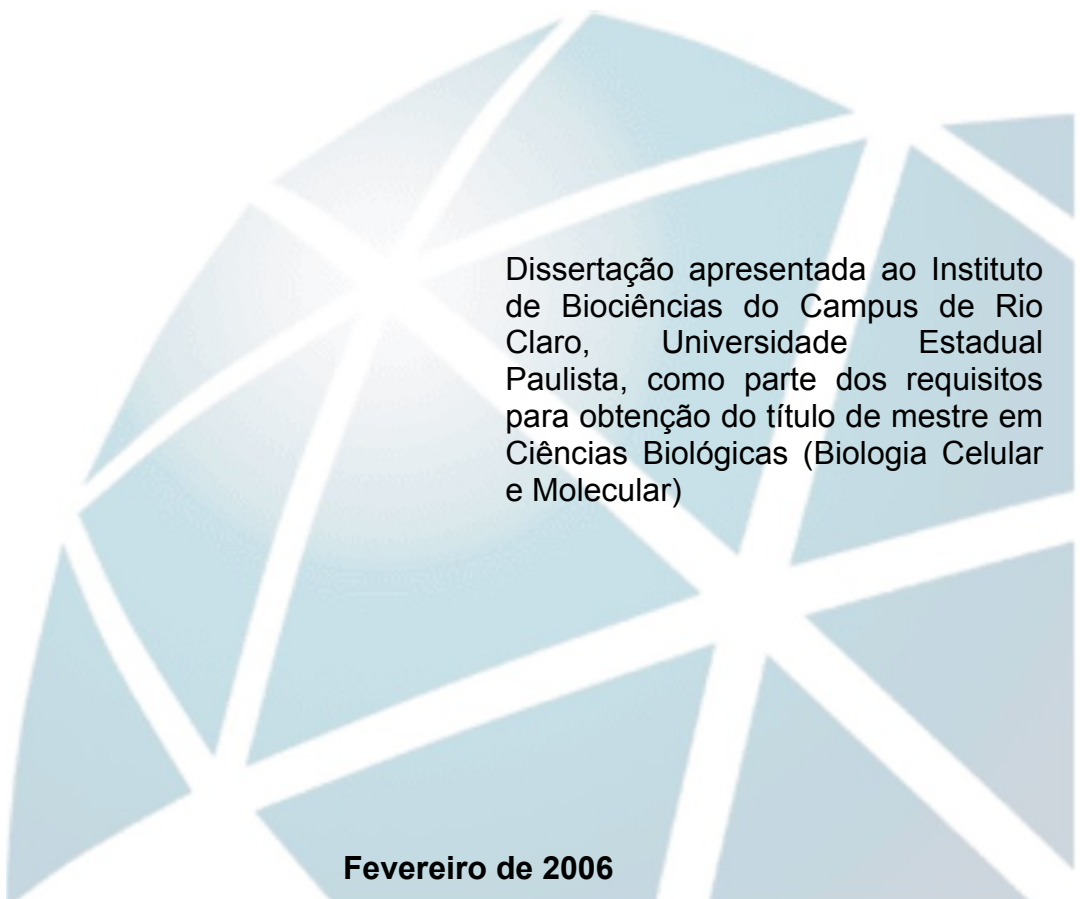

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA BIOLÓGICAS
ÁREA BIOLOGIA CELULAR E MOLECULAR

**Evolução de genes e pseudogenes mitocondriais em
Atta cephalotes (Linnaeus, 1758)**

JOAQUIM MARTINS JÚNIOR

A large, stylized graphic of a globe in the background, composed of light blue and white geometric shapes that form a grid-like pattern over a circular base.

Dissertação apresentada ao Instituto de Biociências do Campus de Rio Claro, Universidade Estadual Paulista, como parte dos requisitos para obtenção do título de mestre em Ciências Biológicas (Biologia Celular e Molecular)

Fevereiro de 2006

**Evolução de genes e pseudogenes mitocondriais
em *Atta cephalotes* (Linnaeus, 1758)**

JOAQUIM MARTINS JUNIOR

Orientador: Prof. Dr. Maurício Bacci Júnior

**Dissertação apresentada ao
Instituto de Biociências da
Universidade Estadual Paulista
“Julio de Mesquita Filho”,
Campus de Rio Claro, para a
obtenção do título de Mestre
em Ciências Biológicas
(Biologia Celular e Molecular)**

Fevereiro de 2006

595.796 Martins Júnior, Joaquim

M386e Evolução de genes e pseudogenes mitocondriais em *Atta cephalotes* (Linnaeus, 1758) / Joaquim Martins Júnior. – Rio Claro: [s.n.], 2006
44 fls. : il., gráfs., tabs.

Dissertação (mestrado) – Universidade Estadual Paulista,
Instituto de Biociências de Rio Claro
Orientador: Maurício Bacci Júnior

1. Formiga. 2. *Numts*. 3. Filogenia. 4. DNA mitocondrial. I.
Título.

Ficha Catalográfica elaborada pela STATI – Biblioteca da UNESP
Campus de Rio Claro/SP

“Quando nada parece dar certo, vou ver o cortador de pedras martelando sua rocha talvez 100 vezes sem que uma única rachadura apareça. Mas na centésima primeira martelada a pedra se abre em duas, e eu sei que não foi aquela que conseguiu isso, mas todas as que vieram antes.”

(Jacob Riis)

Dedicatória

Dedico este trabalho aos meus pais, Joaquim e Geralda por estarem sempre ao meu lado nos momentos mais felizes e naqueles um pouco tristes, me incentivando e apoiando. Vocês são o meu espelho e esteio. A minha namorada Clécia pelo companheirismo, amor, paciência e apoio incondicional em todos os momentos.

Agradecimentos

A Deus, meu guia e mentor espiritual, por nunca me deixar perder o foco: da vida e do trabalho.

À Universidade Estadual Paulista – UNESP, campus Rio Claro e à Coordenação de Pós-Graduação do Instituto de Biociências pela oportunidade de realização do curso;

Ao Prof. Dr. Maurício Bacci Júnior pela orientação e confiança na capacidade de trabalho de seus alunos;

À minha namorada Clécia – *um presente de Deus* – pelo amor, companheirismo e dedicação. Obrigado por me ouvir, me entender e me aconselhar nos momentos em que precisei. Você é meu maior exemplo de que as coisas mais simples da vida são as que mais nos fazem felizes!

Aos meus pais pelo incentivo incondicional, carinho e dedicação e por não terem medido esforços para tornar possível a realização deste sonho. Obrigado e amo vocês! Aos demais familiares que mesmo distantes, sempre estiveram presentes e me incentivando;

Aos velhos amigos “Calangos do Cerrado” por estarem sempre presentes apesar da distância. Especialmente ao João Paulo, amigo e irmão em todos os momentos;

Aos novos amigos do CEIS e do LEM – todos – obrigado por me receberem e acolherem tão bem, fazendo com que me sentisse parte dessa família científica; e pelos momentos de alegria, festinhas, cervejas, risadas e conversas descontraídas;

Ao Scott e ao Sasha, da “*Section of Integrative Biology* – U.T.”, Aline Silva e Adriana Ortiz, pelo material coletado que possibilitou a realização desse trabalho. Principalmente aos amigos e companheiros do LEM e do Laboratório de Microbiologia que, em inúmeras vezes, ajudaram na execução desse trabalho;

E as formigas saúvas que “deram” suas vidas por este trabalho;

À CAPES e FAPESP pelo auxílio financeiro concedido para a execução do trabalho.

ÍNDICE

Apresentação.....	v
Manuscritos.....	1
Filogenia Molecular de <i>Atta cephalotes</i> Linnaeus, 1758 (Myrmicinae: Attini).....	2
Resumo.....	3
Introdução.....	4
Material e Métodos.....	7
Resultados.....	9
Discussão.....	11
Referências.....	14
Anexos.....	18
Primeira evidência de numts em <i>Atta cephalotes</i> Linnaeus, 1758 (Myrmicinae: Attini).....	25
Resumo.....	26
Introdução.....	27
Material e Métodos.....	28
Resultados.....	30
Discussão.....	31
Referências.....	35
Anexos.....	39
Considerações Finais.....	44

APRESENTAÇÃO

Esta dissertação é o resultado de dois anos de pesquisa envolvendo uma espécie de formiga cultivadora de fungos da tribo Attini, conhecida como saúva-da-mata, cujo nome científico é *Atta cephalotes* e a ocorrência se restringe a Mata Atlântica e Amazônia, por ser uma espécie exigente em relação as condições de umidade e temperatura. Durante a realização desta pesquisa encontramos resultados bastante interessantes que se desdobraram em dois trabalhos.

O primeiro trabalho trata da caracterização de uma região do DNA mitocondrial e seu uso como ferramenta na filogenia molecular desta espécie, em que são estudadas populações coletadas na faixa floresta Amazônica que vai desde Colômbia e Venezuela até o norte do Mato Grosso, passando por Amapá e Pará.

O outro trabalho apresenta a descoberta de dois pseudogenes em *Atta cephalotes*, que no decorrer deste estudo, quase nos levaram a formular hipóteses inverídicas em relação a classificação e taxonomia da espécie. Discutimos ainda a importância dos pseudogenes como falseadores de estudos filogenéticos e que, em trabalhos de filogenia, devemos estar seguros de que estamos trabalhando com o gene de interesse correto. A partir desta descoberta foi possível desenhar iniciadores específicos para *Atta cephalotes*, possibilitando a conclusão da proposta inicial que está apresentada no primeiro trabalho.

MANUSCRITOS

Filogenia Molecular de *Atta cephalotes* Linnaeus (1758) (Attini: Myrmicinae)

¹J. Martins Jr, ²S.E. Solomon, ²A. Mycheyev, ²U. Mueller and ¹M. Bacci Jr

¹Centro de Estudos de Insetos Sociais, Instituto de Biociências, Universidade Estadual Paulista, Rio Claro, Brasil.

²Section of Integrative Biology, Patterson Laboratories, University of Texas, Austin, USA.

Corresponding author:

Joaquim Martins Jr., Centro de Estudos de Insetos Sociais, Universidade Estadual Paulista

Av. 24A, n. 1515 – Bela Vista

CEP – 13506-900 – Rio Claro-SP-Brasil

Fone: 55 19 3526 4160

e-mail: martin@rc.unesp.br

Resumo

A formiga saúva-da-mata, *Atta cephalotes*, tem como área de ocorrência a Mata Atlântica e a Floresta Amazônica, pois não consegue viver em regiões quentes e secas como o Cerrado e Caatinga. Como os demais atíneos, corta folhas e flores que servem de substrato para o fungo que cultivam, sendo considerada, na Bahia, uma importante espécie que ataca as plantações de cacau na mata cabrucada. Sua classificação taxonômica sempre foi controversa e difícil devido à dificuldade de separar as subespécies através de caracteres morfológicos, que são polimórficos e pouco informativos, fazendo com todas as subespécies fossem sinonimizadas como *Atta cephalotes*. O presente estudo teve como objetivo caracterizar os *loci* mitocondriais COI, IGS, Leu-tRNA e COII de populações de *Atta cephalotes* e estabelecer uma relação filogenética entre essas populações. Procedeu-se à extração de DNA genômico total, seguida de amplificação com iniciadores específicos e purificação dos amplicons. O produto purificado foi seqüenciado no sentido direto e reverso e consenso das seqüências foi submetido ao alinhamento múltiplo, análise de máxima parcimônia, análise de haplótipos e divergência genética. Os resultados mostraram que existem duas populações geneticamente distintas de *Atta cephalotes*, uma predominantemente de ocorrência amazônica e outra de Mata Atlântica, contrariando a classificação atual. Foram encontradas no Pará e Amapá dois grupos de populações: um se agrupa com o grupo de espécimes de Mata Atlântica (Bahia e Pernambuco) e o outro com os espécimes da Amazônia, sugerindo a existência de populações simpátricas nesses estados. Talvez o uso de outros marcadores, associados a estudos ecológicos e comportamentais, esclareça se nesses dois estados suas populações estão isoladas reprodutivamente. Possivelmente o grupo da Bahia deve ser classificado como uma espécie diferente do grupo da região Norte

Palavras-chave: *Atta cephalotes*, DNA mitocondrial, populações simpátricas, filogenia.

Introdução

A tribo Attini compreende 13 gêneros de distribuição exclusivamente americana (Kempf, 1972; Weber, 1972; Brandão & Nunes, 2001), desde o sul da América do Norte até ao sul da América do Sul (Hölldobler & Wilson, 1990), cuja origem se deu em regiões tropicais, possivelmente no Norte da América do Sul e também em regiões litorâneas, ambas quentes e úmidas (Weber, 1972).

O trabalho de Weber (1972) indica a Região Amazônica como o provável centro de origem das atíneas e das primeiras especiações resultantes da dispersão ou isolamento geográfico, de modo que a presença dessas formigas na Argentina e em países vizinhos sugere que tenha havido o surgimento precoce de alguns gêneros que dispuseram de tempo suficiente para se ajustar às condições climáticas dos habitats secundários.

As rotas de migração sugeridas para os Attini a partir de seu centro de origem na Região Amazônica são (1) para o Norte da América do Sul, através do litoral, onde a Bacia Amazônica se comunica com o litoral do Caribe ou através do interior da Bacia Amazônica; (2) para o Sul através da Bacia do Rio Paraná; e (3) para América Central após o fechamento do istmo do Panamá, por onde as formigas atingiram a América do Norte (Weber, 1972).

Estas rotas supostamente levaram o gênero *Atta* a expandir sua ocupação para o Sul dos Estados Unidos, para o Nordeste, Sudeste e Sul do Brasil, Argentina, Leste do Paraguai e todo o Uruguai, sendo algumas destas regiões no Sul propostas como centros secundários de evolução da tribo na América do Sul (Kusnezov, 1963).

