

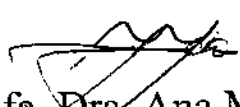
UNIVERSIDADE ESTADUAL DE CAMPINAS

Marcos Túlio de Oliveira

**ESTRUTURA E EVOLUÇÃO DO GENOMA
MITOCONDRIAL NA FAMÍLIA MUSCIDAE (DIPTERA:
CALYPTRATAE)**

Este exemplar corresponde à redação final
da tese defendida pelo(a) candidato (a)
Marcos Túlio de Oliveira
e aprovada pela Comissão Julgadora.

Dissertação apresentada ao Instituto de
Biologia da Universidade Estadual de
Campinas para obtenção do título de
MESTRE em Genética e Biologia
Molecular, na área de Genética Animal e
Evolução.

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

Co-orientadora: Dra. Ana Cláudia Lessinger

Campinas, SP, 2006

UNIDADE	B.C.
Nº CHAMADA	T/ UNICAMP
	OL42e
V	EX
TOMBO BCI	70136
PROC.	16.123/06
C	☐
D	☒
PREÇO	11.00
DATA	25-9-06

BIB ID: 387646

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP**

OL42e

Oliveira, Marcos Túlio de
Estrutura e evolução do genoma mitocondrial na
família Muscidae (Diptera: Calyptratae) / Marcos Túlio de
Oliveira. -- Campinas, SP: [s.n.], 2006.

Orientadora: Ana Maria Lima de Azeredo-Espin.
Co-orientadora: Ana Cláudia Lessinger.
Dissertação (mestrado) – Universidade Estadual de
Campinas, Instituto de Biologia.

1. DNA mitocondrial. 2. Evolução molecular. 3.
Região controle. 4. Muscidae. 5. Diptera. I.
Azeredo-Espin, Ana Maria Lima de. II. Lessinger, Ana
Cláudia. III. Universidade Estadual de Campinas.
Instituto de Biologia. IV. Título.

(rcdt/ib)

Título em inglês: Structure and evolution of the mitochondrial genome in family Muscidae (Diptera: Calyptratae).

Palavras-chave em inglês: Mitochondrial DNA; Molecular evolution; Control region; Muscidae; Diptera.

Área de concentração: Genética Animal e Evolução.

Titulação: Mestre em Genética e Biologia Molecular.

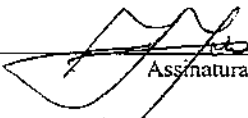
Banca examinadora: Ana Maria Lima de Azeredo-Espin, Cláudia Augusta de Moraes Russo, Sérgio Furtado dos Reis.

Data da defesa: 12/06/2006.

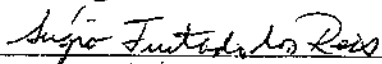
Campinas, 20 de junho de 2006.

BANCA EXAMINADORA

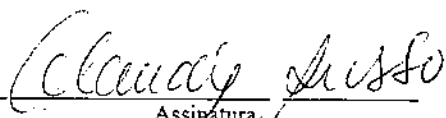
Profa. Dra. Ana Maria Lima de Azeredo-Espin (Orientadora)


Assinatura

Prof. Dr. Sérgio Furtado dos Reis


Assinatura

Profa. Dra. Cláudia Augusta de Moraes Russo


Assinatura

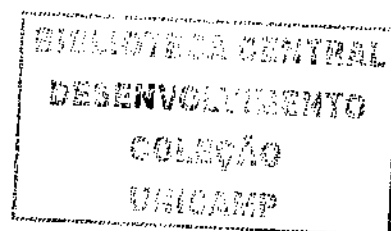
Prof. Dr. Louis Bernard Klaczko

Assinatura

Profa. Dra. Maria Cristina Arias

Assinatura

2006.022413



Por todo trabalho, dedicação, paciência, amor, amizade, determinação, altruísmo, disposição, preocupação, sacrifício, alegria, bondade, educação, fidelidade, compaixão e veracidade.

Por tudo isso, por tudo em que tento me inspirar, é a você, mãe, que dedico esta tese.

“E então Charles Darwin entrou em cena. Em 1859, foi publicada sua Origem das espécies, anunciando ao mundo a inferência cientificamente deduzida de que o homem não fora criado por Deus, mas evoluíra de simples animais pela seleção natural.”

Margulis & Sagan (2002).

“Nothing in Biology makes sense except in the light of evolution.”

Dobzhansky (1970).

*"If one could design a customized genetic marker for use in population and evolutionary biology,
it is quite likely that one would reinvent mitochondrial DNA."*

Rand (1994).

*"Mitochondria are ubiquitous throughout the eukaryotic domain: with very few exceptions,
eukaryotes cannot exist without them."*

Burger et al. (2003).

AGRADECIMENTOS

Primeiramente, gostaria de agradecer àqueles que guiaram os rumos dos meus experimentos, dos meus conhecimentos, da minha formação e conseqüentemente desta tese. Agradeço enormemente à Dra. Ana Cláudia Lessinger, por toda dedicação e confiança depositada em mim ao longo destes quase cinco anos de convívio. Agradeço pela ajuda nos momentos mais embaraçados, pela luz nos momentos de maior escuridão, pela cobrança nos momentos de frouxidão (e às vezes até nos momentos em que eu estava mais sobrecarregado), pelo entusiasmo a cada resultado significativo, pelo conhecimento que tentei absorver de nossas longas conversas (e espero ter conseguido pelo menos um pouco), pela preocupação com o meu destino como biólogo. Obrigado por me ajudar a ser cientista, a ser biólogo, a ser geneticista, a amar o que faço. Obrigado também à Laurinha e ao Paulão, por cederem a atenção da ainda grávida e da recém mamãe Ana Cláudia a um aluno aparentemente e verdadeiramente desesperado com a correção de seu projeto de mestrado.

Agradeço também à Prof. Dra. Ana Maria L. de Azeredo-Espin, pelo apoio e sugestões nas tomadas de decisões, pela confiança que vem desde o primeiro contato na graduação, pela oportunidade de participar como monitor na BG-380, pela preocupação com os rumos dos experimentos no laboratório e por sempre garantir que eles continuariam mesmo em períodos de “entressafra” de projetos.

Agradeço à membro da banca, a Prof. Dra. Cláudia A. M. Russo, com seu estilo espontâneo e sincero, pelos comentários e sugestões pertinentes durante a pré-banca e banca para o enriquecimento desta tese.

Agradeço ao membro da banca, o Prof. Dr. Sérgio F. dos Reis, sempre bem educado e prestativo, pelos comentários referidos à tese e pela oportunidade, junto à Dra. Ana Cláudia R. Alves, de colaboração em outros projetos.

Agradeço a todos os professores que participaram da minha formação durante o curso de Ciências Biológicas e durante o Mestrado, mas em especial, aos professores Dr. Louis Bernard Klaczko, por suas reflexões demoradas de assuntos às vezes complexos, que ele conseguia transpassar através de palavras claras, pausadamente, e que eu adorava ficar ouvindo por horas; ao Dr. Michel G. A. Vincentz, por saber instigar a inteligência e criatividade de seus alunos, com seu sotaque carregado e seu jeito descontraído; e ao Dr. Hernandes F. Carvalho, pela

oportunidade de desenvolver um estudo paralelo à minha tese, em que adquiri conhecimento numa área bem diferente, mas não por isso, menos interessante para mim.

Agradeço a todos os colegas do Laboratório de Genética Animal, aos atuais e aos que já partiram, pela colaboração nos experimentos, pelas dúvidas esclarecidas, pelas histórias compartilhadas, pelas gargalhadas nas horas inapropriadas (dentro do lab!) e apropriadas (num bar ou cinema!). Obrigado Norma, Karina, Jaú, Renato, Mari (Loira), Aline, Pedro, Ana Carolina, Tatiana, Joan, Ize, Patrícia, Fábio, Luciane, Prizinha, Bárbara, Salete, Mário e Lissiene (que o Canadá a tenha!). Em especial, gostaria de agradecer à Mari (Morena), por conduzir meus primeiros passos no laboratório e pelas inesquecíveis e aventureiras coletas em terras nacionais e internacionais; ao Johnny (Gustavo?!), por seus pensamentos sem sentido e pela companhia à viagem para Flórida; ao Palocci (ex-ministro e atual aluno do curso de Biologia da Unicamp), por suas brincadeiras e piadas fora de hora e pela companhia nos jogos vitoriosos do Tricolor; e à Rosângela, pelas dúvidas esclarecidas, pela ajuda experiente no laboratório, a qualquer dia e qualquer hora, e por mostrar como um funcionário público deveria trabalhar.

Agradeço aos amigos uruguaio Pablo Fresia, Santos Gama e sua esposa Patrícia, Stella Lanzeri, Estela Martines e Arnaldo Maggi, pelas desbravadoras coletas durante o verão escaldante do Uruguai e por me mostrarem um grande pequeno país.

Agradeço ao pessoal “lá de casa”: às meninas, Flávia, Fefs e Gabi, por estarem sempre alegres e dispostas; à Iracema, por falar demais nos momentos inadequados, por cuidar da casa e da gente e pela honestidade e sinceridade; à Déa, por sua liderança natural e também por cuidar do folgado do Clau; à Prizinha (de novo!) e ao Plínio, pelas refeições deliciosas que gostam de preparar; ao Jaú (de novo!), pela aquisição do Play Station 2, o que nos fez perder milhares de horas de concentração em deveres e compromissos (inclusive na redação desta tese) e, conseqüentemente, gastá-las na frente da TV; à Taís, por suas manias e tapas doídos; à Priscila, por sua preguiça contagiante; ao Edge e Lutor, por estarem sempre dispostos a brincar e correr atrás das bolinhas e pela felicidade estampada em seus rabinhos devido à chegada de um da casa; ao Kiri, por iniciar nossos dias com seus lindos assovios; ao Márcio, por ser um excelente companheiro de quarto, pela inteligência nas nossas discussões, pela cumplicidade e companheirismo, pelas chatices, mal-humores, atrasos e enrolações e pela dedicação ao trabalho, na qual tento me espelhar; e ao Clau, pelas muitas horas gastas com reflexões sobre a vida, sobre a biologia, sobre a evolução e sobre os rumos da vida acadêmica, pelas muitas horas gastas

jogando Winning Eleven, discutindo sobre os melhores jogadores de futebol do jogo e ajudando, desta forma, a me desconcentrar de minha tese (e a recíproca é verdadeira), pela cumplicidade e companheirismo e por tudo que passamos juntos.

Agradeço aos amigos que fiz na turma Bio00, Alice, Rafael, Rodrigo, Marcelo, Marcela, Garça, Barro, Léo, Daniel, Felipe, Bárbara, Robertinha, Paula, Mirela, Suzana, Antônio Marcos, Valéria, Genésio e Lissiene (de novo).

Agradeço aos “cara” do handebol, pela união e dedicação ao time da Escócia e por mostrar aos outros times da Unicamp o que é ter raça. Obrigado Marcão, Diego, Java, Vareta, Santoro, Bereba, Banana, Fabiano, Diler, Osório, Alessandro, Paraíba, Fipo, Lucão, Palocinho, Frango, Hugo e os professores Rafa e Carla.

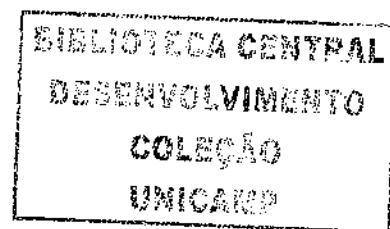
Agradeço à Juliana, por ter sido minha eterna companheira até boa parte do trajeto do mestrado.

Agradeço aos amigos de infância da velha São Sebastião do Paraíso, Marcell, Diego, Bruno, Maurício e Tião, por serem “pau para toda obra”.

Agradeço à minha família, por estarem sempre por perto quando precisei. Aos meus irmãos, Teté e Paçoca, por estarmos sempre juntos, pela cumplicidade e pelas alegrias e tristezas compartilhadas. Ao meu pai, Zé Goiaba, pelo conhecimento geral que sempre tentou nos transmitir. À minha mãe, Vilma, por TUDO que faz por nós, sem descartar um momento sequer. E aos meus tios, primos e avós, por sempre acreditarem que um dia eu seria doutor (com certeza serei, mas não doutor de gente!).

Agradeço a toda comunidade do CBMEG, principalmente ao pessoal do laboratório de Análises Genética e Molecular (Prof. Dra. Anete P. Souza), ao pessoal do laboratório de Genoma (Prof. Dr. Paulo Arruda) e ao pessoal da Administração, pelo empréstimo de máquinas, material e reagentes, pela paciência e pela ajuda na parte técnica e burocrática.

Agradeço aos apoios financeiros da CAPES, pela bolsa concedida durante meu mestrado; do CNPq, através do programa PROFIX da Dra. Ana Cláudia, pela bolsa de iniciação científica e pela maioria dos recursos gastos neste projeto; da FAPESP, pelo auxílio a projetos temáticos da Prof. Dra. Ana Maria L. Azeredo-Espin; e à FAEPEX, pelo auxílio concedido à participação na Reunião Anual da Sociedade Entomológica da América (ESA), realizada em Fort Lauderdale, EUA.



ÍNDICE	PÁGINA
RESUMO	xii
ABSTRACT	xiv
INTRODUÇÃO	
O DNA Mitocondrial Animal	1
O DNA Mitocondrial em Insetos	4
O gene <i>COI</i>	5
A região-controle	6
A família Muscidae	9
CAPÍTULOS - APRESENTAÇÃO	12
CAPÍTULO 1 - Resumo	13
Artigo - “Evolutionary and structural analysis of the cytochrome c oxidase subunit I (COI) gene from <i>Haematobia irritans</i> , <i>Stomoxys calcitrans</i> and <i>Musca domestica</i> (Diptera: Muscidae) mitochondrial DNA.”	14
Metodologia detalhada	19
CAPÍTULO 2 - Resumo	23
Artigo - “Improving access to the control region and tRNA gene clusters of Dipteran mitochondrial DNA.”	24
Metodologia detalhada	28
CAPÍTULO 3 - Resumo	32
Manuscrito - “The mitochondrial DNA control region of Muscidae flies: evolution and structural conservation in a Dipteran context.”	33
Metodologia detalhada	58
CAPÍTULO 4 - Resumo	62
Manuscrito - “The mitochondrial genomes of <i>Haematobia irritans</i> and <i>Stomoxys calcitrans</i> : structure and evolution in Muscidae flies (Diptera: Calyptratae)”	63
Metodologia detalhada	97
DISCUSSÃO	
O gene COI em dípteros	101

Acesso a regiões de instabilidade estrutural do DNAm	102
Controle da replicação/transcrição?	103
Organização estrutural evolução do genoma mitocondrial em Muscidae	105
A evolução dos anticódons	107
CONSIDERAÇÕES FINAIS E PERSPECTIVAS	110
REFERÊNCIAS BIBLIOGRÁFICAS	111

RESUMO

O DNA mitocondrial (DNAm_t) dos metazoários possui uma série de características, tais como organização circular, simples e compacta, conteúdo gênico conservado, herança predominantemente materna e ausência ou baixa taxa de eventos de recombinação, que fazem com que este genoma seja amplamente usado em diversos estudos evolutivos. Geralmente, o DNAm_t animal possui 37 genes (13 codificadores de proteínas, 22 para RNAt e 2 para RNAr) e uma principal região não-codificadora, a região-controle. O objetivo do projeto foi o seqüenciamento e a caracterização estrutural e evolutiva de regiões gênicas e do genoma mitocondrial de espécies da família Muscidae (Diptera: Calyptratae), incluindo a mosca-dos-chifres, *Haematobia irritans*, a mosca doméstica, *Musca domestica*, a mosca menor das frutas, *Atherigona orientalis*, “the black dump fly”, *Ophyra aenescens*, e a mosca-dos-estábulo, *Stomoxys calcitrans*. A estrutura da tese foi dividida em quatro capítulos. No primeiro capítulo, a seqüência completa do gene da subunidade I da citocromo oxidase c (COI) foi caracterizada. COI é o gene mais seqüenciado entre os insetos e tem sido usado em diversos estudos evolutivos e taxonômicos. Este gene mostrou-se estruturalmente semelhante aos demais dípteros, incluindo o uso de um códon de iniciação não usual, TCG (serina). O segundo capítulo descreve um novo conjunto de oligonucleotídeos iniciadores que ajudou a amplificar e seqüenciar regiões do DNAm_t de vários dípteros onde há alta chance de ocorrer rearranjos e/ou duplicações (regiões de “hot spots” ou “instabilidade estrutural” no DNAm_t de artrópodes, que incluem a região-controle e agrupamentos de RNAt). O terceiro capítulo descreve a seqüência completa da região-controle de *H. irritans*, *M. domestica* e *A. orientalis*, define padrões evolutivo-estruturais para moscas da subordem Brachycera e sugere a atuação de elementos *cis*-regulatórios no controle da replicação e transcrição do DNAm_t de insetos. Ao tentar acessar a região-controle de *O. aenescens* e *S. calcitrans*, foi evidenciado um possível evento de rearranjo e/ou duplicação no genoma mitocondrial destas espécies envolvendo os genes que flanqueiam a região-controle. O quarto capítulo descreve uma análise comparativa entre os genomas mitocondriais de *H. irritans* (16078bp) e *S. calcitrans* (~18Kb). Um conjunto de estratégias, incluindo amplificação via Long-PCR, construção de bibliotecas de “shotgun”, montagem, anotação e análises estruturais utilizando recursos de bioinformática, foi otimizada durante este trabalho. Os genomas mitocondriais de *H. irritans* e *S. calcitrans* possuem conteúdo e organização gênicos

conservados, considerando-se o modelo ancestral proposto para artrópodes, exceto pela região-controle de *S. calcitrans* (~2,5Kb), que teve 60% de sua sequência caracterizada, suficientes para indicar a ocorrência de duplicação gênica envolvendo elementos internos da região-controle e o gene de RNAt para isoleucina (I). A duplicação de I no DNAm de *S. calcitrans* é muito semelhante a que ocorre no DNAm de espécies de *Chrysomya* (Diptera: Calliphoridae), reforçando que esta região é um “hot spot” de duplicação no genoma mitocondrial de dípteros. Uma árvore filogenética usando os genes codificadores de proteínas foi inferida para a ordem Diptera e o resultado difere da filogenia tradicional proposta para o grupo.

ABSTRACT

The mitochondrial DNA (mtDNA) of Metazoan has features, such as circular, simple and compact organization, conserved gene content, predominantly maternal inheritance, and lack or low rate of recombination, which have made this genome widely used in a range of evolutionary studies. In general, animal mtDNA presents 37 genes (13 protein-coding genes, 22 genes for tRNA and 2 for rRNA) and a major non-coding region, the control region. The aim of this project was the sequencing and the structural and evolutionary characterization of gene regions and the mitochondrial genome of species from family Muscidae (Diptera: Calyptratae): the horn fly, *Haematobia irritans*; the house fly, *Musca domestica*; the pepper fruitfly, *Atherigona orientalis*; the black dump fly, *Ophyra aenescens*; and the stable fly, *Stomoxys calcitrans*. The structure of the thesis was divided in four chapters. In the first chapter, the complete sequence of the cytochrome c oxidase subunit I (COI) gene was characterized. COI is the most sequenced gene in insects and have been used in a wide range of evolutionary and taxonomic studies. This gene in Muscidae showed a structure similar to other dipterans, including the use of an uncommon start codon, TCG (serine). The second chapter describes a new set of primers that improved the amplification and sequencing of regions from many dipterans mtDNA in which there is a high probability of rearrangements and/or duplications occur (regions involved in structural “hot spots”, including the control region and tRNA clusters). The third chapter describes the complete sequence of *H. irritans*, *M. domestica* and *A. orientalis*, defines evolutionary and structural patterns for Brachyceran flies and suggests *cis*-regulatory elements for the mtDNA replication and transcription control in insects. Trying to access the control region of *O. aenescens* and *S. calcitrans*, a possible event of rearrangement and/or duplication in the mitochondrial genome of these species was evidenced involving the control region flanking genes. The fourth chapter describes a comparative analysis of *H. irritans* (16078bp) and *S. calcitrans* (~18Kb) mitochondrial genomes. A set of strategies, including the amplification via Long-PCR, construction of a shotgun library, assemblage, annotation and structural analysis using bioinformatic tools, was optimized in this work. The mitochondrial genomes of *H. irritans* and *S. calcitrans* showed gene content and organization conserved regarding the ancestral model proposed to arthropods, except for *S. calcitrans* control region (~2.5Kb), which had 60% of its sequence characterized. This characterization indicated the occurrence of duplicated regions

involving control region internal elements and the tRNA gene for isoleucine (I). The duplication of I is very similar to that of *Chrysomya* (Diptera: Calliphoridae) mtDNA, reinforcing that this region is a duplication hot spot in the mitochondrial genome of dipterans. A phylogenetic tree using the protein-coding genes was inferred for the Diptera order and the results differ from the traditional phylogeny proposed for the group.

INTRODUÇÃO

O DNA Mitocondrial Animal

Desde a descoberta nos anos 1960 de que, além do núcleo, algumas organelas celulares - os cloroplastos e as mitocôndrias - também carregavam um material genético próprio, começou-se a investigar os diferentes aspectos deste DNA, tentando inclusive compreender a origem destas organelas e dos seres eucariotos. Este achado fortaleceu a teoria da origem endossimbiótica das mitocôndrias e dos plastídeos (Sagan 1967) e permitiu que os DNAs organelares fossem bem caracterizados estrutural, funcional e evolutivamente, conforme surgiam avanços técnico-metodológicos com o passar dos anos.

Dentre os genomas organelares, o DNA mitocondrial (DNAm) é um dos mais bem estudados entre os seres vivos. O conhecimento acerca do conteúdo e da organização do genoma mitocondrial foi um importante ponto de partida para se começar a compreender os processos evolutivos peculiares a este sistema. Entre os seres vivos, os animais são o grupo que possui mais espécies cujos genomas mitocondriais estão completamente seqüenciados, acumulando um total de 757 representações no banco de dados GenBank (NCBI - Maio/2006). Além disso, existem mais de 350.000 seqüências parciais ou completas relacionadas a genes mitocondriais de espécies animais. O sucesso nesta área, conhecida atualmente como mitogenômica ou genômica mitocondrial, é consequência de dois principais fatores: as propriedades do DNA mitocondrial animal e os avanços tecnológicos atuais da biologia molecular.

O primeiro destes fatores mostra o que levou e continua levando o DNA mitocondrial animal a ser seqüenciado extensivamente para um número cada vez maior de espécies (as propriedades intrínsecas da molécula). Este genoma possui uma série de características, tais como organização circular, simples e compacta, herança predominantemente materna, ausência ou baixa taxa de recombinação e alta taxa de substituição nucleotídica, que tem favorecido seu uso como marcador molecular em estudos evolutivos, estruturais, filogenéticos e populacionais, abrangendo uma ampla diversidade de *taxa* (Wolstenholme 1992, Saccone *et al.* 1999, Boore 1999, Avise 2004, Boore *et al.* 2005). O conteúdo gênico do DNA mitocondrial é bem conservado em Metazoa [geralmente 13 genes codificadores de proteínas (CDS), 2 genes para RNAs ribossômicos (RNAr), 22 genes para RNAs de transferência (RNAt), e uma região não codificadora, a região-controle], mas rearranjos na ordem dos genes têm sido descritos à medida

que novos genomas mitocondriais são caracterizados (revisão em Boore 1999 e Saccone *et al.* 1999).

O segundo fator refere-se principalmente ao desenvolvimento e melhorias das técnicas e metodologias relacionadas à obtenção de seqüências, à genômica e, mais especificamente, à mitogenômica. Primeiramente, com o advento da técnica de PCR (“polymerase chain reaction”), houve um grande aumento na velocidade de geração de seqüências de DNA e uma maior facilidade para seleção de regiões gênicas específicas (Innis *et al.* 1990). O uso de oligonucleotídeos iniciadores (“primers”) chamados *universais* (Kocher *et al.* 1989, Simon *et al.* 1994) permitiu o seqüenciamento de regiões do DNAm de espécies para a qual nenhuma informação sobre seqüências nucleotídicas existia. O sucesso de tais conjuntos de “primers” *universais* é tão evidente que cerca de quinze anos após suas publicações, o trabalho de Kocher *et al.* (1989) já foi citado 1849 vezes e o de Simon *et al.* (1994), 994 vezes (Web of Science - Maio/2006).

Atualmente, além do uso de regiões específicas do DNAm animal em diversos estudos evolutivos, tem se tornado crescente o uso de genomas mitocondriais completos principalmente para abordar questões filogenéticas, de estrutura e de mecanismos de evolução neste genoma. Boore *et al.* (2005) argumenta que a comparação de genomas mitocondriais completos de animais não somente é bem mais informativa que o uso de seqüências curtas e parciais, mas que estes dados também fornecem um conjunto de “caracteres para análise em nível genômico”, tais como estruturas secundárias de RNAs, formas de controle da transcrição e replicação, estruturas físicas do DNAm e, especialmente, o arranjo relativo dos genes no genoma, que pode ser um indicador de ancestralidade comum. Para obtenção do DNAm completo, diferentes métodos têm sido usados, mas três deles se destacam: o isolamento físico do DNAm do DNA nuclear através do uso de ultra-centrifugação em gradiente de cloreto de cério, aproveitando as diferenças em massa e propriedades de moléculas circular e linear (Sambrook *et al.* 1989, Azeredo-Espin *et al.* 1991); o RCA (“Rolling Circle Amplification”), adaptado da genômica tradicional para a mitogenômica, que é capaz de produzir *in vitro* múltiplas cópias em “tandem” de moléculas de DNA circular (Boore *et al.* 2005, Simison *et al.* 2006); e a amplificação do genoma mitocondrial via PCR de longa extensão (Long-PCR). Até pouco tempo atrás, o PCR era capaz de amplificar somente fragmentos de poucos kilobases. Entretanto, metodologias para amplificação de regiões bem maiores estão disponíveis atualmente, incluindo amplificações de genomas mitocondriais

completos em três (+/- 6Kb cada) ou dois (+/- 9Kb cada) fragmentos cujas extremidades se sobrepõem, ou ainda em uma única amplificação (+/- 16Kb) (Roehrdanz 1995, Miya & Nishida 1999, Sorenson *et al.* 1999, Hwang *et al.* 2001, Roehrdanz *et al.* 2002, Boore *et al.* 2004, Yamauchi *et al.* 2004, Barau *et al.* 2005). É notável como a técnica de Long-PCR tem sido empregada com bastante sucesso no desenvolvimento da genômica mitocondrial de artrópodes. Há pouco tempo atrás, uma grande quantidade de DNA inicial era necessária para o seqüenciamento de genomas mitocondriais completos, o que limitava a caracterização deste genoma em organismos de tamanho muito pequeno, como a maioria das espécies de artrópodes. Assim, com o desenvolvimento do Long-PCR, foi possível contornar esta situação e obter quantidade e qualidade suficientes de DNAm de diversas espécies deste filo, que passou de 25 representantes com genomas mitocondriais completos no ano de 2001 (Lessinger 2001) para 108 em 2006 [Arthropodan Mitochondrial Genomes Accessible database (AMiGA, Feijão *et al.* 2006)].

Para o avanço na caracterização de genomas mitocondriais, não bastou apenas a aplicação de novas técnicas de obtenção do DNAm, mas também o desenvolvimento de ferramentas mais específicas para análises comparativas das seqüências. Nos últimos dez anos, bancos de dados especializados na compilação e análise de seqüências mitocondriais foram criados e são de amplo acesso à comunidade científica através da internet. O mais antigo deles, disponível desde 1996, é o GOBASE (<http://megasun.bch.umontreal.ca/gobase>), cujo principal enfoque está relacionado a estudos de mitogenômica comparativa (Shimko *et al.* 2001). O banco de dados GenBank do NCBI possui uma interface para acesso a informações relativas a genomas organelares, incluindo recursos específicos para genomas mitocondriais de metazoários (http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/mztax_short.html). Além destas opções, o Joint Genome Institute (JGI) fornece uma interface para comparações de arranjos gênicos estruturais de genomas mitocondriais no “link” Organelle Genomes (http://evogen.jgi.doe.gov/top_level/organelles.html), onde é possível ter acesso a ferramentas de análises de seqüências para identificação do conteúdo de aminoácidos e nucleotídeos, uso dos códons, viés fita-específica e da estrutura secundária predita de RNAt. Além disso, o desenvolvimento de uma interface para anotação de genomas mitocondriais (e de cloroplastos), o DOGMA - Dual Organellar GenoMe Annotator (Wyman *et al.* 2004), otimizou as análises em mitogenômica, automatizando a anotação de genomas organelares, através da identificação dos

diferentes elementos genômicos (CDS, RNAt ou RNAr) e possibilitando a visualização gráfica, a edição manual da anotação e a submissão do genoma ao GenBank.

Inovações como a submissão de genomas protegidos por senha e a inclusão de ferramentas de análises filogenéticas associados a bancos de dados mitocondriais são recentes iniciativas brasileiras, que contribuem para a otimização de análises comparativas. Com o enfoque em dar suporte para estudos filogenéticos em mamíferos utilizando dados de genomas mitocondriais, pesquisadores da UFRJ e do LNCC criaram o MamMiBase - Mammalian Mitochondrial Genomics Database (Vasconcelos *et al.* 2005). Para artrópodes, nosso grupo de pesquisa (Laboratório de Genética Animal, CBMEG, UNICAMP) desenvolveu o AMiGA - Arthropodan Mitochondrial Genomes Acessible database (Feijão *et al.* 2006) com a proposta principal de promover análises comparativas estruturais de genomas mitocondriais neste diverso grupo de animais.

O DNA Mitocondrial em Insetos

Historicamente, o genoma mitocondrial de insetos tem uma importância fundamental para os estudos sobre estrutura e evolução do DNAm_t animal. Até o ano de 1985, apenas o DNAm_t humano e do anfíbio *Xenopus laevis* estavam completamente seqüenciados entre os animais. Foi com a publicação da seqüência completa do DNAm_t de *Drosophila yakuba* (Clary & Wolstenholme 1985) que as características que agrupam o genoma mitocondrial dos metazoários puderam ser observadas. Estabeleceu-se, a partir deste ponto, um marco na investigação de genes e genomas mitocondriais em invertebrados. Este genoma até hoje serve de modelo para comparação de conteúdo e organização gênicos e taxas evolutivas em outras espécies de insetos e de artrópodes em geral.

Além das características comuns ao DNAm_t animal, em insetos, o genoma mitocondrial possui um marcante enviezamento na composição nucleotídica caracterizado pela elevada quantidade de adenilato (A) e timidilato (T) (revisão em Saccone *et al.* 1999). Wolstenholme (1992) sugere que esta característica deva ser uma solução alternativa à exigência de uma demanda energética alta provocada pelo metabolismo acelerado dos insetos e refletida numa maior eficiência dos processos de replicação e transcrição do genoma mitocondrial. Jermini *et al.* (1994) sugere que a alta porcentagem destes nucleotídeos no DNAm_t de insetos é resultado de um efeito de pressão de mutação direcional ocorrido neste genoma. Dentre as regiões do genoma

mitocondrial de insetos, duas delas, que são tratadas nos próximos tópicos, se destacam por serem respectivamente bem estudada (o gene COI) e funcionalmente importante, mas pouco conhecida (a região-controle).

O gene COI

Regiões específicas do DNAm^t evoluem a diferentes taxas evidenciando padrões evolutivos distintos neste genoma. Em análises evolutivas baseadas na comparação de seqüências ortólogas do DNAm^t, a escolha da região gênica a ser usada é crucial (Simon *et al.* 1994; Lunt *et al.* 1996), uma vez que a sua utilidade como marcador depende do nível taxonômico estudado. Entre os genes codificadores de proteínas do genoma mitocondrial, o mais utilizado em estudos evolutivos em insetos é o gene da subunidade I do complexo citocromo oxidase c (COI) (Zhang & Hewitt 1997; Caterino *et al.* 2000). Sendo um dos maiores genes do DNAm^t de insetos e apresentando diferentes regiões estruturais relacionadas à função da proteína, taxas de substituição nucleotídica diferenciadas ao longo da seqüência do gene COI podem ser evidenciadas. Assim é possível identificar seqüências com variabilidade genética adequada para a investigação de uma determinada questão evolutiva (Lunt *et al.* 1996).

O gene COI tem sido apresentado como um dos genes mitocondriais mais conservados tanto entre os mamíferos (Saccone *et al.* 1999) quanto entre os insetos (Beard *et al.* 1993; Mitchell *et al.* 1993; Crozier & Crozier 1993; Lessinger *et al.* 2000), e suas aplicações como marcador molecular em dípteros têm sido descritas considerando-se diferentes abordagens: filogenética (Wells & Sperling 1999; Bernasconi *et al.* 2000; Stevens *et al.* 2002), forense (Wells & Sperling 2001; Wells *et al.* 2001; Wallman & Donnellan 2001), diagnóstica (Otranto *et al.* 2000; Litjens *et al.* 2001; Van Bortel *et al.* 2002) e, até mesmo, populacional (Brito *et al.* 2002).

Recentemente, tem havido um grande esforço da comunidade científica em produzir um inventário, baseado em uma seqüência de DNA, das espécies existentes no planeta (www.barcodinglife.org). Este projeto, que visa identificar e catalogar a biota atual da Terra, é chamado “Código de Barras da Vida” (Barcode of Life) e tem sido bastante discutido quanto a sua aplicabilidade (Moritz & Cicero 2004, Ebach & Holdrege 2005, Gregory 2005, Scheffer *et al.* 2006). O gene COI mostrou-se muito eficaz na identificação espécie-específica de uma gama enorme de espécies animais (Hebert *et al.* 2003) e foi a região genômica escolhida (648pb da região 5’) para o desenvolvimento deste sistema global de bioidentificação.

