo III - Tabela 8. Valores de suporte (probabilidade *a posteriori*) dos clados que sofreram alterações em suporte ou relações inferidas entre as diferentes estratégias de particionamento utilizadas nas análises Bayesianas - pág. 189

Figuras

Introdução - Figura 1. Árvore segundo McAlpine (1989) mostrando as superfamílias e famílias de Calyptratae e as possíveis relações entre elas - pág. 7

Capítulo I - Figura 1. Localidades amostradas de populações de *C. hominivorax*, as quais foram utilizadas na avaliação da variabilidade intra-específica do ITS2 - pág. 25

Capítulo I - Figura 2. "Boxplot" mostrando a variação na magnitude das distâncias genéticas (distância p não corrigida) para diferentes níveis taxonômicos organizados hierarquicamente - pág. 35

Capítulo I - Figura 3. Algumas das estruturas secundárias preditas para a região do ITS2 em Calliphoridae - pág. 38

Capítulo I - Figura 4. Gráfico mostrando a correlação entre o conteúdo nucleotídico (A+T%) e a distância p (média geral) para cada domínio individual da estrutura secundária do ITS2 em Calliphoridae - pág. 41

Capítulo I - Figura 5. Algumas das árvores filogenéticas inferidas usando a região do ITS2 em Calliphoridae - pág. 46

Capítulo I - Figura B1. Algumas das árvores inferidas usando a região do ITS2 em Calliphoridae - pág. 80

Capítulo II - Figura 1. Representação esquemática da estrutura e organização do "cluster" de DNA ribossomal em eucariotos - pág. 87

Capítulo II - Figura 2. Modelos de estrutura secundária propostos para o ITS2 de *Saccharomyces cerevisiae* - pág. 89

Capítulo II - Figura 3. Gel de agarose mostrando a clonagem da região amplificada do ITS2 em *C. hominivorax* no vetor pHST7.0 - pág. 100

Capítulo II - Figura 4. Gel de agarose mostrando a clonagem da região amplificada do ITS2 de *M. domestica* e *G. morsitans* no vetor pHST7.0 - pág. 100

Capítulo II - Figura 5. Gel de agarose mostrando a reação de transcrição da região completa do ITS2 ($ChITS2_{I,II,III,IV}$) - pág. 101

Capítulo II - Figura 6. Gel de agarose mostrando a reação de transcrição dos domínios isolados do ITS2 de *C. hominivorax* ($ChITS2_{III,IV}$ e $ChITS2_{I,II}$) - pág. 102

Capítulo II - Figura 7. Gel de agarose mostrando a reação de transcrição da região do ITS2 em *M. domestica* ($MdITS2_{I,II,III,IV}$) e *G. morsitans* ($GmITS2_{I,II,III,IV}$) - pág. 103

Capítulo II - Figura A1. Cromatogramas referentes à região transcrita do *ChITS2*_{I,II,III,IV} RNA não submetida à digestão com RNAses (reações de controle negativo) - pág. 121

Capítulo II - Figura A2. Cromatogramas referentes à região transcrita do ITS2 (*ChITS2*_{III,IV}; *ChITS2*_{I,II}; *MdITS2*_{I,II,III,IV}; e *GmITS2*_{I,II,III,IV}) não submetida à digestão com RNAses (reações de controle negativo) - pág. 122

Capítulo II - Figura A3. Cromatogramas obtidos a partir da digestão enzimática do RNA transcrito *ChITS2*_{I,II,III,IV} com diferentes enzimas e seu provável mapeamento na estrutura secundária predita *in silico* pelo programa mfold - pág. 123

Capítulo II - Figura A4. Cromatogramas obtidos a partir da digestão enzimática do RNA transcrito *ChITS2*_{III,IV} com diferentes enzimas e seu provável mapeamento na estrutura secundária predita *in silico* pelo programa mfold - pág. 127

Capítulo II - Figura A5. Cromatogramas obtidos a partir da digestão enzimática do RNA transcrito *ChITS2*_{I,II} com diferentes enzimas e seu provável mapeamento na estrutura secundária predita *in silico* pelo programa mfold - pág. 131

Capítulo II - Figura A6. Cromatogramas obtidos a partir da digestão enzimática do RNA transcrito *MdITS2*_{I,II,III,IV} com diferentes enzimas e seu provável mapeamento na estrutura secundária predita *in silico* pelo programa mfold - pág. 135

Capítulo II - Figura A7. Cromatogramas obtidos a partir da digestão enzimática do RNA transcrito *GmITS2*_{I,II,III,IV} com diferentes enzimas e seu provável mapeamento na estrutura secundária predita *in silico* pelo programa mfold - pág. 139

Capítulo II - Figura A8. Possível interpretação do padrão de digestão enzimática mostrado nas Figuras A3 a A5 para o RNA *ChITS2*_{I,II,III,IV} - pág. 143

Capítulo II - Figura A9. Possível interpretação do padrão de digestão enzimática mostrado na Figura A6 para o RNA *MdITS2*_{I,II,III,IV} - pág. 144

Capítulo II - Figura A10. Possível interpretação do padrão de digestão enzimática mostrado na Figura A7 para o RNA *GmITS2*_{I,II,III,IV} - pág. 145

Capítulo III - Figura 1. Estrutura secundária predita para os domínios IV e V da região 16S de *Calypttratae* - pág. 171

Capítulo III - Figura 2. Estruturas secundárias preditas para a região do ITS2 em *Calypttratae* - pág. 172

Capítulo III - Figura 3. Árvore consenso estrito das 17 árvores igualmente parcimoniosas inferidas pelo programa TnT v1.1 - pág. 178

Capítulo III - Figura 4. Árvore estimada pelo método de Máxima-verossimilhança usando os programas Garli v1.0 e PhyML 3.0 - pág. 179

Capítulo III - Figura 5. Árvore inferida pelo método Bayesiano usando o programa mcmcphase do pacote PHASE e a combinação de partições e modelos S7A/C - pág. 180

Capítulo III - Figura 6. Diferentes hipóteses para as relações interfamiliares em Oestroidea inferidas pelas análises Bayesianas - pág. 182

Capítulo III - Figura 7. Resultado das análises de mapeamento de verossimilhança - pág. 192

Lista de Abreviações Utilizadas na Tese

AIC = Akaike Information Criteria;

AICc = Akaike Information Criteria (corrigido);

BI = Bayesian Inference (Inferência Bayesiana);

BIC = Bayesian Information Criteria;

BS = Bootstrap;

CBC = Compensatory Base Change (Mudança Compensatória de Base);

ITS2 = Internal Transcribed Spacer 2 (Segundo Espaçador Transcrito Interno);

JK = Jackknife;

MB = MrBayes;

MCMC = Markov Chain - Monte Carlo;

ML = Maximum-Likelihood (Máxima-Verossimilhança);

MP = Máxima-Parcimônia;

NJ = Neighbor-Joining;

PP = Posterior Probabilities (Probabilidades *a Posteriori*);

rDNA / rRNA = DNA ribossomal / RNA ribossomal;

s.s. / d.s. = "single-strand" / "double-strand" (conformações estruturais em simples e dupla-fita, respectivamente);

SH = Shimodaira/Hasegawa-like teste para suporte de ramos;

Estratégias de particionamento dos dados (Capítulo III):

NS/NC = "No Structure / No Codon" - Estratégia que não considera informações estruturais e de posição no códon nas análises;

NS/C = "No Structure / Codon" - Estratégia que considera informações de posição no códon mas não estruturais nas análises;

S/NC = "Structure / No Codon" - Estratégia que considera informações estruturais mas não de posição no códon nas análises;

S/C = "Structure / Codon" - Estratégia que considera tanto informações estruturais quanto de posição no códon nas análises.

Estrutura e apresentação da Tese

Os resultados dos experimentos e análises descritos e apresentados nessa tese representam abordagens em diferentes níveis de um mesmo problema: a inferência de relações filogenéticas entre espécies da superfamília Oestroidea do clado Calyptratae, com ênfase na família Calliphoridae, e com uma detalhada avaliação dos métodos utilizados no processo.

Nesse sentido, as diferentes frentes conduzidas ao longo do doutorado foram organizadas e separadas em capítulos, os quais representam possíveis ou já existentes artigos científicos resultantes dessa tese. Cada capítulo lida com um aspecto específico do problema, incluindo: a caracterização e avaliação de um marcador molecular, a região do ITS2, para inferência de filogenias em Calliphoridae, usando informações tanto de sequência primária quanto de estrutura secundária (Capítulo I - artigo publicado no periódico *Genetica*: Volume 139(9): 1189-1207 [2011]); a caracterização estrutural *in vitro* da estrutura secundária do ITS2 em Calyptratae, incluindo sua comparação com modelos preditos *in silico* e uma discussão acerca da precisão dos modelos de estrutura normalmente incorporados em análises genético-evolutivas (Capítulo II - artigo ainda em fase de idealização); e uma análise de filogenia molecular em Oestroidea, utilizando uma abordagem multigênica e com ênfase nos efeitos de diferentes estratégias de particionamento dos dados nas filogenias inferidas (Capítulo III - artigo em fase final de revisão para submissão).

Os três capítulos acima descritos são delimitados, na tese, por uma introdução geral no início, a qual visa fornecer um contexto geral para o problema em estudo e também para

as técnicas empregadas e análises conduzidas, e por um resumo das principais conclusões gerais do trabalho, no fim.

Resumo

A superfamília Oestroidea (Diptera:Brachycera:Calyptratae), com +13.000 espécies descritas, compreende um dos grupos mais numerosos e ecologicamente diversos da ordem Diptera. O grupo possui grande interesse para atividades humanas por englobar espécies de importância médica, veterinária e forense, muitas das quais compõem a família Calliphoridae. Apesar do grande número de estudos disponíveis, as relações evolutivas no grupo, o qual é composto predominantemente por linhagens de rápida diversificação e radiação, ainda são controversas e pouco compreendidas, encorajando a caracterização de novos marcadores moleculares para análises de filogenia molecular. Neste contexto, esta tese foi desenvolvida e organizada em três capítulos descrevendo estudos genético-evolutivos em espécies da superfamília Oestroidea, com ênfase em Calliphoridae. O primeiro capítulo trata da caracterização e avaliação do segundo espaçador transcrito interno (ITS2) do DNA ribossomal como um marcador molecular para análises filogenéticas em Calliphoridae, incorporando informações tanto da sequência primária quanto da estrutura secundária adquirida pela região. A análise do ITS2 revelou um padrão hierarquicamente organizado das distâncias genéticas nos níveis de espécies, gêneros e subfamílias, enquanto pouca variação intra-específica foi encontrada. As árvores inferidas recuperaram muitas das relações comumente aceitas entre os táxons amostrados, sendo que a inclusão da informação estrutural nas análises resultou na recuperação de topologias mais confiáveis. Sendo assim, o potencial da região ITS2 como um marcador molecular para análises evolutivas na família Calliphoridae foi confirmado e seu uso em análises de maior escala, incluindo marcadores de diferentes naturezas de evolução, encorajado. O segundo capítulo da tese descreve a caracterização *in vitro* da estrutura secundária adquirida pelo

ITS2, através de padrões de digestão enzimática e análise dos fragmentos gerados, em espécies representantes das três superfamílias de Calyptratae: *Glossina morsitans*, *Musca domestica* e *Cochliomyia hominivorax*. A análise do padrão de fragmentos gerados pelas enzimas RNase I, A, T1 e V1, quando mapeados na estrutura secundária predita *in silico*, corroborou muitos dos domínios inicialmente preditos pelo método computacional, ressaltando a importância e confiabilidade desses métodos na predição de estruturas secundárias. O terceiro capítulo da tese descreve análises de filogenia molecular na superfamília Oestroidea, com ênfase na amostragem de espécies de Calliphoridae, utilizando quatro marcadores moleculares, dois nucleares (ITS2 e 28S) e dois mitocondriais (COI e 16). As análises, que incluíram uma extensa avaliação dos efeitos de diferentes estratégias de particionamento dos dados em análises de inferência Bayesiana (por conformação estrutural e posição no códon), revelaram a existência de dois clados principais em Oestroidea: Tachinidae + Mesembrinellinae e Oestridae + Rhiniinae + Sarcophagidae + Calliphoridae (definida em senso estrito). O status de família recentemente atribuído à Rhiniinae foi encontrado, enquanto há também evidências para sugerir o mesmo para a subfamília Mesembrinellinae, como proposto anteriormente por outros autores. As diferentes estratégias de particionamento do conjunto de dados amostrados resultaram em diferenças discretas em termos de topologia, comprimentos de ramo e suporte geral das filogenias inferidas. Embora o resultado geral indique uma melhor resolução das análises quando do uso de combinações de partições e modelos mais complexas, as mesmas podem ocasionar também um aumento considerável na incerteza associada às análises.

Palavras-chave: Oestroidea; Mosca-varejeira (Calliphoridae); Filogenia - Aspectos moleculares; DNA espaçador ribossômico; Estrutura secundária de RNA.

Abstract

The Oestroidea superfamily (Diptera: Brachycera: Calyptratae), with +13,000 described species, comprises one of the most numerous and ecologically diverse groups in the Diptera order. The group is actually of great interest for human activities since it includes species of medical, veterinary and forensic importance, most of them included in the Calliphoridae family. Despite the existence of several studies addressing the issue, evolutionary relationships in Oestroidea, a group mainly composed of rapidly diverged lineages, remains contentious and poorly understood, encouraging the characterization of new molecular markers for phylogenetic inference analyses. In this context, this thesis was developed and organized in three chapters describing genetic and evolutionary studies in species of the Oestroidea superfamily, with emphasis in Calliphoridae. The first chapter deals with the characterization and evaluation of the second internal transcribed spacer region (ITS2) of the ribosomal DNA cluster as a molecular marker for phylogenetic inference in Calliphoridae, including information of both primary sequence and secondary structure. The analyses revealed an hierarchically organized pattern of genetic distances in the specific, generic and subfamilial level, while little intraspecific variation was detected. Inferred trees were able to recover most of the commonly accepted relationships among the sampled taxa, with the consideration of structural information resulting in better supported topologies. Thereby, the potential of the ITS2 region as a molecular marker for phylogenetic inference in the Calliphoridae family was corroborated and its use in larger scale analyses, including other markers with different evolutionary patterns, encouraged. Chapter II describes the *in vitro* characterization of the secondary structure of the ITS2 region, through patterns of enzymatic digestion and analysis of the generated fragments, in representative species of the three superfamilies of the Calyptratae clade: *Glossina morsitans*, *Musca*

domestica and *Cochliomyia hominivorax*. Analyses of the patterns of the fragments generated by enzymatic digestions with the RNAses I, A, T1 and V1, when mapped in the *in silico* predicted secondary structure, corroborated the folding of most of the domains predicted by computational methods, highlighting the importance and reliability of these methods in secondary structure prediction. Chapter III describes molecular phylogenetic analyses in the Oestroidea superfamily, with emphasis on the Calliphoridae family, using four different molecular markers, two nuclear (ITS2 and 28S) and two mitochondrial (COI and 16S) regions. The analyses, which included a comprehensive evaluation of the effects of different data partitioning strategies in a Bayesian framework (by structural conformation and codon position), revealed the existence of two main clades in Oestroidea: Tachinidae+Mesembrinellinae and Oestridae+Rhiniinae+Sarcophagidae+Calliphoridae (defined in a strict sense). The recently attributed family status to Rhiniinae was confirmed, and there are evidence to also suggest the same for Mesembrinellinae, as previously pointed out by other studies. The different data partitioning strategies used in the sampled dataset resulted in small differences in terms of inferred topologies, estimated branch lengths and average support. Although the overall results indicate a significant increase in phylogeny resolution when more complex and parameter-rich models / partitions combinations are used, they can also lead to an increased uncertainty in the phylogenetic estimation process.

Key words: Oestroidea; Blowflies; Phylogeny - Molecular aspects; Ribosomal spacer DNA; RNA secondary structure.

Introdução Geral

1. A superfamília Oestroidea e a família Calliphoridae (Diptera: Schizophora: Calyptratae)

Com aproximadamente 150.000 espécies descritas, a ordem Diptera é um dos grupos mais numerosos e provavelmente mais ecologicamente diversos entre os insetos (Grimaldi & Engel 2005; Wiegmann et al. 2011).

Entre os grandes grupos de Diptera, o clado Schizophora possui interesse especial, não apenas por englobar espécies de renomada importância como modelos para estudos gerais em genética e biologia molecular (*Drosophila melanogaster* - Johnston 2002) e em áreas mais específicas da biologia (e.g., a utilização de espécies do gênero *Phormia* em estudos de fisiologia sensorial - Dethier 1976), mas também por apresentar inúmeras espécies associadas a atividades humanas, especialmente nos campos da medicina, veterinária e entomologia forense (Zumt 1965; Guimarães et al. 1983; Hall & Wall 1995; Amendt et al. 2004).

Em termos evolutivos, apesar de haver evidências apontando para a existência de uma diversidade de dípteros considerável no final do período Cretáceo da era Mesozóica (145 a 65 milhões de anos atrás - m.a.a.) (Amorim & Silva 2002; Grimaldi & Engel 2005), com muitos dos grandes grupos de Diptera já possuindo representantes nesse período, acredita-se que foi apenas no início da era Cenozóica (Paleoceno/Eoceno - 65 a 35 m.a.a.), após o período de extinções em massa ocorrido no limite K-T (Cretáceo-Terciário) e paralelamente à proliferação dos mamíferos placentários (Eutheria), que o clado Schizophora alcançou muito de sua diversidade atualmente conhecida, com muitos dos grandes grupos do clado tendo se originado entre 65 e 40 m.a.a. (Wiegmann et al. 2003;

Pape 2006; Kutty et al. 2010) após um rápido processo de divergência de linhagens (Wiegmann et al. 2011). Esse processo de radiação da diversidade na era Cenozóica possui paralelos em muitas outras ordens de insetos (e.g. insetos sociais das ordens Isoptera e Hymenoptera, ectoparasitas das ordens Siphonaptera e Phthiraptera e alguns grupos fitófagos de Lepidoptera e Coleoptera) e provavelmente possui ligações com a diversificação das plantas terrestres e mamíferos (Grimaldi & Engel 2005).

Sistematicamente, o clado Schizophora, o qual compreende quase um terço da diversidade descrita para a ordem Diptera (>50.000 spp. distribuídas em 85 famílias, de um total de 157 - Wiegmann et al. 2011) é comumente dividido em dois grandes grupos: (1) Acalypterae, o qual contém a maioria das famílias, gêneros e espécies descritas para o grupo e que compõe provavelmente um grupo parafilético (Griffiths 1972; Yates & Wiegmann 1999; Wiegmann et al. 2011), embora alguns autores defendam seu possível status monofilético (McAlpine 1989); e (2) Calypterae, o qual contém a maioria das espécies de importância para as atividades humanas e cuja monofilia é bastante aceita e corroborada (Griffiths 1972; Hennig 1973; McAlpine 1989).

O grupo Calypterae, o qual contém algumas espécies de suma importância para as áreas médica e sanitária (e.g., vetores e carreadores de organismos patológicos, como as espécies do gênero *Glossina* - Hippoboscoidea - e a espécie *Musca domestica* - Muscoidea), veterinária (e.g., espécies hematófagas como a mosca-dos-estábulo *Stomoxys calcitrans* e a mosca-dos-chifres *Haematobia irritans* da superfamília Muscoidea e diversas espécies causadoras de miíases das famílias Calliphoridae - especialmente das subfamílias Luciliinae e Chrysomyinae - e Oestridae - a mosca-do-berne *Dermatobia hominis* - da superfamília Oestroidea) e forense (e.g., espécies das famílias Sarcophagidae e Calliphoridae - sobretudo das subfamílias Calliphorinae, Luciliinae e Chrysomyinae) é

comumente dividido em três superfamílias (Figura 1): (1) Hippoboscoidea (4 famílias: Glossinidae, Hippoboscidae; Streblidae e Nycteribiidae); (2) Muscoidea (4 famílias: Muscidae, Fanniidae, Anthomyiidae e Scathophagidae); e (3) Oestroidea (6 famílias: Mystacinobiidae, Oestridae, Rhinophoridae, Tachinidae, Sarcophagidae e Calliphoridae) (McAlpine 1989).

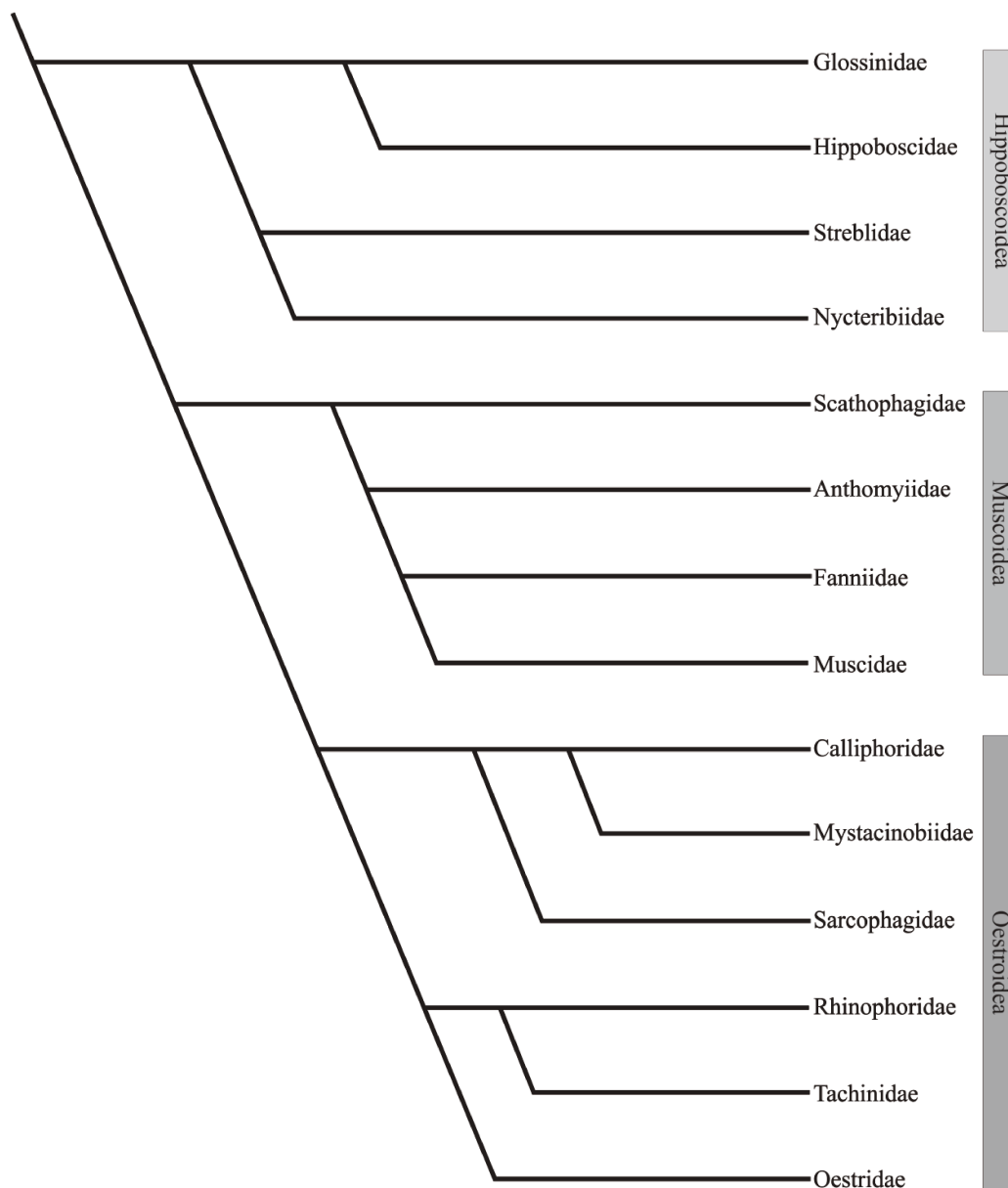


Figura 1. Árvore segundo McAlpine (1989) mostrando as super-famílias e famílias de Calypttratae e as possíveis relações entre elas.

Filogeneticamente, a superfamília Oestroidea, provavelmente um dos grupos mais diverso em termos de estratégias de vida na divisão Calyptratae, ainda possui muitas relações de parentesco pouco compreendidas e resolvidas, com a persistência de algumas hipóteses controversas para as relações inter-familiares no grupo (Lehrer 1970; Hennig 1973; Griffiths 1982; McAlpine 1989; Rognes 1991; Pape 1992; Rognes 1997; Pape & Arnaud 2001; Kutty et al. 2010). A falta de resolução no entendimento das relações filogenéticas no grupo possivelmente se correlaciona ao relativo rápido processo de divergência das linhagens supra-familiares (e de alguns de seus grupos subordinados) ocorrido no início do Cenozóico, o qual pode ter deixado poucas, e muitas vezes controversas, evidências.

Neste contexto, o uso de marcadores moleculares e de análises de inferência de filogenia molecular podem fornecer informações complementares aos estudos tradicionalmente conduzidos com dados morfológicos e comportamentais e, assim, tentar elucidar com maior confiabilidade as relações evolutivas no grupo.

2. Marcadores moleculares e análises de filogenia molecular

O estudo da sistemática de insetos (e dos organismos vivos em geral) presenciou uma grande revolução nas últimas três décadas com a popularização e disseminação do uso de dados e métodos de análise molecular no estudo de relações evolutivas, os quais vieram complementar e adicionar novas informações aos estudos então realizados com dados morfológicos (Caterino et al. 2000, Ronquist & Deans 2010). Apesar do grande número de estudos realizados até o momento, durante muito tempo houve uma predominância do uso de marcadores mitocondriais, em especial das regiões ribossomais 12S e 16S e dos genes do citocromo b (*cytb*) e das subunidades I e II da citocromo oxidase (*coxI* e *coxII*), e do "cluster" nuclear de DNA ribossomal, principalmente as regiões 18S, 28S e os espaçadores

internos, nas análises de reconstrução e inferência de filogenias moleculares (Caterino et al. 2000). Esse viés se deve, principalmente, a maior facilidade técnica na obtenção de sequências dessas regiões, em especial pelo fato de ambas estarem presentes em grande número de cópias por genoma (provavelmente homogeneizadas ou pela evolução em concerto - rDNA: Hillis & Dixon 1991; Elder & Turner 1995; Liao 1999 - ou pela inexistência de recombinação - genoma mitocondrial: Boore 1990; Wolstenholme 1992; Lunt & Hyman 1997).

Apesar da utilidade comprovada das diferentes regiões desses dois conjuntos de marcadores ("cluster" ribossomal nuclear e genoma mitocondrial) em resolver relações em níveis taxonômicos hierarquicamente distintos (Hillis & Dixon 1991), muitas relações evolutivas permaneceram controversas ou não resolvidas, encorajando a caracterização e avaliação de novas fontes de informação. A baixa resolução encontrada nas análises realizadas com o uso de um conjunto limitado de marcadores moleculares é agravada em grupo nos quais a grande diversidade observada nos organismos vivos é provavelmente fruto de um rápido processo de diversificação de linhagens (radiação). Nestes casos, eventos como a seleção incompleta de linhagens (ancestrais ou recentes) e a possível fixação de homoplasias ancestrais podem resultar na inexistência ou existência contraditória de evidências sobre o processo real ocorrido (Pamilo & Nei 1988; Maddison 1997; Wiens 1998a; Takahashi et al. 2001; Degnan & Rosenberg 2006; Maddison & Knowles 2006; Pollard et al. 2006; Koblmüller et al. 2010).

Neste contexto, diferentes regiões gênicas codificadoras de proteína do genoma nuclear têm sido descritas e utilizadas em análises de filogenia molecular (Moulton & Wiegmann 2004; Wahlberg & Wheat 2008). Contudo, a existência de uma única cópia dessas regiões no genoma nuclear torna muitas vezes difícil a amplificação via PCR e

sequenciamento desses marcadores em espécimes fixados por longos períodos de tempo ou preservados com métodos que não garantam a integridade da molécula de DNA. Em análises de filogenia molecular que consideram uma abordagem multigênica para a inferência das relações evolutivas, uma forma de contornar a não disponibilidade de alguns marcadores para a maioria dos táxons é enriquecer a matriz de dados de caracteres para tantos táxons quanto possíveis, enquanto para os demais táxons apenas alguns caracteres são amostrados, com os demais sendo codificados como "dados faltantes" ("missing data" - Wiens 1998b; Burleigh et al. 2009; Cho et al. 2011). Apesar de resultar em um aumento da resolução das relações inferidas em alguns clados, essa abordagem é ainda controversa, assim como também o é a discussão sobre o efeito relativo da amostragem de táxons ou caracteres em análises de inferência filogenética (Graybeal 1998; Hillis 1998; Poe 1998a,b; Bremer et al. 1999; Zwickl & Hillis 2002; Hillis et al. 2003; Poe 2003; Rokas et al. 2003; Sorenson et al. 2003; Rokas & Carroll 2005; Hedtke et al. 2006; Gatesy et al. 2007; Heath et al. 2008).

Em análises de filogenia molecular com uma disponibilidade limitante de caracteres moleculares, uma alternativa para aumentar a resolução sem necessariamente preterir a amostragem taxonômica é a maior exploração da informação contida nos marcadores efetivamente disponíveis e em geral de mais fácil obtenção. No caso de regiões do genoma em que, após a transcrição, as moléculas de RNA resultantes adquirem uma estrutura secundária, determinada primariamente pela informação contida na sequência primária, a consideração da informação estrutural nas análises de inferência filogenética leva, em geral, a um aumento da resolução nas relações inferidas (Telford et al. 2005; Wiemers et al. 2009; Letsch & Kjer 2011). Em sua maioria, esse aumento na resolução e confiabilidade das árvores estimadas se deve a consideração da não-independência na evolução de sítios na

sequência primária que se encontram pareados na estrutura secundária, representando assim uma modelagem mais realística da evolução das regiões estruturalmente funcionais em análise (Dixon & Hillis 1993; Schöniger & von Haeseler 1994; Muse 1995; Rzhetsky 1995; Tillier & Collins 1995; Gutell 1996; Savill et al. 2001; Yu & Thorne, 2006).

Além disso, a análise das estruturas secundárias adquiridas por essas moléculas fornece informações adicionais para o processo de inferência filogenética, incluindo um guia para o alinhamento de regiões mais complexas da sequência primária (Gillespie 2004), levando a um sensível aumento na confiabilidade do estabelecimento da homologia posicional nessas regiões, e de fornecer caracteres adicionais para as análises, como as mudanças compensatórias de base (do inglês "compensatory base changes" - CBCs) (Coleman & Vacquier 2002; Wolf et al. 2005; Müller et al. 2007; Coleman 2009).

Os sucessivos avanços no desenvolvimento de modelos e métodos de reconstrução filogenética que incorporam esse tipo de informação nas análises de inferência evolutiva (Savill et al. 2001; Holder & Lewis 2003; Ronquist & Huelsenbeck 2003; Gowri-Shankar & Jow 2006), e nos métodos de predição dessa estruturas secundárias *in silico* (Hofacker 2003; Zucker 2003; Gardner & Giegerich 2004; Reeder et al. 2006), estimulam a avaliação e posterior utilização da informação estrutural nas análises filogenéticas.

Neste sentido, as análises e resultados descritos nesta tese tiveram como objetivos principais: (1) o estudo das relações filogenéticas entre espécies da superfamília Oestroidea, com ênfase na família Calliphoridae, através do uso de marcadores moleculares; e (2) a avaliação dos métodos de predição e modelagem de estruturas secundárias de moléculas de RNA e os efeitos da inclusão de informações estruturais em análises de filogenia molecular. Para este fim, três abordagens em diferentes âmbitos foram utilizadas, incluindo: (A) a caracterização e avaliação da região do segundo espaçador transcrito interno (ITS2)

do "cluster" de DNA ribossomal nuclear (rDNA) para análises de inferência filogenética em um grupo reduzido de espécies de importância veterinária e forense da família Calliphoridae (Diptera: Calyptratae: Oestroidea) utilizando informações tanto da sequência primária quanto da estrutura secundária (Capítulo I); (B) avaliação e validação do modelo predito *in silico* para a estrutura secundária da região do ITS2 para espécies representativas das três superfamílias de Calyptratae (Hippoboscoidea, Muscoidea e Oestroidea) através de sua comparação com análises realizadas *in vitro* de digestão enzimática e análise dos fragmentos gerados (Capítulo II); e (C) uma análise filogenética utilizando diferentes marcadores moleculares (nucleares: ITS2 e 28S rDNA; mitocondriais: COI e 16S rDNA) para a superfamília Oestroidea (Diptera: Schizophora: Calyptratae), com ênfase em Calliphoridae, e uma avaliação dos efeitos gerados nas topologias resultantes quando da utilização de diferentes estratégias de particionamento dos dados, refletindo tanto a incorporação de informações estruturais quanto de posição no códon, nas análises (Capítulo III).

Capítulo I

Avaliação do segundo espaçador transcrito interno (ITS2) do DNA ribossomal (rDNA) como um marcador molecular para inferência filogenética em Calliphoridae (Diptera: Calyptratae: Oestroidea), utilizando informações de sequência e estrutura secundária.

As análises e resultados descritos neste capítulo encontram-se publicados no periódico

Genetica (2011) Volume 139(9): 1189-1207.

Resumo

O segundo espaçador transcrito interno (ITS2) é uma pequena região não-codificadora localizada no "cluster" nuclear de DNA ribossomal (rDNA). Acredita-se que a variação presente nesta região seja apropriada para estudos de identificação molecular de espécies e também para análises de reconstrução filogenética, as quais podem ser aperfeiçoadas com a consideração de informações estruturais. Neste sentido, o potencial uso do ITS2 como um marcador molecular para análises filogenéticas em Calliphoridae (Diptera: Calyptratae: Oestroidea) foi avaliado utilizando-se diferentes combinações de métodos de inferência e modelos de substituição, incluindo ou não informações estruturais. A análise das sequências revelou um padrão de organização hierárquico da variabilidade e um nível não significativo de saturação nas taxas de substituição. Variação intragenômica foi detectada principalmente na forma de diferentes números de repetições na região poli-T do domínio IV, variação essa que não possui efeitos significativos no sinal filogenético no nível específico. As estruturas secundárias preditas para o ITS2 revelaram que pares GC são encontrados preferencialmente nas regiões flangeadoras de "bojos" e "alças", em regiões estruturalmente conservadas, possivelmente garantindo uma maior estabilidade. Nas análises filogenéticas, o uso de modelos de substituição que levam em consideração informações estruturais resultaram em melhoras significativas no processo de inferência usando ambos os métodos de Neighbor-Joining e análise Bayesiana, apesar do primeiro ter um uso limitado no caso de sequências muito divergentes. Nas análises Bayesianas, foi observado um aumento significativo da verossimilhança das análises com a consideração de informações estruturais, embora com poucas alterações em termos de topologia e suporte das árvores, provavelmente resultante de uma melhor estimativa das taxas de evolução. Baseado nestes resultados, acredita-se que o ITS2 seja um marcador molecular

Capítulo I

apropriado para análises de inferência filogenética em Calliphoridae, em ambos os níveis específico e genérico.

Palavras-chave: ITS2; segundo espaçador transcrito interno; estrutura secundária; Calliphoridae.

1. Introdução

O segundo espaçador transcrito interno (ITS2, do inglês "internal transcribed spacer 2") é uma pequena região não codificadora localizada no "cluster" nuclear de DNA ribossomal (rDNA) entre as regiões 3' da subunidade 5,8S e 5' da subunidade 28S. O ITS2 é transcrito como parte de uma grande molécula precursora de RNA, a qual contém as três subunidades ribossomais, 18S, 5,8S e 26/28S, dois espaçadores internos (ITS1 e 2) e dois espaçadores transcritos externos (ETS - do inglês "external transcribed spacer"), 5'-ETS e 3'-ETS (Veldman et al. 1981; Hillis & Dixon 1991). Embora não esteja presente na molécula de RNA ribossomal (rRNA) madura, sendo digerido durante o processamento do transcrito primário, o ITS2 direciona a clivagem do grande rRNA precursor e, assim, delimita as extremidades 3' da subunidade ribossomal 5,8S e 5' da subunidade ribossomal 28S (Mitchell et al. 1996; Mitchell et al. 1997; Geerlings et al. 2000; Côté et al. 2002).

Modelos desenvolvidos em leveduras sugerem que o reconhecimento e clivagem dos sítios presentes na região do ITS2 estejam relacionados com a estrutura secundária adquirida pela molécula durante a transcrição, possivelmente pela própria estrutura e/ou pela união de sequências distantes, as quais são reconhecidas pelos complexos enzimáticos que processam o ITS2 (van der Sande et al. 1992; Côté et al. 2002). Essa restrição funcional leva a uma grande conservação da região do ITS2 em termos de estrutura secundária, a qual é bastante similar em todos os eucariotos, de leveduras a mamíferos, e pode permitir a comparação do ITS2 entre táxons superiores, acima do nível de família (Joseph et al. 1999; Coleman 2003; Coleman 2007). O atual modelo aceito para a conformação secundária adquirida pelo ITS2 foi descrito inicialmente por Joseph et al. (1999), baseado em modelos computacionais em leveduras e vertebrados, e inclui quatro regiões hélice-alça ("helix-loop") inseridas em um "anel" central. Este modelo estrutural é o

atualmente utilizado na base de dados de estruturas secundárias do ITS2 (Schultz et al. 2006; Selig et al. 2008; Koetschan et al. 2010).

Muitas propriedades do ITS2 atestam para seu uso como um marcador molecular para inferências filogenéticas em uma grande variedade de grupos taxonômicos, incluindo (1) a conservação de sua estrutura secundária em diferentes táxons; (2) a correlação entre a variação em termos de sequência e estrutura secundária do ITS2 com incompatibilidade reprodutiva, uma característica que pode facilitar o processo de identificação de espécies (Coleman & Vacquier 2002; Müller et al. 2007; Coleman 2009); e (3) sua pequena extensão e alto número de cópias, teoricamente homogeneizadas pelo processo de evolução em concerto (Hillis & Dixon 1991; Elder & Turner 1995; Liao 1999), fato este que facilita a amplificação via PCR e sequenciamento da região. Além disso, a implementação de modelos de substituição nucleotídica para moléculas de RNA que consideram a interdependência entre sítios no processo evolutivo (Savill et al. 2001; Gowri-Shankar & Jow 2006) permite uma modelagem mais realística da evolução molecular que deve ocorrer nesta região, dessa forma aprimorando o processo de inferência filogenética. Neste capítulo, a filogenia de Calliphoridae (Diptera: Calyptratae: Oestroidea) foi usada como modelo para avaliar o uso potencial do ITS2 em inferir relações filogenéticas.

A família Calliphoridae, cujos membros são popularmente conhecidos como "moscas varejeiras", é um dos grupos mais importantes de moscas causadoras de miíases, compreendendo mais de 80 espécies descritas como agentes causadores de miíases, isto é, a infestação de humanos ou outros vertebrados por larvas de dípteros, as quais, por pelo menos um de seus períodos de vida, se alimentam de tecidos vivos ou mortos do hospedeiros, substâncias líquidas do seu corpo ou alimentos ingeridos por ele (Zumpt 1965). Parasitas facultativos dos gêneros *Calliphora*, *Chrysomya*, *Cochliomyia*, *Lucilia* e

Phormia estão entre os insetos mais comumente encontrados em carcaças e corpos em decomposição, e muitos deles tem fundamental importância na entomologia forense (Amendt et al. 2004). Além disso, parasitas obrigatórios e facultativos dos gêneros *Chrysomya*, *Cochliomyia* e *Lucilia* estão comumente associados a culturas pecuárias, sendo que elevados níveis de infestação nos animais resultam em um grande impacto econômico (Guimarães et al. 1983; de Azeredo-Espin & Lessinger 2006).

Tradicionalmente, com exceção da região de rDNA nuclear 28S, o DNA mitocondrial (mtDNA) tem sido o marcador molecular mais caracterizado e utilizado para estudos genético-evolutivos na família Calliphoridae, seja através da utilização de regiões gênicas individuais (revisadas por Stevens & Wallman 2006) ou de sequências de genomas completos (Lessinger et al. 2000; Lessinger et al. 2004; Junqueira et al. 2004). Recentemente, o ITS2 foi caracterizado como um marcador molecular para inferência filogenética em Calyptratae (Song et al. 2008a) e para identificação espécie-específica em alguns grupos de Calliphoridae de importância forense (Nelson et al. 2007; Nelson et al. 2008; Song et al. 2008b). Contudo, nenhum desses estudos considerou formalmente a incorporação de informações estruturais da região em suas análises, além de sua consideração durante o processo de alinhamento. A inclusão dessas informações justifica uma avaliação mais profunda do potencial do ITS2 como marcador molecular para análises de inferência filogenética em Calliphoridae usando diferentes combinações de métodos e modelos de substituição que consideram tanto informações da sequência primária quanto da estrutura secundária.

Os resultados descritos aqui podem contribuir para um maior entendimento da evolução do ITS2 em Diptera e, em uma perspectiva mais ampla, para sua avaliação como

um possível marcador molecular universal para identificação de espécies e inferência filogenética, especialmente nos níveis de espécies e gêneros.

2. Materiais e Métodos

2.1. Espécimes e extração de DNA

Para as análises filogenéticas, espécimes congelados das famílias Calliphoridae e Muscidae (Diptera: Calyptratae) foram obtidos da coleção mantida a -70°C do Laboratório de Genética e Evolução Animal (CBMEG / UNICAMP). Espécimes secos e preservados em etanol foram enviadas de museus e coleções entomológicas. A Tabela 1 mostra a lista completa de espécimes utilizadas nas análises filogenéticas.

Tabela 1. Lista das espécies utilizadas nas análises filogenéticas. O comprimento total da região ITS2 amplificada, seu conteúdo nucleotídico e o número de acesso no GenBank são mostrados.

Família	Número de acesso (GenBank)	Espécies	A+T (%)	G+C (%)	Total (pb)
Calliphoridae	EF061800	<i>Calliphora vomitoria</i> ^a	80,8	19,2	318
	EF061799	<i>Calliphora vomitoria</i> ^a	80,1	19,8	312
	AF498021	<i>Calliphora vomitoria</i> ^c	81,1	18,9	312
	AF498020	<i>Calliphora vomitoria</i> ^c	81,1	18,9	312
	AF498019	<i>Calliphora vomitoria</i> ^c	81,1	18,9	312
	AF498018	<i>Calliphora vomitoria</i> ^c	81,1	18,9	312
	EF560179	<i>Calliphora vomitoria</i> ^b	81,2	18,8	314
	EF061801	<i>Aldrichina grahami</i> ^{a,f}	80,4	19,6	337
	EF061802	<i>Aldrichina grahami</i> ^{a,f}	80,4	19,6	336
	EF061803	<i>Calliphora vicina</i> ^a	81,3	18,7	326
	EF560178	<i>Calliphora vicina</i> ^b	81,3	18,7	327
	AF498026	<i>Calliphora vicina</i> ^c	81,3	18,7	327
	AF498025	<i>Calliphora vicina</i> ^c	81,3	18,7	327
	AF498024	<i>Calliphora vicina</i> ^c	81,3	18,7	327
	AF498023	<i>Calliphora vicina</i> ^c	81,3	18,7	327
	AF498022	<i>Calliphora vicina</i> ^c	81,3	18,7	327

Calliphoridae	AF498034	<i>Lucilia caesar</i> ^c	81,8	18,2	312
	AF498033	<i>Lucilia caesar</i> ^c	81,8	18,2	312
	AF498032	<i>Lucilia caesar</i> ^c	81,8	18,2	312
	AF498031	<i>Lucilia caesar</i> ^c	81,8	18,2	312
	AF498038	<i>Lucilia illustris</i> ^c	82,1	17,9	314
	AF498037	<i>Lucilia illustris</i> ^c	82,1	17,9	314
	AF498036	<i>Lucilia illustris</i> ^c	82,1	17,9	314
	AF498035	<i>Lucilia illustris</i> ^c	82,1	17,9	314
	EF061804	<i>Hemipyrellia ligurriens</i> ^a	82,6	17,4	312
	EF560185	<i>Lucilia cuprina</i> ^b	81,6	18,4	336
	EF061805	<i>Lucilia cuprina</i> ^a	81,3	18,7	333
	EF061793	<i>Lucilia sericata</i> ^a	79,6	20,4	318
	EF560187	<i>Lucilia sericata</i> ^b	79,4	20,6	321
	EF061796	<i>Lucilia sericata</i> ^a	79,3	20,7	318
	EF061795	<i>Lucilia sericata</i> ^a	79,3	20,7	318
	EF061794	<i>Lucilia sericata</i> ^a	79,0	21,0	318
	AF498030	<i>Lucilia ampullacea</i> ^c	82,4	17,6	300
	AF498029	<i>Lucilia ampullacea</i> ^c	82,4	17,6	300
	AF498028	<i>Lucilia ampullacea</i> ^c	82,4	17,6	300
	AF498027	<i>Lucilia ampullacea</i> ^c	82,4	17,6	300
	EF061798	<i>Lucilia porphyryna</i> ^a	82,8	17,2	301
	EF061797	<i>Lucilia porphyryna</i> ^a	82,8	17,2	301
	EF061785	<i>Lucilia bazini</i> ^a	82,7	17,3	317
	EF061784	<i>Lucilia bazini</i> ^a	82,8	17,2	315
	EF560186	<i>Lucilia eximia</i> ^b	82,5	17,5	326
	EF061788	<i>Chrysomya megacephala</i> ^a	78,9	21,1	303
	EF061787	<i>Chrysomya megacephala</i> ^a	78,9	21,1	303
	EF061786	<i>Chrysomya megacephala</i> ^a	78,1	21,9	302
	EF071964	<i>Chrysomya megacephala</i> ^d	79,0	21,0	306
	EF071965	<i>Chrysomya megacephala</i> ^d	79,0	21,0	306
	EF071966	<i>Chrysomya megacephala</i> ^d	79,0	21,0	306
	EF560175	<i>Chrysomya megacephala</i> ^b	79,0	21,0	306
	DQ310488	<i>Chrysomya megacephala</i> ^c	79,0	21,0	306
	EF560174	<i>Chrysomya bezziana</i> ^b	79,2	20,8	302
	DQ310490	<i>Chrysomya saffrana</i> ^e	79,3	20,7	310
	EF071971	<i>Chrysomya saffrana</i> ^d	79,3	20,7	309
	EF071970	<i>Chrysomya saffrana</i> ^d	79,3	20,7	309
	EF071969	<i>Chrysomya saffrana</i> ^d	79,3	20,7	309
	EF071968	<i>Chrysomya saffrana</i> ^d	79,3	20,7	309

Calliphoridae	EF071967	<i>Chrysomya saffrana</i> ^d	79,3	20,7	309
	EF071972	<i>Chrysomya saffrana</i> ^d	79,2	20,8	308
	EF061792	<i>Chrysomya pinguis</i> ^a	79,8	20,2	303
	EF061791	<i>Chrysomya pinguis</i> ^a	79,9	20,1	304
	EF071963	<i>Chrysomya latifrons</i> ^d	78,5	21,5	316
	EF071962	<i>Chrysomya latifrons</i> ^d	78,5	21,5	316
	EF071961	<i>Chrysomya latifrons</i> ^d	78,5	21,5	316
	EF071960	<i>Chrysomya latifrons</i> ^d	78,5	21,5	316
	DQ310492	<i>Chrysomya latifrons</i> ^e	78,5	21,5	316
	DQ310493	<i>Chrysomya semimetallica</i> ^e	78,5	21,5	321
	EF071981	<i>Chrysomya semimetallica</i> ^d	78,5	21,5	321
	EF071980	<i>Chrysomya semimetallica</i> ^d	78,5	21,5	321
	EF071979	<i>Chrysomya semimetallica</i> ^d	78,5	21,5	321
	EF071978	<i>Chrysomya semimetallica</i> ^d	78,5	21,5	321
	EF071977	<i>Chrysomya semimetallica</i> ^d	78,5	21,5	321
	EF071976	<i>Chrysomya semimetallica</i> ^d	78,5	21,5	321
	EF071975	<i>Chrysomya semimetallica</i> ^d	78,5	21,5	321
	EF071974	<i>Chrysomya semimetallica</i> ^d	78,5	21,5	321
	EF071973	<i>Chrysomya semimetallica</i> ^d	78,5	21,5	321
	EF560173	<i>Chrysomya albiceps</i> ^b	81,7	18,3	317
	EF560172	<i>Chrysomya albiceps</i> ^b	81,7	18,3	317
	EF061790	<i>Chrysomya rufifacies</i> ^a	80,7	19,3	317
	EF061789	<i>Chrysomya rufifacies</i> ^a	80,7	19,3	317
	DQ310487	<i>Chrysomya rufifacies</i> ^e	80,7	19,3	321
	EF560177	<i>Chrysomya rufifacies</i> ^b	81,0	19,0	320
	DQ310486	<i>Chrysomya incisuralis</i> ^e	80,9	19,1	319
	DQ310489	<i>Chrysomya nigripes</i> ^e	78,5	21,5	325
	EF560176	<i>Chrysomya putoria</i> ^b	79,8	20,2	302
	DQ310497	<i>Chrysomya flavifrons</i> ^e	80,7	19,3	322
	DQ310494	<i>Chrysomya flavifrons</i> ^e	81,2	18,8	329
	DQ310496	<i>Chrysomya flavifrons</i> ^e	80,7	19,3	322
	DQ310495	<i>Chrysomya flavifrons</i> ^e	80,5	19,5	323
	DQ310491	<i>Chrysomya varipes</i> ^e	81,0	19,0	310
	EF560181	<i>Cochliomyia hominivorax</i> ^b	81,2	18,8	351
	EF560182	<i>Cochliomyia macellaria</i> ^b	81,1	18,9	344
	EF560192	<i>Hemilucilia segmentaria</i> ^b	81,2	17,8	327
	EF560180	<i>Chloroprocta idioidea</i> ^b	81,9	18,1	320
	EF061806	<i>Phormia regina</i> ^a	82,7	17,3	324
	EF560190	<i>Phormia regina</i> ^b	82,9	17,1	327

Calliphoridae	EF560193	<i>Protophormia terraenovae</i> ^b	82,8	17,2	331
	EF560188	<i>Mesembrinella peregrina</i> ^b	80,0	20,0	291
	EU076456	<i>Mesembrinella bellardiana</i> (ind. 1) ^b	77,5	22,5	334
	EU076455	<i>Mesembrinella bellardiana</i> (ind. 2) ^b	79,5	20,5	307
Muscidae	EF560191	<i>Stomoxys calcitrans</i> ^b	75,0	25,0	324
	EF560189	<i>Musca domestica</i> ^b	75,9	24,1	365
	EF560184	<i>Haematobia irritans</i> ^b	76,1	23,9	347
Média			80,3	19,7	317

a = sequências produzidas por Song et al. 2008a;

b = produzidas neste trabalho;

c = sequências produzidas por Pickles et al. 2002 (não publicado);

d = sequências produzidas por Nelson et al. 2007;

e = sequências produzidas por Nelson et al. 2008;

f = Rognes (1991) sugeriu o gênero *Aldrichina* como sinônimo de *Calliphora*.

Além disso, com a finalidade de se explorar a variabilidade intra-específica em termos de sequência e estrutura do ITS2, 30 indivíduos congelados de *C. hominivorax* de 10 populações diferentes, compreendendo a atual distribuição geográfica conhecida para a espécie (Tabela 2 e Figura 1), foram obtidas da coleção mantida a -70°C no Laboratório de Genética e Evolução Animal (CBMEG / UNICAMP). A extração do DNA total foi realizada através do protocolo de fenol / clorofórmio (adaptado de Infante & Azeredo-Espin 1995) para os espécimes congelados e utilizando o reagente DNazol (Invitrogen), segundo instruções do fabricante, para os espécimes secos e preservados em etanol.

2.2. Amplificação via PCR, clonagem e sequenciamento

As reações de PCR para amplificação da região ITS2 foram conduzidas com 10mM Tris-HCl (pH 8,8), 50mM KCl, 2mM MgCl₂, 80μM dNTPs, 0,4μM "oligo" 5,8S (5'-ATC ACTCGGCTCGTGGGATTCGAT-3'), 0,4μM "oligo" 28S (5'-GTTAGTTTCTTTTCCT CCCCT-3'), 2,5U *Taq* DNA polimerase (Fermentas) e 1-2μg de DNA total extraído para um volume final de 25μL. Os ciclos de amplificação incluíram um passo inicial de desnaturação a 94°C por 3 minutos, seguido de 35 ciclos de 94°C por 1 minuto, 55°C por 45

segundos e 72°C por 2 minutos, além de um passo final de extensão a 72°C por 3 minutos. Os produtos de PCR foram visualizados em gel de agarose 1X TAE (40mM Tris-acetato, 1mM EDTA) 1,5%, corados com brometo de etídeo.

Tabela 2. Lista dos espécimes de *C. hominivorax*, e suas respectivas regiões de origem, utilizados na avaliação da variabilidade intra-específica do ITS2. O tamanho total do fragmento do ITS2 sequenciado, seu conteúdo nucleotídico e o número de acesso das sequências no GenBank também são mostrados.

População	Nº de indivíduos	ITS2 (pb)	A+T (%)	Número de acesso (GenBank)
Brasil (Centro-Oeste)	3	351-355	81,0-81,4	HQ435567 - HQ435575
Brasil (Sudeste)	3	352-356	81,1-81,5	HQ435593 - HQ435601
Colômbia	3	351-354	81,0-81,3	HQ435558 - HQ435566
Cuba	3	351-358	81,0-81,6	HQ435576 - HQ435583
República Dominicana	3	351-362	81,0-81,8	HQ435611 - HQ435619
Equador	3	355-360	81,1-81,7	HQ435584 - HQ435592
Paraguai	3	351-354	80,8-81,7	HQ435602 - HQ435610
Trinidad e Tobago	3	352-358	81,3-81,6	HQ435620 - HQ435628
Uruguai	3	352-357	81,0-81,4	HQ435629 - HQ435637
Venezuela	3	351-357	80,7-81,5	HQ435638 - HQ435646
Total / Média	30	354,2	81,3	

Com o intuito de investigar a existência de variação intragenômica nas sequências do ITS2 em Calliphoridae, um passo de clonagem foi conduzido previamente ao sequenciamento para todos os espécimes utilizados neste estudo. Neste contexto, os produtos amplificados via PCR foram purificados utilizando-se o kit "QIAquick PCR purification kit" (QIAGEN) e clonados em um vetor pCR2.1 usando o kit TA-Cloning (Invitrogen), de acordo com as instruções do fabricante. Células competentes de *E. coli* linhagem DH10B foram transformadas utilizando-se o protocolo de choque químico com CaCl₂ (adaptado de Sambrook et al. 1989) e plaqueadas em meio LB sólido com 50mg/mL

de X-Gal e 50µg/mL de Ampicilina. As placas foram incubas a 37°C por 12-14 horas. Os plasmídeos foram extraídos das células transformadas utilizando-se o método de hidrólise alcalina (Sambrook et al. 1989) e foram então digeridos com 5U da enzima de restrição *EcoRI* por 2 horas a 37°C. Os resultados das digestões foram visualizados por eletroforese em gel de agarose 1X TAE 1% corado com brometo de etídeo. Pelo menos três clones carregando o inserto de tamanho esperado correto foram submetidos a reação de sequenciamento.



Figura 1. Localidades amostradas de populações de *C. hominivorax*, as quais foram utilizadas na avaliação da variabilidade intra-específica do ITS2 (Cortesia: Mariana Lúcio Lyra).

As reações de sequenciamento foram conduzidas em um sequenciador ABI 3700 (Applied Biosystems) usando o "Big Dye Terminator Cycle Sequencing Ready Reaction Kit" (Applied Biosystems), de acordo com as instruções do fabricante, com os "oligos" universais M13 direto (5'-GTAAAACGACGGCCAG-3') e M13 reverso (5'-CAGGAAACAGCTATGAC-3') e 200-600ng de DNA.

2.3. Análises de sequência e predição de estrutura secundária

Os cromatogramas resultantes do sequenciamento foram analisados e editados (remoção das sequências do vetor e de partes do inserto com má qualidade) com o programa FinchTV 1.3.1 (Geospiza, Inc.). Conteúdo nucleotídico, taxas de transição / transversão e distâncias genéticas foram estimados usando o programa MEGA 4 (Tamura et al. 2007). Testes de saturação de substituição (Xia et al. 2003; Xia & Lemey 2009) foram conduzidos com o programa DAMBE 5.2.31 (Xia & Xie 2001) considerando apenas os sítios completamente resolvidos nas análises (opção padrão). Regiões em simples e dupla-fita foram analisadas separadamente e concatenadas. Análises estatísticas adicionais foram realizadas usando a plataforma R 2.11.1 (R Development Core Team 2010).

As estruturas secundárias do ITS2 foram preditas usando uma combinação de abordagens, incluindo: (1) modelagem por homologia usando como modelo a estrutura descrita para *Drosophila melanogaster* (Young & Coleman 2004); (2) modelagem *in silico* usando o programa mfold 2.3 (versão webserver - Zuker 2003), utilizando os parâmetros padrões para predição da estrutura, com exceção da temperatura, mudada para 25°C, dado que esta temperatura (padrão para temperatura ambiente) é mais plausível de representar a temperatura média corporal de um inseto; e (3) comparação entre as estruturas preditas para todos os espécimes sob análise e estabelecimento de um padrão comum de estrutura secundária. A nomenclatura adotada foi a mesma utilizada para a estrutura de *D.*

melanogaster, com as regiões de hélice-alça ("helix-loop") sendo nomeadas I, II, III e IV, sendo que o domínio III foi subdividido nas ramificações IIIa, IIIb e IIIc, e a hélice proximal desse domínio, antes da junção central, foi nomeada como III-r ("III-root" ou III-raíz). O alinhamento das sequências, levando em consideração informações sobre a estrutura secundária das moléculas, foi conduzido com o programa 4SALE (Seibel et al. 2006), com algumas correções manuais finais.

2.4. Análises de CBCs e identificação espécie-específica

Alinhamentos de sequência e estrutura secundária do ITS2 foram construídos para todos os pares de espécies congêneras mostrados na Tabela 1 (conjunto inter-específico) e, quando disponível, para todas as sequências pertencentes à mesma espécie (conjunto intra-específico), utilizando o programa 4SALE (Seiber et al. 2006). Mudanças compensatórias de bases ("compensatory base changes" - CBCs), isto é, substituições em ambos os nucleotídeos de um sítio pareado na estrutura e que não resultam em uma desestabilização da mesma (e.g., GC↔AT) (Guttel et al. 1994), e mudanças semi-compensatórias de bases (hemi-CBCs), isto é, substituições que ocorrem em apenas um dos nucleotídeos de um sítio pareado na estrutura secundária sem desestabilizar a mesma (e.g., AU↔GU), foram averiguadas e contadas em todos os alinhamentos de ambos os conjuntos inter e intra-específico. O potencial dessas mudanças para servir como um caractere de diagnóstico espécie-específico (Coleman & Vacquier 2002; Müller et al. 2007) foi então avaliado baseando-se em sua ocorrência dentro e entre espécies. Para as análises no conjunto inter-específico nos quais havia duas ou mais sequências por espécie, apenas mutações fixadas foram contadas.

2.5. Análises filogenéticas

As análises filogenéticas foram conduzidas com os métodos de Neighbor-Joining

(NJ), máxima parcimônia (MP), máxima verossimilhança (ML, do inglês "maximum likelihood") e inferência Bayesiana (BI, do inglês "Bayesian inference"), usando informações tanto de sequência quanto de estrutura secundária, quando possível. Espécies da família Muscidae foram usadas como grupo externo nas análises. As espécies do gênero *Mesembrinella* também podem ser vistas como um grupo externo adicional nestas análises de acordo com algumas hipóteses de parentesco para a família Calliphoridae baseadas tanto em marcadores morfológicos quanto moleculares (Guimarães 1977; Rognes 1997; Kutty et al. 2010). Sequências do ITS2 de Calliphoridae disponíveis no GenBank também foram incluídas nas análises.

A seleção do modelo de substituição que melhor se adéqua ao conjunto de dados foi realizada com o programa MrAIC 1.4.4 (Nylander 2004). As análises de NJ baseadas em informação de sequência apenas foram realizadas com o programa MEGA 4 (Tamura et al. 2007) usando o modelo de substituição Tamura-Nei com o parâmetro gama para correção da heterogeneidade de taxas de substituição nos sítios ao longo da sequência (TN93+G), o qual foi escolhido pelo programa MrAIC 1.4.4 como o modelo que melhor se ajusta aos dados. Regiões com lacunas ("gaps") foram totalmente excluídas (opção "complete deletion") ou excluídas apenas nas comparações par-a-par (opção "pairwise deletion"). Análises de NJ considerando informações da estrutura secundária da molécula do ITS2 foram realizadas com o programa ProfDistS 0.9.8 (Wolf et al. 2008) usando um modelo de substituição GTR (do inglês "general time reversible") específico para perfis de sequência-estrutura, o qual encontra-se implementado no programa. As análises conduzidas no programa ProfDistS foram conduzidas utilizando-se três métodos, todos discutidos no manual on-line do programa: (1) método "RNA/DNA structure profile NJ", com a opção de geração de perfis definida como automática; (2) método "Bootstrapping → Distance

Calculation → NJ → Consensus"; e (3) método "Distance → NJ → Consensus". As análises de MP foram conduzidas com o programa TnT v1.1 (Goloboff et al. 2003) usando a opção de novas tecnologias de busca (do inglês, "New Technology Search") ("Sectorial Search", "Ratchet", "Drift" e "Tree Fusing"). Análises de ML foram conduzidas com o programa PhyML 3.0 (Guindon & Gascuel 2003), também utilizando o modelo de substituição TN93+G, com todos os parâmetros estimados durante a análise. Valores de suporte para os nós / ramos nas árvores inferidas pelos métodos de NJ, MP e ML foram estimados usando o método de re-amostragem por "bootstrap" com 1.000 replicatas.

As análises de BI foram conduzidas com o programa mcmcphase do pacote de programas PHASE 2.0 (Gowri-Shankar & Jow 2006) e também com o programa MrBayes v3.1.2 (Ronquist & Huelsenbeck 2003), utilizando em ambos uma gama de diferentes combinações de modelos de substituição e partições, refletindo a conformação secundária adotada pela molécula, conforme resumido na Tabela 3. Todas as corridas de BI foram conduzidas por 5.000.000 de gerações (frequência de amostragem = 1.000) e o "burn-in" (número de amostragens descartadas) foi definido para 25% após checar a convergência das cadeias no final das análises. A estrutura secundária consenso para a região do ITS2 em Calliphoridae, necessária para a implementação dos modelos de substituição baseados em estrutura secundária, foi determinada pelo programa "SecondaryStructConsensus" do pacote de programas PHASE.

2.6 Avaliação das árvores inferidas usando o ITS2

Com o objetivo de avaliar as árvores inferidas utilizando a região do ITS2 com diferentes métodos, as topologias recuperadas foram comparadas com árvores previamente publicadas para o grupo e que utilizaram uma maior amostragem de táxons e/ou caracteres.

Tabela 3. Resumo de todas as combinações de modelos de substituição utilizados nas análises de inferência Bayesiana. O número de parâmetros de cada modelo é dado, bem como o número total de parâmetros incluindo a própria topologia e os comprimentos de ramos. ss = regiões em simples-fita; ds = regiões em dupla-fita.

Programa	ITS2		Nº de parâmetros modelo (total)	Breve descrição
	ss	ds		
MrBayes	GTR+G		9 (208)	O modelo GTR padrão com o parâmetro de correção para heterogeneidade de taxas de substituição ao longo dos sítios.
	GTR+G	Doublet (GTR)+G	31 (230)	Modelo GTR combinado com o modelo "Doublet", no qual as mudanças em bases pareadas ocorrem através de um processo em duas etapas, cada uma delas ocorrendo de acordo com um modelo GTR.
PHASE	TN93+G		6 (205)	O modelo Tamura-Nei padrão (TN93) com o parâmetro de correção para heterogeneidade de taxas de substituição ao longo dos sítios.
	TN93+G	RNA6A+G	28 (227)	Modelo TN93 combinado com o modelo RNA6A, o mais geral na categoria de 6-estados, na qual sítios não pareados ("mismatches") não são considerados.
	TN93+G	RNA7A+G	35 (234)	Modelo TN93 combinado com o modelo RNA7A, o mais geral na categoria de 7-estados, na qual todos os sítios não pareados são considerados em um mesmo estado (MM).
	TN93+G	RNA7D+G	18 (217)	Modelo TN93 combinado com o modelo RNA7D, uma restrição do modelo RNA7A com menos parâmetros de taxas de substituição.
	TN93+G	RNA16A+G	28 (227)	Modelo TN93 combinado com o modelo RNA16A, o mais geral na categoria de 16-estados que é passível de ser usado. Ele tem 5 parâmetros de taxas de substituição ao invés de 120 (RNA16).
	TN93+G	RNA16D+G	15 (214)	Modelo TN93 combinado com o modelo RNA16D, o qual é uma tentativa de transformar um modelo de 4-estados padrão em um modelo de estrutura secundária. Ele permite que os pares GU/UG tenham frequência intermediária entre pares Watson-Crick e "mismatches".

Dado que não existe uma única filogenia publicada para Calliphoridae que inclui todas as espécies utilizadas aqui, subgrupos do conjunto total de espécies amostrado foram escolhidos para as comparações: (1) espécies do gênero *Chrysomya*, cuja relação recuperada foi comparada com a árvore publicada no trabalho de Singh et al. (2011), a qual sugere uma divisão do gênero em três clados - [clado I] *C. semimetallica* e *C. latifrons*; [clado II] *C. bezziana*, *C. pinguis* e *C. megacephala*; e [clado III] *C. putoria*, *C. nigripes*, *C.*

varipes, *C. flavifrons*, *C. incisuralis*, *C. albiceps* e *C. rufifacies*; (2) espécies do gênero *Lucilia*, cuja topologia inferida foi comparada com diferentes árvores previamente descritas (Kurahashi 1966; Shepard 1971; Brown 1979; Stevens & Wall 1996, 1997; Wells et al. 2007), com a maioria desses trabalhos corroborando uma separação entre os clados (*L. cuprina* + *L. sericata*), espécies previamente incluídas em seu próprio gênero *Phaenicia*, e (*L. eximia* + *L. bazini* + *L. ampullacea* + *L. porphyrina* + *L. caesar* + *L. illustris*); (3) gêneros da subfamília Chrysomyinae, cujas relações foram comparadas com a obtida por Singh & Wells (2011), a qual suporta uma relação mais próxima entre os gêneros (*Chrysomya*, (*Phormia*, *Protophormia*)) e (*Cochliomyia*, (*Chloroprocta*, *Hemilucilia*)); e (4) subfamílias de Calliphoridae, entre as quais a relação (Chrysomyinae, (Calliphorinae, Luciliinae)) é corroborada pela análise com marcadores moleculares de Kutty et al. (2010) e também por algumas classificações tradicionais baseadas em caracteres morfológicos (Shewell 1987; McAlpine 1989 - ambos reconhecendo a tribo Luciliini dentro da subfamília Calliphorinae).

