

**UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE  
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA EVOLUTIVA E  
BIOLOGIA MOLECULAR**

**Evolução Molecular, análise multilocus e diferenciação entre espécies  
do grupo *fraterculus* (Diptera, Tephritidae).**

André Luís Andrade Lima

São Carlos  
2011

**UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE  
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA EVOLUTIVA E  
BIOLOGIA MOLECULAR**

**Evolução Molecular, análise multilocus e diferenciação entre espécies  
do grupo *fraterculus* (Diptera, Tephritidae).**

**André Luís Andrade Lima**

Dissertação apresentada ao Programa de Pós-Graduação  
em Genética Evolutiva e Biologia Molecular da  
Universidade Federal de São Carlos como parte dos  
requisitos para a obtenção do Título de Mestre em  
Genética e Evolução

**Orientador: Prof. Dr. Reinaldo Alves de Brito**

São Carlos  
2011

**Ficha catalográfica elaborada pelo DePT da  
Biblioteca Comunitária da UFSCar**

L732em

Lima, André Luís Andrade.

Evolução molecular, análise multilocus e diferenciação entre espécies do grupo *fraterculus* (Diptera, Tephritidae) / André Luís Andrade Lima. -- São Carlos : UFSCar, 2012. 62 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2011.

1. Genética e evolução. 2. Tephritidae. 3. *Anastrepha*. 4. Evolução molecular. 5. Análise multilocus. I. Título.

CDD: 575 (20<sup>a</sup>)

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE  
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA EVOLUTIVA E BIOLOGIA MOLECULAR

**“Evolução Molecular, análise multilocus e diferenciação entre espécies do grupo *fraterculus* (Diptera, Tephritidae)”**

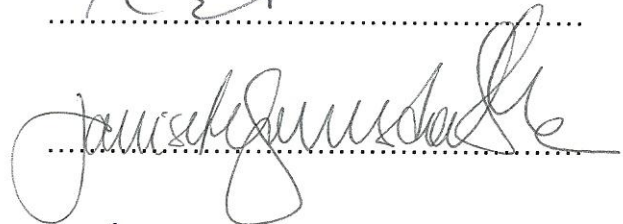
Dissertação de Mestrado de  
**ANDRÉ LUÍS ANDRADE LIMA**

Banca Examinadora

Prof. Dr. Reinaldo Alves de Brito

  
.....

Profa. Dra. Janisete Gomes Silva

  
.....

Prof. Dr. Diogo Meyer

  
.....

Democracia é oportunizar a todos o mesmo ponto de partida. Quanto ao ponto de chegada, depende de cada um.

Fernando Sabino

# AGRADECIMENTO...

Agradeço...

- ✦ **Aos meus queridos pais**, pelo carinho e apoio incondicional em todas as horas; obrigado! Se não fosse por vocês talvez não conseguiria alcançar mais este objetivo.
- ✦ **Ao meu Orientador prof. Dr. Reinaldo**, pela grande oportunidade, grande paciência, dedicação, confiança e por ter acreditado em mim e no meu trabalho.
- ✦ **As minhas queridas irmãs**, por sempre acreditar em mim em todos os momentos, pelo apoio e inquestionável amizade.
- ✦ **A minha amada Naiara**, pela esplendorosa dedicação que teve grande importância na completude deste trabalho; sua presença em minha vida demonstrou ser um alento para os dias difíceis.
- ✦ **Aos membros do laboratório**, pela grande amizade e imensa ajuda para que esta obra pudesse ser concluída. Tenho por vocês a consideração de uma família, Família Científica.
  - ✦ **Ao Júnior (Iderval)**, pela incondicional e inegável ajuda desde o meu primeiro dia em São Carlos. E pelas inúmeras conversas, dúvidas, questões e diversas análises deste manuscrito e outros trabalhos científicos, moldando minha formação profissional e pessoal.
  - ✦ **Ao Daniel**, pelas inúmeras e dedicadas ajudas com os diversos softwares e respectivas análises.
  - ✦ **A Andreia**, pela ajuda incentivos e toda torcida durante a execução desta dissertação.
  - ✦ **A Aline, Baratinha (Isabela), Ferzinha, Nelci, Vanessa, Marcos, Bruno (Montanha) Bruno Sauce, Gustavo e a Caipora (Carol)**, além da imprescindível ajuda nos momentos difíceis, as impagáveis horas de boas risadas e descontração fundamentais para a obtenção de um novo folego durante a jornada rumo ao título de Mestre.
- ✦ **Aos meus amigos(as)-irmãos de Minas Gerais e São Paulo**. Que sempre dedicaram apoio e carinho sem precedentes. Hoje posso dizer que não conquistei muitos amigos, só os melhores....
  - ✦ **Leandro, Alessandro, Ricardo, Olinto, Aline, Gisele, Maila e Charles**, obrigado por sempre me apoiarem e pelas infindáveis reuniões em Passos, Altinópolis, Franca e Ribeirão Preto que contribuíram de maneira sem iguais para a fuga dos pensamentos científicos lembrando que sempre há outros horizontes, além dos diversos dias prazerosos de descontração e descanso.
- ✦ **À FAPESP e CAPES**, pelo apoio financeiro que foi essencial para que esta pesquisa pudesse ser realizada.

