UNIVERSIDADE ESTADUAL DE CAMPINAS

doutorado

INSTITUTO DE BIOLOGIA

UNIVERSIDADE ESTADUAL DE CAMPINAS - UNICAMP

ANA CLÁUDIA LESSINGER

O GENOMA MITOCONDRIAL DE MOSCAS CAUSADORAS DE MIÍASES: SEQUÊNCIA, ORGANIZAÇÃO E EVOLUÇÃO

Tese apresentada para a obtenção do título de DOUTOR junto ao Programa de Pós-Graduação em Genética e Biologia Molecular do Instituto de Biologia da Universidade Estadual de Campinas - UNICAMP

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

Campinas – São Paulo – 2001

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

Lessinger, Ana Cláudia

L566g O genoma mitocondrial de moscas causadoras de Miíases: Seqüência, organização e evolução/ Ana Cláudia Lessinger. -- Campinas, S.P: [s.n.], 2001.

Orientadora: Ana Maria Lima de Azeredo-Espin
Tese (doutorado) – Universidade Estadual de Campinas. Instituto de Biologia.

1. Miíases. 2. DNA mt. 3. Genoma mitocondrial. I. Azeredo-Espin, Ana Maria Lima de. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Campinas, 12 de Dezembro de 2001

BANCA EXAMINADORA

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

Assinatura

Prof.a Dr.a Anete Pereira de Souza

Assinatura

Prof.a Dr.a Cristina Yumi Miyaki

Assinatura

Prof. Dr. Francisco Gorgônio da Nóbrega

Assinatura

Prof. Dr. Sérgio Furtado dos Reis

Assinatura

Prof. Dr. Louis Bernard Klaczko

Assinatura

Prof.a Dr.a Christine Hackel

Assinatura

Aos meus pais e irmãs,

"Mas logo o pequeno pássaro começará a ensaiar seus vôos incertos. Agora não serão mais os braços do pai, arredondados num abraço, que irão definir o espaço do ninho. Os braços do pai terão de se abrir para que o ninho fique maior. E serão os olhos do pai, no espaço que seus braços já não podem conter, que irão marcar os limites do ninho. A criança se sente segura se, de longe, ela vê que os olhos do seu pai a protegem. Olhos também são colos. Olhos também são ninhos. "Não tenha medo. Estou aqui! Estou vendo você": é isso o que eles dizem, os olhos do pai."

Rubem Alves

Por me olharem sempre, por enxergarem quem sou, por me ensinarem a ver...
Pois são os olhos de Laura, Egídio, Júlia e Alessandra que sempre procuro e encontro, sempre.

DEDICO

Ao Paulão

Por me ajudar a descobrir quem sou,
por me transformar enquanto me descubro,
por se encantar enquanto me transformo,
por se revelar enquanto me encanto.

Passo a você as palavras de um homem experiente que disse:

"Uma indicação do zelo da verdade é o fato de que ela não traçou para ninguém uma trilha até ela, e não privou ninguém da esperança de alcança-la, deixando as pessoas correndo nos desertos da perplexidade e se afogando nos mares da dúvida; e aquele que pensa que a atingiu, dela se afasta, e aquele que pensa que se afastou dela, perdeu seu caminho. Portanto, não há como alcança-la e não há como evita-la: ela é inescapável.".

noites das mil e uma noites – Naguib Mahfouz

AGRADECIMENTOS

Segundo Holanda (1986), *ORIENTAÇÃO* define-se conforme segue:
"1. Ato ou *arte* de orientar 2. Direção, *guia*, regra 3. *Impulso*, tendência, inclinação 4. Fase do ciclo docente em que o professor acompanha, utilizando técnicas, recursos e procedimentos adequados, a *marcha* do aprendizado de seus alunos."

Agradeço à Prof.a Dr.a Ana Maria Lima de Azeredo-Espin por esta experiência extraordinária, por me acompanhar nesta marcha, sendo guia nesta aventura, abrindo as portas do laboratório de Genética Animal e do diversificado mundo do ensino e da pesquisa que me conquistou para sempre.

Agradeço à Prof.a Dr.a Christine Hackel, por sua disposição e entusiasmo em participar da avaliação da minha tese, e principalmente por sua dedicação histórica às causas acadêmicas e à formação de alunos, sempre uma referência imprescindível.

Agradeço ao Prof. Dr. Louis Bernard Klaczko, por sua presença marcante, pelo exemplo profissional, por suas palavras sinceras e suas reflexões pausadas que contribuíram e contribuem profundamente na minha formação.

Agradeço ao Prof. Dr. Sérgio Furtado dos Reis, por nossas conversas tão oportunas, por seus comentários e sugestões relevantes tanto para a formação do profissional quanto da pessoa. Por compartilhar seu olhar autêntico sobre a natureza dos padrões e os padrões da natureza.

Agradeço à Prof.a Dr.a Anete Pereira de Souza, pela contribuição contínua na minha formação acadêmica, por seu entusiasmo e dedicação às causas "mitocondriais" e à evolução deste sistema cativante, pela disposição em participar da banca examinadora da minha tese.

Agradeço à Prof.a Dr.a Cristina Yumi Miyaki, pela disposição em contribuir com sua avaliação crítica da minha tese, aceitando de imediato ser parte desta história.

Agradeço ao Prof. Dr. Francisco Gorgônio da Nóbrega, por sua disposição entusiasmada em participar da leitura e avaliação da minha tese, por suas palavras de estímulo e atenção dedicada em diferentes etapas da minha formação acadêmica.

Agradeço aos colegas, estagiários, técnicos e amigos do laboratório de Genética Animal, Ana Carolina Martins Junqueira, Partick Litjens, Tatiana Teixeira Torres, Fábio Frangiote Conte, Mariana Lúcio Lyra, Patrícia Kato, Ize Penhas de Lima, Marcos Túlio de Oliveira, Rosângela Aparecida Rodrigues, Maria Salete do Couto Campos e Mário Constantino Filho, por proporcionarem um ambiente de ricas discussões acadêmicas e científicas, além do privilégio de um convívio saudável e intenso baseado em trabalho, amizade, confiança, respeito, empenho, enfim, uma verdadeira equipe, em que eu me reconheço e a qual muito admiro.

Agradeço à Prof.a Dr.a Maria Elena Infante-Malachias, pelo privilégio de sua amizade, pelos segredos divertidos que sopram no vento superando as distâncias e o tempo, pela cumplicidade que não conhece fronteiras, por acreditar em "*un tiempo nuevo, donde todo es posible*" e compartilhar esta certeza.

Agradeço aos colegas Dr. Edson Luís Kemper, Dr. André Luiz Vettore e Dr. Felipe Rodrigues da Silva do laboratório Genoma, e ao coordenador do CBMEG, Prof. Dr. Paulo Arruda, pela oportunidade de desenvolver trabalhos conjuntos, integrando recursos humanos e técnicos de reconhecida qualidade e excelência, com disposição e bom humor, imprescindíveis na investigação científica.

Agradeço ao Dr. Jeffrey Wells, por sua participação singular na minha formação, por estimular minha produção científica direta e indiretamente, a partir de 1996 (XX Internacional Congress of Entomology), em seguida em 1998 (IV International Congress of Dipterology) e, recentemente, em 2000 (XXI International Congress of Entomology).

Agradeço aos amigos da Turma 90 (Biologia - UNICAMP) pelos encontros constantes, por uma amizade duradoura, pelas tantas histórias e momentos nostálgicos que compartilhamos e pela união que avança no tempo.

Agradeço especialmente aos amigos Alessandra Roque Pancetti (Pan), Lee Wen Wha (LeeMan) e Ricardo Augusto Dante (Dante) pela acolhida calorosa e entusiasmada durante nossas aventuras na "América". Agradecimentos especialíssimos também para o grande amigo Ricardo Hideki Nonaka, por compartilhar intensamente estes momentos inesquecíveis e cultivar esta amizade com dedicação.

Aos amigos Juliana, Eduardo, Giovanna, Heloísa, Alexandre e Marissol, Rogério, Luciana e Letícia, por uma amizade constantemente renovada e preciosa.

Agradeço aos amigos do CBMEG pelo apoio constante, por transformar este centro em um ambiente humano, científico, acadêmico, educacional e cultural, pelos inúmeros auxílios e consultas-relâmpago sempre atendidos com enorme entusiasmo e bom humor.

Agradeço à família Checchia, pela atenção e carinho espontâneos, por me ajudar a crescer, estimulando minhas conquistas e acreditando nas minhas escolhas. Agradeço especialmente à Vó Elvira, por seu carinho sem limites, que aquece nossa memória e nos deixa com saudades profundas.

Agradeço à Sub-Comissão de Pós-Graduação em Genética e Biologia Molecular, pelas reuniões entusiasmadas, pela oportunidade em aprender e contribuir como representante discente. Agradecimentos especiais aos colegas do curso de pós-graduação, que deram sentido a esta representação, proporcionando encontros e discussões inesquecíveis.

Agradeço a toda a equipe da secretaria do CBMEG, pela prestatividade com que sempre se dispõem a atender minhas inúmeras solicitações. Meu *Muito Obrigado(a)* a Tânia, Sandra, Fabiana e Patrícia.

Agradeço especialmente às secretárias do curso de Pós-Graduação, Zaira e Lia, pela disposição em atender a um turbilhão de solicitações acadêmicas, pelas tantas dúvidas respondidas e pela preocupação e zelo com meus direitos e deveres acadêmicos.

Agradeço à Agência Internacional de Energia Atômica (IAEA) pela oportunidade de participar de seu programa de treinamento e fazer parte do seleto grupo SIT, congregando membros de diversas origens e culturas em favor de um objetivo comum: o controle de pragas. Agradecimentos especiais às amigas Laura Marques e Madelaine Chifflet por me apresentarem as reais dimensões do *Uruguay*.

Agradeço a oportunidade de participar dos programas de capacitação docente PED (2000) e PECD (2001) e aos docentes e secretarias envolvidos na realização e avaliação destes estágios. Agradecimentos especiais aos alunos da disciplina BG380 que tornaram esta experiência uma aventura didático-científica emocionante.

Agradeço à FAEP e à FUNCAMP (UNICAMP) pelos auxílios concedidos, que beneficiaram de forma muito significativa a divulgação dos meus trabalhos e do laboratório de Genética Animal no contexto científico internacional.

Agradeço à Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e ao Conselho Nacional de Pesquisa Científica e Tecnológica (CNPq) pelo suporte financeiro para a realização deste trabalho.

Agradeço à Universidade Estadual de Campinas (UNICAMP), local onde foi realizado este trabalho.

ABREVIATURAS

A - adenina

Ala - alanina

AM - Amazonas

Arg - arginina

Asn - asparagina

Asp – ácido aspártico

ATP – adenina trifosfato

AT/12S – região gência contendo as sequências parciais da região rica em A+T e do gene do rRNA 12S

bp – base pair

C - citosina

CAPES – coordenadoria de aperfeiçoamento de pessoal de nível superior

CBMEG – centro de biologia molecular e engenharia genética

CIR – *Chrysomya* intergenic region

CNPq – conselho nacional de desenvolvimento científico e tecnológico

COI – subunidade I do complexo citocromo oxidase c

COII – subunidade II do complexo citocromo oxidase c

COI/II – região gênica que compreende parte dos genes das subunidades I e II do complexo citocromo oxidase

COIII - subunidade III do complexo citocromo oxidase c

CR – control region

CSB – conserved sequence block

Cytb – citocromo b

Cys - cisteína

D-loop – displacement loop

DNA – ácido desoxiribonucléico

DNAmt – DNA mitocondrial

dNTPs - desoxiribonucleotídeos

FAPESP – fundação de amparo à pesquisa do Estado de São Paulo

G - guanina

GEP – gap extension penalty

Gln - glutamina

Gly - glicina

I - isoleucina

Ile - isoleucina

JGI – joint genome institute

Kb - kilobase

Leu - leucina

lrRNA – large ribosomal RNA

Lys - lisina

M - metionina

MAST – motif alignment and search tool

MEGA – molecular evolutionary genetics analysis

Met - metionina

MG – Minas Gerais

MGASG – mitochondrial gene arrangement source guide

min - minutes

mRNA – messenger RNA (ou RNA mensageiro)

MT – Mato Grosso

mtDNA – mitochondrial DNA

NADH – ubiquinona oxidorredutase

NCBI – national center of biotechnology information

ND 1 - subunidade 1 do complexo da NADH desidrogenase

ND 2 - subunidade 2 do complexo da NADH desidrogenase

ND 3 - subunidade 3 do complexo da NADH desidrogenase

ND 4 - subunidade 4 do complexo da NADH desidrogenase

ND 4L - subunidade 4L do complexo da NADH desidrogenase

ND 5 - subunidade 5 do complexo da NADH desidrogenase

ND 6 - subunidade 6 do complexo da NADH desidrogenase

ng - nanograma

NJ – neighbor-joining

NWS – new world screwworm

OGMP – Organelle MegaSequencing Program

OGP – open gap penalty

OR – origin of replication

ORF – open reading frame

PADCT – programa de apoio ao desenvolvimento científico e tecnológico

pb – pares de base

PCR – polymerase chain reaction

Phe - fenilalanina

PR - Paraná

Pro - prolina

RC – região controle

RFLP – restriction fragment length polymorphism

RJ – Rio de Janeiro

RNA – ácido ribonucléico

rRNA – ribosomal RNA (ou RNA ribossomal)

RSCU – relative synonymous codon usage

S - serina

s - seconds

SE – standard error

Ser - serina

SP – São Paulo

srRNA – small ribosomal RNA

SSM – slippage strand mispairing

T - timina

tRNA – transfer RNA (ou RNA transportador)

Tyr - tirosina

tRNA^{Ala} – RNA transportador de alanina

tRNA^{Arg} - RNA transportador de arginina

tRNA^{Asn} - RNA transportador de asparagina

tRNA^{Asp} - RNA transportador de ácido aspártico

tRNA^{Cys} - RNA transportador de cisteina

tRNA^{Gln} - RNA transportador de glutamina

tRNA^{Glu} - RNA transportador de ácido glutâmico

tRNA^{Gly} - RNA transportador de glicina

tRNA^{His} - RNA transportador de histidina

tRNA^{Ile} - RNA transportador de isoleucina

tRNA^{Leu} - RNA transportador de leucina

tRNA^{Lys} - RNA transportador de lisina

tRNA^{Met} - RNA transportador de metionina

tRNA^{Phe} - RNA transportador de fenilalanina

tRNA^{Pro} - RNA transportador de prolina

tRNA^{Ser} - RNA transportador de serina

tRNA^{Trp} - RNA transportador de triptofano

tRNA^{Tyr} - RNA transportador de tirosina

tRNA^{Thr} - RNA transportador de treonina

UNICAMP – Universidade Estadual de Campinas

US$ - United States dollars

ÍNDICE **PÁGINA**

RESUMO xiv

ABSTRACT xvi

INTRODUÇÃO

1. GENOMA MITOCONDRIAL ANIMAL: UM CENÁRIO 01

2. GENOMA MITOCONDRIAL DE ARTHROPODA: HISTÓRICO 05

2.1. PLASTICIDADE DO GENOMA MITOCONDRIAL: tRNAs E REGIÃO CONTROLE 06

2.2. PLASTICIDADE DO GENOMA MITOCONDRIAL: GENES CODIFICADORES DE PROTEÍNAS 07

2.3. GENÔMICA MITOCONDRIAL 07

3. BANCOS DE DADOS 08

4. ANÁLISE DO GENOMA MITOCONDRIAL EM MOSCAS CAUSADORAS DE MIÍASES 09

5. APRESENTAÇÃO DOS ARTIGOS 11

CAPÍTULOS

CAPÍTULO 1 - Evolution and structural organization of the mitochondrial DNA control region of myiasis-causing flies 15

CAPÍTULO 2 - The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae) 26

CAPÍTULO 3 - Caracterization of *Cochliomyia hominivorax* and *Cochliomyia macellaria* by PCR-RFLP of mitochondrial DNA 36

CAPÍTULO 4 - The occurrence of duplication in the mtDNA control region of *Chrysomya* species (Diptera: Calliphoridae) involving two tRNAs 43

DISCUSSÃO

1. REGIÃO CONTROLE DO DNA MITOCONDRIAL 70

2. DUPLICAÇÃO GÊNICA EM Chrysomya 71

3. DIAGNÓSTICO ATRAVÉS DE PCR-RFLP 73

4. O GENOMA MITOCONDRIAL DE C. hominivorax 74

5. CONSIDERAÇÕES GERAIS 75

REFERÊNCIAS BIBLIOGRÁFICAS 77

RESUMO

O genoma mitocondrial animal (DNAmt) em geral apresenta-se com uma ampla diversidade em conteúdo e organização da informação genética quando são analisados os diferentes grupos de organismos que possuem genomas organelares. Entre as espécies causadoras de miíases, a análise de marcadores moleculares no genoma mitocondrial tem contribuído para esclarecer os mecanismos evolutivos envolvidos na manutenção da variabilidade genética e tem se mostrado extremamente útil para avaliar a estrutura de populações de espécies-praga, além de permitir diagnósticos moleculares interespecíficos e a inferência de relações filogenéticas em Calliphoridae. Nesta tese, diferentes aspectos da organização, estrutura e evolução de seqüências do DNA mitocondrial de espécies causadoras de miíases foram investigados conforme descrito em quatro artigos independentes (capítulos 1 ao 4). Um dos artigos mostra que a caracterização do genoma mitocondrial de *Cochliomyia hominivorax* (mosca-da-bicheira) contribui para a compreensão dos padrões gerais de organização e evolução do genoma mitocondrial neste grupo e a investigação da sua diversidade (*Lessinger et al.*, 2000*). O genoma mitocondrial de *C. hominivorax* foi o primeiro a ser completamente seqüenciado no Brasil, ampliando a perspectiva de desenvolvimento de áreas de genômica mitocondrial e comparativa. Em outro artigo, a caracterização da seqüência nucleotídica da região controle do DNAmt de *C. hominivorax* (*Lessinger* & Azeredo-Espin, 2000*) resultou na identificação de domínios conservados e não conservadas e de eventos de reorganização estrutural próximos à origem de replicação, contribuindo na análise global deste genoma. Neste contexto, um terceiro artigo descreve a presença de uma duplicação gênica na região controle de espécies do gênero *Chrysomya* (moscas varejeiras), envolvendo a sequência completa do $tRNA^{Ile}$ e sequências parciais do $tRNA^{Gln}$ e da própria região controle. Esta reorganização gênica é uma evidência da flexibilidade estrutural associada à origem de replicação do DNAmt (*Lessinger et al.* em preparação*). A plasticidade de parte desta região contrasta com a conservação em sequência primária e a presença de elementos estruturais homólogos localizados na extremidade 5', possivelmente envolvidos em processos de replicação e transcrição do genoma (*Lessinger* & Azeredo-Espin, 2000*). E, finalmente, um quarto artigo descreve a utilidade da região controle como eficiente marcador para a identificação espécie-específica de moscas causadoras de miíases, primária e secundária, do gênero *Cochliomyia*, através da análise de padrões de PCR-RFLP (Litjens, *Lessinger* & Azeredo-Espin, 2001*). Esta análise representa uma importante estratégia para o monitoramento de eventos de

invasão e re-colonização de *C. hominivorax* em áreas onde esta praga foi previamente erradicada ou apresenta risco invasor potencial, contribuindo na manutenção e avaliação de programas de controle deste ectoparasita.

* Nota: os trabalhos publicados em Lessinger & Azeredo-Espin, 2000; Lessinger *et al.*, 2000; Litjens, Lessinger & Azeredo-Espin, 2001 e Lessinger *et al.*, em preparação são parte integrante da tese.

ABSTRACT

In general, the mitochondrial genome presents a wide diversity regarding gene content and structural organization of the genetic information for different groups of organisms. The mitochondrial genome sequencing in arthropods contributes for the analysis of general patterns of evolution in invertebrates, providing important data on the phylogenetic utility of specific regions and improving the characterization of molecular markers for population genetics. Characterization of molecular markers in the mtDNA of myiasis-causing flies have been used for the identification of evolutionary mechanisms responsible for specific patterns of genetic variability and are very useful for the analysis of population structures of pest species. Mitochondrial markers have been used for providing species-specific molecular identification and inferring Calliphoridae phylogenetic relationships. Different aspects regarding organization, evolution and potential applications of mitochondrial sequences in myiasis-causing flies were discussed in this thesis, organized in four independent articles (cap. 1 to 4). One of the papers shows that the characterization of the mitochondrial genome of the screwworm fly, *C. hominivorax*, which contributes in the analysis of general patterns of structural organization, evolution, and diversity in myiasis-causing flies (*Lessinger et al.* 2000*). The mitochondrial genome of *C. hominivorax* was the first to be completely sequenced in Brazil, providing an important contribution for the development of further studies on mitochondrial and comparative genomics. In another article, the characterization of mtDNA control region sequences in screwworm (*Lessinger* & Azeredo-Espin, 2000*) provided the identification of hyper-variable and conserved domains. A third paper describes the occurrence of a duplication involving complete $tRNA^{Ile}$ sequences in *Chrysomya* control region (blowflies). This organization indicates different degrees of mtDNA plasticity associated with the replication origin (*Lessinger et al.*, in preparation*). This plasticity contrasts with the conservation of specific sequence elements, possible associated with the replication/transcription process (*Lessinger* & Azeredo-Espin, 2000*). Finally, a fourth paper describes the utility of control region sequences for species-specific identification of myiasis-causing flies from *Cochliomyia*, by PCR-RFLP analysis (Litjens, *Lessinger* & Azeredo-Espin, 2001*). These molecular markers could be useful in monitoring introduction and re-colonization patterns of *C. hominivorax* from regions where it have been previously eradicated or .represents a potential risk of invasion. This 'diagnose' strategy could contribute in evaluating and monitoring screwworm's control programs.

INTRODUÇÃO

1. O genoma mitocondrial animal: um cenário

Através da análise de dados relativos a genomas mitocondriais completos descritos para Arthropoda, compilados a partir de bancos de dados públicos (ver *Bancos de dados*) e da literatura científica, é possível investigar o estado atual da pesquisa envolvendo genomas mitocondriais. Existem atualmente apenas 25 espécies de artrópodes cujos genomas mitocondriais estão descritos (Tabela 1), completando um total de 230 genomas mitocondriais sequenciados (GOBASE/Outubro, 2001). Uma vez que o filo Arthropoda representa uma das categorias taxonômicas mais diversificadas do reino animal, estes dados reforçam a importância de ampliar a caracterização de estudos relativos a genomas mitocondriais neste grupo. De modo geral, das espécies descritas na tabela 1, existe um predomínio de organismos de importância médica/veterinária ou econômica, como os mosquitos transmissores da malária (*Anopheles*), um vetor de tripanosomomíiases (*Triatoma*), pragas agrícolas e de pecuária (*Ceratitis, Cochliomyia, Ixodes*, *Rhipicephalus* e *Tribolium*), além da mosca-varejeira (*Chrysomya*) e do bicho-da-seda (*Bombyx*). Através da visualização gráfica destes dados, é possível verificar a evolução dos estudos relacionados com a descrição de genomas mitocondriais completos em Arthropoda, considerando-se a diversidade de espécies representadas (Gráfico 1) ou o número total de genomas descritos, incluindo redundâncias (Gráfico 2). Este quadro pode ser interpretado considerando-se aspectos qualitativos e quantitativos relacionados à investigação de genomas mitocondriais em artrópodes. É interessante observar que esta evolução no número de genomas descritos está estreitamente associada ao desenvolvimento de novas estratégias técnicas e metodológicas na área e sua aplicação no estudo de insetos e outros artrópodes.

Existe um número significativo de revisões sobre diferentes aspectos do genoma mitocondrial animal. De maneira geral, estes estudos referem-se à caracterização da estrutura e evolução do genoma mitocondrial (Wolstenholme, 1992); a investigação de aspectos evolutivos relacionados com a teoria da endosimbiose (Gray *et al.*, 1999; Lang *et al.*, 1999); a avaliação do potencial filogenético de seqüências mitocondriais completas (Curole & Kocher, 1999; Hwang *et al.*, 2001); a caracterização da ampla diversidade relacionada à organização estrutural deste genoma (Boore, 1999 e Saccone *et al.*, 1999); e a evolução de mecanismos de replicação e transcrição da molécula (Shadel & Clayton, 1997). Revisões sobre aspectos bioquímicos e de biogênese da mitocôndria, seu papel no metabolismo energético celular e a ocorrência de polimorfismos relacionados ao desenvolvimento de doenças e envelhecimento, não serão abordadas neste contexto.

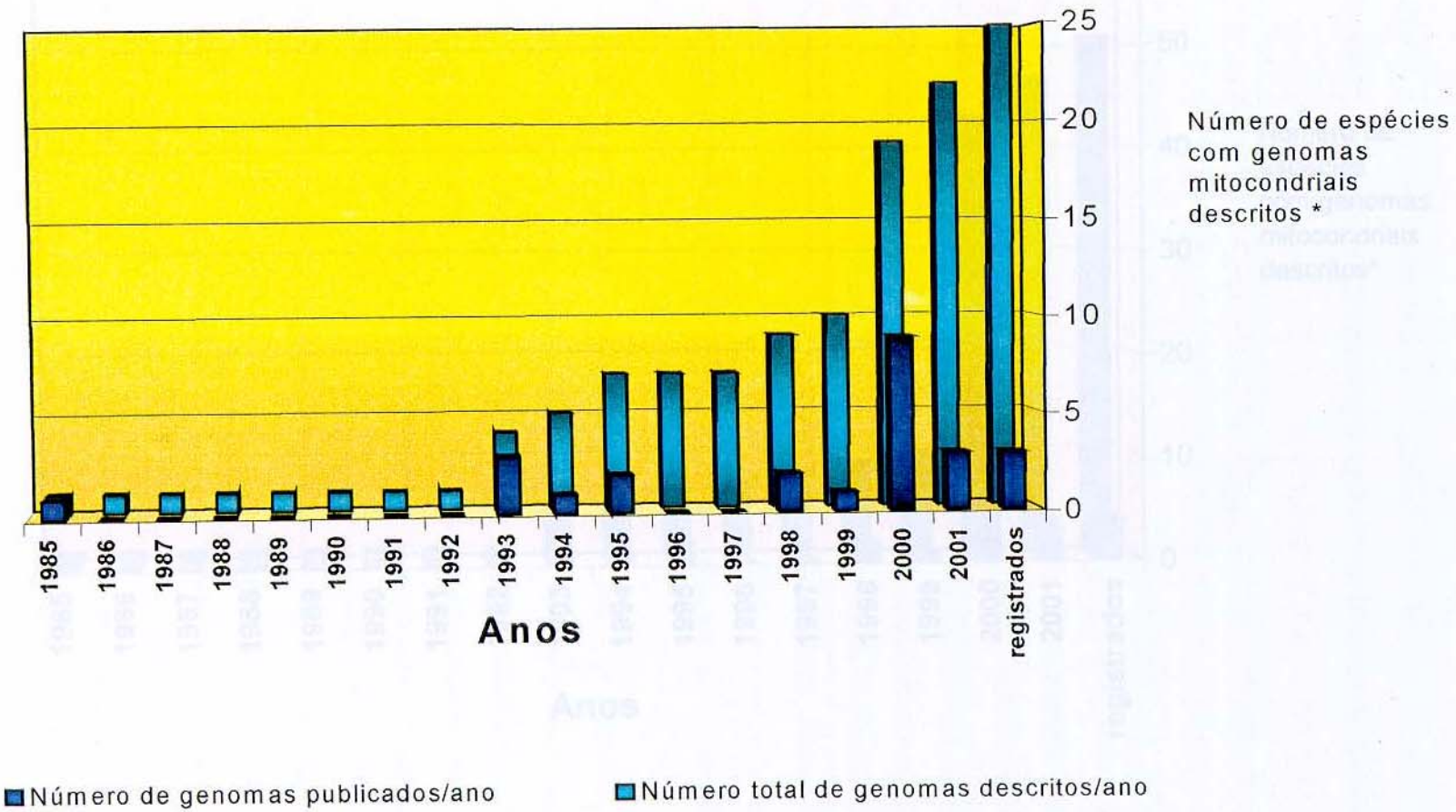
Tabela 1 – Genomas mitocondriais completos descritos em Arthropoda (Outubro/2001)

ESPÉCIES	ACCESSO/ "DATABASE"	REFERÊNCIAS
Anopheles gambiae	NC_002084[1,2,3]	Beard *et al.* (1993)
Anopheles quadrimaculatus	NC_000875[1,2,3]	Mitchell *et al.* (1993)
Apis mellifera	NC_001566[1,2,3]	Crozier & Crozier (1993)
Artemia franciscana	NC_001620[1,2,3]	Valverde *at al.* (1994)
Bombyx mori (2)	NC_002355[1,2,3]	Lee *et al.* unpublished
	AY048187[2]	Lu *et al.* unpublished
Ceratitis capitata	NC_000857[1,2,3]	Spanos *et al.* (2000)
*Chrysomya putoria**	NC_002697[1,2]	Junqueira *et al.* (em preparação)
Cochliomyia hominivorax	NC_002660[1,2]	Lessinger *et al.* (2000)
Daphnia pulex	NC_000844[1,2,3]	Crease (1999)
Drosophila mauritiana (2)	AF200830[2]	Ballard (2000a)
	AF200831[2]	Ballard (2000a)
Drosophila melanogaster (3)	NC_001709[1,2,3]	Lewis *et al.* (1995)
	AF200828[2]	Ballard (2000a)
	AF200829[2]	Ballard (2000a)
Drosophila sechellia	AF200832[2]	Ballard (2000a)
Drosophila simulans (20)	AF200833[2] A AF200854[2]	Ballard (2000b)
Drosophila yakuba	NC_001322[1,2,3]	Clary & Wolstenholme (1985)
Heterodoxus macropus	NC_0002651[1,2]	Shao *et al.* (2001a)
Ixodes hexagonous	NC_002010[1,2,3]	Black & Roehrdanz (1998)
Limulus polyphemus	NC_003057[1,3]	Lavrov *et al.* (2000a)
Lithobius forficatus (2)	NC_002629[1,2]	Lavrov *et al.* (2000b)
	AJ270997[2]	Hwang et al. (2001)
Locusta migratoria	NC_001712[1,2,3]	Flook *et al.* (1995)
Pagurus longicarpus	NC-003058[1,2,3]	Hickerson & Cunningham (2000)
Penaeus monodon	NC_002184[1,2,3]	Wilson *et al.* (2000)
Rhipicephalus sanguineus	NC_002074[1,2,3]	Black & Roehrdanz (1998)
Tetrodontophora bielanensis	NC-002735[1,2]	Nardi *et al.* (2001)
Triatoma dimidiata	NC_002609[1,2]	Dotson & Beard (2001)
Tribolium castaneum	NC_003081[1,2]	Friedrich & Muqim unpublised

*(ou *Chrysomya chloropyga*, sinônimo), NCBI = 1, OGMP = 2, MGASG = 3.

GRÁFICO 1

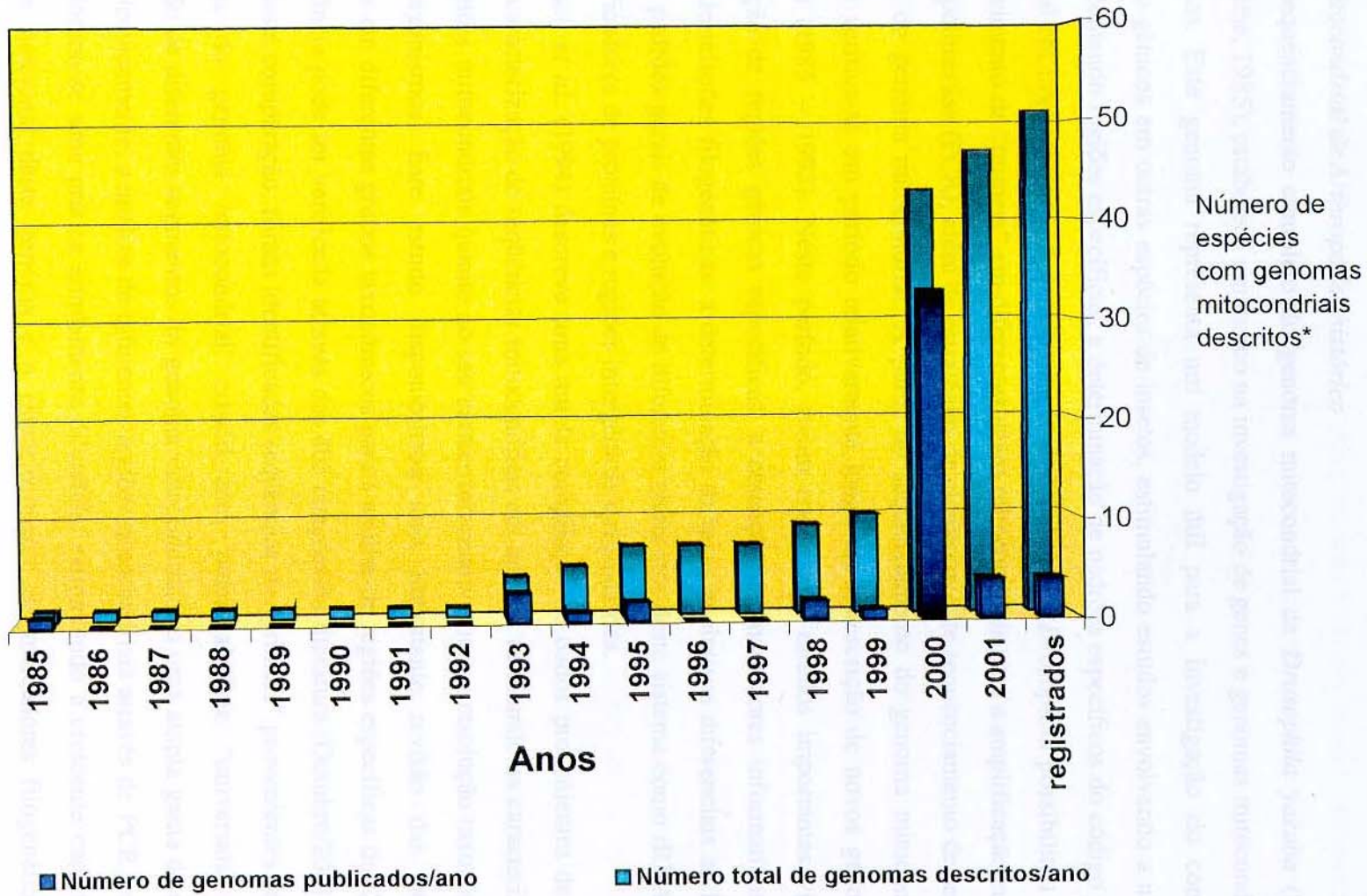
EVOLUÇÃO TEMPORAL DOS ESTUDOS DE GENOMAS MITOCONDRIAIS COMPLETOS EM ARTHROPODA



** espécies com mais de um genoma mitocondrial descrito foram consideradas uma única vez*
*"**registrados**" corresponde aos genomas publicados apenas no GenBank até Novembro/2001*

GRÁFICO 2

EVOLUÇÃO TEMPORAL DOS ESTUDOS DE GENOMAS MITOCONDRIAIS COMPLETOS EM ARTHROPODA



** foram incluídos todos os genomas mitocondriais de Arthropoda descritos no GenBank, segundo a tabela 1.*
*"**registrados**" corresponde aos genomas publicados apenas no GenBank até Novembro/2001*

Genoma mitocondrial de Arthropoda: histórico

O sequenciamento completo do genoma mitocondrial de *Drosophila yakuba* (Clary & Wolstenholme, 1985), estabelece um marco na investigação de genes e genomas mitocondriais em invertebrados. Este genoma representa um modelo útil para a investigação do conteúdo e organização gênicos em outras espécies de insetos, estimulando estudos envolvendo a utilização de sondas contendo regiões específicas, a determinação de padrões específicos do código genético mitocondrial em invertebrados. A caracterização do DNAmt em *Drosophila* possibilitou também o desenvolvimento de "primers" em diferentes genes mitocondriais para a amplificação em cadeia pela DNA polimerase (PCR), além de estratégias de clonagem e de sequenciamento de segmentos específicos do genoma mitocondrial. A partir do sequenciamento do genoma mitocondrial de *Drosophila* verifica-se um período relativamente longo até a descrição de novos genomas em Arthropoda (1985 – 1993). Neste período, foram conduzidos estudos importantes visando a caracterização de regiões gênicas específicas, a obtenção de marcadores informativos para a inferência de relações filogenéticas, a determinação de taxas evolutivas diferenciais ao longo da molécula e padrões gerais de evolução de diferentes elementos deste sistema como rRNA, tRNA, genes codificadores de proteínas e regiões intergênicas e regulatórias.