A Região-Controle

Ao contrário do observado em COI, certas regiões do DNAm_t de insetos são pouco compreendidas, tanto no aspecto evolutivo quanto no funcional. Este é o caso da principal região não-codificadora, também chamada de região rica em A+T em insetos. Para algumas espécies animais, principalmente mamíferos, esta região mostrou conter elementos que controlam a transcrição de genes mitocondriais e a replicação deste genoma, recebendo, assim, o nome de região-controle (revisão em Shadel & Clayton 1997, Taanman 1999). Para a maioria das espécies, assume-se que a maior região não-codificadora é a região-controle (Boore 1999). A natureza estrutural desta região é extremamente complexa, pois sua evolução se dá tanto por substituições nucleotídicas quanto por eventos de inserção e deleção, envolvendo duplicações e translocações e caracterizando-se como a região mais variável do genoma mitocondrial de insetos (Simon *et al.* 1994; Zhang & Hewitt 1997). Além disso, a análise da composição nucleotídica desta região mostra um forte enviezamento a favor do acúmulo de A e T ainda mais evidente que no resto do genoma (Jermiin *et al.* 1994) - de 75 a 96% (Crozier & Crozier 1993; Lewis *et al.* 1994; Mardulyn *et al.* 2003).

A variação em tamanho da região-controle de insetos é bastante ampla, apresentando desde aproximadamente 70pb no piolho *Heterodoxus macropus* (Shao *et al.* 2001) até 130Kb em coleópteros (Boyce *et al.* 1989). Mesmo entre espécies filogeneticamente próximas, a variação em tamanho é relativamente freqüente, encontrando-se regiões que vão de 930 a 4601pb em espécies do gênero *Drosophila* (Lewis *et al.* 1994; Brehm *et al.* 2001). Trechos de seqüências repetidos em série (ou em “tandem”) estão comumente relacionados com a diferença de tamanho entre a região-controle de diferentes espécies de insetos (Zhang & Hewitt 1997) e também estão relacionados à grande maioria da variação em tamanho encontrada entre genomas mitocondriais completos de espécies animais (Boore 1999). Evidenciando esta informação, o Gráfico 1 mostra a relação entre tamanho da região-controle e tamanho do DNAm_t completo (bp) de todas as espécies de insetos registradas no AMiGA (Março/2006). Além disso, a variação entre genomas mitocondriais completos pode envolver a ocorrência de mais de uma região-controle no genoma mitocondrial (revisão em Shao *et al.* 2005) e o abundante número de rearranjos gênicos envolvendo genes de RNAt que flanqueiam esta região (revisão em Boore 1999), sugerindo que a região-controle está ativamente associada a instabilidades estruturais no DNAm_t.

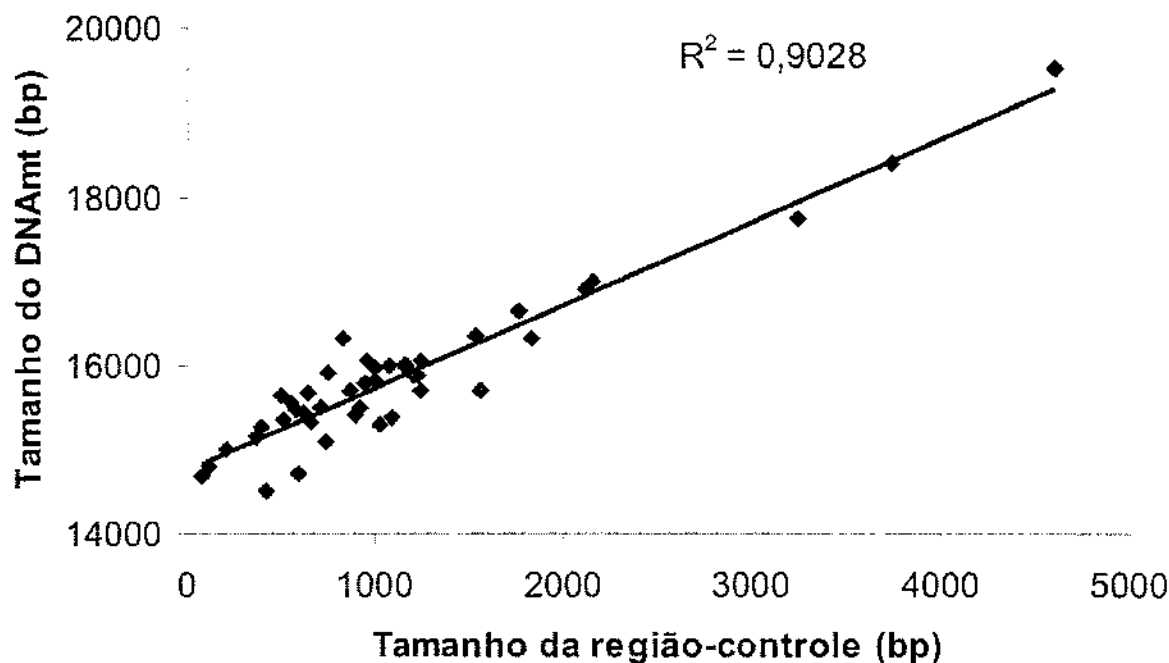


Gráfico 1. Relação entre tamanho do DNAm completo e tamanho da região controle em insetos.

Variações no número de repetições em “tandem” são comumente encontradas em diferentes indivíduos de uma mesma espécie (Dotson & Beard 2001; Snäll *et al.* 2002; Mardulyn *et al.* 2003; Lyra *et al.* 2005) e, somadas a alta variabilidade encontrada na região-controle, podem favorecer o uso desta região em estudos intra-específicos (Zhang & Hewitt 1997), fornecendo um número significativo de caracteres que possam ser informativos na detecção da diferenciação entre populações. Recentes trabalhos mostram a eficiência do uso da região-controle como marcador molecular em análises de estrutura populacional em artrópodes (Chu *et al.* 2003; McMillen-Jackson & Bert 2003; Vandewoestijne *et al.* 2004; Vila & Björklund 2004; Lyra *et al.* 2005). Por outro lado, eventos associados à expansão ou redução de unidades repetidas são comuns à região-controle e devem ser analisados com cautela em estudos com marcadores moleculares, devido à possibilidade da ocorrência de convergência estrutural e, consequentemente, homoplasia.

Além disso, a natureza hiper-variável da região-controle de insetos faz com que seja difícil o reconhecimento da homologia nas seqüências de diferentes espécies. Mesmo assim, Zhang & Hewitt (1997) definiram estruturas homólogas freqüentemente observadas na região-

controle de insetos: longos trechos de seqüências poli-T, seqüências [TA(A)]_n repetidas em série, estruturas secundárias em forma de grampo (“stem and loop”) e curtos trechos ricos em G+A. Para dípteros caliptrados, a caracterização da região-controle de sete espécies identificou oito blocos de seqüência conservada compartilhados entre as famílias Calliphoridae e Oestridae (Lessinger & Azeredo-Espin 2000; Lessinger *et al.* 2004). Estes elementos estruturais conservados descritos na região-controle podem estar envolvidos na regulação do DNAm_t (Zhang & Hewitt 1997). Apesar de ser um grupo numeroso e diversificado, pouco se sabe sobre os mecanismos que regulam os processos de replicação e transcrição do DNAm_t de insetos, embora grandes avanços na caracterização de fatores nucleares envolvidos nestes processos estão sendo realizados no organismo modelo *Drosophila melanogaster* (Kaguni 2004, Garesse and Kaguni 2005). A caracterização das seqüências nucleotídicas da região-controle de um número maior de espécies associada à caracterização destes fatores nucleares pode contribuir na compreensão destes mecanismos e na análise da evolução e função desta região.

O atual avanço na caracterização de genomas mitocondriais completos de espécies de insetos tem algumas vezes encontrado dificuldades técnico-metodológicas devido a instabilidades estruturais deste genoma relacionadas, principalmente, à região-controle e aos genes que a flanqueiam (Boore 1999; Shao *et al.* 2005). Dos 49 genomas mitocondriais completos de insetos disponíveis no AMiGA (Março/2006), 8 não apresentam a seqüência referente à região-controle. Esta é a única região cuja seqüência encontra-se ausente num número significativo de genomas mitocondriais de insetos. Estes dados são um reflexo das dificuldades associadas à caracterização da seqüência nucleotídica desta região do DNAm_t, comprometendo a conclusão do seqüenciamento completo de um genoma. Uma vez caracterizada esta região, a obtenção da seqüência completa do genoma mitocondrial pode ser otimizada, assim como foi realizado nos projetos de seqüenciamento completo dos genomas mitocondriais da mosca-da-bicheira, *Cochliomyia hominivorax* (Lessinger & Azeredo-Espin 2000, Lessinger *et al.* 2000), da mosca-do-berne, *Dermatobia hominis* (Lessinger & Azeredo-Espin 2000, Azeredo-Espin *et al.*, em preparação), e da mosca-dos-chifres, *Haematobia irritans* (Capítulos 3 e 4). Por outro lado, nem sempre a informação referente à região-controle é necessária em estudos de mitogenômica preocupados em analisar variação intra e inter-específica a partir dos CDS, como é o caso em Ballard (2000a, b).

A família Muscidae

A família Muscidae é reconhecida como uma das maiores famílias do táxon Calyptratae (Diptera: Brachycera), e suas espécies podem ser encontradas em todas as regiões zoogeográficas do mundo. Ela inclui cerca de 173 gêneros e mais de 4000 espécies descritas (Carvalho & Couri 2002). Muitas espécies são sinantrópicas, como é o caso de espécies hematófagas da tribo Stomoxyini e espécies saprófagas do gênero *Musca*, *Ophyra* e *Atherigona*.

Duas destas espécies se destacam no continente americano como importantes pragas da pecuária: a mosca-dos-chifres, *Haematobia irritans*, e a mosca-dos-estábulo, *Stomoxys calcitrans* (Linnaeus). Estas espécies possuem aparelhos bucais adaptados para picar e sugar sangue e fluidos teciduais de vertebrados. A mosca-dos-chifres alimenta-se preferencialmente em gado bovino, embora outros rebanhos possam ser eventualmente atacados. Esta espécie tem sido considerada uma das principais pragas da pecuária bovina em vários países do continente Americano, onde é responsável por grandes perdas econômicas para a indústria devido ao estresse dos animais e conseqüente diminuição na produção de carne e leite (Campbell 1976; Stork 1979; Kunz *et al.* 1984; Haufe 1987), além da redução na qualidade do couro (Guglielmone *et al.* 1999). A origem da introdução desta praga no Brasil ainda não é certa, apesar de haver fortes evidências de que esta espécie teria chegado ao Brasil por Roraima vindo da Venezuela (Valério & Guimarães 1983). Entretanto, um recente estudo com marcadores RAPD não dão suporte a esta hipótese (Castiglione & Bicudo 2005) e a ocorrência de um ou mais eventos de introdução desta praga no território brasileiro em datas anteriores à década de 1980 são hipóteses que devem ser consideradas (Valério & Guimarães 1983). Diferentemente de *H. irritans*, *S. calcitrans* possui uma gama mais ampla de hospedeiros, que incluem bovinos, eqüídeos, caprinos, ovinos, cães, gatos e, até mesmo, o homem. Mesmo assim, a redução da produtividade bovina é o principal problema causado por esta espécie (Lancaster & Meisch 1986). Estimativas mostram que cerca de 1,1 bilhões de dólares são gastos anualmente com *S. calcitrans* e *H. irritans* pelos custos com inseticidas e perdas na produção animal (Drummond *et al.* 1981; Arther 1991). Não existem dados oficiais sobre o custo do controle destas pragas no Brasil, mas levando-se em consideração o tamanho do nosso rebanho bovino, tais prejuízos podem ser significativos.

Outra espécie da família Muscidae que merece destaque é a mosca doméstica, *Musca domestica* (Linnaeus). Esta é uma espécie verdadeiramente cosmopolita, associada ao homem e à

produção animal ao longo da história (Lancaster & Meisch 1986). As perdas econômicas na indústria estão relacionadas à higienização dos locais de produção de ovos e leite, que podem ser desqualificados devido ao grau de infestação dessas moscas (Lancaster & Meisch 1986; Learmount *et al.* 2002). Além disso, esta espécie também tem sido descrita como vetor mecânico e/ou biológico de vários agentes patogênicos, incluindo parasitas de animais domésticos e humanos (Greenberg 1973). Para controle biológico de *M. domestica*, outra espécie da família Muscidae, “the black dump fly” *Ophyra aenescens* (Wiedemann), tem sido usada eficientemente. Foi verificado que uma larva de *O. aenescens* é capaz de predação de quinze a vinte larvas de *M. domestica* num só dia (Farkas & Jantyk 1990). *O. aenescens* também é considerada uma espécie de importância forense devido ao hábito saprófago das larvas (Lefebvre & Pasquero, 2004).

Na família Muscidae, existem ainda representantes cuja importância econômica é dada pelos danos causados à agricultura, como é o caso da mosca menor das frutas, *Atherigona orientalis* (Schiner). A larva se alimenta preferencialmente de esterco e material vegetal em decomposição, mas ela pode ter também hábito fitófago atacando melões, tomates, trigo, milho, citros, entre outros (Chaughtal *et al.* 1985).

O desenvolvimento e uso de marcadores moleculares obtidos a partir da análise do DNAmt e do DNA nuclear de moscas causadoras de miíases e ectoparasitos associados à produção animal no Brasil têm sido comumente descritos na literatura (Valle & Azeredo-Espin 1995, Infante-Vargas & Azeredo-Espin 1995, Azeredo-Espin & Madeira 1996, Geurgas *et al.* 2000, Litjens *et al.* 2001, Junqueira *et al.* 2002, Torres *et al.* 2004, Torres & Azeredo-Espin 2005, Lyra *et al.*, 2005, Azeredo-Espin & Lessinger 2006). Estas análises têm se mostrado significativamente importantes na elaboração de programas de controle e erradicação de pragas na América do Sul, como está se tentando estabelecer para a mosca-da-bicheira, *Cochliomyia hominivorax* (IAEA 2000). Exemplos como estes servem para incentivar trabalhos de caracterização do DNAmt de outras espécies-praga, como *H. irritans*, *S. calcitrans*, *M. domestica*, *O. aenescens* e *A. orientalis*, possibilitando a avaliação do potencial informativo destes dados em estudos de variabilidade genética úteis no planejamento de estratégias de controle destas espécies, além de ampliar o conhecimento sobre a estrutura e evolução do genoma mitocondrial de insetos, cujo sequenciamento fornece informações estratégicas para a identificação de “primers” e sítios de restrição, e acrescenta um conjunto de genes ortólogos para

análises comparativas e filogenéticas (Lessinger *et al.* 2000, Junqueira *et al.* 2004, Azeredo-Espin & Lessinger 2006).

CAPÍTULOS - APRESENTAÇÃO

Esta tese apresenta quatro capítulos contendo trabalhos que caracterizam o genoma mitocondrial de espécies de importância econômica da família Muscidae, abordando a estrutura de regiões específicas, assim como do DNAm como um todo, de modo a inferir alguns aspectos relacionados à evolução desta molécula. Os dois primeiros capítulos estão compilados na forma de artigos científicos já publicados e os dois últimos estão na forma de manuscrito, submetidos para publicação. Além disso, todos os capítulos contêm um resumo em português e a metodologia detalhada dos experimentos e análises.

CAPÍTULO 1

Artigo - Oliveira, M.T., Azeredo-Espin, A.M.L. & Lessinger, A.C. (2005). Evolutionary and structural analysis of the cytochrome c oxidase subunit I (COI) gene from *Haematobia irritans*, *Stomoxys calcitrans* and *Musca domestica* (Diptera: Muscidae) mitochondrial DNA. *DNA Sequence* 16: 156-160.

Resumo

Este trabalho descreve a caracterização molecular do gene da subunidade I da citocromo oxidase c (COI) do DNA mitocondrial de três espécies de grande importância médica e veterinária: a mosca-dos-chifres, *Haematobia irritans*, a mosca-dos-estábulo, *Stomoxys calcitrans* e a mosca doméstica, *Musca domestica* (Diptera: Muscidae) (Linnaeus). A região sequenciada nestas espécies tem 1536pb, que codifica para um peptídeo de 512 aminoácidos. O viés na composição nucleotídica para acúmulo de A e T está relacionado a três características: grande conteúdo de A+T em toda sequência, grande conteúdo de A+T na terceira posição do códon, e predominância pelo uso de códons ricos em A+T. Um códon de iniciação anômalo (TCG - serina) foi identificado. Análises comparativas entre membros das famílias Muscidae, Scatophagidae, Calliphoridae e Drosophilidae mostraram grande nível de conservação na sequência. Análises dos aminoácidos divergentes e das topologias da proteína COI entre as três espécies de Muscidae concordam com o modelo evolutivo proposto para a proteína COI de insetos. A caracterização da estrutura e evolução deste gene pode ser informativa para futuros estudos evolutivos em Diptera.

Palavras-chave: DNA mitocondrial, COI, caracterização molecular, mosca-dos-chifres, mosca-dos-estábulo, mosca doméstica.

SHORT COMMUNICATION

Evolutionary and structural analysis of the cytochrome c oxidase subunit I (COI) gene from *Haematobia irritans*, *Stomoxys calcitrans* and *Musca domestica* (Diptera: Muscidae) mitochondrial DNA

MARCOS TÚLIO DE OLIVEIRA¹, ANA MARIA LIMA DE AZEREDO-ESPIN^{1,2}, & ANA CLÁUDIA LESSINGER¹

¹Laboratório de Genética Animal, Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), Campinas, São Paulo, Brazil, and ²Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Campinas, São Paulo, Brazil

(Received 18 August 2004)

Abstract

This work describes the molecular characterization of the cytochrome c oxidase subunit I (COI) gene of the mitochondrial DNA from three species of great medical and veterinary importance: the horn fly, *Haematobia irritans*, the stable fly, *Stomoxys calcitrans* and the house fly, *Musca domestica* (Diptera: Muscidae) (Linnaeus). The nucleotide sequence in all species was 1536 bp in size and coded for a 512 amino acid peptide. The nucleotide bias for an A + T-rich sequence is linked to three features: a high A + T content throughout the entire gene, a high A + T content in the third codon position, and a predominance of A + T-rich codons. An anomalous TCG (serine) start codon was identified. Comparative analysis among members of the Muscidae, Scatophagidae, Calliphoridae and Drosophilidae showed high levels of nucleotide sequence conservation. Analysis of the divergent amino acids and COI protein topologies among these three Muscidae species agreed with the evolutionary model suggested for the insect mitochondrial COI protein. The characterization of the structure and evolution of this gene could be informative for further evolutionary analysis of dipteran species.

Keywords: Mitochondrial DNA, COI, molecular characterization, horn fly, stable fly, house fly

Short communication

Mitochondrial DNA (mtDNA) has been considered a potential molecular marker in phylogenetic, population and evolutionary studies for a wide variety of animal taxa, including the insects. The characterization of mitochondrial genes has contributed to the identification of informative sequences that have improved our understanding of organismal evolution and the diversity of the mitochondrial genome. Among the protein-coding genes in the mitochondrial genome, the cytochrome c oxidase subunit I (COI) has been extensively used for evolutionary studies in insects (Caterino et al. 2000). This gene shows varying degrees of conservation throughout its sequence, and a range of nucleotide substitution rates that can be used for different

evolutionary analysis. The identification of conserved primers able to amplify specific COI regions (Simon et al. 1994, Lunt et al. 1996) has enhanced the usefulness of this gene as a molecular marker. For instance, more than 68% of all Calypttratae (Diptera: Brachycera) mitochondrial sequences deposited in GenBank (www.ncbi.nlm.nih.gov) are related to the COI gene.

Few mtDNA sequences from Muscidae flies have been characterized, despite the occurrence of species of great medical and veterinary importance in this family. Around 50% of all Muscidae mtDNA entries deposited in GenBank describe COI sequences. The aim of this work was to characterize the COI gene sequences of three Muscidae species: the horn fly, *Haematobia irritans*; the stable fly, *Stomoxys calcitrans*; and the house fly, *Musca domestica* (Linnaeus). These species are wide

Correspondence: M.T. Oliveira, Laboratório de Genética Animal, Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), Campinas, CP 6010, São Paulo, CEP 13083-875, Brazil. Tel: 55 19 37881141. Fax: 55 19 37881089. E-mail: mto@unicamp.br

Table I. A + T content of dipteran COI genes.

	Entire sequence (%)	Codon position		
		First (%)	Second (%)	Third (%)
<i>H. irritans</i> *	71.4	58.7	60.6	94.9
<i>S. calcitrans</i> †	70.6	57.5	60.6	94.0
<i>M. domestica</i> ‡	69.6	57.5	60.4	90.6
<i>Cochliomyia hominivorax</i> ¹	68.6	57.3	59.6	88.9
<i>Chrysomya chloropyga</i> ⁵	68.7	58.1	59.4	88.7
<i>Drosophila yakuba</i> ⁱ	69.8	58.4	59.6	91.6
<i>Aedes aegypti</i> ⁱ	68.6	58.2	58.6	88.9

*AY526195, †AY526197, ‡AY526196, ¹AF260826, ⁵AF352790, ⁱX03240, ⁱAF380835.

spread throughout the America and are considered important livestock pests. *H. irritans* and *S. calcitrans* are hematophagous pests that feed on vertebrate blood. *M. domestica* is a well-known cosmopolitan species always found in association with human activities and a potential vector of more than a hundred disease-causing microorganisms (Greenberg 1973). The characterization of specific mtDNA regions in these species could provide informative markers for screening genetic variability and population dynamics, and could also contribute to our understanding of the molecular evolution of mitochondrial genes in the Muscidae family.

The A + T composition of dipteran COI genes is shown in Table I. In general, the A + T content reported for this gene varies from 68.1 to 70.7% (Lunt et al. 1996). The muscid COI gene has one of the highest A + T% values among dipterans, especially that of *H. irritans*, whose A + T content (71.4%) is beyond the previous range. Analysis of the nucleotide content at the third codon position have shown that *H. irritans* and *S. calcitrans* have a higher A + T composition than other dipterans, and even than other insects (Crozier and Crozier 1993). A mutational tendency against G nucleotides in the third codon position was also noted in the muscid COI sequences, as already described for other metazoan mtDNAs (Saccone et al. 1999). Only 9 G nucleotides were identified in the *H. irritans* sequence, and 11 in both *S. calcitrans* and *M. domestica*, while

C nucleotides occurred 17, 20 and 37 times, respectively, in these species.

The putative initiation codon of the COI gene in these flies was suggested to be the triplet TCG (Ser). This anomalous TCG initiation codon has been found in many dipteran species and has been suggested to be the translation start codon for the COI gene (Beard et al. 1993). Recently, Morlais and Severson (2002) isolated and identified the COI transcript of the *Aedes aegypti* gene in which the TCG triplet was confirmed as the initiation codon.

Comparison of the COI genes of the three muscid species studied here showed high levels of nucleotide sequence similarity—around 90% of the nucleotide sites were identical among the three sequences. The nucleotide distances relative to muscid and other dipteran species using Tajima-Nei model are shown in Table II. Genetic distances calculated using the nucleotide sequences of *Haematobia irritans*, *Stomoxys calcitrans* and *Musca domestica* COII gene (Szalanski and Owens 2003) were used to compare with COI data. Analysis of COI sequences show higher genetic distances among these muscid species than those relative to COII sequence analysis, but the relative distances between pairs of species agree in both cases.

Figure 1 shows that most of the substitutions between pairs of COI sequences from different muscid species

Table II. Comparison of the COI genes of *H. irritans*, *S. calcitrans*, *M. domestica* and other dipterans based on nucleotide distances (above the diagonal).

	<i>H. irritans</i>	<i>S. calcitrans</i>	<i>M. domestica</i>	<i>Scathophaga stercoraria</i> (Scathophagidae)	<i>Cochliomyia hominivorax</i> (Calliphoridae)	<i>Chrysomya putoria</i> (Calliphoridae)	<i>Drosophila yakuba</i> (Drosophilidae)
<i>H. irritans</i> *		0.102	0.110	0.114	0.126	0.118	0.128
<i>S. calcitrans</i> †	[0.009]		0.127	0.130	0.147	0.133	0.134
<i>M. domestica</i> ‡	[0.009]	[0.011]		0.119	0.144	0.106	0.138
<i>S. stercoraria</i> ¹	[0.009]	[0.010]	[0.010]		0.114	0.103	0.125
<i>C. hominivorax</i> ⁵	[0.010]	[0.011]	[0.011]	[0.010]		0.103	0.147
<i>C. chloropyga</i> ⁱ	[0.010]	[0.011]	[0.009]	[0.009]	[0.009]		0.151
<i>D. yakuba</i> ⁱ	[0.011]	[0.011]	[0.010]	[0.010]	[0.011]	[0.011]	

The standard errors are given in brackets. The method used was the Tajima-Nei distance model.

*AY526195, †AY526197, ‡AY526196, ¹AF104625, ⁵AF260826, ⁱAF352790, ⁱX03240.

(V) change from *M. domestica* to *H. irritans* and *S. calcitrans* required two nucleotide substitutions involving the first and second positions of the codon.

We performed comparative analysis among these data and other muscid homologous sequence available in GenBank: comparisons of the *M. domestica* COI gene sequence with homologous sequences (AF104616, AF104622 and AF259518) were done to screen for the identification of polymorphic sites. This comparative analysis showed a T $\leftrightarrow$ C transition at position 344 of the nucleotide sequence. This substitution resulted in a methionine (M) $\rightarrow$ threonine (T) amino acid change in the predicted COI protein of *M. domestica* AF259518 sequence. We also compared our data with 11 GenBank entries related to a 218 bp fragment of the *S. calcitrans* COI (AF455120 to AF4551229 and AF452104): Of 31 polymorphic sites identified as synonymous substitutions, five were only found in the *S. calcitrans* sequence reported here. Nucleotide distance analysis showed that the haplotype described here clustered with a previously described haplotype 5, AF455123 (0.023 ± 0.011). Pairwise comparison of our data with the COI partial sequence (1167 bp) of another muscid, *Hydrotaea rostrata* (AB112829), showed an average similarity of 86.3% at the nucleotide level and 97.1% at the amino acid level. There was a predominance of A $\leftrightarrow$ T transversions at the third codon position (44.6% of all nucleotide substitutions), but only two A $\leftrightarrow$ T transversions were *H. rostrata*-specific amino acid changes (K $\rightarrow$ N and I $\rightarrow$ V at homologous positions 406 and 414, respectively). In addition to non-synonymous substitutions at third codon positions, a further five amino acid sites were specific to the *H. rostrata* sequence: N, L, I, F and F at homologous positions 118, 361, 392, 465 and 466, respectively.

The similarities among the predicted amino acid sequences of the three muscid COI proteins were high. Only six divergent sites changed the amino acid polarity and these substitutions occurred at the internal loops I2, I4 and I5 (positions 171, 328 and 402), the external loops E2 and E6 (positions 133 and 447) and the COOH-terminal (position 510) (see Lunt et al. 1996). None of the changes affected the structural profile of the predicted COI protein of *H. irritans*, *S. calcitrans* and *M. domestica*, and this profile agreed with the structural evolutionary model described for the insect COI protein (Lunt et al. 1996).

The amino acid content of the muscid COI protein sequences was examined to assess the A + T bias effect. Crozier and Crozier (1993) suggested a correlation between base composition and codon usage in insect mtDNA, with a preference of protein-coding genes using A + T-rich codons. Comparison of the A + T and G + C-rich codons revealed a significant difference in amino acid bias in the predicted COI proteins of these muscid species (the third codon positions were not considered). Amino acids coded by A + T-rich codons (K, M, N, I, Y and F) were more abundant than those

coded by G + C-rich codons (G, A, R and P) ($\chi^2 = 23.90$, $p < 0.00001$, DF = 1 for *H. irritans*; $\chi^2 = 23.37$, $p < 0.00001$, DF = 1 for *S. calcitrans*; and $\chi^2 = 22.80$, $p < 0.00001$, DF = 1 for *M. domestica*). In the animal mitochondrial genome, leucine (L) is coded by TTA and TTG, which are T-rich codons, and by CTN, which is not. In the muscid COI genes, the use of T-rich codons for L, instead of CTN codons, was more pronounced than in the homologous region of *Apis mellifera*, which has the A + T-richest mitochondrial genome, ($\chi^2 = 86.20$, $p \sim 0$, DF = 1 for *H. irritans*; $\chi^2 = 65.52$, $p \sim 0$, DF = 1 for *S. calcitrans*; and $\chi^2 = 59.29$, $p \sim 0$, DF = 1 for *M. domestica*; versus $\chi^2 = 50.81$, $p \sim 0$, DF = 1 for *A. mellifera*). Foster et al. (1997) suggested that this type of amino acid bias would affect phylogenetic reconstructions based on protein sequences, in a manner similar to that of nucleotide bias in DNA sequences. Evolutionary studies using muscid COI sequence should carefully address this question.

The characterization of mitochondrial genes in the Muscidae family could contribute to increase the knowledge about the diversity and evolution of the mtDNA molecule in the dipteran order. Some taxa such as the Muscoidea superfamily are poorly represented in sequence databases and the few available sequences are from nuclear genes from a low number of species. Phylogenetic studies supported by molecular markers (Bernasconi et al. 2000, Nirmala et al. 2001), including partial COI sequences, have investigated the monophyly of the Muscoidea and Oestroidea taxa, but no monophyletic relationships have been convincingly established. Studies providing mtDNA sequence information could be useful for further evolutionary work on Muscoidea species. A detailed examination of the evolutionary and structural patterns of specific mtDNA regions could provide valuable insights for the effective use of mtDNA as a molecular marker in phylogenetic studies (Lunt et al. 1996).

Acknowledgements

The authors thank Rosângela A. Rodrigues for technical assistance and Patrícia J. Thyssen and João Ricardo Martins for providing specimens. This work, as well as M.T.O. and A.C.L., were supported by the Programa Especial de Estímulo à Fixação de Doutores—Conselho Nacional de Desenvolvimento Científico e Tecnológico (PROFIX/CNPq, grant no. 540.602/01-9). A.M.L.A.E. was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, grant no. 03/01458-9) and CNPq (grant no. 471132/01-2).

References

- Beard CB, Hamm DM, Collins FH. 1993. 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.

- Bernasconi MV, Valsangiacomo C, Piffaretti JC, Ward PI. 2000. Phylogenetic relationships among Muscoidea (Diptera: Calyptratae) based on mitochondrial DNA sequences. *Insect Mol Biol* 9:67–74.
- Caterino MS, Cho S, Sperling FAH. 2000. The current state of insect molecular systematics: A thriving Tower of Babel. *Annu Rev Entomol* 45:1–54.
- Crozier RH, Crozier YC. 1993. The mitochondrial genome of the honeybee *Apis mellifera*: Complete sequence and genome organization. *Genetics* 133:97–117.
- Foster PG, Jermini LS, Hickey DA. 1997. Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. *J Mol Evol* 44:282–288.
- Greenberg B. 1973. Flies and disease. Biology and disease transmission. Vol. II. New Jersey: University Press Princetown.
- Infante ME, Azeredo-Espin AML. 1995. Genetic variability in mitochondrial DNA of screwworm, *Cochliomyia hominivorax* (Diptera: Calliphoridae), from Brazil. *Biochem Genet* 33:737–756.
- Kumar S, Tamura K, Nei M. 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.01. University Park, Pennsylvania: The Pennsylvania State University.
- Lessinger AC, Azeredo-Espin AML. 2000. Evolution and structural organisation of mitochondrial DNA control region of myiasis-causing flies. *Med Vet Entomol* 14:71–80.
- Litjens P, Lessinger AC, Azeredo-Espin AML. 2001. Characterization of the screwworm flies *Cochliomyia hominivorax* and *Cochliomyia macellaria* by PCR-RFLP of mitochondrial DNA. *Med Vet Entomol* 15:183–188.
- Lunt DH, Zhang D-X, Szymura JM, Hewitt GM. 1996. The insect cytochrome oxidase I gene: Evolutionary patterns and conserved primers for phylogenetic studies. *Insect Mol Biol* 5:153–165.
- Morlais I, Severson DW. 2002. Complete mitochondrial DNA sequence and amino acid analysis of the cytochrome c oxidase subunit I (COI) from *Aedes aegypti*. *DNA Seq* 13:123–127.
- Nirmala X, Hypsa V, 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 Mol Biol* 10:475–485.
- 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.
- 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 polymerase chain reaction primers. *Ann Entomol Soc Am* 87:651–701.
- Szalanski AL, Owens CB. 2003. Sequence change and phylogenetic signal in muscoid COII DNA sequences. *DNA Seq* 14:331–334.
- Tajima F, Nei M. 1984. Estimation of evolutionary distance between nucleotide sequences. *Mol Biol Evol* 1:269–285.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25:4876–4882.

Metodologia detalhada

Amostras

Indivíduos adultos de *H. irritans* foram coletados diretamente sobre os hospedeiros bovinos no município de Goiânia, GO, em Outubro/2001, e conservados em álcool absoluto a 4°C. Adultos de *S. calcitrans* foram enviados do município de Eldorado do Sul, RS, pelo pesquisador João Ricardo Martins (FEPAGRO) em Abril/2003. As amostras estavam preservadas em álcool absoluto e foram mantidas a 4°C. Para *M. domestica*, pupas frescas provenientes de uma criação originada a partir de amostras do município de Sarapu, SP, foram fornecidas por Patrícia Thyssen (Depto. Parasitologia, UNICAMP) em Outubro/2002.

