

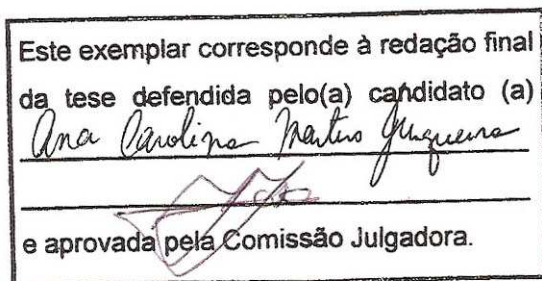
UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA



Ana Carolina Martins Junqueira

**“EVOLUÇÃO MOLECULAR DO GENOMA
MITOCONDRIAL DA FAMÍLIA
CALLIPHORIDAE (DIPTERA: BRACHYCERA)”**



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: Profa. Dra. Ana Maria Lima de Azeredo-Espin

Co-Orientadora: Profa. Dra. Claudia Augusta de Moraes Russo

Campinas, 2008

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

J968e	<p>Junqueira, Ana Carolina Martins Evolução molecular do genoma mitocondrial da família Calliphoridae (Diptera: Brachycera) / Ana Carolina Martins Junqueira. – Campinas, SP: [s.n.], 2008.</p> <p>Orientadores: Ana Maria Lima de Azeredo-Espin, Claudia Augusta de Moraes Russo. Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.</p> <p>1. DNA mitocondrial. 2. Diptera. 3. Mosca-varejeira. 4. Míiase. 5. Evolução molecular. I. Azeredo-Espin, Ana Maria Lima de. II. Russo, Claudia Augusta de Moraes. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.</p>
--------------	---

Título em inglês: Molecular evolution of the mitochondrial genome of the family Calliphoridae (Diptera: Brachycera).

Palavras-chave em inglês: Mitochondrial DNA; Diptera; Blowflies; Myiasis; Molecular evolution.

Área de concentração: Genética Animal e Evolução.

Titulação: Doutora em Genética e Biologia Molecular

Banca examinadora: Ana Maria Lima de Azeredo-Espin, Cristina Yumi Miyaki, João Miguel de Barros Alexandrino, Gonçalo Amarante Pereira, André Victor Lucci Freitas.

Data da defesa: 29/02/2008.

Programa de Pós-Graduação: Genética e Biologia Molecular.

Campinas, 29 de fevereiro de 2008.

BANCA EXAMINADORA

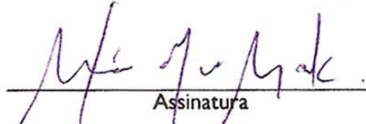
Profa. Dra. Ana Maria L. de Azeredo-Espin (Orientadora)


Assinatura

Profa. Dra. Marie-Anne Van Sluys

Assinatura


Profa. Dra. Cristina Yumi Miyaki


Assinatura

Prof. Dr. Sérgio Furtado dos Reis

Assinatura

Prof. Dr. Gonçalo Amarante Pereira


Assinatura

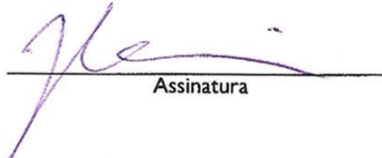
Prof. Dr. Louis Bernard Klaczko

Assinatura

Prof. Dr. André Victor Lucci Freitas


Assinatura

Prof. Dr. João Miguel de Barros Alexandrino


Assinatura



“Eu prefiro os laços firmes, aqueles mais difíceis de se fazer e de desfazer, mas que quando feitos e depois desfeitos podem se orgulhar de si próprios e falar com convicção: ‘Eu fui um grande laço!’”

Do curta “Laços”, de Adriana Falcão

"We live together, we act on, and react to, one another; but always and in all circumstances we are by ourselves. The martyrs go hand in hand into the arena; they are crucified alone. Embraced, the lovers desperately try to fuse their insulated ecstasies into a single self-transcendence; in vain. By its very nature every embodied spirit is doomed to suffer and enjoy in solitude. Sensations, feelings, insights, fancies - all these are private and, except through symbols and at second hand, incommunicable. We can pool information about experiences, but never the experiences themselves. From family to nation, every human is a society of island universes"

Aldous Huxley

**"Neto e neta são netos, no masculino.
Filho e filha são filhos, no masculino.
Pai e mãe são pais, no masculino.
Avô e avó são avós"**

**"Todas as coisas do mundo não cabem numa idéia. Mas
tudo cabe numa palavra, nesta palavra tudo"**

Arnaldo Antunes, em "As Coisas"

A meus pais e meus avós, por tudo.

AGRADECIMENTOS

À Ana Maria por me orientar e desorientar, por me acompanhar na saga das descobertas, por me encorajar nos momentos de dúvida e vacilo, por me ajudar a concretizar sonhos e me mostrar que outros ainda estão por vir, por me suportar nas horas difíceis, por compartilhar as horas de alegria, por me dar liberdade, por confiar tanto e sempre, por me fazer crescer e amadurecer, por me fazer reclamar menos das coisas todas do universo conspiratório e arregaçar as mangas pra fazer de novo... e por sempre, sempre acreditar em mim.

À Claudia Russo, pela co-orientação, pelas discussões valiosas, pelos novos aprendizados e pelas viagens aos congressos da SMBE, que foram divertidíssimas e me propiciaram tantos contatos com as “literaturas ambulantes”.

À Ana Claudia, pelas discussões, pelas viagens, pela companhia sempre-presente, pelas opiniões criteriosas, pelo exemplo que me dá, pela confiança que me inspira, por ser brilhante e me deixar compartilhar um pouquinho disso.

À Rosangela, por ser tão confiável e tão pertinente na bancada do laboratório, sempre tomando conta de todas as sequências e tirando as minhas dúvidas, a qualquer momento do dia ou da noite. Por ser sempre tão providencial e querer ajudar de coração.

Às minhas especiais companheiras e sócias-fundadoras da sala-DDD, Tati e Mari, por me fazerem rir, por me divertirem tirando fotos, por me fazerem companhia, pelas discussões riquíssimas, pela paciência de cada uma, pela ajuda nos momentos de angústia, por me deixarem ligar o ar condicionado mesmo quando estava frio, por serem amigas.

Ao pessoal do Lab, que faz o dia-a-dia ser melhor: Norma, Pablo, Renato, Johnny, Priscila, Karina, Luciane, Alessandra, Salete e Pedro Feijão. Em especial à Bárbara, pelo senso de humor e pelas confidências literárias e cinematográficas; e ao Marco, “meu” aluno-prodígio que hoje mais me ensina do que me deixa ensinar, por me tirar dúvidas de linha de comando, pelos “intervalinhos” da tarde, pelos churrascos e por me deixar participar da sua iniciação.

Ao Jeffrey Wells e Carole Ames, por enviarem moscas tão valiosas de outras partes do mundo. Ao Arício Xavier Linhares, por identificar prontamente as espécies de califorídeos depois das coletas.

Às secretárias do CBMEG Sandra e Tânia, pelo empenho, dedicação e paciência.

Aos membros da pré-banca, André Freitas e João Alexandrino, pelas sugestões e discussão que foram tão proveitosas e pertinentes para a tese.

À minha banca de doutorado: Cristina, João, André e Gonçalo (que também foi meu patrono de formatura!).

Aos meus bons amigos da turma 96D, fiéis e companheiros, que guardo com carinho: Patrick, Carlinha, Michele, Paula, Rachel, Tatiana e Victor.

À família que eu escolhi e que também me escolheu, por estar sempre presente, por me propiciar tantos momentos de pura diversão e alegria, por me acompanhar em todas as fases da vida, na alegria e na tristeza. Sem todos vocês a vida seria menos divertida, apaixonada e intensa. A todos e a cada um, agradeço pelas peculiaridades, singularidades e pelo amor incondicional e recíproco: Jô, Fer, Marina, Léo, Ed, Isa, André, Thiago, Pedro, Gui, Preg, Brunão, Carolina, Alê e às crianças que eu gosto tanto Luca, Leon e Davi.

À família que me acolhe e me abriga, onde e quando eu precisar. À minha mãe Helena, que tenta me entender mesmo quando estou mal-humorada, me dando carinho. Por cuidar de mim, por ter um orgulho particular de cada um dos filhos, por preparar meu almoço e meu jantar, por me acompanhar nas decisões difíceis e por existir tão presentemente na minha vida. Ao meu pai Laurindo, que acredita em mim e me dá todo o suporte necessário para ser quem eu sou, por me deixar crescer e por participar deste processo, pelo interesse, por me dar escolhas e respeitá-las, e por ser um exemplo de perseverança e inteligência. À Laura, minha irmãzinha que me deixa saudades por morar longe, mas que se faz presente mesmo com a distância e com a diferença de fuso. Por me abrigar na sua casa em Vancouver e por cuidar tão bem de mim. Ao Luis, pelo carinho, pela companhia, pelas confidências... À Baba Marfa, minha avó querida, que sempre cuidou de mim desde que eu nasci, por querer satisfazer meu gosto, por ter tanta paciência, atenção e amor, por acender tantas velinhas, por ser um exemplo de vida e por me fazer gostar e ter orgulho das histórias da *minha* família, com os relatos tão ricos e detalhados.

Ao Dani e à Luana, por cuidarem dos meus irmãos e serem companheiros tão especiais, pra mim e pra eles.

Ao Tio Danilo, em especial, pelos múltiplos cuidados ao longo de toda minha vida, pelas pipas singulares que só ele fazia e pela torcida incondicional. A todos os outros tios e tias da família Buracovas, que sempre cuidaram de mim quando pequena.

À Baba Natália, que me acompanha de longe, sendo sempre tão atenciosa, carinhosa e também por me elogiar sempre que me vê! Ao Vovô Laurindo (mesmo em outro plano), pelas histórias do exército, por me levar “no fundão” do mar e me fazer gostar tanto dele. Aos Junqueira, por tornarem as festas de família divertidas e, agora, habitada por crianças.

À Fapesp, pela concessão da bolsa de doutorado, sem a qual eu não faria este trabalho (processo 03/05444-2).

ÍNDICE

	Página
Resumo	I
Abstract	3
Introdução	5
1. A mitocôndria e o genoma mitocondrial	5
2. O mtDNA como marcador molecular em insetos	6
3. A família Calliphoridae (Diptera)	10
4. Estudos moleculares em Calliphoridae	13
5. Genômica mitocondrial da família Calliphoridae: aplicações e perspectivas	14
Objetivos	19
Materiais e Métodos	21
1. Coletas e amostras	21
2. Extração de DNA	21
3. Amplificação por PCR de longa extensão ("long-PCR")	22
4. "Shotgun": quantificação e sonicação dos amplicons	25
5. Extração do mtDNA sonicado do gel de agarose de baixo ponto de fusão	25
6. Biblioteca dos genomas mitocondriais e seleção de clones recombinantes	26
7. Sequenciamento, montagem e anotação dos genomas mitocondriais	26
8. Análise das sequências	28
Resultados	31
1. Extração de DNA	31
2. "Long-PCR"	31
3. Bibliotecas de "shotgun" do genoma mitocondrial	33
4. Caracterização dos genomas mitocondriais de <i>C. idioidea</i> , <i>C. vomitoria</i> e <i>P. regina</i>	34
4.1. Genes codificadores de proteínas	37
4.2. Subunidades do rRNA	46
4.3. Genes para tRNA	46
4.4. Região controle	47
5. Análises filogenéticas	49
Discussão	63
1. Sequenciamento do mtDNA de <i>C. idioidea</i> , <i>C. vomitoria</i> e <i>P. regina</i>	63
2. Arquitetura genômica e composição A+T	64
3. Genes codificadores de proteínas	67
4. Região controle e duplicação de tRNAs	69
5. Análises filogenéticas	71
6. Considerações finais e perspectivas	75

Conclusões	77
Referências bibliográficas	79
Apêndice A Analysis of a conserved duplicated tRNA gene in the mitochondrial genome of blowflies	91
Apêndice B The mitochondrial genome of the blowfly <i>Chrysomya chloropyga</i> (Diptera: Calliphoridae)	99
Apêndice C The mitochondrial control region of blowflies (Diptera: Calliphoridae): a hot spot for mitochondrial genome rearrangements	111
Apêndice D Evaluation of the internal transcribed spacer 2 (ITS2) as a molecular marker for phylogenetic inference using sequence and secondary structure information: an example for the myiasis-causing flies of the Calliphoridae family (Diptera: Brachycera).	123

TABELAS E FIGURAS

<u>Tabelas</u>	Página
Tabela 1. Sequência dos oligonucleoídeos usados no “long-PCR” de califorídeos.	24
Tabela 2. Espécies de dípteros utilizadas nas análises comparativas.	29
Tabela 3. Composição A+T do mtDNA total e medidas de AT e GC-skews.	36
Tabela 4. Códon de iniciação e terminação de <i>C. idioidea</i> , <i>C. vomitoria</i> e <i>P. regina</i> .	38
Tabela 5. “Codon usage” dos GCPs de <i>C. idioidea</i> .	39
Tabela 6. “Codon usage” dos GCPs de <i>C. vomitoria</i> .	40
Tabela 7. “Codon usage” dos GCPs de <i>P. regina</i> .	40
Tabela 8. Frequência de aminoácidos em proteínas mitocondriais.	41
Tabela 9. Distância p , calculada para cada par de espécies de dípteros em cada GCP.	43
Tabela 10. Comparação A+T e tamanho da região controle de espécies de Calyptratae.	48

<u>Figuras</u>	Página
Figura 1. Relações entre as famílias de Oestroidea e as superfamílias de Calyptratae.	11
Figura 2. Esquema de amplificação do “long-PCR”.	24
Figura 3. Gel de agarose com os amplicons amplificados por “long-PCR”.	33
Figura 4. Gel de agarose para verificação do isolamento de fragmentos entre 1 e 2 kb para a biblioteca de mtDNA.	33
Figura 5. Estrutura dos genomas mitocondriais de <i>C. idioidea</i> , <i>C. vomitoria</i> e <i>P. regina</i> .	35
Figura 6. Proporção relativa dos nucleotídeos por posição de códon.	38
Figura 7. Comparação do número absoluto de sítios conservados e variáveis encontrados nas sequências de aminoácidos dos GCPs de dípteros.	42
Figura 8. Comparação da razão C/V nos 13 GCPs de dípteros.	43
Figura 9. Identidade de aminoácidos encontrada ao longo dos 13 GCPs.	44, 45

Figura 10. Estrutura da região controle nas espécies de califorídeos.	48
Figura 11. Árvore filogenética recuperada com o método BI (GTR+G) para o conjunto de dados subnt_12.	52
Figura 12. Árvore de NJ para o conjunto de dados subnt_12 (Tamura-Nei+G).	53
Figura 13. Árvore filogenética recuperada com o método BI (GTR+G) para o conjunto de dados nt_12.	54
Figura 14. Árvore filogenética recuperada com o método ML (mtRev+G+I) para o conjunto de dados sub_aa.	56
Figura 15. Árvore de NJ para o conjunto de dados sub_aa (correção de Poisson).	57
Figura 16. Árvore filogenética recuperada com o método NJ (Tamura-Nei+G) para o conjunto de dados aa_comp.	60
Figura 17. Árvore de BI para o conjunto de dados aa_comp (mtREV+G+I).	61
Figura 18. Esquema da replicação assimétrica no mtDNA de insetos.	65
Figura 19. Consenso das relações entre famílias de Diptera.	71

ABREVIATURAS

A – Adenina

aa – aminoácido

ADP – Adenosina difosfato

ATP – Adenosina trifosfato

atp6, atp8 – Subunidades 6 e 8 da ATP sintetase F₀

BI – Inferência bayesiana (“bayesian inference”)

BSC – Bloco de sequência conservado

C – Citosina

cox1, cox2, cox3 – Subunidades 1, 2 e 3 da citocromo oxidase, respectivamente

cytb – Citocromo b

CAD – Proteína fusionada codificadora da carbamilmfosfato sintetase, aspartato transcarbamilase e dihidroorotase

DNA – Ácido desoxiribonucléico

EF-1 α – Fator de alongação 1 α

EtOH – Etanol

G – Guanina

GCPs – Genes codificadores de proteínas

H3 – Histona H3

IPTG – Isopropyl-beta-D-thiogalactopyranoside

ITS – Espaçador transcrito interno (“internal transcribed spacer”)

kb – Kilobases

LB – Luria Bertani

Long-PCR – Reação em cadeia da polimerase de longa extensão

ML – Máxima verossimilhança (“maximum likelihood”)

MP – Máxima parcimônia

mtDNA – DNA mitocondrial

nd1-nd6 – Subunidades 1 a 6 da NADH desidrogenase

NJ – “Neighbor-joinig”

nt – Nucleotídeo

ORF – Quadro aberto de leitura (“open reading frame”)

pb – Pares de bases

PCR – Reação em cadeia da polimerase

PCR-RFLP – Reação em cadeia da polimerase - Polimorfismo de tamanho de fragmento de restrição (“polymerase chain reaction - restriction fragment length polymorphism”)

RAPD – Polimorfismos de DNA gerados por amplificação aleatória (“random amplified polymorphic DNA”)

RC – Região controle

RFLP – Polimorfismo de tamanho de fragmento de restrição (“restriction fragment length polymorphism”)

RNA – Ácido ribonucléico

rpm – rotações por minuto

rRNA – RNA ribossomal

rrnS*, *rrnL – Anotação das subunidades 12S (“small”) e 16S (“large”) do RNA ribossomal, respectivamente

RSCU – Utilização relativa de códons sinônimos (“relative synonymous codon usage”)

T - Timina

tRNA – RNA transportador

tRNA^{aa} – Anotação dos genes para tRNAs com o aminoácido correspondente em sobrescrito

RESUMO

O DNA mitocondrial (mtDNA) de três espécies da família Calliphoridae foi completamente sequenciado, apresentando 15004 pb em *Chloroprocta idioidea*, 16143 pb em *Calliphora vomitoria* e 16635 pb em *Phormia regina*. Os três genomas mitocondriais apresentaram a ordem gênica ancestral de Pancrustacea, contendo 13 genes codificadores de proteínas (GCPs), duas subunidades de RNA ribossomal e vinte e dois RNAs transportadores (tRNAs). No entanto, *P. regina* apresentou uma duplicação envolvendo a sequência completa dos genes de $tRNA^{Le}$ e $tRNA^{Gln}$, além da sequência parcial do $tRNA^{Met}$. Uma duplicação similar foi previamente descrita para espécies do gênero *Chrysomya* na mesma localização, inserida no domínio hipervariável da região controle do mtDNA. A composição de nucleotídeos dos mtDNAs sequenciados mostrou um alto viés de bases A e T (71.7% em *C. idioidea*, 72.9% em *C. vomitoria* e 75.6% em *P. regina*), principalmente nas terceiras posições do códon e regiões não-codificadoras, onde o conteúdo A+T é >90%. As reconstruções filogenéticas foram conduzidas com todas as espécies de dípteros com mtDNAs completos, consistindo no estudo mais amplo com sequências de mtDNA da ordem Diptera. O emprego de genes individuais para a reconstrução de filogenias resultou em topologias variadas com baixo suporte, enquanto o uso de sequências concatenadas de nucleotídeos e aminoácidos dos GCPs mostrou resolução para as relações internas de Diptera. A monofilia de Muscomorpha não obteve suporte nas análises apresentadas neste trabalho, assim como a de Acalyptratae. A família Calliphoridae foi recuperada como monofilética, mas as relações internas de Oestroidea recuperaram espécies de Muscoidea como um grupo irmão de Calliphoridae. As relações entre as subfamílias de Calliphoridae indicaram que Luciliinae e Calliphorinae são grupos irmãos, relacionados a Chrysomyinae. Devido à importância médica, veterinária, sanitária, econômica e forense da família Calliphoridae, o esclarecimento das relações evolutivas desta família é interessante para guiar futuros estudos acerca da evolução do parasitismo e do hábito de causar miíases, além de contribuir para o diagnóstico espécie-específico. Além disso, a caracterização de genomas mitocondriais completos também pode contribuir na resolução de filogenias da ordem Diptera e estudos em evolução molecular e divergências antigas de insetos.

ABSTRACT

In this study, we present the complete mitochondrial DNA (mtDNA) sequence of three Calliphoridae (Diptera: Brachycera) species: *Chloroprocta idioidea*, *Calliphora vomitoria* and *Phormia regina*, which had 15004, 16143 and 16635 bp, respectively. Each genome was arranged in the same order described for insects and crustaceans, containing 13 protein coding genes (PCGs), two ribosomal RNA subunits (rRNA) and twenty-two transfer RNA (tRNA), with the exception of *P. regina*, which presented a duplication involving the complete sequences of *tRNA^{Ile}* and *tRNA^{Gln}* genes, besides a partial sequence of the *tRNA^{Met}*. A similar duplication has been previously described for *Chrysomya* species in the same location, inserted in the hypervariable domain of the mitochondrial control region. The nucleotide composition was heavily biased toward As and Ts (71.7% for *C. idioidea*, 72.9% for *C. vomitoria* and 75.6% for *P. regina*), mainly when considering third codon positions and non-coding regions, where the A+T content was >90%. The phylogenetic reconstructions were conducted for all available dipteran species in GenBank, this being the most comprehensive study carried out so far with complete mitochondrial genome sequences. The use of single genes has shown that different topologies were obtained with low support, whereas the use of nucleotide and amino-acid data sets with concatenated PCGs usually provided resolution for intraordinal relationships in Diptera. The monophyly of Muscomorpha was not supported in our analyses, as well as the monophyly of Acalyptratae, which is a major clade of Schizophora. The Calliphoridae was a monophyletic family, but the superfamily Oestroidea was disrupted by the inclusion of Muscoidea species as a sister group of Calliphoridae. Within Calliphoridae, the subfamilies Luciliinae and Calliphorinae were clustered together, related to the Chrysomyinae subfamily. In view of its sanitary, medical, economic and forensic importance, knowledge of Calliphoridae relationships is of interest to guide future works on parasitism evolution of the myiasis habit and specific molecular diagnosis of species. In addition, the characterization of complete mitochondrial sequences could provide insights with regard to dipteran relationships and general molecular evolutionary studies on deep-level phylogenies of insects.

INTRODUÇÃO

I. A mitocôndria e o genoma mitocondrial

A mitocôndria é uma organela citoplasmática presente em células eucarióticas que possui sua própria membrana e seu próprio genoma. Esta organela é responsável pela geração de energia nas células, sendo responsável pela conversão do ADP e fosfato em ATP, através de uma cascata enzimática conhecida como fosforilação oxidativa. A idéia de que esta organela descende de uma bactéria endossimbiótica e teve um papel crucial na evolução dos eucariotos é discutida há um longo tempo (Margulis, 1970). Vários trabalhos revisaram e comprovaram a origem simbiótica da mitocôndria (Gray and Doolittle, 1982) através de comparações de sequências de DNA conservadas de subunidades ribossomais organelares e bacterianas (Cedergren et al., 1988), assim como dos genes citocromo b e citocromo oxidase I (Sicheritz-Ponten et al., 1998). As sequências analisadas agruparam a mitocôndria com um grupo de α -proteobactérias simbióticas, que inclui o gênero *Rickettsia*, com uma estimativa de divergência das linhagens equivalente a 1.5-2 bilhões de anos. Estes resultados são consistentes com a estimativa da origem de eucariotos através de outras fontes.

Através da comparação de genomas sequenciados, Lang et al. (1997) apontaram o genoma mitocondrial do protista *Reclinomonas americana* como o mais similar a genomas de bactérias, enquanto Andersson et al. (1998) compararam o DNA mitocondrial (mtDNA) de *R. americana* com o genoma da bactéria *Rickettsia prowazekii* e mostraram regiões de conservação de ordem e conteúdo gênico, apesar dos mais de 1.5 bilhões de anos de divergência, propondo que *R. prowazekii* pode ser próxima do organismo ancestral que deu origem à mitocôndria de eucariotos.

O mtDNA de animais multicelulares (Metazoa) é uma molécula dupla-fita circular com múltiplas cópias por organela (Wolstenholme, 1992), com cerca de 15 a 20 kb, contendo 37 genes e com herança uniparental materna (Wolstenholme, 1992; Avise, 1994; Ballard and Kreitman, 1995; Boore, 1999). Apesar de existirem descrições de genomas mitocondriais maiores entre os metazoários, estas variações de tamanho devem-se principalmente a duplicações ocasionalmente encontradas e não a variações no número de genes. O conteúdo típico de genomas mitocondriais animais codifica 13 subunidades de proteínas relacionadas a fosforilação oxidativa, 22

RNAs transportadores (tRNAs) necessários para a tradução das proteínas codificadas pelo genoma mitocondrial e duas unidades de RNAs ribossomais (rRNAs) mitocondriais. De um modo geral, o conteúdo gênico apresenta uma grande conservação entre os animais, existindo no entanto algumas exceções (Wolstenholme, 1992; Ladoukakis and Zouros, 2001). O mtDNA animal ainda contém uma região não codificadora - denominada de região controle - que foi objeto de alguns estudos funcionais, nos quais foi possível identificar sítios de início de transcrição para cada cadeia (Chang and Clayton, 1984; Carrodegas and Vallejo, 1997) e uma origem de replicação (Shadel and Clayton, 1997). A região controle é a responsável pela maior parte da variação do genoma mitocondrial tanto em tamanho como em sequência (Clary and Wolstenholme, 1987; Lessinger and Azeredo-Espin, 2000).

O mtDNA é, atualmente, um marcador molecular utilizado em diversos estudos evolutivos, incluindo inferência filogenética, identificação da origem de espécies, filogeografia, genética de populações e evolução molecular (Avice, 1994; Zhang and Hewitt, 1997).

2. O mtDNA como marcador molecular em insetos

Esforços consideráveis têm sido feitos para desenvolver novas classes de marcadores moleculares para emprego na sistemática molecular de insetos uma vez que muitas das relações evolutivas, principalmente inter e intra-ordem, não podem ser resolvidas satisfatoriamente com o espectro de marcadores moleculares atualmente otimizados para o sequenciamento em rotina (Cameron et al., 2007). De acordo com Caterino et al. (2000), somente as subunidades dos genes ribossomais nucleares 18S e 28S, algumas subunidades de histona (H3, principalmente), porções de alguns genes mitocondriais (*cox1*, *cox2*, *cytb*, *rrnL* e *rrnS*) e poucos genes relacionados ao desenvolvimento, como o *wingless* e *Hox*, apresentam um grau de conservação suficiente para serem sequenciados em todas as ordens de insetos. Outros genes, como a opsina, o fator de alongação EF-1 α e o gene CAD (carbamilfosfato sintetase - aspartato transcarbamilase - dihidroorotase), foram utilizados em algumas ordens, mas apresentam dificuldades de amplificação e sequenciamento para serem empregados em um amplo espectro.

Neste contexto, o genoma mitocondrial tem sido sequenciado em várias ordens de animais e caracterizado como um potencial marcador molecular para

resolver relações evolutivas de várias ordens. Duas abordagens principais têm sido aplicadas para análises evolutivas envolvendo o sequenciamento completo do mtDNA: a comparação de rearranjos genômicos e o uso do genoma completo para filogenias baseadas em sequências.

Apesar de sequências mitocondriais evoluírem rapidamente através de uma alta taxa de substituição, a disposição dos genes é relativamente estável dentro dos principais grupos animais, embora rearranjos gênicos sejam descritos (Smith *et al.*, 1993; Boore *et al.*, 1995, Boore *et al.*, 1998; Downton and Austin, 1999; Hwang *et al.*, 2001a; Roehrdanz *et al.*, 2002; Shao and Barker, 2003). Os rearranjos mitocondriais estão mais frequentemente associados à reorganização de genes para tRNAs, sendo que rearranjos envolvendo subunidades rRNAs e genes codificadores de proteínas (GCPs) são muito raros (Rawlings *et al.*, 2001). A utilização da “morfologia genômica” para resolver divergências antigas entre insetos mostrou suporte para alguns grupos, como Orthoptera, Hymenoptera e Phthiraptera (Flook *et al.*, 1995; Downton *et al.*, 2003, Covacin *et al.*, 2006), porém a grande semelhança da ordem gênica na maioria dos genomas sequenciados dificultou a resolução de filogenias. No entanto, o emprego de rearranjos gênicos se mostrou eficiente em análises envolvendo a evolução de artrópodes, demonstrando, por exemplo, a polifilia de Hexapoda (Nardi *et al.*, 2003). O mecanismo de rearranjo gênico tem sido bem investigado em vertebrados e sugere-se que os rearranjos de tRNAs podem ser explicados por desapareamento da origem de replicação, levando a duplicações *in tandem*. Os genes duplicados seriam aleatoriamente deletados, levando à reorganização do genoma mitocondrial (Moritz *et al.*, 1987).

Por outro lado, a utilização de sequências completas de mtDNAs não tem sido extensivamente considerada entre ordens de artrópodes para saber como genes individuais e sequências concatenadas podem resolver questões evolutivas em diferentes *taxa*. O tamanho relativamente pequeno do mtDNA, quando comparado a outros genomas, facilita seu sequenciamento completo, representando ao mesmo tempo um maior número de caracteres do que o da maioria dos genes únicos utilizados atualmente em estudos filogenéticos. Sequências mitocondriais completas foram utilizadas em alguns trabalhos de filogenia de artrópodes, desde análises intraespecíficas com linhagens de *Drosophila melanogaster* (Ballard, 2000a) e *D. simulans* (Ballard 2000b) a estudos mais amplos envolvendo o filo Arthropoda (Nardi *et al.*, 2003; Cameron *et al.*, 2004). Devido ao fato de conter genes para rRNAs,

tRNAs, proteínas e uma região não-codificadora com função regulatória, o mtDNA representa um bom sistema no qual a evolução das diferentes categorias de genes pode ser investigada (Saccone et al., 1999). Logo, o sequenciamento de genomas mitocondriais de diversas espécies pode esclarecer questões sobre a estrutura, função e evolução desta molécula.

Com o desenvolvimento de diversas técnicas de biologia molecular, o volume de dados gerados na forma de sequências de DNA e proteínas aumentou exponencialmente na última década. Aliado à informatização, este desenvolvimento possibilitou o acesso das informações geradas para diferentes organismos através de buscas em grandes bancos de dados públicos. Neste contexto, o sequenciamento de genomas mitocondriais também vem apresentando um aumento proeminente nos últimos anos, o que possibilitou o surgimento de uma nova área de pesquisa: a genômica mitocondrial.

Atualmente existem 1214 sequências de genomas mitocondriais completos de animais, plantas e fungos depositadas nos bancos de dados públicos (RefSeq-GenBank, janeiro de 2008). As sequências relativas a metazoários representam 1137 (93.6%) destas entradas, sendo que o filo Chordata, que inclui a espécie humana, tem 839 genomas depositados e é o mais bem representado. O filo Arthropoda possui a segunda maior representação, com 144 sequências de mtDNA. Entre os artrópodes, o sub-filo Hexapoda, que inclui os insetos, é o que possui o maior número de genomas mitocondriais descritos, com 77 depósitos. As outras divisões de Arthropoda são bem menos representadas: 25 Chelicerata (aranhas, escorpiões, carrapatos e límulo), 6 Myriapoda (centípedes e milípedes) e 36 Crustacea (piolhos, camarões, caranguejos, lagostas, etc.).

O primeiro mtDNA de inseto completamente sequenciado foi da espécie *D. yakuba* (Clary and Wolstenholme, 1985), seguido pelo da *D. melanogaster* (Lewis et al., 1994). A ordem de genes descrita para estas duas espécies demonstrou-se ancestral entre hexápodos e crustáceos (que atualmente formam o grupo Pancrustacea), uma vez que representantes destes grupos possuem a ordem gênica conservada (Nardi et al., 2003), com ocasionais rearranjos, duplicações e translocações de tRNAs (Crozier and Crozier, 1993; Flook et al., 1995; Dowton and Austin, 2003; Junqueira et al., 2004). Rearranjos de rRNAs e GCPs são raros entre insetos, a exemplo do que ocorre em outros grupos animais, e estão restritos a espécies de Paraneoptera (Shao et al., 2003).

Os artrópodes representam a maior diversidade de animais conhecida no planeta, com cerca de $\frac{3}{4}$ da biodiversidade animal descrita, tendo representantes em todos os tipos de ambiente, desde mares profundos a picos de montanhas. Os insetos, em particular, são parte do grupo mais diverso de animais terrestres e as relações entre suas ordens têm sido amplamente estudadas ao longo do tempo (Hennig, 1966; Kristensen, 1981; Wheeler et al., 2001; Whitfield and Kjer, 2008). Com a utilização de dados moleculares, esperava-se que as relações problemáticas entre as ordens pudessem ser decifradas, mas os genes nucleares utilizados inicialmente não tiveram sucesso para discriminar todas as relações (Wheeler et al., 2001; Cameron et al., 2007).

Sugeriou-se, a partir destes resultados, que um conjunto de dados mais amplo, utilizando genomas mitocondriais completos, resultasse em pistas a respeito das relações intra e inter-ordem de hexápodos. No entanto, muitos conjuntos de dados de espécies de insetos apresentam substituições que se acumulam em taxas específicas entre linhagens diferentes ou apresentam viés de composição nucleotídica, que resulta em proporções de bases diferentes entre espécies e entre regiões gênicas. Estes problemas frequentemente se combinam à inclusão de dados com taxas de substituição inapropriadas para a questão abordada, fazendo da filogenia de insetos uma empreitada desafiadora (Whitfield and Kjer, 2008).

É importante, portanto, testar se os dados disponíveis são apropriados para resolver relações evolutivas no nível hierárquico que está sendo analisado, determinando quando diferenças de composição e substituição estão interferindo no sinal filogenético (Whitfield and Kjer, 2008). No entanto, a seleção de genes nem sempre é feita de acordo com sua utilidade filogenética, mas sim baseando-se na facilidade de amplificá-lo e na tradição em utilizar determinados marcadores. Por outro lado, escolher os dados que estão evoluindo a uma taxa apropriada ao nível hierárquico abordado não é uma tarefa fácil, já que problemas como atração de ramos longos (que pode agrupar linhagens de modo irreal, de acordo as taxas de substituição nucleotídica de espécies), homoplasias (múltiplas substituições em um mesmo sítio) e determinar o quão “rápido” e quão “devagar” é desejável que o marcador evolua, ainda precisam ser explorados amplamente entre insetos.

Neste contexto, regiões do genoma mitocondrial foram muito utilizadas em trabalhos envolvendo filogenia de insetos, tornando os genes *cox1* e *cox2* os marcadores mais caracterizados, sequenciados e empregados em estudos evolutivos

(Caterino et al., 2000). Cameron et al. (2007), ao examinarem cerca de 100 genomas mitocondriais completos para inferir a filogenia entre ordens de insetos, mostraram que aumentar o número de *taxa* e o tamanho das sequências analisadas leva a uma melhoria na resolução das topologias, quando comparadas a estudos menores.

Por outro lado, quando múltiplos genes são sequenciados e analisados para estimar a mesma filogenia, espera-se que todos, ou a maioria deles, resulte em uma mesma topologia. No entanto, quando analisa-se espécies que divergiram há um curto espaço de tempo, podem existir polimorfismos ancestrais que não foram amostrados e por isso a árvore obtida com o gene escolhido não coincide com a árvore das espécies. Da mesma maneira, a adequação de modelos de evolução para diferentes partições e a incorporação de parâmetros para diferentes taxas de evolução ao longo das sequências são necessárias para minimizar ruídos e maximizar o sinal filogenético das sequências. Sem dúvida, o volume de sequências de mtDNA de metazoários gerados na última década representa o potencial de aplicações biológicas destes dados. Alguns genes mitocondriais certamente são mais apropriados para reconstruir filogenias do que outros, mas é difícil saber qual ou quais deles são melhores, principalmente quando utiliza-se organismos pouco estudados e caracterizados (Russo et al., 1996; Nei, 1996; Russo et al., 1997; Nardi et al., 2003).

3. A família Calliphoridae (Diptera)

Os dípteros da família Calliphoridae são popularmente conhecidos como moscas varejeiras e podem ser considerados como uma das famílias de insetos mais comuns do mundo, tendo em vista sua ampla distribuição e por ocorrerem nos mais diferentes ambientes (Baumgartner and Greenberg, 1984; Esposito, 1999). Pont (1980) registrou a ocorrência de mais de 1000 espécies de califorídeos no mundo, sendo que 80% dessas e 60% dos 150 gêneros reconhecidos são restritos ao Velho Mundo. A região Afro-tropical possui a fauna mais rica de dípteros califorídeos, com 300 espécies e 40 gêneros (Shewell, 1987). Já na região Neotropical, a fauna de califorídeos inclui menos de 100 espécies.

Juntamente com as famílias Sarcophagidae, Tachinidae (a segunda maior família de dípteros com cerca de 8200 espécies), Oestridae e outras famílias pequenas, como Rhinophoridae e Mystacinobiidae, os califorídeos fazem parte da superfamília Oestroidea (Rognes, 1997). O status taxonômico de Calliphoridae tem

vido alvo de diversas discussões, sendo que em sua última e mais ampla revisão, Rognes (1997) questiona a monofilia desta família, o que alteraria a composição de Oestroidea com a inclusão de pequenos subgrupos. Pape (1992) também argumenta que a monofilia de todas as famílias de Oestroidea são bem suportadas, exceto a de Calliphoridae. As relações entre as famílias de Oestroidea e desta superfamília com as outras pertencentes ao grupo Calyptratae podem ser vistas na Figura 1.

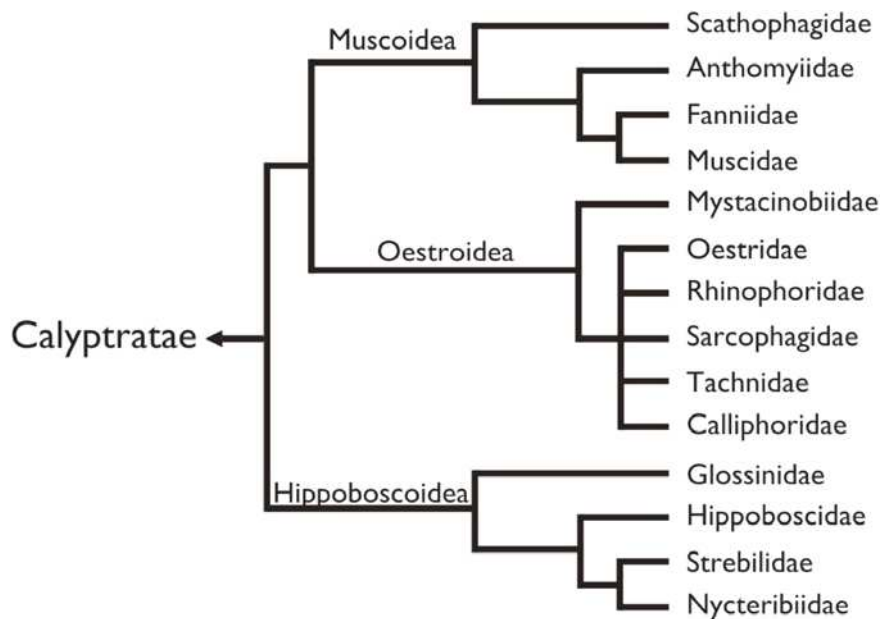


Figura 1: Relações entre as famílias de Oestroidea e as superfamílias de Calyptratae. Fonte: Tree of Life.

As relações entre os grupos de Calliphoridae também são objeto de diversas discussões, com um número de subfamílias que varia entre duas a doze, dependendo do conjunto de dados utilizado (Otranto and Stevens, 2002).

Os califorídeos apresentam grande importância médica, sanitária, econômica e forense por incluir espécies causadoras de miíases. Zumpt (1965) definiu o termo miíase como infestações de vertebrados vivos por larvas de dípteros que, pelo menos por algum período do desenvolvimento, alimentam-se de tecidos vivos ou mortos do hospedeiro, de suas substâncias corporais líquidas ou do alimento por ele ingerido. Espécies causadoras de miíases obrigatórias só podem completar seu ciclo de vida através do parasitismo de hospedeiros vivos, enquanto espécies facultativas podem se desenvolver tanto em matéria orgânica viva quanto em decomposição (Zump, 1965). Deste modo, as miíases primárias são provocadas por larvas que

parasitam feridas recentes em hospedeiros vivos e as secundárias por larvas que atacam os tecidos necrosados e matéria orgânica em decomposição (Hall and Wall, 1995).

Nas áreas tropicais e subtropicais das Américas existem os dois grupos principais de moscas causadoras de miíases pertencentes à família Calliphoridae, incluindo a espécie praga da pecuária *Cochliomyia hominivorax* (mosca da bicheira, responsável por causar miíases primárias), além das espécies causadoras de miíases secundárias.

As famílias Calliphoridae e Sarcophagidae têm espécies obrigatórias que causam miíases de curta duração, na qual as larvas se desenvolvem em feridas ou orifícios corporais durante quatro a sete dias. Já espécies da família Oestridae são parasitas obrigatórios que permanecem no hospedeiro por muitas semanas ou até meses, se desenvolvendo no trato naso-faríngeo, intestino, órgãos internos ou em tecidos subcutâneos (Hall and Wall, 1995). Estas larvas parasitam diversos tipos de animais silvestres e domésticos, representando grandes perdas econômicas por reduzirem a produção de carne e leite, causarem danos ao couro e abortos, além de potencialmente parasitarem seres humanos. Somados ao prejuízo em produção de gado bovino, caprino e suíno, ainda existem os gastos em inseticidas para controle destes parasitas.

As espécies saprófagas da família Calliphoridae podem agir como vetores mecânicos de patógenos por se reproduzirem em lixo, fezes e matéria orgânica em decomposição (Greenberg, 1971), além de ocasionarem miíases secundárias em homens e animais (Zumpt, 1965). As espécies eusinantrópicas transportam patógenos como o poliovírus tipos 1, 2 e 3, *Salmonella*, *Shigella*, *Entamoeba* e ovos de tênia, entre outros (Greenberg, 1971). Devido ao fato de serem os primeiros insetos a detectarem um corpo em decomposição, estas espécies de califorídeos têm sido consideradas de grande importância para a entomologia forense no Brasil (Salviano et al., 1996; Souza & Linhares, 1997) e no mundo (Amendt et al., 2004), representando uma família dominante na colonização faunística pós-morte (Souza and Linhares, 1997; Anderson, 2000; Carvalho et al., 2000; Wells and Sperling, 2001, Stevens and Wall, 2001). Uma vez que o escopo dos estudos em entomologia forense são os insetos associados à decomposição de cadáveres em investigações médico-criminais, a averiguação do estágio de desenvolvimento e da fauna de califorídeos presente em um corpo é de grande importância para estimar o tempo

decorrido após a morte, além de informar sobre possíveis deslocamentos geográficos do corpo e sobre o uso de drogas e substâncias tóxicas ingeridas antes da morte (Amendt, 2004).

4. Estudos moleculares em Calliphoridae

As análises moleculares com califorídeos se concentram em três principais temas: genética de populações, identificação de espécies e filogenia (Azeredo-Espin and Lessinger 2006).

O RFLP (“restriction fragment length polymorphism”) do mtDNA foi utilizado em análises de populações de *C. hominivorax* (Roehrdanz, 1989; Roehrdanz and Johnson, 1988; Infante and Azeredo-Espin, 1995), assim como o PCR-RFLP da região *cox1-cox2* do genoma mitocondrial (Lyra et al., 2005), RAPD (“random amplified polymorphic DNA”; Infante-Malachias et al., 1999) e locus de microssatélites (Torres et al., 2004; Torres and Azeredo-Espin, 2005).

Estudos envolvendo a identificação de espécies através da sequência dos genes *cox1* e *cox2* foram realizados para espécies de importância forense na Austrália (Wallman and Donellan, 2001; Harvey et al., 2003), África do Sul (Harvey et al., 2003) América do Norte (Wells and Sperling, 1999), Japão (Saigusa et al., 2005) e Taiwan (Chen et al., 2004). A diferenciação de *C. hominivorax* e *C. macellaria* por PCR-RFLP da região controle do mtDNA e de *cox1-cox2* foi realizada por Litjens et al. (2001) em populações brasileiras, enquanto Ratcliffe et al. (2003) utilizou PCR-RFLP do ITS1 e ITS2 (“internal transcribed spacer”) do “cluster” de rDNA nuclear para identificar espécies de califorídeos, sarcófagídeos e muscídeos.

A região controle do mtDNA de *C. hominivorax*, *C. macellaria*, *Chrysomya putoria* (= *C. chloropyga*), *C. albiceps*, *C. megacephala*, *Lucilia eximia* e *D. hominis* foi completamente sequenciada (Lessinger and Azeredo-Espin, 2000; Lessinger et al., 2004), caracterizada estruturalmente e testada para avaliar o potencial em estudos filogenéticos. Além disso, Lessinger et al. (2004) e Junqueira et al. (2004) descreveram a duplicação do gene *tRNA^{le}* e suas regiões flangeadoras na região controle das três espécies de *Chrysomya* analisadas (ver Apêndices A e B). Para verificar a presença da duplicação do *tRNA^{le}* em outros califorídeos, o estudo foi ampliado para outras quinze espécies, que também tiveram sua região controle completamente sequenciada e novas duplicações foram encontradas em *Phormia regina*, *Calliphora dubia*, *Chrysomya bezziana* e *Chrysomya rufifacies*. Duarte et al. (2008)

sugerem que a exista um “hot spot” para duplicações de regiões flangeadoras da região controle do mtDNA (ver Apêndice C).

As análises filogenéticas com espécies de Calliphoridae são poucas e envolvem a utilização da subunidade ribossomal 28S do DNA nuclear (Stevens et al., 2002) e os genes mitocondriais *cox1*, *cox2*, *nd4* e *nd4L* (Wallman et al., 2005). A análise filogenética de moscas de importância forense do gênero *Lucilia* foi realizada com o gene *cox1* (Wells et al., 2007). Recentemente, a sequência primária e a estrutura secundária do ITS2 foram caracterizadas e utilizadas em análises filogenéticas da família Calliphoridae (Marinho et al.; submetido. Ver Apêndice D).

Além destes trabalhos, na última década também surgiram as primeiras sequências de genomas mitocondriais completos de espécies de califorídeos, caracterizando-os estruturalmente e utilizando o concatenado de PCGs para inferências filogenéticas (Lessinger et al., 2000; Junqueira et al., 2004 – Apêndice B).

5. Genômica mitocondrial da família Calliphoridae: aplicações e perspectivas

A caracterização de genomas mitocondriais completos está bem estabelecida para vertebrados, mas ainda é incipiente entre os invertebrados. Uma das razões para isso é que muitas das espécies de invertebrados são relativamente pequenas, dificultando a obtenção de quantidade suficiente de mtDNA purificado para o sequenciamento por exigir um grande número de indivíduos (Hwang et al., 2001b). Com a padronização da técnica denominada PCR de longa extensão (“long-PCR”, do inglês “long polymerase chain reaction”), que consiste na amplificação de longos fragmentos de DNA, é possível obter a amplificação do genoma mitocondrial completo de um único indivíduo, em poucos fragmentos, a partir de pequenas quantidades de DNA. Esta abordagem permite que o genoma mitocondrial seja utilizado em diversos estudos de evolução molecular, genética de populações, filogenia molecular, RFLP e análise de rearranjos gênicos (Nelson et al., 1996; Hwang et al., 2001b). Portanto, o estabelecimento de uma metodologia otimizada para a amplificação completa do mtDNA por “long-PCR” de diferentes invertebrados é de extrema importância para o desenvolvimento e aplicação da genômica mitocondrial em estudos genético-evolutivos destes organismos.

Atualmente existem quatro espécies de califorídeos com genomas mitocondriais sequenciados, publicados e depositados no GenBank: as espécies de

importância forense *Chrysomya putoria* (AF352790, Junqueira et al., 2004) e *C. megacephala* (sequência parcial, acesso AJ426041, Stevens et al., 2008), a praga da pecuária *C. hominivorax* (AF260826, Lessinger et al., 2000) e a espécie facultativa (geralmente causa miíase secundária, mas já foi descrita causando miíase primária) *Lucilia sericata* (AJ422212, Stevens et al., 2008).

Devido à importância da família Calliphoridae como representante de espécies-praga e de importância médica, sanitária e forense, a obtenção de sequências do mtDNA pode gerar conjuntos de dados para esclarecer relações evolutivas deste grupo, uma vez que as revisões taxonômicas desta família foram feitas basicamente com dados morfológicos (McAlpine, 1989; Rognes, 1997; Pape, 1992) e apresentam algumas divergências de filogenias obtidas com dados moleculares (Nirmala et al., 2001; Stevens et al., 2002; Wallman et al., 2005). A disponibilidade de sequências completas do mtDNA de califorídeos permite a análise e comparação dos diferentes genes mitocondriais num contexto evolutivo, visando caracterizar comparativamente estes genomas e contribuindo para a elucidação dos diferentes padrões de organização desta molécula não só na família Calliphoridae como também na ordem Diptera. Da mesma maneira, torna-se possível avaliar reconstruções filogenéticas a partir de genes individuais e do concatenado das sequências nucleotídicas e de aminoácidos, criando critérios comparativos para utilizar as diferentes partições de dados para reconstruções filogenéticas.

Com o sequenciamento de genomas mitocondriais de califorídeos, é possível acessar o conteúdo de bases nas diferentes fitas codificadoras, a utilização preferencial de códons, a relação entre transições e transversões, a frequência de aminoácidos, a quantidade de substituições sinônimas e não-sinônimas, criando embasamento para a escolha de sequências para estudos evolutivos. A caracterização molecular dos diferentes genes, associada à utilização diferentes algoritmos para reconstrução de árvores, pode indicar se todos ou a maioria deles apontam para uma mesma topologia. Deste modo, pode-se inferir se as relações obtidas estão sendo geradas por um sinal filogenético e não por um ruído. A geração de sequências do genoma mitocondrial de espécies de califorídeos ainda pode esclarecer a ocorrência de uma duplicação gênica do *tRNA^{le}* encontrada nas espécies *C. putoria* (Junqueira et al., 2004), *C. megacephala* e *C. albiceps* (Lessinger et al., 2004), ajudando no diagnóstico molecular de espécies de importância forense, além de

contribuir nos estudos a respeito dos rearranjos e duplicações no mtDNA de insetos.

Com a intenção de abranger diferentes gêneros e subfamílias de Calliphoridae, as espécies *Chloroprocta idioidea* (Robineau-Devoidy, 1830; subfamília Chrysomyinae), *Calliphora vomitoria* (Linnaeus, 1758; subfamília Calliphorinae) e *Phormia regina* (Meigen, 1826; subfamília Chrysomyinae) foram escolhidas para ter seus genomas mitocondriais completamente sequenciados. Estas espécies representam diferentes gêneros de duas subfamílias de califorídeos, possuindo importância médica, sanitária e forense por se alimentarem e se reproduzirem em matéria orgânica em decomposição (Guimarães and Papavero, 1999).

C. idioidea é uma espécie pouco caracterizada, de bionomia desconhecida, geralmente descrita em estudos de distribuição e ocorrência de dípteros em ambientes naturais, como fragmentos de Mata Atlântica, mata ciliar e áreas de preservação ambiental (Furusawa and Cassino, 2006). Registros de coleta desta espécie foram feitos em Mato Grosso do Sul (Koller et al., 2002), Rio de Janeiro (Furusawa and Cassino, 2006) e São Paulo (A. X. Linhares, comunicação pessoal), sempre em pequena quantidade e em áreas rurais ou de vegetação densa.

A espécie *P. regina* foi muito utilizada para uma técnica medicinal de desinfecção de tecidos necrosados em casos de queimaduras e amputações, denominada de terapia larval. As larvas de espécies causadoras de miíases secundárias secretam uma enzima preoteolítica que ajuda na remoção do tecido morto, ao mesmo tempo que controlam naturalmente a proliferação de bactérias. *P. regina* não é uma espécie de distribuição tropical e portanto nunca foi reportada no Brasil, sendo comum em áreas temperadas (Zumpt, 1965; Greenberg, 1971).

O gênero *Calliphora* é predominantemente paleártico e apenas a espécie *C. vicina* foi reportada no sul do Brasil (Guimarães and Papavero, 1999). A espécie *C. vomitoria* tem pouco grau de associação com centros urbanos, mas é uma das espécies mais coletadas em corpos em decomposição no Reino Unido (Ames et al., 2006) e na Europa. O espécime de *C. vomitoria* é particularmente interessante no desenvolvimento deste trabalho, uma vez que foi enviada fixada em Etanol 96%, após alguns meses de sua coleta, resultando no acesso à informação genética de um inseto preservado em condições sub-ótimas.

Através do sequenciamento do mtDNA destas três espécies e da utilização dos genomas mitocondriais de *C. hominivorax*, *C. putoria* e *L. sericata*, haverá uma

abrangência de seis diferentes gêneros de califorídeos com três subfamílias representadas (Chrysomyinae, Calliphorinae e Lucillinae). Associadas às sequências de dípteros atualmente disponíveis, um total de 28 genomas mitocondriais completos foram analisados para que as comparações fossem mais diversas e robustas.

Sem dúvida, o estudo da genômica mitocondrial pode resultar em contribuições significativas para o entendimento tanto da evolução de organismos como da evolução de genomas, além de potencializar o uso de marcadores moleculares para a filogenia deste grupo de moscas causadoras de miíases e, em uma escala mais ampla, de insetos.

OBJETIVOS

1. Padronizar a metodologia de PCR de longa extensão (“long-PCR”) do mtDNA para diferentes espécies da família Calliphoridae;
2. Sequenciar e caracterizar o mtDNA completo das espécies *Chloroprocta idioidea*, *Calliphora vomitoria* e *Phormia regina*, todas pertencentes à família Calliphoridae;
3. Analisar os genomas mitocondriais das espécies de califorídeos sequenciados neste trabalho e dos já descritos, realizando uma abordagem comparativa da organização, estrutura, utilização preferencial de códons e evolução das sequências de nucleotídeos e aminoácidos de diferentes regiões funcionais da molécula;
4. Inferir a filogenia da família Calliphoridae a partir de diferentes regiões mitocondriais e métodos de reconstrução filogenética, dando ênfase à utilização de diversas partições de dados, como genes individuais e concatenados de sequências, utilizando também sequências da ordem Diptera disponíveis em bancos de dados públicos;
5. Fornecer marcadores moleculares que podem ser utilizados em estudos evolutivos e no diagnóstico espécie-específico de moscas de importância forense, contribuindo para a caracterização destas espécies no Novo Mundo;
6. Investigar a duplicação do *tRNA^{le}* presente no gênero *Chrysomya* através da comparação de sequências com outros gêneros da família Calliphoridae.

MATERIAIS E MÉTODOS

1. Coletas e amostras

As coletas de califorídeos foram realizadas com armadilhas descritas por Linhares (1979, 1981), consistindo em latas pintadas de preto com orifícios na base para a entrada de moscas atraídas pela isca em seu interior. Um cone de cartolina preta e um saco plástico foram presos à parte superior da lata, onde as moscas ficaram presas durante o intervalo de coleta. Ao saco plástico foi amarrado um barbante, que possibilitou que as armadilhas fossem penduradas em galhos de árvores a 60 cm do chão. Além da utilização das armadilhas, também foram realizadas coletas com rede entomológica diretamente nas iscas expostas. Como isca utilizou-se fígado bovino ou peixe, ambos em decomposição. Os adultos foram identificados em microscópio estereoscópico Zeiss STEMI SV8 e acondicionados em tubos para preservação em freezer -70°C.

As coletas foram feitas na Estação Ecológica de Mogi-Guaçu (SP), uma reserva ambiental de mata atlântica e cerrado, e no campus da Universidade Estadual de Campinas (SP), uma área urbana. Realizou-se coletas em dois ambientes diferentes para tentar coletar uma maior diversidade faunística e desta maneira coletar a espécie *C. idioidea*, que normalmente não é encontrada em ambientes antrópicos.

As amostras recebidas de outras regiões do Brasil e de outros países foram enviadas pelo correio em etanol (EtOH) 96% e acondicionadas em freezer -20°C após o recebimento, já identificadas. A espécie *C. vomitoria* foi obtida desta maneira, proveniente de Londres (Reino Unido), enquanto que *P. regina* teve seu DNA liofilizado e enviado pelo correio de West Virginia (Estados Unidos).

2. Extração de DNA

O DNA total de espécies de califorídeos recém-coletadas e de amostras preservadas em freezer -70°C foi extraído de acordo com o protocolo de fenol/clorofórmio (Infante and Azeredo-Espin, 1995), no qual os espécimes foram homogeneizados em 100 µl de tampão de maceração pH7.5 (2M Tris, 5M NaCl, 50% sacarose, 0.5% EDTA) e a seguir lisados com 100 µl de tampão de lise pH 9.0 (2M Tris, 10% SDS, 0.5% EDTA, 2% DEPC). Após 15 minutos de incubação em gelo, igual volume de fenol pH 7.6 foi adicionado, misturado em vortex, incubado em gelo por

2 minutos e os tubos foram centrifugados a 2600 X g por 5 minutos. O sobrenadante foi transferido para um novo tubo e este mesmo passo se repetiu por duas vezes, adicionando-se igual volume de fenol/clorofórmio : álcool isoamílico 24:1 no segundo passo e somente clorofórmio : álcool isoamílico 24:1 no último passo. Após a última centrifugação, 100 µl de 1 X TE, 0.05 volumes de acetato de Sódio 3M (~20 µl) e 2.5 volumes de etanol absoluto gelado (~900 µl) foram adicionados ao sobrenadante. A seguir, os tubos foram incubados a -20°C por duas horas para precipitar o DNA e então centrifugados por 30 minutos a 15000 X g. O sobrenadante foi descartado e o “pellet” diluído em 50 µl de 1 X TE (Tris-HCl 1 mM, EDTA 0.1 mM). Todas as extrações foram mantidas em “freezer” a -20°C até a utilização na reação de PCR.