Simon *et al.* (1994) descreve uma ampla compilação de dados provenientes de estudos relativos à caracterização de seqüências mitocondriais em insetos, resultando na caracterização de genes e regiões mitocondriais quanto ao seu caráter informativo para a resolução taxonômica em diversos organismos. Este estudo disponibilizou uma importante revisão das estratégias empregadas em diferentes grupos taxonômicos para a análise de regiões específicas do DNAmt e sua importância pode ser verificada através das 408 citações na literatura (Outubro/2001). Como resultado desta comparação, foram identificadas seqüências de "primers" provenientes de regiões conservadas do genoma mitocondrial, considerados potencialmente "universais" para a amplificação de diferentes segmentos do genoma mitocondrial em uma ampla gama de espécies. Com isso, tecnicamente, a análise de diferentes regiões mitocondriais através de PCR, com raras exceções, tornou-se uma prática amplamente difundida, estimulando a crescente caracterização estrutural e funcional deste genoma e o estabelecimento de marcadores filogenéticos e de populações em inúmeros organismos.

Entretanto, algumas regiões do DNAmt de insetos, reconhecidamente associadas a aceleradas taxas de variação nucleotídica e estrutural, apresentam dificuldades técnicas e metodológicas quando estudadas, dificultando o acesso à informação genética (Zhang & Hewitt,

1997). A região controle de replicação, ou região rica em A+T em insetos, apresenta níveis elevados de variabilidade, além de seqüências com propriedades termodinâmicas específicas, o que interfere significativamente nos processos de amplificação, clonagem e sequenciamento. Além disso, a ocorrência de rearranjos envolvendo seqüências de tRNAs pode representar uma dificuldade adicional para a caracterização de determinadas regiões quando a organização gênica da molécula não é conhecida.

Plasticidade do genoma mitocondrial: tRNAs e região controle

Interpretações sobre a organização do conteúdo gênico do DNAmt de insetos a partir da caracterização dos genomas mitocondriais da abelha *Apis mellifera* e dos mosquitos *Anopheles gambiae* e *A. quadrimaculatus* (Crozier & Crozier, 1993; Beard *et al.*, 1993; Mitchell *et al.*, 1993), sugerem que o genoma apresenta-se conservado em conteúdo e organização gênica, considerando-se os genes codificadores de proteínas e de rRNAs. Entretanto, foram observados eventos de reorganização gênica significativos associados a transposições e inversões envolvendo seqüências de tRNAs. Devido à organização simples e uniforme do genoma mitocondrial, resultando em um sistema genômico compacto e extremamente eficiente (Wolstenholme, 1992), a ocorrência de seqüências de tRNA, estrategicamente distribuídas entre regiões codificadoras de proteínas, indicam sinais de reconhecimento para o processamento de transcritos no mRNA policistrônico (Ojala *et al.*, 1980). Segundo Saccone *et al.* (1999), a flexibilidade funcional de seqüências de tRNAs pode estar associada à dramática redução em tamanho do genoma mitocondrial animal, favorecendo a aquisição de novos atributos operacionais pelas moléculas de tRNAs. Atualmente o processamento de transcritos mitocondriais vem sendo experimentalmente investigado em alguns organismos, proporcionando uma visão mais ampla do sistema de regulação e expressão gênica deste genoma (Lavrov *et al.*, 2000b, Beagley *et al.*, 1999).

Padrões de variação estrutural também podem ser observados na região controle do DNAmt. O genoma mitocondrial de *Drosophila melanogaster* (Lewis *et al.*, 1995) foi descrito apenas após a caracterização da região controle (Lewis *et al.*, 1994), uma região contendo elementos repetidos e organizados em "tandem", responsáveis pela variação em tamanho da molécula de DNAmt em diferentes grupos de *Drosophila*. Em insetos, o sequenciamento completo do genoma mitocondrial de um organismo pode muitas vezes ser comprometido pela dificuldade de obtenção de seqüências referentes à região controle de replicação (P. Ready, comunicação pessoal). Uma revisão sobre a evolução e organização estrutural desta região em insetos e sua utilidade filogenética foi discutida em Zhang & Hewitt (1997).

Plasticidade do genoma mitocondrial: genes codificadores de proteínas

Uma importante contribuição no estudo de genomas mitochondriais em Arthropoda refere-se à caracterização da seqüência completa de duas espécies de carrapatos (Black & Roehrdanz, 1998), descrevendo a ocorrência de inúmeros rearranjos nestes genomas e divergindo da hipótese de conservação estrutural do DNAmt animal. Nestes genomas ocorre uma drástica reorganização do conteúdo gênico, com rearranjos envolvendo inclusive segmentos contendo genes codificadores de proteínas, além de "tradicionais" rearranjos nas seqüências de tRNAs (Campell & Barker, 1998 e 1999). A análise da evolução e organização estrutural de genomas mitocondriais em artrópodes contribui para uma reavaliação da natureza conservada deste genoma. Evidências de rearranjos em Hymenoptera e Hemiptera (Dowton & Austin, 1999; Shao *et al.*, 2001b) e em outros grupos de artrópodes têm contribuído para a investigação de mecanismos envolvidos na reorganização do conteúdo gênico no genoma mitocondrial, um sistema com comprovada diversidade estrutural. No contexto evolutivo, o compartilhamento de determinados rearranjos por diferentes organismos tem sido utilizado para o estabelecimento de relações filogenéticas consideradas antigas (Boore, 1999; Hwang, 2001). O rearranjo "em comum" representa um caráter conservado, sendo interpretado como homologia estrutural. Entretanto, a ocorrência de regiões de "hot-spots" (Dowton & Austin, 1999) ou eventos múltiplos de reorganização no DNAmt de alguns organismos, não descarta a possibilidade de homoplasias associadas aos rearranjos (Saccone *et al.*, 1999).

A ocorrência de rearranjos no genoma mitocondrial de artrópodes foi associada ao desenvolvimento do hábito de parasitismo em alguns grupos (Shao *et al.*, 2001 a e b; Dowton & Campbell, 2001), como resultado da exposição destes genomas a ambientes ricos na produção de radicais livres, estimulada pelo sistema imune do hospedeiro. Entretanto, ainda não existem registros experimentais comprovando esta hipótese. Além disso, foram verificados também rearranjos em organismos de vida livre e nenhuma reorganização de conteúdo gênico em genomas de algumas espécies de ectoparasitas obrigatórios.

Genômica mitocondrial

Segundo os gráficos 1 e 2, um número significativo de genomas mitocondriais foi descrito nos anos de 2000 e 2001. A crescente inovação tecnológica associada ao estudo de genomas, e a geração de processos mais eficientes para obtenção e análise de seqüências nucleotídicas, teve um papel fundamental no estímulo à pesquisa de genomas mitocondriais. A investigação de padrões gerais de evolução através de análises comparativas de genomas completos oferece uma

abordagem extremamente informativa para estudos intra e interespecíficos. A partir de 1999, verifica-se na literatura o emprego dos termos "mitogenomics", "mitochondrial genomics", "mitochondrial proteome" ou "organelle genomics", relacionados ao conceito de genômica mitocondrial, abordagem baseada na interpretação de padrões de evolução molecular ou de organismos a partir da caracterização de genomas mitocondriais em diferentes níveis taxonômicos (Boore, 1999; Saccone *et al.*, 1999; Curole & Kocher, 1999; Hwang *et al.*, 2001; Shimko *et al.*, 2001). A genômica mitocondrial e comparativa foi um recurso recentemente aplicado em Arthropoda para a investigação de padrões evolutivos entre espécies próximas e populações de *Drosophila*, resultando na caracterização simultânea de 27 genomas mitocondriais (Ballard, 2000 a b; Rand, 2001) e estabelecendo uma nova dimensão no estudo de marcadores moleculares.

Bancos de dados

O sequenciamento de genes e genomas mitocondriais ao longo dos últimos anos produziu um enorme acúmulo de informações, muitas vezes, compiladas aleatoriamente por diferentes fontes o que, de certa forma, pode dificultar ou impedir a consulta rápida e precisa destes dados. O gerenciamento de todas as informações geradas por estudos de genomas mitocondriais em uma estrutura organizada, na forma de amplos bancos de dados com interfaces interativas, representa uma alternativa extremamente eficiente e sofisticada para promover o acesso irrestrito e completo aos diferentes aspectos relacionados ao estudo do genoma mitocondrial.

Nos últimos cinco anos, bancos de dados especializados na compilação de seqüências mitocondriais, foram criados com objetivos, conteúdos e funcionalidades específicos. De maneira geral, um dos principais objetivos destes bancos refere-se à atualização e avaliação das seqüências nucleotídicas depositadas publicamente, além de gerenciar a integração de dados provenientes de outros bancos ou fontes independentes (Shimko, *et al.*, 2001). É importante notar que cada banco de dados apresenta interfaces com diferentes níveis de complexidade, dependendo do seu enfoque principal. Uma das fontes mais completas e precisas de informação sobre seqüências mitocondriais é o GOBASE (http://megasun.bch.umontreal.ca/gobase/), disponível desde 1996, cujo principal objetivo visa à capacidade de resolver questões relacionadas com estudos de genômica comparativa (Shimko, *et al.*, 2001). Recentemente, o banco de dados GenBank (Wheeler *et al.*, 2001), administrado pelo NCBI, tornou disponível uma interface para acesso a informações relativas a genomas organelares (http://www.ncbi.nlm.nih.gov:80/PMGifs/Genomes/organelles.html). Além destas opções, o Mitochondrial Gene Arrangement Source Guide - MGASG

(http://www.jgi.doe.gov/programs/comparative/MGA_Source_Guide.html), organizado pelo Dr. J. Boore - JGI, armazena genomas completos com ênfase em rearranjos estruturais e análises comparativas. Outros bancos de dados, de caráter mais específico no que diz respeito às informações disponibilizadas, também representam uma importante fonte de informação, como por exemplo, o MITOMAP, uma compilação de seqüências mitocondriais relacionadas a polimorfismos humanos e variações do genoma mitocondrial associadas a doenças (Kogelnik *et al.*, 1998) (http://www.gen.emory.edu/mitomap.html); a base de dados MITBase (http://www3.ebi.ac.uk/Research/Mitbase/mitbase.pl), uma interface para acesso de dados sobre a diversidade de genomas mitocondriais (Attimonelli *et al.*, 2000); o AMmtBD (Lanave *et al.*, 2000), onde estão depositados alinhamentos múltiplos de seqüências e genomas mitocondriais de vertebrados (http://bio-www.ba.cnr.it:8000/srsbin/cgi-bin/wgetz?-page+LibInfo+-lib+AMMTDB); o MITOP, outro projeto voltado predominantemente para a análise de seqüências mitocondriais humanas e doenças associadas (http://www.mips.biochem.mpg.de/proj/medgen/mitop/); além de um banco de dados que compila seqüências nucleares humanas codificadoras de proteínas mitocondriais envolvidas na biogênese e função da mitocôndria (http://www-lecb.ncifcrf.gov/mitoDat/) e uma compilação de seqüências mitocondriais da região controle de humanos e primatas (http://db.eva.mpg.de/hvrbase/).

Análise do genoma mitocondrial em moscas causadoras de miíases

Considerando-se espécies de importância econômica e veterinária, a análise de marcadores moleculares no genoma mitocondrial tem contribuído para esclarecer os mecanismos evolutivos envolvidos na manutenção da variabilidade genética, e tem se mostrado extremamente útil para avaliar a estrutura de populações em Calliphoridae e Oestridae (Roehrdanz & Johnson, 1988; Roehrdanz, 1989; Azeredo-Espin, 1993; Azeredo-Espin & Madeira, 1996; Infante & Azeredo-Espin, 1995; Valle & Azeredo-Espin, 1995; Geurgas *et al.*, 2000). A caracterização de seqüências mitocondriais permite também estabelecer padrões de identidade molecular específicos e inferências filogenéticas intra e interespecíficas através da análise de regiões informativas do DNAmt destas espécies (Goldenthal *et al.*, 1991; Sperling *et al.*, 1994; Stevens & Wall, 1997; Smith *et al.*, 1996; Taylor *et al.*, 1996; Wells & Sperling, 1999; Wells & Sperling, 2001; Lessinger & Azeredo-Espin, 2000; Lessinger *et al.*, 2000; Litjens *et al.*, 2001; Wallman & Donnellan, 2001, Hall *et al*, in press; Junqueira *et al.*, in press). Existem apenas dez genomas completos descritos em Diptera (Tabela 1), destes, apenas duas espécies representam o grupo Calyptratae, *Cochliomyia hominivorax* (Lessinger *et al.*, 2000) e *Chrysomya putoria* (Junqueira *et al.*, em preparação). Estudos recentes, envolvendo o

estabelecimento de relações filogenéticas através da análise de genomas mitocondriais, apresentam divergências quanto à posição de *Cochliomyia* em relação às espécies de Acalyptratae, *Drosophila* e *Ceratitis* (Nardi *et al.*, 2001; Hwang *et al.*, 2001; Lessinger *et al.*, 2000).

Perspectivas e tendências

A plasticidade do DNAmt de artrópodes tem sido discutida, conforme novos genomas mitocondriais são descritos, contrastando com a idéia de conservação estrutural do genoma animal. Estes estudos apresentam uma molécula evolutivamente dinâmica e flexível em relação à organização do conteúdo gênico, inclusive contendo duplicações e rearranjos envolvendo regiões codificadoras e/ou seqüências da região controle de replicação (Roerdanz & Black, 1998; Campbell & Barker, 1999; Campbell *et al.*, 1999; Dotson e Beard, 2001; Shao *et al.*, 2001; Nardi *et al.*, 2001; Lessinger *et al.*, em preparação). A caracterização de genomas mitocondriais em grupos ainda pouco representados nos bancos de dados deve contribuir para ampliar o atual quadro de diversidade estrutural deste genoma.

No plano funcional, apesar da escassez de evidências experimentais (Boore *et al.*, 1999; Beagley *et al.*, 1999; Ojala *et al.*, 1980), a análise de transcritos mitocondriais deve esclarecer os mecanismos responsáveis pela edição e o processamento de genes mitocondriais (Lavrov *et al.*, 2000), auxiliando na compreensão dos padrões de expressão gênica da mitocôndria e a interação com componentes nucleares. Recentemente, a questão da reorganização do conteúdo gênico em algumas espécies foi associada à ocorrência de recombinação intramolecular (Dowton & Campbell, 2001; Shao *et al.*, 2001), um tema ainda polêmico devido a poucas evidências experimentais.

A determinação do genoma mitocondrial completo para espécies extintas de aves, a partir da análise de material preservado em museu (Cooper *et al.*, 2001; Haddrath & Baker, 2001), e o sequenciamento do DNAmt de espécies raras ou ameaçadas de extinção conservadas em zoológico, ou "frozen-zoo" (Ryder *et al.*, 2000), representam algumas abordagens de vanguarda da genômica mitocondrial e comparativa.

APRESENTAÇÃO DOS TRABALHOS

A seguir serão apresentados os quatro artigos científicos que constituem esta tese, segundo a ordem cronológica das publicações. Estes artigos tratam da caracterização de seqüências mitocondriais em espécies de moscas causadoras de miíases, especialmente abordando aspectos relacionados com a evolução da organização estrutural do genoma mitocondrial e à obtenção de marcadores espécies-específicos. O estudo do genoma mitocondrial nestas espécies envolveu pesquisas de natureza básica e aplicada. Os três primeiros artigos apresentados estão publicados e o quarto manuscrito será submetido à publicação.

Artigo 1 – ***Lessinger, A. C.*** & Azeredo-Espin, A. M. L. (2000). **Evolution and structural organization of the mitochondrial DNA control region of myiasis-causing flies**. *Medical and Veterinary Entomology*, **14**: 71 – 80.

Resumo

Este estudo descreve a caracterização molecular da região controle (denominada região rica em A+T em insetos) do DNAmt de cinco espécies de dípteros causadores de miíases: *Cochliomyia hominivorax* Coquerel, *Cochliomyia macellaria* Fabricius, *Chrysomya megacephala* Fabricius, *Lucilia eximia* Wiedemann (Diptera: Calliphoridae) e *Dermatobia hominis* (Diptera: Oestridae). A região controle destas espécies varia em tamanho de 1000 a 1600 pb. Dois domínios estruturais com padrões evolutivos específicos foram identificados. Sendo, portanto (1) blocos de seqüência conservada contendo elementos de seqüência primária, incluindo séries dinucleotídicas pirimidina-purina e longos trechos poli-T, localizados na extremidade 5' adjacente ao gene $tRNA^{Ile}$ e (2) um domínio hiper-variável na extremidade 3' caracterizado por elevada divergência nucleotídica e variações em tamanho. Uma elevada freqüência de substituições nucleotídicas do tipo transversões A↔T indica pressão de mutação direcional. A utilidade filogenética da região controle de insetos é discutida

Palavras-chave: Calliphoridae, Oestridae, região rica em A+T, DNA mitocondrial, evolução molecular, região controle do DNAmt, Brasil.

Artigo 2 - ***Lessinger A. C.***; Martins Junqueira A. C.; Lemos T. A., Kemper E. L., da Silva F. R., Vettore A. L., Arruda P., Azeredo-Espin A. M. (2001). **The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae)**. *Insect Molecular Biology*, **9**: 521 - 529.

Resumo

A seqüência completa do genoma mitocondrial da mosca-da-bicheira, *Cochliomyia hominivorax*, foi determinada. Este genoma possui 16.022 pb e corresponde a um genoma mitocondrial típico de Brachycera. Um códon de iniciação Serina para o gene COI e códons incompletos de terminação para os genes COII, NADH5 e NADH4 foram descritos. A composição nucleotídica do DNAmt de *C. hominivorax* é 77% rica em A+T, refletindo a predominância de códons ricos em A+T nos genes codificadores de proteínas. Um "codon-usage" não ótimo é verificado nos genes mitocondriais de *C. hominivorax*. Análises filogenéticas distribuem as espécies Acalyptratae como grupo monofilético e relacionam *C. hominivorax* (Calyptratae) e Acalyptratae em um típico agrupamento Brachycera. A identificação de sítios de restrição diagnósticos no genoma mitocondrial seqüenciado e a correlação com análises anteriores de RFLP são discutidas.

Palavras-chave: *Cochliomyia hominivorax*, mosca-da-bicheira, miíases, DNAmt, Calyptratae

Artigo 3 - Litjens, P., ***Lessinger, A. C.*** & Azeredo-Espin, A. M. L. (2001). **Caracterization of *Cochliomyia hominivorax* and *Cochliomyia macellaria* by PCR-RFLP of mitochondrial DNA.** *Medical and Veterinary Entomology*, **15**, 183 – 188.

Resumo

A mosca-da-bicheira, *Cochliomyia hominivorax* (Coquerel) (Diptera: Calliphoridae), é uma das mais importantes pragas da pecuária na Região Neotropical, enquanto *Cochliomyia macellaria* (Fabricius) (Diptera: Calliphoridae), uma espécie causadora de miíase secundária, possui grande importância médica e sanitária devido ao seu papel na disseminação de patógenos. Estas espécies compartilham semelhanças morfológicas e ambas podem ocorrer em uma mesma miíase, mas em diferentes estágios de desenvolvimento. Neste trabalho, é investigada a utilidade do PCR-RFLP (reação de polimerização em cadeia - polimorfismos de tamanho de fragmentos de restrição) do DNAmt para a identificação precisa de *C. hominivorax* e *C. macellaria*. Duas regiões específicas do DNAmt foram amplificadas: 870 pb da subunidade I do complexo Citocromo Oxidase e 2100 pb da região AT/12S de amostras de *C. hominivorax* e *C. macellaria* de diferentes regiões do Brasil. Resultados consistentes de identificação espécie-específica através de PCR-RFLP foram obtidos para ambas as regiões utilizando as enzimas de restrição *Dra* I e *Ssp* I. Estes resultados confirmam a conservação de sítios de restrição diagnósticos registrados anteriormente e demonstram a utilidade das seqüências da região controle como marcador eficiente para a identificação de populações Brasileiras de *Cochliomyia* através de PCR-RFLP. A ocorrência de padrões de polimorfismo intraespecíficos é discutida baseando-se nas freqüências e possíveis conflitos para a identificação das espécies. A estratégia de PCR-RFLP fornece um método potencialmente informativo para a identificação de amostras provenientes de regiões onde estas espécies são monitoradas.

Palavras-chave: *Cochliomyia hominivorax*, *Cochliomyia macellaria*, identificação molecular, marcadores moleculares, DNAmt, PCR-RFLP, bicheira, Brasil.

Artigo 4 - *Lessinger, A. C.*, Junqueira, A. C. M., Conte, F. F. & Azeredo-Espin, A. M. L. (*em preparação*). **The occurrence of a duplication in the mtDNA control region of *Chrysomya* species (Diptera: Calliphoridae) involving two tRNAs.**

Resumo

A região controle (RC) do DNAmt das moscas varejeiras *C. albiceps*, *C. megacephala* e *C. putoria* (Calliphoridae) foi caracterizada. O aspecto mais incomum encontrado na região controle de *Chrysomya* foi a presença de duas seqüências adicionais de tRNAs, correspondendo aos genes $tRNA^{Ile}$ e $tRNA^{Gln}$, e uma repetição de 19 pb da região controle imediatamente adjacente. O $tRNA^{Gln}$ parcialmente duplicado deve representar um pseudogene, uma vez que a maior parte da seqüência de um $tRNA^{Gln}$ típico de inseto está ausente e estas seqüências apresentam um elevado grau de divergência nucleotídica nas comparações entre as espécies de *Chrysomya*. É interessante notar a elevada divergência especificamente associada ao $tRNA^{Gln}$ duplicado em *C. putoria*, resultando na formação de um elemento de seqüência contendo um padrão de simetria interno. Por outro lado, a cópia adicional do gene de $tRNA^{Ile}$ parece ter preservado sua identidade nucleotídica desde o evento de duplicação e deve representar uma cópia funcional deste gene. A ocorrência de seleção diferencial nos genes duplicados nestas espécies pode estar refletindo a conservação do $tRNA^{Ile}$ extra e a degeneração das seqüências do $tRNA^{Gln}$. A caracterização de elementos de seqüência, associados às extremidades da repetição, pode auxiliar na compreensão dos mecanismos responsáveis pela duplicação. O potencial de formação de estruturas secundárias foi investigado nas extremidades da duplicação, uma vez que existe uma associação freqüente destas estruturas com eventos de duplicação gênica e rearranjos no DNAmt animal. Padrões de evolução específicos estão associados aos diferentes elementos identificados. Estes resultados demonstram e confirmam a plasticidade da molécula de DNAmt em *Chrysomya*, especialmente em relação aos genes de tRNA e seqüências adjacentes à região controle, contribuindo para a investigação da evolução do genoma mitocondrial de insetos. Além disso, a caracterização da RC de *Chrysomya* favorece o acesso às seqüências variáveis do genoma mitocondrial, as quais podem auxiliar na identificação espécie-específica de moscas de importância forense, assim como fornecer dados informativos para análises filogenéticas de espécies próximas e potenciais marcadores para genética de populações.

Palavras-chave: *Chrysomya*, mtDNA, região controle, duplicação, tRNA, rearranjo, evolução molecular.

CAPÍTULO 1

Artigo 1 – ***Lessinger, A. C.*** & Azeredo-Espin, A. M. L. (2000). **Evolution and structural organization of the mitochondrial DNA control region of myiasis-causing flies**. *Medical and Veterinary Entomology*, **14**: 71 – 80.

Medical and Veterinary Entomology (2000) **14**, 71–80

Evolution and structural organisation of mitochondrial DNA control region of myiasis-causing flies

A. C. LESSINGER and A. M. L. AZEREDO-ESPIN
Laboratório de Genética Animal, Universidade Estadual de Campinas, São Paulo, Brazil

Abstract. This study reports the molecular characterization of the mtDNA control region (called the A+T-rich region in insects) of five dipteran species which cause myiasis: *Cochliomyia hominivorax* Coquerel, *Cochliomyia macellaria* Fabricius, *Chrysomya megacephala* Fabricius, *Lucilia eximia* Wiedemann (Diptera: Calliphoridae) and *Dermatobia hominis* Linnaeus Jr (Diptera: Oestridae). The control region in these species varies in length from 1000 to 1600 bp. Two structural domains with specific evolutionary patterns were identified. These were (1) conserved sequence blocks containing primary sequence motifs, including dinucleotide pyrimidine-purine series and long T-stretches, located at the 5′ end adjacent to the $tRNA^{Ile}$ gene and (2) a hypervariable domain at the 3′ end characterized by increased nucleotide divergence and size variation. A high frequency of A↔T transversions at nucleotide substitution level indicated directional mutation pressure. The phylogenetic usefulness of the insect control region is discussed.

Key words. Calliphoridae, Oestridae, A+T-rich region, mitochondrial DNA, molecular evolution, mtDNA control region, myiasis, Brazil.

Introduction

The control region is the major non-coding region of the animal mitochondrial genome and is believed to be involved in the regulation of transcription and control of mtDNA replication (Shadel & Clayton, 1997). The control region (or D-loop) of vertebrate mtDNA has been well characterized at the nucleotide level, resulting in the identification of conserved sequence blocks (CSB) and hypervariable regions (Shadel & Clayton, 1993).

Among insect species, the best characterized control region is that of *Drosophila*. Indeed, *Drosophila* species could be divided into two distinct categories, based on the structure of their control region. The total length of the control region varies between *Drosophila* species and the primary sequence elements differ in their organization, particularly with regard to absence or presence of repetitive motifs (Clary & Wolstenholme, 1985, 1987; Monnerot *et al.*, 1990; Monforte *et al.*, 1993; Inohira *et al.*, 1997). No D-loop structure or any of the conserved sequences identified in the control region of vertebrate mtDNAs have been found in the *Drosophila* A+T-rich region.

Variations in the organization of conserved sequence elements in mtDNA control regions suggest that these sequences follow distinct evolutionary patterns in different insect species or taxa (reviewed in Zhang & Hewitt, 1997). Because conserved sequences may be associated with specific domains or distributed along the entire mtDNA control region, this region may contain phylogenetic information which varies among insect species (Zhang *et al.*, 1995).

Zhang & Hewitt (1997) suggested that the control region can be a suitable marker for studying genetic polymorphism and population structure in species containing a single copy of a highly variable domain. Furthermore, for species in which the control region has a highly conserved domain of reasonable size (>350 bp), phylogenetic studies may be done using the conserved domain as a marker. These observations emphasize the importance of studies that examine the patterns of structural organization in mtDNA control regions as a mean of providing information on the overall evolution of this region in a group as widely diversified as the insects.

Although numerous insect mtDNA sequences have been used as efficient molecular markers of specific taxonomic levels, more information is still required on the structure and

Correspondence: Ana C. Lessinger, Laboratório de Genética Animal, Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), C.P. Box 6109, Campinas, São Paulo, Brazil, CEP 13083-970. E-mail: lessinge@unicamp.br

function of this region in order to establish conclusively its phylogenetic usefulness in insect studies. The polymerase chain reaction (PCR) and sequencing strategies require additional optimization to provide efficient amplification results. Specific proprieties of the mtDNA control region that may increase technical difficulties include the extremely high A + T content, the presence of potential secondary structures, the existence of potential regulatory domains and the occurrence of long thymidylate arrays (poly T), which may decrease the ability of DNA polymerase to produce elongation (Robbins *et al.*, 1996). Methodological difficulties in sequencing the insect mtDNA control region include a lack of information on useful PCR primers, high levels of primary sequence divergence among related insect species and little or no information on the total size of the control region and on potentially useful restriction sites (Simon *et al.*, 1994; Zhang & Hewitt, 1997).

The aim of this paper was to examine the entire A + T-rich region at the nucleotide level of Brachycera taxa other than the Drosophilidae, including calypterate species from the families Calliphoridae and Oestridae. The control regions of *Cochliomyia hominivorax* Coquerel, *Cochliomyia macellaria* Fabricius, *Chrysomya megacephala* Fabricius, *Lucilia eximia* Weidemann and *Dermatobia hominis* Linnaeus Jr. were sequenced and the data then used to determine structural organization of conserved and hypervariable sequences, and to evaluate the phylogenetic usefulness of information provided by the mtDNA control region of these species.

Material and Methods

Flies and materials

The specimens of *C. hominivorax*, *C. macellaria*, *C. megacephala*, *L. eximia* and *D. hominis* used in this study were samples from field collections and laboratory lineages from Brazil. Collected larvae were reared to pupae and then stored at −70°C (Azeredo-Espin, 1987; Valle & Azeredo-Espin, 1995; Yotoko, 1998). *Lucilia eximia* pupae were from Botucatu and *C. megacephala* pupae were from Adamantina, both in São Paulo State (SP). One individual from each of these species was used for sequence analysis. For the *Cochliomyia* and *Dermatobia* species, two individuals from different populations were analysed. *Cochliomyia hominivorax* pupae were from Alfenas (Minas Gerais State, MG) and Manaus (Amazonia State, AM); *C. macellaria* pupae were from Rio de Janeiro (Rio de Janeiro State, RJ) and Porto Urucu (AM); and *D. hominis* pupae were from Alfenas (MG) and Ponta Grossa (Paraná State, PR). Individuals derived from these populations are referred as 1 or 2 in all analyses below. A 1986 dried-pinned adult *L. eximia* (Botucatu, SP) from an entomological collection was also used for amplification of the control region.

Genome DNA extraction

Total individual DNA was extracted from each species as described in Azeredo-Espin (1993) and Infante & Azeredo-Espin (1995). Pinned adult flies were extracted using guanidine isothiocyanate (GIBCO/BRL) to improve the nucleic acid recovery from dried preserved samples (A. C. M. Junqueira, personal communication).

Nested PCR

Amplification of the control region required two separate PCR reactions. This strategy increased product resolution and guaranteed sufficient amounts of specific DNA for cloning and sequencing. The first amplification reaction was performed using primers anchored on $tRNA^{Met}$ sequences (TM-N-193) and the srRNA 12S gene (SR-J-14233). This large PCR product was diluted 1 : 50 and used as a template for the second round of PCR. A second set of primers from the $tRNA^{Ile}$ gene (TI-N-24) and an inner srRNA 12S sequence (SR-J-14776) were used to allow specific amplification of a short A + T-rich fragment in the control region. Table 1 describes the primers used in this work, with additional information from previous studies.

The PCR reactions were done using oligonucleotide primers based on *D. yakuba* mtDNA sequences (Clary & Wolstenholme, 1985; Simon *et al.*, 1994). The reactions were run in a DNA thermal cycler (Perkin Elmer, Foster City, CA,

Table 1. Primers used for amplification of the control region by PCR. Primer identification is that reported by Simon *et al.* (1994) and was also used for SR-J-14776. The oligonucleotide sequences are in the 5′→ 3′ direction.

	Identification	Previously reported for	References
External amplification			
Primers (long A + T-rich reaction)			
5′-TGGGGTATGAACCCAGTAGC-3′	TM-N-193	Lepidoptera	Taylor *et al.* (1993)
5′-AAGAGCGACGGGCGATGTGT-3′	SR-J-14233	*Drosophila virilis*	Simon *et al.* (1991)
Internal amplification			
Prmers (short A + T − rich reaction)			
5′-ATTTACCCTATCAAGGTAA-3′	TI-N-24	*Drosophila virilis*	Simon *et al.* (1991)
5′-GCTGGCACGAATTTTGTC-3′	SR-J-14776	first reported in this work	M. C. Arias (pers. comm.)

U.S.A.), in a volume of 50 µl containing 1.5 mM $MgCl_2$, 200 µM of each dNTP, 0.5 mM of each primer, approximately 1 ng of template DNA and 2.5 units of *Taq* DNA polymerase (Perkin Elmer). The reaction consisted of 30 s at 94°C, 1 min at 58°C (initial PCR) or 56°C (nested PCR) and 2 min at 60°C for a total of 30 cycles, followed by a final cycle in which the elongation step at 60°C was extended to 10 min. The elongation step at 60°C was necessary to provide a reliable control region amplification for these species.

Cloning and sequencing

The amplified DNA product was purified from the PCR mixture by dialysis for 20 min using a VM 0.05 µM (Millipore, Bedford, MA, U.S.A.) filter with T.E. buffer, pH 7.0. An aliquot of the PCR product was used in the cloning reaction. The amplified control region was cloned into the pMOSBlue T-vector system (Amersham, Bucks, U.K.) according to the manufacturer's specifications. The cloning step was used prior to sequencing in order to avoid the presence of amplification by-products that may contribute as a source of interference in sequencing reactions. Two to four clones from independent PCR assays were sequenced for each species in both strands. After preparation of plasmid DNA (Sambook *et al.*, 1989), the samples were dried and sequenced automatically.