## Resumo

Os Tephritidae incluem aproximadamente 4500 espécies já descritas e muitas destas espécies compõem os principais gêneros de importância econômica: *Anastrepha*, *Bactrocera* e *Ceratitis*. O gênero *Anastrepha* é, entre Tephritidae, o que possui maior diversidade nas Américas incluindo 230 espécies descritas, dentre as quais algumas espécies apresentam grande importância econômica por representarem importantes pragas da fruticultura. Há evidências de espécies que constituem o grupo *fraterculus* terem divergido há pouco tempo apresentando ainda algumas espécies crípticas. Devido a esta característica em particular, espécies do grupo *fraterculus* apresentam uma grande vantagem para estudos evolutivos. Para realizar estudos evolutivos torna-se necessário avaliar as taxas de evolução molecular bem como seu papel na identificação de taxas de fluxo gênico e diferenciação populacional, a construção de árvores de espécie e uma possível separação das espécies. Buscando um melhor entendimento do grupo *fraterculus*, escolhemos três espécies constituintes desse grupo, *Anastrepha obliqua*, *Anastrepha fraterculus* e *Anastrepha sororcula*, como modelos para este estudo e cinco genes isolados de biblioteca de cDNA de tecidos reprodutivos de espécies de *Anastrepha*. Diversos genes que são expressos em tecidos reprodutivos apresentam uma taxa de divergência maior do que os expressos em tecidos não reprodutivos, podendo assim dar indícios de isolamento reprodutivo. Neste trabalho encontramos essas características em genes que indicam seleção positiva (*CG11912* e *CG10031*), e os demais genes que embora não tenham apresentado significância estatística para serem considerados possuidores de seleção positiva, no entanto apresentam terem regiões potencialmente sob seleção positiva direcional para *Df31* e *Lcp65Ac* e seleção positiva diversificadora para os genes *Df31* e *CG16712*. Na análise de variância molecular encontramos que a diferenciação espacial é mais bem explicada pelas diferenças entre as espécies do que entre regiões geográficas. As redes haplotípicas e árvores de espécies indicam uma forte coesão entre as populações de *A. obliqua* do Nordeste e Sudeste indicando ausência de estruturação genética para esta espécie enquanto para *A. fraterculus* e *A. sororcula* foi encontrado uma sub estruturação que separa *A. sororcula* do Nordeste e *A. sororcula* do Sudeste, sendo que *A. sororcula* do nordeste pode estar hibridizando com *A. fraterculus* do Nordeste. As análises de migração encontraram migrações diferenciais para estas espécies, colocando *A. obliqua* mais isolada das demais e uma relação mais próxima entre *A. fraterculus* e *A. sororcula*. De frente aos resultados encontrados neste trabalho fica claro a necessidade de uma maior amostragem, principalmente em áreas ainda não coletadas entre as regiões amostradas na tentativa de buscar entender não apenas quais são as possíveis relações entre *A. fraterculus* e *A. sororcula* mas também entre outras espécies do grupo *fraterculus*.