As formigas do gênero *Atta* (Attini: Myrmicinae) (Hölldobler & Wilson, 1990), conhecidas como saúvas, destacam-se pela sua alta capacidade de ocupação territorial relacionada ao seu poder de destruição de lavouras. Cortam e transportam folhas, flores e sementes para seus ninhos, para utilizar como substrato para cultivar o fungo do qual se alimentam (Hölldobler & Wilson, 1990), tornando-as pragas das áreas cultivadas, florestais e de pastagens desde América do Norte até a América do Sul (Della Lucia *et al.*, 1993). Causam prejuízos econômicos e danos que são diretamente proporcionais à área ocupada (Fowler *et al.*, 1986). O sucesso dessas formigas está diretamente associado ao fato de seu fungo crescer sobre praticamente todos dos tipos de vegetação existentes e, como seus ninhos são feitos no solo, acabam evitando predadores e competidores (Weber, 1972). E ainda, trabalhos mostram que as saúvas vem se adaptando bem aos ambientes urbanos. Segundo Weber (1972), praças e parques das grandes cidades são grandes atrativos, uma vez que possuem vegetação e áreas de nidificação em abundancia. Em seu trabalho, Oliveira & Campos-Farinha (2005) verificaram que *Atta sexdens rubropilosa* e *Atta capiguara* eram as principais espécies responsáveis por danos a jardins, hortas, campos de futebol e

praças. Segundo Mariconi (1970), causam danos às construções (prédios, represas, pontes, pontilhões, mausoléus e túmulos) e estradas (rodoviárias e ferroviárias). A origem da dominância, que resultou na massiva presença das saúvas nos diversos ecossistemas das Américas é tema de grandes debates e foi abordada no presente estudo.

No Brasil ocorrem dez espécies e três subespécies do gênero *Atta*: *Atta bisphaerica*, *A. capiguara*, *A. cephalotes*, *A. goiana*, *A. laevigata*, *A. opaciceps*, *A. robusta*, *A. sexdens*, *A. sexdens rubropilosa*, *A. sexdens piriventris*, *A. silvai* e *A. vollenweideri* (Della Lucia *et al.*, 1993). Algumas dessas espécies ocorrem regionalmente, enquanto outras têm uma distribuição mais abrangente. Neste último caso, *Atta laevigata* (F. Smith, 1858), *Atta sexdens* (Forel, 1908) e *Atta cephalotes* (L., 1758) destacam-se por atingirem grandes extensões da América do Sul. Por esta razão estas espécies são modelos importantes para estudos sobre a evolução das atíneas (Weber, 1972).

Atta cephalotes é praga de várias culturas como o cacau e algumas plantas cítricas e vem causando enormes prejuízos no Brasil e nos países em que ocorre. Em Trinidad, causa uma perda média de US\$5,00 por hectare das culturas de plantas cítricas (laranja e limão), gerando uma perda total de 25 % da produção anual (Cherrett & Jutsum, 1982). Na Bahia, as lavouras de Cacau são alvos de sérios ataques que eliminam, na maioria das vezes, uma área foliar considerável, o que no caso de cacauzeiros novos pode levar a morte da planta e à necessidade de replantio. As plantas da mata cabrucada também são alvos de ataques, o que pode prejudicar o sombreamento do cacauzeiro (Abreu & Delabie, 1986).

A distribuição de *Atta cephalotes* abrange a Bolívia, Costa Rica, Colômbia, Equador, Guiana, México, Nicarágua, Panamá, Peru, Suriname, Trinidad e Venezuela (Borgmeier, 1950; Mariconi, 1970) e, no Brasil, nos estados do Amazonas, Rondônia, Roraima, Pará, Amapá, Maranhão, Pernambuco e Sul da Bahia, com provável ocorrência no Acre e Norte do Mato Grosso (Mariconi, 1970; Della-Lucia *et al.*, 1993) e recentemente, teve sua ocorrência verificada em Alagoas (Corrêa, *et al.*, 2005).

As principais características morfológicas de *Atta cephalotes*, popularmente conhecida como “Saúva-da-mata”, são os soldados com um tufo de pelos finos na fronte, vértice da cabeça brilhante e livre de pelos e gáster opaco e coberto de cerdas finas deitados (Mariconi, 1970).

Segundo Gonçalves (1942), *A. cephalotes* é composta pelas seguintes subespécies e variações: *A. cephalotes* var. *opaca* (Forel, 1904), *A. cephalotes* var. *integrior* (Forel, 1904), *A. cephalotes* subsp. *polita* (Emery, 1905), *A. cephalotes* subsp. *isthmicola* (Weber, 1941), *A. cephalotes* subsp. *oaxaquensis* (Gonçalves, 1942) e *A. cephalotes* subsp. *erecta* (Santhschi, 1929). Estas subespécies e variações formam o “Grupo Cephalotes” (Emery, 1922 *apud* Gonçalves, 1942).

Borgmeier (1959) considera *Atta cephalotes* uma única espécie e sinonimizou as outras seis subespécies e variações citadas acima. Entretanto, as características morfológicas que separam cada subespécie apresentam diferenças muito sutis, fazendo com que estas subespécies não sejam de fato elevadas a tal categoria, ficando apenas como sinonímia de *Atta cephalotes* (L.), cuja espécie-tipo foi coletada em Paramaribo, no Suriname (Borgmeier, 1939; Agosti, 2004).

Segundo Gonçalves (1942), a operária máxima da espécie tipo (Suriname) atinge 13 mm de comprimento e a cabeça, 5,5 mm de largura. Já a operária máxima *Atta cephalotes oaxaquensis* (espécime vindo do México) possui uma pontuação densa na parte superior da cabeça que a torna fosca, além da maior altura dos tufos de pêlos frontais, se assemelhando muito a *A. cephalotes opaca*. *A. cephalotes isthmicola* (espécime da Colômbia) tem a operária máxima com os tufos frontais muito altos, a parte superior da cabeça muito pontuada, e as operárias médias e menores com o tórax brilhante e de cor castanha escura, contrastando com a cor castanha clara amarelada da cabeça, do gáster e dos tarsos. *Atta cephalotes opaca* (Colômbia) possui a operária máxima com 13 mm de comprimento, a cabeça mede 3,5 mm de espessura da frente ao occiput e apresenta duas saliências ou bossas frontais bem visíveis de cima, que fazem parte dos lobos cefálicos; nisto difere bastante das outras subespécies e da espécie típica. *A. cephalotes integrrior* (Bahia) possui operárias máximas com tufos frontais pouco pilosos, não atingindo os olhos ao estenderem para os lados; os espinhos mesonotais anteriores são cônicos, pontiagudos, bem mais curtos que os epinotais, sendo estes mais voltados para cima que na espécie típica.

Muito da controvérsia sobre as divisões e denominações dos subgrupos dentro do Grupo “cephalotes” resulta do embasamento da sistemática em caracteres morfológicos, que são polimórficos nas diferentes castas de um mesmo ninho e, considerando-se a mesma casta de ninhos distintos, são muitas vezes escassos. Para a utilização consistente dos caracteres morfológicos na sistemática das atíneas, é necessário analisar as formigas operárias juntamente com suas respectivas formas sexuadas (Borgmeier, 1950). Esta premissa é dificultada pelo fato das castas sexuadas aparecerem uma vez por ano em épocas que variam de acordo com a latitude na grande extensão geográfica ocupada pela espécie.

Esta amplitude de distribuição e conseqüente isolamento geográfico, devido ao Cerrado que funciona como barreira entre as populações amazônicas e de Mata Atlântica de *Atta cephalotes*, é um dos fatores que podem levar a especiação (Futuyma, 1992), de modo que é provável que esta seja um complexo contendo várias espécies que são morfológicamente ainda muito parecidas, dificultando sua diferenciação.

O uso de outros métodos alternativos a análise morfológica tem apresentado bons resultados em elucidar tais questões. O DNA mitocondrial apresenta características interessantes do ponto de vista de estudos evolutivos. Possui herança exclusivamente maternal, o que impede a recombinação, alta taxa evolutiva e polimorfismos intra-específicos, tornando-o uma molécula eficaz em estudos de filogeografia (Avisé *et al.*, 1987). Algumas hipóteses foram lançadas para tentar explicar a alta taxa evolutiva do DNA mitocondrial: (1) relaxamento funcional, uma vez que não codifica proteínas diretamente envolvidas em sua própria replicação e transcrição e por codificar apenas 13 polipeptídeos, é capaz de tolerar certa falta de precisão durante a tradução; (2) a alta taxa de mutação devido à ineficiência dos mecanismos de reparo, exposição a radicais livres no interior da mitocôndria e (3) não estar associado a proteínas histônicas como o DNA nuclear (Avisé, 2001).

Resultados preliminares apontam uma região muito apropriada para a diferenciação e estudos evolutivos de subespécies e populações de *Atta cephalotes*. Trata-se a do espaçador intergênico (IGS), localizado entre os genes mitocondriais codificadores do Citocromo-oxidase I (COI) e de RNA transportador de Leucina (Leu-tRNA). O IGS possui grande variação entre espécies e subespécies de formigas atíneas e é caracterizado por apresentar um alto conteúdo de A/T e baixo índice de homoplasia e tem sido utilizado em nosso laboratório como um marcador molecular que permite encontrar padrões de diversificação biológica e evolução de subespécies de *Atta sexdens*, permitindo inferências filogenéticas (Bacci *et al.*, 2001).

O presente estudo teve como objetivo caracterizar os loci mitocondriais COI, IGS, Leu-tRNA e COII de populações de *Atta cephalotes* e estabelecer uma relação filogenética entre essas populações.

Material e Métodos

Coleta das amostras

Foram coletados 68 espécimes em diferentes localidades no Brasil, Colômbia e uma localidade na Venezuela, identificadas e armazenadas em micro tubos marcados com os códigos de cada ninho e contendo etanol 70-100% em ultra freezer a -80° C. Os locais de coleta e as respectivas coordenadas geográficas bem como os códigos dos ninhos estão descritos na Tabela 1.

Extração de DNA, Amplificação de PCR e Purificação dos Produtos de PCR

Para a extração do DNA total, operárias menores foram trituradas em micro tubos de 1,5 mL contendo 500 uL da solução de lise celular TNES (Tris Base 250 mM pH 7,5; NaCl 2M; EDTA 100 mM; SDS 2%), em seguida foram acrescentados 5 uL de Proteinase K (USB - 76230Y) (20mg/mL) e incubado por 3 horas a 55° C. Em seguida foi feito o tratamento com 5 uL de RNase A (Calbiochem - 556746) (4mg/mL) e incubado por 30 minutos. A precipitação das proteínas foi feita a adição de 200 uL de NaCl 5M. O sobrenadante contendo o DNA foi transferido para um novo micro tubo e precipitado com igual volume de isopropanol 100 % e lavado com etanol 70 %, re-centrifugado e seco em centrífuga a vácuo. O *pellet* de DNA foi dissolvido em 35 uL de TE.

As amostras foram amplificadas usando os iniciadores desenhados para a espécie *Atta cephalotes* ANT-F (5' – ATTCATTCTTATCTTGAAATATTATTTTC – 3') e ANT-RG1.1 (5' – CAATAGTTTGATTTTCTAG – 3') para os *loci* Citocromo Oxidase I (COI), Espaçador Intergênico (IGS) RNA Transportador de Leucina (Leu-tRNA) e Citocromo Oxidase II (COII). A PCR foi feita utilizando o *Kit PureTaq Ready-To-Go PCR Beads* (GE Healthcare – 27-9557-01), utilizando-se 1,5 uL de cada iniciador (6 pmol/uL), 0,5 uL de DNA, 18uL de água ultra-pura, num volume final de 25 uL. O termociclador foi programado para iniciar uma desnaturação inicial 5 min a 94° C seguida de 40 ciclos de 1 min a 94° C, 1 min a 47° C, 2 min a 68° C e terminando com uma extensão final de 15 min a 68° C.

Os produtos de PCR foram purificados utilizando o *Kit GFX PCR DNA and Gel Band Purification* (GE Healthcare – 27-9602-01).

Caracterização dos loci COI, IGS, Leu-tRNA e COII

A caracterização dos *loci* mitocondriais quanto ao número e posição de substituições sinônimas e não sinônimas, inserções e deleções, mudanças de fase de leitura foi feita analisando separadamente cada uma das duas regiões codificadoras (COI e COII) e a região do Leu-tRNA, através do programa DAMBE (Xia & Xie, 2001).

Análises Filogenéticas

As seqüências obtidas foram editadas com o aplicativo BioEdit 7.0.5 (Hall, 1999) e alinhadas com a programa ClustalW (Thompson *et al.*, 1994). Como grupo externo foi utilizada a espécie *Atta colombica*. Para a obtenção das árvores as seqüências foram submetidas à análise de Máxima Parcimônia (MP) com o aplicativo POY 3.0.11 (Gladstein & Wheeler, 1997; Wheeler & Gladstein, 2000; Wheeler *et al.*, 2003) através de 1000 buscas heurísticas, o mesmo peso para todos os caracteres e considerando os *gaps* como quinto estado de caráter. As árvores mais parcimoniosas (AMP) foram identificadas através de busca heurística e o suporte dos nós para as AMP foi obtido através de análise de *bootstrap*

(BS) (Felsenstein, 1985) com 1000 pseudo-réplicas, utilizando o programa PAUP4b10 (Swofford, 2003) e Índice de Bremer (BI) (Bremer, 1994), utilizando o programa POY 3.0.11 (Gladstein & Wheeler, 1997; Wheeler & Gladstein, 2000; Wheeler *et al.*, 2003). A árvore gerada foi visualizada através do programa TreeView (Page, 1998). Para o cálculo de divergência nucleotídica entre as populações foi utilizado o teste de Kimura-2-Parâmetros (K2P) do programa MEGA 3.1 (Kumar *et al.*, 2004). Também foi gerada uma rede de haplótipos por *Median-Joining* através do programa NETWORK 4.0.12 (Bandelt *et al.*, 1999) utilizando a configuração padrão.

Resultados

Caracterização dos loci COI, IGS, Leu-tRNA e COII

Foram obtidos 515 pares de base, sendo 142pb em COI, 160 no IGS, 74 no Leu-tRNA e 140 na região COII.

A região COI apresentou 142 nucleotídeos que codificam 46 aminoácidos e nenhuma substituição não sinônima. A Figura 2A apresenta o gráfico de saturação de transições e transversões em função da divergência (matriz de distância K80) gerado pelo programa DAMBE (Xia & Xie, 2001) em que é possível verificar que as transições tendem a estabilizar devido a saturação enquanto as transversões ainda estão em uma ascendente. Isso denota o menor número de transversões em relação ao de transições. A proteína predita para esta região apresentou 100 % de identidade com a seqüência de *Atta cephalotes* depositada no *GeneBank* (acesso AF016016). Os espécimes coletados na Amazônia colombiana apresentam um substituição de timina por citosina na posição 8 em relação aos demais espécimes coletados.