Os espécimes preservados em EtOH 96% tiveram seu DNA extraído através do reagente DNAzol[®], que demonstrou-se mais eficiente na recuperação e amplificação do DNA proveniente de amostras preservadas (Junqueira *et al.*, 2002). Os adultos foram lavados em água destilada e tiveram suas asas retiradas para minimizar fontes de contaminação. A seguir, as amostras foram separadas individualmente em microtubos de 1,5 ml contendo 0,5 ml de DNAzol[®] e macerados com homogeneizador. Em seguida adicionou-se 0,5 ml de DNAzol[®] e misturou-se por inversão. O homogeneizado foi sedimentado por centrifugação em microcentrífuga a 16.000 X g por 10 minutos a temperatura ambiente e o sobrenadante foi transferido para um novo microtubo de 1,5 ml. A precipitação do DNA foi feita a partir da adição de 0,5 ml de EtOH 100% gelado e cada amostra foi misturada por inversão. Após 5 minutos em temperatura ambiente as amostras foram submetidas a uma nova centrifugação a 16.000 X g por 2 minutos a temperatura ambiente. O “pellet” de DNA obtido foi lavado com a adição de 1,0 ml de EtOH 95% e submetido à centrifugação a 16.000 X g por 1 minuto a temperatura ambiente. Cada “pellet” foi então ressuscitado em 50µl de 1 X TE (1 mM Tris-HCl, 0,1 mM EDTA) e acondicionado em “freezer” -20°C para futuras análises.

3. Amplificação por PCR de longa extensão (“long-PCR”)

A padronização da reação de “long-PCR” em espécies da família Calliphoridae baseou-se inicialmente na escolha de oligonucleotídeos universais que se hibridam em regiões conservadas do genoma mitocondrial de dípteros. Entretanto,

oligonucleotídeos mais longos foram necessários para garantir a estabilidade das reações juntamente com uma combinação de “primers” universais (Simon et al., 1994). Os “primers” utilizados para a obtenção do mtDNA de *C. idiidea*, *P. regina* e *C. vomitoria* foram desenvolvidos para a espécie *Haematobia irritans* (Diptera: Muscidae; Barau et al., 2005), amplificando o genoma mitocondrial completo destas espécies em dois fragmentos sobrepostos de aproximadamente 8000 pb e 9200 pb. A sequência dos oligonucleotídeos e a região de hibridação dos mesmos estão indicadas na Tabela 1, enquanto o esquema de amplificação em duas etapas (“two-steps”) do mtDNA está na Figura 2. Deste modo, a combinação dos “primers” HI6S-A e mtD-22 foi empregada na amplificação do amplicon A, enquanto os “primers” HI6S-B e mtD-19 foram utilizados na obtenção do amplicon B. Entretanto, devido a problemas de amplificação de regiões inespecíficas, a espécie *C. vomitoria* teve seu amplicon A amplificado com dois oligonucleotídeos longos (HCO3-A/HI6S-B), descritos por Barau et al. (2005), garantindo uma maior estabilidade e especificidade na reação.

As reações de “long-PCR” foram integralmente conduzidas em gelo, para um volume final de 50 µl contendo 400 mM de cada “primer”, 300 µM de cada dNTP, 2 mM de MgSO₄, 60 mM de Tris-SO₄ (pH 9.1), 18 mM de (NH₄)₂SO₄ e 2 U de Elongase® (Invitrogen), que contém uma combinação das enzimas Taq DNA polimerase e outras DNA polimerases termo-resistentes de *Pyrococcus sp.* Para cada reação, cerca de 50 ng de DNA foram adicionados. O termociclador foi programado para uma desnaturação inicial a 94°C por 1 minuto, 39 ciclos de desnaturação a 94°C por 20 segundos, hibridação a 48°C por 1 minuto e extensão das cadeias de DNA a 60°C por 16 minutos. Um ciclo adicional de extensão a 60°C por 10 minutos foi adicionado ao final do programa.

Todas as reações foram conduzidas com um controle negativo para garantir que não houve contaminação no procedimento. Este controle consistiu em uma reação contendo todos os reagentes usados para a amplificação, porém sem DNA. Este procedimento foi particularmente importante para a amplificação de amostras preservadas em EtOH 96%, garantindo que o DNA obtido era proveniente do espécime preservado e não de alguma fonte exógena de DNA que pudesse contaminar os reagentes.

Tabela 1: Sequência dos oligonucleotídeos empregados nas reações de “long-PCR” de espécies de califórídeos e suas respectivas regiões de hibridação no mtDNA.

Primer	Sequência 5'→ 3'	Região de hibridação	Amplicon (tamanho)
mtD-22*	TCAACAAAGTGTCAGTATCA	<i>cox3</i>	A (8000 pb)
HI6S-A†	GCTACCTTTGCAGAGTTAGGATACTGCGGCC	<i>rrnL</i>	
HCO3-A†	TTCTGAAAAAATAAATAAAATTATTCTCA	<i>cox3</i>	
mtD-19*	GAAATTTGTGGAGCAAATCATAG	<i>cox2</i>	B (9200 pb)
HI6S-B†	CATATCGATAAAAAAGATTGCGACCTCGATGTTG	<i>rrnL</i>	

† - primers desenhados por Barau et al. (2005). * - primers desenhados por Simon et al. (1994).

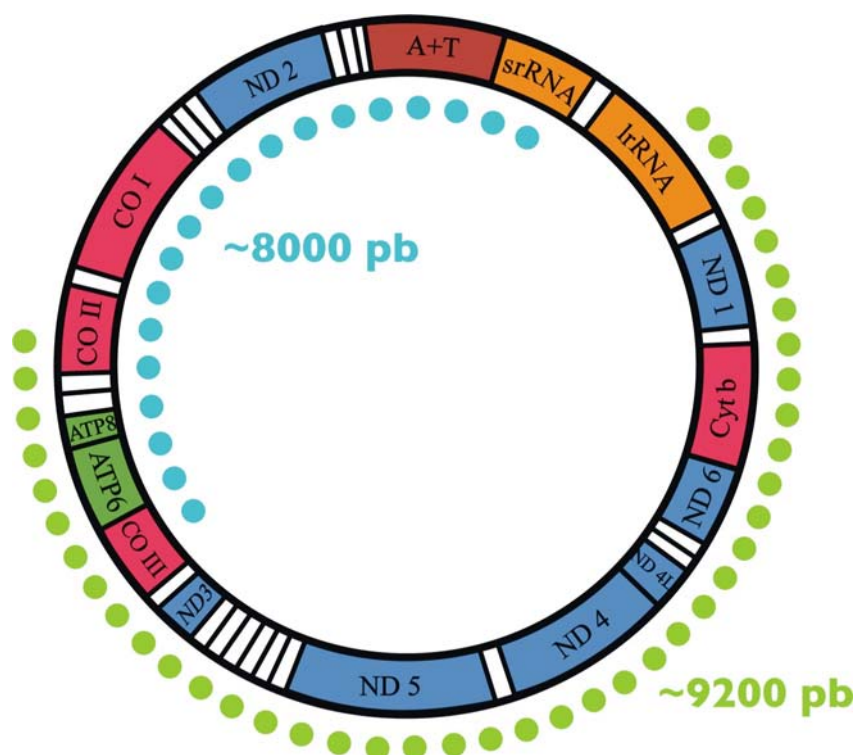


Figura 2: Esquema de amplificação com os oligonucleotídeos utilizados para a amplificação de dois fragmentos longos do mtDNA. Em círculos verdes está representado o amplicon A e em círculos azuis, o amplicon B.

4. “Shotgun”: quantificação e sonicação dos amplicons

Os fragmentos amplificados por “long-PCR” foram visualizados em gel de agarose 1.0% corado com Brometo de Etídeo e quantificados em sistema de análise digital EDAS (Kodak) através da comparação da fluorescência das bandas obtidas no gel com as bandas do marcador de peso molecular λ /HindIII e 1 Kb Plus DNA Ladder, previamente conhecidas e inseridas no programa de quantificação digital.

Após a quantificação, reações de PCR de cada fragmento foram misturadas, totalizando cerca de 500 μ l de solução para a precipitação do mtDNA através da adição de 0.05 volumes de Acetato de Sódio 3M e 2.5 volumes de EtOH absoluto gelado. Após incubação por 2 horas a 20°C, centrifugou-se por 30 minutos a 12000 rpm e o “pellet” foi ressuscitado em 50 μ l de 1 X TE (Tris-HCl 1 mM, EDTA 0.1 mM).

O DNA precipitado foi dividido em alíquotas de 10 μ l para a sonicação. Neste passo, a padronização dos tempos de sonicação foi essencial para a obtenção da maior quantidade possível de fragmentos aleatórios de mtDNA entre 1 e 2 kb para a construção da biblioteca de mtDNA. Bibliotecas com fragmentos abaixo de 1 kb podem resultar no sequenciamento redundante de regiões, dificultando a montagem do genoma mitocondrial. Da mesma maneira, bibliotecas com fragmentos muito grandes podem gerar muitos “gaps” na finalização do genoma. Para a sonicação foi utilizado o sonicador Ultrasonic Homogenizer 4710 Series (Cole Parmer) com o copo de infusão das amostras em gelo e a fonte das ondas de ultra-som no centro. As amostras foram encostadas na fonte de ultra-som e sonicadas por 5, 10 e 15 segundos em intensidade máxima.

5. Extração do mtDNA sonicado do gel de agarose de baixo ponto de fusão

O gel de agarose 0.8% de baixo ponto de fusão (LMP, do inglês “low-melting point”) foi realizado com todo o mtDNA sonicado, afim de recuperar os fragmentos para a condução da biblioteca. O gel foi corrido a 50V a 4°C e a seguir as bandas correspondentes ao intervalo entre 1 e 2 kb foram cortadas do gel e purificadas com fenol e dietil-éter hidratado, de acordo com o protocolo de Sambrook (1989).

Uma vez retirados do gel, os fragmentos foram purificados em colunas com o kit QIAquick (Qiagen), concentrados através da redução de volume com o

liofilizador Vacufuge™ (Eppendorf®) e quantificados em espectrofotômetro GenQuant (Pharmacia-Biotech).

6. Biblioteca dos genomas mitocondriais e seleção de clones recombinantes

Para a condução das bibliotecas de mtDNA, utilizou-se o "kit" de clonagem para fragmentos com extremidades abruptas "Blunt Ended Cloning Kit" (GE Healthcare), com o vetor de clonagem pMOSBlue. Aproximadamente 60 ng de mtDNA purificados do gel de agarose LMP foram utilizados na reação com a polinucleotídeo kinase, que torna as extremidades dos fragmentos sonificados abruptas e os fosforila em um único passo para que a ligação com a enzima T4 DNA ligase seja conduzida. A transformação térmica foi feita em 20 µl de células competentes da linhagem MOSBlue, de acordo com o protocolo sugerido pelo fabricante. As células transformadas foram plaqueadas em meio LB (Luria-Bertani) sólido contendo IPTG (100mM), X-Gal (20 mg/ml) e Ampicilina (50 µg/ml), e incubadas a 37°C invertidas "overnight".

Os clones com a coloração branca foram repicados em placa de segurança e inoculados em meio LB líquido contendo Ampicilina (50 µg/ml). Para isolamento do DNA plasmidial utilizou-se o protocolo de lise alcalina (Sambrook, 1989), seguido de digestão com 1 unidade das enzimas *EcoRI* e *HindIII* por 2 horas a 37°C, para verificação da qualidade da biblioteca e escolha dos clones para sequenciamento.

7. Sequenciamento, montagem e anotação dos genomas mitocondriais

O sequenciamento automático foi feito em sequenciador 377 (gel) e 3700 (capilar) da ABI Prism (Perkin Elmer). As reações foram feitas com o kit "Big Dye Terminator" (ABI-Perkin Elmer), contendo cerca de 500 ng do DNA plasmidial extraído previamente, 3.2 pmol do oligonucleotídeo T7 (5'-TAATACGACTCACTATAGGG), 1 µl de BigDye, 3 µl de tampão para sequenciamento (200mM Tris pH 9.0, 5mM MgCl₂), para um volume final de 10 µl por reação. As reações de sequenciamento foram conduzidas em um termociclador PE 9600 (Perkin-Elmer) programado para uma desnaturação inicial a 96°C por 1 minuto e 30 segundos, seguida de 40 ciclos a 96°C por 12 segundos, 50°C por 6

minutos e 60°C por 4 minutos. Apenas alguns clones escolhidos para fechamento de “gaps” ou melhoria da qualidade das pontas dos contigs montados tiveram o sequenciamento realizado com o oligonucleotídeo reverso U19 (5'-ACGTCGTGACTGGGAAAA).

Os cromatogramas provenientes do sequenciador automático foram analisados em um aplicativo desenvolvido no Laboratório de Genética Animal (CBMEG-Unicamp), que consiste em uma interface gráfica que processa os dados dos cromatogramas em um servidor e interpreta os resultados de qualidades de bases com o programa phred (Ewing et al., 1998), originalmente desenvolvido para utilização em plataforma Unix. Esta ferramenta gráfica está disponível no “website” do banco de dados “Arthropodan Mitochondrial Genomes Accessible Database” (AMiGA), no endereço <http://amiga.cbmeg.unicamp.br> (Feijão et al., 2006). A ferramenta para análise de cromatogramas pode ser acessada no link “Tools” da página inicial do AMiGA e é um aplicativo independente do banco de dados, criado para analisar a qualidade de sequências em qualquer sistema operacional, apenas com a utilização de um navegador de internet.

Após a avaliação das sequências, utilizou-se o programa FinchTV versão 1.3.1 (Geospiza Inc.) para visualizar os cromatogramas, editar manualmente eventuais sítios com sobreposição de picos de leitura e verificar o sinal das bases incorporadas. As sequências de boa qualidade (> 20) e com sinal de leitura acima de 150 foram utilizadas para a montagem do genoma mitocondrial com o pacote de programas phred/phrap/consed (Gordon et al., 1998), disponível para a plataforma Unix. Este pacote de programas é capaz de agrupar as sequências de acordo com a similaridade, formando contigs a partir dos diversos “reads” provenientes do sequenciamento automático. Por sua vez, os contigs formados também foram submetidos à análise de qualidade na interface do AMiGA, para garantir que toda a sequência montada tivesse qualidade >40 e que pelo menos três “reads” passassem por cada região montada do mtDNA das espécies analisadas.

Os genes das subunidades de rRNA foram anotados de acordo com a similaridade encontrada pelo Blastn (Altschul et al., 1990) e por comparação da localização dos mesmos em outros genomas já descritos. A anotação dos genes para tRNAs foi feita com base nas estruturas secundárias encontradas pelo programa tRNA-scanSE (Lowe and Eddy, 1997), que é capaz de prever regiões na sequência

primária de nucleotídeos que têm predisposição a formar a estrutura secundária típica de "clover-leaf" dos tRNAs.

Os GCPs foram anotados com o programa ORF Finder, que procura por quadros abertos de leitura (ORF, do inglês "open reading frames") nas sequências nucleotídicas submetidas à análise, considerando regiões que possuem um códon de iniciação, um códon de terminação e trincas de bases "in frame", que poderiam resultar em proteínas. Cada uma das ORFs encontradas foi submetida ao aplicativo Blastx (Altschul et al., 1990), capaz de fazer buscas por sequências protéicas utilizando a tradução das sequências nucleotídicas com a tabela de código genético para genomas mitocondriais de invertebrados. Após a análise da similaridade encontrada em cada uma das ORFs, os GCPs foram anotados juntamente com os rRNAs e tRNAs utilizando-se o programa Sequin 7.90, disponibilizado pelo NCBI (<ftp://ftp.ncbi.nih.gov/sequin>).

A anotação da região controle das espécies analisadas foi conduzida através da comparação da sequência primária e localização no genoma mitocondrial com as espécies de califorídeos *Cochliomyia hominivorax* e *Chrysomya putoria*, previamente sequenciadas (Lessinger et al., 2000; Junqueira et al., 2004).

8. Análise das sequências

Todas as sequências utilizadas para as análises filogenéticas e comparativas foram obtidas através do banco de dados para genomas mitocondriais completos de artrópodes AMiGA (Feijão et al., 2006) e estão listadas na Tabela 2. Este banco realiza "updates" diários com o NCBI, além de fornecer ferramentas que facilitam a agregação de sequências de genes individuais, permitir a visualização da ordem e conteúdo gênico dos genomas mitocondriais e realizar algumas análises sobre conteúdo dos mtDNAs dentre as espécies selecionadas.

As sequências dos GCPs foram alinhadas com o programa ClustalX 1.83 (Thompson et al., 1997) e MAFFT 6 (Katoh et al., 2005), enquanto a edição dos alinhamentos foi feita com o CLC Free Workbench 4.5 (www.clc.com). As análises de conteúdo nucleotídico, codon usage, relações entre transversões e transições das sequências alinhadas e frequência de aminoácidos nas proteínas foram realizadas com os programas MEGA 4 (Tamura et al., 2007), DAMBE 5.0.5 (Xia and Xie, 2001) e DNAsp 4.0 (Rozas et al., 2003).

Tabela 2: Espécies de dípteros utilizadas nas análises comparativas, com os respectivos números de acesso e grupos taxonômicos a que pertencem.

# Acesso	Espécie					
AMI00076	<i>Chloroprocta idioides</i>	Calliphoridae	Calyptratae	Muscomorpha	Brachycera	
AMI00079	<i>Calliphora vomitoria</i>					
AMI00077	<i>Phormia regina</i>					
NC_002697	<i>Chrysomya putoria</i>					
NC_002660	<i>Cochliomyia hominivorax</i>					
AMI00078	<i>Chrysomya megacephala</i>					
NC_009733	<i>Lucilia sericata</i>					
NC_006378	<i>Dermatobia hominis</i>					Oestridae
NC_007102	<i>Haematobia irritans</i>					Muscidae
AMI00067	<i>Stomoxys calcitrans</i>					
NC_005779	<i>Drosophila mauritiana</i>	Drosophilidae	Acalyptratae	Muscomorpha	Brachycera	
NC_001709	<i>Drosophila melanogaster</i>					
NC_005780	<i>Drosophila sechellia</i>					
NC_005781	<i>Drosophila simulans</i>					
NC_001322	<i>Drosophila yakuba</i>					
NC_009772	<i>Bactrocera carambolae</i>	Tephritidae	Acalyptratae	Muscomorpha	Brachycera	
NC_008748	<i>Bactrocera dorsalis</i>					
NC_005333	<i>Bactrocera oleae</i>					
NC_009770	<i>Bactrocera papayae</i>					
NC_009771	<i>Bactrocera philippinensis</i>					
NC_000857	<i>Ceratitis capitata</i>					
NC_008755	<i>Trichophthalma punctata</i>					
NC_008754	<i>Simosyrphus grandicornis</i>	Nemestrinidae				
NC_008756	<i>Cydistomyia duplonotata</i>	Syrphidae				
NC_006817	<i>Aedes albopictus</i>	Tabanidae		Tabanomorpha		
NC_010241	<i>Aedes aegypti</i>	Culicidae			Nematocera	
NC_002084	<i>Anopheles gambiae</i>					
NC_000875	<i>Anopheles quadrimaculatus</i>					

Nota: As sequências com número de acesso iniciando com “AMI” estão disponíveis no banco de dados AMIGA, uma vez que não estão disponíveis publicamente como sequências completas de mtDNA. As sequências geradas neste trabalho podem ser acessadas através de “login” neste banco de dados (ver item I da Discussão).

As reconstruções filogenéticas com os métodos de máxima parcimônia e máxima verossimilhança foram feitas com o programa PAUP* versão 4.0 beta 10 (Phylogenetic Analysis Using Parsimony and other methods; Swofford, 2000), enquanto as reconstruções com métodos de distância utilizando neighbor-joining (NJ; Saitou and Nei, 1987) foram realizadas com o programa MEGA 4 (Tamura et al., 2007). As análises com estatística bayesiana foram conduzidas com o programa MrBayes v.3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) e as reconstruções utilizando máxima verossimilhança foram feitas com o programa PHYML 2.4.4 (Guindon and Gascuel, 2003).

A escolha dos modelos de substituição nucleotídica e da matriz de substituição de aminoácidos para as análises bayesianas e com máxima

verossimilhança foi conduzida com os programas ModelTest 3.7 (Posada and Crandall, 1998) e ProtTest 1.4 (Abascal et al., 2005), respectivamente.

As análises filogenéticas foram realizadas com a sequência nucleotídica e de aminoácidos individual dos GCPs e também do concatenados delas. Um conjunto de dados inicial, contendo a sequência concatenada de aminoácidos dos 13 GCPs (aa_comp), que foi traduzida a partir do alinhamento das sequências nucleotídicas, foi utilizada nas primeiras reconstruções. Paralelamente, a sequência nucleotídica deste mesmo conjunto de dados foi também usada em reconstruções filogenéticas, mas levando em consideração apenas as primeiras e segundas posições do códon (nt_12), uma vez que as terceiras posições apresentam alta variação e saturação, possuindo um grande número de Adeninas (A) e Timinas (T).

Em segunda instância, foram realizadas reconstruções filogenéticas utilizando apenas os genes que apresentaram uma relação C/V>I (sítios conservados/sítios variáveis) para o alinhamento de aminoácidos, indicando que a sequência apresenta mais sítios conservados do que variáveis. Esta abordagem foi adotada devido ao fato de haver a inclusão de 28 espécies de dípteros no conjunto de dados, garantindo que divergências mais antigas pudessem ser analisadas com maior confiabilidade e menos homoplasias. Deste modo, outros dois sub-conjuntos de dados foram analisados, consistindo nas sequências concatenadas de aminoácidos (sub_aa) e de nucleotídeos (subnt_12).

RESULTADOS

1. Extração de DNA

As amostras recém-coletadas ou preservadas em freezer -70°C tiveram seu DNA eficientemente extraído com o método de Fenol/Clorofórmio, de acordo com a padronização feita por Infante and Azeredo-Espin (1995). As bandas referentes ao DNA total foram visualizadas em gel de agarose 0.7%.

Por outro lado, a verificação das extrações de DNA de amostras preservadas em EtOH 96% não puderam ser visualizadas em gel de agarose. Isto ocorreu provavelmente porque o DNA destas amostras foi degradado por processos hidrolíticos e oxidativos (Pääbo, 1989), reduzindo a quantidade de DNA recuperado e impedindo sua visualização. No entanto, estas amostras foram submetidas à amplificação mesmo sem a avaliação da qualidade das extrações.

O DNA da espécie *P. regina*, que foi extraído do músculo do tórax e então liofilizado para ser enviado por correio, também não pôde ser visualizado em gel de agarose, provavelmente devido a uma menor quantidade de DNA recuperada apenas do músculo torácico. Isto também pode ter ocorrido devido à degradação que o DNA sofreu entre o envio dos EUA e o recebimento no Brasil. Altas temperaturas e a ausência de uma solução-tampão podem ter contribuído na degradação do DNA, assim como danos oxidativos.

2. “Long-PCR”

A amplificação por “long-PCR” do mtDNA de califorídeos foi padronizada utilizando-se a extração de DNA de amostras frescas de *C. idioidea*. O emprego das combinações de oligonucleotídeos universais amplificou eficientemente o mtDNA da espécie *C. idioidea*, mas notou-se uma grande instabilidade nas reações. Mesmo quando as condições da reação padronizada eram repetidas, muitas vezes a amplificação não ocorria. Ao empregar as mesmas combinações de oligonucleotídeos para outras espécies, os resultados não foram satisfatórios, amplificando-se muitos produtos inespecíficos. A instabilidade nas reações pode ter sido causada pelo tamanho dos oligonucleotídeos universais empregados, gerando uma região de hibridação pequena para a manutenção da dupla-fita durante a fase de extensão das cadeias de DNA (Francesco Nardi, comunicação pessoal). Para amplificações longas, geralmente

recomenda-se utilizar “primers” mais longos (>30 pb) dos que geralmente são utilizados em reações de PCR comuns (Nelson *et al.*, 1996; Hwang *et al.*, 2001b).

Deste modo, os “primers” universais e os desenvolvidos por Barau *et al.* (2005) passaram a ser utilizados com a combinação HI6S-A/mtD-22 e HI6Sb/mtD-19 (Tabela I). Ainda assim, para a amplificação do amplicon A (~8000 pb) de *C. vomitoria*, foi necessário realizar combinações de oligonucleotídeos longos para garantir a especificidade da reação (HI6S-A/HCO3-A). A partir de um gradiente na temperatura de hibridação foi possível definir quais as melhores condições para a amplificação de cada um dos fragmentos longos do mtDNA, chegando-se à conclusão de que com o ciclo de hibridação a 48°C obteve-se a maior quantidade de DNA amplificado para os dois amplicons. A Figura 3 mostra a amplificação do mtDNA completo das espécies de califorídeos *C. idioidea*, *P. regina*, *L. eximia* e *C. vomitoria*.

É interessante notar que, além das amostras frescas ou congeladas a -70°C, as extrações de DNA de amostras enviadas em EtOH 96% também foram testadas. As espécies *C. rufifacies*, *C. dubia*, *C. vicina*, *L. sericata* e *M. peregrina* não tiveram os fragmentos amplificados. Isto deve-se à degradação que o DNA sofre através de danos hidrolíticos e oxidativos (Pääbo, 1989; Lindahl, 1993; Dean and Ballard, 2001) após a morte das amostras, quando o processo de reparo do DNA deixa de funcionar (Cano, 1996; Höss *et al.*, 1996). As quebras nas cadeias de DNA são inevitáveis, ocasionando a fragmentação do DNA. Em amostras secas e preservadas em EtOH, é comum obter-se a amplificação de fragmentos pequenos, na maioria das vezes menores que 500 pb (Herrmann and Hummel, 1994; Cano, 1996; Junqueira *et al.*, 2002). Por esta razão, era esperado que os fragmentos longos do mtDNA não fossem recuperados para estas amostras. No entanto, surpreendentemente, a espécie *C. vomitoria* enviada de Londres (Reino Unido) apresentou tanto a banda de ~9200 pb quanto a de ~8000 pb. Este resultado mostra que o mtDNA pode ser recuperado de amostras preservadas, desde que existam algumas cópias intactas. Neste sentido, a utilização do mtDNA provavelmente favoreceu esta recuperação, já que este genoma está presente em grande número de cópias nas células de eucariotos. Este é um resultado inédito, uma vez que ainda não existem descrições de recuperação de genomas mitocondriais completos de espécies de insetos preservadas por alguns meses.

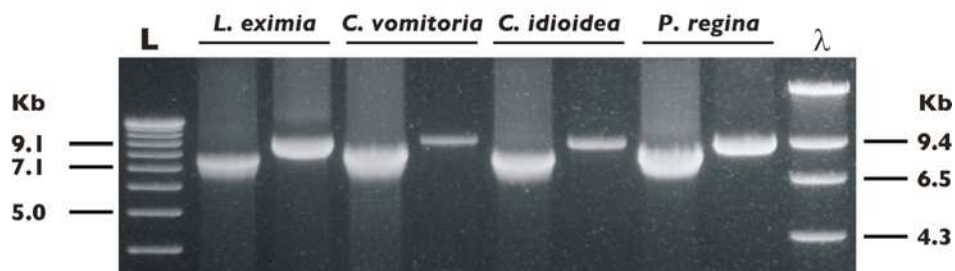


Figura 3: Gel de agarose (1.0%) com os amplicons amplificados por “long-PCR”, representando o genoma mitocondrial completo de quatro espécies da família Calliphoridae. L = marcador de peso molecular 1kb DNA Ladder; λ = marcador de peso molecular Lambda DNA / *HindIII*.

3. Bibliotecas de “shotgun” do genoma mitocondrial

Testes iniciais com o DNA proveniente do “long-PCR” foram feitos para estabelecer o melhor tempo de sonicação, a fim de otimizar a obtenção de fragmentos entre 1 e 2 kb. Entre os tempos de 5, 10 e 15 segundos em intensidade máxima, percebeu-se que 15 segundos são mais apropriados para obter fragmentos dos tamanhos desejados, sendo necessário sonicar algumas amostras por até 25 segundos. A padronização desta etapa é fundamental para o desenvolvimento da biblioteca de mtDNA, uma vez que o isolamento de fragmentos muito menores do que 1 kb ou muito maiores que 2 kb comprometem a qualidade do sequenciamento, levando à necessidade de utilizar um número maior de clones para conseguir finalizar os genomas. Um exemplo de fragmentos sonicados e retirados do gel de agarose de baixo ponto de fusão pode ser visto na Figura 4.

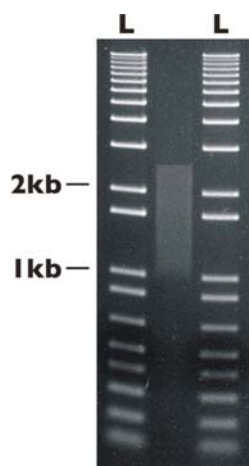


Figura 4: Gel de agarose 1.0% para verificação do isolamento eficiente dos fragmentos entre 1 e 2 kb recortados do gel de agarose LMP. Fragmentos um pouco maiores que 2 kb podem ser visualizados. L = marcador de peso molecular 1 kb DNA Ladder.

4. Caracterização dos genomas mitocondriais de *C. idioidea*, *C. vomitoria* e *P. regina*

As espécies *C. idioidea*, *C. vomitoria* e *P. regina* possuem, respectivamente, 15004 pb, 16143 pb e 16635 pb. É importante notar que a região que compreende a porção 5' da subunidade 16S até a porção 3' da subunidade 12S do rRNA de *C. idioidea* não foi completamente sequenciada, abrangendo uma porção de aproximadamente 800 pb. Assim, o número de bases exato do mtDNA desta espécie ainda é desconhecido, embora esteja em torno de 16000 pb.

Os genes montados e anotados em *C. idioidea* e *C. vomitoria* apresentam a mesma organização descrita para a espécie de califorídeo *C. hominivorax* (Lessinger et al., 2000), que por sua vez é a disposição comumente descrita entre espécies de insetos e artrópodes (Boore, 1999), contendo os 37 genes já previamente mencionados (13 GCPs, 22 tRNAs e 2 rRNAs). Variações na ordem e número de genes geralmente estão associadas a rearranjos e duplicações de tRNAs, como descrito para o califorídeo *Chrysomya putoria* (Junqueira et al., 2004), que apresenta uma duplicação do *tRNA^{Ile}* e suas regiões flanqueadoras, incluindo sequências parciais do *tRNA^{Gln}* e do domínio conservado da região controle. A exemplo do que ocorre para esta espécie de califorídeo, *P. regina* também apresenta a duplicação do *tRNA^{Ile}*, mas também possui o *tRNA^{Gln}* completamente duplicado, além de um pseudogene de *tRNA^{Met}*. Deste modo, ao invés de apresentar os 22 tRNAs usualmente descritos em espécies de dípteros, *P. regina* apresenta 24 genes para tRNAs e um total de 39 genes no mtDNA. O esquema geral dos três genomas mitocondriais sequenciados pode ser visto na Figuras 5, que inclui também os sítios de restrição existentes nestes genomas.

A composição nucleotídica apresenta um alto viés de bases As e Ts, como pode ser visto na Tabela 3, correspondendo a 71.7%, 72.9% e 75.6% em *C. idioidea*, *P. regina* e *C. vomitoria*, respectivamente. O desvio na composição nucleotídica pode ser comparado através do índice "skewness" (Perna and Kocher, 1995), que mede a frequência relativa de As e Ts ($AT_{skew} = [A-T]/[A+T]$) e Gs e Cs ($GC_{skew} = [G-C]/[G+C]$) e é uma medida comumente realizada para caracterização de genomas. A Tabela 3 inclui dados comparativos do AT e GC-skew de 28 espécies de dípteros.

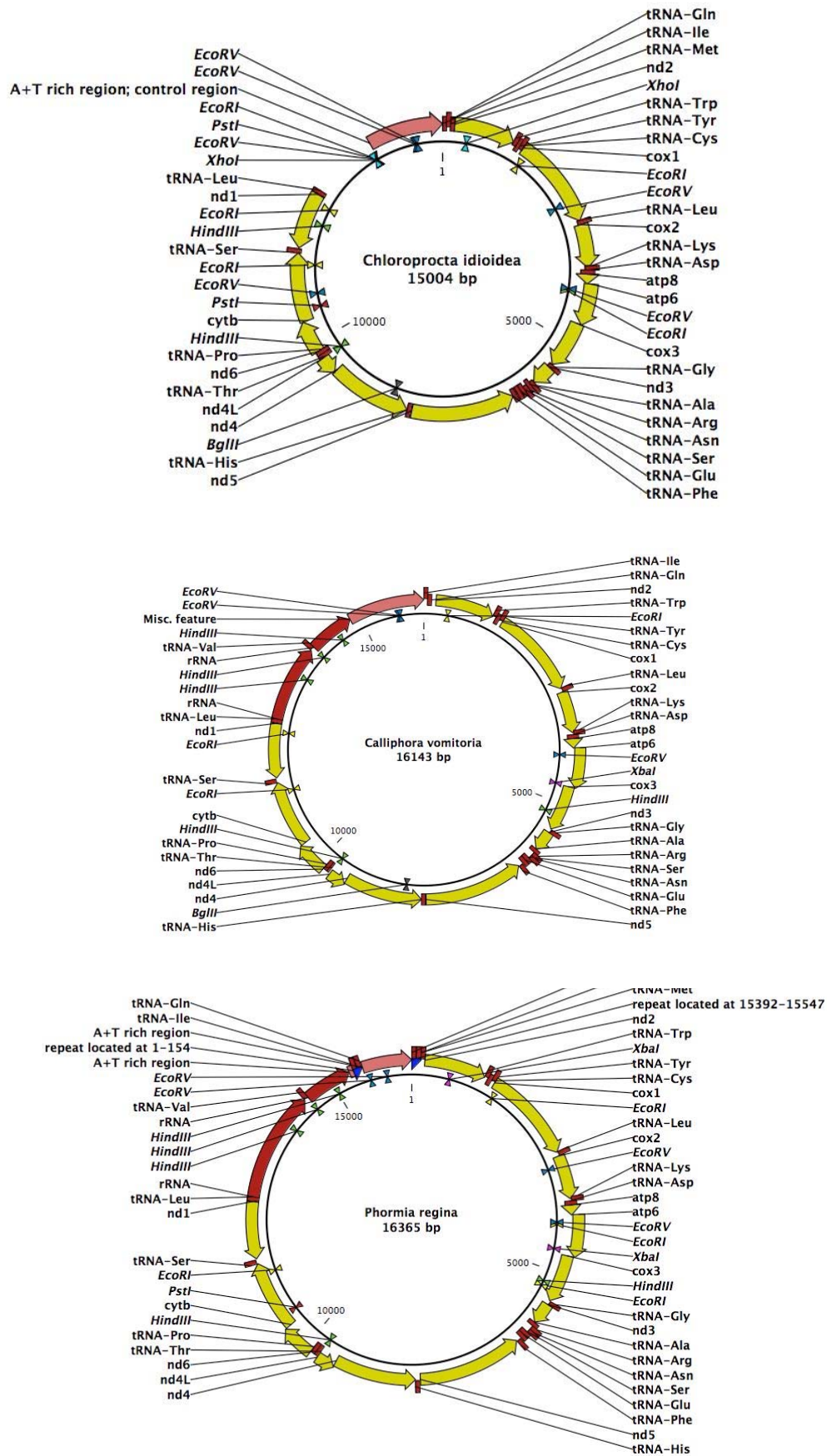


Figura 5: Estrutura dos genomas mitocondriais de *C. idioidea*, *C. vomitoria* e *P. regina*, indicando os genes anotados e os sítios de restrição encontrados em cada uma das espécies. Em *P. regina*, a setas em azul denotam a duplicação envolvendo os tRNAs para Isoleucina e Glutamina.

Tabela 3: Composição A+T do genoma mitocondrial total, medidas de AT e GC-skews dos GCPs, tRNAs e rRNAs das espécies de dípteros que possuem sequência completa disponível no GenBank. Em verde estão ressaltados os genomas sequenciados neste trabalho. O “*” denota sequências que possuem “gaps” ou que não estão completamente anotadas, não permitindo o cálculo de algumas medidas de AT e GC-skew, que por sua vez estão indicadas como “nd” (não disponível).

Espécie	mtDNA - total				PCGs +		PCGs -		tRNAs		rRNAs	
	pb	A+T%	AT skew	GC skew	AT skew	GC skew	AT skew	GC skew	AT skew	GC skew	AT skew	GC skew
<i>Chloroprocta idioidea*</i>	15004	71.7	0.021	-0.157	-0.121	-0.112	-0.225	0.258	-0.018	0.130	nd	nd
<i>Calliphora vomitoria</i>	16143	75.6	0.018	-0.163	-0.107	-0.089	-0.210	0.269	-0.011	0.140	-0.001	0.269
<i>Phormia regina</i>	16365	72.9	0.021	-0.170	-0.112	-0.111	-0.221	0.275	-0.014	0.141	-0.005	0.251
<i>Chrysomya putoria</i>	15837	76.7	0.020	-0.170	-0.116	-0.103	-0.227	0.284	-0.003	0.109	-0.007	0.286
<i>Cochliomyia hominivorax</i>	16022	76.9	0.034	-0.207	-0.094	-0.144	-0.223	0.305	-0.016	0.128	-0.026	0.305
<i>Chrysomya megacephala*</i>	15831	76.5	0.016	-0.174	-0.116	-0.102	-0.213	0.293	nd	nd	-0.009	0.297
<i>Lucilia sericata</i>	15945	77.6	0.016	-0.167	-0.127	-0.091	-0.219	0.281	-0.019	0.120	-0.013	0.270
<i>Dermatobia hominis</i>	16360	77.8	0.043	-0.227	-0.079	-0.160	-0.233	0.330	0.014	0.125	-0.029	0.337
<i>Haematobia irritans</i>	16078	79.1	0.004	-0.125	-0.132	-0.044	-0.186	0.211	-0.010	0.112	-0.023	0.296
<i>Stomoxys calcitrans</i>	16790	78.9	-0.001	-0.140	-0.135	-0.071	-0.177	0.239	-0.018	0.129	-0.009	0.289
<i>Drosophila mauritiana</i>	14964	77.7	0.009	-0.135	-0.127	-0.055	-0.220	0.252	-0.027	0.149	-0.001	0.258
<i>Drosophila melanogaster</i>	19517	82.2	0.017	-0.150	-0.124	-0.058	-0.217	0.246	-0.012	0.141	-0.005	0.288
<i>Drosophila sechellia</i>	14950	77.6	0.009	-0.135	-0.122	-0.067	-0.211	0.220	-0.021	0.152	0.000	0.264
<i>Drosophila simulans</i>	14972	77.9	0.008	-0.135	-0.123	-0.060	-0.211	0.235	-0.018	0.148	-0.001	0.272
<i>Drosophila yakuba</i>	16019	78.6	0.005	-0.136	-0.134	-0.053	-0.196	0.245	-0.023	0.155	0.009	0.276
<i>Bactrocera carambolae</i>	15915	73.5	0.066	-0.224	-0.066	-0.191	-0.278	0.300	-0.028	0.118	-0.079	0.304
<i>Bactrocera dorsalis</i>	15915	73.6	0.068	-0.228	-0.067	-0.196	-0.282	0.312	-0.019	0.132	-0.076	0.287
<i>Bactrocera oleae</i>	15815	72.6	0.088	-0.280	-0.048	-0.243	-0.295	0.378	-0.015	0.086	-0.099	0.347
<i>Bactrocera papayae</i>	15915	73.5	0.066	-0.226	-0.065	-0.196	-0.280	0.306	-0.029	0.123	-0.069	0.287
<i>Bactrocera philippinensis</i>	15915	73.6	0.066	-0.224	-0.067	-0.194	-0.278	0.295	-0.025	0.122	-0.074	0.292
<i>Ceratitis capitata</i>	15980	77.5	0.021	-0.185	-0.123	-0.120	-0.231	0.287	-0.009	0.099	-0.058	0.267
<i>Aedes albopictus</i>	16665	79.5	0.008	-0.181	-0.123	-0.097	-0.203	0.269	0.015	0.128	-0.049	0.293
<i>Aedes aegypti</i>	16655	79.0	0.017	-0.211	-0.120	-0.132	-0.217	0.314	0.034	0.123	-0.050	0.301
<i>Anopheles gambiae</i>	15363	77.6	0.032	-0.154	-0.097	-0.078	-0.227	0.243	-0.003	0.131	-0.053	0.301
<i>Anopheles quadrimaculatus</i>	15455	77.4	0.041	-0.181	-0.089	-0.101	-0.237	0.283	0.011	0.146	-0.056	0.297
<i>Cydistomyia duplonotata</i>	16247	77.9	0.003	-0.177	-0.121	-0.106	-0.192	0.268	-0.007	0.158	-0.010	0.319
<i>Trichophthalma punctata</i>	16396	74.0	0.091	-0.245	-0.045	-0.214	-0.291	0.324	0.012	0.162	-0.116	0.338
<i>Simosyrphus grandicornis</i>	16141	80.8	-0.004	-0.133	0.122	-0.065	-0.164	0.217	0.014	0.141	-0.010	0.287

4.1. Genes codificadores de proteínas

A ordem dos GCPs de *C. idioidea*, *C. vomitoria* e *P. regina* é idêntica à descrita para Arthropoda. Os códons de iniciação e terminação foram identificados por comparação do tamanho das proteínas anotadas por outros dípteros e com a utilização do ORF finder. O códon ATG foi o mais utilizado para a iniciação (6 genes em cada espécie), mas os códons ATT (4 genes em *C. idioidea* e *P. regina* e 3 genes em *C. vomitoria*), ATA (2 genes em cada espécie) e ATC (no gene *atp8* de *C. vomitoria*) também foram detectados (Tabela 4). Além destes, o gene *cox1* apresentou um códon de iniciação não usual, sendo iniciado com uma Serina codificada pelo códon TCG (Tabela 4). Da mesma maneira, códons não usuais de terminação também foram encontrados para os genes *nd5* e *nd4* (Tabela 4). A existência de códons de terminação incompletos é uma característica comum dos genes mitocondriais de metazoários, sendo que o mecanismo mais provável para a criação de um códon de terminação completo é a poliadenilação pós-transcricional, que gera códons TAA completos (Ojala *et al.*, 1981). Por outro lado, o códon de terminação TAA foi o mais utilizado (9, 11 e 10 vezes em *C. idioidea*, *C. vomitoria* e *P. regina*, respectivamente), enquanto TAG foi empregado na terminação de *cytb* e *nd3* de *C. idioidea* e em *cytb* de *P. regina*.

O conteúdo A+T médio dos GCPs de *C. idioidea*, calculado para cada cadeia codificadora, é de 73.1% para a “fita +” e de 77.3% para a “fita -”. A espécie *C. vomitoria* apresenta 74.4% nos GCPs da “fita +” e 77.4% nos da “fita -”, enquanto *P. regina* possui 73.2% para a “fita +” e 77.6% para a “fita -”. A frequência de cada nucleotídeo por posição do códon está na Figura 6.

É evidente que o nucleotídeo T é o mais utilizado em todas as posições do códon, corroborando com os números encontrados de AT-skew, uma vez que estas medidas são negativas para os GCPs de ambas as fitas (Tabela 3). Além disso, há predominância de As e Ts principalmente na primeira e terceira posição do códon. A tendência da utilização destes nucleotídeos principalmente na terceira posição do códon é amplamente observada em genes mitocondriais de invertebrados, devido ao código genético degenerado e à manutenção de taxas altas destes nucleotídeos no mtDNA de insetos. A Figura 6 também mostra que a utilização de Gs e Cs é praticamente equivalente nas três espécies, com a predominância de Gs apenas na primeira posição do códon.

Tabela 4: Códons de iniciação e terminação encontrados nos GCPs mitocondriais de *C. idioidea*, *C. vomitoria* e *P. regina*.

Gene	<i>C. idioidea</i>		<i>C. vomitoria</i>		<i>P. regina</i>	
	iniciação	terminação	iniciação	terminação	iniciação	terminação
<i>atp6</i>	ATG(M)	TAA	ATG(M)	TAA	ATG(M)	TAA
<i>atp8</i>	ATT(M)	TAA	ATC(M)	TAA	ATT(M)	TAA
<i>cox1</i>	TCG*(S)	TAA	TCG*(S)	TAA	TCG*(S)	TAA
<i>cox2</i>	ATG(M)	TAA	ATG(M)	TAA	ATG(M)	TAA
<i>cox3</i>	ATG(M)	TAA	ATG(M)	TAA	ATG(M)	TAA
<i>cytb</i>	ATG(M)	TAG	ATG(M)	TAA	ATG(M)	TAG
<i>nd1</i>	ATA(M)	TAA	ATA(M)	TAA	ATA(M)	TAA
<i>nd2</i>	ATT(M)	TAA	ATT(M)	TAA	ATT(M)	TAA
<i>nd3</i>	ATA(M)	TAG	ATA(M)	TAA	ATA(M)	TAA
<i>nd4</i>	ATG(M)	T	ATG(M)	T	ATG(M)	TA
<i>nd4L</i>	ATG(M)	TAA	ATG(M)	TAA	ATG(M)	TAA
<i>nd5</i>	ATT(M)	TA	ATT(M)	TA	ATT(M)	TA
<i>nd6</i>	ATT(M)	TAA	ATT(M)	TAA	ATT(M)	TAA

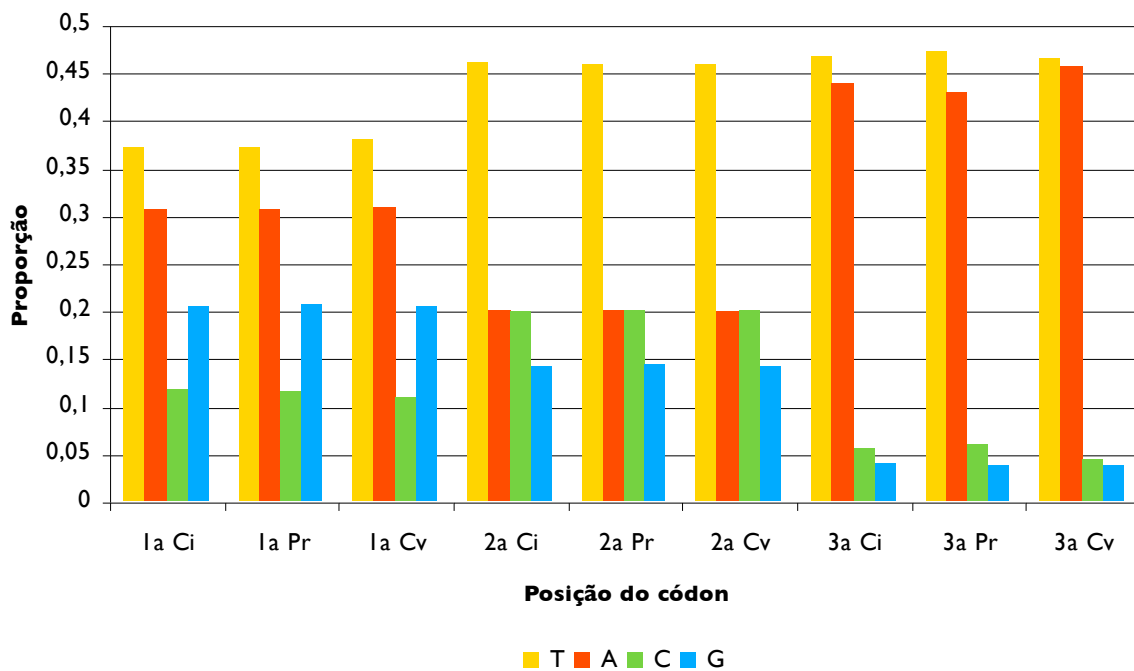


Figura 6: Proporção relativa dos nucleotídeos T, A, C e G por posição de códon nos GCPs mitocondriais de *C. idioidea*, *C. vomitoria* e *P. regina*.

A utilização preferencial de códons ("codon usage") dos GCPs do genoma mitocondrial de *C. idioidea*, *C. vomitoria* e *P. regina*, assim como os valores de RSCU ("relative synonymous codon usage") estão nas Tabelas 5, 6 e 7. Como o esperado, a maior parte dos valores RSCU difere da frequência de equilíbrio (=1) demonstrando que o uso de códons sinônimos é altamente desviado. Os códons CCG (Pro), GCG (Ala), CGC (Arg) e AGG (Ser) não foram utilizados pelo genoma mitocondrial de *C. idioidea*, enquanto o mtDNA de *C. vomitoria* não apresenta utilização dos códons CUC (Leu), CGC (Arg), AGG (Ser) e GGC (Gly). *P. regina* não utiliza os códons CGC (Arg), AGG (Ser) e GGC (Gly). Isto pode ser explicado devido à tendência de acúmulo de bases A e T no genoma mitocondrial de insetos, refletindo na baixa utilização de códons com G e C. Nos exemplos acima, códons com G e/ou C em duas ou três posições do códon não foram utilizados.

A Tabela 8 indicam a frequência dos aminoácidos utilizados nas proteínas mitocondriais das três espécies, evidenciando a alta utilização de Leucina, Serina, Fenilalanina e Isoleucina. Somados, estes quatro aminoácidos representam mais de 40% do total de aminoácidos utilizados em proteínas traduzidas pelos genomas mitocondriais das três espécies.

Tabela 5: "Codon" usage dos GCPs de *C. idioidea*, indicando a frequência (freq.) de cada códon em números absolutos e o valor relativo da utilização de códons sinônimos (RSCU).

Codon (aa)	Freq.	RSCU	Codon (aa)	Freq.	RSCU	Codon (aa)	Freq.	RSCU	Codon (aa)	Freq.	RSCU
UUU (F)	280	1,64	UCU (S)	131	3,1	UAU (Y)	136	1,63	UGU (C)	34	1,89
UUC (F)	61	0,36	UCC (S)	7	0,17	UAC (Y)	31	0,37	UGC (C)	2	0,11
UUA (L)	480	4,91	UCA (S)	88	2,08	UAA (*)	0	0	UGA (W)	91	1,88
UUG (L)	22	0,22	UCG (S)	2	0,05	UAG (*)	0	0	UGG (W)	6	0,12
CUU (L)	41	0,42	CCU (P)	80	2,39	CAU (H)	58	1,55	CGU (R)	10	0,69
CUC (L)	1	0,01	CCC (P)	9	0,27	CAC (H)	17	0,45	CGC (R)	0	0
CUA (L)	42	0,43	CCA (P)	45	1,34	CAA (Q)	67	1,89	CGA (R)	42	2,9
CUG (L)	1	0,01	CCG (P)	0	0	CAG (Q)	4	0,11	CGG (R)	6	0,41
AUU (I)	311	1,86	ACU (T)	113	2,35	AAU (N)	167	1,8	AGU (S)	61	1,44
AUC (I)	23	0,14	ACC (T)	9	0,19	AAC (N)	19	0,2	AGC (S)	3	0,07
AUA (M)	184	1,71	ACA (T)	69	1,44	AAA (K)	50	1,11	AGA (S)	46	1,09
AUG (M)	31	0,29	ACG (T)	1	0,02	AAG (K)	40	0,89	AGG (S)	0	0
GUU (V)	96	1,82	GCU (A)	125	2,73	GAU (D)	50	1,47	GGU (G)	49	0,89
GUC (V)	5	0,09	GCC (A)	12	0,26	GAC (D)	18	0,53	GGC (G)	2	0,04
GUA (V)	107	2,03	GCA (A)	46	1,01	GAA (E)	71	1,87	GGA (G)	151	2,75
GUG (V)	3	0,06	GCG (A)	0	0	GAG (E)	5	0,13	GGG (G)	18	0,33

Tabela 6: “Codon usage” dos GCPs de *C. vomitoria*, indicando a frequência (freq.) de cada códon em números absolutos e o valor relativo da utilização de códons sinônimos (RSCU).

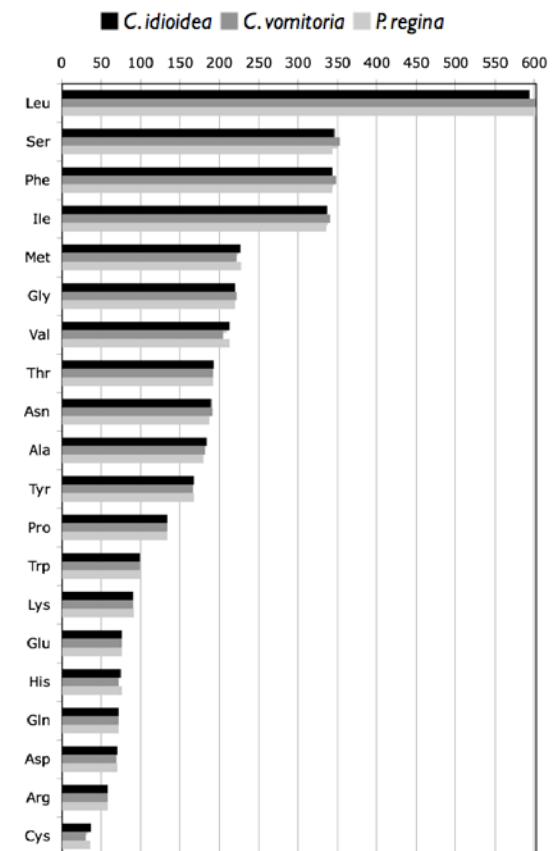
Codon (aa)	Freq.	RSCU	Codon (aa)	Freq.	RSCU	Codon (aa)	Freq.	RSCU	Codon (aa)	Freq.	RSCU
UUU (F)	292	1,72	UCU (S)	125	2,91	UAU (Y)	138	1,66	UGU (C)	29	1,93
UUC (F)	47	0,28	UCC (S)	4	0,09	UAC (Y)	28	0,34	UGC (C)	1	0,07
UUA (L)	505	5,09	UCA (S)	104	2,42	UAA (*)	0	0	UGA (W)	90	1,86
UUG (L)	23	0,23	UCG (S)	2	0,05	UAG (*)	0	0	UGG (W)	7	0,14
CUU (L)	37	0,37	CCU (P)	71	2,12	CAU (H)	62	1,7	CGU (R)	8	0,55
CUC (L)	0	0	CCC (P)	4	0,12	CAC (H)	11	0,3	CGC (R)	0	0
CUA (L)	28	0,28	CCA (P)	57	1,7	CAA (Q)	69	1,94	CGA (R)	49	3,38
CUG (L)	2	0,02	CCG (P)	2	0,06	CAG (Q)	2	0,06	CGG (R)	1	0,07
AUU (I)	333	1,92	ACU (T)	101	1,8	AAU (N)	169	1,8	AGU (S)	52	1,21
AUC (I)	13	0,08	ACC (T)	1	0,2	AAC (N)	19	0,2	AGC (S)	2	0,05
AUA (M)	180	1,74	ACA (T)	87	1,29	AAA (K)	58	1,29	AGA (S)	55	1,28
AUG (M)	27	0,26	ACG (T)	1	0,71	AAG (K)	32	0,71	AGG (S)	0	0
GUU (V)	86	1,69	GCU (A)	105	1,55	GAU (D)	52	1,55	GGU (G)	46	0,83
GUC (V)	2	0,04	GCC (A)	11	0,45	GAC (D)	15	0,45	GGC (G)	0	0
GUA (V)	112	2,21	GCA (A)	61	1,97	GAA (E)	75	1,97	GGA (G)	145	2,61
GUG (V)	3	0,06	GCG (A)	4	0,03	GAG (E)	1	0,03	GGG (G)	31	0,56

Tabela 7: “Codon usage” dos GCPs de *P. regina*, indicando a frequência (freq.) de cada códon em números absolutos e o valor relativo da utilização de códons sinônimos (RSCU).

Codon (aa)	Freq.	RSCU	Codon (aa)	Freq.	RSCU	Codon (aa)	Freq.	RSCU	Codon (aa)	Freq.	RSCU
UUU (F)	274	1,61	UCU (S)	110	2,62	UAU (Y)	142	1,7	UGU (C)	34	1,94
UUC (F)	67	0,39	UCC (S)	7	0,17	UAC (Y)	25	0,3	UGC (C)	1	0,06
UUA (L)	471	4,74	UCA (S)	110	2,62	UAA (*)	0	0	UGA (W)	95	1,96
UUG (L)	30	0,3	UCG (S)	2	0,05	UAG (*)	0	0	UGG (W)	2	0,04
CUU (L)	45	0,45	CCU (P)	80	2,39	CAU (H)	54	1,42	CGU (R)	10	0,69
CUC (L)	2	0,02	CCC (P)	4	0,12	CAC (H)	22	0,58	CGC (R)	0	0
CUA (L)	46	0,46	CCA (P)	49	1,46	CAA (Q)	68	1,92	CGA (R)	43	2,97
CUG (L)	2	0,02	CCG (P)	1	0,03	CAG (Q)	3	0,08	CGG (R)	5	0,34
AUU (I)	310	1,86	ACU (T)	107	2,29	AAU (N)	165	1,76	AGU (S)	56	1,33
AUC (I)	24	0,14	ACC (T)	2	0,04	AAC (N)	23	0,24	AGC (S)	3	0,07
AUA (M)	191	1,79	ACA (T)	73	1,56	AAA (K)	54	1,19	AGA (S)	48	1,14
AUG (M)	23	0,21	ACG (T)	5	0,11	AAG (K)	37	0,81	AGG (S)	0	0
GUU (V)	100	1,9	GCU (A)	118	2,64	GAU (D)	58	1,71	GGU (G)	51	0,93
GUC (V)	2	0,04	GCC (A)	10	0,22	GAC (D)	10	0,29	GGC (G)	0	0
GUA (V)	101	1,91	GCA (A)	47	1,05	GAA (E)	73	1,92	GGA (G)	147	2,67
GUG (V)	8	0,15	GCG (A)	4	0,09	GAG (E)	3	0,08	GGG (G)	22	0,4

Tabela 8: Frequência de aminoácidos em proteínas mitocondriais. O gráfico ao lado indica o valor absoluto de cada aminoácido nos GCPs das três espécies de califorídeos sequenciadas.

Aminoácido	<i>C. idioidea</i>		<i>C. vomitoria</i>		<i>P. regina</i>		Carga
	Frequência	Total	Frequência	Total	Frequência	Total	
Leu	15.92%	594	16.12%	601	16.14%	602	Apolar
Ser	9.28%	346	9.47%	353	9.22%	344	Polar sem carga
Phe	9.22%	344	9.33%	348	9.22%	344	Apolar
Ile	9.03%	337	9.14%	341	9.01%	336	Apolar
Met	6.09%	227	5.95%	222	6.11%	228	Apolar
Gly	5.90%	220	5.95%	222	5.90%	220	Apolar
Val	5.71%	213	5.50%	205	5.71%	213	Apolar
Thr	5.17%	193	5.15%	192	5.15%	192	Polar sem carga
Asn	5.09%	190	5.12%	191	5.04%	188	Polar sem carga
Ala	4.93%	184	4.88%	182	4.83%	180	Apolar
Tyr	4.50%	168	4.48%	167	4.50%	168	Polar sem carga
Pro	3.59%	134	3.59%	134	3.59%	134	Apolar
Trp	2.65%	99	2.65%	99	2.65%	99	Apolar
Lys	2.44%	91	2.44%	91	2.47%	92	Polar - Básico
Glu	2.04%	76	2.04%	76	2.04%	76	Polar - Ácido
His	2.01%	75	1.96%	73	2.04%	76	Polar - Básico
Gln	1.96%	73	1.96%	73	1.96%	73	Polar sem carga
Asp	1.90%	71	1.88%	70	1.90%	71	Polar - Ácido
Arg	1.55%	58	1.56%	58	1.55%	58	Polar - Básico
Cys	0.99%	37	0.83%	31	0.97%	36	Apolar



A análise do número absoluto de sítios conservados e variáveis nas sequências de aminoácidos dos GCPs do genoma mitocondrial de dípteros (Figura 7), assim como a razão entre os mesmos (razão C/V; Figura 8) indica o grau de conservação e/ou variação dos genes. Estes resultados foram particularmente importantes para guiar a escolha do sub-conjunto de dados utilizados nas análises filogenéticas posteriores, embora reconstruções com o concatenado de todos os genes também tenham sido realizadas.