Sequence analysis

The nucleotide sequences of the five species were aligned using the Clustal W software (Thompson *et al.*, 1994) set to default parameters with manual adjustments where necessary. BLASTN version 2.0.8 (Altschul *et al.*, 1997) was used to identify the control region content using conserved $tRNA^{Ile}$ and srRNA sequences to search databases for similar sequences. Aligned sequences were analysed using the MEGA software package (Kumar *et al.*, 1993), the Tamura distance method (Tamura, 1992) and the neighbour-joining method (Saitou & Nei, 1987). MAST (motif alignment and search tool) software version 2.2 was used to provide a statistically valid approach for searching for control region sequence homology (Bailey & Gribskov, 1998). The determination of conserved sequence blocks (CSBs) was based on the identification of a minimum of four consecutive identical sites on the aligned sequences and a minimum length of 10 bp until another series of four divergent sites was reached. The search for energetically stable secondary structures for the control region sequences was done using the software PCfold (Zuker, 1989) and MOLECULE (Lapalme *et al.*, 1982).

GenBank accession numbers

The nucleotide sequences reported in this paper have the following GenBank accession numbers: AF151382, AF151383, AF151384, AF151385, AF151386, AF151387, AF151388 and AF151389.

Results

Access to the mtDNA control region

The nested PCR strategy used here provided sufficient amplified DNA for further cloning and sequencing of the mtDNA control region. The intermediate cloning step was fundamental for improving the sequencing results. However, this procedure may provide inaccurate information as a result of incorrectly incorporated nucleotides by polymerase and may itself be a potential source of sequence variation. The sequencing of multiple clones from independent PCR assays was used to guarantee the sequence identification. The amplification reactions provided diagnostic markers that varied from 1200 to 1800 bp for each species, including sequences from $tRNA^{Ile}$ and rRNA12S genes. The mtDNA control region was recovered efficiently from a dried *Lucilia* collected in 1986, indicating that it is possible to access the genetic information in preserved Calliphoridae species.

Structural organization

The sequenced clones included control region sequences plus 200 bp from the srRNA 12S flanking region on the 3′ end and an additional 42 bp from flanking $tRNA^{Ile}$ gene on the 5′ end of the control region in all the species. A BLASTN (Altschul *et al.*, 1997) search of sequence databases showed that the sequence of these RNA genes was highly conserved, a finding which helped in the identification and alignment of the control region. The sizes of control regions were 1178–1179 bp in *C. hominivorax* ($n=2$); 1222–1226 bp in *C. macellaria* ($n=2$); 986 bp in *C. megacephala* ($n=1$); 1130 bp in *L. eximia* ($n=1$) (Calliphoridae) and 1570–1569 bp in *D. hominis* ($n=2$) (Oestridae).

Alignment of the control region sequences revealed specific patterns of structural organization. Two main subregions (domains A and B) were responsible for the distinct rates of nucleotide substitution. Domain A, which varied in size from 586 to 783 bp, was highly conserved, whereas the B domain sequences were highly divergent between species, varying from 393 to 794 bp in size. To increase the consistency of the data, further analysis was done by considering each domain independently. The A domain sequences aligned consistently in all species, even with the inclusion of *D. hominis* data, which increased the general sequence divergence. Eight conserved sequence blocks (CSB) were distributed near the 5′ end of the aligned A domains (Fig. 1).

Further analysis to identify homologous sequence motifs was done using the unaligned A domain sequences and MAST software (Bailey & Gribskov, 1998). Although this procedure was not as effective in recovering the previously reported CSBs, seven conserved motifs were observed along the control region, with a high degree of overlapping with the CSBs (Table 2). The MAST motifs not shared by all the species were not included in the final diagram (Fig. 2).

Arrangement of the seven motifs on the aligned sequences revealed a significant association with the previously identified

```
                1                                                          60
D.hominis1      -------TTT TT-------- TTTTTTTTTT TTTTATTTTT ATTGTTTT-- ---TTTTTAG
D.hominis2      -------... ..........  ..........  ..........  ..........  ..........
C.macellaria1   AAAAGTAGAC ..CCCATG..  -------...  ....T.....  T..T..A...  ......A..C
C.macellaria2   AAAAGTAGAC ..CCCA....  -------.G.  G...T.....  T..T..A...  ......A..C
C.hominivorax1  AAAAGTAGGA ..TCCATACA  A...---...  ....T.....  T..T..A.GT  ATA...A..C
C.hominivorax2  AAAAGTAGGA ..TCCATACA  A.........  ....T.....  T..T..A.GT  ATA...A..C
L.eximia1       AAAAGTAGG. ..TTCA....  -------.G.  ....T.....  T..T......  .TC...A..C
C.megacephala1  AAGAGTAAGA ..CCCATG..  -------...  ....T.....  T..T..G.TT  TAT...A...
                                                  I

                61                                                         120
D.hominis1      TGAAA-TTTA GAAATATTAT AAATTGATTA ATTTATTTTT AATAATT-TT TAATTAGTAT
D.hominis2      ..........  ..........  ..........  ..........  ..........  ..........
C.macellaria1   .T.TT.....  T...A..A.-  -....A....  T...G.A...  T...T....A  ATT...A..A
C.macellaria2   .T.TT.....  T...A..A.-  -....A....  T...G.A...  T...T....A  ATT...A..A
C.hominivorax1  .ATTT.....  T...A..A.-  -...AT.C..  ...A..A...  T..TT....A  ATT...A..A
C.hominivorax2  .ATTT.....  T...A..AG-  -T..AC.C..  ...A..A...  T..TT....A  ATT...A..A
L.eximia1       .A.......T  T...A.AA.-  --------G.  G...T.....  T..T.....A  .....--..-
C.megacephala1  .A.TTT...T  .TG..T.AT-  ---.AA..A.  .A.A.CA..A  T.A....TAA  AT....A..A

                121                                                        180
D.hominis1      TTT--AATAA AAAAAGATA-  TAATTTTTTA ATTAAAATTT ATGGTTAAAG TTAATTAAAA
D.hominis2      ..........  ..........  .........-  ..........  ..........  ..........
C.macellaria1   ...TA....G  T....A.A..  ...G..G.--  .-----T...  ..TA..----  ...T...G.-
C.macellaria2   ...TA....G  T....A.A.A  ...G..G.--  .-----T...  ..TA..----  ...T...G.-
C.hominivorax1  ...AA.T..G  T..T.A...T  .T.A.-----  .------...  ..TAAA..--  ...T...G.-
C.hominivorax2  ...AA.T..G  T..T.A...T  .T.A.-----  .------...  ..TAAA..--  ...T..GG.-
L.eximia1       ....AC....  T....TT..G  .....A.AAT  .------...  ..AAAA..--  ..TT...T.-
C.megacephala1  ..ATA....G  T...--T..G  .TTA.A---.  .-------..G  .AAAAA..-T  AA.T...G.-

                181                                                        240
D.hominis1      TGTAAATAAT TAAAGTTTGC ATATATATAT ATATATATA- -TTTATATAT ATTAATATAC
D.hominis2      ..........  ..........  ..........  ..........  ..........  ..........
C.macellaria1   ------....  ....A.....  ..........  ..........  A.........  ...--...T.
C.macellaria2   ------....  ....A.....  ..........  ..........  A.........  ...--...T.
C.hominivorax1  ------....  ...TTA....  ..........  ..........  A.........  ....T...T.
C.hominivorax2  ------....  ...TTA....  ..........  ..........  A.........  ....T...T.
L.eximia1       ..........  ...TT.....  ..........  .........T  A.........  ...--...T.
C.megacephala1  ------....  ....AA....  ..........  ..........  A.........  ...--...T.
                           ◄──────────────────────────────────────────────►
                                                 II

                241                                                        300
D.hominis1      ATTCATCTAT ATATATATTA AATTATGTAT ATAAATTTAT ATTTTCATTT AATAAATAAA
D.hominis2      ..........  ..........  ..........  .C........  ..........  ..........
C.macellaria1   ..AT..T.C.  ..T.......  ..A.....T-  -C...--..G  ..A..A....  .....T...-
C.macellaria2   ..AT..T.C.  ..T.......  ..A.......  -C...--..G  ..A..A....  .....T...-
C.hominivorax1  ..AT..T...  ..T.......  ..A....A.-  -.....--.TA  ..A..T....  .....T...-
C.hominivorax2  ..AT..T...  ..T.......  ..G....A.-  -.....--.TA  ..A..T....  .....T...-
L.eximia1       ..AT..T...  ..........  ..A....A.-  -.....--...  ..AAAT..A.  .....T...-
C.megacephala1  ..AT..T...  .-........  .......A..  ....T----G  T.A.AA.A..  .....T...-
                ◄──────────────────────────────►                ◄──────────►
                                                                    III
```

Fig. 1. Alignment of the nucleotide sequences of the A domain in the control region of Calliphoridae and Oestridae species. Differences from the sequence for *D. hominis* 1 are indicated for each species. Conserved sequence blocks are underlined and identified (I–VIII). Conserved Calliphoridae-specific sequences are indicated by an asterisk. A dot indicates site identity with *D. hominis* 1. Gaps are indicated by dashes. A *D. hominis*-specific sequence indicative of size variation is delimited by vertical arrows. The numbers 1 and 2 indicate individuals from distinct populations.

```
                301                                                       360
D.hominis1      TAATTTATAT ATATATAAAT TATTTATATC ATTAA--ATC TTTAAAGTAA AATATAAATA
D.hominis2      .......... .......... .......... .......... .......... ..........
C.macellaria1   ...A....T. .....-.T.. ...AA...AA TG..TCT... .A......TT ....ATTTC.
C.macellaria2   ...A....T. .....-.T.. ...AA...AA TG..TCT... .A......TT ....ATTTC.
C.hominivorax1  ...A.C..T. .....-.... ...AA...AA T...TCT... .A......TT ...GATTTC.
C.hominivorax2  ...A.C..T. .....-.... ...AA...AA T...TCT... .A......TT ...GATTTC.
L.eximia1       ...A....T. .....ATT.A ...AA...A. TA.CCC...A CA.T..-... .T...TTT.T
C.megacephala1  ...A....T. .....-.... ...AA....A ...C.C...A CA.T.T-... .T...TTT.T
                <----------------------------------------->

                361                                                       420
D.hominis1      TATAACATTT A--GGTGG-- --AGAAATAT -ATATAAATA ATAATTAATA AATATTTAAT
D.hominis2      .......... .......... .......... .......... .......... ..........
C.macellaria1   ..ATTG.... ...AT.TCCA CT.T...C.. T....CT... .A...AGT.. ..G.....G.
C.macellaria2   ..ATTG.... ...AT.TCCA CT.T...C.. T....CT... .A...AGT.. ..G.....G.
C.hominivorax1  ..ATTG.... ...AT.TCCA CT.T....T. T...G..... .A...A.T.. ..G.....G.
C.hominivorax2  ..ATTG.... ...AT.TCCA CT.T....T. T...G..... .A...A.T.. ..GG....G.
L.eximia1       .TA.TA.A.C ...AT.TCAA C..T..TA.. T.G....... T.T....... ..G.......
C.megacephala1  ..A....A.C ...AT.TCTA C..T..TA.. T.G......C ..T....... ..........
                           ** **** *                                    <-------->
                                                                          IV

                421                                                       480
D.hominis1      TTAAA-ATAT TCTTAATTGA TCCCCGAAAT TTTTTTTTGA GTCTGGATTA ATAGATAAAT
D.hominis2      .......... .......... AT........ .......... .......... ..........
C.macellaria1   AAC.TG.... A.....C... ATTT-..... ........T. -...AA.... .......T..
C.macellaria2   AAC.TG.... A.....C... ATTT-..... ........T. -...AA.... .......T..
C.hominivorax1  AAC.TG.... C.....C..G ATTT-..... ........T. -...AA.... ........G.
C.hominivorax2  AAC.TG.... C.....C..G ATTT-..... ........-. -...AA.... ........G.
L.eximia1       .AG-TG.... C......... ATTT--.... .......... -...AA.... .......TC.
C.megacephala1  .A.-TG.... ......C..G ATTT--.... ........T. C...AA..... ......T..
                <--------->           ****  <-------------------------------->
                    V

                481                                                       540
D.hominis1      ATTGATTAAT TAAACATTTA TATATTAATA TTTATCTATT AATCCTTATT TGGTATATAG
D.hominis2      .......... ....T..... .......... .C........ .......... ..........
C.macellaria1   ...A...... ....T..... .......... AA........ ...T...... -.......G.
C.macellaria2   ...A...... ....T..... .......... AA........ ...T...... -.........
C.hominivorax1  .......... ....T..... .......... .A........ ......-... -.........
C.hominivorax2  .......... ....T..... .......... .A........ ......-... -.........
L.eximia1       ...A...... ....T..... .......... GA........ ....TA.... -.........
C.megacephala1  ...A...... ....T..... .......... GA........ ...TT.-... ..........
                <------------------------------------------------------------>
                                                  VI

                541                                                       600
D.hominis1      ACCTAAAAAA AATTTCTGTA ACGGTCAATA TAAATTTAT- --GGTATAAT TAAAAATTAA
D.hominis2      .......... .......... .......... .......... .......... ..........
C.macellaria1   ...A....T. .....T..C. CGCT...... ....AA.GGA GA......-- ..T.G..G..
C.macellaria2   ...A....T. .....T..C. CGCT...... ....AA.GGA GA......-- ..T.G..G..
C.hominivorax1  ...A....T. .....T..C. CGCT...... ....AA.GGA GA......-- ..T.G..G..
C.hominivorax2  ...A....T. .....T..C. CGCT...... ....AA.GGA GA......-- .GT.G..G..
L.eximia1       ..TA...TTC .GA..T..C. CGCT...... ....A..TGA GA..C...-- ----------
C.megacephala1  ...A....T. .....T..C. CGCT...... ....A..GGA GA......-- ----------
                <---------->          **** <-------------->** **
                                                  VII
```

Fig. 1. *Continued*

```
                601                                                        660
D.hominis1      TCTTTTTTTT GATTTTTTTA TTTTTAAAGG AAATCTAATT ATCTA--TAA --ATTTTCAT
D.hominis2      .......... .......... .......... .......... .......... ..........
C.macellaria1   ATA.A.A... A....AG... -----.T..A ....-.C... ..A..AG... TT....A.T.
C.macellaria2   ATA.A.A... A....AG... -----.T..A ....-.C... ..A..AG... TT....A.T.
C.hominivorax1  .TA.A.A... A....AA... -----.T.-A GG..-.T... ..A..ATC.. TA....AAT.
C.hominivorax2  .TA.A.A... A....AA... -----.T.-A G...-.T... ..A..ATC.. TA....AAT.
L.eximia1       ---------- A....----- -------..A ....G.T-.. ..A.TA.... T.....--T.
C.megacephala1  ---------- --...----- ---------- ....-G.... ..A.....T. T.-....--T.

                661                                                        720
D.hominis1      AAAAAAAATA ATTAAATTAT ATATATAAAT AATTAATTTT AGACGATTAT AATTAAATTG
D.hominis2      .......... .......... .......... .......... .......... .......A..
C.macellaria1   .TT..T.... ....C..--- ---------- ---------- ---------- ----------
C.macellaria2   .TT..T.... ....C..--- ---------- ---------- ---------- ----------
C.hominivorax1  .T--.T---- ---------- ---------- ---------- ---------- ----------
C.hominivorax2  .T.T.T---- ---------- ---------- ---------- ---------- ----------
L.eximia1       -----T---- ---------- ---------- ---------- ---------- ----------
C.megacephala1  .T.T.T:.A. T--------- ---------- ---------- ---------- ----------

                721                                                        780
D.hominis1      ATATCTTTTT TAAAATTAAA AAAAAA-CAT GAAAATTATA TTGAAAAAAA CAAAATTTTT
D.hominis2      .......... .......... ......A... .......... .......... ..........
C.macellaria1   ---------- ---------- ------.--- ---------- ---------- ----------
C.macellaria2   ---------- ---------- ------.--- ---------- ---------- ----------
C.hominivorax1  ---------- ---------- ------.--- ---------- ---------- ----------
C.hominivorax2  ---------- ---------- ------.--- ---------- ---------- ----------
L.eximia1       ---------- ---------- ------.--- ---------- ---------- ----------
C.megacephala1  ---------- ---------- ------.--- ---------- ---------- ----------

                781                                                827
D.hominis1      AATTATTGAT AAATTTTCAT TTAAAAAAAA AAAAAAAAAA AA-----
D.hominis2      .......... .......... .......... .........-- -------
C.macellaria1   ---------- ---------- -......... .......... .------
C.macellaria2   ---------- ---------- -......... .........-- -------
C.hominivorax1  ---------- ---------- -......... .......... ..AAAAA
C.hominivorax2  ---------- ---------- -......... .......... ..AAAA-
L.eximia1       ---------- ---------- -......... .......... ..-----
C.megacephala1  ---------- ---------- -......... .......... .------
                                                  VIII
```

Fig. 1. *Concluded.*

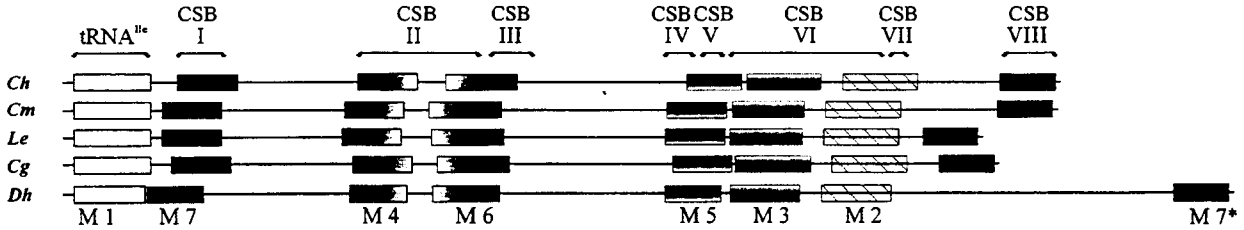


Fig. 2. MAST diagram for A domain sequences for all species investigated. Identified conserved motifs are identified as M1 to M7. Note that M1 corresponds to $tRNA^{Ile}$ gene sequences. * indicates the location of this motif in the complementary strand (poly-A stretch). The nucleotide content of each motif is indicated in Table 2. Overlaps with conserved sequence blocks (CSB) are delimited by arrows. *Ch* = *C. hominivorax*, *Cm* = *C. macellaria*, *Le* = *L. eximia*, *Cg* = *C. megacephala* and *Dh* = *D. hominis*. The numbers 1 and 2 indicate individuals from different populations. The CSB denomination used to indicate conserved elements in these species has no relationship with the vertebrate mtDNA control region nomenclature.

Table 2. Identification of MAST conserved motifs and their association with eight conserved sequence blocks (CSB) from the A domain. The motifs are identified as M1 to M7.

Mast conserved motifs		Corresponds to sequence at
M1	TTA CCC TAT CAA GGT AAC CCT TTT TAT CAG GCA ATT CAT TAA AAG TAG GTT TTC C	$tRNA^{Ile}$ gene
M2	TTG GTA TAT AGA CCA AAA ATA AAT TTT TGC ACG CTT CAA TAT AAA ATT GGA GAG G	CSB VI, VII
M3	TCT AAA TTA ATA GAT ATA TAT TAA TTA ATT AAA TAT TTA TAT ATT AAT ATA TAT C	CSB VI
M4	ATT TAG ATA ATT AAA TTT TGC ATA TAT ATA TAT ATA TAT AAT T	CSB II
M5	TTT AAT TCA TGA TAT TCT TAA CTG AAT TTG AAA TTT TTT TTT	CSB IV, V, VI
M6	CAT ATA TTT ATA TTT ATA TTA AAA TAT GTA TAA ATA TAT ATT TAT TTA ATA ATT A	CSB II, III
M7	TTT TTT TTT TTT TTT TTT TTA	CSB I, VIII

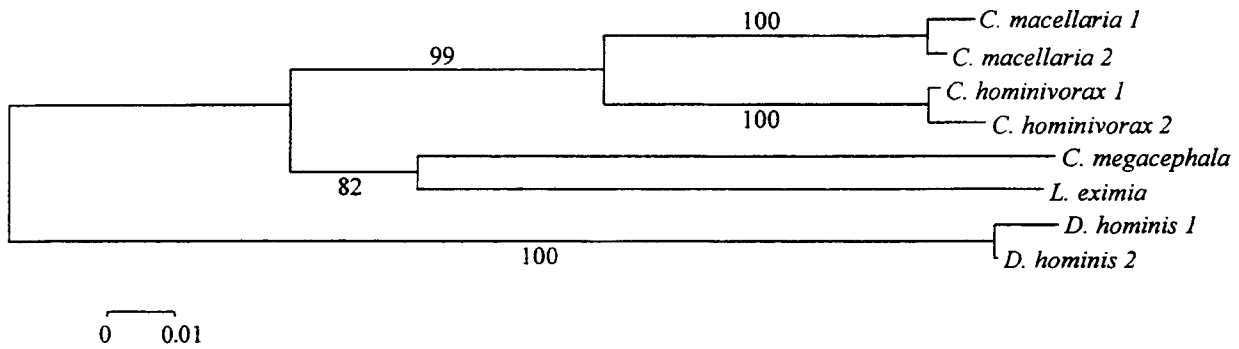


Fig. 3. Neighbour-joining tree obtained with control region A domain sequences using Tamura (1992) distance method. Bootstrap values are indicated. The numbers 1 and 2 indicate individuals from different populations.

CSBs. Motif 1 was used to recover $tRNA^{Ile}$ sequences included in the analysis as an additional parameter. CSB VI was one of the largest conserved elements identified in the alignment and included sequences from motifs 3 and 2. The poly-adenine region of CSB VIII indicates a second T-stretch on the opposite strand, which accurately matched motif 7. Conserved G/C sites were found in CSB II (4), CSB V (2), CSB VI (8) and CSB VII (1), while two consecutive G conserved sites were located after CSB VII in a Calliphoridae-specific G + A-rich conserved element (Fig. 1).

A highly variable B domain was identified in the 3′ sub-region of the mtDNA control region. Most of the length variation in this domain was attributable to insertion/deletion events and the presence of frequent short repetitive sequences in tandem, which were especially evident in *D. hominis* (largest B domain with ~ 793 bp). The hypervariable B domain contained extensive gaps when Oestridade and Calliphoridae species are aligned. However, even with the Calliphoridae, the high level of sequence divergence greatly increased the inconsistency of the resulting alignment of the B domain sequences.

Nucleotide content

The mean AT content was 91% for *C. hominivorax*, 87% for *C. macellaria*, 90.5% for *L. eximia*, 88.5% for *C. megacephala* and 91% for *D. hominis*. Transversions (tv) were more common than transitions (ts) in all comparisons between species, with a maximum ts : tv rate of 1 : 5.2 between *L. eximia* and *C. megacephala* and a minimum of 1 : 1.6 between *C. hominivorax* and *C. macellaria*. Analysis of the A domain nucleotide substitution frequencies showed an average of 75% for transversion events, of which 67–84% resulted from A↔T replacements. The A domain sequences were used for these analyses instead of the total control region information, as the B domain provided low alignment confidence and had not been analysed at this level.

Phylogenetic data

The number of nucleotide substitutions per nucleotide site for the species studied was estimated using the method of Tamura (1992), taking into account the extremely high A + T content of the control region. No B domain sequences were included in this analysis. The distance matrix for Calliphoridae and Oestridae A domain sequences is shown in Table 3. If *D. hominis* is removed from the analysis to calculate distances based on a Calliphoridae-specific alignment of the A domain sequences, the resulting Calliphoridae matrix has slightly higher distance values which yield a very similar relationship to the results with *D. hominis*. Although functional constraints may operate at the conserved sites of the control region, variable sequences within

Table 3. Tamura (1992) distance matrix for Calliphoridae and Oestridae A domain sequences. *C. homi.* = *C. hominivorax*, *C. mace.* = *C. macellaria*, *L. exim.* = *L. eximia*, *C. mega.* = *C. megacephala* and *D. homi.* = *D. hominis*. The upper-right matrix indicates the distance and the lower-left matrix indicates the standard error. Gap sites and missing information were removed from the analysis.

	D. homi. 1	*D. homi.* 2	*C. mace.* 1	*C. mace.* 2	*C. homi.* 1	*C. homi.* 2	*L. exim.* 1	*C. mega.* 1
D. homi. 1		0.0096	0.2801	0.2793	0.2927	0.2994	0.2980	0.3022
D. homi. 2	0.0044		0.2668	0.2662	0.2860	0.2926	0.2914	0.2955
C. mace. 1	0.0298	0.0285		0.0076	0.0981	0.1057	0.2104	0.2120
C. mace. 2	0.0296	0.0283	0.0038		0.0977	0.1052	0.2046	0.2117
C. homi. 1	0.0309	0.0303	0.0153	0.0151		0.0097	0.1995	0.1992
C. homi. 2	0.0316	0.0309	0.0161	0.0160	0.0044		0.2085	0.2018
L. exim. 1	0.0311	0.0305	0.0241	0.0236	0.0232	0.0241		0.1828
C. mega. 1	0.0317	0.0310	0.0240	0.0239	0.0231	0.0234	0.0217	

Table 4. *Cochliomyia* intrageneric analysis: Tamura (1992) distance for B domain sequences. SE indicates standard errors. Gap sites and missing information were removed from the analysis.

Control region B domain Species	Tamura distance (SE)
C. hominivorax 1 × *C. hominivorax* 2	0.0039 (0.0028)
C. hominivorax 1 × *C. macellaria* 1	0.3878 (0.0410)
C. hominivorax 1 × *C. macellaria* 2	0.3920 (0.0416)
C. hominivorax 2 × *C. macellaria* 1	0.3878 (0.0410)
C. hominivorax 2 × *C. macellaria* 2	0.3920 (0.0416)
C. macellaria 1 × *C. macellaria* 2	0.0077 (0.0039)

the A domain were useful for efficient outgroup identification (*D. hominis*) and correctly grouped together *C. hominivorax* and *C. macellaria* in neighbour-joining (NJ) analysis (Fig. 3). However, the NJ tree incorrectly grouped *C. megacephala* and *L. eximia*, which are classified in distinct subfamilies (Chrysomyinae and Luciliinae, respectively).

The B domain sequences were too variable to guarantee consistent alignment and validate homologous sites for phylogenetic inferences at this taxonomic level. Even among the Calliphoridae species, no consensus sequences were accurately identified. However, the B domain sequences could be used to align the *Cochliomyia* species with a high level of confidence. Based on this alignment, the estimated Tamura (1992) distance between *C. hominivorax* and *C. macellaria* was approximately 0.389. The hypervariable B domain evolved at least three times faster, based on nucleotide substitution rates, than the A domain between *Cochliomyia* species (see Tables 3 and 4).

Discussion

Zhang & Hewitt (1997) have divided insect species into two groups based on the distribution of conserved sequence elements in the control region. The structural organization of the control region reported here for blowflies is similar to that for *Drosophila* species, but differs from the model proposed for Lepidoptera (Taylor *et al.*, 1993), Orthoptera (Zhang *et al.*, 1995) and *Anopheles* (Caccone *et al.*, 1996). The genetic variability in the species studied here is apparently restricted to a specific area in the major non-coding region, which suggests that sequences in the A domain may be under functional constraints. These findings agree with previous analyses of the *Drosophila* A + T-rich region (Clary & Wolstenholme, 1987; Monforte *et al.*, 1993; Lewis *et al.*, 1994; Inohira *et al.*, 1997).

The organization observed here suggests that the replication/transcription system may interact with *cis* regulatory elements present in the control region 5′ end of Brachycera species. However, despite extensive sequence analysis, only CSB I and CSB VIII have been described as potentially associated with regulatory domains in insects. Lewis *et al.* (1994) suggested that the positions and orientations of the two conserved thymidylate stretches indicated roles in promoting the transcription and replication of *Drosophila* mtDNA. The search for homologous conserved elements in *Ceratitis capitata* control region sequences (Spanos *et al.*, 2000) revealed the presence of long poly-T stretches close to the $tRNA^{Ile}$ gene and on the opposite strand. This finding strongly supports the idea that highly conserved poly-T sequences may play a fundamental role in the maintenance and function of mtDNA in higher Diptera. At least one of these poly-T stretches (close to the $tRNA^{Ile}$ gene) may have similar regulatory mechanisms in Orthopteran mtDNA (Zhang *et al.*, 1995). It may be noticed that polymerase gamma requires the presence of a single-stranded DNA-binding protein for efficient elongation using poly-(dT) templates, T7 DNA polymerase has a similar requirement (Mikhailov & Bogenhagen, 1996). Other conserved features of the control region include a dinucleotide pyrimidine-purine series in tandem (CSB II and III), which has also been reported for phylogenetically distant insect groups (Zhang & Hewitt, 1997).

Although multiple and extensive intrastrand base pairing was common within conserved domains of the control region (Lewis *et al.*, 1994), no conserved secondary structure could conclusively be assigned to the A domain sequences of the species studied here. The absence of a widely conserved secondary structure has been reported for *D. melanogaster* mtDNA (Inohira *et al.*, 1997).

In addition to control region analysis, the nucleotide composition of the flanking srRNA 12S sequences was also determined in order to allow comparison of the average AT content of this region between Calliphoridae species and other

dipterans. The average A + T composition of 82% found here was similar to the 79% reported for *D. yakuba* srRNA (Clary & Wolstenholme, 1985) and identical to the srRNA A + T content reported for *Anopheles quadrimaculatus* (Mitchell *et al.*, 1993). The nucleotide composition of the control region was comparable to the *Drosophila* A + T content, which varies from 90% to 96% (Clary & Wolstenholme, 1985, 1987; Monnerot *et al.*, 1990; Lewis *et al.*, 1994) and to that of *Anopheles gambiae* (Beard *et al.*, 1993) and *A. quadrimaculatus* (Mitchell *et al.*, 1993), which have an A + T content of 94%.

The A↔T nucleotide substitution pattern of the control region sequences indicates selection for a high AT content. These data agree with the interpretation of Clary & Wolstenholme (1985), who suggested that the maintenance of an AT biased composition may be important for optimizing energy-requiring processes in the replication system of dipteran species. Directional mutation pressure has been suggested to be responsible for the high A + T composition in insect mtDNA (Jermiin *et al.*, 1994)

The incongruent phylogenetic arrangements derived from the present data indicate that the control region has limited usefulness for resolving taxonomic relationships among genera. At this level, control region sequences are influenced considerably by the alignment parameters and distance estimation methodology, which in turn decrease the confidence in homologous site inferences. The inclusion of additional species from diversified Calliphoridae taxa may improve our understanding of phylogenetic usefulness of the control region in these group of organisms.

There are many questions concerning the ecological and evolutionary behaviour of blowflies species that could be elucidated using informative molecular markers. The colonization of the Americas by *Chrysomya* species has reportedly led to a reduction in the native fly fauna (Ferreira, 1982; Prado & Guimarães, 1982; Baumgartner & Greenberg, 1984). The decrease in the genetic variability of *C. macellaria* populations has been associated with the presence of *Chrysomya* (Valle & Azeredo-Espin, 1995). However, the exact source of New World *Chrysomya* remains to be defined. The controversial taxonomic status of *C. albiceps* (Wiedmann) and *C. rufifacies* (Maquart) has recently been investigated using mtDNA markers, which provide an unambiguous approach to species identification (Wells & Sperling, 1999). In Latin America, where the distributions of these species overlap (Tantawi & Greenberg, 1993), the investigation of useful mitochondrial and nuclear DNA markers may be important in ecological, forensic and genetic studies.

Control region sequences can be very informative when traditional mtDNA markers show low or inadequate levels of genetic variability. The recent report of primary myiasis caused by *L. eximia* in Brazil (Azeredo Espin & Madeira, 1996) suggests that it would be important to investigate the evolutionary processes related to these facultative species.

The complete control region sequences provided information about which Calliphoridae and Oestridae species would be useful in designing primers for each domain. This approach could lead to the development of PCR-RFLP diagnostic tests and the screening of suitable markers directly on mtDNA hypervariable sequences. The identification of species-specific restriction sites in these myiasis-causing flies could improve the study of individual variation by providing a useful PCR-RFLP marker for genetic polymorphism. Interspecific analyses of specific mtDNA regions have already been carried out with blowflies in an attempt to identify diagnostic markers for species-specific identification in forensic investigations (Sperling *et al.*, 1994), calliphorid taxonomy (Stevens & Wall, 1997; Wells & Sperling, 1999) and in the monitoring and control of screwworm flies (Taylor *et al.*, 1996).

Our preliminary search for diagnostic restriction sites using the program WebCutter 2.0 (Heiman, 1997), confirmed by experimental analysis, revealed five restriction sites that unambiguously identify *C. hominivorax* and *C. macellaria* in Brazilian screwworm populations (Litjens *et al.*, unpublished). This observation demonstrates the potential applicability of control region sequences to population genetic studies and to the phylogeny of closely related species.

Acknowledgements

The authors thank M. C. Arias for helpful comments and suggestions on sequencing and sequence analysis, R. A. Rodrigues, M. S. do Couto and M. C. Filho for valuable technical assistance and two anonymous referees for their critical review of the manuscript. This research was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP grant 95/8856-1) and PADCT/CNPq grant 620097/95-7 to A.M.L.A.E. A.C.L. was supported by a fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil).

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein data base search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Azeredo-Espin, A.M.L. (1987) *Análise cariotípica, morfométrica e de compatibilidade sexual, em linhagens Brasileiras de* Cochliomyia hominivorax *(Diptera: Calliphoridae).* PhD Thesis, Universidade Estadual de Campinas, Brazil.

Azeredo-Espin, A.M.L. (1993) *Mitochondrial DNA Variability in Geographic Populations of Screwworm Fly from Brazil.* International Atomic Energy Agency, Vienna, pp. 161–165.

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. *Journal of Medical Entomology*, **33**, 839–843.

Bailey, T.L. & Gribskov, M. (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*, **14**, 48–54.

Baumgartner, D.L. & Greenberg, B. (1984) The genus *Chrysomya* (Diptera: Calliphoridae) in the New World. *Journal of Medical Entomology*, **21**, 105–113.

Beard, C.B., Hamm, D.M. & Collins, F.H. (1993) The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization and comparisons with mitochondrial sequence of other insects. *Insect Molecular Biology*, **2**, 103–124.

Caccone, A., Garcia, B.A. & Powell, J.R. (1996) Evolution of the

mitochondrial DNA control region in the *Anopheles gambiae* complex. *Insect Molecular Biology*, **5**, 51–59.

Clary, D.O. & Wolstenholme, D.R. (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. *Journal of Molecular Evolution*, **22**, 252–271.

Clary, D.O. & Wolstenholme, D.R. (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. *Journal of Molecular Evolution*, **25**, 116–125.

Ferreira, M.J.M. (1982) Sinatropia de Calliphoridae (Diptera) em Goiânia, Goiás. *Revista Brasileira de Biologia*, **43**, 199–210.