A região IGS se mostrou variações consideráveis quanto ao tamanho e composição nucleotídica, que poderíamos considerar como sendo uma “assinatura molecular” de cada população, possibilitando agrupar os espécimes segundo essas assinaturas que estão destacadas na Figura 1. É possível separar os espécimes em duas populações principais, uma Amazônica e outra de Mata Atlântica. Entretanto, não foi possível estabelecer uma correlação geográfica perfeita, uma vez que algumas populações Amazônica (Pará e Amapá) formaram um clado irmão ao correspondente aos espécimes de Pernambuco e Bahia (Mata Atlântica), sugerindo possível caso de simpatria.

A região do Leu-tRNA se mostrou totalmente conservada, sem qualquer substituição, composta por 74 nucleotídeos e identidade de 98% com a seqüência de *Atta cephalotes* do Panamá depositada no *GeneBank* (Acesso AF016016).

A região COII 146 nucleotídeos que codificam 49 aminoácidos e apresenta algumas substituições nucleotídicas não sinônimas (Tabela 2). Já a Figura 2B mostra número de

transições e transversões em função da divergência nucleotídica (matriz K80) gerado pelo programa DAMBE (Xia & Xie, 2001), em que a curva que representa as transições começa uma tendência de estabilização e a curva que representa as transversões ainda tem uma tendência de aumento, sugerindo poucas transversões. A proteína predita para esta região apresentou 97% de identidade com a seqüência do Panamá depositada no *Gene Bank* (Acesso AF016016).

As regiões COI e COII apresentaram poucas substituições e, somente em COII, algumas delas não foram sinônimas, o que é esperado para regiões codificadoras (Moritz *et al.*, 1987). E ainda, não apresentaram inserções e/ou deleções que causassem mudança de fase de leitura nem códons de terminação prematura.

Análises Filogenéticas

O cladograma (Figura 3) e rede de haplótipos (Figura 4) mostraram que *Atta cephalotes* está dividida em duas clades principais: a clade A (BS=100 / BI=105) que contém duas outras clades, uma agrupando espécimes da Colômbia e outra agrupando Venezuela, Amazonas, Belém e Alenquer e Alta Floresta (Mato Grosso – Brasil) (BS=69 / BI=93) e a clade B (BS=84 / BI=94) que também contém outras duas clades, uma agrupando espécimes da Bahia, Pernambuco e Macapá (BS=74 / BI=94) e outra agrupando espécimes de Belém, Alenquer, Santarém e Macapá (BS=100 / BI=110). No Amapá parecem existir duas populações isoladas reprodutivamente apesar de estarem localidades bem próximas, uma delas se agrupando com espécimes de Bahia e Pernambuco e outra com espécimes de Santarém, Alenquer e Belém. O mesmo parece ocorrer com os espécimes do Pará, em que uma população se agrupa formando uma clade irmã a clade de Bahia, Pernambuco e Macapá e a outra se agrupa formando uma clade irmã com o grupo de Alta Floresta que por sua vez, juntas, formam uma clade irmã com os espécimes da Colômbia. A Tabela 3 apresenta as divergências nucleotídicas entre as clades, segundo a Figura 3, sendo possível verificar que as clades A e B um alto grau de divergência genética relativamente alto ($4,8 \% \pm 1$). As populações 2 (Venezuela, Alenquer, Belém e Amazonas) e 5 (Alenquer, Belém, Macapá e Santarém) também tiveram um alto grau de divergência ($4,2 \pm 0,9$) quando comparadas com a distância entre elas e suas clades irmãs, que foi de $1,3 \pm 0,5$ e $0,6 \pm 0,3$, respectivamente.

Discussão

As regiões COI e COII apresentaram poucas substituições e, somente em COII, algumas delas não foram sinônimas e de 3ª posição. A baixa quantidade ou ausência de substituições não sinônimas é esperada para regiões codificadoras (Moritz *et al.*, 1987). E ainda, não apresentaram inserções e/ou deleções que causassem mudança de fase de leitura nem códons de terminação prematura, excluindo a possibilidade de relações filogenéticas inverídicas causadas pelo uso inadvertido de pseudogenes (Bensasson, 2001; Nguyen *et al.*, 2002; Benesh *et al.*, 2006). A região IGS apresentou algumas “assinaturas moleculares” que permitiram agrupar os espécimes de acordo com estas assinaturas, entretanto, sem correlacioná-las com sua localização geográfica.

As duas populações distintas apresentadas no cladograma (clades A e B) e na rede de haplótipos, divergem da proposta de Borgmeier (1959), que havia considerado *A. cephalotes* espécie sem denominação subespecífica. O fato de populações tão próximas de Alenquer, Belém e Santarém (Pará) agruparem em clades distintas sugere a existência de algum tipo de isolamento que não permite o fluxo gênico entre elas. Ou ainda, no caso das populações do Macapá (Amapá), estas podem ter sofrido pressões ambientais similares aos espécimes da Bahia e terem derivado para um genótipo parecido. Segundo Avise (2001) e Ballard & Whitlock (2004) o DNA mitocondrial pode apresentar taxas evolutivas variáveis e não esperadas para essa molécula dependendo do tipo de pressão que organismo for submetido no ambiente ao longo do tempo. Segundo Moritz *et al.* (1987) regiões não codificadoras do mtDNA tendem a ser menos conservadas e a apresentar uma taxa evolutiva maior do que regiões codificadoras. Esses dois argumentos talvez justifiquem o fato de a região IGS não ter agrupado os espécimes segundo suas regiões geográficas com precisão. Mesmo o alinhamento feito apenas com COI e COII (não apresentado) formou a mesma topologia obtida na figura 3.

Em sua abordagem sobre “Filogenia Intra-específica e Estrutura Geográfica Populacional”, Avise *et al.* (1987) apresentaram quatro modelos, dos quais dois podem tentar explicar o agrupamento dos espécimes de *A. cephalotes* analisados. O primeiro modelo, categorizado como “Descontinuidade Filogenética e Ausência de Separação Espacial”, embora raro de ser observado, exemplifica um estudo com ratos que apresentavam considerável divergência no mtDNA e ocorriam na mesma localidade. Esse modelo corresponde ao ocorrido com os espécimes de Alenquer, Belém e Santarém, localidades próximas que apresentam populações geneticamente divergentes. O segundo modelo, categorizado como “Continuidade Filogenética e Separação Espacial”, apresenta estudo feito com população de ostras coletadas de Massachusetts até a Geórgia, em que populações relativamente distantes nessa faixa geográfica de amostragem apresentaram

divergência genética muito baixa. Esse modelo corresponderia ao caso observado entre as populações de Macapá e Bahia, distantes geograficamente, porém, geneticamente próximas. Gonçalves (1960) salienta que a distribuição de *A. cephalotes* constitui um problema zoogeográfico, pois esta espécie não consegue viver em regiões secas como o Cerrado e Caatinga, regiões que separam a Mata Atlântica da Floresta Amazônica. Assim sua ocorrência isolada no sul da Bahia seria devido a uma possível uniformidade no clima em eras passadas, possibilitando a migração e o sucesso da espécie na região costeira baiana. Segundo Ab'Saber (2002), na Amazônia existe uma grande diversificação de ecossistemas, com transição simples e linear ou muitas vezes bastante complexa. Talvez os espécimes amostrados no Pará e Amapá estejam em diferentes ecossistemas, com funções ecológicas distintas, causado fluxo gênico baixo ou ausente. Embora também raros, casos de heteroplasmia podem levar a resultados como os obtidos nesse trabalho. Um espécime ou população pode apresentar, ao mesmo tempo, dois genótipos mitocondriais diferentes. Mesmo nestes casos, a diferença entre os genótipos seria muito pequena para causar interferência na análise filogenética (Avise *et al.*, 1987; Avise, 2001).

Estudo recente de Lovato *et al.*, (2005 – não publicado) utilizando 500 pb do gene COI mostrou que a população da Bahia (Mata Atlântica) é geneticamente distinta da população do Norte e da América Central (Floresta Amazônica), embora não tenham analisado espécimes de Pará e Amapá. Segundo Roderick (1996), numa dada população, a frequência de um único haplótipo de mtDNA pode flutuar mais rapidamente do que a dos alelos nucleares, fazendo com que o mtDNA fosse mais sensível a eventos como efeito do fundador e a pequenas populações do que o DNA nuclear. Como consequência, a provável perda ou ganho de um haplótipo de mtDNA será grande para pequenas populações e apenas uma pequena migração seria suficiente para fixar uma linhagem de mtDNA na população. Isso poderia explicar a ocorrência de haplótipos distintos em regiões próximas do Amapá e do Pará. E por se tratarem de populações suficientemente próximas para que haja contato entre elas e, teoricamente, fluxo gênico, é possível que exista alguma barreira ecológica ou microclimática que esteja impedindo ou reduzido drasticamente o fluxo gênico entre essas populações (Slatkin, 1987), de forma a fixar haplótipos diferentes e que conferem melhores características adaptativas

É improvável que tenha havido erro durante a coleta e identificação da localidade bem como durante as análises moleculares. Ainda assim, os espécimes foram processados desde a extração até o sequenciamento por mais de uma vez e sempre apresentaram a mesma seqüência nucleotídica. A contaminação ou uso equivocado de seqüências obtidas a partir de pseudogenes também foi descartado, pois iniciadores específicos para seqüências mitocondriais *Atta cephalotes* foram empregados neste estudo.

Possivelmente, estudos envolvendo outros *loci* mitocondriais e também nucleares, bem como estudos comportamentais e ecológicos, arquitetura de ninho, serão necessários para confirmar se tanto as populações do Macapá (Amapá) quanto as de Belém, Alenquer e Santarém (Pará) estão sofrendo algum tipo de isolamento ou se os genes estudados não são suficientes para diferenciá-las da população da Bahia e agrupá-las juntamente com espécimes de Colômbia e Alta Floresta. Devido aos elevados valores de BS e BI que suportam as duas clades principais, A e B, sugerimos a ocorrência de pelo menos duas populações geneticamente distintas, de *Atta cephalotes*, uma predominante na região Norte outra predominante na Bahia, embora não seja possível estabelecer uma correlação geográfica precisa e também a ocorrência de populações simpátricas no Pará e Amapá. Uma vez que a espécie tipo de *Atta cephalotes* foi coletada no Suriname (Floresta Amazônica), possivelmente a população da Mata Atlântica é uma nova espécie.

Agradecimentos

Agradeço as agências CAPES, CNPq e FAPESP pelo apoio financeiro.

Referências

Abreu, J. M. & Delabie, J. H. C., 1986. Controle das formigas cortadeiras em plantios de cacau. Rev. Theobroma. 16, 199-211.

Ab'Saber, A. N. 2002. Bases para o estudo dos ecossistemas da Amazônia brasileira. Estudos Avançados 16 (45), 7-30.

Agosti, D., 2004, The Social Insects Web.

http://research.amnh.org/entomology/social_insects/. Consultado em 15/01/2006

Avise, J.C.; Arnold, J.; Ball, R.M.; Bermingham, E.; Lamb, T.; Neigel, J.E.; Reeb, C.A. & Saunders, N.C., 1987. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. Ann. Rev. Ecol. Syst. 18, 489-522.

Avise, J.C. Phylogeography: the history and formation of species, 2001, Harvard University Press, Cambridge, Massachusetts

Bacci, M., Silva, A.C.O., Martins, V.G., Pagnocca, F.C., Bueno, O.C., Hebling, M.J.A., 2001. A molecular marker for evolution studies of closely related specimens of higher attine ants. Anais da XXX Reunião da Sociedade Brasileira de Bioquímica e Biologia Molecular, p.37,

Ballard J.W.O. & Whitlock, M. (2004) The incomplete natural history of mitochondria. Mol. Ecol. 13: 729-744.

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37-48

Bensasson D., Zhang D.X, Hartl D.L. & Hewitt G.M., 2001 Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecol.& Evol. 16, 314-321.

Borgmeier, T., 1939. Nova contribuição para o conhecimento das formigas neotrópicas. Rev. de Entomol. 10, 403-428.

Borgmeier, T. 1950. Estudos sobre *Atta* (Hym. Formicidae). Mem. Inst. Oswaldo Cruz. 48, 239-263.

Borgmeier, T., 1959. Revision der gattung *Atta fabricius* (Hymenoptera, Formicidae). *Stu. Entomol.* 2, 321-390.

Brandão, C. R. F. & A. J. Mayhé-Nunes. 2001. A new fungus-growing ant genus, *Mycetagroicus* gen. nov. with the description of three new species and comments on the monophyly of the Attini (Hymenoptera: Formicidae). *Sociobiology* **38**(3B): 639-665.

Bremer K. 1994. Branch support and tree stability. *Cladistics*, 10: 295-304

Corrêa, M.M.; Bieber, A.G.D.; Wirth, R. & Leal, I.R., 2005 Occurrence of *Atta cephalotes* (L.) (Hymenoptera: Formicidae) in Alagoas, Northeastern Brazil, *Neotrop. Entomol.* 34, 695-698.

DELLA LUCIA, T.M.C., FOWLER, H.G. E MOREIRA, D.D.O., 1993. As Formigas Cortadeiras. Viçosa, Universidade Federal de Viçosa.

Emery, C., 1922. Fam. Formicidae, Subfam. Myrmicinae, in *Genera Insectorum*, fasc. 174 B, p. 353, nota 2

Felsenstein J., 1985. Confidence limits on phylogenies: an approach using de bootstrap. *Evol.* 39: 783-791

FOWLER, H.G. E FORTI, L.C., 1986 Economics of grass-cutting ants. *In* C. S. Lofgren and R.K. Vander Meer (eds.), *Fire ants and leaf-cutting ants: biology and management*. West-View Press, Boulder, Colorado, pp. 18-35

Futuyma, D. J., *Biologia Evolutiva*, 2 ed: Ribeirão Preto,

Gladstein D.S. & Wheeler W. C. 1997 POY: the optimization of alignment characters. American Museum of Natural History, New York. Disponível em <ftp://ftp.amnh.org/pub/molecular/poy/>

Gonçalves, C. R., 1942. Contribuição para o conhecimento do gênero *Atta* FABR., das formigas saúvas. *Bol. Soc. Bras. Agronomia.* 5, 333-358.

Gonçalves, C.R. 1960. Distribuição, biologia e ecologia das saúvas. *Divulg. Agronômica* 1, 2-10.

- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nuc. Ac. Sym.. Series. 41, 95-98.
- Hölldobler, B. & Wilson, E.O., 1990. The ants. The Belknap Press of Harvard University Press: Cambridge, MA, USA.
- Kempf, W.W., 1972. Catálogo abreviado das formigas da região Neotropical (Hymenoptera: Formicidae). Stu. Entomol., 1972. 15, 3-344.
- Kumar S., Tamura K., and Nei M., 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief. in Bioinfo. 5,150-163.
- Kusnezov, N., 1963. Zoogeografia de las hormigas en Sudamerica. Acta Zool. Lin. 18, 25-186.
- Lovato L., Vilas-Boas L. A. & Vieira L. G. E. 2005 Caracterização molecular de populações de Atta cephalotes (L. 1758) (Hymenoptera, Formicidae) oriundas da América Central e do Brasil através de análise de sequências do gene mitocondrial COI. Anais do XVII Simpósio de Mirmecologia-Biodiversidade e Bioindicação. pp. 505.
- Mariconi, F.A.M., 1970. As Saúvas.: Editora Agronômica Ceres, São Paulo.
- Moritz, C.; Dowling, T.E. & Brown, W.M., 1987. Evolution of animal mitochondrial DNA: Relevante for population biology and systematics. Ann. Rev. Ecol. Syst., 18, 269-292.
- Nguyen T.T.T., Murphy N.P. & Martin, C. M., 2002. Amplification of multiple copies of mitochondrial Cytochrome b gene fragments in the Australian Freshwater crayfish, *Cherax destructor* Clark (Parastacidae: Decapada). Animal Genetics. 33, 304-308.
- Oliveira M. F. & Campos-Farinha A. E. C., 2005. Formigas Urbanas do Município de Maringá-PR e suas implicações. Arq. do Inst. Biol. 71, 33-39.
- Page, R.D.M., 1998. TreeView: Tree drawing software for Apple Macintosh and Microsoft Windows, versão 1.5.2. Institute of Biomedical & Life Sciences, University of Glasgow, Scotl., UK..

Roderick G. K., 1996. Geographic structure of insect populations: gene flow, phylogeography and their uses. *Annu. Rev. Entomol.* 41, 325-352.

Slatkin M. 1987. Gene flow and the geographic structure of natural populations. *Science.* 236, 787-792

Swofford D.L., 2003. *Phylogenetic Analysis Using Parsimony*, version 4.0b10. Smithsonian Institution, Washington DC.

Thompson, J.D.; Higgins, D.G.; Gibson, T.J., 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gaps penalties and weight matrix choice. *Nuc. Ac. Res.* 22, 4673-4680.

Weber, N.A., 1972. *Gardening Ants, the Attines*. Philadelphia: American Philosophical Society.

Wheeler W., Gladstein D., 2000. POY: the optimization of alignment characters. American Museum of Natural History, New York. Disponível em <ftp://ftp.amnh.org/pub/molecular/poy/>

Wheeler W., Gladstein D. & De Laet J., 2003. POY, phylogeny reconstruction via optimization of DNA and other data version 3.0.11. Command line documentation. Disponível em <ftp://ftp.amnh.org/pub/molecular/poy/>

Xia, X., and Xie. Z., 2001. DAMBE: Data analysis in molecular biology and evolution *J. of Heredity* 92, 371-373.

*** Normas Segundo a revista *Molecular Phylogenetics and Evolution***

Anexos

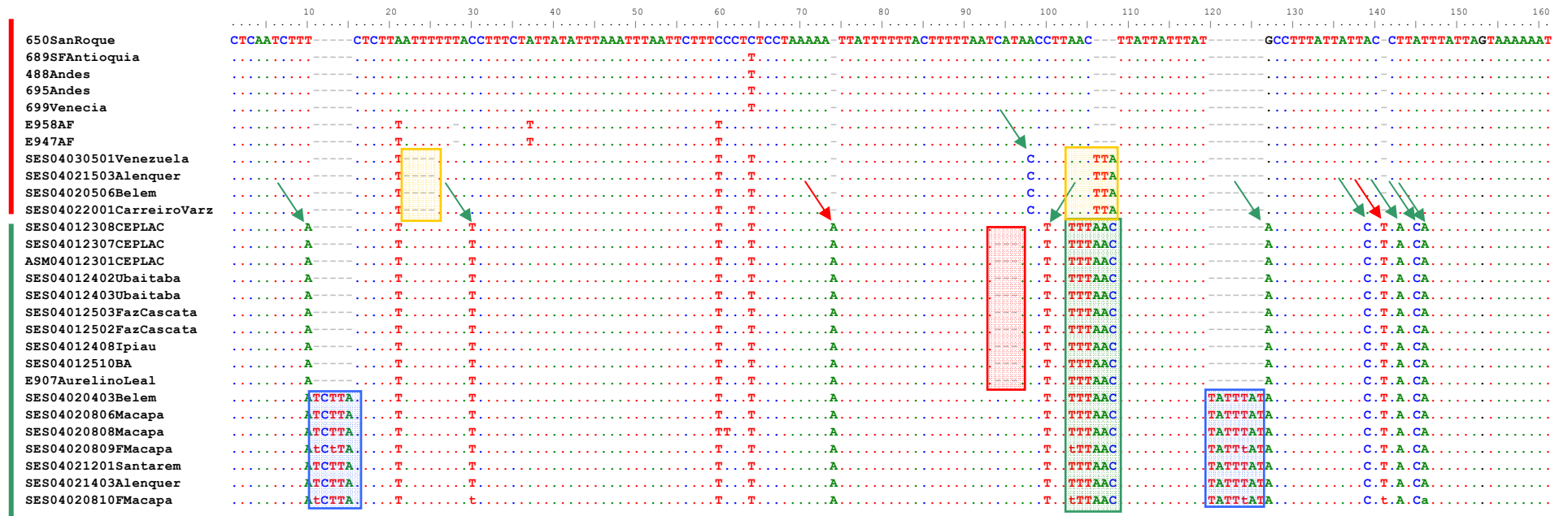


Figura 1: Região IGS, ilustrando as “assinaturas moleculares” de cada população. As barras verticais vermelhas e verdes representam as populações amazônicas e de Mata Atlântica, respectivamente. Os retângulos azuis destacam as “assinaturas” dos espécimes Amazônicos que formam clado irmão ao clado da mata Atlântica. Setas e retângulo verdes representam as substituições que separam os grupos amazônico e de Mata Atlântica e setas vermelhas indicam *indels*. Retângulo vermelho representa deleção de três nucleotídeos exclusiva dos espécimes de Mata Atlântica e retângulos amarelos destacam as “assinaturas” dos espécimes Amazônicos que são simpátricos àqueles que formam o clado irmão ao clado de Mata Atlântica.

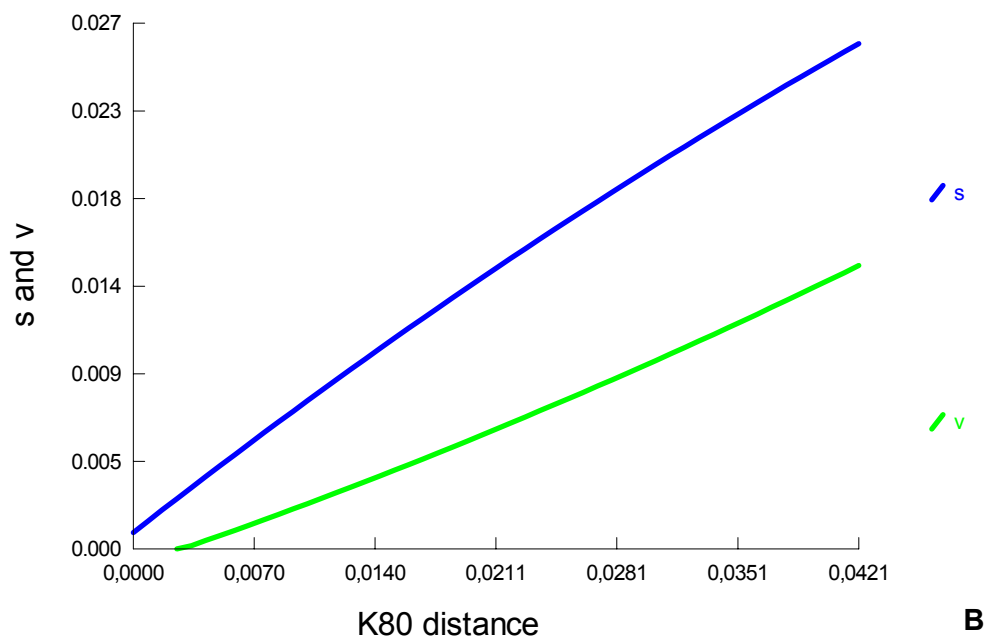
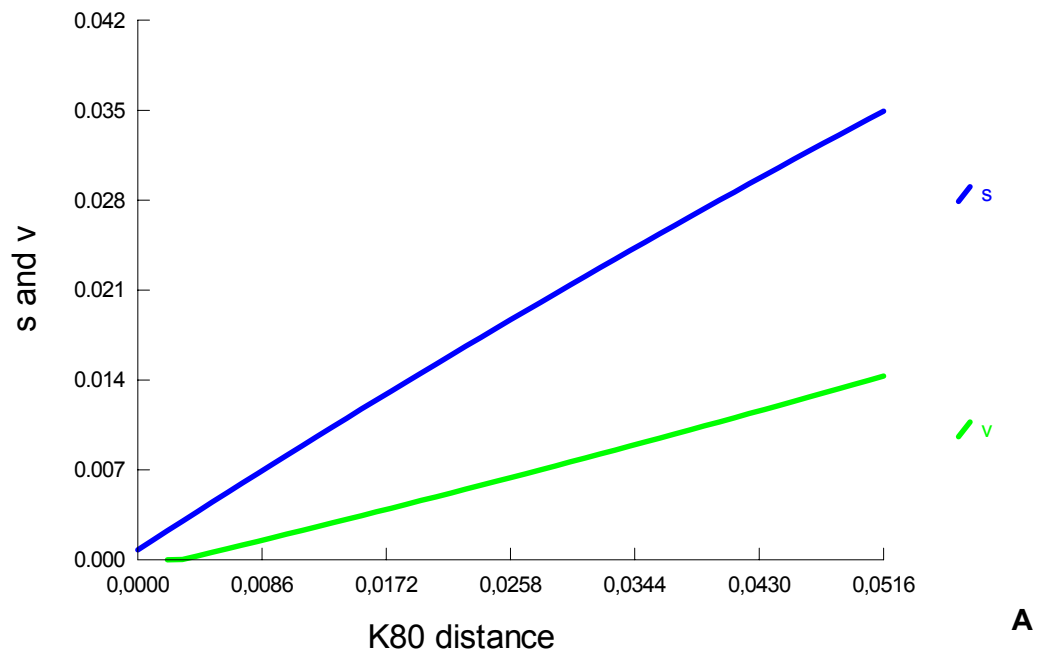
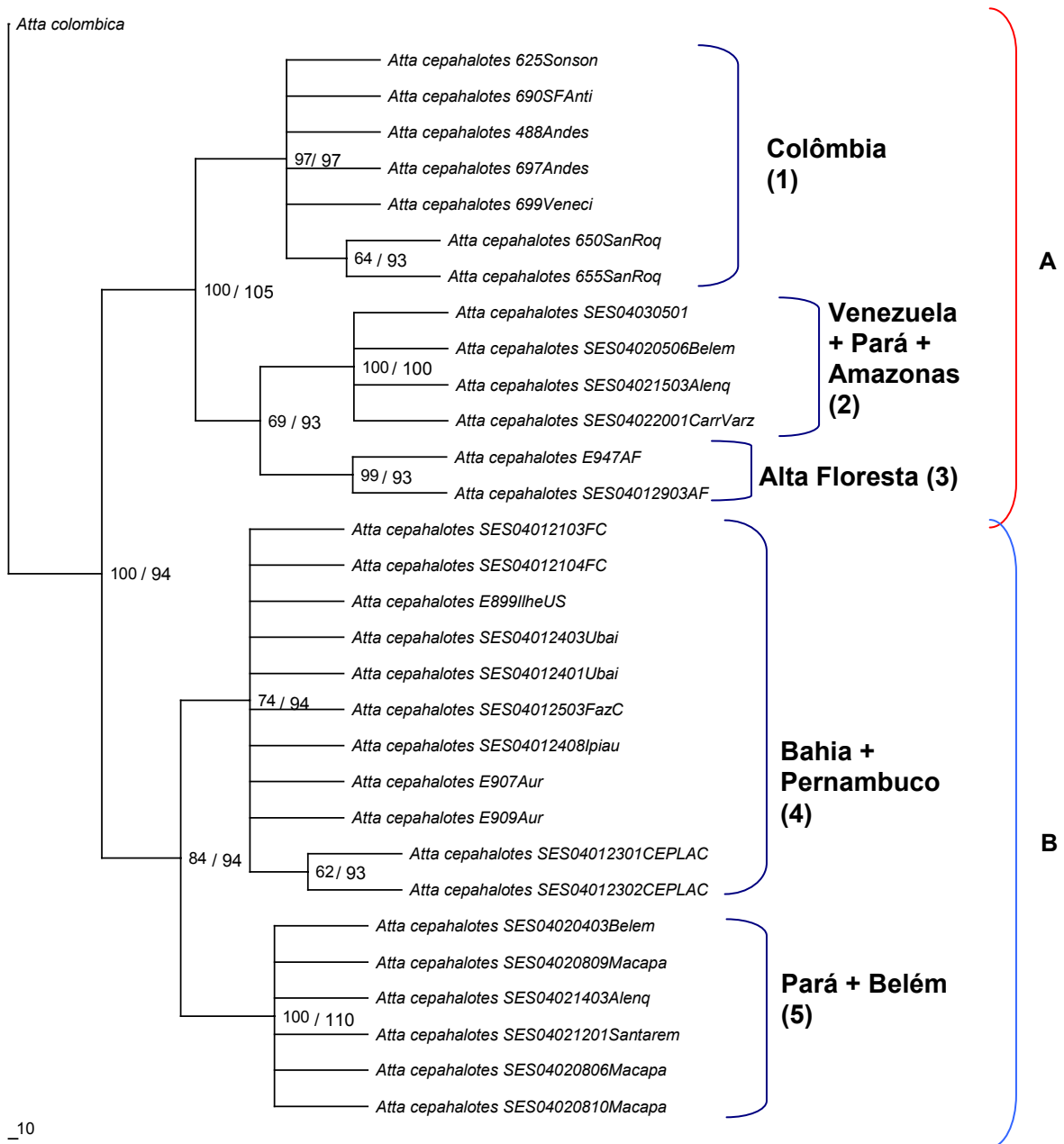


Figura 2: Número de transições (s) e transversões (v) em função da divergência nucleotídica. A – COI e B – COII.