A divergência nucleotídica (calculada a partir da *distância p* para cada par de espécies) também é uma maneira de quantificar a variabilidade ou conservação de cada gene mitocondrial. Um resumo dos resultados obtidos desta análise está na Tabela 9, considerando-se 28 espécies de dípteros. É possível verificar que os genes *nd2*, *atp8* e *nd6* possuem as maiores médias de distância, apresentando também a maior variação quando compara-se grupos proximalmente relacionados com espécies de divergência antiga.

A variabilidade encontrada na sequência de aminoácidos ao longo dos GCPs pode ser avaliada na Figura 9, que mostra a identidade encontrada em cada sítio homólogo de cada GCP do mtDNA. É possível verificar os genes mais conservados, assim como as diferenças de variação ao longo de um mesmo gene, quando compara-se as sequências das 28 espécies de dípteros listadas na Tabela 2.

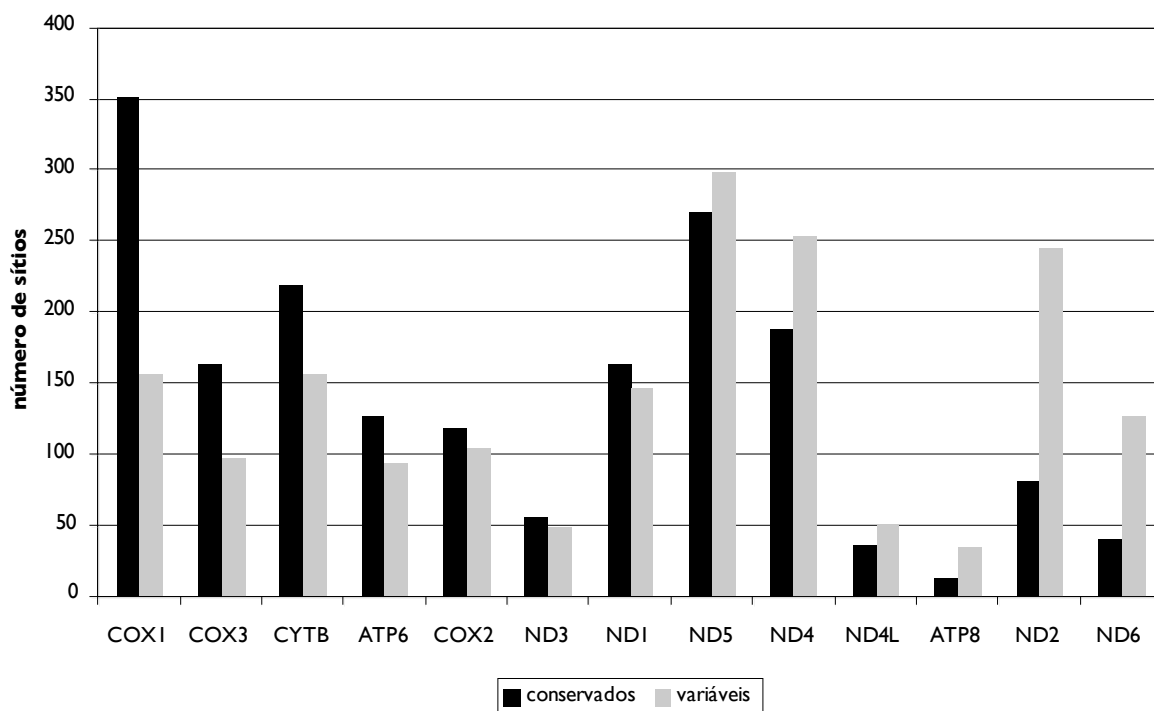


Figura 7: Número absoluto de sítios conservados e variáveis encontrados no alinhamento das sequências de aminoácidos de cada GCP do mtDNA de 28 dípteros.

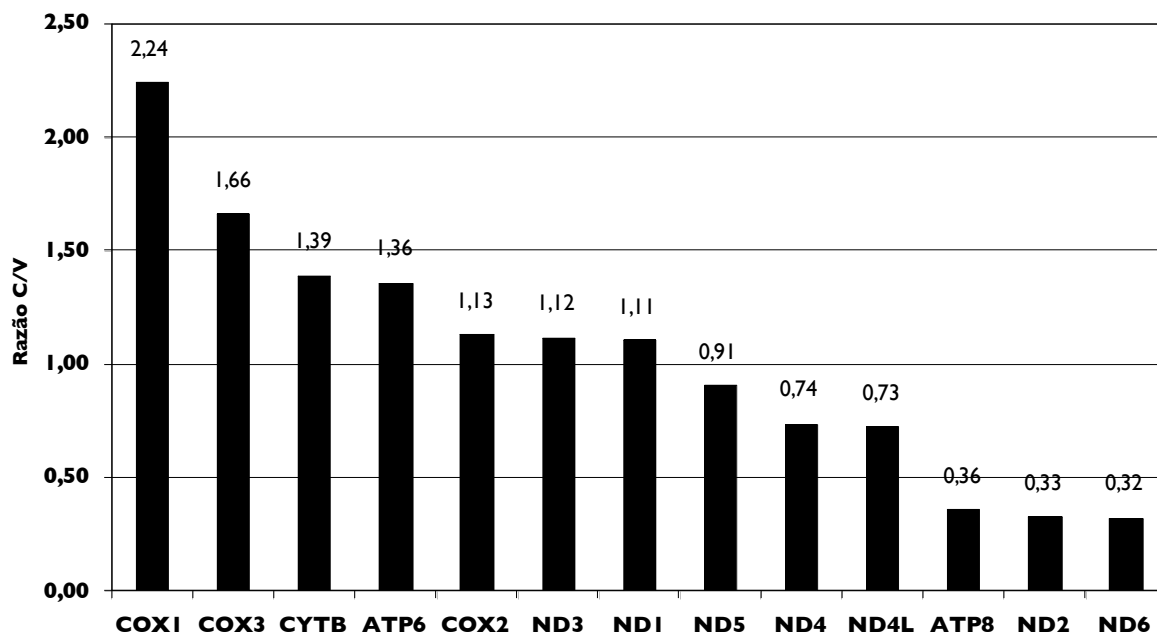


Figura 8: Razão entre sítios conservados e variáveis (C/V), obtida a partir do alinhamento individual de genes mitocondriais, indicando o grau de variação nos 13 GCPs de dípteros. Os genes com a razão C/V < 1.00 foram excluídos das análises filogenéticas posteriores por não resultarem em alinhamentos confiáveis de sítios homólogos.

Tabela 9: Distância *p* calculada para cada GCP mitocondrial, incluindo as 28 espécies de dípteros analisadas. Os números indicam o menor e o maior valor encontrado para cada gene, com o par de espécies relacionado aos mesmos.

Gene	Distância <i>p</i> (par de espécies)	Média
atp6	0.007 (<i>D. sechelia</i> X <i>D. simulans</i>)	0.170
	0.239 (<i>S. grandicornis</i> X <i>B. philippinensis</i>)	
atp8	0.000 (<i>B. papayae</i> X <i>B. philippinensis</i>)	0.257
	0.372 (<i>T. punctata</i> X <i>D. yakuba</i>)	
cox1	0.014 (<i>D. sechelia</i> X <i>D. simulans</i>)	0.148
	0.207 (<i>T. punctata</i> X <i>A. aegypti</i>)	
cox2	0.000 (<i>B. papayae</i> X <i>B. philippinensis</i>)	0.168
	0.257 (<i>B. oeleae</i> X <i>A. gambiae</i>)	
cox3	0.008 (<i>B. papayae</i> X <i>B. philippinensis</i>)	0.154
	0.218 (<i>T. punctata</i> X <i>A. quadrimaculatus</i>)	
cytb	0.011 (<i>B. papayae</i> X <i>B. philippinensis</i>)	0.176
	0.268 (<i>B. dorsalis</i> X <i>A. gambiae</i>)	
nd1	0.008 (<i>B. papayae</i> X <i>B. dorsalis</i>)	0.153
	0.210 (<i>T. punctata</i> X <i>B. carambolae</i>)	
nd2	0.011 (<i>D. sechelia</i> X <i>D. simulans</i>)	0.247
	0.340 (<i>T. punctata</i> X <i>B. philippinensis</i>)	
nd3	0.003 (<i>B. papayae</i> X <i>B. philippinensis</i>)	0.197
	0.284 (<i>B. papayae</i> X <i>C. duplonotata</i>)	
nd4	0.010 (<i>B. papayae</i> X <i>B. carambolae</i>)	0.182
	0.257 (<i>T. punctata</i> X <i>B. oleae</i>)	
nd4L	0.007 (<i>B. philippinensis</i> X <i>B. dorsalis</i>)	0.187
	0.312 (<i>B. papayae</i> X <i>A. albopictus</i>)	
nd5	0.013 (<i>B. papayae</i> X <i>B. dorsalis</i>)	0.187
	0.253 (<i>T. punctata</i> X <i>A. quadrimaculatus</i>)	
nd6	0.002 (<i>B. carambolae</i> X <i>B. dorsalis</i>)	0.269
	0.376 (<i>B. papayae</i> X <i>A. albopictus</i>)	

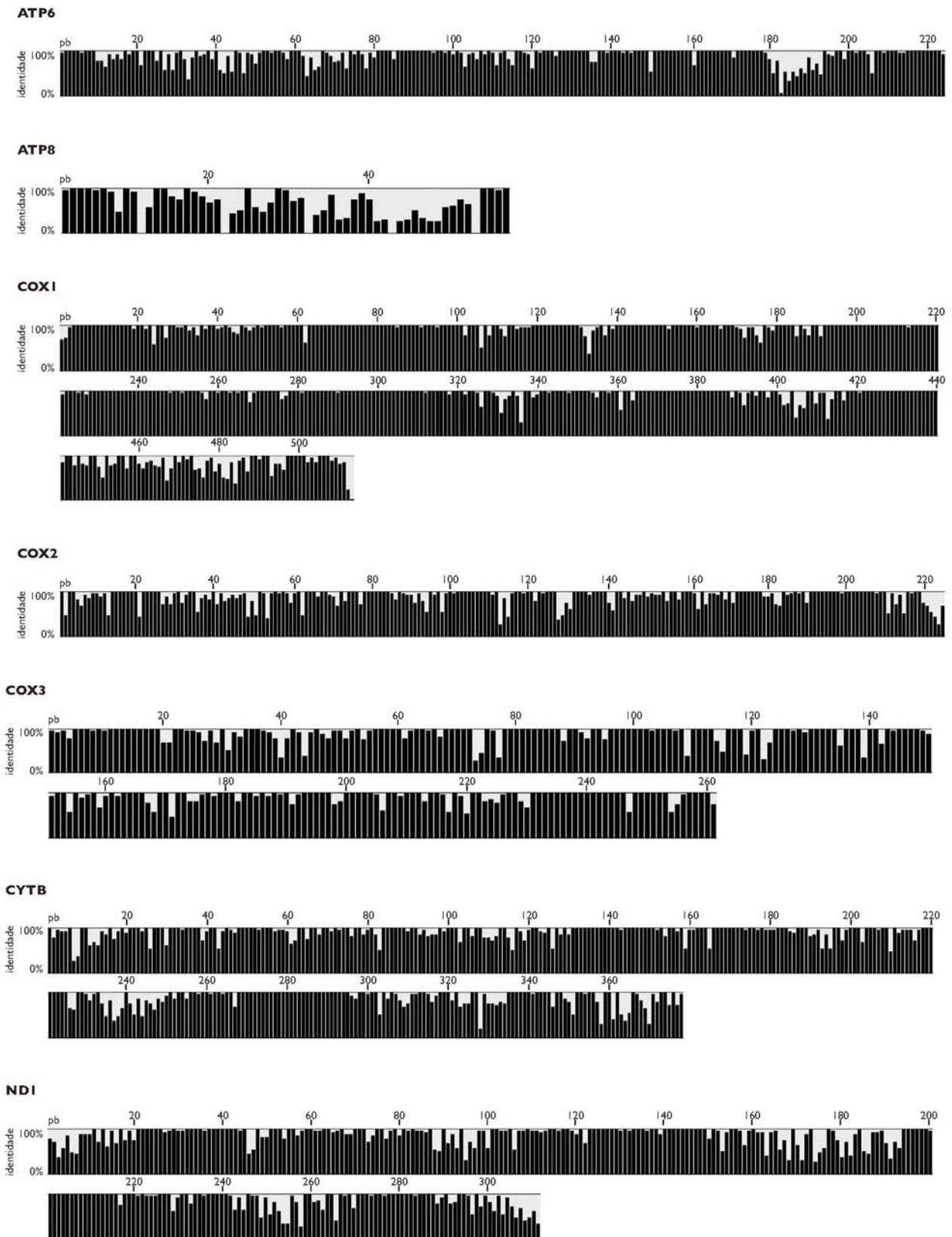
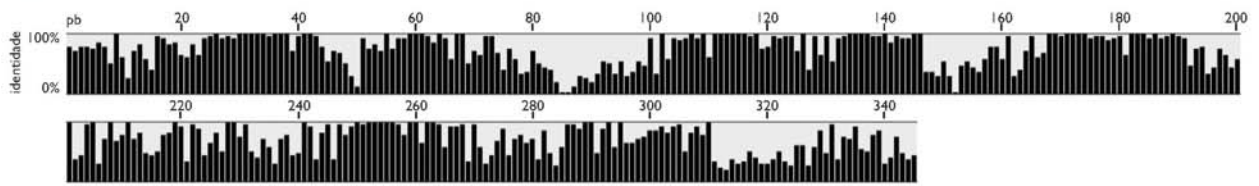
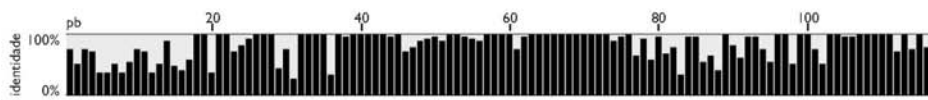
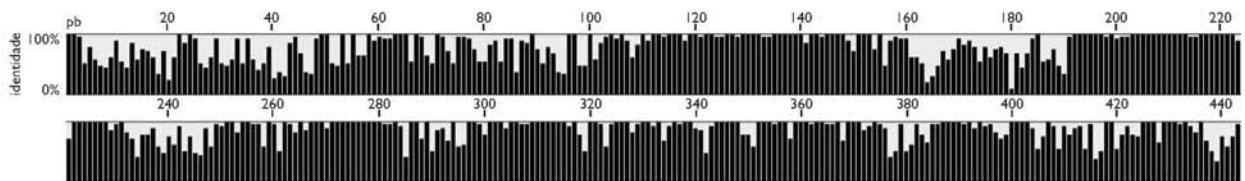
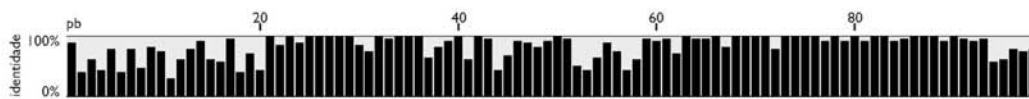
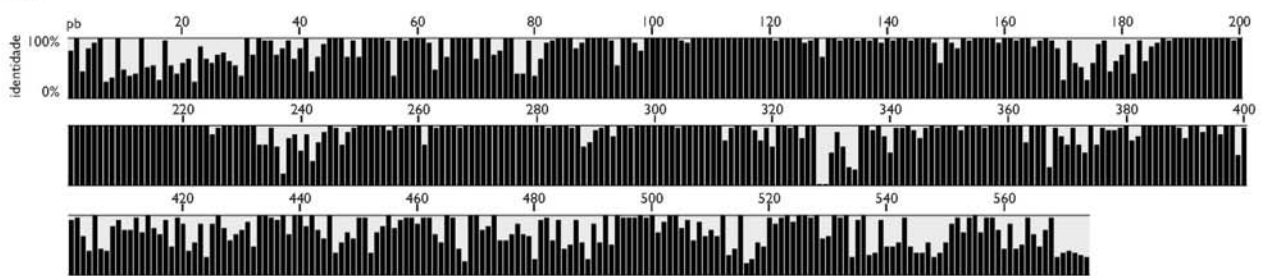
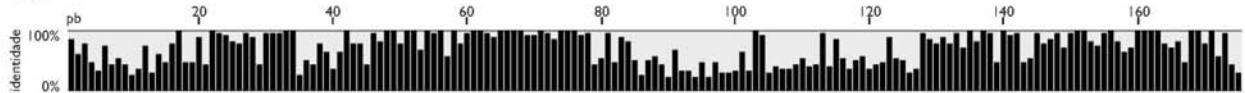


Figura 9: Identidade de aminoácidos encontrada ao longo dos 13 GCPs mitocondriais de 28 espécies de dípteros. As barras indicam a porcentagem de identidade em cada sítio homólogo alinhado.

ND2**ND3****ND4****ND4L****ND5****ND6****Figura 9:** Continuação.

4.2. Subunidades do rRNA

A subunidade 16S do mtDNA (*rrnL*) apresenta um tamanho aproximado de 1300 pb, para *C. idioidea*, considerando um “gap” de cerca de 800 pb que estão sendo sequenciados para fechar completamente este genoma. A subunidade 12S (*rrnS*) tem 786 pb, com um conteúdo A+T de 76.7% e AT e GC-skews de -0.001 e 0.246, respectivamente. Da mesma forma, *C. vomitoria* apresentou o *rrnL* com 1325 pb e o *rrnS* com 782 pb, com o conteúdo A+T de 85% e 78%, respectivamente (AT-skew = -0.001 e GC-skew = 0.269). A espécie *P. regina* tem o *rrnL* e o *rrnS* com 1381 pb e 785 pb, apresentando conteúdo A+T de 84% e 77%, respectivamente (AT-skew = -0.005 e GC-skew = 0.251).

Comparadas às subunidades ribossomais do mtDNA de outros dípteros, o tamanho e a composição nucleotídica do *rrnL* e *rrnS* são conservados. O conteúdo A+T dos genes para rRNA são os mais altos no genoma mitocondrial, quando a região controle não é considerada. Além disso, o *rrnS* demonstrou uma identidade de 85.8% entre as espécies de dípteros analisadas, o *rrnL* possui uma identidade de 86.8%, excluindo-se a sequência de *C. idioidea*, que não tem sequências completas para este gene.

4.3. Genes para tRNA

O genoma mitocondrial de *C. idioidea* e *C. vomitoria* apresentam os 22 genes codificadores de tRNAs comumente descritos em Arthropoda, que puderam ser identificados através da estrutura secundária, com a típica conformação “clover-leaf”, ou folha de trevo. Já a espécie *P. regina* possui 24 tRNAs, havendo uma duplicação completa dos genes *tRNA^{Ile}* e *tRNA^{Gln}*. A estrutura secundária de todos os tRNAs anotados para os três genomas mitocondriais sequenciados neste trabalho possuem os anticódons descritos em outros genomas mitocondriais de insetos, com estruturas secundárias muito semelhantes. A análise das estruturas secundárias demonstrou que apenas o gene *tRNA^{Ser(AGC)}* não possui o braço DHU, sendo uma característica amplamente descrita entre os metazoários.

Em genomas mitocondriais de insetos, é comum a ocorrência do “codon usage” não otimizado, isto é, os anticódons presentes nos genes de tRNAs do mtDNA não são os códons mais utilizados nos GCPs mitocondriais. Isto ocorre devido à pressão de seleção por nucleotídeos A e T em genomas mitocondriais de insetos, ocasionando a utilização de códons mais ricos nestes nucleotídeos do que dos códons presentes no

genoma mitocondrial. Nestes casos, a evolução do “codon usage” é independente da evolução do anticódon, consistindo em uma hipótese de seleção por versatilidade do anticódon (Agris, 2004; Tong and Wong, 2004).

4.4. Região controle

A maior região não codificadora dos mtDNAs de *C. idioidea*, *C. vomitoria* e *P. regina* foi anotada como região controle (RC), levando-se em consideração a posição homóloga quando comparada com outras espécies da família Calliphoridae (entre o *rrnS* e o *tRNA^{Leu}*) e o alto conteúdo de bases A+T. A RC de espécies de califorídeos apresentou dois domínios diferentes: o domínio A, com blocos de sequência conservada (BSCs) e o domínio B, com sequência hipervariável.

As características da RC das espécies sequenciadas neste trabalho encontram-se na Tabela 10, juntamente com os dados de outras espécies de Calyptratae. É evidente o viés de bases As e Ts na RC (justificando seu nome alternativo em insetos de “AT-rich region”, ou região rica em AT), que apresenta um conteúdo próximo de 90% destes nucleotídeos. A variação do tamanho total entre os mtDNAs de diversas espécies frequentemente é atribuída à RC, devido ao tamanho que ela pode assumir, dependendo da espécie. A análise comparativa da RC é uma tarefa difícil, principalmente devido a regiões hipervariáveis presentes entre as espécies. Além disso, é comum encontrar genomas mitocondriais descritos como completamente sequenciados, mas em que a RC está ausente da sequência depositada (como é o caso das espécies de califorídeos *C. megacephala* e *L. sericata*). A RC de fato apresenta problemas durante a amplificação e clonagem que dificultam e algumas vezes inviabilizam sua obtenção e/ou sequenciamento. Uma análise mais abrangente e completa da RC de espécies de califorídeos foi desenvolvida ao longo deste trabalho e pode ser encontrada no Anexo C, em forma de um artigo publicado no periódico especializado *Journal of Medical Entomology*.

A estrutura das RCs de *C. idioidea*, *C. vomitoria* e *P. regina* encontra-se na Figura 10, na qual é possível visualizar a duplicação dos dois tRNAs no genoma mitocondrial de *P. regina*. A estrutura desta duplicação foi descrita previamente para três espécies de *Chrysomya* e também foi detectada nos mtDNAs de *Calliphora dubia* (Duarte et al., 2008) e *Stomoxys calcitrans* (família Muscoidea; Oliveira et al., 2007).

Tabela 10: Comparação A+T e tamanho da região controle de espécies de Calypratae.

Espécies	mtDNA (pb)	Região Controle	
		tamanho (pb)	% A+T
<i>C. idioidea</i>	~16000	1225	88.1
<i>C. vomitoria</i>	16143	1306	90.8
<i>P. regina</i>	16365	815	90.8
<i>C. putoria</i>	15837	818	89.6
<i>C. hominivorax</i>	16022	1175	90.8
<i>D. hominis</i>	16368	1553	91.1
<i>H. irritans</i>	16078	1259	89.4

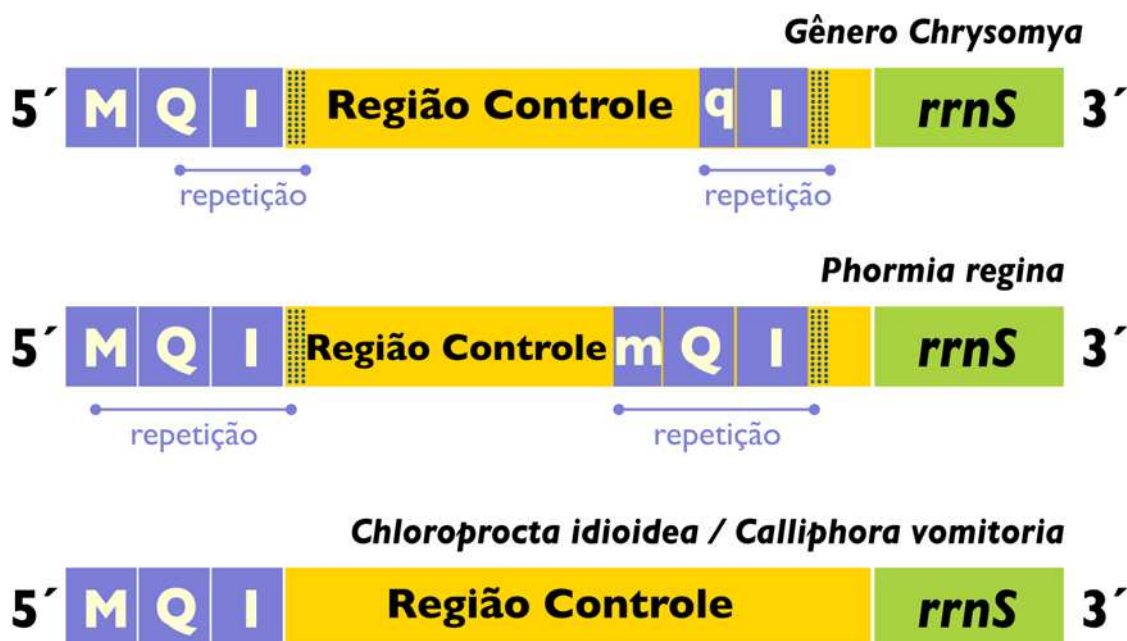


Figura 10: Estrutura da região controle nas espécies de califorídeos sequenciadas neste trabalho, comparadas à descrita para o gênero *Chrysomya*. Em evidência estão as regiões repetidas, que incluem os genes de *tRNA^{le}* (I) e *tRNA^{glu}* (Q) em *P. regina* e do *tRNA^{le}* em espécies de *Chrysomya*. As sequências parciais de tRNAs estão em letra minúscula (m = metionina; q = glutamina).

5. Análises filogenéticas

Foram realizadas reconstruções filogenéticas com diferentes partições de dados, incluindo a reconstrução com genes individuais do mtDNA, com o concatenado de todos os genes e com um sub-conjunto de dados escolhido a partir das análises de comparação de sítios conservados e variáveis (razão C/V, Figura 8). Para as reconstruções com genes individuais, utilizou-se somente a sequência de aminoácidos para minimizar a variação encontrada em muitos dos GCPs mitocondriais (*atp8*, *nd2* e *nd6*, por exemplo), enquanto as reconstruções com o concatenado das sequências foram feitas tanto com a sequência de nucleotídeos (excluindo-se a terceira posição do códon, que apresentou um alto índice de bases As e Ts, podendo introduzir muitas homoplasias nas análises) quanto com a sequência de aminoácidos. Para a obtenção das filogenias obtidas foram utilizados métodos de reconstrução amplamente reconhecidos, como a máxima parcimônia (MP), máxima verossimilhança (ML), neighbor-joining (NJ) e inferência bayesiana (BI).

Um total de 55 árvores filogenéticas foram avaliadas para compor os resultados apresentados neste trabalho, de modo a resumir as topologias obtidas. Atualmente, este é o trabalho que contempla o maior número de genomas mitocondriais de dípteros, incluindo as sequências de 28 espécies.

Ao analisar as topologias geradas com genes individuais, foi possível perceber que há uma tendência à obtenção de diferentes topologias com valores de suporte muito baixos nos ramos, para todos os métodos de reconstrução utilizados. Apesar das topologias variáveis, algumas relações entre clados se mantiveram em várias árvores, como a monofilia das duas sub-ordens de Diptera: Brachycera (moscas) e Nematocera (mosquitos). Ainda que com valores de bootstrap muito variáveis (35 a 100), estes dois grandes grupos foram monofiléticos em 75% das filogenias. A monofilia de Calliphoridae, por outro lado, foi obtida em todas as árvores analisadas. Um resultado inesperado é a inclusão de Muscidae (superfamília Muscoidea de Calyptratae, ver Figura 1) na superfamília Oestroidea em 65% das topologias, formando um clado que agrupa Calliphoridae+Muscidae, com a exclusão de Oestridae. Nas filogenias clássicas de Calyptratae, a superfamília Muscoidea aparece como um grupo irmão de Oestroidea, com o agrupamento (Muscidae(Calliphoridae,Oestridae)). Da mesma forma, Acalyptratae (representados nos conjuntos de dados pelas famílias Tephritidae e Drosophilidae) e Calyptratae (Oestridae, Calliphoridae e Muscidae) foram recuperadas como grupos separados em apenas 14% das árvores obtidas com genes

individuais, sendo que a sistemática morfológica clássica as classifica como dois grupos monofiléticos de Schizophora. Nas topologias obtidas, Drosophilidae e o clado Oestroidea+Muscoidea agrupam-se frequentemente, embora topologias que agrupam Tephritidae a Calyptratae também tenham sido com menor frequência.

Devido à variação de topologias obtidas, principalmente nas relações internas de famílias e gêneros com baixos suportes dos nós, as reconstruções realizadas com GCPs individuais não serão mostradas. As relações citadas acima, que se mantêm na maioria das árvores com os genes individuais, podem ser analisadas também nas árvores obtidas com o concatenado dos genes, porém com um maior suporte nos nós. Por estes motivos, as árvores obtidas com o concatenado de genes é que serão apresentadas com detalhes.

Os resultados obtidos a partir das reconstruções com o concatenado dos GCPs foram, portanto, provenientes de quatro conjunto de dados:

- i. concatenado da sequência nucleotídica (1^{as} e 2^{as} posições do códon) de um sub-conjunto de sete GCPs, incluindo *atp6*, *cox1*, *cox2*, *cox3*, *cytb*, *nd1* e *nd3* (subnt_12);
- ii. concatenado da sequência nucleotídica (1^{as} e 2^{as} posições do códon) de todos os GCPs mitocondriais (nt_12);
- iii. concatenado da sequência de aminoácidos do sub-conjunto de sete GCPs (sub_aa);
- iv. concatenado da sequência de aminoácidos dos 13 GCPs (aa_comp).

Cada conjunto de dados, por sua vez, foi submetido a reconstruções filogenéticas com os métodos MP, ML, NJ e BI.

As reconstruções com MP foram realizadas com a busca heurística CNI (“close neighbor interchange”) e adição randômica de árvores para a busca no espaço amostral, com deleção total dos “gaps” e 1000 réplicas de bootstrap. Um total de 868 sítios parcimoniosamente informativos foram considerados no conjunto subnt_12, 2057 no nt_12, 536 no sub_aa e 1282 no aa_comp.

As árvores de ML foram obtidas com as distâncias provenientes do ProtTest e ModelTest, que resultaram na matriz mtArt para sequências de aminoácidos, com as correções gamma (G; parâmetro para taxas de substituição diferentes entre sítios ao longo da sequência) e proporção de sítios invariáveis (I) estimada e otimizada. Devido

ao fato da matriz mtArt não estar implementada nos programas utilizados, a matriz mtRev+G+I foi utilizada nas análises por ter o valor de verossimilhança próximo ao da melhor opção sugerida pelo programa ProtTest. Para a sequência nucleotídica, utilizou-se o modelo de substituição GTR (“general time reversible”) com a correção gamma.

Em ambas as análises realizou-se 100 réplicas de bootstrap não paramétrico. As árvores de NJ foram obtidas com a distância Tamura-Nei+G para nucleotídeos e distribuição de Poisson para aminoácidos, desconsiderando-se os “gaps” e com 5000 réplicas de bootstrap.

Por fim, as árvores bayesianas foram reconstruídas com 1 milhão de gerações, correndo-se 2 cadeias em paralelo para aminoácidos ou 4 cadeias para nucleotídeos, com amostragem a cada 100 árvores e “burnin” de 25%. Os modelos de substituição foram obtidos da mesma maneira que os utilizados nas reconstruções de ML, com mtRev+G+I para aminoácidos e GTR+G para nucleotídeos. Tanto com o método BI como nos métodos NJ e ML foram considerados 4019 sítios para o conjunto de dados subnt_12, 7337 para nt_12, 2008 para sub_aa e 3652 para aa_comp.

A Figura 11 mostra a topologia obtida com o conjunto subnt_12 utilizando-se o método BI. Esta topologia foi escolhida por ser muito similar às recuperadas com os métodos ML e MP, além de apresentar o maior suporte nos nós. Com o método de NJ (Figura 12), a topologia obtida foi diferente principalmente na divisão do grupo Schizophora, recuperando a monofilia de Acalyptratae e Calyptratae (embora com um valor módico de bootstrap equivalente a 78), que não é obtida com outros métodos. A árvore BI (assim como a de MP e ML) indicam uma maior proximidade de Calyptratae com Drosophilidae.

Por outro lado, a monofilia da subfamília Chrysomyinae, que é recuperada com alto valor de probabilidade posterior (100) pelo método BI, não é suportada com NJ. Foi evidenciado, entretanto, que as relações internas da família Calliphoridae apresentaram baixo suporte com este método.

Nas duas topologias mostradas, o conjunto de dados subnt_12 não recuperou a monofilia de Brachycera, agrupando a espécie *C. duplonotata* às espécies de Nematocera.

A árvore com o método BI, obtida com o conjunto de dados nt_12 está na Figura 13, que agrega os dados de suporte de ramos obtidos com as outras três metodologias de reconstrução filogenética.

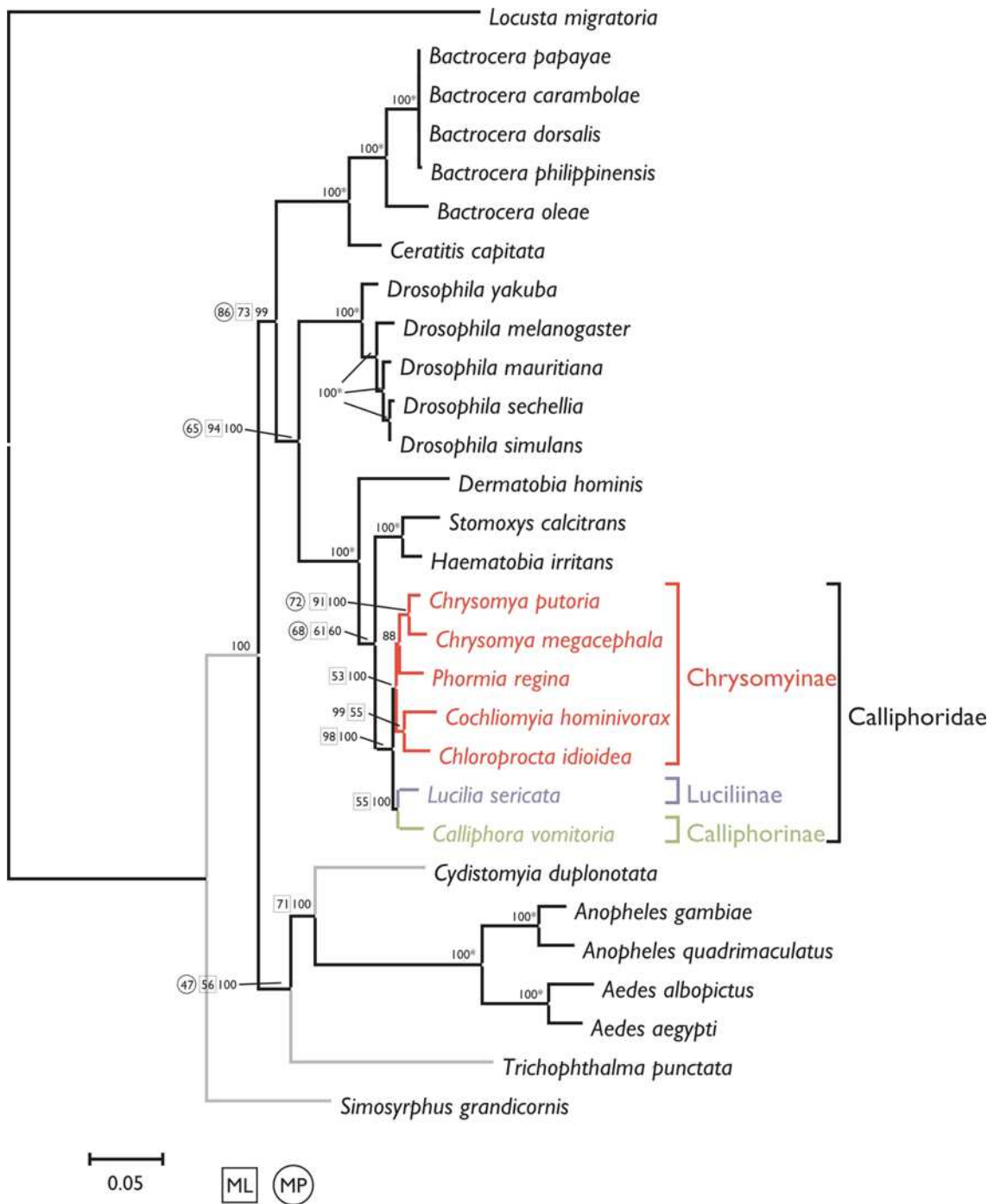


Figura 11: Árvore filogenética recuperada com o método BI (GTR+G) para o conjunto de dados subnt_12. O suporte nos nós é o valor de probabilidade posterior calculado pelo MrBayes, mas o valor do bootstrap não paramétrico obtido por ML (100 réplicas) e o de bootstrap da MP (1000 réplicas) estão representados nos nós em quadrados e círculos, respectivamente. Os valores assinalados com "*" indicam que suportes >95 foram obtidos por outros métodos para este nó, enquanto que a ausência de valores em nós não assinalados com "*" indica que o método não suporta o ramo. Valores de suporte <50 foram omitidos. Os ramos em cinza indicam que outros métodos não recuperaram a mesma relação mostrada nesta topologia e os ramos coloridos mostram a relação obtida entre subfamílias de Calliphoridae. A espécie *Locusta migratoria* (Orthoptera) foi utilizada como grupo externo

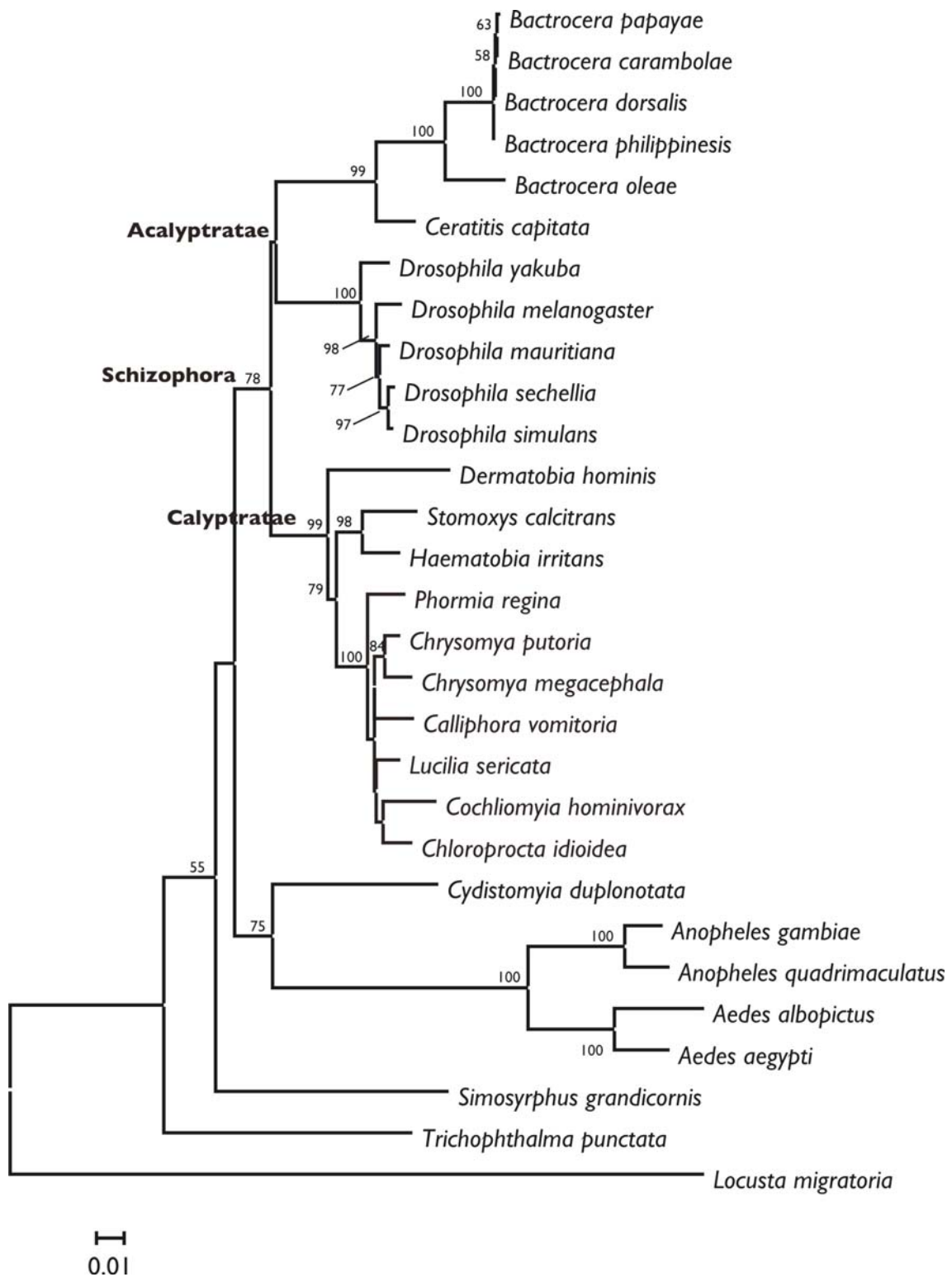


Figura 12: Árvore de NJ para o conjunto de dados subnt_12 (Tamura-Nei+G, 5000 réplicas de bootstrap), indicando a divisão do grupo Schizophora em Calypttratae e Acalyptratae. Os valores de bootstrap <50 foram omitidos. *L. migratoria* foi incluída como grupo externo.

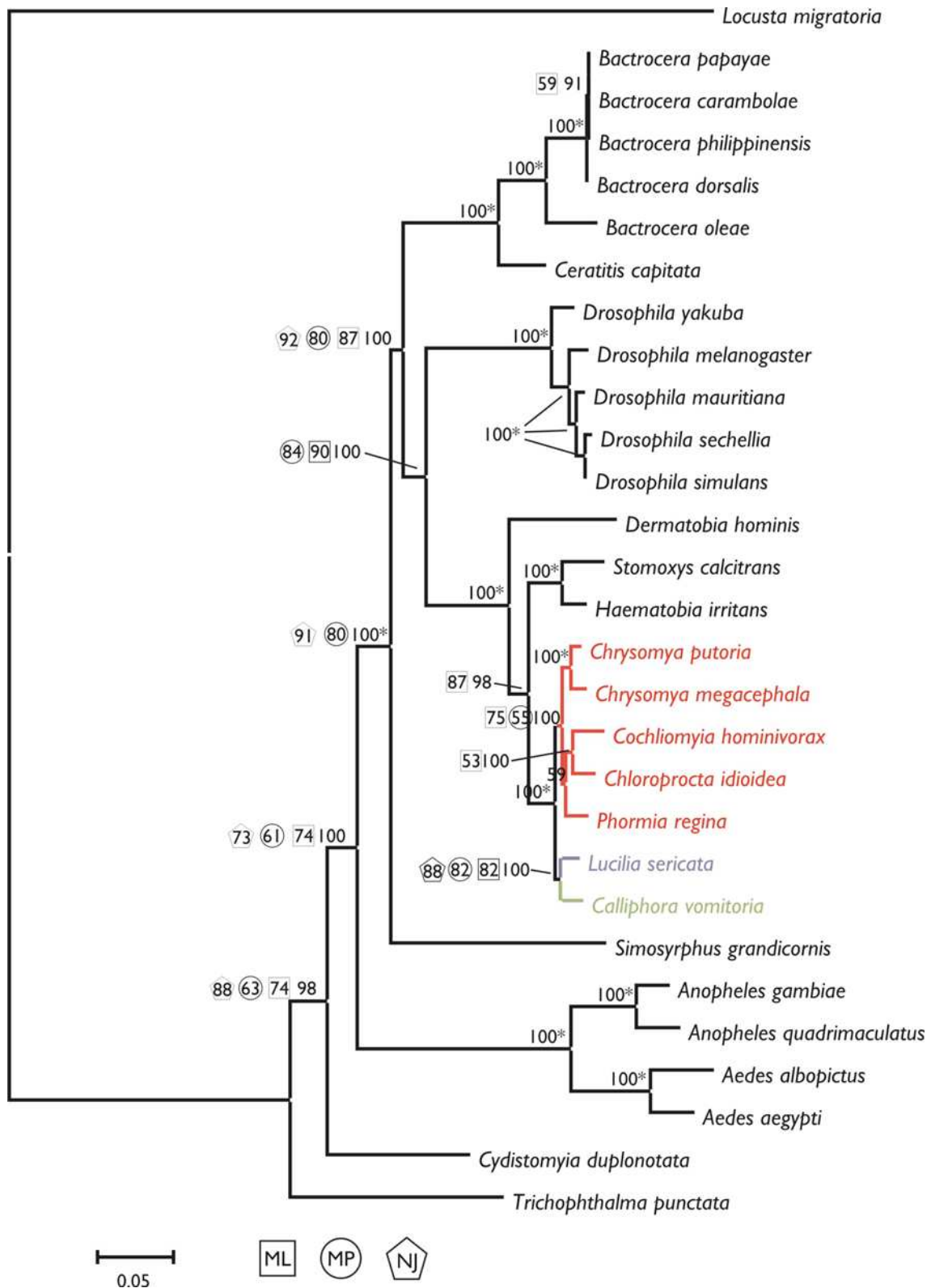


Figura 13: Árvore filogenética recuperada com o método BI (GTR+G) para o conjunto de dados nt_12. O suporte nos nós é o valor de probabilidade posterior, mas o valor do bootstrap não paramétrico obtido por ML (100 réplicas), do bootstrap da MP (1000 réplicas) e de NJ (5000 réplicas) estão representados nos nós em quadrados, círculos e pentágonos, respectivamente. Os valores assinalados com "*" indicam que suportes >95 foram obtidos por outros métodos para este nó, enquanto que a ausência de valores em nós não assinalados com "*" indica que o método não suporta o ramo. Valores de suporte <50 foram omitidos. Os ramos coloridos mostram a relação obtida entre subfamílias de Calliphoridae. A espécie *Locusta migratoria* (Orthoptera) foi utilizada como grupo externo.

Algumas semelhanças entre as árvores reconstruídas com as sequências de nucleotídeos são evidentes. Entre elas, o fato de que ambos os conjuntos de dados não recuperam a monofilia de Brachycera, independentemente do método de reconstrução utilizado. No entanto, apenas os valores de probabilidade posterior gerados pelo MrBayes é que são significativos (98 a 100). Embora os valores de bootstrap de ML, NJ e MP sejam maiores para o conjunto de dados que inclui todos os GCPs mitocondriais, continuam com um valor abaixo do intervalo de confiança desta estatística, que deve ser >95.

Na topologia da Figura 13 (nt_12), a subfamília Chrysomyinae é monofilética, embora os valores de suporte sejam baixos para MP e ML (<95). O método de NJ não suporta esta monofilia, possuindo valor <50. Assim como para o sub-conjunto subnt_12, a espécie *C. vomitoria* e *L. sericata* aparecem como grupos-irmãos, cada uma representando outras subfamílias de Calliphoridae: Calliphorinae e Luciliinae, respectivamente. Este agrupamento apresenta valores muito maiores quando utiliza-se o conjunto de todos os GCPs mitocondriais.

Assim como para a grande maioria das árvores reconstruídas com genes individuais, as espécies de Muscidae (superfamília Muscoidea) *H. irritans* e *S. calcitrans* agrupam-se com Calliphoridae, rompendo a monofilia de Oestroidea.

As relações entre os dois grupos de Schizophora (Calypttratae e Acalypttratae) também repetem o padrão encontrado nas árvores de ML, MP e BI com o conjunto subnt_12, apresentando o agrupamento (Tephritidae(Drosophilidae,Calypttratae)). O método de NJ não suporta esta relação em nenhuma das árvores reconstruídas com sequências nucleotídicas, resultando em uma politomia neste ramo da árvore consenso de bootstrap no conjunto nt_12.

Por outro lado, nos dois conjuntos de sequências nucleotídicas, as relações internas de Drosophilidae são recuperadas com alto valor de suporte em seus nós, assim como as relações de Tephritidae. Apenas os ramos terminais das espécies *B. papayae*, *B. philippinensis*, *B. dorsalis* e *B. carambolae* não apresentam suporte para resolver as relações deste sub-grupo do gênero *Bactrocera*.

As árvores obtidas com o sub-conjunto de dados de aminoácidos estão nas Figuras 14 e 15, representando a topologia obtida pelos métodos de ML (com os suportes de MP e BI) e NJ, respectivamente.

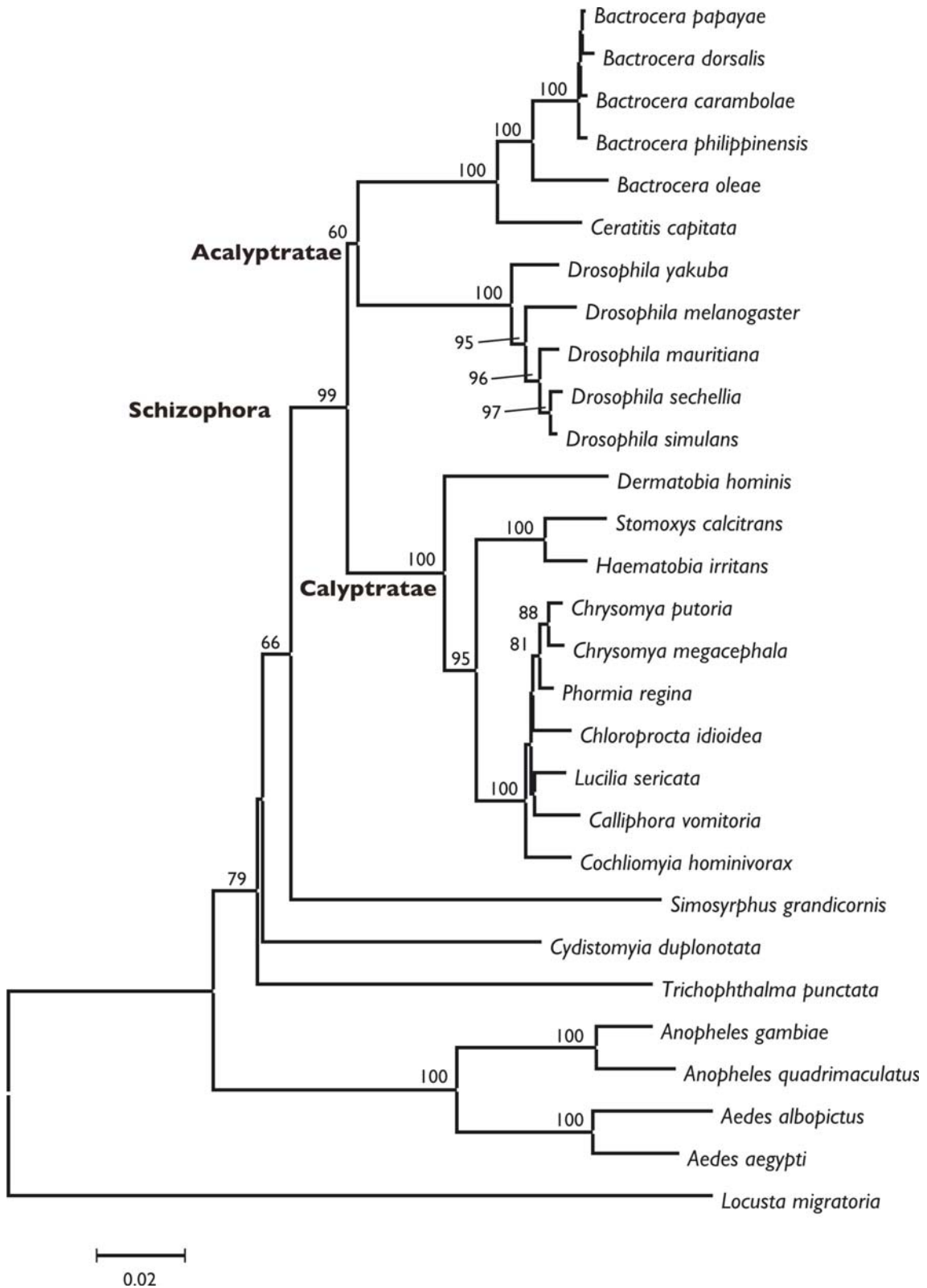


Figura 15: Árvore de NJ para o conjunto de dados sub_aa (correção de Poisson+G, 5000 réplicas de bootstrap), indicando a divisão do grupo *Schizophora* em *Calyptratae* e *Acalyptratae*. Os valores de bootstrap <50 foram omitidos. *L. migratoria* foi incluída como grupo externo.

Comparando-se as árvores obtidas com a sequência de aminoácidos do subconjunto de GCPs (sub_aa), foi evidenciado que as relações entre Brachycera e Nematocera são recuperadas de acordo com a filogenia clássica de Diptera, havendo a divisão em dois grupos monofiléticos. Apenas a árvore de BI não recuperou esta divisão, apresentando uma topologia semelhante às árvores obtidas com o concatenado de nucleotídeos. No entanto, todas as outras metodologias mostraram claramente esta relação.

As reconstruções de MP e ML com os dados sub_aa recuperaram as relações basais de Brachycera, indicando a monofilia de Muscomorpha. Embora com valores de suporte baixo, as relações entre *S. grandicornis*, *T. punctata* e *C. duplonotata* foram recuperadas de acordo com a sistemática morfológica de Brachycera.

A árvore de NJ novamente mostra a subdivisão Schizophora dividida em dois grupos monofiléticos: Calyptratae e Acalyptratae. A exemplo da topologia recuperada com o sub-conjunto subnt_12, as sequências de aminoácidos combinadas ao método de NJ recuperaram a monofilia destes dois subgrupos, mas com maiores valores de bootstrap. Nenhum dos métodos utilizados para reconstruir filogenias com o sub_aa recuperaram a monofilia de Chrysomyinae, embora tenham agrupado Calliphorinae+Lucilinae.

As árvores obtidas com o concatenado de todos os GCPs mitocondriais, consistindo no conjunto aa_comp estão nas Figuras 16 e 17. Na Figura 16, a árvore de NJ evidencia a monofilia de Chrysomyinae e o agrupamento de Calliphorinae+Luciliinae. Assim como na maioria das árvores obtidas com o sub_aa (exceto para o método BI), Nematocera e Brachycera são grupos monofiléticos.

As relações internas entre Drosophilidae e Tephritidae, já mencionadas para as reconstruções obtidas com sequências nucleotídicas, se mantêm nas topologias de sequências de aminoácidos (tanto sub_aa como aa_comp). Da mesma forma, o agrupamento (Oestridae(Muscidae,Calliphoridae)) é recuperado em todas as árvores, com altos valores de suporte em todos os métodos e todos os conjuntos de dados, assim como a monofilia de Calliphoridae.

A subfamília Chrysomyinae apareceu como monofilética em nove das dezesseis árvores apresentadas, com suportes variáveis, mas na maior parte das vezes <95 (exceto para algumas topologias inferidas com o método bayesiano). Em todas as árvores reconstruídas com a sequência nucleotídica dos 13 GCPs, Chrysomyinae foi

recuperada como monofilética. O agrupamento de Luciliinae+Calliphorinae é recuperado em todas as topologias, embora com suportes variáveis.

É interessante notar que a monofilia de Muscomorpha é recuperada em apenas três árvores, todas com sequências de aminoácidos (Figura 14, com os métodos de MP e ML e na Figura 17, com o método BI).

As sub-divisão de Schizophora em Calyptratae e Acalyptratae foram recuperadas em apenas duas árvores de NJ (subnt_12 e sub_aa) das 16 analisadas com sequências concatenadas, ambas envolvendo o sub-conjunto de sete GCPs. Nas outras reconstruções, o agrupamento de Drosophilidae com espécies de Calyptratae (Calliphoridae, Oestridae e Muscidae) foi a relação comumente recuperada.

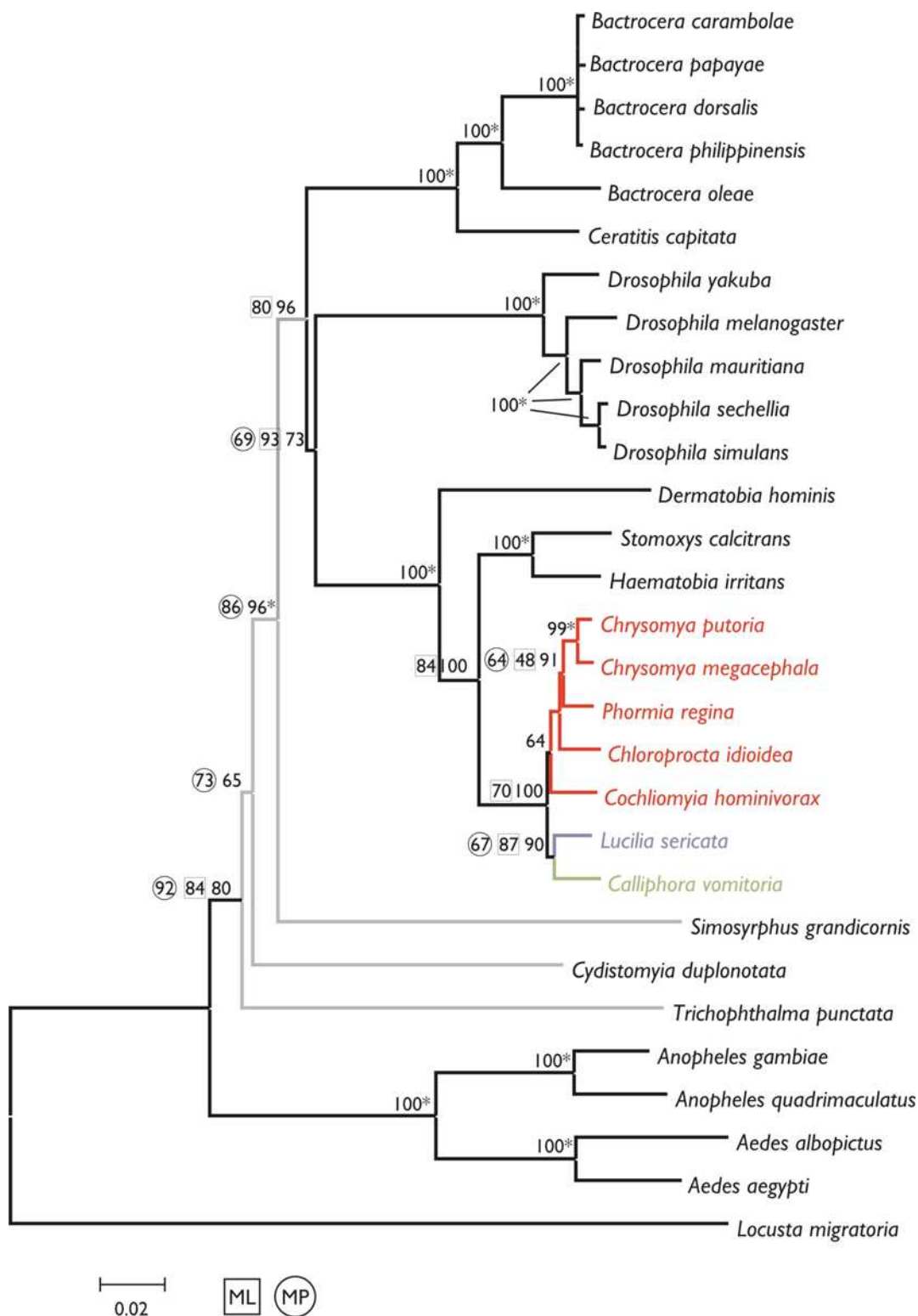


Figura 16: Árvore filogenética recuperada com o método NJ (Poisson+G) para o conjunto de dados aa_comp. O suporte nos nós é o valor de bootstrap (5000 réplicas), mas o valor do bootstrap da MP (500 réplicas) e bootstrap não-paramétrico (100 réplicas) estão representados nos nós em círculos e quadrados, respectivamente. Os valores assinalados com "*" indicam que suportes >95 foram obtidos por outros métodos para este nó, enquanto que a ausência de valores em nós não assinalados com "*" indicam que o método não suporta o ramo. Valores de suporte de nós <50 foram omitidos. Os ramos em cinza indicam que outros métodos não recuperaram a mesma relação mostrada nesta topologia e os ramos coloridos mostram a relação obtida entre subfamílias de Calliphoridae. *L. migratoria* foi utilizada como grupo externo.