Extração do DNA

As extrações de DNA total foram feitas conforme o método fenol/clorofórmio, descrito em Infante-Vargas & Azeredo-Espin (1995), com volumes adaptados para a realização do procedimento em tubos de microcentrífuga. Cada indivíduo foi colocado em um tubo com 150µL de tampão de homogeneização (Tris.Cl 10mM, NaCl 60mM, Sacarose 300mM, EDTA 10mM), pH 7,5, e macerado após a adição de 150µL de tampão de lise (Tris.Cl 300mM, SDS 40mM, EDTA 20mM, 0,7% do volume de DEPC), pH 9,0. Em seguida, as amostras foram incubadas no gelo por 15 minutos e, logo após, foi adicionado igual volume (300µL) de fenol (equilibrado em pH 7,6), seguido de agitação e centrifugação a 4°C por 5 minutos a 5000 rpm. Ao sobrenadante foram adicionados volumes iguais de fenol e clorofórmio:álcool isoamílico (24:1). As amostras foram centrifugadas novamente por 5 minutos a 5000 rpm e ao sobrenadante foram adicionados mesmo volume de clorofórmio:álcool isoamílico (24:1). Logo após, outra centrifugação foi realizada como descrito anteriormente. Ao sobrenadante final foram adicionados 225µL de 1xTE (10mM Tris.Cl, pH 7,4, e 0,1mM EDTA, pH 8,0), 22µL de acetato de sódio 3M e 1200µL de etanol absoluto a -20°C, seguindo-se de precipitação do DNA a -20°C por 2 horas. Após a precipitação, as amostras foram centrifugadas a 9000 rpm por 30 minutos a 4°C. O “pellet” de DNA foi ressuspensionado, novamente, em 150µL de 1xTE, 8µL de acetato de sódio 3M e 450µL de etanol absoluto a -20°C e precipitado por 30 minutos a -70°C. Após uma centrifugação final a 9000 rpm por 30 min, o “pellet” de DNA foi ressuspensionado em 50-100µL de 1xTE e armazenado a -20°C até análises posteriores.

Amplificações

Diferentes combinações de oligonucleotídeos iniciadores (“primers”) *universais* descritos em Simon *et al.* (1994) para o DNAm_t de insetos foram usados para a amplificação via reação em cadeia da polimerase (PCR) do gene COI do DNAm_t das três espécies da família Muscidae. O gene COI foi amplificado em dois fragmentos distintos que se sobrepunham, contendo a região carboxi- (C1-N-2329 + TY-J-1460) ou amino-terminal (TL2-N-3014 + C1-J-2195) (Figura 1). As reações foram realizadas em um volume final de 25µl utilizando os termocicladores PTC-200, MiniCycler (MJ Research) e GeneAmp PCR System 9600 (Perkin Elmer). Foram utilizados os seguintes reagentes (concentrações finais): 20mM Tris-HCl (pH 8.4), 50mM KCl, 1,25U da enzima *Taq* DNA polimerase, 200µM de cada dNTP, 1,5mM MgCl₂, 500nM de cada oligonucleotídeo (Figura 1), e 1µl da amostra de DNA total. As condições das reações foram: desnaturação a 94°C por 5min (1 ciclo); desnaturação a 94°C por 1min, hibridização a 50°C por 1min, e extensão a 72°C por 2min (35 ciclos); extensão final a 72°C por 5min (1 ciclo). Um volume de 5µl de cada reação foi visualizado em gel de agarose 1%, corado com brometo de etídeo.

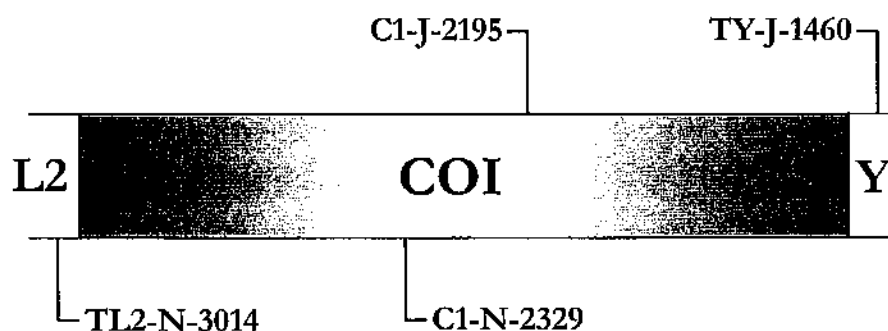


Figura 1. Representação esquemática do gene COI e genes flanqueadores (L2 - RNAt para leucina 2; Y - RNAt para tirosina) do DNAm_t de insetos, mostrando a posição e os “primers” usados na metodologia deste capítulo.

Seqüenciamento

Os produtos das reações de amplificação foram purificados usando o kit QIAQuick PCR Purification (Qiagen), segundo instruções do fabricante. Foi realizado seqüenciamento direto a partir dos produtos da PCR purificados utilizando 2µL do ABI-PRISM Big Dye terminator

sequencing v3.0 (Applied Biosystems), 2µL de tampão de seqüenciamento (Applied Biosystems), 3,5pmol de “primer”, 100ng produto da PCR e água q.s.p. 10µL para cada reação de seqüenciamento. Cada produto da reação de seqüenciamento foi purificado através do método isopropanol/etanol e a obtenção das seqüências foi através do seqüenciador automático modelo 377 (Applied Biosystems), segundo instruções do fabricante.

Análises das Seqüências

Os dois fragmentos seqüenciados do gene COI foram montados manualmente para determinar a seqüência completa das três espécies. Além destas seqüências, as seqüências nucleotídicas do gene COI de outras espécies de dípteros depositadas no GenBank, como *Scathophaga stercoraria* (Scathophagidae) (AF104625), *Cochliomyia hominivorax* (AF260826), *Chrysomya putoria* (Calliphoridae) (AF352790), *Drosophila yakuba* (Drosophilidae) (X03240) e *Aedes aegypti* (Culicidae) (AF380835), também foram alinhadas usando o programa Clustal X (Thompson *et al.* 1997) configurado segundo os parâmetros padrões. As seqüências alinhadas foram analisadas em relação às composições nucleotídicas e de aminoácidos, localização dos sítios variáveis e distâncias genéticas usando o pacote MEGA2 (Kumar *et al.* 2001). O modelo de distância utilizado foi o Tajima-Nei (Tajima & Nei 1984).

As seqüências de aminoácidos preditas da proteína COI das espécies de Muscidae foram submetidas a buscas por padrões de hidrofobicidade através do método de “hidden Markov” utilizando dois diferentes programas: TMHMM Server 2.0 (www.cbs.dtu.dk/services/TMHMM) e HMMTOP 2.0 (www.enzim.hu/hmmtop/index.html). A predição do padrão de hidrofobicidade pode indicar regiões transmembranas na proteína e permitir a comparação com o padrão estrutural proposto para a proteína COI de insetos (Lunt *et al.* 1996, Figura 2).

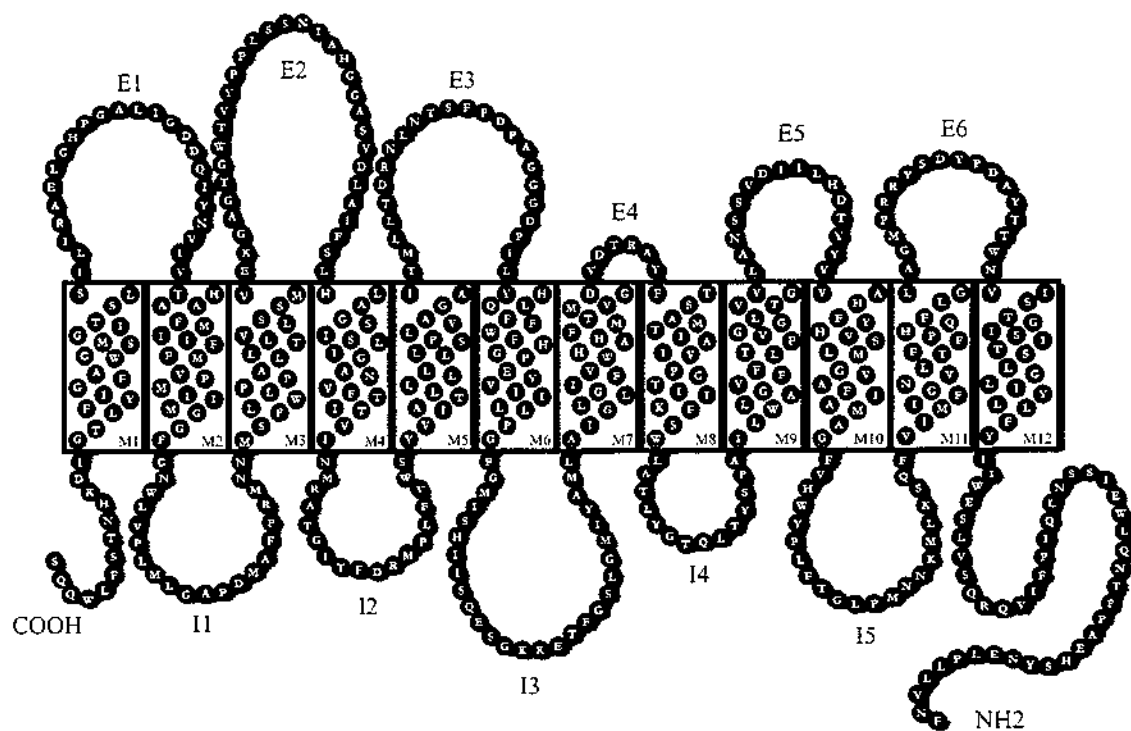


Figura 2. Esquema do modelo proposto para a topologia bidimensional da proteína COI de insetos, mostrando as regiões transmembranas (M1-M12), as alças internas (I1-I5) e externas (E1-E6), e o carboxi- (COOH) e amino-terminal (NH₂). Adaptado de Lunt *et al.* (1996).

CAPÍTULO 2

Artigo - **Oliveira, M.T.**, da Rosa, A.C., Azeredo-Espin, A.M.L. & Lessinger, A.C (2006). Improving access to the control region and tRNA gene clusters of Dipteran mitochondrial DNA. *Journal of Medical Entomology*, 43: 636-639.

Resumo

Apesar do sucesso do uso de “primers” universais para amplificar o DNAm_t de insetos, algumas regiões específicas continuam sendo difíceis de analisar e demandam o uso de “primers” táxon-específicos. Neste trabalho, nós descrevemos um novo conjunto de “primers” para amplificar e seqüenciar eficientemente a região-controle e três agrupamentos de genes de RNAt do DNAm_t de dípteros de importância médica e veterinária, incluindo espécies das famílias Muscidae, Calliphoridae e Oestridae. Estes novos “primers” foram úteis para investigar a informação nucleotídica e a organização estrutural do DNAm_t de dípteros.

Palavras-chave: “primers” conservados, região-controle, Diptera, DNA mitocondrial, RNAt

Improving Access to the Control Region and tRNA Gene Clusters of Dipteran Mitochondrial DNA

M. T. OLIVEIRA,¹ A. C. DA ROSA, A.M.L. AZEREDO-ESPIN, AND A. C. LESSINGER

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, CEP 13083-875, Campinas, São Paulo, Brazil

J. Med. Entomol. 43(3): 636–639 (2006)

ABSTRACT Despite the successful use of universal primers for amplifying insect mtDNA, specific regions remain difficult to recover and demand the use of taxon-specific primers. In this work, we describe a new set of primers for efficiently amplifying and sequencing the mtDNA control region and three tRNA gene clusters of dipterans of medical and veterinary importance, including Muscidae, Calliphoridae, and Oestridae species. These new primers were useful for investigating the nucleotide information and the structural organization of dipteran mtDNA.

KEY WORDS conserved primers, control region, Diptera, mitochondrial DNA, tRNAs

Animal mitochondrial DNA (mtDNA) has been widely used as a molecular marker in evolutionary studies (Boore 1999). A compilation of conserved polymerase chain reaction primers described by Simon et al. (1994) provided a useful tool to assess regions of insect mtDNA. However, there are mtDNA regions that are poorly covered by the conserved insect primers, particularly tRNA gene clusters and the control region (or A+T-rich region). In this report, we describe a new set of primers that are useful for amplifying 1) the control region (CR); 2) the CR-flanking genes, tRNA^{Met}, tRNA^{Gln}, and tRNA^{Ile} (MQI); 3) the gene cluster containing the tRNA^{Tyr}, tRNA^{Cys}, and tRNA^{Trp} (YCW); and 4) the main tRNA cluster of mtDNA (FES2NRA). This set of primers was designed to amplify the mtDNA of important medical and veterinary dipterans, such as the horn fly, *Haematobia irritans* (L.); the house fly, *Musca domestica* (L.); the stable fly, *Stomoxys calcitrans* (L.); *Atherigona orientalis* (Schiner); *Ophyra aenescens* (Wiedemann); (Muscidae) the screwworm, *Cochliomyia hominivorax* (Coquerel); the blowflies *Chrysomya albiceps* (Wiedemann), *Chrysomya chloropyga* (Wiedemann), and *Chrysomya megacephala* (F.) (Calliphoridae); and the botfly, *Dermatobia hominis* (L., Jr.) (Oestridae) (Table 1).

Materials and Methods

In addition to preliminary sequence data for *H. irritans*, *S. calcitrans*, and *M. domestica* mtDNA, the homologous regions from the ND2, ND3, ND5, and 12S rRNA genes of dipterans available in GenBank (www.ncbi.nlm.nih.gov) were aligned using Clustal X (Thompson et al. 1997). The candidate primers were

selected based on conserved aligned regions, except for the primers TI-N-21 and TI-J-38, which were specifically designed based on *H. irritans* sequences. The structural stabilities and thermodynamic properties of all candidate primers were analyzed using Gene Runner software (Hastings Software Inc., Hastings on Hudson, NY). Detailed information for the eight primers reported here is given in Fig. 1.

Total DNA from individual dipterans was extracted according to Infante-Vargas and Azeredo-Espin (1995), with volumes adapted for microcentrifuge tubes. Polymerase chain reaction (PCR) was done using a PTC-200 thermocycler (MJ Research, Watertown, MA), with the following conditions: 1 µl of total DNA was used in a 25-µl reaction with 1.25 U of *Taq* DNA polymerase (Invitrogen, Carlsbad, CA), 200 µM of each dNTP, 1.5 mM MgCl₂, and 500 nM of each primer. After an initial denaturation step at 95°C for 5 min, 35 cycles were run, each consisting of melting at 95°C for 1 min, annealing at 45–52°C for 1 min, and extension at 60°C for 2 min. A final extension step was run at 60°C for 5 min. When appropriate, the purified PCR products were directly sequenced with BigDye Terminator version 3.1 cycle sequencing kit and analyzed in an ABI 377 automated DNA sequencer (Applied Biosystems, Foster City, CA), except for *H. irritans* amplicons, which were cloned before sequencing. The cloning reactions were done using a TA cloning kit (Invitrogen), according to the manufacturer's specifications.

Results and Discussion

The primers N2-N-327, SR-J-14941, and TI-N-21 were designed to provide an efficient amplification of the mtDNA CR. The primer N2-N-327 is perhaps the

¹ E-mail: mto@unicamp.br.

Table 1. Number and origin of samples of dipteran species tested for the amplifications of mtDNA regions

Sample origin	mtDNA region			
	CR ^a	MQI ^b	YCW ^c	FES2NRA ^d
<i>H. irritans</i>				
Pirassununga, Brazil	8	8	8	8
Goiania, Brazil	8	8	8	8
Patos, Brazil	8	8	8	8
Libertad, Uruguay	8	8	8	8
Rafaela, Argentina	8	— ^e	—	—
Barquisimeto, Venezuela	4	—	—	—
Kluang, Malaysia	5	—	—	—
<i>M. domestica</i>				
Surapei, Brazil	8	8	8	8
<i>S. calcitrans</i>				
São Sebastião do Paraíso, Brazil	* ^f	*	*	3
Uberlândia, Brazil	*	*	*	10
Eldorado do Sul, Brazil	*	*	*	10
<i>A. orientalis</i>				
Mogi Guaçu, Brazil	2	2	2	2
<i>O. aeneus</i>				
São Sebastião do Paraíso, Brazil	*	8	*	8
<i>C. hominivorax</i>				
Caiaipônia, Brazil	10	10	10	10
Cerro Largo, Uruguay	10	10	10	10
Barquisimeto, Venezuela	10	10	10	10
<i>C. albicaps</i>				
Rio de Janeiro, Brazil	—	NR	NR	8
Colônia, Uruguay	—	NR	NR	10
<i>C. chloropyga</i>				
Belém, Brazil	8	NR	NR	8
São João da Boa Vista, Brazil	8	NR	NR	8
<i>C. megacephala</i>				
Adamantina, Brazil	8	NR	NR	8
Campanas, Brazil	8	NR	NR	8
<i>D. hominis</i>				
Pirassununga, Brazil	10	10	10	10
Bragança Paulista, Brazil	10	10	10	10
Indaiatuba, Brazil	6	6	6	6

NR, reactions not recommended.

^a Region amplified using primers TI-N-21 + SR-J-14941 (*H. irritans*), and N2-N-327 + SR-J-14941 (all species).^b Region amplified using primers N2-N-327 + TI-J-42 (or TI-J-38 for *H. irritans*).^c Region amplified using primers C1-N-2191 + TI-J-42 (or TI-J-38 for *H. irritans*).^d Region amplified using primers N5-N-6554 + N3-J-5945 and N5-N-6647 + N3-J-5945.^e Not tested.^f No amplification.

most conserved, versatile, and structurally stable primer of this newly described set. In combination with 12S-annealing primers, the N2-N-327 primer provides an alternative for amplifying CR sequences instead of using tRNA-annealing primers. This strategy is particularly useful for the amplification of *Chrysomya* mtDNA, because many species of this genus present two tRNA^{Ile} copies (Lessinger et al. 2004). In addition, tRNA genes adjacent to the CR are frequently involved in structural rearrangements (Boore 1999), and this may contribute to failures in amplifying the CR in species where the mtDNA gene order is unknown. The primer SR-J-14941 is the complementary-reverse sequence of the primer SR-N-14922 (Simon et al. 1994), shifted 2 bp downstream. This primer is located at the 3' end of the 12S rRNA gene (Fig. 1A) and is very useful for improved sequencing of the 3' end of the CR.

The CR sequences of *H. irritans* samples from six localities in South America and one in Asia (Table 1) also could be amplified by combining the primers SR-J-14941 and TI-N-21 (instead of N2-N-327) to extend the CR 5' end sequencing data. The *H. irritans* CR was completely sequenced for eight samples (Table 1; GenBank accession numbers AY929377–AY929384) and ranged from 1256 to 1704 bp. The structural organization of *H. irritans* CR resembles that of myiasis-causing flies (Lessinger and Azeredo-Espin 2000).

We designed new primers to allow a better assessment of the tRNA clusters MQI, YCW, and FES2NRA. The primer N2-N-327, when combined with the new primers TI-J-42 or TI-J-38 described here, efficiently amplified the MQI gene cluster (~300 bp). These new primers are the complementary-reverse sequences of the TI-N-24 (Simon et al. 1994) and TI-N-21 (specific for the *H. irritans* sequence) primers, respectively. We recommend caution when using the primer TI-J-42 for amplifications of *Chrysomya* mtDNA because of the presence of two tRNA^{Ile} copies in the mtDNA of species of this genus (Lessinger et al. 2004).

The newly described primers TI-J-42 and TI-J-38 were used in combination with the primers C1-N-1560 and C1-N-2191 (Simon et al. 1994) to amplify and sequence the tRNA gene cluster YCW. The primer C1-N-1560 failed to amplify the YCW region of calliphorid and oestrid species, but worked well in amplifying muscid mtDNA. The homologous regions of *H. irritans*, *M. domestica*, *A. orientalis*, *C. hominivorax*, and *D. hominis* were successfully amplified using TI-J-42 (TI-J-38 in *H. irritans*) and C1-N-2191 (Simon et al. 1994) and provided an ~2000-bp PCR product, including the ND2 gene.

To amplify and sequence the FES2NRA gene cluster of *H. irritans*, *M. domestica*, and *S. calcitrans*, we designed the primers N3-J-5945, N5-N-6554, and N5-N-6647. PCR provided well-resolved amplicons of ~650 and 750 bp. These primers were successfully used to amplify the homologous regions of all species analyzed (Table 1).

Based on our preliminary analysis of restriction fragment-length polymorphism profiles and sequencing multiple clones, we feel that the primers described here amplify original mtDNA, but we encourage others to validate the amplicons retrieved by these primers before proceeding with any extensive study of a given dipteran species. Contamination of amplification reactions with nuclear copies of mitochondrial genes (NUMTs) is a widespread phenomenon that has not been described in Diptera (Blanchard and Schmidt 1996, Zhang and Hewitt 1996, Bensasson et al. 2000). To avoid the possible amplification of NUMTs, we suggest that amplification of the entire mitochondrial genome by long-PCR (Barau et al. 2005) should be used as the template for nested PCR with the primers described here. Because the insertion of large mitochondrial sequences (~8–9 kb) in the nuclear genome is a rare event (Zhang and Hewitt 1996), this strategy could provide reliable results.

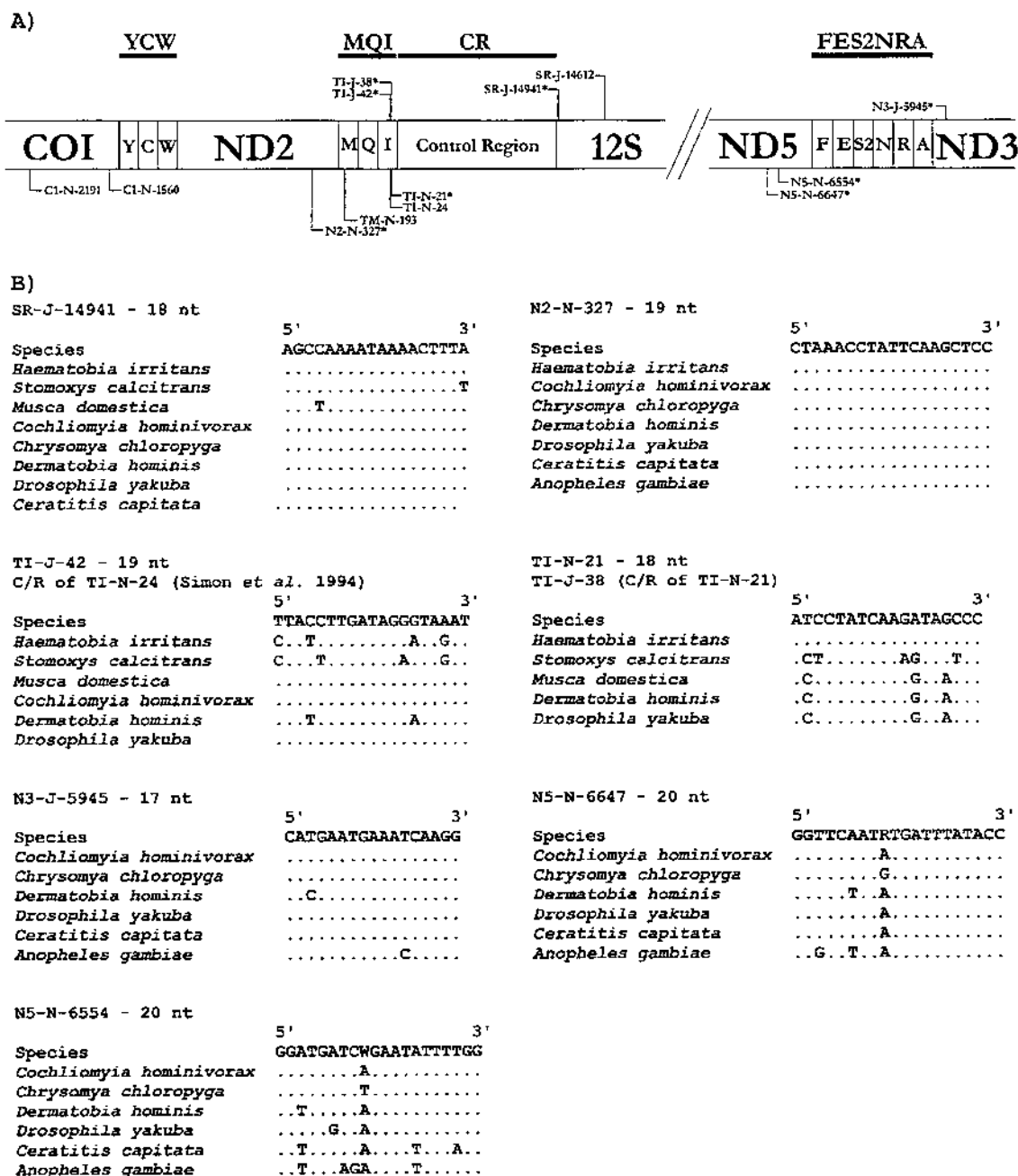


Fig. 1. (A) Schematic model of the mitochondrial genome organization. The primer nomenclature is based on Simon et al. (1994). Target regions are shown in thick bars and include the control region (CR), the CR-flanking tRNA genes (MQI), the YCW tRNA gene cluster (YCW), and the main mtDNA tRNA cluster (FES2NRA). The relative positions of the mtDNA primers described in this study (*) and the primers described by Simon et al. (1994) are given. (B) Primer sequences (5' to 3'), length (nt, nucleotides), and alignment of homologous regions in other dipteran species. Complementary-reverse sequences are indicated by C/R. The following symbols were used in the degenerated primers N5-N-6554 and N5-N-6647: W, A/T and R, A/C. Dots in the alignments indicate identical sites. The complete mitochondrial genomes available in GenBank, in addition to preliminary sequence data for Muscidae species, were included in the primer alignments. Sequences providing low-confidence alignments were excluded from the comparative analysis. The GenBank accession numbers for the *C. hominivorax*, *C. putoria*, *D. hominis*, *Drosophila yakuba* (Miyake), *Ceratitis capitata* (Weidemann), and *Anopheles gambiae* (Giles) mtDNA sequences are NC_002660, NC_002697, NC_006378, NC_001322, NC_000857, and NC_002084, respectively.

In conclusion, eight new primers, which can be efficiently combined with "universal" insect mtDNA primers (Simon et al. 1994), have been developed and evaluated for their usefulness in amplifying four mtDNA regions of muscid, calliphorid, and oestrid species. These primers may be useful for Diptera and may provide informative molecular markers as well as information on the structural diversity of dipteran mitochondrial genomes.

Acknowledgments

We thank R. A. Rodrigues for technical assistance in sequencing; M. L. Lyra and P. Fresia for help in collecting the *H. irritans* samples from Brazil and Uruguay; A. A. Guglielmone, A. Coronado, and P. J. Thyssen for providing the samples of *H. irritans* from Argentina and Venezuela and the *M. domestica* samples, respectively; and C.J.B. Carvalho and A. X. Linhares for the taxonomic identification of *O. aenescens* and *O. orientalis*, respectively. This work was supported by PROFIX/CNPq Grant 540.602/01-9 (to A.C.R. and A.C.L.), FAPESP Grant 03/01458-9 (to A.M.L.A.E.), and a fellowship from CAPES (to M.T.O.).

References Cited

- Barau, J. G., A.M.L. Azeredo-Espin and A. C. Lessinger. 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. *Mol. Ecol. Notes* 5: 885-887.
- Bensasson, D., D.-X. Zhang, and G. M. Hewitt. 2000. Frequent assimilation of mitochondrial DNA by grasshopper nuclear genome. *Mol. Biol. Evol.* 17: 406-415.
- Blanchard, J. L., and G. W. Schmidt. 1996. Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. *J. Mol. Evol.* 13: 537-548.
- Boore, J. L. 1999. Animal mitochondrial genomes. *Nucleic Acid Res.* 27: 1767-1780.
- 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.
- 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: 71-80.
- 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.
- Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, 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.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acid Res.* 25: 4876-4882.
- Zhang, D.-X., and G. M. Hewitt. 1996. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends Ecol. Evol.* 6: 247-251.

Received 23 September 2005; accepted 1 December 2005.

Metodologia detalhada

Seleção dos “primers”

Para a construção dos “primers” para amplificação das regiões de interesse do DNAm_t de dípteros, foi levado em consideração o grau de similaridade entre as seqüências e as propriedades termodinâmicas da molécula. Primeiramente, foram obtidas seqüências parciais do gene de RNAr 12S, do ND2 e do RNAt Ile (I) para as espécies *H. irritans*, *S. calcitrans* e *M. domestica* utilizando “primers” descritos em Simon *et al.* (1994). Estas seqüências foram alinhadas com seqüências de outros dípteros disponíveis no GenBank, tais como *Cochliomyia hominivorax* (NC_002660), *Chrysomya chloropyga* (NC_002697), *Dermatobia hominis* (NC_006378), *Drosophila yakuba* (NC_001322), *Ceratitis capitata* (NC_000857) e *Anopheles gambiae* (NC_002084), usando o programa Clustal X (Thompson *et al.* 1997), para a seleção regiões candidatas de 17 a 20pb altamente conservadas, então testadas quanto as suas propriedades termodinâmicas. O “primer” TI-N-21 e seu complementar-reverso TI-J-38 foram construídos baseados na seqüência idêntica a de *H. irritans* para melhorar a amplificação e o seqüenciamento da região-controle e dos agrupamentos MQI e YCW, pois o “primer” *universal* TI-N-24 (Simon *et al.* 1994) e seu complementar-reverso TI-J-42 não funcionaram bem para esta espécie, apesar de serem usados com sucesso para quase todas as espécies descritas na “Table I” do artigo apresentado neste capítulo. O grau de conservação dos oito “primers” descritos neste trabalho pode ser observado na “Fig. 1B” do artigo.

A avaliação das propriedades termodinâmicas e da estabilidade estrutural dos “primers” candidatos foi realizada através do programa Gene Runner (Hastings Software Inc.) usando os seguintes parâmetros: foram desconsiderados os “primers” que apresentaram dímeros, alças “hairpin”, “bulge” e internas com energia livre menor que -1 kcal/mol. Nenhuma restrição quanto ao conteúdo de G+C e valor de TM dos “primers” foi feita devido ao alto conteúdo de A+T nas seqüências usadas como molde para construção dos mesmos.

Amostras

Os dados sobre coletas e preservação das amostras utilizadas neste capítulo estão apresentados na Tabela 1.

Tabela 1. Dados sobre as amostras das espécies de dípteros utilizados neste trabalho.

Espécie	Localidade	Condição da amostra	Responsável	Data
<i>H. irritans</i>	Pirassununga, SP	Adulto congelado -70°C	M. Constantino Filho	08/2001
	Goiânia, GO	Adulto álcool absoluto	A.M.L. Azeredo-Espin e A.C.M. Junqueira	10/2001
	Patos, PB	Adulto álcool absoluto	M.L. Lyra	08/2002
	Rafaela, Argentina	Adulto álcool absoluto	A. Guglielmone	12/2002
	Libertad, Uruguai	Adulto congelado -70°C	M.T. Oliveira, M.L. Lyra e P. Fresia	01/2003
	Kluang, Malásia	Adulto álcool absoluto	A.M.L. Azeredo-Espin	09/2003
	Barquisimeto, Venezuela	Adulto álcool absoluto	A. Coronado	12/2003
	S. Seb. do Paraíso, MG	Adulto congelado -70°C	M.T. Oliveira, M.L. Lyra e P. Fresia	02/2004
<i>S. calcitrans</i>	Uberlândia, MG	Adulto álcool absoluto	J. Mendes e G.O.R. Motta	06/2003
	Eldorado do Sul, RS	Adulto álcool absoluto	J.R. Martins	04/2003
	Sarapuí, SP	Pupas frescas	P. Thyssen	10/2002
<i>M. domestica</i>	Mogi Guaçu, SP	Adulto congelado -70°C	A.C.M. Junqueira e T.T. Torres	04/2004
<i>A. orientalis</i>	S. Seb. do Paraíso, MG	Pupas frescas	M.T. Oliveira, M.L. Lyra e P. Fresia	02/2004
<i>O. aenescens</i>	Caiapônia, GO	Pupas frescas	R.C. Assis e M.L. Lyra	01/2005
	Cerro Largo, Uruguai	Pupas frescas	M.T. Oliveira, M.L. Lyra e P. Fresia	01/2003
	Barquisimeto, Venezuela	Adulto álcool absoluto	A. Coronado	12/2003
<i>C. albiceps</i>	Rio de Janeiro, RJ	Adulto álcool absoluto	F.F. Conte	05/2000
	Colônia, Uruguai	Pupa fresca	M.T. Oliveira, M.L. Lyra e P. Fresia	01/2003
	Belém, PA	Pupas frescas	A.M.L. Azeredo-Espin e A.C.M. Junqueira	03/2001
<i>C. chloropyga</i>	São João da Boa Vista, SP	Pupas frescas	Depto. Parasitologia, UNICAMP	02/2000
<i>C. megacephala</i>	Adamantina, SP	Adulto álcool absoluto	T.T. Torres	03/2000
	Campinas, SP	Adulto álcool absoluto	T.T. Torres	09/2001
<i>D. hominis</i>	Pirassununga, SP	Pupas frescas	A.A.G. Maia	01/2001
	Bragança Paulista, SP	Larvas frescas	K.S. Yotoko	10/1995
	Indaiatuba, SP	Pupas frescas	S. Geurgas	04/1996

Extração do DNA

O DNA total dos indivíduos foi extraído conforme descrito na Metodologia detalhada do Capítulo 1.

Amplificações

Os novos “primers” foram testados em combinação com “primers” *universais* para o DNAm_t de insetos (Simon *et al.* 1994) conforme esquematizado na “Fig. 1A” do artigo. As reações foram realizadas com um volume final de 25µl utilizando os termocicladores PTC-200, MiniCycler (MJ Research) e GeneAmp PCR System 9600 (Perkin Elmer). As concentrações finais dos reagentes utilizados são as mesmas descritas na Metodologia detalhada do Capítulo 1. As condições das reações foram: desnaturação a 94°C por 5min (1 ciclo); desnaturação a 94°C por 1min, hibridização de 42 a 55°C por 1min, e extensão a 60°C por 2min (35 ciclos); extensão final a 60°C por 5min (1 ciclo). Um volume de 5µl de cada reação foi visualizado em gel de agarose 1%, corado com brometo de etídeo.