Outras topologias usadas nas comparações foram as previamente descritas por: Wallman & Donnellan (2001); Wells & Sperling (2001); Stevens & Wall (2001); Otranto & Stevens (2002); Stevens (2003); Chen et al. (2004); Stevens & Wallman (2006); Stevens et al. (2006); Wells & Williams (2007); e Nelson et al. (2007).

Algumas relações que foram recorrentemente recuperadas tanto nas árvores inferidas com o ITS2 como nas filogenias previamente publicadas não foram incluídas nas comparações, como a recuperação de *H. ligurriens* entre as espécies do gênero *Lucilia*. As relações dentro da subfamília Calliphorinae, aqui representada apenas por um único gênero, dado que o gênero *Aldrichina* foi descrito por Rognes (1991) como sinônimo de *Calliphora*

e o mesmo tem agora status inválido na base de dados "Biosystematic database of World Diptera" (Pape & Thompson 2010), também não foram incluídas nas comparações.

As árvores inferidas a partir da região do ITS2 e as filogenias previamente publicadas para Calliphoridae foram comparadas em termos de topologia usando a medida de distâncias simétricas ("Symmetric Distances" - Robinson & Foulds 1981), implementada no programa TreeDist do pacote PHYLIP 3.67 (Felsenstein 2005). As árvores inferidas com o ITS2 também foram avaliadas baseando-se em: (1) o número de relações não resolvidas (politomias) presentes nas árvores; (2) a correta separação de espécies proximamente relacionadas; (3) a recuperação de clados comumente aceitos como monofiléticos; e (4) o suporte médio das relações recuperadas, estimado como sendo a média aritmética de todos os valores de suporte presentes na árvore.

Para os métodos estatísticos de inferência filogenética, i.e., ML e BI, os critérios de seleção AIC ("Akaike Information Criteria"), AICc (AIC corrigido) e BIC ("Bayesian Information Criteria") foram usados também na avaliação de qual combinação de método e modelos de substituição se ajusta melhor aos dados. Para as análises de inferência Bayesiana, o maior valor de verossimilhança amostrado durante a fase estacionária da corrida MCMC (Markov chain - Monte Carlo) foi utilizado como o estimador de máxima verossimilhança necessário para as análises. O número de parâmetros livres em cada um dos modelos de substituição usados nas análises é mostrado na Tabela 3.

Por último, as diferentes estratégias de particionamento dos dados utilizadas nas análises de BI foram comparadas entre si utilizando o método Bayesiano de comparação de modelos Bayes Factor. Como estimador da verossimilhança marginal, necessária para o cálculo do Bayes Factor, foi utilizada a média harmônica de todas as verossimilhanças amostradas durante a fase estacionária da corrida MCMC (como sugerido por Newton &

Raftery 1994 e Nylander et al. 2004). As interpretações dos valores de Bayes Factor foram baseadas na tabela apresentada por Kass & Raftery (1995).

3. Resultados

3.1. Análises de sequência e estrutura secundária

A caracterização do segmento delimitado pelas regiões 5,8S e 28S revelou a presença de uma subunidade adicional de rDNA, 2S, a qual divide a região do ITS2 em dois segmentos: ITS2a, com aproximadamente 30pb, e ITS2, de tamanho variável (291 a 365pb - Tabela 1). A subunidade 2S foi identificada por sua identidade com sequências previamente publicadas da região ITS2 de *D. melanogaster* (Jordan et al. 1976; Tautz et al. 1988) e *M. domestica* (Schlötterer et al. 1994). Como nenhuma variação foi observada na região ITS2a, a mesma foi excluída das análises.

A análise das sequências da região ITS2 revelou um grande viés em sua composição nucleotídica (A+T%), a qual variou entre 77,5% e 82,9% em Calliphoridae (Tabela 1). A razão geral entre as taxas de transição e transversão ($R = ts/tv$) é 3,1, indicando uma ocorrência de transições três vezes maior que de transversões (Tabela 4). A tabela 2 mostra os mesmos dados para o conjunto de sequências da região ITS2 de *C. hominivorax* obtidos da análise das dez populações amostradas na América do Sul e ilhas do Caribe.

O cálculo da distância p (não corrigida) entre sequências do ITS2 em diferentes níveis taxonômicos revelou a existência de um padrão crescente e com pouca sobreposição da distância genética para os níveis taxonômicos hierarquicamente superiores ao específico, enquanto para o nível sub-específico não foi detectada variação significativa. Neste sentido, pode ser observado um ponto de corte bem definido em termos de distância genética na interface entre os níveis sub e supra-específicos (Figura 2).

Tabela 4. Razão entre as taxas de transição e transversão (R); conteúdo nucleotídico (A+T%) e distância p não-corrigida (média) de cada domínio da estrutura do ITS2 para todas as espécies analisadas, considerando ou não a região de "alça" terminal. Call = espécies de Calliphoridae, excluindo o gênero *Mesembrinella*; Mes = espécies do gênero *Mesembrinella*; Mus = espécies da família Muscidae.

Domínios	R	A+T (%)	Distância p
I (sem "alça")	13,8	80,5	0,218
I (com "alça")	5,6	82,4	0,245
II (sem "alça")	6,1	73,7	0,120
II (com "alça")	5,3	76,4	0,143
III (sem "alça")	1,8	78,9	0,097
III (com "alça")	2,4	79,9	0,118
III r	0,3	78,3	0,093
IIIa (sem "alça")	0,6	79,4	0,110
IIIa (com "alça")	4,2	85,2	0,147
IIIb (sem "alça")	14,8	74,5	0,060
IIIb (com "alça")	16,4	76,1	0,065
IIIc (sem "alça")	4,2	89,0	0,199
IIIc (com "alça")	7,6	86,8	0,274
IV	3,2	92,9	0,312
Média	3,1	80,3	0,141
Média (Call)	3,9	80,4	0,123
Média (Call+Mes)	3,4	80,4	0,134
Média (Call+Mus)	3,4	80,3	0,131

Com relação ao nível de variação intra-específica e intragenômica, pouca variação foi observada no conjunto de sequências de *C. hominivorax* analisado, sendo a mesma composta principalmente de diferenças no número de repetições de sequências de mono e di nucleotídeos nas regiões ricas em A+T, especialmente na região poli-T do domínio IV. Esta variação, contudo, é bem menor em termos de magnitude do que a encontrada no nível inter-específico, dado que o cálculo da distância p (não corrigida) entre os diferentes clones sequenciados de *C. hominivorax* variou entre 0,000 e 0,018 (média = 0,003), enquanto a distância p entre qualquer um desses clones e a sequência de *C. macellaria* apresentou-se

no intervalo de 0,049 a 0,061 (média = 0,051). Sendo assim, a variação intra-específica e intragenômica parece ocorrer em uma escala muito pequena e é improvável que a mesma tenha algum efeito em termos de encobrir o sinal filogenético no nível de espécies.

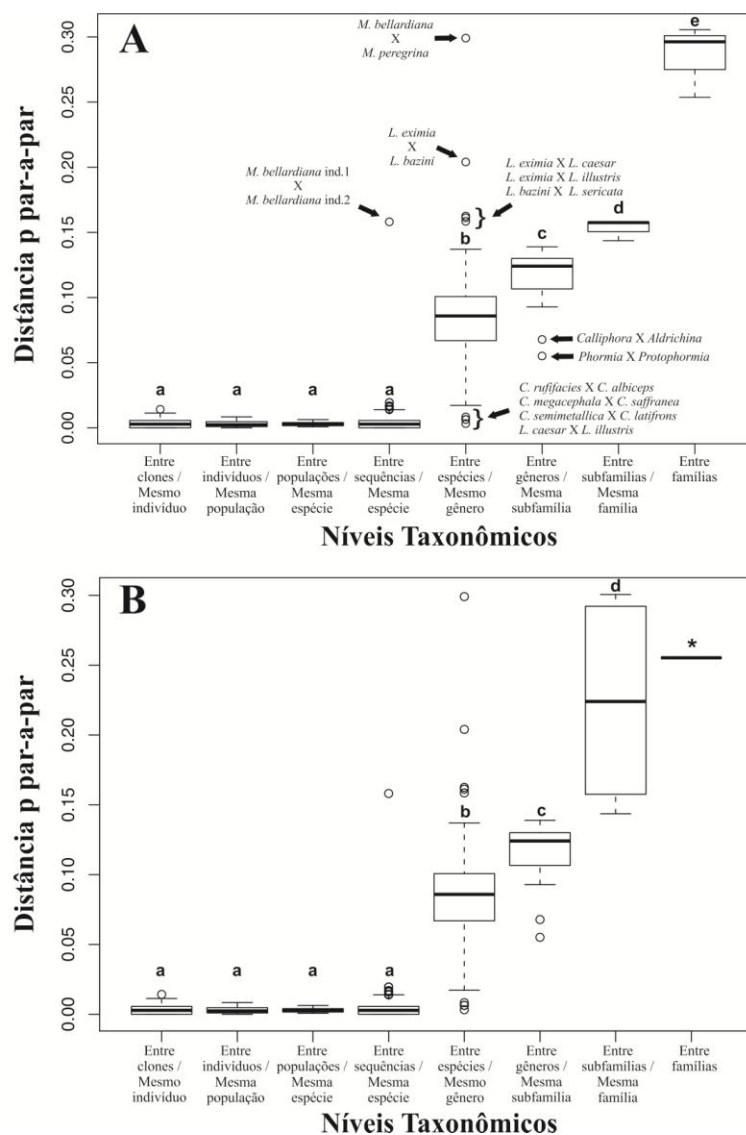


Figura 2. "Boxplot" mostrando a variação na magnitude das distâncias genéticas (distância p não corrigida) para diferentes níveis taxonômicos organizados hierarquicamente. Observações "outliers" estão identificadas. As distâncias entre famílias foram calculadas considerando as espécies do gênero *Mesembrinella* (A) em uma família própria (Mesembrinellidae) ou (B) como uma subfamília de Calliphoridae (Mesembrinellinae). As letras de "a" a "e" indicam classes de valores estatisticamente diferentes (teste T de Welch, $p < 0,05$). "*" indica que há apenas uma observação nessa categoria (distância entre Muscidae X Calliphoridae).

Os testes de saturação de substituições não indicaram a presença de um nível significativo de saturação no conjunto de sequências da região ITS2 para as espécies analisadas, mesmo quando a região foi subdividida entre simples e dupla-fita, apesar de a primeira apresentar níveis de saturação maiores que a segunda (Tabela 5).

Tabela 5. Resultados dos testes de saturação de substituição realizados no programa DAMBE 5.2.31. ss = regiões em simples fita; ds = regiões em dupla fita. Iss = índice de saturação de substituição; Iss.c S ou A = valor crítico para o Iss se a árvore verdadeira é simétrica (S) ou assimétrica (A), respectivamente. Se $Iss > Iss.c$, o conjunto de sequências apresenta altos índices de saturação e possui uso limitado para inferências filogenéticas.

Sequências		Iss	Iss.c S	P	Iss.c A	P
Calliphoridae	ITS2 (ss/ds)	0,095	0,690	0,000	0,375	0,000
	ITS2 (ss)	0,155	1,310	0,000	1,456	0,000
	ITS2 (ds)	0,083	0,712	0,000	0,419	0,000
Calliphoridae + grupos externos	ITS2 (ss/ds)	0,109	0,707	0,000	0,408	0,000
	ITS2 (ss)	0,153	1,508	0,000	1,793	0,000
	ITS2 (ds)	0,107	0,744	0,000	0,478	0,000

As estruturas secundárias preditas com o programa mfold para a região ITS2 de Calliphoridae (Figura 3) são bastante similares ao modelo proposto para a estrutura de *D. melanogaster* e, em um sentido mais amplo, ao padrão comum de estrutura secundária descrito para eucariotos (Joseph et al. 1999; Schultz et al. 2005), a qual inclui quatro domínios hélice-alça emergindo de um anel central. As principais diferenças encontradas entre as estruturas preditas para Calliphoridae em relação à descrita para *D. melanogaster* são a ausência do subdomínio IIa e a presença de ramificações no domínio III (hélices IIIa e IIIc). Para algumas sequências de Calliphoridae, o programa mfold de fato gerou estruturas

sem um domínio III ramificado, mais semelhantes à estrutura de *D. melanogaster*, mas essas estruturas foram sempre preditas como modelos sub-ótimos (apresentando maiores valores de energia livre e, portanto, menos estáveis). Como o modelo de estrutura apresentando um domínio III ramificado foi recuperado para as sequências de todas as espécies analisadas como a estrutura de mínima energia livre, o modelo final adotado para a estrutura secundária da região ITS2 de Calliphoridae possui ramificações no domínio III.

Em termos de variação estrutural na região ITS2 das espécies analisadas, as principais diferenças encontradas foram: (1) presença / ausência de "bojos" ("bulges") e alças internas, as quais geram dobramentos nas regiões de hélice dos domínios do ITS2; (2) presença / ausência de alguns domínios e subdomínios, como a ausência da ramificação IIIc em *L. cuprina*, *L. sericata* e *M. bellardiana* (indivíduo 2); a ausência do domínio IV em *Aldrichina* (= *Calliphora*) *grahami*, *L. bazini*, *M. peregrina* e *M. bellardiana* (indivíduo 1); e a presença de um domínio adicional, IVb, em *M. domestica* e *H. irritans*; e (3) variação no número de pares de bases nas regiões em dupla fita, resultando em hélices maiores ou menores em diferentes grupos, como por exemplo as ramificações do domínio III, as quais são maiores na família Calliphoridae quando comparada a família Muscidae.

O alinhamento das sequências da região ITS2, corrigido posteriormente com informações do modelo predito de estrutura secundária, e o cálculo das distâncias p par-a-par para as espécies analisadas revelou que os domínios II e III da estrutura do ITS2 são os mais conservados, enquanto o domínio IV é o mais variável (Tabela 4). Em termos estruturais, todas as espécies analisadas possuem um nucleotídeo não pareado, uma pirimidina, na porção 5' do domínio II, o qual é uma citosina (C) na maioria das sequências, e um motivo de sequência bastante conservado (5'-GUCUAGCAU-3'), na extremidade distal da porção 5' do domínio III. Ambos os motivos conservados já foram descritos

também em *D. melanogaster* e em outras espécies de *Schizophora* (Young & Coleman 2004).

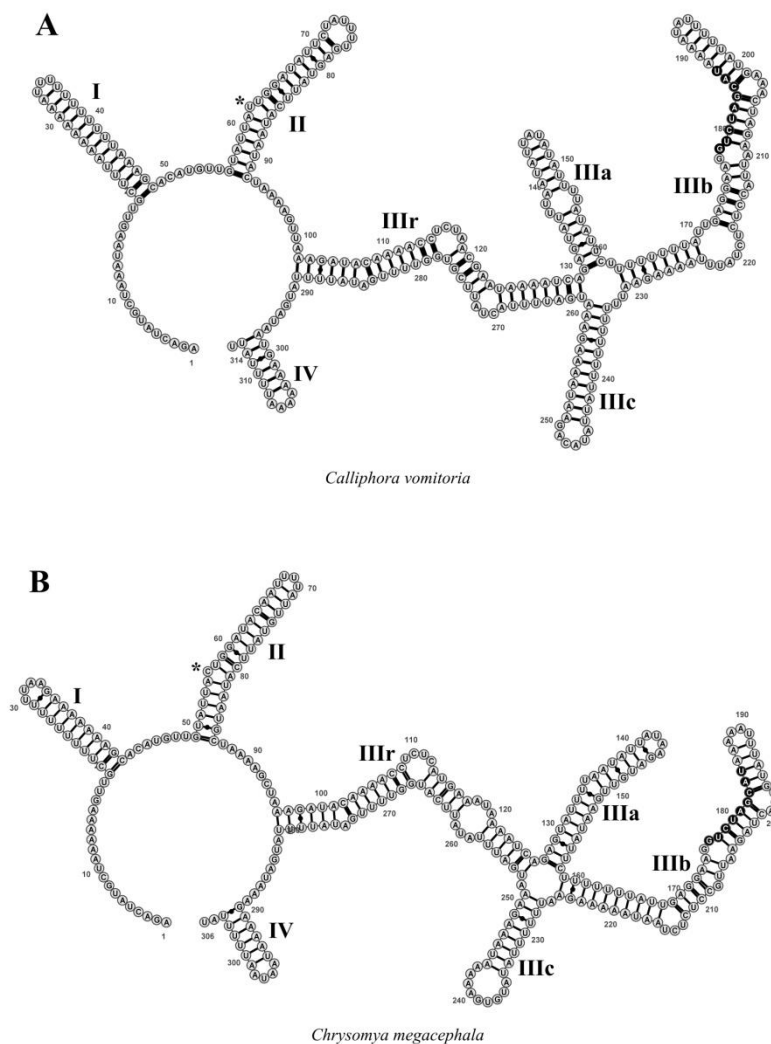


Figura 3. Algumas das estruturas secundárias previstas para a região do ITS2. (A) *C. vomitoria*; (B) *C. megacephala*; (C) *L. sericata*; (D) *C. hominivorax*; (E) *P. regina*; (F) *A. (= Calliphora) grahami*; (G) *M. peregrina*; e (H) *M. domestica*. Os domínios I, II, III e IV estão indicados, bem como as ramificações do domínio III (IIIa, IIIb e IIIc). Em (H), as duas possíveis subdivisões do domínio IV (IVa e IVb) são mostradas. O motivo conservado do domínio III (GUCUAGCAU) é mostrado em círculos pretos enquanto a pirimidina não pareada característica do domínio II é indicada por um "*". Pares de bases "G - C" são mostrados em destaque. As estruturas aqui mostradas foram desenhadas com o programa VARNA 3.3. (Darty et al. 2009).

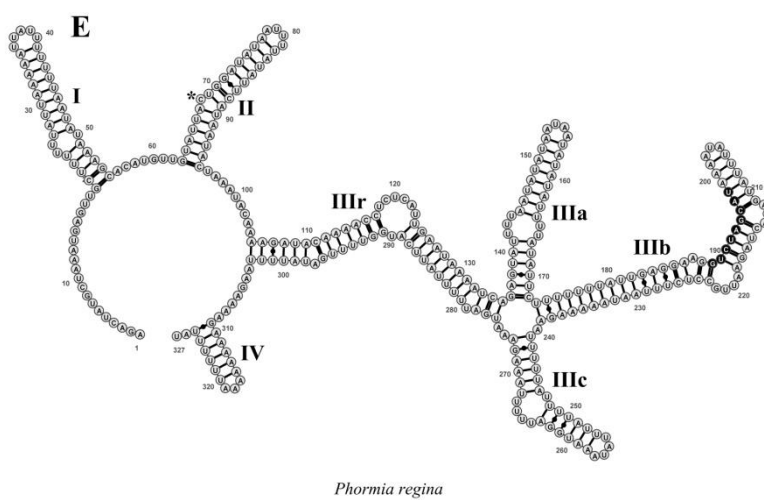
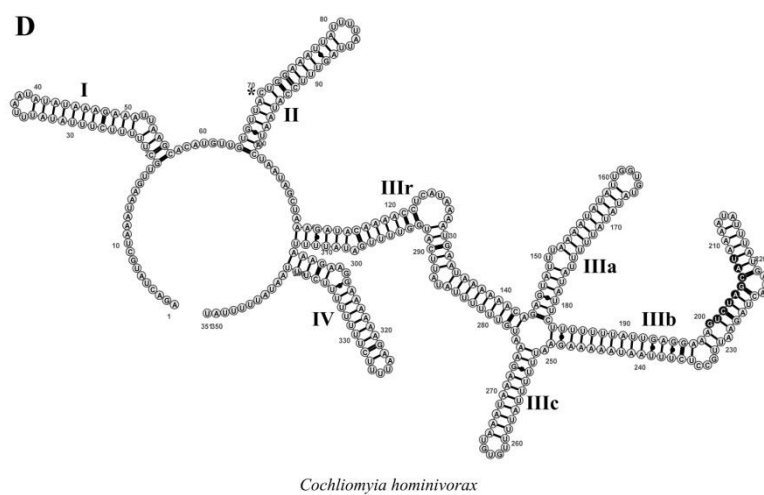
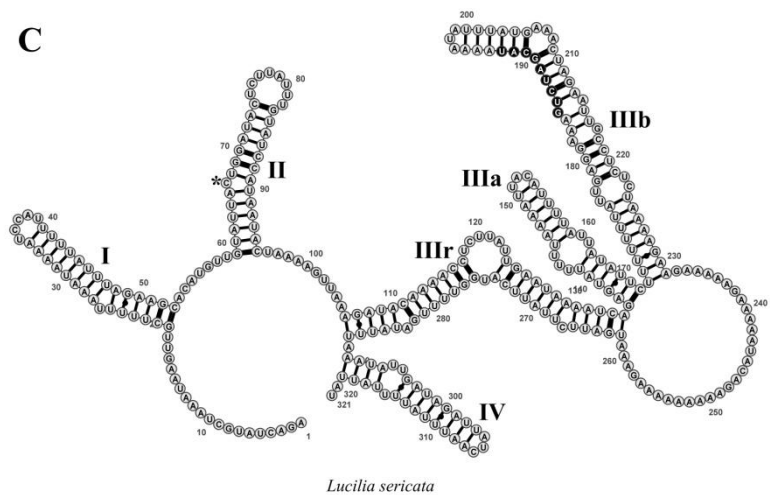


Figura 3. (Continuação)

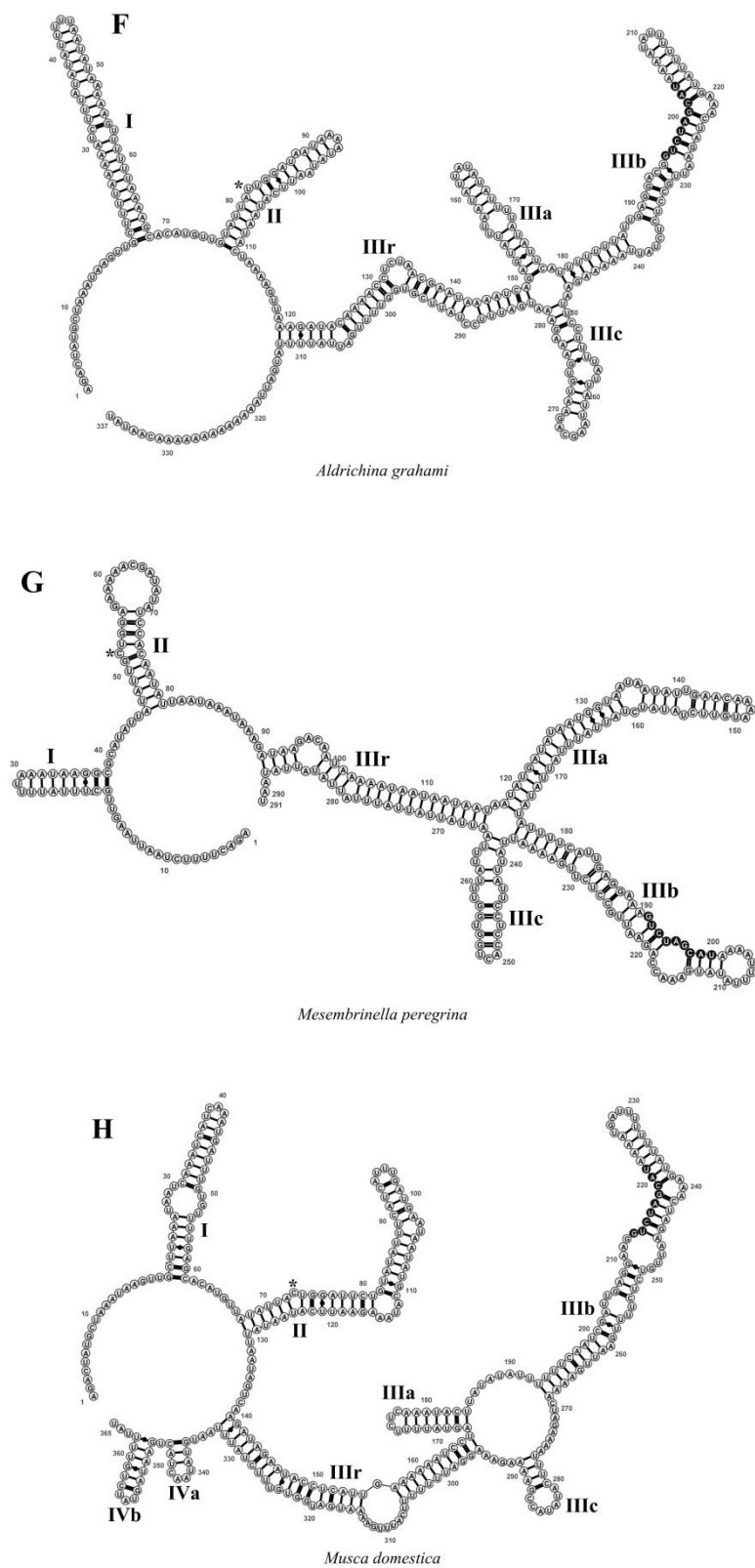


Figura 3. (Continuação)

A tabela 4 mostra os dados de taxas de transição e transversão (R), conteúdo nucleotídico (A+T %) e distâncias p não corrigida (média geral) para as sequências de Calliphoridae, *Mesembrinella* e Muscidae particionadas de acordo com os domínios da estrutura secundária do ITS2, incluindo ou não as alças terminais nas análises. A análise da correlação entre esses valores mostrou a existência de uma correlação positiva estatisticamente significativa entre a composição nucleotídica e os valores de distância p calculados, o qual pode ser extrapolado como uma medida de variabilidade em cada um dos domínios. Essa relação apresentou um coeficiente de Pearson $r = 0,8623$ (F-teste para a regressão linear - $F_{(1,6)} = 17,296$, $p < 0,01$) (Figura 4). Esse resultado mostra que domínios mais variáveis na estrutura secundária do ITS2 tem uma maior porcentagem de A+T que os domínios mais conservados.

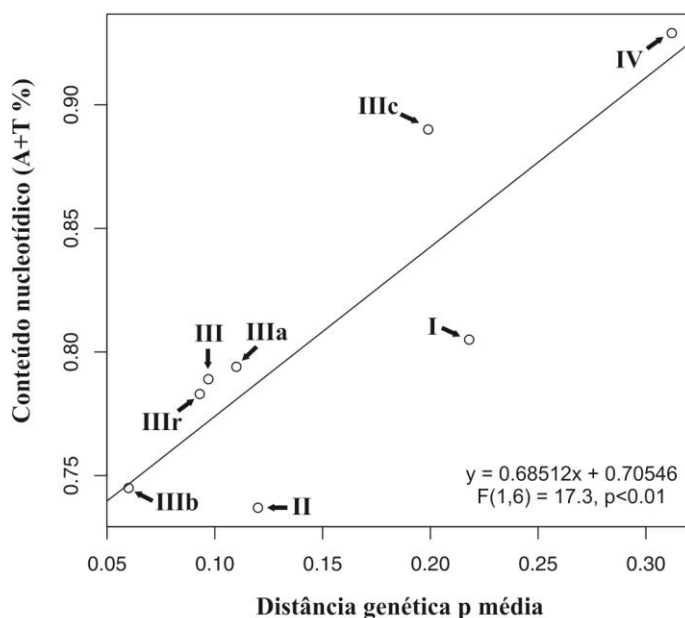


Figura 4. Gráfico mostrando a correlação entre o conteúdo nucleotídico (A+T%) e a distância p (média geral) para cada domínio individual da estrutura secundária do ITS2 em Calliphoridae. As observações para cada domínio são identificadas. A reta da regressão linear é mostrada.

Uma análise mais detalhada das estruturas secundárias mostradas na Figura 3 sugere uma correlação mais profunda entre a composição nucleotídica dos domínios e a organização geral da estrutura do ITS2 em Calliphoridae: pares G-C parecem ser mais frequentemente encontrados na região de interface entre regiões de simples e dupla-fita, como junções, "bojos" e alças internas e terminais. A Tabela A1 (Anexo A) mostra o número e frequência de pares A-U, G-C e G-U encontrados na estrutura secundária do ITS2 em Calliphoridae enquanto a Tabela A2 (Anexo A) mostra sua correlação com junções, "bojos" e alças. O teste qui-quadrado realizado mostrou que pares G-C são encontrados na interface de regiões em simples e dupla-fita em uma frequência maior que a esperada por suas frequências na estrutura secundária do ITS2 (i.e., aleatoriamente) em 11 de 37 espécies analisadas, para $p < 0,05$ (22 de 37 para $p < 0,10$). Na Tabela A3 (Anexo A), o mesmo teste é mostrado mas dessa vez considerando-se dois pares de bases na vizinhança da interface entre regiões em simples e dupla-fita ao invés de um único par (como na Tabela A2). Neste caso, o resultado foi significativo para uma proporção menor das espécies analisadas: 6 de 37 para $p < 0,05$ (17 de 37 para $p < 0,10$). Contudo, se analisadas apenas as regiões de interface internas às hélices ("bojos" e alças internas), os resultados aparecem como significativos para um número considerável das espécies analisadas (Tabelas A4 e A5 - Anexo A): 12 de 37 se apenas um par de bases for considerado na região de interface para $p < 0,05$ (18 de 37 para $p < 0,10$) e 18 de 37 se dois pares de bases forem considerados na região de interface para $p < 0,05$ (24 de 37 para $p < 0,10$).

Esses resultados sugerem, em conjunto, que há uma possível correlação entre a composição nucleotídica e a organização geral da estrutura secundária do ITS2 em pelo menos algumas espécies de Calliphoridae, com domínios mais conservados apresentando

um número maior de pares G-C, os quais estão localizados preferencialmente em posições críticas para garantir a adoção da conformação correta em regiões estruturais provavelmente essenciais para a função desenvolvida pelo ITS2.

3.2. Análises de CBCs e identificação espécie-específica

Na análise do conjunto intra-específico, para as 21 espécies com mais de uma sequência disponível para avaliação, foram encontradas hemi-CBCs em apenas 5 delas - *C. vomitoria*, *L. sericata*, *C. flavifrons*, *C. hominivorax* e *M. bellardiana* - e uma CBC em apenas uma - *M. bellardiana*. Já na análise do conjunto inter-específico, para os 120 alinhamentos produzidos para os pares de espécies congênicas, foram encontradas CBCs e hemi-CBCs em 77 deles, apenas CBCs em 3, apenas hemi-CBCs em 29 e nem CBCs nem hemi-CBCs em 11. A Tabela 6 resume as informações sobre as CBCs e hemi-CBCs encontradas nas comparações, incluindo sua distribuição na estrutura e suas frequências de ocorrência.

Dado que as CBCs ocorrem apenas entre e nunca dentro das espécies, com exceção de *M. bellardiana*, a observação de sua ocorrência parece ser um bom caráter diagnóstico para identificação espécie-específica. Para *M. bellardiana*, as duas sequências analisadas parecem ser muito divergentes e provavelmente não pertencem à mesma espécie, estando uma delas incorretamente identificada.

3.3. Análises filogenéticas e avaliação das árvores inferidas com a região ITS2

Um total de 26 árvores foram obtidas ao final das análises, compreendendo uma para cada combinação de método de inferência e modelo de substituição adotado, com exceção do método de MP, o qual resultou em oito árvores (7 árvores igualmente parcimoniosas e a árvore de consenso estrito), e do método " RNA/DNA structure profile NJ "

Tabela 6. Resumo das CBCs e hemi-CBCs encontradas nos alinhamentos de pares de espécies congêneras em Calliphoridae.

	Domínio	Classes de substituições	Tipos de substituições	Total por tipo	Total por domínio
hemi-CBCs (TOTAL = 316)	I	1	UG ↔ UA	96 (201)	105
		2	UG ↔ CG	9 (20)	
	II	3	GU ↔ AU	52 (70)	96
		4	GU ↔ GC	6 (25)	
		1	UG ↔ UA	31 (201)	
		2	UG ↔ CG	7 (20)	
	IIIa	4	GU ↔ GC	2 (25)	70
		1	UG ↔ UA	62 (201)	
		3	GU ↔ AU	6 (70)	
	IIIb	4	GU ↔ GC	8 (25)	13
		1	UG ↔ UA	5 (201)	
	IIIc	1	UG ↔ UA	3 (201)	7
		2	UG ↔ CG	4 (20)	
	IIIroot	4	GU ↔ GC	9 (25)	13
		1	UG ↔ UA	4 (201)	
	IV	3	GU ↔ AU	12 (70)	12
CBCs (TOTAL = 144)	I	1	CG ↔ UA	52 (58)	60
		2	AU ↔ UA	6 (42)	
		3	UG ↔ AU	2 (6)	
	II	1	CG ↔ UA	6 (58)	55
		4	GC ↔ AU	3 (5)	
		2	AU ↔ UA	18 (42)	
		5	CG ↔ AU	16 (20)	
		6	UA ↔ GC	1 (1)	
		7	CG ↔ GU	8 (8)	
		3	UG ↔ AU	2 (6)	
		8	GU ↔ UG	1 (1)	
	IIIa	2	AU ↔ UA	2 (42)	3
		3	UG ↔ AU	1 (6)	
	IIIb	2	AU ↔ UA	2 (42)	2
	IIIc	9	GU ↔ UA	2 (3)	3
		2	AU ↔ UA	1 (42)	
	IIIroot	2	AU ↔ UA	4 (42)	9
		5	CG ↔ AU	4 (20)	
		3	UG ↔ AU	1 (6)	
	IV	2	AU ↔ UA	9 (42)	12
		9	GU ↔ UA	1 (3)	
		4	GC ↔ AU	2 (5)	

conduzido no programa ProfDistS, o qual resultou em cinco árvores (quatro geradas pelo método de perfis e uma árvore consenso, a qual é idêntica às geradas pelos outros dois métodos implementados no programa). A Figura 5 mostra algumas árvores representativas das análises realizadas, enquanto outras árvores são mostradas na Figura B1 (Anexo B).

A Tabela 7 mostra o resultado das avaliações das árvores inferidas com a região ITS2 para Calliphoridae. A árvore de NJ recuperada com a opção de deleção completa de "gaps" ("complete deletion") não foi incluída nessas avaliações dado que muitas das relações recuperadas estão em completo desacordo com as hipóteses atualmente disponíveis para as relações entre as espécies da família Calliphoridae.

Algumas das relações recuperadas pelas árvores do ITS2 estão em concordância com as filogenias previamente publicadas para o grupo, como a existência de três clados de espécies dentro do gênero *Chrysomya* como proposto por Singh et al. (2011) (embora em algumas árvores inferidas com o ITS2 o clado III tenha sido recuperado como parafilético). Contudo, outras relações não estão, como por exemplo as relações entre os três clados de *Chrysomya*, entre os gêneros de Chrysomyinae (o que explica os relativamente elevados valores calculados para as distâncias simétricas entre as árvores) e entre *L. eximia* e as demais espécies do gênero *Lucilia*. Essas relações conflitantes foram, contudo, recorrentemente recuperadas independentemente do método empregado para inferir árvores com a região ITS2 e, dessa forma, podem refletir uma particularidade da informação evolutiva contida no marcador.

Politomias e parafilias encontradas nas árvores do ITS2 em geral apareceram em clados fracamente suportados quando completamente resolvidos ou monofiléticos nas demais árvores, respectivamente (Figura 5). Os valores de suporte para os clados foram em

geral maiores nas árvores inferidas pelo método Bayesiano e menores nos métodos de MP e NJ (árvore inferida no programa MEGA com deleção par-a-par de "gaps").

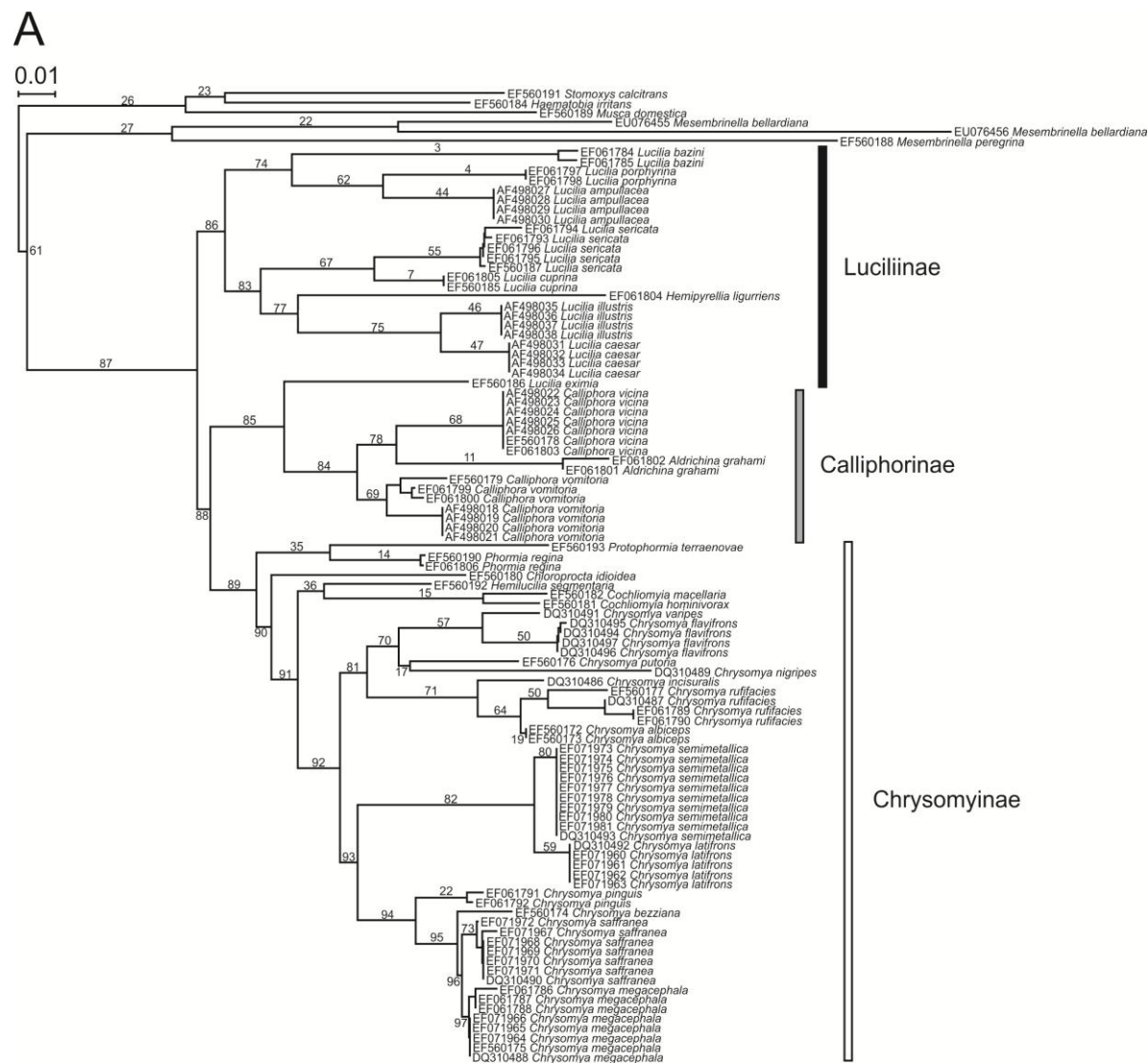


Figura 5. Algumas das árvores filogenéticas inferidas usando a região do ITS2 em Calliphoridae. (A) Árvore de Neighbor-Joining reconstruída considerando-se informações da estrutura secundária da molécula (árvore consenso de regra da maioria [50%] - ProfDistS 0.9.8); (B) Árvore consenso estrito das sete árvores igualmente parcimoniosas inferidas pelo programa TnT v1.1; (C) Árvore de máxima-verossimilhança (PhyML 3.0) ; (D) Árvore Bayesiana inferida usando a combinação de modelos TN93+G (simples-fita) e RNA6A+G (dupla-fita) (PHASE 2.0).

B

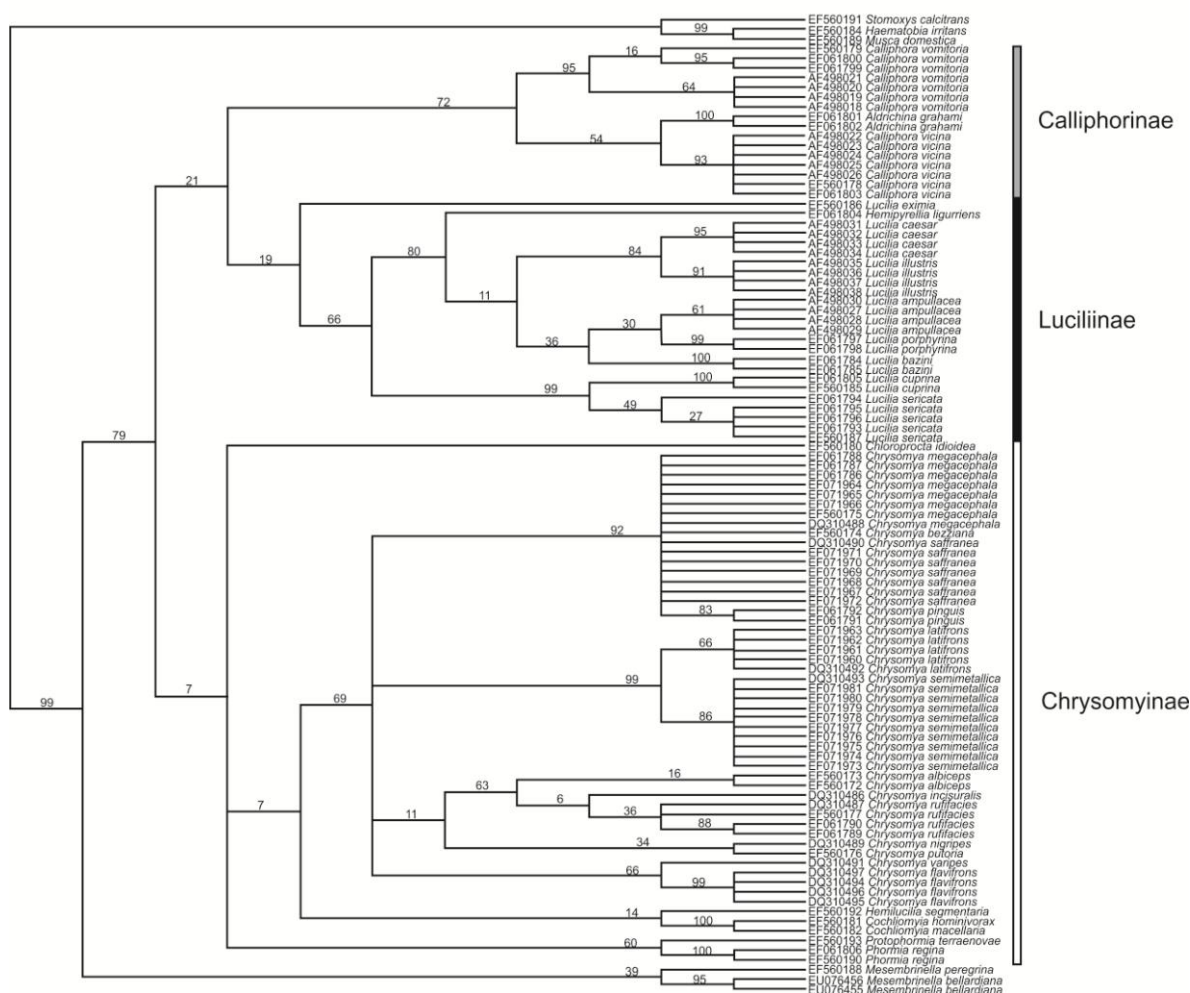


Figura 5. (Continuação)

Em uma análise geral das relações recuperadas por cada método de inferência, as árvores de NJ, incluindo as inferidas com o programa ProfDistS e, assim, considerando informações da estrutura secundária do ITS2, foram as únicas a recuperar uma relação entre as subfamílias de Calliphoridae diferente da hipótese mais aceita atualmente e, também, foram as únicas a alocar incorretamente algumas espécies de *Lucilia*, como *L. eximia*, recuperada dentro do gênero *Calliphora*, e *L. bazini*, recuperada entre as espécies de *Chrysomya* (apenas na árvore de NJ inferida no programa MEGA). Esse comportamento

observado para as espécies de *Lucilia*, aqui encontrado unicamente nas árvores de NJ, está provavelmente relacionado com o valores anormalmente elevados encontrados para as distâncias entre as espécies deste gênero, como mostrado na Figura 2.

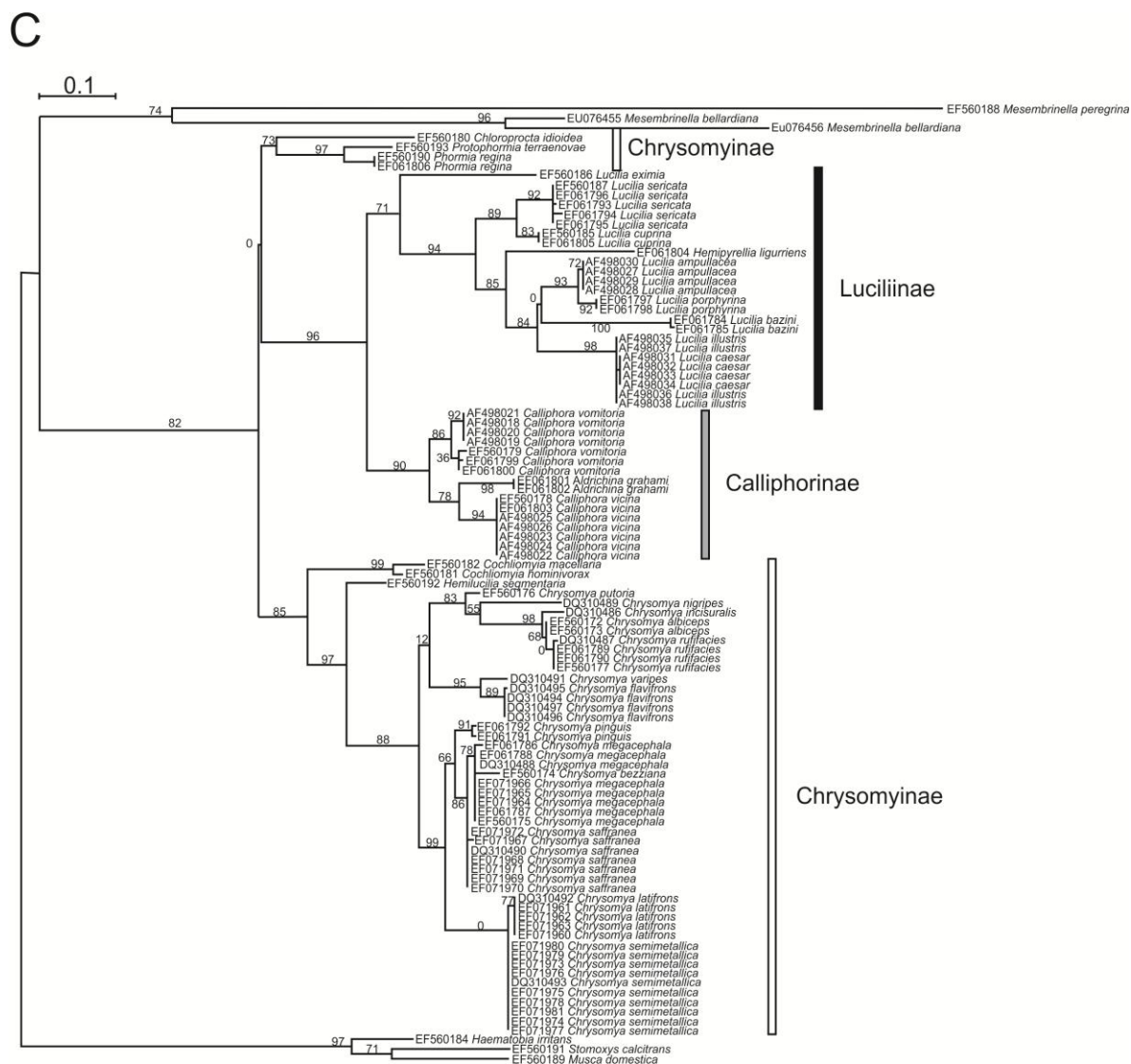


Figura 5. (Continuação).

O uso de modelos de substituição que incorporam informações sobre a estrutura secundária parece ter resultado em uma maior resolução tanto em termos das topologias estimadas quanto do suporte geral das relações inferidas, melhora este que se mostrou mais

pronunciada nas análises de NJ do que nas análises de inferência Bayesiana. Para as análises Bayesianas, embora apenas algumas pequenas mudanças na topologia e no suporte geral das relações tenham sido observadas quando do uso de modelos que incorporam informações estruturais nas análises, ambas as comparações de modelos feitas usando critérios estatísticos, máxima-verossimilhança (Tabela 8) e "Bayes factor" (Tabela 9), parecem favorecer o uso de combinações de modelos mais complexas (os quais lidam diferentemente com regiões de simples e dupla-fita), provavelmente indicando uma melhor estimativa das taxas de evolução na molécula usando estes modelos.



Figura 5. (Continuação).

Tabela 7. Resumo dos resultados encontrados na comparação entre as árvores inferidas usando a região ITS2 e as relações mais recorrentemente encontradas em publicações prévias. SD = distâncias simétricas; Suporte médio = média aritmética dos valores de suporte de todos os nós de cada árvore [0-1]; S = presença ("X") ou ausência ("-") de relações não resolvidas entre espécies próximas; NM = presença ("X") ou ausência ("-") de grupos não monofiléticos que foram recuperados como monofiléticos em estudos anteriores; P = presença ("X") ou ausência ("-") de politomias na árvore.

	Espécies de <i>Chrysomya</i>					Espécies de <i>Lucilia</i>					Gêneros de Chrysomyinae				Subfamílias			
	SD	Suporte médio	S ^b	NM ^c	P ^d	SD	Suporte médio	S ^b	NM ^c	P ^d	SD	Suporte médio	NM ^c	P ^d	SD	Suporte médio	NM ^c	P ^d
NJ (ProfDist)	8	0.76	-	-	-	6	0.76	-	X	-	6	0.68	-	-	4	0.87	-	-
NJ (MEGA)	10	0.56	X	X	-	6	0.42	X	X	-	6	0.22	-	-	10	0.16	-	-
MP	10	0.55	X	X	X	4	0.53	-	-	-	5	0.22	-	X	0	0.39	-	-
ML	9	0.75	X	-	X	3	0.87	X	-	X	7	0.88	-	-	1	0.85	X	X
BI (MB) GTR	6 ^a	0.86	X	-	X	3	0.95	-	-	X	7	0.87	-	-	1	0.92	X	X
BI (MB) Doublet	6 ^a	0.88	X	X	X	4	0.89	-	-	-	7	0.92	-	-	1	0.92	X	X
BI (PH) TN93	8	0.79	X	-	-	4	0.85	-	-	-	8	0.83	-	-	2	0.80	X	-
BI (PH) 6A	10	0.83	-	X	-	4	0.88	-	-	-	8	0.86	-	-	2	0.83	X	-
BI (PH) 7A	10	0.83	-	X	-	4	0.87	-	-	-	6	0.83	-	-	0	0.84	-	-
BI (PH) 7D	10	0.82	-	X	-	4	0.88	-	-	-	6	0.84	-	-	0	0.84	-	-
BI (PH) 16A	10	0.82	-	X	-	4	0.88	-	-	-	8	0.85	-	-	2	0.82	X	-
BI (PH) 16D	10	0.82	-	X	-	4	0.89	-	-	-	8	0.84	-	-	2	0.81	X	-

a = Neste caso, as distâncias simétricas calculadas são pequenas devido à presença de politomias na árvore;

b = As espécies proximamente relacionadas não separadas foram (*C. megacephala*, *C. bezziana* e *C. saffrana*) e (*L. caesar* e *L. illustris*);

c = Clados monofiléticos não recuperados como tal foram: *Chrysomya* = clado III de Singh et al. (2011) / *Lucilia* = *L. eximia* e *L. bazini* com as espécies restantes do gênero *Lucilia* / Subfamílias = Chrysomyinae;

d = Politomias encontradas: *Chrysomya* = entre os clados recuperados por Singh et al. (2011) / *Lucilia* = no clado composto por (*L. caesar*, *L. illustris*) / *L. bazini* / (*L. ampullacea*, *L. porphyrina*) / Gêneros de Chrysomyinae = nas relações entre *Chloroprocta* / (*Phormia*, *Protophormia*) / (*Chrysomya*, (*Hemilucilia*, *Cochliomyia*)) / Subfamílias = nas relações entre (*Luciliinae*, *Calliphorinae*) / Chrysomyinae I (*Chloroprocta*, (*Phormia*, *Protophormia*)) / Chrysomyinae II (*Cochliomyia*, (*Hemilucilia*, *Chrysomya*)).

Tabela 8. Resultado das análises de comparação de modelos baseadas em máxima verossimilhança. Um ranqueamento é fornecido para facilitar a visualização e interpretação dos resultados. ss = regiões em simples-fita; ds = regiões em dupla-fita.

Método	Modelo	Máx. LogL	Nº parâmetros	AIC	AICc	BIC	Rank
ML PhyML	TN93+G (ss e ds)	-4413,822	206	9239,644	9416,215	10174,202	-
MrBayes	GTR+G (ss e ds)	-4563,205	208	9542,41	9723,166	10486,041	2
	GTR+G (ss)	-4396,613	230	9253,226	9484,729	10296,665	1
	Doublet (GTR)+G (ds)						
BI PHASE	TN93+G (ss e ds)	-4553,494	205	9516,988	9691,492	10447,009	6
	TN93+G (ss)	-3969,241	227	8392,483	8616,535	9422,312	1
	RNA6A+G (ds)						
	TN93+G (ss)	-4194,873	234	8857,747	9099,461	9919,332	3
	RNA7A+G (ds)						
	TN93+G (ss)	-4197,87	217	8829,741	9030,19	9814,203	2
	RNA7D+G (ds)						
	TN93+G (ss)	-4329,294	227	9112,589	9336,641	10142,418	5
	RNA16A+G (ds)						
	TN93+G (ss)	-4318,143	214	9064,286	9258,012	10035,138	4
	RNA16D+G (ds)						

Por último, apesar da maioria dos critérios e métodos de seleção de modelos serem descritos como enviesados em prol de modelos mais complexos (i.e., com mais parâmetros), as análises comparativas mostradas aqui para os diferentes modelos de RNA implementados no programa PHASE mostraram que modelos mais simples, neste caso, aqueles com menor número de categorias de substituição - mas não necessariamente com menor número de parâmetros -, apresentaram um ajuste melhor aos dados sob análise (i.e. fornecem uma melhor representação do processo de evolução destas regiões) do que modelos mais complexos. Isso pode ser constatado dado que o modelo RNA6A apresentou um valor de verossimilhança significativamente maior que os modelos RNA7A e RNA7D, os quais, por sua vez, apresentaram valores maiores que os modelos com mais categorias

Tabela 9. Resultados das comparações por "Bayes factor" para as análises realizadas. (A) MrBayes; (B) PHASE. As entradas em cada célula representam duas vezes o log do "Bayes factor" em cada comparação entre os modelos M_0 e M_1 . Valores positivos indicam suporte para M_1 sobre M_0 . Comparações significativas, segundo Kass & Raftery (1995), são destacadas. ss = regiões em simples-fita; ds = regiões em dupla-fita.

A

Log L Marginal (Média harmônica)	Método (MrBayes)	M_0	
		GTR+G (ss e ds)	GTR+G (ss) Doublet (ds)
-4653,600	GTR+G (ss e ds)	0	
-4482,930	GTR+G (ss) Doublet (ds)	<u>314,340</u>	0

B

Log L Marginal (Média harmônica)	Método (PHASE)	M_0					
		TN93+G (ss e ds)	TN93+G (ss) RNA6A+G (ds)	TN93+G (ss) RNA7A+G (ds)	TN93+G (ss) RNA7D+G (ds)	TN93+G (ss) RNA16A+G (ds)	TN93+G (ss) RNA16D+G (ds)
-4604,942	TN93+G (ss e ds)	0					
-4013,288	TN93+G (ss) RNA6A+G (ds)	<u>1183,307</u>	0				
-4240,295	TN93+G (ss) RNA7A+G (ds)	<u>729,293</u>	<u>-454,013</u>	0			
-4248,381	TN93+G (ss) RNA7D+G (ds)	<u>713,121</u>	<u>-470,186</u>	-16,172	0		
-4377,743	TN93+G (ss) RNA16A+G (ds)	<u>454,397</u>	<u>-728,909</u>	<u>-274,896</u>	<u>-258,723</u>	0	
-4370,253	TN93+G (ss) RNA16D+G (ds)	<u>469,378</u>	<u>-713,928</u>	<u>-259,915</u>	<u>-243,742</u>	14,981	0

RNA16A e RNA16D. Nenhuma variação significativa foi encontrada entre modelos da mesma categoria.

4. Discussão

4.1. Análises de sequência e estrutura secundária

A utilidade da região ITS2 como um marcador molecular para análises filogenéticas tem sido bastante discutida desde os primeiros trabalhos resumidos e analisados por Hillis & Dixon (1991), segundo os quais as regiões espaçadoras do "cluster" de rDNA, como o ITS2, seriam informativos no estudo de relações filogenéticas com divergências datando do período Cenozóico (65 milhões de anos atrás - presente). Evidências disponíveis até o momento apontam para o início do período Cenozóico (65 - 35 milhões de anos) como o tempo de divergência para a maioria dos clados de Calypttratae (Kutty et al. 2010; Wiegmann et al. 2011), o que sugere que a região do ITS2 seja adequada para inferir relações evolutivas neste grupo. Apesar de suas altas taxas de evolução (2 a 3 vezes mais rápida que a região do COI - Navajas et al. 1998), o ITS2 em Calliphoridae apresentou um padrão hierarquicamente organizado de divergência em suas sequências e um nível significativamente baixo de saturação relativo ao sinal filogenético. A variabilidade intra-específica, apesar de presente no conjunto de sequências analisado para *C. hominivorax*, é baixa, como previamente mostrado para *C. incisuralis* (Nelson et al. 2008).

Uma das maiores preocupações quando do uso do ITS2 como marcador molecular para análises filogenéticas é a ocorrência comum de variação intragenômica na região, a qual pode levar à violação das premissas de homologia em análises comparativas e, assim, invalidar o sinal filogenético. A ocorrência de variação intragenômica já foi descrita para a região de espaçadores ribossomais em alguns grupos taxonômicos (Harris & Crandall 2000; Álvarez & Wendel 2003; Vollmer & Palumbi 2004; Keller et al. 2006), embora em muitos

desse casos essa variação esteja relacionada com sucessivos eventos de hibridização no grupo. A análise de múltiplas sequências clonadas para cada uma das espécies analisadas aqui (com o máximo de 90 clones analisados em *C. hominivorax*) não revelou a existência de níveis substanciais de variação intragenômica. A mesma, contudo, foi de fato detectada, principalmente na forma de variação no número de repetições de sequências mono e dinucleotídicas nas regiões ricas em A+T, em especial na região poli-T do domínio IV. Harris & Crandall (2000) mostraram que a presença de variação intragenômica associada com variação no número de repetições de sequências pode levar em muitos casos a interpretações errôneas das relações filogenéticas entre populações distintas de uma mesma espécie, mas em muitos casos não é suficiente para perturbar o sinal filogenético no nível específico. A avaliação da variabilidade da região do ITS2 em Calliphoridae, tanto nos níveis intra e inter-específico, parece corroborar essa conclusão.

O viés encontrado na composição nucleotídica do ITS2 (com um excesso no conteúdo A+T), o qual é fortemente correlacionada com a distribuição da variação de sequência em diferentes domínios estruturais e com a localização preferencial de pares G-C em regiões de dupla-fita adjacentes a regiões em simples-fita, sugere a existência de um mecanismo seletivo agindo sobre a dinâmica e evolução estrutural do ITS2. Segundo Friedrich & Tautz (1997), a existência de tal mecanismo é plausível e o mesmo deve atuar mantendo uma composição nucleotídica que forneça um balanço aceitável entre flexibilidade (gerada pelos pares A-T), a qual permite alterações conformacionais dinâmicas biologicamente necessárias, diminuindo assim eventuais restrições impostas ao reconhecimento de motivos estruturais, e estabilidade (gerada pelos pares G-C), a qual garante a formação de estruturas necessárias ao processamento (Guttel et al. 1994). Essa teoria é corroborada pela ampla ocorrência de um viés de A+T na composição nucleotídica

das regiões de rDNA em alguns grupos de Diptera (Schlötterer et al. 1994; Friedrich & Tautz 1997; Young & Coleman 2004) e também pelo papel essencial desempenhado por essa região na célula, a qual depende essencialmente da adoção da conformação estrutural correta. Estudos empíricos de termodinâmica estrutural em moléculas de RNA/DNA também corroboram essa hipótese, dado que regiões em dupla-fita ricas em G-C são consideravelmente mais estáveis que as ricas em A-U, especialmente quando os mesmos estão localizados próximos às extremidades das hélices (Xia et al. 1998; Svozil et al. 2010).

Com relação às estruturas preditas para o ITS2, embora duas diferentes conformações tenham sido preditas como possíveis, com ou sem um domínio III ramificado, tanto os dados termodinâmicos quanto as análises de predição comparativa (análise comparativa das estruturas secundárias preditas para todas as espécies analisadas) favorecem fortemente a estrutura ramificada. Contudo, ambas as estruturas estão de acordo com o modelo geral da estrutura secundária do ITS2 descrito para *D. melanogaster* (Young & Coleman 2004) e para eucariotos em geral (Joseph et al. 1999) e, dado que uma dinâmica conformacional da estrutura do ITS2 pode ocorrer *in vivo* (Côté et al. 2002), é possível que as duas estruturas sejam reais e representem dois possíveis estados conformacionais, passíveis de serem encontrados no organismo.

Por fim, a incorporação da informação de estrutura secundária no processo de alinhamento das sequências teve um efeito bastante positivo, dado que as regiões estruturalmente conservadas, normalmente localizadas nas porções proximais das regiões de hélice-alça, são prontamente alinhadas (Tillier & Collins 1995; Coleman 2003; Coleman 2007). O alinhamento dessas regiões conservadas minimiza a possibilidade de que comparações não-homólogas sejam feitas e ajuda a recuperar um sinal filogenético maior nas regiões mais variáveis, como as alças terminais. No conjunto de sequências de

Calliphoridae analisado, a maioria das sequências não foram divergentes o suficiente para impedir um alinhamento confiável nessas regiões menos conservadas. Os maiores conflitos no alinhamento surgiram quando da comparação das sequências / estruturas secundárias do ITS2 de Calliphoridae e dos grupos externos, embora, mesmo nesses casos, as porções de sequências mais conservadas presentes nos domínios de dupla-fita aumentaram significativamente a confiabilidade no processo de alinhamento.

4.2. Análises de CBCs e identificação espécie-específica

De acordo com Coleman (2000) e Coleman & Vacquier (2002), "quando a distância evolutiva acumulada entre dois táxons é suficiente para gerar no mínimo uma CBC nas posições pareadas relativamente conservadas da estrutura secundária do ITS2, os táxons diferindo nessa CBC são observados experimentalmente como sendo incapazes de realizar reprodução cruzada" ("when sufficient evolutionary distance has accumulated to produce even one CBC in the relatively conserved pairing positions of the ITS2 transcript secondary structure, taxa differing by the CBC are observed experimentally to be totally incapable of intercrossing"). Além disso, Müller et al. (2007) mostraram que, se há pelo menos uma CBC em um alinhamento de duas sequências e estruturas secundárias do ITS2, então essas sequências pertencem a duas espécies diferentes com um índice de confiança de 93%. Os resultados encontrados aqui corroboram essas assertivas a medida que as CBCs encontradas ocorreram apenas entre sequências de diferentes espécies e nunca dentro delas. No que concerne a afirmação sobre isolamento reprodutivo, apesar de dados sobre cruzamentos entre espécies de Calliphoridae serem raros, dois pares de espécies amostradas aqui já foram relatadas como sendo capazes de produzir híbridos: *C. albiceps* / *C. rufifacies* (Ullerich 1963) e *L. cuprina* / *L. sericata* (Mackerras 1933). Em ambos os casos, nenhuma CBC foi encontrada entre essas espécies.

É interessante notar, no entanto, que na comparação das sequências e estruturas secundárias da região do ITS2 em duas espécies distintas, não necessariamente espera-se encontrar uma ou mais CBCs. No conjunto de dados analisado aqui, não foram encontradas CBCs nos alinhamentos de pares de espécies congênicas em 33% dos casos (40 de um total de 120 pares de espécies congênicas). Essas espécies, contudo, ainda podem ser diferenciadas com base em outras características do ITS2, como a ocorrência de hemi-CBCs e substituições nas regiões em simples-fita da molécula. Sendo assim, como previamente mostrado por Nelson et al. (2008) e Song et al. (2008b), a região do ITS2 pode ser usada em análises de identificação espécie-específica em Calliphoridae mesmo na ausência de CBCs. O potencial do ITS2 para identificação taxonômica já foi analisado e confirmado em muitos outros táxons, onde o mesmo mostrou uma performance equivalente ou até mesmo superior quando comparado a outras moléculas normalmente utilizadas como "DNA barcodes" - códigos de barra de DNA - (Sass et al. 2007; Seifert et al. 2007).

4.3. Análises filogenéticas e avaliação das árvores inferidas com a região ITS2

Muitas das relações recuperadas entre os Calliphoridae analisados usando a região ITS2 corroboraram algumas relações previamente descritas para o grupo, apesar de algumas diferenças terem sido encontradas e poderem estar relacionadas a diferenças de amostragem de táxons e regiões gênicas (marcadores). Por exemplo, as relações entre os gêneros de Chrysomyinae encontradas aqui diferiram significativamente das recuperadas por Singh & Wells (2011), os quais amostraram dez dos doze gêneros conhecidos desta subfamília, enquanto nas análises com o ITS2 foram incluídos 6 gêneros. As incongruências encontradas também podem ser devidas a diferenças nas regiões gênicas amostradas, dado que as análises de Singh & Wells (2011) se basearam na análise combinada de duas regiões codificadoras de proteína, COI e CPS, enquanto as filogenias

obtidas aqui foram inferidas com base em uma única região gênica não codificante. Apesar de a caracterização de regiões gênicas adicionais e a inclusão de mais táxons poderem levar a um aumento da resolução das análises filogenéticas, os resultados encontrados para a região do ITS2 podem servir como guia para estudos mais amplos de inferência filogenética na família Calliphoridae usando, além da informação de sequência e estrutura dessa molécula, outros marcadores de origem e taxas de evolução diferentes.