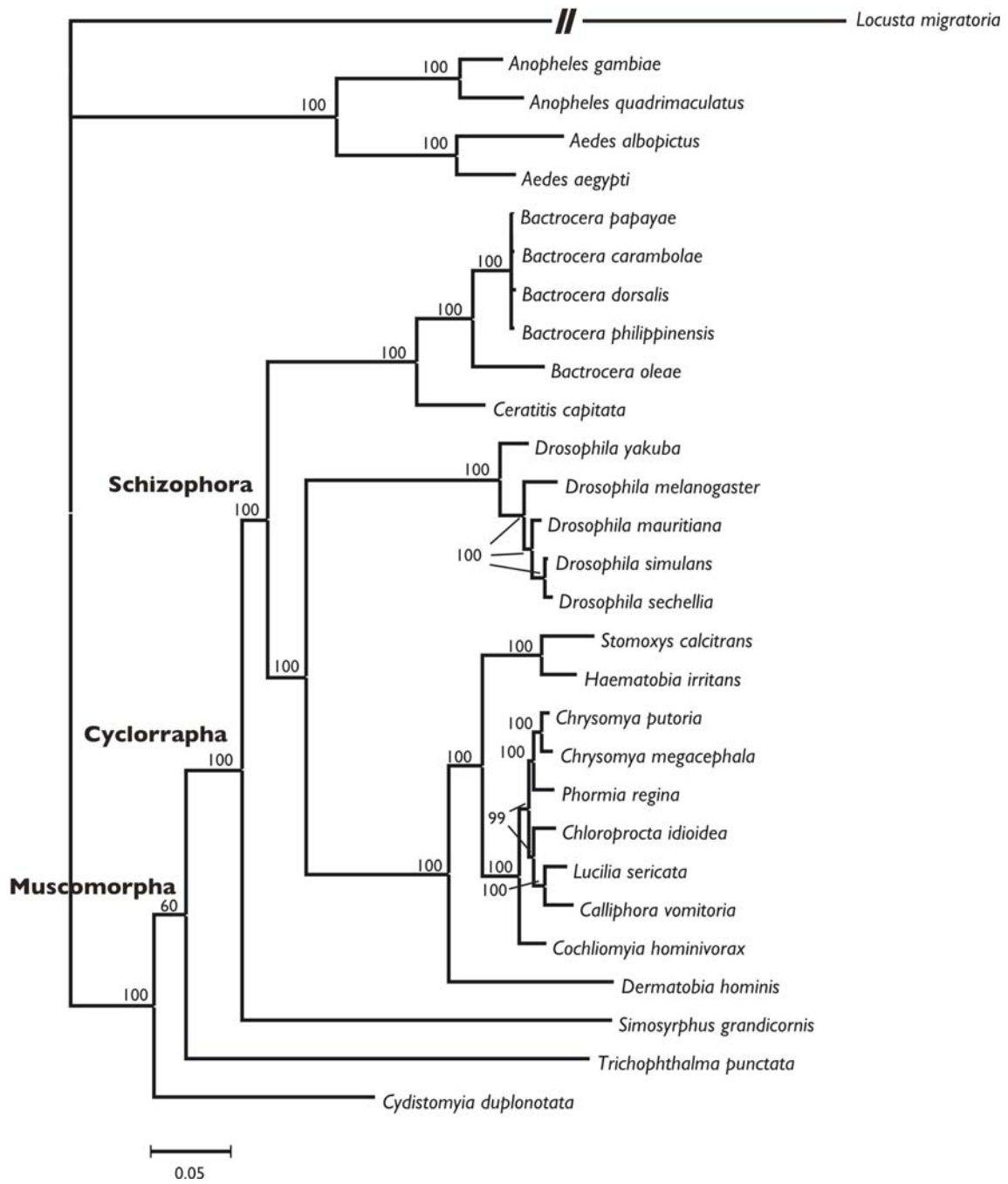


Figura 17: Árvore de BI para o conjunto de dados aa_comp (mtREV+G+I), indicando a divisão do grupo Schizophora em Calyptratae e Acalyptratae. Os valores de bootstrap <50 foram omitidos. *L. migratoria* foi incluída como grupo externo.

DISCUSSÃO

I. Sequenciamento do mtDNA de *C. idioidea*, *C. vomitoria* e *P. regina*

A estratégia de “long-PCR” associada ao sequenciamento por “shotgun” de bibliotecas com 1-2 kb demonstrou-se eficiente para a obtenção das sequências do mtDNA de *C. idioidea*, *C. vomitoria* e *P. regina*, permitindo o fechamento de grande parte de cada um dos genomas com cerca de 200 “reads”, que por sua vez formaram contigs de alta qualidade (>40 ao analisar no programa phred). A amplificação de fragmentos longos foi eficientemente realizada para outras espécies de califorídeos, além das sequenciadas neste trabalho, como para *Lucilia eximia*, *Chrysomya albiceps* e *Cochlomyia macellaria*, indicando que a técnica foi padronizada com eficácia. Espécies de Hexapoda de outras ordens também foram testados por Barau et al. (2005), permitindo o acesso ao mtDNA completo de espécies de diferentes níveis taxonômicos de insetos.

Em particular, foi possível observar que sequências da região controle do mtDNA das três espécies tiveram um baixo índice de sequenciamento a partir das bibliotecas realizadas, provavelmente devido a dificuldades na clonagem desta região, que levaram à necessidade de uma abordagem paralela para seu completo sequenciamento. Assim, foi necessário amplificar e clonar individualmente a região controle para que a mesma pudesse ser sequenciada e montada no restante do genoma. De fato, dificuldades metodológicas de acesso à sequência da região controle são comuns, principalmente devido ao potencial de formação de estruturas secundárias potencialmente envolvidas na origem de replicação do genoma mitocondrial. Estas características podem influenciar a eficiência da amplificação, assim como da ligação em plasmídeos. A constituição rica em bases A e T também pode interferir na recuperação eficiente desta região, devido às propriedades termoestáveis e à maior facilidade de incorporar erros durante a amplificação e sequenciamento. Isto ocorre principalmente em regiões de sequências repetidas *in tandem*, como o domínio de repetições dinucleotídicas TA, e de homopolímeros como os domínios poli-T e poli-A (Levinson and Gutman, 1987).

A região estimada em ~800 pb, que compreende um “gap” na sequência de *C. idioidea*, corresponde ao fim dos amplicons A e B, ambos amplificados com oligonucleotídeos ancorados no gene *rrnL*. A formação de um “gap” de menor

extensão é previsto com a abordagem utilizada, uma vez que a construção dos “primers” no *rrnL* não amplificam fragmentos sobrepostos, como os ancorados na região *cox3* (ver Figura 2 na seção Materiais e Métodos). No entanto, este “gap” criado pela composição de “primers” empregados nas reações de “long-PCR” pode ser facilmente sequenciado com o emprego de “primers” descritos para esta região (Barau et al., 2005). A dificuldade encontrada em *C. idioidea* é que a qualidade baixa da ponta dos contigs compreendia uma região maior do que coberta por esta composição de “primers”, não permitindo sua recuperação completa. Ainda assim, o fechamento desta região está em andamento para que o depósito das sequências seja realizado no GenBank. Por ora, é possível acessar a anotação das sequências realizadas neste trabalho através da página <http://amiga.cbmeg.unicamp.br/>. Na seção “user options”, entrar em “login” e utilizar o usuário “tese” com a senha “2902”.

2. Arquitetura genômica e composição A+T

A ordem e conteúdo gênicos do mtDNA de *C. idioidea* e *C. vomitoria* (Figura 5) são idênticos aos descritos para os insetos basais e crustáceos, possuindo o arranjo ancestral de genomas mitocondriais de Pancrustacea (Nardi et al., 2003; Boore, 1999) e uma região não-codificadora atribuída à região controle do mtDNA. A espécie *P. regina* também possui o mesmo arranjo e conteúdo gênicos, exceto pela existência de uma duplicação existente na região controle envolvendo o *tRNA^{Leu}* e *tRNA^{Gln}*.

A composição A+T dos três genomas sequenciados segue o padrão descrito para insetos, com 71.7%, 72.9% e 75.6% em *C. idioidea*, *P. regina* e *C. vomitoria*, respectivamente. O conteúdo A+T de *C. idioidea* é o mais baixo entre os 28 dípteros analisados (Tabela 3), provavelmente devido ao fato de que parte do gene *rrnL* não está incluído nesta análise. Comparando-se o conteúdo A+T dos GCPs da “fita +” e da “fita -” desta espécie com todas as outras espécies de dípteros, os padrões de composição A+T são compatíveis com os descritos na literatura, reforçando a hipótese de que o valor ligeiramente menor de sua composição A+T na molécula inteira está relacionada à não inclusão do *rrnL*, que é altamente enviesado para a composição destes nucleotídeos (84% em *P. regina* e 85% em *C. vomitoria*).

As análises das frequências relativas entre As e Ts (AT-skew) e Gs e Cs (GC-skew) no mtDNA total (Tabela 3) ressaltam que os valores de AT-skew de 0.021 para *C. idioidea* e *P. regina* e de 0.018 são baixos e congruentes com as outras espécies de Diptera, indicando que As e Ts ocorrem praticamente na mesma proporção, embora

com uma quantidade maior de As. Por outro lado, o GC-skew negativo (-0.157, -0.163 e -0.170 em *C. idioidea*, *C. vomitoria* e *P. regina*, respectivamente) indica a presença de um maior número de Cs do que Gs. As análises realizadas com o mtDNA completo foram feitas com a “fita +” do mtDNA, indicando, portanto, uma maior quantidade de As e Cs ao longo de toda esta fita. De acordo com Saccone et al. (1999), o mtDNA de metazoários apresenta um desbalanceamento entre estes dois nucleotídeos, sendo comumente reportada uma menor frequência de Gs e Ts na "fita +". Este viés no acúmulo As e Cs seria causado por características intrínsecas à replicação assimétrica e unidirecional do mtDNA (Goddard and Wolstenholme, 1978), em que a fita “lagging” (“fita +”) inicia sua replicação quando 97% da fita “leading” (“fita -”) já está replicada, ocasionando a exposição da "fita -" parental por muito tempo na forma de simples fita. Esta exposição, por sua vez, deixa os nucleotídeos mais suscetíveis à desaminação. Deste modo, a desaminação preferencial de C→U na "fita -" parental ocasiona o aumento de As e a diminuição de Gs na fita complementar ("fita +" filha). Da mesma maneira, o nucleotídeo A é alvo de desaminação, tornando-se uma hipoxantina que é complementada por C na fita “+” filha. Um esquema da replicação assimétrica do mtDNA pode ser vista na Figura 18.

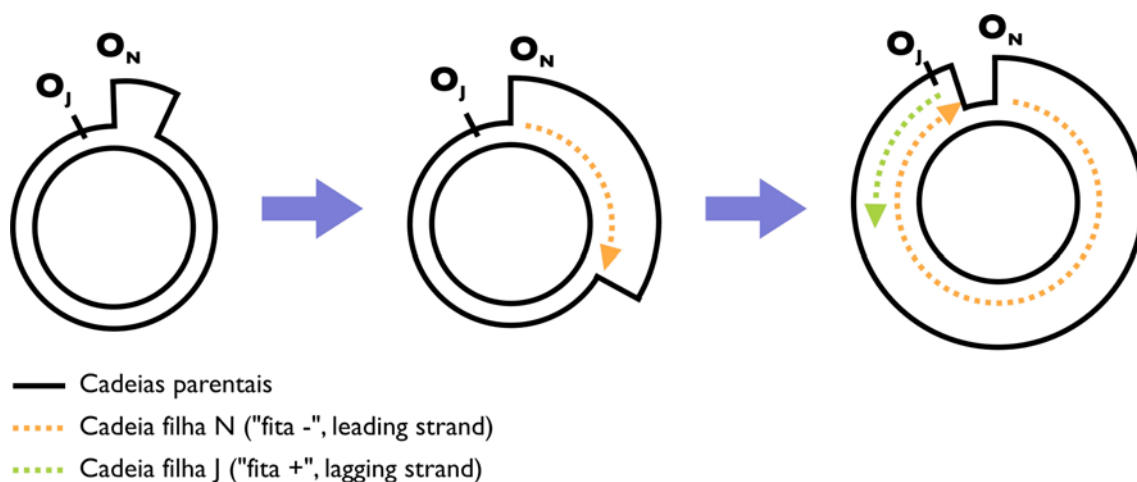


Figura 18: Esquema da replicação assimétrica no mtDNA de insetos (adaptado de Saccone et al., 1999).

Para abordar a hipótese de manutenção de um maior índice de As e Cs na "fita +" do mtDNA, Stewart and Beckenbach (2006) realizaram uma análise interessante, em que consideraram apenas os nucleotídeos na terceira posição do

códon de aminoácidos codificados por um conjunto de quatro códons degenerados (“four-fold degenerate codons”). Desta maneira, qualquer um dos nucleotídeos A, T, G ou C na terceira posição do códon resultariam no mesmo aminoácido. Ao analisarem o mtDNA completo de dezenove espécies de insetos (incluindo a espécie de californídeo *C. putoria*), Stewart and Beckenbach (2006) descreveram um maior acúmulo de As e Cs na terceira posição de códons 4 vezes degenerados da "fita +", mas os conjuntos de genes que são posicionados mais longe da origem de replicação na região controle (e que portanto ficam mais tempo expostos como simples-fita) mostraram um aumento significativo na frequência de As e Cs nos códons analisados. Mais do que apenas descrever um viés de nucleotídeos, este trabalho corrobora indiretamente a hipótese proposta por Goddard and Wolstenholme (1978) de replicação unidirecional e assimétrica do genoma mitocondrial em insetos, que é uma área pouco estudada e permeada por contradições, uma vez que os trabalhos referentes à replicação do mtDNA de insetos foram realizados, principalmente, com microscopia eletrônica e endonucleases de restrição.

Com relação aos outros valores de AT e GC-skew encontrados em partições de genes mitocondriais (GCPs+, GCPs-, tRNAs e rRNAs), as espécies *C. idioidea*, *C. vomitoria* e *P. regina* seguem os padrões de composição nucleotídica encontradas em todos os dípteros analisados (Tabela 3).

Ao avaliar os GCPs codificados pela "fita +", é possível verificar a maior presença de Cs, devido aos valores negativos encontrados no GC-skew, porém os valores também negativos de AT-skew indicam a maior quantidade de Ts nos GCPs desta fita. Estes valores não anulam a hipótese descrita acima, uma vez que as análises não foram realizadas para códons 4 vezes degenerados. Os GCPs da "fita -" indicam, entretanto, uma maior frequência de Ts e Gs (Tabela 3).

Apesar da proposta de assimetria relativa entre nucleotídeos A/T e G/C medidas pelos valores de “skewness”, é inegável que, de um modo geral, o genoma mitocondrial de insetos é rico em A+T. Análises detalhadas dos padrões dos nucleotídeos A e T em regiões codificadoras e não codificadoras, da função das posições dos códons e de estruturas secundárias levaram alguns autores a sugerirem a existência de uma pressão seletiva para um alto conteúdo das bases A e T (Clary and Wolstenholme, 1985; Crozier and Crozier, 1993). Esta pressão seletiva parece ser mais acentuada em posições nucleotídicas não codificadoras, sendo evidente na região controle do mtDNA.

Zhang and Hewitt (1997) sugerem que um alto conteúdo de A+T apresentado por todos os insetos cuja região controle já foi sequenciada pode ser uma característica filogenética comum e ancestral deste taxon, decorrente de mutação direcional. De fato os insetos são organismos de metabolismo acelerado, necessitando de uma grande eficiência na produção e replicação da fonte de energia celular: a mitocôndria. Como a região controle está relacionada à origem de replicação e início da transcrição do mtDNA, a rápida desnaturação da dupla cadeia nesta região é fundamental para a otimização metabólica nas células. Esta interpretação está intimamente relacionada à hipótese proposta por Kondo *et al.* (1990) e reforçada por Ballard (2000), em que o desvio de bases A e T pode ser resultado de uma seleção, já que estes genomas replicam-se mais rapidamente e requerem menos energia durante as etapas de replicação e transcrição do que genomas ricos em G+C.

Uma outra explicação para o desvio de composição nucleotídica no genoma mitocondrial foi proposta por Xia (1996), que sugeriu uma hipótese de maximização da eficiência transcricional em que a utilização dos códons depende do nucleotídeo que é mais abundante na matriz celular, uma vez que ele deve ser mais utilizado no mRNA. Na mitocôndria, códons terminados com A são utilizados mais frequentemente em substituições sinônimas, uma vez que o ATP é muito mais abundante do que os outros três ribonucleotídeos, representando a principal forma de armazenamento de energia celular.

Martin (1995) também propõe uma hipótese para explicar as diferenças na composição nucleotídica em genomas mitocondriais de insetos. Esta tendência à alta composição em bases A e T pode ser causada pela maior suscetibilidade do dGTP e do dCTP a danos oxidativos e conseqüente incorporação preferencial do dATP pela polimerase em sítios sem bases. A tendência também pode ser causada devido à conversão de citosinas em uracilas através da desaminação hidrolítica (Ballard, 2000).

3. Genes codificadores de proteínas

Os códons de iniciação e terminação descritos na Tabela 4 demonstram algumas peculiaridades do genoma mitocondrial animal, em que códons de terminação incompletos são comumente descritos. Acredita-se que os transcritos policistrônicos com os pré-RNAs são reconhecidos por uma maquinaria de processamento ainda desconhecida, mas que teoricamente seguem um sistema de pontuação por tRNA (Ojala *et al.*, 1980). Neste sistema, os grandes transcritos sofrem a remoção e

maturação dos tRNAs, que liberam moléculas de rRNA e mRNA (Ojala et al., 1981). A poliadenilação dos mRNAs é necessária, através da adição de uma cauda poli-A, que completaria os códons de terminação que possuem apenas os nucleotídeos T ou TA no quadro de leitura (como no caso dos genes *nd4* e *nd5*).

A ocorrência de uma Serina ao invés da Metionina como o primeiro aminoácido de COXI tem sido descrita para um grande número de espécies de insetos. No entanto, foi possível reconhecer a sequência ATTTAA em uma região "upstream" ao códon TCG nas espécies *C. idioidea*, *C. vomitoria* e *P. regina*. A presença deste hexanucleotídeo tem sido sugerida como um elemento possivelmente relacionado à sinalização do início da transcrição de *cox1*, já detectado em mosquitos (Mitchell et al., 1993; Beard et al., 1994) e no inseto basal *Tetradontophora bielensis* (Nardi et al., 2001). Uma avaliação da presença deste hexanucleotídeo em sequências do gene *cox1* mostrou que os dípteros *C. capitata*, *B. oleae*, *H. irritans*, *Stomoxys calcitrans* e sete espécies de califorídeos (*Chrysomya rufifacies*, *C. albiceps*, *C. megacephala*, *Cochliomyia hominivorax*, *Calliphora vicina*, *Lucilia caesar*, *L. cuprina* e *L. sericata*) apresentam a sequência ATTTAA flanqueando o códon de iniciação TCG, podendo atuar como um sinalizador do início da transcrição deste gene. Estudos com transcritos policistrônicos do mtDNA são poucos e inconclusivos com respeito à iniciação de *cox1*.

A Figura 6 mostra a proporção relativa de cada nucleotídeo nas primeiras, segundas e terceiras posições dos códons dos genes mitocondriais das espécies sequenciadas. A alta representação de Ts na segunda posição do códon sugere uma seleção para a manutenção de aminoácidos apolares e hidrofóbicos em proteínas associadas à membrana (Asakawa et al., 1991), já que os aminoácidos Fenilalanina, Leucina, Isoleucina, Metionina e Valina são apolares e hidrofóbicos, apresentando o nucleotídeo T na segunda posição do códon (Helfenbein et al., 2001). Ao analisar as Tabelas 5, 6 e 7, é possível verificar que há uma grande utilização dos códons com T na segunda posição, quando comparados às outras posições. Essa análise se completa com a Tabela 8, que mostra a frequência de aminoácidos em proteínas mitocondriais. É evidente que o aminoácido mais utilizado em todas as espécies é a Leucina, seguido por Serina, Fenilalanina e Isoleucina, que juntos representam mais de 40% dos aminoácidos utilizados nas proteínas mitocondriais. Os aminoácidos apolares são os mais utilizados, corroborando a hipótese de seleção por Ts na segunda posição do

códon (Asakawa *et al.*, 1991), por se tratarem da composição de proteínas com domínio transmembrânicos.

A utilização preferencial de códons terminados em A e T (Tabelas 5, 6 e 7) está de acordo com a Figura 6, que indica a presença destes nucleotídeos em ~90% das terceiras posições dos códons. Da mesma maneira, códons com composição G e C tendem a ser pouco ou não utilizados pelo genoma mitocondrial de insetos.

O grau de conservação de sequências de aminoácidos dos GCPs mitocondriais de dípteros foi abordado de diferentes maneiras. A Figura 7 mostra o número absoluto de sítios conservados e variáveis nos GCPs, enquanto a Figura 8 evidencia a relação entre estes sítios (razão C/V). Ambas indicam os genes que possuem mais sítios conservados do que variáveis. Estas análises podem guiar a escolha de conjuntos de dados ou de genes individuais, para adequar a variabilidade do marcador escolhido ao problema taxonômico de interesse.

A Tabela 9, por outro lado, evidencia a amplitude de divergência nucleotídica encontrada para cada gene, ao analisar uma matriz de distância p com comparações par a par das 28 espécies de dípteros consideradas neste trabalho. Estes dados indicam as menores divergências encontradas, que são medidas intra-gêneros, e as maiores divergências, entre sub-grupos de Brachycera. A variação da divergência também pode ser útil para guiar escolhas de genes individuais ou combinações de genes para abordar questões nos mais diferentes níveis hierárquicos.

Por um outro ângulo, a identidade de aminoácidos para cada sítio ao longo dos GCPs mitocondriais, além de providenciar um panorama geral a respeito da variação de cada gene individualmente, também torna possível visualizar a variação ao longo de diferentes regiões das moléculas, permitindo a avaliação de regiões que se adequem ao problema a ser abordado. Esta análise também pode ser particularmente útil para a construção de “primers” que amplifiquem regiões específicas de cada gene.

4. Região controle e duplicação de tRNAs

Baseando-se na comparação de sequências da região controle (RC), Zhang and Hewitt (1997) dividiram os insetos em dois grupos. O Grupo 1 compreende os insetos em que esta região contém dois domínios diferentes, como entre espécies de *Drosophila*. Um dos domínios é conservado (adjacente ao gene *tRNA^{Leu}*) e o outro é variável tanto em sequência como em tamanho. Repetições *in tandem* ocorrem em ambos os domínios. O Grupo 2 compreende os insetos em que a região controle não

pode ser dividida em domínios conservados e variáveis distintos, como em louva-a-deus, mosquitos e borboletas.

Em espécies da família Calliphoridae foi possível identificar duas subregiões com taxas de substituição nucleotídica distintas: um domínio com blocos de sequência conservada (domínio A) que possivelmente abrigam a origem de replicação do genoma mitocondrial; e outra com sequências altamente variáveis (domínio B; Lessinger and Azeredo-Espin, 2000). Deste modo, os califorídeos pertencem ao Grupo I.

A RC de insetos possui um alto viés em bases A+T, podendo apresentar >90% de composição destes nucleotídeos (Tabela 10). Uma análise detalhada da RC de califorídeos está no Apêndice C, que fez parte de uma extensão das análises geradas no início deste trabalho.

Uma característica muito peculiar encontrada na espécie *P. regina* é a duplicação de uma região adjacente à RC, incluindo a sequência completa dos genes para $tRNA^{Ile}$ e $tRNA^{Gln}$ e sequências parciais do início do domínio A da RC e do $tRNA^{Met}$ (Figura 10). A mesma estrutura desta duplicação foi encontrada previamente no domínio B (hipervariável) nas espécies de califorídeos *Chrysomya putoria*, *C. albiceps* e *C. megacephala*, mas envolvendo a duplicação completa apenas do $tRNA^{Ile}$ (Lessinger et al., 2004; Junqueira et al., 2004). A caracterização da RC de 15 espécies de califorídeos também indicou a presença desta duplicação na espécie *Calliphora dubia*, *Chrysomya rufifacies* e *C. bezziana* (Duarte et al., submetido. Anexo C). Além destas, a espécie de Muscidae *Stomoxys calcitrans* também apresenta a duplicação do $tRNA^{Ile}$ (Oliveira et al., 2007).

A análise das duas cópias dos dois tRNAs duplicados de *P. regina* indicou que os genes são idênticos. Os mesmos resultados foram encontrados em outros califorídeos que apresentam as duplicações, com eventuais substituições pontuais sendo mantidas entre as cópias da mesma espécie. Isto indica que algum mecanismo de homogeneização, como evolução em concerto, pode estar mantendo as cópias idênticas e funcionais.

A ocorrência de duplicações no gênero *Chrysomya*, em *P. regina*, *C. dubia* e na espécie de muscídeo *S. calcitrans* (pertencente a superfamília Muscoidea) pode sugerir que estes eventos ocorreram paralela e independentemente, o que indicaria o domínio B da RC como um “hot spot” para duplicações mitocondriais.

5. Análises filogenéticas

A partir das árvores filogenéticas reconstruídas com os métodos ML, MP, NJ e BI, foi possível notar que as partições de dados têm mais efeitos na obtenção das topologias do que os métodos utilizados, ou seja, a escolha de conjuntos de genes parece ser um fator mais determinante na topologia do que o método de reconstrução utilizado, apesar de algumas exceções terem sido demonstradas nas Figuras 11 a 17. Apesar disso, os suportes dos nós nas topologias variaram bastante, sendo que o método de inferência bayesiana apresentou os maiores suportes, mesmo quando recuperou relações aparentemente improváveis, como no caso do agrupamento de Nematocera (*Aedes*+*Anopheles*) com espécies de Brachycera basais (*S. grandicornis*, *C. duplonotata* e *T. punctata*) nas árvores obtidas com as sequências nucleotídicas (conjuntos de dados subnt_12 e nt_12), cujas topologias estão nas Figuras 11, 12 e 13.

Para embasar as discussões acerca das topologias obtidas com diferentes métodos e conjuntos de dados, será utilizada a comparação das mesmas com uma topologia clássica de Diptera (Figura 19), resumida de estudos sistemáticos com dados morfológicos realizados por Mc Alpine (1989) e Yeates and Wiegmann (2005).

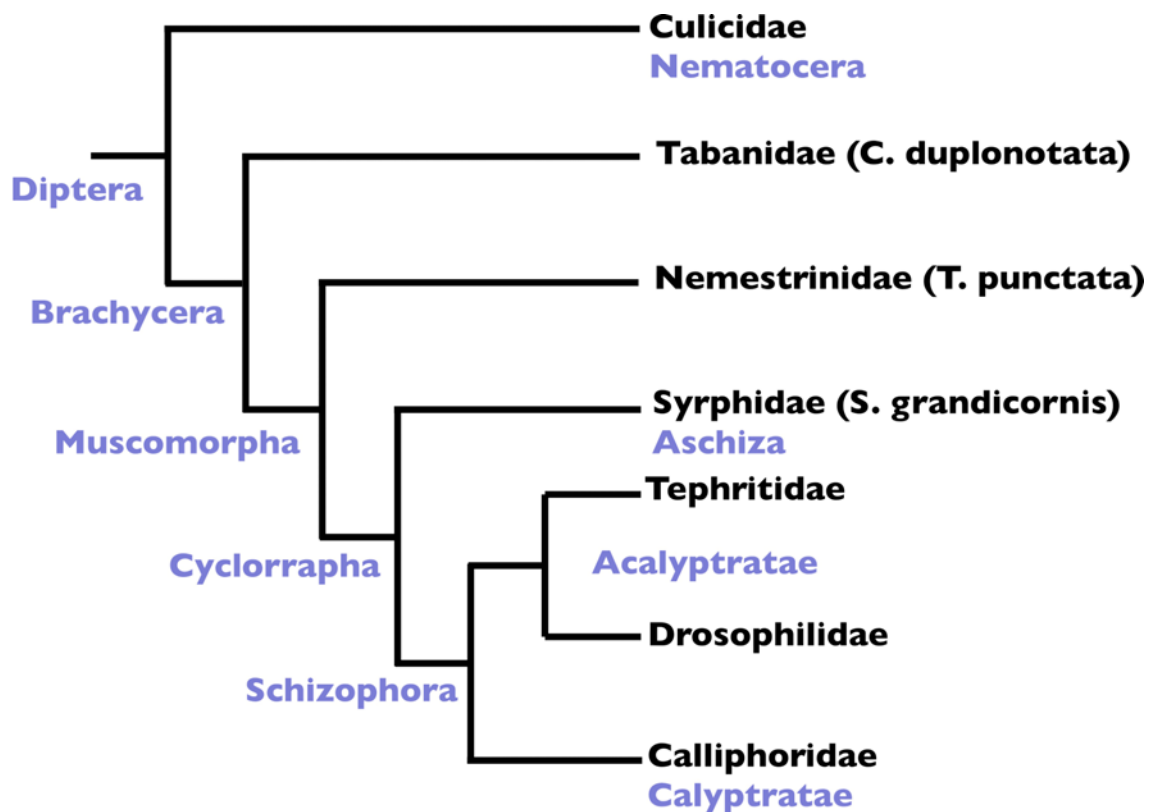


Figura 19: Consenso das relações entre famílias de Diptera, representando seus grupos principais.

Como já mencionado acima, as topologias obtidas a partir de sequências nucleotídicas agrupou espécies fora do grupo Schizophora com Nematocera, rompendo a monofilia de Brachycera. Este agrupamento é improvável, uma vez a monofilia de Brachycera é amplamente aceita, apresentando um grande número de sinapomorfias indiscutíveis (Yeates and Wiegmann, 1999). Embora a monofilia de Nematocera seja discutível, a pequena representação de sequências completas de mtDNA de espécies desta sub-ordem no GenBank não permite a avaliação da hipótese parafilética do grupo. Por outro lado, Brachycera e Nematocera foram recuperados como grupos monofiléticos em todas as árvores obtidas com sequências de aminoácidos (Figuras 14 a 17). É possível que o grau de variação das sequências nucleotídicas, ainda que empregando apenas a primeira e segunda posições dos códons, tenha gerado ruído e por esta razão o agrupamento de espécies basais de Brachycera com Nematocera ocorreu nestas análises. Já as sequências de aminoácidos, por serem menos variáveis, geraram um sinal filogenético capaz de recuperar as duas sub-ordens de Diptera como grupos-irmãos.

A monofilia de Muscomorpha, provavelmente pelas mesmas razões citadas acima, também só foi obtida com sequências de aminoácidos (Figuras 14 e 17). Ainda assim, os suportes dos nós deste agrupamento foi variável e a espécie *C. duplonotata* (infra-ordem Tabanomorpha de Brachycera) foi agrupada dentro da infra-ordem Muscomorpha, na maioria das vezes com o nemestrínídeo *T. punctata*. Logo, a monofilia da infra-ordem Muscomorpha obteve pouco suporte nas análises apresentadas neste trabalho. De acordo com Woodley (1989), a família Nemestrinidae é geralmente parte de Muscomorpha. Yeates (2002) também encontrou um alto suporte para a monofilia de Muscomorpha incluindo Nemestrinidae, mas discute o possível caráter homoplásico de algumas sinapomorfias utilizadas. Por outro lado, Colless and Mc Alpine (1991) colocam Nemestrinidae como grupo irmão de Tabanidae, assim como Krzeminski and Krzeminska (2003) argumentam sobre a necessidade de reavaliar o agrupamento de Nemestrinidae dentro de Muscomorpha. O único estudo molecular realizado com o rRNA 28S contemplando infra-ordens de Brachycera tem um suporte baixo (>50) para a monofilia de Muscomorpha (Collins and Wiegmann, 2002; Wiegmann et al., 2003).

O único conjunto de dados que não recuperou a monofilia de Cyclorrapha foi o subnt_12, agrupando a espécie *S. grandicornis* ao grupo formado por Nematocera+Nemestrinidae+Tabanidae. Este agrupamento, como já discutido, é

improvável. Os dados nt_12 e aa_comp resultaram em maiores suportes para Cyclorrapha do que o conjunto reduzido de aminoácidos, indicando que um maior número de caracteres aumenta a resolução do sinal filogenético neste nível taxonômico.

O grupo Schizophora apresentou-se monofilético em todas as árvores reconstruídas, com suportes altos. Apenas nas árvores de ML com o sub-conjunto de sete genes é que os suportes tiveram valores medianos. Com os dados de GCPs completos (tanto nucleotídeos como aminoácidos), os suportes para a monofilia deste grupo foram altas.

Ao analisar os grupos internos de Schizophora, no entanto, a maior parte das topologias indicaram a parafilia de Acalyptratae com alto suporte (Figuras 11, 13, 14, 16 e 17), apresentando o agrupamento recorrente de Drosophilidae com Calliphoridae. Apenas as árvore de NJ com os subconjuntos de genes mitocondriais recuperaram a monofilia de Acalyptratae (Figuras 12 e 15). A monofilia de Acalyptratae é um assunto altamente contraditório nas filogenias de Schizophora. A redução da caliptra, que é a sinapomorfia que dá nome ao grupo, é discutível e alguns autores sugerem que trata-se um caractere muito variável para unir Acalyptratae (Yeates and Wiegmann, 2005). As filogenias moleculares, por outro lado, não incluíram um número representativo de acaliptrados para obter uma resolução confiável sobre esta questão, mas o grupo tem sido recuperado como parafilético em vários trabalhos que utilizaram o rRNA 28S (Collins and Wiegmann, 2002), o *rrnS*, *rrnL* e *cox2* (Han and Ro, 2005) e genomas mitocondriais completos (Nardi et al., 2003; Cameron et al., 2007; Carapelli et al., 2007).

Dentro de Acalyptratae, as relações entre as espécies de tefritídeos *B. papayae*, *B. carambolae*, *B. dorsalis* e *B. philippinensis* não pôde ser resolvida, mas todas as árvores obtidas indicam que elas formam um grupo embasado por *B. oleae*. O gênero *Bactrocera*, por sua vez, aparece como grupo-irmão de *Ceratitis*, também da família Tephritidae. Da mesma maneira, as relações internas de Drosophilidae apresentam suportes muito altos, estando em acordo com as relações descritas previamente para este grupo, usando marcadores moleculares múltiplos e morfologia (Russo et al., 1995; Cameron et al., 2007). A monofilia de Calyptratae foi obtida em todas as topologias com um alto valor de suporte nos ramos. De acordo com Yeates and Wiegmann (1999), Calyptratae de fato tem sido reconhecida como uma linhagem monofilética entre os Diptera, sendo que o suporte morfológico para este clado é forte.

Em todas as topologias obtidas, as espécies *S. calcitrans* e *M. domestica*, pertencentes à família Muscidae, da superfamília Muscoidea, agruparam-se com os califorídeos, formando um clado que inclui Muscoidea dentro de Oestroidea. De acordo com Mc Alpine (1989), Calyptratae é formada por três superfamílias monofiléticas: Oestroidea, Muscoidea e Hippoboscoidea. Com as reconstruções obtidas a partir do concatenado de genes mitocondriais neste trabalho, a monofilia de Oestroidea é rompida dentro dos Calyptratae, contrariando a taxonomia clássica. Em outros estudos que utilizaram marcadores moleculares, esta relação também foi obtida. Nirmala et al. (2001) recuperou uma filogenia que indica o agrupamento de Muscoidea dentro de Oestroidea, porém mais próxima da família Tachnidae (ver Figura 1). Suas análises com os genes para rRNA 28S e *rnl* nunca recuperaram Muscidae como um grupo irmão de Oestroidea. Bernasconi (2000), em um estudo com os genes mitocondriais *cox1* e *cox2* também não encontrou suporte para separar Muscoidea de Oestroidea. Isto pode indicar que as duas superfamílias representam agrupamentos polifiléticos de Calyptratae. A inclusão de dados adicionais das três superfamílias, em especial de espécies de Hippoboscoidea, que ainda não possui nenhum genoma mitocondrial completamente sequenciado até o momento, será de extrema importância para o esclarecimento de relações evolutivas em Calyptratae.

A monofilia da família Calliphoridae teve um alto suporte em todas as topologias reconstruídas, não importando o conjunto de dados ou o método utilizado. As relações internas de califorídeos indicam que o suporte para a monofilia da subfamília Chrysomyinae é baixo, embora tenha sido recuperada em nove das dezesseis árvores analisadas com o concatenado de genes mitocondriais. O agrupamento de Calliphorinae com Luciliinae ocorre na maioria das árvores reconstruídas, embora com valores estatísticos de suporte muito variáveis em todos os conjuntos de dados.

O agrupamento de Calliphorinae+Luciliinae não está de acordo com o maior estudo morfológico realizado acerca da família Calliphoridae (Rognes et al., 1997), que identificou Chrysomyinae como grupo irmão de Calliphorinae. No entanto, estudos recentes com reconstruções filogenéticas utilizando marcadores moleculares têm indicado uma maior proximidade de Calliphorinae e Luciliinae. Wallman et al. (2005), num estudo com 32 espécies de califorídeos coletados da Austrália (utilizando os genes mitocondriais *cox1*, *cox2*, *nd4* e *nd4L*), recuperaram esta relação com um suporte moderado nos nós. Utilizando uma árvore linearizada com uma calibração

prévia feita por Brower (1994), os mesmos autores sugerem que o nó que separa o clado Chrysomyinae do de Calliphorinae+Luciliinae é estimado em 21.7 milhões de anos. A divergência de Calliphorinae e Luciliinae ocorreu há cerca de 19 milhões de anos, enquanto divergências entre gêneros podem ter ocorrido há 15 milhões de anos. Entre espécies do mesmo gênero, a estimativa de divergência chega, em sua maioria, a 500 mil anos ou menos.

Stevens (2003), em análises baseadas nas sequências de *cox1*, *cox2* e do rRNA 28S também encontrou uma maior proximidade entre Calliphorinae e Luciliinae, assim como Duarte (2007) em análises baseadas na região controle do mtDNA. No Apêndice D, Marinho et al. (submetido) também encontraram o agrupamento de Calliphorinae com Luciliinae, à exclusão de Chrysomyinae na maioria das reconstruções filogenéticas realizadas com o ITS2 do rDNA. As bases morfológicas de associação de Calliphorinae com Chrysomyinae, como descrito por Rognes (1997), foram detalhadas com caracteres da terminália de fêmeas e os dados moleculares sugerem a necessidade de revisão dos caracteres morfológicos utilizados para suportar este agrupamento.

6. Considerações finais e perspectivas

O sequenciamento e caracterização de genomas mitocondriais completos de insetos tem disponibilizado uma quantidade de dados que pode ser aplicada nas mais diferentes questões biológicas, desde estudos acerca da estrutura e arranjo do conteúdo gênico à resolução de filogenias de grandes grupos.

A utilização dos dados gerados através do sequenciamento completo do mtDNA tem resolvido questões evolutivas que abordam a origem endossimbiótica da mitocôndria como uma organela citoplasmática, além de evidenciar diferentes ordens do conteúdo gênico, contribuindo para o estudo de rearranjos em grupos animais.

O aparato molecular envolvido na transcrição do genoma mitocondrial é amplamente desconhecido, com poucos estudos feitos *in vitro*. O processamento e maturação dos rRNAs se baseia em teorias desenvolvidas para vertebrados, podendo não ser mantidas para os invertebrados. A transcrição policistrônica e o modelo de pontuação por tRNA são utilizados, mas pouco se sabe a respeito do funcionamento destes processos em mtDNA de insetos. Um dos exemplos mais interessantes compete ao início do gene *cox1*, que não possui um códon de iniciação no quadro de leitura, de modo que ainda não foi descrita a maneira que o transcrito é reconhecido

no fragmento policistrônico, ou se há algum mecanismo de sinalização para o início de sua transcrição.

Os mecanismos de replicação do mtDNA são igualmente intrigantes, havendo poucos estudos funcionais e bioquímicos com relação ao modo de replicação desta molécula em insetos. Alguns complexos enzimáticos foram descritos, mas estudos com a origem de replicação são poucos e precários, levando ao desconhecimento da maquinaria completa envolvida nas etapas de replicação deste genoma.

Estudos envolvendo a região controle podem culminar em descobertas importantes a respeito da estrutura, função e evolução do genoma mitocondrial. O envolvimento em duplicações (como as descritas neste trabalho), repetições *in tandem*, a composição altamente enviesada em A+T, a presença de blocos conservados de sequências (provavelmente relacionados ao início da replicação e à regulação da transcrição) e a existência de sequências hipervariáveis tornam esta região particularmente interessante. A procura por duplicações do mesmo tipo que foram descritas para espécies de califorídeos neste trabalho pode ajudar a explicar a maneira com que elas ocorreram e se podem ser resultado de eventos paralelos numa região altamente suscetível a duplicações, ou uma característica ancestral que foi perdida em algumas espécies.

Por fim, a geração de sequências completas do mtDNA de espécies de insetos é uma fonte importante de informação para estudos evolutivos, podendo melhorar a resolução de filogenias de diferentes grupos, ajudar na escolha de marcadores apropriados para questões taxonômicas em diferentes níveis hierárquicos, melhorar o acesso a diferentes regiões da molécula através do desenho apropriado de “primers”, além de contribuir no diagnóstico espécie-específico. No âmbito da família Calliphoridae, o mtDNA pode contribuir no estudo da evolução do hábito de parasitismo, disponibilizando informações evolutivas que guiem estudos mais abrangentes em expressão diferencial entre espécies e fases do desenvolvimento.

CONCLUSÕES

1. As técnicas de amplificação por “long-PCR” e “shotgun” dos fragmentos longos do DNA mitocondrial são abordagens apropriadas para a obtenção de sequências completas deste genoma, incluindo o acesso a informação genética de indivíduos preservados em Etanol e contribuindo para a ampliação da genômica mitocondrial de invertebrados;
2. O mtDNA de *Chloroprocta idioidea*, *Calliphora vomitoria* e *Phormia regina* está organizado de acordo com a ordem ancestral de genes descrito para Pancrustacea (Hexapoda+Crustacea);
3. O conteúdo nucleotídico e “codon usage” mostram um evidente viés na composição de adeninas e timinas, como previamente descrito em genomas mitocondriais de insetos;
4. A utilização de genes individuais para a reconstrução filogenética de dípteros gera uma grande variedade de topologias com baixo suporte;
5. O concatenado das sequências nucleotídicas e de aminoácidos dos 13 genes codificadores de proteínas mitocondriais produz resultados mais robustos do que genes individuais ou do concatenado de um subconjunto dos sete genes mais conservados em reconstruções filogenéticas;
6. A escolha do conjunto de dados parece ser a variável determinante na obtenção de topologias. Diferentes métodos de reconstrução filogenética tendem a gerar topologias similares, embora com diferentes suportes nos ramos;
7. A sequência de nucleotídeos dos conjuntos concatenados de genes não recuperou as relações mais basais da ordem Diptera, provavelmente devido à variabilidade introduzida. Por outro lado, as sequências de aminoácidos demonstraram-se mais apropriadas para a recuperação destas relações, com maior suporte no nós. Assim, os clados principais de Diptera (Nematocera e

Brachycera) são monofiléticos em todas as topologias obtidas com aminoácidos;

8. As cinco famílias de dípteros representadas por múltiplas espécies são monofiléticas (Culicidae, Tephritidae, Drosophilidae, Muscidae e Calliphoridae);
9. Sub-grupos de Brachycera, como Cyclorrapha e Schizophora, são monofiléticos. A monofilia de Muscomorpha não teve robustez nas análises realizadas, indicando um agrupamento de Nemestrinidae+Tabanomorpha;
10. A monofilia de Acalyptratae não foi suportada, enquanto a de Calyptratae apresentou robustez. Dentro de Calyptratae, as topologias indicaram a inclusão de Muscidae como grupo-irmão de Calliphoridae, não recuperando a relação entre as superfamílias Muscoide e Oestroidea;
11. As relações entre sub-famílias de Calliphoridae indicam que Calliphorinae e Luciliinae são grupos irmãos, enquanto a monofilia de Chrysomyinae foi recuperada na maioria das árvores, embora com baixo suporte;
12. A duplicação do *tRNA^{le}* presente no gênero *Chrysomya* também está presente em *P. regina*. Estudos complementares sugerem que a região controle do mtDNA pode atuar como um “hotspot” para duplicações em Calliphoridae.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abascal F, Zardoya R, Posada, D. (2005). ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics*, 21: 2104-2105.
- Agris, P.F., (2004). Decoding the genome: a modified view. *Nucleic Acids Research*, 32, 223–238.
- Altshul, S. F.; Gish, W.; Miller, W.; Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, 215: 403-410.
- Amendt, J.; Krettek, R. and Zehner, R. (2004). Forensic entomology. *Naturwissenschaften*, 91: 51-65.
- Ames C.; Turner, B. and Daniel B. (2006) The use of mitochondrial cytochrome oxidase I gene (COI) to differentiate two UK blowfly species - *Calliphora vicina* and *Calliphora vomitoria*. *Forensic Science International*, 164:179-182.
- Anderson, G. S. (2000). Minimum and maximum development rates of some forensically important Calliphoridae (Diptera). *Journal of Forensic Sciences*, 45: 824-832.
- Andersson, S. G. E.; Zomorodipour, A.; Andersson, J. O.; Sicheritz-Pontén, T.; Alsmark, U. C. M.; Podowski, R. M.; Näslund, A. K.; Eriksson, A. S.; Winkler, H. H. and Kurland, C. G. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*, 396: 133-140.
- Asakawa, S., Kumazawa, Y., Araki, T., Himeno, H., Miura, K. and Watanabe, K. (1991). Strand specific nucleotide bias in echinoderm and vertebrate mitochondrial genomes. *Journal of Molecular Evolution*, 32: 511-520.
- Avise, J. C. (1994). *Molecular Markers, Natural History and Evolution*. Chapman and Hall, New York, NY.
- Azeredo-Espin, A. M. L. and Lessinger, A. C. (2006). Genetic approaches for studying myiasis-causing flies: molecular markers and mitochondrial genomics. *Genetica*, 126: 111–131.
- Ballard, J. W. O. and Kreitman, M. (1995). Is mitochondrial DNA a strictly neutral marker? *Trends in Ecology and Evolution*, 10: 485-488.
- Ballard, J. W. O. (2000a). Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. *Journal of Molecular Evolution*, 51: 48-63.
- Ballard, J. W. O. (2000b). Comparative genomics of mitochondrial DNA in *Drosophila simulans*. *Journal of Molecular Evolution*, 51: 64-75.

- Barau, J. G.; Azeredo-Espin, A. M. L. and Lessinger, A. C. (2005) Conservation and versatility of a new set of primers for long-PCR amplification of complete insect mitochondrial genomes based on *Haematobia irritans* mtDNA sequences. *Molecular Ecology Notes*, 5: 885-887.
- Baumgartner, D. L. and Greenberg, B. (1984). The genus *Chrysomya* (Diptera: Calliphoridae) in the New World. *Journal of Medical Entomology*, 21: 105-113.
- Beard, C.B., Hamm, D.M., Collins, F.H. (1993). The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Molecular Biology*, 2: 103-124.
- Bernasconi, M. V.; Pawlowski, J.; Valsagiacomo, C.; Piffaretti, J. C.; Ward, P.I (2000). Phylogeny of the scathophagidae (Diptera, Calyptratae) based on mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, 16: 308-315.
- Boore, J. L.; Collins, T. M.; Stanton, D.; Daehler, L. L. and Brown, W. M. (1995). Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature*, 376: 163-165.
- Boore, J. L.; Lavrov, D. V. and Brown, W. M. (1998). Gene translocation links insects and crustaceans. *Nature* 392: 667-668.
- Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research*, 27: 1767-1780.
- Brower, A. V. Z. (1994). Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the USA*, 91: 6491-6495.
- Cameron, S. L.; Miller, K. B.; D'Haese, C. A.; Whitinf, M. F. and Barker, S. C. (2004). Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea *sensu lato* (Arthropoda). *Cladistics*, 20: 534-557.
- Cameron, S. L.; Lambkin, C. L.; Barker, S. C. and Whiting, M. F. (2007). A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. *Systematic Entomology*, 32: 40-59.
- Cano, R. J. (1996). Analysing Ancient DNA. *Endeavour*, 20: 162-167.
- Carapelli, A.; Liò, P.; Nardi, F.; van der Wath, E. and Frati, F. (2007). Phylogenetics analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. *BMC Evolutionary Biology*, 7 (Suppl 2): S8.
- Carrodeguas, J. A. and Vallejo, C. G. (1997). Mitochondrial transcription initiation in the crustacean *Artemia franciscana*. *European Journal of Biochemistry*, 250: 514-23.

- Carvalho, L. M. L.; Thyssen, P. J.; Linhares, A. X. and Palhares, F. A. B. (2000). A checklist of arthropods associated with pig carrion and human corpses in Southeastern Brazil. *Memórias do Instituto Oswaldo Cruz*, 95: 135-138.
- Caterino, M. S.; Cho, S. and Sperling, F. A. H. (2000). The current state of insect molecular systematics: a thriving Tower of Babel. *Annual Reviews of Entomology*, 45: 1-54.
- Cedergren, R.; Gray, M. W.; Abel, Y. and Sankoff, D. (1988). The evolutionary relationships among known life forms. *Journal of Molecular Evolution*, 28: 98-112.
- Chang, D. D. and Clayton, D. A. (1984). Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. *Cell*, 36: 635-643.
- Chen, W. Y.; Hung, T. H. and Shiao, S. F. (2004). Molecular Identification of Forensically Important Blow Fly Species (Diptera: Calliphoridae) in Taiwan. *Journal of Medical Entomology*, 41: 47-57
- Clary, D. O. and Wolstenholme D. R. (1985) The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *Journal of Molecular Evolution*, 22: 252-271.
- Clary, D. O. and Wolstenholme, D. R., (1987). *Drosophila* mitochondrial DNA: conserved sequences in the A+T- rich region and supporting evidence for a secondary model of the small ribosomal RNA. *Journal of Molecular Evolution*, 25: 116-125.
- Coless, D. H. and Mc Alpine, D. K. (1991). *Diptera. The insects of Australia (CSIRO)*. Melbourne University Press, Melbourne.
- Collins, K. P. and Wiegmann, B. M. (2002). Phylogenetic relationships of lower Cyclorhapha (Diptera: Brachycera) based on 28S rDNA sequences. *Insect Systematics and Evolution*, 33: 445-456.
- Covacin, C.; Shao, R.; Cameron, S. and Barker, S. C. (2006). Extraordinary number of gene rearrangements in the mitochondrial genomes of lice (Phthiraptera: Insecta). *Insect Molecular Biology*, 15: 63-68.
- Crozier, R. H. and Crozier, Y. C. (1993). The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* 133: 97-117.
- Dean, M. D. and Ballard, W. O. (2001). Factors affecting mitochondrial DNA quality from museum preserved *Drosophila simulans*. *Entomologia Experimentalis et Applicata*, 98: 279-283.
- Dowton, M. and Austin, A. D. (1999). Evolutionary Dynamics of a Mitochondrial Rearrangement “Hot Spot” in the Hymenoptera. *Molecular Biology and Evolution*, 16: 298–309.

Dowton, M.; Castro, L. R.; Campbell, S. L., Bargon, S. D. and Austin, A. D. (2003). Frequent mitochondrial gene rearrangements at the hymenopteran nad3-nad5 junction. *Journal of Molecular Evolution*, 56:517-526.

Duarte, G. T. (2007). Caracterização da região controle do genoma mitocondrial em espécies da família Calliphoridae (Insecta: Diptera) e perspectivas filogenéticas. Tese de Mestrado, Unicamp, Campinas, SP.

Esposito, M. C. (1999). A fauna de moscas varejeiras (Diptera: Calliphoridae) da Amazônia e sua ecologia na região de Caxiuana e cidade de Portel/PA. Tese de Doutorado, Manaus, AM.

Ewing, B.; Hillier, L.; Wendl, M. C. and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. assessment. *Genome Research*, 8: 175-185.

Feijão, P. C.; Neiva, L. S.; Azeredo-Espin, A. M. L. and Lessinger, A. C. (2006). AMiGA: the arthropodan genomes accessible database. *Bioinformatics*, 22 (7): 902-903.

Flook, P. K.; Rowell, C. H.; Gellissen, G. (1995). The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. *Journal of Molecular Evolution*, 41: 928-941.

Furusawa, G. P. and Cassino, P. C. R. (2006). Ocorrência e Distribuição de Calliphoridae (Diptera, Oestroidea) em um Fragmento de Mata Atlântica Secundária no Município de Engenheiro Paulo de Frontin, Médio Paraíba, RJ. *Revista de Biologia e Ciências da Terra*, 6: 152-164.

Goddard, J. M. and Wolstenholme, D. R. (1978). Origin and direction of replication in mitochondrial DNA molecules from *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences USA*, 8: 3886-3890.

Goddard, J. M. and Wolstenholme, D. R. (1980). Origin and direction of replication in mitochondrial DNA molecules from the genus *Drosophila*. *Nucleic Acids Research*, 25: 741-757.

Gordon, D.; Abajian, C. and Green, P. (1998) CONSED: a graphical tool for sequence finishing. *Genome Research*, 8: 195-202.

Greenberg, B. (1971). Ecology, classifications and biotic associations. In: *Flies and Disease*. Vol I, Princeton University Press, Princeton.

Grey, M. W. and Doolittle, W. F. (1982). Has the endosymbiont hypothesis been proven? *Microbiology Review*, 46: 1-42.

Guimarães, J. H. and Papavero, N. (1999). Myiasis in man and animals in the neotropical region. Ed. Plêiade.

Guindon, S. and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematics Biology*, 52: 696-704.

Hall, M. and Wall, R. (1995) Myiasis of humans and domestic animals. *Advances in Parasitology*, 35:257-334.

Harvey, M. L., M. W. Mansell, M. H. Villet, and R. Dadour. (2003). Molecular identification of some forensically important blowflies of southern Africa and Australia. *Medical and Veterinary Entomology*, 17: 363-369.

Han, H.-Y and Ro, K.-E. (2005). Molecular phylogeny of the superfamily Tephritoidea (Insecta: Diptera): new evidence from mitochondrial 12S, 16S and COII genes. *Molecular Phylogenetics and Evolution*, 34: 416-430.

Helpenbein, K.G., Brown, W.M. and Boore, J.L. (2001). The complete mitochondrial genome of the articulate brachiopod *Terebratalia transversa*. *Molecular Biology and Evolution*, 18: 1734-1744.

Hennig, W. (1966). *Phylogenetic Systematics*. Univ. of Illinois Press, Urbana.

Herrmann, B.; Hummel, S. (1994). *Ancient DNA*. Springer-Verlag, New York, NY.

Höss, M.; Jaruga, P., Zastawny, T. H., Dizdaroglu, M. and S. Pääbo. (1996). DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Research*, 24: 1304-1307.

Huelsenbeck, J. P. and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, 17: 754-755.

Hwang, U. W.; Friederich, M.; Tautz, D.; Park, C. J. and Kim, W. (2001a). Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature* 413: 154-157.

Hwang, U. W.; Park, C. J.; Yong, T. S. and Kim, W. (2001b). One-step PCR amplification of complete arthropod mitochondrial genomes. *Molecular Phylogenetics and Evolution*, 19: 345-352.

Infante, M. E. And Azeredo-Espin, A. M. L. (1995) Genetic variability in mitochondrial DNA of screwworm, *Cochliomyia hominivorax* (Diptera: Calliphoridae), from Brazil. *Biochemical Genetics*, 33: 737-756.

Infante-Malachias, M. E.; Yotoko, K. S. C. and Azeredo-Espin, A. M. L. (1999). Random amplified polymorphic DNA of screwworm ssy populations (Diptera: Calliphoridae) from southeastern Brazil and northern Argentina. *Genome*, 42: 772-779.

Junqueira, A. C. M., Lessinger, A. C. and Azeredo-Espin, A. M. L. (2002). Methods for recovery of mtDNA sequence from museum specimens of myiasis-causing-flies. *Medical and Veterinary Entomology*, 16: 39-45.

Junqueira, A. C. M.; Lessinger, A. C.; Torres, T. T.; Vettore, A. L.; da Silva, F.; Arruda, P. and Azeredo-Espin, A. M. L. (2004). The complete mitochondrial genome of the blowfly *Chrysomya putoria* (Diptera: Calliphoridae), *Gene* (339): 7-15.