Heiman, M. (1997) *Webcutter 2.0*. copyright 1997. http.//www.firstmarket.com/cutter/cut2.htm

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

Inohira, K., Hara, T. & Matsuura, E.T. (1997) Nucleotide sequence divergence in the A+T-rich region of mitochondrial DNA in *Drosophila simulans* and *Drosophila mauritiana*. *Molecular Biology and Evolution*, **14**, 814–822.

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. *Journal of Molecular Evolution*, **39**, 160–173.

Kumar, S., Tamura, K. & Nei, M. (1993) *MEGA: Molecular Evolutionary Genetics Analysis, Version 1.01*. The Pennsylvania State University, University Park, Pennsylvania.

Lapalme, G., Cedergren, R.J. & Sankoff, D. (1982) An algorithm for the display of nucleic acid secondary structure. *Nucleic Acids Research*, **10**, 835–836.

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. *Molecular Biology and Evolution*, **11**, 523–538.

Mikhailov, V.S. & Bogenhagen. D.F. (1996) Termination within oligo (dT) tracts in template DNA by DNA polymerase gamma occurs with formation of a DNA triplex structure and is relieved by mitochondrial single-stranded DNA-binding protein. *Journal of Biological Chemistry*, **271**, 30774–30780.

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.

Monforte, A., Barrio, E. & Latorre, A. (1993) Characterisation of the length polymorphism in the A+T-rich region of the *Drosophila obscura* group species. *Journal of Molecular Evolution*, **36**, 214–223.

Monnerot, M., Solignac, M. & Wolstenholme, D.R. (1990) Discrepancy in divergence of the mitochondrial and nuclear genomes of *Drosophila teissieri* and *Drosophila yakuba*. *Journal of Molecular Evolution*, **30**, 500–508.

Prado, A.P. & Guimarães, J.H. (1982) Estado atual da distribuição e dispersão das espécies do gênero *Chrysomya* R-D na região Neotropical (Diptera: Calliphoridae). *Revista Brasileira de Entomologia*, **26**, 225–231.

Robbins, C.M., Hsu, E. & Gillevet, P.M. (1996) Sequencing homopolymer tracts and repetitive elements. *Biotechniques*, **20** (862–4), 866–8.

Saitou, N. & Nei, M. (1987) The neighbor–joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Sambook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Shadel, G.S. & Clayton, D.A. (1993) Mitochondrial transcription initiation: variation and conservation. *Journal of Biological Chemistry*, **268**, 16083–16086.

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

Simon, C. (1991) Molecular systematics of the species boundary: exploiting conserved and variable regions of the mitochondrial genome of animals via direct sequencing from amplified DNA. *Molecular Techniques in Taxonomy* (ed. by G. M. Hewitt, A. W. B. Johnston and J. P. W. Young), pp. 33–71. Springer Verlag, Berlin.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution, weighting and phylogenetic utility of mitochondrial DNA sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Spanos, L., Koutroumbas, G., Kotsyfakis, M. & Louis, C. (2000) *Ceratitis capitata* sequence. *Insect Molecular Biology*, in press.

Sperling, F.A.H., Anderson, G.S. & Hickey, D.A. (1994) A DNA-based approach to the identification of insect species used for postmorten interval estimation. *Journal of Forensic Sciences*, **39**, 418–427.

Stevens, J. & Wall, R. (1997) The evolution of ectoparasitism in the genus *Lucilia* (Diptera: Calliphoridae). *International Journal of Parasitology*, **27**, 51–59.

Tamura, K. (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution*, **9**, 678–687.

Tantawi, T.I. & Greenberg, B. (1993) *Chrysomya albiceps* and *C. rufifacies* (Diptera: Calliphoridae) contribution to an ongoing taxonomic problem. *Journal of Medical Entomology*, **30**, 646–648.

Taylor, M.F.J., McKechnie, S.W., Pierce, N. & Kreitman, M. (1993) The lepidopteran mitochondrial control region: structure and evolution. *Molecular Biology and Evolution*, **10**, 1259–1272.

Taylor, D.B., Szalanski, A.L. & Peterson, R.D. II (1996) Identification of screwworm species by polymerase chain reaction–restriction fragment length polymorphism. *Medical and Veterinary Entomology*, **10**, 63–70.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Wells, J.D. & Sperling, F.A.H. (1999) Molecular phylogenetics of *Chrysomya albiceps* and *Chrysomya rufifacies* (Diptera: Calliphoridae). *Journal of Medical Entomology*, **36**, 222 – 226.

Valle, J.S. & Azeredo-Espin, A.M.L. (1995) Mitochondrial DNA variation in two Brazilian populations of *C. macellaria* (Diptera: Calliphoridae). *Brazilian Journal of Genetics*, **18**, 521–526.

Yotoko, K.S.C. (1998) *Caracterização da variabilidade genetica de mosca do berne* Dermatobia hominis *(Diptera: Oestridae) através da técnica de RAPD-PCR*. Masters Dissertation, Universidade Estadual de Campinas, Brazil.

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

Zhang, D.-X., Szymura, J.M. & Hewitt, G.M. (1995) Evolution and structural conservation of the control region of insect mitochondrial DNA. *Journal of Molecular Evolution*, **40**, 382–391.

Zuker, M. (1989) Computer prediction of RNA structure. *Methods in Enzymology*, **180**, 262–288.

Accepted 26 October 1999

CAPÍTULO 2

Artigo 2 - *Lessinger A. C.*; Martins Junqueira A. C.; Lemos T. A., Kemper E. L., da Silva F. R., Vettore A. L., Arruda P., Azeredo-Espin A. M. (2001). **The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae).** *Insect Molecular Biology*, **9**: 521 - 529.

Insect Molecular Biology (2000) 9(5), 521–529

The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae)

A. C. Lessinger,[1] A. C. Martins Junqueira,[1] T. A. Lemos,[1] E. L. Kemper,[2] F. R. da Silva,[2] A. L. Vettore,[2] P. Arruda[2,3] and A. M. L. Azeredo-Espin[1,3]

[1]*Laboratório de Genética Animal,* [2]*Laboratório Genoma, Centro de Biologia Molecular e Engenharia Genética,* [3]*Departamento de Genética e Evolução/DGE – Instituto de Biologia, Universidade Estadual de Campinas – UNICAMP, Campinas, São Paulo, Brazil*

Abstract

The complete sequence of the mitochondrial genome of the screwworm *Cochliomyia hominivorax* was determined. This genome is 16 022 bp in size and corresponds to a typical Brachycera mtDNA. A Serine start codon for COI and incomplete termination codons for COII, NADH 5 and NADH 4 genes were described. The nucleotide composition of *C. hominivorax* mtDNA is 77% AT-rich, reflected in the predominance of AT-rich codons in protein-coding genes. Non-optimal codon usage was commonly observed in *C. hominivorax* mitochondrial genes. Phylogenetic analysis distributed the Acalypterate species as a monophyletic group and assembled the *C. hominivorax* (Calyptratae) and the Acalyptratae in a typical Brachycera cluster. The identification of diagnostic restriction sites on the sequenced mitochondrial genome and the correlation with previous RFLP analysis are discussed.

Keywords: *Cochliomyia hominivorax*, screwworm, myiasis, mtDNA, Calyptratae.

Introduction

Mitochondrial genomes have been studied increasingly because of the ease of recovering genetic information that may be useful for investigating molecular and organismal evolution. Complete sequences of the mitochondrial genomes are available for several animal taxa and provide insights into the patterns of mtDNA evolution. In general, metazoan mtDNA contains thirteen protein-coding genes, two rRNA genes, twenty-two tRNA genes and a major noncoding (A + T-rich in insects) region, with an average size of 16 kb, indicating a very compact circular genome. The predominance of maternal inheritance, lack of extensive recombination, and accelerated rates of nucleotide substitution are features that have favoured the use of mtDNA as an informative evolutionary marker.

Animal mtDNA sequences have been the subject of recent reviews that have provided comparative analysis of mitochondrial systems (Boore, 1999), and of evolutionary genomics (Saccone *et al.*, 1999), and have assessed the phylogenetic information over a wide taxonomic range (Curole & Kocher, 1999). Boore (1999) referred to 'mitochondrial genomics' as those approaches related to understanding the diversity of mitochondrial systems. Because this genome contains information for protein-coding, ribosomal and transfer RNA genes, as well as regulatory noncoding sequences, it provides a reliable system for analysis of the evolution of these genetic elements.

Currently, the complete mtDNA sequences of only eleven species of Arthropoda are known, including the Hexapoda (Clary & Wolstenholme, 1985; Beard *et al.*, 1993; Crozier & Crozier, 1993; Mitchell *et al.*, 1993; Flook *et al.*, 1995; Lewis *et al.*, 1995; and Spanos *et al.*, 2000), Chelicerata (Black & Roehrdanz, 1998) and Crustacea (Valverde *et al.*, 1994 and Crease, 1999) taxa. The five dipteran mitochondrial genomes sequenced are from the suborders Brachycera, subsection Acalyptratae (*Drosophila* and *Ceratitis*), and Nematocera (*Anopheles*).

The information on calypterate species in GenBank corresponds to six main regions of the mitochondrial genome (fifty-seven entries), including partial sequences of the genes encoding subunits I and II of cytochrome oxidase, the rRNA 16S and 12S genes, subunit I of cytochrome b and the control region. The most common sequence is that of subunit I of cytochrome oxidase, and reflects the fact that this gene has become one of the most used molecular markers applied for insects evolutionary studies (Zhang & Hewitt, 1997).

Most studies of dipteran species are related to the occurrence of flies of medical and economic importance, such

Received 5 May 2000; accepted following revision 4 July 2000. Correspondence: Ana Maria Lima de Azeredo-Espin, Laboratório de Genética Animal, CBMEG – UNICAMP, PO Box 6010, 13083–970 Campinas, São Paulo, Brazil. Fax: 055–19–7881089; e-mail: azeredo@unicamp.br

as the Anopheline vector-species of malaria (Beard *et al.*, 1993; Mitchell *et al.*, 1993) and the fruit fly *Ceratitis capitata*, a serious agricultural pest (Spanos *et al.*, 2000). The importance of *Drosophila* as a model organism lead to this species being the first invertebrate for which mtDNA was completely sequenced (Clary & Wolstenholme, 1985). The family Calliphoridae contains members of great medical and veterinary importance, especially species associated with myiasis. This family includes free-living saprophagous insects that serve as vectors for pathological agents and parasitic species responsible for serious infestations of livestock as obligate or facultative parasites. Approximately 50% of the calypterate mtDNA sequences in GenBank are from Calliphoridae species. These sequences have been derived from studies dealing with the identification of species of forensic importance (Sperling *et al.*, 1994), the resolution of controversial taxonomic relationships (Ballard *et al.*, 1992; Wells & Sperling, 1999), and the evolution of mtDNA control region (Lessinger & Azeredo-Espin, 2000).

The New World screwworm fly (NWS), *Cochliomyia hominivorax* (Coquerel, 1859), is one of the most important agents of traumatic myiasis throughout the neotropical region (Guimarães *et al.*, 1983). *Cochliomyia hominivorax* is an obligate ectoparasite and can infest almost all warm-blooded animals, including humans. Before its eradication from North America (Graham, 1985), this screwworm ranged from Central United States to central Argentina and Chile (Hall & Wall, 1995). Because of its importance as a serious livestock pest, the mtDNA of *C. hominivorax* has been the subject of many studies aimed at establishing the genetic polymorphism and population structure of this species (Roehrdanz & Johnson, 1988; Roehrdanz, 1989; Azeredo-Espin, 1993; Infante & Azeredo-Espin, 1995; Taylor *et al.*, 1996; Infante-Malachias, 1999). The current GenBank entries for this species refer to the mtDNA control region and partial sequences for the 12S rRNA and $tRNA^{Ile}$ genes (Lessinger & Azeredo-Espin, 2000).

In this work, we describe the complete sequence of the mitochondrial genome of *C. hominivorax*. These results should allow the identification of species-specific genetic markers in the primary mtDNA sequence, which may be useful for monitoring and controlling this ectoparasite (Taylor *et al.*, 1996; Litjens *et al.*, 1999).

The complete sequence of *C. hominivorax* mitochondrial genome is also interesting as a source of sequence information for general Diptera molecular and evolutionary approaches. These data could contribute in primer selection at specific mtDNA regions, optimization of PCR-RFLP assays and analysis of phylogenetic information for understanding dipteran evolution, especially from a Calyptrate perspective.

Results and discussion

MtDNA gene content and general organization

The mitochondrial genome of *C. hominivorax* is a circular molecule of 16 022 bp. The gene content and general organization pattern corresponds to typical Brachycera mtDNA when compared with species for which the complete mtDNA sequence is known (Clary & Wolstenholme, 1985; Lewis *et al.*, 1995; Spanos *et al.* 2000). A usual gene content of thirteen protein-coding genes, twenty-two tRNA genes and two rRNA genes was found for *C. hominivorax*. The main noncoding region, identified as the mtDNA control region located between the $tRNA^{Ile}$ and rRNA 12S genes, was 1175 bp in size and was previously characterized by Lessinger & Azeredo-Espin (2000). The differences in the total length of dipteran mtDNA molecule were apparently related to variation in the size of the control region sequences, particularly since noncoding sequences and overlapping regions are infrequent events in this genome. As reported for other insects, *C. hominivorax* mtDNA had a bias in nucleotide composition which led to an AT-rich genome, this bias is increased in the codon's third base of protein coding genes as predicted by lower selection and mutational pressure (Jermiin *et al.*, 1994).

As shown in Table 1, the mitochondrial rRNA genes of *C. hominivorax* were conserved in size and nucleotide composition when compared with other dipterans. In addition to control region sequences, the ribosomal RNA genes had a high AT content, especially 16S rRNA, and this

Table 1. Comparison of dipteran mitochondrial genomes showing the size and A + T content of the complete mtDNA, the control region and the rRNA genes

	mtDNA		Control region		lrRNA		srRNA		
Taxa	Size	AT%	Size	AT%	Size	AT%	Size	AT%	References
C. hominivorax	16.022	77.0	1.175	90.8	1.323	81.5	785	77.0	This work
C. capitata	15.980	77.5	1.004	91.1	1.335	81.8	788	77.7	Spanos *et al.* (2000)
D. yakuba	16.019	78.6	1.077	92.9	1.326	83.4	789	79.3	Clary & Wolstenholme (1985)
D. melanogaster	19.517	82.2	4.061	96.0	1.325	82.9	786	80.2	Lewis *et al.* (1994) Lewis *et al.* (1995)
A. gambiae	15.363	77.6	520	94.0	1.325	82.5	800	82.5	Beard *et al.* (1993)
A. quadrimaculatus	15.455	77.4	625	94.0	1.321	80.5	794	82.2	Mitchell *et al.* (1993)

Table 2. Non-coding regions and overlapping gene sequences in the *C. hominivorax* mitochondrial genome. The signs + and – correspond to gene orientation relative to the plus and minus strands, respectively

Nucleotide position	Gene	Strand	Size (bp)
Non-coding regions			
1239–1251	$tRNA^{Ile}/tRNA^{Gln}$	+/–	13
1321–1336	$tRNA^{Gln}/tRNA^{Met}$	–/+	16
2545	$tRNA^{Cys}/tRNA^{Tyr}$	–/–	1
4210–4213	$tRNA^{Leu}$/CO II	+/+	4
6662–6667	CO III/$tRNA^{Gly}$	+/+	6
7087	ND 3/$tRNA^{Ala}$	+/+	1
7217–7231	$tRNA^{Arg}/tRNA^{Asn}$	+/+	15
7434–7451	$tRNA^{Glu}/tRNA^{Phe}$	+/–	18
9220–9252	ND 5/$tRNA^{His}$	–/–	15
10948–1949	ND 4L/$tRNA^{Thr}$	–/+	2
11081–11082	$tRNA^{Pro}$/ND 6	–/+	2
12809–12824	$tRNA^{Ser}$/ND 1	+/–	16
13764–13773	ND 1/$tRNA^{Leu}$	–/–	10
13839	$tRNA^{Leu}$/lrRNA 16S	–/–	1
	Total		120
Overlapping sequences			
2422–2421	ND 2/$tRNA^{Trp}$	+/+	2
2488–2481	$tRNA^{Trp}/tRNA^{Cys}$	+/–	8
4148–4144	CO I/$tRNA^{Leu}$	+/+	5
5202–5196	ATPase8/ATPase6	+/+	7
6732–6730	$tRNA^{Gly}$/ND 3	+/+	3
10657–10651	ND 4/ND 4L	–/–	7
12743–12742	Cytb/$tRNA^{Ser}$	+/+	2
	Total		34

contributes to the increase in the average AT composition of the mtDNA molecule. Analyses based on sequence similarity showed that the 16S and 12S rRNA genes had, respectively, 89% and 88% sequence identity with homologous regions of *Drosophila yakuba*.

An initial attempt to characterize A + T-rich region polymorphism was done based on alignment with available *C. hominivorax* control regions from GenBank (AF151382/AF151383). In this analysis, we identified thirteen nucleotide substitutions (eight transitions and five transversions) and two deletions (one single bp and a 5 bp extension gap). Both *C. hominivorax* control regions previously reported have an ATATA-sequence between positions 137 and 138 described in this work. The occurrence of insertion/deletion events is a common feature of insect control regions, which are known to have repetitive sequences.

In addition to the control region, fourteen noncoding regions varying in size from 1 to 18 bp (Σ = 120 bp) were also identified in the *C. hominivorax* mitochondrial genome (Table 2). Several noncoding regions have been reported for other arthropod mtDNAs, with total length varying from 43 in *Anopheles* to 620 in *Apis*.

The assignment of noncoding regions was directly related to the annotation criteria used for identifying the mitochondrial genes, but may occasionally conflict with undefined start and stop codons in protein coding genes as discussed above.

Termination in protein-coding genes

The identification of ORFs in animal mtDNA, especially insects, is not conclusively established, because information on unusual initiation and termination codons is increasing as more mitochondrial genomes are described (Campbell & Barker, 1999; Crease, 1999; Boore & Brown, 2000; Spanos *et al.*, 2000). Among dipterans, incomplete termination codons or even the absence of recognizable stop codons has been reported for a variety of protein-coding genes (Beard *et al.*, 1993; Clary & Wolstenholme, 1985; Mitchell *et al.*, 1993; Spanos *et al.*, 2000).

An initial attempt to identify overlapping regions in *C. hominivorax* mtDNA led to three conflicts with the predicted transcripts required to agree with the standard TAA or TAG termination codons. Large overlaps of 20 bp and 35 bp occurred between the $tRNA^{Phe}$ and ND5 genes and between the $tRNA^{Lys}$ and COII genes, respectively. An even more extensive overlap involving the entire $tRNA^{His}$ gene sequence and the 3′ end of the ND4 gene was necessary in order to agree with a final TAG codon. This resulted in a predicted protein thirty-one amino acids longer (477 residues) than the corresponding *Drosophila* and *Ceratitis* ND4 proteins (446 residues).

These major conflicts in overlapping involved sequences immediately adjacent to the 5′ terminal nucleotide from the sense strand of a tRNA gene. No overlapping regions in the *C. hominivorax* mitochondrial genome were longer than 8 bp when an alternative tRNA termination was assumed as an authentic mechanism with the occurrence of T/TA ends.

The characterization of tRNA-terminated genes and the few studies dealing with the expression of mtDNA genes could be responsible for potential conflicts among predicted gene transcripts and the originally expressed proteins. Based on the annotation criteria described in *Methods*, and considering the possibility of conceptual conflicts, we have identified seven overlapping regions (involving at least 2 bp) in the mtDNA of *C. hominivorax* (Table 2).

The genes for $tRNA^{Trp}$ and $tRNA^{Cys}$ overlapped by 8 bp, but because they are encoded by opposite strands, complete transcripts of each gene could be produced independently. There were two overlaps involving adjacent protein-coding genes. The overlaps between the ATPase 8 and ATPase 6 genes are a common feature of metazoan mtDNA (Campbell & Barker, 1999), and are translated from the same bicistronic mRNA (Fearnley & Walker, 1986). In arthropods, these genes also overlapped in *Locusta*, *Ceratitis*, *Drosophila*, *Anopheles*, *Apis* and *Boophilus* (Flook *et al.*, 1995; Clary & Wolstenholme, 1985; Crozier & Crozier, 1993; Mitchell *et al.*, 1993; Campbell & Barker, 1999; Spanos *et al.*, 2000). Overlapping regions such as seen in the

Table 3. Recognizable initiation and termination signals of *C. hominivorax* protein-coding genes. The *denotes an anomalous start codon. (I) = Isoleucine (M) = Methionine (S) = Serine

Genes	Position	strand	Initiation	Termination
ND 2	1.406–2.422	plus	ATT (I)	TAA
CO I	2.610–4.148	plus	TCG* (S)	TAA
CO II	4.214–4.901	plus	ATG (M)	T or $tRNA^{Lys}$
ATPase 8	5.038–5.202	plus	ATT (I)	TAA
ATPase 6	5.196–5.873	plus	ATG (M)	TAA
CO III	5.873–6.661	plus	ATG (M)	TAA
ND 3	6.730–7.086	plus	ATA (M)	TAA
ND 5	7.518–9.237	minus	ATT (I)	T(A) or $tRNA^{Phe}$
ND 4	9.319–10.657	minus	ATG (M)	T(A) or $tRNA^{His}$
ND 4L	10.651–10.947	minus	ATG (M)	TAA
ND 6	11.083–11.607	plus	ATT (I)	TAA
Cyt b	11.607–12.743	plus	ATG (M)	TAG
ND 1	12.825–13.763	minus	ATA (M)	TAA

C. hominivorax ND4 and ND4L genes (7 bp) were also reported for *Locusta*, *Anopheles* and *Boophilus*.

Experiments using mtRNA transcripts of human mtDNA have suggested that tRNA genes play a major role as recognition signals in the processing of the primary transcript, being responsible for the precise release of individual transcripts from the polycistronic precursor RNA (Ojala *et al.*, 1981). A post-transcriptional mechanism could be responsible for the polyadenylation of mtDNA transcripts by providing complete TAA stop codons and recovering the termination signal for efficient translation (Yokobori & Pääbo, 1997).

A cleavage-polyadenylation mechanism seems to operate in *Drosophila* mitochondria, because polyadenylated mRNAs and mRNAs with multiple gene transcripts have been obtained from *D. melanogaster* and *D. virilis* (Spradling, 1977; Battey & Clayton, 1978; Merten & Pardue, 1981). However, most of the considerations concerning the identification of gene boundaries were based on previous reports from model organisms rather than being confirmed by experimental evidence from different insect species.

The Cyt.b termination codon (TAG) provided a valid condition for testing the incomplete termination hypothesis for *C. hominivorax* mitochondrial genes based on a post-transcriptional polyadenylation mechanism. This gene has a 2 bp overlap with $tRNA^{Ser}$ which, during mRNA editing, may be converted to a TAA termination codon. Such analysis was not conducted for *C. hominivorax*, however a similar approach based on the ND1 cDNA sequence of amphioxus (*Branchiostoma lanceolatum*) confirmed the construction of a TAA stop codon from an TAG overlap in the original mtDNA sequence (Delarbre *et al.*, 1997).

Table 3 summarizes the protein-coding genes and their respective initiation/termination codons as assumed for the *C. hominivorax* mitochondrial genome described here.

Initiation in protein-coding genes

In addition to termination conflicts, initiation codons could also provide an interpretation slightly different from the standard invertebrate mtDNA code.

The initiation of insect COI genes does not agree perfectly with the invertebrate mitochondrial code and, in Diptera, the initiation in a TCG codon was reported for species from Calliphoridae (Sperling *et al.*, 1994; Wells & Sperling, 1999), Tephritidae (Spanos *et al.*, 2000) and Culicidae (Beard *et al.*, 1993; Mitchell *et al.*, 1993) families. A potential TCG initiation codon (Ser) was also identified in the *C. hominivorax* COI gene. This putative initiation codon can be consistently assigned in other insects (Beard *et al.*, 1993; Lunt *et al.*, 1996) and, recently, Spanos *et al.* (2000) suggested that the occurrence of a serine start in COI genes is a usual feature in dipteran mtDNA. We agree with Mitchell *et al.* (1993) who suggested that further analysis of protein and/or mRNA transcripts would be required to provide an accurate interpretation of COI initiation.

Inferences of the initiation codon for ND1 and ND6 genes were less certain because two commonly used start codons were 'in-frame' and adjacent at both the 5′ ends of both genes, with no overlapping conflicts. We have adopted the first start codon available (ATA for ND1 and ATT for ND6) because the conceptual transcript corresponded to proteins more similar to those described for other dipterans.

Protein-coding genes

Protein-coding genes had an A + T content of approximately 75%, and the distribution of this composition over the first, second and third nucleotides of the codon was, respectively, 68.5%, 66.1% and 89.6%. This agrees with the suggestion that there is an evident bias for a high proportion of AT nucleotides in the wobble position of several insect mtDNA codons.

Table 4 provides a general overview of the *C. hominivorax* mtDNA coding sequences compared with *D. yakuba* and *C. capitata* genomes. Individual gene sequences were described in terms of predicted protein size and A + T content. This analysis showed that *C. hominivorax* mitochondrial genes were very similar to their homologues in other dipteran species. Considering the variation in predicted protein length, the ND1 gene gave the more variable results. This variation was partially caused by the occurrence of a longer *C. capitata* ND1 gene at the 3′ end, at least under hypothetical translation. *Cochliomyia hominivorax* had a shorter predicted size for ND1 (311 residues) than the other flies. The predicted ND1 peptide for *C. hominivorax* was most similar in length to the *Anopheles* (316 residues) and *Apis* (305 residues) proteins, indicating that this gene has preserved its functionality over a wide range in size in Diptera.

Distances were calculated for individual genes in order to examine variable patterns of evolution (Table 4). The Kimura two-parameter model (Kimura, 1980) was used to

Table 4. Comparison of the protein-coding genes of *C. hominivorax*, *D. yakuba* and *C. capitata* mtDNA based on predicted protein size, A + T content and genetic distance. The nucleotide distance was based on the Kimura 2-P model (Kimura, 1980). The initiation codons for the *C. hominivorax* genes were excluded from the analysis

	C. hominivorax		*D. yakuba*			*C. capitata*		
Protein-coding genes	Protein **size (a.a.)**	AT%	Protein **size (a.a.)**	AT%	Nucleotide **distance**	Protein **size (a.a.)**	AT%	Nucleotide **distance**
ND 2	337	79.0	341	81.5	0.2586	340	79.9	0.2668
CO I	511	68.6	511	69.8	0.1426	511	69.3	0.1448
CO II	228	72.5	227	73.9	0.1615	228	72.1	0.1781
ATPase 8	53	78.2	53	82.7	0.2994	52	77.8	0.2275
ATPase 6	224	74.8	224	75.9	0.1687	225	72.7	0.2147
CO III	261	69.1	268	71.2	0.1624	262	70.2	0.1768
ND 3	117	76.2	117	79.4	0.2183	117	79.7	0.2689
ND 5	572	77.9	574	77.7	0.2119	573	78.1	0.2033
ND 4	445	76.7	446	79.5	0.1962	446	78.5	0.2199
ND 4L	97	80.8	96	83.8	0.1819	96	81.8	0.1954
ND 6	173	81.1	174	84.8	0.2605	174	91.9	0.2971
Cyt b	377	71.2	378	73.9	0.1740	378	73.3	0.1788
ND 1	311	71.6	324	79.0	0.1758	343	76.7	0.1922
Total	–	75.2	–	**77.9**	**0.2006**	–	**76.3**	**0.2124**

Table 5. Number of occurrences and RSCU of the 3707 codons in the protein-coding genes of *C. hominivorax* mtDNA. The anticodon of the corresponding tRNA is shown in brackets. Start and stop codons were excluded from the analysis

Amino acid	Codon	*n*	RSCU	Amino acid	Codon	*n*	RSCU	Amino acid	Codon	*n*	RSCU	Amino acid	Codon	*n*	RSCU
F (Phe)	UUU	281	1.64	S (Ser)	UCU	118	2.42	Y (Tyr)	UAU	145	1.00	C (Cys)	UGU	34	1.00
[UUC]	UUC	62	0.36	[UCA]	UCC	15	0.31	[UAC]	UAC	23	1.00	[UGC]	UGC	3	0.09
L (Leu)	UUA	466	4.64		UCA	93	1.90	*	UAA	–	–	W (Trp)	UGA	88	2.59
[UUA]	UUG	40	0.40		UCG	3	0.06		UAG	–	–	[UGA]	UGG	10	1.43
L (Leu)	CUU	38	1.12	P (Pro)	CCU	74	2.18	H (His)	CAU	49	0.44	R (Arg)	CGU	9	1.00
[CUA]	CUC	4	0.57	[CCA]	CCC	4	1.00	[CAC]	CAC	24	1.00	[CGA]	CGC	0	0
	CUA	55	0.55		CCA	54	0.79	Q (Gln)	CAA	63	0.93		CGA	42	2.00
	CUG	0	0		CCG	2	0.29	[CAA]	CAG	9	1.00		CGG	7	1.00
I (Ile)	AUU	315	2.81	T (Thr)	ACU	105	1.85	N (Asn)	AAU	175	2.66	S (Ser)	AGU	52	1.33
[AUC]	AUC	30	0.26	[ACA]	ACC	11	0.17	[AAC]	AAC	18	2.57	[AGC]	AGC	5	0.10
M (Met)	AUA	203	1.74		ACA	75	1.14	K (Lys)	AAA	63	0.56		AGA	56	1.15
[AUG]	AUG	20	0.35		ACG	2	0.03	[AAG]	AAG	26	0.67		AGG	1	0.14
V (Val)	GUU	81	1.00	A (Ala)	GCU	102	1.79	D (Asp)	GAU	57	1.73	G (Gly)	GGU	40	1.86
[GUA]	GUC	5	0.09	[GCA]	GCC	12	0.36	[GAC]	GAC	11	0.26	[GGA]	GGC	1	0.02
	GUA	103	1.81		GCA	61	1.85	E (Glu)	GAA	74	1.74		GGA	151	2.53
	GUG	6	0.11		GCG	2	0.06	[GAA]	GAG	3	0.14		GGG	27	0.45

calculate distances taking into account differences between transitions and transversions. The nucleotide sequences of the COI and COII genes are among the most conserved among these dipteran species. The range of nucleotide divergence varied from 0.1426 to 0.1448 (for *Drosophila* and *Ceratitis* COI genes) to 0.2994–0.2971 (for *Drosophila* ATPase 8 and *Ceratitis* ND6 genes).

The average distance obtained in the analysis of the protein-coding genes yielded a very similar result for *Drosophila* and *Ceratitis* when compared with *C. hominivorax*. Major divergences among the *Drosophila* and *Ceratitis* genes distances were found for ATPase 8 (0.2994/0.2275), ATPase 6 (0.1687/0.2147) and ND3 (0.2183/0.2689) genes (Table 4).

Codon usage

The codon usage of the *C. hominivorax* mitochondrial genome is shown in Table 5. Comparison with the codon usage of the mosquito *A. quadrimaculatus* and *D. yakuba* revealed very similar patterns. Only two codons, CUG and CGC, were not represented in the *C. hominivorax* mtDNA coding sequences, reflecting the influence of a strong biased codon usage. As reported for other metazoan mtDNA (Wolstenholme, 1992), the most commonly

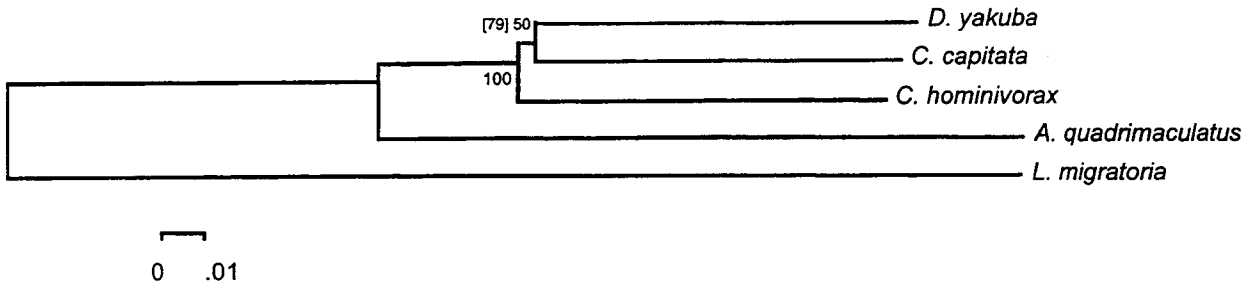


Figure 1. Phylogenetic relationships among Diptera using neighbour-joining analysis (Saitou & Nei, 1987), of aligned amino acid sequences. The bootstrap (1000 replicates) supports are indicated. The value in brackets indicates the bootstrap support for the Acalyptratae node based on a reduced data set analysis (the ATPase 8 and ND6 genes were removed).

used codon in degenerate codon families often does not match the anticodon. The tolerance of mismatches is a common feature in metazoan mtDNA, but its mechanisms have not been conclusively established. There is a significant correlation between codon usage and the nucleotide composition of insect mtDNA, as seen by the increased use of A + T-rich codons (Foster *et al.*, 1997).

Significant variation in nucleotide composition exists in the different mitochondrial genomes, therefore another approach for studying codon usage is to examine the relationships between the base composition of codon families and amino acid occurrence. A simple method that has been used to measure this is to calculate the number of G + C-rich codons (Pro, Ala, Arg and Gly) and A + T-rich codons (Phe, Ile, Met, Tyr, Asn and Lys) and then determine their ratio (Crozier & Crozier, 1993). This value was 0.43 in *C. hominivorax*, which agrees with the dipteran value of 0.42–0.43. The overall amino acid composition of *C. hominivorax* mtDNA resembled that of *Drosophila*, which over-represented the following amino acids: Leu (16.3%), Ile (9.3%), Phe (9.3%), Ser (9.3%), Met (6.0%) and Gly (5.9%); the less abundant being Cys (1.0%), Arg (1.6%), Asp (1.8%) and Gln (1.9%).

The occurrence of C or G at the third codon positions corresponded to only 10% of the 3707 codons analysed. Despite the evolution towards an A + T-rich genome, there was a slight bias, particularly against G nucleotides, which were represented only by 158 occurrences in the third codon position. A mutational tendency against a third codon position with G has been suggested for some metazoan mtDNA (Boore & Brown, 2000).

Phylogenetic perspectives

To reduce eventual inconsistencies resulting from increased variation in individual gene sequences, further phylogenetic analyses were performed with the predicted amino acid sequences rather than with the nucleotide data. Aligned amino acid sequences for dipterans and the locust (as an outgroup) were analysed to provide a general overview of the phylogenetic relationships recovered by mitochondrial coding genes. To minimize eventual homoplasic events and/or ambiguous phylogenetic information provided by inconsistent alignments, we also performed the same phylogenetic analysis with a reduced data set from which the ATPase 8 and ND6 sequences were absent.