Neste contexto, a maioria dos métodos parecem ter uma performance aceitável no estudo da região do ITS2, embora os métodos de MP e de NJ, o último considerando apenas informações da sequência primária, podem não ser aconselháveis devido à baixa resolução e baixo suporte dado a grupos relativamente bem estabelecidos. Por sua vez, os métodos de NJ, considerando ou não informações estruturais nas análises, são aconselháveis apenas quando as distâncias genéticas entre os grupos em estudo não são muito divergentes, dado que distâncias anormalmente baixas ou altas podem levar a erros no processo de reconstrução de árvores, como ocorreu para as espécies de *Lucilia* nas análises aqui mostradas. Este comportamento é, contudo, esperado, já que o método de NJ se baseia apenas no cálculo de distâncias evolutivas, com muita informação sendo perdida no processo de compressão da informação contida nas sequências em distâncias (Holder & Lewis 2003; Felsenstein 2004).

Para os métodos estatísticos de inferência filogenética, isto é, ML e BI, ambos parecem ter uma performance geral aceitável. Apesar do fato que para alguns clados recuperados nas análises Bayesianas os valores de suporte dados pelas probabilidades *a posteriori* sejam maiores que os valores de suporte de "bootstrap" estimados pela análise de ML, isto não necessariamente implica em filogenias mais bem resolvidas, dado que os valores de probabilidade *a posteriori* são tidos como super estimativas do real suporte para

os clados (Suzuki et al. 2002; Yang 2007) e, portanto, isso não se caracteriza como uma propriedade desejável dos métodos de inferência filogenética. As árvores Bayesianas inferidas usando o programa MrBayes (árvores consenso de regra da maioria - 50%) apresentaram muitas politomias, fato este que pode estar relacionado com a incerteza em torno da inferência tanto da topologia quanto dos parâmetros, a qual é incorporada pelo programa durante o processo de estimação das árvores. Dado que para a região do ITS2 o número de sítios é pequeno e o número de parâmetros é grande nos modelos mais complexos avaliados, isso pode levar a uma grande incerteza durante o processo de inferência filogenética (Nylander 2004).

Por fim, parece haver um consenso de que o uso de modelos que incorporam informações estruturais nas análises em geral levam a uma melhora significativa no processo de inferência de árvores filogenéticas (Telford et al. 2005; Letsch & Kjer 2011), uma vez que eles representam uma forma mais realística de modelar a história evolutiva de moléculas estruturalmente funcionais (Dixon & Hillis 1993; Schöniger & von Haeseler 1994; Muse 1995; Rzhetsky 1995; Tillier & Collins 1995; Gutell 1996; Savill et al. 2001). Apesar disso, o uso de modelos que consideram informações estruturais nas análises Bayesianas usando o ITS2 resultaram apenas em leves melhoras em termos de topologia e suporte geral dos clados inferidos. As diferenças foram significativas, contudo, nas filogenias inferidas como um todo, como mostrado pelo significativo aumento nos valores de verossimilhança. Este fato deve-se, provavelmente, a uma melhora na estimativa das taxas de evolução e dos comprimentos de ramo nas árvores inferidas. Isso destaca a importância do uso desses modelos mais complexos quando das estimativas de processos evolutivos nos quais mais do que a topologia é necessária. Nesses casos, o uso de modelos mais simples que levam em consideração informações estruturais, como o modelo RNA6A,

podem ser favorecidos, como previamente mostrado por Letsch & Kjer (2011), dado que eles não possuem uma grande quantidade de parâmetros desnecessários.

5. Conclusões

O uso do ITS2 como um marcador molecular para inferência filogenética tem sido bastante discutido ultimamente e seu uso em análises genético-evolutivas em Arthropoda vem apresentando um crescimento considerável nos últimos cinco anos. Até o momento, há mais de 350 artigos publicados disponíveis em bases de dados de publicações descrevendo o uso da região do ITS2 em estudos de identificação de espécies, análises filogenéticas e de genética de populações apenas em artrópodes. Apesar dos estudos de identificação perfazerem a maioria dos artigos publicados, análises filogenéticas com esse marcador compreendem atualmente mais de 30% desses artigos. Contudo, menos de 10% desses estudos evolutivos incluem uma etapa de predição da estrutura secundária do ITS2 e sua incorporação nas análises filogenéticas. Com raras exceções (e.g. Wiemers et al. 2009), a informação estrutural dessa molécula tem sido utilizada apenas para guiar o processo de alinhamento e auxiliar na discussão final dos resultados.

Neste sentido, o grande número de árvores filogenéticas inferidas neste estudo teve como intenção prover uma avaliação profunda da performance do ITS2 como um marcador molecular para estudos filogenéticos em Calliphoridae usando uma gama de métodos de inferência e modelos de substituição. Futuras análises usando esta região podem se beneficiar dos resultados aqui obtidos e discutidos, incluindo a relativa melhora no processo de inferência de filogenias quando da consideração de informações estruturais nas análises. Contudo, com o objetivo de inferir árvores filogenéticas que mais se assemelham à provável árvore real, é aconselhável o uso de uma variedade maior de marcadores moleculares, compreendendo diferentes padrões de evolução de sequência. Em grupos nos

quais parece ter ocorrido uma rápida divergência das linhagens, como o grupo dos Calyptratae, a região do ITS2 pode ser útil nas análises de reconstrução filogenética nos níveis específico e genérico.

Os resultados aqui descritos para a evolução da sequência e da estrutura secundária da região do ITS2 podem contribuir para um melhor entendimento da evolução de regiões não-codificantes estruturais em Diptera e, em uma maior escala, em diferentes grupos de Arthropoda.

6. Referências

- Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Mol Phylogenet Evol* 29(3):417-434
- Amendt J, Krettek R, Zehner R (2004) Forensic entomology. *Naturwissenschaften* 91:51-65
- Brown KR (1979) Multivariate assessment of phenetic relationships within the tribe Luciliini. *Aust J Zool* 27:465-477
- Chen WY, Hung TH, Shiao SF (2004) Molecular identification of forensically important blow fly species (Diptera: Calliphoridae) in Taiwan. *J Med Entomol* 41:47-57
- Coleman A W (2000) The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. *Protist* 151:1-9
- Coleman AW (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet* 19:370-375
- Coleman AW (2007) Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Res* 35:3322-3329
- Coleman AW (2009) Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. *Mol Phylogenet Evol* 50:197-203
- Coleman AW, Vacquier VD (2002) Exploring the phylogenetic utility of its sequences for animals: a test case for abalone (haliotis). *J Mol Evol* 54:246-257
- Côté CA, Greer CL, Peculis BA (2002) Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. *RNA* 8:786-797
- Darty K, Denise A, Ponty Y (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25:1974-1975

- de Azeredo-Espin AM, Lessinger AC (2006) Genetic approaches for studying myiasis-causing flies: molecular markers and mitochondrial genomics. *Genetica* 126:111-131
- Dixon MT, Hillis DM (1993) Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. *Mol Biol Evol* 10: 256-267
- Elder JF Jr, Turner BJ (1995) Concerted evolution of repetitive DNA sequences. *Q Rev Biol* 70:297-320
- Felsenstein J (2004) *Inferring Phylogenies*. Sinauer Associates, Sunderland, Massachusetts
- Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
- Friedrich M, Tautz D (1997) An episodic change of rDNA nucleotide substitution rate has occurred during the emergence of the insect order Diptera. *Mol Biol Evol* 14: 644-653
- Geerlings TH, Vos JC, Raué HA (2000) The final step in the formation of 25S rRNA in *Saccharomyces cerevisiae* is performed by 5'→3' exonucleases. *RNA* 6:1698-1703
- Goloboff PA, Farris JS, Nixon K (2003) TNT: Tree analysis using New Technology. Version 1.0, version Beta test v. 0.2. Program and documentation available at <http://www.zmuc.dk/public/phylogeny/TNT/>
- Gowri-Shankar V, Jow H (2006) PHASE: a software package for phylogenetics and sequence evolution. Available at: <http://www.bioinf.manchester.ac.uk/resources/phase>
- Guimarães JH (1977) A systematic revision of the Mesembrinellidae, stat. nov. (Diptera, Cyclorrhapha). *Arquivos de Zoologia* 29:1-109
- Guimarães JH, Papavero N, Prado AP (1983) As míases na região Neotropical: identificação, biologia e bibliografia. *Rev Bras Zool* 1:239-416
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696-704
- Gutell RR (1996) Comparative sequence analysis and the structure of 16S and 23S rRNA. p. 111–128. In *Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis*. (R. A. Zimmermann & A. E. Dahlberg, eds.). CRC Press, New York
- Gutell RR, Larsen N, Woese CR (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Rev* 58:10-26
- Harris DJ, Crandall KA (2000) Intragenomic variation within ITS1 and ITS2 of freshwater crayfishes (Decapoda: Cambaridae): implications for phylogenetic and microsatellite studies. *Mol Biol Evol* 17(2):284-291
- Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *Q Rev Biol* 66:411-453

- Holder H, Lewis PO (2003) Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet* 4:75-284
- Infante MEV, Azeredo-Espin AML (1995) Genetic variability in mtDNA of *Cochliomyia hominivorax* (Diptera: Calliphoridae) from Brazil. *Biochem Genet* 33:237-256
- Jordan BR, Jourdan R, Jacq B (1976) Late steps in the maturation of *Drosophila* 26 S ribosomal RNA: generation of 5.8 S and 2 S RNAs by cleavages occurring in the cytoplasm. *J Mol Biol* 101:85-105
- Joseph N, Krauskopf E, Vera MI, Michot B (1999) Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. *RNA* 8:786-797
- Junqueira AC, Lessinger AC, Torres TT, da Silva FR, Vettore AL, Arruda P, Azeredo-Espin AM (2004) The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). *Gene* 339:7-15
- Kass RE, Raftery AE (1995) Bayes factor. *J Am Stat Assoc* 90:773-795
- Keller I, Chintauan-Marquier IC, Veltsos P, Nichols RA (2006) Ribosomal DNA in the grasshopper *Podisma pedestris*: escape from concerted evolution. *Genetics* 174(2):863-874
- Koetschan C, Förster F, Keller A, Schleicher T, Ruderisch B, Schwarz R, Müller T, Wolf M, Schultz J (2010) The ITS2 Database III: sequences and structures for phylogeny. *Nucleic Acids Res* 38(Database issue):D275-279
- Kurahashi H (1966) Studies on the calyptratae muscoid flies from Japan. V. Revision of the tribe Luciliini (Diptera: Calliphoridae). *Pac Insects* 11:105-124
- Kutty SN, Pape T, Wiegmann BM, Meier R (2010) Molecular phylogeny of the Calyptratae (Diptera: Cyclorrhapha) with an emphasis on the superfamily Oestroidea and the position of Mystacinobiidae and McAlpine's fly. *Syst Entomol* 35:614-635
- Lessinger AC, Junqueira ACM, Conte FF, Azeredo-Espin AM (2004) Analysis of a conserved duplicated tRNA gene in the mitochondrial genome of blowflies. *Gene* 339:1-6
- Lessinger AC, Junqueira ACM, Lemos TA, Kemper EL, da Silva FR, Vettore AL, Arruda P, Azeredo-Espin AM (2000) The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Mol Biol* 9(5):521-529
- Letsch HO, Kjer KM (2011) Potential pitfalls of modelling ribosomal RNA data in phylogenetic tree reconstruction: Evidence from case studies in the Metazoa. *BMC Evol Biol* 11:146
- Liao D (1999) Concerted evolution: molecular mechanisms and biological implications. *Am J Hum Genet* 64:24-30
- Mackerras MJ (1933) Observations on the life-histories, nutritional requirements and fecundity of blowflies. *Bull Entomol Res* 24:353-362

- McAlpine JF (1989) Manual of nearctic Diptera 3. Agriculture Canada Monograph 32. Canadian Gov. Publ. Center, Quebec, Canada
- Mitchell P, Petfalski E, Shevchenko A, Mann M, Tollervey D (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'=>5' exoribonucleases. *Cell* 91:457-466
- Mitchell P, Petfalski E, Tollervey D (1996) The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes Dev* 10:502-513
- Müller T, Philippi N, Dandekar T, Schultz J, Wolf M (2007) Distinguishing species. *RNA* 13:1469-1472
- Muse SV (1995) Evolutionary analyses of DNA sequences subject to constraints of secondary structure. *Genetics* 139:1429-1439
- Navajas M, Lagnel J, Gutierrez J, Boursot P (1998) Species wide homogeneity of nuclear ribosomal ITS2 sequences in the spider mite *Tetranychus urticae* contrasts with excessive mitochondrial COI polymorphism. *J Hered* 80:742-752
- Nelson LA, Wallman JF, Dowton M (2007) Using COI barcodes to identify forensically and medically important blowflies. *Med Vet Entomol* 21:44-52
- Nelson LA, Wallman JF, Dowton M (2008) Identification of forensically important *Chrysomya* (Diptera: Calliphoridae) species using the second ribosomal internal transcribed spacer (ITS2). *Forensic Sci Int* 177:238-247
- Newton MA, Raftery AE (1994) Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *J R Stat Soc B Met* 56:3-48
- Nylander JA (2004) MrAIC.pl. Program distributed by the author. Evolutionary Biology Centre, Uppsala University
- Nylander JA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL (2004) Bayesian phylogenetic analysis of combined data. *Syst Biol* 53(1):47-67
- Pape T, Thompson FC (editors) (2010) *Systema Dipteriorum*, Version 1.0. <http://www.diptera.org/>, accessed on 2 July 2011
- R Development Core Team (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>
- Robinson DR, Foulds LR (1981) Comparison of phylogenetic trees. *Math Biosci* 53:131-147
- Rognes K (1991) Blowflies (Diptera, Calliphoridae) of Fennoscandia and Denmark. *Fauna Entomologica Scandinavica* 24:1-272

- Rognes K (1997) The Calliphoridae (blowflies) (Diptera. Oestroidea) are not a monophyletic group. *Cladistics* 13:27-66
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574
- Rzhetsky A (1995) Estimating substitution rates in ribosomal RNA genes. *Genetics* 141: 771-783
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning - A Laboratory Manual*, 2nd Edition. Cold Spring Harbour Laboratory Press, New York
- Sass C, Little DP, Stevenson DW, Specht CD (2007) DNA barcoding in the cycadales: testing the potential of proposed barcoding markers for species identification of cycads. *PLoS ONE* 2: e1154
- Savill NJ, Hoyle DC, Higgs PG (2001) RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods. *Genetics* 157:399-411
- Schlötterer C, Hauser MT, von Haeseler A, Tautz D (1994) Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. *Mol Biol Evol* 11:513-522
- Schöniger M, von Haeseler A (1994) A stochastic model for the evolution of autocorrelated DNA sequences. *Mol Phylogenet Evol* 3:240-247
- Schultz J, Maisel S, Gerlach D, Müller T, Wolf M (2005) A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA* 11:361-364
- Schultz J, Müller T, Achtziger M, Seibel PN, Dandekar T, Wolf M (2006) The internal transcribed spacer 2 database: a web server for (not only) low level phylogenetic analyses. *Nucleic Acids Res* 34(Web Server issue):W704-707
- Seibel PN, Müller T, Dandekar T, Schultz J, Wolf M (2006) 4SALE: a tool for synchronous RNA sequence and secondary structure alignment and editing. *BMC Bioinformatics* 7:498
- Seifert KA, Samson RA, Dewaard JR, Houbraken J, Lévesque CA, Moncalvo JM, Louis-Seize G, Hebert PD (2007) Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. *Proc Natl Acad Sci USA* 104:3901-3906.
- Selig C, Wolf M, Müller T, Dandekar T, Schultz J (2008) The ITS2 Database II: homology modelling RNA structure for molecular systematics. *Nucleic Acids Res* 36(Database issue):D377-380
- Shepard JH (1971) A phenetic analysis of the Luciliini (Diptera: Calliphoridae). *Syst Zool* 20:223-232
- Shewell G (1987) 106 Calliphoridae. In "Manual of Nearctic Diptera." (J. F. McAlpine, Ed.), Vol. 2. Research Branch, Agriculture Canada, Monograph 28, pp 1133-1145

- Singh B, Kurahashi H, Wells JD (2011) Molecular phylogeny of the blowfly genus *Chrysomya*. *Med Vet Entomol* 25(2):126-134
- Singh B, Wells JD (2011) Chrysomyinae (Diptera: Calliphoridae) is monophyletic: a molecular systematics analysis. *Syst Entomol* 36(3):415-420
- Song Z, Wang X, Liang G (2008a) Molecular Evolution and Phylogenetic Utility of the Internal Transcribed Spacer 2 (ITS2) in Calyptratae (Diptera: Brachycera). *J Mol Evol* 67:448-464
- Song Z, Wang X, Liang G (2008b) Species identification of some common necrophagous flies in Guangdong province, southern China based on the rDNA internal transcribed spacer 2 (ITS2). *Forensic Sci Int* 175:17-22
- Stevens J, Wall R (1996) Classification of the genus *Lucilia* (Diptera: Calliphoridae): a preliminary parsimony analysis. *J Nat Hist* 30:1087-1094
- Stevens J, Wall R (1997) The evolution of ectoparasitism in the genus *Lucilia* (Diptera: Calliphoridae). *Int J Parasitol* 27:51-59
- Stevens J, Wall R (2001) Genetic relationships between blowflies (Calliphoridae) of forensic importance. *Forensic Sci Int* 120:116-123
- Stevens JR (2003) The evolution of myiasis in blowflies (Calliphoridae). *Int J Parasitol* 33:1105-1113
- Stevens JR, Wallman JF (2006) The evolution of myiasis in humans and other animals in the Old and New Worlds (part I): phylogenetic analyses. *Trends Parasitol* 22:129-136
- Stevens JR, Wallman JF, Otranto D, Wall R, Pape T (2006) The evolution of myiasis in humans and other animals in the Old and New Worlds (part II): biological and life-history studies. *Trends Parasitol* 22:181-188
- Suzuki Y, Glazko GV, Nei M (2002) Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc Natl Acad Sci U S A* 99(25):16138-16143
- Svozil D, Hobza P, Sponer J (2010) Comparison of intrinsic stacking energies of ten unique dinucleotide steps in A-RNA and B-DNA duplexes. Can we determine correct order of stability by quantum-chemical calculations? *J Phys Chem B* 114:1191-203; Erratum in: *J Phys Chem B* (2010) 114:2547
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599
- Tautz D, Hancock JM, Webb DA, Tautz C, Dover GA (1988) Complete sequences of the rRNA genes of *Drosophila melanogaster*. *Mol Biol Evol* 5:366-376

- Telford MJ, Wise MJ, Gowri-Shankar V (2005) Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the bilateria. *Mol Biol Evol* 22:1129-1136
- Tillier E, Collins R (1995) Neighbor-Joining and maximum likelihood with RNA sequences: addressing the inter-dependence of sites. *Mol Biol Evol* 12:7–15.
- Ullerich FH (1963) Geschlechtschromosomen und Geschlechtsbestimmung bei einigen Calliphorinen (Calliphoridae, Diptera). *Chromosoma* 14:45-110
- van der Sande CA, Kwa M, van Nues RW, van Heerikhuizen H, Raué HA, Planta RJ (1992) Functional analysis of internal transcribed spacer 2 of *Saccharomyces cerevisiae* ribosomal DNA. *J Mol Biol* 223:899-910
- Veldman GM, Klootwijk J, van Heerikhuizen H, Planta RJ (1981) The nucleotide sequence of the intergenic region between the 5.8S and 26S rRNA genes of the yeast ribosomal RNA operon. Possible implications for the interaction between 5.8S and 26S rRNA and the processing of the primary transcript. *Nucleic Acids Res* 9:4847-4862
- Vollmer SV, Palumbi SR (2004) Testing the utility of the internally transcribed spacer sequences in coral phylogenetics. *Mol Ecol* 13(9):2763-2772
- Wallman JF, Donnellan SC (2001) The utility of mitochondrial DNA sequences for the identification of forensically important blowflies (Diptera: Calliphoridae) in southeastern Australia. *Forensic Sci Int* 120:60-67
- Wells JD, Sperling FA (2001) DNA-based identification of forensically important Chrysomyinae (Diptera: Calliphoridae). *Forensic Sci Int* 120:110-115
- Wells JD, Wall R, Stevens JR (2007) Phylogenetic analysis of forensically important *Lucilia* flies based on cytochrome oxidase I sequence: a cautionary tale for forensic species determination. *Int J Legal Med* 121(3):229-233
- Wells JD, Williams DW (2007) Validation of a DNA-based method for identifying Chrysomyinae (Diptera: Calliphoridae) used in a death investigation. *Int J Legal Med* 121:1-8
- Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim JW, Lambkin C, Bertone MA, Cassel BK, Bayless KM, Heimberg AM, Wheeler BM, Peterson KJ, Pape T, Sinclair BJ, Skevington JH, Blagoderov V, Caravas J, Kutty SN, Schmidt-Ott U, Kampmeier GE, Thompson FC, Grimaldi DA, Beckenbach AT, Courtney GC, Friedrich M, Meier R, Yeates DK (2011) Episodic radiations in the fly tree of life. *Proc Natl Acad Sci U S A* 108(14):5690-5695

- Wiemers M, Keller A, Wolf M (2009) ITS2 secondary structure improves phylogeny estimation in a radiation of blue butterflies of the subgenus *Agrodiaetus* (Lepidoptera: Lycaenidae: Polyommatus). BMC Evol Biol 9:300
- Wolf M, Ruderisch B, Dandekar T, Schultz J, Müller T (2008) ProfDistS: (profile-) distance based phylogeny on sequence-structure alignments. Bioinformatics 24:2401-2402
- Xia T, SantaLucia JJr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. Biochemistry 37:14719-14735
- Xia X, Lemey P (2009) Assessing substitution saturation with DAMBE. Pp. 615-630 in Philippe Lemey, Marco Salemi and Anne-Mieke Vandamme, eds. The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny. 2nd edition Cambridge University Press.
- Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. J Hered 92(4):371-373
- Xia X, Xie Z, Salemi M, Chen L, Wang Y (2003). An index of substitution saturation and its application. Mol Phylogenet Evol 26:1-7
- Yang Z (2007) Fair-balance paradox, star-tree paradox and Bayesian phylogenetics. Mol Biol Evol 24(8):1639-1655
- Young I, Coleman AW (2004) The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a Drosophila example. Mol Phylogenet Evol 30:236-242
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406-3415
- Zumpt F (1965) Myiasis in man and animals in the Old World. London, Butterworths. p. 267

Anexo A

Tabela A1. Número total de pares de bases encontrados nas estruturas secundárias do ITS2 para cada uma das espécies analisadas. O número e frequência dos pares G-C, G-U e A-U são mostrados. Nos casos em que há mais de uma sequência por espécie, apenas uma delas foi sorteada e analisada.

Espécies	Nº pares	Nº pares (por tipo)			Frequência dos pares (por tipo)		
		G-C	G-U	A-U	G-C	G-U	A-U
<i>Calliphora vomitoria</i>	108	18	7	83	0,166666667	0,064814815	0,768518519
<i>Aldrichina grahami</i>	108	20	7	81	0,185185185	0,064814815	0,75
<i>Calliphora vicina</i>	108	15	10	83	0,138888889	0,092592593	0,768518519
<i>Lucilia caesar</i>	108	17	7	84	0,157407407	0,064814815	0,777777778
<i>Lucilia illustris</i>	103	16	8	79	0,155339806	0,077669903	0,766990291
<i>Hemipyrellia ligurriensis</i>	99	15	5	79	0,151515152	0,050505051	0,797979798
<i>Lucilia cuprina</i>	98	18	3	77	0,183673469	0,030612245	0,785714286
<i>Lucilia sericata</i>	100	18	6	76	0,18	0,06	0,76
<i>Lucilia ampullacea</i>	90	16	4	70	0,177777778	0,044444444	0,777777778
<i>Lucilia porphyria</i>	94	17	3	74	0,180851064	0,031914894	0,787234043
<i>Lucilia bazini</i>	88	16	6	66	0,181818182	0,068181818	0,75
<i>Lucilia eximia</i>	109	17	5	87	0,155963303	0,04587156	0,798165138
<i>Chrysomya megacephala</i>	104	17	11	76	0,163461538	0,105769231	0,730769231
<i>Chrysomya bezziana</i>	103	17	10	76	0,165048544	0,097087379	0,737864078
<i>Chrysomya saffrana</i>	106	17	11	78	0,160377358	0,103773585	0,735849057
<i>Chrysomya pinguis</i>	103	17	10	76	0,165048544	0,097087379	0,737864078
<i>Chrysomya latifrons</i>	106	18	14	74	0,169811321	0,132075472	0,698113208
<i>Chrysomya semimetallica</i>	108	17	15	76	0,157407407	0,138888889	0,703703704
<i>Chrysomya albiceps</i>	111	18	5	88	0,162162162	0,045045045	0,792792793
<i>Chrysomya rufifacies</i>	108	18	7	83	0,166666667	0,064814815	0,768518519
<i>Chrysomya incisuralis</i>	113	18	6	89	0,159292035	0,053097345	0,787610619
<i>Chrysomya nigripes</i>	110	17	13	80	0,154545455	0,118181818	0,727272727
<i>Chrysomya putoria</i>	102	18	7	77	0,176470588	0,068627451	0,754901961
<i>Chrysomya flavifrons</i>	109	17	8	84	0,155963303	0,073394495	0,770642202
<i>Chrysomya varipes</i>	104	17	9	78	0,163461538	0,086538462	0,75
<i>Cochliomyia hominivorax</i>	124	18	9	97	0,14516129	0,072580645	0,782258065
<i>Cochliomyia macellaria</i>	125	17	10	98	0,136	0,08	0,784
<i>Hemilucilia segmentaria</i>	112	17	8	87	0,151785714	0,071428571	0,776785714
<i>Chloroprocta idioidea</i>	109	17	5	87	0,155963303	0,04587156	0,798165138
<i>Phormia regina</i>	113	16	10	87	0,14159292	0,088495575	0,769911504
<i>Protophormia terraenovae</i>	108	16	6	86	0,148148148	0,055555556	0,796296296
<i>Mesembrinella peregrina</i>	95	17	4	74	0,178947368	0,042105263	0,778947368
<i>Mesembrinella bellardiana I</i>	108	24	7	77	0,222222222	0,064814815	0,712962963
<i>Mesembrinella bellardiana II</i>	99	20	6	73	0,202020202	0,060606061	0,737373737
<i>Stomoxys calcitrans</i>	96	18	10	68	0,1875	0,104166667	0,708333333
<i>Musca domestica</i>	115	26	9	80	0,226086957	0,07826087	0,695652174
<i>Haematobia irritans</i>	107	21	12	74	0,196261682	0,112149533	0,691588785

Tabela A2. Número total de pares de bases encontrados nas áreas de interface entre regiões em simples e dupla-fita (junções, "bojos", alças internas e terminais). Apenas um par de bases na vizinhança dessas regiões foi considerado. Os valores observados e esperados para os pares G-C, G-U e A-U nessas regiões são mostrados, incluindo os resultados de um teste qui-quadrado comparando-os. Pares G-U e A-U foram agrupados na mesma categoria dado que ambos são menos estáveis que os pares G-C e porque as substituições $UG \leftrightarrow UA$ e $GU \leftrightarrow AU$ são as mais comuns entre as hemi-CBCs observadas (Tabela 6). Valores em negrito indicam resultados significativos para $p < 0,05$ (**) e $p < 0,10$ (*).

Espécies	Nº pares interface	Total de pares (observado)				Total de pares (esperado)				Qui-quadrado			
		GC	GU	AU	GU+AU	GC	GU	AU	GU+AU	GC	GU+AU	Total	p-valor (1 g.l.)
<i>Calliphora vomitoria</i>	32	10	0	22	22	5,33333	2,07407	24,5926	26,6667	4,08333	0,81667	4,9	p<0,05 **
<i>Aldrichina grahami</i>	40	12	3	25	28	7,40741	2,59259	30	32,5926	2,84741	0,64714	3,49455	p<0,07 *
<i>Calliphora vicina</i>	34	8	2	24	26	4,72222	3,14815	26,1296	29,2778	2,27516	0,36696	2,64213	p<0,15
<i>Lucilia caesar</i>	40	11	2	27	29	6,2963	2,59259	31,1111	33,7037	3,51394	0,65645	4,17039	p<0,05 **
<i>Lucilia illustris</i>	37	10	2	25	27	5,74757	2,87379	28,3786	31,2524	3,14622	0,57862	3,72484	p<0,06 *
<i>Hemipyrellia ligurriens</i>	38	7	2	29	31	5,75758	1,91919	30,3232	32,2424	0,2681	0,04788	0,31598	p<0,60
<i>Lucilia cuprina</i>	34	9	1	24	25	6,2449	1,04082	26,7143	27,7551	1,21549	0,27348	1,48897	p<0,30
<i>Lucilia sericata</i>	34	10	1	23	24	6,12	2,04	25,84	27,88	2,45987	0,53997	2,99984	p<0,09 *
<i>Lucilia ampullacea</i>	34	11	2	21	23	6,04444	1,51111	26,4444	27,9556	4,06283	0,87845	4,94128	p<0,03 **
<i>Lucilia porphyrina</i>	36	12	1	23	24	6,51064	1,14894	28,3404	29,4894	4,62829	1,02183	5,65011	p<0,02 **
<i>Lucilia bazini</i>	28	10	1	17	18	5,09091	1,90909	21	22,9091	4,73377	1,05195	5,78571	p<0,02 **
<i>Lucilia eximia</i>	40	9	0	31	31	6,23853	1,83486	31,9266	33,7615	1,22236	0,22587	1,44823	p<0,30
<i>Chrysomya megacephala</i>	32	9	4	19	23	5,23077	3,38462	23,3846	26,7692	2,71606	0,53073	3,24679	p<0,08 *
<i>Chrysomya bezziana</i>	32	9	4	19	23	5,28155	3,1068	23,6117	26,7184	2,61795	0,5175	3,13545	p<0,08 *
<i>Chrysomya saffrana</i>	32	9	3	20	23	5,13208	3,32075	23,5472	26,8679	2,91516	0,55683	3,47199	p<0,07 *
<i>Chrysomya pinguis</i>	30	9	1	20	21	4,95146	2,91262	22,1359	25,0485	3,31028	0,65436	3,96464	p<0,05 **
<i>Chrysomya latifrons</i>	28	9	1	18	19	4,75472	3,69811	19,5472	23,2453	3,79043	0,77532	4,56575	p<0,04 **
<i>Chrysomya semimetallica</i>	30	9	2	19	21	4,72222	4,16667	21,1111	25,2778	3,87516	0,72393	4,5991	p<0,04 **
<i>Chrysomya albiceps</i>	36	10	1	25	26	5,83784	1,62162	28,5405	30,1622	2,96747	0,57435	3,54182	p<0,06 *

Tabela A2. (Continuação).

Espécies	Nº pares interface	Total de pares (observado)				Total de pares (esperado)				Qui-quadrado			
		GC	GU	AU	GU+AU	GC	GU	AU	GU+AU	GC	GU+AU	Total	p-valor (1 g.l.)
<i>Chrysomya rufifacies</i>	37	10	3	24	27	6,16667	2,39815	28,4352	30,8333	2,38288	0,47658	2,85946	p<0,10 *
<i>Chrysomya incisuralis</i>	40	11	1	28	29	6,37168	2,12389	31,5044	33,6283	3,36196	0,637	3,99896	p<0,05 **
<i>Chrysomya nigripes</i>	38	8	3	27	30	5,87273	4,49091	27,6364	32,1273	0,77056	0,14086	0,91142	p<0,40
<i>Chrysomya putoria</i>	38	11	4	23	27	6,70588	2,60784	28,6863	31,2941	2,74974	0,58923	3,33897	p<0,07 *
<i>Chrysomya flavifrons</i>	34	9	1	24	25	5,30275	2,49541	26,2018	28,6972	2,57784	0,47634	3,05418	p<0,09 *
<i>Chrysomya varipes</i>	32	10	2	20	22	5,23077	2,76923	24	26,7692	4,34842	0,84969	5,19811	p<0,03 **
<i>Cochliomyia hominivorax</i>	34	7	1	26	27	4,93548	2,46774	26,5968	29,0645	0,86359	0,14665	1,01024	p<0,40
<i>Cochliomyia macellaria</i>	34	7	1	26	27	4,624	2,72	26,656	29,376	1,22089	0,19218	1,41306	p<0,30
<i>Hemilucilia segmentaria</i>	36	9	1	26	27	5,46429	2,57143	27,9643	30,5357	2,28782	0,4094	2,69721	p<0,09 *
<i>Chloroprocta idioidea</i>	38	9	1	28	29	5,92661	1,74312	30,3303	32,0734	1,59379	0,2945	1,88829	p<0,20
<i>Phormia regina</i>	34	9	1	24	25	4,81416	3,00885	26,177	29,1858	3,63953	0,60033	4,23986	p<0,04 **
<i>Protophormia terraenovae</i>	36	8	2	26	28	5,33333	2	28,6667	30,6667	1,33333	0,23188	1,56522	p<0,30
<i>Mesembrinella peregrina</i>	40	10	0	30	30	7,15789	1,68421	31,1579	32,8421	1,12848	0,24595	1,37443	p<0,30
<i>Mesembrinella bellardiana</i> I	38	8	5	25	30	8,44444	2,46296	27,0926	29,5556	0,02339	0,00668	0,03008	p<0,90
<i>Mesembrinella bellardiana</i> II	40	8	4	28	32	8,08081	2,42424	29,4949	31,9192	0,00081	0,0002	0,00101	p<0,99
<i>Stomoxys calcitrans</i>	34	8	7	19	26	6,375	3,54167	24,0833	27,625	0,41422	0,09559	0,5098	p<0,50
<i>Musca domestica</i>	38	12	3	23	26	8,5913	2,97391	26,4348	29,4087	1,35244	0,39509	1,74753	p<0,40
<i>Haematobia irritans</i>	38	10	5	23	28	7,45794	4,26168	26,2804	30,5421	0,86647	0,21158	1,07804	p<0,80

Tabela A3. Número total de pares de bases encontrados nas áreas de interface entre regiões em simples e dupla-fita (junções, "bojos", alças internas e terminais). Dois pares de bases na vizinhança dessas regiões foi considerado. Os valores observados e esperados para os pares G-C, G-U e A-U nessas regiões são mostrados, incluindo os resultados de um teste qui-quadrado comparando-os. Pares G-U e A-U foram agrupados na mesma categoria dado que ambos são menos estáveis que os pares G-C e porque as substituições UG↔UA e GU↔AU são as mais comuns entre as hemi-CBCs observadas (Tabela 6). Valores em negrito indicam resultados significativos para $p < 0,05$ (**) e $p < 0,10$ (*).

Espécies	Nº pares interface	Total de pares (observado)				Total de pares (esperado)				Qui-quadrado			
		GC	GU	AU	GU+AU	GC	GU	AU	GU+AU	GC	GU+AU	Total	p-valor (1 g.l.)
<i>Calliphora vomitoria</i>	62	17	0	45	45	10,3333	4,01852	47,6481	51,6667	4,30108	0,86022	5,16129	p<0,025 **
<i>Aldrichina grahami</i>	75	19	4	52	56	13,8889	4,86111	56,25	61,1111	1,88089	0,42747	2,30836	p<0,15
<i>Calliphora vicina</i>	68	15	5	48	53	9,44444	6,2963	52,2593	58,5556	3,26797	0,52709	3,79507	p<0,06 *
<i>Lucilia caesar</i>	73	16	3	54	57	11,4907	4,73148	56,7778	61,5093	1,76955	0,33057	2,10012	p<0,15
<i>Lucilia illustris</i>	68	16	3	49	52	10,5631	5,28155	52,1553	57,4369	2,7984	0,51465	3,31305	p<0,07 *
<i>Hemipyrellia ligurriens</i>	70	15	3	52	55	10,6061	3,53535	55,8586	59,3939	1,82035	0,32506	2,14541	p<0,15
<i>Lucilia cuprina</i>	66	17	1	48	49	12,1224	2,02041	51,8571	53,8776	1,96252	0,44157	2,40408	p<0,15
<i>Lucilia sericata</i>	66	17	3	46	49	11,88	3,96	50,16	54,12	2,2066	0,48438	2,69097	p<0,15
<i>Lucilia ampullacea</i>	62	16	3	43	46	11,0222	2,75556	48,2222	50,9778	2,24803	0,48606	2,73409	p<0,10 *
<i>Lucilia porphyria</i>	66	17	1	48	49	11,9362	2,10638	51,9574	54,0638	2,14829	0,4743	2,62259	p<0,15
<i>Lucilia bazini</i>	54	15	1	38	39	9,81818	3,68182	40,5	44,1818	2,73485	0,60774	3,34259	p<0,07 *
<i>Lucilia eximia</i>	73	17	0	56	56	11,3853	3,34862	58,2661	61,6147	2,76888	0,51164	3,28052	p<0,07 *
<i>Chrysomya megacephala</i>	62	15	6	41	47	10,1346	6,55769	45,3077	51,8654	2,33575	0,45641	2,79217	p<0,10 *
<i>Chrysomya bezziana</i>	62	15	4	43	47	10,233	6,01942	45,7476	51,767	2,22068	0,43897	2,65965	p<0,15
<i>Chrysomya saffrana</i>	62	15	5	42	47	9,9434	6,43396	45,6226	52,0566	2,57148	0,49118	3,06266	p<0,08 *
<i>Chrysomya pinguis</i>	58	15	3	40	43	9,57282	5,63107	42,7961	48,4272	3,07687	0,60822	3,68509	p<0,06 *
<i>Chrysomya latifrons</i>	55	16	5	34	39	9,33962	7,26415	38,3962	45,6604	4,74972	0,97153	5,72126	p<0,025 **
<i>Chrysomya semimetallica</i>	59	16	6	37	43	9,28704	8,19444	41,5185	49,713	4,85234	0,90648	5,75882	p<0,025 **
<i>Chrysomya albiceps</i>	68	16	1	51	52	11,027	3,06306	53,9099	56,973	2,24271	0,43407	2,67679	p<0,15

Tabela A3. (Continuação).

Espécies	Nº pares interface	Total de pares (observado)				Total de pares (esperado)				Qui-quadrado			
		GC	GU	AU	GU+AU	GC	GU	AU	GU+AU	GC	GU+AU	Total	p-valor (1 g.l.)
<i>Chrysomya rufifacies</i>	69	16	4	49	53	11,5	4,47222	53,0278	57,5	1,76087	0,35217	2,11304	p<0,15
<i>Chrysomya incisuralis</i>	75	17	1	57	58	11,9469	3,9823	59,0708	63,0531	2,13727	0,40496	2,54223	p<0,15
<i>Chrysomya nigripes</i>	70	14	7	49	56	10,8182	8,27273	50,9091	59,1818	0,93583	0,17107	1,10689	p<0,30
<i>Chrysomya putoria</i>	71	17	4	50	54	12,5294	4,87255	53,598	58,4706	1,59514	0,34182	1,93696	p<0,20
<i>Chrysomya flavifrons</i>	63	16	2	45	47	9,82569	4,62385	48,5505	53,1743	3,87984	0,71693	4,59677	p<0,03 **
<i>Chrysomya varipes</i>	61	16	3	42	45	9,97115	5,27885	45,75	51,0288	3,64521	0,71228	4,3575	p<0,04 **
<i>Cochliomyia hominivorax</i>	66	15	2	49	51	9,58065	4,79032	51,629	56,4194	3,06549	0,52056	3,58605	p<0,06 *
<i>Cochliomyia macellaria</i>	65	13	2	50	52	8,84	5,2	50,96	56,16	1,95765	0,30815	2,2658	p<0,15
<i>Hemilucilia segmentaria</i>	66	15	2	49	51	10,0179	4,71429	51,2679	55,9821	2,47775	0,44339	2,92114	p<0,09 *
<i>Chloroprocta idioidea</i>	69	15	3	51	54	10,7615	3,16514	55,0734	58,2385	1,6694	0,30848	1,97787	p<0,20
<i>Phormia regina</i>	65	16	4	45	49	9,20354	5,75221	50,0442	55,7965	5,01892	0,82786	5,84679	p<0,02 **
<i>Protophormia terraenovae</i>	68	15	2	51	53	10,0741	3,77778	54,1481	57,9259	2,40863	0,41889	2,82753	p<0,10 *
<i>Mesembrinella peregrina</i>	69	17	0	52	52	12,3474	2,90526	53,7474	56,6526	1,75317	0,3821	2,13527	p<0,15
<i>Mesembrinella bellardiana</i> I	72	17	5	50	55	16	4,66667	51,3333	56	0,0625	0,01786	0,08036	p<0,80
<i>Mesembrinella bellardiana</i> II	67	16	5	46	51	13,5354	4,06061	49,404	53,4646	0,44879	0,11362	0,5624	p<0,50
<i>Stomoxys calcitrans</i>	59	13	9	37	46	11,0625	6,14583	41,7917	47,9375	0,33934	0,07831	0,41764	p<0,60
<i>Musca domestica</i>	74	22	6	46	52	16,7304	5,7913	51,4783	57,2696	1,65975	0,48487	2,14462	p<0,15
<i>Haematobia irritans</i>	72	17	7	48	55	14,1308	8,07477	49,7944	57,8692	0,58256	0,14225	0,72481	p<0,40

Tabela A4. Número total de pares de bases encontrados nas áreas de interface entre regiões em simples e dupla-fita (apenas "bojos" e alças internas). Apenas um par de bases na vizinhança dessas regiões foi considerado. Os valores observados e esperados para os pares G-C, G-U e A-U nessas regiões são mostrados, incluindo os resultados de um teste qui-quadrado comparando-os. Pares G-U e A-U foram agrupados na mesma categoria dado que ambos são menos estáveis que os pares G-C e porque as substituições $UG \leftrightarrow UA$ e $GU \leftrightarrow AU$ são as mais comuns entre as hemi-CBCs observadas (Tabela 6). Valores em negrito indicam resultados significativos para $p < 0,05$ (**) e $p < 0,10$ (*).

Espécies	Nº pares interface	Total de pares (observado)				Total de pares (esperado)				Qui-quadrado			
		GC	GU	AU	GU+AU	GC	GU	AU	GU+AU	GC	GU+AU	Total	p-valor (1 g.l.)
<i>Calliphora vomitoria</i>	18	6	0	12	12	3	1,16667	13,8333	15	3	0,6	3,6	p<0,06 *
<i>Aldrichina grahami</i>	28	9	3	16	19	5,18519	1,81481	21	22,8148	2,80661	0,63787	3,44448	p<0,07 *
<i>Calliphora vicina</i>	20	5	2	13	15	2,77778	1,85185	15,3704	17,2222	1,77778	0,28674	2,06452	p<0,20
<i>Lucilia caesar</i>	26	7	1	18	19	4,09259	1,68519	20,2222	21,9074	2,06544	0,38585	2,4513	p<0,15
<i>Lucilia illustris</i>	23	6	1	16	17	3,57282	1,78641	17,6408	19,4272	1,6489	0,30325	1,95215	p<0,20
<i>Hemipyrellia ligurriens</i>	24	5	2	17	19	3,63636	1,21212	19,1515	20,3636	0,51136	0,09131	0,60268	p<0,50
<i>Lucilia cuprina</i>	22	6	0	16	16	4,04082	0,67347	17,2857	17,9592	0,94991	0,21373	1,16364	p<0,30
<i>Lucilia sericata</i>	22	6	1	15	16	3,96	1,32	16,72	18,04	1,05091	0,23069	1,2816	p<0,30
<i>Lucilia ampullacea</i>	20	7	1	12	13	3,55556	0,88889	15,5556	16,4444	3,33681	0,72147	4,05828	p<0,05 **
<i>Lucilia porphyrina</i>	20	7	0	13	13	3,61702	0,6383	15,7447	16,383	3,16408	0,69856	3,86264	p<0,05 **
<i>Lucilia bazini</i>	16	6	0	10	10	2,90909	1,09091	12	13,0909	3,28409	0,7298	4,01389	p<0,05 **
<i>Lucilia eximia</i>	26	6	0	20	20	4,05505	1,19266	20,7523	21,945	0,93287	0,17238	1,10525	p<0,30
<i>Chrysomya megacephala</i>	18	6	1	11	12	2,94231	1,90385	13,1538	15,0577	3,1776	0,62091	3,79851	p<0,06 *
<i>Chrysomya bezziana</i>	18	6	1	11	12	2,97087	1,74757	13,2816	15,0291	3,08852	0,61052	3,69904	p<0,06 *
<i>Chrysomya saffrana</i>	18	6	1	11	12	2,88679	1,86792	13,2453	15,1132	3,35738	0,6413	3,99868	p<0,05 **
<i>Chrysomya pinguis</i>	16	6	1	9	10	2,64078	1,5534	11,8058	13,3592	4,27313	0,84469	5,11782	p<0,03 **
<i>Chrysomya latifrons</i>	14	6	0	8	8	2,37736	1,84906	9,77358	11,6226	5,52022	1,12914	6,64935	p<0,01 **
<i>Chrysomya semimetallica</i>	16	6	1	9	10	2,51852	2,22222	11,2593	13,4815	4,81264	0,89906	5,7117	p<0,02 **
<i>Chrysomya albiceps</i>	22	7	0	15	15	3,56757	0,99099	17,4414	18,4324	3,30242	0,63918	3,94159	p<0,05 **

Tabela A4. (Continuação).

Espécies	Nº pares interface	Total de pares (observado)				Total de pares (esperado)				Qui-quadrado			
		GC	GU	AU	GU+AU	GC	GU	AU	GU+AU	GC	GU+AU	Total	p-valor (1 g.l.)
<i>Chrysomya rufifacies</i>	23	7	0	16	16	3,83333	1,49074	17,6759	19,1667	2,61594	0,52319	3,13913	p<0,08 *
<i>Chrysomya incisuralis</i>	26	8	0	18	18	4,14159	1,38053	20,4779	21,8584	3,59458	0,68108	4,27566	p<0,04 **
<i>Chrysomya nigripes</i>	24	5	0	19	19	3,70909	2,83636	17,4545	20,2909	0,44929	0,08213	0,53141	p<0,50
<i>Chrysomya putoria</i>	24	8	1	15	16	4,23529	1,64706	18,1176	19,7647	3,34641	0,71709	4,06349	p<0,05 **
<i>Chrysomya flavifrons</i>	20	6	0	14	14	3,11927	1,46789	15,4128	16,8807	2,66044	0,4916	3,15205	p<0,08 *
<i>Chrysomya varipes</i>	18	7	0	11	11	2,94231	1,55769	13,5	15,0577	5,5959	1,09345	6,68935	p<0,01 **
<i>Cochliomyia hominivorax</i>	20	4	1	15	16	2,90323	1,45161	15,6452	17,0968	0,41434	0,07036	0,4847	p<0,50
<i>Cochliomyia macellaria</i>	20	4	0	16	16	2,72	1,6	15,68	17,28	0,60235	0,09481	0,69717	p<0,50
<i>Hemilucilia segmentaria</i>	22	6	1	15	16	3,33929	1,57143	17,0893	18,6607	2,12003	0,37937	2,49941	p<0,15
<i>Chloroprocta idioidea</i>	24	6	1	17	18	3,74312	1,10092	19,156	20,2569	1,36077	0,25145	1,61221	p<0,30
<i>Phormia regina</i>	20	6	0	14	14	2,83186	1,76991	15,3982	17,1681	3,54436	0,58464	4,12899	p<0,05 **
<i>Protophormia terraenovae</i>	22	6	1	15	16	3,25926	1,22222	17,5185	18,7407	2,30471	0,40082	2,70553	p<0,15
<i>Mesembrinella peregrina</i>	28	8	0	20	20	5,01053	1,17895	21,8105	22,9895	1,78364	0,38874	2,17238	p<0,15
<i>Mesembrinella bellardiana</i> I	26	6	3	17	20	5,77778	1,68519	18,537	20,2222	0,00855	0,00244	0,01099	p<0,99
<i>Mesembrinella bellardiana</i> II	28	6	2	20	22	5,65657	1,69697	20,6465	22,3434	0,02085	0,00528	0,02613	p<0,90
<i>Stomoxys calcitrans</i>	18	6	1	11	12	3,375	1,875	12,75	14,625	2,04167	0,47115	2,51282	p<0,15
<i>Musca domestica</i>	22	7	2	13	15	4,97391	1,72174	15,3043	17,0261	0,82531	0,2411	1,06641	p<0,40
<i>Haematobia irritans</i>	22	5	2	15	17	4,31776	2,46729	15,215	17,6822	0,1078	0,02632	0,13412	p<0,80

Tabela A5. Número total de pares de bases encontrados nas áreas de interface entre regiões em simples e dupla-fita (apenas "bojos" e alças internas). Dois pares de bases na vizinhança dessas regiões foi considerado. Os valores observados e esperados para os pares G-C, G-U e A-U nessas regiões são mostrados, incluindo os resultados de um teste qui-quadrado comparando-os. Pares G-U e A-U foram agrupados na mesma categoria dado que ambos são menos estáveis que os pares G-C e porque as substituições $UG \leftrightarrow UA$ e $GU \leftrightarrow AU$ são as mais comuns entre as hemi-CBCs observadas (Tabela 6). Valores em negrito indicam resultados significativos para $p < 0,05$ (**) e $p < 0,10$ (*).

Espécies	Nº pares interface	Total de pares (observado)				Total de pares (esperado)				Qui-quadrado			
		GC	GU	AU	GU+AU	GC	GU	AU	GU+AU	GC	GU+AU	Total	p-valor (1 g.l.)
<i>Calliphora vomitoria</i>	34	11	0	23	23	5,66667	2,2037	26,1296	28,3333	5,01961	1,00392	6,02353	p<0,025 **
<i>Aldrichina grahami</i>	53	14	3	36	39	9,81481	3,43519	39,75	43,1852	1,78463	0,4056	2,19022	p<0,15
<i>Calliphora vicina</i>	40	10	3	27	30	5,55556	3,7037	30,7407	34,4444	3,55556	0,57348	4,12903	p<0,05 **
<i>Lucilia caesar</i>	48	11	1	36	37	7,55556	3,11111	37,3333	40,4444	1,57026	0,29335	1,86361	p<0,20
<i>Lucilia illustris</i>	43	11	1	31	32	6,67961	3,33981	32,9806	36,3204	2,79444	0,51392	3,30836	p<0,07 *
<i>Hemipyrellia ligurriens</i>	46	11	3	32	35	6,9697	2,32323	36,7071	39,0303	2,33057	0,41617	2,74674	p<0,15
<i>Lucilia cuprina</i>	42	11	0	31	31	7,71429	1,28571	33	34,2857	1,39947	0,31488	1,71435	p<0,20
<i>Lucilia sericata</i>	42	11	2	29	31	7,56	2,52	31,92	34,44	1,56529	0,3436	1,90889	p<0,20
<i>Lucilia ampullacea</i>	36	11	2	23	25	6,4	1,6	28	29,6	3,30625	0,71486	4,02111	p<0,05 **
<i>Lucilia porphyria</i>	36	11	0	25	25	6,51064	1,14894	28,3404	29,4894	3,09561	0,68345	3,77905	p<0,06 *
<i>Lucilia bazini</i>	30	11	0	19	19	5,45455	2,04545	22,5	24,5455	5,63788	1,25286	6,89074	p<0,01 **
<i>Lucilia eximia</i>	48	12	0	36	36	7,48624	2,20183	38,3119	40,5138	2,72153	0,50289	3,22442	p<0,08 *
<i>Chrysomya megacephala</i>	34	10	1	23	24	5,55769	3,59615	24,8462	28,4423	3,55077	0,69383	4,2446	p<0,04 **
<i>Chrysomya bezziana</i>	34	10	1	23	24	5,61165	3,30097	25,0874	28,3883	3,43172	0,67836	4,11008	p<0,05 **
<i>Chrysomya saffrana</i>	34	10	1	23	24	5,45283	3,5283	25,0189	28,5472	3,79193	0,7243	4,51623	p<0,04 **
<i>Chrysomya pinguis</i>	30	10	1	19	20	4,95146	2,91262	22,1359	25,0485	5,14753	1,01754	6,16507	p<0,02 **
<i>Chrysomya latifrons</i>	27	11	1	15	16	4,58491	3,56604	18,8491	22,4151	8,97585	1,83597	10,8118	p<0,01 **
<i>Chrysomya semimetallica</i>	31	11	2	18	20	4,87963	4,30556	21,8148	26,1204	7,67659	1,43409	9,11068	p<0,01 **
<i>Chrysomya albiceps</i>	42	12	0	30	30	6,81081	1,89189	33,2973	35,1892	3,95367	0,76523	4,71889	p<0,03 **

Tabela A5. (Continuação).

Espécies	Nº pares interface	Total de pares (observado)				Total de pares (esperado)				Qui-quadrado			
		GC	GU	AU	GU+AU	GC	GU	AU	GU+AU	GC	GU+AU	Total	p-valor (1 g,l,)
<i>Chrysomya rufifacies</i>	43	12	1	30	31	7,16667	2,78704	33,0463	35,8333	3,25969	0,65194	3,91163	p<0,05 **
<i>Chrysomya incisuralis</i>	50	13	0	37	37	7,9646	2,65487	39,3805	42,0354	3,18349	0,60319	3,78668	p<0,06 *
<i>Chrysomya nigripes</i>	45	10	2	33	35	6,95455	5,31818	32,7273	38,0455	1,33363	0,24378	1,57741	p<0,30
<i>Chrysomya putoria</i>	46	13	1	32	33	8,11765	3,15686	34,7255	37,8824	2,93649	0,62925	3,56573	p<0,06 *
<i>Chrysomya flavifrons</i>	38	12	1	25	26	5,92661	2,78899	29,2844	32,0734	6,22382	1,15005	7,37387	p<0,01 **
<i>Chrysomya varipes</i>	34	12	0	22	22	5,55769	2,94231	25,5	28,4423	7,46773	1,45921	8,92694	p<0,01 **
<i>Cochliomyia hominivorax</i>	38	9	1	28	29	5,51613	2,75806	29,7258	32,4839	2,20034	0,37364	2,57398	p<0,15
<i>Cochliomyia macellaria</i>	38	9	0	29	29	5,168	3,04	29,792	32,832	2,84137	0,44725	3,28863	p<0,07 *
<i>Hemilucilia segmentaria</i>	40	12	1	27	28	6,07143	2,85714	31,0714	33,9286	5,78908	1,03594	6,82502	p<0,01 **
<i>Chloroprocta idioidea</i>	43	10	1	32	33	6,70642	1,97248	34,3211	36,2936	1,6175	0,29889	1,91639	p<0,20
<i>Phormia regina</i>	38	12	1	25	26	5,38053	3,36283	29,2566	32,6195	8,14369	1,34329	9,48698	p<0,01 **
<i>Protophormia terraenovae</i>	41	12	1	28	29	6,07407	2,27778	32,6481	34,9259	5,78139	1,00546	6,78685	p<0,01 **
<i>Mesembrinella peregrina</i>	49	13	0	36	36	8,76842	2,06316	38,1684	40,2316	2,04213	0,44508	2,48721	p<0,15
<i>Mesembrinella bellardiana</i> I	50	12	3	35	38	11,1111	3,24074	35,6481	38,8889	0,07111	0,02032	0,09143	p<0,40
<i>Mesembrinella bellardiana</i> II	51	14	3	34	37	10,303	3,09091	37,6061	40,697	1,32656	0,33584	1,6624	p<0,20
<i>Stomoxys calcitrans</i>	30	10	2	18	20	5,625	3,125	21,25	24,375	3,40278	0,78526	4,18803	p<0,05 **
<i>Musca domestica</i>	43	14	4	25	29	9,72174	3,36522	29,913	33,2783	1,88274	0,55001	2,43276	p<0,15
<i>Haematobia irritans</i>	41	10	2	29	31	8,04673	4,59813	28,3551	32,9533	0,47414	0,11578	0,58992	p<0,50

Anexo B

80

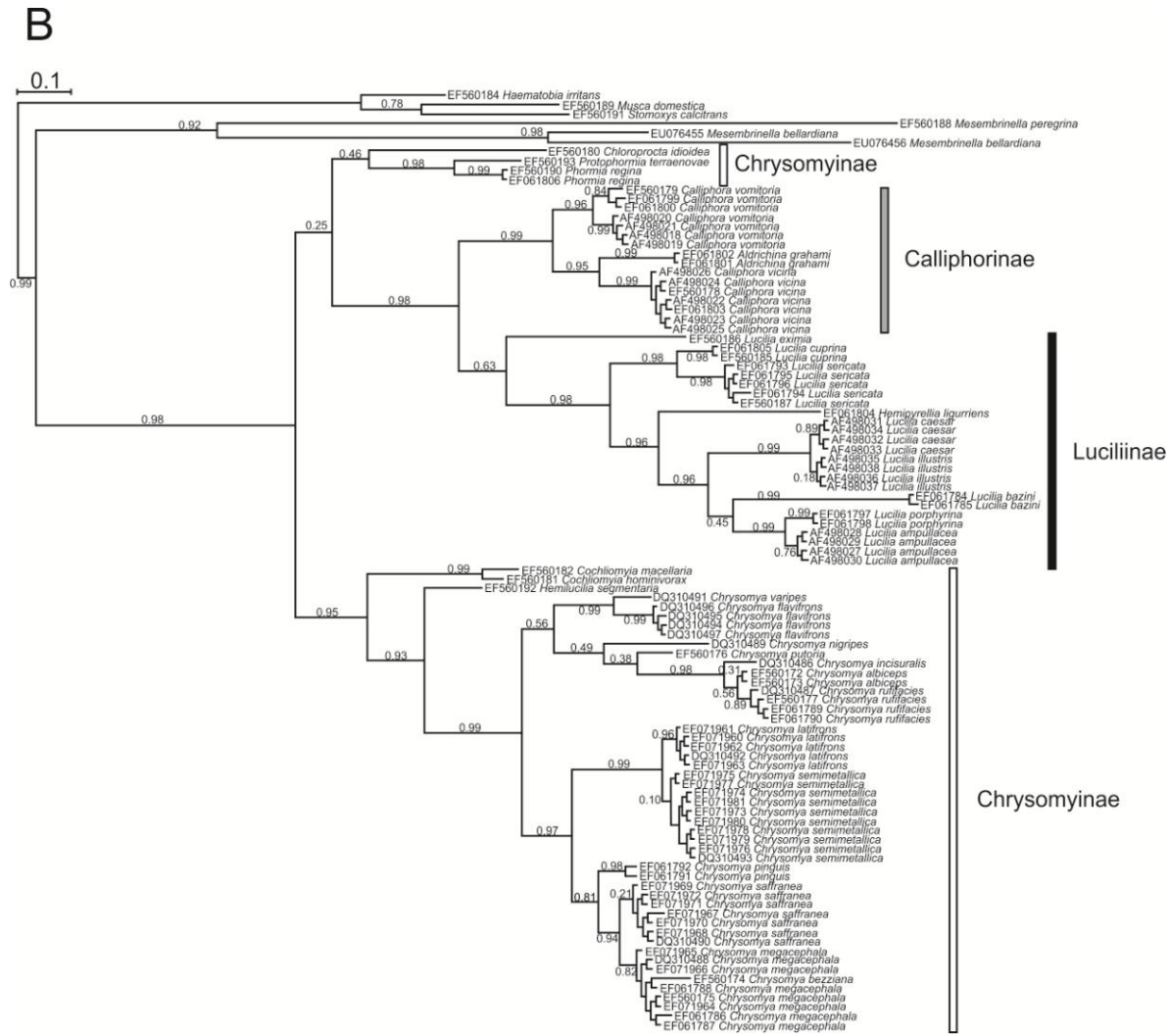


Figura B1. (Continuação).

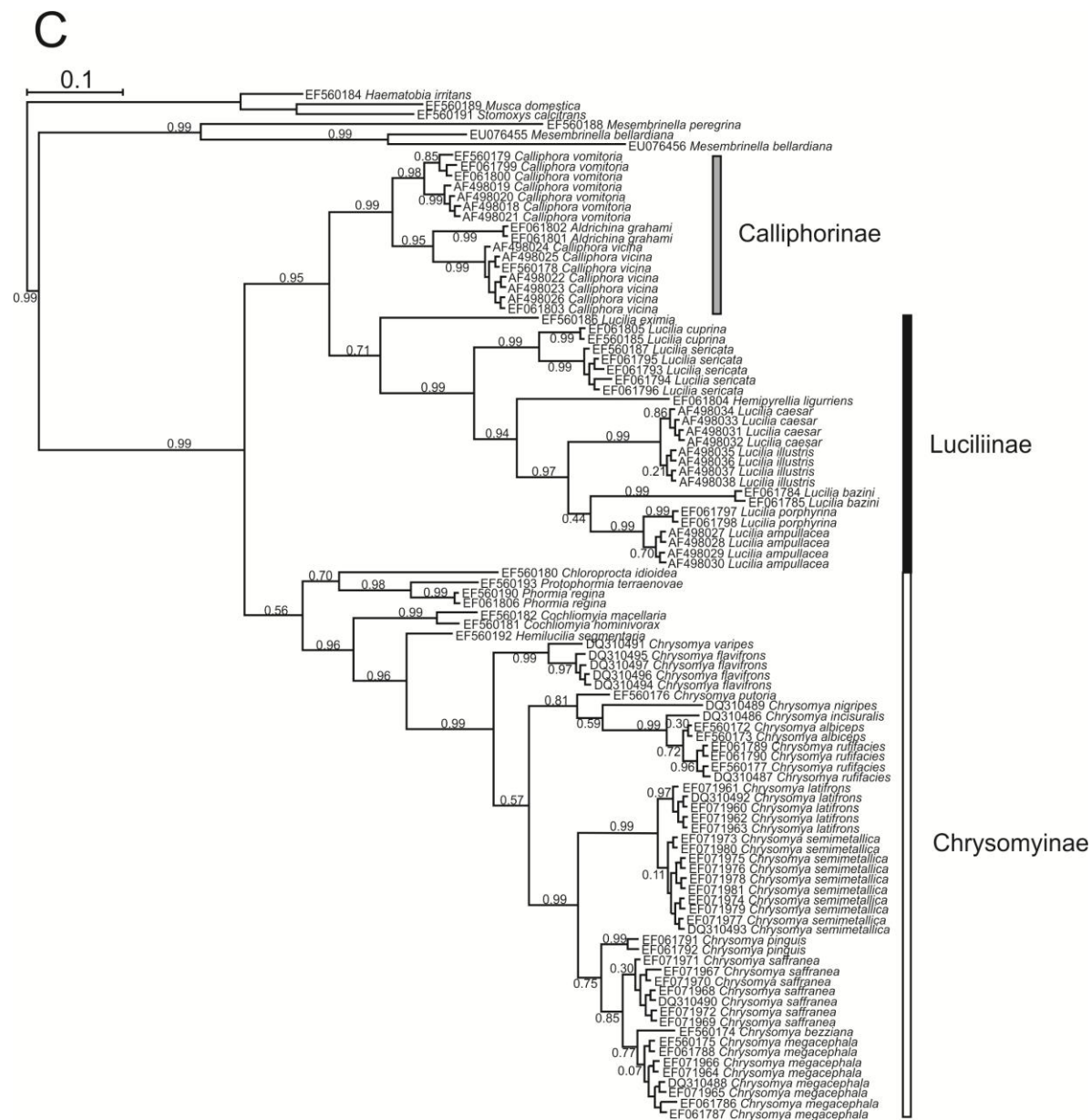


Figura B1. (Continuação).

Capítulo II

Caracterização estrutural *in vitro* do segundo espaçador transcrito interno (ITS2) do cluster de DNA ribossomal (rDNA) em Calyptratae (Diptera: Schizophora): implicações em estudos genético-evolutivos.

Resumo

O segundo espaçador transcrito interno (ITS2) é uma pequena região não codificante localizada no "cluster" de DNA ribossomal nuclear em eucariotos. Apesar de ser prontamente clivado após a transcrição, o ITS2 possui função essencial no processamento dos RNAs ribossomais uma vez que nele residem os sítios de reconhecimento para os elementos responsáveis por esse processo, uma função altamente dependente da estrutura secundária adquirida pela molécula. O uso crescente do ITS2 em análises genético-evolutivas, incluindo a consideração de informações estruturais, encoraja a avaliação da precisão dos modelos estruturais atualmente considerados para a molécula, os quais baseiam-se em uma abordagem combinada de predição *in silico* e modelagem comparativa. Com esse objetivo, a estrutura secundária da região do ITS2 de três espécies representativas das superfamílias de Calyptratae (Diptera: Schizophora): *Glossina morsitans*, *Musca domestica* e *Cochliomyia hominivorax*; foi analisada usando métodos de transcrição *in vitro*, digestão com RNases sítio/estrutura específicas (V1, I, A e T1) e eletroforese em capilar para visualização do padrão de fragmentos resultante, o qual foi mapeado na(s) estrutura(s) preditas previamente *in silico*. A precisão do modelo computacional para a estrutura do ITS2 foi então avaliada. Apesar de apresentar baixa resolução, o modelo estabelecido *in vitro* parece corroborar a maior parte dos elementos estruturais inicialmente preditos. Uma provável dinâmica conformacional pode existir na região do cerne da estrutura do ITS2, como inferido pela análise das múltiplas estruturas preditas *in silico* e de dados previamente disponíveis em análises *in vivo* em leveduras. Com relação ao modelo usualmente considerado e incorporado em análises genético-evolutivas, apesar de não considerar a possível dinâmica existente, o mesmo parece conter todos os elementos evolutivamente conservados na molécula, os quais são provavelmente

essenciais para o processamento do ITS2. A consideração de um modelo estático, embora resulte em uma pequena perda da informação contida na molécula, representa um ganho em termos de poder e tempo computacional necessários para a condução das análises e, portanto, parece ser justificável.

Palavras-chave: Calyptratae; ITS2; estrutura secundária; *in silico*; *in vitro*; RNAses; análise de fragmentos.

1. Introdução

O DNA ribossomal (rDNA) em eucariotos organiza-se em múltiplas cópias dispostas *in tandem* e que são transcritas em grandes unidades contendo, além dos RNAs para as subunidades 18S, 5,8S e 28S, dois espaçadores internos (ITS1 e 2) e dois espaçadores externos (ETS) um a 5' e outro a 3' do transcrito (Veldman et al. 1981; Hillis & Dixon 1991) (Figura 1).

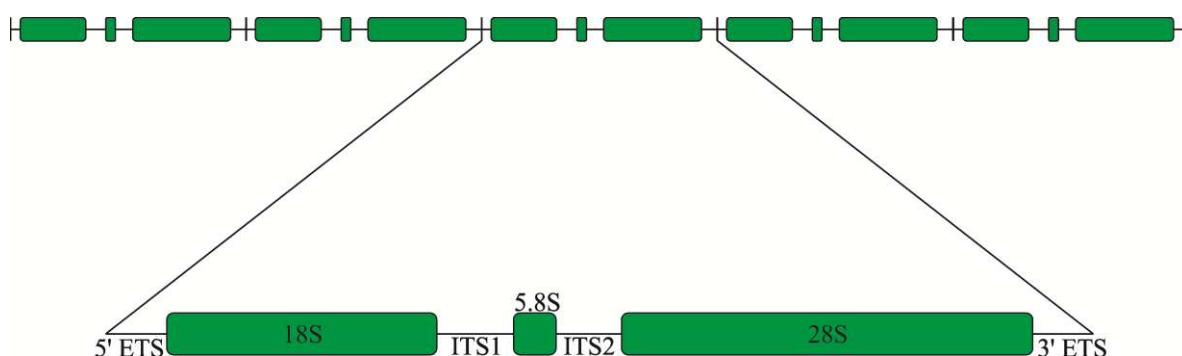


Figura 1. Representação esquemática da estrutura e organização do cluster de DNA ribossomal em eucariotos. ETS = espaçador transcrito externo; ITS = espaçador transcrito interno.