_10

Figura 3: Árvore mais parcimoniosa (com 201 passos) obtidas a partir de 57 caracteres informativos para parcimônia dos *loci* COI, IGS, Leu-tRNA e COII. Os números à frente de cada uma das clades representam os valores de *bootstrap* (BS) e índice de Bremer (BI), respectivamente, obtidos a partir de 1000 buscas heurísticas, cada uma delas com 1000 pseudo-réplicas. Os colchetes indicam as duas principais clades (A – vermelho e B – Azul) e as setas indicam populações próximas geograficamente e que ocupam clades distintas.

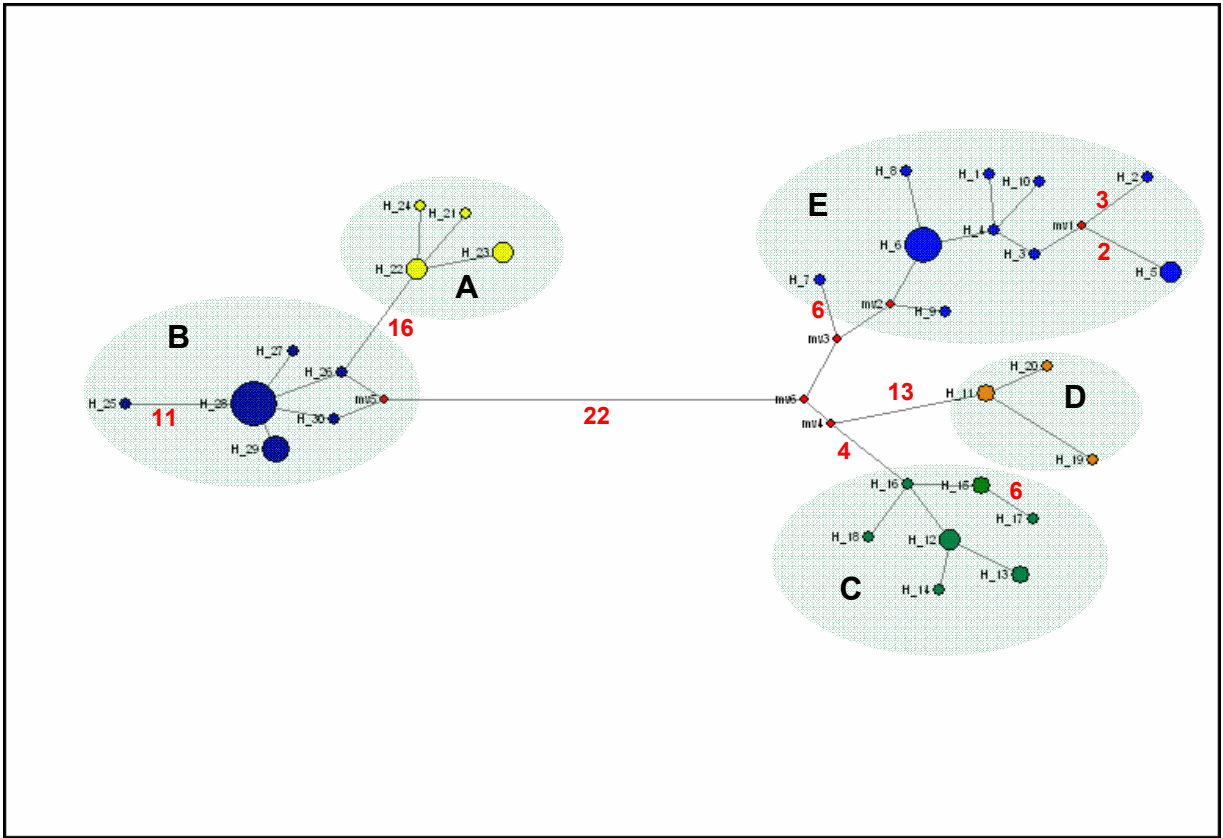


Figura 4: Rede de haplótipos. A – população Santarém + Alenquer + Macapá + Belém; B – Bahia + Pernambuco; C – Alta Floresta; D – Amazonas + Belém + Venezuela + Alenquer; E – Colômbia. Os números entre os haplótipos indicam o número de mutações. Ramos com apenas uma mutação tiveram os valores omitidos.

Tabela 1: Lista de espécimes e a respectiva localidade

Código	Cidade	Estado	Latitude (°)	Longitude (°)
SES040120-01	Frei Caneca	Pernambuco	-8,720383333	-35,84425
SES040120-04	Frei Caneca	Pernambuco	-8,720383333	-35,84425
SES040121-01	Frei Caneca	Pernambuco	-8,720383333	-35,84425
SES040121-03	Frei Caneca	Pernambuco	-8,720383333	-35,84425
SES040121-04	Frei Caneca	Pernambuco	-8,720383333	-35,84425
ASM040123-01	CEPLAC	Bahia	-14,75536111	-39,23255556
SES040123-01	CEPLAC	Bahia	-14,75536111	-39,23255556
SES040123-02	CEPLAC	Bahia	-14,75536111	-39,23255556
SES040123-03	CEPLAC	Bahia	-14,75536111	-39,23255556
SES040123-06	CEPLAC	Bahia	-14,75536111	-39,23255556
SES040123-07	CEPLAC	Bahia	-14,75536111	-39,23255556
SES040123-08	CEPLAC	Bahia	-14,75536111	-39,23255556
SES040124-01	Ubaitaba	Bahia	-14,83975	-39,02688889
SES040124-02	Ubaitaba	Bahia	-14,83975	-39,02688889
SES040124-03	Ubaitaba	Bahia	-14,83975	-39,02688889
SES040124-08	Ipiau	Bahia	-14,09619444	-39,78102778
SES040125-01	Fazenda de Cascata	Bahia	-14,41355556	-39,33088889
SES040125-02	Fazenda de Cascata	Bahia	-14,41355556	-39,33088889
SES040125-03	Fazenda de Cascata	Bahia	-14,41355556	-39,33088889
SES040125-09	farm	Bahia	-14,20027778	-39,81586111
SES040125-10	farm	Bahia	-14,20027778	-39,81586111
SES040129-03	Alta Floresta	Mato Grosso	-9,75642	-55,86272
SES040129-06	Alta Floresta	Mato Grosso	-9,75642	-55,86272
SES040130-02	Alta Floresta	Mato Grosso	-9,59214	-56,02422
SES040131-03	Alta Floresta	Mato Grosso	-10,05337	-55,43224
SES040131-06	Alta Floresta	Mato Grosso	-10,05337	-55,43224
SES040131-12	Alta Floresta	Mato Grosso	-9,86249	-56,07603
SES040131-13	Alta Floresta	Mato Grosso	-9,86249	-56,07603
SES040131-17	Alta Floresta	Mato Grosso	-9,86249	-56,07603
SES040204-03	Belem	Para	-1,68755	-48,54977
SES040205-02	Belem	Para	-1,68755	-48,54977
SES040205-06	Belem	Para	Não informada	
SES040208-06	Macapa	Amapa	0,62059	-51,69178
SES040208-08	Macapa	Amapa	0,61917	-51,70161
SES040208-09	Macapa	Amapa	0,60029	-51,75318
SES040208-10	Macapa	Amapa	0,60082	-51,75435
SES040212-01	Santerem	Para	-2,55767	-54,72733
SES040214-03	Alenquer	Para	-1,90357	-54,64113
SES040214-11	Alenquer	Para	-1,91676	-54,62662
SES040215-03	Alenquer	Para	-1,92631	-54,63676
SES040220-01	Carreiro da Várzea	Amazonas	-3,65538	-60,26173
E958	Alta Floresta	Mato Grosso	-09,55657	-55,99700
E947	Alta Floresta	Mato Grosso	-10,03715	-55,43497
E907	Aurelino Leal	Bahia	-14,4133333	-39,3308333
E899	Ilhéus	Bahia	-14,83975	-39,0268889
E909	Aurelino Leal	Bahia	-14,2017778	-39,82375
SES040131-04	Alta Floresta	Mato Grosso	-10,05337	-55,43224
SES0400305-01	Kukenan Camp (Venezuela)	Bolívar	5,10819	-60,8299
650	San Roque (Colômbia)		6,4683333	-75,0355556
651	San Roque (Colômbia)		6,4691667	-75,0347222
693	Santafé de Antioquia (Colômbia)		6,5352778	-75,8838889
690	Santafé de Antioquia (Colômbia)		6,5333333	-75,8852778

691	Santafé de Antioquia (Colômbia)	6,5333333	-75,8833333
694	Santafé de Antioquia (Colômbia)	6,5358333	-75,8847222
695	Andes (Colômbia)	5,6672222	-75,8508333
697	Andes (Colômbia)	5,6672222	-75,8502778
699	Venecia (Colômbia)	5,9672222	-75,7177778
692	Santafé de Antioquia	6,5333333	-75,8833333
688	Santafé de Antioquia	6,5347222	-75,8844444
689	Santafé de Antioquia	6,5358333	-75,885
625	Sonsón (Colômbia)	5,7008333	-75,4013889
652	San Roque (Colômbia)	6,4691667	-75,0347222
655	San Roque (Colômbia)	6,4691667	-75,0355556
656	San Roque (Colômbia)	6,4691667	-75,0336111
657	San Roque (Colômbia)	6,4691667	-75,0336111
659	San Roque (Colômbia)	6,485	-75,0180556
488	Andes (Colômbia)	5,6669444	-75,8855556
626	Sonsón (Colômbia)	5,7025	-75,4008333

* indica as amostras utilizadas na obtenção da árvore filogenética

Tabela 2: substituições não sinônimas em COII

Código	Posição	Códon	Posição do Códon	Códon	Aminoácido
SES040214-03	16	6	1 ^a	UUC	ILE para PHE
SES040208-08	16	6	1 ^a	UUC	ILE para PHE
SES040204-03	16	6	1 ^a	UUC	ILE para PHE
SES040208-06	16	6	1 ^a	UUC	ILE para PHE
SES040205-02	16	6	1 ^a	UUC	ILE para PHE
SES040212-01	16	6	1 ^a	UUC	ILE para PHE
SES040121-01	16	6	1 ^a	UUC	ILE para PHE
E818	16	6	1 ^a	UUC	ILE para PHE
E899	16	6	1 ^a	UUC	ILE para PHE
E907	16	6	1 ^a	UUC	ILE para PHE
E909	16	6	1 ^a	UUC	ILE para PHE
SES040120-01	16	6	1 ^a	UUC	ILE para PHE
SES040121-03	16	6	1 ^a	UUC	ILE para PHE
SES040123-06	16	6	1 ^a	UUC	ILE para PHE
SES040125-03	16	6	1 ^a	UUC	ILE para PHE
SES040124-02	16	6	1 ^a	UUC	ILE para PHE
SES040120-04	16	6	1 ^a	UUC	ILE para PHE
SES040124-03	16	6	1 ^a	UUC	ILE para PHE
SES040123-08	16	6	1 ^a	UUC	ILE para PHE
SES040125-01	16	6	1 ^a	UUC	ILE para PHE
SES040125-10	16	6	1 ^a	UUC	ILE para PHE
SES040123-01	16	6	1 ^a	UUC	ILE para PHE
SES040123-03	16	6	1 ^a	UUC	ILE para PHE
SES040123-02	16	6	1 ^a	UUC	ILE para PHE
SES040124-01	16	6	1 ^a	UUC	ILE para PHE
SES040208-10	16	6	1 ^a	UUC	ILE para PHE
SES040124-08	16	6	1 ^a	UUC	ILE para PHE
SES040121-04	16	6	1 ^a	UUC	ILE para PHE
SES040125-09	16	6	1 ^a	UUC	ILE para PHE
SES040208-09	16	6	1 ^a	UUC	ILE para PHE
SES040125-02	16	6	1 ^a	UUC	ILE para PHE
ASM040123-01	16	6	1 ^a	UUC	ILE para PHE
SES040123-07	16	6	1 ^a	UUC	ILE para PHE
SES040204-03	50	17	2 ^o	GUC	ASP para VAL
SES040131-06	133	45	1 ^a	GUG	MET para VAL
SES040131-12	133	45	1 ^a	GUG	MET para VAL

* Substituição de ILE para PHE; ** Substituição de ASP para VAL; *** Substituição de MET para VAL

Tabela 3: Divergência nucleotídica entre as clades apresentadas na Figura 3

	1	2	3	4
Colômbia (1)				
Alta Floresta (2)	1,9 ± 0,6			
Venezuela + Amazonas + Pará (3)	1,6 ± 0,5	1,3 ± 0,5		
Pará + Amapá (4)	4,4 ± 0,9	4,5 ± 1,0	4,2 ± 0,9	
Bahia + Pernambuco + Amapá (5)	4,8 ± 1,0	4,9 ± 1,0	4,5 ± 1,0	0,6 ± 0,3

Primeira ocorrência de *Numts* em *Atta cephalotes* (Attini: Myrmicinae)

¹J. Martins Jr, ²S.E. Solomon, ²A. Mycheyev, ²U. Mueller and ¹M. Bacci Jr

¹Centro de Estudos de Insetos Sociais, Instituto de Biociências, Universidade Estadual Paulista, Rio Claro, Brasil.

²Section of Integrative Biology, Patterson Laboratories, University of Texas, Austin, USA.