Clonagem e Seqüenciamento

Para verificação da amplificação das regiões alvo (região-controle, MQI, YCW e FES2NRA), os produtos da PCR de um indivíduo de *H. irritans* de cada localidade foram clonados e seqüenciados. A purificação dos produtos amplificados foi realizada através de diálise por 20 min utilizando membranas Millipore 0,05µm e tampão T.E., pH 7,0, segundo Lessinger & Azeredo-Espin (2000). A clonagem dos fragmentos em plasmídeos foi feita com o kit TA Cloning (INVITROGEN), segundo instruções do fabricante: para a ligação, foi usado 3µL do produto de amplificação, 2µL do plasmídeo pCR 2.1, 1µL de tampão, 1µL de DNA ligase e água q.s.p. 10µL (14°C por 16 horas); a transformação foi feita através de choque químico utilizando tampão KCM/PEG (KCl 1M, CaCl₂·2H₂O 0,3M, MgCl₂ 1M e 10% PEG 6000) em células DH10b; e a seleção das colônias recombinantes foi verificada por digestão dos plasmídeos recuperados através de lise alcalina.

As reações de seqüenciamento foram realizadas utilizando o método Big Dye (Perkin Elmer): 2µL do ABI-PRISM Big Dye terminator sequencing v3.0 (Applied Biosystems), 2µL de tampão de seqüenciamento (Applied Biosystems), 3,5pmol de “primer” (M13D ou M13R), 400ng plasmídeos recuperados e água q.s.p. 10µL para cada reação de seqüenciamento. Cada

produto da reação de seqüenciamento foi purificado através do método isopropanol/etanol e a obtenção das seqüências foi através do seqüenciador automático modelo 377 (Applied Biosystems), segundo instruções do fabricante.

Análises das Seqüências

Os genes de RNAt foram identificados nas seqüências obtidas utilizando o programa tRNA-Scan (Lowe & Eddy 1997). A organização estrutural dos genes de RNAt foi comparada àquela encontrada no genoma mitocondrial de *D. yakuba* (Clary & Wolstenholme 1985). A estrutura da região-controle de *H. irritans* foi analisada conforme descrito a seguir no Capítulo 3.

Capítulo 3

Manuscrito - **Oliveira, M.T.**, Azeredo-Espin, A.M.L. & Lessinger, A.C. "The mitochondrial DNA control region of Muscidae flies: evolution and structural conservation in a dipteran context". Submetido para publicação.

Resumo

A estrutura e evolução da região-controle e de seus genes flanqueadores no DNAm_t de dípteros de importância econômica da família Muscidae (Brachycera: Calyptratae), *Haematobia irritans*, *Musca domestica* e *Atherigona orientalis* estão apresentadas neste trabalho, junto à descrição de regiões intergênicas não-codificadores relacionadas a seqüências flanqueadoras da região-controle no DNAm_t de *Stomoxys calcitrans* e *Ophyra aenescens* (ScIR e OaIR, respectivamente). A presença destas pequenas regiões intergênicas entre o gene de RNAr 12S e o agrupamento de genes RNAt^{Met}, RNAt^{Gln} e RNAt^{Ile} sugere uma organização diferente do DNAm_t de ambas as espécies. A caracterização das seqüências da região-controle de *H. irritans*, *M. domestica* e *A. orientalis* levou a identificação de sete blocos de seqüências conservadas homólogos aos elementos previamente descritos para dípteros Calliphoridae e Oestridae (Brachycera: Calyptratae). Análises comparativas com espécies de *Drosophila* (Brachycera: Acalyptratae) revelaram quatro regiões conservadas em espécies de Brachycera. Os papéis funcionais putativos destes elementos conservados na regulação dos processos de replicação e transcrição são discutidos. A natureza molecular da organização da região-controle do genoma mitocondrial demonstra a plasticidade da molécula de DNAm_t na família Muscidae.

Palavras-chave: DNAm_t, região-controle, região rica em A+T, organização estrutural, Muscidae, Diptera.

The Mitochondrial DNA Control Region of Muscidae Flies: Evolution and Structural Conservation in a Dipteran Context

Marcos T. Oliveira^{1,}, Ana M. L. Azeredo-Espin^{1,2}, Ana C. Lessinger¹*

¹ Laboratório de Genética Animal, Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), Campinas, São Paulo, Brazil.

² Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Campinas, São Paulo, Brazil.

* Corresponding author: 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.

Phone: +55 19 37881141; Fax: +55 19 37881089; E-mail: mto@unicamp.br

Abstract

The structure and evolution of the mtDNA control region (CR) and its flanking genes in economically important dipterans from family Muscidae (Brachycera: Calyptratae), *Haematobia irritans*, *Musca domestica* and *Atherigona orientalis*, are presented in this paper, along with the description of short non-coding intergenic regions possibly related to CR flanking sequences in *Stomoxys calcitrans* and *Ophyra aenescens* mtDNAs (ScIR and OaIR, respectively). The presence of short non-coding intergenic regions between the 12S rRNA gene and the tRNA^{Met}, tRNA^{Gln} and tRNA^{Ile} gene cluster suggests a different organization in the mtDNA of both species. The characterization of *H. irritans*, *M. domestica* and *A. orientalis* control region sequences led to the identification of seven conserved sequence blocks homologous to the elements previously described for Calliphoridae and Oestridae dipterans (Brachycera: Calyptratae). Comparative analysis with *Drosophila* species (Brachycera: Acalyptratae) revealed four conserved regions in Brachyceran species. The putative functional roles of the conserved elements in the regulation of replication and transcription processes are addressed. The molecular nature of the organization of the mitochondrial genome control region demonstrates the plasticity of the mtDNA molecule in family Muscidae.

Key words: mtDNA, control region, A+T-rich region, structural organization, Muscidae, Diptera

Introduction

The mitochondrial DNA (mtDNA) of Metazoan species has a circular, compact organization, conserved gene content, predominantly maternal inheritance, a lack of recombination events, and a high rate of nucleotide substitutions. These features have favored its use as a molecular marker in evolutionary, phylogenetic and population genetics studies concerning a wide range of *taxa* (Wolstenholme 1992, Saccone *et al.* 1999, Boore 1999). Animal mtDNA usually contains 37 genes [13 protein-coding genes, 2 ribosomal RNA (rRNA) genes and 22 transfer RNA (tRNA) genes] and a major non-coding region (Boore 1999). The major non-coding region of vertebrate mtDNA contains mitochondrial transcription and mtDNA replication origin promoters (Tracy and Stern 1995, Shadel and Clayton 1997, Taanman 1999). Consequently, it is known as the control region (CR). Despite the fact that the major non-coding region of other animal mtDNA lacks homology with vertebrate regulatory elements, they are generally known as the control region as well (Boore 1999).

For insects, although much less is known about the mtDNA replication and transcription mechanism, advances in the characterization of nuclear factors involved in these processes are ongoing in the model organism *Drosophila melanogaster* (Kaguni 2004, Garesse and Kaguni 2005). The finding that the origin of replication of *Drosophila* mtDNA maps within the CR (Goddard and Wolstenholme 1978, 1980) led to the detection of conserved elements in the CR of insect mtDNA (also known as the A+T-rich region), which may be involved in replication and transcription processes of this genome (Zhang and Hewitt 1997). Recently, Saito *et al.* (2005) demonstrated the relationship between CR conserved elements and the precise position of the replication origin of seven insect mtDNAs.

Zhang and Hewitt (1997) proposed two insect CR groups : group I, containing two domains - a conserved region near the tRNA^{Met}, tRNA^{Gln} and tRNA^{Ile} (MQI) clustered genes, and

a variable region adjacent to the 12S rRNA gene - and group 2, with conserved elements distributed along the entire CR sequence. Those structural elements seem to be conserved even in more basal insect orders (Schultheis *et al.* 2002).

Aside from genus *Drosophila*, from which the CR sequences of 12 species have been partially or completely sequenced (Clary and Wolstenholme 1985, 1987, Monnerot *et al.* 1990, Monforte *et al.* 1993, Lewis *et al.* 1994, Inohira *et al.* 1997, Brehm *et al.* 2001, Tsujino *et al.* 2002), only two other dipteran *taxa* have a fully characterized CR: genus *Anopheles* (Nematocera), in which the entire CR sequence is very well conserved (Caccone *et al.* 1996), and Calyptratae flies from families Calliphoridae and Oestridae (Lessinger and Azeredo-Espin 2000, Lessinger *et al.* 2004), whose CR structure resembles that of *Drosophila* species (or group 2). Lessinger and Azeredo-Espin (2000) identified eight conserved sequence blocks (CSB) in the CR conserved domain of seven myiasis-causing flies (Calyptratae).

Rearrangements and duplications involving mtDNA regions have been increasingly reported in animals. One of the most common events observed in insect mtDNA has been the occurrence of tandem repeats as part of CR sequences. Tandem repeat unit size and copy number variation is responsible for size variation among CR sequences from different species, or even from individuals of the same species (Zhang and Hewitt 1997). Analysis of this in tandem variation among different geographic samples may reveal important data on the population structure of an insect species (Dotson & Beard 2001; Snäll *et al.* 2002; Mardulyn *et al.* 2003).

Replication errors may lead to the increase or decrease of in tandem unit number and even to the duplication of CR-adjacent sequences. Recently, Lessinger *et al.* (2004) described the presence of duplicated tRNA sequences corresponding to the tRNA^{Ile} gene and a portion of tRNA^{Gln} in the mitochondrial genome of three species from genus *Chrysomya* (Diptera: Calliphoridae). These duplicated genes are located near the 12S rRNA gene, at the opposite end

of the CR from where the “original” tRNA genes are located. These common structural changes involving insect CR sequences are usually responsible for failures in amplification reactions and limit access to CR information.

In this paper, we analyze the mtDNA CR and adjacent genes of economically important Calypttratae flies from family Muscidae (superfamily Muscoidea), comparing them with previously described sequences from families Calliphoridae and Oestridae (superfamily Oestroidea), and from *Drosophila* species (Acalypttratae). The CRs of 11 individuals of the horn fly, *Haematobia irritans*, two of the house fly, *Musca domestica*, and two of the pepper fruit fly, *Atherigona orientalis*, were entirely sequenced, and the evolution and structural organization of Brachycera mtDNA CR are discussed. In the stable fly, *Stomoxys calcitrans*, and in the black dump fly, *Ophyra aenescens*, the comparative analysis of putative CR sequences suggests a different organization of the mtDNA CR and its flanking genes, and the presence of an intergenic region. The structural characterization of Muscidae CR will contribute to increase knowledge about animal mitochondrial genome diversity, evolution and function.

Materials and Methods

Samples and DNA extraction

Muscidae samples consisted of 11 *H. irritans* individuals from different geographic localities in South America and Malaysia, two *M. domestica*, two *A. orientalis*, one *O. aenescens* and one *S. calcitrans* individual (Table 1). Total individual DNA was extracted from pupae or adults as described in Infante-Vargas and Azeredo-Espin (1995), using a phenol/chloroform procedure, with volumes adapted for microcentrifuge tubes.

Table 1. Primer combinations used in the amplification reactions for the recovery of mtDNA CR and flanking genes from Muscidae samples.

Sample (Origin)	Primer combinations	Estimated amplicon size (bp) [£]	Size of the non-coding region between I and 12S (bp) [£]
<i>H. irritans</i>			
7-11 (Caraguatatuba, Brazil)	TI-N-21 [*] + SR-J-14941 [*]	1800	1702
7-12 (Caraguatatuba, Brazil)	TI-N-21 [*] + SR-J-14941 [*]	1300	1251
Hi 16 (Pirassununga, Brazil)	TM-N-193 [#] + SR-J-14233 [#]	2000	1257
	TM-N-193 [#] + SR-J-14612 [#]	1700	
Hi 51 (Goiânia, Brazil)	TM-N-193 [#] + SR-J-14233 [#]	2000	1256
	TM-N-193 [#] + SR-J-14612 [#]	1700	
P1 (Patos, Brazil)	TM-N-193 [#] + SR-J-14233 [#]	2000	1259
	TM-N-193 [#] + SR-J-14612 [#]	1700	
Hi 65 (Rafaela, Argentina)	TI-N-21 [*] + SR-J-14941 [*]	1300	1257
Hi 69 (Libertad, Uruguay)	TI-N-21 [*] + SR-J-14941 [*]	1300	1258
Hi 76 (Libertad, Uruguay)	TI-N-21 [*] + SR-J-14941 [*]	1800	1703
Hi 88 (Barquisimeto, Venezuela)	TI-N-21 [*] + SR-J-14941 [*]	1800	1704
Hi 89 (Barquisimeto, Venezuela)	TI-N-21 [*] + SR-J-14941 [*]	1300	1259
Hi 84 (Kluang, Malaysia)	TI-N-21 [*] + SR-J-14941 [*]	1400	1315
<i>M. domestica</i>			
Md 01 (Sarapuí, Brazil)	TM-N-193 [#] + SR-J-14233 [#]	2000	1287
	TM-N-193 [#] + SR-J-14612 [#]	1700	
Md 02 (Sarapuí, Brazil)	TI-N-24 [#] + SR-J-14941 [*]	1300	1289
<i>A. orientalis</i>			
X41 (Mogi Guaçu, Brazil)	TM-N-193 [#] + SR-J-14612 [#]	1600	1183
	TM-N-193 [#] + SR-J-14941 [*]	1300	
Ao 01 (Mogi Guaçu, Brazil)	TI-N-24 [#] + SR-J-14941 [*]	1200	1073
<i>S. calcitrans</i>			
X35 (S. S. do Paraíso, Brazil)	N2-N-327 [*] + SR-J-14612 [#]	750	96
	TM-N-193 [#] + SR-J-14612 [#]	650	

<i>O. aenescens</i>			
Oa 01 (S. S. do Paraíso, Brazil)	TM-N-193 [#] + SR-J-14612 [#]	900	382

* Primers described in Oliveira *et al.* (in press).

[#] Primers described in Simon *et al.* (1994).

[£] The estimates were done based on the analysis of 1% agarosis gel electrophoreses.

[€] I - tRNA gene for isoleucine; 12S - rRNA gene for 12S subunit; the non-coding regions between I and 12S in the mitochondrial genome of *H. irritans*, *M. domestica* and *A. orientalis* are considered to be the control region (see text for details).

Amplification reactions

The primers used in the amplification reactions for CR sequences and flanking regions for the Muscidae mtDNA are listed in Table 1 and indicated in Figure 1A. The PCRs were done according to conditions described in Oliveira *et al.* (in press). Except for *S. calcitrans* reactions, PCR templates were an aliquot of total individual DNA. To improve the amplification of *S. calcitrans* sequences, the DNA template was a 1:500 dilution from the Long-PCR amplification product (~9.5 Kb) of *S. calcitrans* mtDNA (Barau *et al.* 2005).

Cloning and Sequencing

The 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 using the TA Cloning® kit (Invitrogen), according to manufacturer specifications. The clones were sequenced by automated DNA sequencing in an ABI 377 DNA sequencer using the ABI-PRISM BigDye® terminator sequencing v3.1 (Applied Biosystems). Eight sequencing primers were designed based on *H. irritans* and *M. domestica* CR sequences: HCR-A (5' GTAGATAATTTTTTTTGGCG 3'), HCR-A2 (5' ACTGATAATTTTCCATTA 3'), HCR-B (5' TAAAAATTGGTACTATCTCC 3'), HCR-B2 (5' AAAACAAGCTATTTCTAT 3'), MCR-A (5' TTCTCTATATAAAATTAC 3'), MCR-A2 (5' AATGTTATTATTTTTTTC 3'), MCR-B (5' ATTTATAAAACATACCCC 3'), and MCR-B2 (5' AAAAAAAAAAATTTTCGAGA 3'). Three to five clones were sequenced for each amplified region of the Muscidae species.

Comparative and Structural Analysis

The nucleotide sequences were submitted to similarity searches in GenBank using BLASTN (Altschul *et al.* 1997). The tRNA gene sequences were identified using tRNAscan-SE

software (Lowe and Eddy 1997). The non-coding sequences identified between tRNA^{Ile} and 12S rRNA genes were analyzed by searching for structural elements and conserved regions already described for insects (Lewis *et al.* 1994, Zhang and Hewitt 1997, Inohira *et al.* 1997, Lessinger and Azeredo-Espin 2000, Brehm *et al.* 2001, Tsujino *et al.* 2002), using the softwares Clustal X (Thompson *et al.* 1997), MEME - motif discovery tool (Bailey and Elkan 1994), and Mfold (Zucker 2003) set to default parameters, with manual adjustments when appropriate. The software Dotlet (Junier and Pagni 2000) was used to identify tandem repeats in CR sequences. Other comparative and statistical data were obtained using MEGA3 software (Kumar *et al.* 2004).

Long-PCR amplifications and Restriction Endonucleases Analyses

The whole mitochondrial genomes of *H. irritans*, *S. calcitrans* and *O. aenescens* were amplified via Long-PCR as described in Barau *et al.* (2005). mtDNA sizes were evidenced by electrophoresis in 0.8% agarose gel. For *S. calcitrans*, an aliquot of the Long-PCR products was digested with four restriction endonucleases (*HindIII*, *DdeI*, *ClaI* and *EcoRV*), in different reactions, for 4 hours at 37°C, to predict the location of the extra amount of DNA in this mitochondrial genome. The digested fragments were separated by electrophoresis in 2% agarose gel.

GenBank Accession Numbers

The nucleotide sequences reported in this paper have the following GenBank accession numbers: AY929377 - AY929384 and DQ377072 - DQ377074 for *H. irritans* samples; DQ377075 and DQ377076 for *M. domestica*; DQ377077 and DQ377078 for *A. orientalis*; DQ377079 for *O. aenescens*; and DQ377080 for *S. calcitrans*.

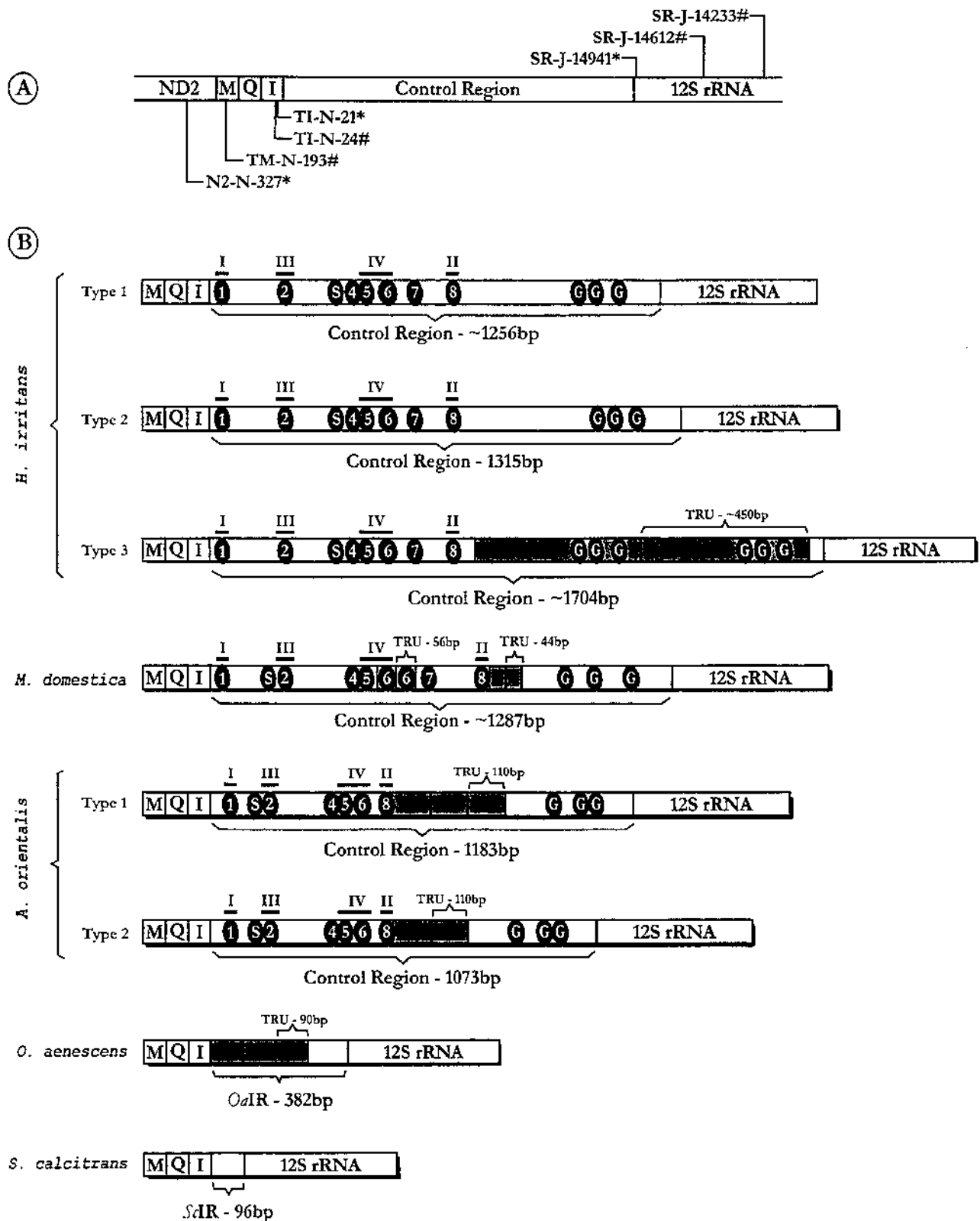


Figure 1. A) Schematic organization of insect mtDNA CR and adjacent genes, indicating the primers used in this study (see Table 1 for primer combinations). The primer nomenclature

follows Simon *et al.* (1994). * refers to primers described in Oliveira *et al.* (in press), and # to primers described in Simon *et al.* (1994). ND2, M, Q, I and 12S rRNA respectively represent genes for subunit 2 of NADH dehydrogenase, tRNA^{Met}, tRNA^{Gln}, tRNA^{Ile} and subunit 12S of rRNA. B) Structural organization of the mitochondrial genome of the five Muscidae species studied here. The conserved structural elements reported in the CR sequences of *H. irritans*, *M. domestica* and *A. orientalis* homologous to the Conserved Sequence Blocks (CSB) of myiasis-causing flies are shown in dark circles and numbered from 1 to 2 and from 4 to 8. S and G elements represent the positions of the predicted secondary structures and of the G islands, respectively. TRU - tandem repeat unit. The bold lines above the CR sequences indicate Brachyceran Conserved Regions (BCR) I - IV. *H. irritans* Type 1 - 3 and *A. orientalis* Type 1 -2 refer to the intraspecific CR size variations. *OaIR* and *ScIR* respectively represent intergenic non-coding regions in the *O. aenescens* and *S. calcitrans* mtDNAs.

Results and Discussion

Structural conservation of Muscidae CR

Figure 1B schematically shows the organization of the Muscidae CR sequences. The size of *O. aenescens* and *S. calcitrans* amplicons were substantially shorter relative to the CR amplicons of the other species. However, Long-PCR analyses indicate mitochondrial genomes of ~16 and 18Kb, respectively (see forthcoming discussion). Searches for CR sequence similarities among *H. irritans*, *M. domestica*, *A. orientalis* and the blowfly species *Cochliomyia hominivorax*, *Co. macellaria*, *Chrysomya megacephala*, *Ch. chloropyga*, *Ch. albiceps*, *Lucilia eximia* (Calliphoridae), and *Dermatobia hominis* (Oestridae) allowed the identification of seven conserved sequence blocks (CSB) out of eight conserved elements previously described in Calyptratae flies (Lessinger and Azeredo-Espin 2000; Lessinger *et al.* 2004). CSB 3 was not

present in these Muscidae species, possibly representing an Oestroidea-specific motif. Interestingly, CSB 7 was not identified in *A. orientalis*. The structural organization of the Muscidae CR resembles that of blowflies, with two domains presenting distinct rates of nucleotide substitutions and indel events (Figure 1B): 1) a conserved domain, containing sequence motifs homologous to blowfly CSBs; and 2) a variable domain, lacking homology with any other mtDNA sequence and showing interspecific length variation. However, we report in the variable domain of the Muscidae CR conserved stretches of four or five Gs, previously described by Brehm *et al.* (2001) in *Drosophila* species. These G stretches (or G islands) also occur in the blowfly CR variable domain, but they vary in number among species. The Muscidae CR most commonly presents three G islands.

A poly-T stretch near the MQI genes, TA(A) tandem repeats and a stem-and-loop secondary structure were insect CR conserved features (Zhang & Hewitt, 1997) also identified in Muscidae. The poly-T stretch (~ 20 Ts) corresponds to CSB 1, and the TA(A) region is homologous to CSB 6. The predicted Muscidae secondary structures (Figure 1B, S element) are represented in detail in Figure 2. In Muscidae species, the most energetically stable predicted structures do not correspond to any CSB, in contrast to the predicted structures described for blowfly CR (Lessinger *et al.* 2004). Moreover, *H. irritans*, *M. domestica* and *A. orientalis* CR secondary structures are not conserved in primary sequence or relative position (Figure 1B). Nevertheless, these elements are structurally very similar to those of other dipterans (Figure 2). The involvement of secondary structure in the replication initiation of insect mtDNA was experimentally supported in Orthoptera species (Saito *et al.* 2005).

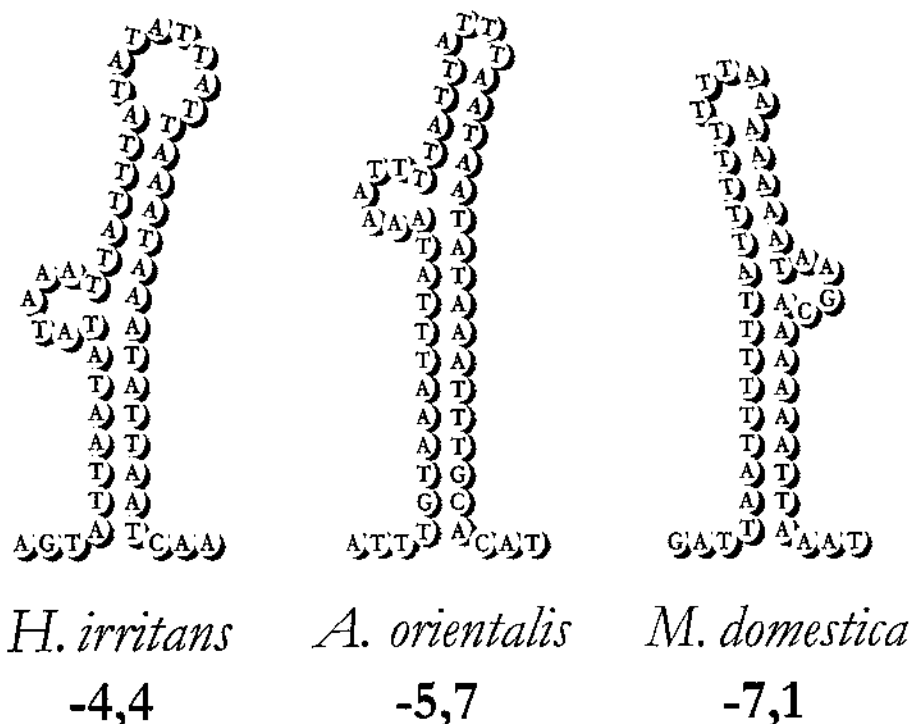


Figure 2. The predicted secondary structures identified in the CR sequences of *H. irritans*, *M. domestica* and *A. orientalis*. The relative positions of these structures are shown as an S in Figure 1. The free energy values (kcal.mol⁻¹) are shown below each structure.

In order to extend our analysis to other non-Calyptratae dipterans, we compared the mtDNA CR of Muscidae, Calliphoridae and Oestridae species to those of seven *Drosophila* species, *D. melanogaster* (GenBank accession number: J01404), *D. yakuba* (NC_001322), *D. subobscura* (AJ132900), *D. madeirensis* (AJ132902), *D. guanche* (AJ132901), *D. teissieri* (X54011) and *D. virilis* (X05914). In addition to the previously described G island motifs (Brehm *et al.* 2001), four other conserved regions were identified. These regions were named “Brachyceran Conserved Regions” (BCR), due to their potential to be recognizable in other brachyceran CR. BCR I and II are extensive arrays of Ts located on opposite DNA strands, which correspond to CSB 1 and 8, respectively (Figure 1B). BCR III and IV seem to be homologous to

a 300bp conserved region that was identified in several *Drosophila* species with short CR (Clary and Wolstenholme 1987; Monnerot *et al.* 1990; Monforte *et al.* 1993; Brehm *et al.* 2001), and in the type II repeat element of *Drosophila* species with longer CR (Lewis *et al.* 1994, Inohira *et al.* 1997, Tsujino *et al.* 2002). Figure 3 shows the alignment and the consensus sequence of BCR III and IV.

The identification of conserved sequences among Muscidae, Calliphoridae, Oestridae and Drosophilidae could indicate the similar functional roles of these elements in regulating higher Diptera mtDNA transcription and replication processes. Clary and Wolstenholme (1987) argued that the location of the initial poly-T (BCR I) near the MQI genes might promote the recognition of a transcription origin oriented to the flanking tRNA cluster. Similarly, BCR II might regulate transcription through the 12S rRNA direction. Recently, Saito *et al.* (2005) precisely mapped the replication origin of both strands of *Drosophila* mtDNA to the first nucleotides downstream of both poly-T elements (BCR I and II). Other evidence from crosslinking protection experiments (Potter *et al.* 1980, Pardue *et al.* 1984), S1 nuclease assays (Monforte *et al.* 1993, Tsujino *et al.* 2002) and electron microscopy (Goddard and Wolstenholme 1978, 1980) in *Drosophila* species mapped the mtDNA replication origin of *Drosophila* CR in the 300bp conserved region. We suggest that BCR III and IV (equivalent to the CSBs 2, 5 and 6 of Calypttratae species) are homologous to the *Drosophila* 300bp conserved region, reinforcing the suggestion of Lewis *et al.* (1994) on the site-specific protein-DNA interaction properties of this region. It was remarkable that the predicted secondary structures were consistently located within or near BCR III and IV in all Brachyceran species (Figure 1B, Clary and Wolstenholme 1987, Monforte *et al.* 1993, Inohira *et al.* 1997, Tsujino *et al.* 2002, Lessinger *et al.* 2004).

Intraspecific variability of CR sequences

The analysis of *H. irritans* intraspecific variability based on mtDNA CR sequences identified few nucleotide substitutions among South American samples from the same species: uncorrected p-distances ranging from 0 to 0.32% for *H. irritans*; 0.16% for *M. domestica*; and 0% for *A. orientalis*. In contrast, extreme sequence divergences were found for *H. irritans* between South American and Malaysian CR haplotypes (from 9.68 to 9.93%).

Significant CR length variation was evident in *H. irritans*, and was mainly due to duplicated elements. Most individuals analyzed presented a Type 1 CR (Figure 1B), with ~1256bp. Amplification reactions of large numbers of South America samples of *H. irritans* (Oliveira *et al.* in press) identified the Type 1 CR as the most frequent haplotype. The Type 2 CR was specific to the Malaysian mtDNA, with a CR larger than Type 1 (1315bp), due to multiple insertions in the variable domain. The *H. irritans* Type 3 CR was shared by three geographically distant South American samples: Barquisimeto (Venezuela), Libertad (Uruguay) and Caraguatatuba (Brazil) (Table 1), despite its low frequency in these populations. The largest-sized Type 3 CR (~1704bp) was the result of a ~450bp in tandem duplication which has taken place in the variable domain of *H. irritans* CR. Tandem duplications are a common feature of insect CR (Zhang and Hewitt 1997). Variation in tandem repeat numbers was observed in the CR sequences of the two individuals of *A. orientalis* (110bp unit), Types 1 and 2 (Figure 1B). Regarding *M. domestica* samples, tandem duplications of 56bp were also identified involving CSB 6, despite its putative function in mtDNA replication control.

The low variability among South American *H. irritans* CRs, and their significant divergence in comparison to the Malaysian type, are expected considering the extreme geographic distance between these samples. Historically, *H. irritans* is a new invader species in the South American continent (Valério and Guimarães, 1983), and founder effects could be

responsible for the low levels of genetic variation found in these samples. In spite of that, the occurrence of multiple introductions could not be discarded; in this case, other factors may be responsible for reduced horn fly variability.

Muscidae mitochondrial genome organization

Extreme size variations among Muscidae species and among individuals from the same species (Table 1 and Figure 1B) were found. The species *H. irritans*, *M. domestica* and *A. orientalis* presented a structural organization and gene order of the CR and flanking genes which was the same as described for the mitochondrial genome of *D. yakuba* (Clary and Wolstenholme 1985), considered to be ancestral for arthropod mtDNA. In these genomes, the CR is flanked by tRNA gene cluster MQI and the 12S rRNA gene. Figure 1B illustrates the mtDNA CR sequences and flanking genes of the Muscidae species.

For *O. aenescens* and *S. calcitrans* mtDNA, we consistently amplified an extremely short CR amplicon, which could be the result of rearrangements involving CR flanking sequences – and primer hybridization sites, consequently – or PCR artifacts. In these species, the CR flanking genes MQI and 12S were easily identified by tRNA structure prediction and local alignment with the other Muscidae homologous sequences described here, as well as those of other dipterans: *Co. hominivorax* (GenBank accession number: AF260826), *Ch. chloropyga* (AF352790) and *D. yakuba* (X03240). The gene order found in *O. aenescens* and *S. calcitrans* CR flanking genes was the same described for other dipterans (Figure 1B). However, 382 and 96bp non-coding regions were identified between MQI and 12S genes in the mtDNA of *O. aenescens* and *S. calcitrans*, respectively. These non-coding regions could not be assigned as the putative mtDNA CR, as they lack the conserved elements described for insect CR in general (Zhang and Hewitt 1997), for Calypttratae dipterans (Lessinger and Azeredo-Espin 2000, Lessinger *et al.* 2004), and for

Muscidae species (discussed above). Searches for sequence similarities in GenBank provided no identity match regarding these non-coding regions.