O ITS2 é uma região pequena (normalmente <1000pb em eucariotos), localizada entre as extremidades 3' e 5' das subunidades 5,8S e 28S, respectivamente (Veldman et al. 1981), e que possui função essencial na maturação das subunidades ribossomais, uma vez que nele residem os sítios de reconhecimento e clivagem que, após a ação de um conjunto de endonucleases ainda não totalmente conhecidas, delimitarão as extremidades das moléculas de rRNA (Mitchell et al. 1996, 1997; Geerlings et al. 2000; Côte et al. 2002).

Modelos de estrutura secundária do ITS2 propostos em levedura indicam que o reconhecimento desses sítios de clivagem estão relacionados à conformação secundária adquirida pela molécula durante o processamento do transcrito primário, sugerindo que a

clivagem do rRNA requer o reconhecimento de sequências específicas (que podem ser aproximadas na conformação secundária adquirida) e/ou elementos estruturais (Veldman et al. 1981; van der Sande et al. 1992; Côte et al. 2002). O fato da estrutura secundária do ITS2 ser conservada entre os filos de eucariotos indica a importância de sua estrutura no reconhecimento e processamento das subunidades ribossomais (van der Sande et al. 1992; Joseph et al. 1999; Côte et al. 2002; Schultz et al. 2005).

O primeiro modelo de estrutura secundária da molécula, proposto a partir de análises de restrição enzimática e modificações químicas de bases *in vitro* do ITS2 de *S. cerevisiae*, foi sugerido por Yeh & Lee (1990) e consistia em um “grampo” (do inglês “hairpin”) com seis regiões de hélice-alça (do inglês “helix-loop”) principais denominadas como domínios (Figura 2A). Esse modelo foi posteriormente sugerido também para o ITS2 de *D. melanogaster* por Schlötterer et al. (1994), utilizando métodos de predição *in silico*. Análises posteriores de Joseph et al. (1999), novamente com métodos computacionais, compararam sequências e estruturas secundárias do ITS2 de leveduras a vertebrados e chegaram a um modelo diferente, único e geral entre os eucariotos, o qual apresenta um “anel” central (do inglês “ring”) do qual ramificam-se quatro regiões de hélice-alça, também denominadas como domínios (Figura 2B).

Estudos funcionais demonstraram posteriormente que sub-estruturas de ambos os modelos eram necessárias para o correto processamento das subunidades ribossomais em leveduras, embora o modelo em “anel” possuísse mais estruturas essenciais para o processamento do que o modelo de “grampo” (Côte et al. 2002). Um modelo de dinâmica conformacional foi então proposto, no qual ambas as estruturas co-existem *in vivo* durante o processamento, com o modelo em “anel” presente em estágios iniciais e o modelo de “grampo” em estágios tardios (Côte et al. 2002).

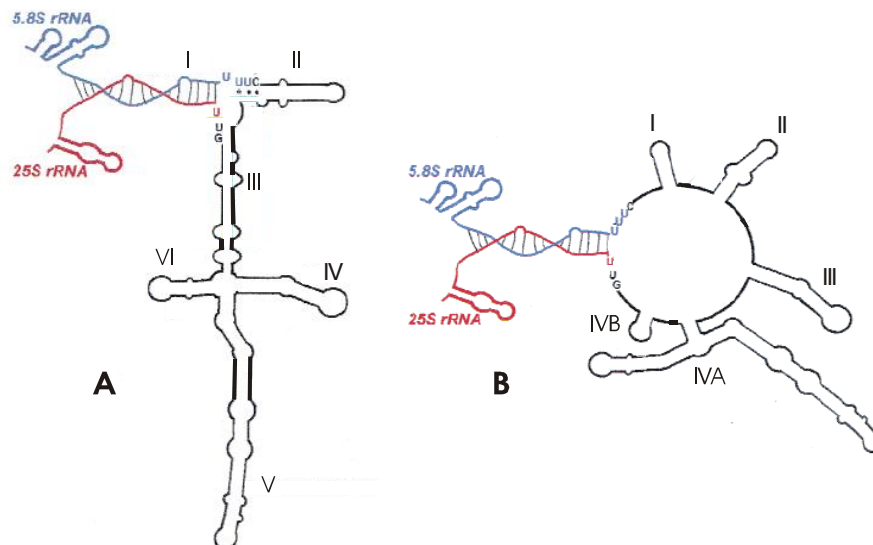


Figura 2. Modelos de estrutura secundária propostas para o ITS2 de *Saccharomyces cerevisiae* (A) em “grampo” (Yeh & Lee, 1990) e (B) em “anel” (Joseph et al. 1999). (Figura retirada de Côte et al. 2002).

Até o momento, o principal método de predição de estruturas secundárias em sequências de DNA e RNA é sua modelagem *in silico* (Hofacker 2003; Zucker 2003; Gardner & Giegerich 2004; Reeder et al. 2006), sendo que na maioria dos programas a predição é realizada utilizando-se algoritmos que computam a estrutura com a menor energia livre global (MFE, do inglês “minimum free energy”) (i.e., estrutura mais estável, com o número de pareamentos maximizado). No caso de múltiplas sequências disponíveis para um conjunto de organismos aparentados, as estruturas obtidas para cada organismo podem ser comparadas e o padrão de estrutura secundária comum para todos é estabelecido por homologia e modelagem comparativa. Nesses casos, a constatação de alterações na sequência primária que não alteram a conformação global adquirida pela molécula (mudanças compensatórias de base - CBC, do inglês “Compensatory Base Changes”) pode ser tomada como uma evidência corroborando o modelo proposto (Mai & Coleman 1997; Coleman 2003). A estrutura secundária dos rRNAs de procariotos e eucariotos e do ITS2 de

diversos táxons de eucariotos tem sido modelada por essa combinação de abordagens (Guttel et al. 1994; Young & Coleman 2004; Schultz et al. 2005; Wolf et al. 2005).

Apesar da confiabilidade aparente dos modelos computacionais, estima-se que os algoritmos utilizados na predição da estrutura secundária apresentem uma precisão de 50 a 70% de bases corretamente pareadas quando se analisam sequências isoladamente (Eddy 2004), enquanto métodos de modelagem simultâneos podem apresentar precisões variando de 60 a 100% (Gardner & Giegerich 2004). Apesar da alta precisão em alguns casos, a modelagem computacional apresenta uma multiplicidade de estruturas resultantes para cada sequência ou conjunto de sequências fornecido, sendo que a estrutura correta, caso tenha sido alcançada, pode se encontrar em um intervalo de 5 a 10% da mínima energia livre computada (Eddy 2004). Isso implica que a estrutura computada como a de menor energia livre não necessariamente é a correta, cabendo ao usuário julgar, dentre as muitas estruturas possíveis, qual se aproximaria mais a provável estrutura adquirida *in vivo*. Além disso, a não consideração de alguns parâmetros biológicos, como a existência de proteínas e outras macromoléculas que auxiliam na aquisição da conformação secundária correta, a existência de nucleotídeos modificados que alteram o padrão de pareamento dentro da molécula e a multiplicidade de orientações nas quais pode haver o pareamento das bases (que podem levar à formação de pareamentos não convencionais, em geral devido a interações de ordem terciária), são outros pontos pouco considerados nas análises computacionais (Holbrook et al. 2001; Leontis & Westhof 2001; Favaretto et al. 2005; Holbrook 2008).

Como uma alternativa para os modelos computacionais, modelos *in vitro* podem ser obtidos através de técnicas de cristalografia, ressonância magnética nuclear (NMR – “nuclear magnetic resonance”) ou de análises *in vitro* utilizando-se RNAses e modificações químicas de nucleotídeos para determinação de resíduos pareados ou não-pareados. Embora

os métodos cristalográficos e de NMR tenham evoluído muito nos últimos anos, moléculas de RNA são de difícil manipulação e cristalização e estas técnicas têm sido utilizadas na determinação estrutural em três dimensões de moléculas de até 150 nucleotídeos (Tzakos et al. 2006). Para moléculas maiores, técnicas *in vitro* utilizando modificações químicas de nucleotídeos e digestão com nucleases específicas para resíduos em simples ou dupla-fita possuem maior aplicabilidade. Estas técnicas permitiram a caracterização estrutural em duas dimensões do ITS2 em *S. cerevisiae* (Yeh & Lee. 1990) e, mais recentemente, de alguns segmentos de expansão das subunidades de rRNA em levedura, camundongo e trigo (Alkemar & Nygard 2006).

Apesar dos avanços técnicos, estas metodologias ainda possuem limitações. A principal delas é o fato de, após isolada, a molécula de ácido nucléico adquirir sua conformação final *in vitro*, onde muitos parâmetros biológicos não são reproduzidos. Mesmo assim, a criação de um modelo de estrutura secundária através de diferentes técnicas é interessante por permitir a comparação entre diferentes estruturas propostas, fornecendo um maior volume de dados para análises.

O estabelecimento de modelos *in vitro* para estruturas secundárias de ácidos nucléicos, incluindo o ITS2, visa a obtenção de bases mais sólidas e confiáveis para a modelagem computacional, permitindo também o desenvolvimento de análises genético-evolutivas com maior robustez e precisão.

2. Materiais e Métodos

2.1. Extração de DNA e amplificação via PCR

Indivíduos das espécies *C. hominivorax* (Calypttratae: Oestroidea: Calliphoridae), *Musca domestica* (Calypttratae: Muscoidea: Muscidae) e *Glossina morsitans* (Calypttratae: Hippoboscoidea: Glossinidae) foram obtidas da coleção disponível no Laboratório de

Genética e Evolução Animal do CBMEG/UNICAMP. O DNA total das espécies *C. hominivorax* e *Musca domestica* foi extraído segundo o protocolo de fenol/clorofórmio (Infante & Azeredo-Espin 1995), enquanto para *G. morsitans* o DNA total foi extraído utilizando-se o kit "Spin Tissue Mini-Kit" (Invitex). Após o procedimento, os DNAs extraídos foram armazenados em freezer a -20°C.

Para as reações de amplificação via PCR, oligonucleotídeos foram desenvolvidos baseados na análise das sequências da região do ITS2 de *C. hominivorax* (GenBank nº EF560181), *M. domestica* (GenBank nº EF560189) e *G. morsitans* (GenBank nº JQ246601) com o objetivo de amplificar a região completa do ITS2 nessas espécies, além de 25 nucleotídeos das subunidades 2S (5' do ITS2) e 28S (3' do ITS2). Os oligos utilizados na amplificação do ITS2 completo em *C. hominivorax* (*Ch*ITS2_{I,II,III,IV}) e *G. morsitans* (*Gm*ITS2_{I,II,III,IV}) foram: 2S-D (5'-GAAAGATCTGGACTACATATGGTTGAGGGTTG-3'), complementar a região 2S do "cluster" ribossomal (com sítio de restrição para a enzima *Bgl*II adicionado à extremidade 5' da sequência - sublinhado), e 28S-R (5'-CGGAATTCG GTAATCCCATATGAGTTGAGG-3'), que se hibridiza na extremidade 5' da região 28S no "cluster" ribossomal (com sítio de restrição para *Eco*RI - sublinhado). Para *M. domestica* (*Md*ITS2_{I,II,III,IV}), devido à existência de um sítio de restrição para a enzima *Eco*RI na sequência, oligonucleotídeos foram desenvolvidos especificamente para a amplificação da região nessa espécie: 2S-D-*Hind*III (5'-CCCAAAGCTTGGACTACATATGGTTGAGG GTTG-3'), complementar a região 2S do "cluster" ribossomal (com sítio de restrição para a enzima *Hind*III adicionado à extremidade 5' da sequência - sublinhado), e 28S-R-Md-*Bam*HI (5'-CGCGGATCCGGTAGTCCCATATGAGTTGAGG-3'), que se hibridiza na extremidade 5' da região 28S no "cluster" ribossomal (com sítio de restrição para a enzima *Bam*HI - sublinhado).

Adicionalmente, para a espécie *C. hominivorax*, a amplificação da região ITS2 foi dividida de modo a amplificar individualmente as regiões correspondentes aos domínio I e II (*ChITS2_{I,II}*) e III e IV (*ChITS2_{III,IV}*). Para tanto, os oligos utilizados foram: 2S-D e ITS2-R (5'-CGGAATTCGAGGTTTTGTATCTTTAGC-3'), capaz de se hibridizar internamente à região ITS2 (com sítio de restrição para *EcoRI* -sublinhado) (*ChITS2_{I,II}*); e ITS2-D (5'-GAA GATCTGCTAAAGATACAAAACCTC-3'), que se hibridiza internamente a região do ITS2 (com sítio de restrição para *BglII* - sublinhado) e 28S-R (*ChITS2_{III,IV}*).

As reações de amplificação via PCR foram conduzidas para um volume total de 50 µl com 20mM Tris-HCl (pH 8,4), 50mM KCl, 2mM MgCl₂, 0,2mM dNTPs, 0,5µM do oligo direto, 0,5 µM do oligo reverso, 1,25 U da enzima *Taq* DNA polimerase e 1-2 µg de DNA. Os ciclos de amplificação incluíram um passo de desnaturação inicial a 94°C por 3 minutos, seguido de 35 ciclos a 94°C por 1 minuto para desnaturação, 55°C por 45 segundos para hibridização dos oligonucleotídeos, 60°C por 2 minutos para extensão das cadeias de DNA e um passo final a 60°C por 3 minutos para extensão final das cadeias de DNA. Os produtos de amplificações foram verificados em gel de agarose, corados em brometo de etídio, visualizados em luz ultravioleta (300nm) e fotografados em sistema digital EDAS (Kodak®).

2.2. Clonagem

2.2.1. Ligação

Os produtos das reações de PCR foram purificados utilizando-se o “kit” QIAquick® PCR Purification Kit (QIAGEN), seguindo instruções do fabricante. Após a purificação, os produtos das amplificações foram digeridos a 37°C por 2 horas utilizando-se as enzimas de restrição *EcoRI* e *BglII* (Invitrogen) (*ChITS2_{I,II,III,IV}*, *ChITS2_{I,II}*, *ChITS2_{III,IV}* e *GmITS2_{I,II,III,IV}*) e

*Bam*HI e *Hind*III (Invitrogen) (*Md*ITS2_{I,II,III,IV}). Os fragmentos digeridos foram então clonados em vetor de expressão pHST 7.0 (que contém a região promotora para a RNA polimerase T7), previamente digerido com as mesmas enzimas utilizadas na digestão do fragmento amplificado, utilizando a enzima T4 DNA Ligase do kit de clonagem pGEM T-Easy Vector (Promega).

2.2.2. Transformação

Os vetores construídos foram transformados em células de *E. coli* linhagem DH5 α através de choque químico com tampão de transformação (0,1M de KCl; 0,03M de CaCl₂; 0,05M de MgCl₂; 1,5% PEG - Sambrook et al. 1989). As bactérias recombinantes foram plaqueadas em meio LB contendo o antibiótico Ampicilina (50 μ g/ml). As placas foram incubadas invertidas em estufa a 37°C por 12-16 horas.

2.2.3. Midipreparação plasmidial

A midipreparação plasmidial foi realizada utilizando-se protocolo adaptado de Sambrook et al. (1989). Para verificar se o inserto clonado correspondia ao produto de PCR amplificado, ~150ng de plasmídeo foram digeridos com as enzimas *Eco*RI / *Bgl*II e *Bam*HI / *Hind*III. As reações foram incubadas a 37°C por 2 horas e o resultado da digestão foi aplicado em gel de agarose, corado com brometo de etídio e visualizado em luz ultravioleta (300 nm).

2.3. Transcrição *in vitro*

Para as reações de transcrição *in vitro*, os vetores pHST 7.0 contendo os fragmentos obtidos via PCR foram linearizados através de digestão com a enzima *Eco*RI (Fermentas) [*Ch*ITS2_{I,II,III,IV}, *Ch*ITS2_{I,II}, *Ch*ITS2_{III,IV} e *Gm*ITS2_{I,II,III,IV}] e *Bam*HI (Invitrogen) [*Md*ITS2_{I,II,III,IV}], a 37°C por 4 horas. Os vetores linearizados foram aplicados em gel de agarose, corados com brometo de etídeo e visualizados em luz UV (300nm). As bandas de

interesse foram excisadas do gel e purificadas utilizando-se o kit Invisorb DNA CleanUp (Invitek), seguindo instruções do fabricante. O DNA resultante foi quantificado utilizando-se o sistema “Qubit Quantitation Platform” (Invitrogen).

Para a transcrição *in vitro*, utilizou-se o kit “TranscriptAid™ T7 High Yield Transcription” (Fermentas) seguindo as instruções do fabricante. As reações foram conduzidas a 37°C por 2-3 horas contendo: 1µg DNA molde linearizado; 10mM cada NTP; 2µl TranscriptAid Enzyme Mix; 4µl 5X TranscriptAid Reaction Buffer e H₂O DEPC para um volume total de 20µl. Uma alíquota de 1µl foi reservada para posterior verificação em gel. Em seguida, foram adicionados 2µl de DNase I (para digestão do DNA molde) (a enzima DNase I acompanha o kit de transcrição) e a reação foi incubada a 37°C por 15 minutos. A digestão com DNase I foi interrompida através da adição de 2µl 0,5M EDTA (pH 8,0) e em seguida incubada a 65°C por 10 minutos. Uma nova alíquota de 1µl foi separada para verificação em gel.

O produto da transcrição *in vitro* foi então purificado por protocolo de fenol/clorofórmio modificado a partir do descrito no manual do kit “TranscriptAid”. O processo consistiu na adição, à reação de transcrição, de 160µl de H₂O DEPC, 20µl (1/10 volume total) de acetato de potássio 3M (pH 5,3) e 200µl (1 volume) de uma solução 1:1 fenol/clorofórmio. A solução foi então vortexada, centrifugada por 10 minutos a 13.000 rpm e a fase aquosa (superior) foi retirada e transferida para um novo tubo. Em seguida, adicionou-se 200µl (1 volume) de clorofórmio à solução, a qual foi vortexada e centrifugada por 10 minutos a 13.000 rpm. A fase aquosa foi novamente retirada e o passo anterior repetido (1 volume de clorofórmio). Uma alíquota de 1µl foi novamente retirada para verificação em gel.

Após a purificação por fenol/clorofórmio, o RNA purificado foi precipitado através da adição de 600µl (3 volumes) de etanol 100% gelado, incubado em freezer -20°C por 1 hora e então centrifugado por 30 minutos (13.000 rpm). O RNA precipitado foi lavado com etanol 75% e o precipitado restante foi incubado em banho seco a 37°C por 5 minutos (para evaporação do etanol residual) e ressuspendido em 20µl H₂O DEPC.

2.4. Desnaturação e renaturação (“refolding”) do RNA transcrito

Para o protocolo de desnaturação e “refolding”, uma solução contendo ~10µg do RNA transcrito, 16µg de RNA de levedura fragmentado (“sheared”) (Ambion – Applied Biosystems), 1X RNA Structure Buffer (10mM Tris pH 7,0; 0,1M KCl; 10mM MgCl₂) e água DEPC para 144µl (conforme protocolo RNA-Grade Ribonucleases - Ambion) foi submetida a uma temperatura de 70°C para desnaturação das fitas de RNA, resfriada 1°C por minuto até 25°C e então estabilizada a 25°C por 5 minutos em um termociclador MJ Research PTC-200. Após esse processo, a solução foi imediatamente colocada em gelo por 5 minutos e então procedeu-se os tratamentos com RNAses sítio e estrutura específicas.

2.5. Tratamento do RNA transcrito com RNAses sítio e estrutura específicas

A solução contendo o RNA desnaturado e re-estruturado foi aliquoteada em 16 tubos (9µl por tubo, com aproximadamente 600ng de RNA em cada) e submetida as análises de digestão com as RNAses sítio e estrutura específicas (Ambion – Applied Biosystems): RNase I (clivagem à 3' de As, Us, Cs e Gs em simples fita); RNase A (clivagem à 3' de Us e Cs em simples fita); RNase V1 (clivagem de regiões em dupla fita) e RNase T1 (clivagem à 3' de Gs em simples fita). A cada tubo foi adicionado 1µl de enzima, sendo que para cada enzima foram utilizadas diferentes concentrações/diluições:

[1] RNase I (concentração inicial: 100U/µl): 1:10; 1:100; 1:1000

[2] RNase A (concentração inicial: 1ng/µl): 1; 1:10; 1:100; 1:1000

[3] RNase V1 (concentração inicial: 0,1U/μl): 1; 1:10; 1:100; 1:1000

[4] RNase T1 (concentração inicial: 1U/μl): 1; 1:10; 1:100; 1:1000

As reações foram incubadas a temperatura ambiente por 15 minutos e em seguida interrompidas através da adição de 20μl do tampão de inativação/precipitação (Ambion). Uma solução de RNA ao qual não foi adicionada enzima (controle negativo) foi submetido ao mesmo tempo de incubação que os RNAs tratados com RNases.

Após a adição do tampão de inativação/precipitação, as soluções foram incubadas a -20°C por 15 minutos e centrifugadas a 13.000rpm por 15 minutos. O precipitado formado foi “lavado” com etanol 75%, incubado em banho seco a 37°C por 5 minutos para evaporação do etanol residual e ressuspendido em 10μl de H₂O DEPC.

2.6. “Primer extension” (síntese do cDNA)

O RNA purificado e ressuspendido foi então submetido a reação de síntese da fita de DNA complementar (cDNA - “primer extension”). A reação de síntese do cDNA foi realizada segundo protocolo que acompanha a enzima M-MLV Reverse Transcriptase (Invitrogen), o qual consiste na preparação de uma solução inicial contendo aproximadamente 600ng de RNA, 0,5mM de dNTPs (concentração final em 20μl), 2pmol do oligo complementar à extremidade 3' da molécula de RNA sintetizada, marcado com o fluoróforo (“dye”) VIC (Applied Biosystems), o qual possui espectro de coloração verde, e água para 12μl. Os "oligos" utilizados na síntese das moléculas de cDNA foram: (1) 28S-R (5'-[VIC]GGTAATCCCATATGAGTTGAGG-3') [para os RNAs *ChITS2*_{I,II,III,IV}; *ChITS2*_{III,IV} e *GmITS2*_{I,II,III,IV}]; (2) 28S-R-*Md* (5'-[VIC]GGTAGTCCCATATGAGTTGAGG-3') [*MdITS2*_{I,II,III,IV}]; e (3) ITS2-R-*Ch* (5'-[VIC]GAGGTTTTGTATCTTTAGC-3') [*ChITS2*_{I,II}]. Essa solução inicial foi aquecida a 65°C por 5 minutos e em seguida foram adicionados 50mM Tris-HCl pH8,3, 75mM KCl, 5mM MgCl₂, 10mM DTT e 40U de RNase OUT

(Invitrogen), com a solução sendo incubada em seguida a 37°C por 2 minutos. Ao final desse período, 200U de M-MLV Reverse Transcriptase (Invitrogen) foram adicionadas à solução e a mesma foi incubada a 37°C por 50 minutos, sendo a reação interrompida por aquecimento a 70°C por 15 minutos.

Após a reação de síntese da fita de cDNA, as soluções foram purificadas segundo protocolo de purificação para amostras de sequenciamento em capilar, a qual consiste na precipitação do DNA com etanol 80% seguida por “lavagem” do precipitado com etanol 70%.

2.7. Eletroforese em capilar e análise dos cromatogramas

O cDNA precipitado foi ressuspendido em 17µl de Hi-Di formamida [metanamida] (Applied Biosystems) e, em seguida, foram adicionados 3µl do marcador de tamanho molecular GeneScan 600LIZ (Applied Biosystems), com espectro de coloração laranja. As amostras foram então submetidas à eletroforese em capilar em um sequenciador automático ABI 3130XL.

Os cromatogramas resultantes foram analisados utilizando-se o programa Peak Scanner v1.0 (Applied Biosystems), o qual realiza a determinação dos tamanhos dos fragmentos submetidos à eletroforese em capilar. Nos casos em que o programa não foi capaz de determinar o tamanho dos fragmentos, esse procedimento foi realizado manualmente baseando-se nos tamanhos conhecidos dos picos gerados pelo padrão GeneScan 600LIZ.

Os dados referentes aos tamanhos dos fragmentos bem como da altura (“height”) de cada pico, o qual é uma medida da intensidade do sinal de cada fragmento, foram transferidas para uma planilha excel, classificados de acordo com a intensidade do sinal de cada fragmento e então mapeados na sequência e na estrutura secundária predita *in silico* do

ITS2 para as espécies consideradas. As estruturas secundárias *in silico* de *C. hominivorax*, *M. domestica* e *G. morsitans* foram preditas através do programa Mfold v2.3 (Zuker 2003) utilizando-se os parâmetros padrões para predição de estrutura, com exceção da temperatura, alterada para 25°C. A estrutura de menor energia livre global foi selecionada e utilizada no mapeamento do padrão de digestão obtido *in vitro*. O programa RNAfold (Hofacker 2003) também foi utilizado para avaliar a existência de modelos alternativos para a estrutura secundária adquirida pelo ITS2. Ambos os programas utilizam como base o critério de mínima energia livre para predição das estruturas secundárias. No programa RNAfold, as informações de energia livre são complementadas pelo cálculo de funções de partição (funções que descrevem as propriedades estatísticas de sistemas em equilíbrio termodinâmico) e das probabilidades de pareamento de bases na predição da estrutura, resultando em uma única estrutura no final do processo.

As estruturas apresentadas foram desenhadas com auxílio do programa VARNA (Darty et al. 2009). A nomenclatura dos motivos de estrutura secundária seguiu a utilizada por Gesteland & Atkins (1993).

3. Resultados

3.1. Amplificação da região ITS2 e clonagem no vetor pHST7.0

As clonagens das regiões do ITS2, após amplificação via PCR, de *C. hominivorax* ($ChITS2_{I,II,III,IV}$, $ChITS2_{I,II}$ e $ChITS2_{III,IV}$), *M. domestica* ($MdITS2_{I,II,III,IV}$) e *G. morsitans* ($GmITS2_{I,II,III,IV}$) são mostradas nas Figuras 3 e 4.

3.2. Transcrição *in vitro*

As reações de transcrição *in vitro* realizadas primeiramente para a obtenção de moléculas de RNA referentes à região amplificada do $ChITS2_{I,II,III,IV}$ (Figura 5) revelaram a

presença de um padrão de bandas de tamanho inferior ao esperado para o transcrito completo com base no DNA molde (~400 bases) (Figura 5A).

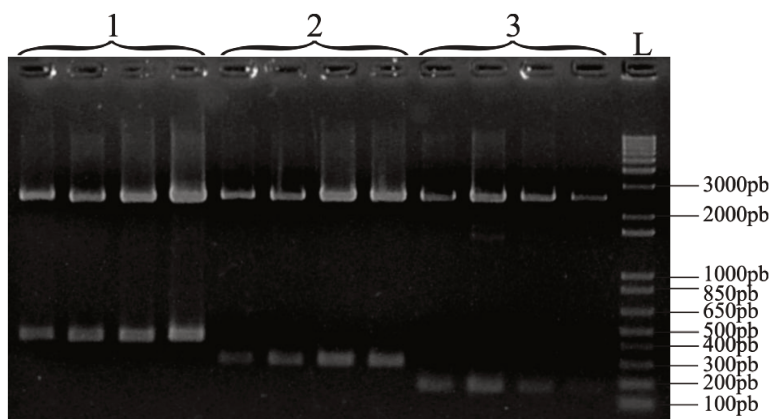


Figura 3. Gel de agarose 1,5%, corado em brometo de etídeo, mostrando a clonagem da região amplificada do ITS2 em *C. hominivorax* no vetor pHST7.0 (~2600pb). L = 1Kb Plus DNA Ladder (Invitrogen); 1 = *Ch*ITS2_{I,II,III,IV}; 2 = *Ch*ITS2_{I,II}; e 3 = *Ch*ITS2_{III,IV}.

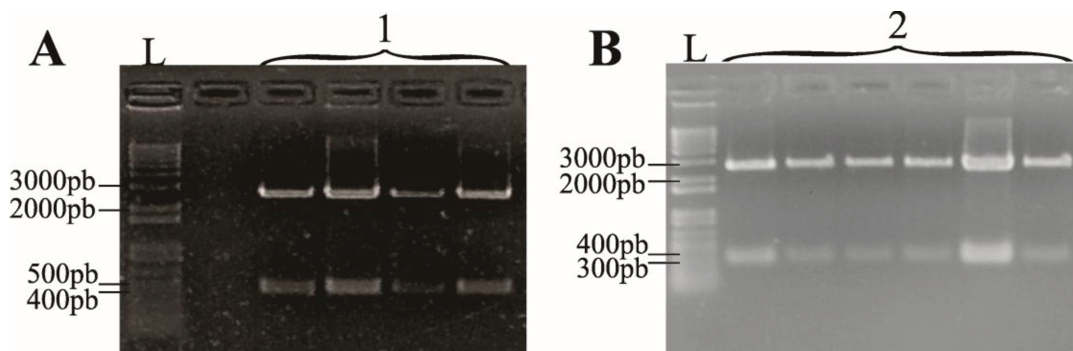


Figura 4. Gel de agarose 1,5%, corado em brometo de etídeo, mostrando a clonagem da região amplificada do ITS2 de (A) *M. domestica* e (B) *G. morsitans* no vetor pHST7.0 (~2600pb). L = 1Kb Plus DNA Ladder (Invitrogen); 1 = *Md*ITS2_{I,II,III,IV}; 2 = *Gm*ITS2_{I,II,III,IV}.

A geração de transcritos abortados durante o processo de transcrição *in vitro* (e também *in vivo*) de moléculas com uma estrutura secundária estável associada é bastante conhecida (e.g. *in vivo*: Jeng et al. 1990; Uptain et al. 1997; Pan & Sosnick 2006) e, embora

existam alterações no protocolo de transcrição que podem ser utilizadas para minimizar sua formação (“Practical Tips for *in vitro* transcription – generating full-length transcripts”, boletim técnico nº 182 - Ambion / Applied Biosystems - disponível em http://www.ambion.com/techlib/tb/tb_182.html), incluindo alterações no tempo e temperatura de incubação da reação, nenhuma das modificações testadas resultou na eliminação dos transcritos abortados sem comprometer a eficiência da reação de transcrição *in vitro*.

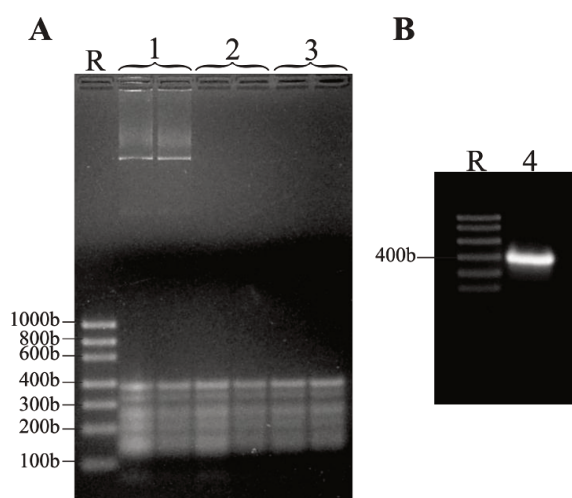


Figura 5. Gel de agarose 2%, corado em brometo de etídeo, mostrando a reação de transcrição da região completa do ITS2 (C_h ITS2_{I,II,III,IV}) (A) antes e (B) depois do isolamento e purificação a partir de gel de agarose. R = RiboRuler RNA Ladder Low Range (Fermentas); 1 = reação de transcrição antes da digestão com DNase; 2 = reação após o tratamento com DNase (eliminação do DNA molde); 3 = reação já tratada com DNase após a purificação por fenol/clorofórmio; 4 = verificação da banda isolada e purificada a partir de gel de agarose referente ao transcrito completo do C_h ITS2_{I,II,III,IV}.

Embora a natureza das técnicas utilizadas nas etapas seguintes tornem irrelevantes a presença desses transcritos abortados, uma vez que o oligo marcado com fluorescência utilizado na etapa de síntese da molécula de cDNA se hibridiza na extremidade 3' das

moléculas de RNA transcritos, sendo que a mesma estará presente somente nas moléculas de RNA transcritas por completo, uma tentativa de purificação da banda referente ao RNA transcrito completo a partir de gel de agarose foi realizada, utilizando-se para isso o kit de purificação "Invisorb DNA CleanUp" (Invitek). Apesar de uma aparente integridade do RNA purificado a partir do gel de agarose ter sido observada (Figura 5B), a análise dos resultados das etapas subsequentes revelou uma significativa degradação do RNA total utilizado nas análises (Figura A1 - Anexo A). Sendo assim, as reações de transcrição *in vitro* para as demais regiões do ITS2 isoladas (Figuras 6 e 7) e as etapas subsequentes foram realizadas sem a etapa de isolamento e purificação da banda do RNA transcrito completo do gel de agarose.

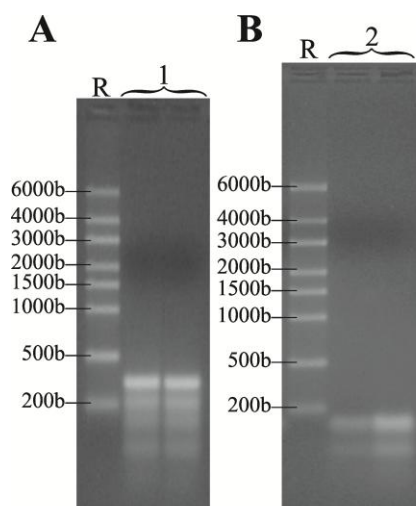


Figura 6. Gel de agarose 2%, corado em brometo de etídeo, mostrando a reação de transcrição dos domínios (A) III e IV e (B) I e II do ITS2 de *C. hominivorax*, isoladamente. R = RiboRuler RNA Ladder High Range (Fermentas); 1 = *ChITS2*_{III,IV}; 2 = *ChITS2*_{I,II}; ambos após purificação por método de fenol/clorofórmio

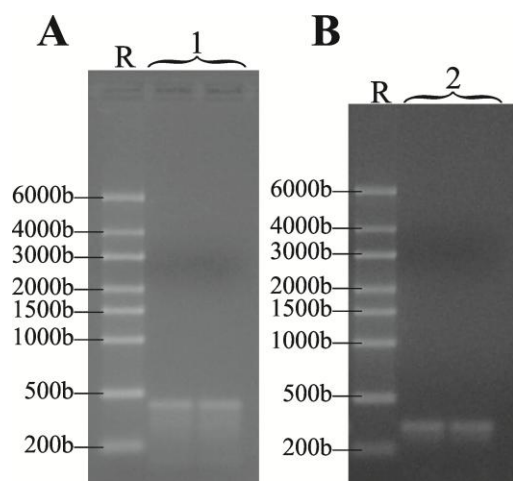


Figura 7. Gel de agarose 2%, corado em brometo de etídeo, mostrando a reação de transcrição da região do ITS2 em (A) *M. domestica* e (B) *G. morsitans*. R = RiboRuler RNA Ladder High Range (Fermentas); 1 = *MdITS2*_{I,II,III,IV}; 2 = *GmITS2*_{I,II,III,IV}; ambos após purificação por método de fenol/clorofórmio.

3.3. Eletroforese em capilar e análise dos cromatogramas

A análise dos cromatogramas apresentados nas Figuras A1-A7 revelou a existência de múltiplos picos referentes ao fragmento de RNA completo (e íntegro) transcrito para cada espécie, sendo que o tamanho inferido desses fragmentos diferia, em algumas bases, do tamanho esperado do RNA transcrito com base no sequenciamento do DNA molde clonado (Tabela 1). Sendo assim, antes do mapeamento dos fragmentos resultantes da digestão com RNases ser efetuado, foi necessário inferir qual fragmento, referente ao RNA transcrito completo, apresenta-se em maior proporção na solução e a partir do qual o padrão de fragmentos gerado a partir da digestão com as RNases foi mais provavelmente obtido.

Para tanto, a análise dos fragmentos gerados pelas RNases de padrão de digestão mais simples, com um número menor de sítios de clivagem esperado a partir das estruturas secundárias preditas *in silico* (e consequentemente de fragmentos e picos presentes nos

cromatogramas) (e.g. RNase T1) é bastante informativo. Além disso, o conhecimento de que pode existir uma significativa variação no número de repetições nucleotídicas em regiões homopoliméricas, como a região poli-T presente no domínio IV de *C. hominivorax* (Marinho et al. 2011), fato este provavelmente relacionado à ocorrência de um fenômeno de deslizamento ("slippage") das enzimas polimerases (tanto de DNA quanto de RNA - Chamberlin & Berg; 1962; Uptain et al. 1997), torna possível também inferir qual região da molécula é responsável pela variação de tamanho observado e na qual, espera-se, exista um padrão conflitante quando do mapeamento dos dados de clivagem a partir das RNases nas estruturas preditas *in silico*.

Tabela 1. Tamanhos esperados (com base no sequenciamento do DNA molde clonado), observados e considerados no mapeamento dos fragmentos gerados a partir da digestão dos RNAs transcritos da região do ITS2 nas espécies *C. hominivorax* (*Ch*), *M. domestica* (*Md*) e *G. morsitans* (*Gm*). O valor mostrado na coluna "Considerado" foi estabelecido a partir da análise *a posteriori* do padrão de digestões enzimáticas e do conhecimento dos possíveis pontos de expansão/contração da sequência do ITS2 (maiores detalhes são dados no texto).

RNA transcrito	Esperado	Observado	Considerado
<i>Ch</i> ITS2 _{I,II,III,IV}	412	404-410	405
<i>Ch</i> ITS2 _{I,II}	281	273-277	275
<i>Ch</i> ITS2 _{III,IV}	155	148-151	150
<i>Md</i> ITS2 _{I,II,III,IV}	422	423-426	424
<i>Gm</i> ITS2 _{I,II,III,IV}	308	303-305	304

Sendo assim, a partir dessas duas informações, foi possível determinar o tamanho de molécula referente ao RNA completo transcrito, para cada um dos RNAs considerados, mais plausível de ter gerado o padrão de fragmentos observados a partir da digestão com RNases (Tabela 1). Os dados das reações de digestão dos RNAs transcritos e estruturados

foram então mapeados na estrutura secundária predita *in silico* do ITS2 (Figuras A3 a A7 - Anexo A), utilizando um código de cores que reflete a intensidade (altura) dos picos observados nos cromatogramas, os quais são proporcionais à frequência dos fragmentos gerados pela ação de cada uma das RNAses utilizadas em diferentes diluições (nesse passo, apenas os fragmentos maiores que 20 bases foram considerados, devido à baixa resolução observada na porção inicial dos cromatogramas):

[1] $X < 500$ = círculos azuis com a base em branco (nessa categoria foram considerados apenas os picos com intensidade $X < 500$ que se apresentavam maiores que os picos circundantes e que apareciam consistentemente em diferentes diluições da enzima);

[2] $500 < X < 1000$ = círculos azuis com a base em preto;

[3] $1000 < X < 2000$ = círculos verdes com a base em preto;

[4] $2000 < X < 3000$ = círculos amarelos com a base em preto;

[5] $3000 < X < 4000$ = círculos laranja claro com a base em preto;

[6] $4000 < X < 5000$ = círculos laranja escuro com a base em preto;

[7] $5000 < X < 6000$ = círculos vermelhos com a base em preto.

Nesta etapa, as intensidades dos picos referentes aos fragmentos também encontrados nos cromatogramas dos RNAs não-tratados com RNAses (controle negativo) foram subtraídas das intensidades dos picos encontrados nos cromatogramas dos RNAs tratados com RNAses (experimentais).

O padrão de fragmentos (e suas intensidades) gerados nas análises de digestão com as diferentes RNAses utilizadas e em suas diferentes diluições foram combinados e utilizados na confecção de um modelo da provável estabilidade da estrutura secundária do ITS2 nas espécies em estudo, conforme a predição realizada *in silico* pelo programa mfold, com algumas extrapolações feitas para as bases não clivadas por nenhuma enzima (Figuras

A8-A10 - Anexo A). Nesses modelos, os dados obtidos *in vitro* que não condizem com o predito *in silico* podem representar possíveis pontos conflitantes entre ambos os modelos de estrutura secundária (*in vitro* e *in silico*), os quais podem ser devidos a: (1) possíveis regiões de dinâmica conformacional na molécula (clivagem concomitante por enzimas específicas para simples-fita - RNAses I, A e T1 - e dupla-fita - RNase V1); ou (2) possíveis contradições reais entre os modelos (padrão de digestão observado diferente do esperado pela estrutura predita *in silico*).

Nas Figuras A8 a A10 (Anexo A), a escala de tons de cinza indica a probabilidade de as regiões consideradas estarem em (1) dupla-fita (cinza escuro - clivagem apenas pela RNase V1 ou RNase V1 > RNases I, A e T1); (2) em dinâmica mais acentuada entre simples e dupla-fita (cinza médio - clivagem de igual intensidade tanto pela RNase V1 quanto pelas RNases I, A e T1); e (3) simples-fita (cinza claro - clivagem apenas pelas RNases I, A e T1 ou RNases I, A e T1 > RNase V1). As regiões preditas como estando em simples-fita na estrutura *in silico* foram deixadas propositalmente em branco, a não ser quando clivadas também pela RNase V1.

Em *C. hominivorax* (Figura A8 - Anexo A), o padrão de fragmentos gerados pelas análises de digestão *in vitro* parece bastante condizente com o modelo predito *in silico*. Os fragmentos gerados pela RNase T1 (que cliva à 3' de Gs em simples-fita), parece corroborar a existências das ramificações IIIa e IIIc do domínio III, uma vez que os Gs em simples-fita presentes nas alças terminais desses domínios (G₁₉₃, G₁₉₄ e G₁₉₆ - IIIa; G₂₉₃ e G₂₉₅ - IIIc) foram consistentemente clivados por essa enzima, gerando fragmentos com altas intensidades (X>5000) nos cromatogramas das Figuras A3 e A4 (Anexo A).

Das hélices presentes na estrutura do ITS2 predita *in silico* (Figura A8A), aquelas presentes nos domínios definidos na estrutura secundária gerada segundo o modelo em

"anel" do ITS2 (Marinho et al. - 2011 - Figura A8C) parecem ser as mais estáveis, com exceção da hélice H8 (porção proximal do domínio III) e da hélice H18 (domínio IV), embora esta última se localize na porção da estrutura onde há uma provável ocorrência de heterogeneidade no número de repetições nucleotídicas nas regiões homopoliméricas (explicando o padrão contraditório dos picos mapeados nesta região). Em geral, nestas hélices, os pontos mais estáveis se encontram nas imediações de pares G-C (mais estáveis em termos de pareamento - Tinoco & Bustamante 1999), enquanto os pontos mais instáveis, onde provavelmente há uma maior alternância entre os estados pareado e não-pareado, se encontram nas imediações de motivos de simples-fita, como junções, "bojos" e alças internas e terminais.

Por fim, o ponto da estrutura que parece menos estável se localiza na região da junção 1 (J1), onde uma alternância entre os estados pareado e não-pareado das hélices H1 e H8, com um eventual rearranjo dos pareamentos nessa hélice, pode gerar a estrutura mostrada na Figura A8B, a qual foi predita *in silico* pelo programa RNAfold. Como essa dinâmica de pareamentos pode também ter um efeito destabilizador na hélice H2, acarretamento em uma possível união das junções J1 e J2, é possível que a estrutura predita pelo modelo em "anel" do ITS2 (estabelecida a partir da modelagem comparativa e por homologia da região do ITS2 em diferentes espécies de Calliphoridae) surja como um estado intermediário entre os dois modelos de estrutura preditos pelas análises *in silico*.

Em *M. domestica*, apesar do padrão de fragmentos gerados pelas análises de digestão do RNA não ter sido consistente o suficiente para gerar um modelo de estabilidade e dinâmica de pareamento como o de *C. hominivorax*, o modelo geral obtido (Figura A9A) parece corroborar os domínios preditos através de modelagem comparativa e por homologia pelo modelo em "anel" do ITS2 (Figura A9C). O padrão de estabilidade das

hélices, as quais são mais estáveis nas imediações de pareamentos G-C, e de possível alternância entre estados pareados e não-pareados, nas regiões em dupla-fita adjacentes a elementos em simples fita (alças, "bojos" e junções), foi semelhante ao encontrado em *C. hominivorax*.

Assim como também encontrado em *C. hominivorax*, uma possível dinâmica conformacional das hélices adjacentes à junção J1 (hélices H10, H22 e H23), acompanhada de uma reestruturação dos pareamentos nessas hélices, pode gerar, em um estágio intermediário, o "anel" central do qual emergem os domínios mostrados na Figura A9C, como constatado a partir da análise das duas estruturas possíveis preditas *in silico* pelos programas mfold (Figura A9A) e RNAfold (Figura A9B). Essa dinâmica e consequente reestruturação das hélices pode levar a alterações significativas no domínio IV da molécula, o qual pode se apresentar como uma hélice única (Figura A9B) ou subdividido em duas hélices (Figura A9A). A análise das duas estruturas preditas *in silico* aponta também para a provável existência de uma dinâmica conformacional envolvendo as junções J4 e J5 e a hélice formada entre elas (H14), sendo que as hélices H15 (subdomínio IIIa) e H20 (subdomínio IIIc) têm composições diferentes nos dois modelos de estrutura propostos (Figuras A9A e B).

Para *G. morsitans*, a estrutura predita *in silico* pelo programa mfold (Figura A10A), na qual foram mapeados os fragmentos gerados nas análises de digestão *in vitro*, possui grande similaridade com a estrutura predita através de modelagem comparativa segundo o modelo em "anel" (Figura A10C). Com exceção da hélice H2, formada pela extremidade 3' da subunidade ribossomal 2S e 5' do ITS2, a estrutura é bastante semelhante também à descrita para *D. melanogaster* por Young & Coleman (2004), incluindo a presença do subdomínio IIa, o qual é provavelmente estável (segundo o modelo de estabilidade das

regiões em dupla-fita inferido a partir do padrão de digestão do RNA *in vitro*) e se encontra ausente em *C. hominivorax* e *M. domestica*. Ambas as estruturas (subdomínio IIa e hélice H2), encontram-se ausentes, contudo, na estrutura alternativa predita *in silico* pelo programa RNAfold (Figuras A10B). No caso da hélice H2, como sua formação envolve a extremidade 3' da subunidade ribossomal 2S, a qual é uma subdivisão da subunidade 5,8S (sendo homóloga da porção 3' dessa subunidade quando a mesma não é subdividida - Jordan et al. 1976), e o seu pareamento com a extremidade 5' da subunidade 28S tem importância fundamental no processamento do transcrito primário de rRNA (Peculis & Greer 1998), é provável que esta hélice não esteja presente ou se forme apenas de maneira instável e efêmera *in vivo*.

Além disso, assim como observado nas estruturas preditas para *C. hominivorax* e *M. domestica*, a junção J1 e suas hélices adjacentes parece ser a região de maior instabilidade da molécula e a possível ocorrência de uma dinâmica conformacional nessa região, com uma constante alteração entre os estados pareados e não-pareados preditos nas Figuras A10A e A10B, pode resultar na formação da estrutura predita pelo modelo em "anel" (Figura A10C) como estado intermediário.

4. Discussão

4.1. Eletroforese em capilar e análise dos cromatogramas

A análise do padrão de digestão dos RNAs referentes à região ITS2 de diferentes espécies representativas do clado Calyptratae, embora de baixa resolução e insuficiente para gerar, isoladamente, um modelo de estrutura secundária, parecem corroborar os modelos de estruturas secundárias preditos *in silico* para as moléculas, com muitos dos elementos estruturais essenciais presentes nestes modelos sendo comprovados *in vitro*. Nas regiões da estrutura secundária onde parece haver maior discordância entre o padrão de

fragmentos gerados pelas RNAses *in vitro* e as estruturas preditas *in silico*, a provável dinâmica conformacional e consequente alternância dos estados pareados e não-pareados que pode ocorrer na estrutura formada *in vitro* é uma explicação plausível.

Essa possível dinâmica conformacional na região do cerne da estrutura secundária do ITS2 (de onde emergem os quatro domínios comumente considerados na molécula, localizado logo após a hélice proximal formada pelo pareamento das extremidades 3' da subunidade 2S e 5' da subunidade 28S) envolveria a reorganização dos pareamentos das hélices adjacentes a primeira junção encontrada na molécula, resultando na alternância entre duas (ou mais) estruturas secundárias possíveis, ambas seguindo o modelo em "grampo" para o ITS2 e com números similares de pareamento (e, consequentemente, bastante similares em termos de estabilidade termodinâmica). Nesse modelo de dinâmica conformacional, o modelo em "anel" da estrutura secundária do ITS2 surgiria como um estágio intermediário entre as duas estruturas em "grampo".

Essa região do cerne do ITS2 parece ser também a base para o modelo de dinâmica conformacional proposto *in vivo* para a molécula em leveduras (Côté et al. 2002), embora nesse modelo a alternância se dê entre os modelos em "grampo" e em "anel" da estrutura secundária da molécula, havendo uma separação temporal da presença de ambos no processamento do transcrito primário do rDNA. Nesse caso, a presença e atuação de fatores externos ao ITS2 e ao próprio "cluster" de rDNA (como, por exemplo, snoRNAs - Michot et al. 1999) podem atuar na estabilização do modelo em "anel", o qual possui menor número de sítios pareados e seria, termodinamicamente, mais instável.

Em termos funcionais, tanto as estruturas preditas segundo o modelo em "grampo" quanto o modelo em "anel" do ITS2 parecem conter as regiões mais conservadas, em termos de sequência primária e estrutura secundária, nas mesmas conformações estruturais.

Isto inclui a pirimidina não pareada na região medial do domínio II (C₁₀₁ em *C. hominivorax*, C₁₀₄ em *M. domestica* e U₉₃ em *G. morsitans*) e o motivo de sequência primária GUCUAGCAU na porção distal do domínio III (lado 5' das hélices H15/H16 em *C. hominivorax*, H18/19 em *M. domestica* e H14/15 em *G. morsitans*), ambos descritos por Young & Coleman (2004) em *D. melanogaster* e nas demais espécies do clado Schizophora analisadas. Como a conservação de motivos de sequência e estrutura secundária em geral sugere uma implicação funcional essencial dos elementos conservados (Gesteland & Atkins 1993), é possível assumir que ambos os modelos de estrutura secundária apresentem um papel no processamento da molécula. As hélices e subdomínios estruturais formados que são exclusivos de cada modelo são provavelmente necessários em momentos específicos do processamento da molécula, visto a dinâmica conformacional temporalmente definida sugerida por Côté et al. (2002) em leveduras.

Neste sentido, a grande variabilidade encontrada no domínio IV da estrutura secundária do ITS2 parece sugerir uma pequena, senão ausente, implicação funcional dessa região no processamento da molécula. Como este domínio é basicamente composto pelos nucleotídeos A e T (U), muitas vezes organizados na forma de repetições homopoliméricas consecutivas, comumente apresentando diferenças em termos de número de repetições, é possível que a dinâmica de evolução desse domínio se dê muito mais pela expansão/contração das sequências homopoliméricas, com eventuais substituições que possibilitem o pareamento entre as bases sendo fixadas, do que pelo acúmulo de substituições que visem manter uma relação de equilíbrio entre flexibilidade e estabilidade na estrutura secundária (como sugerido por Friedrich & Tautz 1997), como deve ocorrer com os demais domínios do ITS2. Nessa relação de equilíbrio, pares G-C, mais estáveis (Xia et al. 1998; Svozil et al. 2010; resultados obtidos aqui), deveriam estar presentes em

locais "estratégicos" da molécula, onde garantiriam a formação de elementos estruturais locais necessários para o correto processamento da mesma (Marinho et al. 2011). Por outro lado, pares A-U e G-U, menos estáveis, poderiam estar presentes em regiões onde uma possível dinâmica conformacional, na qual há uma alternância entre estados pareados e não-pareados na estrutura, auxiliariam no melhor reconhecimento e acomodação dos fatores que atuam no processamento do ITS2, diminuindo assim as restrições funcionais atuando sobre a molécula.

Sendo assim, em análises genético-evolutivas que considerem informações estruturais durante sua elaboração, em especial análises de inferência filogenética utilizando dados moleculares, a utilização do modelo em "anel" para a estrutura secundária do ITS2, o qual é em geral obtido a partir de uma combinação entre predição *in silico* e um posterior refinamento do modelo obtido usando abordagens de modelagem comparativa e por homologia, parece ser uma simplificação justificável. Isto porque o mesmo representa um modelo mais "estático" da estrutura secundária, em contrapartida ao modelo em "grampo", sugerido pela análises das estruturas preditas exclusivamente *in silico* e pelas análises realizadas *in vitro*, o qual é mais "dinâmico" e pode apresentar dois ou mais estados alternativos possíveis.

Além disso, os programas atualmente disponíveis que são capazes de incorporar informações estruturais em análises de inferência filogenética não utilizam as estruturas individuais preditas para cada táxon, mas ao invés disso empregam um modelo de estrutura secundária que é um consenso das estruturas preditas individualmente. Deste modo, a incorporação de um modelo estático nas análises representaria uma vantagem em termos de tempo e processamento das análises uma vez que não é necessário estimar e calcular

parâmetros para diferentes estados conformacionais possíveis, embora no processo uma quantidade sensível de informação possa ser perdida.

No que concerne à evolução estrutural da molécula do ITS2 em Calyptratae, a partir da hipótese atual mais comumente aceita para a relação entre suas superfamílias - (Hippoboscoidea, (Muscoidea, Oestroidea)), sendo Hippoboscoidea o grupo mais próximo do clado Acalyptratae e, conseqüentemente, de *D. melanogaster* (McAlpine 1989; Kutty et al. 2010) - é possível sugerir que a principal alteração conformacional sofrida pela molécula foi o aumento considerável, em termos de número de pareamentos e extensão, do domínio III, com a conseqüente perda do subdomínio IIa, encontrado apenas em Hippoboscoidea e em *D. melanogaster* (Young & Coleman 2004). Essa expansão do domínio III, a qual pode ter resultado tanto do acúmulo de substituições no subdomínio IIa, desfazendo pareamentos locais e permitindo novos pareamentos com regiões já presentes no domínio III, quanto de eventos de inserção e/ou eventuais duplicações de porções locais neste domínio, parece ter sido acompanhada do surgimento de uma possível zona de instabilidade na porção mediana do domínio III. Essa região de instabilidade se apresenta como uma possível área de dinâmica conformacional na molécula, onde diferentes hélices laterais podem surgir como ramificações do domínio III (subdomínio IIIa e IIIc), o qual era até então formado por uma única "haste" (do inglês "stem"). Em Oestroidea, o acúmulo e fixação de substituições nessa região de dinâmica conformacional parece ter se direcionado no sentido de estabilização das ramificações ao invés da recuperação do padrão de uma única "haste".

Estudos funcionais do processamento do ITS2 *in vivo* em Calyptratae poderiam definir se esta estabilização das ramificações do domínio III resultou em algum aperfeiçoamento ou melhoria no processamento da molécula em Oestroidea (caso tenha sido acompanhado por mutações e alterações conformacionais correlacionadas nas

proteínas que processam a molécula na célula) ou se foi alcançado apenas no sentido de estabilizar a molécula como um todo, garantindo que elementos e motivos estruturais, já presentes nas estruturas secundárias formadas nos demais grupos de Calyptratae, sejam formados. Como o processamento *in vivo* do ITS2 é pouco conhecido na maioria das espécies de eucarioto, com exceção de leveduras, qualquer discussão nesse sentido é apenas especulativa.

5. Conclusões

Atualmente, o uso de métodos de predição *in silico* para estabelecimento de modelos de estrutura secundária de moléculas de DNA e RNA é bastante difundido. Quando há a disponibilidade de múltiplas sequências da mesma região para diferentes táxons relacionados, a combinação das abordagens de predição *in silico* e modelagens comparativa e por homologia tem resultado em um refinamento significativo dos modelos estruturais propostos.

No que concerne a região do ITS2, sua utilização crescente como um marcador molecular em análises genético-evolutivas, associada com o estabelecimento de um modelo de estrutura secundária bastante robusto, que parece ser conservado entre todos os táxons de eucariotos, estimula a incorporação da informação estrutural presente na molécula nas análises de inferência filogenética conduzidas em diferentes grupos. Acredita-se que a consideração de informações estruturais em análises de inferência filogenética, utilizando para isso modelos de substituição nucleotídica adequados, em geral resulta em um aumento da resolução das análises, uma vez que representa uma forma mais realista de modelar a evolução de regiões que possuem uma estrutura secundária associada bastante estável.

Para tanto, apesar da relativa confiança geralmente depositada nos modelos preditos *in silico*, é necessário estimar o quão precisos e acurados esses modelos são. Neste sentido,

as análises de digestão enzimática e posterior mapeamento dos fragmentos gerados nas possíveis estruturas preditas *in silico* conduzidas aqui *in vitro*, apesar de possuírem baixa resolução, permitiram confirmar a maioria dos elementos e motivos estruturais presentes nos modelos inicialmente estimados usando abordagens combinadas de predição computacional e modelagem comparativa.

Apesar do modelo em "grampo" da estrutura secundária do ITS2 ser termodinamicamente mais estável e, por consequência, mais frequentemente estabelecido pelos métodos de predição *in silico*, o modelo em "anel", estabelecido principalmente nas análises de modelagem comparativa e por homologia, é mais comumente considerado quando da incorporação de informações estruturais em análises genético-evolutivas. Por se tratar de um modelo "estático", sua adoção preferencial nessas análises se justificaria principalmente por eliminar o problema de lidar com eventuais dinâmicas conformacionais e uma multitude de estruturas possíveis, as quais são comuns no modelo em "grampo" (mais dinâmico), o que resultaria em um aumento considerável em termos de poder e tempo computacionais necessários para as análises. Essa simplificação, embora acarrete por outro lado na perda significativa de informação, uma vez que essa dinâmica conformacional de fato ocorre *in vivo* e *in vitro* com a molécula, é uma decisão bastante justificável.

Referências

- Alkemar G, Nygard O (2006) Probing the secondary structure of expansion segment ES6 in 18S ribosomal RNA. *Biochemistry* 45: 8067-8078
- Chamberlin M, Berg P (1962) Deoxyribonucleic acid-directed synthesis of ribonucleic acid by an enzyme from *Escherichia coli*. *Proc Natl Acad Sci USA* 15;48:81-94
- Coleman AW (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet* 19:370-375
- Côté CA, Greer CL, Peculis BA (2002) Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. *RNA* 8:786-797

- Darty K, Denise A, Ponty Y (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25:1974-1975
- Eddy SR (2004) How do RNA folding algorithms work? *Nature Biotechnology*. 22(11): 1457-1458
- Favaretto P, Bhutkar A, Smith TF (2005) Constraining ribosomal RNA conformational space. *Nucleic Acids Res* 33(16): 5106-5111
- Friedrich M, Tautz D (1997) An episodic change of rDNA nucleotide substitution rate has occurred during the emergence of the insect order Diptera. *Mol Biol Evol* 14: 644-653
- Gardner PP, Giegerich R (2004) A comprehensive comparison of comparative RNA structure prediction approaches. *BMC Bioinformatics*. 30(5): 140-158
- Geerlings TH, Vos JC, Raué HA (2000) The final step in the formation of 25S rRNA in *Saccharomyces cerevisiae* is performed by 5'=>3' exonucleases. *RNA* 6:1698-1703
- Gesteland, R.F.; Atkins, J.F. (editors) (1993). *The RNA World*. Cold Spring Harbor Press, NY;
- Gillespie JJ (2004) Characterizing regions of ambiguous alignment caused by the expansion and contraction of hairpin-stem loops in ribosomal RNA molecules. *Mol Phylogenet Evol* 33: 936-943.
- Gutell RR, Larsen N, Woese CR (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Rev* 58:10-26
- Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *Q Rev Biol* 66:411-453
- Hofacker IL (2003) Vienna RNA secondary structure server. *Nucleic Acids Research*. 31(13):3429-31;
- Holbrook SR (2008) Structural principles from large RNAs. *Annu Rev Biophys* 37:445-464
- Holbrook SR, Holbrook EL, Walukiewicz HE (2001) Crystallization of RNA. *Cellular and Molecular Life Sciences*. 58: 234-243.
- Infante MEV, Azeredo-Espin AML (1995) Genetic variability in mitochondrial DNA of screwworm, *Cochliomyia hominivorax* (Diptera: Calliphoridae), from Brazil. *Biochemical Genetics*. 33: 737-756
- Jeng ST, Gardner JF, Gumpert RI (1990) Transcription termination by bacteriophage T7 RNA polymerase at rho-independent terminators. *J Biol Chem* 265(7): 3823-3830
- Jordan BR, Jourdan R, Jacq B (1976) Late steps in the maturation of *Drosophila* 26 S ribosomal RNA: generation of 5.8 S and 2 S RNAs by cleavages occurring in the cytoplasm. *J Mol Biol* 101:85-105
- Joseph N, Krauskopf E, Vera MI, Michot B (1999) Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. *RNA* 8:786-797

- Kutty SN, Pape T, Wiegmann BM, Meier R (2010) Molecular phylogeny of the Calyptratae (Diptera: Cyclorrhapha) with an emphasis on the superfamily Oestroidea and the position of Mystacinobiidae and McAlpine's fly. *Syst Entomol* 35:614-635
- Leontis NB, Westhof E (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* 7:499-512
- Mai JC, Coleman AW (1997) The internal transcribed spacer 2 exhibits a common core secondary structure in greens algae and flowering plants. *J Mol Evol* 44: 258-271
- Marinho MAT, Junqueira ACM, Azeredo-Espin AML (2011) Evaluation of the internal transcribed spacer 2 (ITS2) as a molecular marker for phylogenetic inference using sequence and secondary structure information in blow flies (Diptera: Calliphoridae). *Genetica* 139(9): 1189-1207
- McAlpine JF (1989) Manual of nearctic Diptera 3. Agriculture Canada Monograph 32. Canadian Gov. Publ. Center, Quebec, Canada
- Mitchell P, Petfalski E, Shevchenko A, Mann M, Tollervey D (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'=>5' exoribonucleases. *Cell* 91:457-466
- Mitchell P, Petfalski E, Tollervey D (1996) The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes Dev* 10:502-513
- Pace NR, Thomas BC, Woese CR (1999) Probing RNA structure, function and history by comparative analysis. In "The RNA World, Second Edition". Gesteland RF, Cech TR, Atkins JF (eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp.113-141
- Pan T, Sosnick T (2006) RNA folding during transcription. *Annu Rev Biophys Biomol Struct* 35:161-175
- Peculis BA, Greer CL (1998) The structure of the ITS2-proximal stem is required for pre-rRNA processing in yeast. *RNA*. 4(12):1610-1622
- Reeder J, Höchsmann M, Rehmsmeier M, Voss B, Giegerich R (2006) Beyond Mfold: recent advances in RNA bioinformatics. *Journal of Biotechnology*. 124(1): 41-55.
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning - A Laboratory Manual*, 2nd Edition. Cold Spring Harbour Laboratory Press, New York
- Schlötterer C, Hauser MT, von Haeseler A, Tautz D (1994) Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. *Mol Biol Evol* 11:513-522
- Schultz J, Maisel S, Gerlach D, Muller T, Wolf M (2005) A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA*. 11: 361-364

- Svozil D, Hobza P, Sponer J (2010) Comparison of intrinsic stacking energies of ten unique dinucleotide steps in A-RNA and B-DNA duplexes. Can we determine correct order of stability by quantum-chemical calculations? *J Phys Chem B* 114:1191-203; Erratum in: *J Phys Chem B* (2010) 114:2547
- Tinoco IJr, Bustamante C (1999) How RNA folds. *J Mol Biol* 293(2):271-81
- Tzakos AG, Grace CR, Lukavsky PJ, Riek R (2006) NMR techniques for very large proteins and RNAs in solution. *Annual Review of Biophysics and Biomolecular Structure*. 35: 319-242.
- Uptain SM, Kane CM, Chamberlin MJ (1997) Basic mechanisms of transcript elongation and its regulation. *Annu Rev Biochem* 66:117-72
- van der Sande CA, Kwa M, van Nues RW, van Heerikhuizen H, Raué HA, Planta RJ (1992) Functional analysis of internal transcribed spacer 2 of *Saccharomyces cerevisiae* ribosomal DNA. *J Mol Biol* 223:899-910
- Veldman GM, Klootwijk J, van Heerikhuizen H, Planta RJ (1981) The nucleotide sequence of the intergenic region between the 5.8S and 26S rRNA genes of the yeast ribosomal RNA operon. Possible implications for the interaction between 5.8S and 26S rRNA and the processing of the primary transcript. *Nucleic Acids Res* 9:4847-4862
- Wolf M, Achtziger M, Schultz J, Dandekar T, Müller T (2005) Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) structures. *RNA*. 11:1616-1623
- Xia T, SantaLucia JJr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*. 37:14719-14735
- Yeh LCC, Lee JC (1990) Structural analysis of the internal transcribed spacer 2 of the precursor ribosomal RNA from *Saccharomyces cerevisiae*. *Journal of Molecular Biology*. 211: 699-712.
- Young I, Coleman AW (2004) The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a *Drosophila* example. *Mol Phylogenet Evol* 30:236-242
- Yu J, Thorne JL (2006) Dependence among sites in RNA evolution. *Mol Biol Evol* 23(8): 1525-1537.
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406-3415

Anexo A

Legenda do código de cores utilizado nas Figuras de A3 a A7:

[1] $X < 500$ = círculos azuis com a base em branco (nessa categoria foram considerados apenas os picos com intensidade $X < 500$ que se apresentavam maiores que os picos circundantes e que apareciam consistentemente em diferentes diluições da enzima);

[2] $500 < X < 1000$ = círculos azuis com a base em preto;

[3] $1000 < X < 2000$ = círculos verdes com a base em preto;

[4] $2000 < X < 3000$ = círculos amarelos com a base em preto;

[5] $3000 < X < 4000$ = círculos laranja claro com a base em preto;

[6] $4000 < X < 5000$ = círculos laranja escuro com a base em preto;

[7] $5000 < X < 6000$ = círculos vermelhos com a base em preto.