The complete data set with sequences from *Cochliomyia*, *Drosophila*, *Ceratitis*, *Anopheles* and *Locusta* was sufficient to provide an appropriate level of resolution for dipteran relationships (Fig. 1), correctly distributing the Acalyptratae species as a monophyletic cluster (*Drosophila* and *Ceratitis*). *Cochliomyia hominivorax* (Calypterate) and the Acalypterate assembled in a typical Brachycera cluster, with *A. quadrimaculatus* (Nematocera) in a basal position in the dipteran topology. The orthopteran outgroup was included to support the interpretation of dipteran phylogenetic relationships. When the ATPase 8 and ND6 sequences were absent, this analysis resulted in the same pattern of relationships, although the bootstrap support for the *Drosophila*/*Ceratitis* node provided a more confident score for this branch (Fig. 1).

Cochliomyia hominivorax *RFLP analysis*

The restriction sites in the *C. hominivorax* sequenced mitochondrial genome were identified to provide a restriction map of this genome for previously identified polymorphic endonucleases. The characterization of Brazilian populations of *C. hominivorax* by RFLP (restriction fragment length polymorphism) of mtDNA, have provided an overall scenario of *C. hominivorax* variability and population structure (Azeredo-Espin, 1993; Infante & Azeredo-Espin, 1995). Figure 2 summarizes the distribution of the diagnostic enzymes along the 'linearized' mtDNA. We identified four *Hae*III restriction sites providing digestion fragments of 7.2, 7.0, 1.0, 0.8 kb, six *Hin*dIII restriction sites with fragments of 6.7, 4.1, 3.6, 0.5 (two) and 0.4 kb, six restriction sites for *Msp*I producing the fragments of 5.0, 4.9, 4.2, 1.4, 0.18 and 0.15 kb and five restriction sites for *Pvu*II responsible for fragments of 8.3, 3.1, 3.0, 0.9 and 0.6 kb.

Based on the restriction patterns produced by each enzyme (B for *Hae*III, A for *Hin*dIII, A for *Msp*I and C for

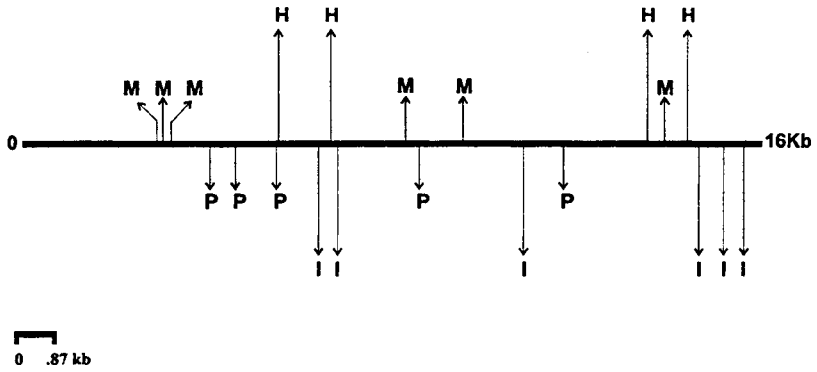


Figure 2. Polymorphic restriction sites mapped on *C. hominivorax* mtDNA based on RFLP analysis (Infante & Azeredo-Espin, 1995). M = *Msp*I, *P* = *Pvu*II, H = *Hae*III and I = *Hin*dIII. The mitochondrial genome is displayed schematically according to the gene order from the control region to the 12S rRNA gene.

*Pvu*II) we were able to characterize the 'mitochondrial haplotype' represented by the sequenced genome. Infante & Azeredo-Espin (1995) described the 'BAAC' haplotype as haplotype 2. This haplotype accounted for about 44% of the individuals analysed in previous work and had a wide distribution over the sampled populations. In a recent analysis using multiple molecular markers, Infante-Malachias (1999) suggested that South America is the probable centre of origin of this ectoparasitic species. Since the mtDNA haplotype 2 is the most frequent in screwworm populations, is geographically the most widespread and occupies a central position in the cladogram distribution, it is a likely candidate for the 'ancestral' haplotype of *C. hominivorax* mtDNA.

The identification of potential regions for PCR-RFLP screening could provide a reliable strategy for immediate access to genetic variation in *C. hominivorax* populations, thereby providing an efficient method for monitoring this species. The COI gene sequences, for example, have three restriction sites for *Msp*I and one site for *Pvu*II, which makes this gene a suitable marker for PCR-RFLP analysis.

Experimental procedures

Fly strains

The *C. hominivorax* used in this study were from a laboratory lineage originally collected at Caraguatatuba, S.P., Brazil. The screwworm larvae were reared and maintained in the laboratory as described elsewhere (Infante & Azeredo-Espin, 1995).

mtDNA extraction

mtDNA was obtained as described by Azeredo-Espin *et al.* (1991). An amount of 60 ng provided enough DNA for the construction of a mitochondrial genomic library (Sambrook *et al.*, 1989).

Sequencing

The complete sequence of the *C. hominivorax* mtDNA was determined by the shotgun approach (Anderson *et al.*, 1982). Intact mtDNA was sheared by sonication, repaired with the Klenow fragment of *E. coli* DNA polymerase I and T4 polynucleotide kinase and fractionated by electrophoresis on low melting agarose. Fragments of 0.8–2.0 kb were purified and inserted into *Sma*I-digested pUC 18 (Amersham Pharmacia Biotech). Plasmid clones were sequenced using BigDye terminator (Perkin-Elmer) chemistry with forward and/or reverse universal primers (Stratagene) in ABI PRISM 377 sequencer (PE Applied Biosystems) following the manufacturers instructions. The amount of clones sequenced was sufficient to provide a five times coverage of the mitochondrial genome sequence. Assembling and final consensus sequences were obtained using the phred/phrap/consed package (Gordon *et al.*, 1998).

Annotation

The software GLIMMER (Salzberg *et al.*, 1998) was used for gene finding in the final sequence. Similarities between the primary nucleotide and translated open reading frame (ORF) sequences and already described genes and proteins were determined with BLAST using the GenBank nonrandom database (Altschul *et al.*, 1997). Possible sequence polymorphisms are indicated as 'variation' in the annotated genome.

Possible conflicts between the ORF results and peculiar features of the mitochondrial transcription mechanism of metazoans meant that the annotation criteria for gene description assigned incomplete termination codons to genes with regions overlapping adjacent tRNA sequences in the same orientation, based on the data reported by Ojala *et al.* (1981). Despite the occurrence of short overlapping sequences involving the genes COI (5 bp), ND2 (2 bp) and Cyt.b (2 bp), we chose to preserve complete termination codons (TAA and TAG) because of the high sequence identity of the hypothetical transcript with several predicted protein sequences described for other dipterans.

Another deviation from the usual approach to determining gene sequences was also adopted for the COI gene. Based on numerous reports of an alternative initiation codon (TCG) for the insects' COI gene, this criterion was also taken into consideration in the annotation of the *C. hominivorax* mitochondrial genome.

The limits of some of the mitochondrial genes described in this work may require modification later as new transcript sequence information emerges.

Comparative analysis

The alignment of homologous mtDNA gene sequences and statistical and phylogenetic analysis of dipteran species were performed using CLUSTALW (Thompson *et al.*, 1994) and the MEGA package (Kumar *et al.*, 1993). The nucleotide distances

used in comparisons of the protein-coding genes of *C. hominivorax*, *D. yakuba* and *C. capitata* was based on the Kimura 2 parameter model (Kimura, 1980). A phylogeny was constructed based on aligned amino acid sequences for *C. hominivorax* and on GenBank sequences for *Drosophila yakuba* (NC_001322), *Ceratitis capitata* (NC_000857), *Anopheles quadrimaculatus* (NC_000875) and *Locusta migratoria* (NC_001712). Inferred amino acid sequences were multiple aligned among all taxa for each of the thirteen genes. Individual alignments were then assembled to provide a single alignment file. Alternatively, a limited data set was designed in which genes providing low confident alignments (ATPase 8 and ND6 genes) were removed. For these multiple alignments, the parameters forming open gap (OGP) and gap extension (GEP) penalties were set at 15 and 6.66, respectively, using the BLOSUM30 matrix, as described by Black & Roehrdanz (1998).

GenBank accession number

The GenBank accession number of the *C. hominivorax* mitochondrial genome is AF260826.

Acknowledgements

The authors thank Dr Christos Louis for providing a copy of the *C. capitata* mtDNA manuscript prior to publication, Rosangela A. Rodrigues, Maria Salete do Couto Campos, Mário C. Filho, Fabiana Kuhne and Kelly Santos for valuable technical assistance, and two anonymous referees for useful comments. This research was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) grant 99/07355–0 and PADCT/CNPq grant 620097/95–7 to A.M.L.A.E. A.C.L. was supported by a fellowship from Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil). Other authors were supported by FAPESP.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein data base search programs. *Nucl Acids Res* **25**: 3389–3402.

Anderson, S., de Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G. (1982) Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian genome. *J Mol Biol* **156**: 683–717.

Azeredo-Espin, A.L.M. (1993) *Mitochondrial DNA Variability in Geographic Populations of Screwworm Fly from Brasil.* International Atomic Energy Agency (IAEA), Vienna, pp. 161–165.

Azeredo-Espin, A.L.M., Schroder, R.F.W., Huettel, M.D., Sheppard, W.S. (1991) Mitochondrial DNA variation in geographic population of Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *Experientia* **47**, 483–485.

Ballard, J.W., Olsen, G.J., Faith, D.P., Odgers, W., Rowell, D.M. and Atkinson, P.W. (1992) Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods. *Science* **258**: 1345–1348.

Battey, J. and Clayton, D.A. (1978) The transcription map of mouse mitochondrial DNA. *Cell* **14**: 143–156.

Beard, C.B., Hamm, D.M. and Collins, F.H. (1993) The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization and comparisons with mitochondrial sequence of other insects. *Insect Mol Biol* **2**: 103–124.

Black, W.C. and 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. *Nucl Acids Res* **27**: 1767–1780.

Boore, J.L. and Brown, W.M. (2000) Mitochondrial genomes of *Galathealinum*, *Helobdella* and *Platynereis*: sequence and gene arrangement comparisons indicates that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. *Mol Biol Evol* **17**: 87–106.

Campbell, N.J.H. and Barker, S.C. (1999) The novel mitochondrial gene arrangement of the cattle tick, *Boophilus microplus*: fivefold tandem repetition of a coding region. *Mol Biol Evol* **16**: 732–740.

Clary, D.O. and Wolstenholme, D.R. (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. *J Mol Evol* **22**: 252–271.

Clary, D.O. and Wolstenholme, D.R. (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.

Coquerel, C. (1859) Note sur une larvae d'oestridae extraite du bras d'um homme à Cayenne. *Rev Mag Zool* **2**: 356.

Crease, T.J. (1999) The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). *Gene* **233**: 89–99.

Crozier, R.H. and Crozier, Y.C. (1993) The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* **133**: 97–117.

Curole, J.P. and Kocher, T.D. (1999) Mitogenomics: digging deeper with complete mitochondrial genomes. *TREE* **14**: 394–398.

Delarbre, C., Barriel, V., Tillers, S., Janvier, P. and Gachelin, G. (1997) The main features of the craniate mitochondrial DNA between the ND1 and the COI genes were established in the common ancestor with the lancelet. *Mol Biol Evol* **14**: 807–813.

Fearnley, I.M. and Walker, J.E. (1986) Two overlapping genes in bovine mitochondrial DNA encode membrane components of ATP synthase. *EMBO J* **5**: 2003–2008.

Flook, P.K., Rowell, C.H.F. and Gellissen, G. (1995) The sequence organization, and evolution of the *Locusta migratoria* mitochondrial genome. *J Mol Evol* **41**: 928–941.

Foster, P.G., Jermiin, L.S. and Hickey, D.A. (1997) Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. *J Mol Evol* **44**: 282–288.

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

Graham, O.H. (1985) Introductions, conclusions. *Misc Publ Entomol Soc Am* **621**: 67.

Guimarães, J.H., Papavero, N. and Prado, A.P. (1983) As miíases na região neotropical (identificação, biologia, bibliografia). *Rev Bras Zool* **1**: 239–416.

Hall, M. and Wall, R. (1995) Myiasis of humans and domestic animals. *Adv Parasitol* **35**: 256–333.

Infante, M.E. and Azeredo-Espin, A.M.L. (1995) Genetic variability in mitocondrial DNA of screwworm, *Cochliomyia hominivorax* (Diptera: Calliphoridae), from Brazil. *Biochem Genet* **33**: 737–756.

Infante-Malachias, M.E. (1999) *Estrutura genética de populações de* Cochliomyia hominivorax *(Diptera: Calliphoridae) da região sudeste do Brasil: análise atraves de três tipos de marcadores genéticos.* PhD Thesis, State University of Campinas.

Jermiin, L.S., Graur, D., Lowe, R.M. and Crozier, R.H. (1994) Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome b genes. *J Mol Evol* **39**: 160–173.

Kimura, M. (1980) A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J Mol Evol* **16**: 111–120.

Kumar, S., Tamura, K. and Nei, M. (1993) *MEGA: Molecular Evolutionary Genetics Analysis, Version 1.01.* The Pennsylvania State University, University Park, Pennsylvania.

Lessinger, A.C. and Azeredo-Espin, A.M.L. (2000) Evolution and structural organisation of mitochondrial DNA control region of myiasis-causing flies. *Med Vet Entomol* **14**: 71–80.

Lewis, D.L., Farr, C.L., Farquhar, A.L. and Kaguni, L.S. (1994) Sequence, organization, and evolution of the A + T region of *Drosophila melanogaster* mitochondrial DNA. *Mol Biol Evol* **11**: 523–538.

Lewis, D.L., Farr, C.L. and Kaguni, L.S. (1995) *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. *Insect Mol Biol* **4**: 263–278.

Litjens, P., Lessinger, A.C. and Azeredo-Espin, A.M.L. (1999) Identification of *Cochliomyia hominivorax* (Coquerel) and *Cochliomyia macellaria* (Fabricius) by PCR-RFLP with mitochondrial DNA. *Genet Mol Biol* **22** (Suppl.): 183.

Lunt, D.H., Zhang, D.-X., Szymura, J.M. and Hewitt, G.M. (1996) The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies. *Insect Mol Biol* **5**: 150–165.

Merten, S.H. and Pardue, M.L. (1981) Mitochondrial DNA in *Drosophila*. An analysis of genome organization and transcription in *Drosophila melanogaster* and *Drosophila virilis*. *J Mol Biol* **153**: 1–21.

Mitchell, S.E., Cockburn, A.F. and Seawright, J.A. (1993) The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. *Genome* **36**: 1058–1073.

Ojala, D., Montoya, J. and Attardi, G. (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**: 470–474.

Roehrdanz, R.L. (1989) Intraspecific genetic variability in mitochondrial DNA of the screwworm fly (*Cochliomyia hominivorax*). *Biochem Genet* **27**: 551–569.

Roehrdanz, R.L. and Johnson, D.A. (1988) Mitochondrial DNA variation among geographical populations of the screwworm fly, *Cochliomyia hominivorax*. *J Med Entomol* **25**: 136–141.

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.

Saccone, C., Pesole, G. and Sbisà, E. (1991) The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. *J Mol Evol* **33**: 83–91.

Saitou, N. and Nei, M. (1987) The neighbor – joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.

Salzberg, S.L., Salzberg, A.J., Kerlavage, A.R. and Tomb, J.F. (1998) Skewed oligomers and origins of replication. *Gene* **217**: 57–67.

Sambook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Spanos, L., Koutroumbas, G., Kotsyfakis, M. and Louis, C. (2000) The mitochondrial genome of the Mediterranean fruitfly, *Ceratitis capitata*. *Insect Mol Biol* **9**: 139–144.

Sperling, F.A.H., Anderson, G.S. and Hickey, D.A. (1994) A DNA-based approach to the identification of insect species used for postmortem interval estimation. *J Forensic Sci* **39**: 418–427.

Spradling, A., Pardue, M.L. and Penman, S. (1977) Messenger RNA in heat-shocked *Drosophila* cells. *J Mol Biol* **109**: 559–587.

Taylor, D.B., Szalanski, A.L. and Peterson, R.D.I.I. (1996) Identification of screwworm species by polymerase chain reaction-restriction fragment length polymorphism. *Med Vet Entomol* **10**: 63–70.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucl Acids Res* **22**: 4673–4680.

Valverde, J.R., Batuecas, B., Moratilla, C., Marco, R. and Garesse, R. (1994) The complete mitochondrial DNA sequence of the crustacean *Artemia franciscana*. *J Mol Evol* **39**: 400–408.

Wells, J.D. and Sperling, F.A.H. (1999) Molecular phylogenetics of *Chrysomya albiceps* and *Chrysomya rufifacies* (Diptera: Calliphoridae). *Med Vet Entomol* **36**: 222–226.

Wolstenholme, D.R. (1992) Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol* **141**: 173–216.

Yokobori, S. and Pääbo, S. (1997) Polyadenylation creates the discriminator nucleotide of chicken mitochondrial tRNA (Tyr). *J Mol Biol* **265**: 95–99.

Zhang, D.-X. and 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.

CAPÍTULO 3

Artigo 3 - Litjens, P., ***Lessinger, A. C.*** & Azeredo-Espin, A. M. L. (2001). **Caracterization of *Cochliomyia hominivorax* and *Cochliomyia macellaria* by PCR-RFLP of mitochondrial DNA.** *Medical and Veterinary Entomology*, **15**, 183 – 188.

Medical and Veterinary Entomology (2001) **15**, 183–188

Characterization of the screwworm flies *Cochliomyia hominivorax* and *Cochliomyia macellaria* by PCR-RFLP of mitochondrial DNA

P. LITJENS*, A. C. LESSINGER† and A. M. LIMA DE AZEREDO-ESPIN*†

*Laboratório de Genética Animal, Universidade Estadual de Campinas, São Paulo, and †Departamento de Genética e Evolução, Universidade Estadual de Campinas, São Paulo, Brazil

Abstract. The primary screwworm fly *Cochliomyia hominivorax* (Coquerel) (Diptera: Calliphoridae) is one of the most important insect pests of livestock in neotropical regions, whereas *Cochliomyia macellaria* (Fabricius) (Diptera: Calliphoridae), the secondary screwworm, is of medical and sanitary importance because of its role in the dissemination of pathogens. These two species share morphological similarities and both may occur in the same myiasis, but in different developmental stages. In this work, the usefulness of PCR-RFLP (polymerase chain reaction – restriction fragment length polymorphism) of mitochondrial DNA (mtDNA) for the unambiguous identification of *C. hominivorax* and *C. macellaria* was investigated. Two specific regions of mtDNA were amplified: 870 bp from Cytochrome oxidase subunit I and 2100 bp from the A+T rich/12S region from *C. hominivorax* and *C. macellaria* specimens from different areas of Brazil. Reliable species-specific PCR-RFLP results were obtained for the CO I region and the A+T rich/12S region using the restriction enzymes *Dra* I and *Ssp* I. These results confirm the conservation of CO I diagnostic restriction sites previously reported and demonstrate the usefulness of the control region sequences as an efficient marker for PCR-RFLP identification of Brazilian screwworm flies. The occurrences of intraspecific polymorphic patterns are discussed based on frequencies and potential conflicts for species identification. PCR-RFLP provides a potentially useful method for identifying samples from the areas where these species are monitored.

Key words. *Cochliomyia hominivorax*, *Cochliomyia macellaria*, molecular identification, molecular markers, mtDNA, PCR-RFLP, screwworm, Brazil.

Introduction

The primary screwworm fly, *Cochliomyia hominivorax* (Coquerel), is one of the most important insect pests of livestock in neotropical regions. The larvae of this species infest open wounds in humans and warm-blooded vertebrates, causing severe myiasis and substantial profit losses for cattle breeders. The distribution of this species initially ranged from the southern United States to northern Argentina, with the highest abundance in neotropical regions (Guimarães *et al.*, 1983).

Cochliomyia hominivorax is morphologically similar to the secondary screwworm, *Cochliomyia macellaria* (Fabricius), a blowfly that feeds primarily on carrion. *Cochliomyia macellaria* is a secondary agent of myiases and a purely saprophagous species. The adults are attracted to a wide variety of substrates for food and reproduction, including urban garbage, human and livestock faeces, carcasses and wounds, whether or not already infested with dipteran larvae (Hall, 1948; Zumpt, 1965; Paraluppi, 1992). Because of its synanthropic behaviour, *C. macellaria* has been reported to be a mechanical vector of human and animal diseases. This species is very common in neotropical America, ranging from

Correspondence: Ana Maria Lima de Azeredo-Espin, Laboratório de Genética Animal, CBMEG – UNICAMP, PO Box 6010, 13083-970 Campinas, São Paulo, Brazil. E-mail: azeredo@unicamp.br

southern Canada to northern Argentina (Ferreira, 1983; Baumgartner & Greenberg, 1985; Dear, 1985).

The geographical overlap between *C. macellaria* and *C. hominivorax*, the morphological similarity, especially between earlier developmental stages, the coexistence of both species in the same infested wound and the potential losses that *C. hominivorax* represents for cattle breeders, means that it is important to investigate reliable new methods for the characterization of these two species (Pomonis, 1989; Taylor & Peterson, 1994). Molecular markers have been used in the investigation of screwworm population genetics (Roehrdanz, 1989; Infante & Azeredo-Espin, 1995; Valle & Azeredo-Espin, 1995; Infante-Malachias *et al.*, 1999; Infante-Malachias, 1999). Mitochondrial DNA provides a suitable marker for studies on genetic variability and molecular characterization. Given its simple and uniform organization, the absence of recombination, maternal inheritance and the high rate of nucleotide sequence evolution, mtDNA is a useful marker for studying molecular evolution and making phylogenetic inferences (Simon *et al.*, 1994).

PCR-RFLP analysis of molecular markers in closely related species has been used to elucidate ambiguous taxonomic status and to identify taxa of importance to forensic science and pest control. The insect groups studied include Diptera (Sperling *et al.*, 1994; Wells & Sperling, 1999), Coleoptera (Tuda *et al.*, 1995; Hidayat *et al.*, 1996; Szalanski & Powers, 1996), Hymenoptera (Taylor *et al.*, 1997) and Lepidoptera (Sperling & Hickey, 1995; Roehrdanz, 1997). Diagnostic PCR-RFLP patterns were previously demonstrated for screwworm species based on the analysis of specific mtDNA regions (Taylor *et al.*, 1996), supporting the potential of this approach for taxonomic identification.

RFLP analysis of the mtDNA of both screwworm species has allowed the characterization of population structure and the levels of genetic variability of each species. Roehrdanz & Johnson (1988) and Roehrdanz (1989) analysed screwworm populations from the United States, Mexico, Jamaica, Guatemala and Costa Rica, and suggested an evolutionary scenario with reduced gene flow, based on the frequency of unique mitochondrial haplotypes. Taylor *et al.* (1991) applied RFLP markers to analyse the introduction of *C. hominivorax* in Libya in an attempt to identify the origin of the infestation.

In Brazil, Azeredo-Espin (1993) and Infante & Azeredo-Espin (1995) detected significant genetic heterogeneity in *C. hominivorax* on mtDNA RFLP studies. However, the mitochondrial haplotype identified in the Libyan infestation was not shared by any of the Brazilian populations analysed. Beesley (1991) suggested that the infestation originated from Uruguay (South America). These analyses indicated that the genetic structure of *C. hominivorax* in American mainland regions was similar and suggested that populations were interconnected but had reduced gene flow.

Based on RFLP of mtDNA, Valle & Azeredo-Espin (1995) and Valle (1997) concluded that Brazilian populations of *C. macellaria* showed genetic discontinuity with spatial separation, a situation similar with the data reported for *C. hominivorax*, but with less intraspecific variability.

In the work reported here, PCR-RFLP of specific mtDNA regions is used to address two main questions. The CO I marker is analysed to corroborate previous results (Taylor *et al.*, 1996), while also including the characterization of widely distributed Brazilian populations of screwworm flies. The usefulness of the mtDNA control region (A + T rich) as an efficient marker for distinguishing *C. hominivorax* and *C. macellaria* samples is also investigated.

Materials and Methods

Fly strain

Seventy-seven *C. hominivorax* samples from different Brazilian localities were analysed: 15 from Alfenas, Minas Gerais State (MG); 15 from Caraguatatuba, São Paulo State (SP); 15 from Manaus, Amazonas State (AM); 15 from Poconé, Mato Grosso State (MT) and 15 from Valinhos, SP. Additionally, two *C. hominivorax* individuals from Uruguay were included in the analysis. Seventy-three *C. macellaria* samples were analysed: 15 from Campinas, SP; 15 from Caraguatatuba, SP; 15 from Manaus, AM; 13 from Porto Urucu, AM and 15 from Rio de Janeiro, Rio de Janeiro State (RJ). Six individuals of *C. macellaria* from Porto Urucu were selected to represent the six mtDNA haplotypes previously described for this locality (Valle, 1997). This selection provided information from a population with significant genetic variability and allowed validation of the screening for diagnostic enzymes. A total of 150 individual flies were analysed in this work.

DNA extraction

Cochliomyia hominivorax and *C. macellaria* samples were derived from previous DNA extractions used in RFLP analysis (Infante & Azeredo-Espin, 1995; Valle & Azeredo-Espin, 1995). The samples were stored at −20°C.

PCR amplification

Two specific mtDNA regions were amplified: 870 bp of cytochrome oxidase I (CO I) and 2100 bp including the complete control region and partial rRNA 12S sequences (A + T rich/12S). The UBC-insect mitochondrial DNA oligonucleotide set, described by Simon *et al.* (1994), was used for CO I amplification using the primers C1 − J − 2195 and L2 − N − 3014; and for A + T rich/12S amplification, the primers TM − N − 193 and SR − J − 14233 were used.

Standard PCR conditions were used in the amplification reactions, which contained 10× PCR buffer (Gibco-BRL, Rockville, MD), 0.2 mM of dNTPs, 1.5 mM $MgCl_2$, 0.5 μM of UBC primers, 1.25 units of *Taq* DNA polymerase (Gibco-BRL), and 1.0 μL of extracted DNA in a final reaction volume of 25 μL. Temperature cycling was done in a Perkin-Elmer (Norwalk, CT) thermal cycler set for an initial denaturing cycle of 3 min at 94°C followed by 35 cycles of 1 min at 94°C, annealing for 1 min at 42–56°C, and a 2-min extension at

Table 1. Diagnostic restriction patterns for the species-specific identification of *C. hominivorax* and *C. macellaria* by PCR-RFLP analysis. * PCR product not digested. A and B indicate polymorphic restriction patterns.

		mtDNA patterns/species	
Region	Enzyme	*C. hominivorax*	*C. macellaria*
CO I			
	Dde I	A 450, 250, 170 B 620, 250	620, 250
	Dra I	870*	520, 350
	Ssp I	740, 130	870*
A+T/12S			
	Dra I	A 800, 350, 220 B 1000, 350, 220	600, 510, 350, 320, 220
	EcoR V	1480, 620	A 2100* B 1450–650
	Ssp I	450, 370, 225	700, 410, 350, 190

72°C. The final cycle was identical except for a 7-min extension step at 72°C. The A+T rich/12S region was amplified as described by Lessinger & Azeredo-Espin, 2000). A lower extension temperature (60°C) was used because of the high A+T composition of this region (Lessinger & Azeredo-Espin, 2000; Lessinger *et al.*, 2000).

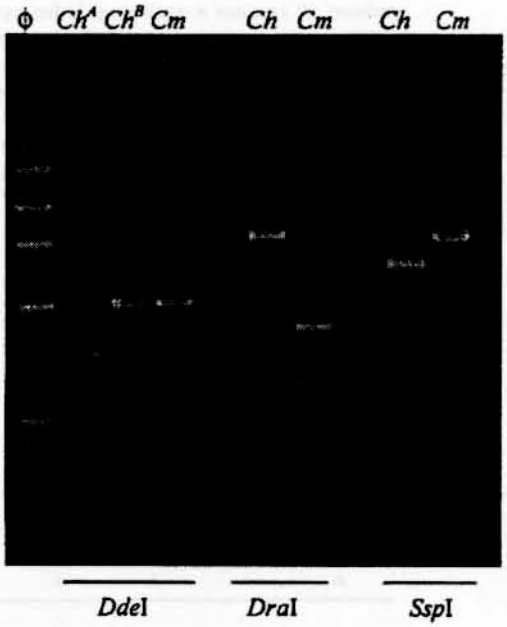


Fig. 1. CO I diagnostic patterns for *C. macellaria* (*Cm*) and *C. hominivorax* (*Ch*). ϕ is the ϕX174/*Hind* III molecular size standard. [A] and [B] indicate polymorphic restriction patterns.

Restriction enzymes

The PCR products of CO I were digested with nine restriction endonucleases: *Bam*H I, *Cla* I, *Dde* I, *Dra* I, *Eco*R I, *Hae* III, *Pvu* II, *Ssp* I and *Taq* I. These enzymes were selected based on previous RFLP analysis and PCR-RFLP results for *C. macellaria* and *C. hominivorax* (Infante & Azeredo-Espin, 1995; Valle & Azeredo-Espin, 1995; Taylor *et al.*, 1996; Valle, 1997). Enzymes yielding intraspecific polymorphism in these previous population genetics studies were avoided.

The A+T rich/12S sequences available for *Cochliomyia* species (Lessinger & Azeredo-Espin, 2000) were analysed using the WebCutter 2.0 software (Heiman, 1997). Thirty-three enzymes were screened using the default parameters. The enzymes yielding diagnostic patterns for species-specific identification were used to provide experimental confirmation of the predicted restriction sites and to test the confidence of the diagnostic results for all of the samples analysed.

Digestions were done as specified by the enzyme suppliers (Gibco-BRL and Pharmacia, Peapack, NJ). The digested fragments were separated by electrophoresis in 2.0% agarose gels, stained with ethidium bromide (EtBr) and photographed on Polaroid film. ϕX174 digested with *Hae* III and a 1-kb LADDER (Gibco-BRL) were used as molecular size markers.

Selection criteria

Two DNA samples of each species were amplified and digested to provide initial screening information on the restriction sites present in the mtDNA regions analysed and on their usefulness for species-specific identification. Enzymes showing at least one restriction site were used in the final analysis with the complete sample set. Enzymes producing distinct restriction patterns for each species were considered diagnostic.

Results and discussion

The diagnostic restriction enzymes and respective mtDNA regions are shown in Table 1. A few inconsistencies were related to the sizes of the originally amplified PCR products and to the total size recovered from the sum of the digested fragments; these findings may reflect the occurrence of an unknown number of small fragments which were unresolved in the agarose gels.

Analysis of predicted restriction sites in the A+T rich/12S region revealed four enzymes (*Dra* I, *Mse* I, *Ssp* I and *Vsp* I) with restriction sites for *C. macellaria* and five (*Dra* I, *Eco*R V, *Mse* I, *Ssp* I and *Vsp* I) for *C. hominivorax*. The enzymes *Vsp* I and *Msp* I provided too many restriction fragments of short sizes, resulting in inaccurate species identification. These enzymes were therefore removed from the final analysis.

The enzymes listed in Table 1 were potential diagnostic markers; however, the occurrence for polymorphic patterns in some populations may represent identification conflicts. Pattern B of the *C. hominivorax* CO I/*Dde* I restriction analysis was similar to the *C. macellaria* pattern for the same analysis (Fig. 1) and restricted the use of *Dde* I for the

Table 2. Haplotypes of geographical samples of *C. hominivorax* and *C. macellaria*. *n* indicates the number of individuals analysed. A and B are species-specific restriction patterns. Monomorphic restriction patterns from Table 1 are indicated as A in this table.

		COI			A+T/12S		
Samples	*n*	*Dde* I	*Dra* I	*Ssp* I	*Dra* I	*Eco*RV	*Ssp* I
C. hominivorax							
ALFENAS – MG	(15)	A	A	A	A	A	A
CARAGUATATUBA – SP	(15)	A	A	A	A	A	A
MANAUS – AM	(11)	A	A	A	A	A	A
	(4)	A	A	A	B	A	A
POCONÉ – MT	(15)	A	A	A	A	A	A
VALINHOS – SP	(15)	B	A	A	A	A	A
URUGUAY	(2)	A	A	A	A	A	A
C. macellaria							
CAMPINAS – SP	(15)	A	A	A	A	A	A
CARAGUATATUBA – SP	(15)	A	A	A	A	A	A
MANAUS – AM	(15)	A	A	A	A	A	A
PORTO URUCU – AM	(12)	A	A	A	A	A	A
	(1)	A	A	A	A	B	A
RIO DE JANEIRO – RJ	(15)	A	A	A	A	A	A

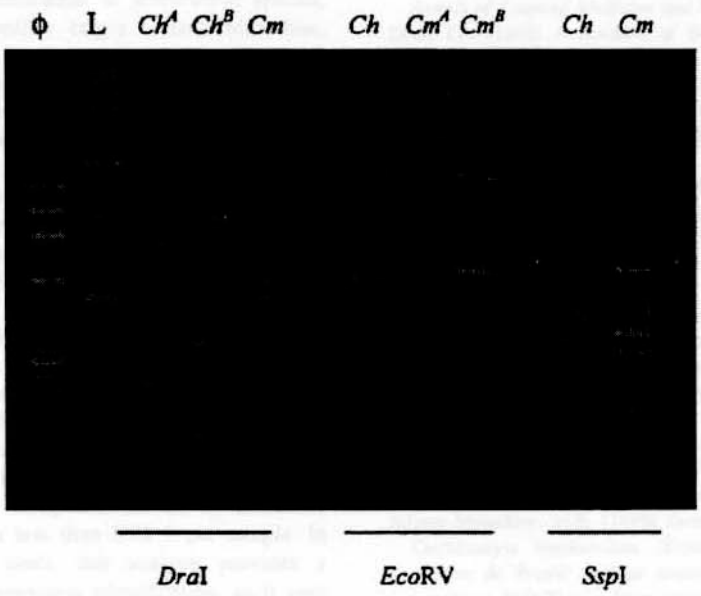


Fig. 2. A+T rich/12S region diagnostic patterns for *C. macellaria* (*Cm*) and *C. hominivorax* (*Ch*). L and ϕ are the 1 kb ladder and ϕX174/*Hind* III molecular size standards, respectively. [A] and [B] indicate polymorphic restriction patterns.

unambiguous identification of these species. The CO I/*Dde* I restriction pattern B was fixed and was exclusive of *C. hominivorax* samples from Valinhos (S. P.) (Table 2).