Based on these analyses, we suggest that these non-coding regions should be addressed as mtDNA ‘intergenic regions’ (*OaIR* and *ScIR* for *O. aenescens* and *S. calcitrans*, respectively), in spite of the occurrence of other insect CR sequences shorter than 96 bp (Shao *et al.* 2001). Due to the compactness of animal mtDNA, intergenic regions are generally short and infrequent, but large intergenic regions have already been described in the mtDNA of some insect species, such as *Triatoma dimidiata* (310bp; Dotson and Beard 2001), *Apis mellifera* (193bp; Crozier and Crozier 1993), *Bemisia tabaci*, and *Trialeurodes vaporariorum* (170 and 136bp, respectively; Thao *et al.* 2004).

In combination with these analyses, we performed Long-PCR amplification of the complete mtDNA of *O. aenescens* and *S. calcitrans*, according to Barau *et al.* (2005). The Long-PCR products of *O. aenescens* showed that the size of its mitochondrial genome was similar to other Muscidae species (~16Kb). Long-PCR amplification of *S. calcitrans* samples revealed a mtDNA molecule 1.5 - 2 Kb larger than other Muscidae mitochondrial genomes. Restriction endonuclease mapping of *S. calcitrans* mtDNA suggests that the extra ~2Kb sequence was inserted between the ND2 and 12S rRNA genes (data not shown). Size differences among insect mtDNAs are frequently related to differences in CR length (Zhang and Hewitt 1997), however, duplications involving complete or partial gene sequences have also been reported (Lessinger *et al.*, 2004; Shao *et al.* 2005).

The complete sequencing of *O. aenescens* and *S. calcitrans* mtDNA is being conducted to investigate the structural and evolutionary aspects of its general organization, and to fully characterize potential CR sequences.

Conclusions

Although the PCR strategies used to recover mtDNA CR sequences of the Muscidae species *H. irritans*, *M. domestica* and *A. orientalis* were successful, structural features related to *O. aenescens* and *S. calcitrans* mtDNA limited access to CR data. The identification of *OaIR* and *ScIR*, and the analysis of Long-PCR amplifications of the complete mitochondrial genomes of *O. aenescens* and *S. calcitrans*, suggest that their CR and flanking genes (ND2, M, Q, I and 12S) may be involved in duplications and/or rearrangements that occurred during the evolution of the mtDNA structural organization. These results imply a structural plasticity and diversity in the mtDNA molecule of family Muscidae, and should contribute to further understanding of insect mitochondrial genome evolution.

The recognition of previously described CSBs of myiasis-causing flies (Brachycera: Calyptratae: Calliphoridae and Oestridae) in the mtDNA CR of *H. irritans*, *M. domestica* and *A. orientalis* (Brachycera: Calyptratae: Muscidae), in addition to the identification of conserved regions shared with *Drosophila* species (Brachycera: Acalyptratae: Drosophilidae), support the suggestion that these conserved sequences are maintained by selective functional constraints, due to their potential *cis*-regulatory roles in mtDNA replication and/or transcription processes. The identification of these structural homologies among Muscidae, Calliphoridae, Oestridae and Drosophilidae implies that Brachyceran species have a similar mechanism for the regulation of these processes.

Acknowledgements

The authors thank those who provided Muscidae samples: Mariana L. Lyra, Pablo Fresia, Santos Gama, Stella Lanzzeri, Estela Martines, Arnaldo Maggi, Alberto Guglielmone, Alfredo Coronado, Patrícia Thyssen and Ana C. M. Junqueira. We also thank Rosângela A. Rodrigues, Cristina F. Abreu and Fabio O. Dias for technical assistance; and Cláudio J. B. Carvalho and Arício X. Linhares for the taxonomic identification of *O. aenescens* and *A. orientalis*, respectively. This work, as well as A.C.L., were supported by the Programa Especial de Estímulo à Fixação de Doutores sponsored by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (PROFIX/CNPq, grant no. 540.602/01-9). A.M.L.A.E. was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, grant no. 03/01458-9), and M.T.O. by a fellowship from Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

References

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein data base search programs. *Nucleic Acid Res* 25: 3389-3402.
- Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymer. *Proceedings of the Second International Congress on Intelligent System for Molecular Biology*, pp. 28-36, AAAI Press, Menlo Park, California.
- Barau JG, Azeredo-Espin AML, Lessinger AC (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. *Mol Ecol Notes*, DOI: 10.1111/j.1471-8286.2005.011003.x.
- Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Res* 27: 1767-1780.
- Brehm A, Harris DJ, Hernández M, Cabrera VM, Larruga JM, Pinto FM, González AM (2001) Structure and evolution of the mitochondrial DNA complete control region in the *Drosophila subobscura* subgroup. *Insect Mol Biol* 10: 573-578.
- Caccone A, García BA, Powell JR (1996) Evolution of mitochondrial DNA control region in the *Anopheles gambiae* complex. *Insect Mol Biol* 5: 51-59.
- Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. *J Mol Evol* 22: 252-271.
- Clary DO, Wolstenholme DR (1987) *Drosophila* mitochondrial DNA: conserved sequences in the A+T-rich region and supporting evidence for a secondary structure model of the small ribosomal RNA. *J Mol Evol* 25:116-125.
- Crozier RH, Crozier YC (1993) The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* 133: 97-117.

- Dotson EM, Beard CB (2001) Sequence and organisation of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. *Insect Mol Biol* 10: 205-215.
- Garesse R, Kaguni LS (2005) A *Drosophila* model of mitochondrial DNA replication: proteins, genes and regulation. *IUBMB Life* 57: 555-561.
- Goddard JM, Wolstenholme DR (1978) Origin and direction of replication in mitochondrial DNA molecules from *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 75: 3886-3890.
- Goddard JM, Wolstenholme DR (1980) Origin and direction of replication in the mitochondrial DNA molecules from the genus *Drosophila*. *Nucleic Acids Res* 8: 741-757.
- Infante-Vargas ME, Azeredo-Espin AML (1995) Genetic variability in mitochondrial DNA of screwworm, *Cochliomyia hominivorax* (Diptera: Calliphoridae), from Brazil. *Biochem Genet* 33: 237-256.
- Inohira K, Hara T, Matsuura ET (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.
- Junier T, Pagni M (2000) Dotlet: diagonal plots in a Web browser. *Bioinfo Appl Note* 16: 178-179.
- Kaguni LS (2004) DNA polymerase gamma, the mitochondrial replicase. *Annu Rev Biochem* 73: 293-320.
- Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinfo* 5: 150-163.
- Lessinger AC, Azeredo-Espin AML (2000) Evolution and structural organization of mitochondrial DNA control region of myiasis-causing flies. *Med Vet Entomol* 14: 71-80.
- Lessinger AC, Junqueira ACM, Conte FF, Azeredo-Espin AML (2004) Analysis of a conserved duplicated tRNA gene in the mitochondrial genome of blowflies. *Gene* 339: 1-6.

- Lewis DL, Farr CL, Farquhar AL, Kaguni LS (1994) Sequence, organization, and evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA. *Mol Biol Evol* 11: 523-538.
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequences. *Nucleic Acids Res* 25: 955-964.
- Mardulyn P, Termonia A, Milinkovitch, MC (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.
- Monforte A, Barrio E, Latorre A (1993) Characterization of the length polymorphism in the A + T-rich region of the *Drosophila obscura* group species. *J Mol Evol* 36: 214-223.
- Monnerot M, Solignac M, Wolstenholme DR (1990) Discrepancy in divergence of the mitochondrial and nuclear genomes of *Drosophila teissieri* and *Drosophila yakuba*. *J Mol Evol* 30: 500-508.
- Oliveira MT, da Rosa AC, Azeredo-Espin AML, Lessinger AC. Improving access to the control region and tRNA gene clusters of dipteran mitochondrial DNA. *J Med Entomol*, in press.
- Pardue ML, Fostel JM, Cech TR (1984) DNA-protein interactions in the *Drosophila virilis* mitochondrial chromosome. *Nucleic Acids Res* 12: 1991-1999.
- Potter DA, Fostel JM, Berninger M, Pardue ML, Cech TR (1980) DNA-protein interactions in the *Drosophila melanogaster* mitochondrial genome as deduced from trimethylpsoralen crosslinking patterns. *Proc Natl Acad Sci USA* 77: 4118-4122.
- 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.
- Saito S, Tamura K, Aotsuka T (2005) Replication origin of mitochondrial DNA in insects. *Genetics* 171: 1695-1705.

- Schultheis AS, Weigt LA, Hendricks AC (2002) Arrangement and structural conservation of 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 GS, Clayton DA (1997) Mitochondrial DNA maintenance in vertebrates. *Annu Rev Biochem* 66: 409-435.
- Shao R, Campbell NJ, Barker SC (2001) Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). *Mol Biol Evol* 18: 858-865.
- Shao R, Mitani H, Barker SC, Takahashi M, Fukunaga M (2005) Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the chigger mite, *Leptotrombidium pallidum*. *J Mol Evol* 60: 764-773.
- Snäll N, Huoponen K, Savontaus M-L, Ruohomäki, K (2002) Tandem repeats and length variation in the mitochondrial DNA control region of *Epirrita autumnata* (Lepidoptera: Geometridae). *Genome* 45: 855-861.
- Taanman JW (1999) The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* 1410: 103-123.
- Thao ML, Bauman L, Bauman P (2004) Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). *BMC Evol Biol* 4: 25.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876-4882.
- Tracy RL, Stern DB (1995) Mitochondrial transcription initiation: promoter structures and RNA polymerases. *Curr Genet* 28: 205-216.

- Tsujino F, Kosemura A, Inohira K, Hara T, Otsuka YF, Obara MK, Matsuura ET (2002) Evolution of the A+T-rich region of mitochondrial DNA in the *melanogaster* species subgroup of *Drosophila*. J Mol Evol 55: 573-583.
- Valério JR, Guimarães JH (1983) Sobre a ocorrência de uma nova praga, *Haematobia irritans* (L.) (Diptera: Muscidae), no Brasil. Revta Bras Zool 1: 417-418.
- Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. Int Rev Cytol 141: 173-216.
- Zhang D-X, Hewitt GM (1997) Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochem Syst Ecol 25: 99-120.
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 1-10.

Metodologia detalhada

Amostras

As amostras de *H. irritans*, *M. domestica*, *A. orientalis*, *O. aenescens* e *S. calcitrans* usadas neste capítulo foram as mesmas usadas no Capítulo 2 (Tabela 1), além de dois indivíduos adultos (congelados -70°C) de *H. irritans* de Caraguatatuba, SP, coletados em Novembro/2002 por mim e M.L. Lyra.

Extração do DNA

O DNA total dos indivíduos foi extraído conforme descrito na Metodologia detalhada do Capítulo 1.

Amplificações

As reações de amplificação da região-controle e genes flanqueadores do DNAm_t das espécies de Muscidae foram realizadas conforme descrito na Metodologia detalhada do Capítulo 2. Os “primers” utilizados estão descritos na “Table 1” e esquematizados na “Figure 1A” do manuscrito.

Para *S. calcitrans*, ao invés de uma alíquota (1µl) do DNA total do indivíduo, foi usada na reação uma alíquota do produto de amplificação da reação de Long-PCR com os “primers” H16SC e HCOIII (Barau *et al.* 2005), diluído 1:500, contendo um fragmento de cerca de 9-10Kb do DNAm_t de *S. calcitrans*. As reações de Long-PCR foram realizadas como descrito a seguir.

Clonagem e Seqüenciamento

A clonagem e o seqüenciamento dos fragmentos amplificados neste capítulo foram realizados como descrito na Metodologia detalhada do Capítulo 2.

Long-PCR

Para avaliar o tamanho do DNAm_t em espécies de Muscidae, o genoma mitocondrial completo de *H. irritans*, *S. calcitrans* e *O. aenescens* foi amplificado via Long-PCR em dois fragmentos de 8-10Kb cada (Figura 3), assim como descrito em Barau *et al.* (2005). As reações de Long-PCR foram realizadas com um volume final de 50µl utilizando os termocicladores PTC-200, MiniCycler (MJ Research) e GeneAmp PCR System 9600 (Perkin Elmer). Foram utilizados

os seguintes reagentes (concentrações finais): 10µL tampão B (2mM MgSO₄, 60mM Tris-SO₄ e 18mM (NH₄)₂SO₄), 2U da enzima *Elongase* (Invitrogen), 300µM de cada dNTP, 400nM de cada “primer”, e 1µl da amostra de DNA total. As condições das reações foram: desnaturação a 92°C por 2min (1 ciclo); desnaturação a 92°C por 20s, hibridização a 42°C por 1min, e extensão a 60°C por 16 min (35 ciclos); extensão final a 60°C por 6min (1 ciclo). Um volume de 1µl de cada reação foi visualizado em gel de agarose 0,8%, corado com brometo de etídeo.

Uma diluição (1:500) do Produto A de Long-PCR, no qual teoricamente se encontra a região-controle de *S. calcitrans* (Figura 3), foi usada como DNA molde para as reações descritas no item *Amplificações* deste capítulo.

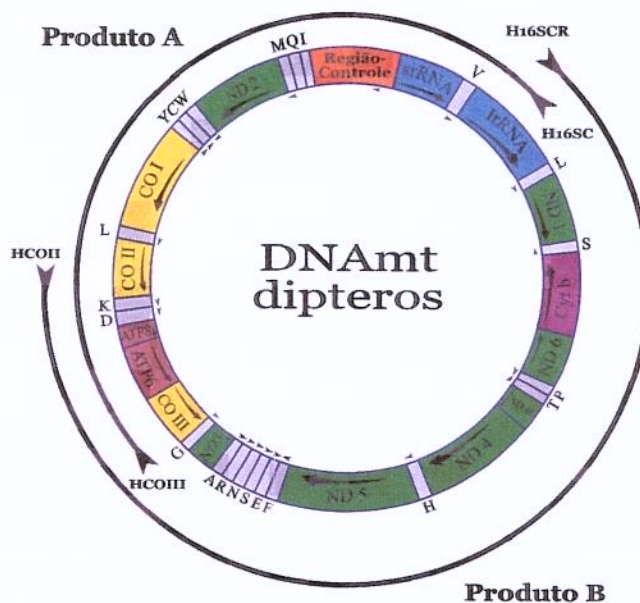


Figura 3. Representação esquemática do genoma mitocondrial de dípteros mostrando os genes amplificados em cada produto, A e B, através das reações de Long-PCR descritas em Barau *et al.* (2005). H16C, H16CR, HCOII e HCOIII são os “primers” utilizados nas reações.

Análises do Padrão de Restrição

Para estimar em que região do genoma mitocondrial de *S. calcitrans* está inserido a quantidade de DNA que o caracteriza como 2Kb maior (~18Kb) que o genoma mitocondrial de *H. irritans*, o Produto A da amplificação via Long-PCR foi digerido com quatro endonucleases de restrição (*Hind*III, *Dde*I, *Cla*I e *Eco*RV) em diferentes reações por ~4 horas a 37°C. As reações de digestão foram realizadas num volume final de 15µL, usando 10U de enzima

(Invitrogen), 1,5µL de tampão (segundo o fabricante), 3µL do produto de Long-PCR e água q.s.p. 15µL. Os fragmentos gerados pelas digestões foram separados em gel de agarose 2% e corados com brometo de etídeo (Figura 4).

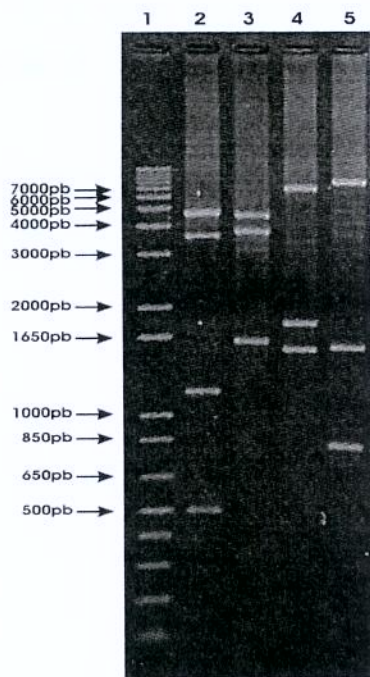


Figura 4. Resultado da digestão do Produto A (Long-PCR) amplificado a partir do DNAm de *S. calcitrans*. As colunas de 1 a 5 representam, respectivamente: marcador de peso molecular Ladder 1Kb PLUS, Produto A digerido com *HindIII*, *DdeI*, *ClaI* e *EcoRV*.

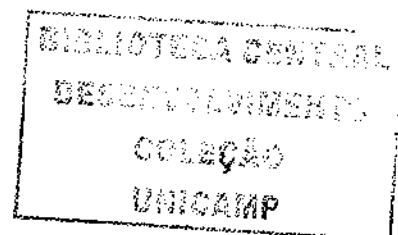
Análises das Seqüências

As seqüências nucleotídicas obtidas através desta estratégia de amplificação (“Table 1” e “Figure 1A” do manuscrito) foram submetidas a buscas por similaridade no GenBank, utilizando o aplicativo BLASTN (Altschul *et al.* 1997). Este tipo de busca nem sempre é eficiente no que se refere a análises da região-controle. Por se tratar de uma região não codificadora e muito rica em A e T, estas análises podem encontrar seqüências homoplásicas. Juntamente a esta análise, elementos estruturais descritos para a região-controle de insetos (Zhang & Hewitt 1997) e de outros dípteros Calyptratae (Lessinger & Azeredo-Espin 2000), tais como estruturas secundárias e regiões conservadas, serviram de base para a identificação de homologias e caracterização da estrutura. As seqüências dos genes de RNAt flanqueadores a região-controle foram identificados através do programa tRNA-Scan (Lowe & Eddy 1997)

As análises comparativas entre as seqüências de *H. irritans*, *S. calcitrans*, *M. domestica*, *A. orientalis*, *O. aenescens* e seqüências de outras espécies de dípteros (Brachycera: Calyptratae) já caracterizadas foram realizadas inicialmente com base no resultado de alinhamentos múltiplos

obtidos com o auxílio do programa CLUSTAL X (Thompson *et al.* 1997), com eventuais ajustes manuais. Muitas vezes, o alinhamento “clássico” não é suficiente para identificar regiões homólogas entre as seqüências das regiões-controle analisadas. Para contornar este problema, o aplicativo MEME (Bailey & Elkan 1994), uma ferramenta de identificação de motivos conservados, foi utilizado nas análises. Regiões cujas seqüências eram similares e de tamanho maior que 10pb foram analisadas quanto à posição e repetição na seqüência da região-controle.

O aplicativo MFOLD (Zucker 2003) foi utilizado para a verificação de estruturas secundárias presentes nas seqüências das regiões-controle. A ocorrência de elementos repetidos nesta região foi investigada com o auxílio do programa DOTLET (Junier & Pagni 2000), baseado no alinhamento diagonal de pares de seqüências nucleotídicas. As análises referentes à natureza das substituições nucleotídicas (transições X transversões), ao conteúdo de A e T das seqüências de nucleotídeos e outras inferências estatísticas foram realizadas através dos programas MEGA2 e MEGA3 (Kumar *et al.* 2001, 2004).



Capítulo 4

Manuscrito - **Oliveira, M.T.**, Barau, J.G., Feijão, P.C., da Rosa, A.C., Abreu, C.F., Azeredo-Espin, A.M.L. & Lessinger, A.C. “The mitochondrial genomes of *Haematobia irritans* and *Stomoxys calcitrans*: structure and evolution in Muscidae flies (Diptera: Calyptratae)”. Em preparação.

Resumo

Neste trabalho, nós apresentamos os dois primeiros genomas mitocondriais de dípteros da família Muscidae, a mosca-dos-chifres *Haematobia irritans* e a mosca-dos-estábulo *Stomoxys calcitrans*. Estes genomas são típicos DNAmT de insetos na maioria dos aspectos analisados, tais como alto conteúdo de A+T da sequência completa (79,1 e 78,9%, respectivamente) e dos genes codificadores de proteína (77,7 e 77,0; maiores valores entre os dípteros), preferência pelo uso de códons ricos em A+T, e uso de códons não otimizado. Outras pressões de mutação além da pressão a favor de uma composição rica em A+T foram verificadas no DNAmT de Muscidae e uma associação com a evolução dos anticódons dos RNAt de metazoários foi proposta. Considerando o tamanho, conteúdo gênico e estrutura, o genoma mitocondrial de *H. irritans* (16078pb) é idêntico ao arranjo genômico ancestral inferido para artrópodes e sua região-controle é consistente com o modelo descrito para dípteros Brachycera. Entretanto, o genoma mitocondrial de *S. calcitrans* é ~2Kb maior, caracterizado pela presença de repetições em “tandem” de ~550pb na região-controle e por uma cópia extra do RNAt-Ile, restando ainda a caracterização de uma sequência de ~1,5Kb contida na região-controle. O sinal filogenético dos genes codificadores de proteínas mitocondriais foi utilizado para inferir as relações em Diptera.

Palavras-chave: DNAmT, mitogenômica, evolução molecular, duplicação de gene de RNAt, região-controle, Muscidae.

The mitochondrial genomes of *Haematobia irritans* and *Stomoxys calcitrans*: structure and evolution in Muscidae flies (Diptera: Calyptratae)

Key words: mtDNA, mitogenomics, molecular evolution, tRNA gene duplication, control region, Muscidae.

Marcos Túlio Oliveira^{1,*}, Joan Grande Barau¹, Pedro Cipriano Feijão¹, Aline Coelho da Rosa¹, Cristina Feix Abreu¹, Ana Maria L. Azeredo-Espin^{1,2} and Ana Cláudia Lessinger¹

¹ *Laboratório de Genética Animal, Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), Campinas, São Paulo, Brazil.*

² *Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Campinas, São Paulo, Brazil.*

* Corresponding author: 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.

Phone: +55 19 37881141; Fax: +55 19 37881089; E-mail: mto@unicamp.br

Abbreviations: *12S* and *16S rRNA*, genes for 12S and 16S subunits of ribosomal RNA; *ATP6* and *8*, ATPase subunits 6 and 8; *CDS*, protein-coding genes; *COI-III*, cytochrome c oxidase subunits I-III; *CSB*, conserved sequence blocks; *CYTB*, cytochrome b; *NDI-6* and *4L*, NADH dehydrogenase subunits 1–6 and 4L; *ScIR*, *Stomoxys calcitrans* intergenic region; *tRNA*, transfer RNA gene.

Abstract

In this paper, we present the first two mitochondrial genomes of Muscidae dipterans, the horn fly *Haematobia irritans* and the stable fly *Stomoxys calcitrans*. Typical insect mtDNA aspects were evident, such as the high A+T content of the entire genome (79.1 and 78.9%, respectively) and of the protein-coding genes (77.7 and 77.0; highest values among dipterans), the preference for A+T-rich codons, and the evidence of a non-optimal codon usage. Other mutation pressures besides the A+T enrichment were verified in Muscidae mtDNA and an association with *tRNA* anticodon evolution in metazoans is proposed. Regarding the total size, gene content and structural organization, *H. irritans* mitochondrial genome (16078bp) is similar to the hypothetical ancestral arthropod mitochondrial genome and its control region structure is consistent with that described for Brachycera dipterans. In the other hand, the mitochondrial genome of *S. calcitrans* is ~2Kb longer (18Kb), characterized by the presence of ~550bp tandem repeats in the control region and an extra copy of *tRNA-Ile*. The final characterization of ~1Kb from *S. calcitrans* control region is still missing. A selected set of mitochondrial protein-coding genes were used for inferring Diptera relationships.

1. Introduction

In animals, mitochondrial DNA (mtDNA) is generally a small and compact genome (~16Kb) containing 37 genes [13 protein-coding genes (*CDS*), 2 ribosomal RNA genes (*rRNA*) and 22 transfer RNA genes (*tRNA*)] and a main non-coding region, which is usually called the control region. The predominance of maternal inheritance, lack of extensive recombination, and accelerated rates of nucleotide substitution are features that have favored the use of the mtDNA as an informative evolutionary marker (Saccone et al., 1999).

During the past few years, mitochondrial genomics of arthropods has extensively increased due to the investigative interests in the mitochondrial genome organization and evolution (Boore, 1999) and in the use of complete sequences in phylogenetic inferences (Hwang et al., 2001a; Nardi et al., 2003b; Hassanin, 2006). The increasing amount of data about complete mitochondrial genomes of arthropods have led to the possibility of comparing a set of “genome-level characters” (concatenated *CDS* alignment, *RNA* secondary structures, modes of control of replication and transcription, and gene order) (Boore et al., 2005), to new methods for mtDNA isolation, such as Long PCR-based strategies (Hwang et al., 2001b; Roehrdanz et al., 2002; Yamauchi et al., 2004; Barau et al., 2005), and to the development of web-based softwares for mtDNA annotation (Wyman et al., 2004) and comparative analysis (Feijão et al., 2006).

The mitochondrial genome of the fruit fly *Drosophila yakuba* (Clary and Wolstenholme, 1985) is considered to be ancestral based on its structural organization. Rearrangements relative to this hypothetical ancestral organization are frequently found in the insect orders Hemiptera, Phthiraptera and Hymenoptera (Crozier and Crozier, 1993; Shao et al., 2001; Thao et al., 2004; Covacin et al., 2006), in the Decapoda crustaceans (Hickerson and Cunningham, 2000; Miller et al., 2004; Sun et al., 2005), and in the Ixodidae arachnids (Black IV and Roehrdanz, 1998; Shao et al., 2005a). In the order Diptera (Insecta), the mtDNA gene order of Brachycera differs from

that of Nematocera species only by the inversion of *tRNA-Ser^{AGN}* (Mitchell et al., 1993). The main structural differences among dipteran mitochondrial genomes, as well as among insects, are related to the size and number of control regions and the arrangements of its structural motifs. For blowflies from the genus *Chrysomya* (Brachycera: Calliphoridae), a unique gene content was recently described (Lessinger et al., 2004; Junqueira et al., 2004). These mtDNAs were characterized by the duplication of a region containing the *tRNA-Gln* and *tRNA-Ile* gene, and a few nucleotides from the control region. Among arthropods, the control region and its flanking genes are known to be involved in numerous *tRNA* gene translocations (Boore, 1999), which are associated to technical problems concerning the amplification, cloning and sequencing of mitochondrial genomes (Boore et al., 2005).

The dipteran species whose mitochondrial genomes are completely sequenced are flies of great medical and economic importance, such as three Anopheline species (Mitchell et al., 1993; Beard et al., 1993; Krzywinski et al., 2006), the Mediterranean fruitfly, *Ceratitis capitata* (Spanos et al., 2000), the olive fruitfly, *Bactrocera oleae* (Nardi et al., 2003a), the primary screwworm fly, *Cochliomyia hominivorax* (Lessinger et al., 2000), the blowfly, *Chrysomya chloropyra* (Junqueira et al., 2004), and the botfly, *Dermatobia hominis* (**NC 006378**), and model organisms, such as *Drosophila yakuba* (Clary and Wolstenholme, 1985), *D. melanogaster* (Lewis et al., 1995), *D. simulans*, *D. mauritiana* and *D. sechellia* (Ballard, 2000). In this work, we describe the complete mitochondrial genome of the horn fly *Haematobia irritans* and the almost complete mitochondrial genome of the stable fly *Stomoxys calcitrans* (L.), both veterinary and economic important flies from the family Muscidae (Brachycera: Calyptratae). These species are common blood-sucking ectoparasites of bovine livestock throughout the New World and their economic relevance is related to the stress caused in the hosts and the loss in meat and milk production (Haufe, 1987). Our analyses focused on the characterization of the evolution and

structural organization of *H. irritans* and *S. calcitrans* mitochondrial genomes, and the evaluation of complete mtDNA sequences for dipteran phylogeny. These results may contribute in promoting molecular marker analysis in horn fly and stable fly populations, which could be useful for monitoring and controlling these parasites; moreover, the characterization of Muscidae mitochondrial genomes increases the knowledge about the structure and evolution of animal mitochondrial genomes.

2. Materials and Methods

2.1. Samples, DNA extraction, amplifications and sequencing

Adult individuals of *H. irritans* from Caraguatatuba, SP, Brazil, and *S. calcitrans* from São Sebastião do Paraíso, MG, Brazil, were collected in farms directly from the hosts. Total individual DNA was extracted as described in Infante-Vargas and Azeredo-Espin (1995), using a phenol/chloroform procedure, with volumes adapted for microcentrifuge tubes.

The whole mitochondrial genome of one individual of each species were amplified via Long PCR as described in Barau *et al.* (2005), using the primers H16SA + HCOIII (Amplicon A) and H16SB + HCOII (Amplicon B) for *H. irritans*, and H16SC + HCOIII (Amplicon A) and H16SCR + HCOII (Amplicon B) for *S. calcitrans*. An 800bp fragment of the *16S rRNA* gene including the annealing sites of the *16S* Long PCR primers was amplified and sequenced according to Barau *et al.* (2005) to fill the gap between these primers. To improve the sequencing of the *ND2/control* region and *ND6/ND4* sequences of *S. calcitrans* mitochondrial genome, the primers C1-N-2191 + SR-J-14233 (Simon *et al.*, 1994) (Amplicon C) and CB-N-10920 (Simon *et al.*, 1994) + N4-J-9403 (first described here: 5'TAAGAAATTATATCTATACC3') were used, respectively. New primers designed to anneal in the control region sequence of *S. calcitrans* [ScCR-1 (5'ATAGAAAGATAGAGGGAC3'), ScCR-2 (5'GTTAAGGTTATTAGTCCG3'), and

MuscCR (5'ATTTTTTTTGGCTYTRGAT3') were used in combination with *ND2* and *12S rRNA* annealing primers to extend the sequencing of control region and flanking genes. The amplification reactions of the mtDNA of these two Muscidae species are schematized in Figure 1.

Approximately thirteen micrograms of mtDNA were used to construct genomic libraries by the shotgun approach (Anderson et al., 1982) for these species. The Long PCR products were shared by sonication and fragments of 1-2Kb were isolated from a low-melting point agarose electrophoresis gel. The selected fragments were blunted and cloned using the pMOS*Blue* blunt ended cloning kit (Amersham Biosciences), according to the manufacturer's instructions. The clones were sequenced automatically in the ABI Prism 377 and 3700 sequencers using the Big Dye terminator kit (PE Applied Biosystems).

2.2. Assembling, annotation, sequence analysis and phylogenetic inferences

The final consensus sequences of the mitochondrial genomes of *H. irritans* and *S. calcitrans* were assembled using the phred/phrap/consed package (Gordon et al., 1998). The *CDS*, *rRNA* and *tRNA* genes were identified and annotated using DOGMA (Wyman et al., 2004) and Sequin Application (<ftp://ftp.ncbi.nih.gov/sequin>). Predicted tRNA secondary structures were confirmed by tRNAScan-SE 1.1 program (Lowe and Eddy, 1997). The control region analyses were conducted according to Oliveira et al. (submitted).

Comparative analysis of mtDNA structure, nucleotide composition, amino acid content, codon usage biases and sequence alignments were done using AMiGA database tools (Feijão et al., 2006). Amino acid and nucleotide sequences were analyzed as a set of concatenated *CDS* from all dipteran species and used for phylogenetic inferences. Phylogenetic trees were reconstructed with maximum parsimony (MP) and neighbor-joining (NJ) tree-building methods, using MEGA3 (Kumar et al., 2004). The NJ trees using nucleotide sequences were inferred with the p-distance and Kimura-2-parameters, 1000 replicates for bootstrap statistics.

2.3. Genbank accession numbers

The nucleotide sequences of *H. irritans* and *S. calcitrans* mitochondrial genomes have the following GenBank accession numbers: NC_007102 and DQ533708, respectively.

3. Results and Discussion

3.1. Muscidae mitochondrial genomes

The mitochondrial genome of *H. irritans* was completely sequenced (16078bp - Figure 1) and resembles that of *D. yakuba* in terms of its structural organization and composition. The size of the mitochondrial genome of *S. calcitrans* was estimated to be around 18Kb, but by the sequencing of *S. calcitrans* mtDNA, we were able to provide the characterization of 16790bp (~94% of the genome), which include 13 *CDS*, 2 *rRNA*, 23 *tRNA* genes [including an extra copy of the *tRNA-Ile* gene (*tRNA-Ile2*)] and partial sequences of the control region. The total size of the *S. calcitrans* mtDNA was experimentally confirmed by independent analyses of standard PCR products including the control region and LongPCR-RFLP results from additional *S. calcitrans* samples (Oliveira et al., submitted).

The mitochondrial genomes of both Muscidae species are heavily biased towards a high A+T content. The overall A+T content was 79.1% for *H. irritans* and 78.9% for *S. calcitrans*, two of the A+T-richest genomes among dipterans (Table 1). Regarding only the *CDS* genes, *H. irritans* has the A+T-richest mtDNA in Diptera with 77.7% and *S. calcitrans* mtDNA has 77.0%. The base composition measured by the AT and GC skewness (Perna and Kocher, 1995) showed virtually no AT skew when the major strands (which encodes for 9 *CDS* and 14 *tRNA* genes) of *H. irritans* and *S. calcitrans* mitochondrial genomes were considered (0.004 and -0.001, respectively), indicating no preference for one of both nucleotides. Conversely, the GC skews registered the occurrence of more Cs than Gs (-0.125 and -0.140, respectively), a common feature of animal mitochondrial genomes (Saccone et al., 1999).

Despite control region sequences, noncoding sequences are generally rare and characterized by few nucleotides in metazoan mtDNA. The Table 2 shows the noncoding and overlapping regions of *H. irritans* and *S. calcitrans* mtDNAs. It is important to notice the presence of a relatively large noncoding region (96bp) in the *S. calcitrans* mtDNA, which is consistently recovered in PCR analysis in spite of the expected ~3Kb control region amplicon, and has none of the potential regulatory elements described for insect mtDNA (Zhang and Hewitt, 1997; Lessinger and Azeredo-Espin, 2000; Oliveira et al., submitted).

Table 1. A+T content, AT and GC skew values of mitochondrial genomes of dipteran species. 3CDS+ and 3CDS- respectively represent the third codon position of major and minor strand CDS.