Corresponding author:

Joaquim Martins Jr., Centro de Estudos de Insetos Sociais, Universidade Estadual Paulista

Av. 24A, n. 1515 – Bela Vista

CEP – 13506-900 – Rio Claro-SP-Brasil

Fone: 55 19 3526 4160

e-mail: martin@rc.unesp.br

Resumo

Os pseudogenes nucleares de origem mitocondrial, ou *numts*, originados a partir de cópias de DNA mitocondrial que migraram para o núcleo estão sendo relatados em um número crescente de trabalhos e para uma diversidade cada vez maior de organismos. Como um desdobramento de nosso estudo de filogenia molecular e sistemática baseados em DNA mitocondrial, nós descrevemos a primeira descoberta de pseudogenes homólogos a Citocromo Oxidase I e II em *Atta cephalotes* e discutimos sua implicação em estudos evolutivos envolvendo esses marcadores. A partir de produtos não específicos e ambigüidade no seqüenciamento direto, amplificamos esses genes com novos iniciadores específicos e detectamos a presença de três grupos gênicos distintos, presentes em cada um dos 68 espécimes analisados. O grupo 1 corresponde ao gene mitocondrial, enquanto os grupos 2 e 3 parecem ser pseudogenes, sendo que o grupo 2 provavelmente é transcrito. O grupo 3 apresenta uma inserção em COI seguida de um duplo códon de terminação, sugerindo que seja um pseudogene que tenha perdido sua função e não seja mais transcrito. As árvores filogenéticas obtidas mostram que esses grupos compõem clades distintas, em que o grupo 3 apresenta uma irradiação mais basal em relação aos espécimes do grupo 1. O grupo 2 aparece como grupo irmão do grupo 1, o qual é considerado o gene mitocondrial funcional de *Atta cephalotes*. Os grupos 2 e 3 apresentam uma taxa evolutiva menor em relação ao grupo 1. Esses resultados, associados às outras descobertas recentes de pseudogenes nucleares de origem mitocondrial em outros organismos, reforçam a necessidade de um cuidado durante a interpretação de resultados de amplificação e seqüenciamento utilizados em estudo evolutivos em *Atta cephalotes*.

Palavras-chave: *Atta cephalotes*, Citocromo Oxidase, pseudogenes, mtDNA, Filogenia

Introdução

O DNA mitocondrial (mtDNA) vem sendo amplamente utilizado em estudos filogenéticos populacionais por apresentar características únicas, como grande número de cópias, alta taxa de mutação, neutralidade, pouca ou nenhuma recombinação, haploidia e herança exclusivamente materna (Avise *et al.*, 1987; Moritz *et al.*, 1987). Por outro lado, problemas como heteroplasmia, ligação paterna, casos de não-neutralidade e uma taxa variável de evolução de suas seqüências podem ser considerados fatores que limitam aplicação do mtDNA para fins de estudos evolutivos (Avise, 2001; Ballard & Whitlock, 2004). Outro problema está nos Pseudogenes, que são seqüências de DNA bastante semelhantes estruturalmente aos genes funcionais e que contêm algumas alterações, como a perda do códon de início, presença de códons de terminação adicionais e ausência ou anormalidade nas seqüências regulatórias que os flanqueiam, tornando-os não funcionais (D'Errico *et al.*, 2004). Nos últimos anos, uma grande diversidade de pseudogenes mitocondriais ou *Numts* (*Nuclear mitochondrial-like sequences*) (Lopez *et al.*, 1994) vem sendo observada em um número cada vez maior de organismos (Bensasson *et al.*, 2001).

Os *Numts* são seqüências nucleares de origem mitocondrial, que apresentam um alto grau de similaridade com as seqüências de mtDNA, que podem ter migrado para o núcleo por transferência direta de DNA (Lopez *et al.*, 1994) ou mediada por RNA (Nugent & Palmer, 1991) em que, possivelmente, elementos transponíveis ou virais estão envolvidos (Farrelly & Butow, 1983; Zullo *et al.*, 1991; Ossario *et al.*, 1991; D'Errico *et al.*, 2004). A maioria dos *Numts* vem sendo relatada em vertebrados, particularmente em mamíferos e aves e relatos sobre sua presença em peixes são praticamente inexistentes (D'Errico *et al.*, 2004), embora isso provavelmente seja devido à atenção desproporcionalmente maior dispensada a esses dois primeiros táxons em relação aos demais grupos. Mesmo organismos largamente estudados como a *Drosophila melanogaster* e o nematódeo *Caenorhabditis elegans*, aparentemente apresentavam poucos *Numts* (Bensasson *et al.*, 2001). Recentemente, Harrison *et al.* (2003) identificaram aproximadamente 100 pseudogenes em *Drosophila melanogaster*, dos quais pelo menos seis seriam pseudogenes processados (originados por transferência de RNA). Assim, a documentação desses pseudogenes é de fundamental importância para a comunidade científica em geral, uma vez que os *Numts* e o mtDNA apresentam diferentes graus de funcionalidade e estão localizados em locais diferentes na célula, sendo, portanto, submetidos à pressões evolutivas distintas. Assim, se seqüências de *Numts* e de DNA mitocondrial forem utilizadas simultaneamente em análises filogenéticas, relações errôneas podem ser propostas (Sorenson & Quinn, 1998). Portanto, é necessário que os *Numts* sejam identificados e evitados em estudos filogenéticos utilizando mtDNA.

O grupo das formigas cortadeiras pertencentes à tribo Attini vem sendo cada vez mais estudado do ponto de vista filogenético, tendo como principal marcador o mtDNA (Wetterer *et al.*, 1998; Silva-Pinhati *et al.*, 2004). Em nosso laboratório, durante estudos de filogenia molecular baseados nos *loci* mitocondriais Citocromo Oxidase I (COI), Espaçador Intergênico (IGS), RNA Transportador de Leucina (Leu-tRNA) e Citocromo Oxidase II (COII) da espécie *Atta cephalotes* Linnaeus (1758) (Myrmicinae, Attini), enfrentamos problemas no seqüenciamento direto. Após a clonagem e seqüenciamento dos produtos de PCR ficou evidente que múltiplas seqüências estavam sendo amplificadas e três tipos de seqüências, então denominadas G1, G2 e G3, foram encontrados, sugerindo a presença de pseudogenes. Por isso novos iniciadores foram desenhados para amplificar cada uma das três seqüências especificamente. Assim, o presente estudo tem como objetivo apresentar a primeira evidência sugerindo a ocorrência de *Numts* em formigas da tribo Attini e discutir suas implicações em estudo de filogenia molecular.

Material e Métodos

Coleta das amostras

As formigas foram coletadas em diferentes localidades no Brasil, Colômbia e uma localidade na Venezuela, identificadas e armazenadas em micro tubos marcados com os códigos de cada ninho e contendo etanol 70-100% em ultra freezer a -80° C. Os locais de coleta e as respectivas coordenadas geográficas bem como os códigos dos ninhos estão descritos na Tabela 1.

Extração de DNA, Amplificação de PCR e Purificação dos Produtos de PCR

Para a extração do DNA genômico total as amostras foram trituradas em micro tubos de 1,5 mL contendo 500 uL da solução de lise celular TNES (Tris Base 250 mM pH 7,5; NaCl 2M; EDTA 100 mM; SDS 2%), em seguida foram acrescentados 5 uL de Proteinase K (USB - 76230Y) (20mg/mL) e incubado por 3 horas a 55° C. Em seguida foi feito o tratamento com 5 uL de RNase A (Calbiochem - 556746) (4mg/mL) e incubado por 30 minutos. A precipitação das proteínas foi feita se acrescentado 200 uL de NaCl 5M. O sobrenadante contendo o DNA foi transferido para um novo micro tubo e precipitado com isopropanol 100 % e lavado com etanol 70 %, re-centrifugado e seco em centrífuga a vácuo. O *pellet* de DNA foi dissolvido em 35 uL de TE.

Inicialmente, as amostras foram amplificadas usando os iniciadores desenhados para a espécie *Atta sexdens* ANT-F (5' – ATTCATTCTTATCTTGAAATATTATTTTC – 3') e ANT-R (5' – TTCATAAGTTTCAGTATCATTGGTG – 3') para os *loci* Citocromo Oxidase I (COI), Espaçador Intergênico (IGS) RNA Transportador de Leucina (Leu-tRNA) e Citocromo

Oxidase II (COII). A reação de PCR foi feita utilizando o *Kit PureTaq Ready-To-Go PCR Beads* (GE Healthcare – 27-9557-01), utilizando-se 1,5 uL de cada iniciador (6 pmol/uL), 0,5 uL de DNA, 18uL de água ultra-pura, num volume final de 25 uL. O termociclador foi programado para iniciar uma desnaturação inicial 5 min a 94° C seguida de 40 ciclos de 1 min a 94° C, 1 min a 47° C, 2 min a 68° C e terminando com uma extensão final de 15 min a 68° C. Devido a inespecificidade dos iniciadores ANT-F e ANT-R, novos foram desenhados a partir das seqüências dos clones em que foram escolhidas regiões (Figura 1) das seqüências que possibilitassem a amplificação específica dos grupos G1, G2 e G3. As seqüências dos novos iniciadores estão descritas na Tabela 2. Também foram utilizados os mesmos kits e condições citados acima, sendo que para o grupo gênico G3 o anelamento dos iniciadores foi de 30 s a 50° C. Os produtos de PCR foram purificados utilizando o *Kit GFX PCR DNA and Gel Band Purification* (GE Healthcare – 27-9602-01.).

Clonagem e Seqüenciamento

Os produtos purificados, inicialmente amplificados com os iniciadores ANT-F e ANT-R, foram ligados ao pGEM-T Vector (Promega), transformados em células de *Escherichia coli* DH10B e analisados em seqüenciamento bidirecional, utilizando-se os iniciadores T7 e SP6. Foram seqüenciados de 2 a 5 clones de 20 indivíduos diferentes em seqüenciador automático ABI377 (Applied Biosystems) utilizando o Kit ABI BigDye Terminator (Applied Biosystems), representando diferentes os locais de coleta (veja Tabela 1). Os produtos de PCR dos 68 espécimes, obtidos com os novos iniciadores, foram submetidos a sequenciamento direto e bidirecional, utilizando o método descrito acima.

Caracterização dos grupos gênicos G1, G2 e G3

A caracterização dos grupos gênicos quanto ao número e posição de substituições sinônimas e não sinônimas, inserções e deleções, mudança de fase de leitura e foi feita analisando separadamente cada uma das duas regiões codificadoras (COI e COII) e a região do Leu-tRNA, através dos aplicativos DNAsp (Rozas *et al.*, 2003), DAMBE (Xia & Xie, 2001) e MEGA 3.1 (Kamur *et al.*, 2004)

Análises Filogenéticas

As seqüências obtidas foram editadas com o aplicativo BioEdit 7.0.5 (Hall, 1999) e alinhadas com a ferramenta ClustalW (Thompson *et al.*, 1994) integrada a este aplicativo, juntamente com seqüências de *Atta cephalotes* de outras localidades. Como grupo externo foi utilizada a espécie *Acromyrmex octospinosus*. Para a obtenção das árvores as seqüências foram submetidas à análise de Máxima Parcimônia (MP) com o aplicativo POY 3.0.11 (Gladstein & Wheeler, 1997; Wheeler & Gladstein, 2000; Wheeler *et al.*, 2003), com o

mesmo peso para todos os caracteres e considerando os *gaps* como quinto estado de caráter. As árvores mais parcimoniosas (AMP) foram identificadas através de busca heurística e o suporte dos nós para as AMP foi obtido através de análise de *bootstrap* (BS) (Felsenstein, 1985) com 1000 pseudo-réplicas, com o programa PAUP4b10 (Swofford, 2003) e Índice de Bremer (BI) (Bremer, 1994) com o programa POY 3.0.11 (Gladstein & Wheeler, 1997; Wheeler & Gladstein, 2000; Wheeler *et al.*, 2003). A árvore gerada foi visualizada através do programa TreeView (Page, 1998).

Resultados

Amplificação específica e características do seqüenciamento

A amplificação com os iniciadores específicos confirmou que os amplicons do grupos G1, G2 e G3 estavam presentes simultaneamente em um mesmo espécime, resultado que se repetiu em todos os 68 espécimes coletados. O seqüenciamento direto realizado a partir dos produtos de PCR obtidos com os iniciadores específicos confirmou a identidade dessas seqüências, correspondendo as mesmas obtidas no seqüenciamento dos clones.

Caracterização dos grupos gênicos G1, G2 e G3

O grupo G1 apresentou as seqüências dos genes COI e COII com poucas substituições nucleotídicas entres os espécimes das diferentes regiões estudadas (Tabela 1) e apenas uma, em COII, não foi sinônima (Aspartico mudando para Valina, Isoleucina mudando para Fenilalanina e Metionina mudando para Fenilalanina) (Figura 1A e B). Ainda em relação a G1, não foram verificadas inserções ou deleções que causassem *frameshifts* nem códons de terminação prematura no meio da seqüência. As seqüências deste grupo submetidas ao *Blastn* apresentaram identidade entre 95 e 99 % em relação a seqüências de *Atta cephalotes* já depositadas (Acesso AF016016) e quando submetidas ao *Blastp* todas apresentaram identidade de 100% com a seqüência de aminoácidos já depositada.

O grupo G2 não apresentou qualquer variação na composição nucleotídica nos diferentes espécimes coletados, sendo idênticos em todos os 68. Quanto aos resultados de *Blastn* e *Blastp*, G2 apresentou 92% de identidade para ambas as seqüências nucleotídicas e de aminoácidos para a região COI e 87 e 95 % para a região COII. Para o gene COI (Figura 1A) foram verificadas 9 substituições em relação as seqüências de G1, sendo 1 transversão e 8 transições, das quais duas não foram sinônimas e para COII (Figura 1B), 14 substituições, sendo 12 transições e 2 transversões, das quais 6 não foram sinônimas.

O grupo G3 também não apresentou qualquer variação nucleotídica nos 68 espécimes analisados. Os resultados de *Blastn* apresentaram identidade de 82 % com a região COI e 85 % com a região COII e para os de *Blastp*, apenas COII apresentou

identidade, que foi de 80 %. Ainda foram verificadas 21 substituições de nucleotídeos em COI (Figura 1A) em relação a G1, sendo 9 transições e 12 transversões, das quais 9 não foram sinônimas, além de uma inserção de 33 nucleotídeos contíguos seguida de um duplo códon de parada. Para o gene COII (Figura 2B), houve 22 substituições, sendo 16 transições e 6 transversões, das quais 10 não foram sinônimas.

Quanto a região do Leu-tRNA, as seqüências de G1 foram 100 % idênticas em todos os espécimes coletados. Comparando G2 e G3 em relação a G1, houveram 2 substituições em G2 (1 transição e 1 transversão) e 6 substituições em G3 (2 transições e 4 transversões) (Figura 2).

Quanto a região IGS, verificou-se apenas o seu tamanho, que foi de 152 pb em G1, 186pb em G2 e 111pb em G3.

Conjuntamente, esses dados sugerem que G1 representa os *loci* mitocondriais enquanto G2 e G3 parecem ser os pseudogenes.