Katoh, K.; Kuma, K.; Miyata, T. and Toh, H. (2005). Improvement in the accuracy of multiple sequence alignment program MAFFT. *Genome Informatics*, 16:22-33.

Koller, W. W.; Gomes, A.; Gomes, P. R.; Umaki, A.; Santos, S. T. P.; Carvalho, C. J. B. (2002). Dípteros Calliphoridae em mata ciliar remanescente no interior de pastagem cultivada, em Campo Grande, MS, Brasil. In: Congresso Brasileiro de Parasitologia Veterinária, 12., Rio de Janeiro, 2002. Programas e Resumos, Rio de Janeiro: CBPV/UFRRJ/PJ Eventos, R400.pdf (CD-ROM).

Kondo, R.; Satta, Y.; Matsuura, E. T., Ishiwa, H.; Takahato, N. and Chigusa, S. I. (1990). Incomplete maternal transmission of mtDNA in *Drosophila*. *Genetics*, 126: 657-663.

Kristensen, N. P. (1981). Phylogeny of insect orders. *Annual Reviews of Entomology*, 26: 135–157.

Krzeminski, W. and Krzeminska, E. (2003) Triassic Diptera: descriptions, revisions and phylogenetic relations. *Acta Zoologica Cracoviensia*, 46S: 153–184.

Ladoukakis E. D. and Zouros, E. (2001) Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. *Molecular Biology and Evolution* 18: 1168-75.

Lang, B. F.; Burger, G.; O'Kelly, C. J.; Cedergren, R.; Golding, G. B.; Lemieux, C.; Sankoff, D.; Turmel, M. and Gray, M. W. (1997). An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature*, 387: 493-497.

Lessinger, A. C. and Azeredo-Espin, A.M.L. (2000). Evolution and structural organisation of mitochondrial DNA control region of myiasis-causing flies. *Medical and Veterinary Entomology*, 14: 1-10.

Lessinger, A. C., Martins Junqueira, A. C., Lemos, T. A., Kemper, E. L., da Silva, F. R., Vettore, A. L., Arruda, P. and Azeredo-Espin, A. M. L. (2000) The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Molecular Biology* 9: 521-529.

Lessinger, A. C.; Junqueira, A. C. M.; Conte, F. F. and Azeredo-Espin, A. M. L. (2004). Analysis of a conserved duplicated tRNA gene in the mitochondrial genome of blowflies. *Gene*. 339: 1-6.

Levinson, G. and Gutman, G. A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*, 4 (3): 203-221.

Lewis, D. L.; Farr, C. L.; Farquhar, A. L. and Kaguni, L. S. (1994). Sequence, organization, and evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA. *Molecular Biology and Evolution*, 11: 523-538.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, 362: 709-715.

- Linhares, A. X. (1979). Sinantropia de dípteros muscóides de Campinas, São Paulo, Brasil. *Revista Brasileira de Entomologia*, 25: 189-215.
- Linhares, A. X. (1981). Sinanthropy of Calliphoridae and Sarcophagidae (Diptera) in the city of Campinas, São Paulo, Brasil. *Revista Brasileira de Entomologia*, 35: 189-215.
- Litjens, P.; Lessinger, A. C. and Azeredo-Espin A. M. L. (2001). Characterization of the screwworm flies *Cochliomyia hominivorax* and *Cochliomyia macellaria* by PCR-RFLP of mitochondrial DNA. *Medical and Veterinary Entomology*, 15:183-188.
- Lowe, T. D. And Eddy, S. R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25: 955-964.
- Lyra, M. L.; Fresia, P.; Gama, S.; Cristina, J.; Klaczko, L. B. and Azeredo-Espin, A. M. L. (2005). Analysis of mitochondrial DNA variability and genetic structure in populations of New World screwworm flies (Diptera: Calliphoridae) from Uruguay. *Journal of Medical Entomology*, 42: 589-595.
- Margulis, L. (1970). *Origin of Eukaryotic Cells*. Yale University Press, New Haven.
- Martin, A. P. (1995). Metabolic rate and directional nucleotide substitution in animal mitochondrial DNA. *Molecular Biology and Evolution*, 12: 1124-1131.
- McAlpine, J. F. (1989) Phylogeny and classification of the Muscomorpha. In: *Manual of Nearctic Diptera*. 3: 1397-1518. Research Branch, Agriculture Canada.
- Mitchell, S.E., Cockburn, A.F., Seawright, J.A. (1993). The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. *Genome*, 36: 1058-1073.
- Moritz, C.; Dowling, T. E. and Brown, W. M. (1987). Evolution of Animal Mitochondrial DNA: Relevance for Population Biology and Systematics. *Annual Review of Ecology and Systematics*, 18: 269-292.
- Nardi, F., Carapelli, A., Fanciulli, P.P., Dallai, R., Frati, F. (2001). The complete mitochondrial DNA sequence of the basal hexapod *Tetradontophora bielensis*: evidence for heteroplasmy and tRNA translocations. *Molecular Biology and Evolution*, 18: 1293-1304.
- Nardi, F.; Spinsanti, G.; Boore, J. L.; Carapelli, A.; Dallai, R. and Frati, F. (2003). Hexapod origins: monophyletic or paraphyletic? *Science*, 299: 1887-1889.
- Nei, M. (1996). Phylogenetic analysis in molecular evolutionary genetics. *Annual Reviews of Genetics*, 30: 371-403.
- Nelson, W. S.; Prodöhl, P. A. and Avise, J. C. (1996). Development and application of long-PCR for the assay of full-length animal mitochondrial DNA. *Molecular Ecology*, 5: 807-810.

- Nirmala, X.; Hypsa, V and Zurovec, M. (2001) Molecular phylogeny of Calyptratae (Diptera: Brachycera): the evolution of 18S and 16S ribosomal rDNAs in higher dipterans and their use in phylogenetic inference. *Insect Molecular Biology*, 10: 475-485.
- Ojala, D.; Merkel, C.; Gelfand, R. and Attardi, G. (1980). The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. *Cell*, 22:393-403.
- Ojala, D.; Montoya, J. and Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature*, 290: 470-474.
- Oliveira, M. T. O.; Azeredo-Espin, A. M. L. and Lessinger, A. C. (2007). The mitochondrial DNA control region of Muscidae flies: evolution and structural conservation in a dipteran context. *Journal of Molecular Evolution*, 64:519–527.
- Otranto, D. and Stevens , J. R. (2002) Molecular approaches to study of myiasis-causing larvae. *International Journal for Parasitology*, 32: 1345-1360.
- Pääbo, S.; Higuchi, R. G. and Wilson, A. C. (1989). Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *Journal of Biological Chemistry*, 264: 9709-9712.
- Pape, T. (1992). Phylogeny of the Tachinidae family-group (Diptera: Calyptratae). *Tijdschrift voor Entomologie* 135: 43–86.
- Perna, N. T. And Kocher, T. D. (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of Molecular Evolution*, 41: 353-358.
- Pont, A. C. (1980). Family Calliphoridae, in *Catalogue of the Diptera of the Afrotropical region*. London, Flectcher and Ltd. Norwich.
- Posada, D. and Crandall, K.A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics*, 14: 817-818.
- Ratcliffe, S. T.; Webb, D. W.; Weinzievr, R. A. and Robertson,H. M. (2003). PCR-RFLP Identification of diptera (Calliphoridae, Muscidae and Sarcophagidae) - a generally applicable method. *Journal of Forensic Science*, 48: 1-3.
- Rawlings, T. A.; Collins, T. M. And Bieler, R. (2001). A Major Mitochondrial Gene Rearrangement Among Closely Related Species. *Molecular Biology and Evolution*, 18: 1604-1609.
- Roehrdanz, R. L. and Johnson, D. A. (1988) Mitochondrial DNA variation among geogrgaphical populations of the screwworm fly, *Cochliomyia hominivorax*. *Journal of Medical Entomology*, 25: 136-141.
- Roehrdanz, R. L. (1989) Intrataspecific genetic variability in mitochondrial DNA of the screwworm fly (*Cochliomyia hominivorax*). *Biochemical Genetics*, 27: 551-569.

- Roehrdanz, R. L.; Degrugillier, M. E. and Black IV, W. C. (2002). Novel rearrangements of arthropod mitochondrial DNA detected with Long-PCR: applications to arthropod phylogeny and evolution. *Molecular Biology and Evolution*, 19: 841-849.
- Rognes, K. (1997) The Calliphoridae (Blowflies) (Diptera: Oestroidea) are Not a Monophyletic Group. *Cladistics*, 13:27-66.
- Ronquist, F. and Huelsenbeck, J. P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19: 1572-1574.
- Rozas, J.; Sanchez-DelBarrio, J. C.; Messeguer, X. and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 19: 2496-2497.
- Russo, C. A. M.; Takezaki, N. and Nei, M. (1995). Molecular phylogeny and Divergence times of drosophilid species. *Molecular Biology and Evolution*, 12: 391-404.
- Russo, C. A. M.; Takezaki, N. and Nei, M. (1996). Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. *Molecular Biology and Evolution*, 13: 525-536.
- Russo, C. A. M. (1997). Efficiencies of different statistical tests in supporting a known vertebrate phylogeny. *Molecular Biology and Evolution*, 14: 1078-1080.
- Saccone, C.; de Giorgi, C.; Gissi, C.; Pesole, G. and Reyes, A. (1999). Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene*, 238: 195-209.
- Saigusa, K.; Takamiya, M. and Aoki, Y. (2005). Species identification of the forensically important flies in Iwate prefecture, Japan based on mitochondrial cytochrome oxidase gene subunit I (COI) sequences. *Legal Medicine*, 7: 175–178.
- Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4: 406-425.
- Salviano, R. J. ; Mello, R.P. ; Santos, R. F. S. ; Beck L. C. N. H. ; Ferreira A. (1996). Calliphoridae (Diptera) associated with human corpses in Rio de Janeiro, Brazil. *Entomologia y Vectores*, 3: 145-146.
- Sambrook, J., Maniatis, T. and Fritsch, E. F. (1989). *Molecular cloning: a laboratory manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Simon, C.; Frati, F.; Beckenbach, A.; Crespi, B.; Lin, H. and Flook, P. (1994). Evolution, weighting and phylogenetic utility of DNA sequences and compilation of conserved polymerase chain reaction primers. *Annals of Entomological Society of America*, 87: 651-701.

Shadel, G. S. and Clayton, D. A. (1997). Mitochondrial DNA maintenance in vertebrates. *Annual Reviews of Biochemistry*, 66: 409-435.

Shao, R.; Campbell, N. J. H.; Schmidt, E. R. and Barker, S. C. (2001). Increased Rate of Gene Rearrangement in the Mitochondrial Genomes of Three Orders of Hemipteroid Insects. *Molecular Biology and Evolution*, 18: 1828–1832.

Shao, R. and Barker, S. C. (2003). The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. *Molecular Biology and Evolution*, 20: 362-70.

Shewell, G. E. (1987). Calliphoridae, in *Manual of Nearctic Diptera*. Research Branch Agriculture Canada, 2: 1133-1145.

Sicheritz-Pontén T.; Kurland C. G. and Andersson, S. G. (1998). A phylogenetic analysis of the cytochrome b and cytochrome c oxidase I genes supports an origin of mitochondria from within the Rickettsiaceae. *Biochimica et Biophysica Acta*, 1365: 545-551.

Smith, M. J.; Arndt, A.; Gorski, S. and Fajber, E. (1993). The phylogeny of echinoderm classes based on mitochondrial gene rearrangements. *Journal of Molecular Evolution*, 36: 545-554.

Souza, A. M. de & Linhares, A. X. (1997). Diptera and Coleoptera of potential forensic importance in southeastern Brazil: relative abundance and seasonality. *Medical and Veterinary Entomology*, 11: 8-12.

Stevens, J. and Wall, R. (2001). Genetic relationships between blowflies (Calliphoridae) of forensic importance. *Forensic Science International*, 120: 116-123.

Stevens, J. R.; Wall, R. and Wells, J. D. (2002). Paraphyly in Hawaiian hybrid blowfly populations and the evolutionary history of anthropophilic species. *Insect Molecular Biology*, 11: 141–148.

Stevens, J. R. (2003). The evolution of myiasis in blowflies (Calliphoridae). *International Journal of Parasitology*, 33: 1105-1113.

Stevens, J. R.; West, H. and Wall, R. (2008). Mitochondrial genomes of the sheep blowfly, *Lucilia sericata*, and the secondary blowfly, *Cherysomya megacephala*. *Medical and Veterinary Entomology*, 22: 89-91.

Stewart, J. B. and Beckenbach, A. (2006). Insect mitochondrial genomics 2: the complete mitochondrial genome sequence of a giant stonefly, *Pteronarcys princeps*, asymmetric directional mutation bias, and conserved plecopteran A+T-region elements. *Genome*, 49: 815-824.

Swofford, D. L. (2000). *Phylogenetic Analysis Using Parsimony and Other Methods* (software). Sinauer Associates, Sunderland, MA.

Tamura, K.; Dudley, J.; Nei, M. and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology and Evolution*, 24: 1596-1599.

Thomas, R. (2003). Wingless Insects and plucked chickens. *Science*, 299: 1854-1855.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 24: 4876-4882.

Tong, K.L., Wong, J.T., (2004). Anticodon and wobble evolution. *Gene* 333, 169-177.

Torres, T. T.; Brondani, R.P.V.; Garcia, J. E. and Azeredo-Espin, A.M.L. (2004). Isolation and characterization of microsatellite markers in the new world screw-worm *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Molecular Ecology Notes*, 4: 182-184.

Torres, T. T. and Azeredo-Espin, A.M.L. (2005). Development of new polymorphic microsatellite markers for the New World screw-worm *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Molecular Ecology Notes*, 5: 815-817.

Wallman, J. F. and Donnellan, S. C. (2001). The utility of mitochondrial DNA sequences for the identification of forensically important blowflies (Diptera: Calliphoridae) in southeastern Australia. *Forensic Science International*, 120: 60-67.

Wallman, J. F.; Leys, L. and Hogendoorn, K. (2005). Molecular systematics of Australian carrion-breeding blowflies (Diptera: Calliphoridae) based on mitochondrial DNA. *Invertebrate Systematics*, 19: 1-15.

Wells, J. D. and Sperling, F. A. H. (2001). DNA-based identification of forensically important Chrysomyinae (Diptera: Calliphoridae). *Forensic Science International*, 120: 110-115.

Wells, J. D.; Wall, R. and Stevens, J. R. (2007). Phylogenetic analysis of forensically important *Lucilia* flies based on cytochrome oxidase I sequence: a cautionary tale for forensic species determination. *International Journal of Legal Medicine*, 121:229–233.

Wheeler, W. C.; Whiting, M.; Wheeler, Q. D. and Carpenter, J. M (2001). The Phylogeny of the Extant Hexapod Orders. *Cladistics*, 17: 113–169.

Whitfield, J. B. and Kjer, K. M. (2008). Ancient rapid radiations of insects: challenges for phylogenetic analysis. *Annual Reviews of Entomology*, 53:449-472.

Wiegmann, B. M.; Yeates, D. K.; Thorne, J. L. and Kishino, H. (2003). Time flies, a new molecular time scale for brachyceran fly evolution without a clock. *Systematic Biology*, 52: 745-756.

Wolstenholme, D. R. (1992). Animal mitochondrial DNA: structure and evolution. *International Review of Cytology*, 141: 173-216.

Woodley, N. E. (1989). Phylogeny and classification of the 'Orthoraphous' Brachycera. *Manual of Nearctic Diptera*, Vol. 3. Research Branch, Agriculture Canada, Ottawa.

Yeates, D. K. and Wiegmann, B. M. (1999). Congruency and controversy: toward a higher-level phylogeny of Diptera. *Annual Reviews of Entomology*, 44: 397-428.

Yeates, D. K. (2002). Relationships of extant Brachycera (Diptera): a quantitative synthesis of morphological characters. *Zoologica Scripta*, 31: 105-121.

Yeates, D. K. and Wiegmann, B. M. (2005). Phylogeny and evolution of Diptera: recent insights and new perspectives. *The Evolutionary Biology of Flies*, Columbia University Press, New York.

Xia, X. (1996). Maximizing transcription efficiency causes codon usage bias. *Genetics*, 144: 1309-1320.

Xia, X.; Xie, Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution. *Journal of Heredity*, 92:371-373.

Zhang, D.-X and Hewitt, G. M. (1997). Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology*, 25 (2): 99-120.

Zumpt, F. (1965). *Myiasis in Man and Animals in the Old World*. Butterworths, London.

APÊNDICE A

Analysis of a conserved duplicated tRNA gene in the mitochondrial genome of blowflies

Ana Cláudia Lessinger, Ana Carolina M. Junqueira, Fábio F. Conte and Ana Maria L. Azeredo Espin

Gene 339 (2004): 1-6

RESUMO

A região controle e genes flanqueadores das moscas varejeiras *Chrysomya albiceps*, *Chrysomya megacephala* e *Chrysomya chloropyga* (Calliphoridae) foram caracterizadas. A principal característica encontrada foi a presença de sequências duplicadas de tRNA correspondentes a *trnI* e a uma porção do *trnQ*. A duplicação parcial *trnQ* foi muito semelhante a um pseudogene uma vez que a maior parte da sequência do *trnQ* típico de insetos foi perdida. Por outro lado, o gene *trnI* tem uma sequência primária conservada seguida de um evento de duplicação e pode representar uma cópia funcional. Esses resultados demonstram a plasticidade da molécula do DNAm em *Chrysomya*, especialmente para genes de tRNA e as sequências adjacentes da região controle.

Analysis of a conserved duplicated tRNA gene in the mitochondrial genome of blowflies

Ana Cláudia Lessinger*, Ana Carolina M. Junqueira, Fábio F. Conte,
Ana Maria L. Azeredo Espin

Laboratório de Genética Animal, Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP),
C.P. 6010, CEP 13083-875, Campinas, SP, Brazil

Received 14 January 2004; received in revised form 20 May 2004; accepted 21 June 2004
Available online 18 August 2004
Received by G. Pesole

Abstract

The mtDNA control region (CR) and flanking genes of the blowflies *Chrysomya albiceps*, *Chrysomya megacephala* and *Chrysomya chloropyga* (Calliphoridae) were characterized. The most unusual feature found was the presence of duplicated tRNA sequences corresponding to *trnI* and a portion of *trnQ*. The partially duplicated *trnQ* was very likely a pseudogene since most of the sequence of the typical insect *trnQ* gene was missing. In contrast, the *trnI* gene had a conserved primary sequence following the duplication event and may represent a functional copy. These results demonstrate the plasticity of the mtDNA molecule in *Chrysomya*, especially for tRNA genes and the adjacent control region sequences.

© 2004 Elsevier B.V. All rights reserved.

Keywords: mtDNA; Control region; Rearrangement; Molecular evolution; *Chrysomya*

1. Introduction

Rearrangements and duplications have been reported in animal mtDNA and occur independently of the coding or non-coding nature of the duplicated sequences and of the relative position in the mitochondrial genome, although duplication events associated with or adjacent to control region (CR) sequences are more frequent (Boore, 1999).

Abbreviations: A, adenine; bp, base pair(s); C, cytosine; CIR, *Chrysomya* intergenic region; CSB, conserved sequence block(s); CR, control region; G, guanine; IQM, clustered *trnI*, *trnQ* and *trnM* genes; mtDNA, mitochondrial DNA; *nad1–6*, NADH dehydrogenase subunits 1–6; *rns*, small ribosomal RNA; T, thymine; *trn* or tRNA, transfer RNA gene; *trnQ*, glutamine *trn*; *trnI*, isoleucine *trn*; *trnM*, methionine *trn*; *trnF*, phenylalanine *trn*; *trnP*, proline *trn*; *trnS*, serine *trn*; *trnY*, tyrosine *trn*; U, uridine.

* Corresponding author. Tel.: +55 19 3788 1141; fax: +55 19 3788 1089.

E-mail address: lessinge@unicamp.br (A.C. Lessinger).

0378-1119/\$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.06.041

Duplicated genes are rarely observed among mitochondrial DNAs (mtDNAs) and the disintegration of supernumerary genes is typically rapid.

The structural organization of the insect mtDNA control region (also known as the A+T-rich region) has been reviewed by Zhang and Hewitt (1997). The lack of structural conservation of the control region and adjacent genes, even among closely related species, limits the use of universal insect mtDNA primers and requires special care in the interpretation of homologous elements in comparative analyses. Lessinger and Azeredo-Espin (2000) described conserved and variable sequences arranged in specific structural domains in the mtDNA control region of myiasis-causing flies (myiasis is an infestation produced by dipterous larvae). Comparative analysis led to the identification of eight conserved sequence blocks (CSB) in control region sequences near to the clustered *trnI*, *trnQ* and *trnM* (IQM) and highly variable sequences near the *rns*.

Table 1
Primers used for the amplification of *Chrysomya* mtDNA CR sequences

Reaction	Direct	Reverse	Product/region
1	CMEG A ATGATATTCTTAAGTGGATT	SR-J-14776 GCTGGCACGAATTTTGTC	Duplicated region Amplicon 1
2	TI-N-24 ATTACCCCTATCAAGGTAA	CMEG AR AATCCAGTTAAGAATATCAT	Conserved region Amplicon 2
3	TI-N-24 ^a ATTACCCCTATCAAGGTAA	SR-J-14776 GCTGGCACGAATTTTGTC	300 bp region
4	N2-N-1128 ^b CGTAATAATTTAATATAAAAGC	SR-J-14776 GCTGGCACGAATTTTGTC	NESTED PCR ^c : Long target
5	TM-N-193 TGGGGTATGAACCCAGTAGC	CMEG AR AATCCAGTTAAGAATATCAT	NESTED PCR ^d : 550 bp region

The primers are oriented in the 5'→3' direction. The terms *Direct* and *Reverse* are used arbitrarily to indicate orientation. See Fig. 1.

^a Annealing site in the duplicated gene near the *rnrS* gene.

^b Primer designed based on *Haematobia irritans* sequences, using the nomenclature of Simon et al. (1994).

^c First round of amplification reactions for nested-PCR.

^d Second round of amplification reactions for nested-PCR.

A review of the published *Chrysomya megacephala* sequence (AF151386) identified duplicated genes incorrectly assigned to part of the CR. To address this question, in this work we analyzed additional data from *C. megacephala* mtDNA and included the characterization of *Chrysomya albiceps* and *Chrysomya chloropyga* CR sequences and adjacent genes. The analysis of *C. megacephala*, *C. albiceps* and *C. chloropyga* mtDNA identified two copies of *trnI* and *trnQ* near the *rnrS* gene. The CR and flanking genes of *Chrysomya* mtDNA were compared in order to determine the level of conservation of duplicated *trn* sequences in these species, and to investigate the patterns of molecular evolution that were operating in the repeated region.

2. Material and methods

2.1. *Chrysomya* samples

Adults of *C. albiceps* Wiedemann were collected in Rio de Janeiro, Rio de Janeiro State, Brazil, and stored at -70 °C. Samples of *C. chloropyga* Wiedemann (= *C. putoria*) from Pirassununga, São Paulo State, Brazil, were reared as a laboratory strain, and the pupae were stored at -70 °C until used. The *C. megacephala* Fabricius specimens used for

comparative analysis were from Adamantina, São Paulo State, Brazil.

2.2. DNA extraction

Total individual DNA was extracted from pupae or adults as described in Infante and Azeredo Espin (1995), using a phenol-chloroform procedure.

2.3. Amplification reactions

The PCR reactions were done as described by Lessinger and Azeredo-Espin (2000). The primers used in the amplification reactions of CR sequences and flanking regions for *C. albiceps*, *C. megacephala* and *C. chloropyga* are described in Table 1 and Fig. 1. The sequences of the primers CMeg A and CMeg AR were based on the *C. megacephala* CR sequences (AF151386) and recognized complementary annealing sites with opposite orientations in the mtDNA molecule. The complete sequences of the *Chrysomya* CR were amplified in two separate sections (products are referred to as *amplicons* 1 and 2) to avoid competition between alternative annealing sites for the primer TI-N-24 in the duplicated *trnI* sequence. The clustered *trnI*, *trnQ* and *trnM* genes (IQM) at the 3' end of the *Chrysomya* CR (Fig. 1) were also amplified as a

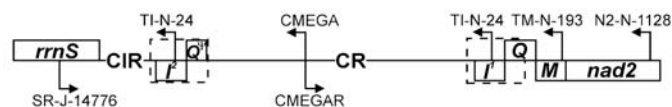


Fig. 1. Schematic organization of the *Chrysomya* mtDNA control region (CR) and adjacent genes indicating the primer annealing sites. Two PCR products (approximately 300 and 1200 bp) were potentially amplified using the primers TI-N-24 and SR-J-14776. Two distinct amplification reactions (TI-N-24 and CMEGAR; CMEGA and SR-J-14776) provided complete amplification of the CR. The *trnQ* and *trnI* from the IQM gene cluster were also amplified using the primers N2-N-1128 and SR-J-14776, followed by a *nested* amplification reaction with the primers TM-N-193 and CMEGAR (see Table 1). I¹, Q and M represent the *trnI*, *trnQ* and *trnM* genes, respectively; I² and Q² represent the *trnI* copy and the *trnQ* partial copy, respectively. CIR indicates a *Chrysomya* intergenic region. The position of the genes indicates the respective coding strand. Duplicated regions are boxed.

separate amplicon using a nested PCR approach. The “external” primers N2-N-1128 and SR-J-14776 were used for the first step of the PCR reactions (long target) followed by nested amplifications (specific target) with the “internal” primers TM-N-193 and CMEG AR. A 1 µl aliquot of the “long target” PCR product was used as a DNA template for the nested reactions. The annealing temperatures for the PCR amplification in the nested reactions were 40 °C for the initial reactions (long target) and 45 °C for the nested

amplifications. The N2-N-1128 primer was designed based on the *Haematobia irritans* (Diptera: Muscidae) *nad2* sequence.

2.4. Cloning and sequencing

The amplified PCR products were purified using 0.05 µm filters as described in Lessinger and Azeredo-Espin (2000). An aliquot of the purified PCR product was cloned into the

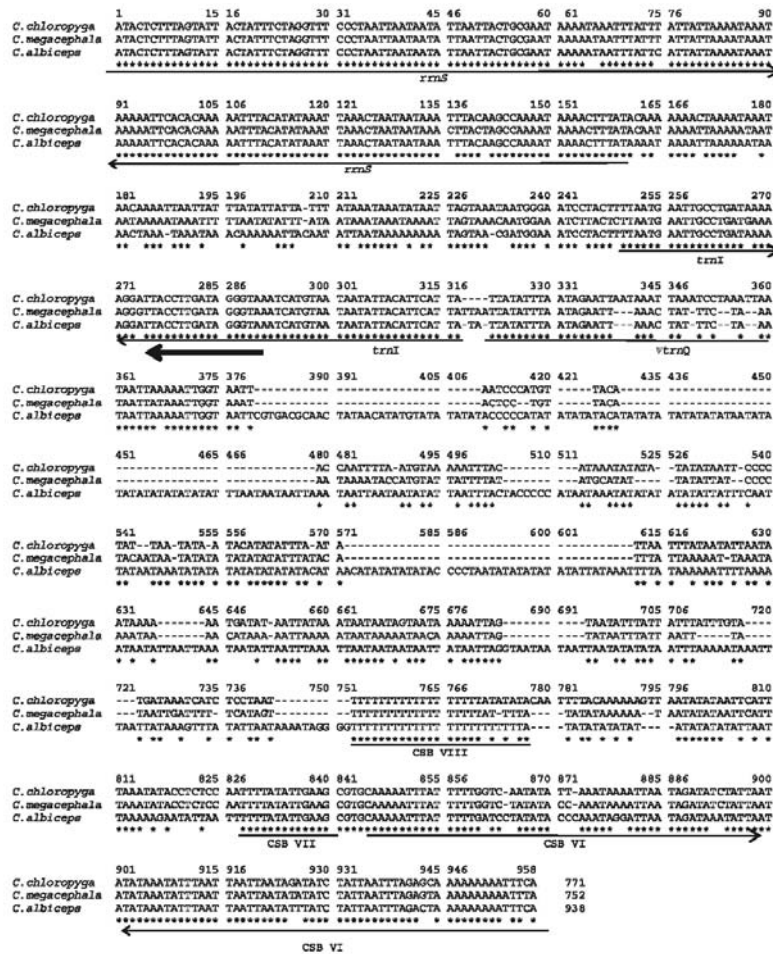


Fig. 2. Alignment of the mtDNA CR and adjacent genes of *Chrysomya* species showing partial sequences from *rns* (1–160), the duplicated region (233–361), and partial sequences from the Calliphoridae CR conserved domain—CSB VI, VII and VIII (751–958). The thick arrow identifies the alternative annealing site of the primer TN-I-24 (Table 1). The primary sequence elements identified in the alignment were the 19 bp repeat (233–251), the *trnI* copy (252–317), the *rns* pseudogene (322–361) and an 18 bp conserved sequence (362–379). The high level of nucleotide divergence (380–750) provides no alignment confidence in this interval.

pUC18—Sure Clone system (Amersham Pharmacia Biotech) according to the manufacturer's specifications. The clones were sequenced by automated DNA sequencing in an ABI 377 DNA sequencer (Amersham-Biosciences) using the ABI-PRISM BigDye terminator sequencing v.3 (Applied Biosystems). The sequencing of cloned PCR products was an efficient strategy for retrieving high quality sequences. Both strands of 3–5 clones were sequenced for each target region in each *Chrysomya* species.

2.5. Comparative and structural analysis

The CR nucleotide sequences of *C. albiceps*, *C. chloropyga* and *C. megacephala* were aligned using Clustal W (Thompson et al., 1994) set to default parameters with manual adjustments where required. The identification of conserved sequence blocks was based on Lessinger and Azeredo-Espin (2000). The tRNA secondary structure was determined using tRNAscan-SE software (Lowe and Eddy, 1997). Other secondary structures and their thermodynamic

properties were determined using Mfold software (Santa-Lucia, 1998).

The nucleotide sequences reported in this paper have the following GenBank accession numbers: AY220131, AY220132, AY220133 and AY220134. The AF151386 file of the *C. megacephala* CR (Lessinger and Azeredo-Espin, 2000) was updated in order to include additional sequence information for the *trnI* and *trnQ* mitochondrial genes.

3. Results and discussion

3.1. Structural organization of blowflies control region and flanking genes

Fig. 1 schematically shows the *Chrysomya* CR and its flanking genes, including the tRNA repeats (Q, I¹ and Q^ψ, I²). The alignment of the *C. albiceps*, *C. chloropyga* and *C. megacephala amplicon 1* sequences is shown in Fig. 2. Size variations in the CR sequences were evident for *C. albiceps*

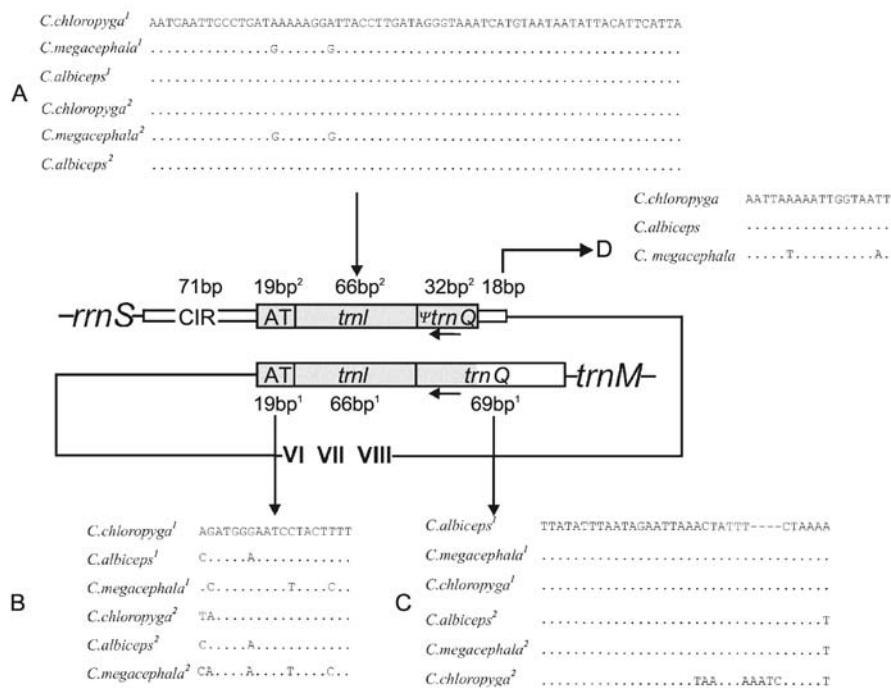


Fig. 3. The structural organization of duplicated elements close to the *Chrysomya* CR. Comparative analyses of homologous regions indicated different levels of conservation among these elements. ¹ and ² represent duplicated elements close to the *trnM* end and *rrnS* end, respectively. *trnQ* lies in the opposite strand relative to *trnI* and *trnM* (thick arrows). A—Alignment of the *Chrysomya trnI* genes (both copies). B—Alignment of the 19 bp element presenting the initial levels of nucleotide divergence between copies. C—Alignment of *Chrysomya trnQ* with the *trnQ* pseudogene ($\Psi trnQ$). D—Alignment of the conserved 18 bp element. A 71 bp region immediately adjacent to the *rrnS* gene is referred to as the *Chrysomya* intergenic region (CIR) in the text. Conserved sequence blocks (CSB) are indicated as VI, VII and VIII.

relative to the other two species and were partially related to the presence of numerous short tandem repeats $[AT]_n$. Amplification of the control region and adjacent genes in two separate sections (1 and 2 in Table 1) allowed the recovery of complete CR sequences for each species. The efficiency of the primers (CMEG-A, CMEG-AR and N2-N-1128) used to amplify *Chrysomya* CR sequences and flanking regions was confirmed.

A general overview of the structural organization of the duplicated regions of *Chrysomya* mtDNA is shown schematically in Fig. 3. Analysis of the CR flanking sequences from the IQM gene cluster allowed characterization of the *trnI* (66 bp), *trnQ* (69 bp) and *trnM* (partial sequence—19 bp) and two intergenic regions (*trnM/trnQ* and *trnQ/trnI*). The analysis of *amplicon 1* revealed duplicated sequences of *trnI* and *trnQ* in *C. albiceps*, *C. chloropyga* and *C. megacephala*. This duplication was conserved among these species, as indicated by amplification analysis of numerous individuals from each species (data not shown). The alignment of *amplicon 2* sequences of these species confirmed the conserved features associated with the CR of calliphorid and oestrid flies (Lessinger and Azeredo-Espin, 2000).

A BLAST search for sequences similar to the 71 bp sequence immediately adjacent to the *rns* gene provided no information about the identity or possible origin of this region. This 71 bp region between the duplicated *trnI* and the *rns* genes is referred to here as the *Chrysomya* intergenic region (CIR).

3.2. Duplicated sequences: conserved copy versus pseudogene

The entire duplicated region consisted of approximately 117 bp (Fig. 3), including 19 bp derived from CR sequences, 66 bp duplicated from the *trnI* gene and a 32 bp sequence with a *trnQ* origin. Two A→G transitions were identified in both *C. megacephala trnI* copies and represented the only divergent sites among these species. These transitions did not map at stem-forming sequences. Except for *C. megacephala*, the *trnI* copies of *Chrysomya* mtDNA were identical to the homologous single-copy gene of *Cochliomyia hominivorax* (Lessinger et al., 2000). When compared to the other *Chrysomya* (Fig. 3A) the divergent sites of the *C. megacephala trnI* copies represent a potential conflict for the hypothesis that the duplication was a single event that occurred before the diversification of this genus. In this case, gene conversion could account for the homogenization of the *trnI* sequences after the nucleotide substitutions in one copy. Nevertheless, the duplicated state of the *trnI* gene seems to be a stable feature of the *Chrysomya* mitochondrial genome.

The duplicated *trnQ* sequence close to the *rns* in *Chrysomya* was an incomplete or degenerated copy of this gene (pseudogene) as shown by comparison with the *trnQ* copy located in the IQM cluster. The partially duplicated gene corresponded to the *trnQ* 3' end, and consisted of the

acceptor stem, the TΨC and variable loops, and two initial nucleotides of the anticodon stem. The *trnM/trnQ* intergenic region was 8 bp in size (ATTATAAA) in *C. megacephala*, and 5 bp (ATTTA) in *C. chloropyga*. The *C. albiceps trnM/trnQ* junction overlapped by 1 bp. The *trnQ/trnI* intergenic sequences were also duplicated and conserved in both copies in each *Chrysomya* species. There were no intergenic nucleotides in *C. chloropyga*, while *C. albiceps* and *C. megacephala* had 2 (TA) and 4 (TTAA) bp, respectively (see Fig. 2). Homogenization of duplicated sequences by gene conversion could be invoked.

3.3. Conserved sequence blocks (CSB) in the control region of blowflies

All previously described CSBs for the mtDNA CR of myiasis-causing flies were assigned to the *Chrysomya* species analyzed, except for CSB V, which corresponds to the annealing site of the primers CMEG-A and CMEG-AR and, as expected, ought to be highly conserved in these species. The predicted secondary structure for the putative origin of *Chrysomya* mtDNA replication (OR—Fig. 4) was

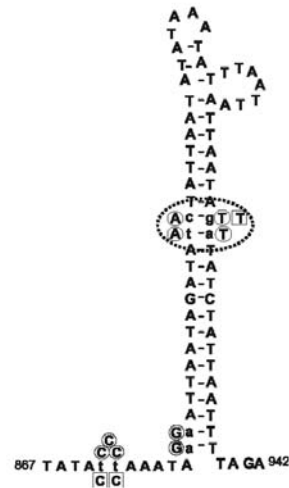


Fig. 4. The predicted secondary structure formed by conserved CR sequences in *Chrysomya*. This conserved structure is energetically stable (free energy values vary from -12.2 to -15.2 in these species) and represents a putative origin of replication (OR) for *Chrysomya* mtDNA. The nucleotide positions from 867 to 942 correspond to the relative positions of this structure in the alignment in Fig. 2. Conserved nucleotides among *Chrysomya* species are shown in capital letters; lowercase letters indicate nucleotide divergence in the corresponding *C. chloropyga* sequence; squares and circles indicate substitutions in the *C. megacephala* and *C. albiceps* sequences, respectively. The dotted circle indicates compensatory nucleotide substitutions resulting in the replacement of consecutive TA and CG pairs in *C. chloropyga* for two AT pairs in *C. albiceps*.

conserved as in other Calliphoridae species and its sequence included the largest conserved region identified so far in the CR of myiasis-causing flies.

3.4. Functional duplicated genes in animal mitochondrial genomes

The prediction of functional duplicated coding regions in animal mtDNA, as suggested for *Chrysomya trnI*, is restricted to the *trn* genes in *Mytilus* (Beagley et al., 1999), in an amphibiaenian (Macey et al., 1998), in snakes (Kumazawa et al., 1998), and in *Thrips* (Shao and Barker, 2003), and corresponds to the *trnM*, *trnP*, *trnF* and *trnS* genes, respectively. Recently, Hwang et al. (2001) re-examined non-annotated regions in published mitochondrial genome sequences and identified a “coding probability” for a second *trnY* in *Lumbricus terrestris*.

4. Conclusions

Although specific proprieties of the mtDNA CR may increase technical and methodological difficulties, the amplification of *Chrysomya* CR and flanking regions was efficiently optimized using the primers CMEG A, CMEG AR and N2-N-1128. This may represent an important contribution to further studies of blowfly evolution and DNA-based species identification.

The duplicated *trnI* genes of the blowflies *C. albiceps*, *C. chloropyga* and *C. megacephala* are conserved and may represent functional copies. However, it is still unclear whether gene conversion or selective constraints operate to maintain conserved gene copies.

Conserved sequence blocks (CSB) previously described for myiasis-causing flies (Lessinger and Azeredo-Espin, 2000) were also identified in the CR of these *Chrysomya* species and may play a role in mtDNA replication/transcription processes. We suggest that the duplication event occurred before diversification of *Chrysomya* genus, and that this feature may therefore be conserved in other *Chrysomya* species.

These results demonstrate and confirm the plasticity of the mtDNA molecule in *Chrysomya*, especially the *trn* genes and adjacent CR sequences, and should contribute to our understanding of insect mitochondrial genome evolution.

Acknowledgements

The authors thank Dr. Mark Dowton for his valuable suggestions on an earlier version of the manuscript. Two anonymous referees also provided helpful comments and

suggestions. We thank Rosângela A. Rodrigues, Cristina Feix de Abreu, Maria Salete C. Campos and Mário C. Filho for help in the DNA sequencing, and in collecting and rearing the *Chrysomya* samples. A.C.L. was supported by a PhD fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the other authors were supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

References

- Beagley, C.T., Okimoto, R., Wolstenholme, D.R., 1999. Mytilus mitochondrial DNA contains a functional gene for a trnS (UCN) with a dihydrouridine arm-replacement loop and a pseudo-trnS (UCN) gene. *Genetics* 152, 641–652.
- Boore, J.L., 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27, 1767–1780.
- Hwang, U.W., Friedrich, M., Tautz, D., Park, C.J., 2001. Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature* 413, 154–157.
- Infante, M.E., Azeredo Espin, A.M.L., 1995. Genetic variability in mitochondrial DNA of screwworm, *Cochliomyia hominivorax* (Diptera: Calliphoridae), from Brazil. *Biochem. Genet.* 33, 737–756.
- Kumazawa, Y., Ota, H., Nishida, M., Ozawa, T., 1998. The complete nucleotide sequence of a snake *Dinodon semicarinatus* mitochondrial genome with two identical control regions. *Genetics* 150, 313–329.
- Lessinger, A.C., Azeredo-Espin, A.M.L., 2000. Evolution and structural organization of mitochondrial DNA control region of myiasis-causing flies. *Med. Vet. Entomol.* 14, 71–80.
- Lessinger, A.C., Martins Junqueira, A.C., Lemos, T.A., Kemper, E.L., da Silva, F.R., Vettore, A.L., Arruda, P., Azeredo-Espin, A.M.L., 2000. The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Mol. Biol.* 9, 521–529.
- Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequences. *Nucleic Acids Res.* 25, 955–964.
- Macey, J.R., Schulte II, J.A., Larson, A., Papenfuss, T.J., 1998. Tandem duplication via light-strand synthesis may provide a precursor for mitochondrial genomic rearrangement. *Mol. Biol. Evol.* 15, 71–75.
- SantaLucia, J., Jr., 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U. S. A.* 95, 1460–1465.
- Shao, R., Barker, S.C., 2003. The highly rearranged mitochondrial genome of the plague thrips, *Thrips imuginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. *Mol. Biol. Evol.* 3, 362–370.
- Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved PCR primers. *Ann. Entomol. Soc. Am.* 87, 651–701.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Zhang, D.X., Hewitt, G.M., 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochem. Syst. Ecol.* 25, 99–120.

APÊNDICE B

The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae)

Ana Carolina M. Junqueira, Ana Cláudia Lessinger, Tatiana Teixeira Torres, Felipe Rodrigues da Silva, André Luiz Vettore, Paulo Arruda and Ana Maria L. Azeredo Espin

Gene 339 (2004): 7-15

RESUMO

Devido à importância médica, sanitária e forense das espécies de *Chrysomya*, a geração de sequências pode ser útil para a caracterização molecular desse gênero, além de ajudar no desenho de primers e na otimização da identificação molecular das espécies de califorídeos. Nesse trabalho, o genoma mitocondrial da mosca varejeira *Chrysomya chloropyga* (Diptera: Calliphoridae) foi completamente sequenciado. A molécula de DNA mitocondrial completa possui 15.837 pb e foi sequenciada usando a metodologia de “shotgun”. A composição nucleotídica total apresentou um viés alto para As e Ts, com um total de 76,7% do genoma. O gene *cox I* tem uma serina como códon de iniciação, enquanto códons de terminação incompletos mediados pela sinalização de tRNAs foram encontrados para *cox I*, *nd4* e *nd5*. Os genes de *C. chloropyga* estão na mesma ordem e orientação que os observados para outras espécies de dípteros, exceto pela ocorrência de uma região de 123 pb que inclui uma duplicação completa dos genes *tRNA^{Leu}* e uma duplicação parcial do *tRNA^{Gln}*. *C. chloropyga* é a primeira espécie de Diptera com 23 genes de tRNA descrita, ao invés dos 22 comumente descritos. Uma análise filogenética mostrou a subdivisão de Brachycera em Calyptrate e Acalyptrate. A sequência completa do mtDNA de *C. chloropyga* descrita neste trabalho será uma fonte útil de informações de sequências para estudos moleculares e evolutivos em Diptera.

The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae)

Ana Carolina M. Junqueira^{a,*}, Ana Cláudia Lessinger^a, Tatiana Teixeira Torres^a, Felipe Rodrigues da Silva^b, André Luiz Vettore^c, Paulo Arruda^d, Ana Maria L. Azeredo Espin^a

^aLaboratório de Genética Animal, Centro de Biologia Molecular e Engenharia Genética and Departamento de Genética e Evolução, Universidade Estadual de Campinas (UNICAMP), Campinas, SP, Brazil

^bEmbrapa Recursos Genéticos e Biotecnologia (CENARGEN), Brasília, DF, Brazil

^cLaboratório de Genética do Câncer, Instituto Ludwig de Pesquisa sobre o Câncer, São Paulo, SP, Brazil

^dLaboratório Genoma, Centro de Biologia Molecular e Engenharia Genética and Departamento de Genética e Evolução, Universidade Estadual de Campinas (UNICAMP), Campinas, SP, Brazil

Received 14 January 2004; received in revised form 20 May 2004; accepted 17 June 2004

Received by G. Pesole

Abstract

In view of the medical, sanitary and forensic importance of *Chrysomya* species, a knowledge of their nucleotide sequences would be useful for the molecular characterization of this genus, and would help in designing primers and in improving the molecular identification of Calliphoridae species. In this work, the mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae) was completely sequenced. The entire mitochondrial DNA (mtDNA) molecule was 15,837 bp long and was sequenced using the shotgun approach. The overall nucleotide composition was heavily biased towards As and Ts, which accounted for 76.7% of the whole genome. The *cox1* gene had a serine as the start codon, while incomplete termination codons mediated by tRNA signals were found for *cox2*, *nd4* and *nd5*. The *C. chloropyga* genes were in the same order and orientation as the mitochondrial genome of other dipteran species, except for the occurrence of a 123 bp region that included a complete duplication of *tRNA^{Asp}* and a partial duplication of *tRNA^{Gln}* genes. *C. chloropyga* is the first species of Diptera with 23 tRNA genes instead of the usual 22 already described. A phylogenetic analysis showed a split of Brachycera into Calyptratae and Acalyptratae subdivisions. The complete sequence of *C. chloropyga* mtDNA described here will be a useful source of sequence information for general molecular and evolutionary studies in Diptera.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Mitochondrial DNA; Forensic entomology; Myiasis; Calyptratae

1. Introduction

Mitochondrial DNA (mtDNA) has been extensively used in studies of phylogenetics, phylogeography, population structure and dynamics, and molecular evolution (Zhang and Hewitt, 1997; Boore, 1999). In general, animal mtDNA is a small (15–20 kb), circular, double-stranded molecule that encodes a set of 37 genes for 22 tRNAs, 13 protein coding genes (PCGs) and 2 rRNAs (Wolstenholme, 1992). Animal mtDNA also contains a major non-coding region related to the control of replication (Shadel and Clayton, 1993) and the regulation of transcription (Chang and

Abbreviations: A, adenine; *atp6* and 8, ATPase subunits 6 and 8; bp, base pairs; C, cytosine; *cox1–3*, cytochrome oxidase c subunits 1, 2 and 3; CSB, conserved sequence blocks; *cytb*, cytochrome b; G, guanine; IR, intergenic region; mtDNA, mitochondrial DNA; MP, maximum parsimony; *nd1–6* and *4L*, NADH dehydrogenase subunits 1–6 and 4L; NJ, neighbor-joining; ORF, open reading frame; PCG, protein coding gene; *rrnS*, small subunit of rRNA gene; *rrnL*, large subunit of rRNA gene; T, thymine; *tRNA*, transfer RNA gene.

* Corresponding author. Tel.: +55 19 37881141; fax: +55 19 37881089.

E-mail address: anacmj@unicamp.br (A.C.M. Junqueira).

0378-1119/\$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.06.031

Clayton, 1984; Carrodegua and Vallejo, 1997). In insects, this portion is also known as the A+T-rich region because of the high content of these two nucleotides. The widespread use of mtDNA as an informative evolutionary marker is based on its maternal inheritance, accelerated rates of nucleotide substitution, the presence of polymorphisms, and lack of recombination (Wolstenholme, 1992).

Complete mitochondrial genome sequences are available for more than 400 metazoans, thus confirming the versatility of this genetic marker for molecular and evolutionary studies. Among the animal mtDNA sequences, there are more than 50 complete genomes for the Arthropoda, only eight of which are for dipterans: *Drosophila yakuba* (X03240; Clary and Wolstenholme, 1987), *Drosophila melanogaster* (U37541; Clary et al., 1982; Lewis et al., 1995), *Anopheles gambiae* (L20934; Beard et al., 1994), *Anopheles quadrimaculatus* (L04272; Mitchell et al., 1993), *Ceratitidis capitata* (AJ242872; Spanos et al., 2000), *Cochliomyia hominivorax* (AF260826; Lessinger et al., 2000), *Bactrocera oleae* (AY210702 and AY210703; Nardi et al., 2003a), and *Chrysomya chloropyga* (AF352790, this work). Most of the dipteran mitochondrial genomes that have been sequenced are for species of medical and economic importance.

C. chloropyga (= *Chrysomya putoria*) is a tropical and subtropical blowfly native to Africa that was introduced into the New World around 1975 and showed a rapid dispersal (Baumgartner and Greenberg, 1984; Wells, 1991). Subsequent investigations have shown that three other Old World *Chrysomya* species (*C. albiceps*, *C. megacephala* and *C. ruffifacies*) have also been introduced into New World. These invader species have displaced native species through competition for substrate and ecological niches (Laurence, 1988).

C. chloropyga and the other *Chrysomya* species are of considerable medical and sanitary importance since they occur in garbage, feces and carcasses, and may act as a vector of pathogens. These species are also secondary agents of myiasis in man and animals (Baumgartner and Greenberg, 1984) and, together with other members of the family Calliphoridae, are of potential forensic importance because they feed and breed on exposed corpses (Souza and Linhares, 1997).

Of the 1020 Calliphoridae species already described, only 45 are listed in GenBank, with a total of almost 500 entries. Of these, 96 sequences are related to the genus *Chrysomya*, including 64 entries for mtDNA regions. The genes for subunits 1 and 2 of cytochrome oxidase (*cox1* and *cox2*) are the most represented, with 191 entries for Calliphoridae species and 57 for *Chrysomya* (GenBank, December 2003). According to Caterino et al. (2000), the *cox1* gene is the most commonly used mitochondrial marker for evolutionary studies across a broad range of insects.

In this study, we report the entire mtDNA sequence for the blowfly *C. chloropyga*. The mtDNA genome of this species is an important source of sequence information for

improving primer selection and design, and for the identification of diagnostic regions suitable for species-specific identification.

2. Material and Methods

2.1. Fly strain

Specimens of *C. chloropyga* were collected at São João da Boa Vista, São Paulo State, southeastern Brazil, and maintained in the laboratory as described by Infante and Azeredo-Espin (1995).

2.2. DNA extraction and sequencing

Purified mtDNA for sequencing was obtained as described by Azeredo-Espin et al. (1991), using a cesium chloride gradient. Eighteen micrograms of mtDNA were used to construct a mitochondrial genomic library by the shotgun approach (Anderson et al., 1982). The mtDNA was sheared by sonication and the fragments were blunted using *Klenow* fragment of DNA polymerase and T4 DNA polymerase. Fragments of 1–2 kb were isolated from a low-melting agarose electrophoresis gel and cloned into the vector pUC18 *SmaI*/BAP using a Ready-to-Go kit (Amersham Biosciences, Piscataway, NJ), according to the manufacturer's instructions. The two strands of the clones were sequenced automatically in an ABI Prism 3700 sequencer using the Big Dye terminator kit (PE Applied Biosystems, Foster City, CA).

2.3. Assembling, annotation and phylogenetic analysis

The final consensus sequence was assembled using the phred/phrap/consed package (Gordon et al., 1998). Open reading frames (ORFs) were determined using the ORF finder of Sequin Application (version 3.2) from NCBI ([ftp://ftp.ncbi.nih.gov/sequin/](http://ftp.ncbi.nih.gov/sequin/)) and the similarities between the primary nucleotide sequences and described genes were determined with BLASTX (Altschul et al., 1997). The complete sequence annotation submitted to GenBank was done on Sequin Application 3.2. The tRNA secondary structures were predicted using the tRNAscan-SE 1.1 program (Lowc and Eddy, 1997). The codon usage bias was analyzed with DnaSP 3.51 (Rozas and Rozas, 1999).

The amino acid sequences of all the PCGs were aligned individually using Clustal X version 1.83 (Chenna et al., 2003). After the alignment, a single sequence, corresponding to a set of concatenated PCGs was obtained and used for phylogenetic reconstruction. Phylogenetic trees were reconstructed with the neighbor-joining (NJ) tree-building method, using MEGA (Kumar et al., 2001) and PAUP* 4.0b10 (Swofford, 2000). The NJ trees were reconstructed with the K2P distance, using 5000 replicates for bootstrap statistics.

The complete mitochondrial genome sequence of *C. chloropyga* is available through the NCBI/EMBL/DDBJ nucleotide databases under the accession number AF352790.

3. Results and Discussion

3.1. Genome organization and skewness

The entire mitochondrial genome of *C. chloropyga* was 15837 base pairs (bp) long, with a nucleotide composition heavily biased towards As and Ts. The overall A+T content was 76.7% (A=39.1%, C=13.6%, G=9.7%, T=37.6%), similar to that of *Co. hominivorax* (77%) and other dipterans (Table 1). The genome content also included the same set of 13 PCGs and 2 rRNAs. However, the *C. chloropyga*

mtDNA had a total of 23 tRNAs instead of the 22 usually described for Arthropoda.

The base composition of nucleotide sequences can be described by skewness (Perna and Kocher, 1995), which measures the relative number of As to Ts (AT skew=[A-T]/[A+T]) and Gs to Cs (GC skew=[G-C]/[G+C]). As shown in Table 1, there was virtually no AT skew (0.02) when the major strand of the entire mtDNA of *C. chloropyga* and other dipterans was considered, indicating that As and Ts occurred in almost the same proportion. In contrast, the GC skew is slightly negative (-0.17) for the entire mtDNA, indicating the occurrence of more Cs than Gs. According to Saccone et al. (1999), a bias against the use of Gs in this strand is characteristic of the metazoan mitochondrial genome.

When considering the whole set of PCGs, the GC skews had a value close to zero (0.03 for *C. chloropyga*),

Table 1
Comparison of the A+T content and the AT and GC skews of different classes of genes in the mitochondrial genome of dipterans

Sequence class	Species	Total length (bp)	A+T%	AT skew	GC skew	
Whole mtDNA	<i>C. chloropyga</i>	15837	76.7	0.020	-0.170	
	<i>Co. hominivorax</i>	16022	77.0	0.034	-0.207	
	<i>B. oleae</i>	15815	72.6	0.088	-0.280	
	<i>Ce. capitata</i>	15980	77.5	0.021	-0.185	
	<i>D. melanogaster</i>	19517	82.2	0.017	-0.150	
	<i>D. yakuba</i>	16019	78.6	0.005	-0.136	
	<i>A. gambiae</i>	15363	77.6	0.032	-0.154	
	<i>A. quadrimaculatus</i>	15455	77.4	0.041	-0.181	
	PCGs (+)	<i>C. chloropyga</i>	6843	73.4	-0.118	-0.105
		<i>Co. hominivorax</i>	6843	73.1	-0.096	-0.147
		<i>B. oleae</i>	6892	67.7	-0.048	-0.243
		<i>Ce. capitata</i>	6837	73.9	-0.125	-0.124
		<i>D. melanogaster</i>	6837	75.8	-0.125	-0.061
<i>D. yakuba</i>		6837	75.2	-0.135	-0.055	
<i>A. gambiae</i>		6846	74.7	-0.098	-0.079	
<i>A. quadrimaculatus</i>		6849	74.3	-0.090	-0.103	
PCGs (-)		<i>C. chloropyga</i>	4275	77.3	-0.228	0.282
		<i>Co. hominivorax</i>	4275	77.3	-0.225	0.304
		<i>B. oleae</i>	4297	74.1	-0.295	0.378
		<i>Ce. capitata</i>	4362	78.1	-0.233	0.286
		<i>D. melanogaster</i>	4272	79.3	-0.218	0.244
	<i>D. yakuba</i>	4305	78.9	-0.197	0.243	
	<i>A. gambiae</i>	4313	78.0	-0.229	0.241	
	<i>A. quadrimaculatus</i>	4299	77.1	-0.238	0.282	
	rRNAs	<i>C. chloropyga</i>	2114	80.0	-0.007	0.286
		<i>Co. hominivorax</i>	2110	79.8	-0.026	0.305
		<i>B. oleae</i>	2116	77.1	-0.099	0.347
		<i>Ce. capitata</i>	2123	80.2	-0.058	0.267
		<i>D. melanogaster</i>	2111	81.9	-0.005	0.288
<i>D. yakuba</i>		2115	81.8	0.009	0.276	
<i>A. gambiae</i>		2125	81.4	-0.053	0.301	
<i>A. quadrimaculatus</i>		2115	81.6	-0.056	0.297	
rRNA		<i>C. chloropyga</i>	1537	76.1	-0.009	0.136
		<i>Co. hominivorax</i>	1470	76.6	-0.016	0.128
		<i>B. oleae</i>	1466	74.8	-0.015	0.086
		<i>Ce. capitata</i>	1472	76.8	-0.009	0.099
		<i>D. melanogaster</i>	1457	77.1	-0.012	0.141
	<i>D. yakuba</i>	1465	76.6	-0.023	0.155	
	<i>A. gambiae</i>	1478	77.9	-0.003	0.131	
	<i>A. quadrimaculatus</i>	1478	78.6	0.009	0.146	

(+)=major strand; (-)=minor strand.

while the AT skew was negative (−0.16). These findings indicate that there were generally more Ts than As in the PCGs, while Gs and Cs were almost equivalent, with a slight dominance of Cs.