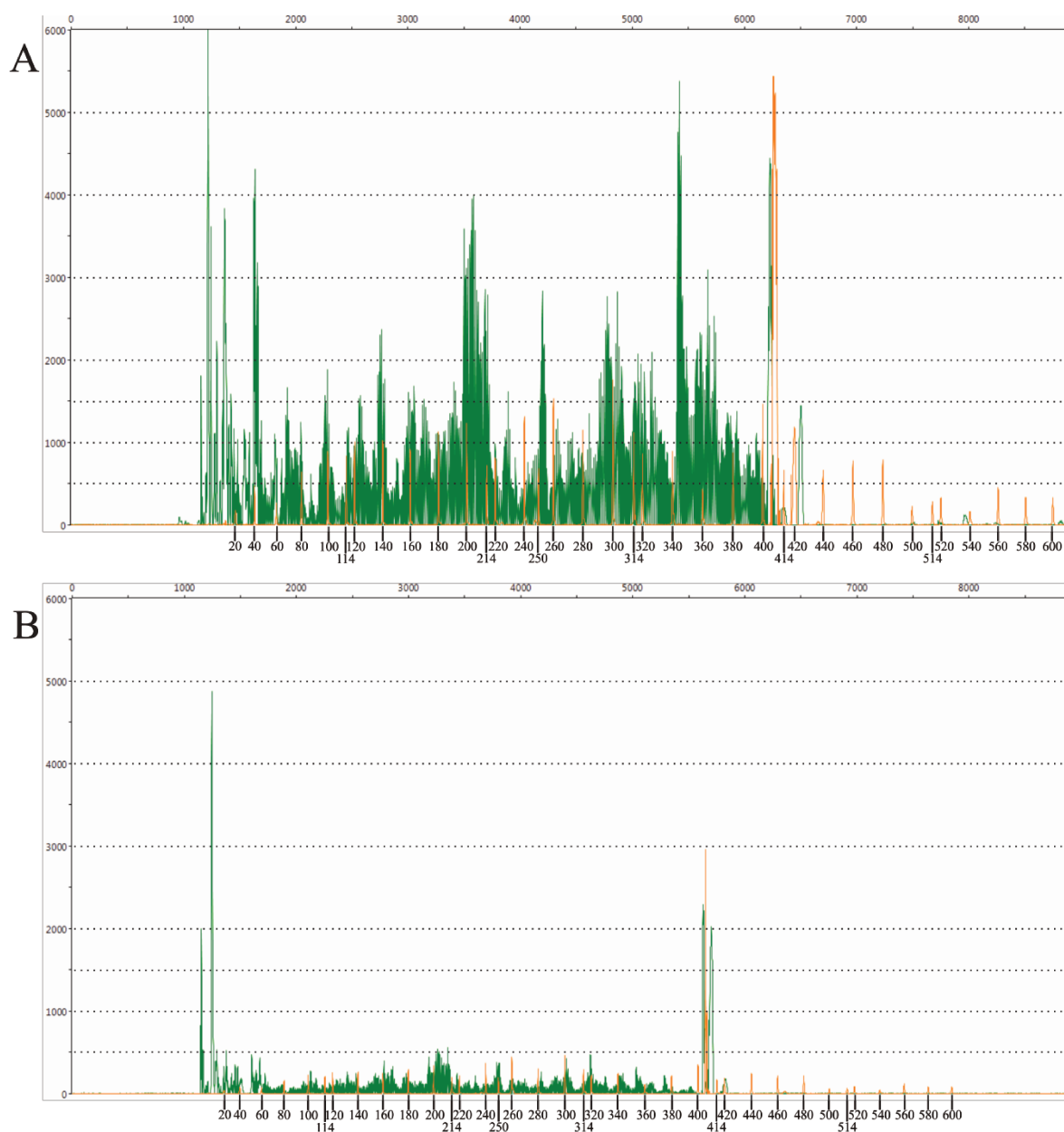


Figura A1. Cromatogramas referentes à região transcrita do *ChITS2_{I,II,III,IV}* não submetida a digestão com RNases (reações de controle negativo). (A) Transcrito completo purificado a partir de gel de agarose; (B) Reação de transcrição purificada exclusivamente com método de fenol/clorofórmio, sem isolamento do RNA transcrito completo a partir de gel de agarose.

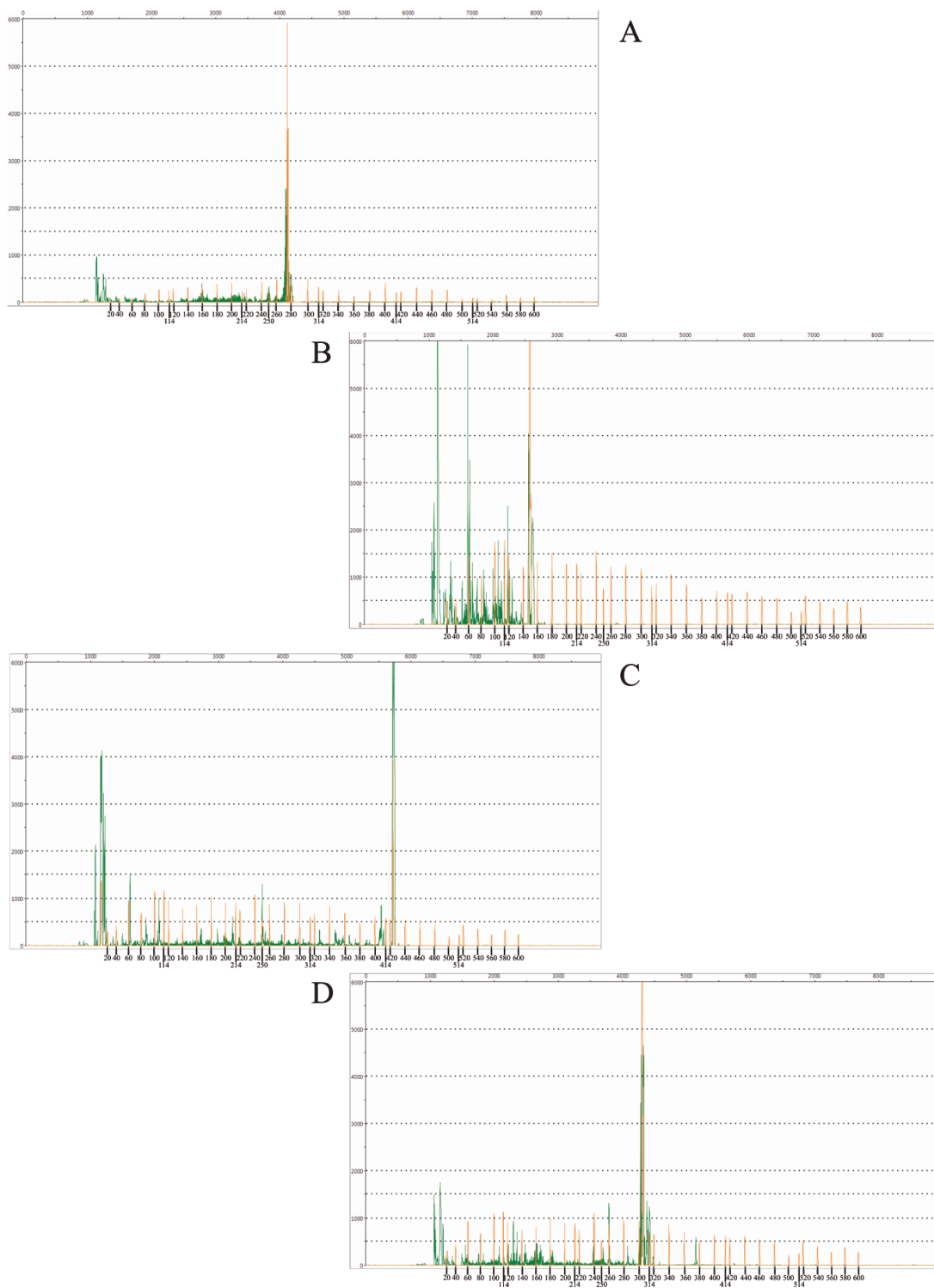


Figura A2. Cromatogramas referentes à região transcrita do ITS2 não submetida a digestão com RNases (reações de controle negativo). (A) *Ch*ITS2_{III,IV}; (B) *Ch*ITS2_{I,II}; (C) *Md*ITS2_{I,II,III,IV}; (D) *Gm*ITS2_{I,II,III,IV}.

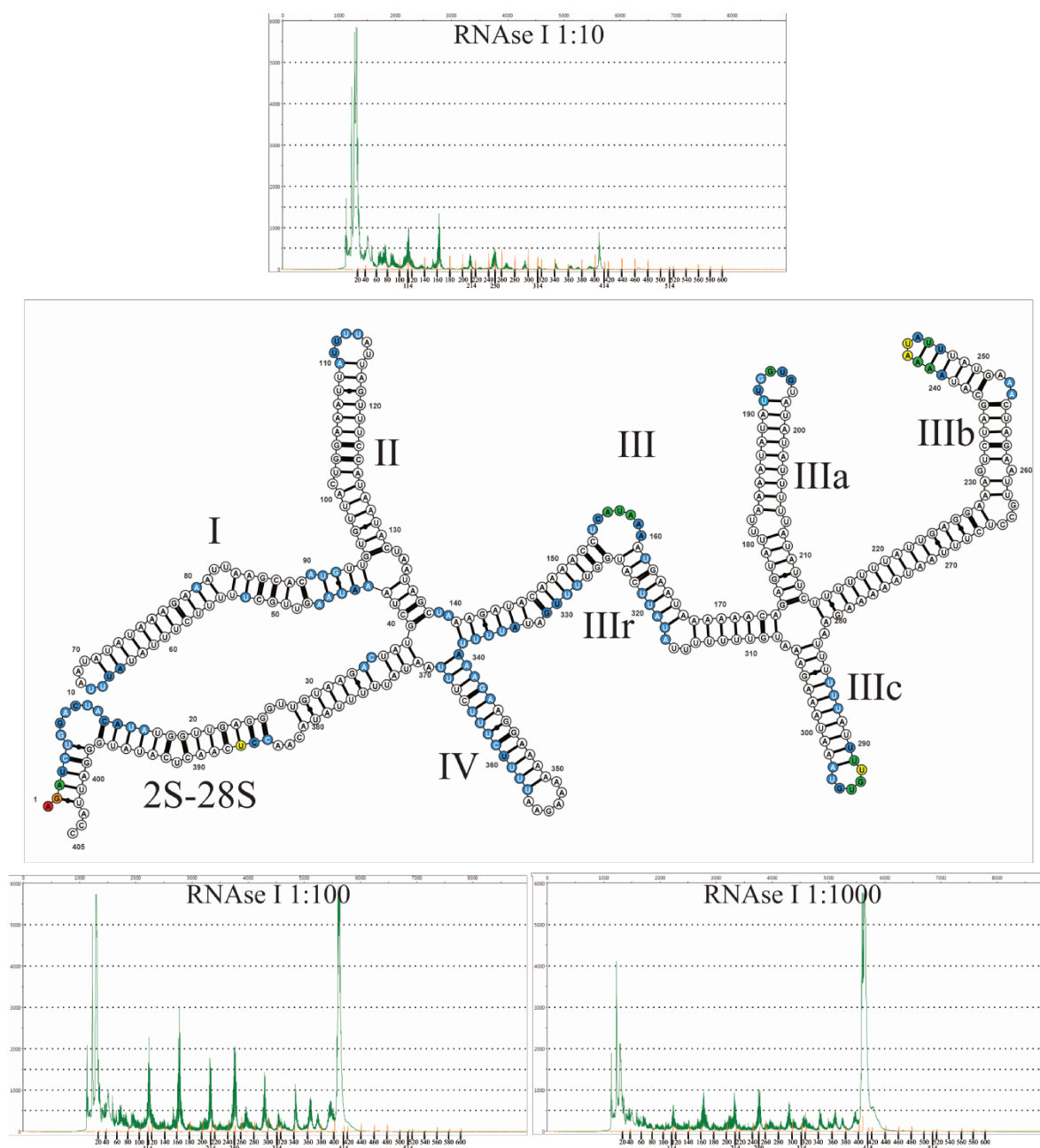


Figura A3. Cromatogramas obtidos a partir da digestão enzimática do RNA transcrito $chITS2_{I,II,III,IV}$ com diferentes enzimas e seu provável mapeamento na estrutura secundária predita *in silico* pelo programa mfold.

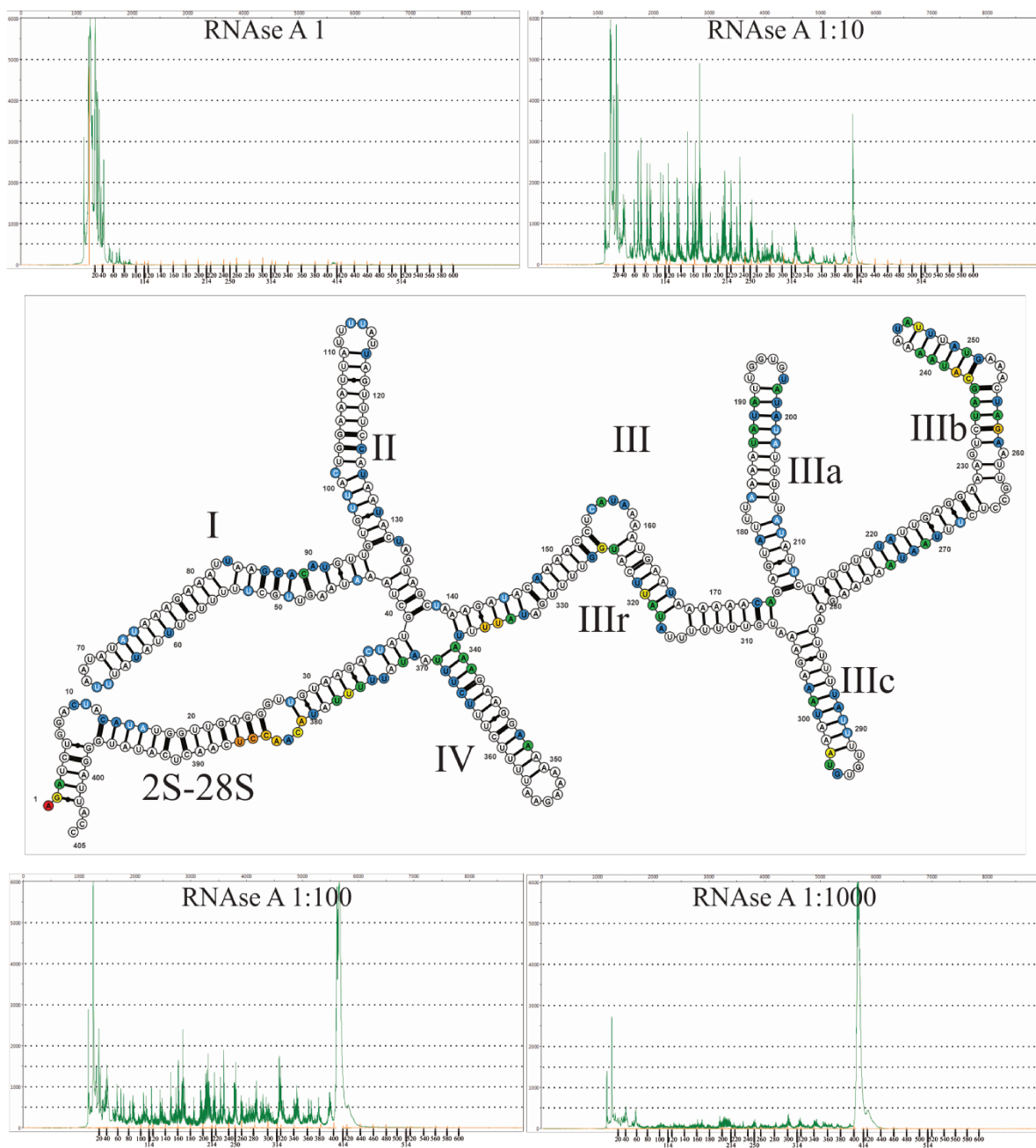


Figura A3. (Continuação).

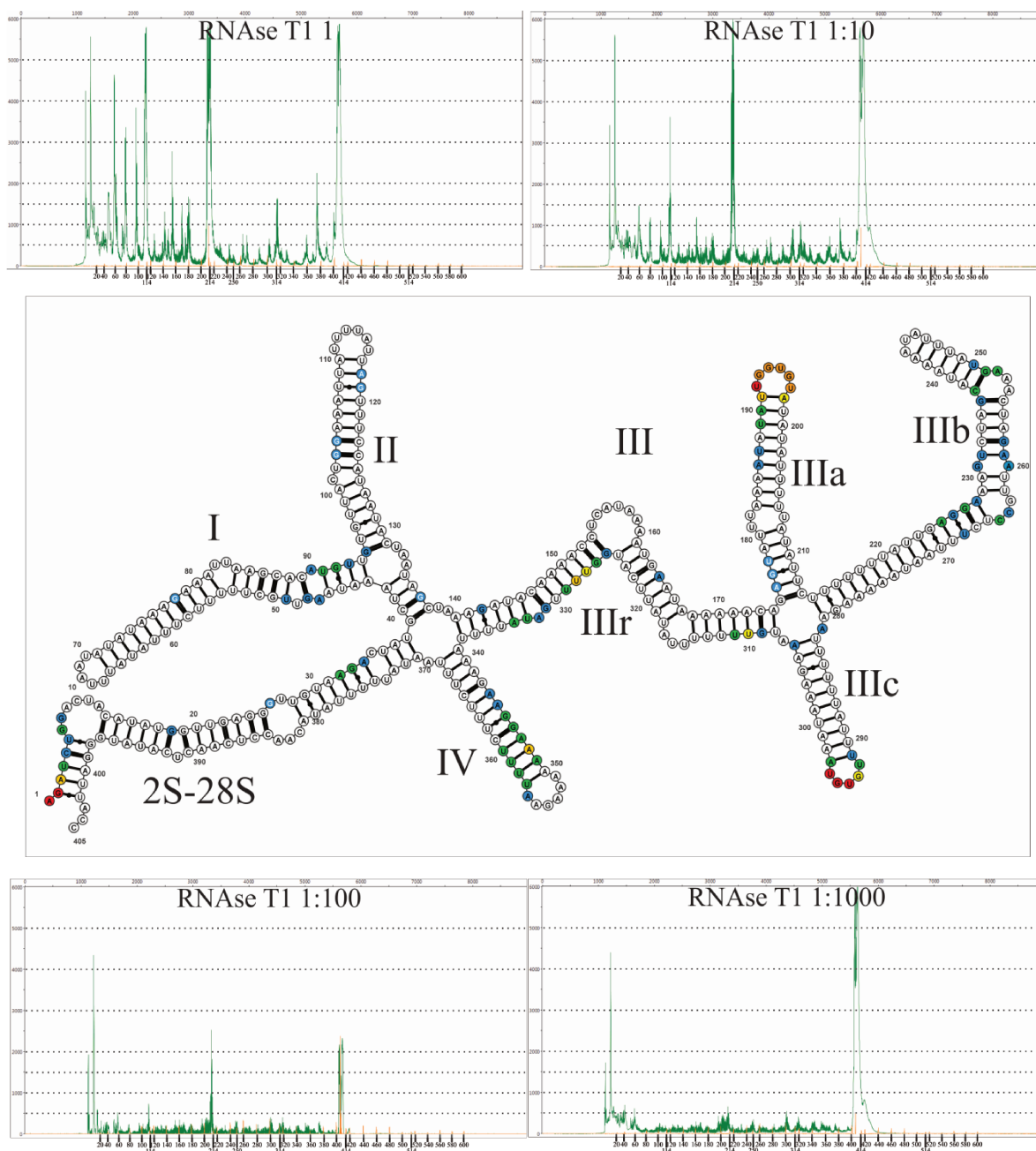


Figura A3. (Continuação).

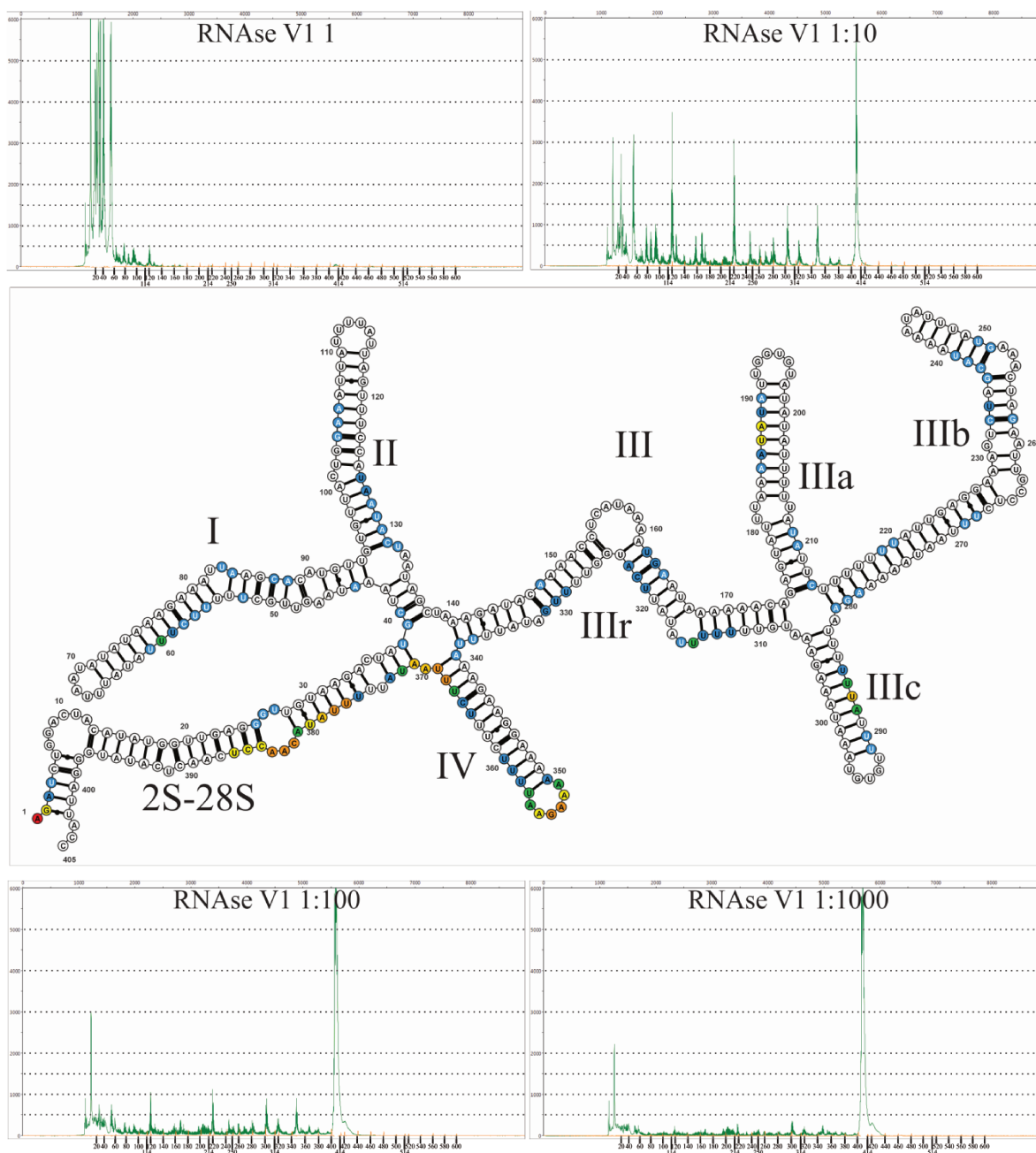


Figura A3. (Continuação).

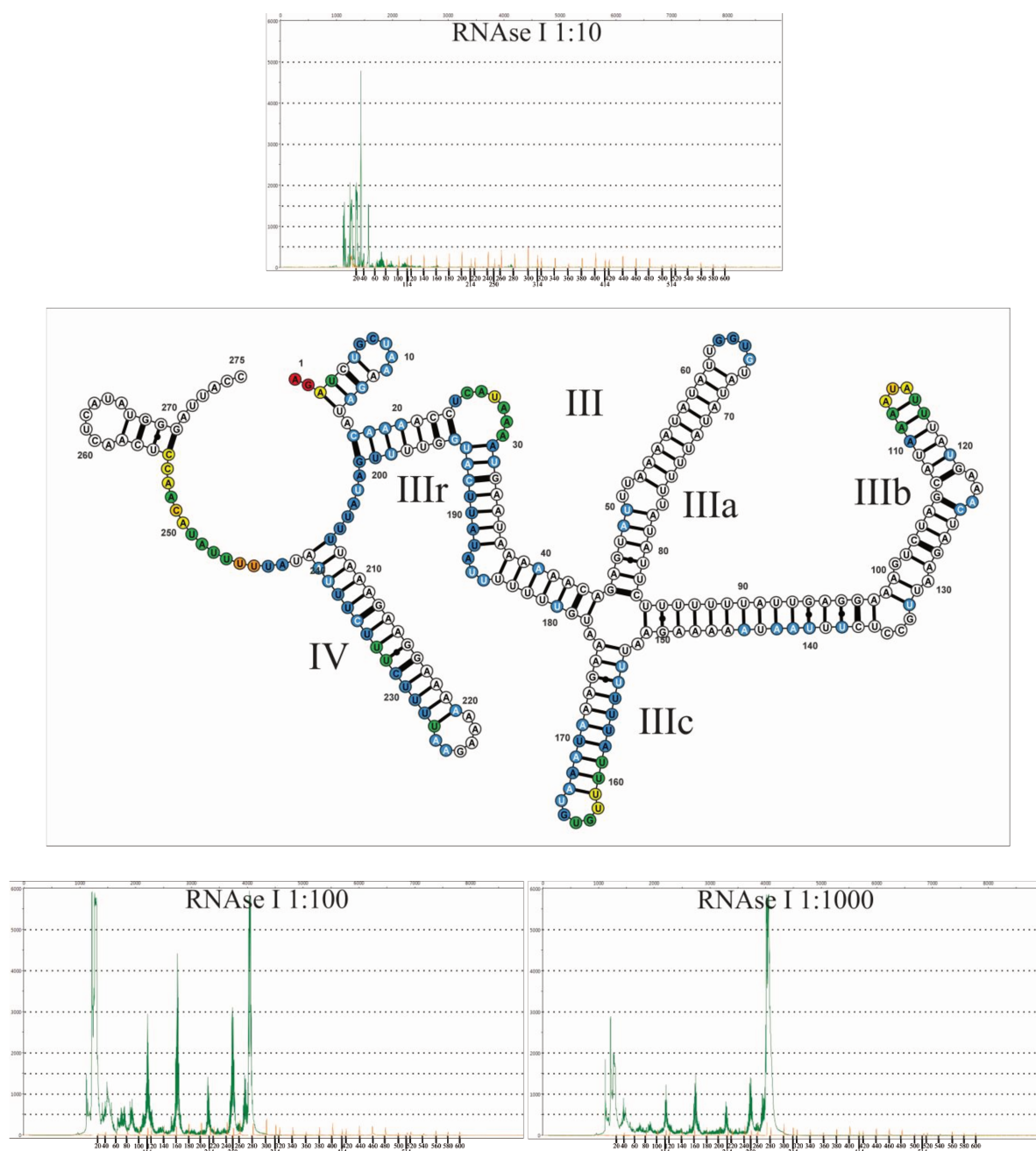


Figura A4. Cromatogramas obtidos a partir da digestão enzimática do RNA transcrito *chITS2*_{III,IV} com diferentes enzimas e seu provável mapeamento na estrutura secundária predita *in silico* pelo programa mfold.

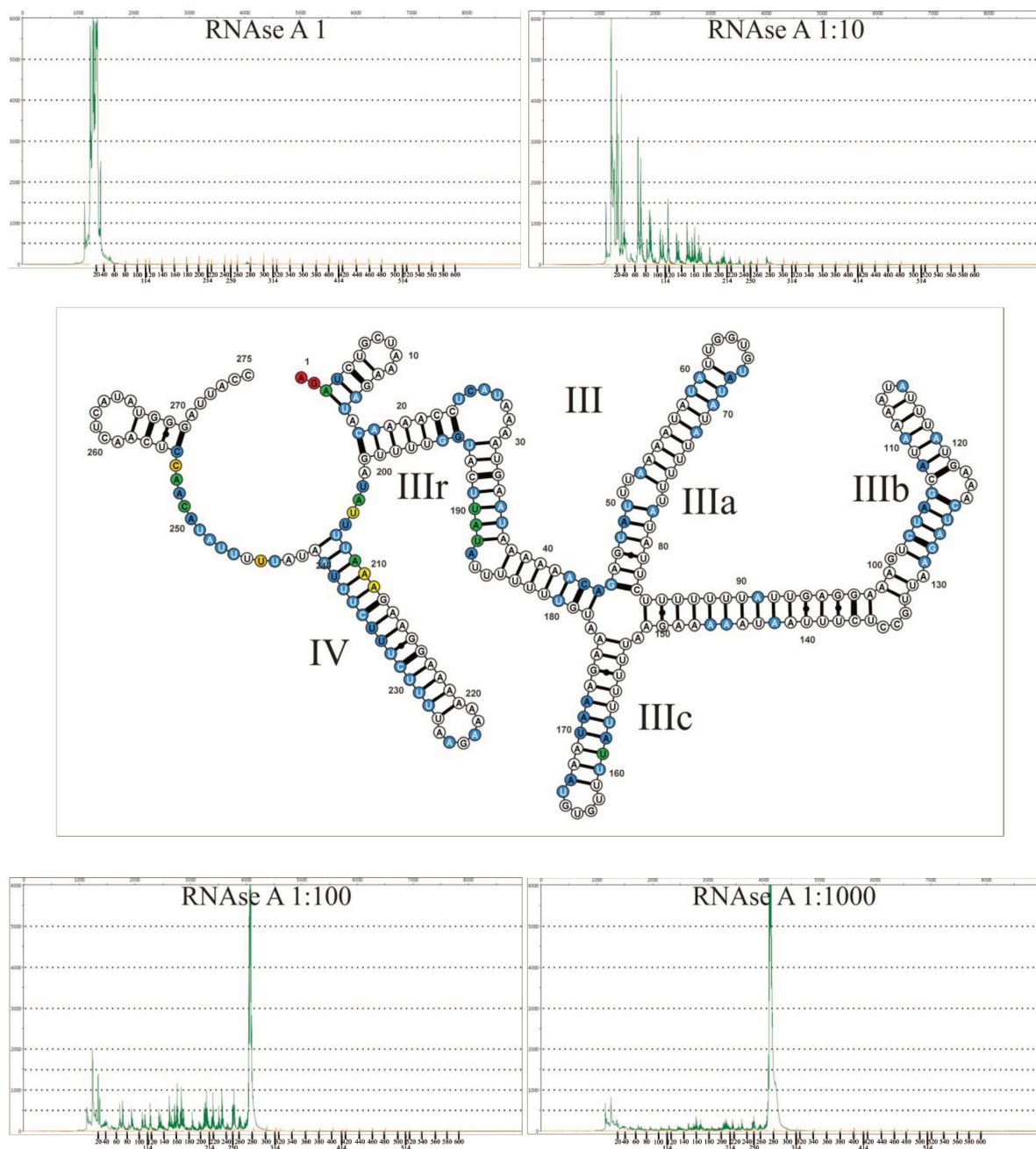


Figura A4. (Continuação).

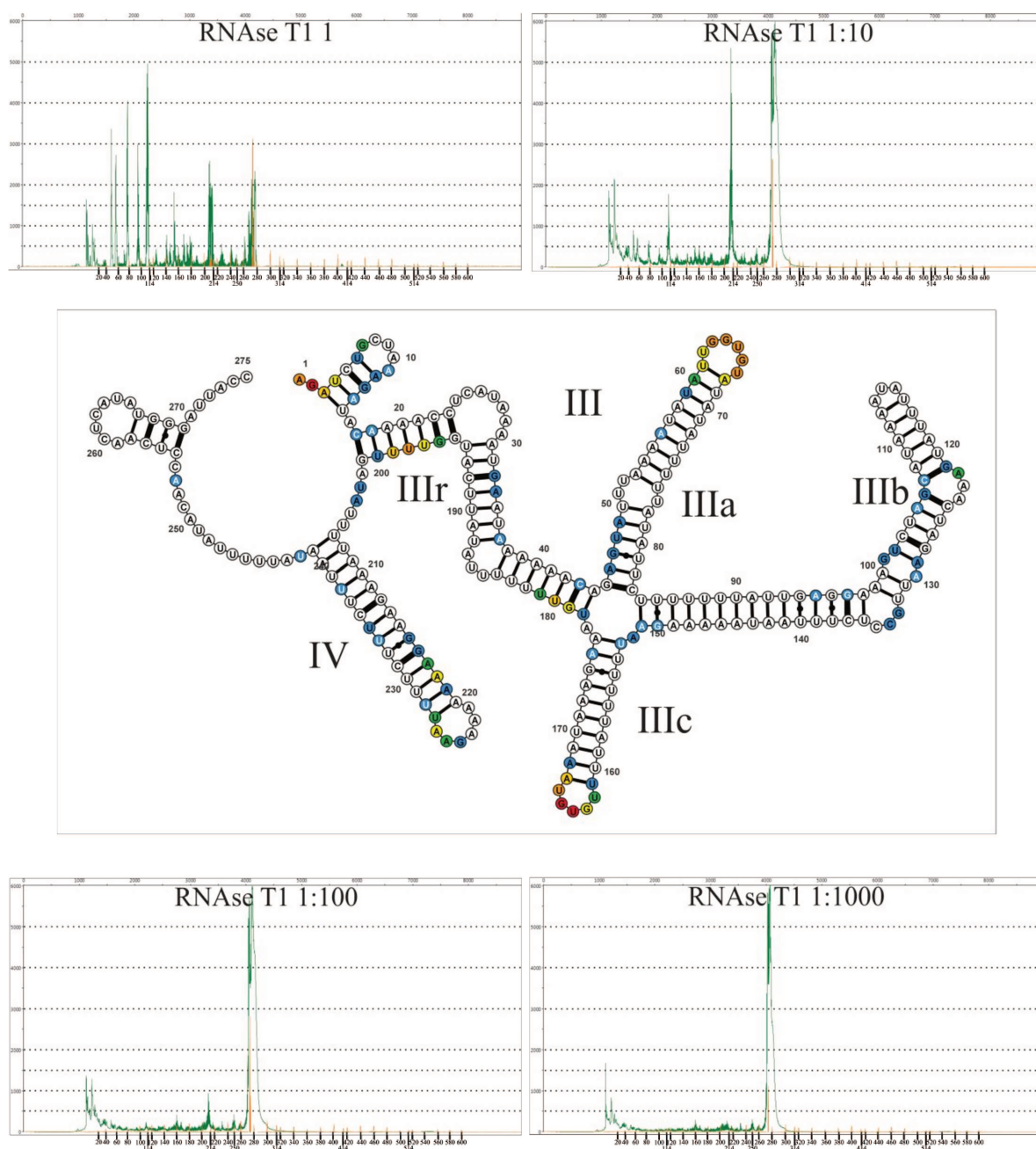


Figura A4. (Continuação).

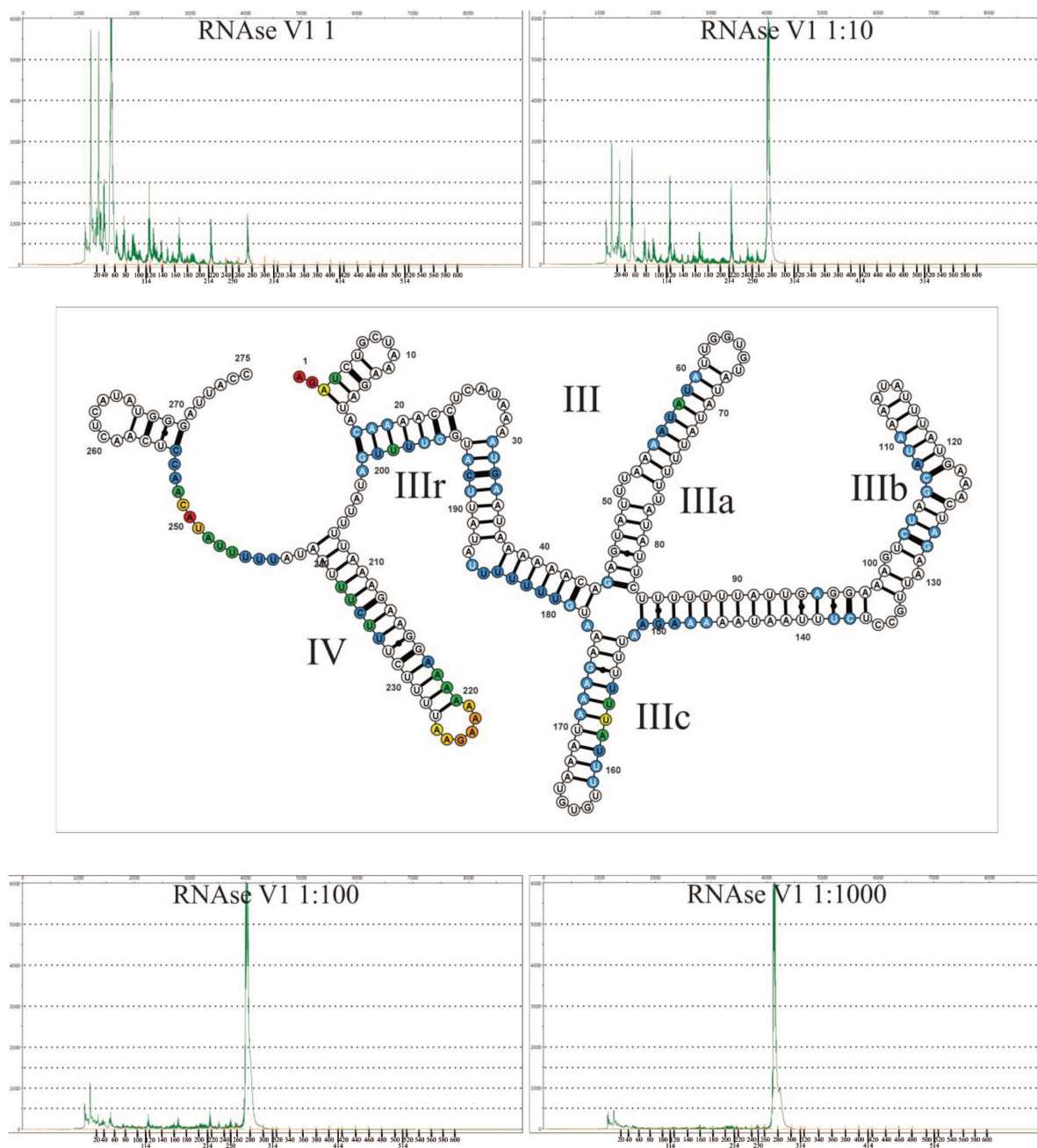


Figura A4. (Continuação).

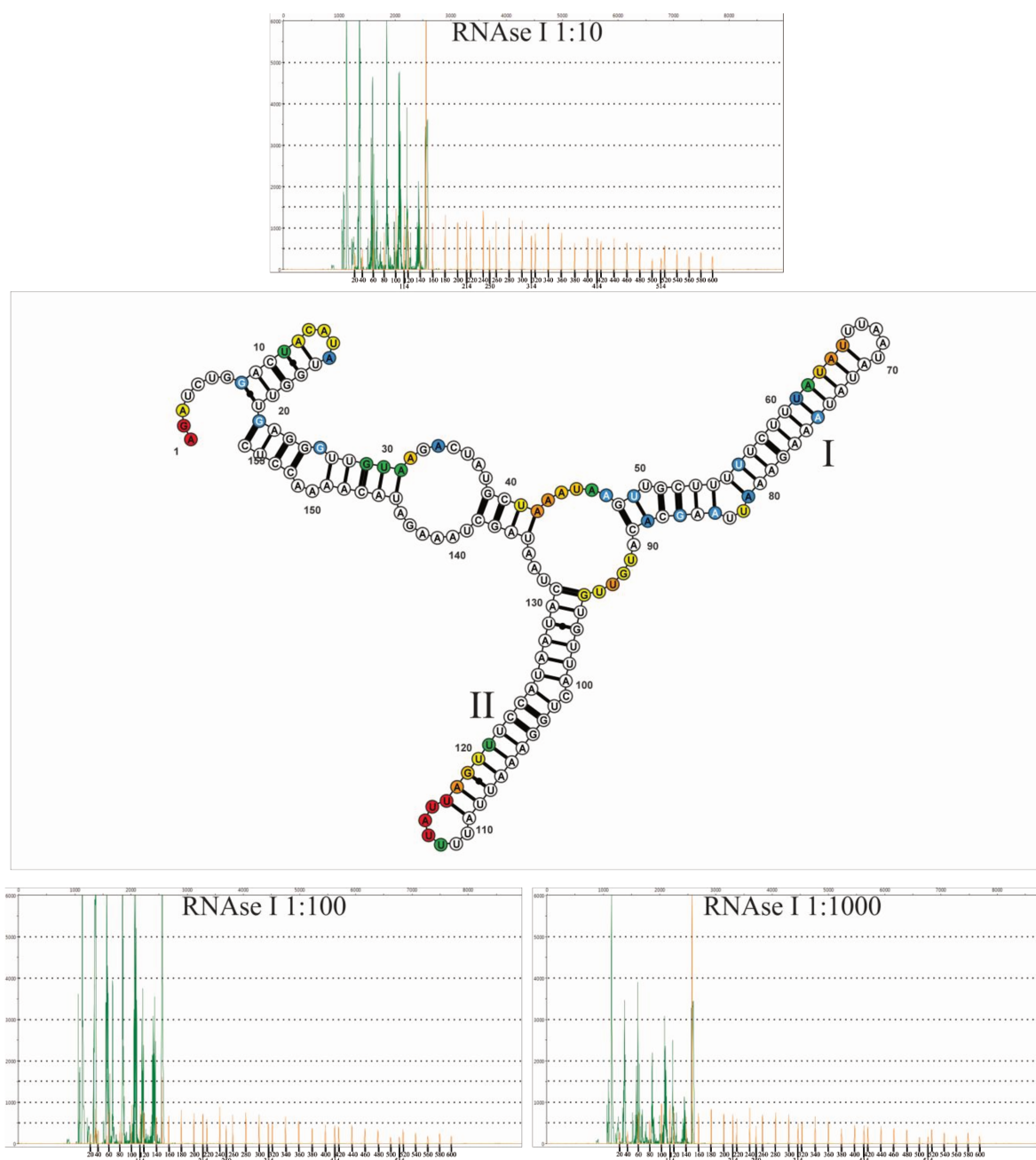


Figura A5. Cromatogramas obtidos a partir da digestão enzimática do RNA transcrito $chITS2_{I,II}$ com diferentes enzimas e seu provável mapeamento na estrutura secundária predita *in silico* pelo programa mfold.

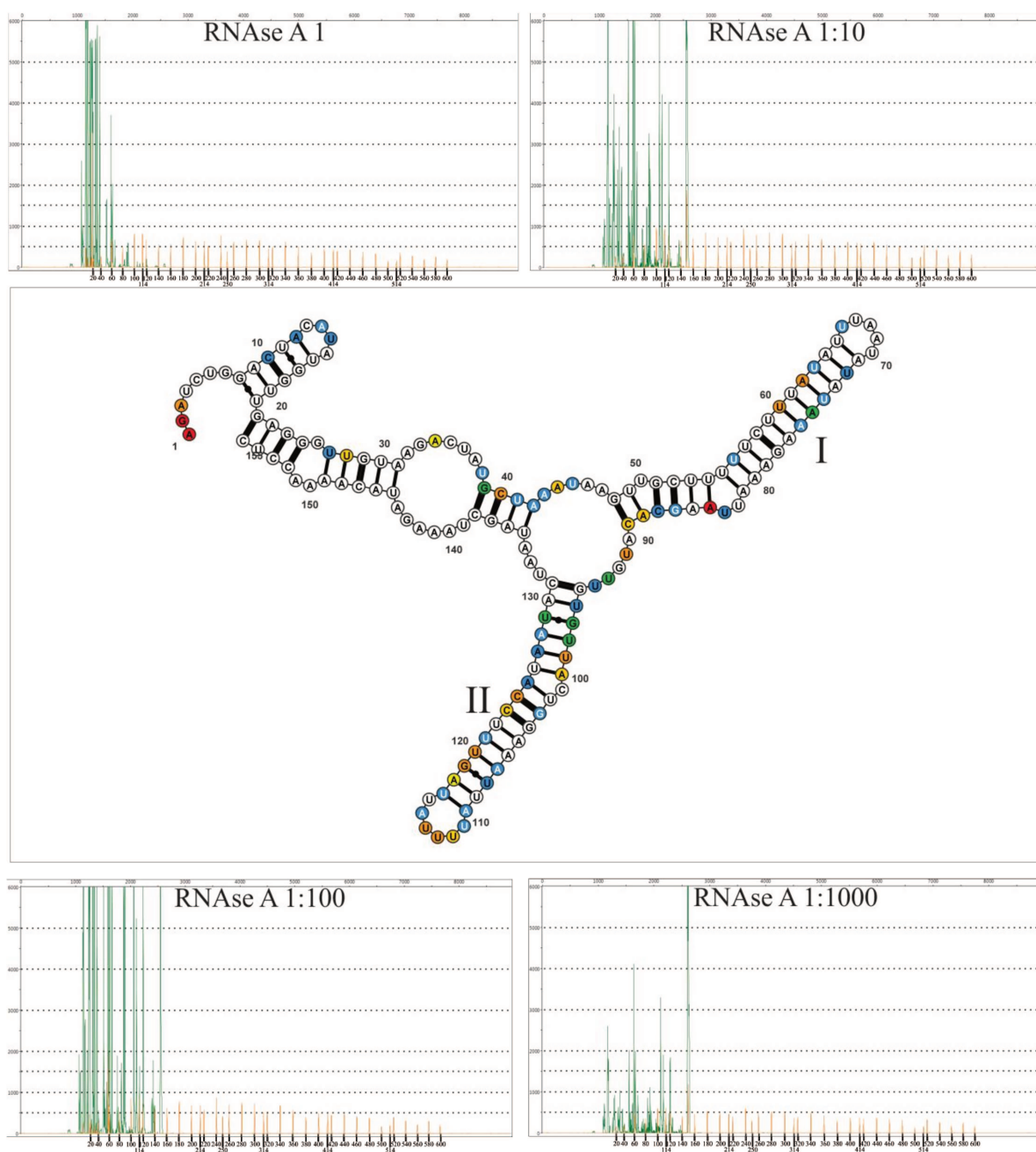


Figura A5. (Continuação).

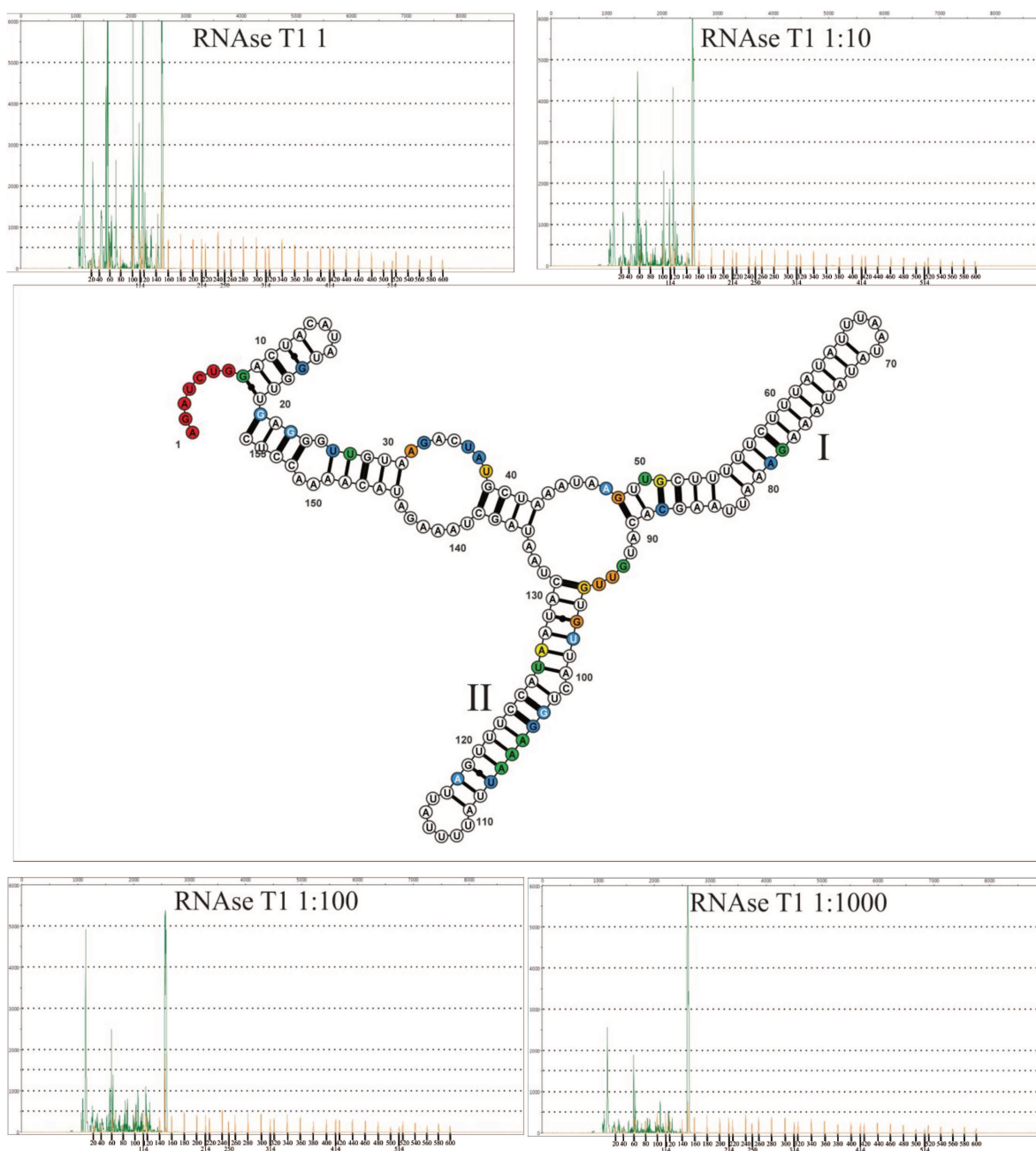


Figura A5. (Continuação).

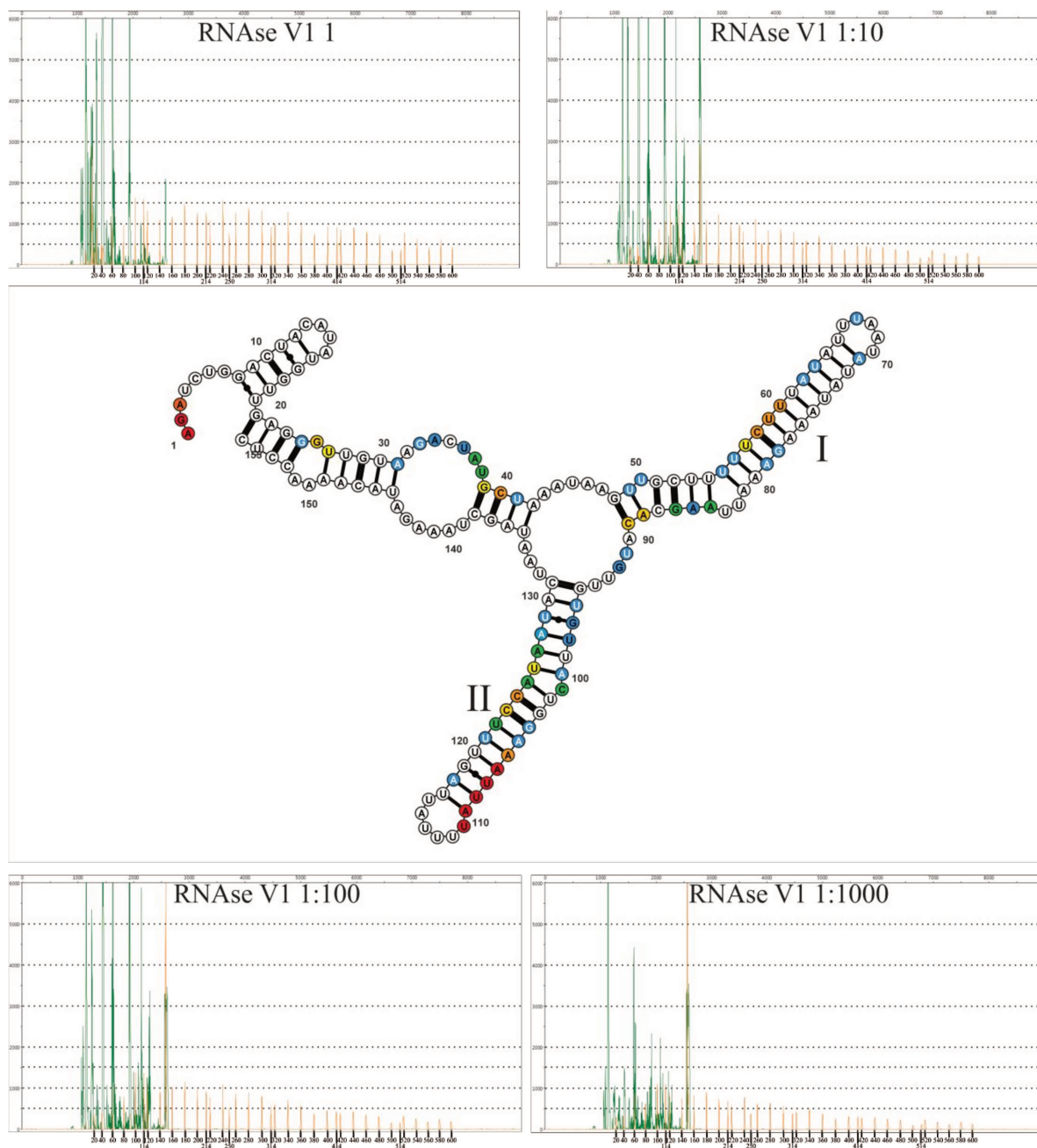


Figura A5. (Continuação).

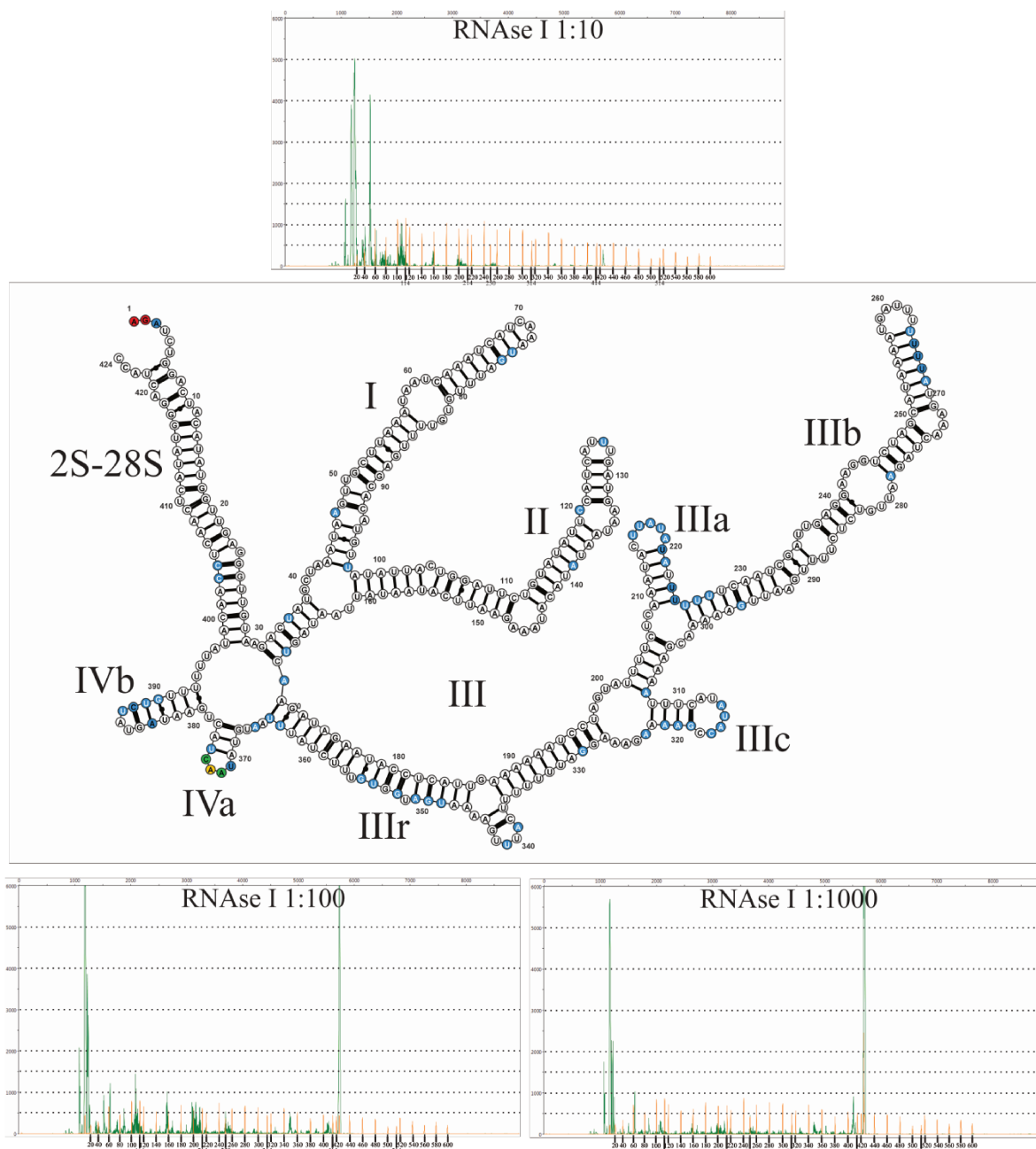


Figura A6. Cromatogramas obtidos a partir da digestão enzimática do RNA transcrito $MdITS2_{I,II,III,IV}$ com diferentes enzimas e seu provável mapeamento na estrutura secundária predita *in silico* pelo programa mfold.

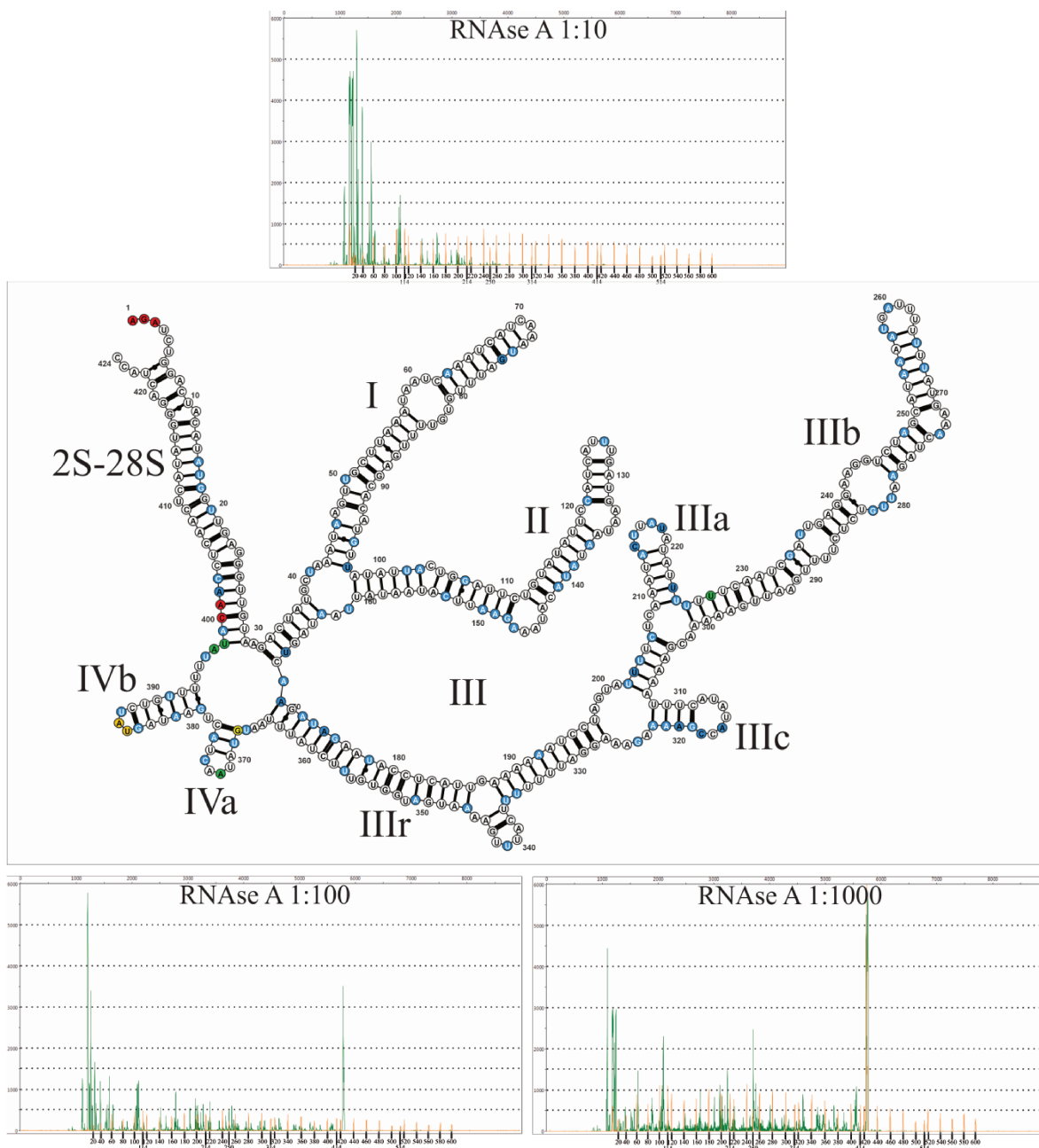


Figura A6. (Continuação).

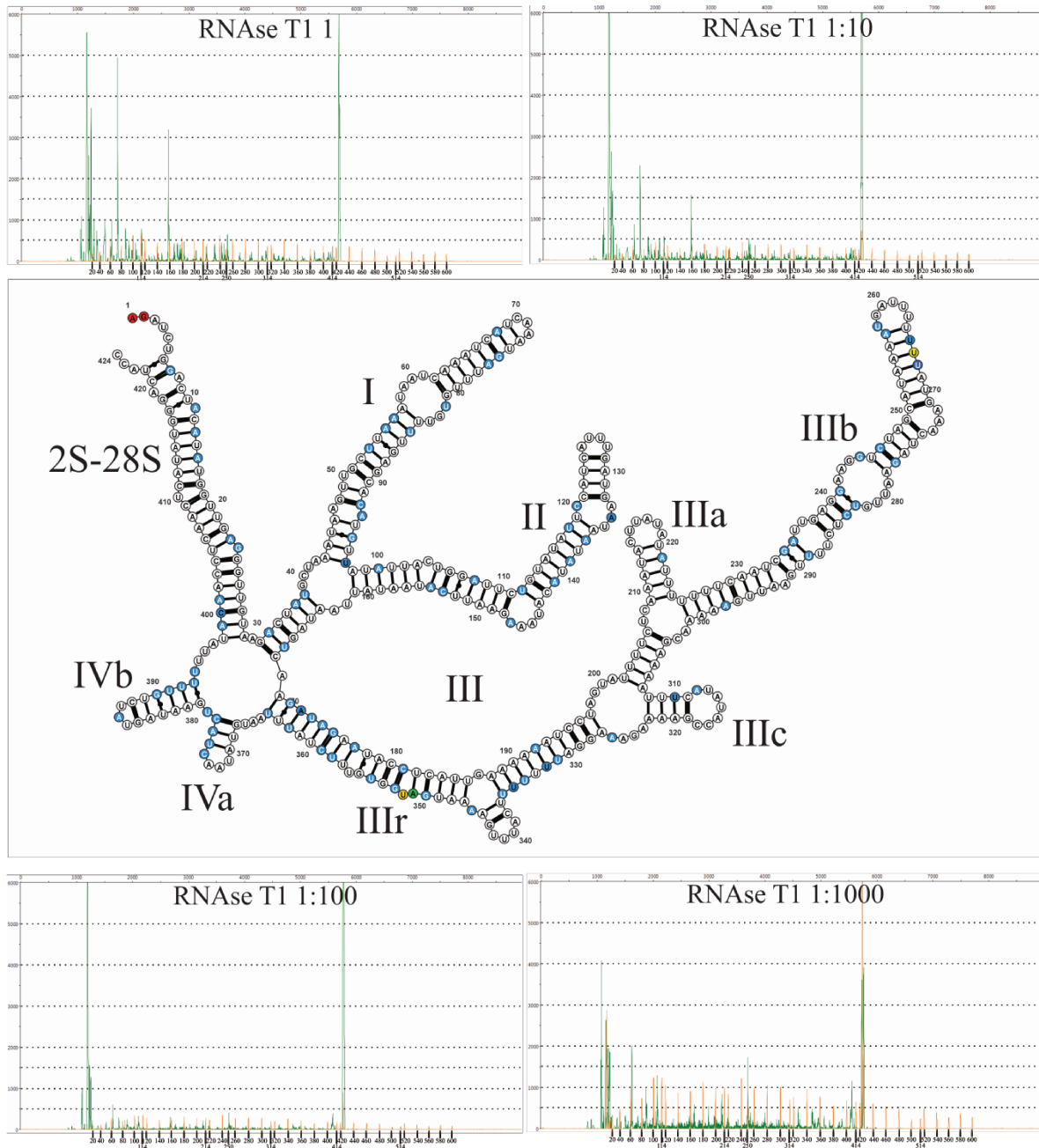


Figura A6. (Continuação).