The *C. macellaria* samples from Porto Urucu (AM), representing six distinct mtDNA haplotypes (Valle, 1997), yielded a different restriction pattern for only one individual when *Eco*R V were used to digest the A+T rich/12S region (Table 2). This pattern differed by one *Eco*R V restriction site and may be a useful marker for future population genetic studies. The polymorphic patterns described above were associated with the variable domain of the control region (Lessinger & Azeredo-Espin, 2000). We suggest that the enzyme *Eco*R V should be avoided in identification analysis because its alternative restriction pattern in *C. macellaria* mtDNA produced ambiguous results (Fig. 2).

The low frequency of polymorphisms was expected because enzymes yielding polymorphic patterns in previous RFLP studies were avoided as an initial criterion. However, the occurrence of polymorphism in a particular species would not always represent an identification conflict: the polymorphic patterns for *Dra* I restriction analysis of *C. hominivorax* A+T rich/12S region provide an efficient identification of screwworm species (Fig. 2 and Table 2).

Taylor *et al.* (1996) analysed *Cochliomyia* species by PCR-RFLP and identified the enzymes *Dra* I and *Ssp* I as diagnostic markers for a 348-bp CO I sequence. This region was previously reported as an informative marker for blowfly identification (Sperling *et al.*, 1994) and is included in the larger 870 bp CO I region analysed in this work. Previous PCR-RFLP analysis included limited Brazilian samples and requires more extensive characterization of this marker to validate its usefulness for Brazilian screwworm populations. Combining the restriction analysis of CO I and A+T rich/12S regions with the enzymes *Dra* I and *Ssp* I, we were able to correctly recognize the species *C. hominivorax* and *C. macellaria*. The characterization of additional diagnostic markers (such as the mtDNA control region) and the analysis of geographically separated Brazilian populations provide a basis for the efficient identification of screwworm species, their distribution and possible origin (Infante-Malachias, 1999).

For practical purposes, individual or combined PCR-RFLP tests could be applied for screwworm identification regarding the relative costs for screening samples. The analysis of an unidentified sample by the amplification of a single mtDNA region (COI or A+T/12S) followed by restriction analysis (*Dra* I or *Ssp* I) could provide accurate identification of screwworm species requiring no further procedures; indeed, morphological and ecological aspects could contribute as additional evidences.

The operational sensitivity of PCR-RFLP tests could be demonstrated based on the consistent reproducibility of the results (based on approximately 1000 digestions with less than 2% revisal) and enzymes relative costs (*Taq* DNA polymerase and restriction endonucleases). Target amplification plus *Dra* I digestion would be less than US$ 1 per sample, while increased confidence levels on species identity by combined PCR-RFLP tests may costs less than US$ 3 per sample. In addition to relative low costs, this analysis provides a convenient strategy for screwworm identification, as it uses universal primers applied in general insect molecular research and requires basic molecular biology devices.

The results described here demonstrate that PCR-RFLP of mtDNA provides a simple and reliable method for identifying *C. hominivorax* and *C. macellaria* samples and for monitoring the expansion and control of these species. This technique may also be useful to study intraspecific genetic variability and could help in determining the origin of *C. hominivorax* samples in (re)colonizations or invasions. Additional population studies may contribute to evaluation of the extent to which of the diagnostic restriction patterns are conserved. However, this work confirms the potential usefulness of these markers for studying Brazilian screwworm populations.

Acknowledgements

The authors thank Dr J. D. Wells for helpful comments on sampling analysis, Rosângela A. Rodrigues, Maria Salete Couto and Mário C. Filho for technical and laboratory assistance and two anonymous referees for their valuable suggestions on the manuscript. P.L. was supported by a scholarship from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), A.C.L. was supported by Coordenadoria de Aperfeiçoamento Pessoal do Ensino Superior (CAPES) and A.M.L.A.E. was supported by FAPESP (grant no. 99/07355-0) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/PADCT).

References

Azeredo-Espin, A.M.L. (1993) Mitochondrial DNA variability in geographic populations of screwworm fly from Brazil. *International Atomic Energy Agency, Vienna*, **327**, 161–165.

Baumgartner, D.L. & Greenberg, B. (1985) Distribution and medical ecology of the blow flies (Diptera: Calliphoridae) of Peru. *Annals of the Entomological Society of America*, **78**, 565–587.

Beesley, W.N. (1991) The New World screwworm fly in North Africa. *Annals of Tropical Medicine and Parasitology*, **85**, 173.

Dear, J.P. (1985) A revision of New World Crysomini (Diptera: Calliphoridae). *Revista Brasileira de Zoologia*, **3**, 109–169.

Ferreira, M.J.M. (1983) Sinantropia de Calliphoridae (Diptera) em Goiânia, Goiás. *Revista Brasileira de Biologia*, **43**, 199–210.

Guimarães, J.H., Papavero, N. & do Prado, A.P. (1983) As miíases na região neotropical (identificação, biologia, bibliografia). *Revista Brasileira de Zoologia*, **1**, 239–416.

Hall, D.G. (1948) *The Blow Flies of North America*. Thomas Say Foundation, Entomological Society of America, Washington, DC.

Heiman, M. (1997) *Software Webcutter 2.0*. www.firstmarket.com/cutter/cut2.html.

Hidayat, F., Phillips, T.W. & French-Constant, H. (1996) Molecular and morphological characters discriminate *Sitophilus oryzae* and *S. zeamais* (Coleoptera: Curculionidae) and confirm reproductive isolation. *Annals of the Entomological Society of America*, **89**, 645–652.

Infante, M.E. & Azeredo-Espin, A.M.L. (1995) Genetic variability in mitochondrial DNA of *Cochliomyia hominivorax* (Diptera: Calliphoridae) from Brazil. *Biochemical Genetics*, **33**, 237–256.

Infante-Malachias, M.E. (1999) *Estrutura genética de populações de* Cochliomyia hominivorax *(Diptera: Calliphoridae) da região sudeste do Brasil: análise através de três tipos de marcadores genéticos*. PhD Thesis, State University of Campinas, Brazil.

Infante-Malachias, M.E., Yotoko, K. & Azeredo-Espin, A.M.L. (1999) Random amplified polymorphic DNA of screwworm fly populations (Diptera: Calliphoridae) from Southeastern Brazil and Northern Argentina. *Genome*, **42**, 772–779.

Lessinger, A.C. & Azeredo-Espin, A.M.L. (2000) Evolution and structural organisation of the mitochondrial DNA control region of myiasis-causing flies. *Medical and Veterinary Entomology*, **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 Molecular Biology*, **9**, 521–529.

Paraluppi, N.D. (1992) *Calliphoridae (Diptera) de Manaus: levanta-*

mento taxonômico, sazonalidade, atividade de vôo, desenvolvimento ovariano e potencialidade como vetores de bactérias nas feiras livres. Doctoral Thesis, INPA, Manaus, AM, Brazil.

Pomonis, J.G. (1989) Cuticular hydrocarbons of the screwworm, *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Journal of Chemical Ecology*, **15**, 2301–2317.

Roehrdanz, R.L. (1989) Intraspecific genetic variability in mitochondrial DNA of the screwworm fly (*Cochliomyia hominivorax*). *Biochemical Genetics*, **27**, 551–569.

Roehrdanz, R.L. (1997) Identification of tobacco budworm and corn earworm (Lepidoptera: Noctuidae) during early developmental stages by polymerase chain reaction and restriction fragment length polymorphism. *Annals of the Entomological Society of America*, **90**, 329–332.

Roehrdanz, R.L. & Johnson, D.A. (1988) Mitochondrial DNA variation among geographical populations of the screwworm fly *Cochliomyia hominivorax*. *Journal of Medical Entomology*, **25**, 136–141.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Sperling, F.A.H., Anderson, G.S. & Hickey, D.A. (1994) DNA-based approach to the identification of insect species used for postmortem interval estimation. *Journal of Forensic Sciences*, **39**, 418–427.

Sperling, F.A.H. & Hickey, D.A. (1995) Amplified mitochondrial DNA as a diagnostic marker for species of conifer-feeding *Choristoneura* (Lepidoptera: Tortricidae). *Canadian Entomologist*, **127**, 277–288.

Szalanski, A.L. & Powers, T.O. (1996) Molecular diagnostics of three *Diabrotica* (Coleoptera: Crysomelidae) pest species. *Journal of Kansas Entomological Society*, **69**, 260–266.

Taylor, D.B., Hammack, L. & Roehrdanz, R.L. (1991) Reproductive compatibility and mitochondrial DNA restriction site analysis of the New World screwworm fly *C. hominivorax* from North Africa and Central America. *Medical and Veterinary Entomology*, **5**, 145.

Taylor, D.B. & Peterson, R.D.I.I. (1994) Population genetics and gene variation in primary and secondary screwworm (Diptera: Calliphoridae). *Annals of the Entomological Society of America*, **87**, 626–633.

Taylor, D.B., Peterson II, R.D., Szalanski, A.L. & Petersen, J.J. (1997) Mitochondrial DNA variation among *Muscidifurax* spp. (Hymenoptera: Pteromalidae), pupal parasitoids of filth flies (Diptera). *Annals of the Entomological Society of America*, **90**, 814–824.

Taylor, D.B., Szalanski, A.L. & Peterson II, R.D. (1996) A polymerase chain reaction – restriction fragment length polymorphism technique for identification of screwworms (Diptera: Calliphoridae). *Medical and Veterinary Entomology*, **10**, 63–70.

Tuda, M., Fukatsu, T. & Shimada, M. (1995) Species differentiation of bruchis beetles (Coleoptera: Bruchidae) analyzed by mitochondrial DNA polymorphism. *Applied Entomological Zoology*, **30**, 377–380.

Valle, J.S. (1997) *Análise da variabilidade genética de* Cochliomyia macellaria *(Diptera: Calliphoridae); avaliação através do DNA mitocondrial e análise cariotípica.* Masters Thesis, State University of Campinas, SP, Brazil.

Valle, J.S. & Azeredo-Espin, A.M.L. (1995) Mitochondrial DNA variation in two Brazilian populations of *Cochliomyia macellaria* (Diptera: Calliphoridae). *Genetics and Molecular Biology*, **18**, 521–526.

Wells, J.D. & Sperling, F.A.H. (1999) Molecular phylogeny of *Chrysomya albiceps* and *C. rufifacies* (Diptera: Calliphoridae). *Journal of Medical Entomology*, **36**, 222–226.

Zumpt, F. (1965) *Myiasis in Man and Animals in the Old World.* Butterworths, London, p. 267.

Accepted 3 February 2001

CAPÍTULO 4

Artigo 4 - *Lessinger, A. C.*, Junqueira, A. C. M., Conte, F. F. & Azeredo-Espin, A. M. L. (*Em preparação*). The occurrence of a duplication in the mtDNA control region of *Chrysomya* species (Diptera: Calliphoridae) involving two tRNAs.

The occurrence of a duplication in the mtDNA control region of *Chrysomya* species (Diptera: Calliphoridae) involving two tRNAs

A. C. Lessinger, A. C. M. Junqueira, F. F. Conte, A. M. L. Azeredo-Espin

Lab. Genética Animal, CBMEG/UNICAMP

Affiliations: DGE/CBMEG, Universidade Estadual de Campinas (UNICAMP),

C. P. 6010, CEP 13083-970, Campinas, SP, Brazil

Corresponding author:

Ana Cláudia Lessinger

Laboratório de Genética Animal, CBMEG, Universidade Estadual de Campinas (UNICAMP),

C.P. 6010, CEP 13083-970, Campinas, S.P., Brazil

Phone: 55 – 19 – 3788-1141

Fax: 55 – 19 – 3788-1089

E-mail: lessinge@unicamp.br

Keywords: *Chrysomya*, mtDNA, control region, duplication, tRNA, rearrangement, molecular evolution

Running head: tRNA duplication in *Chrysomya* control region

ABSTRACT

The mtDNA control regions (CR) of the blowflies *C. albiceps*, *C. megacephala* and *C. putoria* (Calliphoridae) were characterized. The most unusual feature found was the presence of two extra tRNA sequences corresponding to the $tRNA^{Ile}$ and $tRNA^{Gln}$ genes and a 19 bp repeat of adjacent CR sequences. The partially duplicated $tRNA^{Gln}$ may correspond to a pseudogene since most of the sequence of the typical insect $tRNA^{Gln}$ is missing and interspecific comparisons showed a significant degree of sequence divergence among *Chrysomya* species. The $tRNA^{Gln}$ gene showed marked divergence in *C. putoria* and resulted in a specific sequence element with an internal symmetric pattern. In contrast, the additional $tRNA^{Ile}$ gene had a conserved primary sequence following the duplication event and may have represented a functional copy of the gene. Differential selection against duplicated genes appears to have been operating in these species as reflected in the conserved extra $tRNA^{Ile}$ copy and the degeneration of the $tRNA^{Gln}$ sequences. The characterization of CR sequence elements close to the duplication endpoints could provide a better understanding of the mechanism responsible for the duplication event. The potential to form stable stem-and-loop structures was investigated in duplication endpoints because of the common association of secondary structures with duplication events and rearrangements in animal mtDNA. Duplication of the tRNA-coding sequences adjacent to the mtDNA control region was observed, including that of a noncoding sequence repeat. Specific patterns of evolution were assigned to the different sequences identified. These results demonstrated and confirmed the plasticity of the mtDNA molecule in *Chrysomya*, especially for the tRNA genes and adjacent control region sequences. The characterization of the *Chrysomya* CR provides access to variable sequences in the mitochondrial genome which should be useful in the identification of forensically important flies, and could provide data for phylogenetic analyses of closely related species, as well as potential markers for population genetics.

INTRODUCTION

Rearrangements in gene order are well documented in a number of organisms and assumptions on the conservation of animal mitochondrial genomes have to deal with the increasing evidence for a more "flexible" and dynamic molecule (Black IV & Roehrdanz, 1998; Dowton & Austin, 1999; Campbell *et al.*, 1999; Shao *et al.*, 2001a,b). Studies of tRNA positions in animal mitochondrial genomes suggest their accelerated mobility relative to other mitochondrial genes (Saccone *et al.*, 1999). The control region (CR) and adjacent sequences are frequently associated with structural divergence in animal mtDNA, indicating that the driving mechanisms of structural organization may be closely related to errors in the replication process (Zevering, *et al.*, 1991; Macey *et al.*, 1998).

The structural organization of the insect mtDNA control region (or A+T-rich region) consists of a variety of sequence elements with specific patterns of evolution in different species (Zhang & Hewitt, 1997). In *Drosophila*, the occurrence of repetitive or duplicated motifs is usually associated with the control region, considering both variable and conserved domains (Monforte, *et al.*, 1993; Lewis *et al*, 1994). Rearrangements of tRNAs associated with sequences flanking the control region are a common feature of insect mtDNA (Wolstenholme, 1992). The analysis of control region evolution is strongly compromised by the species-specific patterns of structural organization. This lack of structural conservation, even among closely related species, limits the use of universal insect mtDNA primers in amplification reactions and requires special care in the interpretation of homologous elements in comparative analyses.

The blowfly mtDNA control region contains conserved and variable sequences arranged in specific domains (Lessinger & Azeredo-Espin, 2000). The A domain shows multiple conserved sequence blocks when Calliphoridae sequences are aligned. In contrast, the B domain sequences are too variable to allow consistent alignment and validation of homologous sites in myiasis-causing flies. Although the control region sequences of *C. megacephala* have been described (Lessinger & Azeredo-Espin, 2000) there has been no detailed characterization of the repeated region in this species. To address this question, we reviewed the *C. megacephala* sequence (AF151386) and included an analysis of the mtDNA control region of *Chrysomya albiceps* and *Chrysomya putoria*.

A revised analysis of the *Chrysomya* control region identified the $tRNA^{Ile}$ and $tRNA^{Gln}$ gene sequences near to the rRNA 12S end of the control region. The control region sequences of *C. megacephala*, *C. albiceps* and *C. putoria* mtDNA were established in order to confirm the occurrence of duplicated tRNA sequences in these species, and to determine which patterns of molecular evolution were operating in the repeated region, as well as the molecular mechanism that could give rise to this duplication in *Chrysomya*.

RESULTS AND DISCUSSION

Amplification of the Chrysomya control region

Figure 1 schematically shows the structural organization of the *Chrysomya* control region based on the presence of a tRNA repeat adjacent to the 3' end of the control region as predicted by a review of the *C. megacephala* AF151386 sequence. Annealing sites for the universal and *Chrysomya*-specific primers are shown in order to provide a general overview of the possible amplification products generated by alternative PCR reactions. Amplification of the control region sequences with the primers *Cmeg*A and *Cmeg*AR (Table 1) combined with the primers TI-N-24 (Simon *et al.*, 1994) and SR-J-14776 (Lessinger & Azeredo-Espin, 2000), in two independent reactions, allowed recovery of the complete control region sequences in each species (*amplicons* 1 and 2, figure 2). Size variations in the homologous amplified regions were evident for *C. albiceps* relative to the other two species and were partially related to numerous short tandem repeats distributed along the entire region (as poly-Ts, poly-As, $[TA]_n$ and $T[A]_n$ stretches). The amplification of a 300 bp fragment (Table 1 and Figure 2 lane 8) is indirect evidence of the presence of the a tRNA copy and could be used as a control parameter to search for this duplication in other *Chrysomya* and related species. The analysis of amplified control region-specific fragments provided an estimate of the total size for *C. albiceps* (~ 1300 bp), *C. megacephala* and *C. putoria* (~ 1150 bp each), indicating the potential application of control region markers for species-specific identification, especially considering their combination with PCR-RFLP analysis (Litjens *et al.* 2001). The amplification of the 300 bp product for *Chrysomya* was initially misinterpreted as an non-specific PCR product and has now been recognized as a legitimate amplification product.

The usefulness of newly described primers for the efficient amplification of *Chrysomya* control region sequences was confirmed. These primers were also effective in recovering control

region sequences from other forensically important blowflies (unpublished observations) and from preserved museum specimens of myiasis-causing flies (Junqueira *et al.*, *in press*).

Structural organization of the Chrysomya control region

Analysis of the mtDNA control region of *Chrysomya* revealed duplicated tRNA sequences in *C. albiceps*, *C. megacephala* and *C. putoria*. This duplication was conserved among these species, as indicated by amplification analysis of numerous individuals of each species (data not shown). The *amplicon 1* sequences confirmed the identity of both tRNAs genes ($tRNA^{Ile}$ and partial $tRNA^{Gln}$), as well as the initial 19 bp from the 5' end of the control region (adjacent to the original $tRNA^{Ile}$ gene). However, interspecific sequence analyses have suggested that specific sequence elements associated with the duplicated region may be under differential mutation pressure and may have evolved independently since duplication.

The alignment of control region sequences (Figure 3A, B) of *C. albiceps*, *C. megacephala* and *C. putoria* provided an improved interpretation of the *Chrysomya* 'variable' domain B and confirmed the conserved features associated with the control region of the Calliphoridae A domain. Several sequence elements were identified and compared among these species, including the conserved sequence blocks (CSBs) previously described for myiasis-causing flies (Lessinger & Azeredo-Espin, 2000), the predicted secondary structures potentially associated with the origin of mtDNA replication (OR), duplicated sequences, adjacent regions and stem-and-loop-forming sequences at the duplication endpoints. A general overview of the structural organization of the *Chrysomya* control region is shown schematically in Figure 4.

Calliphoridae conserved sequences

All previously described CSBs for the mtDNA control region of myiasis-causing flies were assigned for the *Chrysomya* species analyzed, except for CSB V which corresponds to the annealing site of the *CMeg* primers and, as expected, should be highly conserved in these species. The evolution of the *Chrysomya* CR A domain and specific elements is consistent with the analysis reported for other Calliphoridae and Oestridae, as described in Lessinger & Azeredo-Espin (2000).

The predicted secondary structure for the *Chrysomya* mtDNA replication origin (Figure 5) is conserved as in other Calliphoridae species (data not shown) and its primary sequence included

CSB VI, the largest conserved element identified in the control region sequences of myiasis-causing flies. Some, but not all, of the conserved elements described for the insect control region (Zhang & Hewitt, 1997) could be addressed in the suggested *Chrysomya* OR sequences, including a TATA motif (from 712 to 715 in figure 5), a proximal CG-rich element - GAAGCGTGC (from 682 to 690 in figure 3A) and a prior $[TA]_n$ stretch (from 168 to 210 in figure 3B). Compensatory mutations were identified in *C. albiceps* and *C. putoria* sequences indicating possible functional constraints associated with the maintenance of the secondary structure. These nucleotide substitutions resulted in the replacement of consecutive CG and TA stem-forming pairs in *C. putoria* with two TA pairs in the corresponding *C. albiceps* sequence. The disruption of basal stem pairing in *C. albiceps* because of two adjacent A→G transitions did not prevent the formation of an energetically stable stem-and-loop structure. The same was true for the *C. megacephala* structure, based on the nucleotide divergences shown in figure 5.

Duplicated region and adjacent sequences

The entire duplicated region comprised approximately 120 bp, including 19 bp derived from control region sequences, 68 bp duplicated from the $tRNA^{Ile}$ gene and a 32 bp sequence from a $tRNA^{Gln}$ origin. A BLAST search for sequence similarity for the 71 bp sequence immediately adjacent to the rRNA 12S gene provided no information about the identity or origin of this region. Because of the lack of primary sequence conservation among *Chrysomya* species, and the lack of correspondence with any known sequence described in GenBank, the 71 pb region was assigned as an intergenic region than as a remnant control region sequence. Nevertheless, there are insufficient evidence to validate any candidate origin for these sequences in the analyzed species. In this work, the 90 bp region between the extra $tRNA^{Ile}$ and the rRNA 12S genes is referred to as the *Chrysomya* intergenic region (CIR).

The secondary structure of the duplicated $tRNA^{Ile}$ gene of *Chrysomya* is shown in figure 6A. Two A→G transitions in the *C. megacephala* copy are indicated and represent the only divergent sites among these species. Comparative analysis of the orthologous and paralogous $tRNA^{Ile}$ sequences provided a highly significant nucleotide identity among *Chrysomya* species and the homologous single-copy *C. hominivorax* gene (Figure 4A). The high degree of sequence conservation observed in the $tRNA^{Ile}$ copies contrasted with the variation found in adjacent duplicated sequences. This differential mutation pattern could be indicative of functional

constraints responsible for preserving the original tRNA activity, or could result from the homogenization of gene copies. The duplicated state of the $tRNA^{Ile}$ gene seems to be a stable feature in the *Chrysomya* mitochondrial genome. As suggested by Campbell & Barker (1999), stabilizing selection, in which duplicated genes persist without any adjacent loss of functionality could be invoked to explaining the evolution of $tRNA^{Ile}$ sequences in these species.

The duplicated $tRNA^{Gln}$ sequence in *Chrysomya* is an incomplete or degenerated copy of this gene (pseudogene) as shown in figure 6B, based in the original tRNA cloverleaf structure retrieved from the *C. putoria* mitochondrial genome (AF352790). The repeated sequence corresponded to the tRNA 3' end, and consisted of the acceptor stem, the TΨC and variable loops and two initial nucleotides of the anticodon stem. Numerous nucleotide substitutions were observed among *Chrysomya* $tRNA^{Gln}$ repeated sequences, however accelerated evolution from the original duplicated sequence was only evident in the *C. putoria* repeat, which was significantly divergent, and included a 4 bp insertion and several substitutions. The *C. albiceps* and *C. megacephala* copies agreed with the original paralogous region of the $tRNA^{Gln}$ gene.

The *Chrysomya* duplication endpoints (Figure 7A, B) were examined for the presence of common sequence features that could play a role in duplication. The occurrence of sequences able to produce stable stem-and-loop structures has been described as a feature associated with gene rearrangements and duplications in animal mtDNA (Stanton *et al.*, 1994). Furthermore, the nature of the intergenic sequences surrounding the repeat could reveal aspects of the mechanism of rearrangement (Dowton & Austin, 1999). In particular, the occurrence of tandem or inverted repeats might provide sufficient plasticity to allow a reorganization of the mitochondrial genome. Low complexity sequence elements and simple repeats (such as [TA]n, [TAA]n, [A]n) are abundant motifs in the surrounding repeat area and may be involved in duplication. The initial characterization of endpoint sequences and the possible patterns of evolution for related structures are provided below. However, little is known about the specific role of these elements in the duplication process.

As indicated in figure 7A, partial sequences of the $tRNA^{Ile}$ copy and the 19 bp element could be arranged in an energetically stable stem-and-loop structure that was highly conserved between *C. albiceps* and *C. putoria,* while in *C. megacephala* there was a more divergent pattern, including four transitions in paired sites of the stem formation. The occurrence of four GC pairings in the *Chrysomya* structures (one of which was lost in the *C. megacephala* sequence)

contributed to the stabilization of the stem-and-loop configuration. The inclusion of additional sequence data from different *Chrysomya* species would provide a better identification of possibly regulatory motifs in the mtDNA control region.

As described above, there was a significant divergence between the *C. putoria* sequences for the $tRNA^{Gln}$ pseudogene and adjacent sequences and those of the other *Chrysomya* species. This variation was related to the occurrence of a motif containing a symmetric internal repeat (33 bp element) in the *C. putoria* duplication endpoint (from 335 to 367 in Figure 3A). This element consisted of sequences of the $tRNA^{Gln}$ repeat and of an 18 bp conserved sequence of the *Chrysomya* control region (Figure 4C). Although there was no significant evidence supporting any regulatory importance for this region in the mtDNA, especially considering the rate of accelerated evolution (13 divergent sites, including five transversions and numerous gaps), a stem-and-loop structure could be formed by this *C. putoria* 33 bp element (Figure 7B). The discrepancy between the *C. putoria* copy and theother species, may have resulted from accelerated evolution after duplication and *Chrysomya* divergence.

Evolution of duplication in animal mtDNA

Rearrangements and duplications are a frequent feature of animal mtDNA, occurring independently of the coding or noncoding nature of the duplicated sequences and of the relative position in the mitochondrial genome, although duplication events associated with or adjacent to control region sequences are more frequent (Zhang & Hewitt, 1997; Black IV & Roehrdanz, 1998; Saccone *et al.*, 1999; Dotson & Beard, 2001).

The resulting copies of duplication events have been characterized as complete genes, pseudogenes and repeated elements in the control region (or even the entire CR). Most definitions concerning functional/non-functional assumptions for gene copies are based on predicted transcribed sequences or inferred secondary structures of the duplicated regions. Recently, direct experimental evidence of the expressed gene product lead to determination of the non-functional nature of a previously misinterpreted duplicated gene in *Mytilus* (Beagley *et al.*, 1999). The occurrence of duplicated regions involving complete, partial or specific elements of control region sequences has frequently been reported for the arthropod mitochondrial genome (Rand & Harrison, 1989; Zhang & Hewitt, 1997; Black IV & Roehrdanz, 1998; Dotson & Beard, 2001; Nardi *et al.*, 2001).

The presence of pseudogenes, mainly tRNAs, in the animal mitochondrial genome may represent a possible transient or intermediate state from the degeneration of duplicated regions because of the lack of functional constraints on redundant copies (Moritz, 1987; Macey, 1998; Kumazawa *et al.*, 1998; Campbell & Barker, 1999; Beagley *et al.*, 1999; Eberhard *et al.*, 2001; Dorner *et al.*, 2001). The degeneration of original or recently duplicated sequences occurs randomly and contributes to the rearrangement process (Macey, 1997).

Duplication of mitochondrial coding regions other than those for tRNAs, has been reported only for the protein-coding ND6 gene in parrots (Eberhard *et al.*, 2001) and the ND1 gene in *Boophilus* ticks (Campbell & Braker, 1999), both of which represent non-functional copies. The prediction of functional duplicated coding regions in animal mtDNA, as suggested for *Chrysomya* $tRNA^{Ile}$, is restricted to the tRNA genes in *Mytilus* (Hoffman, 1992; Beagley *et al.*, 1999), in an anphisbaenian (Macey, 1998) and in snakes (Kumazawa, 1998), and correspond to the $tRNA^{Met}$, $tRNA^{Pro}$ and $tRNA^{Phe}$ genes, respectively. Recently, Hwang *et al.* (2001) re-examined non-annotated regions in published mitochondrial genome sequences and identified a "coding probability" for a second $tRNA^{Tyr}$ in *Lumbricus terrestris*.

The duplicated copies of coding sequences which are free from functional constraints are prone to high levels of sequence divergence, even among closely related species. The maintenance of conserved copies of coding or non-coding regions in the animal mitochondrial genome is a challenging subject that is raising new questions about the strength and nature of stabilizing selection in mitochondrial genomes (Kumazawa *et al.*, 1998; Black IV & Roehrdanz, 1998; Campbell & Barker, 1999; Eberhard *et al.*, 2001).

Dowton & Austin (1999) proposed a correlation between mtDNA rearrangements and the evolution of a parasitic lifestyle in hymenopterans. An extension of this association, including the mitochondrial dynamics of rearrangements in obligate ectoparasites such as ticks, wallaby lice, mosquitoes and others, suggests that parasitic lineages could be prone to mtDNA rearrangements as a result of selective pressure from the host-parasite interaction (Shao *et al.*, 2001a; Dowton & Campbell, 2001). The Calliphoridae contains many species of myiasis-causing flies that are obligate ectoparasites of great medical and veterinary importance, including species of the genera *Cochliomyia* (the New World screwworm), *Chrysomya* (the Old World screwworm) and *Lucilia* (the sheep blowfly). The structural organization of the mtDNA of *C. hominivorax*, a primary agent of myiasis, resembles that proposed for *Drosophila* species. Moreover, no correlation was

found with the initial levels of structural divergence in the mitochondrial genome of *Chrysomya* since the species analyzed were all primarily free-living organisms. The possibility of multiple evolution as pathways for the parasitic habit from more generalist species in this group (Wall & Hall, 1995) apparently did not directly affect the Calliphoridae mitochondrial gene order.

Mechanisms of structural organization in animal mtDNA

Duplications of mitochondrial sequences may represent an intermediate state for mtDNA with further structural organization produced by random the loss of duplicated sequences and relaxed selection on redundant copies (Macey, 1997). This duplication/random loss model has been used to explain numerous rearrangements among mitochondrial genomes of related groups (Moritz *et al.*, 1987; Macey, 1997, 1998). On the other hand, various examples of conserved duplicated functional or nonfunctional regions have been interpreted as evidence for gene conversion or concerted evolution (Zhang *et al.*, 1995; Zhang & Hewitt, 1997; Black IV & Roehrdanz, 1998; Eberhard *et al.*, 2001). Intramitochondrial recombination has also been suggested as a legitimate mechanism for mediating some specific rearrangements in animal mtDNA (Dowton & Campbell, 2001), in addition to other frequently described mechanisms such as slippage-strand-mismatch, tRNA mediated integration, replication errors, stem-and-loop regulatory structures and direct/inverted repeats interactions (Levinson & Gutman, 1987; Stanton, 1994; Moritz, 1987; Cantatore *et al.*, 1987).

The most reasonable explanation for the present data is that the duplication process in *Chrysomya* mtDNA was mediated by stem-and-loop structures (Saccone *et al.*, 1999), since one duplication endpoint was within the $tRNA^{Gln}$ sequence while the other lay near a stable secondary structure formed by $tRNA^{Ile}$ and 19 bp-element sequences (Figure 7A). The mechanism whereby site-specific integration occurs remains unknown (Stanton, 1994). An alternative hypothesis would involve a tandem repeat duplication process via a slippage-strand mismatch or mispairing (SSM) mechanism in which the secondary structures may have played a role in stabilizing a single-strand region implicated in the slippage during replication. If an SSM event did occur, a drastic reorganization of the duplicated control region sequences must have taken place in order to produce the arrangement observed in the *Chrysomya* control region sequences, including extensive deletions and degeneration of CR-like duplicated sequences, despite the preservation of the $tRNA^{Ile}$ copy.

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

EXPERIMENTAL PROCEDURES

Chrysomya samples

Adults of *C. albiceps* (Wiedemann, 1819) were collected in Rio de Janeiro, Rio de Janeiro State, Brazil, and stored at -70°C. Samples of *C. putoria* (Wiedemann, 1830) or *C. chloropyga* (synonym), from Pirassununga, São Paulo State, Brazil, were reared as a laboratory strain, and the pupae were stored at –70°C prior to analysis. The *C. megacephala* (Fabricius, 1794) used for comparative analysis (AF151386) was form Adamantina, São Paulo State, Brazil.

Genomic extraction

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

Amplification reactions

The PCR reactions were done as described by Lessinger & Azeredo-Espin (2000). The primers used in the amplification reactions of control region sequences for *C. albicpes*, *C. megacephala* and *C. putoria* are described in table 1. The primers *CMeg A* and *CMeg AR* were based on the C. *megacephala* control region sequence (AF151386) and recognized complementary annealing sites with opposite orientations in the mtDNA molecule (Figure 1). The complete sequences of the *Chrysomya* control region were amplified in two independent reactions (products are referred to as *amplicon*s 1 and 2) to avoid competition between alternative annealing sites for the primer TI-N-24 in the duplicated $tRNA^{Ile}$ sequence. This PCR reaction preferentially amplified a short 300 bp product instead of the expected 1200 bp (approximately) control region product (Figure 1).

Cloning and sequencing

The amplified PCR products were purified using 0.05 μm filters as described in Lessinger & Azeredo-Espin (2000). An aliquot of the purified PCR product was cloned into the pUC18 – Sure Clone system (Amersham Pharmacia Biotech) according to the manufacturer's specification. Two to four clones from independent PCR reactions were sequenced for each species in both strands.

Comparative and structural analysis

The CR nucleotide sequences of *C. albiceps*, *C. putoria* and *C. megacephala* (AF151386) species were aligned using Clustal W (Thompson *et al.*, 1994) set to default parameters with manual adjustments where required. The identification of conserved sequence blocks was based on the description by Lessinger & Azeredo-Espin (2000) for blowflies. The tRNA secondary structure was determined using tRNAscan-SE software (Lowe & Eddy, 1997). Other secondary structures and their thermodynamic properties were determined using Mulfold software (SantaLucia, 1998). The *C. hominivorax* (AF260826) and *C. putoria* (AF352790) mitochondrial sequences were used as additional parameters for comparison to provide information on the primary sequences and predicted secondary structures of tRNA.

General definitions

Homology was defined based on sequence similarity, relative position and structural organization of the mtDNA control region elements in *Chrysomya* species. The definition of homology in non-coding and/or regulatory sequences is complex and should be addressed carefully as indicated by Fitch (2000) and Mindell & Meyer (2001). The observation that shared variation in mitochondrial gene organization represents highly reliable evidence of a common ancestry (Boore *et al.*, 1998) supports the assumptions of orthology and paralogy defined for the *Chrysomya* mtDNA duplicated sequences in this work. Note that the term 'original' in the text refers to homologous genes and mtDNA sequences in the corresponding *C. hominivorax* and *C. putoria* mtDNA structural organization (Lessinger *et al.*, 2000; Junqueira *et al.*, in preparation).