Species	A+T content (%)		AT skew		GC skew	
	Whole mtDNA	CDS	Major strand	3CDS+	Major strand	3CDS-
<i>H. irritans</i>	79.1	77.7	0.004	-0.026	-0.125	-0.455
<i>S. calcitrans</i>	78.9	77.0	-0.001	-0.037	-0.140	-0.457
<i>C. hominivorax</i>	76.9	74.7	0.034	0.015	-0.207	-0.753
<i>C. chloropyga</i>	76.7	75.0	0.020	0.027	-0.170	-0.610
<i>D. hominis</i>	77.8	75.3	0.043	0.073	-0.227	-0.675
<i>C. capitata</i>	77.5	75.6	0.021	-0.027	-0.185	-0.642
<i>D. yakuba</i>	78.6	76.7	0.005	-0.019	-0.136	-0.408
<i>A. gambie</i>	77.6	76.0	0.032	-0.041	-0.154	-0.467

Table 2. Noncoding regions and overlapping gene sequences in *H. irritans* and *S. calitrans* mitochondrial genomes.

Noncoding regions			Overlapping sequences		
Between genes	Strands	Size (bp)	Between genes	Strands	Size (bp)
<i>H. irritans</i>					
<i>tRNA-Ile / tRNA-Gln</i>	major / minor	3	<i>tRNA-Trp / tRNA-Cys</i>	major / minor	8
<i>tRNA-Gln / tRNA-Met</i>	minor / major	8	<i>tRNA-Tyr / COI</i>	minor / major	2
<i>tRNA-Cys / tRNA-Tyr</i>	major / major	1	<i>ATP8 / ATP6</i>	major / major	7
<i>tRNA-Leu^{TTR} / COII</i>	major / major	4	<i>ATP6 / COIII</i>	major / major	1
<i>COIII / tRNA-Gly</i>	major / major	6	<i>tRNA-Ala / tRNA-Arg</i>	major / major	1
<i>ND3 / tRNA-Ala</i>	major / major	1	<i>tRNA-Asn / tRNA-Ser^{AGN}</i>	major / major	1
<i>tRNA-Arg / tRNA-Asn</i>	major / major	7	<i>ND4 / ND4L</i>	minor / minor	7
<i>tRNA-Glu / tRNA-Phe</i>	major / minor	18	<i>ND6 / CYTB</i>	major / major	1
<i>ND5 / tRNA-His</i>	minor / minor	15	Total		28
<i>ND4L / tRNA-Thr</i>	minor / major	2			
<i>tRNA-Pro / ND6</i>	minor / major	2			
<i>CYTB / tRNA-Ser^{TCN}</i>	major / major	2			
<i>tRNA-Ser^{TCN} / ND1</i>	major / minor	19			
<i>ND1 / tRNA-Leu^{CTN}</i>	minor / minor	1			
<i>tRNA-Leu^{CTN} / 16S rRNA</i>	minor / minor	3			
<i>16S rRNA / tRNA-Val</i>	minor / minor	1			
<i>tRNA-Val / 12S rRNA</i>	minor / minor	1			
Total		94			

<i>S. calcitrans</i>				
<i>tRNA-Gln / tRNA-Met</i>	minor / major	3	<i>tRNA-Ile / tRNA-Gln</i>	major / minor 3
<i>tRNA-Cys / tRNA-Tyr</i>	major / major	2	<i>tRNA-Trp / tRNA-Cys</i>	major / minor 8
<i>tRNA-Leu^{TTR} / COII</i>	major / major	4	<i>tRNA-Tyr / COI</i>	minor / major 2
<i>tRNA-Lys / tRNA-Asp</i>	major / major	4	<i>ATP8 / ATP6</i>	major / major 7
<i>COIII / tRNA-Gly</i>	major / major	6	<i>ATP6 / COIII</i>	major / major 1
<i>tRNA-Arg / tRNA-Asn</i>	major / major	3	<i>tRNA-Ala / tRNA-Arg</i>	major / major 1
<i>tRNA-Glu / tRNA-Phe</i>	major / minor	18	<i>tRNA-Asn / tRNA-Ser^{AGN}</i>	major / major 1
<i>ND5 / tRNA-His</i>	minor / minor	15	<i>tRNA-Ser^{AGN} / tRNA-Glu</i>	major / major 1
<i>ND4L / tRNA-Thr</i>	minor / major	2	<i>ND4 / ND4L</i>	minor / minor 7
<i>tRNA-Pro / ND6</i>	minor / major	7	Total	31
<i>ND6 / CYTB</i>	major / major	4		
<i>tRNA-Ser^{TCN} / ND1</i>	major / minor	16		
<i>ND1 / tRNA-Leu^{CTN}</i>	minor / minor	1		
<i>tRNA-Leu^{CTN} / 16S rRNA</i>	minor / minor	6		
<i>16S rRNA / tRNA-Val</i>	minor / minor	1		
<i>tRNA-Val / 12S rRNA</i>	minor / minor	1		
<i>12S rRNA / tRNA-Ile2 (ScIR)</i>	minor / major	96		
Total		189		

3.1.1. *rRNA and protein-coding genes*

The *rRNA* genes of these Muscidae species are highly conserved within dipterans, although they are among the shortest genes (1321bp for *16S rRNA* and 784bp for *12S rRNA* in *H. irritans*; and 1325 and 782bp in *S. calcitrans*, respectively). The nucleotide sequence identities are more than 95% for both genes between muscid species and more than 78% among other dipterans. The A+T contents are: 82.7% in the *16S rRNA* and 78.9% in the *12S rRNA* of *H. irritans*; 82.6 and 79.5% for *S. calcitrans* 16SrRNA and 12S rRNA, respectively.

As well as *rRNA* genes, the *CDS* genes of Muscidae mitochondrial genomes are conserved: 95% of amino acid and 91% of nucleotide sequences identities. When all dipteran are compared, the similarities showed an average of 61% for amino acid and 54% for nucleotide sequences. The genes *COI* and *COIII* are the most conserved in nucleotide sequences (61 and 60%, respectively), and the genes *ND6*, *ATP8* and *ND2* are the most variable (37, 40 and 41%, respectively).

The initiation of *CDS* genes in the Muscidae mitochondrial genomes uses the standard invertebrate mtDNA code in almost all *CDS*, except one (*COI*). Start codons are conserved in both species: 7 genes initiates with ATG, 5 with ATT and 1 with TCG (Table 3). In contrast to other dipteran mtDNAs (Clary and Wolstenholme, 1985; Spanos et al., 2000; Lessinger et al., 2000; Junqueira et al., 2004), the recognized start codons ATC and ATA were not used. The putative initiation site for the gene *COI* was the same non-standard TCG triplet described for other muscid species in Oliveira et al. (2005). This gene has no identifiable standard start codon in many species and alternative start signals for insect *COI* translation have not been experimentally confirmed in most species, relying only on predicted amino acid sequences and comparative sequence analysis (Clary and Wolstenholme, 1985; Flook et al., 1995; Hickerson and Cunningham, 2000; Nardi et al., 2001). However, Morlais and Severson (2002) and

Krzywinski et al. (2006) recently confirmed that the most probable translation initiation site of the messenger RNA of *COI* is the TCG triplet in mosquito species.

These Muscidae species have both complete stop codons (TAA or TAG) and incomplete stop codons adjacent to *tRNA* genes as termination signals for mitochondrial *CDS* (Table 3). The mtDNA of *H. irritans* and *S. calcitrans* shared eleven stop codons; only *ND3* and *CYTB* have distinct terminations. It's assumed that polyadenylation of mtDNA mRNAs may restore stop codon information after tRNA releases from the polycistronic transcript (Ojala et al., 1981).

3.1.2. Codon usage, *tRNA* anticodons and nucleotide bias

The codon usage of *H. irritans* and *S. calcitrans* mitochondrial genomes (Table 4) are similar to that of other insects. First, the strong bias in nucleotide composition towards an A+T accumulation is clearly evident at the neutral (or nearly neutral) third codon position. In average, 95% of the nucleotides at this site are A or T. Second, amino acids coded by A+T-rich codon families (Phe, Ile, Met, Tyr, Asn and Lys) are more abundant in Muscidae *CDS* than amino acids coded by G+C-rich codon families (Pro, Ala, Arg and Gly). The ratio G+C / A+T-rich codons was 0.30 in both species, which is lower than the 0.42-0.43 ratio of dipteran mtDNA (Crozier and Crozier, 1993; Lessinger et al., 2000). Finally, the absence of some G+C-rich codons is commonly reported in insect mtDNA: in *S. calcitrans*, as well as in most dipteran mtDNA, the codons AGG, ACG, TGC and CTG were not used, however, in *H. irritans* mitochondrial genome, the number of non-used codons is exceptionally high (GGC, GAG, GTC, AGG, CGG, CGC, CAG, CTG, CTC and CCG). Comparative analysis showed that only the hymenopteran *Melipona bicolor* mtDNA (Silvestre, 2002), which is the A+T-richest genome described in insects, has more non-used codons than *H. irritans* (total = 13).

Table 3. Recognized initiation and termination signals of mitochondrial CDS of Muscidae flies. AA represents the corresponding amino acid.

Genes	Initiation (AA)		Termination	
	<i>H. irritans</i>	<i>S. calcitrans</i>	<i>H. irritans</i>	<i>S. calcitrans</i>
<i>ND2</i>	ATT(I)	ATT(I)	T + <i>tRNA-Trp</i>	T + <i>tRNA-Trp</i>
<i>COI</i>	TCG(S)	TCG(S)	T + <i>tRNA-Leu^{TTR}</i>	T + <i>tRNA-Leu^{TTR}</i>
<i>COII</i>	ATG(M)	ATG(M)	T + <i>tRNA-Lys</i>	T + <i>tRNA-Lys</i>
<i>ATP8</i>	ATT(I)	ATT(I)	TAA	TAA
<i>ATP6</i>	ATG(M)	ATG(M)	TAA	TAA
<i>COIII</i>	ATG(M)	ATG(M)	TAA	TAA
<i>ND3</i>	ATT(I)	ATT(I)	TAA	T + <i>tRNA-Ala</i>
<i>ND5</i>	ATT(I)	ATT(I)	T + <i>tRNA-Phe</i>	T + <i>tRNA-Phe</i>
<i>ND4</i>	ATG(M)	ATG(M)	T + <i>tRNA-His</i>	T + <i>tRNA-His</i>
<i>ND4L</i>	ATG(M)	ATG(M)	TAA	TAA
<i>ND6</i>	ATT(I)	ATT(I)	TAA	TAA
<i>CYTB</i>	ATG(M)	ATG(M)	TAA	T + <i>tRNA-Ser^{TCN}</i>
<i>NDI</i>	ATG(M)	ATG(M)	TAG	TAG

Table 4. Codon usage of *H. irritans* (Hi) and *S. calcitrans* (Sc) CDS. The anticodon of the corresponding tRNA is shown in non-capitalized letters in parentheses. The anticodon that matches to the most used codon in a family is underlined.

AA	Codon	Hi	Sc	AA	Codon	Hi	Sc	AA	Codon	Hi	Sc	AA	Codon	Hi	Sc
Phe	TTT	312	315	Ser	TCT	123	112	Tyr	TAT	163	158	Cys	TGT	32	32
(gaa)	TTC	22	28	(tga)	TCC	6	7	(gta)	TAC	15	13	(gca)	TGC	1	0
Leu	TTA	541	514		TCA	97	106	Stop	TAA	7	5	Trp	TGA	96	93
(taa)	TTG	11	27		TCG	1	3		TAG	1	1	(tca)	TGG	1	4
Leu	CTT	26	44	Pro	CCT	82	80	His	CAT	64	64	Arg	CGT	12	10
(tag)	CTC	0	1	(tgg)	CCC	2	3	(gtg)	CAC	8	8	(tgg)	CGC	0	1
	CTA	22	22		CCA	51	51	Gln	CAA	71	66		CGA	44	42
	CTG	0	0		CCG	0	1	(ttg)	CAG	0	4		CGG	0	3
Ile	ATT	351	337	Thr	ACT	114	110	Asn	AAT	198	194	Ser	AGT	42	38
(gat)	ATC	6	14	(tgt)	ACC	1	7	(gtt)	AAC	9	14	(gcu)	AGC	2	3
Met	ATA	220	210		ACA	73	73	Lys	AAA	72	80		AGA	58	61
(cat)	ATG	13	19		ACG	2	0	(ctt)	AAG	19	15		AGG	0	0
Val	GTT	95	85	Ala	GCT	122	107	Asp	GAT	62	59	Gly	GGT	60	56
(tac)	GTC	0	2	(tgc)	GCC	3	2	(gtc)	GAC	5	7	(tcc)	GGC	0	1
	GTA	98	109		GCA	55	66	Glu	GAA	75	74		GGA	148	139
	GTG	1	2		GCG	1	1	(ttc)	GAG	0	1		GGG	10	20

The fact that C nucleotides are overrepresented (relative to Gs) in the major strands of *H. irritans* and *S. calcitrans* mtDNAs, and, consequently, minor strands are G-rich implies in the identification of another composition bias of dipteran codon usage. We separately analyzed CDS genes coded by the major and the minor strand and calculated the AT and GC skews using only nucleotides from the third codon position (Table 1). The AT skew analysis evidenced no significant differences in A related to T content at neutral sites, despite slightly more Ts than As in the minor strand. However, the GC skews indicated that there is 3.8 times more C than G, in average, in major strand neutral sites, and 5.4 times more G than C at minor strand neutral sites. Codon usage bias was also verified in the vertebrate mitochondrial genome (Reyes et al., 1998; Xia, 2005), in which a mutation pressure towards the accumulation of A and C (A+C mutation pressure) in the major strand (or L-strand) was suggested. This phenomenon occurs due to spontaneous deaminations that frequently occur when the mtDNA minor strand is left single-stranded for an extended period of time during mtDNA replication (Shadel and Clayton, 1997), promoting the accumulation of the A→G and C→T mutations in this strand and responding to the strand-specific biases (Reyes et al., 1998; Xia, 2005).

Codon-anticodon adaptations are characteristic of vertebrate mtDNA (also known as optimal codon usage), i.e., the most used codon in a degenerate codon family perfectly matches the anticodon of the corresponding *tRNA* (Bulmer, 1987). Xia (2005) proposed the *selection hypothesis of anticodon adaptation* to explain *tRNA* anticodon evolution in vertebrates (since the leading hypothesis, the *mutation hypothesis of anticodon evolution*, would predict non-optimal codon usage under the A+C mutation pressure of the vertebrate mtDNA). Interestingly, the *tRNA* anticodons of *H. irritans* and *S. calcitrans* were the same reported to vertebrate mtDNA (Table 4), excluding that of *tRNA-Lys* (CTT). Species from other insect orders, such as the hymenopteran *Apis mellifera* (Crozier and Crozier, 1993), the hemipteran *Philaenus spumarius*

(Stewart and Beckenbach, 2005), the phthirapteran *Heterodoxus macropus* (Shao et al., 2001) and the thysanopteran *Thrips imagines* (Shao and Barker, 2003), also share this conserved set of anticodons except for the *tRNA-Lys* (CTT or TTT). This conservation may indicate that anticodon evolution in metazoan mitochondrial genomes could be under the same operational forces, but this argument potentially conflicts with Xia selection hypothesis.

Different from the vertebrate mitochondrial genome, non-optimal codon usage is a common feature of insect mtDNA, where the most used codon does not perfectly matches the corresponding *tRNA* anticodon in 15 of the 22 degenerate codon families (Table 4). This pattern is clearly caused by the strong A+T mutation pressure of insect mitochondrial genomes.

Our data agrees with the *selection hypothesis of anticodon versatility* (Agris, 2004; Tong and Wong, 2004), which we believe that better explains the evolution of metazoan mtDNA anticodons, since it could possibly explain both vertebrate and invertebrate mitochondrial genomes. According to this hypothesis, the versatility of one single tRNA in translating codons from degenerate codon families (as is the case for the mtDNA) is essential for the translation machinery, so more versatile anticodons (T or G in the wobble position) would have been selected. This hypothesis could also explain the vertebrate anticodon evolution as emphasized by Xia (2005). In the insect mtDNA, the A+T mutation pressure disrupts the optimization of codon usage, what is not necessarily prejudicial for the translation process if it evolved based on versatile anticodons that efficiently translate the mitochondrial genes. In *H. irritans* and *S. calcitrans* mtDNA, the wobble position is a T in 8 out of 9 anticodons for tRNAs of four-fold degenerate codon families and in 4 out of 6 of two-fold degenerate codon families ended by a purine; a G is in 7 out of 8 anticodons used to translate two-fold degenerate codon families ended by a pyrimidine. These data strongly agrees with the *selection hypothesis of anticodon versatility*. An A+T-rich mitochondrial genome (efficient in replication and transcription processes) is a

better choice for insects than an optimal codon usage, regarding the nature of the selective pressures operating in these organisms. In insect mitochondrial genomes, the codon usage (via AT-enrichment) seems to evolve independently of the anticodon (versatility evolution).

3.1.3. The control region and genome organization

The control region of *H. irritans* mtDNA was 1259bp in size and corresponds to the Type 1 control region previously described by Oliveira et al. (submitted). The *H. irritans* control region is typical of Brachycera dipterans, containing four brachyceran conserved regions (BCR), a predicted secondary structure and three G islands (Figure 2). These elements were suggested to be involved in the regulation of the transcription and/or replication processes of insect mtDNA (Lewis et al., 1994; Zhang and Hewitt, 1997; Lessinger and Azeredo-Espin, 2000; Saito et al., 2005; Oliveira et al., submitted). For *S. calcitrans*, 938bp of the 5' end of the control region (flanked by the *tRNA-Met*, *-Gln* and *-Ile* gene cluster) and 839bp of the 3' end (flanked by the *tRNA-Ile2* and *12S rRNA* genes) were characterized. Approximately 65% of the predicted ~2.5Kb control region of *S. calcitrans* was sequenced but, despite being incomplete, the identification of conserved elements in *S. calcitrans* control region was possible (Figure 2A and B) and revealed a structure similar to that of Brachycera (Oliveira et al., submitted). New features, such as tandem repeats of about 550bp containing the BCR III and IV elements, were described in the *S. calcitrans* control region. The exact number of repeats is unknown, but they could account for the difference in size of *H. irritans* and *S. calcitrans* mitochondrial genomes. Long tandem repeats have already been described in *Drosophila melanogaster* control region resulting in a 19517bp mtDNA (Lewis et al., 1994, 1995).

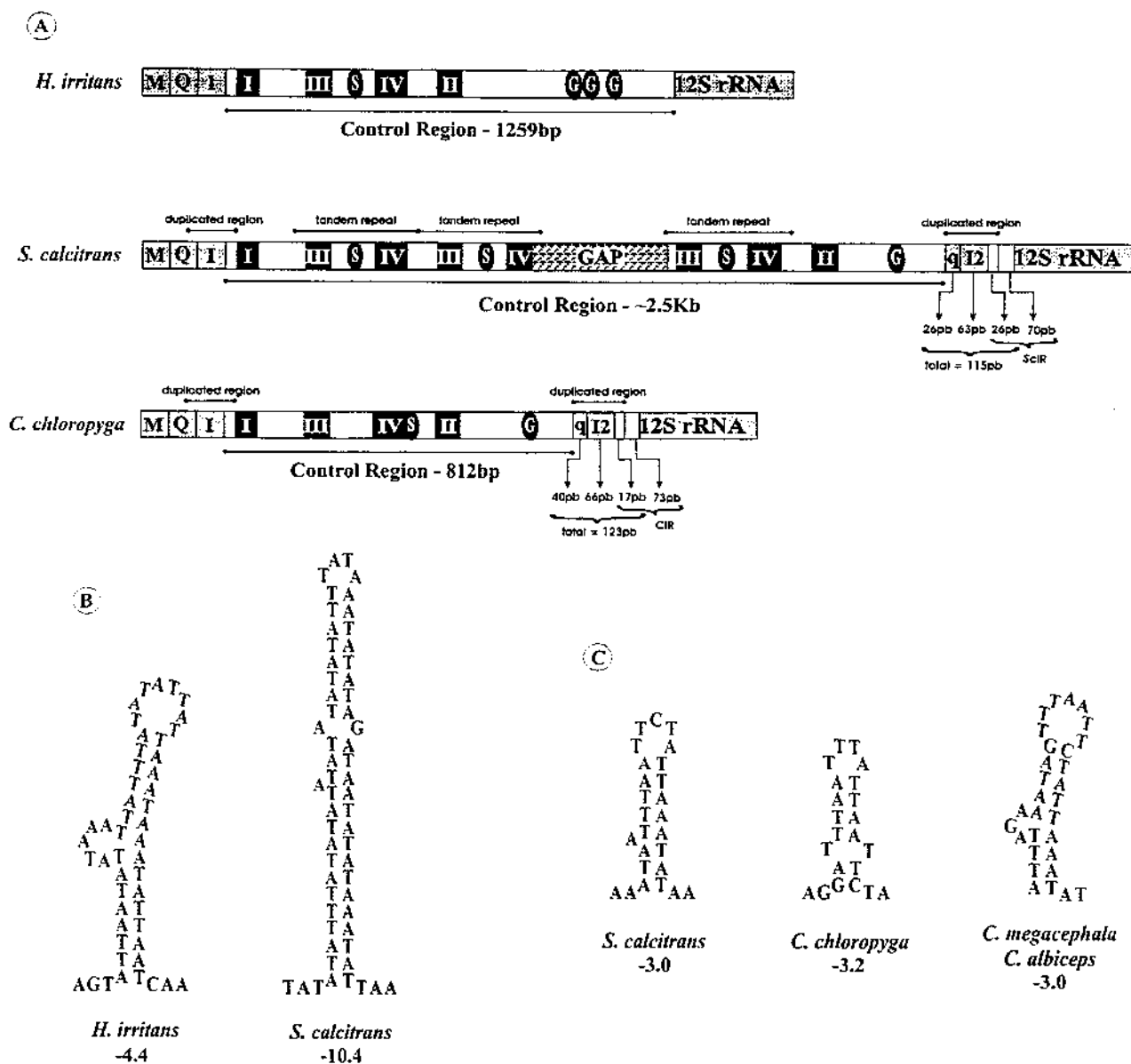


Figure 2. A) Comparison of the structural organization of the mtDNA control region of *H. irritans*, *S. calcitrans* and *Chrysomya chloropyga*. The control region flanking genes *tRNA-Met* (M), *tRNA-Gln* (Q), *tRNA-Ile* (I) and *12S rRNA* are represented in grey boxes. Black boxes numbered from I to IV indicate the brachyceran conserved regions (BCR; Oliveira et al., submitted), and the black circles indicate predicted secondary structures (S) and the G elements. The duplication of *tRNA-Ile* (I2) and associated sequences (q = partial *tRNA-Gln*; ScIR or CIR = *S. calcitrans* or *Chrysomya* intergenic region, respectively) are presented. The *S. calcitrans* gap on control region sequence is shown in the spotted box. B) In detail, the secondary structure of Muscidae control regions. The numbers below each structure indicate the free energy values (kcal.mol^{-1}). C) Predicted secondary structures formed at the duplication endpoints in the *S. calcitrans*, *C. chloropyga*, *C. megacephala* and *C. albiceps* mtDNA.

Another unusual feature of *S. calcitrans* mtDNA is the presence of an additional copy of the *tRNA-Ile* gene flanking the control region 3' end, near to the *12S rRNA* gene (Figure 2A). The *tRNA-Ile2* gene is part of an 115bp duplicated region, including 26bp from *tRNA-Gln* (partial), 63bp from *tRNA-Ile* and 26bp from the control region 5' end. There is a 96bp sequence, named *ScIR*, between *tRNA-Ile2* and *12S rRNA* genes (Oliveira et al., submitted). The gene arrangement described in Oliveira et al. (submitted) for *S. calcitrans* mtDNA (*ND2*, *tRNA-Met*, *-Gln*, *-Ile*, *ScIR* and *12S rRNA*) disagrees with the partial control region results retrieved from the complete mitochondrial genome analysis based on shotgun strategies as described here.

Curiously, the duplication identified in *S. calcitrans* mtDNA is similar to that described for *Chrysomya* species (Lessinger et al., 2004; Junqueira et al., 2004). Comparative analysis of these duplicated regions and flanking sequences revealed a convergent structure: partial sequence of *tRNA-Gln*, including the T Ψ C loop; the entire sequence of *tRNA-Ile*; partial sequence of the control region, which includes the sequence upstream the poly-T element (BCR I); and the presence of an intergenic region of unknown origin between the duplicated region and the *12S rRNA* gene. Comparing the endpoints of the duplicated region of *S. calcitrans* and *Chrysomya* species, we observed that the sequence adjacent to the control region (partial *tRNA-Gln*) has the ability to form a putative hairpin structure (Figure 2C). Sequences with potential to form stem-and-loop secondary structures have already been associated with duplications in animal mitochondrial genomes (Stanton et al., 1994). The duplicated sequence endpoint adjacent to the intergenic regions does not form any evident stable secondary structure, but it is notable that the motif associated with this duplication endpoint (directly upstream to BCR I) is similar to the precise origin of major strand replication of *Drosophila* mtDNA (Saito et al., 2005). In *Drosophila*, the major strand replication initiates after 97% of the minor strand synthesis is complete (Goddard and Wolstenholme, 1980), in a sequence between *tRNA-Ile* and BCR I.

Based on these evidences, we suggest that the convergent organization of *S. calcitrans* and *Chrysomya* species mtDNAs results from the formation of a secondary structure in the minor strand (template) during mtDNA replication, which may interrupt the synthesis of the new major strand and excise to double-stranded fragment. In this case, illegitimate inter-mtDNA recombination (Shao et al., 2005b) could be invoked to account for the insertion of the duplicated region adjacent to control region 3' end. A mechanism operated by tandem duplication and random loss of redundant copies (Macey et al., 1998) would have required a drastic reorganization of duplicated control region sequences and flanking genes, in addition to extensive deletions and degeneration of the duplicated sequences coincident with the conservation of *tRNA-Ile* copies.

Some questions about this unusual organization shared by *S. calcitrans* and *Chrysomya* mtDNAs remain unclear: why was the duplicated region inserted at the exactly same position in the mitochondrial genome of both dipteran lineages? Would this region be more prone to breaks? Moreover, what is the origin of the ~70bp intergenic sequence between the duplicated region and the *12S rRNA* gene? Is there a functional role associated to it? Could the duplicated region provide an alternative origin of replication?

3.2. Phylogenetic perspectives

Aligned nucleotide and amino acid sequences for 15 dipterans and the lepidopteran *Bombyx mori* (included as an outgroup; **AB070264**) were analyzed and used to evaluate the phylogenetic relationships recovered using mitochondrial *CDS* data. The use of a complete set of mitochondrial *CDS* sequences as a reliable molecular marker for inferring phylogenetic trees has increased in recent years (Boore et al., 2005). Two criteria were used to select a set of candidate genes for inferring different phylogenies: conserved/variable amino acid ratio > 0.5 (Nardi et al.,

2003b), which in this case excluded *ATP8*, *ND2* and *ND6*, and ratio > 1 (Junqueira et al., 2004), excluding more four genes (*ND3*, *ND4*, *ND4L* and *ND5*).

The topology showed in Figure 3 was retrieved using distance (p-distance or Kimura-2-parameter) and tree-building methods (NJ or MP) for both nucleotide and amino acid sequences, for which the results diverge only in relation to the bootstrap value of the Acalyptratae clade. These analyses divided Diptera in the suborders Brachycera (flies) and Nematocera (mosquitoes), as expected. Within Brachycera, Acalyptratae and Calyptratae species were distributed as distinct monophyletic clusters. The grouping of Acalyptratae species in a monophyletic clade is contradictory (Yeates and Wiegmann, 1999; Nardi et al., 2003b), but our data agree with other phylogenies based on mtDNA sequences (Lessinger et al., 2000; Junqueira et al., 2004). The distribution of Calyptratae species disrupts the monophyly of Oestroidea. According to McAlpine and Wood (1989), the Calyptratae section has three morphologically supported monophyletic superfamilies, including Muscoidea (Muscidae: *H. irritans* and *S. calcitrans*), Oestroidea (Calliphoridae: *C. chloropyra* and *C. hominivorax*; Oestridae: *D. hominis*) and Hippoboscoidea (no mtDNA available so far). However, these phylogenetic analyses clustered Calliphoridae and Muscidae species, excluding *De. hominis* (Oestroidea: Oestridae), that was placed as a separate branch. Similar results were found in other works using single mtDNA and/or nuclear gene sequences (Bernasconi et al., 2000; Nirmala et al. 2001).

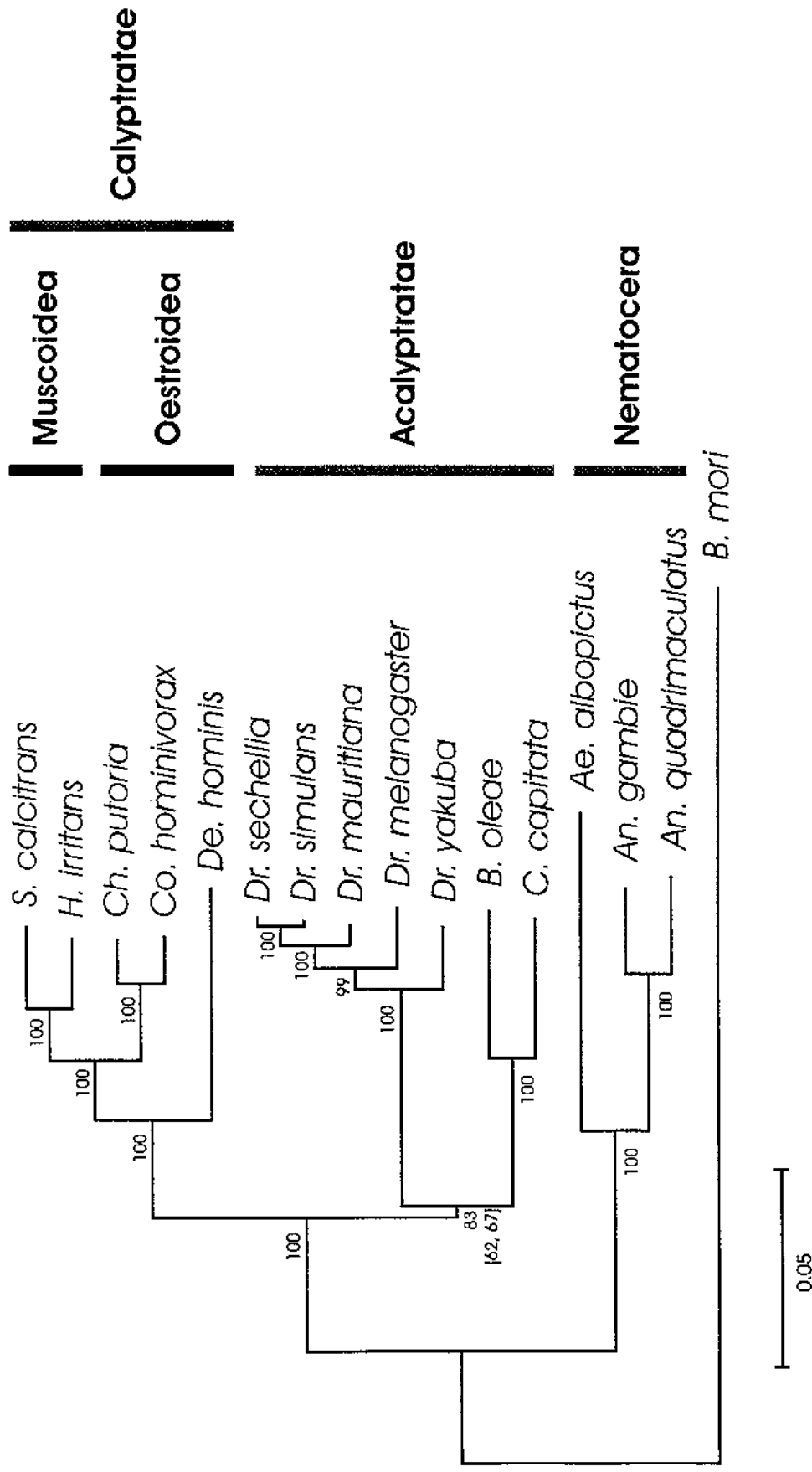


Figure 3. Phylogenetic tree based on the analysis of a concatenated alignment of amino acid sequence from a selected set of CDS (conserved/variable ratio > 1) of dipteran mtDNAs. This topology resulted from NJ analysis, using p-distance. Bootstrap values are given for each node. For the Acalypttratae clade, the bootstrap values showed in brackets resulted from the analysis of nucleotide sequences (using p-distance and Kimura-2-parameters, respectively). The same results were obtained considering genes with conserved/variable ratio > 0.5.

The analysis of *ND5* gene sequences was sufficient to reconstruct the same topology shown in Figure 3. Since this gene is the longest *CDS* of the mitochondrial genome, it was expected that it may be contributing with a great number of sites for inferring that topology. However, the analysis of most other genes also supports clustering Calliphoridae and Muscidae as sister groups. The inclusion of additional mtDNA data from Hippoboscoidea, Oestroidea and Muscoidea would provide more confident results regarding Calyptratae phylogeny and evolution.