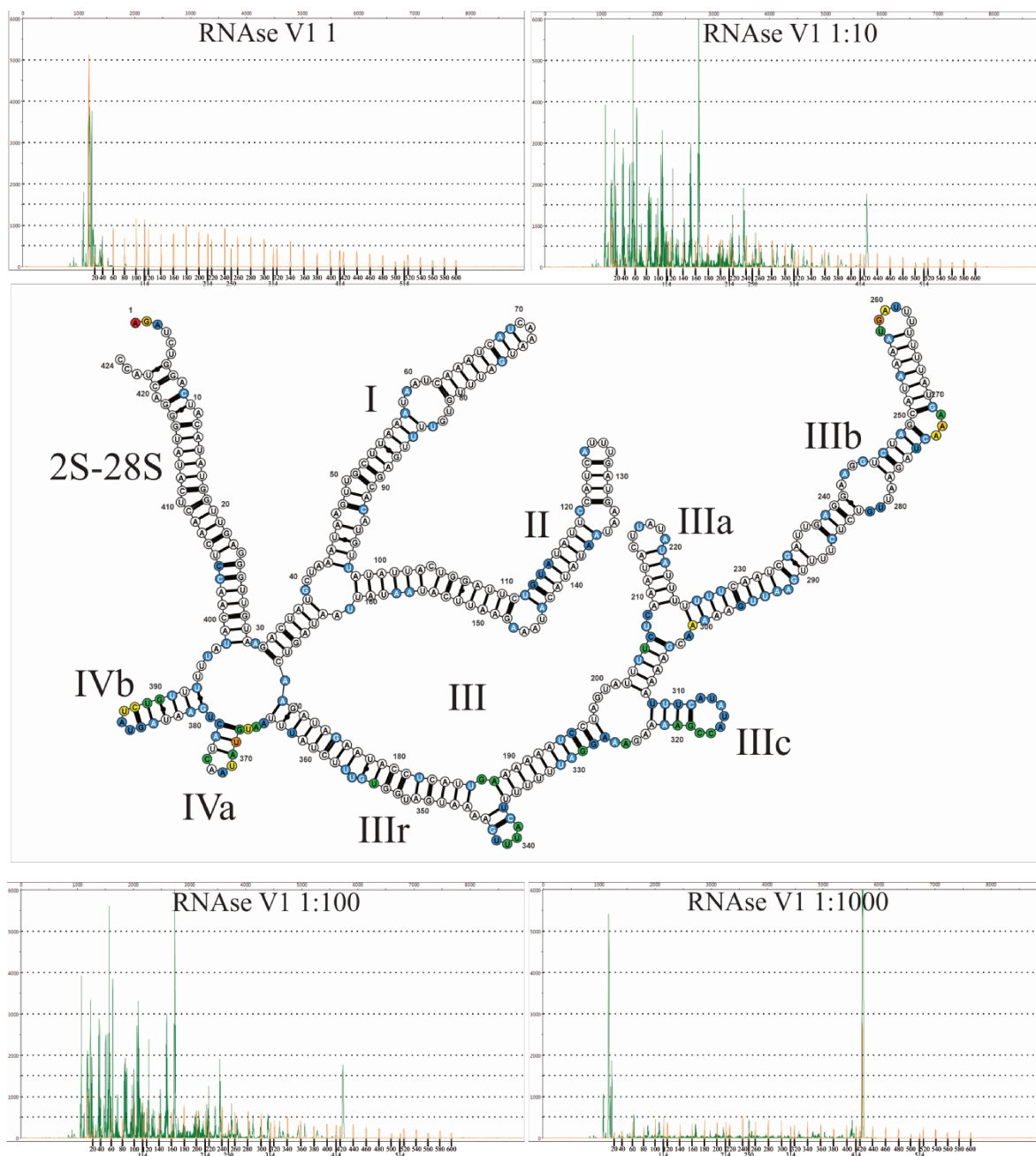


Figura A6. (Continuação).

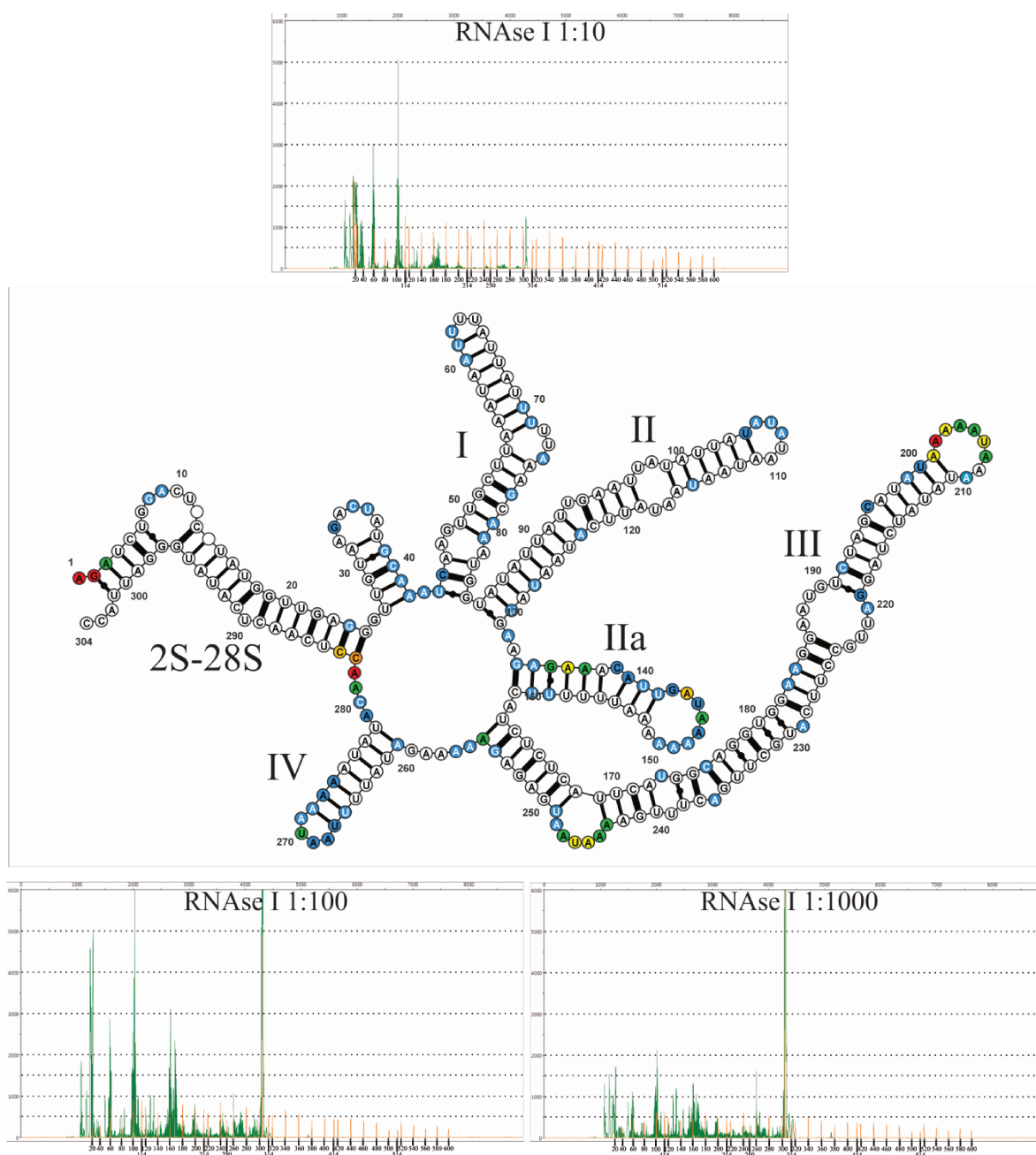


Figura A7. Cromatogramas obtidos a partir da digestão enzimática do RNA transcrito *GmITS2*_{I,II,III,IV} com diferentes enzimas e seu provável mapeamento na estrutura secundária predita *in silico* pelo programa mfold.

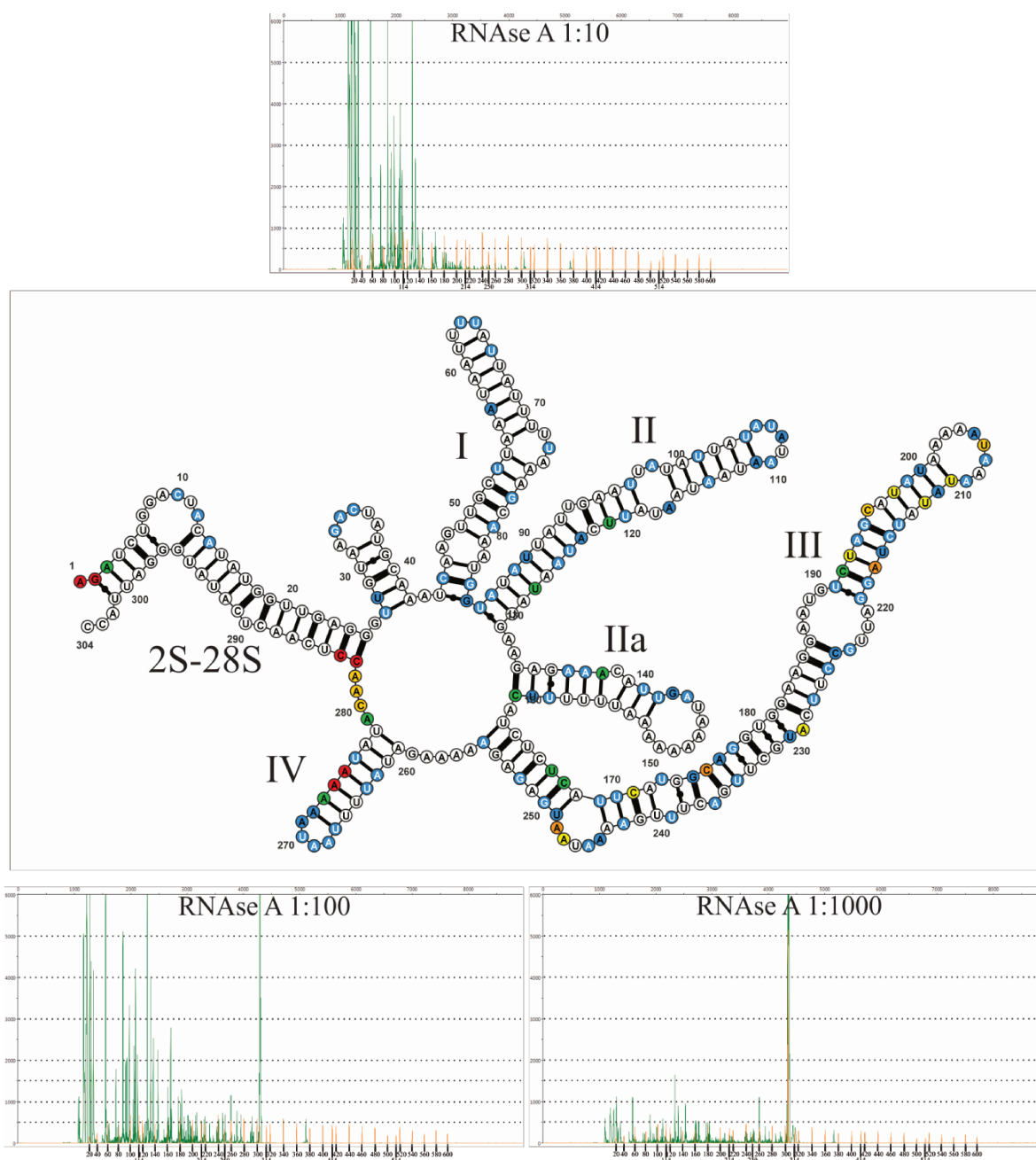


Figura A7. (Continuação).

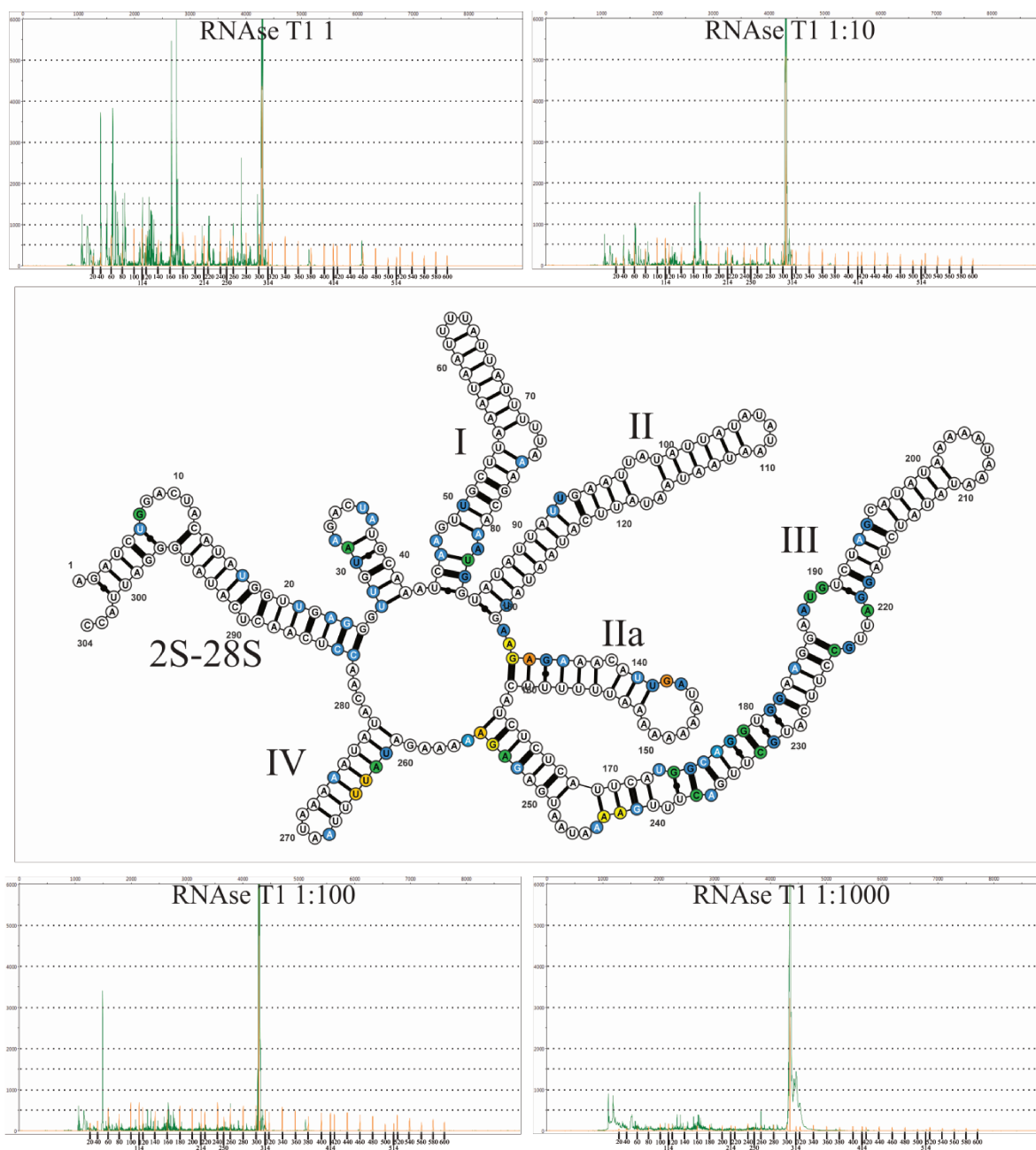


Figura A7. (Continuação).

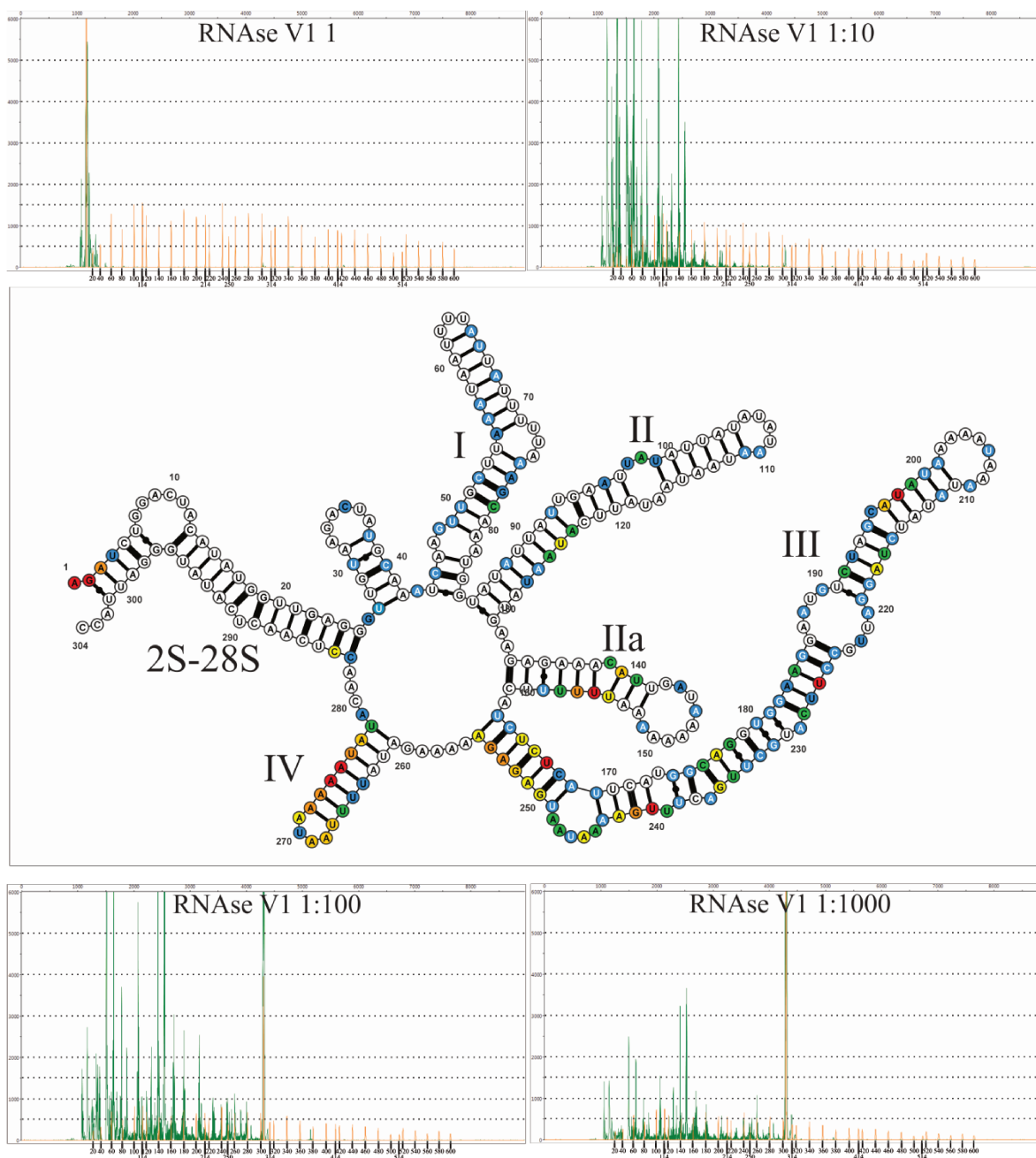


Figura A7. (Continuação).

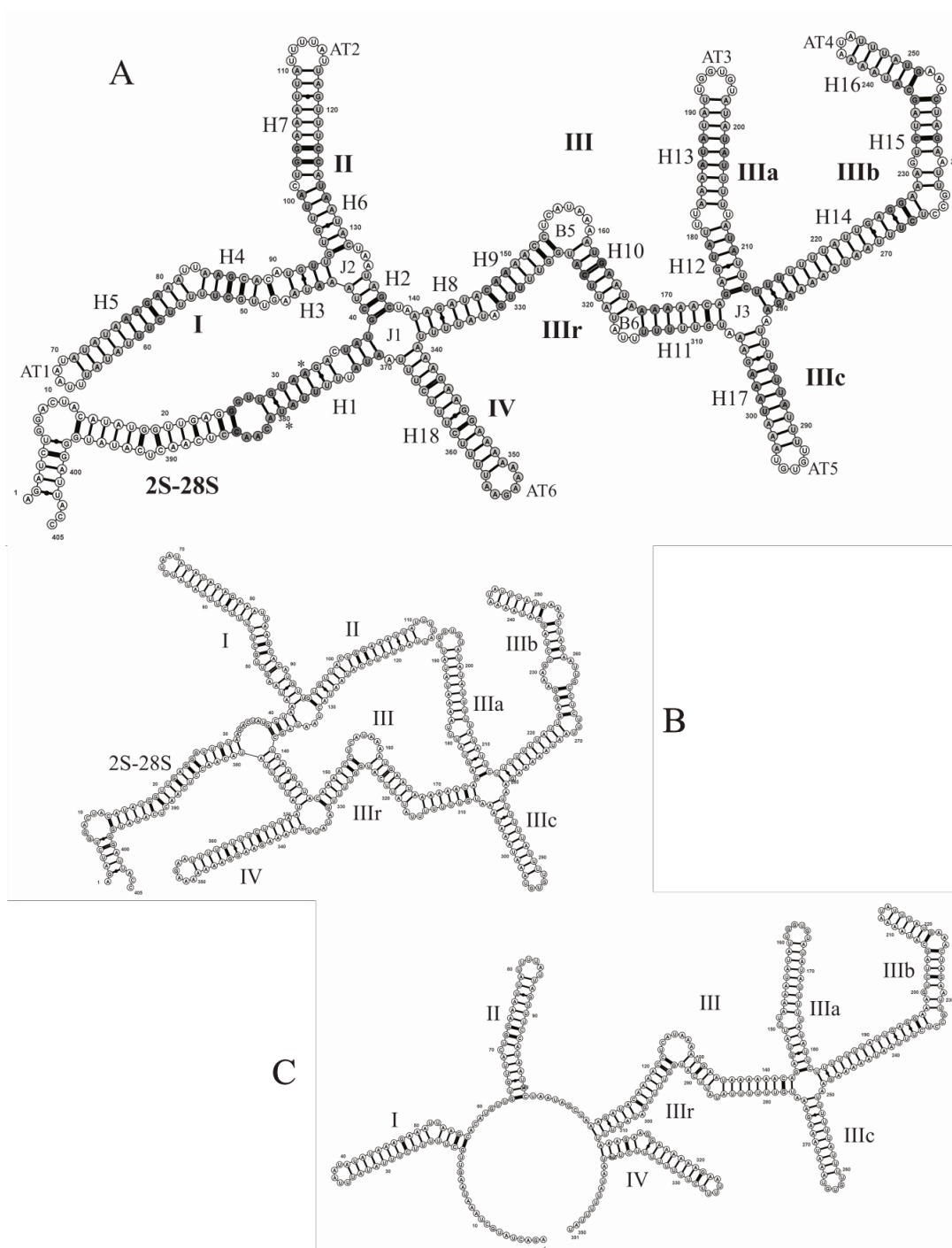


Figura A8. Possível interpretação do padrão de digestão enzimática mostrado nas Figuras A3 a A5 para o RNA $ChITS2_{I,II,III,IV}$. (A) Estrutura predita *in silico* pelo programa mfold (modelo em "grampo"); (B) Estrutura alternativa predita pelo software RNAfold; (C) Modelo de estrutura em "anel" recuperado pela abordagem conjunta de modelagem por homologia, comparativa e predição *in silico*. Os "*" indicam os pontos de início e fim da região do ITS2. H = hélice; J = junção; AT = alça terminal; B = "bojo".

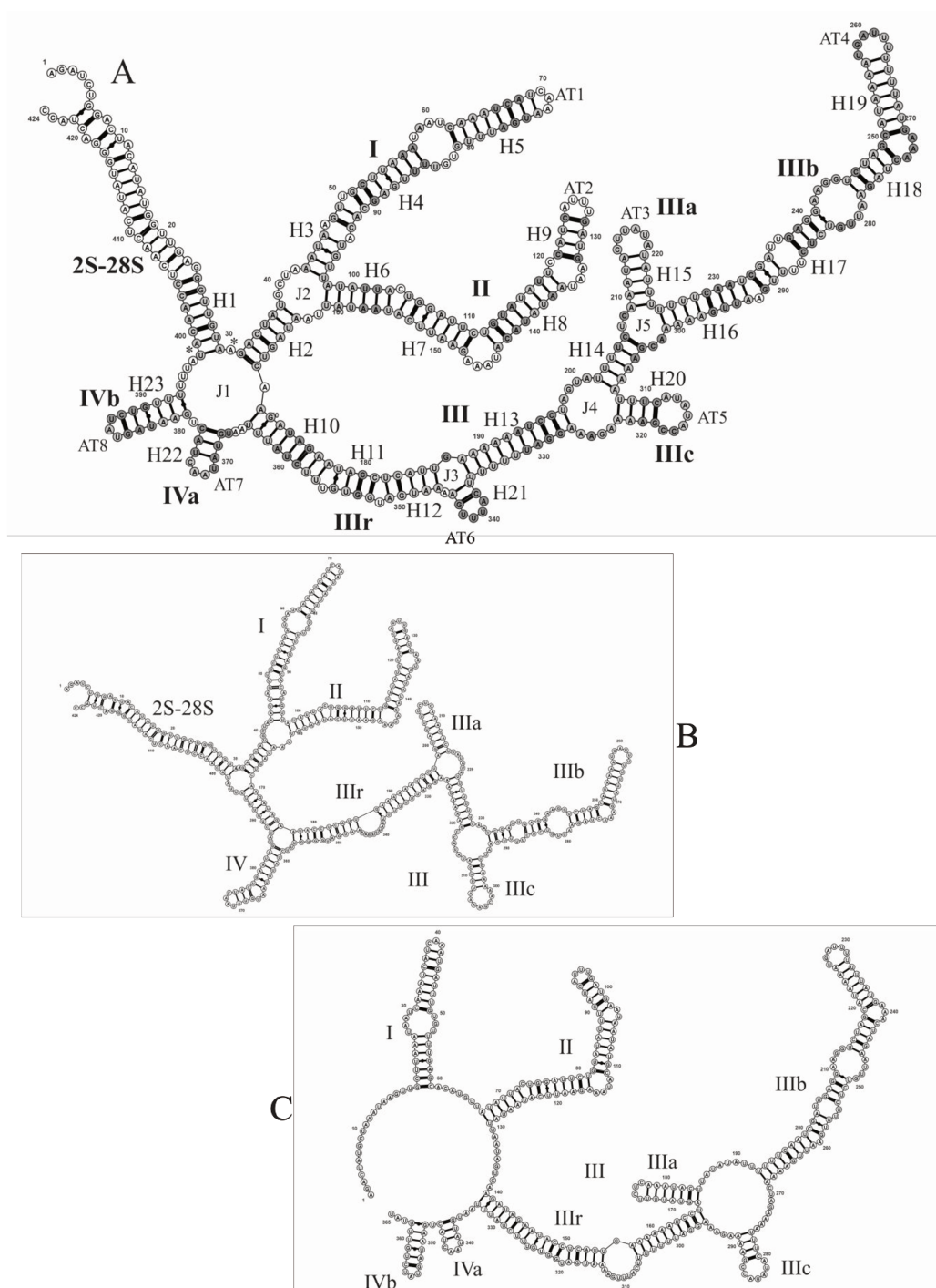


Figura A9. Possível interpretação do padrão de digestão enzimática mostrado na Figura A6 para o RNA *MdITS2*_{I,II,III,IV}. (A) Estrutura predita *in silico* pelo programa mfold (modelo em "grampo"); (B) Estrutura alternativa predita pelo software RNAfold; (C) Modelo de estrutura em "anel" recuperado pela abordagem conjunta de modelagem por homologia, comparativa e predição *in silico*. Os "*" indicam os pontos de início e fim da região do ITS2. H = hélice; J = junção; AT = alça terminal.

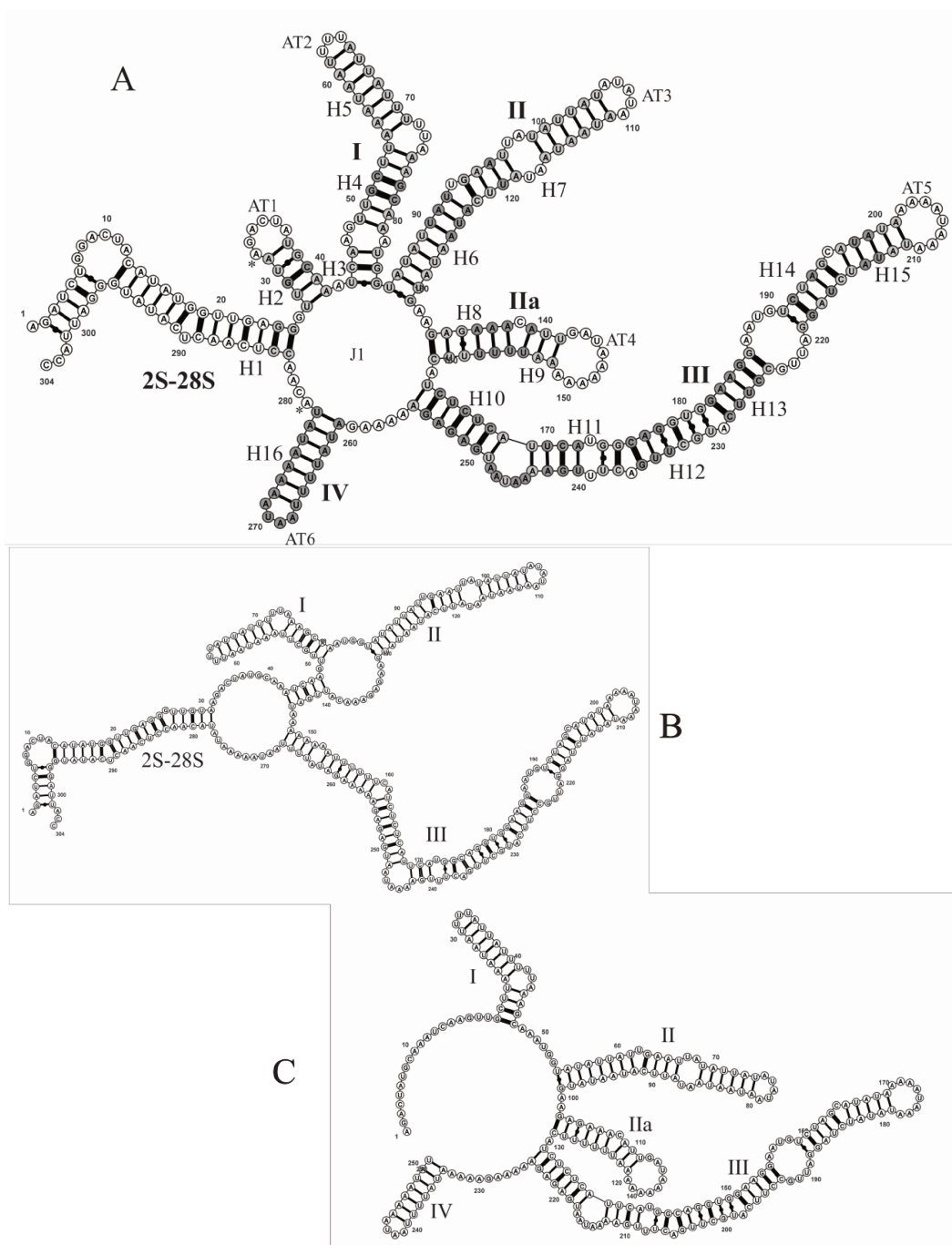


Figura A10. Possível interpretação do padrão de digestão enzimática mostrado na Figura A7 para o RNA *GmITS2*_{I,II,III,IV}. (A) Estrutura predita *in silico* pelo programa mfold (modelo em "grampo"); (B) Estrutura alternativa predita pelo software RNAfold; (C) Modelo de estrutura em "anel" recuperado pela abordagem conjunta de modelagem por homologia, comparativa e predição *in silico*. Os "*" indicam os pontos de início e fim da região do ITS2. H = hélice; J = junção; AT = alça terminal.

Capítulo III

Análises de filogenia molecular em Oestroidea (Diptera: Schizophora: Calyptratae) com evidências adicionais para a não-monofilia de Calliphoridae.

Resumo

A superfamília Oestroidea (Diptera: Calyptratae) é um grupo bastante diverso em termos de estratégia de vida. Das seis famílias comumente reconhecidas no grupo, a mais genericamente definida, Calliphoridae, parece ser crucial para o entendimento das relações evolutivas dentro da superfamília, uma vez que seu status monofilético é questionável. Com o objetivo de investigar mais a fundo as relações interfamiliares dentro de Oestroidea, as regiões gênicas do ITS2, 28S, COI e 16S foram utilizadas para inferir filogenias moleculares no grupo através dos métodos de máxima-parcimônia (MP), máxima-verossimilhança (ML) e inferência Bayesiana (BI). Nas análises Bayesianas, uma avaliação mais profunda dos efeitos de diferentes estratégias de particionamento dos dados foi conduzida, incluindo a consideração de informações sobre conformação estrutural (ITS2 e 16S) e posição no códon (COI). Os resultados das análises sugerem a existência de dois clados principais em Oestroidea: (Tachinidae + Mesembrinellinae) e (Rhiniinae, (Sarcophagidae + Calliphoridae *sensu stricto* [s.s.])). A família Oestridae foi recuperada como grupo irmão dois demais Oestroidea nas árvores de MP enquanto as análises de ML e BI sugerem a proximidade deste grupo com o clado composto por Rhiniinae + Sarcophagidae + Calliphoridae s.s.. O status não monofilético de Calliphoridae foi confirmado, assim como o status de família recém atribuído para Rhiniinae. A subfamília Mesembrinellinae, atualmente incluída em Calliphoridae, provavelmente pode ser incluída em uma família própria, mas mais estudos são necessários para confirmar essa proposição. A consideração de informações estruturais e de posição no códon levaram a um aumento significativo na verossimilhança das análises, embora esse aumento tenha sido acompanhado apenas por pequenas mudanças nas topologias, comprimentos de ramos e suportes de probabilidade *a posteriori* nas árvores inferidas. O uso de modelos mais

complexos levou também a um aumento na incerteza acerca das filogenias estimadas, incluindo as topologias de árvores, e filogenias inferidas com modelos ricos em parâmetros podem ser menos confiáveis mesmo quando apresentando maiores valores de verossimilhança.

Palavras-chave: Filogenia molecular, Oestroidea, Calliphoridae, Mesembrinellinae, ITS2, COI, 28S, 16S, estrutura secundária.

1. Introdução

A família Calliphoridae (Diptera: Calyptratae: Oestroidea), cujos membros são popularmente conhecidos como moscas-varejeiras, é um grupo bastante diverso e heterogêneo que compreende mais de 1.000 espécies descritas com distribuição mundial. A família é mais conhecida e estudada por seus membros sarcosaprófagos e causadores de miíase das subfamílias Chrysomyinae, Luciliinae e Calliphorinae (de Azeredo-Espin & Lessinger 2006; Stevens & Wallman 2006; Stevens et al. 2006), os quais possuem hábito sinantrópico e grande importância em questões forenses, veterinárias, médicas e econômicas (Zumpt 1965; Guimarães et al. 1983; Hall & Wall 1995; Amendt et al. 2004). Não obstante, o grupo possui na verdade uma grande variedade em termos de hábito e ambientes de reprodução, os quais incluem o hábito hematófago em aves e mamíferos (e.g., espécies de *Protocalliphora*, *Trypocalliphora* e *Auchmeromyia*), o parasitismo de gastrópodes terrestres (e.g., espécies de *Melanomya*, *Melinda* e a maioria das espécies da subfamília Ameniinae) e anelídeos (e.g., espécies de *Bellardia* e *Pollenia*) e a associação com colônias de formigas e cupins (e.g., espécies de *Bengalia*, *Tricylea*, *Hemigymnochaeta*, *Termitocalliphora* e muitas espécies de Rhiniinae).

Em um grupo tão diverso, não é de todo surpreendente que ainda exista muita controvérsia em relação a monofilia e composição de Calliphoridae. Com relação à monofilia, enquanto Lehrer (1970), Rognes (1991), McAlpine (1989) e Pape (1992) forneceram evidências corroborando-a (apesar de alguns reconhecerem implicitamente a fragilidade do grupo), Hennig (1973) e Griffiths (1982) notaram falta de suporte para o grupo e sugeriram um provável status para ou polifilético. Posteriormente, tanto Rognes (1997) quanto Kutty et al. (2010) forneceram evidências da não-monofilia da família baseando-se na análise de dados morfológicos e moleculares, respectivamente.

A composição da família em termos de tribos e subfamílias é também bastante controversa e o número de subfamílias normalmente atribuídas a Calliphoridae é variável, abrangendo de duas (Shewell 1987: Calliphorinae, contendo as tribos Calliphorini, Polleniini, Angioneurini e Luciliini; e Chrysomyinae, contendo as tribos Chrysomyini, Rhiniini e Phormiini) a treze (Rognes 1986, 1991, 1997: Chrysomyinae, Calliphorinae, Luciliinae, Toxotarsinae, Melanomyinae, Auchmeromyinae, Bengaliinae, Polleniinae, Mesembrinellinae, Phumosiinae, Rhiniinae, Helicoboscinae e Ameniinae), com algumas classificações intermediárias (e.g., Hennig 1973: Calliphorinae, Chrysomyinae, Mesembrinellinae, Ameniinae e Rhiniinae).

A controvérsia existente em torno do status monofilético de Calliphoridae e das relações interfamiliares em Oestroidea sugere a possibilidade de que a família, como atualmente compreendida, possa ser desmembrada em grupos menores com posições ainda não estabelecidas na filogenia de Oestroidea. Isso torna a família Calliphoridae provavelmente a chave para entender a evolução e a filogenia da superfamília Oestroidea (McAlpine 1989; Rognes 1997).

A superfamília Oestroidea (Diptera: Calyptratae) é um grupo comumente aceito como monofilético (Griffiths 1972; Hennig 1973; McAlpine 1989; Pape 1992; Rognes 1997) e é formalmente reconhecido como sendo composto por seis famílias (McAlpine 1989; Pape & Thompson 2010): Calliphoridae, Mystacinobiidae, Rhinophoridae, Oestridae, Tachinidae e Sarcophagidae. Este número foi recentemente acrescido pela inclusão da família Rhiniidae, anteriormente uma subfamília de Calliphoridae, a qual possui agora status reconhecido de família na base de dados "Systematic Database of World Diptera" (Pape & Thompson 2010). Outras subfamílias de Calliphoridae, a exemplo de Rhiniinae, já foram propostas como famílias distintas, como por exemplo Bengaliidae (Lehrer 2005) e

Mesembrinellidae (Guimarães 1977), embora ainda exista muita controvérsia sobre essas classificações e estudos adicionais são necessários para comprovar sua validade.

Os trabalhos de Rognes (1997) e Kutty et al. (2010) forneceram evidências acerca da não monofilia de Calliphoridae e propuseram diferentes posicionamentos para os grupos para/polifiléticos da família. Contudo, ainda há um grande número de hipóteses filogenéticas diferentes para as relações entre as famílias de Oestroidea e as subfamílias de Calliphoridae, algumas delas muito pouco corroboradas, enfatizando a necessidade de estudos adicionais.

Nesse contexto, as análises apresentadas aqui visam um estudo molecular abrangente das relações filogenéticas interfamiliares em Oestroidea, com ênfase na família Calliphoridae e no posicionamento de seus possíveis grupos não monofiléticos, usando marcadores moleculares tanto de origem nuclear quanto mitocondrial e sob diferentes métodos de inferência filogenética. Além disso, aproveitando-se da robustez do método de inferência Bayesiana no que concerne ao uso de modelos complexos, diferentes estratégias de particionamento dos dados foram empregadas, incluindo a utilização de modelos de substituição baseados em informação tanto de sequência quanto de estrutura secundária, e avaliadas em termos de seus impactos nas topologias inferidas e no suporte geral das filogenias recuperadas. As implicações dessas diferentes abordagens de tratamento dos dados na acurácia e precisão das análises de inferência filogenética foram também acessadas.

2. Materiais e Métodos

2.1. Espécimes e extração de DNA

A lista completa dos 56 espécimes utilizados nas análises de filogenia molecular é mostrada na Tabela 1. A classificação das subfamílias de Calliphoridae seguiu a proposta

por Rognes (1997), incluindo a subfamília Rhiniinae, recentemente elevada ao status de família. O termo Calliphoridae *sensu stricto* (Calliphoridae *s.s.*) será utilizado aqui para se referir à família Calliphoridae excluindo-se as subfamílias Rhiniinae e Mesembrinellinae, enquanto o termo Calliphoridae *sensu lato* (Calliphoridae *s.l.*) será utilizado para designar o clado composto por (Calliphoridae *s.s.* + Rhiniinae + Mesembrinellinae).

Assim, a amostragem taxonômica incluiu oito das treze subfamílias de Calliphoridae reconhecidas por Rognes (1997) e quatro das 6 famílias tradicionalmente reconhecidas em Oestroidea (excluindo-se Rhiniidae) (McAlpine 1989). Espécies das superfamílias Muscoidea e Hippoboscoidea foram usadas como grupo externo.

A extração de DNA total para espécimes frescos e congelados foi realizada segundo o protocolo de fenol / clorofórmio (Infante & Azeredo-Espin 1995), enquanto para os espécimes secos e preservados em etanol foram utilizados o reagente DNAzol (Invitrogen) e posteriormente o kit "Spin Tissue Mini-Kit" (Invitex), ambos seguindo as instruções do fabricante.

2.2. Amplificação via PCR, clonagem e sequenciamento

Para as análises de filogenia molecular, quatro regiões gênicas foram amplificadas via PCR e sequenciadas: (1) a região completa do segundo espaçador transcrito interno (ITS2) e (2) a porção 5' do gene da subunidade ribossomal 28S, ambos do "cluster" de DNA ribossomal nuclear; (3) a porção 5' do gene da subunidade I da citocromo oxidase c (COI) e (4) a porção 3' da subunidade ribossomal 16S, ambos do genoma mitocondrial.

As reações de PCR para amplificação da região ITS2 foram conduzidas com 10mM Tris-HCl (pH 8,8), 50mM KCl, 2mM MgCl₂, 80µM dNTPs, 0,4µM "oligo" 5,8S (5'-ATC

Tabela 1. Lista das espécies utilizadas nas análises de filogenia molecular. A classificação em subfamílias de Calliphoridae adotada segue a descrita em Rognes (1997). N/A = dado não disponível.

Superfamília	Família	Subfamília	Espécies	Sequências			
				ITS2	28S	COI	16S
Oestroidea	Calliphoridae	Chrysomyinae	<i>Chloroprocta idioidea</i> (Robineau-Desvoidy 1830)	EF560180	JQ246603	JQ246658	JQ246708
			<i>Chrysomya albiceps</i> (Wiedemann 1819)	EF560173	JQ246604	JQ246659	JQ246709
			<i>Chrysomya bezziana</i> (Villeneuve 1914)	EF560174	JQ246605	JQ246660	JQ246710
			<i>Chrysomya chloropyga</i> (Wiedemann 1818)	JQ246571	JQ246606	JQ246661	JQ246711
			<i>Chrysomya megacephala</i> (Fabricius 1794)	EF560175	JQ246607	JQ246662	-
			<i>Chrysomya putoria</i> (Wiedemann 1830)	EF560176	JQ246608	JQ246663	JQ246712
			<i>Chrysomya rufifacies</i> (Macquart 1843)	EF560177	JQ246609	JQ246664	JQ246713
			<i>Cochliomyia hominivorax</i> (Coquerel 1858)	EF560181	JQ246610	JQ246665	JQ246714
			<i>Cochliomyia macellaria</i> (Fabricius 1775)	EF560182	JQ246611	JQ246666	JQ246715
			<i>Hemilucilia segmentaria</i> (Fabricius 1805)	EF560192	JQ246612	JQ246667	JQ246716
			<i>Hemilucilia semidiaphana</i> (Rondani 1850)	JQ246572	JQ246613	JQ246668	JQ246717
			<i>Phormia regina</i> (Meigen 1826)	EF560190	JQ246614	JQ246669	JQ246718
			<i>Protophormia terraenovae</i> (Robineau-Desvoidy 1830)	EF560193	JQ246615	JQ246670	JQ246719
		Calliphorinae	<i>Calliphora croceipalpis</i> (Jaenicke 1867)	JQ246573	JQ246616	JQ246671	JQ246720
			<i>Calliphora vicina</i> (Robineau-Desvoidy 1830)	EF560178	JQ246617	JQ246672	JQ246721
			<i>Calliphora vomitoria</i> (Linnaeus 1758)	EF560179	JQ246618	JQ246673	JQ246722
		Toxotarsinae	<i>Sarconesia chlorogaster</i> (Wiedemann 1830)	JQ246574	JQ246619	JQ246674	JQ246723
		Luciliinae	<i>Hemipyrellia</i> sp. (Townsend 1918) ⁽¹⁾	JQ246575	JQ246620	JQ246675	JQ246724
			<i>Hemipyrellia ligurriens</i> (Wiedemann 1830)	JQ246576	JQ246621	JQ246676	JQ246725
			<i>Lucilia cuprina</i> (Wiedemann 1830)	EF560185	JQ246622	JQ246677	JQ246726

Superfamília	Família	Subfamília	Espécies	Sequências			
				ITS2	28S	COI	16S
Oestroidea	Calliphoridae	Luciliinae	<i>Lucilia eximia</i> (Wiedemann 1819)	EF560186	JQ246623	JQ246678	JQ246727
			<i>Lucilia sericata</i> (Meigen 1826)	EF560187	JQ246624	JQ246679	JQ246728
			<i>Lucilia</i> sp. (Robineau-Desvoidy 1830) ⁽²⁾	JQ246577	JQ246625	JQ246680	JQ246729
		Auchmeromyiinae	<i>Auchmeromyia bequaerti</i> (Roubaud 1913)	JQ246578	JQ246626	-	-
			<i>Cordylobia anthropophaga</i> (Blanchard & Berenger-Feraud 1872)	JQ246579	JQ246627	JQ246681	JQ246730
			<i>Hemigymnochaeta unicolor</i> (Bigot 1888)	JQ246580	JQ246628	JQ246682	JQ246731
			<i>Pachychoeromyia praegrands</i> (Austen 1910)	JQ246581	JQ246629	JQ246683	JQ246732
			<i>Tricyclea</i> sp. (Wulp 1884)	JQ246582	JQ246630	JQ246684	JQ246733
			<i>Bengalia peuhi</i> (Villeneuve 1914)	JQ246583	JQ246631	JQ246685	JQ246734
		Mesembrinellinae	<i>Eumesebrinella benoisti</i> (Séguy 1925)	JQ246584	JQ246632	JQ246686	JQ246735
			<i>Eumesebrinella quadrilineata</i> (Fabricius 1805)	JQ246585	JQ246633	JQ246687	JQ246736
			<i>Mesembrinella bellardiana</i> (Aldrich 1922) ind.1	JQ246586	JQ246635	JQ246688	JQ246738
			<i>Mesembrinella bellardiana</i> (Aldrich 1922) ind.2	EU076455	JQ246634	-	JQ246737
			<i>Mesembrinella</i> sp. (Giglio-Tos 1893) ⁽³⁾	EU076456	JQ246636	-	JQ246739
			<i>Mesembrinella bicolor</i> (Fabricius 1805)	JQ246587	JQ246637	JQ246689	JQ246740
			<i>Mesembrinella peregrina</i> (Aldrich 1922)	EF560188	JQ246638	JQ246690	JQ246741
		Rhiniinae ⁽⁴⁾	<i>Cosmina fuscipennis</i> (Robineau-Desvoidy 1830)	JQ246588	JQ246639	JQ246691	JQ246742
			<i>Rhinia</i> sp. (Robineau-Desvoidy 1830)	JQ246589	JQ246640	JQ246692	JQ246743
			<i>Rhyncomya soyauxi</i> (Karsch 1886)	JQ246590	JQ246641	JQ246693	JQ246744
			<i>Thoracites</i> sp. (Brauer & Bergenstamm 1891)	JQ246591	JQ246642	JQ246694	JQ246745
		Sarcophagidae	<i>Oxysarcodexia thornax</i> (Walker 1849)	-	-	JQ246695	JQ246746
			<i>Peckia ingens</i> (Walker 1849)	JQ246592	JQ246643	-	JQ246747
			<i>Sarcophaga bullata</i> (Parker 1916)	JQ246593	JQ246644	JQ246696	JQ246748

Superfamília	Família	Subfamília	Espécies	Sequências			
				ITS2	28S	COI	16S
Oestroidea	Tachinidae	Exoristinae	<i>Chetogena</i> sp. 1 (Rondani 1856)	JQ246594	JQ246645	JQ246697	JQ246749
			<i>Chetogena</i> sp. 2 (Rondani 1856)	JQ246595	JQ246646	JQ246698	JQ246750
			<i>Tachinidae</i> sp. (N/A)	JQ246597	JQ246648	-	JQ246752
	Oestridae	Dexiinae	<i>Prophorostoma pulchrum</i> (Townsend 1927)	JQ246596	JQ246647	JQ246699	JQ246751
			<i>Cuterebra</i> sp. (Clark 1815)	JQ246598	JQ246649	JQ246700	JQ246753
		Cuterebrinae	<i>Dermatobia hominis</i> (Linnaeus 1781)	EF560183	JQ246650	JQ246701	JQ246754
Muscoidea	Muscidae	Muscinae	<i>Haematobia irritans</i> (Linnaeus 1758)	EF560184	JQ246651	JQ246702	JQ246755
			<i>Musca domestica</i> (Linnaeus 1758)	EF560189	JQ246652	JQ246703	JQ246756
			<i>Stomoxys calcitrans</i> (Linnaeus 1758)	EF560191	JQ246653	JQ246704	JQ246757
		Cyrtoneurinae	<i>Cyrtoneuropsis maculipennis</i> (Macquart 1843)	JQ246599	JQ246654	-	JQ246758
	Fannidae	Fanniinae	<i>Fannia</i> sp. (Robineau-Desvoidy 1830)	JQ246600	JQ246655	JQ246705	JQ246759
Hippoboscoidea	Glossinidae	N/A	<i>Glossina morsitans</i> (Westwood 1851)	JQ246601	JQ246656	JQ246706	JQ246760
	Hippoboscidae	Ornithomyinae	<i>Ornithoetona erythrocephala</i> (Leach 1817)	JQ246602	JQ246657	JQ246707	JQ246761

(1) Uma fêmea de *H. ligurriens* (Wiedemann 1830) ou *H. tagaliana* (Bigot 1877);

(2) Identificação ambígua como *L. cuprina* (Wiedemann 1830) ou *L. sericata* (Meigen 1826), com incongruência entre marcadores nucleares e mitocondriais (Tourle et al. 2009);

(3) Previamente identificada como *M. bellardiana* (Aldrich 1922), necessita revisão;

(4) A subfamília de Calliphoridae Rhiniinae foi recentemente elevada ao status de família, Rhiniidae (Pape & Thompson 2010).

ACTCGGCTCGTGGGATTTCGAT-3'), 0,4µM "oligo" 28S (5'-GTTAGTTTCTTTTCCTCCCCT-3'), 2,5U de *Taq* DNA polimerase (Fermentas) e 1-2µg de DNA total extraído para uma reação com volume final de 25µl. As reações de PCR foram realizadas com um passo inicial de desnaturação a 95°C por 3 minutos, seguido por 35 ciclos de 95°C por 1 minuto, 55°C por 45 segundos e 72°C por 2 minutos, com um passo final de extensão da cadeia a 72°C por 3 minutos.

As reações de PCR para as demais regiões (28S, COI e 16S) foram conduzidas com 10mM Tris-HCl (pH 8,8), 50mM KCl, 1mM MgCl₂, 80µM dNTPs, 0,4µM "oligo" direto, 0,4µM "oligo" reverso, 2,5U *Taq* DNA polimerase (Fermentas) e 1-2µg de DNA total extraído. As reações de PCR foram realizadas com um passo inicial de desnaturação a 95°C por 3 minutos, seguido por 35 ciclos de 95°C por 30 segundos, 52°C por 1 minuto e 72°C por 2 minutos, com um passo final de extensão da cadeia a 72°C por 2 minutos. As combinações de "oligos" direto e reverso utilizados para amplificar cada uma das regiões foram: C1-N2320 (5'-AATCCTAATAATCCAATAGC-3') e TW-J-1287 (5'-ACTAATAGCCTTCAAAGC-3') para o COI; 28S-F1 (5'-GGGAGGAAAAGAACTAACAAGG-3') e 28S-R1 (5'-CTGTTTCGGTCTTCCATCAGGG-3') para o 28S; e LR-N-13398 (5'-CGCCTGTTTAACAAAAACAT-3) e L1R (5'-CCATTGCACTAATCTGCC-3') para o 16S. Todos os produtos de PCR foram visualizados em gel de agarose 1X TAE (40mM Tris-acetato, 1mM EDTA) corados com brometo de etídeo.

Os produtos de PCR foram purificados utilizando-se o kit "Invisorb Fragment Clean-up Kit" (Invitex) e clonados em um vetor pGEM-T Easy (Promega), seguindo as instruções do fabricante. Células competentes de *E. coli* linhagem DH5α foram transformadas utilizando-se o protocolo de choque químico com CaCl₂ (adaptado de Sambrook et al.1989) e plaqueadas em meio LB sólido com 50mg/mL de X-Gal e 50µg/mL

de Ampicilina. As placas foram incubadas a 37°C por 12-14 horas. Os plasmídeos foram extraídos das células transformadas utilizando-se o método de hidrólise alcalina (Sambrook et al. 1989) e foram digeridos com 5U da enzima de restrição *EcoRI* por 2 horas a 37°C. O resultado das digestões foram visualizados por eletroforese em gel de agarose 1X TAE corado com brometo de etídeo. Pelo menos três clones carregando o inserto de tamanho esperado correto foram submetidos à reação de sequenciamento.

As reações de sequenciamento foram conduzidas em um sequenciador ABI 3700 (Applied Biosystems) usando o "Big Dye Terminator Cycle Sequencing Ready Reaction Kit" (Applied Biosystems), de acordo com as instruções do fabricante, com os "oligos" universais M13 direto (5'-GTAAAACGACGGCCAG-3') e M13 reverso (5'-CAGGAAACAGCTATGAC-3') e 200-600ng de DNA.

2.3. Predição das estruturas secundárias e alinhamento de sequências

Estruturas secundárias previamente descritas e publicadas para as regiões 16S (Guttel et al. - não publicado; de Rijk et al. 1997; Buckley et al. 2000) e ITS2 (Young & Coleman 2004) de *D. melanogaster* foram utilizadas para modelar por homologia ambas as regiões em Calyptratae. Dado que a região do 16S sequenciada é bastante conservada entre os táxons amostrados e, em uma maior escala, entre os Insecta em geral (Buckley et al. 2000), a estrutura secundária dessa região foi determinada manualmente para cada espécie usando a estrutura de *D. melanogaster* como modelo. As estruturas individuais das duas hélices mais variáveis, H75 e H84 (Buckley et al. 2000) foram preditas utilizando o programa mfold v3.0, versão "webserver" (Zuker 2003).

Para a região do ITS2, a modelagem da estrutura secundária foi baseada em (1) modelagem por homologia usando como modelo a estrutura descrita para *Drosophila melanogaster* (Young & Coleman 2004) e para algumas espécies de Calliphoridae

(Marinho et al. 2011); (2) modelagem *in silico* com o programa mfold 2.3 (versão webserver - Zuker 2003), utilizando os parâmetros padrões para predição da estrutura, com exceção da temperatura, mudada para 25°C, dado que esta temperatura (padrão para temperatura ambiente) é mais plausível de representar a temperatura média corporal de um inseto; e (3) comparação entre as estruturas preditas para todos os espécimes sob análise e estabelecimento de um padrão comum de estrutura secundária. Este procedimento mais elaborado foi utilizado dado que a região do ITS2 é mais variável tanto em termos de sequência quanto de estrutura secundária, embora a conformação geral adotada pela molécula seja conservada entre todos os eucariotos (Joseph et al. 1999). O nível de variação presente na molécula não é, contudo, alto o suficiente para obscurecer o sinal filogenético da região, como mostrado por Coleman (2003, 2007, 2009) e Muller et al. (2007). As estruturas inferidas tanto para a região do ITS2 quanto do 16S foram anotadas no formato ponto-parêntese ("dot-bracket") e visualizadas no programa VARNA 3.3 (Darty et al. 2009). As estruturas secundárias consenso para ambas as regiões foram determinadas com o programa SecondaryStructConsensus do pacote PHASE 2.0 (Gowri-Shankar & Jow 2006).

O alinhamento das sequências das regiões do 16S, COI e 28S foram conduzidos com o programa ClustalX 2.0 (Larkin et al. 2007), enquanto para a região ITS2 o alinhamento das sequências utilizando informações estruturais foi realizado com o programa 4SALE (Seibel et al. 2006). Com exceção do alinhamento da região do COI, alguns ajustes manuais finais foram realizados nos alinhamentos.

2.4. Análises de sinal filogenético, saturação de substituição e congruência filogenética.

O sinal filogenético presente em cada uma das quatro regiões gênicas sequenciadas foi avaliado usando o método baseado em parcimônia de Steel et al. (1993) e o método de informação baseado em entropia ("information entropy-based") de Xia et al. (2003) e Xia & Lemey (2009), ambos implementados no programa DAMBE 5.2.38 (Xia 2001; Xia & Xie 2001). O teste de Steel foi realizado amostrando-se todos os possíveis quartetos de espécies ("quartets of species") no conjunto de dados de sequências e o nível de sinal presente em cada sequência foi avaliado de acordo com os valores de $\phi_{\text{médio}}$ individuais, o qual é baseado no número de quartetos no qual a sequência forneceu informações suficientes para sua resolução. De acordo com a interpretação proposta por Xia & Lemey (2009), sequências com $\phi_{\text{médio}} < 0,04$ podem ser consideradas como pouco informativas no contexto do conjunto de sequências sob análise. O teste de Xia, o qual é baseado no cálculo de um Índice de Saturação de Substituição crítico (ISS.c) através de simulações com amostragens aleatórias de 4, 8, 16 e 32 OTUs (do inglês "Operational Taxonomic Units" - Unidades Taxonômicas Operacionais) do conjunto de dados e a comparação do mesmo com um ISS calculado a partir do conjunto completo de sequências, foi conduzido considerando-se apenas os sítio completamente resolvidos (sem "gaps") e com 10.000 repetições. Segundo Xia & Lemey (2009), se $\text{ISS} > \text{ISSc}$, então o sinal filogenético no conjunto de dados é baixo e pouco útil para análises filogenéticas.

A congruência filogenética entre os quatro marcadores utilizados foi avaliada usando o programa Concaterpillar 1.4 (Leigh et al. 2008), o qual usa agrupamentos hierárquicos e testes de razão de verossimilhanças para detectar congruência entre conjuntos distintos de sequências e avaliar se eles podem ser combinados por concatenação em um mesmo conjunto de dados (hipótese nula do teste de congruência topológica) e, se combináveis, se os mesmos podem compartilhar valores estimados de parâmetros (hipótese

nula do teste de congruência dos comprimentos de ramo) ou se os conjuntos devem ter parâmetros estimados e otimizados separadamente. O programa Concaterpillar 1.4 usa o programa RaxML-VI-HPC (Stamatakis 2006) para os cálculos de máxima-verossimilhança.

2.5. Análises filogenéticas

A seleção de modelos de substituição foi realizada com o programa MrAIC 1.4.4 (Nylander 2004) tanto para o concatenado final das quatro regiões gênicas sequenciadas quanto para cada uma dessas regiões gênicas separadas e para suas subpartições utilizadas nas análises subsequentes.

A análise filogenética realizada utilizando-se o método de máxima-parcimônia (MP) foi conduzida com o programa TnT v1.1 (Goloboff et al. 2003) utilizando a opção de "new technology search" (parâmetros: "search at level 50"; "initial addseqs = 15"; "find minimum tree length ten times") e tratando os sítios com "gaps" como dados ausentes ("missing data"). Os valores de suporte para os clados foram estimados tanto pelo método de re-amostragem por "bootstrap" (BS) (1.000 replicatas) quanto pelo de jackknife (JK) (1.000 replicatas, 36% de probabilidade de remoção independente de caracteres), usando para ambos as mesmas opções de busca utilizadas na análise original.

As análises de máxima-verossimilhança (ML) foram conduzidas utilizando dois programas: (1) Garli v1.0 (Zwickl 2006), cuja implementação é baseada no uso de algoritmos genéticos para a busca de árvores, a qual foi realizada empregando-se os seguintes parâmetros: "three independent replicates"; "20.000.000 generations"; "50 individuals per generation"; "default parameters for automated stopping"; e (2) PhyML 3.0 (Guindon & Gascuel 2003), o qual implementa métodos mais tradicionais na busca de árvores, a qual foi realizada com os seguintes parâmetros: "analyze multiple datasets = yes"; "number of datasets = 4"; "all parameters estimated"; "tree topology search operations

= SPR moves". Para ambos os programas, o modelo de substituição nucleotídica utilizado foi o modelo GTR+I+G, escolhido pelo programa MrAIC 1.4.4 como o melhor modelo que se ajusta ao conjunto de dados final com as quatro regiões gênicas concatenadas. O valor de suporte dos clados foi estimado usando o método de re-amostragem por "bootstrap" (BS) (1.000 replicatas em ambos os programas) e também o teste tipo-Shimoidara/Hasegawa ("SH-like test" - SH) (apenas no programa PhyML 3.0).

As análises de inferência Bayesiana (BI) foram realizadas com os programas MrBayes v3.1.2 (Ronquist & Huelsenbeck 2003) e mcmcphase do pacote de programas PHASE 2.0 (Gowri-Shankar & Jow 2006). Como padrão, cada uma das quatro regiões gênicas foi considerada como uma partição diferente, evoluindo de acordo com o melhor modelo de substituição definido pelo programa MrAIC. As regiões do ITS2 e do 16S foram posteriormente sub-particionadas de acordo com suas estruturas secundárias adquiridas em regiões de simples e dupla-fita, enquanto a região do COI foi dividida de acordo com as três posições dos códons (primeira - 1st, segunda - 2nd e terceira - 3rd). Uma vez que o programa PHASE implementa uma série de modelos que lidam com regiões pareadas em estruturas secundárias (modelos de RNA), considerando assim a interdependência entre os sítios no processo de evolução da sequência, o modelo RNA7A (modelo mais geral em sua categoria) foi escolhido para a realização das análises iniciais, dado que os modelos com 7-estados (que combinam todos os possíveis "mismatches" - duplas de nucleotídeos não pareados na estrutura - em uma única categoria) são um intermediário interessante entre os modelos mais restritos de 6-estados (que não lidam explicitamente com os "mismatches") e os modelos mais gerais de 16-estados (os quais consideram cada possível estado de "mismatch" em sua própria categoria).

Após a avaliação dos resultados das análises iniciais, a combinação de partições e modelos empregada no programa PHASE que resultou no maior valor de verossimilhança marginal foi analisada de forma mais extensa, substituindo-se o modelo de estrutura secundária pelos modelos mais gerais nas categorias de 6-estados (RNA6A) e 16-estados (RNA16A). Além disso, o modelo RNA16I também foi testado e avaliado dado que o mesmo considera uma racionalização diferente dos processos de substituição nucleotídica que ocorrem em regiões de dupla-fita. O modelo RNA16I considera o processo de substituição dos pares de bases formados na estrutura em duas etapas, de forma semelhante ao modelo "Doublet" implementado no programa MrBayes. Nos demais modelos de estrutura implementados no programa PHASE, as substituições de pares em regiões de hélices é modelado em uma única etapa. A Tabela 2 resume todas as análises conduzidas com o método de inferência Bayesiana.

Em todas as análises, cada partição foi definida com um conjunto de parâmetros estimados e otimizados separadamente (comandos no MrBayes: "unlink statefreq=(all) revmat=(all) shape=(all) pinvar=(all)" e "prset applyto=(all) ratepr=variable"). Para cada combinação de partições e modelos, duas análises independentes em cada programa, conduzidas com diferentes pontos iniciais ("random seeds"), foram realizadas. Cada análise foi realizada por 20.000.000 de gerações, com frequência de amostragem a cada 1.000 e, 25% das observações amostradas foram descartadas após checar pela convergência das cadeias. O valor de suporte de clados foi estimado usando os valores de probabilidade *a posteriori* (PP) nas árvores consenso resultantes (regra da maioria de 50% estendida, na qual clados com probabilidade < 0,50 são considerados na árvore final se os mesmos não são incompatíveis com as relações já consideradas - opção padrão no pacote PHASE e comandos "sumt contype=allcompat" no programa MrBayes).

Tabela 2. Resumo de todas as estratégias de particionamento empregadas nas análises de inferência Bayesiana.

Programa	Estratégia de particionamento	Nº de partições	Regiões							Nº de parâmetros (modelo / partições)	Nº de parâmetros (Total)	
			ITS2		COI			28S	16S			
			ss	ds	1st	2nd	3rd		ss			ds
MrBayes	Não estrutura / códon (NS/C)	6	HKY+I+G		GTR+I+G	HKY	HKY+G	GTR+I+G	GTR+I+G		51	160
	Não estrutura / não códon (NS/NC)	4	HKY+I+G		GTR+I+G			GTR+I+G	GTR+I+G		40	149
	Estrutura / códon (S/C)	8	HKY+I+G	Doublet (GTR)+I+G	GTR+I+G	HKY	HKY+G	GTR+I+G	GTR+I+G	Doublet (GTR)+I+G	97 ⁽¹⁾	206
	Estrutura / não códon (S/NC)	6	HKY+I+G	Doublet (GTR)+I+G	GTR+I+G			GTR+I+G	GTR+I+G	Doublet (GTR)+I+G	86 ⁽¹⁾	195
PHASE	Não estrutura / códon (NS/C)	6	HKY+I+G		TN93+I+G	HKY	HKY+G	GTR+I+G	GTR+I+G		48	157
	Não estrutura / não códon (NS/NC)	4	HKY+I+G		GTR+I+G			GTR+I+G	GTR+I+G		40	149
	Estrutura / códon (S7A/C)	8	HKY+I+G	RNA7A+I+G	TN93+I+G	HKY	HKY+G	GTR+I+G	GTR+I+G	RNA7A+I+G	106	215
	Estrutura / não códon (S7A/NC)	6	HKY+I+G	RNA7A+I+G	GTR+I+G			GTR+I+G	GTR+I+G	RNA7A+I+G	98	207
	Estrutura / códon (S6A/C)	8	HKY+I+G	RNA6A+I+G	TN93+I+G	HKY	HKY+G	GTR+I+G	GTR+I+G	RNA6A+I+G	91 ⁽²⁾	200
	Estrutura / códon (S16A/C)	8	HKY+I+G	RNA16A+I+G	TN93+I+G	HKY	HKY+G	GTR+I+G	GTR+I+G	RNA16A+I+G	91 ⁽²⁾	200
	Estrutura / códon (S16I/C)	8	HKY+I+G	RNA16I+I+G	TN93+I+G	HKY	HKY+G	GTR+I+G	GTR+I+G	RNA16I+I+G	94 ⁽¹⁾	203

(1) Tanto o modelo "Doublet" do MrBayes quanto o modelo RNA16I do pacote PHASE consideram o processo de mudanças nos pares de bases em regiões de dupla-fita como ocorrendo em duas etapas. Ambos possuem 16 parâmetros de frequências e 6 parâmetros de taxas de substituição do modelo GTR (20 parâmetros livres);

(2) O modelo RNA6A tem 15 parâmetros de taxas de substituição e 6 de frequências (com 2 fixados) resultando em 19 parâmetros livre, o mesmo número de parâmetros do modelo RNA16A, o qual tem 16 parâmetros de frequências e 5 de taxas (com 2 fixados).