Duplicated sequences are described as a "copy", "repeat", "duplication" or related term. An approximately 90 bp sequence, immediately adjacent to the rRNA 12S gene, is suggest to be recognized as a "*Chrysomya* intergenic region" (CIR) based on its relative position between the extra $tRNA^{Ile}$ and rRNA 12S genes and on the fact that it shares no significant primary sequence

identity with any known control region element. The further characterization of *Chrysomya* CR sequences could provide additional evidence for a better understanding of the origin and evolution of this noncoding intergenic region.

Duplication endpoints were identified by aligning the sequences flanking the internal duplication junction with the corresponding original sequences. However, their exact boundaries could not be assigned precisely because of the high sequence variability at the terminations of the duplicated region.

Functional and non-functional interpretations were based in the analysis of primary sequence elements and predicted secondary structures, and need to be confirmed by further experiments. The discussion of mitochondrial duplicated sequences and pseudogenes was limited to the structural organization of the organellar genome. The occurrence of complete mitochondrial genomes, as well as partial sequences and pseudogenes that have been found transferred to the nuclear genome in numerous organisms (Bensasson *et al.*, 2000 and Willians & Knoulton, 2001) was not addressed here.

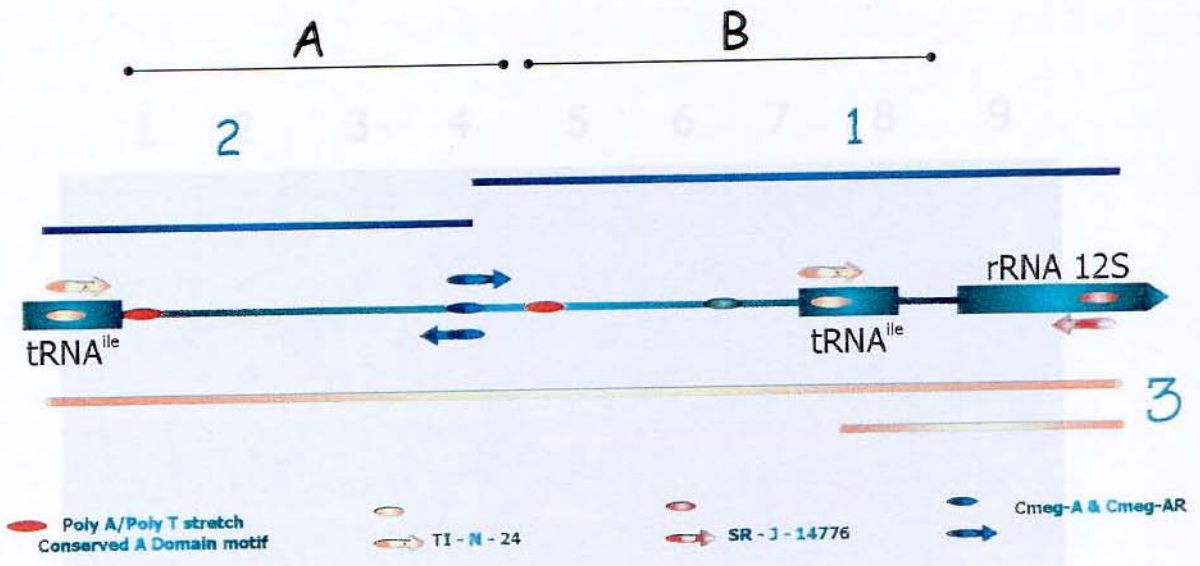


Figure 1 - Schematic organization of the *Chrysomya* mtDNA control region and adjacent genes (tRNA^{Ile} and rRNA 12S) indicating the primer sites and related PCR amplification products. Amplification results are indicated for reactions 1 to 3. Numbers 1 and 2 indicate *Amplicons* 1 and 2. Two PCR products (approximately 300 bp and 1200 bp) are potentially amplified using a universal primer set (3). A and B indicate control region-specific domains (Lessinger & Azeredo-Espin, 2000). Experimental results are shown in Figure 2.

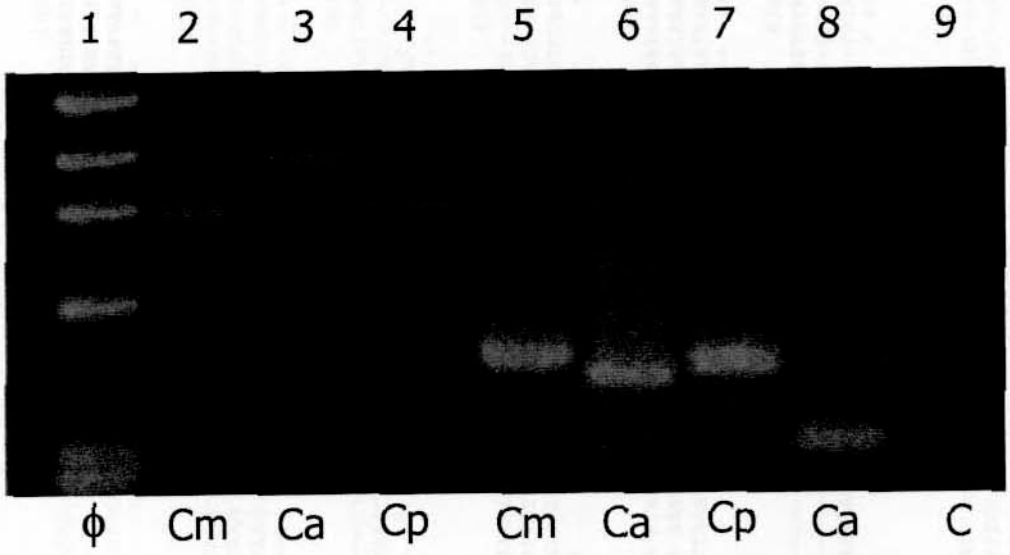


Figure 2 - Control region amplification of *Chrysomya* species using universal (Simon *et al.*, 1994) and *Chrysomya*-specific primers. Cm = *C. megacephala*, Ca = *C. albiceps* and Cp = *C. putoria*. C = PCR control reaction (no DNA). Lanes 2 to 4 and 5 to 7 are related to PCR reactions 1 and 2 (see Figure 1). These PCR products are referred to as *amplicons* 1 and 2 in the text. Lane 8 is a 300 bp product produced by the primer TI-N-24 annealing at the tRNA copy (3 in Figure 1). This product predominates in the PCR reaction, overcoming a potentially amplifiable 1200 bp product. The molecular marker Φ X 174 is indicated in lane 1

FIGURE 3

A

```
                1              15 16              30 31              45 46              60 61              75 76              90
1 C.putoria     ATACTCTTTAGTATT ACTATTTCTAGGTTT CCCTAATTAATAATA TTAATTACTGCGAAT AAAATAAATTTATTT ATTATTAAAATAAAT      90
2 C.megacephala ATACTCTTTAGTATT ACTATTTCTAGGTTT CCCTAATTAATAATA TTAATTACTGCGAAT AAAAATAATTTATTT ATTATTAAAATAAAT      90
3 C.albiceps    ATACTCTTTAGTATT ACTATTTCTACGTTT CCCTAATTAATAATA TTAATTACTGCGAAT AAAAATAATTTATTC ATTATTAAAATAAAT      90
                *************** *************** *************** *************** **** ******** ***************

                91            105 106            120 121            135 136            150 151            165 166            180
1 C.putoria     AAAAATTCACACAAA AATTTACATATAAAT TAAACTAATAATAAA TTTACAAGCCAAAAT AAAACTTTATACAAA AAAACTAAAATAAAT     180
2 C.megacephala AAAAATTCACACAAA AATTTACATATAAAT TAAACTAATAATAAA CTTACTAGCCAAAAT AAAACTTTATACAAT AAAATTAAAAATAAT     180
3 C.albiceps    AAAAATTCACACAAA AATTTACATATAAAT TAAACTAATAATAAA TTTACAAGCCAAAAT AAAACTTTATAAAAT AAAATTAAAAAATAA     180
                *************** *************** *************** **** ********** **********  ** **** ***** *

                181           195 196            210 211            225 226            240 241            255 256            270
1 C.putoria     AACAAAATTAATTAT TTATATTATTA-TTT ATAAATAAATATAAT TAGTAAATAATGGGA ATCCTACTTTTAATG AATTGCCTGATAAAA     269
2 C.megacephala AATAAAAATAAATTT TTAATATATTT-ATA ATAAATAAATAAAAT TAGTAAACAATGGAA ATCTTACTCTTAATG AATTGCCTGATGAAA     268
3 C.albiceps    AACTAAA-TAAATAA ACAAAAAATTACAAT ATTAATAAAAAAAAA TAGTAA-CGATGGAA ATCCTACTTTTAATG AATTGCCTGATAAAA     268
                **  *** *** *     *    ***       ** ****** * **  ******  **** * *** **** ****** ********** ***

                271           285 286            300 301            315 316            330 331            345 346            360
1 C.putoria     AGGATTACCTTGATA GGGTAAATCATGTAA TAATATTACATTCAT TA----TTATATTTA ATAGAATTAATAAAT TAAATCCTAAATTAA     355
2 C.megacephala AGGGTTACCTTGATA GGGTAAATCATGTAA TAATATTACATTCAT TATTAATTATATTTA ATAGAATTAAACTAT TT----CTAAAT---     350
3 C.albiceps    AGGATTACCTTGATA GGGTAAATCATGTAA TAATATTACATTCAT TAT--ATTATATTTA ATAGAATTAAACTAT TT----CTAAAT---     348
                *** ********** *************** *************** **    ********* **********  ** *    ******

                361           375 376            390 391            405 406            420 421            435 436            450
1 C.putoria     TAATTAAAAATTGGT AATTAATCCCATGTT ACAACCAATTTTA-A TGTAAAAATTTACAT AAATATATATATAT- AATTCCCCTAT---T     440
2 C.megacephala -AATTATAAATTGGT AAATACTCC--TGTT ACAAATAAAATACCA TGTATTATTTTATAT GCATAT---TATAT- TAT-CCCCTACA-AT     432
3 C.albiceps    -AATTAAAAATTGGT AATT---CG--TGAC GCAACTATAACAT-A TGTATATATATACCC CCATATATATATACA TATATATATATATAT//     432
                 ***** ******** ** *   *   **    *** *       * **** * **      ****  ****   **     **    *

                451           465 466            480 481            495 496            510 511            525 526            540
1 C.putoria     AATATA-ATACATAT ATTTA------ATAT TAATTTAT-AATATT AATA-ATAAAAAATG ATATAATTATAAATA A-TAATAGTAATAAA     520
2 C.megacephala AATATATATATATAT ATTTAT-----ACAT TTATTAAA-AAT-TA AATA-AAATAAAACA TAAAAATTAAAAATA A-TAAAAATAACAAA     513
3 C.albiceps    AATTTT-ATAAAAAA TTTTAAAAATAATAT TAATTAAATAATATT AATTTAAATTAATAA TAATAATTATAATTA GGTAATAATAATTAA     521
                *** *  *** * *   ****     * ** * *** *  *** *  ***  * *  **      * ***** ** **   *** * ***  **

                541           555 556            570 571            585 586            600 601            615 616            630
1 C.putoria     AATTAGTAATATTTA TTATTTATTTGTATG ATAAATCATCTCCTA AT----------TTT TTTTTTTTTTTTTTT ATATATACAATTTTA     600
2 C.megacephala AATTAGTATAATTTA TTAATT-----TATA ATTGATTTT-TCATA GT----------TTT TTTTTTTTTTTTTTT AT-TTTA----TATA     582
3 C.albiceps    TAT--ATATAATTTA AAAATAAATTTAATT ATAAAGTTTATATTA ATAAAATAGGGGTTT TTTTTTTTTTTTTTT TTTTTA----TATA     605
                 **   **  *****   * *       **  **  *   * *  **  *           *** ***************  * * **    * **
                                                                                                  CSB VII

                631           645 646            660 661            675 676            690 691            705 706            720
1 C.putoria     CAAAAAAGTTAATAT ATAATTCATTTAAAT ATACCTCTCCAATTT TATATTGAAGCGTGC AAAAATTTATTTTTG GTC-AATATATT-AA     688
2 C.megacephala TAAAAAA--TAATAT ATAATTCATTTAAAT ATACCTCTCCAATTT TATATTGAAGCGTGC AAAAATTTATTTTTG GTC-TATATACC-AA     668
3 C.albiceps    TATATAT---A-TAT ATATATTAATTAAAA AGAATATTAATTTTT TATATTGAAGCGTGC AAAAATTTATTTTTG ATCCTATATACCCAA     691
                 * * *    * *** ***  * * *****  * *    *   *** ***************** ***************  **  *****  **
                                                                                  CSB VII

                721           735 736            750 751            765 766            780 781            795 796            810
1 C.putoria     ATAAAATTAATAGAT ATCTATTAATATATA AATATTTAATTAATT AATAGATATCTATTA ATTTAGAGCAAAAAA AAA-TTTCA     771
2 C.megacephala ATAAAATTAATAGAT ATCTATTAATATATA AATATTTAATTAATT AATATATATCTATTA ATTTAGAGTAAAAAA AAAATTT-A     751
3 C.albiceps    ATAGGATTAATAGAT AAATATTAATATATA AATATTTAATTAATT AATATTTATCTATTA ATTTAGACTAAAAAA AAA-TTTCA     774
***  ********** *  ************ *************** **** ********* *******
CSB VI
```

FIGURE 3

B

```
                1           15 16          30 31          45 46          60 61          75 76          90
1 C.albiceps    GTTATTAAAACTTTA TTAATAATGATTTAT ATAATAATTATGTAG AAATTAATT-TATTA ATAAAATTATTTATA AATGTATATGAATTA      89
2 C.putoria     GTTATTAAAACTTTA TTAATAATGATTTAT ATAATAATTATGTAG AAATTAATT-TATTA ATAAAATTATTTATA AATGTATATGAATTA      89
3 C.megacephala TTAATTAAATATTTA TTAATAATGATTTAT CTAATTATTATGTAG AAATTGATTGTTTTA AAAAAATAATTTAAT AATGTATGTGAATTA      90
                * ******  **** ***************  **** ********* ***** *** * *** * ***** *****   ******* *******
                  CSB IV                                                                                  →

                91          105 106         120 121         135 136         150 151         165 166         180
1 C.albiceps    TATTTATAATTTATA ATAAATAATTTATTA ATTATTAAATTTATA TATATT-TATTCATA ATTTAATAAATAATA AATATATGAATATAT    178
2 C.putoria     TATTTATAATTTATA ATAAATAATTTATTA ATTATTAAATTTATA TATATT-TATTCATA ATTTAATAAATAATA AATATATGAATATAT    178
3 C.megacephala TATTTATAATTTTAT ATAAATAATTTATTA ATTATTAATTTTATA -ACATTATATTCATA ATTTAATATATTATA AATATATGAATAAAT    179
                ************    *************** ******** ******  * *** ******** ******** ** *** ************ **
                ←──────────────────────CSB III──────────────────●  ●──────────────────CSB II──────────────────→

                181         195 196         210 211         225 226         240 241         255 256         270
1 C.albiceps    ATATATA---TATAT AT--TTATATATAT- CACAATATAATAATA ACAATATAATACT-- --AATTT-------- --TTATTATTTAAAT    248
2 C.putoria     ATATATA---TATAT ATATTTATATATAT- CACAATATAATAATA ACAATATAATACT-- --AATTT-------- --TTATTATTTAAAT    250
3 C.megacephala ATATATAAATTATAT ATATATATATATATG CAATTTTTAATTATC TAAATTTATTTTTTT TCAATTTATAAACTA ATTTACTATTTATAA    269
                *******   ***** **   *********  **      *** **    *** ** * *    *****           *** ****** *
                ←──────────────────────CSB II──────────────────●

                271         285 286         300 301         315 316         330 331         345 346         360
1 C.albiceps    AAATTAATTATTAAA CTAAATAAT------ ----TTAATAATTCT TAAAAAAATAAGTAT AAAGAAAAA--AAAA AAAAAAAAAAAACCG    326
2 C.putoria     AAATTAATTATTAAA CTAAATAAT------ ----TTAATAATTCT TAAAAAAATAAGTAT AAAGAAAAA--AAAA AAAAAAAAAA--CCG    326
3 C.megacephala TTATTAATATTTAAA TTTTATAATGTTATT TTATTTAATAAATCA CAAAAAAATTACTAT AAAATAAAACAAAAA AAAAAAAAAAAAAAC    359
                  ******  *****  *  *****          ******* **   ******** * *** ***  ****  **** **********
                                                                ●──────────────────CSB I──────────────────●

                361         375 376         390 391
1 C.albiceps    ATGGAAATCCTACTT TTAATGAATTGCCTG ATAAAAAGGGA    367
2 C.putoria     ATGGAAATCCTACTT TTAATGAATTGCCTG ATAAAAAGGGA    367
3 C.megacephala ATGGGAATCTTACTC TTAATGAATTGCCTG ATGAAAAGGGT    400
                **** **** ****  **************  ** *******
```

Figure 3 - Alignment of the mtDNA control region of *Chrysomya* species. **A** – Alignment of control region *amplicon* 1 that contains partial sequences from rRNA 12S (1 - 160), the duplicated region (233 - 357), and partial sequences from the Calliphoridae conserved domain A (598 - 803). A red arrow identifies the alternative annealing site of the primer TN-I-24 (table 1). Boxes identify sequences capable of forming stable stem–and-loop structures. Primary sequence elements identified in the alignment were the 19 bp repeat (233 - 251), the $tRNA^{Ile}$ copy (252 - 317), the $tRNA^{Gln}$ pseudogene (322 - 357) and the 18 bp-conserved sequence (358 - 379). The *C. putoria* 33 bp-symmetric element spanned from 335 to 367. The // indicates variable size for *C. albiceps* due to $[AT]_n$ extension. **B** – Alignment of control region *amplicon* 2 with the identification of "conserved sequence blocks" (CSBs) described in Lessinger & Azeredo-Espin (2000). The original $tRNA^{Ile}$ sequences adjacent to the CR ranged from 378 to 401. See figure 4.

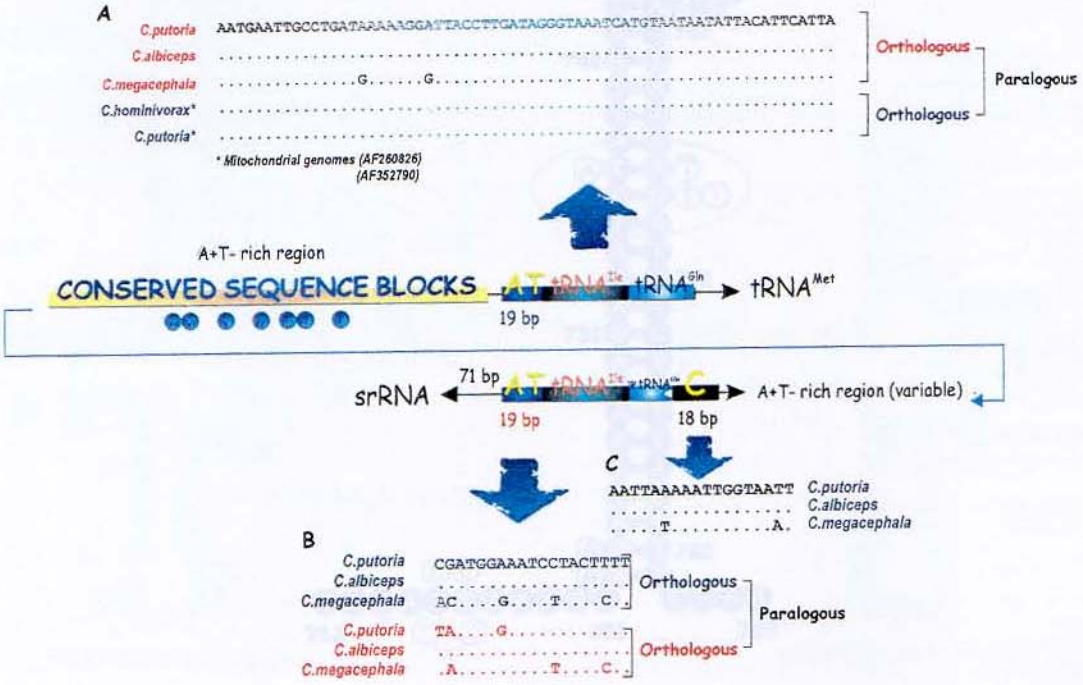


Figure 4 - Structural organization of duplicated elements in the *Chrysomya* control region. Comparative analyses of homologous regions indicate different levels of conservation among these elements (A, B and C). **A** – Alignment of *Chrysomya* orthologous and paralogous $tRNA^{Ile}$ genes, including the *Cochliomyia* homologous sequence as an additional parameter. **B** – Alignment of the 19 bp element (repeated CR sequence) presenting initial levels of nucleotide divergence between copies. **C** – Alignment of the conserved 18 bp element. A $tRNA^{Gln}$ pseudogene was identified in the duplicated region followed by the conserved 18 bp sequence. An approximately 90 bp region, including 71 bp immediately adjacent to the rRNA 12S gene plus the 19 bp repeated element, is referred to as the *Chrysomya* intergenic region (CIR) in the text.

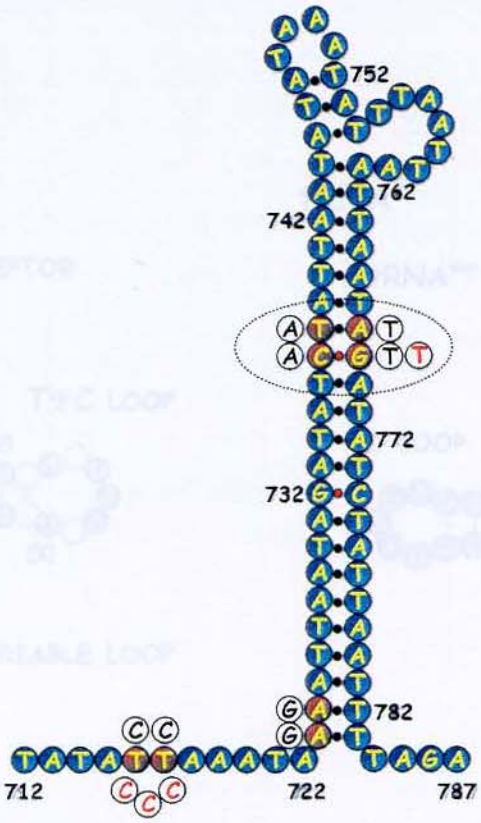


Figure 5 – The predicted secondary structure formed by the A domain control region sequences in *Chrysomya.* This conserved structure is energetically stable (free energy values vary from –12.2 to –15.2 in these species) and represents the potential origin of replication (OR) of the *Chrysomya* mtDNA (blowfly CSB VI). Nucleotide positions from 712 to 787 correspond to the relative positions of alignment in Figure 3A. Conserved nucleotides among *Chrysomya* species are shown in blue circles; red circles indicate nucleotide divergence in the corresponding *C. putoria* sequence; white circles with red or black characters indicate substitutions in the *C. megacephala* and *C. albiceps* sequences, respectively. Compensatory nucleotide substitutions resulting in the replacement of consecutive CG and TA pairs in *C. putoria* for two AT pairs in *C. albiceps* indicated by the dotted oval.

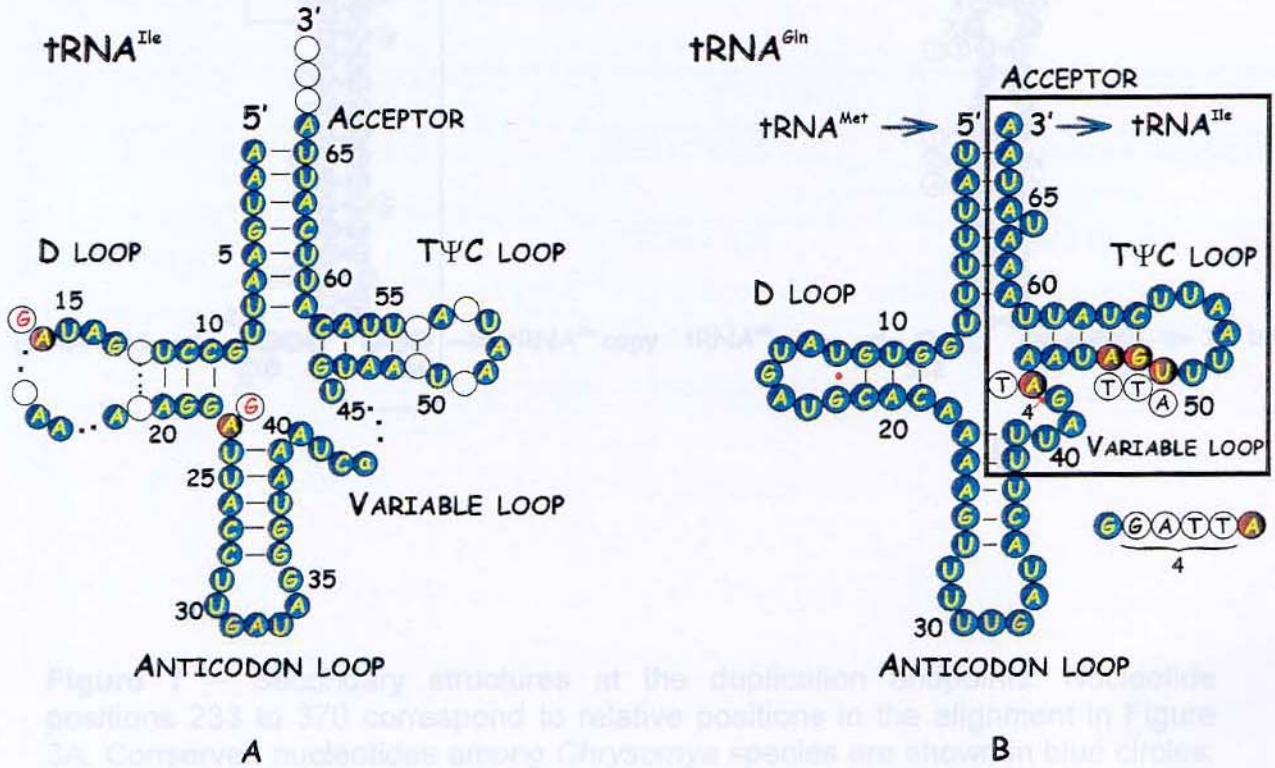


Figure 6 – Secondary structure of duplicated tRNA sequences. Conserved nucleotides among *Chrysomya* species are shown in blue circles; red circles indicate nucleotide divergence in the *C. putoria* sequence. **A** – "Clover-leaf" configuration predicted for the duplicated *Chrysomya* $tRNA^{Ile}$. Blank circles indicate overlap with a typical tRNA structure; white circles with red characters indicate substitutions in the *C. megacephala* sequences. **B** – The predicted secondary structure of *C. putoria* $tRNA^{Gln}$ (AF352790) used for comparison of duplicated sequences. The box indicates the $tRNA^{Gln}$-duplicated region in *Chrysomya*. The main sequence in the box is conserved for both *C. albiceps* and *C. megacephala* repeats. White circles with black characters indicate nucleotide substitutions and a 4 bp insertion (GATT) between position 42 (G) and 43 (A) in the repeated element of *C. putoria*.

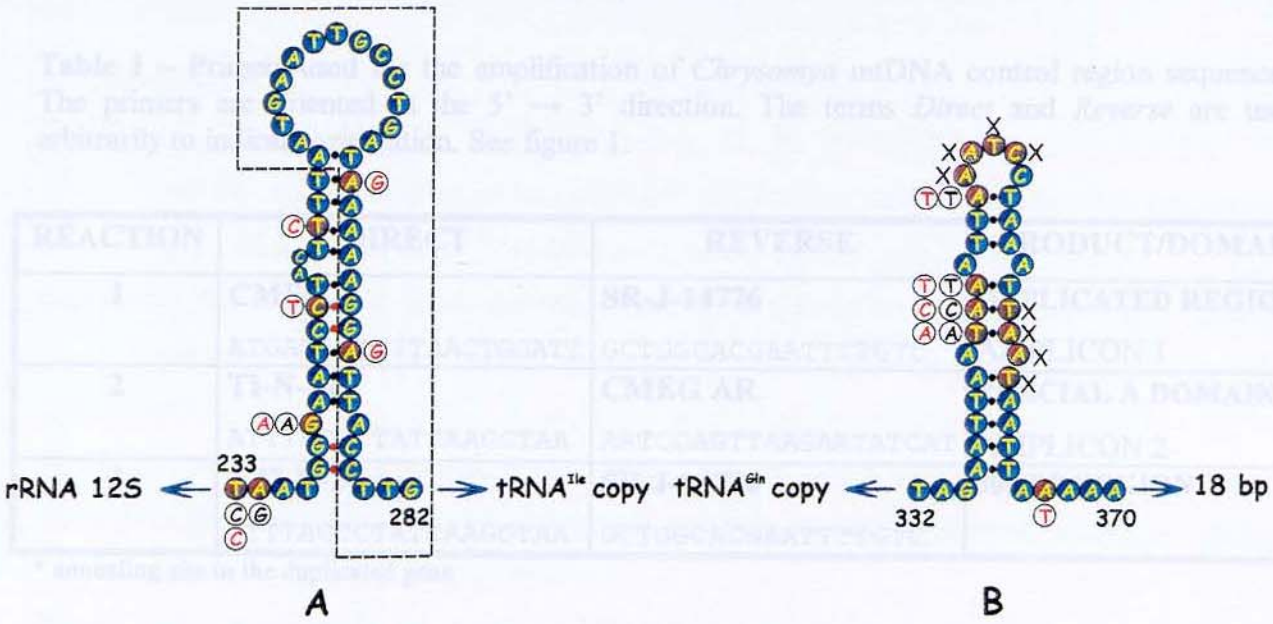


Figure 7 – Secondary structures at the duplication endpoints. Nucleotide positions 233 to 370 correspond to relative positions in the alignment in Figure 3A. Conserved nucleotides among *Chrysomya* species are shown in blue circles; red circles indicate nucleotide divergence in the corresponding *C. putoria* sequence (note the occurrence of a 33 bp symmetric sequence element in B); white circles with red or black characters indicate substitutions in *C. megacephala* and *C. albiceps* sequences, respectively. **A** - Predicted secondary structure formed by partial $tRNA^{Ile}$ and 19 bp element sequences in the duplication endpoint near the rRNA12S gene (free energy -6.1). $tRNA^{Ile}$ sequence is boxed. **B** – Predicted secondary structure formed by partial sequences of the pseudo-$tRNA^{Gln}$ and the 18 bp conserved element (free energy -3.2). X indicates gaps in both species.

Table 1 – Primers used for the amplification of *Chrysomya* mtDNA control region sequences. The primers are oriented in the 5' → 3' direction. The terms *Direct* and *Reverse* are used arbitrarily to indicate orientation. See figure 1.

REACTION	DIRECT	REVERSE	PRODUCT/DOMAIN
1	**CMEG A** ATGATATTCTTAACTGGATT	**SR-J-14776** GCTGGCACGAATTTTGTC	**DUPLICATED REGION** AMPLICON 1
2	**TI-N-24** ATTTACCCTATCAAGGTAA	**CMEG AR** AATCCAGTTAAGAATATCAT	**PARCIAL A DOMAIN** AMPLICON 2
3	*** TI-N-24** ATTTACCCTATCAAGGTAA	**SR-J-14776** GCTGGCACGAATTTTGTC	**300 BP REGION**

* annealing site in the duplicated gene

REFERENCES

Beagley, C. T.; Okimoto, R. &Wolstenholme, D. R. (1999). *Mytilus* mitochondrial DNA contains a functional gene for a $tRNA^{Ser}$ (UCN) with a dihydrouridine arm-replacement loop and a pseudo-$tRNA^{Ser}$ (UCN) gene. *Genetics*, **152**: 641 – 652.

Bensasson, D.; Zhang, D. X. & Hewitt, G. M. (2000). Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Molecular Biology and Evolution*, **17**: 406 – 415.

Black, IV, W. C. & Roehrdanz, R. L. (1998). Mitochondrial gene order is not conserved in arthropods: prostriate and metastriate tick mitochondrial genomes. *Molecular Biology and Evolution,* **15**: 1772 - 1785.

Boore, J. L.; Lavrov, D. V. & Brown, W. M. (1998). Gene translocation links insects and crustaceans. *Nature*, **392**: 667 – 668.

Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research*, **27**: 1767 – 1780.

Campbell, N. J. & Barker, S. C. (1998). An unprecedented major rearrangement in an arthropod mitochondrial genome. *Molecular Biology and Evolution*, **15**: 1786 –1787.

Campbell, N. J. H. & Barker, S. C. (1999). The novel mitochondrial gene arrangement of the cattle tick, *Boophilus microplus*: fivefold tandem repetition of a coding region. *Molecular Biology and Evolution,* **16**: 732 – 740.

Cantatore, P.; Gadaleta, M. N.; Roberti, M.; Saccone, C. & Wilson, A. C. (1987). Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. *Nature*, **329**: 853 – 855.

Dorner, M.; Altmann, M.; Pääbo, S. & Morl, M. (2001). Evidence for import of a Lysyl-tRNA into marsupial mitochondria. *Molecular Biology of the Cell*, **12**: 2688 – 2698.

Dotson E. M. & Beard C. B. (2001). Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. *Insect Molecular Biology*, **10**: 205 – 215.

Dowton, M. & Austin, A. D. (1999). Evolutionary dynamics of a mitochondrial rearrangement "hot spot" in Hymenoptera. *Molecular Biology and Evolution,***16**: 298 – 309.

Dowton M. & Campbell N. J. (2001). Intramitochondrial recombination – is it why some mitochondrial genes sleep around? *Trends in Ecology and Evolution,* **16**: 269 – 271.

Eberhard, J. R.; Wrigth, T. F. & Bermingham, E. (2001). Duplication and concerted evolution of the mitochondrial control region in the parrot genus *Amazona*. *Molecular Biology and Evolution*, **18**: 1330 – 1342.

Fitch, W. M. (2000). Homology a personal view on some of the problems. *Trends in Genetics*, **16:** 227 – 231.

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

Hoffmann, R. J.; Boore, J. L. & Brown, W. M. (1992). A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. *Genetics*, **131**: 397 – 412.

Hwang, U. H.; Friedrich, M.; Tautz, D.; Park, C. J. & Kim, W. (2001). Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature*, **413**: 154 – 157.

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

Junqueira, A. C. M.; Lessinger, A. C. & Azeredo-Espin, A. M. L. Methods for the recovery of mitochondrial DNA sequences from museum specimens of myiasis-causing flies. *Medical and Veterinary Entomology, in press.*

Kumazawa, Y.; Ota, H.; Nishida, M. & Ozawa, T. (1998). The complete nucleotide sequence of a snake *Dinodon semicarinatus* mitochondrial genome with two identical control regions. *Genetics*, **150**: 313 – 329.