Comparison of the AT and GC skews calculated for dipterans showed that *C. chloropyga* followed the same pattern of bias for all classes of genes and for the entire mtDNA.

3.2. Protein coding genes

The order of PCGs in *C. chloropyga* was identical to that described for a large number of Arthropoda species. The initiation and termination codons were identified using the ORF finder and by comparison with the complete mitochondrial sequences of other dipterans. ATG was the most used initiation codon (6), but ATT (3), ATA (2) and ATC (1) are also used (Table 2). For the *cox1* gene, neither these codons nor the proposed ATAA initiation in *Drosophila* and *Locusta* were found in *C. chloropyga*. Nevertheless, the hexanucleotide ATTTAA involved in initiation signaling was found immediately adjacent to a TCG codon. Therefore, the TCG for a serine was assigned as the *cox1* start codon (Table 2). The occurrence of serine as the first amino acid instead of methionine has been described for a number of species. This hexanucleotide signal was also detected in mosquitoes (Mitchell et al., 1993; Beard et al., 1994) and in the basal hexapod *Tetrodontophora bielanensis* (Nardi et al., 2001). Of the 13 species of Brachycera with complete *cox1* sequences available in GenBank, only four species of *Drosophila* have the ATAA initiation, while *Ce. capitata*, *B. oleae* and the remaining seven species of the family Calliphoridae [*C. rufifacies* (AF083658), *C. albiceps* (AF083657), *Co. hominivorax* (AF260826), *Calliphora vicina* (AJ417702), *Lucilia caesar* (AJ417703), *L. cuprina* (AJ417711), and *L. sericata* (AJ417717)] had the ATTTAA sequence flanking the TCG start codon.

Similarly, the usual TAA and TAG termination codons were found for nine PCGs and one PCG, respectively. The

Table 3
Nucleotide frequencies for the first, second and third codon positions in the major and minor strands of *C. chloropyga* PCGs

		Codon position		
		1st	2nd	3rd
Major strand	T	0.35	0.44	0.44
	C	0.14	0.22	0.08
	A	0.31	0.20	0.46
	G	0.21	0.13	0.02
Minor strand	T	0.42	0.48	0.52
	C	0.08	0.16	0.01
	A	0.30	0.20	0.40
	G	0.20	0.16	0.08

cox2, *nd4* and *nd5* genes had incomplete termination codons (Table 2). These codons are commonly found in metazoan mitochondrial genes and the tRNAs may play a major role in the correct release of transcripts during mRNA processing. Post-transcriptional polyadenylation is the most likely mechanism for the creation of a complete TAA stop codon (Ojala et al., 1981).

The A+T content calculated from the coding strand of each gene was 74.9% (T=43.5%, C=12.2%, A=31.4%, G=12.9%) in *C. chloropyga*. The A+T content of the minor strand was slightly higher (77.3%) than for the major strand (73.4%).

The frequency of each nucleotide by codon position is shown in Table 3. Thymine was the most frequent nucleotide used in the PCGs, except for the third codon position of the major strand, where A was more used. These results agreed with the AT skew numbers since the PCG skews for both strands were negative (Table 1). There was a predominance of As and Ts mainly in the first and third codon positions. A bias towards these nucleotides is widely observed in invertebrate mtDNA protein genes, especially in the third codon position associated with a degenerated genetic code. Table 3 also shows a low frequency for C at the third codon position of the minor strand and a low frequency of Gs in the third codon position of the major strand, which also agreed with the GC skews in Table 1.

Table 2
Initiation and termination codons for mitochondrial PCGs in *C. chloropyga*

Genes	Position	Initiation	Termination	Length (aa)	Accession numbers
<i>nd2</i>	210–1226	ATT (M)	TAA	338	AAK21318
<i>cox1</i>	1423–2921	TCG (S) ^a	TAA	512	AAK21319
<i>cox2</i>	3029–3716	ATG (M)	T	229	AAK21320
<i>atp8</i>	3854–4018	ATC (M)	TAA	54	AAK21322
<i>atp6</i>	4012–4689	ATG (M)	TAA	225	AAK21321
<i>cox3</i>	4689–5477	ATG (M)	TAA	262	AAK21323
<i>nd3</i>	5546–5902	ATA (M)	TAA	118	AAK21324
<i>nd5</i>	6322–8041	ATT (M)	T	573	AAK21325
<i>nd4</i>	8122–9460	ATG (M)	T	446	AAK21326
<i>nd4L</i>	9454–9750	ATG (M)	TAA	98	AAK21327
<i>nd6</i>	9886–10410	ATT (M)	TAA	174	AAK21328
<i>cytb</i>	10410–11546	ATG (M)	TAG	378	AAK21329
<i>nd1</i>	11629–12567	ATA (M)	TAA	312	AAK21330

^a The hexanucleotide ATTTAA flanking the start codon was assigned as an initiation signal.

Table 4
C. chloropyga codon usage of PCGs

Codon(aa)	n(RSCU)	Codon(aa)	n(RSCU)	Codon(aa)	n(RSCU)	Codon(aa)	n(RSCU)
UUU(F)	266.0(1.56)	UCU(S)	102.0(2.39)	UAU(Y)	139.0(1.64)	UGU(C)	34.0(1.94)
UUC(F)	75.0(0.44)	UCC(S)	8.0(0.19)	UAC(Y)	30.0(0.36)	UGC(C)	1.0(0.06)
UUA(L)	501.0(5.01)	UCA(S)	119.0(2.79)	UAA(*)	0.0(0.00)	UGA(W)	96.0(1.96)
UUG(L)	23.0(0.23)	UCG(S)	2.0(0.05)	UAG(*)	0.0(0.00)	UGG(W)	2.0(0.04)
CUU(L)	41.0(0.41)	CCU(P)	83.0(2.48)	CAU(H)	56.0(1.47)	CGU(R)	14.0(0.95)
CUC(L)	1.0(0.01)	CCC(P)	4.0(0.12)	CAC(H)	20.0(0.53)	CGC(R)	0.0(0.00)
CUA(L)	34.0(0.34)	CCA(P)	47.0(1.40)	CAA(Q)	69.0(1.94)	CGA(R)	44.0(2.98)
CUG(L)	0.0(0.00)	CCG(P)	0.0(0.00)	CAG(Q)	2.0(0.06)	CGG(R)	1.0(0.07)
AUU(I)	324.0(1.92)	ACU(T)	107.0(2.25)	AAU(N)	173.0(1.81)	AGU(S)	70.0(1.64)
AUC(I)	13.0(0.08)	ACC(T)	4.0(0.08)	AAC(N)	18.0(0.19)	AGC(S)	3.0(0.07)
AUA(M)	181.0(1.72)	ACA(T)	77.0(1.62)	AAA(K)	36.0(0.78)	AGA(S)	37.0(0.87)
AUG(M)	30.0(0.28)	ACG(T)	2.0(0.04)	AAG(K)	56.0(1.22)	AGG(S)	0.0(0.00)

A total of 3706 codons were analyzed, excluding the initiation and termination codons. RSCU=relative synonymous codon usage; n=frequency of each codon.

* Stop codons.

The over-representation of Ts in the second codon position (Table 3) may be related to a preference for nonpolar and hydrophobic amino acids in membrane-associated proteins (Asakawa et al., 1991) since phenylalanine, leucine, isoleucine, methionine and valine are nonpolar and hydrophobic amino acids with codons that have a T at the second position (Helfenbein et al., 2001).

The codon usage of *C. chloropyga* mtDNA PCGs and the relative synonymous codon usage (RSCU) values are given in Table 4. As expected, most of the values differed from the equilibrium frequency and the codon bias index (CBI) of 0.726 clearly showed that the use of synonymous codons was heavily distorted. Comparisons with the codon usage of other dipterans revealed similar patterns. The codons CTG (Leu), CGC (Arg), CCG (Pro) and AGG (Ser) were not used by *C. chloropyga* mitochondrial PCGs. A comparative analysis among dipterans using the Organellar Genome Retrieval database (OGRe; Jameson et al., 2003), indicated that CTG was also not used by *A. quadrimaculatus*, *Co. hominivorax* and *D. melanogaster*. The codon CGC was not used in PCGs of *A. gambiae*, *A. quadrimaculatus*, *D. melanogaster* and *Co. hominivorax*, whereas the CCG codon was used by all dipterans, except for *Ce. capitata* and *C. chloropyga*. Finally, only *Co. hominivorax* used ACG. Since insect mtDNAs show a high bias against C and G, this could explain the lack of use of these codons.

3.3. rRNA genes

The large mitochondrial rRNA subunit (*rnl*) had a total length of 1329 bp, with an A+T content of 81.8% and AT

and GC skews of -0.01 and 0.31 , respectively. The small subunit (*rns*) was 785 bp long, with an A+T content of 76.9% and AT and GC skews of 0 and 0.26, respectively.

Compared to other dipterans, the length, skewness and nucleotide composition of *rnl* and *rns* were conserved. The A+T content of the rRNA genes was the highest in the mitochondrial genome, excluding the control region. The sequence identity between these genes and the rRNA genes of other dipterans was >80% (Table 5).

3.4. tRNA genes and the control region

C. chloropyga mtDNA encoded 23 tRNAs instead of the usual 22 tRNAs described for other Arthropoda. In addition to *tRNA^{Ser}* and *tRNA^{Leu}*, which usually have two copies in mtDNA, the *C. chloropyga* mitochondrial genome also had a second tRNA for the isoleucine flanking the variable domain of the control region, near the *rns* end. The entire duplicated region was 123 bp long in *C. chloropyga*, including 17 bp from the conserved domain of the control region, 66 bp from *tRNA^{Ile}* and a further 40 bp from a partial sequence of the *tRNA^{Gln}* gene. The two copies of *tRNA^{Ile}* had identical sequences, including the GAU anticodon. Analysis of the control region sequence of *C. megacephala* (AF151386; Lessinger and Azeredo-Espin, 2000) confirmed the occurrence of the *tRNA^{Ile}* duplication in this *Chrysomya* species as well. The same analysis for control region sequences in other genera of Calliphoridae, such as *Cochliomyia* and *Lucilia*, detected no such duplication.

Secondary structure analysis revealed the absence of a DHU arm in the *tRNA^{Ser(AGN)}* gene, where nucleotide sequence pairing was unlikely. This is a common feature

Table 5
Comparison of the rRNA sequence identities between *C. chloropyga* and other dipterans with completely sequenced mtDNA

CHRCHL	COCHOM	DROYAK	DROMEL	CERCAP	BACOLE	ANOQUA	ANOGAM
rnl	96%	89%	88%	86%	84%	82%	82%
rns	96%	88%	87%	86%	85%	82%	81%

CHRCHL=*C. chloropyga*; COCHOM=*Co. hominivorax*; DROYAK=*D. yakuba*; DROMEL=*D. melanogaster*; CERCAP=*Ce. capitata*; BACOLE=*B. oleae*; ANOQUA=*A. quadrimaculatus*; ANOGAM=*A. gambiae*.

already described in metazoan mtDNA (Wolstenholme, 1992). The remaining tRNA genes resembled the structure described for other arthropod mitochondrial genomes. Seventeen non-Watson-and-Crick pairings and five mismatches were identified in the tRNA genes, and were located in the DHU (seven), T ψ C (five), anticodon (four) and acceptor (six) arms.

The major noncoding region apparently corresponded to the control region because of its location (and the MQI-tRNA cluster), high A+T content (89.5%) and the presence of conserved sequence blocks that have been described in the control regions of other Calliphoridae species (Lessinger and Azeredo-Espin, 2000) and may control gene transcription and mtDNA replication.

Excluding the 123 bp duplicated region and 73 bp with no sequence similarity in GenBank, the control region of *C. chloropyga* had 812 bp. The 73 bp immediately adjacent to the *rrnS* were assigned as an intergenic region (IR) because of the lack of sequence conservation among *Chrysomya* species and the lack of information about the identity of this region. The comparison of a *D. yakuba*-like mtDNA control region with the *C. chloropyga* control region is shown in Fig. 1. The *C. chloropyga* control region was shorter than that described for Brachycera species, which ranges from 930 bp in *D. guanche* (Brehm et al., 2001) to 4601 bp in *D. melanogaster* (Lewis et al., 1995). Differences in the total length of insect mitochondrial sequences mainly reflect variations in the size of the control region (Zhang and Hewitt, 1997).

The control region has been particularly difficult to characterize because of its variable sequence and high A+T content, which tend to reduce the number of efficient annealing sites. Nevertheless, the alignment of conserved sequence blocks (CSB) described by Lessinger and Azeredo-Espin (2000) has revealed a potential site for primer annealing that consists of a region located on the CSB V. This primer could be combined with universal primers on *rrnS* and *tRNA* genes to recover two different fragments from the control region. The CSB V is conserved when aligned with the homologous region of another Calliphoridae species deposited in GenBank (Lessinger et al., 2004).

Table 6

Noncoding and overlapping regions in the mitochondrial genome of *C. chloropyga*

Nucleotide position	Genes	Strands	Size (bp)
<i>Noncoding regions</i>			
135–141	<i>tRNA^{Gln}/tRNA^{Met}</i>	–/+	5
1348–1358	<i>tRNA^{Cys}/tRNA^{Tyr}</i>	–/–	9
3022–3029	<i>tRNA^{Leu}/cox2</i>	+/+	8
5477–5484	<i>cox3/tRNA^{Gly}</i>	+/+	5
5902–5905	<i>nd3/tRNA^{Ala}</i>	+/+	2
6032–6034	<i>tRNA^{Arg}/tRNA^{Asn}</i>	+/+	1
6168–6170	<i>tRNA^{Ser}/tRNA^{Glu}</i>	+/+	1
6236–6255	<i>tRNA^{Glu}/tRNA^{Phe}</i>	+/–	18
8041–8057	<i>nd5/tRNA^{Tsu}</i>	–/–	15
9750–9753	<i>nd4L/tRNA^{Tyr}</i>	–/+	2
9883–9886	<i>tRNA^{Pro}/nd6</i>	–/+	2
11612–11629	<i>tRNA^{Ser}/nd1</i>	+/–	16
12567–12578	<i>nd1/tRNA^{Leu}</i>	–/–	10
12642–12645	<i>tRNA^{Leu}/rRNA</i>	–/–	2
	Total		96
<i>Overlapping regions</i>			
1226–1225	<i>nd2/tRNA^{Tyr}</i>	+/+	2
1292–1285	<i>tRNA^{Tyr}/tRNA^{Cys}</i>	+/–	8
1424–1423	<i>tRNA^{Tyr}/cox1</i>	–/+	2
2961–2957	<i>cox1/tRNA^{Leu}</i>	–/+	5
4018–4012	<i>atp8/atp6</i>	+/+	7
4689	<i>atp6/cox3</i>	+/+	1
5548–5546	<i>nd3/tRNA^{Glu}</i>	+/+	3
5969	<i>tRNA^{Ala}/tRNA^{Arg}</i>	+/+	1
6099	<i>tRNA^{Asn}/tRNA^{Ser}</i>	+/+	1
9460–9454	<i>nd4/nd4L</i>	–/–	7
10410	<i>nd6/cytb</i>	+/+	1
11546–11545	<i>cytb/tRNA^{Ser}</i>	+/+	2
14045	<i>tRNA^{Val}/srRNA</i>	–/–	1
	Total		41

(+) major strand; (–) minor strand.

3.5. Noncoding and overlapping regions

In addition to the control region and the 73 bp related to the IR, another 14 noncoding regions varying from 1 to 18 bp were also identified in the mitochondrial genome of *C. chloropyga* (Table 6). Several noncoding regions have been reported for other arthropod mtDNAs, with a total length

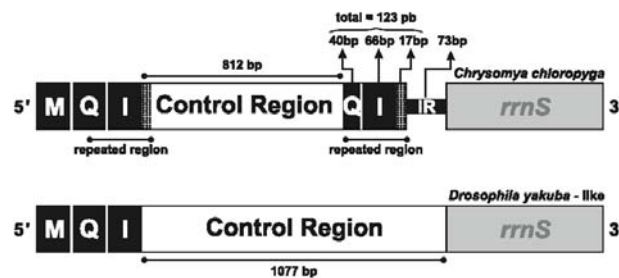


Fig. 1. Comparison of the *C. chloropyga* and a *D. yakuba*-like control region showing the duplication of *tRNA^{His}*.

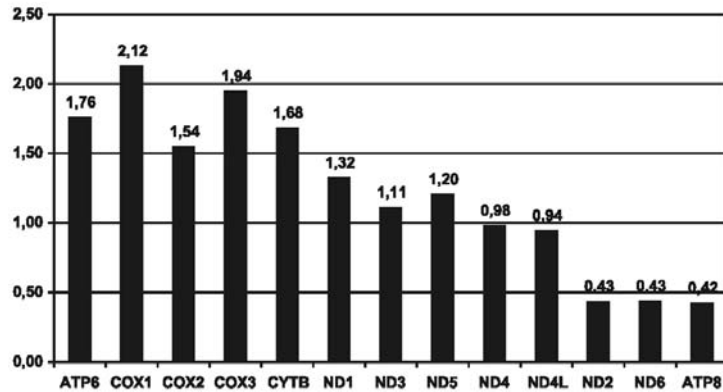


Fig. 2. The C/V ratio of aligned mitochondrial PCG amino acid sequences. Only eight genes with a C/V>1 were used for phylogenetic analysis.

ranging from 43 bp in *Anopheles* to 620 bp in *Apis* (Crozier and Crozier, 1993).

Forty-one base pairs were identified as overlapping sequences ranging from 1 to 8 bp in 13 regions (Table 6). The 13 PCGs of *C. chloropyga* showed four overlaps with the reading frames involving the *atp8/atp6*, *atp6/cox 3*, *nd4/nd4L* and *nd6/cytb* genes.

3.6. Phylogenetic perspectives

The 13 aligned amino acid PCG sequences were analyzed and divided in two groups according to the conserved (C) and variable (V) sites ratio. The genes with a C/V amino acid ratio >1 (Fig. 2) were concatenated and used for phylogenetic analysis. Thus, only eight PCGs (ATP6, COX1, COX2, COX3, CYTB, ND1, ND3 and ND5) were used for further phylogenies. The PCGs with a C/V ratio <1 (Fig. 2) were excluded because of the high variation and unreliable alignments. In these genes, the rate of amino acid change could cause distortions in the phylogenetic reconstruction. Of the five excluded PCGs, at least two (ATP8 and ND6) are commonly excluded from

phylogenies using mitochondrial genomes. The eight individual amino acid sequences were concatenated with the sequences of other dipterans, and the PCG sequences of *Locusta migratoria* (X80245; Flook et al., 1995) were used as an outgroup. A total of 2597 sites were used for distance method reconstructions after removing the gaps. Of these, 975 sites were variable.

Despite the low number of taxa sampled, the mitochondrial PCGs were used for phylogenetic analyses to assess their applicability as reliable molecular markers since the use of concatenated mtDNA sequences for resolving phylogenetic problems has increased in recent years. The topology obtained (Fig. 3) showed two clusters of Diptera consisting of Brachycera (flies) and Nematocera (mosquitoes). The low number of complete mtDNA sequences for Diptera species in GenBank meant that these relationships could not be confirmed, but there was a phylogenetic signal indicating the splitting of this order into two clusters.

The suborder Brachycera has been divided into Calyptratae (Calliphoridae) and Acalyptratae (Drosophilidae+Tephritidae), which agrees with the reconstruction described by Lessinger et al. (2000). According to Yeates

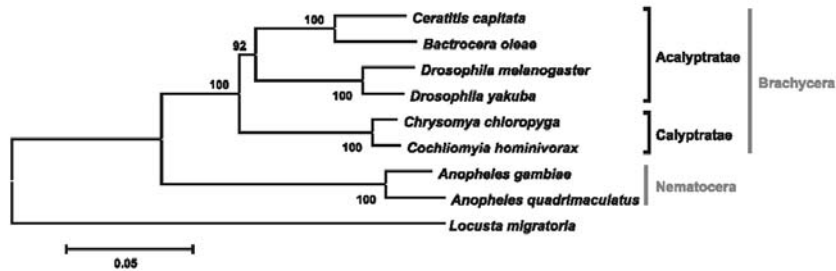


Fig. 3. Phylogenetic reconstruction of dipterans for which the mtDNA has been completely sequenced. The NJ bootstrap for 5000 replicates is indicated in each node. The ATP8, DN2, ND4, ND4L and ND6 amino acid sequences have been removed.

and Wiegmann (1999), the Calypterae have been recognized as a major lineage of higher Diptera, and morphological support for this clade is strong. However, the monophyly of the Acalypterae is contradictory and the relationships between the superfamilies of this taxon remain obscure. Indeed, other molecular phylogenetic studies using mitochondrial genomics have indicated a different relationship, which disagrees with the classic interpretation of dipteran phylogeny (Nardi et al., 2001, 2003b).

The CYTB, ND1 and ND5 PCGs individually reconstructed the same topology shown in Fig. 3, but the remaining PCGs showed the clustering of Drosophilidae and Calliphoridae as sister groups, thus rejecting the Acalypterae taxa. This is particularly important since COX1 and COX2 have different topologies, even though they are the most used genes for molecular systematics studies in Diptera and Hexapoda. When the complete set of PCGs was used, the tree presented here was not recovered. When the ND6 and ATP8 sequences, which are normally excluded in such analyses, were removed, the NJ method reconstructed the same topology shown in Fig. 3.

4. Conclusions

The results presented here show a typical metazoan mtDNA molecule with standard length, nucleotide composition, organization and set of genes. However, *C. chloropyga* has a duplicated region that includes a complete *tRNA* gene flanking the control region. This is the first description of a dipteran species with 23 *tRNA* genes instead of the usual 22. Among the Arthropoda, only *Thrips imaginis* (AF335993; Shao et al., 2001) has 23 *tRNA* genes, with three copies of the *tRNA^{Ser}* gene. Characterization of the mtDNA of another *Chrysomya* species indicates that the duplicated region could be a useful molecular marker for these Calliphoridae species, particularly for larval identification because of the lack of diagnostic morphological traits in this developmental stage.

Molecular phylogenetic trees for dipteran species usually yield contradictory reconstructions, depending on the number of species included in the study and the genes used in the analyses. As indicated here, the tree obtained using the genes that showed more conserved than variable sites yielded two clusters of Brachycera and separated the Calypterae from the Acalypterae. This criterion provided a phylogenetic reconstruction with high bootstrap values that supported each node, but a wide range of species must be included in this analysis to support the relationships within Diptera. For a better understanding of Brachycera evolution, the mitochondrial genome of another species needs to be completely sequenced. In addition, phylogenetic reconstructions with individual genes may indicate appropriate genes to clarify lower phylogenies. Other nuclear

molecular markers should also be used to address this question in order to expand the data set used in dipteran molecular phylogenies.

For *C. chloropyga* in particular, the nucleotide sequence data provided here should be useful for the molecular characterization and analysis of species of medical, sanitary and forensic importance. The complete sequence of the mitochondrial genome of this species is also of particular interest as a source of information for general molecular and evolutionary studies of Diptera.

Acknowledgements

The authors thank Dr. A.X. Linhares for providing the flies that were reared in the laboratory, and R.A. Rodrigues and M.S. Couto for valuable technical assistance. We also thank the two anonymous referees for their helpful suggestions. This research was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, grant 00/13814-6) to A.M.L.A.E. A.C.M.J. is supported by a PhD fellowship from FAPESP (grant 03/05444-2).

References

- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein data base search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Anderson, S., de Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F., Young, I.G., 1982. Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian genome. *J. Mol. Evol.* 15, 683–717.
- Asakawa, S., Kumazawa, Y., Araki, T., Himeno, H., Miura, K., Watanabe, K., 1991. Strand specific nucleotide bias in echinoderm and vertebrate mitochondrial genomes. *J. Mol. Evol.* 32, 511–520.
- Azeredo-Espin, A.M.L., Schroder, R.F.W., Huettel, M.D., Sheppard, W.S., 1991. Mitochondrial DNA variation in geographic populations of Colorado potato beetle. *Leptinotarsa decemlineata* (Coleoptera; Chrysomelidae). *Experientia* 47, 483–485.
- Baungartner, D.L., Greenberg, B., 1984. The genus *Chrysomya* (Diptera: Calliphoridae) in the New World. *J. Med. Entomol.* 21, 105–113.
- Beard, C.B., Mills Hamm, D., Collins, F.H., 1994. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol. Biol.* 2, 103–124.
- Boore, J.L., 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27, 1767–1780.
- Brehm, A., Harris, D.J., Hernandez, M., Cabrera, V.M., Larruga, J.M., Pinto, F.M., Gonzalez, A.M., 2001. Structure and evolution of the mitochondrial DNA complete control region in the *Drosophila subobscura* subgroup. *Insect Mol. Biol.* 10, 573–578.
- Carrodeguas, J.A., Vallejo, C.G., 1997. Mitochondrial transcription initiation in the crustacean *Artemia franciscana*. *Eur. J. Biochem.* 250, 514–523.
- Caterino, M.S., Cho, S., Sperling, F.A.H., 2000. The current state of insect molecular systematics: a thriving Tower of Babel. *Annu. Rev. Entomol.* 45, 1–54.
- Chang, D.D., Clayton, D.A., 1984. Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. *Cell* 36, 635–643.

- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., Thompson, J.D., 2003. Multiple sequence alignment with the clustal series of programs. *Nucleic Acids Res.* 31, 3497–3500.
- Clary, D.O., Wolstenholme, D.R., 1987. *Drosophila* mitochondrial DNA: conserved sequences in the A+T-rich region and supporting evidence for a secondary model of the small ribosomal RNA. *J. Mol. Evol.* 25, 116–125.
- Clary, D.O., Goddard, J.M., Martin, S.C., Faaron, S.C., Wolstenholme, D.R., 1982. *Drosophila* mitochondrial DNA: a novel gene order. *Nucleic Acids Res.* 10, 6619–6637.
- Crozier, R.H., Crozier, Y.C., 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* 133, 97–117.
- Flook, P.K., Rowell, C.H.F., Gelissen, G., 1995. The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. *J. Mol. Evol.* 41, 928–941.
- Gordon, D., Abajian, C., Green, P., 1998. CONSED: a graphical tool for sequence finishing. *Genome Res.* 8, 195–202.
- Helfenbein, K.G., Brown, W.M., Boore, J.L., 2001. The complete mitochondrial genome of the articulate brachiopod *Terebratalia transversa*. *Mol. Biol. Evol.* 18, 1734–1744.
- Infante, M.E.V., Azeredo-Espin, A.M.L., 1995. Genetic variability in mitochondrial DNA of screwworm, *Cochliomyia hominivorax* (Diptera: Calliphoridae), from Brazil. *Biochem. Genet.* 33, 737–756.
- Jameson, D., Gibson, A.P., Hudelot, C., Higgs, P.G., 2003. OGRE: a relational database for comparative analysis of mitochondrial genomes. *Nucleic Acids Res.* 31, 202–206.
- Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17, 1244–1245.
- Laurence, B.R., 1988. The tropical African latrine blowfly. *Chrysomya putoria* (Wiedemann). *Med. Vet. Entomol.* 2, 285–291.
- Lessinger, A.C., Azeredo-Espin, A.M.L., 2000. Evolution and structural organisation of mitochondrial DNA control region of myiasis-causing flies. *Med. Vet. Entomol.* 14, 71–80.
- Lessinger, A.C., Junqueira, A.C.M., Lemos, T.A., Kemper, E.L., da Silva, F.R., Vettore, A.L., Arruda, P., Azeredo-Espin, A.M.L., 2000. The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Mol. Biol.* 9, 521–529.
- Lessinger, A.C., Junqueira, A.C.M., Conte, F.F., Azeredo-Espin, A.M.L., 2004. Analysis of a conserved duplicated tRNA gene in the mitochondrial genome of blowflies. *Gene* 339, 1–6.
- Lewis, D.L., Farr, C.L., Kaguni, L.S., 1995. *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. *Insect Mol. Biol.* 4, 263–278.
- Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequences. *Nucleic Acids Res.* 25, 955–964.
- Mitchell, S.E., Cockburn, A.F., Seawright, J.A., 1993. The mitochondrial genome of *Anopheles quadrimaculatus* species A. Complete nucleotide sequence and gene organization. *Genome* 36, 1058–1073.
- Nardi, F., Carapelli, A., Fanciulli, P.P., Dallai, R., Frati, F., 2001. The complete mitochondrial DNA sequence of the basal hexapod *Tetradontophora bislanensis*: evidence for heteroplasmy and tRNA translocations. *Mol. Biol. Evol.* 18, 1293–1304.
- Nardi, F., Carapelli, R., Dallai, R., Frati, F., 2003a. The mitochondrial genome of the olive fly *Bactrocera oleae*: two haplotypes from distant geographical locations. *Insect Mol. Biol.* 12, 605–611.
- Nardi, F., Spinsanti, G., Boore, J., Carapelli, A., Dallai, R., Frati, F., 2003b. Hexapod origins: monophyletic or paraphyletic? *Science* 299, 1887–1889.
- Ojala, D., Montoya, J., Attardi, G., 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290, 470–474.
- Perna, N.T., Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* 41, 353–359.
- Rozas, J., Rozas, R., 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15, 174–175.
- Saccone, C., de Giorgi, C., Gissi, C., Pesole, G., Reyes, A., 1999. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene* 238, 195–209.
- Shadel, G.S., Clayton, D.A., 1993. Mitochondrial transcription initiation: variation and conservation. *J. Biol. Chem.* 268, 16083–16086.
- Shao, R., Campbell, N.J., Schmitt, E.R., Barker, S.C., 2001. Increased rate of gene rearrangement in the mitochondrial genomes of three orders of hemipteroid insects. *Mol. Biol. Evol.* 18, 1828–1832.
- Souza, A.M., Linhares, A.X., 1997. Diptera and Coleoptera of potential forensic importance in southeastern Brazil: relative abundance and seasonality. *Med. Vet. Entomol.* 11, 8–12.
- Spanos, L., Koutroumbas, G., Kotsyfakis, M., Louis, C., 2000. The mitochondrial genome of the Mediterranean fruit fly, *Ceratitis capitata*. *Insect Mol. Biol.* 9, 139–144.
- Swofford, D.L., 2000. *Phylogenetic Analysis Using Parsimony and Other Methods (Software)*. Sinauer Associates, Sunderland, MA.
- Wells, J.D., 1991. *Chrysomya megacephala* (Diptera: Calliphoridae) has reached the continental United States: review of its biology, pest status, and spread around the world. *J. Med. Entomol.* 28, 471–473.
- Wolstenholme, D.R., 1992. Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol.* 141, 173–216.
- Yeates, D.K., Wiegmann, B.M., 1999. Congruency and controversy: toward a higher-level phylogeny of Diptera. *Annu. Rev. Entomol.* 44, 397–428.
- Zhang, D.-X., Hewitt, G.M., 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochem. Syst. Ecol.* 25, 99–120.

APÊNDICE C

The mitochondrial control region of blowflies (Diptera: Calliphoridae): a hot spot for mitochondrial genome rearrangements

Gustavo Turqueto Duarte, Ana Maria L. de Azeredo-Espin and Ana Carolina M. Junqueira

Journal of Medical Entomology 45 (2008): 667-676

RESUMO

A família Calliphoridae compreende moscas causadoras de miíases, incluindo espécies de importância econômica, médica e forense. Neste trabalho, a região controle completa do DNA mitocondrial de 15 espécies de califorídeos foi sequenciada e estruturalmente caracterizada. A região controle tem um alto conteúdo de Adeninas (A) e Timinas (T) e varia em tamanho de 854 pb a 2018 pb, mostrando variações intraespecíficas na sequência e tamanho. Foram identificados dois domínios principais: um domínio conservado contendo blocos de sequências conservadas e estruturas cis-regulatórias que podem estar relacionadas a transcrição e a origem de replicação do DNA mitocondrial, e um domínio variável, contendo alta variação de sequências e tamanho. Dentro do domínio variável, a duplicação do gene *tRNA^{leu}*, previamente descrita para três espécies de *Chrysomya*, foi identificada em duas outras espécies desse gênero e em duas espécies de dois outros gêneros. A caracterização estrutural mostra a plasticidade do genoma mitocondrial em Diptera. Neste trabalho, as similaridades da organização da região duplicada encontrada nas diferentes espécies e a possível origem dos genes duplicados são discutidas.

The Mitochondrial Control Region of Blowflies (Diptera: Calliphoridae): A Hot Spot for Mitochondrial Genome Rearrangements

GUSTAVO TURQUETO DUARTE, ANA MARIA L. DE AZEREDO-ESPIN,¹ AND ANA CAROLINA M. JUNQUEIRA

Laboratório de Genética Animal, Centro de Biologia Molecular e Engenharia Genética (CBMEG) and Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), CP 6010, 13083-875 Campinas, SP, Brazil

J. Med. Entomol. 45(4): 667-676 (2008)

ABSTRACT The family Calliphoridae consists of myiasis-causing flies, including species of economic, forensic, and medical importance. In this study, the complete control regions (CRs) of mitochondrial DNA from 15 calliphorid species were sequenced and structurally characterized. The CRs had a high content of adenines (A) and thymines (T) and varied in length from 854 to 2,018 bp, showing intraspecific variations in sequence and length. Two major domains were identified: the conserved domain containing conserved sequence blocks and *cis*-regulatory structures that may be related to the transcription and the origin of replication of mitochondrial DNA, and the variable domain, containing high sequence and length variation. Within the variable domain, duplication of the *tRNA^{Leu}* gene, previously reported for three *Chrysomya* species, was identified in two more species of this genus and in two species of two other genera. The structural characterization shows the plasticity of the mitochondrial genome in dipterans. The organizational similarities of the duplicated region found in different species and the possible origin of the duplicated genes are discussed.

KEY WORDS A+T-rich region, gene duplication, gene order, illegitimate recombination, Insecta

Metazoan mitochondrial DNA (mtDNA) has been used for evolutionary studies involving population structure (Lyra et al. 2005), species identification (Wells and Sperling 2001, Waugh 2007), and phylogenetics (Moore 1995, Waugh 2007). The simple genetic structure of mtDNA, its predominantly maternal inheritance, and its higher rate of nucleotide substitution compared with nuclear DNA (Avice et al. 1987, Moore 1995) are some of the features that make the mtDNA a valuable source of genotypic characters. The high copy number of mtDNA enhances the access to genetic information through polymerase chain reaction (PCR) amplification and is a useful feature for forensic analyses and studies with preserved specimens, because it minimizes the loss of DNA caused by the oxidative and hydrolytic damage that normally occurs after death (Pääbo et al. 1989, Junqueira et al. 2002).

Typical animal mtDNA is a circular, double-stranded molecule with 16–20 kb, generally containing 13 protein-coding genes, two subunits of ribosomal RNA (rRNA), and 22 transfer RNA (tRNA) genes. The product of these genes, together with nuclear proteins and RNAs, provides the machinery for mitochondrial replication, transcription, mRNA processing, and protein translation (Boore 1999). The major

noncoding region of mtDNA is known as the control region (CR). In insects, vertebrates, and crustaceans (Carrodeguas and Vallejo 1997, Shadel and Clayton 1997), the presence of conserved sequence blocks (Zhang and Hewitt 1997) and secondary structure (Inohira et al. 1997), as well as data from ligation-mediated PCR (Saito et al. 2005) and electron microscopy (Goddard and Wolstenholme 1980), indicate that the CR is related to the transcription and replication of the mitochondrial genome.

The CR generally shows two patterns of organization among insects. In fruit flies (Zhang and Hewitt 1997), blowflies (Lessinger and Azeredo-Espin 2000, Junqueira et al. 2004, Lessinger et al. 2004), and Muscidae flies (Oliveira et al. 2007), the CR has two distinct domains: a conserved domain with conserved sequence blocks (CSBs) that harbors the putative origin of replication, and a variable domain that varies markedly in sequence and length. Both of these domains may contain [A]_n[T]_n tandem repeats. In contrast, in butterflies, moths, stoneflies, grasshoppers, locusts, and mosquitoes, the CR cannot be divided into distinct domains, although it may contain tandem sequence repeats and CSBs dispersed throughout its extension (Schultheis et al. 2002, Vila and Björklund 2004).

The family Calliphoridae is well known for its economic, medical, sanitary, and forensic importance

¹ Corresponding author. e-mail: azeredo@unicamp.br.

Table 1. Sampling localities, fixation methods, and amplified fragments in the calliphorid species studied

Species	Locality	Collection date	Preservation	Accession no.	Reference
<i>Chrysomya albiceps</i> (Wiedemann) ^c	Brazil (Rio de Janeiro)	N/A	N/A	AY220131	Lessinger et al. (2004)
<i>Chrysomya bezziana</i> (Villeneuve) ^{ab}	Malaysia (Kluang)	Sep. 2003	Alcohol	EF423877	This work
<i>Chrysomya bezziana</i> ^b	Malaysia (Kluang)	Sep. 2003	Alcohol	EF423876	This work
<i>Chrysomya megacephala</i> (F.) ^b	Australia (Perth)	Dec. 2003	Alcohol	EF423850	This work
<i>Chrysomya megacephala</i> ^a	Brazil (São Paulo)	N/A	Frozen -70°C	AF151386	Lessinger and Azeredo-Espin (2000)
<i>Chrysomya putoria</i> (Wiedemann) ^c	Brazil (São Paulo)	N/A	Fresh	AF352790	Junqueira et al. (2004)
<i>Chrysomya rufifacies</i> (Macquart) ^{ab}	Australia (Perth)	Dec. 2003	Alcohol	EF423851	This work
<i>Chrysomya rufifacies</i> ^a	Australia (Perth)	Dec. 2003	Alcohol	EF423854	This work
<i>Chrysomya rufifacies</i> ^{ab}	USA (Texas)	Jun. 1993	Dried	EF423852	This work
<i>Chrysomya rufifacies</i> ^b	Australia (Perth)	Dec. 2003	Alcohol	EF423853	This work
<i>Calliphora dubia</i> (Macquart) ^{ab}	Australia (Perth)	Dec. 2003	Alcohol	EF423878	This work
<i>Calliphora vomitoria</i> (L.) ^{ab}	England (London)	Apr. 2004	Alcohol	EF423855	This work
<i>Chloroprocta idiotidea</i> (Robineau-Desvoidy) ^{ab}	Brazil (São Paulo)	Apr. 2004	Fresh	EF423879	This work
<i>Cochliomyia hominivorax</i> (Coquerel) ^c	Brazil (Minas Gerais)	N/A	Frozen -70°C	AF151382	Lessinger and Azeredo-Espin (2000)
<i>Cochliomyia macellaria</i> (F.) ^c	Brazil (Amazonas)	N/A	Frozen -70°C	AF151384	Lessinger and Azeredo-Espin (2000)
<i>Hemilucilia flavifacies</i> (Engel) ^{ab}	Brazil (R. Grande do Sul)	Jun. 1994	Alcohol	EF423856	This work
<i>Hemilucilia segmentaria</i> (F.) ^a	Brazil (São Paulo)	Apr. 2001	Frozen -20°C	EF423857	This work
<i>Hemilucilia semidiaphana</i> (Rondani) ^{ab}	Brazil (São Paulo)	Apr. 2002	Fresh	EF423858	This work
<i>Lucilia eximia</i> (Wiedemann) ^{ab}	Brazil (R. Grande do Sul)	Jun. 1994	Alcohol	EF423859	This work
<i>Lucilia sericata</i> (Meigen) ^{ab}	France (South)	Apr. 2004	Alcohol	EF423890	This work
<i>Phormia regina</i> (Meigen) ^{ab}	USA (West Virginia)	Apr. 2004	Alcohol	EF423892	This work
<i>Phormia regina</i> ^b	USA (California)	Jul. 1997	Alcohol	EF423891	This work

^a Amplicon A sequenced.

^b Amplicon B sequenced.

^c GenBank sample.

N/A, no data available in GenBank flat file.

(Baumgartner and Greenberg 1984, Amendt et al. 2004, Azeredo-Espin and Lessinger 2006). This family includes myiasis-causing flies that have a wide geographic distribution and synantropic habits. Myiasis are infestations on living vertebrates, caused by dipteran larvae that feed on the host's dead or living tissue, body substances, or ingested food (Zumpt 1965). Blowflies are among the first insects to colonize corpses, being important for forensic science (Amendt et al. 2004) also acting as pathogen vectors, whereas the New World screwworm fly, *Cochliomyia hominivorax*, is one of the most important livestock pests in South America, causing substantial economic losses (Hall and Wall 1995).

The aim of this study was the characterization of the structure and organization of the CR in seven calliphorid genera for the potential usefulness of this region as a molecular marker for evolutionary studies. In addition, we analyzed the CR's flanking sequences and its role as a hot spot for tRNA genes duplication, leading to mtDNA rearrangements. This report extends previous studies on the CR as a molecular marker and is the most comprehensive comparative study of myiasis-causing flies performed thus far.

Materials and Methods

Samples, DNA Extraction, and PCR Amplification. Adult flies of Calliphoridae were collected using traps and decomposed fish and liver as baits and stored at -70°C. Ethanol-preserved adults were provided by

collaborators. Table 1 summarizes the species, sampling locations, collection dates, preservation methods, and GenBank accession numbers of all specimens used in this work.

DNA from fresh and frozen flies was extracted using the phenol/chloroform method, as described by Infante Vargas and Azeredo-Espin (1995). Total DNA was extracted from dried/pinned or alcohol-preserved specimens using DNAzol (Invitrogen, Carlsbad, CA), because this reagent was the most efficient method for recovering DNA from preserved specimens, as described by Junqueira et al. (2002).

The CR was amplified using the nested PCR strategy with two sets of primers in two distinct reactions for amplicons A and B. An initial reaction was done with the primers C1-N-1560 and SR-J-14612 (Simon et al. 1994), and the products of these amplifications were used as templates for the nested PCR, in which the second reaction was carried out with the primers TM-N-193 (Simon et al. 1994) and CMegAR (Lessinger et al. 2004) to obtain the amplicon A, or with the primers CMegA (Lessinger et al. 2004) and SR-J-14941 (Oliveira et al. 2006) to obtain amplicon B. The reactions consisted of denaturation at 94°C for 3 min, followed by 35 cycles at 94°C for 1 min, 40°C for 1 min, and 60°C for 2 min. A final elongation step at 60°C for 10 min was added after the last cycle. Both PCR reactions were performed identically, except that the second reaction (nested PCR) had an annealing temperature of 45°C.

Table 2. Primers used for PCR amplification of the control region

Primer	Sequence	Source	Annealing region
CI-N-1560	5'-TGTTCTACTATTCCGGCTCA-3'	Simon et al. (1994)	<i>coxI</i>
SR-J-14612	5'-AGGGTATCTAATCTAGTTT-3'	Simon et al. (1994)	<i>srRNA</i>
TM-N-193	5'-TGGGGTATGAAACCCAGTAGC-3'	Simon et al. (1994)	<i>tRNA^{Met}</i>
CRc-J-BVII	5'-TTTTATATTGAAGCGTCG-3'	This work	CR conserved domain
CRc-N-BVII-R	5'-TTGCACGGTTCAATATAA-3'	This work	CR conserved domain
CMegAR	5'-AATCCAGTTAA GAATATCAT-3'	Lessinger et al. (2004)	CR conserved domain
CMegA	5'-ATGATATTCTAACTGGATT-3'	Lessinger et al. (2004)	CR conserved domain
CRv-J-Cadu	5'-TTGGTAAATCCTACTGT-3'	This work	CR variable domain
SR-J-14941	5'-AGCCAAAATAAACTTTA-3'	Oliveira et al. (2006)	<i>srRNA</i>

The amplicon A of *Chloroprocta idioidea* was amplified with the CRc-J-BVII primer. This primer, combined with TM-N-193, recovers a 130-bp overlapping sequence with amplicon B. The large size of the variable domain of *Calliphora dubia* required additional primer combination; therefore, CRv-J-Cadu was designed based on the partial sequence of amplicon B and used with CRc-N-BVII-R. Table 2 lists the primers used in this work.

The final reagent concentrations for all reactions were 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.25 U of *Taq* DNA polymerase (Invitrogen), 1.5 mM MgCl₂, 500 nM of each primer, and 200 μM of each dNTP. The amplicons were separated by electrophoresis on 1.0% agarose gels using TAE 1× buffer (40 mM Tris-acetate, pH 8.0, 1 mM EDTA), and the products were visualized under UV light in the presence of ethidium bromide.

Cloning and Sequencing. The PCR products were purified by dialysis using 0.05-μm Millipore filters and TE buffer (1 M Tris-HCl, pH 7.4, 0.5 M EDTA). The products were ligated into a pCR2.1-TOPO vector (Invitrogen TOPO-TA Cloning kit), according to the manufacturer's instructions, followed by chemical transformation in *E. coli* DH5α competent cells (Sambrook et al. 1989). Positive clones were sequenced automatically on a Perkin Elmer-ABI Prism 377 sequencer using the Big Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA), according to the manufacturer's instructions.

Sequence Analysis. The chromatograms were analyzed and edited with the Chromas 2.21 software (Technelysium Pty, Tewantin, QLD, Australia). Multiple alignments were performed using the ClustalW algorithm (Thompson et al. 1994) with gap penalties reduced to 12 because of differences in sequence length. The nucleotide diversity across the conserved domain was computed by the sliding window method with DnaSP 4.0 (Rozas et al. 2003), using seven sites for the window length and step size, without gaps. A putative stem-loop structure of the A+T-rich region and its folding free energy were inferred with the Mfold web server version 3.1 using default parameters (Zucker 2003). Repeats in the CR sequences were identified with the Dotlet web server (Junier and Pagni 2000).

Conserved sequence blocks were assigned based on the alignment of the 15 complete conserved domain

sequences of calliphorid species, according to Lessinger and Azeredo-Espin (2000) and the following criteria: (1) a CSB should contain at least 15 sites, (2) the start of a CSB was assigned based on the presence of similarity for at least four sites in 13 of the 15 species studied, and (3) the end of a CSB was considered after four variable consecutive sites in more than two species.

Results

The CR was flanked by *srRNA* and *tRNA^{Met}* genes (Fig. 1) and varied in length (excluding primers) from 854 bp in *Chrysomya megacephala* to 1,905 bp in *Ca. dubia* (Table 3). The nucleotide composition showed a high content of A and T that varied from 86.0% in *Ch. idioidea* to 91.6% in *Co. hominivorax* (Table 3). The alignment and sequence analysis evidenced the presence of two domains in CR, as described by Lessinger et al. (2004) and Junqueira et al. (2004). A conserved domain is located adjacent to the *tRNA^{Met}* gene, presenting conserved sequence blocks between the species of family Calliphoridae, as well as other conserved structures characterized for other insect groups. However, adjacent to the *srRNA* gene, there is a variable domain, with high sequence and length differences between the species.

Table 3. Length variation and nucleotide bias of the CRs in calliphorid species

Species	CR (bp)	Conserved domain (bp)	Variable domain (bp)	A+T (%)
<i>C. bezziiana</i>	855	574	281	90.5
<i>C. megacephala</i>	854	576	278	90.8
<i>C. putoria</i>	858	582	306	90.4
<i>C. ruffifacies</i> 1	988	562	426	90.4
<i>C. ruffifacies</i> 2	1,026	556	470	88.6
<i>C. albiceps</i>	1,008	542	466	91.3
<i>L. sericata</i>	1,103	565	538	90.3
<i>P. regina</i>	869	570	229	90.9
<i>Co. hominivorax</i>	1,158	619	539	91.6
<i>H. semidiaphana</i>	1,171	612	559	86.9
<i>H. flavifuscus</i>	1,180	615	565	86.6
<i>Co. macellaria</i>	1,206	617	589	87.1
<i>L. eximia</i>	1,247	568	679	91.0
<i>Ch. idioidea</i>	1,234	666	568	86.0
<i>Ca. vomitoria</i>	1,303	595	708	90.8
<i>Ca. dubia</i>	1,905	600	1,305	90.7

C. ruffifacies 1 and 2 correspond to specimens from Australia and the United States, respectively.

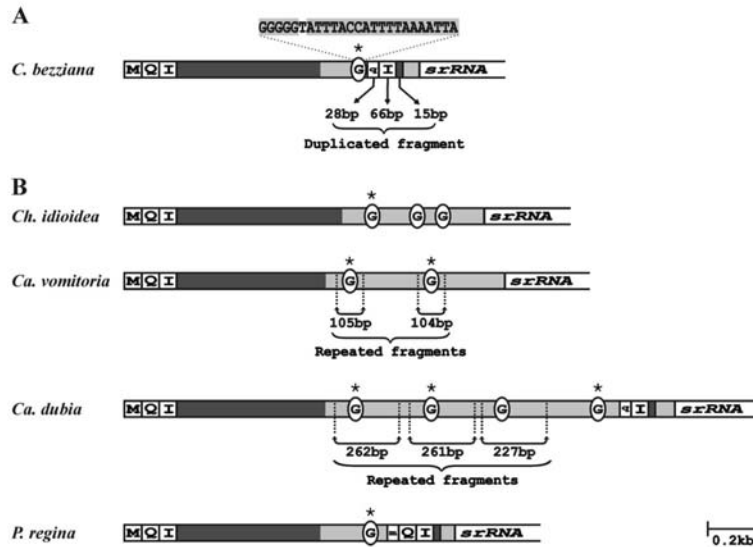


Fig. 1. (A) *Chrysomya*-like organization of the mtDNA CR showing gene duplication. M, Q, and I correspond to tRNAs methionine, glutamine, and isoleucine genes, respectively, whereas q and m denotes glutamine and methionine pseudogenes. Dark gray represents the conserved domain of CR, whereas light gray denotes the variable domain. The ellipse with a G and an asterisk indicates a conserved element of the variable domain, represented above. (B) Other patterns of CR organization showing tRNA duplication in *Ca. dubia* and *P. regina* and the repeated fragments in *Calliphora*. The ellipse with a G but no asterisk indicates G islands not associated with the 19-bp motif. In the last repeated fragment of *Ca. dubia*, unlike the previous two, the G island is not associated with the motif, which seems to be degenerated.

The conserved domain of blowflies analyzed in this work varied in length from 542 bp in *Chrysomya albiceps* to 666 bp in *Ch. idioidea* (Table 3), with A↔T substitutions accounting for 60.6% of the total substitutions. Transversions (tv) predominated over transitions (ts; ts/tv = 0.32), corroborating the suggestion of a mutation pressure to maintain the high content of these two nucleotides in the CR. The ts and tv analysis against the pairwise distance showed that transversions saturate as more distantly related sequences are

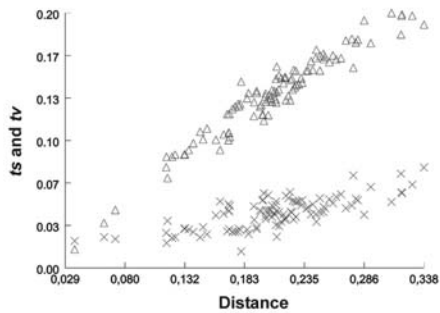


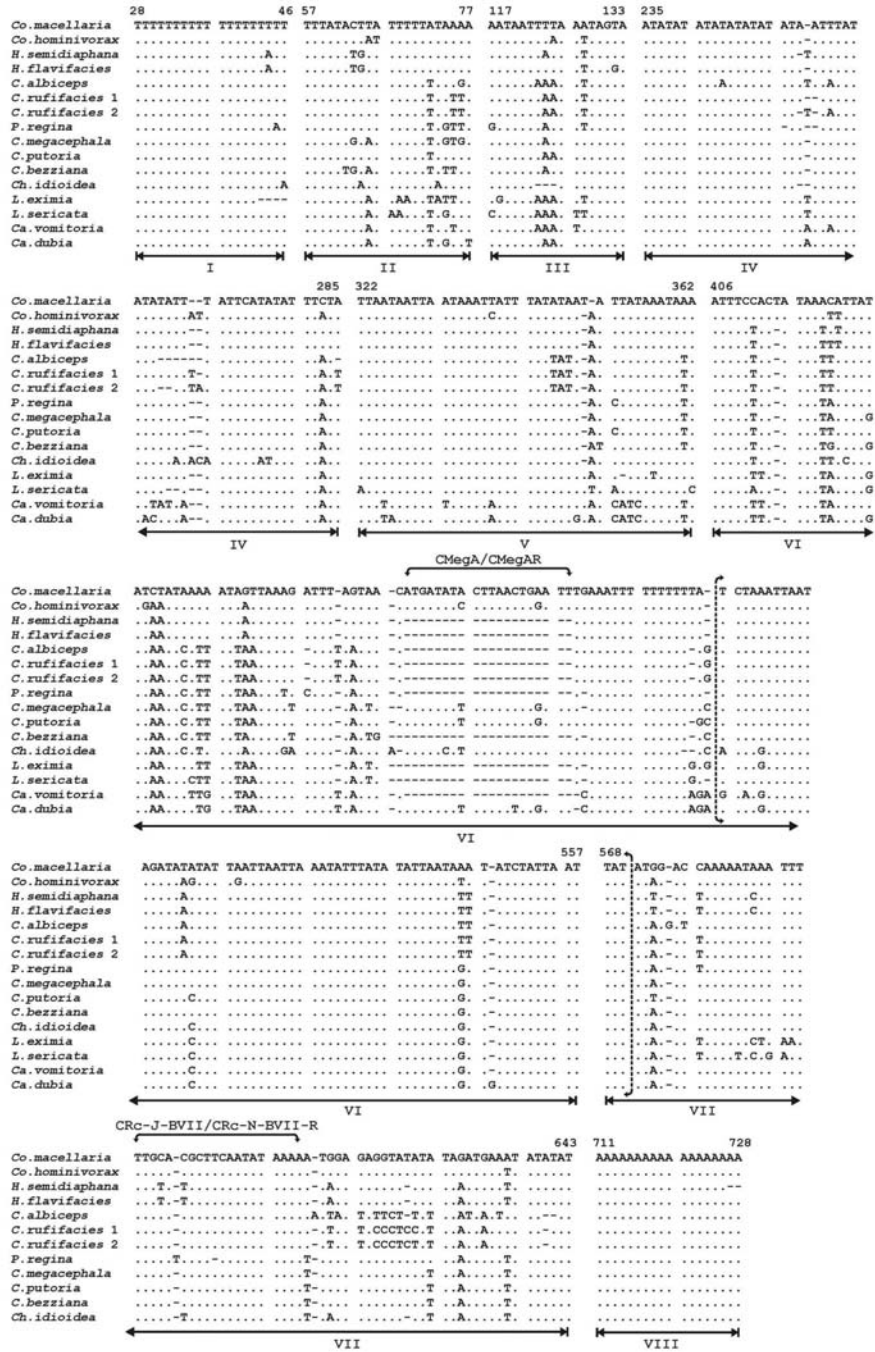
Fig. 2. *ts* and *tv* analysis against the pairwise distance, showing that transversions saturate as more distantly related sequences are compared. Triangles denotes *tv*, whereas crosses denotes *ts*.

compared, as shown in Fig. 2. The uncorrected p-distances of the conserved domain ranged from 2.2 to 13.9% for intrageneric comparisons, 9.1 to 19.1% between genera of the same subfamily, and 12.7 to 24.1% between subfamilies.

Despite the size variation, alignment of the conserved domain showed the presence of eight CSBs (Fig. 3) that corresponded to 395 sites. The criteria used to assemble the CSBs were supported by the estimate of nucleotide diversity (P_i). Although the conserved domain had a mean P_i of 0.16 (range, 0–0.54), the CSBs were located within regions of lower diversity with a mean P_i of 0.09 (range, 0–0.36).

The CSB I is a T-stretch, also present upstream of *tRNA^{Ile}* in *Drosophila* spp., Muscidae, Orthoptera, and Plecoptera (Clary and Wolstenholme 1987, Zhang et

Fig. 3. (Opposite page) Alignment of the conserved domain sequences. Dots indicate nucleotide identities with the sequence of *Co. macellaria*; letters indicate substitutions and dashes represent gaps. *C. rufifacies* one and two represent species from Australia and the United States, respectively. CSB I–CSB VIII are denoted below the sequences. The annealing sites for the primers CMegA/CMegAR and CRc-J-BVII/CRc-N-BVII-R are indicated above their respective sequences. The dashed brackets delimit the sequence of the putative stem-loop structure inferred with Mfold and associated with the origin of replication for the L-strand of mtDNA.



al. 1995, Brehm et al. 2001, Schultheis et al. 2002, Oliveira et al. 2007). This motif varied in size from 15 bp in *Lucilia eximia* and *Cochliomyia macellaria* to 23 bp in *Co. hominivorax*, presenting intraspecific size variation. The CSBs II and III were the shortest blocks (21 and 17 bp, respectively) and preceded the most variable region of the conserved domain that was located between CSBs III and IV; this variable region ranged in length from 24 bp in *C. albiceps* to 98 bp in *Ch. idioidea*, which made its alignment difficult. CSBs IV and V contained a purine-pyrimidine [TA(A)]_n tandem repeat that has also been reported between the T-stretch and the conserved secondary structure in Lepidoptera (Taylor et al. 1993), Orthoptera and Coleoptera (Zhang and Hewitt 1997), Plecoptera (Schultheis et al. 2002), and Muscidae (Oliveira et al. 2007). Like the A and T stretches, tandem repeats also showed inter- and intraspecific variation.

The CSB VI was the largest block, with 152 bp, containing the annealing sites for the CMegA/CMegAR primers. A conserved stem-loop structure that has been associated with the origin of replication for the major strand of mtDNA in Diptera, Orthoptera, Hymenoptera, Plecoptera, and Phthiraptera (Clary and Wolstenholme 1987, Zhang et al. 1995, Schultheis et al. 2002, Lessinger et al. 2004, Covacin et al. 2006, Oliveira et al. 2007) was also inferred for the Calliphoridae and was located at the end of CSB VI and beginning of CSB VII. The predicted structures for each species and their free energy are represented in Fig. 4. CSB VII contains the sites used to design the primers CRc-J-BVII and CRc-N-BVII-R. Located after a region of high variation in sequence and length, CSB VIII was an A-stretch that ranged in size from 16 bp in *Hemilucilia semidiaphana* and *L. eximia* to 26 bp in *Chrysomya rufifacies* and marked the end of the conserved domain.

However, the variable domain showed a high variation in sequence and length, ranging from 229 bp in *Phormia regina* to 1,305 bp in *Ca. dubia* (Table 3). An interesting feature of this hypervariable region was a conserved 19-bp motif, preceded by stretches of four to seven Gs (G islands), found in all calliphorid species analyzed (Fig. 1A). Additional "G islands" were identified in the variable domain, unrelated to the conserved motif, with their number and location varying among species (Fig. 1B).