2.6. Comparação de modelos por critérios de máxima-verossimilhança e Bayesianos.

As diferentes combinações de partições e modelos usadas nas análises Bayesianas foram avaliadas usando os métodos de máxima-verossimilhança baseados em teoria da informação AIC (Akaike Information Criteria), AICc (AIC corrigido) e BIC (Bayesian Information Criteria), com o valor de máxima-verossimilhança amostrado durante a fase estacionária da corrida MCMC (Markov chain - Monte Carlo) sendo usado como o estimador de máxima-verossimilhança necessário para as análises. O número de parâmetros livres em cada modelo é mostrado na Tabela 2. A comparação de modelos usando o método Bayesiano de "Bayes factor" também foi utilizada na avaliação dos resultados das análises Bayesianas. Como estimador do valor de verossimilhança marginal, necessária para o cálculo do "Bayes factor", foi utilizada a média harmônica dos valores de verossimilhança amostrados durante a fase estacionária da corrida MCMC, como sugerido por Nylander et al. (2004). As interpretações dos resultados foram feitas baseando-se na tabela apresentada por Kass & Raftery (1995).

Com o objetivo de avaliar se as mudanças significativas nos valores de verossimilhança entre as diferentes análises foram acompanhadas de alterações significativas nas topologias estimadas, no suporte geral das árvores e nos comprimentos de ramos inferidos, todas as árvores Bayesianas foram: (1) comparadas usando a medida de distâncias simétricas entre topologias proposta por Robinson & Foulds (1981), a qual se baseia no número de passos necessários para transformar uma topologia em outra, e que se encontra implementada no programa TreeDist do pacote PHYLIP 3.67 (Felsenstein 2005); (2) avaliadas em termos de alterações nos valores de suporte de clado específicos e também do suporte geral das árvores, estimado pela média aritmética dos valores de probabilidade a

posteriori de todos os clados na filogenia; e (3) comparadas pelo tamanho total da árvore ("total tree length" - TL), o qual é a soma de todos os comprimentos de ramo presentes na topologia e que foi estimado utilizando-se a média aritmética dos comprimentos de ramo de todas as árvores amostradas durante a fase estacionária da corrida MCMC.

2.7. Análises de mapeamento de verossimilhança

Após a conclusão das análises filogenéticas, hipóteses conflitantes entre os diferentes métodos empregados foram avaliadas usando-se mapas de verossimilhança (Strimmer & Von Haeseler 1997), como implementado no programa Tree-Puzzle 5.2 (Schmidt et al. 2002). O método é baseado na amostragem aleatória de quartetos de espécies no conjunto de dados completo de sequências e na inferência de árvores para esses táxons amostrados. O número de topologias completa e parcialmente resolvidas e completamente não resolvidas é então contado e, apesar da interpretação dos resultados ser variável, em geral valores maiores que 5% de árvores completamente não resolvidas é tido como indicador de sinal filogenético fraco para as sequências sendo analisadas (Schmidt et al. 2002). O mapa de verossimilhança também pode ser aplicado estipulando-se previamente conjuntos de sequências que pertençam, a princípio, ao mesmo clado e em seguida avaliando-se o número de topologias total ou parcialmente resolvidas que favorecem cada uma das três árvores possíveis para quatro OTUs. Ambos os modos de análise foram realizados, sendo que para o segundo modo quatro grupos foram definidos baseados nos resultados das análises filogenéticas previamente conduzidas. As análises de mapeamento de verossimilhança foram conduzidas amostrando-se todos os quartetos de sequências possíveis no conjunto de dados geral, com a opção de estimação de parâmetros mais precisa (método mais lento) usando a opção "Quartet sampling + NJ tree" para inferência das árvores.

3. Resultados

3.1. Análises de sequências e estruturas secundárias

O tamanho total dos fragmentos amplificados via PCR e o conteúdo nucleotídico de cada uma das quatro regiões gênicas amostradas são mostrados na Tabela 3. Em ambos os marcadores nucleares (ITS2 e 28S) foi encontrado uma considerável variação no tamanho das sequências amplificadas, a qual é mais pronunciada nas espécies da família Oestridae. Todas as regiões apresentaram um viés de composição nucleotídica no sentido de excesso de A+T, o qual é maior na região do ITS2.

As estruturas secundárias previstas para as regiões ITS2 e 16S são mostradas nas Figuras 1 e 2. Com exceção das hélices H75 e H84, seguindo a nomenclatura de Buckley et al. (2000), as regiões de hélice-alça do 16S são bastante conservadas e foram modeladas de forma manual diretamente a partir da estrutura descrita para *D. melanogaster*. Enquanto para a hélice H75 as estruturas previstas pelo programa mfold foram bastante consistentes entre os táxons amostrados (Figura 1C), a modelagem da estrutura da hélice H84, mais variável, apresentou maiores dificuldades, sendo que para todas as estruturas previstas apenas dois pares de bases são comuns (mostrados na Figura 1A). Algumas das conformações alternativas para essa hélice são mostradas na Figura 1B.

Para a região do ITS2, as estruturas previstas corroboraram o modelo de 4 domínios, amplamente aceito para a região em Eukaryota (Joseph et al. 1999), como mostrado na Figura 2. A organização estrutural geral, contudo, pode ser variável em cada região de hélice-alça, com destaque para o domínio III, o qual é provavelmente ramificado em espécies das superfamílias Oestroidea e Muscoidea mas não nas de Hippoboscoidea e Acalyptratae (baseado na estrutura descrita para *D. melanogaster*). As porções iniciais das regiões de hélice dos domínios I, II e III são bastante conservadas, assim como são os

motivos de sequência GUCUAGCAUA na porção terminal do domínio III e o resíduo de pirimidina não pareado no domínio II, o qual é um C na maioria das espécies (motivos estes descritos previamente por Young & Coleman (2004) e Marinho et al. (2011).

O alinhamento final do conjunto concatenado de sequências das quatro regiões gênicas, corrigidos por informação de estrutura secundária para as regiões ITS2 e 16S, possui 3355 nucleotídeos.

Tabela 3. Tamanho total do fragmento amplificado via PCR e sequenciado (em pares de bases - pb) e conteúdo nucleotídico (A+T %) de cada uma das quatro regiões gênicas usadas nas análises filogenéticas.

Espécies	ITS2		28S		COI		16S	
	pb	A+T (%)	pb	A+T (%)	pb	A+T (%)	pb	A+T (%)
<i>Chloroprocta idioidea</i>	320	81,88	1010	67,43	752	67,42	540	73,52
<i>Chrysomya albiceps</i>	317	81,70	1009	67,20	752	69,02	491 ³	75,76
<i>Chrysomya bezziana</i>	302	79,14	1011	66,96	752	67,82	540	73,52
<i>Chrysomya chloropyga</i>	298	78,86	1008	67,16	752	66,89	539	73,84
<i>Chrysomya megacephala</i>	306	79,08	1007	67,03	752	68,88	N/A	N/A
<i>Chrysomya putoria</i>	302	79,80	1007	67,13	752	68,09	539	73,84
<i>Chrysomya rufifacies</i>	320	80,94	1009	67,10	752	68,48	538	73,98
<i>Cochliomyia hominivorax</i>	351	81,20	1009	67,89	752	67,69	537	73,00
<i>Cochliomyia macellaria</i>	344	81,10	1009	67,59	752	67,69	538	73,23
<i>Hemilucilia segmentaria</i>	327	82,26	1012	67,49	752	69,28	516 ³	74,81
<i>Hemilucilia semidiaphana</i>	321	82,55	1010	67,33	752	67,95	538	73,61
<i>Phormia regina</i>	327	82,87	1010	67,33	752	67,82	421 ³	76,48
<i>Protophormia terraenovae</i>	331	82,78	1008	67,56	752	67,15	522 ³	74,33
<i>Calliphora croceipalpis</i>	327	81,35	1010	67,43	752	68,35	538	73,23
<i>Calliphora vicina</i>	327	81,35	1010	67,43	752	68,22	538	73,23
<i>Calliphora vomitoria</i>	314	81,21	1010	67,23	752	69,55	526 ³	74,14
<i>Sarconesia chlorogaster</i>	309	79,61	1007	67,53	752	69,81	528 ³	74,24
<i>Hemipyrellia sp.</i>	315	82,86	1010	67,62	752	67,55	538	72,49
<i>Hemipyrellia ligurriens</i>	315	82,86	1010	67,62	752	67,29	538	72,30
<i>Lucilia cuprina</i>	336	81,55	1006	67,50	752	69,81	513 ³	73,88
<i>Lucilia eximia</i>	326	82,52	1009	67,20	752	66,36	538	73,42
<i>Lucilia sericata</i>	321	79,44	1006	67,30	752	69,68	520 ³	74,04

Capítulo III

Espécies	ITS2		28S		COI		16S	
	pb	A+T (%)	pb	A+T (%)	pb	A+T (%)	pb	A+T (%)
<i>Lucilia sp.</i>	337	80,71	1006	67,50	752	69,55	514 ³	73,54
<i>Auchmeromyia bequaerti</i>	330	83,03	1009	66,80	N/A	N/A	N/A	N/A
<i>Cordylobia anthropophaga</i>	336	83,33	1011	67,26	752	67,15	539	73,28
<i>Hemigymnochaeta unicolor</i>	322	81,99	1010	67,33	752	69,68	539	74,21
<i>Pachychoeromyia praegrans</i>	342	83,33	1010	67,23	752	68,88	521 ³	74,47
<i>Tricyclops sp.</i>	333	81,38	1013	67,82	752	69,28	539	74,40
<i>Bengalia peuhi</i>	298	83,22	1009	67,59	752	69,02	537	73,74
<i>Eumesembrinella benoisti</i>	356	82,87	1069	68,48	752	67,95	519 ³	77,07
<i>Eumesembrinella quadrilineata</i>	341	81,23	1049	68,16	752	68,35	536	76,12
<i>Mesembrinella bellardiana ind.1</i>	313	80,19	1055	68,53	752	66,22	516 ³	74,81
<i>Mesembrinella bellardiana ind.2</i>	307	79,48	1055	68,44	N/A	N/A	536	74,07
<i>Mesembrinella sp.</i>	334	77,54	1045	66,12	N/A	N/A	513 ³	74,46
<i>Mesembrinella bicolor</i>	304	81,25	1062	68,55	752	69,41	526 ³	77,00
<i>Mesembrinella peregrina</i>	291	80,07	1066	68,76	752	69,68	535	75,89
<i>Cosmina fuscipennis</i>	328	79,88	1014	66,77	755	70,07	538	75,28
<i>Rhinia sp.</i>	335	83,28	1008	67,36	752	70,74	539	74,77
<i>Rhyncomyia soyauxi</i>	345	79,42	1015	67,09	752	68,22	537	74,30
<i>Thoracites sp.</i>	300	81,67	1013	66,83	752	69,02	538	74,77
<i>Oxysarcodexia thornax</i>	N/A	N/A	N/A	N/A	752	67,02	481 ³	75,68
<i>Peckia ingens</i>	336	80,06	1009	67,20	N/A	N/A	538	72,49
<i>Sarcophaga bullata</i>	346	80,35	1007	67,33	752	67,42	526 ³	74,52
<i>Chetogena sp. 1</i>	311	80,39	1008	66,37	752	70,08	542	74,54
<i>Chetogena sp. 2</i>	311	80,39	1008	66,47	752	70,21	542	74,54
<i>Tachinidae sp.</i>	315	80,63	1014	67,55	N/A	N/A	524 ³	76,72
<i>Prophorostoma pulchrum</i>	384	77,08	1019	67,62	752	68,62	517 ³	76,02
<i>Cuterebra sp.</i>	587 ¹	81,94	1203	68,25	752	68,35	536	75,37
<i>Dermatobia hominis</i>	632 ¹	80,70	1103	69,08	752	67,29	537	74,67
<i>Haematobia irritans</i>	347	76,08	1003	64,21	752	70,74	537	74,30
<i>Musca domestica</i>	365	75,89	1004	63,94	752	68,62	538	72,68
<i>Stomoxys calcitrans</i>	324	75,00	1005	65,17	752	68,88	538	73,98
<i>Cyrtoneuropsis maculipennis</i>	313	81,15	1009	66,80	N/A	N/A	539	73,28
<i>Fannia sp.</i>	378	85,45	1021	67,09	752	68,22	538	74,54
<i>Glossina morsitans</i>	251	78,88	1029	67,83	752	67,69	537	75,23
<i>Ornithoctona erythrocephala</i>	401 ²	85,54	1061	70,69	752	69,28	537	74,12
Média	337	80,95	1022	67,37	752	68,53	529 (538 ⁴)	74,33 (73,97 ⁴)

- (1) As duas espécies de Oestridae possuem uma grande inserção no subdomínio IIIc, o qual é ramificado;
(2) *O. erythrocephala* possui uma grande inserção no domínio I, o qual é ramificado;
(3) Sequências da região 16S que não correspondem ao tamanho total do fragmento amplificado via PCR devido à retirada de uma região com má qualidade de sequenciamento no início do cromatograma;
(4) Média calculada usando apenas as sequências que correspondem ao tamanho total do fragmento amplificado via-PCR.

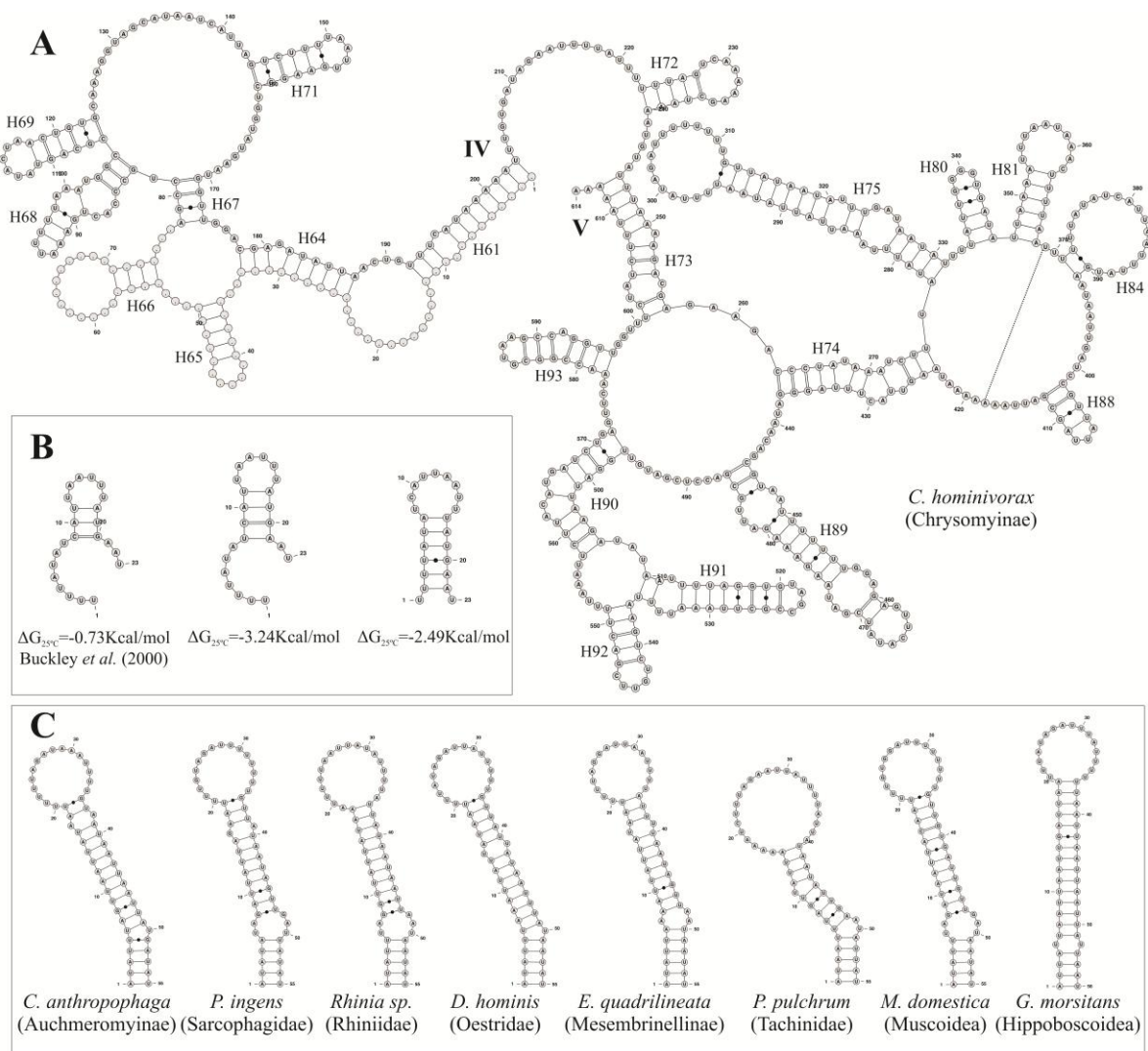


Figura 1. Estrutura secundária predita para os domínio IV e V da região 16S de Calyptratae. (A) Estrutura secundária da região 16S de *C. hominivorax*, como utilizada nas análises filogenéticas, incluindo a nomenclatura sugerida em Buckley *et al.* (2000); (B) Conformações alternativas possíveis para a hélice H84 em *C. hominivorax* com seus respectivos valores de mínima energia livre: esquerda = modelada de acordo com o modelo proposto em Buckley *et al.* (2000); meio = estrutura com o menor valor de energia livre predita pelo programa mfold; direita = conformação alternativa (sub-ótima) sugerida pelo programa mfold. Em (A), apenas os dois pares de bases comuns nas estruturas preditas da hélice H84 para todas as espécies em análise são mostrados. (C) Conformação predita pelo programa mfold para a hélice H75 de algumas espécies representativas em Calyptratae.

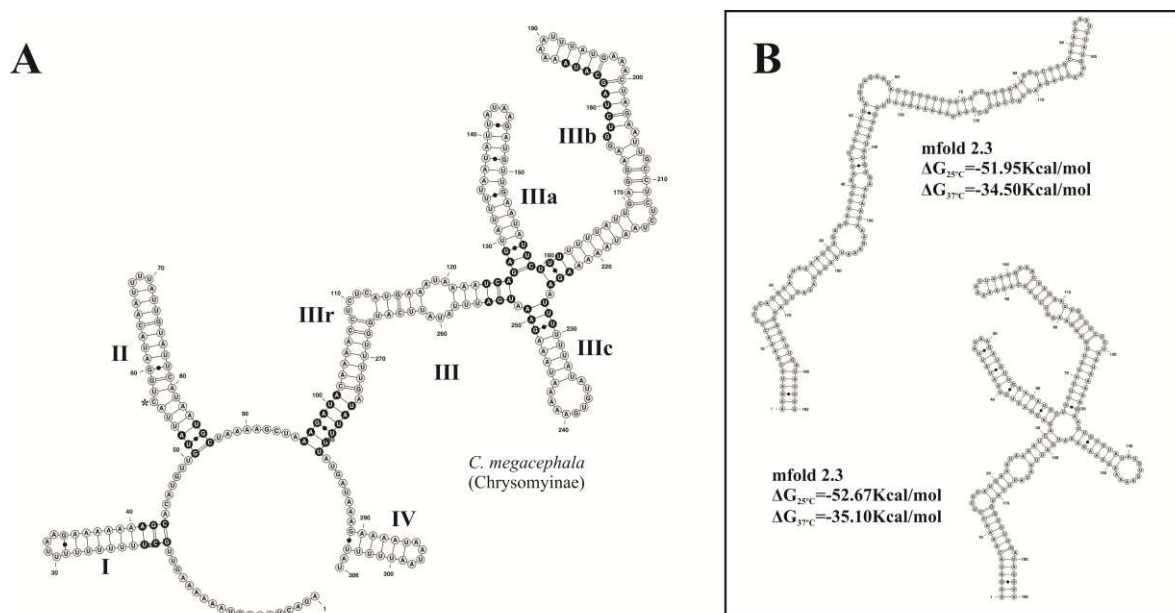


Figura 2. (A) Estrutura secundária predita para a região ITS2 de *C. megacephala*. Alguns motivos conservados de sequência primária são mostrados em círculos pretos. O resíduo de pirimidina não pareado no domínio II é mostrado com um "*". A porção inicial do domínio III é anotada como IIIr, como explicado no texto. (B) Conformações alternativas previstas pelo programa mfold para o domínio III com seus respectivos valores de energia livre: Acima = estrutura não ramificada, menos estável e não predita em todas as espécies; Abaixo = estrutura ramificada, mais estável e predita em todas as espécies; (C) e (D) Conformações previstas para os domínios I, II e III para algumas espécies representativas dos grupos Calliphoridae s.s., Sarcophagidae e Rhiniinae. (E) Estrutura secundária predita para a região do ITS2 em alguns grupos de Oestroidea, Muscoidea e Hippoboscoidea. A estrutura secundária de *D. melanogaster*, como descrita por Young & Coleman (2004), é mostrada na porção inferior da figura.

3.2. Análises de sinal filogenético, saturação de substituição e congruência filogenética.

As análises de sinal filogenético mostraram baixos níveis de saturação nas quatro regiões gênicas analisadas, indicando a possibilidade do uso dessas regiões em análises de filogenia molecular, sobretudo na reconstrução das relações entre as espécies da família Calliphoridae s.s. (Tabela 4).

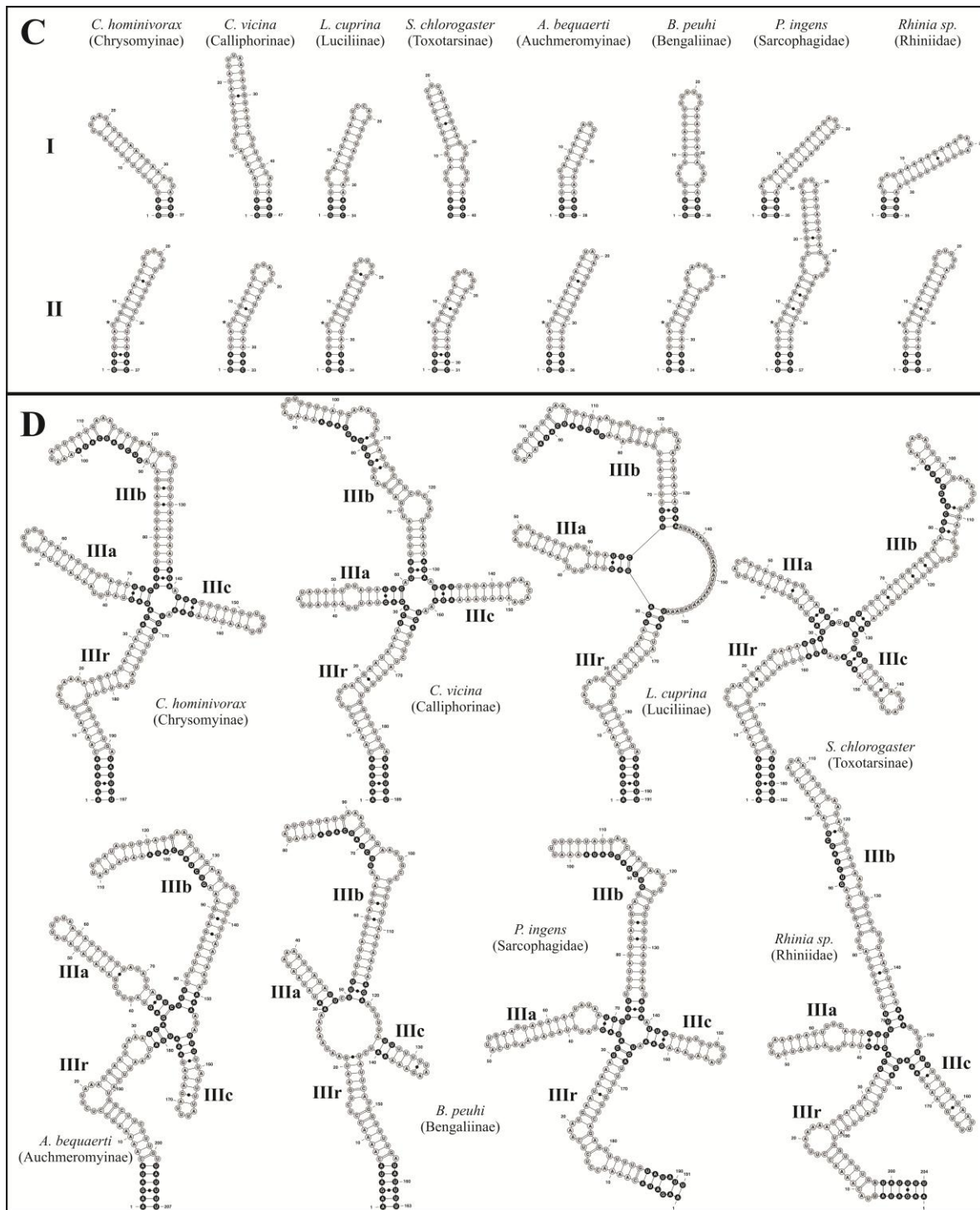


Figura 2. (Continuação).

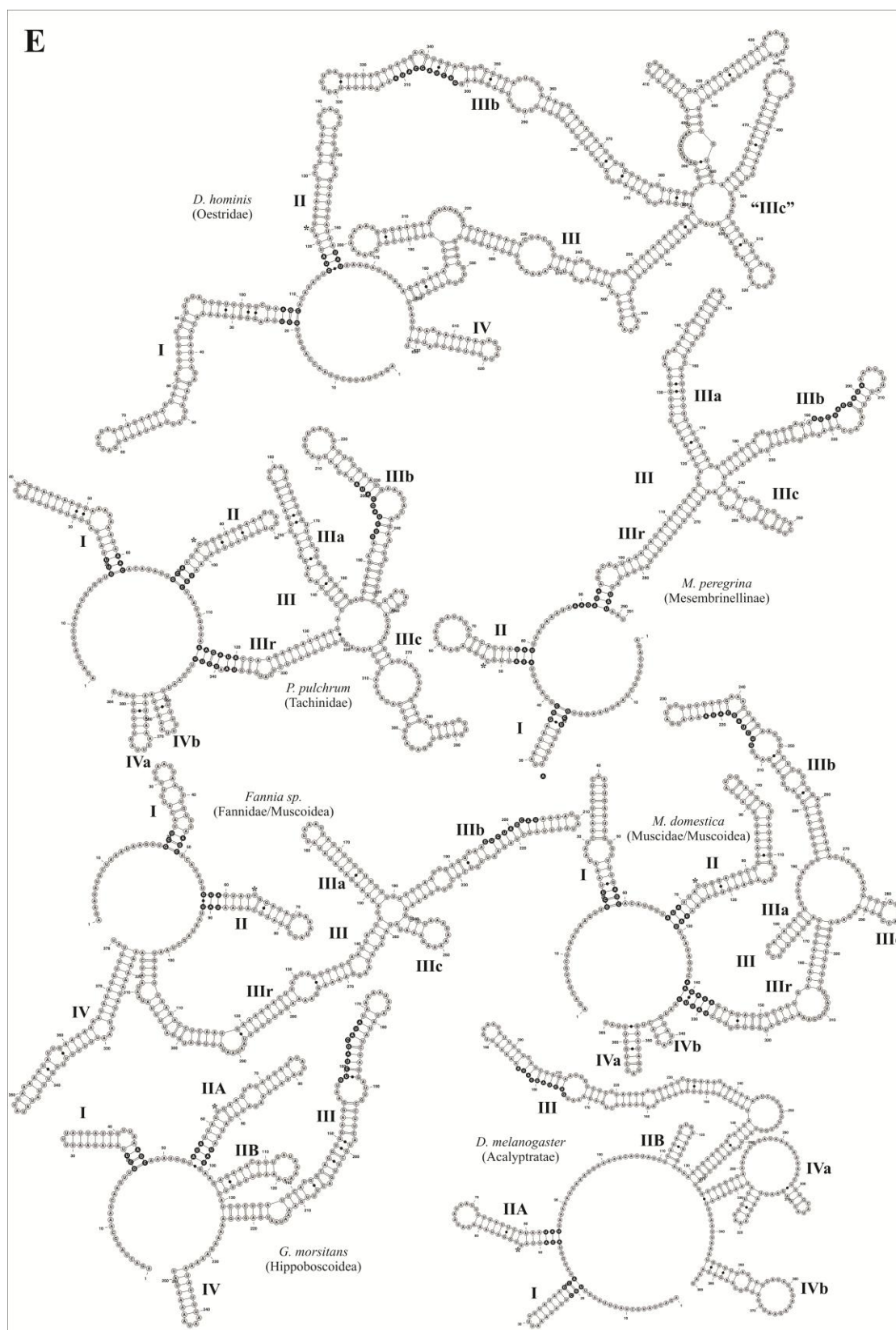


Figura 2. (Continuação).

Para os níveis taxonômicos superiores, Oestroidea e Calyptratae, as regiões do 16S e 28S parecem ser mais adequadas ($\phi_{\text{médio}}$ nos testes de Steel $> 0,04$ para a maioria das espécies), apesar das regiões do COI e do ITS2 não apresentarem níveis de saturação significativo nesses níveis hierárquicos (analisando-se somente os sítios totalmente resolvidos no teste de Xia).

O sub-particionamento das regiões ITS2 e 16S de acordo com a conformação estrutural de seus domínios (Tabela 4) levou a um aumento significativo no sinal filogenético detectado nas regiões em dupla-fita, enquanto para as regiões em simples-fita o sinal presente permaneceu bastante similar ao da molécula analisada em sua totalidade (análise do $\phi_{\text{médio}}$ no teste de Steel). As regiões em simples-fita também apresentaram valores maiores para os índices de saturação de substituição do que as regiões em dupla-fita (teste de Xia). Já o sub-particionamento da região COI de acordo com as posições do códon também levaram a um aumento significativo do sinal filogenético na primeira e segunda posição, enquanto na terceira os níveis permaneceram quase inalterados. A terceira posição do códon também apresentou níveis significativos de saturação de substituição ($I_{ss} > I_{ss.c}$ A no teste de Xia), sugerindo um uso limitado dessa região em análises filogenéticas. A combinação desses resultados sugere que o sinal filogenético presente em cada uma das regiões gênicas analisadas é limitado por sua sub-partição menos informativa.

Os testes de congruência topológica mostraram que as quatro regiões gênicas podem ser concatenadas em um mesmo conjunto de dados ($p = 0,1198$ para concatenação das regiões ITS2 e 28S; $p = 0,0644$ para concatenação das regiões ITS2-28S e 16S; e $p = 0,1824$ para concatenação das regiões ITS2-28S-16S COI), enquanto os testes de congruência dos comprimentos de ramo indicaram que as quatro regiões devem ter seus parâmetros estimados e otimizados separadamente ($p < 0,05$ em todas as comparações).

Tabela 4. Resultados dos testes de sinal filogenético. Os resultados do teste de Steel são dados em função da amplitude dos valores de $\phi_{\text{médio}}$ para cada sequência. Os resultados do teste de Xia mostrados são baseados nas simulações com 32 OTUs (10.000 replicatas). Iss = Índice de Saturação de Substituição; Iss.c S/A = ISS crítico se a verdadeira árvore é simétrica (S) ou assimétrica (A). Um nível significativo de saturação é encontrado quando $\text{Iss} > \text{Iss.c}$. ds = regiões em dupla-fita; ss = regiões em simples-fita; 1st, 2nd e 3rd = posições nos códons da região COI.

Partições		Grupo taxonômico	Teste de Steel ($\phi_{\text{médio}}$)	Teste de Xia (32 OTUs)				
				Iss	Iss.c S	P	Iss.c A	P
ITS2	ds + ss	Calyptratae	0,0254 - 0,0342	0,1220	0,8280	0,0000	0,6280	0,0000
		Oestroidea	0,0284 - 0,0415	0,1140	0,8210	0,0000	0,6150	0,0000
		Calliphoridae <i>s.s</i>	0,0544 - 0,0734	0,0948	0,6276	0,0000	0,3610	0,0000
	ds	Calyptratae	0,0438 - 0,0545	0,0950	1,4430	0,0000	1,6830	0,0000
		Oestroidea	0,0514 - 0,0653	0,0910	1,4130	0,0000	1,6320	0,0000
		Calliphoridae <i>s.s.</i>	0,0957 - 0,1089	0,0573	0,5839	0,0000	0,4491	0,0000
	ss	Calyptratae	0,0249 - 0,0334	0,1420	1,0990	0,0000	1,0970	0,0000
		Oestroidea	0,0278 - 0,0399	0,1320	1,0860	0,0000	1,0740	0,0000
		Calliphoridae <i>s.s.</i>	0,0541 - 0,0730	0,1359	0,5839	0,0000	0,4491	0,0000
28S	Calyptratae	0,0329 - 0,0431	0,0420	0,7440	0,0000	0,4350	0,0000	
	Oestroidea	0,0411 - 0,0533	0,0350	0,7450	0,0000	0,4370	0,0000	
	Calliphoridae <i>s.s.</i>	0,0832 - 0,1043	0,0203	0,7570	0,0000	0,4772	0,0000	
COI	1st + 2nd + 3rd	Calyptratae	0,0174 - 0,0283	0,1370	0,7280	0,0000	0,4060	0,0000
		Oestroidea	0,0207 - 0,0340	0,1310	0,7280	0,0000	0,4060	0,0000
		Calliphoridae <i>s.s.</i>	0,0417 - 0,0667	0,1044	0,7439	0,0000	0,4566	0,0000
	1st	Calyptratae	0,0299 - 0,0424	0,0650	0,6830	0,0000	0,3580	0,0000
		Oestroidea	0,0352 - 0,0507	0,0610	0,6830	0,0000	0,3580	0,0000
		Calliphoridae <i>s.s.</i>	0,0584 - 0,0818	0,0425	0,6446	0,0000	0,3664	0,0000
	2nd	Calyptratae	0,0327 - 0,0501	0,0070	0,6830	0,0000	0,3580	0,0000
		Oestroidea	0,0397 - 0,0559	0,0060	0,6830	0,0000	0,3580	0,0000
		Calliphoridae <i>s.s.</i>	0,0316 - 0,0668	0,0025	0,6446	0,0000	0,3664	0,0000
	3rd	Calyptratae	0,0175 - 0,0289	0,4590	0,6830	0,0000	0,3580	0,0002
		Oestroidea	0,0216 - 0,0348	0,4390	0,6830	0,0000	0,3580	0,0030
		Calliphoridae <i>s.s.</i>	0,0438 - 0,0676	0,3615	0,6441	0,0000	0,3662	0,8619
16S	ds + ss	Calyptratae	0,0351 - 0,0505	0,0960	0,6920	0,0000	0,3640	0,0000
		Oestroidea	0,0436 - 0,0601	0,0860	0,6920	0,0000	0,3640	0,0000
		Calliphoridae <i>s.s.</i>	0,0941 - 0,1173	0,0463	0,6961	0,0000	0,4129	0,0000
	ds	Calyptratae	0,0413 - 0,0557	0,0840	0,6880	0,0000	0,3710	0,0000
		Oestroidea	0,0519 - 0,0650	0,0720	0,6880	0,0000	0,3710	0,0000
		Calliphoridae <i>s.s.</i>	0,0929 - 0,1200	0,0391	0,6196	0,0000	0,3614	0,0000
	ss	Calyptratae	0,0378 - 0,0502	0,1100	0,6900	0,0000	0,3750	0,0000
		Oestroidea	0,0478 - 0,0603	0,1020	0,6900	0,0000	0,3750	0,0000
		Calliphoridae <i>s.s.</i>	0,0927 - 0,1146	0,0544	0,6152	0,0000	0,3600	0,0000

Esses resultados, combinados com os resultados dos testes de sinal filogenético e de saturação de substituição, indicam que o particionamento e sub-particionamento dos dados com o uso de diferentes modelos para cada região é a forma mais apropriada de explorar mais profundamente a informação contidas nas moléculas analisadas.

3.3. Análises filogenéticas

As árvores filogenéticas inferidas resultaram na recuperação consistente de várias relações entre e dentro dos grupos de Calyptratae amostrados (Figuras 3 a 6). Um clado monofilético de Oestroidea foi recuperado, inserido dentro de um clado parafilético de Muscoidea, com alto suporte nas análises Bayesianas (PP = 1,00) (Figuras 5 e 6), com suporte moderado nas análises de ML (BS >75, SH = 0,97) (Figura 4) e com baixo suporte na análise de MP (BS = 48 / JK = 67) (Figura 3). Com exceção de Calliphoridae *s.l.*, todas as famílias de Oestroidea foram recuperadas como monofiléticas com altos valores de suporte (MP - BS > 95; JK > 99 / ML - BS > 95; SH > 0,97/ BI - PP > 0,99). A monofilia de Calliphoridae *s.s.* foi recuperada com suporte alto ou moderado nas análises Bayesianas (PP média = 0,92) e de ML (SH = 0,86), embora nesta última os valores de BS tenham sido significativamente menores (Garli = 18 / PhyML = 61).

As relações entre as famílias de Oestroidea foram mais variáveis, sobretudo no que concerne ao posicionamento das duas espécies da família Oestridae, as quais foram recuperadas como (1) grupo irmão dos demais Oestroidea nas análises de MP (BS = 48, JK = 67); (2) como grupo irmão do clado Rhiniinae + Sarcophagidae + Calliphoridae *s.s.* nas análises de ML e em algumas de BI (ML - BS = 44, SH = 0,25; BI - PP média PHASE = 0,90); e (3) como grupo irmão de Rhiniinae na maioria das árvores de BI (PP média = 0,85).

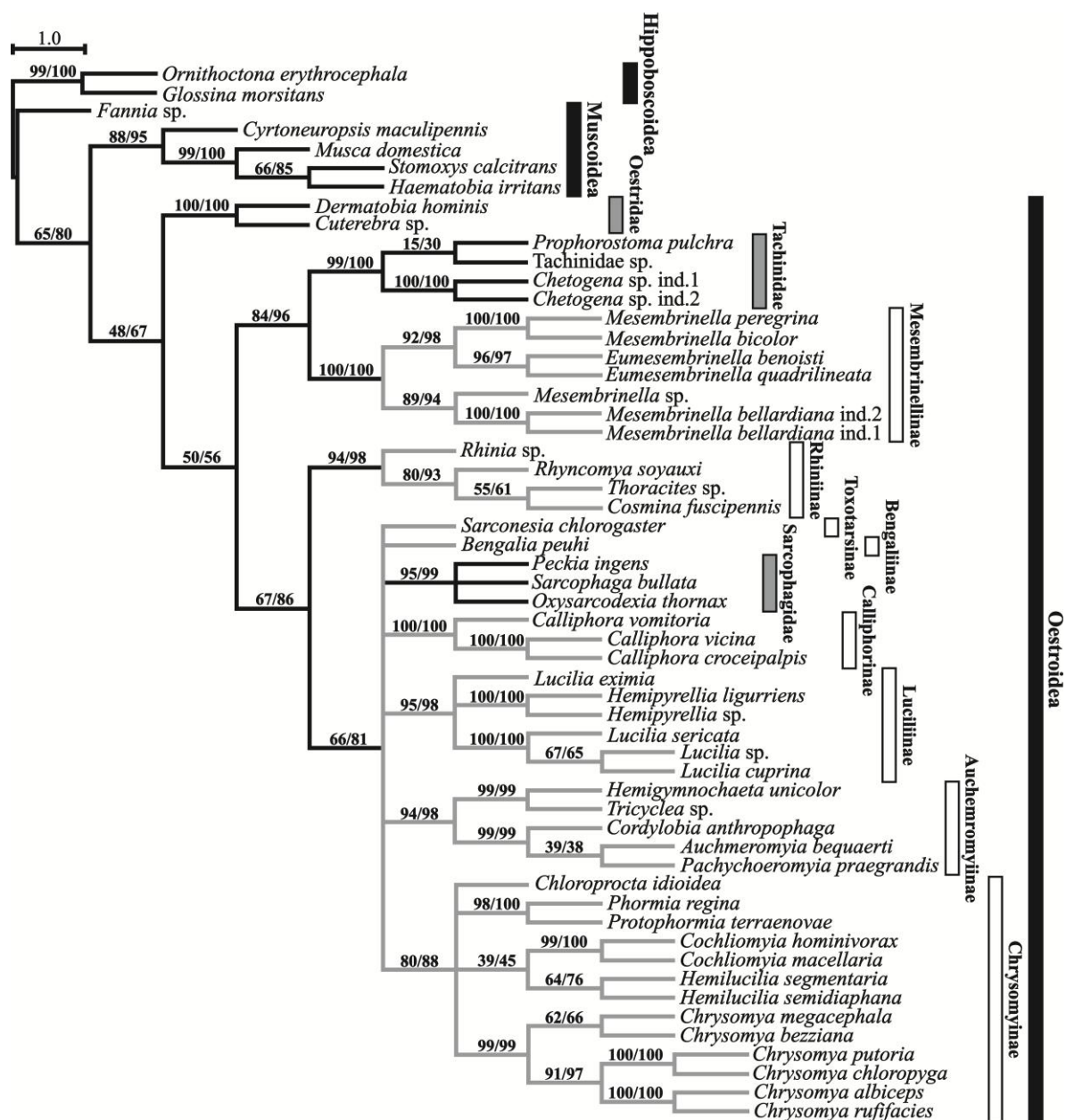


Figura 3. Árvore consenso estrito das 17 árvores igualmente parcimoniosas inferidas pelo programa TnT v1.1. Os valores de suporte de bootstrap (esquerda) e jackknife (direita) são mostrados acima dos respectivos ramos. Os clados formalmente incluídos na família Calliphoridae são mostrados em cinza.

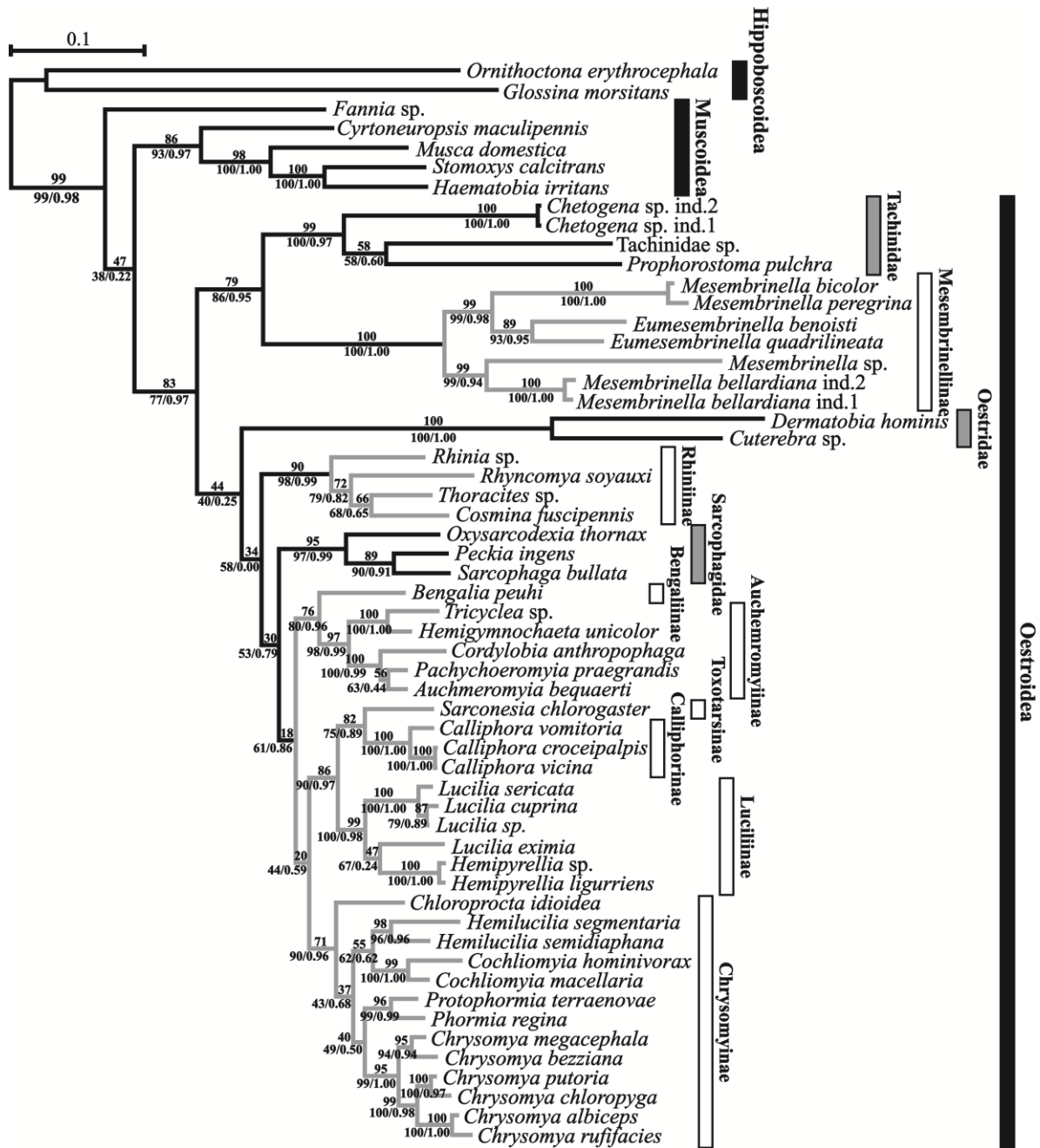


Figura 4. Árvore estimada pelo método de Máxima-verossimilhança usando os programas Garli v1.0 e PhyML 3.0. Os valores de suporte de bootstrap do programa Garli v1.0 são mostrados acima dos respectivos ramos enquanto os valores do programa PhyML 3.0 são mostrados abaixo e a esquerda. Valores de suporte inferidos pelo teste "SH-like" são mostrados abaixo dos ramos a direita. Os clados formalmente incluídos na família Calliphoridae são mostrados em cinza.

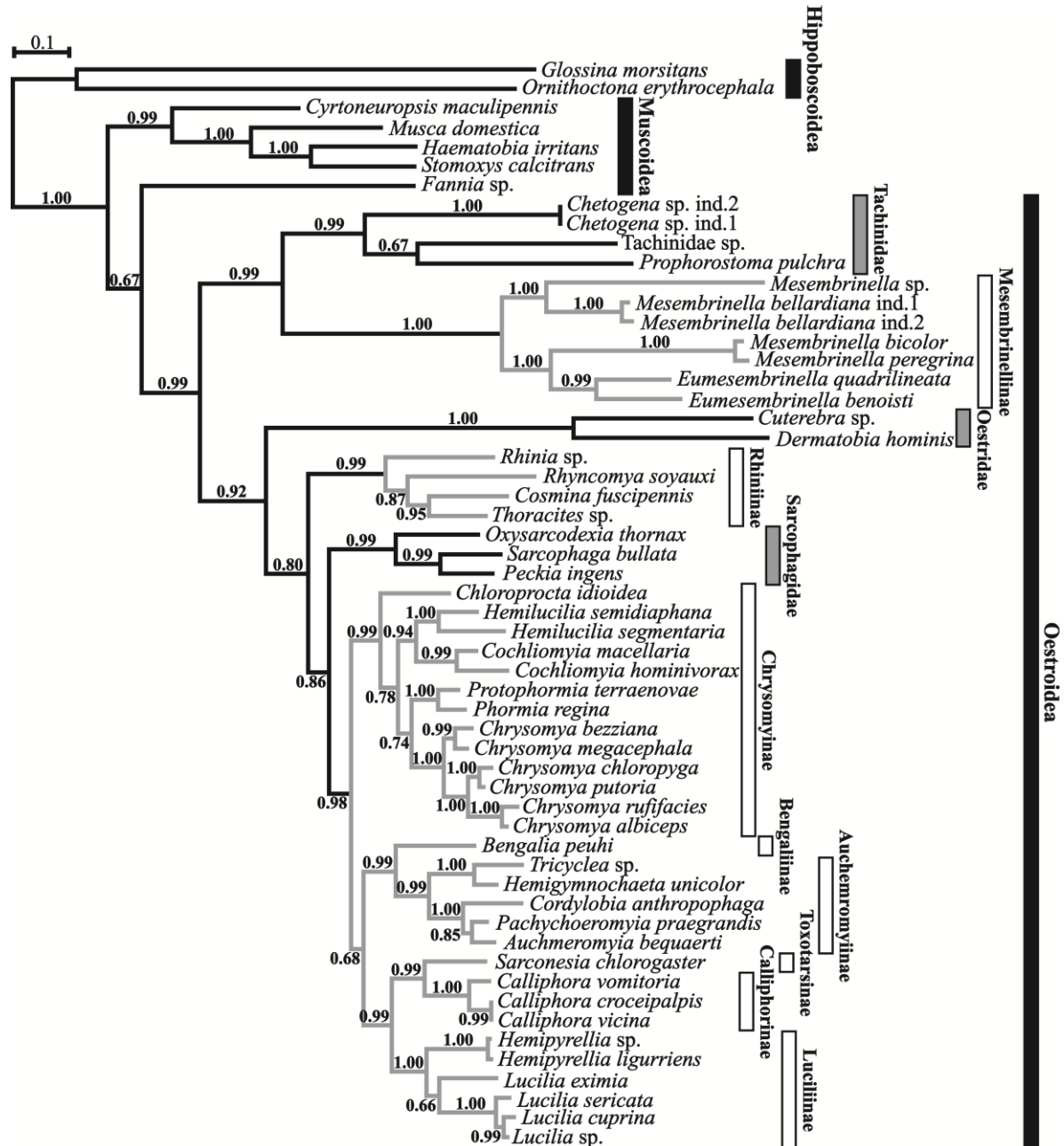


Figura 5. Árvore inferida pelo método Bayesiano usando o programa mcmcphase do pacote PHASE e a combinação de partições e modelos S7A/C (Tabela 2 para maiores detalhes). Os valores de probabilidade *a posteriori* são mostrados sobre os respectivos ramos. Os clados formalmente incluídos na família Calliphoridae são mostrados em cinza.

A família Sarcophagidae foi recuperada com suporte baixo ou moderado como grupo irmão de Calliphoridae *s.s.* nas análises de MP, ML e em todas menos uma das análises de BI (análise S/C - corrida B), na qual ela foi recuperado como grupo irmão de Oestridae + Rhiniinae + Calliphoridae *s.s.*. Das subfamílias de Calliphoridae *s.l.* não incluídas no clado Calliphoridae *s.s.*, Mesembrinellinae foi consistentemente recuperada como grupo irmão de Tachinidae (MP - BS = 84, JK = 96; ML - BS > 79, SH = 0,95; BI - PP > 0,99), enquanto Rhiniinae foi recuperada como grupo irmão de Sarcophagidae + Calliphoridae *s.s.* nas análises de MP, ML e em algumas análises de BI (MP - BS = 67, JK = 86; ML - BS < 58, SH = 0; BI - PP média PHASE = 0,84) e como grupo irmão de Oestridae na maioria das árvores de BI (PP média = 0,85).

Todas as subfamílias de Calliphoridae *s.l.* foram recuperadas como monofiléticas (grupos representados por mais de dois gêneros: MP - BS = >80, JK > 88; ML - BS > 90, SH > 0,96; BI - PP > 0,99). Para as relações entre as subfamílias de Calliphoridae *s.s.*, as análises de ML e BI recuperaram uma relação próxima entre as subfamílias Bengaliinae e Auchmeromyiinae e entre Luciliinae, Calliphorinae e Toxotarsinae, ambas com suporte moderado ou alto. O clado Luciliinae + Calliphorinae + Toxotarsinae foi recuperado como grupo irmão de Chrysomyinae nas árvores de ML e de Bengaliinae + Auchmeromyiinae nas árvores de BI, ambas as relações com baixo suporte.

Em Chrysomyinae, a subfamília de Calliphoridae *s.s.* melhor amostrada nas análises, as relações recuperadas nas análises de ML e BI sugerem uma relação próxima entre o gênero *Chrysomya* e os gêneros de Phormiini *Phormia* + *Protophormia*, com suporte de baixo a moderado, e entre os gêneros *Cochliomyia* e *Hemilucilia*, com suporte baixo nas análises de ML e alto nas Bayesianas. O gênero *Chloroprocta* foi recuperado como grupo irmão dos demais Chrysomyinae nas análises de ML e na maioria das análises

de BI, enquanto em algumas das árvores Bayesianas (MrBayes NS/NC e S/NC - ambas não incorporando informações sobre posição no códon nas análises) ele foi recuperado como grupo irmão do clado composto por *Cochliomyia* + *Hemilucilia*.

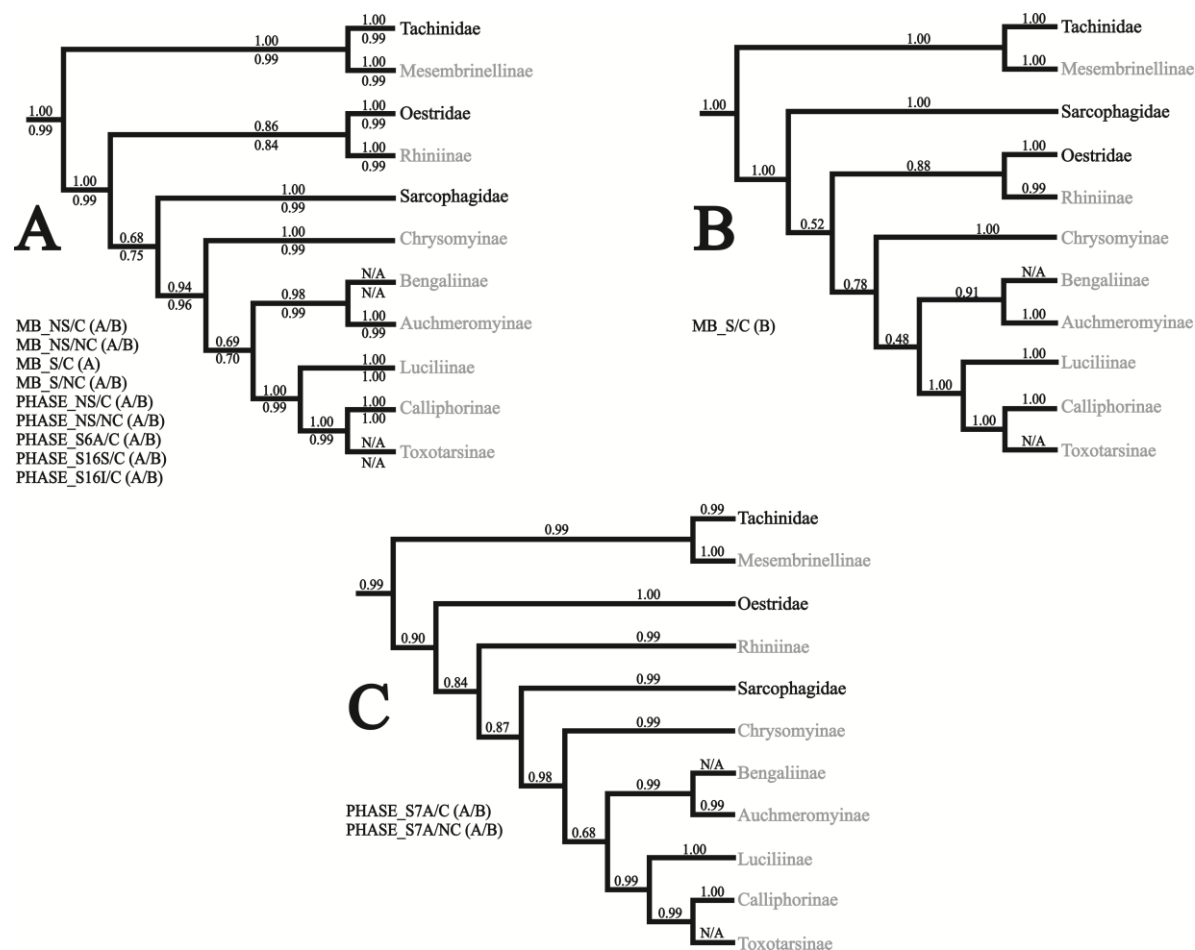


Figura 6. Diferentes hipóteses para as relações interfamiliares em Oestroidea inferidas pelas análises Bayesianas. (A) Topologia recuperada pela maioria das estratégias de particionamento usadas (detalhadas a direita). Os valores médios de probabilidade *a posteriori* estimados pelos programas PHASE e MrBayes são mostrados sob e sobre os ramos, respectivamente. (B) Relações inferidas em uma das duas corridas independentes do programa MrBayes usando a combinação de partições e modelos S/C (Tabela 2 para detalhes). (C) Topologia recuperada usando a combinação de partições e modelos S7A/C (Tabela 2) no programa PHASE. Os nomes em cinza correspondem aos clados formalmente atribuídos à família Calliphoridae.

Em Auchmeromyinae, todas as filogenias inferidas corroboram, com alto suporte, relações próximas entre os gêneros *Hemigymnochaeta* e *Tricyclea* e entre os gêneros *Cordylobia*, *Auchmeromyia* e *Pachychoeromyia*.

3.4. Comparação de modelos por critérios de máxima-verossimilhança e Bayesianos

As comparações de modelos realizadas tanto por critérios de verossimilhança quanto Bayesianos indicaram um aumento significativo das verossimilhanças com o aumento da complexidade das combinações de modelos, sendo que as análises que incluíram partições estruturais e por posição no códon apresentaram os maiores valores de verossimilhança (Tabelas 5, 6A e 6B). Nas análises expandidas realizadas no programa PHASE, as quais consideravam partições estruturais e por posição no códon e que diferiam entre si apenas no modelo de RNA usado, um padrão oposto foi observado, com os valores de verossimilhança decrescendo significativamente com o aumento da complexidade dos modelos. (Tabelas 5 e 6C).

Essas alterações significativas nos valores de verossimilhança foram acompanhadas por poucas modificações nas topologias inferidas, como revelado pela análise das distâncias simétricas entre as árvores estimadas (Tabela 7), as quais não necessariamente estavam diretamente correlacionadas com as mudanças nos valores de verossimilhança. Enquanto nas árvores Bayesianas estimadas com o programa MrBayes as alterações topológicas envolveram diferentes posições para a família Sarcophagidae e para o gênero de Chrysomyinae *Chloroprocta* (Tabela 8A), nas análises iniciais feitas no pacote PHASE a única mudança topológica observada foi a posição relativa entre a família Oestridae e a subfamília de Calliphoridae Rhiniinae (Tabela 8B). Nas análises posteriores realizadas no pacote PHASE, novas mudanças topológicas foram observadas, incluindo diferentes

relações entre as espécies de Tachinidae e a posição relativa de Oestroidea como grupo irmão das famílias de Muscoidea Muscidae ou Fanniidae (Tabela 8C).

Tabela 5. Resultado das análises de comparação de modelos baseado em máxima-verossimilhança. Os valores de AIC, AICc e BIC são mostrados. O ranqueamento mostrado na última coluna tem propósitos comparativos. O número de parâmetros em cada combinação de partições e modelos é mostrado na Tabela 2. A e B referem-se às duplicatas independentes de cada análise.

Método		Max Log L (amostrada)	AIC	AICc	BIC	Rank
MrBayes	NS/C (A)	-26938,9740	54197,9480	54214,0782	54442,0588	3
	NS/C (B)	-26941,7330	54203,4660	54219,5962	54447,5768	
	NS/NC (A)	-27200,9100	54699,8200	54713,7669	54927,1482	4
	NS/NC (B)	-27201,7890	54701,5780	54715,5249	54928,9062	
	S/C (A)	-26569,6470	53551,2940	53578,3854	53865,5867	1
	S/C (B)	-26575,8130	53563,6260	53590,7174	53877,9187	
	S/NC (A)	-26830,5640	54051,1280	54075,3255	54348,6380	2
	S/NC (B)	-26825,4870	54040,9740	54065,1715	54338,4840	
PHASE	NS/C (A)	-26628,0170	53570,0340	53585,5523	53809,5677	3
	NS/C (B)	-26628,5945	53571,1890	53586,7073	53810,7227	
	NS/NC (A)	-27166,7201	54631,4402	54645,3871	54858,7684	4
	NS/NC (B)	-27168,3993	54634,7986	54648,7455	54862,1268	
	S7A/C (A)	-26003,2352	52436,4704	52466,0594	52764,4943	1
	S7A/C (B)	-26002,7359	52435,4718	52465,0608	52763,4957	
	S7A/NC (A)	-26538,5830	53491,1660	53518,5292	53806,9844	2
	S7A/NC (B)	-26536,7807	53487,5614	53514,9246	53803,3798	
	S6A/C (A)	-25434,3465	51268,6930	51294,1844	51573,8315	1
	S6A/C (B)	-25436,8435	51273,6870	51299,1784	51578,8255	
	S7A/C (A)	-26003,2352	52436,4704	52466,0594	52764,4943	2
	S7A/C (B)	-26002,7359	52435,4718	52465,0608	52763,4957	
	S16A/C (A)	-26282,2935	52964,5870	52990,0784	53269,7255	3
	S16A/C (B)	-26287,9219	52975,8438	53001,3352	53280,9823	
	S16I/C (A)	-26291,9917	52989,9834	53016,2683	53299,6990	4
	S16I/C (B)	-26293,6741	52993,3482	53019,6331	53303,0638	

Tabela 6. Resultado das comparações de modelo usando métodos Bayesianos. (A) MrBayes. (B) PHASE (análises iniciais). (C) PHASE (análises expandidas usando diferentes modelos de RNA implementados no programa). Valores significativos nas comparações são mostrados em negrito. Valores positivos indicam suporte para M_1 (modelos nas linhas) sobre M_0 (modelos nas colunas).

A

Log L Marginal (Média harmônica)	Método	M_0							
		NS/C (A)	NS/C (B)	NS/NC (A)	NS/NC (B)	S/C (A)	S/C (B)	S/NC (A)	S/NC (B)
-26975,96863	NS/C (A)	0							
-26975,97567	NS/C (B)	-0,0140	0						
-27232,69321	NS/NC (A)	-513,4491	-513,4350	0					
-27232,40883	NS/NC (B)	-512,8804	-512,8663	0,5687	0				
-26611,45897	S/C (A)	729,0193	729,0334	1242,4684	1241,8997	0			
-26611,42615	S/C (B)	729,0849	729,0990	1242,5341	1241,9653	0,0656	0		
-26868,27385	S/NC (A)	215,3895	215,4036	728,8387	728,2699	-513,6297	-513,6954	0	
-26868,47684	S/NC (B)	214,9835	214,9976	728,4327	727,8639	-514,0357	-514,1013	-0,4059	0

B

Log L Marginal (Média harmônica)	Método	M_0							
		NS/C (A)	NS/C (B)	NS/NC (A)	NS/NC (B)	S7A/C (A)	S7A/C (B)	S7A/NC (A)	S7A/NC (B)
-26662,21019	NS/C (A)	0							
-26661,64489	NS/C (B)	1,1306	0						
-27197,83687	NS/NC (A)	-1071,2533	-1072,3839	0					
-27197,78172	NS/NC (B)	-1071,1430	-1072,2736	0,1103	0				
-26044,03051	S7A/C (A)	1236,3593	1235,2287	2307,6127	2307,5024	0			
-26043,16646	S7A/C (B)	1238,0874	1236,9568	2309,3408	2309,3405	1,7281	0		
-26575,05535	S7A/NC (A)	174,3096	173,1790	1245,5630	1245,4527	-1062,0496	-1063,7777	0	
-26573,22752	S7A/NC (B)	177,9653	176,8347	1249,2187	1249,1084	-1058,3940	-1060,1221	3,6556	0

Tabela 6. (Continuação).**C**

Log L Marginal (Média harmônica)	Método	M ₀							
		S6A/C (A)	S6A/C (B)	S7A/C (A)	S7A/C (B)	S16A/C (A)	S16A/C (B)	S16I/C (A)	S16I/C (B)
-25470,42685	S6A/C (A)	0							
-25474,80717	S6A/C (B)	-8,7606	0						
-26044,03051	S7A/C (A)	-1147,2073	-1138,4466	0					
-26043,16646	S7A/C (B)	-1145,4792	-1136,7185	1,7281	0				
-26320,51977	S16A/C (A)	-1700,1858	-1691,4252	-552,9785	-554,7066	0			
-26324,70656	S16A/C (B)	-1708,5594	-1699,7987	-561,3521	-563,0802	-8,3735	0		
-26333,09095	S16I/C (A)	-1725,3282	-1716,5675	-578,1208	-579,8489	-25,1423	-16,7687	0	
-26333,65605	S16I/C (B)	-1726,4584	-1717,6977	-579,2510	-580,9791	-26,2725	-17,8989	-1,1302	0

Tabela 7. Suporte médio (média aritmética), comprimento total da árvore (TL) e alterações topológicas (distâncias simétricas) para as diferentes estratégias de particionamento utilizadas nas análises Bayesianas. (A) MrBayes. (B) PHASE (análises iniciais). (C) PHASE (análises expandidas). Os valores foram ordenados de acordo com os valores de verossimilhança marginal, do maior para o menor. SD = desvio padrão.

A

Método	Log L Marginal (Média harmônica)	Suporte		TL		Distâncias simétricas							
		Média	SD	Média	SD	S/C (B)	S/C (A)	S/NC (A)	S/NC (B)	NS/C (A)	NS/C (B)	NS/NC (B)	NS/NC (A)
S/C (B)	-26611,42615	0,93622	0,13666	6,29843	0,32293	0							
S/C (A)	-26611,45897	0,93528	0,13965	6,29899	0,31340	2	0						
S/NC (A)	-26868,27385	0,94679	0,12306	7,59315	0,54169	4	2	0					
S/NC (B)	-26868,47684	0,94641	0,12367	7,57710	0,53970	4	2	0	0				
NS/C (A)	-26975,96863	0,94981	0,10746	6,24202	0,30551	2	0	2	2	0			
NS/C (B)	-26975,97567	0,94981	0,10735	6,24223	0,31016	2	0	2	2	0	0		
NS/NC (B)	-27232,40883	0,95528	0,10687	7,70588	0,56779	4	2	0	0	2	2	0	
NS/NC (A)	-27232,69321	0,95528	0,10564	7,69718	0,56275	4	2	0	0	2	2	0	0

Tabela 7. (Continuação).**B**

Método	Log L Marginal (Média harmônica)	Suporte		TL		Distâncias simétricas							
		Média	SD	Média	SD	S7A/C (B)	S7A/C (A)	S7A/NC (B)	S7A/NC (A)	NS/C (B)	NS/C (A)	NS/NC (B)	NS/NC (A)
S7A/C (B)	-26043,16646	0,94358	0,10675	12,88786	0,60814	0							
S7A/C (A)	-26044,03051	0,94716	0,09975	12,89292	0,61602	0	0						
S7A/NC (B)	-26573,22752	0,95188	0,10740	12,75023	0,58508	0	0	0					
S7A/NC (A)	-26575,05535	0,95226	0,10819	12,68646	0,58440	0	0	0	0				
NS/C (B)	-26661,64489	0,94433	0,10208	12,24829	0,59192	2	2	2	2	0			
NS/C (A)	-26662,21019	0,94698	0,10045	12,29612	0,59988	2	2	2	2	0	0		
NS/NC (B)	-27197,78172	0,94867	0,10650	12,06721	0,59154	2	2	2	2	0	0	0	
NS/NC (A)	-27197,83687	0,94886	0,10631	12,11434	0,57262	2	2	2	2	0	0	0	0

C

Método	Log L Marginal (Média harmônica)	Suporte		TL		Distâncias simétricas							
		Média	SD	Média	SD	S6A/C (A)	S6A/C (B)	S7A/C (B)	S7A/C (A)	S16A/C (A)	S16A/C (B)	S16I/C (A)	S16I/C (B)
S6A/C (A)	-25470,42685	0,94169	0,11164	12,87040	0,62955	0							
S6A/C (B)	-25474,80717	0,93943	0,11400	12,91066	0,59739	0	0						
S7A/C (B)	-26043,16646	0,94358	0,10675	12,88786	0,60814	4	4	0					
S7A/C (A)	-26044,03051	0,94716	0,09975	12,89292	0,61602	4	4	0	0				
S16A/C (A)	-26320,51977	0,93811	0,11604	12,90833	0,59932	4	4	4	4	0			
S16A/C (B)	-26324,70656	0,93792	0,11752	12,90462	0,59702	4	4	4	4	0	0		
S16I/C (A)	-26333,09095	0,93867	0,11681	12,92475	0,62246	2	2	2	2	2	2	0	
S16I/C (B)	-26333,65605	0,93547	0,11923	12,90985	0,59485	2	2	2	2	2	2	0	0

O valor de suporte médio geral para as árvores também sofreu poucas alterações com o uso de modelos de maior ou menor complexidade. Nas análises realizadas no programa MrBayes, o suporte médio das árvores mostrou um leve decréscimo à medida que a complexidade dos modelos aumentava e, conseqüentemente, seus valores de verossimilhança marginal (Tabela 7A). Esse comportamento é provavelmente um reflexo do aumento na incerteza em torno da topologia estimada (dado que a própria topologia é considerada um parâmetro nas análises Bayesianas), o que pode ser observado no número crescente de topologias incluídas no intervalo de confiança de 99% à medida que a complexidade do modelo aumenta: (1) NS/NC [A] = 4008, [B] = 3952; (2) NS/C [A] = 4338, [B] = 4412; (3) S/NC [A] = 4378, [B] 4479; (4) S/C [A] = 4773, [B] 4735. Já nas análises conduzidas no pacote de programas PHASE, as alterações nos valores de suporte médio geral não foram diretamente correlacionadas com o aumento de complexidade das combinações de modelos e partições, uma vez que modelos com complexidade intermediária e com valores de verossimilhança intermediários apresentaram os maiores valores de suporte médio (combinações S7A/NC nas análises iniciais do PHASE - Tabela 7B - e S7A/C nas análises expandidas - Tabela 7C).