Análise filogenética

O consenso estrito das 2 AMP com 243 passos contendo 17 táxons terminais, está descrito na Figura 4, onde o grupo G3 apareceu como o mais basal em relação aos espécimes de *Atta cephalotes* representados por G1, estando mais próximo do grupo externo representado por *Acromyrmex octospinosus* (com valores de BS/BI de 100 e 407). Em seguida, aparece o grupo G2 como grupo irmão de G1 (gene mitocondrial correto) que representa as duas populações apresentadas no capítulo anterior. (com BS/BI de 100 e 407, respectivamente). Considerando as seqüências de G1 como o *loci* mitocondrial, *Atta cephalotes* esta dividida em dois grupos populacionais, um representando a Colômbia, Venezuela e outro representado o Brasil.

Discussão

Em estudos de filogenia molecular é muito importante ter certeza de estar analisando a seqüência correta homóloga a região alvo de estudo. Ficou evidente que o uso de iniciadores desenhados especificamente para *Atta sexdens* não foi ideal para *Atta cephalotes*, pois o seqüenciamento direto revelou ambigüidades bem como a inespecificidade desses iniciadores, embora já tenha amplificado a seqüência correta, sem ambigüidades, em *Atta laevigata*. Ainda, alguns trabalhos descrevem que os pseudogenes podem ser preferencialmente amplificados em detrimento do mtDNA, pois costumam estar presentes em grande número de cópias, o que no caso de pequenos organismos como os insetos, torna muito difícil separar os *numts* do mtDNA (Bensasson *et al.*, 2000), além de também terem sido descritos para uma gama de outros organismos (Benesh *et al.*, 2006). Ainda, este mesmo autor verificou que ao utilizar iniciadores universais para amplificar o

mtDNA de helmintos, a provável presença de mutações na região de anelamento dos iniciadores, fez com que seu anelamento fosse inespecífico, preferindo os pseudogenes. Em nosso estudo, quando os espécimes foram amplificados com os novos iniciadores especificamente desenhados para cada um dos três grupos gênicos, verificou que cada um dos 68 espécimes eram portadores dos três grupos gênicos. Conforme descrito por Sorenson & Quinn (1998), utilizar iniciadores específicos é um bom método para evitar *numts*, desde que estes e as seqüências mitocondriais possam ser separadas. Em nosso estudo, o uso de iniciadores específicos não só ajudou a separar os grupos gênicos como também permitiu identificar cada um deles através de seqüenciamento de DNA. E mesmo que alguns trabalhos considerem os pseudogenes com um “DNA lixo”, trabalhos recentes demonstram que os pseudogenes podem ser funcionais ou mesmo, a partir de si mesmos, formar novos genes ou assumir novas funções, seja na expressão e regulação gênicas, seja na geração de diversidade genética. (Balakirev & Ayala, 2003)

Várias outras técnicas têm sido propostas para se evitar *numts* e atingir de fato o mtDNA. RT-PCR é uma técnica bastante efetiva para evitá-los (Collura *et al.*, 1996; Williams & Knowlton, 2001; Benesh *et al.*, 2006), embora alguns *numts* sejam transcritos (Blanchard & Schmidt, 1996). Esta técnica pode ser utilizada em estudos futuros para verificar se G2, e mesmo G3, ainda são transcritos. Ainda, como forma mais de evitar os *numts* seja isolar e purificar a mitocôndria antes da amplificação (Zhang & Hewitt, 1996). Também, o uso de tecidos ricos em mtDNA, como o muscular, pode aumentar a proporção de mtDNA num extrato de DNA genômico (Greenwood & Paabo, 1999). Infelizmente, ao se trabalhar com insetos pequenos como as formigas, mesmos as do gênero *Atta* que são relativamente maiores, aumentar a proporção de mtDNA amplificado pode ser extremamente difícil (Benesh *et al.*, 2006) devido ao seu tamanho e a necessidade de usar o organismo inteiro na extração.

O grupo G1 não apresentou qualquer característica descrita para os pseudogenes, o que faz acreditar que esse grupo de seqüências seja realmente o mitocondrial e que o iniciador foi realmente efetivo em amplificá-lo. Todas as substituições e variações nucleotídicas verificadas nos 68 espécimes foram sinônimas, não houve qualquer códon de terminação prematura nem *frameshifts*. As proteínas codificadas por COI e COII apresentam 100 % de identidade com as proteínas preditas para esse genes que estão depositadas no *GeneBank* (Acesso AF016016).

G2 apresenta algumas substituições em COI e COII que causam alterações no aminoácido predito para esses genes, embora não apresente códons de terminação prematura nem *frameshifts*. Isso, associado a uma irradiação mais recente desse grupo, sugere que as seqüências de G2 ainda podem ser transcritas, até mesmo em proteínas funcionais (Blanchard & Schmidt, 1996). Outra hipótese, embora a localização assumida

para os *numts* seja o núcleo, seria a de que esse grupo reflita a existência de heteroplasmia, isto é, genomas mitocondriais distintos em um mesmo indivíduo (Avise, 2001). Entretanto, por ser todas idênticas nos 68 espécimes, essas seqüências não corresponderiam a alta taxa evolutiva prevista para o mtDNA (Moritz *et al.*, 1987; Avise *et al.*, 1987; Avise, 2001), além de que múltiplas cópias funcionais do genoma mitocondrial poderia levar a replicação desemparelhada, o que provavelmente seria eliminado rapidamente pela seleção natural (Moritz *et al.*, 1987; Mirol *et al.*, 2000).

G3 foi o grupo que apresentou menor similaridade nucleotídica e de aminoácidos em relação a G1 e também as seqüências de *Atta cephalotes* do *GeneBank*. Não apresentou identidade com a proteína predita para a região COI e a similaridade nucleotídica para essa região foi de 82 %. Ainda em COI apresentou uma inserção contígua de 33 nucleotídeos seguida por um duplo códon de parada. Em COII, a identidade com a proteína predita foi de apenas 80 % e a nucleotídica foi de 82 %, além de apresentar uma deleção contígua de 4 nucleotídeos que causou *frameshifts* e alteração no aminoácido codificado. Esses dados, associados a irradiação basal desse grupo em relação ao grupo G1 (com *bootstrap* de 100) na topologia apresentada suporta a teoria de que esse grupo seja um pseudogene mais antigo e de baixa taxa evolutiva e que já não é mais transcrito, embora isso precise ser confirmado.

Embora, a princípio, os *numts* se tornem um inconveniente para estudos filogenéticos, eles podem ser utilizados como fósseis moleculares. Em sua revisão, Bensasson *et al* (2001) apontam que em humanos, a taxa de mutação no DNA nuclear é bem menor do que a mitocondrial e que os *numts*, por estarem no núcleo, parecem estar congelados em relação a seus homólogos mitocondriais. Salienta ainda que em insetos, essa diferença mutacional não é tão evidente, podendo os *numts* variar numa maior, menor ou similar taxa evolutiva em relação ao mtDNA, dependendo da pressão seletiva. Talvez G3 possa ser utilizado com um marcador fóssil, pois apresenta uma irradiação basal em relação aos espécimes de *Atta cephalotes* correspondentes a G1 e mesmo em relação ao gênero *Atta* e se sua presença for constatada em outras espécies do gênero, ele poderia ser utilizado como um marcador evolutivo.

A presença desses pseudogenes em populações de *Atta cephalotes* separadas por uma distância geográfica considerável sugere que, caso estejam núcleo, essa integração tenha ocorrido no princípio da expansão do gênero. Segundo a Figura 3, G2 e G3 apareceram em momentos diferentes, sugerindo que G3 esteja numa posição mais basal e mais relacionado ao grupo externo (*Acromyrmex octospinosus*) e G2 surgindo posteriormente, no grupo de *Atta cephalotes*.

Assim, reconhecer a presença de pseudogenes é de fundamental importância para se evitar propostas evolutivas e conclusões erradas. Se isso tivesse acontecido em nosso

estudo filogenético de *Atta cephalotes*, poderíamos ter considerado o grupo formado como uma linhagem mitocondrial distinta, o que poderia ser considerado uma nova espécie, fazendo com que biogeografia e o relacionamento com as outras espécies fosse mal interpretado. Visto que o mtDNA é uma ferramenta que provavelmente ainda será bastante utilizada, é aconselhável o uso de iniciadores específicos e de técnicas como o RT-PCR para se evitar os *numts*.

Mesmo sendo indesejáveis em estudos evolutivos, os *numts* podem ser interessantes em estudos de ancestralidade, sendo utilizados como marcadores fósseis.

Agradecimentos

Agradeço as agências CAPES, CNPq e FAPESP pelo apoio financeiro.

Referências

Avise J.C., Arnold J., Ball R.M., Bermingham E., Lamb T., Neigel J.E., Reeb C.A. & Saunders N.C. (1987) Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics, *Annual Review of Ecology Systematics*. **18**: 489-522.

Avise J.C. (2001) *Phylogeography: the history and formation of species*. 447p. Harvard University Press, London.

Balakirev E. S. & Ayala F. J. (2003) Pseudogenes: are they “junk” or functional DNA? *Annual Review Genetics*. **37**: 123-151

Ballard J.W.O. & Whitlock, M. (2004) The incomplete natural history of mitochondria. *Molecular Ecology*. **13**: 729-744.

Benesh D. P., Hasu T., Suomalaine L. R., Valtonen E. T. & Tirola M. (2006) Reliability of mitochondrial DNA in an acanthocephalan: The problem of pseudogenes. *International Journal of Parasitology*. **36** (2): 247-254.

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.

Bensasson D., Zhang D.X, Hartl D.L. & Hewitt G.M. (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution*. **16**: 314-321.

Blanchard J.F. & Schmidt G.W. (1996) Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. *Molecular Biology and Evolution*. **13**: 537-548.

Bremer K. 1994. Branch support and tree stability. *Cladistics*, 10: 295-304

Collura R.V., Auerbach M.R. & Stewart C.B. (1996) A quick direct method that can differentiate expressed mitochondrial genes from their nuclear pseudogenes. *Current Biology*. **6**: 1337-1339.

D'Errico I., Gadaleta G. & Saccone, C. (2004) Pseudogenes in metazoa: Origin and features. Briefings in Functional Genomics and Proteomics. **3**: 157-167.

Farrelly F. & Buttow R.A. (1983) Rearranged mitochondrial genes in the yeast nuclear genome. Nature. **301**: 296-301.

Felsenstein J., 1985. Confidence limits on phylogenies: an approach using de bootstrap. Evol.. 39: 783-791

Gladstein D.S. & Wheeler W. C. (1997) POY: the optimization of alignment characters. American Museum of Natural History, New York. Disponível em <ftp://ftp.amnh.org/pub/molecular/poy/>

Greenwood A. & Paabo S. (1999) Nuclear insertions sequences of mitochondrial DNA predominate in hair but not in blood of elephants. Molecular Ecology. **8**: 133-137.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium. Series. **41**: 95-98.

Harrison P.M., Milburn D., Zhang Z., Bertone P. & Gerstein M. (2003) Identification of pseudogenes in the *Drosophila melanogaster* genome. Nucleic Acids Research **31**: 1033-1037.

Kumar S., Tamura K., & Nei M. (2004) MEGA 3.1. Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics **5**:150-163.

Lopez J.V., Yuhki N., Masuda R., Modi W. & O'Brien S.J. (1994) Numt, a recent transfer and tandem amplifications of mitochondrial DNA to the nuclear genome of the domestic cat. Journal of Molecular Evolution. **39**: 174-190.

Mirol P.M., Mascheretti S. & Searle J.B. (2000) Multiple nuclear pseudogenes of mitochondrial cytochrome b in *Ctenomys* (Caviomorpha. Rodentia) with either great similarity to or high divergence from the true mitochondrial sequence. Heredity. **84**: 538-547.

Moritz C., Dowling T.E. & Brown W.M. (1987) Evolution of animal mitochondrial DNA: Relevante for population biology and systematics. *Annual Review of Ecology Systematics*. **18**: 269-292.

Nugent J.M. & Palmer J.D. (1991) RNA-mediated transfer of the gene COXII from the mitochondrion to the nucleus during flowering plant evolution. *Cell*. **66**: 473-481.

Ossario P.N., Sibley D.L. & Boothroyd J.C. (1991) Mitochondrial-like DNA sequences flanked by direct and inverted repeats in the nuclear genome of *Toxoplasma gondii*. *Journal of Molecular Biology*. **22**: 525-536.

PAGE, R.D.M. (1998) TreeView: Tree drawing software for Apple Macintosh and Microsoft Windows, versão 1.5.2. Institute of Biomedical & Life Sciences, University of Glasgow, Scotl., UK.

Rozas J., Sánchez-DelBarrio, J.C., Messeguer X. & Rozas R. (2003) DnaSP, DNA polymorphism analices by the coalescent and other methods. *Bioinformatics*. **19**: 2496-2497.

Silva-Pinhatti A.C.O., Bacci Jr M, Hinkle G, Sogin M.L., Pagnocca FC, Martins V.G., Bueno O.C. & Hebling M.J.A. (2004) Low variation in ribosomal DNA and internal transcribed spacers of the symbiotic fungi of leaf-cutting ants (Attini: Formicidae). *Brazilian Journal of Medical and Biological Research*. **37**: 1463-1472.

Sorenson M.D. & Quinn T.W. (1998) Numts: a challenge for avian systematics and population biology. *The Auk*. **115**: 214-221.

SWOFFORD D.L. (2003) *Phylogenetic Analysis Using Parsimony*, version 4.0b10. Smithsonian Institution, Washington DC..

THOMPSON J.D., HIGGINGS D.G., & GIBBSON T.J. (1994) CusltalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gaps penalties and weight matrix choice. *Nucleic Acid Research*. **22**: 4673-4680.

Wheeler W., Gladstein D. (2000) POY, POY: the optimization of alignment characters. American Museum of Natural History, New York. Disponível em <ftp://ftp.amnh.org/pub/molecular/poy/>

Wheeler W., Gladstein D. & De Laet J. (2003) POY, phylogeny reconstruction via optimization of DNA and other data version 3.0.11. Comand line documentation. Disponível em <ftp://ftp.amnh.org/pub/molecular/poy/>

Wetterer J.K., Schultz T.R. & Meier R. (1998) Phylogeny of fungus-growing ants (Tribe Attini) based on mtDNA sequence and Morphology. *Molecular Phylogenetics and Evolution*. **9**: 42-47.

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

Willians 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.

Zhang D-X., Hewitt G.M. (1996) Nuclear integration: challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution*. **11**: 247-251.