**Palavra-Chave:** Tephritidae. *Anastrepha*. Evolução Molecular. Análise multilocus.

## Abstract

The Tephritidae includes approximately 4500 species described to date and many of these species make up the main genera of economic importance: *Anastrepha*, *Bactrocera* and *Ceratitis*. The genus *Anastrepha* is among the Tephritidae with greatest diversity in the Americas including 230 described species, among which some species with great economic importance because they represent important fruit pests. There is evidence that species in the *fraterculus* group have diverged recently, even having some cryptic species. Because of this, species in the *fraterculus* group are a great system for evolutionary studies. For evolutionary studies, it is necessary to evaluate the rates of molecular evolution and their role in identifying rates of gene flow and population differentiation, the construction of tree species and a possible separation for the species. Seeking a better understanding of the *fraterculus* group, we chose three species of this group, *Anastrepha obliqua*, *A. fraterculus* and *A. sororcula* as models for this study and five genes isolated from a cDNA library of reproductive tissues of *Anastrepha*. Several genes that are expressed in reproductive tissues have a higher divergence rate THAN those expressed in non-reproductive tissues, may thus give evidence of reproductive isolation. In this work, we found evidence of positive diversifying and positive selection for a couple of genes (*CG11912* and *CG10031*), and high levels of polymorphism, although they failed to meet statistical significance for positive selection, in other genes here studied, such as *Lcp65Ac* and *CG16712*, whereas *Df31* was more conserved. An analysis of molecular variance found that levels of genetic polymorphism are best explained by differences between species THAN between geographic regions. Haplotype networks for each gene failed to differentiate the species here studied and showed high levels of shared polymorphism among the species, with some rare exceptions. On the other hand, a joint analysis of these data to try to infer a species tree indicate a strong cohesion for populations of *A. obliqua* from the Brazilian Northeast and Southeast while for *A. fraterculus* and *A. sororcula* we observed a substructure that separated Southeast populations of *A. sororcula*, which was well defined, from Northeast populations, which may be hybridizing with populations of *A. fraterculus* from the Northeast, due to great number of shared polymorphisms. Other THAN populations in the Brazilian northeast, *A. fraterculus* also forms a separate, though short, lineage. Considering these findings, a broader sampling is needed, especially in areas not yet collected particularly in the Northeast, seeking to untangle the phylogenetic relationships not only among these species, but also other species of the *fraterculus* group.

**Keywords:** Tephritidae. *Anastrepha*. Molecular Evolution. multilocus analysis.



## Lista de figuras

<b>FIGURA 1.</b> CARACTERES MORFOLÓGICOS UTILIZADOS PARA IDENTIFICAÇÃO DAS MOSCAS. ....	2
<b>FIGURA 2.</b> DISTRIBUIÇÃO DAS ESPÉCIES ESTUDADAS NO PRESENTE TRABALHO.. ....	3
<b>FIGURA 3.</b> MAPA DO BRASIL EVIDENCIANDO OS AGRUPAMENTOS FORMADOS.. ....	14
<b>FIGURA 4.</b> REGIÕES AMPLIFICADAS COM <i>ÍNTRONS</i> E <i>ÉXONS</i> DE CADA GENE EM <i>ANASTREPHA</i> SP.. ....	25
<b>FIGURA 5.</b> REDE HAPLOTÍPICA PARA A REGIÃO CODIFICANTE DO GENE <i>CG16713</i> . ....	30
<b>FIGURA 6.</b> REDE HAPLOTÍPICA PARA A REGIÃO CODIFICANTE DO GENE <i>CG11912</i> . ....	31
<b>FIGURA 7.</b> REDE HAPLOTÍPICA PARA A REGIÃO CODIFICANTE DO GENE <i>CG10031</i> . ....	32
<b>FIGURA 8.</b> REDE HAPLOTÍPICA PARA A REGIÃO CODIFICANTE DO GENE <i>Df31</i> . ....	33
<b>FIGURA 9.</b> REDE HAPLOTÍPICA PARA A REGIÃO CODIFICANTE DO GENE <i>LCP65Ac</i> . ....	34
<b>FIGURA 10.</b> ÁRVORE DE ESPÉCIES GERADA PELO PROGRAMA BUCKY. ....	35
<b>FIGURA 11.</b> RESULTADOS DO FLUXO GÊNICO INFERIDO POR MÁXIMA VEROSSIMILHANÇA. ....	36
<b>FIGURA 12.</b> FST PAR-A-PAR MÉDIO PARA TODOS OS GENES.. ....	37