The variable domain of *Ca. dubia* had a fragment that was found in triplicate, which explained the greater length of this domain (Fig. 1B). Despite the variation in fragment length (262, 261, and 227 bp), they presented a similarity >90%. Interestingly, the variable domain of *Ca. vomitoria* also had a repeated fragment of 104 and 103 bp (Fig. 1B) that shared 83.7% similarity. This fragment also showed high similarity with the first half of the repeated fragments of *Ca. dubia* (e.g., the first repeated fragment of *Ca. vomitoria* is 73.3% similar to the first half of the first repeated fragment of *Ca. dubia*).

Another peculiarity of the variable domain involved *P. regina*, *Ca. dubia*, and *Chrysomya* species. In *Ca. dubia* and in the five species of *Chrysomya* examined,

the variable domain was interrupted by a duplication of *tRNA^{Ile}*, whereas in *P. regina*, the duplication also included the *tRNA^{Gln}* gene (Fig. 1). In *Chrysomya* species and *Ca. dubia*, the entire duplication occurred immediately after the 19-bp motif and included regions flanking the *tRNA^{Ile}* gene, which corresponded to partial sequences of the *tRNA^{Gln}* and CR. The duplicated fragment in *P. regina* showed the same pattern but included the complete sequence of *tRNA^{Gln}* and *tRNA^{Ile}* and the partial sequence of the *tRNA^{Met}* and CR. The duplicated tRNA genes were identical to their "original" copies in all of the species. *C. megacephala* and *C. rufifacies*, which present substitutions in their *tRNA^{Ile}* compared with *Co. hominivorax*, also presented the same substitutions in their duplicated tRNAs. However, the only exception was *P. regina* from California, for which only the amplicon B could be sequenced. The duplicated *tRNA^{Ile}* of this sample presents a substitution compared with the "original" tRNA copy of the sample from West Virginia. Conserved sequences, as observed for the duplicated tRNAs, were not found for the flanking regions of the repeated fragments of any species. These regions may be degenerated, being observed differences even among samples from the same species (e.g., *C. megacephala* and *P. regina*).

Discussion

Composition of the Control Region. A high A+T content in the mitochondrial genome, especially evident in the CR, has been widely described for insects (Zhang and Hewitt 1997, Lessinger and Azeredo-Espin 2000, Lessinger et al. 2000, 2004, Mardulyn et al. 2003, Shao and Barker 2003, Vila and Björklund 2004, Covacin et al. 2006). Such nucleotide composition may be related to a role in energy-dependent processes during replication of the mitochondrial genome (Clary and Wolstenholme 1985). Hence, nucleotide substitutions may not occur randomly, suggesting the prevalence of directional mutation pressure for the A+T composition in insect mtDNA (Jermin et al. 1994).

The Conserved Domain. Among the eight CSBs identified, conserved structures reported for Diptera, Lepidoptera, Orthoptera, Coleoptera, and Plecoptera (Clary and Wolstenholme 1987, Taylor et al. 1993, Lewis et al. 1994, Zhang et al. 1995, Zhang and Hewitt 1997, Brehm et al. 2001, Schultheis et al. 2002, Saito et al. 2005, Oliveira et al. 2007) could also be identified in Calliphoridae. Short tandem repeats, A/T-stretches, and the regions of high variability between CSBs III and IV and CSBs VII and VIII are the main features responsible for length variation of the conserved domain. However, although the alignment of some regions of this domain was not trivial, the presence of conserved and variable sequences in the same marker may provide an interesting starting point for phylogenetic analysis of recently diverged taxa, identifying calliphorid species, and for population genetics studies.

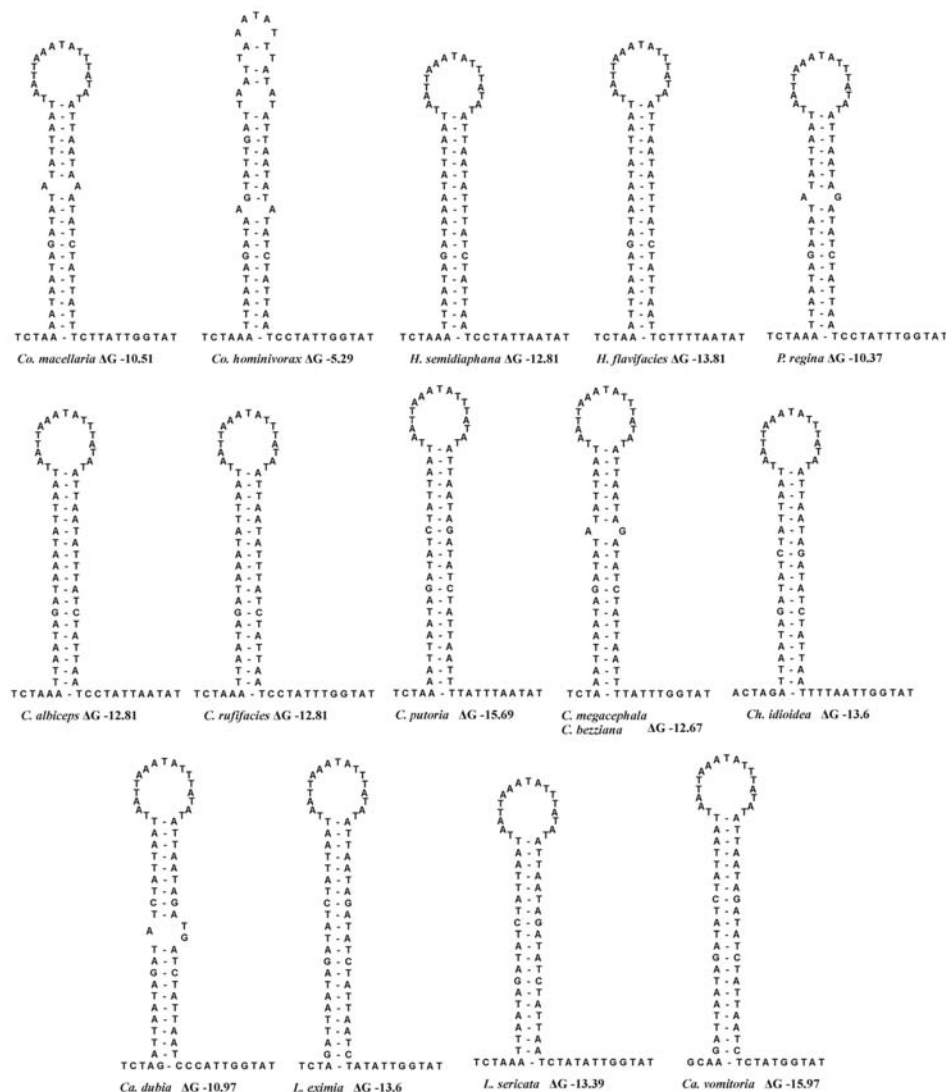


Fig. 4. Stem-loop structures and their free energy values (kcal/mol) predicted for the Calliphoridae species examined. Their relative positions are represented in Fig. 3.

The regulation of transcription of the mitochondrial genome through protein-DNA interactions was related to a T-stretch located adjacent to the *tRNA^{Ile}* in *Drosophila* spp. (Clary and Wolstenholme 1987); this structure was also described for Muscidae flies (Oliveira et al. 2007) and Orthoptera and Plecoptera (Zhang et al. 1995, Schultheis et al. 2002). Similarly, a T-stretch near the *tRNA^{Ile}* gene is present in the calliphorid species examined, located in the CSBI, and its

conservation among different groups is an evidence for such a functional role. T-stretches in the CR have also been related to the replication of the mitochondrial genome (Lewis et al. 1994). In a recent work, the origin of replication for the major coding strand of mtDNA was mapped downstream of the T-stretch adjacent to the *tRNA^{Ile}* gene in *Drosophila* spp. (Saito et al. 2005), whereas the origin or replication for the minor coding strand was mapped downstream to the

T-stretch located on the complementary. Such elements correspond, respectively, to CSBs I and VIII of the control region of family Calliphoridae, thus indicating that also in this family the CR might be related to the replication of the mtDNA, possibly being the origins of replication inside the limits of the conserved domain. Another conserved element that refers the CR to functional activities of the mitochondrial genome is the predicted secondary structures represented in Fig. 4. Saito et al. (2005) suggested that this structure might be involved in the replication initiation in Orthoptera. Similar structures have been identified in Orthoptera, Hymenoptera, Plecoptera, Phthiraptera, and other Diptera groups (Clary and Wolstenholme 1987, Zhang et al. 1995, Schultheis et al. 2002, Lessinger et al. 2004, Covacin et al. 2006, Oliveira et al. 2007), although their primary sequences are not conserved. Therefore, the data reinforce that, although there is sequence divergence, these structural elements and their locations seem to be conserved in Insecta, some of which relate the CRs to the replication and transcription of the mitochondrial DNA also in Calliphoridae. Additional functional evidence for the stem-loop structure was the occurrence of compensatory mutations, also identified in *C. albiceps* and *Chrysomya putoria* (Lessinger et al. 2004).

The Variable Domain. This domain was the most responsible not only for the variation in CR length among species, but also for differences in the mitochondrial genome size. Large repeated fragments, tandem repeats, and indels were the main cause of the variability of this domain and can be explained by slipped-strand mispairing (Moritz et al. 1987) during replication. Slipped-strand mispairing may also explain the duplicated fragments observed in the *Calliphora* species studied; however, analysis of additional species and samples of this genus is necessary to confirm the origin of these duplications. The G islands identified in Calliphoridae were also described for *Drosophila* and Muscidae species and have been associated with the termination of mtDNA replication (Brehm et al. 2001, Oliveira et al. 2007) but not with conserved motifs. Nevertheless, it seems that these G stretches are a feature of the CR of Diptera.

Although conserved elements were identified in this domain, the significant interspecific variation in size and sequence did not allow a confident alignment because of the lack of homology between nucleotide sites, therefore limiting the usefulness of this domain for phylogenetic studies. Such variation have been reported for *Drosophila* species and restricted its use to closely related species (Brehm et al. 2001). Nevertheless, the variable size of the domain and the presence of sequence polymorphisms could be a potential marker for populational studies. Indeed, PCR restriction fragment-length polymorphism analysis of this region has been used to identify screwworm species (Litjens et al. 2001), and a recent study with *Co. hominivorax* populations indicated that polymorphisms in the CR and srRNA region are useful for screening intraspecific variation and population structure (Lyra et al. 2005).

The Control Region as a Hot Spot for Mitochondrial Rearrangement. Duplication of the *tRNA^{Ile}* (Fig. 1A) has previously been reported for three *Chrysomya* species: *C. megacephala*, *C. putoria*, and *C. albiceps* (Junqueira et al. 2004, Lessinger et al. 2004). In this work, *tRNA^{Ile}* gene duplication was also found in *Chrysomya bezziana*, *C. rufifacies*, and *Ca. dubia*. A larger duplication involving the *tRNA^{Ile}* and *tRNA^{Gln}* genes was also found in *P. regina* (Fig. 1B). Because both copies of the duplicated tRNA genes are identical in these species, a mechanism for maintaining the homogeneity of the two copies, such as concerted evolution, may be active in both copies of the genes but not in the regions flanking the duplicated fragment, explaining the degeneration observed in these bound regions. The substitution observed in the duplicated *tRNA^{Ile}* of *P. regina* from California may have occurred during a recent event, so that homogenization may not have occurred yet, because the original copy does not possess such a substitution. The region between the duplicated genes and the *srRNA* gene, previously referred to as the *Chrysomya* intergenic region (Junqueira et al. 2004, Lessinger et al. 2004), was compared with the final region of the variable domain in the 15 calliphorid species studied, and we suggest here that this is a continuation of the variable domain, changing the current annotation of the CR variable domain. Molecular and morphological phylogenies (Rognes 1997, Wells and Sperling 2001, Harvey et al. 2003) indicate that the genera *Chrysomya* and *Calliphora* belong to the subfamilies Chrysomyinae and Calliphorinae, respectively. The subfamily Chrysomyinae also contains the genera *Chloroprocta*, *Cochliomyia*, *Hemilucilia*, and *Phormia*. However, of these genera, duplicated tRNAs were identified only in *Phormia*. This finding suggests that gene duplication and genomic rearrangement occurred independently on three occasions: in *Ca. dubia*, in *P. regina*, and also before the divergence of *Chrysomya*.

The gene arrangement observed in species with tRNA gene duplications was not explained by the current models for mitochondrial rearrangements: (1) tandem duplication followed by random loss caused by deletion that would lead to two adjacent copies of the duplicated fragment (Moritz et al. 1987); (2) homologous recombination that does not modify the original gene arrangement (Tsukamoto and Ikeda 1998, Rokas et al. 2003); and (3) intramolecular recombination that does not result in gene duplication (Lunt and Hymann 1997). Our results provide additional evidence for the illegitimate inter-mtDNA model previously reported in animals for the chigger mite *Leptotrobidium pallidum* (Shao et al. 2005). Because the gene duplication in *P. regina*, *Ca. dubia*, and the *Chrysomya* species occurred at the same genomic location, the variable domain of the CR seems to be a hot spot for gene rearrangements in mtDNA of Calliphoridae species. In addition, *Stomoxys calcitrans* sequences (Diptera: Muscidae) indicate that this species also has a *tRNA^{Ile}* duplicated into its CR (Oliveira et al. 2007). The finding that the duplicated fragments always involved the tRNA cluster formed by *tRNA^{Met}*,

tRNA^{Cln}, and *tRNA^{Ile}* probably reflects the hypothesis that, in all of the calliphorid species studied and in *Drosophila* species (Saito et al. 2005), the origin of replication for the major strand in the CR is located upstream of the *tRNA^{Ile}* gene, immediately adjacent to a T-stretch. According to the asymmetric replication model, the synthesis of the major coding strand begins after 97% of the minor coding strand has been synthesized, as occurs in *Drosophila* (Goddard and Wolstenholme 1980). This arrangement leaves the minor strand (the template for major strand synthesis) single-stranded for a lengthy period. Considering that transfer RNAs are structures known to form stem-loop (Macey et al. 1997), if, during the replication of the minor strand, a stem-loop structure is formed by any of the tRNA genes in the major strand, the synthesis of the new strand may be interrupted, and the short fragment is cleaved giving rise to a fragment that could be inserted in a hot spot region.

These results imply a structural plasticity and diversity in the mtDNA molecule of the family Calliphoridae and should contribute to further understanding of insect mitochondrial genome evolution.

Acknowledgments

We thank R. A. Rodrigues for technical assistance; A. X. Linhares for taxonomic identifications; B. Greenberg, M. L. Harvey, C. Ames, R. Mahon, B. Chauvet, É.E.S. Vianna, and J. D. Wells for providing specimens; and the anonymous referee for critical review of the manuscript. This work was supported by CNPq (Grant 133982/2005-0 to G.T.D. and Grant 450818/2004-5 to A.M.L.A.E.) and FAPESP (Grant 03/05444-2 to A.C.M.J. and Grant 05/57690-7 to A.M.L.A.E.).

References Cited

- Amendt, J., R. Krettek, and R. Zehner. 2004. Forensic entomology. *Naturwissenschaften* 91: 51–65.
- Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reece, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* 18: 489–522.
- Azeredo-Espin, A.M.L., and A. C. Lessinger. 2006. Genetic approaches for studying myiasis-causing flies: molecular markers and mitochondrial genomics. *Genetica* 126: 111–131.
- Baumgartner, D. L., and B. Greenberg. 1984. The genus *Chrysomya* (Diptera: Calliphoridae) in the New World. *J. Med. Entomol.* 21: 105–113.
- Boore, J. L. 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27: 1767–1780.
- Brehm, A., D. J. Harris, M. Hernández, V. M. Cabrera, J. M. Larruga, F. M. Pinto, and A. M. González. 2001. Structure and evolution of the mitochondrial DNA complete control region in the *Drosophila subobscura* subgroup. *Insect Mol. Biol.* 10: 573–578.
- Carrrodegua, J. A., and C. G. Vallejo. 1997. Mitochondrial transcription initiation in the crustacean *Artemia franciscana*. *Eur. J. Biochem.* 250: 514–523.
- Clary, D. O., and D. R. Wolstenholme. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. *J. Mol. Evol.* 22: 252–271.
- Clary, D. O., and D. R. Wolstenholme. 1987. *Drosophila* mitochondrial DNA: conserved sequences in the A + T-rich region and supporting evidence for a secondary structure model of the small RNA. *J. Mol. Evol.* 25: 116–125.
- Covacin, C., R. Shao, S. Cameron, and S. C. Barker. 2006. Extraordinary number of gene rearrangements in the mitochondrial genomes of lice (Phthiraptera: Insecta). *Insect Mol. Biol.* 15: 63–68.
- Goddard, J. M., and D. R. Wolstenholme. 1980. Origin and direction of replication in mitochondrial DNA molecules from the genus *Drosophila*. *Nucleic Acids Res.* 25: 741–757.
- Hall, M., and R. Wall. 1995. Myiasis of human and domestic animals. *Adv. Parasitol.* 35: 257–334.
- Harvey, M. L., M. W. Mansell, M. H. Villet, and R. Dadour. 2003. Molecular identification of some forensically important blowflies of southern Africa and Australia. *Med. Vet. Entomol.* 17: 363–369.
- Infante Vargas, M. E., and A.M.L. Azeredo-Espin. 1995. Genetic variability in mitochondrial DNA of screwworm, *Cochliomyia hominivorax* (Diptera: Calliphoridae), from Brazil. *Biochem. Genet.* 33: 237–256.
- Inohira, K., T. Hara, and E. T. Matsuura. 1997. Nucleotide sequence divergence in the A+T-rich region of mitochondrial DNA in *Drosophila simulans* and *Drosophila mauritiana*. *Mol. Biol. Evol.* 14: 814–822.
- Jermin, L. S., D. Graur, R. M. Lowe, and R. H. Crozier. 1994. Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome b genes. *J. Mol. Evol.* 39: 190–173.
- Junier, T., and M. Pagni. 2000. Dotlet: diagonal plotting in a Web browser. *Bioinfo. Appl. Note.* 16: 175–179.
- Junqueira, A.C.M., A. C. Lessinger, and A.M.L. Azeredo-Espin. 2002. Methods for recovery of mtDNA sequence from museum specimens of myiasis-causing flies. *Med. Vet. Entomol.* 16: 39–45.
- Junqueira, A.C.M., A. C. Lessinger, T. T. Torres, F. R. da Silva, A. L. Vettore, P. Arruda, and A.M.L. Azeredo-Espin. 2004. The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). *Gene* 339: 7–15.
- Lessinger, A. C., and A.M.L. Azeredo-Espin. 2000. Evolution and structural organization of mitochondrial DNA control region of myiasis-causing flies. *Med. Vet. Entomol.* 14: 1–10.
- Lessinger, A. C., A.C.M. Junqueira, T. A. Lemos, E. L. Kemper, F. R. da Silva, A. L. Vettore, P. Arruda, A.M.L. and Azeredo-Espin. 2000. The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Mol. Biol.* 9: 521–529.
- Lessinger, A. C., A.C.M. Junqueira, F. F. Conte, and A.M.L. Azeredo-Espin. 2004. Analysis of a conserved duplicated tRNA gene in the mitochondrial genome of blowflies. *Gene* 339: 1–6.
- Lewis, D. L., C. L. Farr, A. L. Farquhar, and L. S. Kaguni. 1994. Sequence, organization, and evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA. *Mol. Biol. Evol.* 11: 523–538.
- Litjens, P., A. C. Lessinger, and A.M.L. Azeredo-Espin. 2001. Characterization of the screwworm flies *Cochliomyia hominivorax* and *Cochliomyia macellaria* by PCR-RFLP of mitochondrial DNA. *Med. Vet. Entomol.* 15: 183–185.
- Lunt, D. H., and B. C. Hymann. 1997. Animal mitochondrial DNA recombination. *Nature (Lond.)* 387: 247.
- Lyra, M. L., P. Fresia, S. Gama, J. Cristina, L. B. Klaczko, and A.M.L. Azeredo-Espin. 2005. Analysis of mitochondrial DNA variability and genetic structure in populations of

- New World screwworm flies (Diptera: Calliphoridae) from Uruguay. *J. Med. Entomol.* 42: 589–595.
- Macey, J. R., A. Larson, N. B. Ananjeva, Z. Fang, and T. J. Papenfuss. 1997. Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Mol. Biol. Evol.* 14: 91–104.
- Mardulyn, P., A. Termonia, and M. C. Milinkovitch. 2003. Structure and evolution of the mitochondrial control region of leaf beetles (Coleoptera: Chrysomelidae): a hierarchical analysis of nucleotide sequence variation. *J. Mol. Evol.* 56: 38–45.
- Moore, W. S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49: 718–726.
- Moritz, C., T. E. Dowling, and W. M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst.* 18: 269–292.
- Oliveira, M. T., A. C. da Rosa, A.M.L. Azeredo-Espin, and A. C. Lessinger. 2006. Improving access to the control region and tRNA gene clusters of dipteran mitochondrial DNA. *J. Med. Entomol.* 43: 636–639.
- Oliveira, M. T., A.M.L. Azeredo-Espin, and A. C. Lessinger. 2007. The mitochondrial DNA control region of Muscidae flies: evolution and structural conservation in a dipteran context. *J. Mol. Evol.* 64: 519–527.
- Pääbo, S., R. G. Higuchi, and A. C. Wilson. 1989. Ancient DNA and the polymerase chain reaction. *J. Biol. Chem.* 264: 9709–9712.
- Rognes, K. 1997. The Calliphoridae (blowflies) (Diptera: Oestroidea) are not a monophyletic group. *Cladistics* 13: 27–68.
- Rokas, A., E. Ladoukakis, and E. Zouros. 2003. Animal mitochondrial DNA recombination revisited. *Trends Ecol. Evol.* 18: 411–417.
- Rozas, J., J. C. Sánchez-DelBarrio, X. Messegyer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496–2497.
- Saito, S., K. Tamura, and T. Aotsuka. 2005. Replication origin of mitochondrial DNA in insects. *Genetics* 171: 1685–1705.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Preparation and transformation of competent *E. coli* pp. 1.52–1.84. In J. Sambrook, E. F. Fritsch, and T. Maniatis (eds.), *Molecular cloning—a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory, NY.
- Schultheis, A. S., L. A. Weigt, and A. C. Hendricks. 2002. Arrangement and structural conservation on the mitochondrial control region of two species of Plecoptera: utility of tandem repeat-containing regions in studies of population genetics and evolutionary history. *Insect Mol. Biol.* 11: 605–610.
- Shadel, C. S., and D. A. Clayton. 1997. Mitochondrial DNA maintenance in vertebrates. *Annu. Rev. Biochem.* 66: 409–435.
- Shao, R., and S. C. Barker. 2003. The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. *Mol. Biol. Evol.* 20: 362–370.
- Shao, R., H. Mitani, S. C. Barker, M. Takahashi, and M. Fukunaga. 2005. Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the chigger mite, *Leptotrobidium pallidum*. *J. Mol. Evol.* 60: 764–773.
- Simon, C., F. Frati, A. Bechenbach, B. Crespi, H. Liu, H., and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* 87: 651–701.
- Taylor, M.F.J., S. W. McKechnie, N. Pierce, and M. Kreitman. 1993. The lepidopteran control region: structure and evolution. *Mol. Biol. Evol.* 10: 1259–1272.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673–4680.
- Tsukamoto, Y., and H. Ikeda. 1998. Double-strand break repair mediated by DNA end-joining. *Genes Cells* 3: 135–144.
- Vila, M., and M. Björklund. 2004. The utility of the neglected mitochondrial control region for evolutionary studies in Lepidoptera (Insecta). *J. Mol. Evol.* 58: 280–290.
- Waugh, J. 2007. DNA barcoding in animal species: progress, potential, and pitfalls. *BioEssays* 29: 188–197.
- Wells, J. D., and F.A.H. Sperling. 2001. DNA-based identification of forensically important Chrysomyinae (Diptera: Calliphoridae). *Forensic Sci. Int.* 120: 110–115.
- Zhang, D.-X., and G. M. Hewitt. 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochem. Syst. Ecol.* 25: 99–120.
- Zhang, D.-X., J. M. Szymura, and G. M. Hewitt. 1995. Evolution, and structural conservation of the control region of insect mitochondrial DNA. *J. Mol. Evol.* 40: 382–391.
- Zucker, 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*. Butterworths, London, United Kingdom.

Received 17 January 2008; accepted 19 March 2008.

APÊNDICE D

Evaluation of the internal transcribed spacer 2 (ITS2) as a molecular marker for phylogenetic inference using sequence and secondary structure information: an example for the myiasis-causing flies of the Calliphoridae family (Diptera: Brachycera).

Marco Antônio Tônus Marinho, Ana Carolina Martins Junqueira and Ana Maria Lima de Azeredo-Espin

Artigo submetido à Molecular Phylogenetics and Evolution

RESUMO

O ITS2 é uma região não codificadora localizada no “cluster” de DNA ribossomal, entre os genes 5.8S e 28S. A variabilidade encontrada na sequência do ITS2 é indicada para diferenciar espécies e para a reconstrução filogenética em *taxa* com divergência recente, enquanto sua estrutura secundária, conservada em Eukaryota, permite utilizar este marcador em grupos de divergência antiga. O potencial do ITS2 como um marcador molecular para análises filogenéticas da família Calliphoridae (Diptera: Brachycera) foi avaliado utilizando neighbor-joining, máxima parcimônia, máxima verossimilhança e inferência bayesiana, incluindo a informação da sequência primária e da estrutura secundária. As análises com a sequência primária revelaram um viés na composição nucleotídica ($A+T>75\%$) e na taxa ts/tv (<0.5), que são mais evidentes em hélices do que em regiões simples-fita, sugerindo a ocorrência de uma pressão seletiva. As mudanças compensatórias de bases foram analisadas e mostraram potencial para o diagnóstico espécie-específico. As comparações das topologias utilizando métodos de máxima verossimilhança revelaram que os métodos de inferência que consideram a informação da estrutura secundária apresentaram melhores resultados. As relações filogenéticas dentro de Calliphoridae corroboram estudos moleculares prévios. A relação entre Calliphoridae e *Mesembrinella* não foi resolvida, sendo a hipótese de que este gênero pode formar uma família separada (Mesembrinellidae) não foi rejeitada. Os padrões evolutivos do ITS2 descritos neste trabalho podem ser aplicados para a caracterização deste marcador para inferências filogenéticas na ordem Diptera e em grupos de insetos.

Evaluation of the internal transcribed spacer 2 (ITS2) as a molecular marker for phylogenetic inference using sequence and secondary structure information: an example for the myiasis-causing flies (Diptera: Calliphoridae).

Research Article

Marinho, M. A. T.¹, Junqueira, A. C. M.¹ & Azeredo-Espin, A. M L^{1,2}.

¹. Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), CP 6010, 13083-875, Campinas, SP, Brazil

². Departamento de Genética e Evolução (DGE), Instituto de Biologia (IB), Universidade Estadual de Campinas (UNICAMP), CP 6010, 13083-875, Campinas, SP, Brazil.

Corresponding author:

Marco Antonio Tonus Marinho

Laboratório de Genética Animal / Centro de Biologia Molecular e Engenharia Genética – CBMEG / Universidade Estadual de Campinas - UNICAMP

CP: 6010 CEP: 13083-875

Campinas, São Paulo, Brazil

TEL: +55 19 37881141

E-mail: mmarinho@unicamp.br

Keywords: ITS2. rRNA. secondary structure. Calliphoridae. myiasis.

Abstract

The internal transcribed spacer 2 is a small non-coding region located inside the ribosomal DNA cluster between 5.8S and 28S genes. ITS2 sequence variability is appropriate to differentiate species and for phylogenetic reconstruction in low order *taxa*, whereas its secondary structure, conserved among Eukaryota, could allow phylogenetic comparison in high order groups. We evaluated the potential of ITS2 as a molecular marker for phylogenetic analyses in the Calliphoridae family (Diptera: Brachycera) using neighbor-joining, maximum-parsimony, maximum-likelihood and bayesian inference methods including information of primary sequence alone and mixed with gaps and secondary structure information. ITS2 sequence analyses revealed nucleotide composition ($A+T>75\%$) and transition/transversion ($ts/tv<0.5$) biases. Comparisons between tree topologies using likelihood-based tests reveal that statistical methods of phylogenetic inference with models that account for secondary structure information perform better than other methods. Gap consideration reduces the likelihood, but had different effects on each phylogenetic method. Implications for the genetic relationships of the blowflies are discussed, including the status of the *Mesembrinella* group, for which no previous molecular phylogenetic data is available. Results described here could be applied to the characterization of this molecular marker for phylogenetic inference in other Diptera and Insecta groups.

1. Introduction

Molecular phylogenetics is a science of choices. Even after one decides which molecular marker use and which *taxa* study, there are still choices to be made. Which algorithm performs the best alignment with minimal need of manual correction? Which substitution model has the best fit to the data and corrects the evolutionary events that are not observable? Which tree reconstruction method will recover the correct tree topology, branch lengths and parameters of the evolutionary process for the data with higher confidence?

Much has been discussed in the last decades about the multiple steps of phylogenetic inference, from alignment (e.g., Thompson et al., 1999; Hickson et al., 2000; Lutzoni et al., 2000) to model selection (Sullivan and Joyce, 2005; Kelchner and Thomas, 2006) and the basis of phylogenetic inference (see Holder and Lewis, 2003, and Felsenstein, 2004, for a review) but the complexity of molecular phylogenetics as a science keeps discussion far from an end.

The evolutionary process has already occurred, i.e., species have diverged from common ancestors in a pattern that, as we believe, assumes a tree shape, and this process is unique. The objective of phylogenetics is to recover this pattern, but how to be confident that the true evolutionary history of a group has been recovered if sometimes different methods give us different answers? It is well known that a tree reconstructed from a nucleotide or protein sequence does not necessarily correspond to the evolutionary history of the group and that, supposing that all homologous sites have been perfectly aligned, the same sequences might provide different topologies with different phylogenetic inference methods, since they are based on different assumptions and try to optimize distinct functions.

Our goal in this study is not to give a definitive answer to questions mentioned above, neither to define which tree reconstruction method is the best. Our objective is to characterize the internal transcribed spacer 2 (ITS2) as a molecular marker for phylogenetic inference, considering some of the multiple approaches possible in the phylogenetic context, using a particular species group comprising the myiasis-causing flies of the Calliphoridae family.

The ITS2 is a small non-coding sequence located in the nuclear ribosomal cluster between the 3' end of 5.8S ribosomal DNA (rDNA) and 5' end of 28S rDNA. It is transcribed into a larger precursor RNA molecule, containing the three ribosomal RNA subunits 18S, 5.8S and 26S / 28S, two internal spacers, ITS1 and ITS2, and two

external transcribed spacers (ETS), 5'-ETS and 3'-ETS (Veldman et al., 1981; Hillis and Dixon, 1991). Although it is not present in any of the mature ribosomal RNA (rRNA) molecules, being digested during the primary transcript processing, the ITS2 has an essential function in the maturation process since it drives the cleavage of the great RNA precursor and, therefore, delimitates the 3' and 5' end of 5.8S and 28S rRNAs, respectively (Mitchell et al., 1996, 1997; Geerlings et al., 2000; Côté et al., 2002).

Yeast models suggest that recognition of cleavage sites inside ITS2 is related with its secondary structure acquired during transcription, maybe because of the structure itself and/or for bringing together sequences that are recognized by enzymatic complexes (van der Sande et al., 1992; Côté et al., 2002). This biological constraint causes ITS2 secondary structure to be conserved among all eukaryotic *taxa*, from yeast to mammals, and allows the comparison of ITS2 between high order *taxa*, such as families and beyond, since the structure is maintained, despite sequence variation (Joseph et al., 1999; Coleman, 2003; Coleman, 2007).

The current model for ITS2 secondary structure was first described by Joseph et al. (1999), based on computational modeling in yeasts and vertebrates, and includes four stem-loop (helix) regions inserted in a central ring. These stem-loop regions can be ramified or subdivided, as stated by many works with ITS2 (e.g., the subdivided regions II and IV of *Drosophila melanogaster* and the putative “Y” shaped region IV for *Musca domestica*, both described by Young and Coleman, 2004). In this article, we use this four-domains model, which is adopted in the ITS2 secondary structure database (Schultz et al., 2006; Selig et al., 2008).

According to some authors, the conservation of ITS2 secondary structure, allied with evidences that its sequence variation is correlated with sexual incompatibility and therefore allowing species differentiation (Coleman and Vacquier, 2002; Müller et al., 2007), could make ITS2 one of the most important and widespread molecular markers used for phylogenetics in a broad range of taxonomic levels (Joseph et al., 1999. Coleman, 2003; Schultz et al., 2005; Schultz et al., 2006; Coleman, 2007; Selig et al., 2007). Perspectives for the use of ITS2 as a molecular marker for phylogenetic inference are vast and growing, especially because of its particular features and the constant development of new methods of analysis. Its short extension and high-copy number (theoretically homogenized by concerted evolution – Elder and Turner, 1995; Liao, 1999) make easy the recovery and sequencing of ITS2.

In addition, the implementation of substitution models for RNA molecules that consider evolutionary dependence among sites (Savill et al., 2001; Gowri-Shankar and Jow, 2006) encourage deeper studies with ITS2 using more than the primary sequence information only.

The ITS2 has already been characterized as a molecular marker for species-specific diagnostic and phylogenetic inference in Diptera (Severini et al., 1996; Marinucci et al., 1999; Young and Coleman, 2004; Kampen, 2005; Li and Wilkerson, 2005; Gomulski et al., 2006; Song et al., 2008a, 2008b) and other Arthropoda orders (Gómez-Zurita et al., 2000; Gallego and Galián, 2001; Malloch et al., 2001; de Rojas et al., 2002; Shaw et al., 2002; Becerra, 2004; Dumont et al., 2005) but many of these works did not deal with secondary structure information. The need for a deeper characterization of the processes underlying ITS2 sequence and secondary structure evolution in the Diptera group led to the evaluation of this region as a molecular marker for phylogenetic inference in Calliphoridae.

The blowflies of the Calliphoridae family (Diptera: Brachycera) are one of the most important groups of myiasis-causing flies, comprising at least 80 species recorded as agents of myiasis, which is defined as the infestation of live human and vertebrate animals with dipterous larvae, that at least for a period, feed on the host's dead or living tissue, liquid body substances, or ingested food (Zumpt, 1965). In a parasitological context, Hall and Wall (1995) divided myiasis-causing species in three main groups: obligatory parasites, which need a live host to develop; primary facultative parasites, which can develop on both living and dead organic matter and are able to initiate myiasis; and secondary facultative parasites, which can also develop in living and dead organic matter but cannot initiate myiasis by themselves

Facultative parasites of the genera *Calliphora*, *Chrysomya*, *Cochliomyia*, *Lucilia* and *Phormia* are among the most common insects found in dead bodies and carcasses, and many of them have summary importance in forensic entomology (Amendt et al., 2004). Additionally, obligatory and facultative parasites of the genera *Chrysomya*, *Cochliomyia* and *Lucilia* are often associated with cattle raising activities and the infestation of these species in live animals has great economic impact (Guimarães et al., 1983; de Azeredo-Espin and Lessinger, 2006).

Studies with Calliphoridae species are mainly focused in: (I) forensic entomology, with emphasis in species diagnosis and influence of abiotic components and toxic substances in the development of eggs and larvae; (II) population structure

and variability, especially for pest control and eradication projects; (III) phylogenetic inference; and (IV) morphological and ethological studies to describe population dynamics of species (reviewed by de Azeredo-Espin and Lessinger, 2006; Stevens and Wallman, 2006; Stevens et al., 2006).

Until now, the mitochondrial DNA (mtDNA) has been the most characterized molecular marker for genetic and evolutionary studies of Calliphoridae family. In addition, the nuclear 28S ribosomal subunit also has been used for molecular phylogenetics by Stevens and Wall (2001) and Stevens (2003). In this work, we evaluated the ITS2 as a molecular marker for recovering genetic relationships among blowflies and, in a large extension view, the use of ITS2 as a molecular marker for phylogenetic inference based on different approaches using both sequence and secondary structure information in Calliphoridae. The results achieved here will contribute to a deeper understanding of ITS2 evolution in the Arthropoda group and, in a broader perspective, to its evaluation as a universal molecular marker for phylogenetic inference, especially at the species level.

2. Materials and Methods

2.1. Specimens and DNA extraction

Frozen specimens of Calliphoridae, Muscidae and Oestridae families (Diptera: Brachycera) were obtained from the collection maintained at -70°C in the Laboratório de Genética Animal (CBMEG/UNICAMP). Pinned and ethanol-preserved specimens were sent from museums and entomological collections. The DNA of frozen specimens was extracted using phenol/chloroform protocol (Infante and Azeredo-Espin, 1995) whereas the DNA of dried and ethanol preserved specimens was extracted using DNAzol[®] (Invitrogen) reagent, as described by Junqueira et al. (2002). Sequences of Calliphoridae species available on GenBank were included in the analyses. Table 1 shows a complete list of species analyzed in this study.

2.2. PCR amplification, DNA cloning and sequencing

PCR reactions were conducted with 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2 mM MgCl₂, 0.2 mM dNTPs, 0.5 μM 5.8S primer (5'-ATCACTCGGCTCGTGGGAT TCGAT-3'), 0.5 μM 28S primer (5'-GTTAGTTTC TTTTCCTCCCCT-3'), 1.25 U *Taq* DNA polymerase (Invitrogen[®]) and 1 to 2 μg of extracted DNA for a 25 μL reaction. PCR reactions were performed with an initial denaturation step of 94°C for 3 minutes, followed by 35 cycles of 94°C for 1 minute, 55°C for 45 seconds and 72°C

for 2 minutes, and a final elongation step of 72°C for 3 minutes. PCR products were visualized by electrophoresis in TAE (40mM Tris-acetate, 1mM EDTA) agarose gels 1.5%, stained with ethidium bromide.

PCR products were purified using the QIAquick PCR purification kit (QIAGEN[®]) and then cloned into a pCR[®]2.1 vector using the TA-Cloning Kit (Invitrogen[®]), according to the supplier's protocols. Competent DH10B *E. coli* cells were transformed using the CaCl₂ protocol (Sambrook et al., 1989) and then plated on Luria-Bertani (LB) medium with 50mg/mL X-Gal and 50µg/mL Ampicillin. The plates were grown at 37°C overnight. Plasmids were extracted using alkaline lyses protocol (Sambrook et al., 1989) and then digested with 5U *Eco* RI restriction enzyme at 37°C for 2 hours. The digestions were visualized by electrophoresis in TAE agarose gels 1.0%, stained with ethidium bromide and photographed under UV-illumination (300nm) with EDAS digital system (Kodak[®]). At least three clones were submitted to sequencing reactions.

Cycle sequencing reactions were conducted in an ABI 3700 (Perkin-Elmer) using the “Big Dye[™] Terminator Cycle Sequencing Ready Reaction Kit” (Prism[™], Perkin-Elmer), according to the supplier's protocol, with the universal primers M13 Forward (5'-GTAAAACGACGGCCAG-3') and M13 Reverse (5'-CAGGAAACAGCTATGAC-3') and 200-600 ng of DNA.

2.3. Sequence and secondary structure analyses

Chromatograms were visualized and edited using the softwares Chromas 2.21 (Technelysium Pty Ltd) and FinchTV 1.3.1 (Geospiza, Inc). All sequences were aligned using the program Clustal X 1.83 (Thompson et al., 1997) and final adjustments were made manually, according to secondary structure information, using the 4SALE software (Seibel et al., 2006). This software was also used for compensatory base changes (CBCs) visualization.

Initial sequence analyses, including nucleotide composition, transition and transversion rates (ts/tv) and genetic distances, were carried out using MEGA 4 (Tamura et al., 2007) and DAMBE (Xia and Xie, 2001) softwares.

ITS2 secondary structure was predicted using the mfold 2.3 webserver (Zuker, 2003) with default parameters for folding and temperature set to 25°C. The secondary structures were then homology modeled and the common pattern of folding was established.

2.4. Phylogenetic analyses

Phylogenetic analyses were carried out using neighbor-joining (NJ), maximum-parsimony (MP), maximum likelihood (ML) and bayesian inference (BI), using both primary sequence and secondary structure information with different approaches for dealing with gaps. Because of its large size and unreliable alignment, the ITS2 of *Dermatobia hominis* was excluded from phylogenetic analyses. The species *M. domestica*, *S. calcitrans* and *H. irritans* (Diptera: Muscidae) were used as outgroups.

Model selection was performed using both Modeltest 3.7 (Posada and Crandall, 1998), which implements the hierarchical likelihood ratio tests (hLRTs) and Akaike information criterion (AIC) frameworks, and MrAIC 1.4.3 (Nylander, 2004), implementing the AIC, AIC correction (AICc), Bayesian information criteria (BIC) and Akaike weights (AW).

NJ analyses were conducted in PAUP 4.0b10 (Swofford, 2001) using the Tamura-Nei substitution model with gamma correction for rate heterogeneity across the sites (TN93+G – Tamura & Nei, 1993), given by MrAIC as the best model which fits the data. Modeltest 3.7 showed the TN93+G model as the third best model. Since the first (TVM+G – transversional model) and the second (TIM+G – transitional model) models are not implemented in any phylogenetic program, we used the TN93+G in our analyses. Gap regions, reflecting insertion-deletion (indels) events, were complete and pair-wise deleted.

MP analyses were also conducted in PAUP 4.0b10, with gaps deleted, treated as a fifth character state or coded as a binary character and added as a new partition to the sequence data. Gaps were coded using the program FastGap 1.0.8 (Borchsenius, 2007), which implements the coding method of Simmons and Ochoterena (2000).

ML analysis was conducted in PhyML 2.4.4 (Guindon and Gascuel, 2003), also using the TN93+G model, with all parameters estimated. NJ, MP and ML analyses were performed with 1,000 bootstrap replicates.

The BI analyses were performed with PHASE 2.0 package (Gowri-Shankar and Jow, 2006) and MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003). All BI analyses were conducted for 2,000,000 generations (sample frequency = 1,000. burn-in = 25%).

Since MrBayes does not implement the TN93 model, analyses were conducted using the general time reversible model (GTR or REV, Tavaré, 1986). Four separate

analyses were performed in MrBayes using: (1) the GTR+G model alone; (2) the GTR+G model for nucleotides and a binary model for gaps (Felsenstein81-like model, as described by Ronquist et al., 2005), which were added as a new partition to the sequence data; (3) the GTR+G model for single strand regions and the doublet model (Ronquist et al., 2005) for double strand regions, with pairs determined by the program SecondaryStructConsensus in the PHASE package; and (4) a combination of analyses (2) and (3), with the GTR+G model for single strand regions, binary model for coded gaps and doublet model for pairs.

The PHASE package was used since it implements substitution models that are suitable for dealing with nucleic acid molecules that presents a secondary structure. These models are more fully discussed by Savill et al. (2001) and can be divided in three main categories, with 6-, 7- and 16-states, reflecting the treatment of mismatched states. The 7-state model puts all mismatches in the same category in the substitution rate matrix, whereas the 16-state model establishes a category for every mismatch state. The 6-state model does not account for mismatches. Six separate analyses were performed with the mcmphase software of the PHASE package using: (1) the TN93+G model alone, (2) the RNA6A+G model for double strand regions mixed with the TN93+G for single strand regions, (3) RNA7A+G mixed with TN93+G, (4) RNA7D+G mixed with TN93+G, (5) RNA16A+G mixed with TN93+G and (6) RNA16D+G mixed with TN93+G. Models 6A, 7A and 16A are the most general models in each category. The 7D model is a biological plausible restriction of 7A model, as discussed in the PHASE manual. The 16D model is not a restriction from 16A, being more likely an attempt to convert 4-state models to accounting for pairing constraints (also discussed in the PHASE manual and Savill et al., 2001).

All tree topologies were compared against each other using: (1) the Robinson and Foulds metric (RF - Robinson and Foulds, 1981), implemented in the TreeDist software of the PHYLIP 3.67 package (Felsenstein, 2005); (2) the Shimodaira-Hasegawa test (SH - Shimodaira and Hasegawa, 1999) implemented in CONSEL v0.1i (Shimodaira and Hasegawa, 2001); (3) the Approximately Unbiased test (AU - Shimodaira, 2002), also implemented in CONSEL v0.1i; and (4) the Expected Likelihood Weights (ELW - Strimmer and Rambaut, 2001), implemented in Tree-Puzzle v5.2 (Schmidt et al., 2002).

3. Results

3.1. Sequence and secondary structure analyses

The ITS2 sequences showed a high A+T content, ranging from 77.5 to 82.9% in Calliphoridae and from 75 to 76.1% in the Muscidae family. The sequences ranged in length from 291 bp to 351 bp in the Calliphoridae family, and from 324 bp to 354 bp Muscidae. The ITS2 region of *Dermatobia hominis* (Oestridae) is 632 bp long with 80.7% of A+T. The A+T content and length of ITS2 of these three families are shown in Table 1. The ts/tv rate, excluding *D. hominis*, is 0.4, indicating an occurrence of transversions 2.5 times greater than transitions. The ts/tv rates plotted against the genetic distance corrected by the TN93+G substitution model are shown in Figure 1.

Characterization of ITS2 sequences revealed the presence of an additional ribosomal subunit 2S, which splits the region in two parts: ITS2a with 30bp in the species analyzed in this study and ITS2 with variable sequence length (Table 1). The 2S subunit was identified by identity with previous released sequences of the ITS2 region of the dipterans species *Drosophila melanogaster* (Jordan et al., 1976; Tautz et al., 1988) and *Musca domestica* (Schlötterer et al., 1994). Since no sequence variation was found in the ITS2a region, it was excluded from phylogenetic analysis.

ITS2 secondary structures predicted for Calliphoridae, Muscidae and Oestridae species (Figure 2) presented the same general patterns described for eukaryotic ITS2, with a central ring with four helix-loop domains, named I, II, III and IV (Joseph et al., 1999. Young and Coleman, 2004).

Alignment of ITS2 sequences, corrected by the base-pairing information of secondary structure, have shown that helixes I and II, including the pyrimidine mismatch in the later are the most conserved among all species analyzed and helix IV is the most variable, as described by Young and Coleman (2004) for other Diptera species. The helix III had conserved regions, as the GTCTAGCAT-motif (Figure 2) in its 5' portion, even though the variable organization of this domain among families. For Calliphoridae species, domain III had a pronounced branching pattern (Figure 2) with two lateral branches for almost all species, one in the 5'-end (IIIa) and other in the 3'-end (IIIc). Exceptions of this pattern were found for *L. cuprina*, *L. sericata* and *M. bellardiana*. In *L. cuprina* and *L. sericata*, there was only one branch in the 5'-end and a bulge in the 3'-end (Figure 2C). In *M. bellardiana*, domain III had a similar pattern to that described for Muscidae species, but with shorter branches (Figure 2F), while *M. peregrina* presented large branches in domain III and the most uncommon

pattern is the lack of the domain IV (Figure 2D). The species *D. hominis* have shown a different branching pattern for helix III, with multiple branches in its 3'-end (Figure 2E). Indeed, the ITS2 of *D. hominis* is 281 bp longer than the longest ITS2 sequences in Calliphoridae family, mostly because of the multiple branching pattern in the third domain and also larger domains I and II.

The CBCs and hemi-CBCs (Gutell et al., 1994) were detected in helices I, II and III. Domain IV, because of its sequence and structure variability, had no reliably assigned CBCs. The presence of CBCs in three of four domains was used to confirm the secondary structure previously predicted.

The potential of CBCs to distinguish species was evaluated in the cases that two or more sequences of the same species were available (*C. vicina*, *C. vomitoria*, *C. albiceps*, *C. megacephala*, *C. rufifacies*, *P. regina*, *L. cuprina*, *L. sericata* and *M. bellardiana*). No CBCs were found between sequences from same species, except for *M. bellardiana*, which has four potential CBCs in the terminal region of helix II (Figure 3). Since the terminal region of stem-loop domains contains many indel events, these CBCs cannot be assigned with high confidence.

3.2. Phylogenetic analyses

Four representative trees recovered in the analyses (one for each phylogenetic method) are shown in Figure 4. To evaluate how different tree topologies are from each other, we used the symmetric distance between pairs of topologies, which implements the Robinson-Foulds (RF) metric (Table 2). According to Robinson and Foulds (1981), the RF distance measures the number of operations necessary to transform one topology into another. NJ and MP methods are the most distant from the others, with a RF distance usually higher than 15, whereas the BI trees using the TN93+G model mixed with the RNA models 6A+G, 7A+G, 16A+G and 16D+G are identical.

In order to make a direct comparison between tree topologies, we used three likelihood based tests: SH, AU and ELW. Results are shown in Tables 3 and 4. MP and NJ trees shown the lowest likelihood values, whereas the BI trees with mixed models (TN93+G mixed with a RNAmode+G) had the highest values, been considered the best trees according to all tests. The ML tree has a lower likelihood value when compared with the BI trees, although this value is not significantly different from those of the BI trees.

When comparing different approaches for dealing with gaps, results vary. For NJ analyses, complete exclusion of gaps resulted in a slightly improvement, although not significant, of the likelihood value when compared with pair-wise deletion. For BI analyses, the use of the GTR+G model alone (complete deletion) resulted in a significantly higher likelihood value when compared to the use of the GTR+G model mixed with the binary model for gaps coded. For MP analyses, complete deletion of gaps resulted in lower likelihood values. Additionally, including gap information in MP analyses reduced the number of most parsimonious trees recovered (260 in complete deletion approach to 12 and 6, in gaps coded and gaps treated as a fifth state approaches, respectively), whereas the use of complete deletion of gaps resulted in trees with lower lengths (898 in complete deletion approach compared to 1519 and 1962 in gaps coded and gaps treated as a fifth state approaches, respectively).

When considering secondary structure information, evaluated only for BI analyses, the TN93+G model mixed with an RNA model resulted in higher likelihood values when compared with the TN93+G alone, the GTR+G model alone and the GTR+G model mixed with the doublet model implemented in MrBayes.

3.3. Genetic relationships of blowflies

The overall relationships between subfamilies, *genera* and species were variable, depending upon the method and parameters used. In general, trees presented a good resolution at the species and *genera* level (terminal branches), with most topologies corroborating the same relationships, and a low resolution at the subfamily level.

Mesembrinellinae and Calliphorinae subfamilies are monophyletic in all trees, always with high bootstrap (BS) or posterior probability (PP) support, whereas Chrysomyinae and Luciliinae were not recovered as monophyletic clades in some analyses, or with low BS or PP support. Relationships between subfamilies are summarized in Table 5.

The *Calliphora* genus was not recovered as monophyletic in any tree, since *Aldrichina grahami* (Calliphorinae) is recovered as a sister *taxa* of *C. vicina*, with high BP and PP support (>80 and 0.8, respectively). *Lucilia* genus also was not monophyletic in trees recovered, mostly because of the paraphyletic status of *L. eximia* (as in tree 4A). Moreover, when all *Lucilia* species are recovered in the same clade, *Hemypirellia ligurriens* (Luciliinae) groups inside it, in different positions, with high PP support (~1.00) but low BP support (<60), therefore suggesting the

paraphyletic status of this *genus*. Chrysomyinae always appear as a monophyletic subfamily with high BP and PP support (~100 and 1.00, respectively). The sister *taxa* status of *Phormia* and *Protophormia* is well supported and *C. idioidea* seems to be closer related to this group than to the *Chrysomya* *genus*, as recovered in some MP trees. *H. segmentaria* is close related to the *Cochliomyia* *genus*, with BP support varying between 50 and 100 and PP varying between 0.39 and 1.00. *Mesembrinella* *genus* was always recovered as monophyletic with high BP and PP support (100 and 1.00, respectively), and it could be a subfamily (Mesembrinellinae) or a sister family (Mesembrinellidae) of Calliphoridae, as stated by some authors (Guimarães, 1977. Furusawa and Cassino, 2006).

4. Discussion

4.1. The ITS2 as a molecular marker: overall characterization

The use of ITS2 as a molecular marker for phylogenetic inference has been broadly discussed recently. The short length of this region allows its easy recovery even in museum preserved-specimens, the sequence variation allows species differentiation and the presence of a conserved core of secondary structure helps the alignment of distant-related *taxa* (Tillier and Collins, 1995 Coleman, 2003. Coleman, 2007). Since ITS2 is present in all kind of eukaryotic organisms and has a bi-parental inheritance, it could be used for molecular and evolutionary studies of protists, fungi, animals and plants, differing from current markers used as barcode, such as mitochondrial and plastidial genes, which are not present in all organisms. In some cases, ITS2 shows equal or even better performances when compared with other barcoding markers (Sass et al., 2007; Seifert et al., 2007).

However, before proceeding with phylogenetic and species diagnosis, processes underlying ITS2 sequence evolution should be well characterized. The results described in this work are similar to previous studies concerning the evolution of ribosomal DNA sequences. The nucleotide composition bias, which shows an excess of A+T, corroborates with results of Friedrich and Tautz (1997) for the Diptera order and with Song et al. (2008b) for the Calyptratae group. According to Friedrich and Tautz (1997), this bias appeared after an increase in substitution rates during the divergence of basal Diptera lineage (~250 million years ago). Possibly because of a directional mutation pressure, A+T frequency became higher, and the variability in the ITS2 A+T content found today among Diptera groups, ranging from 74-82% in

Drosophila (Young and Coleman, 2004) to 30-50% in species complex of *Anopheles* genera (Wilkerson et al., 2004; Marrelli et al., 2005), might be a reflection of different mutation rates in each group after the divergence from the Diptera common ancestor. The same does not hold for other Arthropoda groups, which shows lower A+T contents, frequently below 50% (Schlötterer et al., 1994; Friedrich and Tautz, 1997).

The reason for this bias to be preserved in Diptera is unknown. Friedrich and Tautz (1997) discuss this nucleotide composition bias in terms of equilibrium between flexibility and stability of the secondary structure of ribosomal RNA molecules, since G-C pairs, with three hydrogen bonds, confer more stability to the secondary structure, whereas A-U and the non-canonical G-U pairs, with two hydrogen bonds, confer more flexibility. Secondary structure flexibility might be important to decrease the constraint for recognition of structural motifs by proteins that interact with them, allowing conformational changes that are biologically necessary (Friedrich and Tautz, 1997; Gutell et al., 1994). The correct processing of ITS2 after transcription ensures the availability of rRNA subunits to the cell, since ITS2 mutations which compromise the correct assemblage of its secondary structure will result in reduced amounts of 5.8S and 28S rRNAs molecules (Côté et al., 2002). Once this recognition relies mostly on structural conformation, flexibility at the structural level could allow sequence variation to occur without affecting processing. For ribosomal rRNA subunits, Friedrich and Tautz (1997) also argue that adenine rich regions are related, at least for the 18S rRNA, to the recognition of hydrophobic ribosomal proteins, since adenine is the nucleotide with the lowest polarity. The reason why we do not find this A+T bias in other Arthropoda groups, since it seems to be advantageous, is unknown.

Song et al. (2008b) also discussed the A+T bias in the ITS2 region of Calypttratae flies, proposing a different explanation. According to them, mono and dinucleotide repeats, mainly A, T or AT, are common in ITS2 sequences of the Calypttratae group and a polymerase slippage mechanism could explain the maintenance or the increasing of this nucleotide composition bias. However, if this is the case, we should expect a strong correlation between mono or dinucleotide repeats and length variation between Calypttratae species, which was not found by them. Moreover, nucleotide composition variation due to a slippage mechanism could explain the accentuated A+T bias in ribosomal spacers, where this kind of event seems to be prone to occur, but not in the ribosomal genes (18S, 5.8S and 28S), where sequence and structure conservation are under selective pressures. It is possible that

nucleotides repeats play a significant role in ITS2 sequence evolution, but structural constraints (flexibility/stability) may have major influences in the rDNA cistron as a whole, thus explaining the nucleotide composition bias along the entire molecule.

An unusual feature of the ITS2 is the prevalence of transversions over transitions ($ts/tv = 0.4$). Chemical similarity, modification of nucleotides (such as methylation of cytosines), codon usage and metabolic rate are the most common factors that cause ts/tv bias, although the most frequently rate is towards higher transition values in mitochondrial and plastidial genomes (Brown and Simpson, 1982; Brown et al., 1982; Tamura, 1992a; Tamura, 1992b; Stepien et al., 1999; GuhaMajumdar and Sears, 2005; Keller et al., 2007). None of these effects seems to explain ts/tv bias to an excess of transversion in ribosomal DNA spacers, as ITS2. Schlötterer et al. (1994) found a similar result for the ts/tv ratio of ITS2 sequences in *Drosophila* ($ts/tv = 0.5$), and suggested that this value could be due to a sampling error, since the number of transversions is not significantly different from the number of transitions when gaps are excluded. In our analyses, the number of observed transversions (34) is significantly higher than the number of transitions (13) if we include gaps. If we exclude all gaps, the number of transitions becomes higher than the number of transversions (5 and 4, respectively). Since we believe that there is no biological reason to exclude gaps from this alignment, as gaps reflect indel events that are expected to occur in looped and bulged regions, the ts/tv ratio found when gaps are excluded could be viewed as the source of the sampling error. Moreover, the higher rates of A \leftrightarrow T transversions when compared to other classes of transversions (26 A \leftrightarrow T transversion against 8 of all other types of transversions when averaged for all species) indicate a strong correlation with the A+T bias, suggesting that this ts/tv ratio is biased and is probably not originated by sample error. It is possible that the same forces related to maintenance of nucleotide composition are responsible for transition/transversion bias in the ITS2 molecule.

One of the most interesting features of the ITS2 as a molecular marker is the existence of a primary sequence associated to a secondary structure. The common core of ITS2 secondary structure described for eukaryotes (Joseph et al., 1999) was identified for Calliphoridae and related families in this study and corroborates previous results of Song et al. (2008b). When compared with the secondary structure model described for *Drosophila* species by Young and Coleman (2004), helix I and II

are very similar, including the pyrimidine mismatch found in the 5' side of helix II (Figure 2). A small putative helix, named Ia by Song et al. (2008b), was found near the 5' side of ITS2. Since this helix lacks CBCs and therefore has no proofs of existence, we considered this helix an artifact, as was interpreted by Song et al. (2008b). The helix IIa is not present in dipterans analyzed in this study, as described for *Drosophila* (referencia). Helix III is split in branches in Calliphoridae, Muscidae and Oestridae species, and they also bear the conserved nucleotide stretch GTCTAGCAT on the 5' terminal portion of this helix, which is conserved for all species analyzed and possibly plays an important role in recognition of proteins which bind and process ITS2 in Diptera. Helix IV is not subdivided, as described for *Drosophila* (Young and Coleman, 2004), but is the most variable one.