Acknowledgements

The authors thank Rosângela A. Rodrigues and Lissiene S. Neiva for technical assistance. This work, as well as A.C.L., were supported by the Programa Especial de Estímulo à Fixação de Doutores sponsored by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (PROFIX/CNPq, grant no. 540.602/01-9). A.M.L.A.E. was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, grant no. 04/09654-4), and M.T.O. by a fellowship from Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

References

- Agris, P.F., 2004. Decoding the genome: a modified view. *Nucleic Acids Res.* 32, 223–238.
- 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.* 156, 683-717.
- Ballard, J.W., 2000. Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. *J. Mol. Evol.* 51, 48-63.
- Barau, J.G., Azeredo-Espin, A.M.L., 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. *Mol. Ecol. Notes* 5, 885-887.
- 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 Mol. Biol.* 2, 103-124.
- Bernasconi, M.V., Valsangiacomo, C., Piffaretti, J.C., Ward, P.I., 2000. Phylogenetic relationships among Muscoidea (Diptera: Calyptratae) based on mitochondrial DNA sequences. *Insect Mol. Biol.* 9, 67-74.
- Black IV, W.C., Roehrdanz, R.L. 1998. Mitochondrial gene order is not conserved in arthropods: prostriate and metastriate tick mitochondrial genomes. *Mol. Biol. Evol.* 15, 1772-1785.
- Boore, J.L., 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27, 1767-1780.
- Boore, J.L., Macey, J.R., Medina, M., 2005. Sequencing and comparing whole mitochondrial genomes of animals. *Methods Enzymol.* 395, 311-348.
- Bulmer, M., 1987. Coevolution of codon usage and transfer RNA abundance. *Nature* 325, 728-730.

- Clary, D.O., Wolstenholme, D.R., 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. *J. Mol. Evol.* 22, 252-271.
- Covacin, C., Shao, R., Cameron, S., Barker, S.C., 2006. Extraordinary number of gene rearrangements in the mitochondrial genomes of lice (Phthiraptera: Insecta). *Insect Mol. Biol.* 15, 63-68.
- Crozier, R.H., Crozier, Y.C., 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* 133, 97-117.
- Feijão, P.C, Neiva, L.S, Azeredo-Espin, A.M.L., Lessinger, A.C., 2006. AMiGA: the arthropodan mitochondrial genomes accessible database. *Bioinformatics* 22, 902-903.
- Flook, P.K., Rowell, C.H., Gellissen, G., 1995. The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. *J. Mol. Evol.* 41, 928-941.
- Goddard, J.M., Wolstenholme, D.R., 1980. Origin and direction of replication in mitochondrial DNA molecules from genus *Drosophila*. *Nucleic Acids Res.* 8, 741-757.
- Gordon, D., Abajian, C., Green, P., 1998. CONSED: a graphical tool for sequence finishing. *Genome Res.* 8, 195-202.
- Hassanin, A., 2006. Phylogeny of Arthropoda inferred from mitochondrial sequences: strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. *Mol. Phylogenet. Evol.* 38, 100-116.
- Haufe, W.O., 1987. Host-parasite interaction of blood feeding dipterans in health and productivity of mammals. *Int. J. Parasitol.* 17, 607-614.
- Hickerson, M.J., Cunningham, C.W. 2000. Dramatic mitochondrial gene rearrangements in the hermit crab *Pagurus longicarpus* (Crustacea, anomura). *Mol. Biol. Evol.* 17, 639-644.

- Hwang, U.W., Friedrich, M., Tautz, D., Park, C.J., Kim, W., 2001a. Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature* 413, 154–157.
- Hwang, U.W., Park, C.J., Yong, T.S., Kim, W., 2001b. One-step PCR amplification of complete arthropod mitochondrial genomes. *Mol. Phylogenet. Evol.* 19, 345-352.
- Infante-Vargas, M.E., Azeredo-Espin, A.M.L., 1995. Genetic variability in mitochondrial DNA of screwworm, *Cochliomyia hominivorax* (Diptera: Calliphoridae), from Brazil. *Biochem. Gen.* 33, 237-256.
- Junqueira, A.C.M., Lessinger, A.C., Torres, T.T., da Silva, F.R., Vettore, A.L., Arruda, P., Azeredo-Espin, A.M.L., 2004. The complete mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). *Gene* 339, 7-15.
- Krzywinski, J., Grushko, O.G., Besansky, N.J., 2006. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Mol. Phylogenet. Evol.* 39, 417-423.
- Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinfo.* 5, 150-163.
- 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., 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., Farquhar, A.L., Kaguni, L.S., 1994. Sequence, organization, and evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA. *Mol. Biol. Evol.* 11, 523-538.
- 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.
- 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.
- McAlpine, J.F., Wood, D.M., 1989. Manual of Nearctic Diptera Volume 3. pp.1492-1502. Research Branch Agriculture Canada, Ottawa, Ontario.
- Miller, A.D., Nguyen, T.T., BurrIDGE, C.P., Austin, C.M., 2004. Complete mitochondrial DNA sequence of the Australian freshwater crayfish, *Cherax destructor* (Crustacea: Decapoda: Parastacidae): a novel gene order revealed. *Gene* 331, 65-72.
- 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.
- Morlais, I., Severson D.W., 2002. Complete mitochondrial DNA sequence and amino acid analysis of the cytochrome c oxidase subunit I (COI) from *Aedes aegypti*. *DNA Seq* 13: 123-127.

- Nardi, F., Carapelli, A., Fanciulli, P.P., Dallai, R., Frati, F., 2001. The complete mitochondrial DNA sequence of the basal hexapod *Tetrodontophora bielanensis*: evidence for heteroplasmy and tRNA translocations. *Mol. Biol. Evol.* 18, 1293-1304.
- Nardi, F., Carapelli, A., 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.L., Carapelli, A., Dallai, R., Frati, F., 2003b. Hexapod origins: monophyletic or paraphyletic?. *Science* 299, 1887-1889.
- Nirmala, X., Hypsa, V., 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 Mol. Biol.* 10, 475-485.
- Ojala, D., Montaya, J., Attardi, G., 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290, 470-474.
- Oliveira, M.T., Azeredo-Espin, A.M.L., Lessinger, A.C. The mitochondrial DNA control region of Muscidae flies: evolution and structural conservation in a dipteran context. *J. Mol. Evol.*, submitted.
- Oliveira, M.T., Azeredo-Espin, A.M.L., Lessinger, A.C., 2005. Evolutionary and structural analysis of the cytochrome c oxidase subunit I (COI) gene from *Haematobia irritans*, *Stomoxys calcitrans* and *Musca domestica* mitochondrial DNA. *DNA Seq.* 16, 156-160.
- Oliveira, M.T., da Rosa, A.C., Azeredo-Espin, A.M.L., Lessinger, A.C., 2006. Improving access to the control region and tRNA gene clusters of dipteran mitochondrial DNA. *J. Med. Entomol.* 43, 636-639.
- Perna, N.T., Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerated sites of animal mitochondrial genomes. *J. Mol. Evol.* 41, 353-359.

- Reyes, A., Gissi, C., Pesole, G., Saccone, C., 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Mol. Biol. Evol.* 15, 957–966.
- Roehrdanz, R.L., Degrugillier, M.E., Black IV, W.C., 2002. Novel rearrangements of arthropod mitochondrial DNA detected with long-PCR: applications to arthropod phylogeny and evolution. *Mol. Biol. Evol.* 19, 841-849.
- 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.
- Saito, S., Tamura, K., Aotsuka, T., 2005. Replication origin of mitochondrial DNA in insects. *Genetics* 171, 1695-1705.
- Shadel, G.S., Clayton, D.A., 1997. Mitochondrial DNA maintenance in vertebrates. *Annu. Rev. Biochem.* 66, 409–435.
- Shao, R., 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. *Mol. Biol. Evol.* 20, 362-370.
- Shao, R., Barker, S.C., Mitani, H., Aoki, Y., Fukunaga, M., 2005a. Evolution of duplicate control regions in the mitochondrial genomes of Metazoa: a case study with Australasian *Ixodes* ticks. *Mol. Biol. Evol.* 22, 620-629.
- Shao, R., Mitani, H., Barker, S.C., Takahashi, M., Fukunaga, M., 2005b. Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the chigger mite, *Leptotrombidium pallidum*. *J Mol Evol* 60, 764-773.
- Shao, R., Campbell, N.J., Barker, S.C., 2001. Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthitaptera). *Mol. Biol. Evol.* 18, 858-865.

- Silvestre, D., 2002. Sequenciamento e análise do genoma mitocondrial de *Melipona bicolor* (Hymenoptera, Apidae, Meliponini). Master thesis, Universidade de São Paulo, São Paulo, Brazil.
- 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 polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* 87, 651-701.
- Spanos, L., Koutroumbas, G., Kotsyfakis, M., Louis, C., 2000. The mitochondrial genome of the Mediterranean fruitfly, *Ceratitis capitata*. *Insect Mol. Biol.* 9, 139-144.
- Stanton, D.J., Daehler, L.L., Moritz, C.C., Brown, W.M., 1994. Sequences with the potential to form stem-and-loop structures are associated with coding-region duplications in animal mitochondrial DNA. *Genetics* 137, 233-241.
- Stewart, J.B., Beckenbach, A.T., 2005. Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae). *Genome* 48, 46-54.
- Sun, H., Zhou, K., Song, D. 2005. Mitochondrial genome of the Chinese mitten crab *Eriocheir japonica sinensis* (Brachyura: Thoracotremata: Grapsoidea) reveals a novel gene order and two target regions of gene rearrangements. *Gene* 349, 207-217.
- Thao, M.L., Bauman, L., Bauman, P. 2004. Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). *BMC Evol. Biol.* 4, 25.
- Tong, K.L., Wong, J.T., 2004. Anticodon and wobble evolution. *Gene* 333, 169-177.
- Wyman, S.K., Jansen, R.K., Boore, J.L., 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252-3255.
- Xia, X., 2005. Mutation and selection on the anticodon of tRNA genes in vertebrate mitochondrial genomes. *Gene* 345, 13-20.

- Yamauchi, M.M., Miya, M.U., Nichida, M., 2004. Use of a PCR-based approach for sequencing whole mitochondrial genomes of insects: two examples (cockroach and dragonfly) based on the method developed for decapod crustaceans. *Insect Mol. Biol.* 13, 435-442.
- 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.

Metodologia detalhada

Amostras

As amostras de *H. irritans* e *S. calcitrans* foram coletadas em Caraguatatuba, SP e São Sebastião do Paraíso, MG, respectivamente (Capítulos 2 e 3).

Extração do DNA

O DNA total dos indivíduos foi extraído conforme descrito na Metodologia detalhada do Capítulo 1.

Obtenção do Genoma Mitocondrial

Para obter quantidades suficientes de DNAm para construção da biblioteca de “shotgun”, os genomas mitocondriais de *H. irritans* e *S. calcitrans* foram amplificados via Long-PCR gerando dois produtos de amplificação de 8-10Kb cada, assim como descrito em Barau *et al.* (2005) e na Metodologia detalhada do Capítulo 3.

Para aumentar a qualidade do seqüenciamento de alguns trechos destes genomas mitocondriais e completar trechos não seqüenciados, foram feitos PCRs “padrões” sob as condições descritas na Metodologia detalhada do Capítulo 2 para a amplificação destes trechos. Como a região-controle de *S. calcitrans* não foi representada na biblioteca de “shotgun” construída a partir destes produtos de Long-PCR, foi construída uma nova biblioteca a partir do produto obtido via Long-PCR usando os “primers” C1-N-2191 e SR-J-14233 (Simon *et al.* 1994). Foi presumido que esta região (~5,5Kb) conteria os genes COI/Y/C/W/ND2/M/Q/I/região-controle/ND2. Além disso, “primers” elaborados com base em seqüências internas da região-controle de *S. calcitrans* foram eficientes para a obtenção da maior parte da seqüência desta região. Estes novos “primers” estão descritos na seção “Materials and Methods” do manuscrito. As estratégias de amplificação realizadas neste capítulo, bem como todos os “primers” utilizados, estão mostrados na “Figure 1” do manuscrito.

Construção das Bibliotecas de “Shotgun” e Seqüenciamento

Os produtos de Long-PCR do genoma mitocondrial de *H. irritans* e *S. calcitrans* (Amplicons A e B, “Figure 1” do manuscrito) foram utilizados para a construção da primeira biblioteca de DNAm para cada espécie. Eles foram submetidos a quebras aleatórias por

sonicação com ondas ultra-sônicas utilizando o sonicador Cole Parmer, série 4710, por 15-20 segundos. Os fragmentos sonicados foram separados em corrida eletroforética a 47V em gel de agarose de baixo ponto de fusão 0,8%, durante aproximadamente 4,5h a 4°C. Os fragmentos entre 1 e 2Kb foram recortados do gel e purificados segundo o protocolo de fenol/dietil-éter. Amostras recortadas do gel contendo os fragmentos selecionados foram aquecidas a 65°C por cerca de 10min, com agitação periódica. 0,1 volumes de NaCl 4M foram adicionados à amostra para dissociar o DNA do brometo de etídeo. Foi adicionado 1 volume de fenol pH 8.0 à amostra, seguida de uma centrifugação de 10 min a 13000 rpm (20°C) e recuperação do sobrenadante. Repetiu-se este passo 2 vezes. Foram adicionados 2 volumes de dietil-éter/H₂O, seguido de uma centrifugação por 1 min a 13000 rpm para recuperação da fase inferior. Repetiu-se este passo 2 vezes. Para precipitação do DNA, foram adicionados 0,1 volumes de Acetato de Sódio 3M pH 5.2 e 2 volumes de Etanol absoluto gelado. A amostra foi mantida a -20°C por ~15 horas. A amostra foi então centrifugada por 30 min a 13000 rpm e 4°C. Aos “pellets”, foi adicionado Etanol 70% seguido de centrifugação por 5min. Os “pellets” foram então dissolvidos em água milliQ. A solução foi purificada com o kit da QiaQuick (Qiagen) para eliminar vestígios de fenol ou éter que possam ter ficado na amostra e concentrar o DNA em 20µL finais. Uma alíquota foi analisada em gel de agarose 0,8% para verificação da recuperação dos fragmentos.

A clonagem dos fragmentos de 1 a 2Kb foi realizada com o kit pMOSBlue blunt ended (Amersham Biosciences), como segue: a recuperação das extremidades dos fragmentos utilizou ~60ng do DNA purificado a 1µL de tampão 10x, 0,5µL de 100mM DTT, 1µL do mix de enzimas e 4,5µL de água, seguido de incubação a 22°C por 40min. Após a desnaturação das enzimas por temperatura (10min a 75°C), procedeu-se a ligação dos fragmentos ao plasmídeo de clonagem, adicionados 1µL de plasmídeo e 1µL de DNA ligase à reação. A reação foi mantida a 22°C por 15 horas. Após este período de incubação, a transformação foi feita através de choque químico utilizando tampão KCM/PEG em células MOSBlue (Amersham Biosciences); a seleção das colônias recombinantes foi verificada por digestão dos plasmídeos após isolamento através de lise alcalina.

A segunda biblioteca do DNAmT de *S. calcitrans* foi construída a partir do Amplicon C (“Figure 1” do manuscrito), seguindo o mesmo protocolo usado nas primeiras bibliotecas. As reações de seqüenciamento foram realizadas conforme descrito na Metodologia detalhada do Capítulo 2, exceto pelo uso dos “primers” de seqüenciamento T7 e U-19 e do seqüenciador

automático modelo 3700 (Applied Biosystems). A clonagem e o seqüenciamento dos fragmentos amplificados via PCRs “padrões” foram realizados segundo a Metodologia detalhada do Capítulo 2.

Análises Estruturais e Comparativas

As seqüências obtidas a partir da análise das bibliotecas de DNAm de *H. irritans* e *S. calcitrans* foram montadas utilizando o pacote de aplicativos phred/phrap/consed (Gordon *et al.* 1998). Os genes codificadores de proteínas, RNAr e RNAt foram anotados com o auxílio dos aplicativos DOGMA (Wyman *et al.* 2004) e Sequin (<ftp://ftp.ncbi.nih.gov/sequin>). Os codons de iniciação e de terminação de todos os CDS foram verificados manualmente para a correta anotação dos genes. As seqüências e as estruturas secundárias dos genes de RNAt foram confirmadas utilizando o programa tRNA-Scan (Lowe & Eddy 1997), exceto o tRNA para arginina, recuperado apenas no DOGMA. As seqüências referentes à região-controle foram analisadas como descrito na Metodologia detalhada do Capítulo 3.

Dados comparativos entre genomas mitocondriais de dípteros, tais como conteúdo de A+T, AT e GC “skew”, tabela com uso dos codons, dados de estrutura do DNAm e similaridade entre seqüências, foram obtidos utilizando as ferramentas disponíveis no banco de dados AMiGA (Feijão *et al.* 2006).

Análises Filogenéticas

As seqüências de aminoácidos e de nucleotídeos dos CDS mitocondriais de todos os dípteros representados nos bancos de dados públicos até o momento (*C. hominivorax* - NC_002660; *C. chloropyga* - NC_002696; *D. hominis* - NC_006378; *Ceratitis capitata* - NC_000857; *Bactrocera oleae* - NC_005333; *D. yakuba* - NC_001322; *D. melanogaster* - NC_001709; *Drosophila simulans* - NC_005781; *Drosophila mauritiana* - NC_005779; *Drosophila sechellia* - NC_005780; *Anopheles gambiae* - NC_002084; *Anopheles quadrimaculatus* - NC_000875; *Aedes albopictus* - NC_006817), incluindo as seqüências de *H. irritans* e de *S. calcitrans*, foram alinhadas individualmente usando o AMiGA e analisadas quanto à consistência destes alinhamentos. Dois critérios foram usados para compor a seqüência de genes concatenados, a qual foi utilizada nas análises filogenéticas posteriores. A razão número de sítios conservados / número de sítios variáveis (C/V) foi calculada para a seleção de genes

com C/V maior que 0,5 (Nardi *et al.* 2003), excluindo os genes ATP8, ND2 e ND6. Um segundo conjunto de CDS selecionado considerando C/V maior que 1 (Junqueira *et al.* 2004), excluindo também os genes ND3, ND4, ND4L e ND5. Nestes genes, taxas de substituição significativamente diferentes poderiam causar distorções nas inferências filogenéticas. As seqüências escolhidas foram concatenadas aplicativo do AMiGA.

As inferências filogenéticas em Diptera foram realizadas com o auxílio do programa MEGA3 (Kumar *et al.* 2004). Dois métodos de reconstrução filogenética foram utilizados tanto para as seqüências de nucleotídeos quanto para a de aminoácidos, a máxima parcimônia (MP) e o “neighbor-joining” (NJ). No caso da inferência baseada nas seqüências de aminoácidos usando NJ, o método de distância utilizado foi a distância p. Para as inferências baseadas nas seqüências de nucleotídeos também usando NJ, dois métodos de distância foram usados, a distância p e Kimura-2-parâmetros. Em todas as análises, o valor de “bootstrap” foi calculado baseado em 1000 replicações.

DISCUSSÃO

O gene COI em dípteros

O trabalho descrito no Capítulo 1 visou a caracterização estrutural e a identificação do padrão de variação ao longo da sequência completa do gene COI em três espécies da família Muscidae. O gene COI é um dos genes mais utilizados em estudos evolutivos em insetos (Caterino *et al.* 2000), além de ser o marcador escolhido para o desenvolvimento do projeto “Barcode of Life” (www.barcodinglife.org). Nós entendemos que a informação gerada pelo sequenciamento do gene COI de *H. irritans*, *S. calcitrans* e *M. domestica* pode contribuir com futuros trabalhos de filogenia e identificação taxonômica de espécies da família Muscidae, já que nosso artigo foi o primeiro a apresentar a caracterização deste gene em espécies desta família. Antes do sequenciamento do gene COI destas três espécies, poucas sequências parciais de genes mitocondriais (COI, COII e 12S) estavam disponíveis no GenBank para Muscidae, apesar da importância econômica de algumas espécies no cenário agropecuário.

A caracterização do gene COI foi também importante para acessar algumas informações relevantes para a caracterização do genoma mitocondrial completo de *H. irritans* e *S. calcitrans* (Capítulo 4). Por exemplo, a caracterização do gene COI já tinha evidenciado que esta duas espécies possuíam sequências bem ricas em A+T, o que se mostrou verdadeiro para o genoma mitocondrial como um todo (*H. irritans* e *S. calcitrans* estão entre os dípteros cujos DNAm_t são os mais ricos em conteúdo de A+T).

O gene COI de muscídeos mostrou ser bem similar, estrutural e evolutivamente, ao de outros dípteros e insetos em geral, mas talvez o aspecto mais interessante destas sequências seja a identificação de um códon de iniciação da tradução não reconhecido para o DNAm_t animal, o trinucleotídeo TCG (aminoácido: serina). Os códons de iniciação ATG, ATT, ATA e ATC são característicos do DNAm_t de metazoários (Wolstenholme 1992), mas exceções a esta regra parecem ser comuns ao gene COI de artrópodes. Nos primeiros genomas mitocondriais de artrópodes descritos (*D. yakuba* e *D. melanogaster*), nenhum ATN seguindo o quadro de leitura foi encontrado próximo ao presumido sítio de iniciação deste gene. Isto fez com que um tetranucleotídeo ATAA logo à jusante do TCG fosse sugerido como o iniciador de COI (de Bruijn 1983, Clary & Wolstenholme 1985), o que nunca foi testado experimentalmente. Alternativamente, para o caranguejo *Pagurus longicarpus*, o tetranucleotídeo ATCA foi sugerido

para a iniciação de COI (Hickerson & Cunningham 2000) e, para o colêmbola *Tetrodontophora bielanensis*, o hexanucleotídeo ATTTAA (Nardi *et al.* 2001). Em dípteros, assim que novos genomas e seqüências iam sendo caracterizados, mais se especulava sobre o papel do códon TCG na iniciação da tradução do gene COI. Recentemente, dois artigos caracterizando o gene COI nos mosquitos *Aedes aegypti* (Morlais & Severson 2002) e *Anopheles funestus* (Krzywinski *et al.* 2006) mostraram, através do seqüenciamento de cDNA, que a extremidade 5' do transcrito de COI começava com o trinucleotídeo TCG, dando suporte à hipótese de iniciação com o TCG em dípteros.

Acesso a regiões de instabilidade estrutural do DNAm

Apesar de ser uma das principais referências para quem trabalha com DNAm de insetos (ver Introdução - O DNA Mitocondrial Animal), a compilação de “primers” descrita no trabalho de Simon *et al.* (1994) foi baseada nas poucas seqüências disponíveis naquele momento, o que fez com que, para algumas regiões do DNAm de insetos, poucos “primers” *universais* fossem disponibilizados. Dentre estas regiões estão os locais em que é registrada maior ocorrência de eventos de translocação e/ou duplicação de trechos de seqüências mitocondriais de artrópodes, segundo Boore (1999). Estas regiões são comumente chamadas de regiões de “hot spot” estrutural do DNAm e incluem a região-controle, o agrupamento de RNAt que a flanqueia, o agrupamento de RNAt entre COI e ND2 e o agrupamento entre ND3 e ND5. A Figura 4 mostra um esquema indicando estas regiões em diferentes artrópodes.

O desenvolvimento de novos “primers” conservados e seu relativo sucesso para acessar estas regiões do DNAm de dípteros via PCR é uma importante contribuição para estudos que envolvam a caracterização da organização dos genes de RNAt e da região-controle no genoma mitocondrial e também a análise de variabilidade intra e interespecífica em espécies de insetos. Exemplos da utilidade destes “primers” já podem ser observados no Laboratório de Genética Animal, CBMEG, UNICAMP, onde eles têm ajudado o desenvolvimento de outros projetos, principalmente envolvendo a amplificação da região-controle de dípteros de importância econômica e forense.

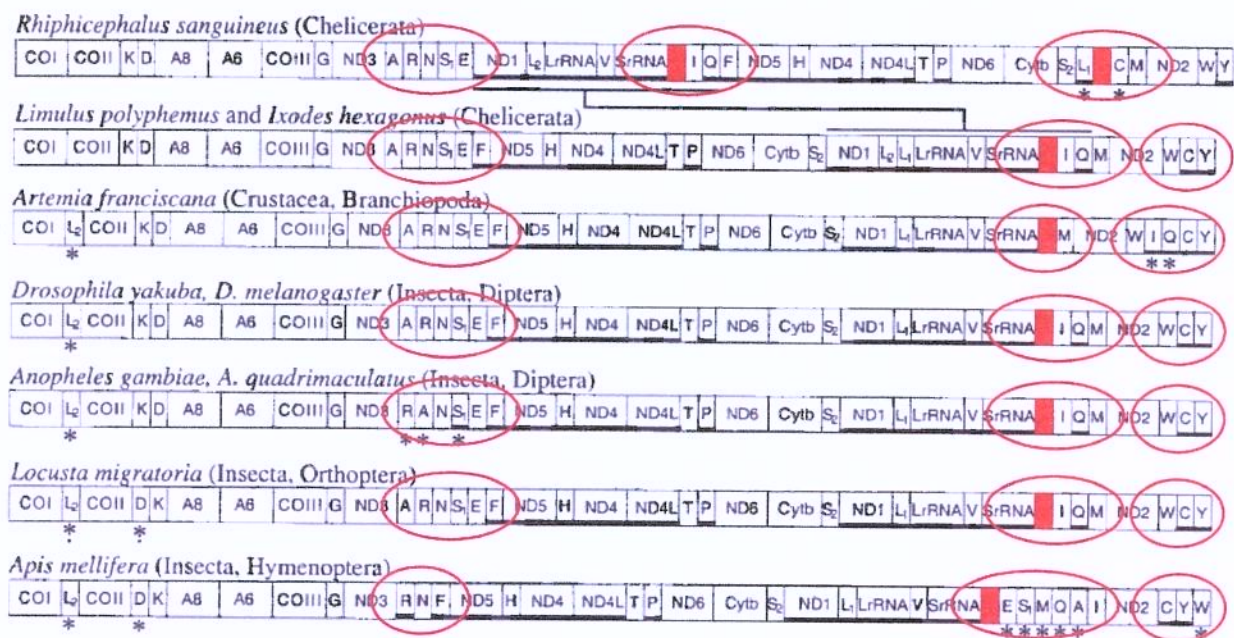


Figura 5. Representação do genoma mitocondrial linearizado de espécies de artrópodes. A região do DNAm_t preenchida em vermelho é considerada a região-controle e os círculos vermelhos indicam as regiões de maior instabilidade estrutural quando estes genomas são comparados. Adaptado de Boore (1999).

Controle da replicação/transcrição?

A caracterização molecular da região-controle de três espécies da família Muscidae, como mostrado no Capítulo 3, amplia o conhecimento acerca da estrutura e evolução desta região em dípteros e sugere possíveis formas de regulação dos mecanismos de replicação e transcrição do DNAm_t de insetos, baseado nas possíveis “assinaturas moleculares” identificadas na região-controle de dípteros braquíceros, como as regiões de sequência conservada. Zhang & Hewitt (1997), comparando seqüências de região-controle de espécies de *Drosophila*, de ortópteros e de lepidópteros, tentaram determinar o que chamaram de homologias estruturais e propor que estes elementos estariam presentes na região-controle de toda classe dos insetos. Estas homologias estruturais nem sempre são reconhecidas por alinhamentos de seqüências (busca comum por similaridade), por exemplo, Zhang & Hewitt (1997) consideram homologias estruturais seqüências ricas em G+A e trechos [TA(A)]_n repetidos em “tandem”. Isto resulta no não reconhecimento de tais elementos indicados como conservados para algumas espécies de insetos (Nardi *et al.* 2001, Dotson & Beard 2001, Snäll *et al.* 2002).

É importante ressaltar que com a identificação de seqüências primárias e estruturas secundárias conservadas entre espécies de quatro famílias distintas de dípteros (Muscidae, Calliphoridae, Oestridae e Drosophilidae), uma subordem toda, a Brachycera (Figura 5), pôde ser agrupada no que se refere à regulação de processos de replicação/transcrição do DNAm, através do compartilhamento das possíveis seqüências *cis*-regulatórias da região-controle. As mesmas análises foram conduzidas no sentido de englobar também espécies de mosquitos (subordem Nematocera), mas apenas uma região conservada pôde ser identificada, o poli-T central (BCR II - “Figure 1” do manuscrito do Capítulo 3), que é o mesmo elemento claramente identificado em outros insetos holometábolos. Este elemento foi reconhecido recentemente como o ponto preciso de início da replicação da fita “minor” (codifica para 4 CDS, 8 RNAt e 2 RNAr) do DNAm de holometábolos - definição segundo Simon *et al.* (1994). O início da replicação da fita “major” (9 CDS e 14 RNAt) apenas pôde ser determinado para *Drosophila*, sendo precisamente mapeado à montante do outro elemento poli-T (BCR I - “Figure 1” do manuscrito do Capítulo 3). Insetos hemimetábolos e ametábolos, em sua maioria, possuem elementos poli-T mais curtos (menos de 10 nucleotídeos), os quais parecem não estar envolvidos no início da replicação de nenhuma fita do DNAm (Saito *et al.* 2005).

Fica evidenciado, assim, que pelo menos três mecanismos diferentes de replicação do DNAm são encontrados em insetos: 1) ambas as fitas iniciam a replicação na região-controle, com envolvimento de ambos os elementos poli-T (dípteros braquíceros); 2) a replicação da fita “minor” é iniciada na região-controle, com envolvimento do elemento poli-T central (insetos holometábolos); e 3) a replicação da fita “minor” é iniciada na região-controle, sem envolvimento dos elementos poli-T (insetos hemimetábolos). Desta forma, é fácil visualizar a dificuldade de se encontrar regiões conservadas na região-controle de insetos que sejam resultado de restrições funcionais seletivas. A caracterização da região-controle de espécies de Brachycera e a identificação de seqüências primárias homólogas reafirmam os possíveis papéis regulatórios destas seqüências para o genoma mitocondrial destas espécies.

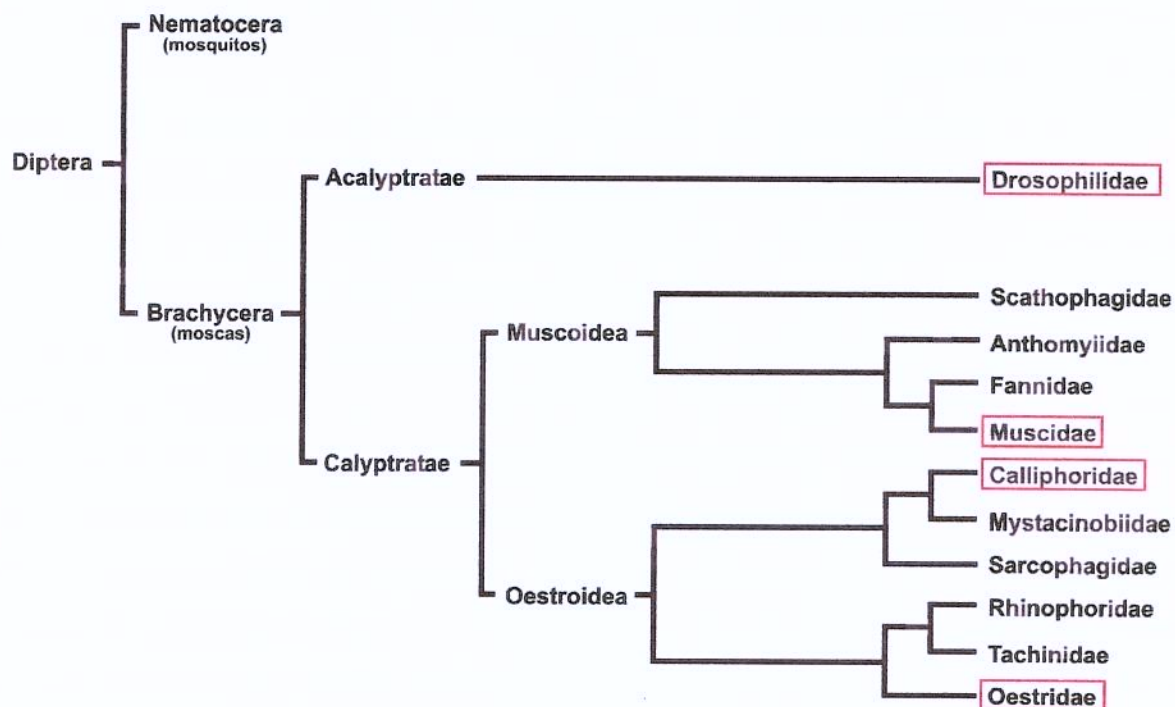


Figura 6. Filogenia simplificada da ordem Diptera, mostrando os *taxa* de interesse para o presente trabalho. Destacadas em vermelho estão as famílias cujas espécies foram analisadas no manuscrito do Capítulo 3. Adaptado de McAlpine & Wood (1989).

Organização estrutural e evolução do genoma mitocondrial em Muscidae

No Capítulo 3, já se tinha evidenciado, por tentativas de amplificar a região-controle das espécies *S. calcitrans* e *O. aenescens*, que o genoma mitocondrial destas espécies possuía uma estrutura diferente daquela dos outros muscídeos. Embora os genes que flanqueiam a região-controle tenham sido facilmente identificados, as regiões amplificadas (*ScIR* e *OaIR*) não possuíam nenhuma similaridade com a região-controle de dípteros braquíceros. Duas possibilidades foram levantadas: ou a regulação do DNAm_t nestas duas espécies teria sofrido uma alteração significativa o suficiente para mudar a estrutura da região-controle (a ponto de transformá-las naquelas seqüências de 96 e 382pb), ou estaríamos diante de um evento de rearranjo e/ou duplicação de elementos gênicos no genoma mitocondrial de *S. calcitrans* e *O. aenescens*, que limitaram, inclusive, o sucesso das metodologias usadas para acessar a região-controle. Primeiramente, foi com a análise de amplificação via Long-PCR e digestão com enzimas de restrição que verificamos que os DNAm_t destas espécies eram de mesmo tamanho ou ainda maiores que o DNAm_t de outros dípteros. Se *ScIR* e *OaIR* fossem realmente as regiões-controle de *S. calcitrans* e de *O. aenescens*, respectivamente, estes genomas deveriam ser de 700-

1100bp menores que o DNAm de *H. irritans*, o que não ocorre (o DNAm de *S. calcitrans*, por exemplo, é ~2Kb maior).