Apesar das pequenas alterações observadas nos valores médios de suporte das probabilidades *a posteriori*, o suporte estimado para alguns clados sofreu alterações mais significativas com o aumento ou diminuição da complexidade dos modelos (Tabela 8), especialmente nas análises feitas no programa MrBayes. Enquanto para alguns clados os valores de suporte diminuíram com o aumento da complexidade do modelo, e.g., o suporte para o clado monofilético Calliphoridae *s.s.* nas análises do MrBayes (Tabela 8A) e para o status de grupo irmão entre os clados Oestroidea + *Fannia* sp. nas análises em ambos os programas (Tabelas 8A e 8B), seguindo assim o mesmo comportamento observado para o

Tabela 8. Valores de suporte (probabilidade *a posteriori* - PP) dos clados que sofreram alterações em suporte ou relações inferidas entre as diferentes estratégias de particionamento usadas nas análises Bayesianas. Apenas relações com menos de 0,90 de suporte em pelo menos uma das estratégias são mostradas. (A) MrBayes. (B) PHASE (análises iniciais). (C) PHASE (análises extendidas). Os valores de suporte mostrados correspondem aos nós representados por "+" em cada uma das relações analisadas e foram ponderados (média aritmética) entre as duas replicatas independentes feitas para cada análise, com exceção da análise S/C no programa MrBayes, a qual resultou em topologias diferentes. Os valores de suporte são mostrados em tons de cinza do maior (mais escuro) para o menor (mais claro). As análises Bayesianas foram organizadas segundo seus valores de verossimilhança marginal, do menor (esquerda) para o maior (direita).

A

Alterações	Relações	PP				
		NS/NC	NS/C	S/NC	S/C(A)	S/C(B)
Suporte apenas	((<i>Chrysomya</i>) + (<i>Phormia</i> , <i>Protophormia</i>))	0,960	0,820	0,960	0,820	0,830
	(<i>L. eximia</i> + (<i>L. sericata</i> , (<i>L. cuprina</i> , <i>Lucilia</i> sp.)))	0,915	0,665	0,920	0,740	0,730
	((<i>Luciliinae</i> , (<i>Toxotarsinae</i> , <i>Calliphorinae</i>)) + (<i>Auchmeromyiinae</i> , <i>Bengaliinae</i>))	0,805	0,605	0,770	0,470	0,480
	Calliphoridae s.s.	0,990	0,960	0,940	0,790	0,780
	Rhiniinae + Oestridae	0,870	0,920	0,785	0,870	0,880
	Tachinidae sp. + <i>P. pulchra</i>	0,775	0,715	0,575	0,540	0,540
	Oestroidea + <i>Fannia</i> sp.	0,760	0,635	0,610	0,500	0,510
Topologia e suporte	(<i>Chloroprocta</i> , ((<i>Cochliomyia</i> , <i>Hemilucilia</i>) + (<i>Chrysomya</i> , (<i>Phormia</i> , <i>Protophormia</i>))))	-	0,880	-	0,850	0,850
	((<i>Chloroprocta</i> + (<i>Cochliomyia</i> , <i>Hemilucilia</i>)), (<i>Chrysomya</i> , (<i>Phormia</i> , <i>Protophormia</i>))))	0,395	-	0,430	-	-
	((<i>Calliphoridae</i> s.s. + <i>Sarcophagidae</i>), (<i>Rhiniinae</i> , <i>Oestridae</i>))	0,780	0,695	0,660	0,490	-
	(<i>Sarcophagidae</i> , (<i>Calliphoridae</i> s.s. + (<i>Rhiniinae</i> , <i>Oestridae</i>))))	-	-	-	-	0,520

Tabela 8. (Continuação).**B**

Alterações	Relações	PP			
		NS/NC	NS/C	S7A/NC	S7A/C
Suporte apenas	((<i>Chrysomya</i>) + (<i>Phormia</i> , <i>Protophormia</i>))	0,940	0,755	0,930	0,740
	((<i>Chrysomya</i> , (<i>Phormia</i> , <i>Protophormia</i>)) + (<i>Cochliomyia</i> , <i>Hemilucilia</i>))	0,415	0,800	0,370	0,780
	(<i>L. eximia</i> + (<i>L. sericata</i> , (<i>L. cuprina</i> , <i>Lucilia</i> sp.)))	0,915	0,800	0,860	0,660
	((<i>Luciliinae</i> , (<i>Toxotarsinae</i> , <i>Calliphorinae</i>)) + (<i>Auchmeromyinae</i> , <i>Bengaliinae</i>))	0,730	0,765	0,700	0,655
	Calliphoridae s.s. + Sarcophagidae	0,830	0,790	0,895	0,850
	Tachinidae sp. + <i>P. pulchra</i>	0,710	0,650	0,730	0,670
	Oestroidea + <i>Fannia</i> sp.	0,750	0,740	0,735	0,695
Topologia e suporte	((<i>Oestridae</i> + <i>Rhiniinae</i>) + (<i>Calliphoridae</i> s.s., <i>Sarcophagidae</i>))	0,85/0,99	0,90/0,99	- / -	- / -
	(<i>Oestridae</i> + (<i>Rhiniinae</i> + (<i>Calliphoridae</i> s.s., <i>Sarcophagidae</i>)))	- / -	- / -	0,88/0,93	0,92/0,75

C

Alterações	Relações	PP			
		S16I/C	S16A/C	S7A/C	S6A/C
Suporte apenas	(<i>L. eximia</i> + (<i>L. sericata</i> , (<i>L. cuprina</i> , <i>Lucilia</i> sp.)))	0,755	0,665	0,640	0,670
	((<i>Luciliinae</i> , (<i>Toxotarsinae</i> , <i>Calliphorinae</i>)) + (<i>Auchmeromyinae</i> , <i>Bengaliinae</i>))	0,690	0,665	0,655	0,645
	Calliphoridae s.s. + Sarcophagidae	0,685	0,760	0,855	0,660
Topologia e suporte	((<i>Oestridae</i> + <i>Rhiniinae</i>) + (<i>Calliphoridae</i> s.s., <i>Sarcophagidae</i>))	0,80/0,99	0,73/0,99	- / -	0,91/0,99
	(<i>Oestridae</i> + (<i>Rhiniinae</i> + (<i>Calliphoridae</i> s.s., <i>Sarcophagidae</i>)))	- / -	- / -	0,92/0,75	- / -
	(<i>Chetogena</i> , (<i>Tachinidae</i> sp. + <i>Prophorostoma</i>))	0,540	0,595	0,670	-
	(<i>Prophorostoma</i> , (<i>Tachinidae</i> sp. + <i>Chetogena</i>))	-	-	-	0,520
	Oestroidea + <i>Fannia</i> sp.	0,565	-	0,695	0,710
	Oestroidea + Muscidae	-	0,540	-	-

valor de suporte médio geral da árvore, outros clados parecem ter sido mais sensíveis à adição de sub-partições para diferentes posições do códon. Neste contexto, o valor de suporte para alguns clados foi maior quando as diferentes posições do códon não foram permitidas de evoluir separadamente sob diferentes modelos [e.g., (*Chrysomya* + (*Phormia*, *Protophormia*)) e (*L. eximia* + (*L. sericata*, *L. cuprina*, *Lucilia* sp.))] enquanto para outros o suporte foi maior quando sub-partições individuais para cada posição foram consideradas [e.g., ((*Chrysomya*, (*Phormia*, *Protophormia*)) + (*Cochliomyia*, *Hemilucilia*))].

Permitir que as três posições do códon evoluam independentemente também teve um efeito mais significativo sobre os comprimentos de ramos e, conseqüentemente, no comprimento total das árvores, do que o aumento/diminuição da complexidade dos modelos por si só, pelo menos nas análises no programa MrBayes (Tabela 7A). Nessas análises, aquelas que consideravam diferentes partições para cada posição do códon individualmente resultaram em árvores mais curtas, com a única alteração topológica diretamente relacionada a esse fato sendo o posicionamento de *C. idioidea* como grupo irmão dos demais Chrysomyinae (nas análises NS/C e S/C). Nas análises que não consideram as posições do códon separadamente (NS/NC e S/NC), *C. idioidea* foi recuperado como grupo irmão do clado *Cochliomyia* + *Hemilucilia*, sendo que essa relação teve suporte menor que a anterior. Nas análises conduzidas no programa PHASE, embora alterações discretas tenham sido observadas no comprimento total das árvores, elas parecem ser mais diretamente influenciadas pela complexidade geral do modelo, com modelos mais complexos resultando em árvores ligeiramente mais longas.

3.5. Análises de mapeamento de verossimilhança

Os mapas de verossimilhança construídos (Figura 7) mostraram que, apesar do conjunto de dados de sequências concatenadas final possuir sinal filogenético suficiente

para resolver relações evolutivas nos três níveis hierárquicos considerados (5% ou menos de quartetos completamente não resolvidos) (Figuras 7A, B e E), ainda há um conflito no

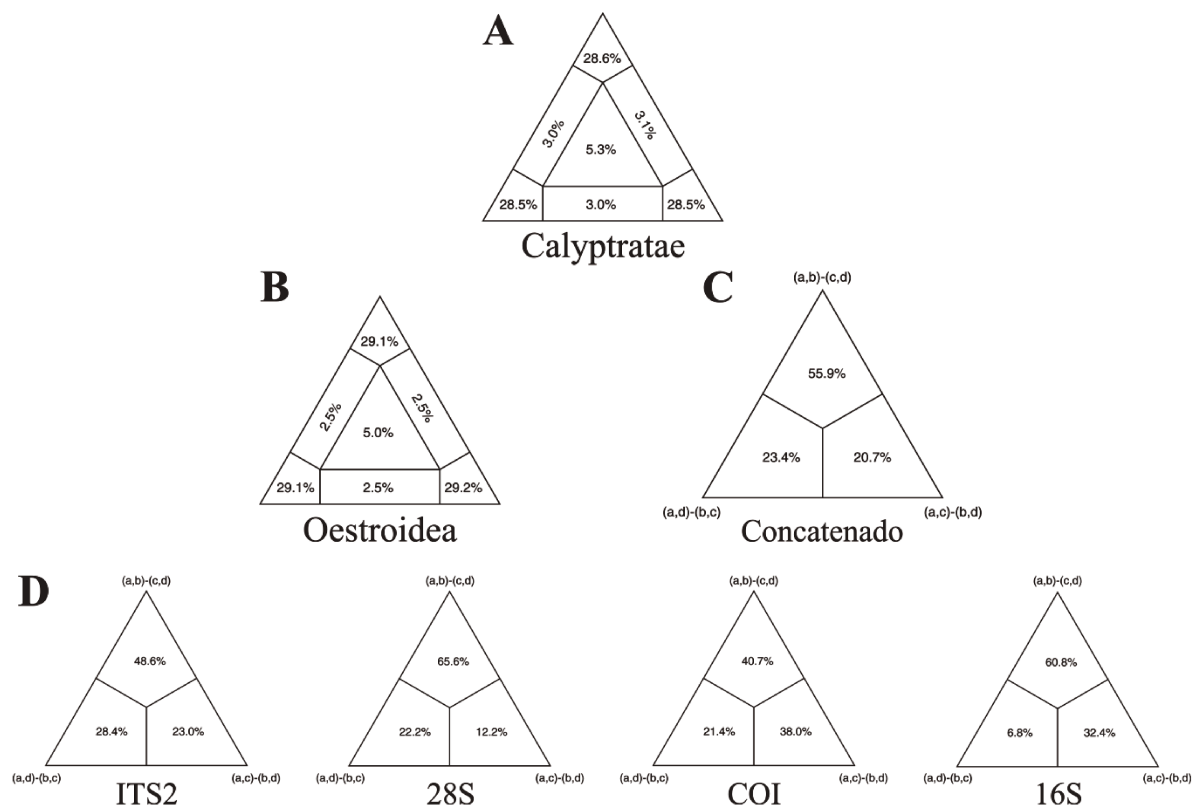


Figura 7. Resultado das análises de mapeamento de verossimilhança. (A), (B) e (E) poder de resolução do conjunto de seqüências concatenado final para os grupos Calypttratae, Oestroidea e Calliphoridae s.s., respectivamente, como estimado analisando-se todos os possíveis quartetos de espécies presente no conjunto total. O triângulo central indica a porcentagem de quartetos completamente não resolvidos enquanto os vértices apresentam a porcentagem de clados totalmente resolvidos. (C) Porcentagem dos quartetos analisados favorecendo cada uma das três possíveis relações entre os grupos previamente definidos: a = Calliphoridae s.s. + Sarcophagidae; b = Rhiniinae; c = Oestridae; d = Tachinidae + Mesembrinellinae, considerando nas análises as quatro regiões gênicas amostradas concatenadas. (D) A mesma análise de (C), mas usando-se cada uma das 4 regiões gênicas separadamente. (F) Porcentagem dos quartetos analisados favorecendo cada uma das três possíveis relações entre os grupos previamente definidos: a = Chrysomyinae; b = Calliphorinae + Luciliinae + Toxotarsinae; c = Auchmeromyinae + Bengaliinae; d = Sarcophagidae. (G) A mesma análise de (F), mas usando-se cada uma das 4 regiões gênicas separadamente.

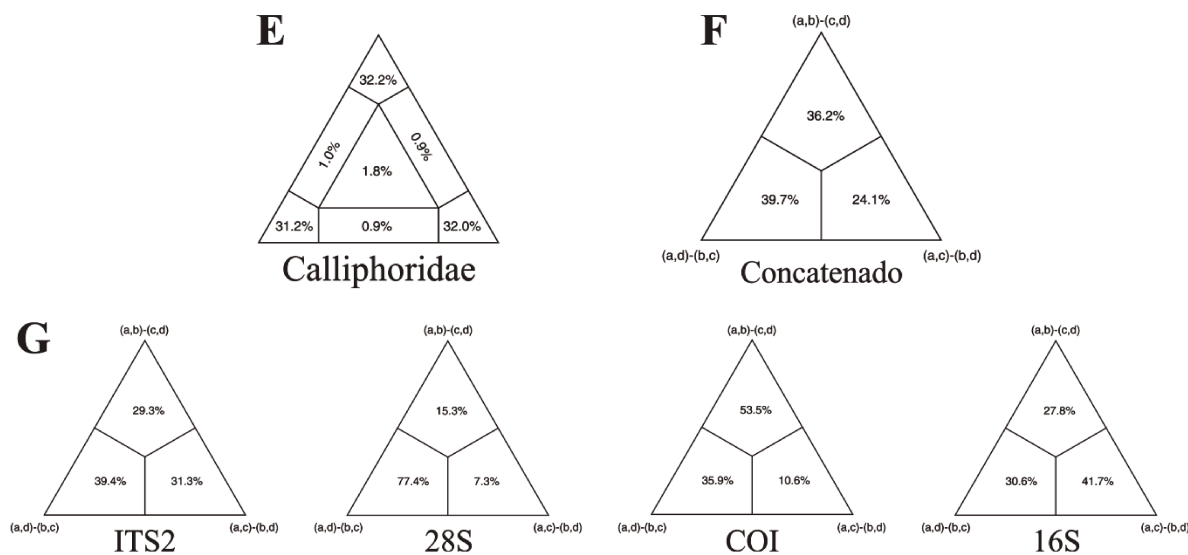


Figura 7. (Continuação).

sinhal considerando as posições relativas dos clados Oestridae, Rhiniinae, Calliphoridae *s.s.* + Sarcophagidae e Tachinidae + Mesembrinellinae (Figura 7C) e dos clados Luciliinae + Toxotarsinae + Calliphorinae, Chrysomyinae, Bengaliinae + Auchmeromyinae e Sarcophagidae (Figura 7F). No segundo caso, o conflito observado pode ser parcialmente explicado por uma incongruência na informação contida nos quatro marcadores moleculares analisados (Figura 7G), embora em ambos os casos o conflito é mais apropriadamente explicado pela falta de resolução do conjunto de sequências para resolver essas duas relações em particular.

4. Discussão

4.1. Análises filogenéticas

A monofilia de Oestroidea foi confirmada, como previamente apontada por Griffiths (1972), Hennig (1976), McAlpine (1989), Pape (1992) e Rognes (1997), bem como sua posição dentro da superfamília Muscoidea, não monofilética, como sugerido recentemente por Kutty et al. (2010). Com exceção de Calliphoridae, todas as famílias de Oestroidea

foram recuperadas como monofiléticas, apesar do tamanho amostral reduzido para a maioria desses grupos não permitir maiores assertivas e conclusões.

Com relação às relações inter-familiares em Oestroidea, a maioria das topologias inferidas propõem novas hipóteses para as relações evolutivas dentro do grupo, apesar de algumas das relações recuperadas já terem sido encontradas em análises previamente publicadas. Além disso, devido ao fato de muitos estudos disponíveis até o momento envolvendo análises genético-evolutivos em grupos de Oestroidea terem diferenças sensíveis em termos de amostragem de táxons e de número e natureza de marcadores utilizados (especialmente quando se compara análises com caracteres morfológicos versus moleculares), alguns dos resultados aqui obtidos são difíceis de serem discutidos e interpretados baseado no conhecimento filogenético atual disponível para o grupo.

As árvores recuperadas pelos métodos de ML e BI sugerem a existência de uma divisão basal entre as famílias de Oestroidea em dois clados, um deles composto pela família Tachinidae e pela subfamília de Calliphoridae Mesembrinellinae, enquanto o outro é composto pelas famílias Oestridae, Sarcophagidae e Calliphoridae s.s. e pela subfamília de Calliphoridae Rhiniinae. Já a análise de MP recuperou relações bastante similares as de ML e BI, com exceção da posição da família Oestridae como grupo irmão dos demais Oestroidea e da politomia recuperada entre a família Sarcophagidae e as subfamílias de Calliphoridae s.s., sugerindo uma relação muito próxima entre esses dois grupos.

A relação próxima entre Tachinidae e Mesembrinellinae, como recuperada aqui e também sugerida nas análises moleculares de Kutty et al. (2010), não corrobora as relações filogenéticas inferidas usando dados morfológicos previamente publicadas para o grupo, sendo que nessas filogenias a família Tachinidae em geral foi recuperada como: (1) grupo irmão de Rhinophoridae (família sem representantes nesse trabalho) e próxima a Oestridae

(McAlpine 1989); (2) grupo irmão de Sarcophagidae, provavelmente próxima de Rhiniinae (Rognes 1997); (3) grupo irmão de Sarcophagidae ou Oestridae (Pape 1992 - o clado Tachinidae + Oestridae foi recuperado apenas na análise que desconsiderava o caráter 32 - Processo fálico dorsolateral, configuração apical - definido no estudo).

O posicionamento de Mesembrinellinae como um grupo separado de Calliphoridae não é corroborado pela maioria dos estudos baseados em morfologia já realizados até o momento, embora alguns autores já tenham reconhecido o grupo como bastante aberrante entre os califorídeos (Hall 1948). Guimarães (1977) argumentou sobre a possibilidade do grupo ser considerado em uma família própria (Mesembrinellidae), baseando-se, principalmente, na presença de ovipositor não-telescópico, no sub-escutelo desenvolvido e na reprodução macro (ou uni) larvípara existente no grupo (Toma & Carvalho 1995). Esta última característica, em especial, é compartilhada com as subfamílias Ameninae e Phumosiinae, as quais provavelmente formam um clado monofilético com Mesembrinellinae (Rognes 1997). Contudo, Rognes (1986) argumentou em suas análises que Mesembrinellinae possuía todas as apomorfias de Calliphoridae usadas em suas análises e, sendo assim, não havia evidências suficientes para propor uma nova família.

A medida que novos dados moleculares se tornam disponíveis, as recém inferidas hipóteses de relações filogenéticas para o grupo parecem corroborar a proposição original de Guimarães (1977), embora em uma posição diferente da inicialmente sugerida por ele (como grupo irmão de Calliphoridae). Estudos adicionais, incluindo uma amostragem maior dos gêneros da subfamília e uma análise combinada de caracteres morfológicos e moleculares, são necessários antes de conclusões serem tomadas com relação a esse assunto, embora a necessidade de uma revisão sistemática neste grupo já seja recomendada.

O correto posicionamento de Oestridae na filogenia de Oestroidea pode ser um problema particularmente difícil, uma vez que o grupo parece ter passado por um processo de diversificação e diferenciação bastante acentuado, incluindo alterações tanto morfológicas quanto moleculares, o qual levou a um hábito parasitário extremamente especializado e que pode tornar qualquer comparação entre caracteres neste grupo e em grupos próximos bastante complicada. Nas topologias inferidas neste estudo, a família Oestridae foi a que apresentou o maior número de posicionamentos possíveis quando diferentes métodos de inferência foram utilizados. Seu status como grupo irmão dos demais Oestroidea, como recuperado nas análises de MP, foi também encontrado por Kutty et al. (2010) em suas análises moleculares com os métodos de ML e BI. As posições alternativas para este grupo recuperadas neste trabalho (grupo irmão de Rhiniinae e de Rhiniinae + Sarcophagidae + Calliphoridae *s.s.*) foram consistentemente diferentes das previamente propostas: (1) McAlpine 1989 - grupo irmão de Tachinidae + Rhinophoridae; (2) Pape 1992 - grupo irmão de Calliphoridae; (3) Rognes (1997) e Pape & Arnaud (2001) - entre as subfamílias de Calliphoridae; (4) Kutty et al. (2010) - dentro de um clado composto por Mesembrinellinae + Tachinidae.

O recentemente proposto status de família para o grupo Rhiniinae (Rhiniidae) foi confirmado aqui, embora ambas as possíveis relações com outros clados de Oestroidea inferidos aqui (grupo irmão de Oestridae ou de Sarcophagidae + Calliphoridae *s.s.*) sejam significativamente diferentes das atualmente propostas (Pape 1992 - dentro de Calliphoridae, próxima das subfamílias Toxotarsinae e Chrysomyinae; Rognes 1997 - grupo irmão de Sarcophagidae + Tachinidae; Pape & Arnaud 2001 - grupo irmão de Rhinophoridae; Kutty et al. 2010 - grupo irmão dos demais Oestroidea excluindo Sarcophagidae e Mystacinobiidae).

A relação de proximidade entre as famílias Sarcophagidae e Calliphoridae s.s., embora não amplamente aceita, foi proposta previamente por McAlpine (1989), com ambas as famílias incluídas em um clado contendo também a família monotípica Mystacinobiidae (não amostrada neste estudo). A maioria dos estudos publicados previamente sugerem para Sarcophagidae uma relação mais próxima com a família Tachinidae (Pape 1992; Rognes 1997, Tachi & Shima 2010) ou Mystacinobiidae, neste caso ambas formando um clado que é grupo irmão dos demais Oestroidea (Kutty et al. 2010). Para a família Calliphoridae, a maioria das relações filogenéticas previamente publicadas para o grupo sugerem sua proximidade evolutiva com as famílias Oestridae e Rhinophoridae (Tsochorsnig 1985; Pape 1992; Rognes 1997) ou uma posição basal como grupo irmão dos demais Oestroidea (Griffiths 1972), embora em todos estes trabalhos a composição considerada da família tenha sido bastante diversa.

O clado monofilético de Calliphoridae encontrado nas análises de ML e BI é composto de três clados bem definidos e consistentes: (1) Chrysomyinae; (2) Bengaliinae + Auchmeromyinae; e (3) Luciliinae + Toxotarsinae + Calliphorinae. A relação geral entre esses três clados é, no entanto, conflitante, sendo que a relação (1,(2,3)) é favorecida por classificações morfológicas mais tradicionais para a família Calliphoridae (nas quais as subfamílias Auchmeromyinae e Bengaliinae são incluídas dentro da subfamília Calliphorinae, definida neste caso em um sentido amplo - Hennig 1973, Shewel 1987), enquanto algumas análises de caracteres morfológicos na família corroboram a relação (2,(1,3)) (incluindo ou não a família Oestridae no clado de Calliphoridae - Pape 1992; Rognes 1997; Pape & Arnaud 2001) e as análises moleculares de Kutty et al. (2010) recuperando a relação (3,(1,2)) (de fato, a única espécie de Bengaliinae amostrada foi recuperada dentro de um clado parafilético de Chrysomyinae). Nas análises conduzidas

aqui, enquanto as análises de BI sugerem a relação (1,(2,3)), as análises de ML sugerem (2,(1,3)), ambas com baixos valores de suporte. Como mostrado nas análises de mapeamento de verossimilhança, embora o conjunto de sequências concatenadas forneça um leve suporte para uma relação mais próxima entre os clados (2) e (3) (Figura 7F), a análise individual das regiões gênicas revelaram a existência de um conflito de sinal considerável e poucas conclusões podem ser feitas neste sentido.

A monofilia do clado Auchmeromyinae + Bengaliinae é bem corroborada por estudos prévios (Rognes 1997), assim também como as relações entre os gêneros de Auchmeromyinae recuperadas neste estudo (*Cordylobia* + *Auchmeromyia* + *Pachychoeromyia* e *Tricyclea* + *Hemigymnochaeta* - Zumpt 1956).

A relação de proximidade entre as subfamílias Calliphorinae e Luciliinae é também bastante corroborada por estudos prévios (Kutty et al. 2010), sendo que em algumas classificações ambas as subfamílias são incluídas em uma única subfamília, Calliphorinae, definida em sentido amplo (Hennig 1973; Shewell 1987; Kurahashi 1989). A inclusão de Toxotarsinae neste clado parece ser apenas corroborada por estudos moleculares na família (Kutty et al. 2010), dado que a maioria dos estudos morfológicos conduzidos em Calliphoridae sugerem uma maior proximidade entre as subfamílias Toxotarsinae e Chrysomyinae (Boyes & Shewell 1983; Pape 1992 - incluindo Rhiniinae; Rognes, 1997).

O status monofilético de Chrysomyinae foi recuperado, como recentemente proposto e confirmado por Singh & Wells (2011) usando marcadores moleculares, apesar das análises moleculares de Kutty et al. (2010) terem recuperado um grupo de Chrysomyinae não-monofilético devido à inclusão de *B. peuhi* neste clado. As relações intergenéricas estão em sua maioria de acordo com as propostas por Singh & Wells (2011), sendo que em ambas não foram encontradas evidências corroborando a atualmente aceita

classificação em tribos dessa subfamília (Chrysomyini e Phormiini - Hall 1948), dado que o gênero *Chrysomyia* foi recuperado mais próximo dos gêneros de Phormiini do que dos demais gêneros de Chrysomyini (como inicialmente sugerido por Rognes 1991). A única relação conflitante encontrada foi a posição de *C. idioidea* como grupo irmão dos demais Chrysomyinae, uma vez que Singh & Wells (2011) recuperaram o gênero *Chloroprocta* no mesmo clado dos demais gêneros de Chrysomyini (excluindo *Chrysomya*). De fato, o gênero *Chloroprocta* foi recuperado como grupo irmão do clado composto pelos gêneros *Hemilucilia* e *Cochliomyia*, de forma semelhante ao recuperado por Singh & Wells (2011), apenas nas análises Bayesianas conduzidas no programa MrBayes nas quais as três posições do códon não foram modeladas independentemente (sendo que nas análises de Singh & Wells 2011 as três posições do códon também não foram modeladas independentemente nas duas regiões gênicas consideradas, COI e CPS), embora essa relação seja consideravelmente menos suportada do que a posição de *Chloroprocta* como grupo irmão dos demais Chrysomyinae (PP < 0,50 - Tabela 8A).

4.2. Comparação de modelos por critérios de máxima-verossimilhança e Bayesianos

Como mostrado nas análises de sinal filogenético e saturação de substituição (Tabela 4), a informação filogenética contida em cada um dos marcadores moleculares analisados parece ser limitada pela região menos informativa presente na molécula, ao menos quando a evolução da mesma é modelada em conjunto, dado que a adição de sub-partições às regiões gênicas (seja por conformação estrutural ou posição no códon) e sua modelagem independente resultou em um aumento significativo do sinal filogenético em todas as sub-partições com exceção da menos informativa, na qual o nível de sinal permaneceu inalterado. Sendo assim, a consideração de sub-partições nas partições gênicas

amostradas e a modelagem independente de cada uma delas, com o modelo que melhor se adéqua a mesma, resultou em um melhor ajuste geral da combinação de partições e modelos ao conjunto de dados de sequência amostrado, o que explica o aumento significativo tanto da verossimilhança máxima quanto da marginal quando do uso de modelos mais complexos em sua modelagem (Tabelas 5 e 6).

Contudo, com o aumento da complexidade dos modelos utilizados, aumenta também a incerteza acerca dos parâmetros estimados na análise de inferência, incluindo a própria topologia, a qual é levada em consideração nas análises Bayesianas. Neste contexto, o aumento no número de topologias consideradas no intervalo de confiança estipulado pelo programa MrBayes com o aumento da complexidade do modelo utilizado é esperado (Nylander et al. 2004). Com mais árvores incluídas e consideradas no intervalo de confiança, maior a possibilidade de topologias conflitantes serem consideradas e, assim, árvores inferidas sob combinações de partições e modelos mais complexas resultam, em geral, em árvore com clados menos suportados e muitas vezes mais instáveis. Isto pode explicar o decréscimo nos valores de suporte de probabilidade *a posteriori* médio geral das árvores com o aumento da complexidade dos modelos (e consequentemente, neste caso, dos valores de verossimilhança marginal) nas análises Bayesianas (Tabela 7).

Inesperado, contudo, foi o fato de que nas análises realizadas no programa MrBayes o número de árvores incluídas no intervalo de confiança aumentou proporcionalmente ao número de partições utilizadas nas análises e não ao número de parâmetros (como mostrado por Nylander et al. 2004). Nessas análises, as combinações de partições e modelos de complexidade intermediária (S/NC e NS/C) possuem o mesmo número de partições e números similares de árvores contidas no intervalo de confiança considerado, mesmo que o

método S/NC seja consideravelmente mais rico em parâmetros (86 parâmetros livres) do que o método NS/C (51 parâmetros livre).

No caso de modelos e combinações de modelos muito ricas em parâmetros, a incerteza acerca da topologia inferida pode resultar em árvores com clados bastante instáveis. Neste contexto, duas corridas independentes realizadas com o mesmo conjunto de sequências e com as mesmas condições de análise no mesmo programa (contudo, com diferentes pontos de início - "random seeds") podem gerar árvores de topologia bastante diferentes mesmo que seus valores de verossimilhança sejam bem similares (como no método de análise S/C no programa MrBayes).

Com relação à consideração de informações estruturais e / ou de posição do códon nas análises de inferência filogenética, ambas parecem ter pequenos efeitos específicos em termos de mudanças nas topologias inferidas, comprimento total da árvore e suporte médio geral de PP nas árvores recuperadas, ao menos quando comparados os efeitos gerais resultantes do aumento ou diminuição da complexidade das combinações de modelos utilizados. Exceções podem ser observadas na alteração dos valores de suporte de PP em alguns clados específicos e nas modificações topológicas associadas com as diferentes posições possíveis ocupadas pela família Oestridae (quando da inclusão de partições estruturais) e pela espécie *C. idioidea* (quando da inclusão de partições por posição no códon), com a última incluindo alterações também nos comprimentos de ramos estimados. Esses pequenos efeitos específicos observados contradizem, de certa forma, a crença comum de que a adoção de estratégias de particionamento dos dados que mais realisticamente retratam e modelam a evolução de diferentes regiões gênicas teriam um efeito significativo na melhora do processo de inferência filogenética, resultando em

árvores estimadas mais próximas da provável árvore real (ou mais plausível de ser observada, como sugerido pelos valores de verossimilhança).

Contudo, como mostrado recentemente por Letsch & Kjer (2011) em suas análises dos efeitos do uso de diferentes modelos combinados de DNA/RNA em análises de inferência filogenética de grandes grupos de Metazoa utilizando sequências de diferentes rDNA, na ausência de níveis significativos de saturação no conjunto de sequências (especialmente nas regiões de simples-fita), o uso de modelos mais simples de DNA ou combinados de modelos de DNA/RNA em geral resultam em árvores bastante similares, potencialmente corretas. Contudo, mesmo nesses casos, a utilização de modelos combinados de DNA/RNA ainda seria preferível por, em geral, fornecerem estimativas mais realistas de suporte para os clados na árvore estimada. É possível que algumas dessas conclusões se apliquem também ao uso de partições independentes para as posições no códon.

Como também mostrado por Letsch & Kjer (2011), nos casos em que o conjunto de sequências apresenta um nível significativo de saturação, o uso de modelos de substituição de DNA ou de modelos combinados de DNA/RNA com modelos de RNA mais simples (e.g., modelos de RNA com 6-estados) devem ser preferidos ao uso de modelos combinados de DNA/RNA com modelos de RNA muito ricos em parâmetros (e.g., modelos de RNA com 7- e 16-estados). Embora este não seja exatamente o cenário encontrado aqui, os valores de verossimilhança nas análises expandidas realizadas com o pacote PHASE indicam uma melhor performance do modelo RNA6A sobre os modelos RNA7A, RNA16A e RNA 16I (Tabelas 5 e 6C). Contudo, apesar de possuir valores de verossimilhança menores, o modelo RNA7A ainda merece uma análise especial, uma vez que o mesmo, entre os modelos de RNA utilizados e testados, foi o único que resultou em

alterações topológicas nas árvores obtidas (com relação à posição de Oestroidea) e as árvores inferidas com esse modelo foram as que apresentaram os maiores valores de suporte de PP (Tabela 7C). Esses resultados podem ser indicativos da existência de algum sinal filogenético nas moléculas de RNA que são considerados apenas quando sítios não pareados ("mismatches") são formalmente incluídos nas análises, o que não ocorre no modelo RNA6A. Contudo, a medida que a consideração dos estados não-pareados nas análises é expandida, atribuindo cada um destes estados a uma categoria própria (como ocorre nos modelos de RNA de 16-estados), a super-parametrização e a multiplicação do erro devido a mesma pode obscurecer o sinal filogenético, o que pode explicar a performance inferior desses modelos mais complexos (Tabelas 5 e 6C).

A complexidade do cenário geral observado nos resultados alcançados torna difícil fazer conclusões sobre qual estratégia de particionamento resultaria em melhores estimativas filogenéticas usando o método de análise Bayesiano com o conjunto de sequências considerado. Do que foi observado, os efeitos do uso de diferentes combinações de partições e modelos nas análises Bayesianas parece depender de um balanço entre (A) a adoção de estratégias de particionamento mais complexas para melhor modelar a evolução de moléculas não-codificadoras estruturais e regiões gênicas codificadoras de proteínas; e (B) os efeitos finais da super-parametrização e aumento da incerteza nas árvores filogenéticas inferidas. Apesar disso, a análise mais profunda e abrangente realizada aqui, envolvendo diferentes estratégias de particionamento para a análise de conjuntos de sequências de múltiplas regiões gênicas, tem sua importância uma vez que possibilita identificar e compreender quais relações nas árvores filogenéticas recuperadas são instáveis e podem ser modificadas de acordo com o método de análise utilizado e, por consequência,

quais clados são mais sensíveis à super-parametrização e ao aumento da incerteza nas análises filogenéticas decorrentes do uso de modelos mais complexos.

5. Conclusões

As relações inter-familiares inferidas em Oestroidea ainda permanecem longe de um consenso. A medida que novos estudos de filogenia molecular se tornam disponíveis para o grupo, mais e novas hipóteses de relações dentro do grupo são propostas, algumas delas de certa forma contradizendo hipóteses filogenéticas propostas previamente baseadas na análise de caracteres morfológicos. Apesar dessa contradição poder ser explicada em parte pelas diferenças marcantes de amostragem de táxons entre os estudos normalmente conduzidos com caracteres morfológicos quando comparados aos moleculares, algumas relações inferidas podem de fato ser diferentes devido a natureza distinta dos processos que caracterizam a evolução de ambos os tipos de marcadores. Essa diferença básica ressalta a importância de uma hipótese filogenética unificada, incluindo tanto informações morfológicas quanto moleculares, para as relações neste grupo.

A inferência de filogenias confiáveis em Oestroidea pode ser particularmente complicada uma vez que o clado Schizophora (o grupo que contém Calyptratae e a superfamília Oestroidea) parece ter sofrido um evento de radiação episódica entre 65 e 45 milhões de anos atrás (Wiegmann et al. 2011) e existe a possibilidade de que poucas evidências desse processo tenham permanecido nos dados moleculares. Além disso, dependendo dos marcadores moleculares amostrados e considerados, essas evidências podem até mesmo ser conflitantes. Outro fato relevante é que a existência em Oestroidea de grupos com hábitos parasitários extremamente especializados pode ter levado alguns desses grupos a experimentar taxas de mudanças evolutivas bastante distintas, cada qual relacionada a co-evolução com seus respectivos hospedeiros. Essas taxas evolutivas muito

distintas em Oestroidea podem resultar em filogenias estimadas bastante instáveis e em muitos casos não confiável com relação ao posicionamento de alguns clados devido a existência de ramos longos e a possível atração entre eles (dependendo do método de inferência). A família Oestridae, em particular, parece se enquadrar neste cenário. Nestes casos, o uso de modelos e métodos que considerem a heterogeneidade entre taxas evolutivas no processo de inferência filogenética podem contornar esse problema, embora limitações em termos de poder computacional possam tornar esse tipo de análise inaplicável em conjuntos de sequência de tamanho moderado a grande.

Com base no conhecimento atual, as análises aqui apresentadas fornecem evidências adicionais a respeito da não-monofilia de Calliphoridae na forma como a família é atualmente definida. O status de família recentemente atribuído à subfamília Rhiniinae (Rhiniidae) foi recuperado, embora seu posicionamento na filogenia de Oestroidea ainda seja incerto. A inclusão da subfamília Mesembrinellinae dentro de Calliphoridae é duvidosa e análises adicionais mais bem direcionadas em resolver essa questão são necessárias. A possibilidade de que o grupo componha uma família própria, distinta de Calliphoridae, é considerável, como originalmente proposto por Guimarães (1977).

A avaliação dos efeitos das diferentes estratégias de particionamento testadas no processo de inferência filogenética mostrou que o uso de modelos mais complexos (com maior número de parâmetros), apesar de apresentar em geral um melhor ajuste aos dados e assim resultar em um aumento significativo nos valores de verossimilhança associados, não necessariamente resultam em uma melhora significativa nas filogenias inferidas, uma vez que com modelos mais complexos há também um aumento da incerteza acerca dos parâmetros estimados, incluindo a própria topologia das árvores. Além disso, quando não há níveis significativos de saturação de substituição no conjunto de sequências em análise,

pouca ou nenhuma alteração é observada nas topologias inferidas sob modelos com diferentes complexidades. Nestes casos, topologias inferidas com a maioria das estratégias de particionamento consideradas aqui devem resultar em topologias muito semelhantes, embora com diferentes níveis de suporte para os clados recuperados. Sendo assim, o uso de modelos de complexidade intermediária, embora não necessariamente levem a melhoras no processo de inferência filogenética (sobretudo em termos de topologias), devem resultar em árvores mais apropriadamente suportadas, eliminando assim os efeitos de super e sub estimação dos valores de suporte dos ramos por probabilidade *a posteriori*.

6. Referências

- Amendt J, Krettek R, Zehner R (2004) Forensic entomology. *Naturwissenschaften* 91:51-65
- Boyes JW, Shewell GE (1975) Cytotaxonomy of Calliphoridae (Diptera). *Genetica* 45:435–488
- Buckley TR, Simon C, Flook PK, Misof B (2000) Secondary structure and conserved motifs of the frequently sequenced domains IV and V of the insect mitochondrial large subunit rRNA gene. *Insect Mol Biol* 9(6):565-580
- Coleman AW (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet* 19:370-375
- Coleman AW (2007) Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Res* 35:3322-3329
- Coleman AW (2009) Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. *Mol Phylogenet Evol* 50:197-203
- Darty K, Denise A, Ponty Y (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25:1974-1975
- de Azeredo-Espin AM, Lessinger AC (2006) Genetic approaches for studying myiasis-causing flies: molecular markers and mitochondrial genomics. *Genetica* 126:111-131
- de Rijk P, Van De Peer Y, de Wachter R (1997) Database on the structure of the large ribosomal RNA. *Nucleic Acids Res* 25:117-123
- Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

- Goloboff PA, Farris JS, Nixon K (2003) TNT: Tree analysis using New Technology. Version 1.0, version Beta test v. 0.2. Program and documentation available at <http://www.zmuc.dk/public/phylogeny/TNT/>
- Gowri-Shankar V, Jow H (2006) PHASE: a software package for phylogenetics and sequence evolution. Available at: <http://www.bioinf.manchester.ac.uk/resources/phase>
- Griffiths GCD (1972) The phylogenetic classification of Diptera Cyclorrhapha, with special reference to the structure of the male postabdomen. *Series Entomologica*, 8:1-340
- Griffiths GCD (1982) On the systematic position of *Mystacinobia* (Diptera: Calliphoridae). *Memoirs Entomol Soc Washington* 10:70-77
- Guimarães JH (1977) A systematic revision of the Mesembrinellidae, stat. nov. (Diptera, Cyclorrhapha). *Arquivos de Zoologia* 29:1-109
- Guimarães JH, Papavero N, Prado AP (1983) As mífases na região Neotropical: identificação, biologia e bibliografia. *Rev Bras Zool* 1:239-416
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696-704
- Gutell RR, Subashchandran S, Schnare M, Du Y, Lin N, Madabusi L, Muller K, Pande N, Yu N, Shang Z, Date S, Konings D, Schweiker V, Weiser B, Cannone JJ (unpublished) Comparative sequence anal prediction RNA structure and the web. <<http://www.rna.icmb-utexas.edu/>>
- Hall DG (1948) The blowflies of North America. Washington: Thomas Say Foundation
- Hall M, Wall R (1995) Myiasis of humans and domestic animals. *Adv Parasitol* 35:257-334
- Hennig W (1973) 31. Diptera (Zweiflügler). *Handbuch der Zoologie* 4 (2) 2/31: 1–337 (Lieferung 20)
- Infante MEV, Azeredo-Espin AML (1995) Genetic variability in mtDNA of *Cochliomyia hominivorax* (Diptera: Calliphoridae) from Brazil. *Biochem Genet* 33:237-256
- Joseph N, Krauskopf E, Vera MI, Michot B (1999) Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. *RNA* 8:786-797
- Kass RE, Raftery AE (1995) Bayes factor. *J Am Stat Assoc* 90:773-795
- Kurahashi H (1989) Family Calliphoridae. In “Catalog of the Diptera of the Australasian and Oceanian Regions.” (N. L. Evenhuis, Ed.). Bishop Museum Serial Publication 86, Honolulu
- Kutty SN, Pape T, Wiegmann BM, Meier R (2010) Molecular phylogeny of the Calyptratae (Diptera: Cyclorrhapha) with an emphasis on the superfamily Oestroidea and the position of Mystacinobiidae and McAlpine’s fly. *Syst Entomol* 35:614-635

- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947-2948
- Lehrer AZ (1970) Considerations phylogenetiques et taxonomiques sur la famille Calliphoridae (Diptera). *Annotationes zoologicae et botanicae. Slovenske Norodne Muzeum v Bratislave* 61:1- 51
- Lehrer AZ (2005) Bengaliidae du Monde (Insecta, Diptera). *Pensoft Series Faunistica No. 50*, Sofia, Moscow
- Leigh JW, Susko E, Baumgartner M, Roger AJ (2008) Testing congruence in phylogenomic analysis. *Syst Biol* 57(1): 104-15
- Letsch HO, Kjer KM (2011) Potential pitfalls of modelling ribosomal RNA data in phylogenetic tree reconstruction: Evidence from case studies in the Metazoa. *BMC Evol Biol* 11:146
- Marinho MAT, Junqueira ACM, Azeredo-Espin AML (2011) Evaluation of the internal transcribed spacer 2 (ITS2) as a molecular marker for phylogenetic inference using sequence and secondary structure information in blow flies (Diptera: Calliphoridae). *Genetica* 139(9): 1189-1207
- McAlpine JF (1989) *Manual of nearctic Diptera 3. Agriculture Canada Monograph 32*. Canadian Gov. Publ. Center, Quebec, Canada
- Müller T, Philippi N, Dandekar T, Schultz J, Wolf M (2007) Distinguishing species. *RNA* 13:1469-1472
- Nylander JA (2004) MrAIC.pl. Program distributed by the author. Evolutionary Biology Centre, Uppsala University
- Nylander JA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL (2004) Bayesian phylogenetic analysis of combined data. *Syst Biol* 53(1):47-67
- Pape T (1992) Phylogeny of the Tachinidae family-group (Diptera: Calyptratae). *Tijdschrift voor Entomologie* 135:43–86
- Pape T, Arnaud PH (2001) *Bezzimyia* – a genus of native New World Rhinophoridae (Insecta, Diptera). *Zoologica Scripta* 30:257–297
- Pape T, Thompson FC (editors) (2010) *Systema Dipteriorum*, Version 1.0. <http://www.diptera.org/>, accessed on October 2011
- Robinson DR, Foulds LR (1981) Comparison of phylogenetic trees. *Math. Biosci.* 53, 131-147
- Rognes K (1986) The systematic position of the genus *Helicobosca* Bezzi with a discussion of the monophyly of the calyptrate families Calliphoridae, Rhinophoridae, Sarcophagidae and Tachinidae (Diptera). *Entomologica Scandinavica* 17, 75–92.

- Rognes K (1991) Blowflies (Diptera, Calliphoridae) of Fennoscandia and Denmark. *Fauna Entomologica Scandinavica* 24:1-272
- Rognes K (1997) The Calliphoridae (blowflies) (Diptera, Oestroidea) are not a monophyletic group. *Cladistics* 13:27-66
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning - A Laboratory Manual*, 2nd Edition. Cold Spring Harbour Laboratory Press, New York
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502-504
- Seibel PN, Müller T, Dandekar T, Schultz J, Wolf M (2006) 4SALE: a tool for synchronous RNA sequence and secondary structure alignment and editing. *BMC Bioinformatics* 7:498
- Shewell G (1987) 106 Calliphoridae. In "Manual of Nearctic Diptera." (J. F. McAlpine, Ed.), Vol. 2. Research Branch, Agriculture Canada, Monograph 28, pp.1133-1145
- Singh B, Wells JD (2011) Chrysomyinae (Diptera: Calliphoridae) is monophyletic: a molecular systematics analysis. *Syst Entomol* 36(3):415-420
- Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688-2690
- Steel MA, Lockhart PJ, Penny D (1993) Confidence in evolutionary trees from biological sequence data. *Nature* 364, 440-442
- Stevens JR, Wallman JF (2006) The evolution of myiasis in humans and other animals in the Old and New Worlds (part I): phylogenetic analyses. *Trends Parasitol* 22:129-136
- Stevens JR, Wallman JF, Otranto D, Wall R, Pape T (2006) The evolution of myiasis in humans and other animals in the Old and New Worlds (part II): biological and life-history studies. *Trends Parasitol* 22:181-188
- Strimmer K, von Haeseler A (1997) Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. USA* 94:6815-6819
- Tachi T, Shima H (2010) Molecular phylogeny of the subfamily Exoristinae (Diptera, Tachinidae), with discussions on the evolutionary history of female oviposition strategy. *Syst Entomol* 35:148-163
- Toma R, Carvalho CJB (1995) Estudo filogenético de Mesembrinellinae com ênfase no gênero *Eumesembrinella* Townsend (Diptera: Calliphoridae). *Revta Bras Zool* 12(1): 127-144

- Tourle R, Downie DA, Villet MH (2009) Flies in the ointment: a morphological and molecular and comparison of *Lucilia cuprina* and *Lucilia sericata* (Diptera: Calliphoridae) in South Africa. *Med Vet Entomol* 23(1):6-14
- Tschorsnig HP (1985) Die Struktur des männlichen Postabdomens der Rhinophoridae (Diptera). *Stuttgarter Beiträge zur Naturkunde, Serie A (Biologie)* 375:1–18
- Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim JW, Lambkin C, Bertone MA, Cassel BK, Bayless KM, Heimberg AM, Wheeler BM, Peterson KJ, Pape T, Sinclair BJ, Skevington JH, Blagoderov V, Caravas J, Kutty SN, Schmidt-Ott U, Kampmeier GE, Thompson FC, Grimaldi DA, Beckenbach AT, Courtney GC, Friedrich M, Meier R, Yeates DK (2011) Episodic radiations in the fly tree of life. *Proc Natl Acad Sci U S A* 108(14):5690-5695
- Xia X (2001) Data analysis in molecular biology and evolution. Kluwer Academic Publishers, Boston
- Xia X, Lemey P (2009) Assessing substitution saturation with DAMBE. Pp. 615-630 in Philippe Lemey, Marco Salemi and Anne-Mieke Vandamme, eds. *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny*. 2nd edition Cambridge University Press
- Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. *J Hered* 92(4):371-373
- Xia X, Xie Z, Salemi M, Chen L, Wang Y (2003). An index of substitution saturation and its application. *Mol Phylogenet Evol* 26:1-7
- Young I, Coleman AW (2004) The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a *Drosophila* example. *Mol Phylogenet Evol* 30:236-242
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406-3415
- Zumpt F (1956) Calliphoridae (Diptera Cyclorrhapha). Part I: Calliphorini and Chrysomyiini. *Exploration du Parc National Albert Mission G. F. de Witte (1933–1935)* 87:1–200
- Zumpt F (1965) Myiasis in man and animals in the Old World. London, Butterworths. p. 267
- Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD dissertation, University of Texas at Austin, Austin, Texas

Conclusões Gerais

Capítulo I:

- O potencial da região do ITS2 como um marcador molecular para análises de inferência filogenética em Calliphoridae, como previamente sugerido em outros grupo, foi confirmado;
- A variação em termos de sequência da região ITS2 parece estar hierarquicamente organizada nos níveis de espécies, gêneros e subfamílias, enquanto pouca variação foi encontrada nos níveis intra-específico e intra-genômico;
- A evolução da sequência primária do ITS2, incluindo seu conteúdo nucleotídico, parece estar intimamente relacionada à evolução da estrutura secundária, com domínios mais conservados apresentando maior conteúdo G+C (estabilidade);
- Avaliação baseada em performance corroborou o potencial do ITS2 como marcador molecular para análises filogenéticas sob diferentes métodos de análise;
- O uso de modelos que incorporam informação de estrutura secundária nas análises filogenéticas, embora tenham mostrado um maior ajuste aos dados, resultaram em mudanças discretas nas árvores recuperadas (topologia e suporte).

Capítulo II:

- A análise do padrão de fragmentos gerados após a digestão do RNA transcrito permitiu a obtenção de um modelo de baixa resolução para a estrutura do ITS2, o qual parece corroborar muitos dos motivos estruturais preditos *in silico*;
- A análise das estruturas preditas *in silico* para o ITS2, juntamente com alguns padrões observados nas análises *in vitro*, sugerem a existência de uma dinâmica conformacional na região central da estrutura do ITS2, sendo que a alternância nos estados

pareados de duas estruturas preditas segundo o modelo em "grampo" do ITS2 pode gerar a estrutura predita segundo o modelo em "anel" como estado intermediário;

- O uso da estrutura segundo o modelo em "anel" quando da incorporação de dados estruturais em análises genético-evolutivas, embora desconsidere a dinamicidade existente na molécula *in vitro* e *in vivo*, parece uma simplificação razoável, uma vez que a estrutura contém todos os domínios gerais e de possível importância funcional da molécula;
- Em termos de evolução da estrutura do ITS2 em Calyptratae, a principal alteração parece ter sido relacionada a um aumento significativo da extensão (número de pareamentos) do domínio III, o qual foi acompanhado do surgimento de uma zona de instabilidade em sua porção central, a qual permitiu o surgimento de pequenas hélices laterais (em Muscoidea). Essas hélices laterais foram estabilizadas em Oestroidea (subdomínios IIIa e IIIc);

Capítulo III:

- A monofilia de Oestroidea e da maioria de suas famílias foram recuperadas, com exceção de Calliphoridae;
- As relações inter-familiares em Oestroidea ainda estão longe de um consenso;
- A família Calliphoridae, como atualmente definida, é provavelmente não-monofilética;
- O recém atribuído status de família para Rhiniinae foi recuperado;
- O mesmo pode também ser sugerido para Mesembrinellinae, com o grupo podendo ser mais próximo da família Tachinidae do que do grupo monofilético de Calliphoridae;

- O conflito entre estudos morfológicos e moleculares é evidente em algumas relações (e.g. relações de parentesco de Toxotarsinae), ressaltando a necessidade de um estudo integrado entre dados morfológicos e moleculares;
- Avaliação de diferentes estratégias de particionamento dos dados mostrou que modelos mais complexos e mais ricos em parâmetros apresentam melhor ajuste, como inferido pelo aumento significativo dos valores de verossimilhança nas análises, embora estejam também relacionados com um aumento significativo da incerteza nas topologias e parâmetros inferidos;
- Apesar das diferenças significativas no ajuste do modelo aos dados, não foram observadas alterações significativas em termos de topologia e comprimentos de ramos, com apenas algumas alterações discretas sendo observadas no suporte geral das árvores recuperadas.

Referências

- Amendt J, Krettek R, Zehner R (2004) Forensic entomology. *Naturwissenschaften* 91:51-65
- Amorim DS, Silva VC (2002) How far advanced was Diptera evolution in the Pangaea? *Annales de la Société Entomologique de France*. 32:177–200
- Boore JL (1990) Animal Mitochondrial Genomes. *Nucleic Acids Res* 27(8):1767-1780
- Bremer B, Jansen RK, Oxelman B, Backlund M, Lantz H, Kim KJ (1999) More characters and more taxa for a robust phylogeny – case study from the coffee family (Rubiaceae). *Syst Biol* 48: 413-435
- Burleigh JG, Hilu KW, Soltis DE (2009) Inferring phylogenies with incomplete data sets: a 5-gene, 567-taxon analysis of angiosperms. *BMC Evol Biol* 9:61
- Caterino MS, Cho S, Sperling FAH (2000) The current state of insect molecular systematics: A thriving tower of babel. *Annl Rev Entomol* 45:1-54
- 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)? *Syst Biol* 60(6): 782–796
- Coleman AW (2009) Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. *Mol Phylogenet Evol* 50:197-203
- Coleman AW, Vacquier VD (2002) Exploring the phylogenetic utility of its sequences for animals: a test case for abalone (haliotis). *J Mol Evol* 54:246-257
- Degnan JH, Rosenberg NA (2006) Discordance of species trees with their most likely gene trees. *PLoS Genet* 2:762–768
- Dethier VG (1976) *The Hungry Fly*. Harvard University Press; Cambridge, Massachusetts; 489 pp
- Dixon MT, Hillis DM (1993) Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. *Mol Biol Evol* 10: 256-267
- Elder JF Jr, Turner BJ (1995) Concerted evolution of repetitive DNA sequences. *Q Rev Biol* 70:297-320
- Gardner PP, Giegerich R (2004) A comprehensive comparison of comparative RNA structure prediction approaches. *BMC Bioinformatics*. 30(5): 140-158
- Gatesy J, DeSalle R, Wahlberg N (2007) How many genes should a systematist sample? Conflicting insights from a phylogenomic matrix. *Syst Biol* 56: 355-363

Referências

- Gillespie JJ (2004) Characterizing regions of ambiguous alignment caused by the expansion and contraction of hairpin-stem loops in ribosomal RNA molecules. *Mol Phylogenet Evol* 33: 936-943.
- Gowri-Shankar V, Jow H (2006) PHASE: a software package for phylogenetics and sequence evolution. Available at: <http://www.bioinf.manchester.ac.uk/resources/phase>
- Graybeal A (1998) Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst Biol* 47:9-17
- Griffiths GCD (1972) The phylogenetic classification of Diptera Cyclorrhapha, with special preference to the structure of the male postabdomen. *Series Entomologica* 8:1-340
- Griffiths GCD (1982) On the systematic position of *Mystacinobia* (Diptera: Calliphoridae). *Memoirs Entomol Soc Washington* 10:70-77
- Grimaldi D, Engel MS (2005) *Evolution of the Insects* (Cambridge Univ. Press, New York)
- Guimarães JH, Papavero N, Prado AP (1983) As míases na região Neotropical: identificação, biologia e bibliografia. *Rev Bras Zool* 1:239-416
- Gutell RR (1996) Comparative sequence analysis and the structure of 16S and 23S rRNA. p. 111-128. In *Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis*. (R. A. Zimmermann & A. E. Dahlberg, eds.). CRC Press, New York
- Hall M, Wall R (1995) Myiasis of humans and domestic animals. *Adv Parasitol* 35:257-334
- Heath TA, Hedtke SM, Hillis DM (2008) Taxon sampling and the accuracy of phylogenetic analyses. *J Syst Evol* 46 (3): 239-257
- Hedtke SM, Townsend TM, Hillis DM (2006) Resolution of Phylogenetic Conflict in Large Data Sets by Increased Taxon Sampling. *Sist Biol* 55(3):522-529
- Hennig W (1973) 31. Diptera (Zweiflügler). *Handbuch der Zoologie* 4 (2) 2/31: 1-337 (Lieferung 20)
- Hillis DM (1998) Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Syst Biol* 47:3-8
- Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *Q Rev Biol* 66:411-453
- Hillis DM, Pollock DD, McGuire JA, Zwickl DJ (2003) Is sparse taxon sampling a problem for phylogenetic inference? *Syst Biol* 52:124-126
- Hofacker IL (2003) Vienna RNA secondary structure server. *Nucleic Acids Res* 31(13):3429-31
- Holder H, Lewis PO (2003) Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet* 4:75-284

- Johnston SD (2002) The art and design of genetic screens: *Drosophila melanogaster*. Nat Rev Genet 3(3): 176-188
- Koblmüller S, Egger B, Sturmbauer C, Sefc KM (2010) Rapid radiation, ancient incomplete lineage sorting and ancient hybridization in the endemic Lake Tanganyika cichlid tribe Tropheini. Mol Phylogenet Evol 55:318–334
- Kutty SN, Pape T, Wiegmann BM, Meier R (2010) Molecular phylogeny of the Calyptratae (Diptera: Cyclorrhapha) with an emphasis on the superfamily Oestroidea and the position of Mystacinobiidae and McAlpine's fly. Syst Entomol 35:614-635
- Lehrer AZ (1970) Considerations phylogenetiques et taxonomiques sur la famille Calliphoridae (Diptera). Annotationes zoologicae et botanicae. Slovenske Narodne Muzeum v Bratislave 61:1- 51
- Letsch HO, Kjer KM (2011) Potential pitfalls of modelling ribosomal RNA data in phylogenetic tree reconstruction: Evidence from case studies in the Metazoa. BMC Evol Biol 11:146
- Liao D (1999) Concerted evolution: molecular mechanisms and biological implications. Am J Hum Genet 64:24-30
- Lunt DH, Hyman BC (1997) Animal Mitochondrial DNA recombination. Nature 387: 247
- Maddison WP (1997) Gene trees in species trees. Syst Biol 46:523-536
- Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting . Syst Biol 55(1):21-30
- McAlpine JF (1989) Manual of nearctic Diptera 3. Agriculture Canada Monograph 32. Canadian Gov. Publ. Center, Quebec, Canada
- Moulton JK, Wiegmann BM (2004) Evolution and phylogenetic utility of cad (rudimentary) among Mesozoic-age Eremoneuran Diptera (Insecta). Mol Phylogenet Evol 31:363-378
- Müller T, Philippi N, Dandekar T, Schultz J, Wolf M (2007) Distinguishing species. RNA 13:1469-1472
- Muse SV (1995) Evolutionary analyses of DNA sequences subject to constraints of secondary structure. Genetics 139:1429-1439
- Pamilo P, Nei M (1988) Relationships between gene trees and species trees. Mol Biol Evol 5:568-583
- Pape T (1992) Phylogeny of the Tachinidae family-group (Diptera: Calyptratae). Tijdschrift voor Entomologie 135:43–86
- Pape T (2006) Phylogeny and evolution of the bot flies. The Oestrid Flies: Biology, Host–Parasite Relationships, Impact and Management (ed. by D. Colwell, P. Scholl and M. Hall), pp. 20–50. CABI Publishers, Wallingford

Referências

- Pape T, Arnaud PH (2001) *Bezzimyia* – a genus of native New World Rhinophoridae (Insecta, Diptera). *Zoologica Scripta* 30:257–297
- Poe S (1998a) The effect of taxonomic sampling on accuracy of phylogeny estimation: Test case of a known phylogeny. *Mol Biol Evol* 15:1086-1090
- Poe S (1998b) Sensitivity of phylogeny estimation to taxonomic sampling. *Syst Biol* 47:18-31
- Poe S (2003) Evaluation of the strategy of long-branch subdivision to improve the accuracy of phylogenetic methods. *Syst Biol* 52:423- 428
- Pollard DA, Iyer VN, Moses AM, Eisen MB (2006) Widespread Discordance of Gene Trees with Species Tree in *Drosophila*: Evidence for Incomplete Lineage Sorting. *PLoS Genet* 2(10):e173
- Reeder J, Höchsmann M, Rehmsmeier M, Voss B, Giegerich R (2006) Beyond Mfold: recent advances in RNA bioinformatics. *Journal of Biotechnology*. 124(1): 41-55.
- Rognes K (1991) Blowflies (Diptera, Calliphoridae) of Fennoscandia and Denmark. *Fauna Entomologica Scandinavica* 24:1-272
- Rognes K (1997) The Calliphoridae (blowflies) (Diptera, Oestroidea) are not a monophyletic group. *Cladistics* 13:27-66
- Rokas A, Carroll SB (2005) More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Mol Biol Evol* 22:1337-1344
- Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425:798-804
- Ronquist F, Deans AR (2010) Bayesian phylogenetics and its influence on insect systematics. *Ann Rev Entomol* 55:189-206
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574
- Rzhetsky A (1995) Estimating substitution rates in ribosomal RNA genes. *Genetics* 141: 771-783
- Savill NJ, Hoyle DC, Higgs PG (2001) RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods. *Genetics* 157:399-411
- Schöniger M, von Haeseler A (1994) A stochastic model for the evolution of autocorrelated DNA sequences. *Mol Phylogenet Evol* 3:240-247
- Sorenson MD, Oneal E, García-Moreno J, Mindell DP (2003) More taxa, more characters: the hoatzin problem is still unresolved. *Mol Biol Evol* 20: 1484-1499

- Takahashi K, Terai Y, Nishida M, Okada N (2001) Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. *Mol Biol Evol* 18(11):2057-2001
- Telford MJ, Wise MJ, Gowri-Shankar V (2005) Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the bilateria. *Mol Biol Evol* 22:1129-1136
- Tillier E, Collins R (1995) Neighbor-Joining and maximum likelihood with RNA sequences: addressing the inter-dependence of sites. *Mol Biol Evol* 12:7–15.
- Wahlberg N, Wheat CW (2008) Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Syst Biol* 57: 231-242
- Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim JW, Lambkin C, Bertone MA, Cassel BK, Bayless KM, Heimberg AM, Wheeler BM, Peterson KJ, Pape T, Sinclair BJ, Skevington JH, Blagoderov V, Caravas J, Kutty SN, Schmidt-Ott U, Kampmeier GE, Thompson FC, Grimaldi DA, Beckenbach AT, Courtney GC, Friedrich M, Meier R, Yeates DK (2011) Episodic radiations in the fly tree of life. *Proc Natl Acad Sci U S A* 108(14):5690-5695
- Wiegmann BM, Yeates DK, Thorne JL, Kishino H (2003) Time flies, a new molecular time-scale for brachyceran fly evolution without a clock. *Syst Biol* 52: 745–756
- Wiemers M, Keller A, Wolf M (2009) ITS2 secondary structure improves phylogeny estimation in a radiation of blue butterflies of the subgenus *Agrodiaetus* (Lepidoptera: Lycaenidae: Polyommatus). *BMC Evol Biol* 9:300
- Wiens JJ (1998a) Combining data sets with different phylogenetic histories. *Syst Biol* 47:568–581
- Wiens JJ (1998b) Does adding characters with missing data increase or decrease phylogenetic accuracy? *Syst Biol* 47:625–640
- Wolf M, Achtziger M, Dandekar T, Müller T (2005) CBCAnalyzer: inferring phylogenies based on compensatory base changes in RNA secondary structures. *In silico Biology*. 5(3): 291-294
- Wolstensolme DR (1992) Animal Mitochondrial DNA: Structure and Evolution. *IntRev Cytol* 141:173-216
- Yeates DK, Wiegmann BM (1999) Congruence and controversy: toward a higher-level phylogeny of Diptera. *Ann Rev Entomol* 44:397-428
- Yu J, Thorne JL (2006) Dependence among sites in RNA evolution. *Mol Biol Evol* 23(8): 1525-1537.
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406-3415

Referências

Zumpt F (1965) Myiasis in man and animals in the Old World. London, Butterworths. p. 267

Zwickl D J, Hillis DM (2002) Increased taxon sampling greatly reduces phylogenetic error. Syst Biol 51:588-598