Whereas distal portions of helix regions could not be alignment with confidence because of gaps, proximal portions of helix domains could be confidently aligned and, in most cases, used as a guide to global alignment procedure of distant-related *taxa*.

The occurrence of CBCs was determined in helices I, II and III and their existence was used to corroborate our secondary structure model. Since CBCs could be reliably assigned in proximal portions of helix regions only, the potential CBCs found in helix II of *M. bellardiana* could be due to indel events that are common in the terminal portion of helix regions rather than for substitution events that occurred and preserved the secondary structure. For the remaining species with two or more sequences available, no CBCs were found when comparing the sequences and secondary structures for the same species. Therefore, we could not reject the proposition that CBCs can be used to distinguish species, as proposed recently by Müller et al. (2007), since CBCs appear only between species, but never between individuals of the same species.

The evolutionary patterns of ITS2 were characterized for the Calliphoridae family and have shown a great potential to be used as a reliable molecular marker for phylogenetic inference and species diagnosis, when its associated secondary structure was considered. The positional homology of sites could be enhanced by using the proximal portions of helix regions as a guide for multiple alignments, thus extending the use of ITS2 for phylogenetic inference in higher order *taxa*.

4.2 Phylogenetic analyses using ITS2

Major problems related to phylogenetic analyses using ITS2 and ribosomal rRNA regions are often related to alignment reliability. Sequence length variation may occur in single strand regions, resulting in a larger amount of gaps that difficult the alignment. In addition, since these regions are not involved with secondary structure architecture and the evolutionary pressure is low or absent, substitutions rates are higher, resulting in the loss of positional homology even between close related *taxa*. While stem regions are evolving under high selective constraint, single strand regions evolve near neutrally, with high propensity for indel events (Schlötterer et al., 1994).

Considering gap information in phylogenetic analyses is a discussable approach, since it introduces unknown events that are not easily modeled. There is no objective biological reason to simple exclude gaps in this case, since indel events that gave rise to them are a reflection of evolutionary changes, but the best way to dealing with gaps is not fully understood. The problems of dealing with gaps in multiple sequence alignment, with special attention to the ribosomal DNA region and non-coding spacers is discussed by Lutzoni et al. (2000), Gillespie (2004) and Egan and Crandall (2007). In order to compare the dealing with gaps, in this work we used four different approaches: (i) complete deletion of gaps, (ii) pair-wise deletion of gaps, (iii) coding gaps as a binary character and (iv) treating gaps as a fifth state of character. In (iii), we coded gaps as if they had occurred one nucleotide at a time because it seems to be the most biological plausible way to do this. Indel events of larger regions could have disruptive effects on secondary structure stability and maintenance. Although this kind of event had already been described in ticks (Murrell et al., 2001), this is probably an exception rather than a rule.

The inclusion of gaps in phylogenetic analyses showed different results for different phylogenetic methods. For NJ, the complete deletion approach resulted in a slightly but not significant likelihood improvement when compared with the pair-wise approach. The two NJ trees differ from each other only in the position of *L. eximia* and *H. ligurriens*, which have little BS support, suggesting that gap consideration added little or even no information to the phylogenetic analysis with ITS2. For BI analyses, including gap information (GTR+G mixed with the Binary-model) significantly reduces the likelihood when compared to the complete deletion approach (GTR+G), although both trees retrieved almost the same topology, differing only in

some Chrysomyinae branches, which are polytomic in the mixed-model tree. Therefore, gap consideration performs worst in BI analyses with ITS2. For MP trees, considering gap information seems to improve the resolution of the method by drastically reducing the number of most parsimonious trees found, which allow a more detailed evaluation of the robustness of each branch when combined in a strict consensus tree. This improvement in resolution also seems to fit the tree to better explain the data, thus causing the likelihood increasing. This is corroborated by the fact that the MP tree with gaps as fifth state of character has both the highest likelihood score between the MP trees and the fewest number of MP trees found (6).

Regarding the use of secondary structure information to model ITS2 sequence evolution, this seems to be an additional challenge in phylogenetic analysis. Dependence among sites in sequence evolution, where the substitution pattern in a particular site is linked to another that is structural related to it, is a process that has been considered and evaluated only recently (Schöniger and von Haeseler, 1994; Muse, 1995; Rzhetsky, 1995; Tillier and Collins, 1995). The RNA secondary structures evolve at lower rates than underlying sequences (Dixon and Hillis, 1993; Gutell, 1996) and substitution rates in paired sites are higher when substitutions tend to preserve the acquired secondary structure. This feature is not properly modeled by standard, 4 by 4, nucleotide models. For this reason, more appropriate models have been developed for use with statistical methods of phylogenetic inference, including the doublet-model implemented in MrBayes and the RNA models implemented in the PHASE package (Savill et al., 2001). These models are based on 6 by 6, 7 by 7 or even 16 by 16 substitution matrices, which accounts for paired and mismatched sites with different weights for each element. The major problems with these models are associated with the rising of parameters to estimate, which increases the variance and thus the reliability of the analyses. Effects of over-parameterization vary in phylogenetic reconstruction methods and there is a consensus that statistical methods tend to perform better than others (Holder and Lewis, 2003; Felsenstein, 2004).

In our analyses, consideration of secondary structure information significantly improved the likelihood of phylogenetic analyses, as previously noted by Telford et al. (2005) with ribosomal genes. The use of the mixed models TN93+G for single strand regions and a RNA-model+G for helix regions have the highest likelihood scores and are considered the best in all tests, from the most conservative (the SH-test) to the most stringent one (the ELW test).

In an overall comparison using likelihood phylogenetic methods, some patterns could be found. MP trees had the lowest likelihood values, which could be due to the fact that MP has some undesirable properties when analyzed by a statistical point of view (Felsenstein, 2004). The low performance of parsimony could also be explained by the assumptions of the method. Maximum parsimony favors simpler hypothesis to explain an observed amount of data, which is undesirable if the more complex hypothesis is the true one. ITS2 sequence evolution is attached to its secondary structure. Loops evolve at near neutral rates, with high propensity of indel events that are allowed to occur unless they cause a structural disruption, whereas helices have their evolution constrained to allow only substitutions or indel events that preserve the pairing pattern. Thus, it is reasonable to assume that the simplest explanation could not be the best in this case. Besides, the three trees recovered using MP are not consistent among them, each one retrieving a different history. These facts suggest that MP may be not the best phylogenetic method to use with ITS2.

As described for MP, NJ has also some undesirable properties when analyzed by a statistical point of view, which could explain the comparatively lower likelihood values of the NJ trees. The most serious one is probably the loss of information in the process of simplifying evolutionary distances to observed differences (Holder and Lewis, 2003). The complexity underlying ITS2 sequence and secondary structure evolution may not be easily simplified in genetic distances. Besides, NJ method builds a tree by clustering species based on similarity rather than on evolutionary information, which could result in topologies that not necessarily correspond to a phylogenetic tree. This could explain the performance of NJ in the AU and ELW tests, being excluded from the confidence intervals of trees. Thus, NJ also seems to be not the best choice to conduce phylogenetic analyses using ITS2.

On the other hand, statistical methods (ML and BI) are claimed to make a full recovery of the information contained in the data under a given model and also to be robust against over-parameterization (Holder and Lewis, 2003; Felsenstein, 2004). Our findings corroborate these facts, since ML and BI trees had the highest likelihood values, especially those that considered secondary structure information, which are chosen as the best in all tests. Except for the monophyletic status of Chrysomyinae, which was recovered on the ML tree but not on the BI trees, both methods tell us the same history. These results suggest that the best way to access ITS2 information for phylogenetic reconstruction seems to be using statistical methods of tree

reconstruction under evolutionary models that take into account secondary structure information. However, this may not be a generalized rule, and one has to have in mind that choosing the best tree topology relies, after all, on the judgment with a biological plausible answer.

There are some regards about tree topologies comparisons using likelihood, many of them involving the complexity of tree parameters space and the distinct parameters involved in each method of phylogenetic reconstruction. However, according to Felsenstein (2004), “likelihoods of trees of different topologies” are “the probability of the same event (the data), computed conditional on different topologies” and “as such, they are on the same scale” and thus can be compared. More regards are about the nature of the trees chosen to be tested. According to Goldman et al. (2000), some tests commonly used for tree topologies comparison, specially the Kishino-Hasegawa test (KH-test, Kishino and Hasegawa, 1989), are designed to evaluate trees that are chosen *a priori*, and some bias could be encountered when using *a posteriori* defined trees, including the ML tree. The SH-test is then argued to be a more reliable test to use when the ML tree is among the compared topologies.

We think that results found in our analyses could give important clues about phylogenetic inference using ITS2, such as the importance of using models that take into account secondary structure information and phylogenetic methods that can properly deal with this kind and amount of data.

4.3. Implications for relationships of blowflies

Phylogenetic relationships of blowflies varied according to the reconstruction method used. The most accepted relationship between Calliphoridae subfamilies was (Mesembrinellinae,(Chrysomyinae,(Luciliinae,Calliphorinae))), with Chrysomyinae being paraphyletic. Relationships intra and inter *genera* recovered using ITS2 are mostly in agreement with previous studies on the Calliphoridae family that were conducted with mitochondrial and ribosomal genes (Wallman and Donnellan, 2001; Wells and Sperling, 2001; Stevens and Wall, 2001; Otranto and Stevens, 2002; Stevens, 2003; Chen et al., 2004; Junqueira et al., 2004; de Azeredo-Espin and Lessinger, 2006; Stevens and Wallman, 2006; Stevens et al., 2006; Wells and Williams, 2007; Nelson et al., 2007). Our results significantly differed from those of Song et al. (2008b), which used ITS2 for phylogenetic reconstruction of the Calyptratae group and recovered a paraphyletic relationship between Calliphorinae,

Chrysomyinae and Luciliinae subfamilies, with Calliphoridae being not a monophyletic group. Since they used a reduced Calliphoridae dataset, sampled to perform comparisons at higher taxonomic levels, our results could not be directly compared to theirs.

The paraphyletic status of the genera *Calliphora* and *Lucilia* corroborate previous works for Calliphoridae. The insertion of *H. ligurriens* inside the *Lucilia* genus (Figure 4) was also recovered by Chen et al. (2004) using the gene cytochrome c oxidase subunit I (*cox1*) of the mitochondrial genome. They argued that some species of the *Hemipyrellia* genus have been previously described as *Lucilia* group, including the type species *Hemipyrellia curriei* (Townsend 1918), which was originally designated as *Lucilia fernandica* (Macquart 1855). Fan (1997) described these two genera with similar morphologies and habitats. Maybe, as suggested by Chen et al. (2004), a more detailed taxonomical revision is necessary. The same could be true for *A. grahami*, which was initially described as *Calliphora grahami* (Aldrich 1930) and then included in the *Aldrichina* genus (Townsend 1937), being *A. grahami* the type species (Kurahashi, 1971). Kurahashi (1971) described some characters that could be used to distinguish the *Aldrichina* genus from other Australian and Oriental Calliphorinae genera. Its inclusion in the *Calliphora* genus was also recovered using ITS2 by Song et al. (2008a, 2008b), thus suggesting the need of a more broadly taxonomical revision of Calliphoridae using both morphological and molecular data.

The *Cochliomyia* and *Chloroprocta* genera seem to be sister-groups of *Hemilucilia* and (*Phormia*,*Protophormia*), respectively (Figure 4), although *Chloroprocta* could also be closer related to the *Chrysomya* genus. The sister taxa status of (*Hemilucila*,*Cochliomyia*) and (*Chloroprocta*,*Chrysomya*) corroborate a previous work using the control region of the mitochondrial DNA (Gustavo Turqueto Duarte, personal communication).

The exact position of *Mesembrinella* group could not be resolved. Guimarães (1977) and Furusawa & Cassino (2006) elevated this group to family status (Mesembrinellidae), as a sister taxa of Calliphoridae, based on a few morphological characters. This classification is not accepted for many authors, such as MacAlpine (1989) and Rognes (1997), which consider the *Mesembrinella* a subfamily of Calliphoridae (Mesembrinellinae), although they recognized the paraphyletic status of the Calliphoridae family. Rognes (1997) suggests that the clade composed by (Lucilinae+Melanomyinae+Calliphorinae+Toxotarsinae+Chrysomyinae) “contains all

the sarcosaprophogus flies and is a good candidate for the take-over of the name Calliphoridae, should it prove a reliable group". In this statement, he excluded Mesembrinellinae from the group, thus giving some support to the family status of Mesembrinellidae given by Guimarães (1977). Based on the trees recovered in this work, we could not accept neither reject the classification proposed by Guimarães (1977), since *Mesembrinella* grouped with the Calliphoridae family but is distantly related to other subfamilies (Chrysomyinae, Luciliinae and Calliphorinae). There are few studies on the Mesembrinellinae group, most of them describing new species, since the habits of these flies are barely known. *Mesembrinella* species are found only in natural, non-disturbed environments, restricted to the rain forest habitat of Neotropical region, and little is known about their natural history, reproductive cycle and other habits (Guimarães, 1977).

Genetic relationships of the Calliphoridae family are almost resolved with great confidence, at least for species having forensic and veterinary importance (Stevens and Wallman, 2006; Stevens et al., 2006). More attention should be dedicated to species that inhabit natural and non-synantropic areas, since these groups seem to present a greater biological diversity, thus providing insights into the evolution of the family and particularly to the parasitism habit, since larvae from Calliphoridae species can be free-living, facultative or parasites. A true understanding about the evolution of parasitism in this family will only be achieved when more data have been collected and deeper and wider evolutionary analyses have been performed.

Conclusions

The ITS2 region has shown potential as a molecular marker for species diagnostic and phylogenetic inference of species at higher order *taxa* level. Its primary sequence variability allows species differentiation and phylogenetic reconstruction of relationships between species, *genera* and family level, whereas its secondary structure information allows alignment of distant-related *taxa*, thus enhancing its potential for phylogenetic reconstruction beyond the family level. Compensatory base changes information also seems to be a possible way to distinguish between species, but deeper studies must be performed.

Authors who intend to characterize the ITS2 as a molecular marker for phylogenetic inference must consider nucleotide and ts/tv biases, since substitution models must account for this kind of information. The dependence between sites in

sequence evolution seems to play a more important role in phylogenetic analyses than consideration of gaps. Mixed models using a standard nucleotide model for single strand regions coupled with a RNA-model for helix regions are desirable and statistical methods of phylogenetic inference (ML and BI) performed better for analyses using ITS2, since they are more robust to parameter-rich models.

For the Calliphoridae family, phylogenetic trees inferred using ITS2 corroborates previous studies with other molecular markers, thus confirming its potential as a good molecular marker for phylogenetic inference. The monophyletic status of Luciliinae and Calliphorinae subfamilies is supported, as the paraphyletic status of Chrysomyinae. For relationships between *genera*, *Calliphora* and *Lucilia* paraphyletism corroborates previous results using mitochondrial markers. The suggestion for a taxonomic review in Calliphoridae groups incorporating molecular data seems to be reasonable and urgent. Further analyses using broader taxonomic samples will indicate the relationship between the *Mesembrinella* and Calliphoridae groups, as will elucidate the potential of ITS2 to recover evolutionary history through higher *taxa*. Fully knowledge about evolution of the parasitism habit in Calyptratae groups will only be achieved after phylogenetic analyses that include a broader representation of Calliphoridae species, including other groups than only those with veterinary and forensic importance.

The evolutionary processes underlying ITS2 sequence and structure evolution in the Calliphoridae family agree with previous studies carried out using ribosomal gene sequences in other Diptera, Insecta and Arthropoda groups. Therefore, patterns of sequence evolution in the ITS2 region that are characterized in this work probably could be applied in other groups of the Arthropoda *phyla*.

Acknowledgements

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, grant 05/50100-5). The authors would like to thank Rosângela Rodrigues for technical assistance, Alexandre Ururahy Rodrigues, Bernard Greenberg, Michelle L. Harvey, Carole Ames, Rod Mahon, Bernard Chauvet, Élvia E. S. Vianna and Jeffrey D. Wells for providing specimens, Arício X. Linhares for taxonomic identifications and Mariana Lúcio Lyra, Luciane Mendes Hatadani and Karina Lucas Silva Brandão for valuable discussions on molecular phylogenetics.

Literature cited

- Amendt, J., Krettek, R., Zehner, R., 2004. Forensic entomology. *Naturwissenschaften*. 91(2), 51-65.
- Azeredo-Espin, A.M., Lessinger, A.C., 2006. Genetic approaches for studying myiasis-causing flies: molecular markers and mitochondrial genomics. *Genetica*. 126(1-2), 111-131.
- Becerra, J.X., 2004. Molecular systematics of Blepharida beetles (Chrysomelidae: Alticinae) and relatives. *Mol. Phylogenet. Evol.* 30(1), 107-117.
- Borchsenius, F., 2007. FastGap 1.0.8. Department of Biological Sciences, University of Aarhus. Published online at http://www.aubot.dk/fb/FastGap_home.htm.
- Brown, G.G., Simpson, M.V., 1982. Novel features of animal mtDNA evolution as shown by sequences of two rat cytochrome oxidase subunit II genes. *Proc. Natl. Acad. Sci. USA*. 79(10), 3246-3250.
- Brown, W.M., Prager, E.M., Wang, A., Wilson, A.C., 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J. Mol. Evol.* 18(4), 225-239.
- Chen, W.Y., Hung, T.H., Shiao, S.F., 2004. Molecular identification of forensically important blow fly species (Diptera: Calliphoridae) in Taiwan. *J. Med. Entomol.* 41(1), 47-57.
- Coleman, A.W., Vacquier, V.D., 2002. Exploring the phylogenetic utility of ITS2 sequences for animals: a test case for abalone (*Haliotis*). *J. Mol. Evol.* 54(2), 246-257.
- Coleman, A.W., 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet.* 19(7), 370-375.
- Coleman, A.W., 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Res.* 35(10), 3322-3329.
- Côté, C.A., Greer, C.L., Peculis, B.A., 2002. Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. *RNA* 8(6), 786-797.
- Dixon, M.T., Hillis, D.M., 1993. Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. *Mol. Biol. Evol.* 10(1), 256-267.
- Dumont, H.J., Vanfleteren, J.R., De Jonckheere, J.F., Weekers P.H.H., 2005. Phylogenetic relationships, divergence time estimation, and global biogeographic patterns of calopterygoid damselflies (Odonata, Zygoptera) inferred from ribosomal DNA sequences. *Syst. Biol.* 54(3), 347-362.
- Egan, A.N., Crandall, K.A., 2007. Incorporating gaps as phylogenetic characters across eight DNA regions: Ramifications for North American Psoraleae (Leguminosae). *Mol. Phylogenet. Evol.* 46(2), 532-546.
- Elder, J.F.Jr., Turner, B.J., 1995. Concerted evolution of repetitive DNA sequences. *Q. Rev. Biol.* 70(3), 297-320.
- Fan, Z.D., 1997. *Fauna sinica, Insecta Vol. 6, Diptera: Calliphoridae*. Beijing, Science Press.
- 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(6), 644-653.

- Furusawa, G.P., Cassino, P.C.R., 2006. Ocorrência e Distribuição de Calliphoridae (Díptera, Oestroidea) em um Fragmento de Mata Atlântica Secundária no Município de Engenheiro Paulo de Frontin, Médio Paraíba, RJ. *Revista de Biologia e Ciências da Terra* 6, 152-164.
- Gallego, D., Galián, J., 2001. The internal transcribed spacers (ITS1 and ITS2) of the rDNA differentiates the bark beetle forest pests *Tomicus destruens* and *T. piniperda*. *Insect Mol. Biol.* 10(5), 415-420.
- Geerlings, T.H., Vos, J.C., Raué, H.A., 2000. The final step in the formation of 25S rRNA in *Saccharomyces cerevisiae* is performed by 5'→3' exonucleases. *RNA* 6(12), 1698-1703.
- Gillespie, J.J., 2004. Characterizing regions of ambiguous alignment caused by the expansion and contraction of hairpin-stem loops in ribosomal RNA molecules. *Mol. Phylogenet. Evol.* 33(3), 936-943.
- Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* 49(4), 652-70.
- Gómez-Zurita, J., Juan, C., Petitpierre, E., 2000. Sequence, secondary structure and phylogenetic analyses of the ribosomal internal transcribed spacer 2 (ITS2) in the *Timarcha* leaf beetles (Coleoptera: Chrysomelidae). *Insect Mol. Biol.* 9(6), 591-604.
- Gomulski, L.M., Meiswinkel, R., Delécolle, J.C., Goffredo, M., Gasperi, G., 2006. Phylogeny of the subgenus *Culicoides* and related species in Italy, inferred from internal transcribed spacer 2 ribosomal DNA sequences. *Med. Vet. Entomol.* 20(2), 229-238.
- 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>.
- Guhamajumdar, M., Sears, B.B., 2005. Chloroplast DNA base substitutions: an experimental assessment. *Mol. Genet. Genomics.* 273(2), 177-183.
- Guimarães, J.H., Papavero, N., Prado, A.P., 1983 As miíases na região Neotropical: identificação, biologia e bibliografia. *Rev. Bras. Zool.* 1, 239-416.
- Guimarães, J.H., 1977. A systematic revision of the Mesembrinellidae, stat. nov. (Diptera, Cyclorrhapha). *Arquivos de Zoologia* 29(1), 1-109.
- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52(5), 696-704.
- Gutell, R.R., Larsen, N., Woese, C.R., 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Rev.* 58(1), 10-26.
- Gutell, R.R., 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.
- Hickson, R.E., Simon, C., Perry, S.W., 2000. The performance of several multiple-sequence alignment programs in relation to secondary-structure features for an rRNA sequence. *Mol. Biol. Evol.* 17, 530-539.
- Hillis, D.M., Dixon, M.T., 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.* 66(4), 411-453.
- Holder, H., Lewis, P.O., 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nat. Rev. Genet.* 4(4), 275-284.

- Infante, M.E.V., Azeredo-Espin, A.M.L., 1995. Genetic variability in mtDNA of *Cochliomyia hominivorax* (Diptera: Calliphoridae) from Brazil. *Biochem. Genet.* 33(7/8), 237-256.
- Jordan, B.R., 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(1), 85-105.
- Joseph, N., Krauskopf, E., Vera, M.I., Michot, B., 1999. Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. *RNA* 8(6), 786-797.
- Junqueira, A.C., Lessinger, A.C., Azeredo-Espin, A.M., 2002. Methods for the recovery of mitochondrial DNA sequences from museum specimens of myiasis-causing flies. *Med. Vet. Entomol.* 16(1), 39-45.
- Junqueira, A.C., Lessinger, A.C., Torres, T.T., da Silva, F.R., Vettore, A.L., Arruda, P., Azeredo-Espin, A.M., 2004. The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). *Gene* 339, 7-15.
- Kampen, H., 2005. The ITS2 ribosomal DNA of *Anopheles beklemishevi* and further remarks on the phylogenetic relationships within the *Anopheles maculipennis* group of species (Diptera: Culicidae). *Parasitol. Res.* 97(2), 118-128.
- Kelchner, S.A., Thomas, M.A., 2007. Model use in phylogenetics: nine key questions. *Trends Ecol Evol* 22(2), 87-94.
- Keller, I., Bensasson, D., Nichols, R.A., 2007. Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes. *PLoS Genet.* 3(2), e22.
- Kishino, H., Hasegawa, M., 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J. Mol. Evol.* 29(2), 170-179.
- Kurahashi, H., 1971. The tribe Calliphorini from Australian and Oriental regions. *Pacific Insects* 13(1), 141-204.
- Li, C., Wilkerson, R.C., 2005. Identification of *Anopheles* (*Nyssorhynchus*) *albitarsis* complex species (Diptera: Culicidae) using rDNA internal transcribed spacer 2-based polymerase chain reaction primers. *Mem. Inst. Oswaldo Cruz.* 100(5), 495-500.
- Liao, D., 1999. Concerted evolution: molecular mechanisms and biological implications. *Am. J. Hum. Genet.* 64(1), 24-30.
- Lutzoni, F., Wagner, P., Reeb, V., Zoller, S., 2000. Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. *Syst. Biol.* 494, 628-651.
- MacAlpine, J.F., 1989. *Manual of nearctic Diptera 3*. Agriculture Canada Monograph 32. Canadian Gov. Publ. Center, Quebec, Canada.
- Malloch, G., Fenton, B., Goodrich, M.A., 2001. Phylogeny of raspberry beetles and other Byturidae (Coleoptera). *Insect Mol. Biol.* 10(3), 281-291.
- Marinucci, M., Romi, R., Mancini, P., Di Luca, M., Severini, C., 1999. Phylogenetic relationships of seven palearctic members of the *maculipennis* complex inferred from ITS2 sequence analysis. *Insect Mol. Biol.* 8(4), 469-480.
- Marrelli, M.T., Floeter-Winter, L.M., Malafrente, R.S., Tadei, W.P., Lourenço-de-Oliveira, R., Flores-Mendoza, C., Marinotti, O., 2005. Amazonian malaria vector anopheline relationships interpreted from ITS2 rDNA sequences. *Med. Vet. Entomol.* 19(2), 208-218.

- 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(4), 502-513.
- 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(4), 457-466.
- Müller, T., Philippi, N., Dandekar, T., Schultz, J., Wolf, M., 2007. Distinguishing species. *RNA* 13(9), 1469-1472.
- Murrell, A., Campbell, N.J., Barker, S.C., 2001. Recurrent gains and losses of large (84-109 bp) repeats in the rDNA internal transcribed spacer 2 (ITS2) of rhipicephaline ticks. *Insect Mol. Biol.* 10(6), 587-596.
- Muse, S.V., 1995. Evolutionary analyses of DNA sequences subject to constraints of secondary structure. *Genetics.* 139(3), 1429-1439.
- Nelson, L.A., Wallman, J.F., Dowton, M., 2007. Using COI barcodes to identify forensically and medically important blowflies. *Med. Vet. Entomol.* 21(1), 44-52.
- Nylander, J.A.A., 2004. MrAIC.pl. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.
- Otranto, D., Stevens, J.R., 2002. Molecular approaches to the study of myiasis-causing larvae. *Int. J. Parasitol.* 32(11), 1345-1360.
- Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics.* 14(9), 817-818.
- Robinson DR, Foulds LR 1981. Comparison of phylogenetic trees. *Math. Biosci.* 53, 131-147.
- Rognes, K., 1997. The Calliphoridae (blowflies) (Diptera. Oestroidea) are not a monophyletic group. *Cladistics* 13, 27-66.
- Rojas, M., Mora, M.D., Ubeda, J.M., Cutillas, C., Navajas, M., Guevara, D.C., 2002. Phylogenetic relationships in rhinonyssid mites (Acari: Rhinonyssidae) based on ribosomal DNA sequences: insights for the discrimination of closely related species. *Parasitol. Res.* 88(7), 675-681.
- Ronquist, F., Huelsenbeck, J.P., van der Mark, P., 2005. MrBayes 3.1 manual, Draft 5/17/2005. Available at: http://mrbayes.csit.fsu.edu/mb3.1_manual.pdf.
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12), 1572-1574.
- Rzhetsky, A., 1995. Estimating substitution rates in ribosomal RNA genes. *Genetics* 141(2), 771-783.
- Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. *Molecular Cloning - A Laboratory Manual*, 2nd Edition. Cold Spring Harbour Laboratory Press, New York.
- Sass, C., Little, D.P., Stevenson, D.W., Specht, C.D., 2007. DNA barcoding in the cycadales: testing the potential of proposed barcoding markers for species identification of cycads. *PLoS ONE* 2(11), e1154.
- Savill, N.J., Hoyle, D.C., Higgs, P.G., 2001. RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods. *Genetics* 157(1), 399-411.
- Schlötterer, C., Hauser, M.T., von Haeseler, A., Tautz, D., 1994. Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. *Mol. Biol. Evol.* 11(3), 513-522.
- Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18, 502-504.

- Schöniger, M., von Haeseler, A., 1994. A stochastic model for the evolution of autocorrelated DNA sequences. *Mol. Phylogenet. Evol.* 3(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(4), 361-364.
- Schultz, J., Müller, T., Achtziger, M., Seibel, P.N., 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, P.N., 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, K.A., Samson, R.A., Dewaard, J.R., Houbraken, J., Lévesque, C.A., Moncalvo, J.M., Louis-Seize, G., Hebert, P.D., 2007. Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. *Proc. Natl. Acad. Sci. USA* 104(10), 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.
- Severini, C., Silvestrini, F., Mancini, P., La Rosa, G., Marinucci, M., 1996. Sequence and secondary structure of the rDNA second internal transcribed spacer in the sibling species *Culex pipiens* L. and *Cx. quinquefasciatus* Say (Diptera: Culicidae). *Insect Mol. Biol.* 5(3), 181-186.
- Shaw, M., Murrell, A., Barker, S.C., 2002. Low intraspecific variation in the rRNA internal transcribed spacer 2 (ITS2) of the Australian paralysis tick, *Ixodes holocyclus*. *Parasitol. Res.* 88(3), 247-252.
- Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114-1116.
- Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17(12), 1246-1247.
- Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* 51(3), 492-508.
- Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49(2), 369-381.
- Song, Z., Wang, X., Liang, G., 2008a. 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(1), 17-22.
- Song, Z., Wang, X., Liang, G., 2008b. Molecular Evolution and Phylogenetic Utility of the Internal Transcribed Spacer 2 (ITS2) in Calypterae (Diptera: Brachycera). *J. Mol. Evol.* 67(5), 448-464.
- Stepien, C.A., Hubers, A.N., Skidmore, J.L., 1999. Diagnostic genetic markers and evolutionary relationships among invasive dreissenoid and corbiculoid bivalves in North America: phylogenetic signal from mitochondrial 16S rDNA. *Mol. Phylogenet. Evol.* 13(1), 31-49.
- Stevens, J., Wall, R., 2001. Genetic relationships between blowflies (Calliphoridae) of forensic importance. *Forensic Sci. Int.* 120(1-2), 116-123.
- Stevens, J.R., 2003. The evolution of myiasis in blowflies (Calliphoridae). *Int. J. Parasitol.* 33(10), 1105-1113.
- Stevens, J.R., Wallman, J.F., 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(4), 181-188.

- Stevens, J.R., Wallman, J.F., 2006. The evolution of myiasis in humans and other animals in the Old and New Worlds (part I): phylogenetic analyses. *Trends Parasitol.* 22(3), 129-136.
- Strimmer, K., Rambaut, A., 2002. Inferring confidence sets of possibly misspecified gene trees. *Proc. Bio. Sci.* 269(1487), 137-142.
- Sullivan, J., Joyce, P., 2005. Model selection in phylogenetics. *Ann. Rev. Ecol. Evol. Syst.* 36, 445-466.
- Swofford, D.L., 2001. PAUP: Phylogenetic Inference Using Parsimony, ver. 4.0b10. Sinauer Associates, Sunderland, Massachusetts.
- Tamura, K., 1992a. The rate and pattern of nucleotide substitution in *Drosophila* mitochondrial DNA. *Mol. Biol. Evol.* 9(5), 814-825.
- Tamura, K., 1992b. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol. Biol. Evol.* 9(4), 678-687.
- Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10(3), 512-526.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24(8), 1596-1599.
- Tautz, D., Hancock, J.M., Webb, D.A., Tautz, C., Dover, G.A., 1988. Complete sequences of the rRNA genes of *Drosophila melanogaster*. *Mol. Biol. Evol.* 5(4), 366-376.
- Tavaré, S., 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. *Lect. Math. Life Sci.* 17, 57-86.
- Telford, M.J., Wise, M.J., Gowri-Shankar, V., 2005. Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the bilateria. *Mol. Biol. Evol.* 22(4), 1129-1136.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25(24), 4876-4882.
- Thompson, J.D., Plewniak, F., Poch, O., 1999. A comprehensive comparison of multiple sequence alignment programs. *Nucleic Acids Res.* 27(13), 2682-2690.
- 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.
- van der Sande, C.A., Kwa, M., van Nues, R.W., van Heerikhuizen, H., Raué, H.A., Planta, R.J., 1992. Functional analysis of internal transcribed spacer 2 of *Saccharomyces cerevisiae* ribosomal DNA. *J. Mol. Biol.* 223(4), 899-910.
- Veldman, G.M., Klootwijk, J., van Heerikhuizen, H., Planta, R.J., 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(19), 4847-4862.
- Wallman, J.F., Donnellan, S.C., 2001. The utility of mitochondrial DNA sequences for the identification of forensically important blowflies (Diptera: Calliphoridae) in southeastern Australia. *Forensic Sci. Int.* 120(1-2), 60-67.

- Wells, J.D., Sperling, F.A., 2001. DNA-based identification of forensically important Chrysomyinae (Diptera: Calliphoridae). *Forensic Sci. Int.* 120(1-2), 110-115.
- Wells, J.D., Williams, D.W., 2007. Validation of a DNA-based method for identifying Chrysomyinae (Diptera: Calliphoridae) used in a death investigation. *Int. J. Legal Med.* 121(1), 1-8.
- Wilkerson, R.C., Reinert, J.F., Li, C., 2004. Ribosomal DNA ITS2 sequences differentiate six species in the *Anopheles crucians* complex (Diptera: Culicidae). *J. Med. Entomol.* 41(3), 392-401.
- Xia, X., Xie, Z., 2001. DAMBE: software package for data analysis in molecular biology and evolution. *J. Hered.* 92(4), 371-373.
- Young, I., Coleman, A.W., 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(1), 236-242.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31(13), 3406-3415.
- Zumpt, F., 1965. *Myiasis in man and animals in the Old World*. London, Butterworths. p. 267.

Figure Legends

Figure 1. Transition/transversion rates plotted against genetic distance corrected by the Tamura-Nei model. Dotted lines represent the least-square best fit line. Δ = transversions; X = transversions.

Figure 2. Predicted secondary structures of ITS2 for (A) *C. megacephala*, (B) *C. hominivorax*, (C) *L. cuprina*, (D) *M. peregrina*, (E) *D. hominis* and (F) *M. domestica*. Domains I, II, III and IV are indicated. IIIa, IIIb and IIIc indicate branches found in helix III. IVa and IVb are possible subdivisions of the domain IV in *M. domestica*. The conserved nucleotide region GUCUAGCAU is shown in black circles and the characteristic pyrimidine mismatch of helix II is indicated by an arrowhead.

Figure 3. Sequence and secondary structure variation of the helix II for *M. bellardiana*. Brackets show the four potential CBCs in the distal portion of this helix.

Figure 4. Phylogenetic trees reconstructed using ITS2. (A) Neighbor-Joining tree with complete deletion of gaps. (B) Strict consensus of the 6 Maximum Parsimony trees with gaps treated as a fifth state of character. (C) Maximum Likelihood tree recovered with the TN93+G model. (D) Bayesian tree recovered using the TN93+G model mixed with the RNA6A+G model for pairs.

Figure 1

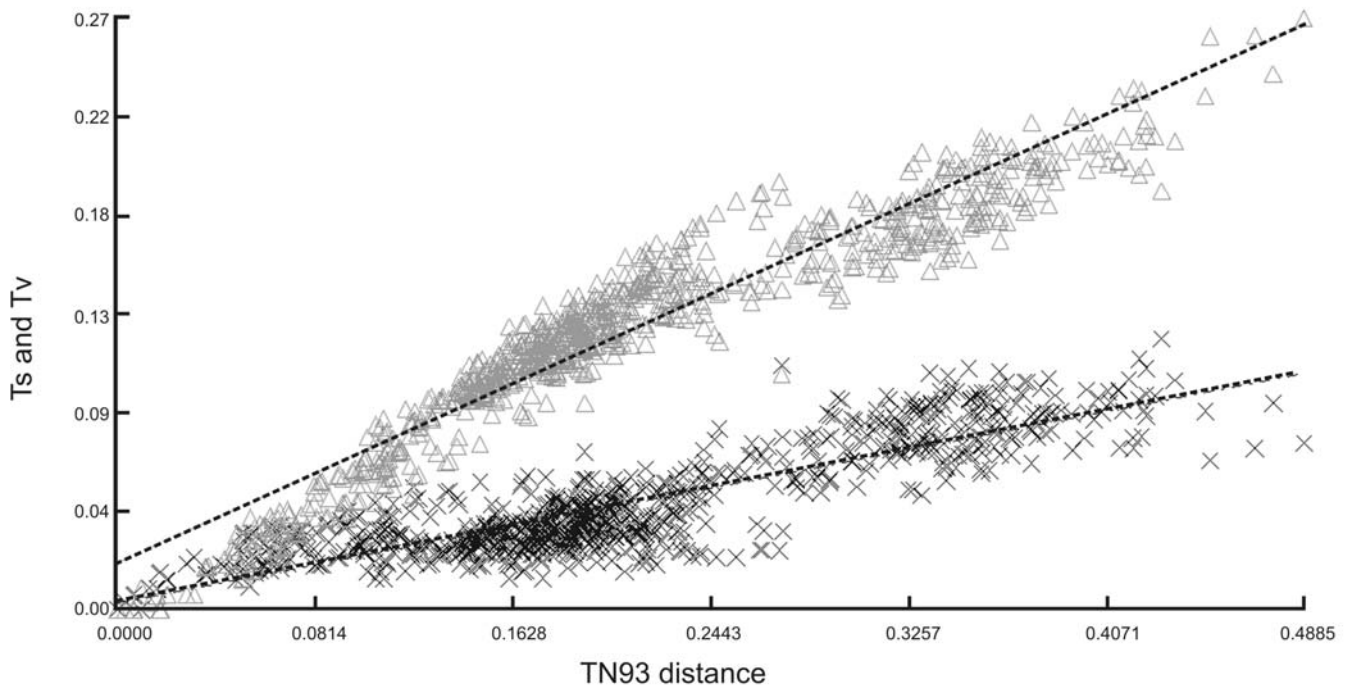


Figure 2

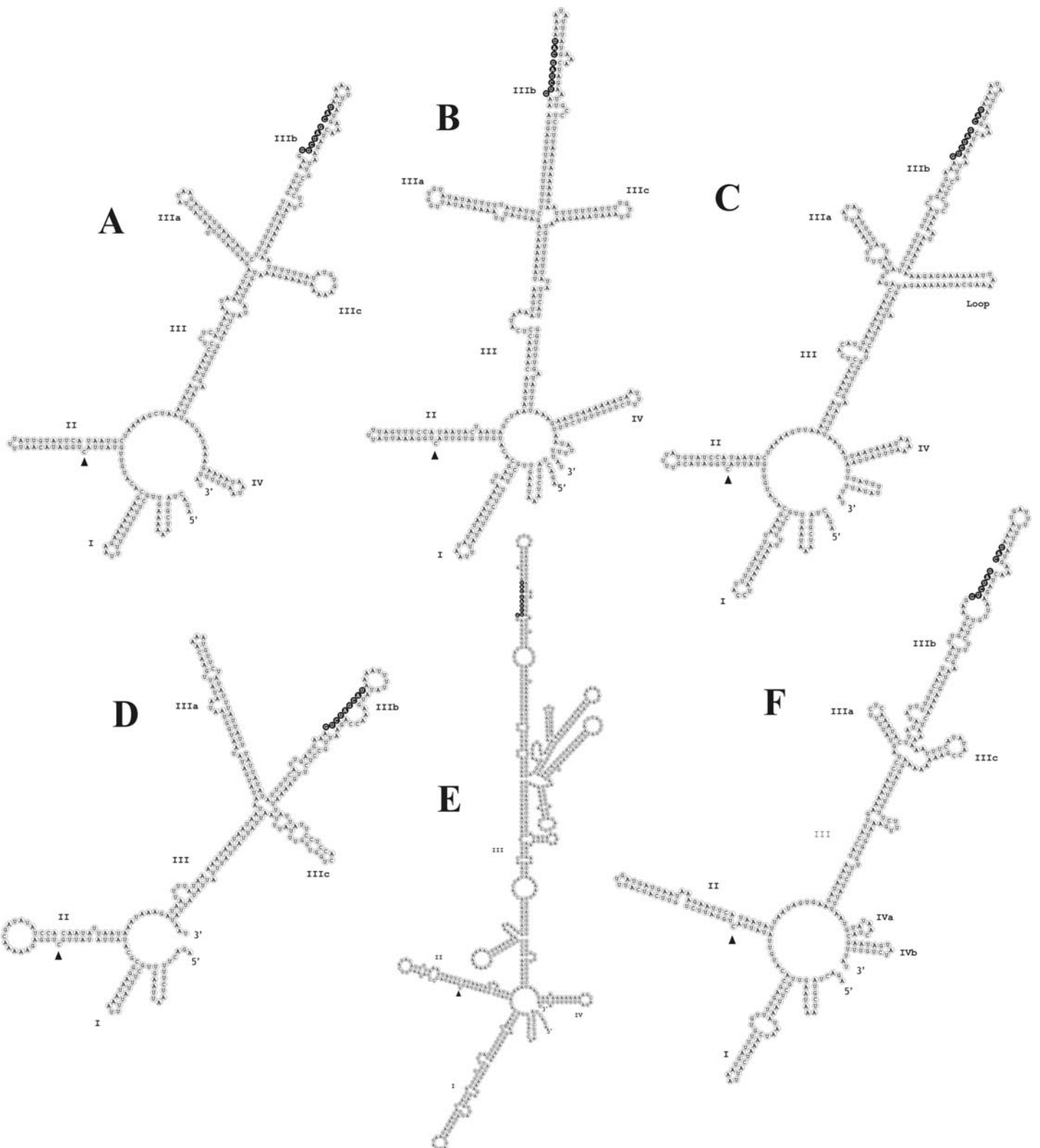


Figure 3

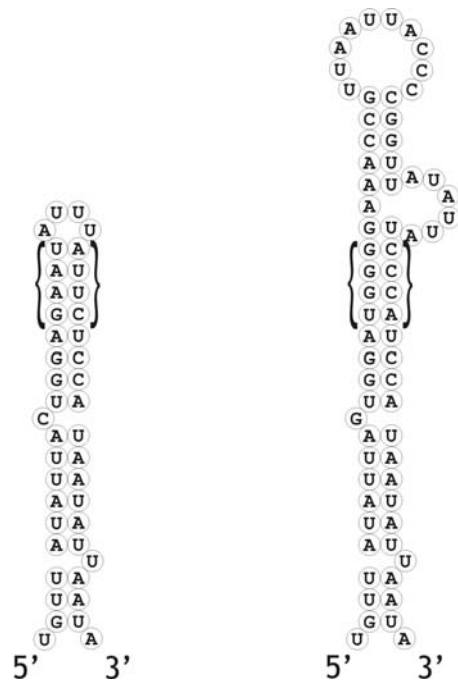


Figure 4

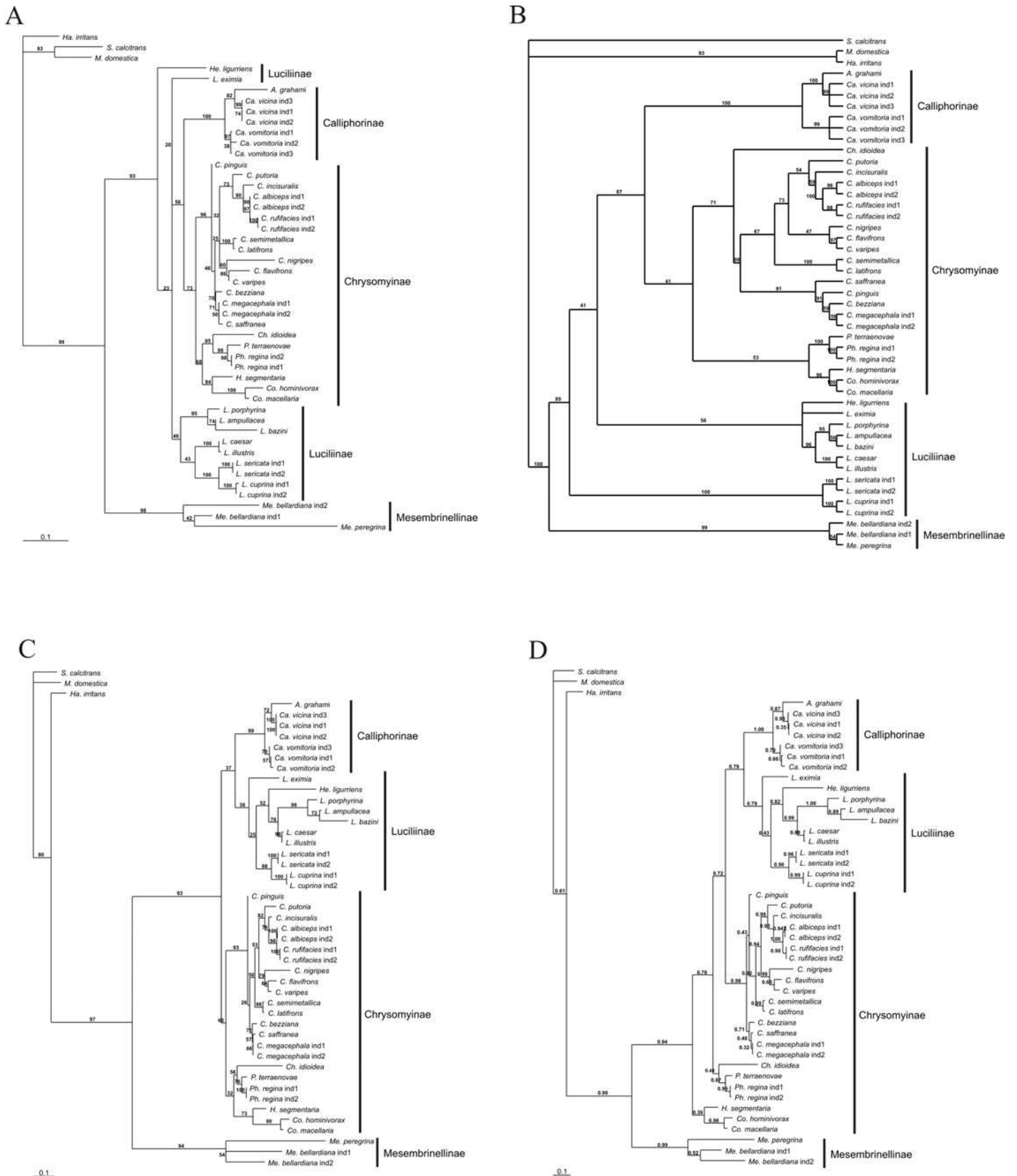


Table 1. List of species analyzed in this study.

Family	Species	GenBank Accession Number	ITS2 length (bp)	A+T content (%)
Calliphoridae	<i>Aldrichina grahami</i>	EF061802 ^a	336	80.4
	<i>Calliphora vicina</i> ind1	EF560178 ^b	327	81.3
	<i>Calliphora vicina</i> ind2	AF498026 ^c	327	81.3
	<i>Calliphora vicina</i> ind3	EF061803 ^a	326	81.3
	<i>Calliphora vomitoria</i> ind1	EF560179 ^b	314	81.2
	<i>Calliphora vomitoria</i> ind2	AF498021 ^c	312	81.1
	<i>Calliphora vomitoria</i> ind3	EF061800 ^a	318	80.8
	<i>Chloroprocta idioidea</i>	EF560180 ^b	320	81.8
	<i>Chrysomya albiceps</i> ind1	EF560172 ^b	317	81.7
	<i>Chrysomya albiceps</i> ind2	EF560173 ^b	317	81.7
	<i>Chrysomya bezziana</i>	EF560174 ^b	302	79.2
	<i>Chrysomya flavifrons</i>	DQ310497 ^d	322	80.8
	<i>Chrysomya incisuralis</i>	DQ310486 ^d	319	80.8
	<i>Chrysomya latifrons</i>	DQ310492 ^d	316	78.5
	<i>Chrysomya megacephala</i> ind1	EF560175 ^b	306	79.1
	<i>Chrysomya megacephala</i> ind2	EF061788 ^a	303	78.8
	<i>Chrysomya nigripes</i>	DQ310489 ^d	325	78.5
	<i>Chrysomya pinguis</i>	EF061792 ^a	303	79.9
	<i>Chrysomya putoria</i>	EF560176 ^b	302	79.8
	<i>Chrysomya rufifacies</i> ind1	EF560177 ^b	320	80.9
	<i>Chrysomya rufifacies</i> ind2	EF061790 ^a	317	80.7
	<i>Chrysomya saffranaea</i>	DQ310490 ^d	310	79.3
	<i>Chrysomya semimetallica</i>	DQ310493 ^d	321	78.5
	<i>Chrysomya varipes</i>	DQ310491 ^d	310	81.0
	<i>Cochliomyia hominivorax</i>	EF560181 ^b	351	81.2
	<i>Cochliomyia macellaria</i>	EF560182 ^b	344	81.1
	<i>Hemilucilia segmentaria</i>	EF560192 ^b	327	82.3
	<i>Hemipyrellia ligurriens</i>	EF061804 ^a	312	82.7
	<i>Lucilia ampullacea</i>	AF498030 ^c	300	82.3
	<i>Lucilia bazini</i>	EF061785 ^a	317	82.6
	<i>Lucilia caesar</i>	AF498034 ^c	312	81.7
	<i>Lucilia cuprina</i> ind1	EF560185 ^b	336	81.6
	<i>Lucilia cuprina</i> ind2	EF061805 ^a	333	81.4
	<i>Lucilia eximia</i>	EF560186 ^b	326	82.5
	<i>Lucilia illustris</i>	AF498038 ^c	314	82.1
	<i>Lucilia porphyrina</i>	EF061798 ^a	301	82.7
	<i>Lucilia sericata</i> ind1	EF560187 ^b	321	79.4
	<i>Lucilia sericata</i> ind2	EF061796 ^a	318	79.3

	<i>Mesembrinella bellardiana</i> ind1	<u>EU076455</u> ^b	307	79.5
	<i>Mesembrinella bellardiana</i> ind2	<u>EU076456</u> ^b	334	77.5
	<i>Mesembrinella peregrina</i>	<u>EF560188</u> ^b	291	80.0
	<i>Phormia regina</i> ind1	<u>EF061806</u> ^a	324	82.7
	<i>Phormia regina</i> ind2	<u>EF560190</u> ^b	327	82.9
	<i>Protophormia terraenovae</i>	<u>EF560193</u> ^b	331	82.8
	<i>Haematobia irritans</i>	<u>EF560184</u> ^b	347	76.1
Muscidae	<i>Musca domestica</i>	<u>EF560189</u> ^b	354	75.7
	<i>Stomoxys calcitrans</i>	<u>EF560191</u> ^b	324	75.0
Oestridae	<i>Dermatobia hominis</i>	<u>EF560183</u> ^b	632	80.7

a = sequences from Song et al., 2008;

b = new sequences;

c = sequences from Pickles et al., 2002 (unpublished)

d = sequences from Nelson et al., 2008.

Table 2. Symmetric distances between pairs of topologies. GTR = general time reversible model TN93 = the Tamura-Nei model. Numbers on the upper row refer to trees recovered using the method on the second column.

Trees		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Neighbor-Joining	TN93+G Complete Deletion (1)	-															
	TN93+G Pair-wise Deletion (2)	10	-														
Bayesian Inference	GTR+G (3)	23	19	-													
	GTR+G and Binary Model (4)	24	22	13	-												
	GTR+G; Binary and Doublet Models (5)	24	22	19	6	-											
	GTR+G and Doublet Model (6)	26	22	3	12	18	-										
	TN93+G (7)	34	28	13	22	28	16	-									
	TN93+G and RNA6A+G (8)	24	20	5	12	18	6	16	-								
	TN93+G and RNA7A+G (9)	24	20	5	12	18	6	16	0	-							
	TN93+G and RNA7D+G (10)	24	20	5	12	18	6	16	0	0	-						
	TN93+G and RNA16A+G (11)	24	20	5	12	18	6	16	0	0	0	-					
	TN93+G and RNA16D+G (12)	26	22	7	10	16	4	20	4	4	4	4	-				
Maximum Likelihood	TN93+G (13)	18	16	9	12	16	10	22	6	6	6	6	8	-			
	Gaps as Fifth State (14)	25	27	24	15	13	23	33	25	25	25	25	21	23	-		
Maximum Parsimony	Complete Deletion (15)	20	20	13	12	14	12	22	16	16	16	16	16	16	19	-	
	Gaps Coded as Binary (16)	27	27	24	13	13	23	33	23	23	23	23	21	21	16	15	-

Table 3. Results of the Shimodaira-Hasegawa (SH) and Approximately Unbiased (AU) tests given by the program CONSEL. Bolded numbers indicate values that fall inside the confidence interval ($p > 0.05$). Log L = Likelihood score of the tree; ΔL = score difference when compared to the best tree. Other abbreviations used are the same of Table 2.

	Trees	log L	ΔL	Rank	p-SH	p-AU
Neighbor-Joining	TN93+G Complete Deletion	-4312.62	33.53	14	0.080	0.005
	TN93+G Pair-wise Deletion	-4312.88	33.79	15	0.067	0.001
	GTR+G	-4280.17	1.08	5	0.961	0.389
	GTR+G and Binary Model	-4296.14	17.05	11	0.343	0.048
	GTR+G; Binary and Doublet Models	-4296.00	16.92	10	0.348	0.059
	GTR+G and Doublet Model	-4280.93	1.84	7	0.905	0.433
Bayesian Inference	TN93+G	-4284.32	5.23	9	0.736	0.125
	TN93+G and RNA6A+G	-4279.08	0.00	1	0.962	0.836
	TN93+G and RNA7A+G	-4279.08	0.00	2	0.962	0.836
	TN93+G and RNA7D+G	-4279.08	0.00	4	0.962	0.836
	TN93+G and RNA16A+G	-4279.08	0.00	3	0.962	0.836
	TN93+G and RNA16D+G	-4280.79	1.70	6	0.905	0.567
Maximum Likelihood	TN93+G	-4282.93	3.85	8	0.879	0.233
	Gaps as Fifth State	-4301.09	22.01	12	0.227	0.057
Maximum Parsimony	Complete Deletion	-4380.16	101.08	16	0.000	0.000
	Gaps Coded as Binary	-4307.79	28.71	13	0.100	0.015

Table 4. Results of the cumulative-Expected Likelihood Weight (ELW) test given by the program Tree-Puzzle. Bolded numbers indicate values that fall inside the confidence interval. S.E. = standard error. Other abbreviations used are the same of Table 3.

	Trees	log L	ΔL	S.E.	c-ELW
Neighbor-Joining	TN93+G Complete Deletion	-4333.90	31.61	13.075	0.0015
	TN93+G Pair-wise Deletion	-4334.57	32.28	11.950	0.0003
	GTR+G	-4302.77	0.48	1.237	0.0862
	GTR+G and Binary Model	-4321.60	19.31	11.346	0.0012
	GTR+G; Binary and Doublet Models	-4321.43	19.14	12.017	0.0091
	GTR+G and Doublet Model	-4303.49	1.20	5.433	0.1402
Bayesian Inference	TN93+G	-4309.04	6.75	4.881	0.0312
	TN93+G and RNA6A+G	-4302.29	0.00	best	0.1200
	TN93+G and RNA7A+G	-4302.29	0.00	0.000	0.1200
	TN93+G and RNA7D+G	-4302.29	0.00	0.000	0.1200
	TN93+G and RNA16A+G	-4302.29	0.00	0.000	0.1200
	TN93+G and RNA16D+G	-4303.27	0.99	5.537	0.1812
Maximum Likelihood	TN93+G	-4306.09	3.81	3.564	0.0543
	Gaps as Fifth State	-4326.18	23.89	14.461	0.0096
Maximum Parsimony	Complete Deletion	-4410.47	108.18	23.584	0.0000
	Gaps Coded as Binary	-4332.66	30.37	14.646	0.0052

Table 5. General evolutionary relationships recovered for Calliphoridae subfamilies. “+” indicates a monophyletic group; “-” indicates a non-monophyletic group. Bolded numbers inside parentheses indicate the bootstrap / posterior probability support when a clade is monophyletic. Abbreviations used are the same of Table 2.

Trees		Monophyletic subfamilies				Relationships between subfamilies
		Calliphorinae	Luciliinae	Chrysomyinae	Mesembrinellinae	
Neighbor-Joining	TN93+G Complete Deletion	+ (100)	-	+ (73)	+ (98)	((Luciliinae,(Chrysomyinae,Calliphorinae))
	TN93+G Pair-wise Deletion	+ (100)	+ (47)	+ (75)	+ (97)	((Luciliinae,(Chrysomyinae,Calliphorinae))
GTR+G		+ (1.00)	+ (0.88)	-	+ (1.00)	((Chrysomyinae,(Luciliinae,Calliphorinae))
GTR+G and Binary Model		+ (1.00)	+ (0.87)	-	+ (1.00)	((Chrysomyinae,(Luciliinae,Calliphorinae))
GTR+G; Binary and Doublet Models		+ (1.00)	+ (0.76)	+ (0.60)	+ (0.98)	((Luciliinae,(Chrysomyinae,Calliphorinae))
GTR+G and Doublet Model		+ (1.00)	+ (0.81)	-	+ (0.86)	((Chrysomyinae,(Luciliinae,Calliphorinae))
Bayesian Inference	TN93+G	+ (1.00)	+ (1.00)	-	+ (1.00)	((Chrysomyinae,(Luciliinae,Calliphorinae))
	TN93+G and RNA6A+G	+ (1.00)	+ (0.79)	-	+ (0.99)	((Chrysomyinae,(Luciliinae,Calliphorinae))
	TN93+G and RNA7A+G	+ (1.00)	+ (0.73)	-	+ (0.99)	((Chrysomyinae,(Luciliinae,Calliphorinae))
	TN93+G and RNA7D+G	+ (1.00)	+ (0.54)	-	+ (0.99)	((Chrysomyinae,(Luciliinae,Calliphorinae))
	TN93+G and RNA16A+G	+ (1.00)	+ (0.78)	-	+ (0.98)	((Chrysomyinae,(Luciliinae,Calliphorinae))
	TN93+G and RNA16D+G	+ (1.00)	+ (0.71)	-	+ (0.99)	((Chrysomyinae,(Luciliinae,Calliphorinae))
Maximum likelihood	TN93+G	+ (99)	+ (38)	+ (60)	+ (94)	((Chrysomyinae,(Luciliinae,Calliphorinae))
Maximum parsimony	Gaps as Fifth State	+ (100)	-	+ (41)	+ (99)	((Luciliinae,(Chrysomyinae,Calliphorinae))
	Complete Deletion	+ (95)	-	-	+ (95)	(Calliphorinae,Luciliinae,Chrysomyinae)
	Gaps Coded as Binary	+ (100)	+ (52)	+ (64)	+ (99)	((Calliphorinae,(Luciliinae,Chrysomyinae))