Zullo S., Leang C.S., Slighton J.L., Hadler H.I. & Eisenstadt J.M. (1991) Mitochondrial D-loop sequence are integrated in the rat nuclear genome. *Journal of Molecular Biology*. **221**: 1223-1235

*** Normas segundo a revista *Animal Genetics***

Tabela 1: Lista de espécimes e a respectiva localidade

Código	Cidade	Estado	Latitude	Longitude
SES040120-01	Frei Caneca	Pernambuco	-8,720383333	-35,84425
SES040120-04	Frei Caneca	Pernambuco	-8,720383333	-35,84425
SES040121-01	Frei Caneca	Pernambuco	-8,720383333	-35,84425
SES040121-03	Frei Caneca	Pernambuco	-8,720383333	-35,84425
SES040121-04	Frei Caneca	Pernambuco	-8,720383333	-35,84425
ASM040123-01	CEPLAC	Bahia	-14,75536111	-39,23255556
SES040123-01	CEPLAC	Bahia	-14,75536111	-39,23255556
SES040123-02	CEPLAC	Bahia	-14,75536111	-39,23255556
SES040123-03	CEPLAC	Bahia	-14,75536111	-39,23255556
SES040123-06	CEPLAC	Bahia	-14,75536111	-39,23255556
SES040123-07	CEPLAC	Bahia	-14,75536111	-39,23255556
SES040123-08	CEPLAC	Bahia	-14,75536111	-39,23255556
SES040124-01	U baitaba	Bahia	-14,83975	-39,02688889
SES040124-02	U baitaba	Bahia	-14,83975	-39,02688889
SES040124-03	U baitaba	Bahia	-14,83975	-39,02688889
SES040124-08	Ipiau	Bahia	-14,09619444	-39,78102778
SES040125-01	Fazenda de Cascata	Bahia	-14,41355556	-39,33088889
SES040125-02	Fazenda de Cascata	Bahia	-14,41355556	-39,33088889
SES040125-03	Fazenda de Cascata	Bahia	-14,41355556	-39,33088889
SES040125-09	farm	Bahia	-14,20027778	-39,81586111
SES040125-10	farm	Bahia	-14,20027778	-39,81586111
SES040129-03	Alta Floresta	Mato Grosso	-9,75642	-55,86272
SES040129-06	Alta Floresta	Mato Grosso	-9,75642	-55,86272
SES040130-02	Alta Floresta	Mato Grosso	-9,59214	-56,02422
SES040131-03	Alta Floresta	Mato Grosso	-10,05337	-55,43224
*SES040131-06	Alta Floresta	Mato Grosso	-10,05337	-55,43224
SES040131-12	Alta Floresta	Mato Grosso	-9,86249	-56,07603
SES040131-13	Alta Floresta	Mato Grosso	-9,86249	-56,07603
SES040131-17	Alta Floresta	Mato Grosso	-9,86249	-56,07603
SES040204-03	Belem	Para	-1,68755	-48,54977
SES040205-02	Belem	Para	-1,68755	-48,54977
SES040205-06	Belem	Para	-1,68755	-48,54977
SES040208-06	Macapa	Amapa	0,62059	-51,69178
SES040208-08	Macapa	Amapa	0,61917	-51,70161
SES040208-09	Macapa	Amapa	0,60029	-51,75318
SES040208-10	Macapa	Amapa	0,60082	-51,75435
SES040212-01	Santerem	Para	-2,55767	-54,72733
SES040214-03	Alenquer	Para	-1,90357	-54,64113
SES040214-11	Alenquer	Para	-1,91676	-54,62662
SES040215-03	Alenquer	Para	-1,92631	-54,63676
SES040220-01	Carreiro da Varzea	Amazonas	-3,65538	-60,26173
E958	Alta Floresta	Mato Grosso	-09,55657	-55,99700
E947	Alta Floresta	Mato Grosso	-10,03715	-55,43497
E907	Aurelino Leal	Bahia	-14,4133333	-39,3308333
E899	Ilhéus	Bahia	-14,83975	-39,0268889
E909	Aurelino Leal	Bahia	-14,2017778	-39,82375
SES040131-04	Alta Floresta	Mato Grosso	-10,05337	-55,43224
SES0400305-01	Kukenan Camp (Venezuela)	Bolívar	5,10819	-60,8299
650	San Roque (Colômbia)		6,4683333	-75,0355556
651	San Roque (Colômbia)		6,4691667	-75,0347222
693	Santafé de Antioquia (Colômbia)		6,5352778	-75,8838889
690	Santafé de Antioquia (Colômbia)		6,5333333	-75,8852778

691	Santafé de Antioquia (Colômbia)	6,5333333	-75,8833333
694	Santafé de Antioquia (Colômbia)	6,5358333	-75,8847222
695	Andes (Colômbia)	5,6672222	-75,8508333
697	Andes (Colômbia)	5,6672222	-75,8502778
699	Venecia (Colômbia)	5,9672222	-75,7177778
692	Santafé de Antioquia	6,5333333	-75,8833333
688	Santafé de Antioquia	6,5347222	-75,8844444
689	Santafé de Antioquia	6,5358333	-75,885
625	Sonsón (Colômbia)	5,7008333	-75,4013889
652	San Roque (Colômbia)	6,4691667	-75,0347222
655	San Roque (Colômbia)	6,4691667	-75,0355556
656	San Roque (Colômbia)	6,4691667	-75,0336111
657	San Roque (Colômbia)	6,4691667	-75,0336111
659	San Roque (Colômbia)	6,485	-75,0180556
488	Andes (Colômbia)	5,6669444	-75,8855556
626	Sonsón (Colômbia)	5,7025	-75,4008333

* Código sinônimo de E999g3, nas figuras de 1 a 3. Amostras em azul foram usadas para representar G1, em vermelho para representar G2 e em verde para representar G3

Tabela 2: Iniciadores utilizados para amplificar G1, G2 e G3

Grupo Gênico	Par de Iniciadores	Seqüência
G1	ANTF *ANTRG1.1	5 – ATTCATTCTTATCTTGAAATATTATTTTC - 3 5 – CAATAGTTTGATTTTCTAG - 3
G2	*ANTFG2 ANTR	5 – CTCTTCTATACCCAATTTCTATAACCC – 3 5 - TTCATAAGTTCAGTATCATTGGTG - 3
G3	ANTF *ANTRG3.2	5 – ATTCATTCTTATCTTGAAATATTATTTTC - 3 5 – CAATGATTATAAATCATGGAAAAAATC - 3

* indica os iniciadores desenhados para amplificar cada um dos grupos gênicos



A



B

Figura 1: A) Região COI. O retângulo vermelho indica a inserção em G3 seguida de duplo códon de terminação e os retângulos amarelos as substituições não sinônimas. B) Região COII. A seta indica a única substituição não sinônima em G1. O retângulo amarelo indica a inserção de uma timina causando *frameshift*, mudança de Histidina para Prolina e de Asparagina para Triptofano e o retângulo vermelho indica uma deleção contígua de 4

nucleotídeos em G3, causando perda do aminoácido codificado. Os retângulos azuis indicam substituições não sinônimas.

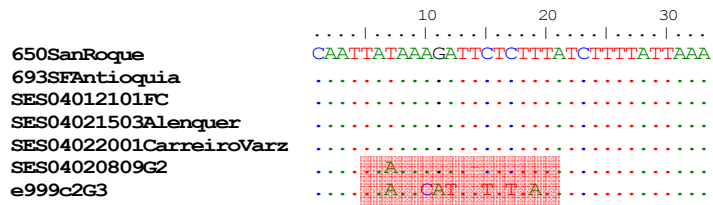
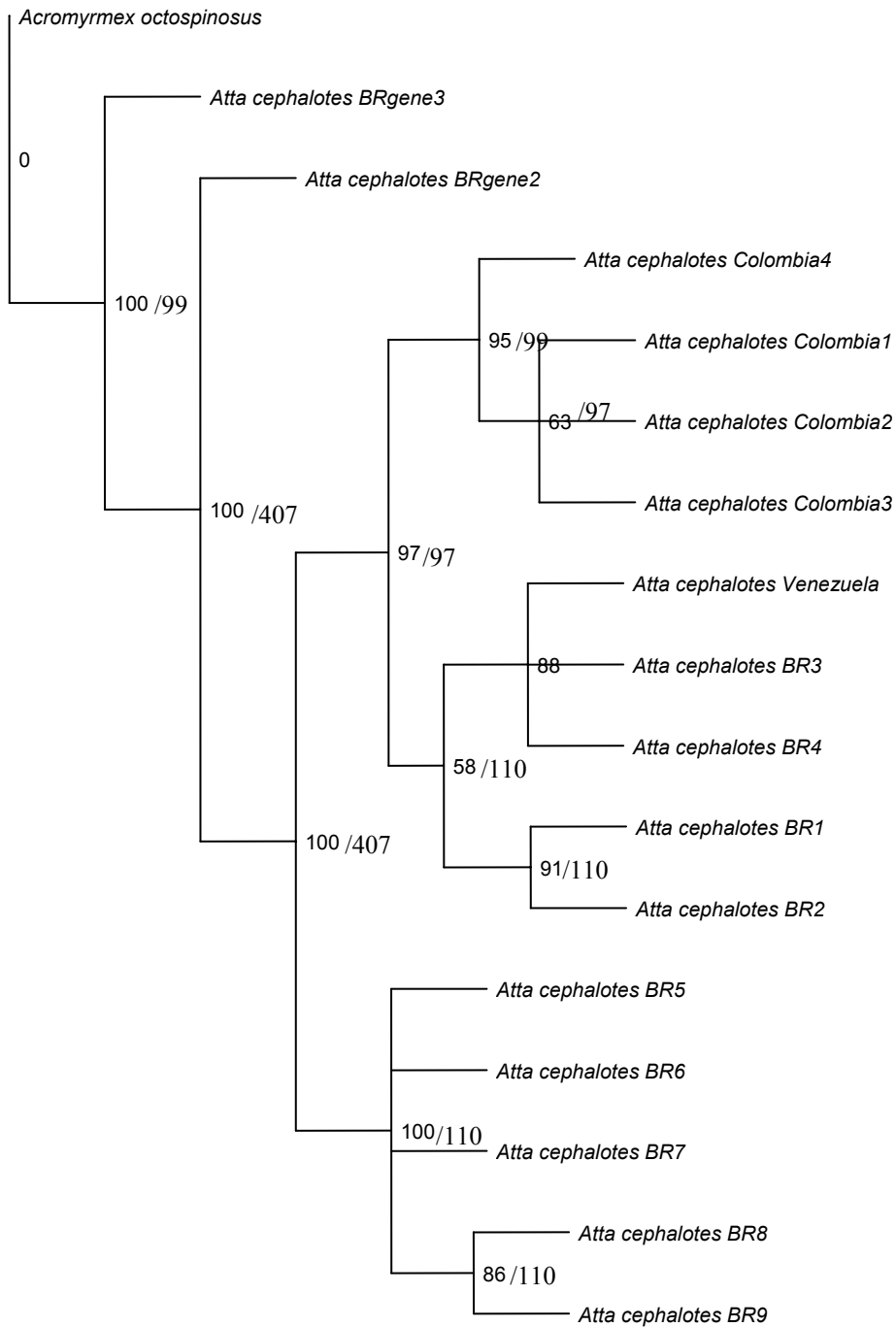


Figura 2: Região Leu-tRNA. O retângulo indica o polimorfismo na seqüência em G2 e G3



_10

Figura 3: Consenso estrito entre as 2 árvores mais parcimoniosas 243 passos obtidas a partir de 90 caracteres informativos para parcimônia dos *loci* COI, IGS, Leu-tRNA e COII. Os números à frente de cada uma das clades representam os valores de *bootstrap* (BS) e Índice de Bremer (BI), respectivamente, obtidos a partir de 1000 buscas heurísticas, cada uma delas com 1000 pseudo-réplicas.

Considerações Finais

Atta cephalotes apresenta três grupos gênicos (G1, G2 e G3) que ocorrem concomitantemente num mesmo espécime e, devido as características apresentadas, sugerimos que G2 e G3 sejam pseudogenes nucleares de origem mitocondrial ou *numts*. É muito importante ter segurança de estar trabalhando com o gene mitocondrial correto para evitar falsas predições e propostas filogenéticas. Recentemente, inúmeros trabalhos descrevem a existência de pseudogenes em organismos em que antes eles eram desconhecidos ou considerados como inexistentes. O uso de técnicas como clonagem e RT-PCR minimizam a utilização de seqüências erradas, pois possibilita identificá-las durante o sequenciamento e selecionam apenas genes transcritos, respectivamente. Como acreditá-se que os pseudogenes não sejam transcritos, o técnica de RT-PCR seria bastante eficiente em preteri-los. Ainda, G3 parece ser um pseudogene de origem basal em relação a G1 e talvez até em relação ao gênero *Atta* enquanto G2 tem uma origem mais tardia, aparecendo como grupo irmão de G1. O grupo G2 ainda apresenta bastante conservada em relação a G1 a proteína predita para os genes COI e COII sugerindo que ele ainda possa ser funcional ou até mesmo ser um segundo tipo de genoma mitocondrial (heteroplasmia), embora seja uma caso bastante raro. O grupo G1 não apresentou qualquer das características próprias de um pseudogene e foi considerado o representante correto do *loci* mitocondrial em estudo.

O estudo de filogenia de *Atta cephalotes* mostrou que existem duas grandes populações geneticamente distintas. Embora pareça haver casos de simpatria, a população descrita na Bahia, devido ao seu isolamento geográfico em relação a população Amazônica (exceto os casos mencionados anteriormente) poderia ser considerada um espécie nova, visto que a espécie tipo de *Atta cephalotes* foi coletada na região amazônica e a origem proposta para os Attini também seja essa região. A Amazônia apresenta uma transição muitas vezes complexa e sutil em seus ecossistemas. Talvez as populações do Pará e Amapá que se agruparam com as populações Atlânticas do Nordeste estejam isoladas das populações demais populações Amazônicas por fatores próprios de seus ecossistemas, diminuindo muito ou impedindo o fluxo gênico.

O DNA mitocondrial se mostrou bastante informativo, principalmente a região IGS, que apresentou assinaturas moleculares características para cada população. Entretanto, a possível existência de populações simpátricas ocorrendo na Amazônia deve ser confirmada pela utilização de outros marcadores e também estudos ecológicos e comportamentais.