Lessinger, A. C. & Azeredo-Espin, A. M. L. (2000). Evolution and structural organization of mitochondrial DNA control region of myiasis-causing flies. *Medical and Veterinary Entomology*, **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 mitiochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Molecular Biology*, **9**: 521 – 529.

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

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. *Medical and Veterinary Entomology*, **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 Research*, **25**: 955 – 964.

Macey, J. R.; Larson, A.; Ananjeva, N. B.; Fang, Z. & Papenfuss, T. J. (1997). Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Molecular Biology and Evolution*, **14**: 91 – 104.

Macey, J. R; Schulte, J. A. 2nd; Larson, A. & Papenfuss, T. J. (1998). Tandem duplication via light-strand synthesis may provide a precursor for mitochondrial genomic rearrangement. *Molecular Biology and Evolution*, **15**: 71 – 75.

Mindell, D. P. & Meyer, A. (2001). Homology evolving. *Trends in Ecology and Evolution*, **16**: 434 – 440.

Moritz, C. & Brown, W. M. (1987). Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. *Proceedings of the National Academy of Sciences* U. S. A., **84**: 7183 – 7187.

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. *Molecular Biology and Evolution*, **18**: 1293 – 1304.

Rand, D. M. & Harrison, R. G. (1989). Molecular population genetics of mtDNA size variation in crickets. *Genetics*, **121**: 551 – 569.

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.

SantaLucia, J. Jr (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceeding of the National Academy of Sciences* USA, **95**: 1460 – 1465.

Shao, R.; Campbell, N. J. & Barker, S. C. (2001a). Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). *Molecular Biology and Evolution*, **18**: 858 – 865.

Shao, R; Campbell, N. J; Schmidt, E. R. & Barker, S. C. (2001b). Increased rate of gene rearrangement in the mitochondrial genomes of three orders of hemipteroid insects. *Molecular Biology and Evolution*, **18**: 1828 – 1832.

Simon, C.; Frati, F.; Beckenbach, A.; Crespi, B.; Liu, H. & Flook, P. (1994). *Annals of the Entomological Society of America*, **87**: 651 – 701.

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.

Thompson, J. D.; Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**: 4673 – 4680.

Williams, S. T. & Knowlton, N. (2001). Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus *Alpheus*. *Molecular Biology and Evolution*, **18**: 1484 – 1493.

Zevering, C. E.; Moritz, C.; Heideman, A. & Sturm, R. A. (1991). Parallel origins of duplications and the formation of pseudogenes in mitochondrial DNA from parthenogenetic lizards (*Heteronotia binoei*; Gekkonidae). *Journal of Molecular Evolution*, **33**: 431 – 441.

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

Zhang, D. X.; Szymura, J. M. & Hewitt, G. M. (1995). Evolution and structural conservation of the control region of insect mitochondrial DNA. *Journal of Molecular Evolution*, **40**: 382 – 391.

DISCUSSÃO

Região controle do genoma mitocondrial

A caracterização molecular da região controle do DNAmt em espécies causadoras de miíases (Lessinger & Azeredo-Espin, 2000) enfoca principalmente a identificação de elementos conservados entre membros de duas famílias de Oestroidea (Dipetra: Calyptratae), Calliphoridae e Oestridae. Supõe-se que estes elementos desempenham um papel importante na regulação dos processos de replicação e transcrição da molécula de DNAmt, mas a identificação da provável função destes elementos, individualmente ou integrados na maquinaria de regulação dos processos de replicação e transcrição, ainda não está determinada. Zhang & Hewitt (1997), revisaram aspectos evolutivos da organização estrutural da região controle de insetos com a intenção de determinar homologias funcionais ou estruturais dentro deste grupo, e investigar o potencial desta região como marcador filogenético. A caracterização da região controle para novas espécies, entretanto, nem sempre resulta no reconhecimento dos elementos indicados como estruturas conservadas para insetos (Nardi *et al.*, 2001; Dotson & Beard, 2001).

A natureza da informação genética proveniente da região controle obedece a diferentes padrões de evolução, contendo domínios conservados e regiões com intensa variabilidade genética quando analisadas espécies próximas. Entretanto, o desvio de composição nucleotídica em favor de bases A e T nesta região, e a manutenção desta condição através do predomínio de substituições nucleotídicas do tipo transversões A↔T (Lessinger & Azeredo-Espin, 2000), aumenta o risco de homoplasias nas análises interespecíficas causadas por reversões nas seqüências hipervariáveis desta região. Deste modo, a dificuldade em estabelecer homologias consistentes, devido à presença de seqüências hipervariáveis, associada à extrema conservação em seqüência de domínios supostamente regulatórios, são aspectos que podem comprometer a utilidade desta região como marcador filogenético. Por outro lado, a análise estrutural da região controle, no contexto da caracterização de genomas completos, vem ganhando evidência ultimamente, principalmente devido à investigação de aspectos evolutivos relacionados a fenômenos de heteroplasmia, duplicação em "tandem", evolução em concerto, rearranjos, pressão de mutação, entre outros (Lewis *et al.*, 1995; Black & Roerhandz, 1998; Shao *et al.*, 2001; Nardi et al., 2001; Dotson & Beard, 2001).

Duplicação gênica em Chrysomya

Em Calliphoridae, a análise da região controle do DNAmt de três espécies do gênero *Chrysomya*, *C. albiceps*, *C. megacephala* e *C. putoria*, resultou na identificação de uma duplicação gênica envolvendo as seqüências de dois tRNAs e da região controle. Espécies deste gênero são consideradas importantes evidências na investigação forense e a análise de seqüências hipervariáveis da região controle pode representar um recurso eficiente para a identificação de diferentes espécies. Entretanto, o acesso a essa informação através de estratégias de PCR (Lessinger & Azeredo-Espin, 2000) produz resultados ambíguos ou inespecíficos em muitos casos (J. Wells, comunicação pessoal). A revisão e caracterização de elementos estruturais da região controle de *Chrysomya* permite a eficiente recuperação de seqüências domínio-específico neste gênero e em espécies próximas, ampliando a utilidade deste marcador (Lessinger *et al.*, em preparação).

Recentemente, Wells & Sperling (2001) descreveram as relações filogenéticas entre membros da subfamília Chrysomyinae (Diptera: Calliphoridae), baseadas na análise de marcadores mitocondriais envolvendo sequências das subunidades gênicas COI e COII, com a finalidade de caracterizar espécies de importância forense (Figura 1). Segundo estes autores, esta filogenia diverge significativamente do modelo tradicionalmente aceito para o grupo, baseado em características morfológicas, principalmente em relação à natureza polifilética da tribo Phormini. De modo a integrar a duplicação gênica encontrada na região controle de *Chrysomya* na história evolutiva do grupo, foi realizada uma análise preliminar considerando-se o mapeamento deste evento relacionado à evolução molecular sobre esta hipótese filogenética.

O sistema taxonômico tradicional de Chrysomyinae agrupa os gêneros *Protocalliphora* + *Phormia* + *Protophormia* na tribo Phormini e *Chrysomya* + *Compsomyiops* + *Cochliomyia* na tribo Chrysomyini (Dear, 1985 e McAlpine & Wood, 1989). Entretanto, apesar dos conflitos na interpretação das relações filogenéticas indicarem a necessidade de revisão taxonômica, algumas considerações em relação à origem da duplicação gênica são descritas a seguir com base na figura 1. É importante notar que estão indicadas as espécies cuja região controle foi previamente caracterizada (*), com especial atenção às espécies de *Chrysomya*, onde a duplicação gênica foi detectada. No sentido de tornar esta abordagem mais informativa, é relevante citar que não foram encontrados elementos duplicados na região controle das espécies *D. hominis* (Oestridae) e *L. eximia*, o que poderia indicar a ausência deste caráter em *H. lineatum* (Oestridae), *L. sericata* e *L. illustris*, representados na filogenia da figura 1.

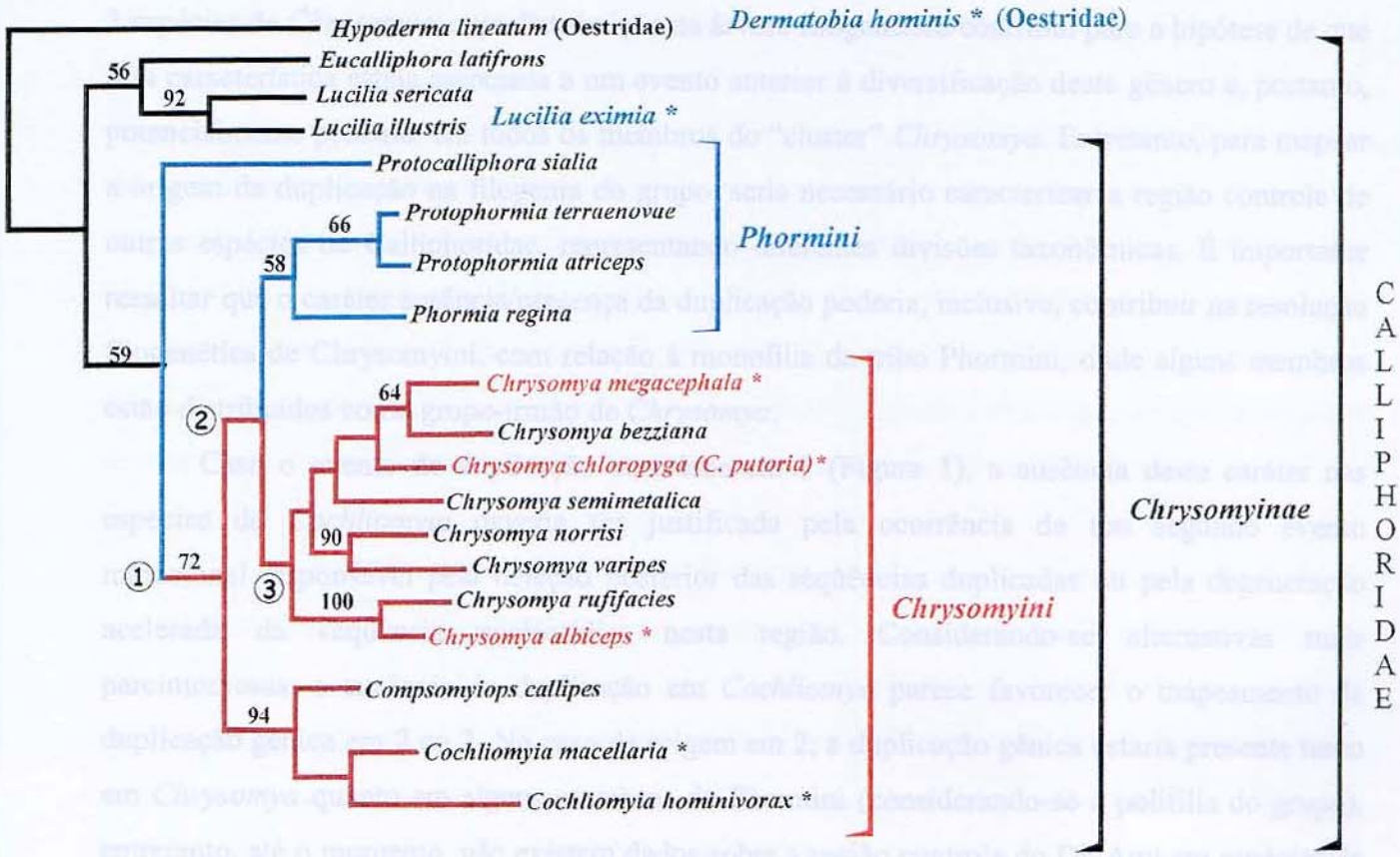


Figura 1 – Relações filogenéticas na subfamília Chrysomyinae obtidas pela análise dos marcadores mitocondriais COI/II (adaptado de Wells & Sperling, 2001). O asterisco indica espécies cuja região controle do DNAmt foi previamente analisada (Lessinger & Azeredo-Espin, 2000 e Lessinger *et al.*, em preparação). Em vermelho estão indicadas as espécies onde foi verificada a presença da duplicação gênica na região controle (Lessinger *et al.* em preparação). Os números 1, 2 e 3 indicam possíveis origens da duplicação (ver discussão). As espécies *D. hominis* (Oestridae) e *L. eximia* (Luciliinae) não estão incluídas na filogenia, mas são importantes parâmetros na investigação sobre a origem da duplicação na região controle do genoma mitocondrial.

No que diz respeito à tribo Chrysomyini, a duplicação gênica está presente em pelo menos 3 espécies de *Chrysomya*, cuja distribuição na árvore filogenética contribui para a hipótese de que esta característica esteja associada a um evento anterior à diversificação deste gênero e, portanto, potencialmente presente em todos os membros do "cluster" *Chrysomya*. Entretanto, para mapear a origem da duplicação na filogenia do grupo, seria necessário caracterizar a região controle de outras espécies de Calliphoridae, representando diferentes divisões taxonômicas. É importante ressaltar que o caráter ausência/presença da duplicação poderia, inclusive, contribuir na resolução filogenética de Chrysomyini, com relação à monofilia da tribo Phormini, onde alguns membros estão distribuídos como grupo-irmão de *Chrysomya*.

Caso o evento de duplicação ocorresse em 1 (Figura 1), a ausência deste caráter nas espécies de *Cochliomyia* deveria ser justificada pela ocorrência de um segundo evento mutacional responsável pela deleção posterior das seqüências duplicadas ou pela degeneração acelerada da seqüência nucleotídica nesta região. Considerando-se alternativas mais parcimoniosas, a ausência da duplicação em *Cochliomya* parece favorecer o mapeamento da duplicação gênica em 2 ou 3. No caso da origem em 2, a duplicação gênica estaria presente tanto em *Chrysomya* quanto em alguns membros de Phormini (considerando-se a polifilia do grupo), entretanto, até o momento, não existem dados sobre a região controle do DNAmt em espécies de *Phormia* e *Protophormia* que poderiam confirmar esta alternativa. A hipótese mais razoável seria a origem da duplicação em 3, na linhagem ancestral do grupo *Chrysomya*, neste contexto, explica-se a ausência do genótipo em *Cochliomyia*, além de corroborar com o padrão de evolução recente deste genótipo, inferido através do grau de conservação das seqüências duplicadas em relação às originais (Lessinger *et al.*, em preparação). É interessante observar que esta última alternativa independe do conflito filogenético das análises morfológica e molecular, sendo válida em ambos cenários, considerando-se a natureza monofilética das relações em *Chrysomya*.

Diagnóstico através de PCR-RFLP

Do ponto de vista aplicado, a região controle tem se mostrado um marcador eficiente na análise de padrões moleculares PCR-RFLP para o diagnóstico espécie-específico, reconhecendo a identidade das espécies *Cochliomyia hominivorax* e *C. macellaria* considerando-se uma ampla distribuição geográfica (Litjens, Lessinger & Azeredo-Espin, 2001). Padrões polimórficos de PCR-RFLP foram observados em baixa freqüência nestas espécies, resultando em uma identificação taxonômica ambígua, entretanto a combinação ou associação de marcadores independentes (diferentes seqüências-alvo ou diferentes endonucleases de restrição) é uma

abordagem eficiente e econômica para assegurar a identidade da amostra analisada. A padronização de testes-diagnóstico em espécies causadoras de miíases é uma alternativa extremamente útil para recuperar a informação taxonômica a partir de amostras parcialmente degradadas ou em estágios imaturos de desenvolvimento, onde as evidências morfológicas não são confiáveis ou estão ausentes. A rápida caracterização de espécies-praga através do diagnóstico molecular tem contribuído em programas de controle e no monitoramento de eventos de introdução e erradicação destas espécies (Taylor *et. al.* 1996). Este trabalho amplia esta perspectiva, incluindo uma amostragem significativa de populações brasileiras de *Cochliomyia* e validando as seqüências da região controle como um eficiente marcador espécie-específico, além de outras regiões codificadoras do DNAmt como as subunidades COI e COII.

O genoma mitocondrial de C. hominivorax

No que se refere à caracterização de genomas completos, o genoma mitocondrial de *C. hominivorax* (Lessinger *et al.*, 2000) foi o primeiro a ser seqüenciado no Brasil. A importância deste trabalho pode ser verificada indiretamente através do número de citações relativas ao artigo que descreve o genoma mitocondrial da mosca-da-bicheira (Nardi, *et al.* 2001; Dotson & Beard 2001; Litjens, Lessinger & Azeredo-Espin 2001; Shao, Campbell & Barker, 2001; Wells & Sperling, 2001; Shao *et al*, 2001, Junqueira *et al.*, *in press*).

O genoma mitocondrial de *C. hominivorax* apresenta a mesma organização estrutural descrita para outras espécies de dípteros superiores, *Drosophila* e *Ceratitis* (Clary & Wolstenholme, 1985; Spanos *et al.*, 2000). De modo geral, a terminação incompleta de alguns genes codificadores de proteínas, possivelmente mediada por sequências de tRNAs imediatamente adjacentes, e a indefinição do códon de iniciação correspondente ao gene da subunidade I do complexo Citocromo Oxidase (COI), são questões freqüentemente abordadas pelos estudos que descrevem genomas mitocondriais em diferentes grupos de Arthropoda.

Entretanto, a ausência de estudos experimentais, relacionados à análise da expressão gênica de transcritos mitocondriais neste grupo, restringe interpretações mais amplas a respeito destas questões. A análise de transcritos mitocondriais nestes estudos, refere-se normalmente à inferência da provável seqüência de aminoácidos a partir da caracterização de seqüências nucleotídicas, da mesma forma para RNAs ribossomais e de transferência.

O estabelecimento de relações filogenéticas, a partir de genomas mitocondriais completos, tem sido realizado através da análise das seqüências previstas de aminoácidos, de modo a diminuir a interferência de eventuais inconsistências associadas a regiões muito variáveis ou ao

acentuado desvio de composição nucleotídica dos genomas mitocondriais de insetos. O alinhamento individual de cada um dos 13 genes codificadores de proteínas e a integração dos resultados (concatenado) produzindo um único conjunto de dados ("mitochondrial proteome", segundo Hwang *et al.*, 2001) tem sido amplamente empregado. Eventualmente, a remoção de seqüências que produzem alinhamentos inconsistentes, ou pouco confiáveis, pode evitar o comprometimento da inferência filogenética, como é o caso das seqüências que codificam os genes ATPase 8 e ND6.

Em relação à ordem Diptera, poucos estudos de filogenia molecular envolvem as relações evolutivas em Calyptratae (Diptera: Brachycera) (Nirmala *et al.*, 2001). Têm sido estudadas, principalmente, as relações filogenéticas entre membros de uma mesma superfamília (Bernasconi *et al.*, 2000a), subfamília (Stevens & Wall, 2001; Wells & Sperling, 2001, Bernasconi *et al*, 2000b) ou gênero (Wells & Sperling, 1999; Stevens & Wall, 1997). Recentemente, o estabelecimento de relações filogenéticas através da análise de genomas mitocondriais completos tem apresentado divergências quanto à posição de *C. hominivorax* (Calyptratae) em relação às espécies *D. yakuba* e *C. capitata* (Acalyptratae). Segundo Nardi *et al.* (2001) e Hwang *et al.* (2001), a posição de *C. hominivorax* não concorda com a divisão Acalyptratae/Calyptratae, estando esta espécie distribuída entre *Drosophila* e *Ceratitis*. Lessinger *et al.* (2000), descreve a manutenção da divisão taxonômica nestas espécies, agrupando *Ceratitis* e *Drosophila* no "cluster" Acalyptratae. A análise das relações filogenéticas em Brachycera, através da caracterização de genomas mitocondriais em outras espécies de Calyptratae, pode representar uma contribuição importante na investigação da hipótese de evolução monofilética do taxon Acalyptratae.

Considerações gerais

A análise de seqüências completas de genomas mitocondriais em insetos conta apenas com 15 espécies descritas no GenBank (www.ncbi.nlm.nih.gov) de um total de 50 registros em Arthropoda, representando 25 diferentes espécies. Há um evidente predomínio de seqüências completas com relação à ordem Diptera. Este fato se justifica em parte pela presença de espécies de significativa importância econômica (*Cochliomyia* e *Ceratitis*) e médica (*Anopheles* e *Chrysomya*), além de espécies de comprovada relevância como modelo biológico (*Drosophila*). A importância da descrição do genoma mitocondrial de *Drosophila* pode ser ilustrada considerando-se o crescente número de citações do artigo de Clary & Wolstenholme (1985), que compreende aproximadamente 596 citações (Outubro/2001).

Diferentes aspectos relacionados à análise do genoma mitocondrial em espécies causadoras de miíases foram abordados neste trabalho. A resolução de conflitos taxonômicos através de marcadores mitocondriais merece especial atenção, tanto em relação à resolução filogenética em grupos proximamente relacionados, como Calliphoridae, quanto na investigação de relações entre grandes divisões taxonômicas, como no caso de Brachycera (Diptera). Conflitos entre hipóteses filogenéticas alternativas em Calliphoridae também podem, potencialmente, ser avaliados considerando-se a presença ou ausência da duplicação gênica da região controle de replicação do genoma mitocondrial.

Os estudos apresentados nesta tese demonstram que a caracterização da organização estrutural e dos padrões de evolução do DNAmt, através da investigação de genomas completos e regiões específicas, amplia significativamente o conhecimento sobre a diversidade deste sistema em insetos. Deste modo, contribuindo tanto para a avaliação da utilidade filogenética de potenciais marcadores moleculares, quanto para a análise de mecanismos evolutivos responsáveis por diferentes padrões de variabilidade genética, tanto em estrutura quanto em seqüência, no DNAmt. Sendo importante ressaltar, portanto, que a caracterização preliminar de seqüências mitocondriais recupera informações estratégicas em estudos envolvendo evolução molecular, além de orientar a aplicação de marcadores na investigação da história evolutiva de organismos, evidenciando assim os diferentes níveis de complexidade deste genoma.

REFERÊNCIAS BIBLIOGRÁFICAS

Attimonelli M., Altamura N., Benne R., Brennicke A., Cooper J. M., D'Elia D., Montalvo A., Pinto B., De Robertis M., Golik P., Knoop V., Lanave C., Lazowska J., Licciulli F., Malladi B. S., Memeo F., Monnerot M., Pasimeni R., Pilbout S., Schapira A. H., Sloof P. & Saccone C. (2000). MitBASE : a comprehensive and integrated mitochondrial DNA database. The present status. *Nucleic Acids Research*, **28**: 148 - 152.

Azeredo Espin, A. L. M. (1993). Mitochondrial DNA variability in geographic populations of screwworm fly from Brasil. *International Atomic Energy Agency*, p. 161 - 165.

Azeredo Espin, A. L. M. & Madeira, N. G. (1996). Primary myiasis in dog caused by *Phaenicia eximia* (Diptera: Calliphoridae) and preliminary mitochondrial DNA analysis of the species in Brazil. *Journal of Medical Entomology*, **33**, 839 – 843.

Ballard J. W. (2001 a) Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. *Journal of Molecular Evolution,* **51**: 48 - 63.

Ballard J. W. (2001 b). Comparative genomics of mitochondrial DNA in *Drosophila simulans*. *Journal of Molecular Evolution*, **51**: 64 - 75.

Beagley C. T., Okimoto R. & Wolstenholme D. R. (1999). *Mytilus* mitochondrial DNA contains a functional gene for a tRNA (Ser) (UCN) with a dihydrouridine arm - replacement loop and a pseudo – tRNA (Ser) (UCN) gene. *GENETICS,* **152**: 641 – 652.

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 sequence of the other insects. *Insect Molecular Biology,* **2,** 103 - 124.

Bernasconi M. V., Valsangiacomo C., Piffaretti J. C. & Ward P. I. (2000 a). Phylogenetic relationships among muscoidea (Diptera: Calyptratae) based on mitochondrial DNA sequences. *Insect Molecular Biology*, **9**: 67 - 74.

Bernasconi M. V., Pawlowski J., Valsangiacomo C., Piffaretti J. C. & Ward P. I. (2000 b). Phylogeny of the scathophagidae (Diptera, Calyptratae) based on mitochondrial DNA sequences. *Molecular and Phylogenetic Evolution*, **16**: 308 - 315.

Black, W. C. & Roerhdanz, R. L. (1998). Mitochondrial gene order is not conserved in arthropods: prostriate and metastriate tick mitochondrial genomes. *Molecular Biology and Evolution* **15**: 1772 – 1785.

Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research*, **27**: 1767 – 1780.

Brown, W. M. (1985). The mitochondrial genome of animals. 95 – 130. In: *Molecular Evolutionary Genetics*, ed. R.J. Macintyre. Plenum Press, N. Y.

Campbell, N. J. H. & Barker, S. C. (1998). An unprecedented major rearrangement in an arthropod mitochondrial genome. *Molecular Biology and Evolution*, **15**: 1786 – 1787.

Campbell N. J. H. & Barker S. C. (1999). The novel mitochondrial gene arrangement of the cattle tick, *Boophilus microplus*: Fivefold tandem repetition of a coding region. *Molecular Biology and Evolution*, **16**: 732 - 740.

Clary, D. O. & Wolstenholme, D. R. (1985). The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. *Journal of Molecular Evolution,* **22,** 252 - 271.

Cooper A., Lalueza-Fox C., Anderson S., Rambaut A., Austin J. & Ward R. (2001). Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature*, **409**: 704 - 707.

Crease, T. J. (1999). The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). *Gene*, **233**, 89 – 99.

Crozier, R. H. & Crozier, Y. C. (1993). The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics*, **133,** 97 - 117.

Curole A. P. & Kocher T. D. (1999). Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends in Ecology and Evolution*, **14**: 394 - 398.

Dear, J. P. (1985). A revision of the New World Chrysomyini (Diptera: Calliphoridae). *Revista Brasileira de Zoologia*, **3**: 109 – 169.

Dotson E. M. & Beard C. B. (2001). Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata. Insect Molecular Biology*, **10**: 205 – 215.

Dowton M. & Austin A. D. (1999). Evolutionary dynamics of a mitochondrial rearrangement "hot spot" in the hymenoptera. *Molecular Biology and Evolution,* **16**: 298 – 309.

Dowton M. & Campbell N. J. H. (2001). Intramitochondrial recombination - is it why some mitochondrial genes sleep around? *Trends in Ecology and Evolution*, **16**: 269 – 271.

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

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. *Annais of the Entomological Society of America,* **93**: 1085 – 1094.

Goldenthal, M. J.; Mckenna, K. A. & Joslyn, D. J. (1991). Mitochondrial DNA of the blow fly *Phormia regina*: restriction analysis and gene localization. *Biochemical Genetics*, **29**, 1 – 11.

Gray M. W., Burger G. & Lang B. F. (1999). Mitochondrial evolution *Science*, **283**: 1476 – 1481.

Haddrath O. & Baker A. J. (2001). Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. Proceedings of the Royal Society of London B Biological Sciences, **268**: 939 - 945.

Hickerson M. J. & Cunningham C. W. (2000). Dramatic mitochondrial gene rearrangements in the hermit crab *Pagurus longicarpus* (Crustacea, anomura). *Molecular Biology and Evolution,* **17**: 639 - 644.

Hwang U. W., Friedrich M., Tautz D., Park C. J. & Kim W. (2001). Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature*, **413**: 154 – 157.

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

Infante-Malachias, M. E.; Yotoko, K. & Azeredo-Espin, A. M. L. (1999). Random amplified polymorphic DNA of screwworm fly populations (Diptera: Calliphoridae) from Southeastern Brazil and Northern Argentina. *Genome*, **42**, 772 – 779.

Kogelnik A. M., Lott M. T., Brown M. D., Navathe S. B. & Wallace D. C. (1998). MITOMAP: a human mitochondrial genome database--1998 update. *Nucleic Acids Research*, **26**: 112 - 115.

Lanave C., Liuni S., Licciulli F. & Attimonelli M. (2000). Update of AMmtDB: a database of multi-aligned metazoa mitochondrial DNA sequences. *Nucleic Acids Research*, **28**: 153 - 154.

Lang B. F., Gray M. W. & Burger G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annual Review in Genetics*, **33**: 351 - 397.

LaRoche, J.; Snyder, M.; Cook, D. I.; Fuller, K. & Zouros, E. (1990). Molecular characterization of a repeat element causing large-scale size variation in the mtDNA of the Sea Scallop *Placopecten magellanicus*. *Molecular Biology and Evolution*, **7**, 45 – 64.

Lavrov D. V., Boore J. L. & Brown W. M. (2000 a). The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus. Molecular Biology and Evolution*, **17**: 813 – 824.

Lavrov, D. V., Brown, W. M. & Boore, J. L. (2000 b). A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus. Proceedings of the National Academy of Sciences, U.S.A.* **97**, 13738 – 13742.

Lessinger, A. C. & Azeredo-Espin, A. M. L. (2000). Evolution and structural organisation of mitochondrial DNA control region of myiasis-causing flies. *Medical and Veterinary Entomology*, **14**, 71 – 80.

Lessinger A. C.; Martins Junqueira A. C.; Lemos T. A., Kemper E. L., da Silva F. R., Vettore A. L., Arruda P., Azeredo-Espin A. M. (2001). The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Molecular Biology*, **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. *Molecular Biology and Evolution*, **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 Molecular Biology*, **4**: 263 - 278

Litjens, P.; Lessinger, A. C. & Azeredo-Espin, A. M. L. (2001). Caracterization of *Cochliomyia hominivorax* and *Cochliomyia macellaria* by PCR-RFLP of mitochondrial DNA. *Medical and Veterinary Entomology*, **15**, 183 – 188.

McAlpine, J. F. (Editor) & Wood, D. M. (1989). Manual of Neartic Diptera. Research Branch Agriculture Canada. Series: Monograph N° 32.

Mitchell, S. E.; Cockburn, A. F. & Seawright, J. A. (1993). The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. *Genome*, **36,** 1058 - 1073.

Nardi F., Carapelli A., Fanciulli P. P., Dallai R. & Frati F. (2001) The Complete Mitochondrial DNA Sequence of the Basal Hexapod *Tetrodontophora bielanensis*: Evidence for Heteroplasmy and tRNA Translocations. *Mol Biol Evol.*, **18**: 1293 - 1304.

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 Molecular Biology*, **10**: 475 – 486.

Ojala D., Merkel C., Gelfand R. & Attardi G. (1980). The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. *Cell*, **22**: 393 - 403.

Okimoto, R.; Macfarlane, J. L.; Clary, D. O. & Wolstenholme, D. R. (1992). The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics*, **130**, 471 – 498.

Rand D. (2001) Mitochondrial genomics flies high. *Trends in Ecology and Evolution,* **16**: 2 - 4.

Roehrdanz, R. L.; Johnson & D. A. (1988). Mitochondrial DNA Variation Among Geographical Populations of the Screwworm Fly, *Cochliomyia hominivorax*. *Journal of Medical Entomology* **25** (2): 136-141.

Roehrdanz, R. L. (1989). Intraspecific genetic variability in mitochondrial DNA of the screwworm fly (*Cochliomyia hominivorax*). *Biochemical Genetics*, **27**: 551-569.

Ryder O. A., McLaren A., Brenner S., Zhang Y.-P. & Benirschke K. (2000). DNA banks for endangered animal species. *Science*, **288**: 275 - 277.

Saccone C., De Giorgi C., Gissi C., Pesole G. & Reyes A. (1999). Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene*, **238**: 195 – 209.

Shadel G. S., Clayton D. A. (1997). Mitochondrial DNA maintenance in vertebrates. *Annual Review in Biochemistry*, **66**: 409 – 435.

Shao R., Campbell N. J. & Barker S. C. (2001 a) Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). *Molecular Biology and Evolution*, **18**: 858 - 865.

Shao R., Campbell N. J., Schmidt E. R. & Barker S. C. (2001 b). Increased rate of gene rearrangement in the mitochondrial genomes of three orders of hemipteroid insects. *Molecular Biology and Evolution*, **18**: 1828 – 1832.

Shimko N., Liu L., Lang B. F. & Burger G. (2001). GOBASE: the organelle genome database. *Nucleic Acids Research*, **29**: 128 - 132.

Simon, C.; Frati, F.; Beckenbach, A.; Crespi, B.; Liu, H. & Flook, P. (1994). Evolution, weighting and phylogenetic utility of mitochondrial DNA sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651 - 701.

Smith, S. M.; Fuerst, P. & Meckenburg, K. L (1996). Mitochondrial DNA sequence of cytochrome oxidase II from *Calliphora erythrocephala*: evolution of blow flies (Diptera: Calliphoridae). *Annals of the Entomological Society of America*, **89**, 28 – 36.

Spanos, L; Koutroumbas, G.; Kotsyfakis, M. & Louis, C. (2000). The mitochondrial genome of the mediterranean fruitfly, *Ceratitis capitata*. *Insect Molecular Biology*, **9**, 139 – 144.

Sperling, F. A. H.; Anderson, G. S. & Hickey, D. A. (1994). A DNA-based approach to the identification of insect species used for postmorten interval estimation. *Journal of Forensic Sciences*, **39**, 418 - 427.

Stevens J. & Wall, R. (1997). The evolution of ectoparasitism in the genus *Lucilia* (Diptera: Calliphoridae). *International Journal of Parasitology*, **27**, 51 - 59.

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

Taylor, D. B.; Szalanski, A. L. & Peterson II, R. D. (1996). Identification of Screwworm Species by Polymerase Chain Reaction-Restriction Fragment Length Polymorphism. *Medical and Veterinary Entomology* **10**: 63-70.

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**, 521 – 526.

Valverde J. R.; Batuecas B.; Moratilla, C.; Marco, R. & Garesse, R. (1994). The Complete Mitochondrial DNA Sequence of the Crustacean *Artemia franciscana*. *Journal of Molecular Evolution*, **39**: 400-408.

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 Science International*, **120**: 60 – 67.

Wells, J. D. & Sperling, F. A. H. (1999). Molecular phylogenetics of *Chrysomya albiceps* and *Chrysomya rufifacies* (Diptera: Calliphoridae). *Journal of Medical Entomology*, **36**, 222 - 226.

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

Wheeler D. L., Church D. M., Lash A. E., Leipe D. D., Madden T. L., Pontius J. U., Schuler G. D., Schriml L. M., Tatusova T. A., Wagner L. & Rapp B. A. (2001). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, **29**: 11 - 16.

Wilson K., Cahill V., Ballment E. & Benzie J. (2000). The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: are malacostracan

crustaceans more closely related to insects than to branchiopods? *Molecular Biology and Evolution*, **17**: 863 - 874.

Wolstenholme D. R. (1992). Animal Mitochondrial DNA - Structure and Evolution. *International Review in Cytology*, **141**: 173 - 216.

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