Com o sequenciamento dos genomas mitocondriais de *H. irritans* e *S. calcitrans*, pudemos comparar e identificar a presença de uma região-controle em *S. calcitrans* estruturada semelhantemente à região-controle de braquíceros, exceto pela presença de grandes trechos repetidos em “tandem”. Entretanto, a organização do genoma mitocondrial descrita no Capítulo 3 não foi confirmada pelos resultados do projeto de sequenciamento do DNAm de *S. calcitrans* (Capítulo 4), exceto pela identificação da ScIR. O agrupamento de RNAt que geralmente flanqueia a região-controle no DNAm de insetos (MQI), também o faz em *S. calcitrans*. A duplicação observada em *S. calcitrans* já foi descrita para o DNAm de outros dípteros do gênero *Chrysomya* (Calliphoridae) (Lessinger *et al.* 2004, Junqueira *et al.* 2004) e envolve seqüências parciais de RNAt para glutamina e da região-controle, e a seqüência completa do RNAt para isoleucina. À montante da região duplicada, entre os genes I2 e RNAr 12S, é onde se encontram a ScIR e a CIR (no caso das espécies de *Chrysomya*).

A convergência de estrutura e conteúdo gênico nos genomas mitocondriais de *S. calcitrans*, *C. chloropyga*, *C. megacephala* e *C. albiceps* sugere que este tipo de evento está associado a um “hot spot” de duplicação no DNAm de dípteros. Com base no atual conhecimento sobre a origem de replicação do DNAm em dípteros Brachycera (Goddard & Wolstenholme 1980, Saito *et al.* 2005, Capítulo 3), na idéia de que estruturas secundárias no DNA podem estar envolvidas com duplicações no genoma mitocondrial (Stanton *et al.* 1994) e em evidências da ocorrência de recombinação ilegítima entre moléculas do DNAm animal (Shao *et al.* 2005b), nós propomos o mecanismo explicado no Capítulo 4 e esquematizado na Figura 6. Mais evidências experimentais devem ser levadas em consideração para a consolidação deste modelo mecanístico, mas esta proposta ajuda a entender como a estrutura do genoma mitocondrial dos metazoários pode evoluir, inclusive gerando arranjos semelhantes em linhagens evolutivas distintas. A questão do compartilhamento de padrões de organização gênica como indicador de ancestralidade comum tem sido sugerida (revisão em Boore 1999 e Boore *et al.* 2005), mas também bastante discutida, já que a convergência da ordem dos genes mitocondriais em organismos de diferentes linhagens evolutivas não parece ser um evento tão raro (Dowton & Austin 1999, Dowton *et al.* 2003).

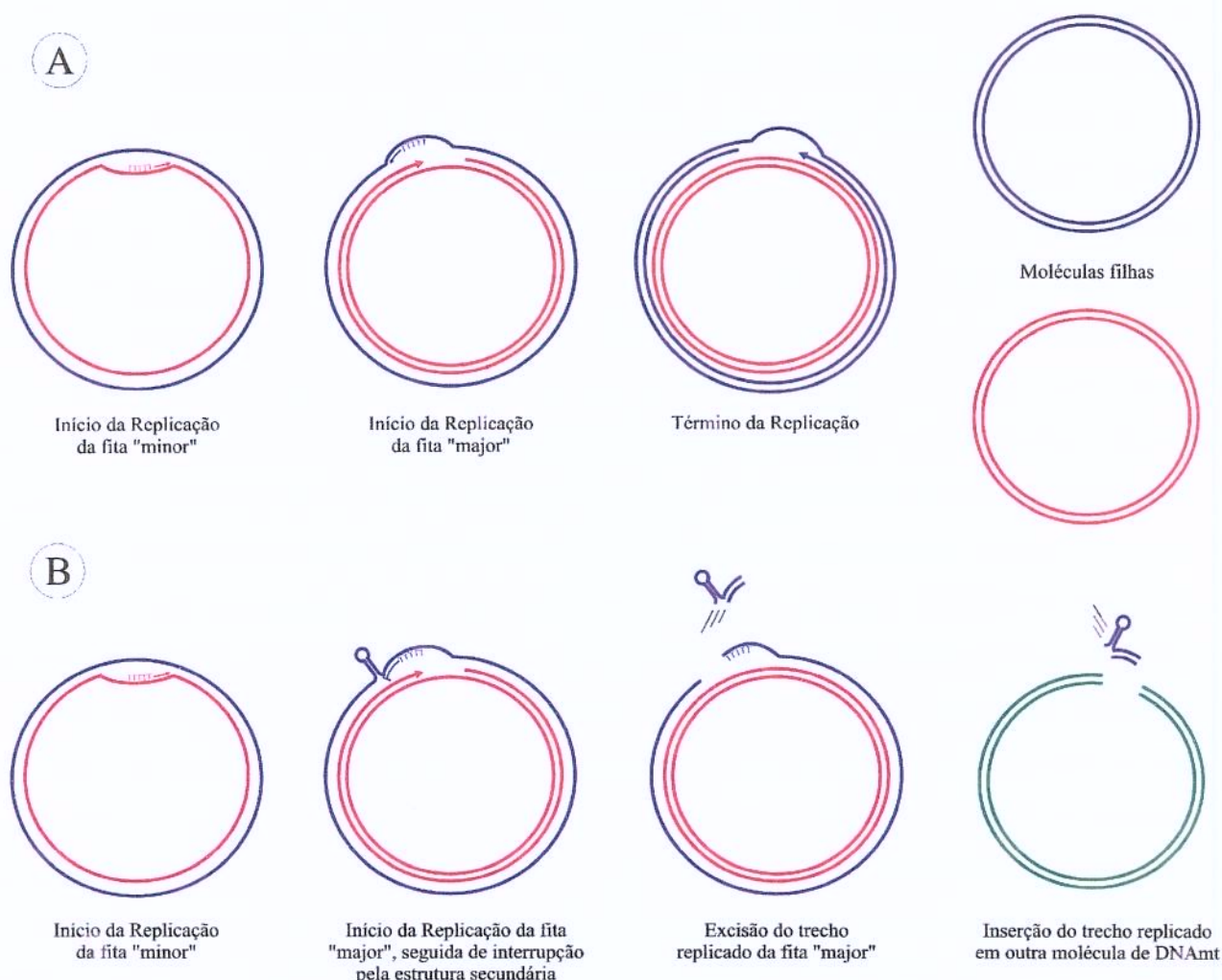


Figura 7. Esquema do modelo mecânico de ocorrência da duplicação de rRNAt para glutamina e isoleucina e parte da região-controle verificado nos genomas mitocondriais de *S. calcitrans*, *C. chloropyga*, *C. megacephala* e *C. albiceps*. A) Representação de uma replicação “bem sucedida” do DNAm, segundo Goddard & Wolstenholme (1980) e Saito *et al.* (2005). TTTT em vermelho e azul representam os “brachyceran conserved regions” I e II (início da replicação das fitas “minor” and “major”, respectivamente). B) Possíveis etapas de formação da região duplicada durante o processo de replicação.

A evolução dos anticódons

O fato dos genomas mitocondriais de espécies de insetos serem caracterizados pelo uso de códons não-ótimos foi sempre considerado um aspecto intrigante deste sistema genômico. Desde a primeira evidência de que os códons mais utilizados dentre uma família de códons degenerados

de um genoma eram aqueles que se complementavam precisamente com o anticódon do RNAt correspondente (Ikemura 1981), surgiu a hipótese da adaptação códon-anticódon ou uso de codons ótimos (“optimal codon-usage”). Neste caso, a evolução para a preferência do uso dos códons levaria a uma adaptação destes ao anticódon disponível (ou o mais abundante) e que, neste último, a evolução da posição variável seria governada pelas pressões de mutação que ocorrem no genoma (“*hipótese de mutação para a evolução do anticódon*”). Recentemente, duas novas hipóteses surgiram: primeiro, a “*hipótese de seleção para a versatilidade do anticódon*” (Agris 2004, Tong & Wong 2004), que diz que RNAt com anticódons possuindo T ou G na posição variável seriam mais versáteis por serem eficientes na tradução de todos os códons de uma família de códons degenerados e seriam selecionados, inclusive em situações onde evoluir para um “optimal codon-usage” seria desfavorável (quando há seleção por genomas ricos em A+T, por exemplo); e segundo, a “*hipótese de seleção para a adaptação do anticódon*” (Xia 2005), proposta exclusivamente para o DNAm de vertebrados, que diz que a posição variável do anticódon evolui para se adaptar ao uso de códons (e não o contrário). Segundo Xia (2005), a predominância de A e C na posição degenerada dos codons (que ocorre devido à pressão de mutação no DNAm de vertebrados) selecionaria T ou G como base variável do anticódon, promovendo a otimização da tradução. De fato, todos os anticódons dos RNAt mitocondriais de vertebrados, exceto o RNAt para metionina (ver justificativa em Xia, 2005), possuem T ou G na posição variável, caracterizando a preferência pelo uso de códons ótimos. Estas duas últimas hipóteses de seleção explicam igualmente bem os dados encontrados para o genoma mitocondrial de vertebrados, entretanto apenas a *hipótese de seleção para a versatilidade do anticódon* poderia, também, explicar a manutenção de um uso preferencial por codons não-ótimos no genoma mitocondrial de insetos.

Comparando os RNAt do DNAm de insetos e vertebrados, incluindo os muscúdeos, verificamos que os anticódons são conservados (exceto pela variação no RNAt para lisina). Esta conservação sugere que os anticodons de ambos os grupos respondem a uma mesma força de seleção, permanecendo iguais desde a divergência entre vertebrados e invertebrados. Contrariamente, a preferência pelo uso de codons diverge entre vertebrados e invertebrados (codon ótimo vs não-ótimo, respectivamente) devido às diferentes pressões de mutação que atuam nestes genomas, A+C/assimétrica e A+T/generalizada. Concluímos, com base na conservação dos anticodons dos tRNAs e na abrangência da *hipótese de seleção para a*

versatilidade do anticódon, que o uso preferencial de códons não orienta diretamente a evolução dos anticodons no DNAm animal, divergindo da proposta de Xia (2005) sobre evolução de anticodons do DNAm de vertebrados.

CONSIDERAÇÕES FINAIS E PERSPECTIVAS

Nós entendemos que o conjunto dos trabalhos apresentados nesta dissertação colaborou imensamente para a compreensão de alguns aspectos evolutivos relacionados à estrutura do genoma mitocondrial em espécies de dípteros da família Muscidae, e nos metazoários em geral. Além disso, os dados gerados neste trabalho podem servir de base para futuros estudos que visem ampliar a caracterização de genomas mitocondriais de insetos (seqüências parciais ou genoma completo), utilizar o DNAm_t como marcador molecular em estudos de variabilidade intra e interespecífica considerando a distribuição geográfica de moscas-praga como *H. irritans*, *S. calcitrans*, *M. domestica*, *A. orientalis* e *O. aenescens*, e aprofundar o entendimento das relações evolutivas entre as famílias de Diptera. Durante o desenvolvimento deste estudo, nós depositamos no GenBank quase 58Kb de seqüências mitocondriais de espécies da família Muscidae, contribuindo para o grande crescimento no número de seqüências nucleotídicas depositadas em banco de dados públicos relacionadas a estas moscas-praga.

Como perspectivas, três principais focos são considerados importantes. O primeiro relaciona-se ao estudo bioquímico das interações entre as proteínas reguladoras importadas do núcleo e as seqüências conservadas descritas para a região-controle dos braquíceros. Como fatores de transcrição ou proteínas associadas à origem da replicação interagem entre si e com o DNAm_t de insetos para promover o início destes processos que são tão fundamentais para a expressão e a transmissão da informação genética mitocondrial? O segundo foco está relacionado à elaboração de diferentes estratégias para obtenção da seqüência completa da região-controle do DNAm_t de *S. calcitrans*, já que os métodos utilizados neste trabalho não foram suficientes para se obter toda a informação que gostaríamos referente a esta peculiar região. Com a integralização da seqüência completa do DNAm_t de *S. calcitrans*, poderíamos verificar como as organizações estruturais apresentadas nos Capítulos 3 e 4 estão relacionadas ou se há chance de artefatos de PCR estarem envolvidos. O terceiro e último foco associa-se com a abrangência filogenética da duplicação encontrada no genoma mitocondrial de *S. calcitrans*. Outras espécies do gênero *Stomoxys* também possuem tal duplicação ou este é um evento mais recente? E a organização de *O. aenescens* é fruto de um evento ancestral à separação das linhagens *Stomoxys* e *Ophyra*, ou é um exemplo de mais uma convergência estrutural? Ficam aqui estas questões...

REFERÊNCIAS BIBLIOGRÁFICAS

- Agris, P.F. (2004). Decoding the genome: a modified view. *Nucleic Acids Res* 32: 223-238.
- 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 Acid Res* 25: 3389-3402.
- Arther, R.G. (1991). Management of horn fly resistance. 40th Annual Florida Beef Cattle Short Course – Proceedings, Gainesville, Florida.
- Awise, J.C. (2004). Molecular markers, natural history, and evolution. 2^a edição. Sinauer Associates, Inc., Publishers. Plumtree Road, EUA.
- Azeredo-Espin, A.M.L. & Lessinger, A.C. (2006). Genetic approaches for studying myiasis-causing flies: molecular markers and mitochondrial genomics. *Genetica* 126: 111-131.
- Azeredo-Espin, A.M.L. & Madeira, N.G. (1996). Primary myiasis in dog caused by *Phaenicia eximia* (Diptera: Calliphoridae) and preliminary mitochondrial DNA analysis of the species in Brazil. *J Med Entomol* 33: 839-843.
- Azeredo-Espin, A.M.L., Schroder, R.F.W., Huettle, M.D. & Sheppard, W.S. (1991). Mitochondrial DNA variation in geographic population of Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *Experientia* 47: 483-485.
- Bailey, T.L. & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymer. Proceedings of the Second International Congress on Intelligent System for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California.
- Ballard, J.W. (2000a). Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. *J Mol Evol* 51: 48-63.
- Ballard, J.W. (2000b). Comparative genomics of mitochondrial DNA in *Drosophila simulans*. *J Mol Evol* 51: 64-75.
- Barau, J.G., Azeredo-Espin, A.M.L. & 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. *Mol Ecol Notes* 5: 885-887.
- 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 Mol Biol* 2: 103-124.

- Bernasconi, M.V., Valsangiacomo, C., Piffaretti, J.C. & Ward, P.I. (2000). Phylogenetic relationships among Muscoidea (Diptera: Calyptratae) based on mitochondrial DNA sequences. *Insect Mol Biol* 9: 67-74.
- Boore, J.L. (1999). Animal mitochondrial genomes. *Nucleic Acids Res* 27: 1767-1780.
- Boore, J.L., Macey, J.R. & Medina, M. (2005). Sequencing and comparing whole mitochondrial genomes of animals. *Methods Enzymol* 395: 311-348.
- Boore, J.L., Medina, M. & Rosenberg, L.A. (2004). Complete sequences of two highly rearranged molluscan mitochondrial genomes, those of the scaphopod *Graptacme eborea* and of the bivalve *Mytilus edulis*. *Mol Biol Evol* 21: 1492-1503.
- Boyce, T.M., Zwick, M.E. & Aquadro, C.F. (1989). Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. *Genetics* 123: 825-836.
- Brehm, A., Harris, D.J., Hernández, M., Cabrera, V.M., Larruga, J.M., Pinto, F.M. & González, A.M. (2001). Structure and evolution of the mitochondrial DNA complete control region in the *Drosophila subobscura* subgroup. *Insect Mol Biol* 10: 573-578.
- Brito, R.A., Manfrin, M.H. & Sene, F.M. (2002). Mitochondrial DNA phylogeography of Brazilian populations of *Drosophila buzzatii*. *Genetics Mol Biol* 25: 161-171.
- Burger, G., Gray, M.W. & Lang, B.F. (2003). Mitochondrial genomes: anything goes. *Trends Genetics* 19: 709-716.
- Campbell, J.B. (1976). Effect of horn fly control on cows as expressed by increased weaning weight of calves. *J Econ Entomol* 69: 711-712.
- Carvalho, C.J.B. & Couri, M.S. (2002). Part I, Basal Groups. *In*: Carvalho, CJB (ed.) Muscidae (Diptera) of Neotropical Region: Taxonomy. Editora Universidade Federal do Paraná, Curitiba.
- Castiglione, L. & Bicudo, H.E.M.C. (2005). Molecular characterization and relatedness of *Haematobia irritans* (horn fly) populations, by RAPD-PCR. *Genetica* 124: 11-21.
- 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.
- Chaughtal, G.H., Khan, S. & Baloch, U.K. (1985). A new record of infestation of melon fruits by an anthomyiid fly in Indus River Beach Areas of D.I. Khan, Pakistan. *Pakistan J Zool* 17: 165-168.

- Chu, K.H., Li, C.P., Tam, Y.K. & Lavery, S. (2003). Application of mitochondrial control region in population genetic studies of the shrimp *Penaeus*. *Mol Ecol Notes* 3: 120-122.
- Clary, D.O. & Wolstenholme, D.R. (1985). The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. *J Mol Evol* 22: 252-271.
- Crozier, R.H. & Crozier, Y.C. (1993). The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* 133: 97-117.
- de Bruijn, M.H. (1983). *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code. *Nature* 304: 234-241.
- Dobzhansky, T. (1970). *Genetics of the evolutionary process*. Columbia University Press, New York, N.Y. pp. 5-6.
- Dotson, E.M. & Beard, C.B. (2001). Sequence and organisation of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. *Insect Mol Biol* 10: 205-215.
- Dowton, M. & Austin, A.D. (1999). Evolutionary dynamics of a mitochondrial rearrangement “hot spots” in the Hymenoptera. *Mol Biol Evol* 16: 298-309.
- Dowton, M., Castro, L.R., Campbell, S.L., Bargon, S.D. & Austin, A.D. (2003). Frequent mitochondrial gene rearrangements at the hymenopteran nad3-nad5 junction. *J Mol Evol* 56: 517-526.
- Drummond, R.O., Lambert, G., Smalley, H.E. & Terril, C.E. (1981). Estimated losses of livestock to pests. *In*: Pimentel, D (ed.) *CRC Handbook of pest management in agriculture*. CRC, Boca Raton, FL.
- Ebach, M.C. & Holdrege, C. (2005). DNA barcoding is no substitute for taxonomy. *Nature* 434: 697-697.
- Farkas, R. & Jantyk, J. (1990). Laboratory studies on *Hydrotaea aenescens* (Wiedemann, 1830) as a predator of house fly larvae. *Parasitol Hung* 23: 103-108.
- Feijão, P.C, Neiva, L.S, Azeredo-Espin, A.M.L. & Lessinger, A.C. (2006). AMiGA: the arthropodan mitochondrial genomes accessible database. *Bioinformatics* 22: 902-903.
- Garesse, R. & Kaguni, L.S. (2005). A *Drosophila* model of mitochondrial DNA replication: proteins, genes and regulation. *IUBMB Life* 57: 555-561.

- Geurgas, S.R., Infante-Malachias, M.E. & Azeredo-Espin, A.M.L. (2000). Extreme mitochondrial DNA variability and lack of genetic structure in populations of *Dermatobia hominis* (Diptera: Cuterebridae) from Brazil. *Ann Entomol Soc Am* 93: 1085-1094.
- Goddard, J.M. & Wolstenholme, D.R. (1980). Origin and direction of replication in mitochondrial DNA molecules from genus *Drosophila*. *Nucleic Acids Res* 8: 741-757.
- Gordon, D., Abajian, C. & Green, P. (1998). CONSED: a graphical tool for sequence finishing. *Genome Res* 8: 195-202.
- Greenberg, B. (1973). *Flies and Disease. Biology and Disease Transmission*, vol. II. Univ Press Princeton, New Jersey.
- Gregory, T.R. (2005). DNA barcoding does not compete with taxonomy. *Nature* 434: 1067-1067.
- Guglielmone, A.A., Gimeno, E., Idiart, J., Fisher, W.F., Volpogni, M.M., Quaino, O., Anziani, O.S., Flores, S.G. & Warnke, O. (1999). Skin lesions and cattle hide damage from *Haematobia irritans* infestation in cattle. *Med Vet Entomol* 13: 323-328.
- Haufe, W.O. (1987). Host-parasite interaction of blood feeding dipterans in health and productivity of mammals. *Int J Parasitol* 17: 607-614.
- Hebert, P.D.N., Ratnasingham, S. & deWaard, J.R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc Lond B* 270 suppliment 03BL0066, 1-4.
- Hickerson, M.J. & Cunningham, C.W. (2000). Dramatic mitochondrial gene rearrangements in the hermit crab *Pagurus longicarpus* (Crustacea, anomura). *Mol Biol Evol* 17: 639-644.
- Hwang, U.W., Park, C.J., Yong, T.S. & Kim, W. (2001). One-step PCR amplification of complete arthropod mitochondrial genomes. *Mol Phylogenet Evol* 19: 345-352.
- IAEA – International Atomic Energy Agency (2000). Genetic sexing and population genetics of screwworms. Aug. 7-11, 2000, Vienna, Austria. 19pp.
- Ikemura, T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J Mol Biol* 151: 389-409.
- Infante-Vargas, M.E. & Azeredo-Espin, A.M.L. (1995). Genetic variability in mitochondrial DNA of screwworm, *Cochliomyia hominivorax* (Diptera: Calliphoridae), from Brazil. *Biochem Gen* 33: 237-256.

- Innis, M.A., Gelfand, D.H., Sninsky, J.J. & White, T.J. (1990). PCR protocols. Academic, San Diego.
- Jermiin, L.S., Graur, D., Lowe, R.M. & Crozier, R.H. (1994). Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome b genes. *J Mol Evol* 39: 191-197.
- Junier, T. & Pagni, M. (2000). Dotlet: diagonal plots in a Web browser. *Bioinfo Appl Note* 16: 178-179.
- Junqueira, A.C.M., Lessinger, A.C. & Azeredo-Espin, A.M.L. (2002). Methods for recover of mitochondrial DNA sequences from museum specimens of myiasis-causing flies. *Med Vet Entomol* 16: 39-45.
- Junqueira, A.C.M., Lessinger, A.C., Torres, T.T., da Silva, F.R., Vettore, A.L., Arruda, P. & Azeredo-Espin, A.M.L. (2004). The complete mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). *Gene* 339: 7-15.
- Kaguni, L.S. (2004). DNA polymerase gamma, the mitochondrial replicase. *Annu Rev Biochem* 73: 293-320.
- Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X. & Wilson, A.C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc Natl Acad Sci USA* 86: 6196-6200.
- Krzywinski, J., Grushko, O.G. & Besansky, N.J. (2006). Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Mol Phylogenet Evol*, doi:10.1016/j.ympev.2006.01.006.
- Kumar, S., Tamura, K., Jakobsen, I.B. & Nei, M. (2001). MEGA2: Molecular Evolutionary Genetics Analysis. Arizona State University, Tempe.
- Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinfo* 5: 150-163.
- Kunz, S.E., Miller, J.A., Sims, P.L. & Meyerhoeffler, D.C. (1984). Economics of controlling horn flies (Diptera: Muscidae) in range cattle management. *J Econ Entomol* 77: 657-660.
- Lancaster, J.L. & Meisch, M.V. (1986). Arthropods in livestock and poultry production. John Wiley & Sons, New York.

- Learmount, J., Chapman, P. & MacNicoll, A. (2002). Impact of an insecticide resistance strategy for house fly (Diptera: Muscidae) control in intensive animal units in the United Kingdom. *J Econ Entomol* 95: 1245-1250.
- Lefebvre, F. & Pasquerault, T. (2004). Temperature-dependent development of *Ophyra aenescens* (Wiedemann, 1830) and *Ophyra capensis* (Wiedemann, 1818) (Diptera: Muscidae). *Forensic Sci Int* 139: 75-79.
- Lessinger, A.C. (2001). O genoma mitocondrial de moscas causadoras de miíases: sequência, organização e evolução. Tese de doutorado, Instituto de Biologia, Universidade Estadual de Campinas.
- 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., 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.
- 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.
- Lewis, D.L., Farr, C.L., Farquhar, A.L. & Kaguni, L.S. (1994). Sequence, organization, and evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA. *Mol Biol Evol* 11: 523-538.
- Litjens, P., Lessinger, A.C. & Azeredo-Espin, A.M.L. (2001). Characterization of the screwworm flies *Cochliomyia hominivorax* and *Cochliomyia macellaria* by PCR-RFLP of mitochondrial DNA. *Med Vet Entomol* 15: 183-188.
- 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.
- Lunt, D.H., Zhang, D.-X., Szymura, J.M. & Hewitt, G.M. (1996). The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetics studies. *Insect Mol Biol* 5: 153-165.
- Lyra, M.L., Fresia, P., Gama, S., Cristina, J., Klaczko, L.B. & 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. *J Med Entomol*, 42: 589-595.
- Mardulyn, P., Termonia, A. & Milinkovitch, M.C. (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.
- Margulis, L. & Sagan, D. (2002). *O que é vida?* Jorge Zahar Editor Ltda, Rio de Janeiro, RJ, pp. 53.
- McAlpine, J.F. & Wood, D.M. (1989). *Manual of Nearctic Diptera Volume 3*. pp.1492-1502. Research Branch Agriculture Canada, Ottawa, Ontario.
- McMillen-Jackson, A.L. & Bert, T.M. (2003). Disparate patterns of population genetic structure and population history in two sympatric penaeid shrimp species (*Farfantepenaeus aztecus* and *Litopenaeus setiferus*) in the eastern United States. *Mol Ecol* 12: 2895-2905.
- 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.
- Miya, M. & Nishida, M. (1999). Organization of the mitochondrial genome of a deep-sea fish, *Gonostoma gracile* (Teleostei: Stomiiformes): first example of transfer RNA gene rearrangements in bony fishes. *Mar Biotechnol* 1: 416-426.
- Moritz, C. & Cicero, C. (2004). DNA barcoding: promise and pitfalls. *PloS Biol* 2: e354.
- Morlais, I. & Severson D.W. (2002). Complete mitochondrial DNA sequence and amino acid analysis of the cytochrome c oxidase subunit I (COI) from *Aedes aegypti*. *DNA Seq* 13: 123-127.
- 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. *Mol Biol Evol* 18: 1293-1304.
- Nardi, F., Spinsanti, G., Boore, J.L., Carapelli, A., Dallai, R. & Frati, F. (2003). Hexapod origins: monophyletic or paraphyletic?. *Science* 299: 1887-1889.
- Otranto, D., Tarsitano, E., Giangaspero, A. & Puccini, V. (2000). Differentiation by polymerase chain reaction – restriction fragment length polymorphisms of some Oestridae larvae causing myiasis. *Vet Parasitol* 90: 305-313.

- Rand, D.M. (1994). Thermal habit, metabolic rate and the evolution of mitochondrial DNA. *Trends Ecol Evol* 9: 125-131.
- Roehrdanz, R.L. (1995). Amplification of complete insect mitochondrial genome in two easy pieces. *Insect Mol Biol* 4: 169-172.
- Roehrdanz, R.L., Degrugillier, M.E. & Black IV, W.C. (2002). Novel rearrangements of arthropod mitochondrial DNA detected with long-PCR: applications to arthropod phylogeny and evolution. *Mol Biol Evol* 19: 841-849.
- 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.
- Sagan, L. (1967). On the origin of mitosing cells. *J Theor Biol* 14: 255-274.
- Saito, S., Tamura, K. & Aotsuka, T. (2005). Replication origin of mitochondrial DNA in insects. *Genetics* 171: 1695-1705.
- Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular cloning: a laboratory manual*. 2^a edição. Cold Spring Harbor Laboratory Press.
- Scheffer, S.J., Lewis, M.L. & Joshi, R.C. (2006). DNA barcoding applied to invasive leafminers (Diptera: Agromyzidae) in the Philippines. *Ann Entomol Soc Am* 99: 204-210.
- Shadel, G.S. & Clayton, D.A. (1997). Mitochondrial DNA maintenance in vertebrates. *Annu Rev Biochem* 66: 409-435.
- Shao, R., Barker, S.C., Mitani, H., Aoki, Y. & Fukunaga, M. (2005a). Evolution of duplicate control regions in the mitochondrial genomes of Metazoa: a case study with Australasian *Ixodes* ticks. *Mol Biol Evol* 22: 620-629.
- Shao, R., Mitani, H., Barker, S.C., Takahashi, M. & Fukunaga, M. (2005b). Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the chigger mite, *Leptotrombidium pallidum*. *J Mol Evol* 60: 764-773.
- Shao, R., Campbell, N.J. & Barker, S.C. (2001). Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthitaptera). *Mol Biol Evol* 18: 858-865.
- Shimko, N., Liu, L. Lang, B.F. & Burger, G. (2001). GOBASE: the organellar genome database. *Nucleic Acids Res* 29: 128-132.
- Simison, W.B., Lindberg, D.R. & Boore, J.L. (2006). Rolling circle amplification of metazoan mitochondrial genomes. *Mol Phylogenet Evol* 39: 562-567.

- 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 polymerase chain reaction primers. *Ann Entomol Soc Am* 87: 651-701.
- Snäll, N., Huoponen, K., Savontaus, M.-L. & Ruohomäki, K. (2002). Tandem repeats and length variation in the mitochondrial DNA control region of *Epirrita autumnata* (Lepidoptera: Geometridae). *Genome* 45: 855-861.
- Sorenson, M.D., Ast, J.C., Dimcheff, D.E., Yuri, T. & Midell, D.P. (1999). Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Mol Phylogenet Evol* 12: 105-114.
- Stanton, D.J., Daehler, L.L., Moritz, C.C. & Brown, W.M. (1994). Sequences with the potential to form stem-and-loop structures are associated with coding-region duplications in animal mitochondrial DNA. *Genetics* 137: 233-241.
- Stevens, J.R., Wall, R. & Wells, J.D. (2002). Paraphyly in Hawaiian hybrid blowfly populations and the evolutionary history of anthropophilic species. *Insect Mol Biol* 11: 141-148.
- Stork, M.G. (1979). The epidemiological and economic importance of fly infestation of meat and milk producing animals in Europe. *Vet Rec* 105:341-343.
- Taanman, J.W. (1999). The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta*. 1410: 103-123.
- Tajima, F. & Nei, M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Mol Biol Evol* 1: 269-285.
- 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: 4876-4882.
- Tong, K.L. & Wong, J.T. (2004). Anticodon and wobble evolution. *Gene* 333: 169-177.
- Torres, T.T. & Azeredo-Espin, A.M.L. (2005). Development of new polymorphic microsatellite markers for the New World screw-worm *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Mol Ecol Notes* 5: 825-817.
- Torres, T.T., Brondani, R.P.V., Garcia, J.E. & Azeredo-Espin, A.M.L. (2004). Isolation and characterization of microsatellite markers in the new world screw-worm *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Mol Ecol Notes* 4: 182-184.

- Valério, J.R. & Guimarães, J.H. (1983). Sobre a ocorrência de uma nova praga, *Haematobia irritans* (L.) (Diptera: Muscidae), no Brasil. *Revta Bras Zool* 1: 417-418.
- Valle, J.S. & Azeredo-Espin, A.M.L. (1995). Mitochondrial DNA variation in two Brazilian populations of *Cochliomyia macellaria* (Diptera: Calliphoridae). *Braz J Genetics* 18: 382-391.
- Van Bortel, W., Sochanta, T., Harbach, R.E., Socheat, D., Roelants, P., Backeljau, T. & Coosemans, M. (2002). Presence of *Anopheles culicifacies* B in Cambodia established by the PCR-RFLP assay developed for the identification of *Anopheles minimus* species A and C and four related species. *Med Vet Entomol* 16: 329-334.
- Vandewoestijne, S., Baguette, M., Brakefield, P.M. & Saccheri, I.J. (2004). Phylogeography of *Aglaia urticae* (Lepidoptera) based on DNA sequences of the mitochondrial COI gene and control region. *Mol Phylogenet Evol* 31: 630-646.
- Vasconcelos, A.T., Guimarães, A.C., Castelletti, C.H., Caruso, C.S., Ribeiro, C., Yokaichiya, F., Armoa, G.R., Pereira, G. da S., da Silva, I.T., Schrago, C.G., Fernandes, A.L., da Silveira, A.R., Carneiro, A.G., Carvalho, B.M., Viana, C.J., Gramkow, D., Lima, F.J., Correa, L.G., Mudado, M. de A., Nehab-Hess, P., Souza, R., Correa, R.L. & Russo, C.A. (2005). MamMiBase: a mitochondrial genome database for mammalian phylogenetic studies. *Bioinformatics* 21: 2566-2567.
- Vila, M. & Björklund, M. (2004). The utility of the neglected mitochondrial control region for evolutionary studies in Lepidoptera (Insecta). *J Mol Evol* 58: 280-290.
- 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: 60-67.
- Wells, J.D. & Sperling, F.A.H. (1999). Molecular phylogeny of *Chrysomya albiceps* and *C. rufifacies* (Diptera: Calliphoridae). *J Med Entomol* 36: 222-226.
- Wells, J.D. & Sperling, F.A.H. (2001). DNA-based identification of forensically important Chrysomyinae (Diptera: Calliphoridae). *Forensic Sci Int* 120: 110-115.
- Wells, J.D., Pape, T. & Sperling, F.A.H. (2001). DNA-based identification and molecular systematics of forensically important Sarcophagidae (Diptera). *J Forensic Sci* 46: 1098-1102.

- Wolstenholme, D.R. (1992). Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol* 141: 173-216.
- Wyman, S.K., Jansen, R.K. & Boore, J.L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252-3255.
- Xia, X. (2005). Mutation and selection on the anticodon of tRNA genes in vertebrate mitochondrial genomes. *Gene* 345: 13-20.
- Yamauchi, M.M., Miya, M.U. & Nichida, M. (2004). Use of a PCR-based approach for sequencing whole mitochondrial genomes of insects: two examples (cockroach and dragonfly) based on the method developed for decapod crustaceans. *Insect Mol Biol* 13: 435-442.
- 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.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 1-10.