## Lista de Tabelas

<b>TABELA 1.</b> LOCALIDADES DAS COLETAS.....	14
<b>TABELA 2.</b> RELAÇÃO DOS PRIMERS TESTADOS.....	15
<b>TABELA 3.</b> TESTES DE SELEÇÃO POSITIVA REALIZADOS NO PAML (CODEML).....	26
<b>TABELA 4.</b> TESTES DE SELEÇÃO POSITIVA REALIZADOS NO HYPHY.....	26
<b>TABELA 5.</b> TESTES DE NEUTRALIDADE E ÍNDICES DE POLIMORFISMOS.....	27
<b>TABELA 6.</b> RESULTADOS DAS ANÁLISES DA AMOVA.....	28

## Sumário

<b>1. INTRODUÇÃO</b>	<b>1</b>
<b>1.1 SUPER FAMÍLIA TEPHRITOIDEA</b> .....	<b>1</b>
<b>1.2 BIOLOGIA DO GRUPO <i>FRATERCULUS</i></b> .....	<b>3</b>
<b>1.3 BUSCA POR MARCADORES EVOLUTIVOS</b> .....	<b>5</b>
<b>1.4 BASES PARA TESTES DE NEUTRALIDADE</b> .....	<b>6</b>
<b>1.5 GENES</b> .....	<b>8</b>
<b>2 OBJETIVOS</b>	<b>12</b>
<b>2.1 OBJETIVO GERAL</b> .....	<b>12</b>
<b>2.2 OBJETIVOS ESPECÍFICOS</b> .....	<b>12</b>
<b>3. MATERIAIS E MÉTODOS</b>	<b>13</b>
<b>3.1 COLETA E MANUTENÇÃO DAS ESPÉCIES</b> .....	<b>13</b>
<b>3.2 LOCALIDADES AMOSTRADAS</b> .....	<b>13</b>
<b>3.3 DADOS MOLECULARES</b> .....	<b>15</b>
<b>3.4 CONSTRUÇÕES DE <i>PRIMERS</i></b> .....	<b>15</b>
<b>3.5 EXTRAÇÃO DE DNA</b> .....	<b>16</b>
<b>3.6 OBTENÇÃO E PREPARAÇÃO DAS SEQUÊNCIAS PARA O SEQUENCIAMENTO</b> .....	<b>16</b>
<b>3.7 SEQUENCIAMENTO E PROCESSAMENTO DAS SEQUÊNCIAS</b> .....	<b>17</b>
<b>3.8 ANÁLISES DOS DADOS</b> .....	<b>17</b>
<b>3.8.1 Testes de Neutralidade e Índice de Polimorfismo</b>	<b>17</b>
<b>3.8.2 Análise Populacional</b>	<b>21</b>
<b>3.8.4 Análise de Migração</b>	<b>22</b>
<b>3.8.5 Árvore de Espécies</b>	<b>22</b>
<b>4 RESULTADOS</b>	<b>24</b>
<b>4.1 <i>PRIMERS</i> E ALINHAMENTO</b> .....	<b>24</b>
<b>4.2 ANÁLISES</b> .....	<b>25</b>
<b>4.2.1 Testes de Neutralidade e Índice de Polimorfismo</b>	<b>25</b>
<b>4.2.2 Análise populacional</b>	<b>28</b>
<b>4.2.3 Rede Haplotípica, análise de migração e árvore de espécies</b>	<b>29</b>
<b>5. DISCUSSÃO</b>	<b>38</b>
<b>6. CONCLUSÕES</b>	<b>46</b>
<b>7. BIBLIOGRAFIA</b>	<b>48</b>
<b>8. ANEXOS</b>	<b>54</b>



## 1. Introdução

### 1.1 Super Família Tephritoidea

A superfamília Tephritoidea inclui as famílias Lonchaeidae, Piophilidae, Pallopteridae, Richardiidae, Ulidiidae (= Otitidae = Pterocallidae), Platystomatidae, Pyrgotidae e Tephritidae. As relações entre algumas famílias pertencentes a esta superfamília apresentam discordâncias (plano básico de caracteres e taxonomia) entre taxonomistas contemporâneos. Entre os problemas não resolvidos estão as relações dos Tephritidae e famílias próximas (Platystomatidae, Pyrgotidae e Ulidiidae) que nunca foram completamente avaliados (KORNEYEV, 1999a).

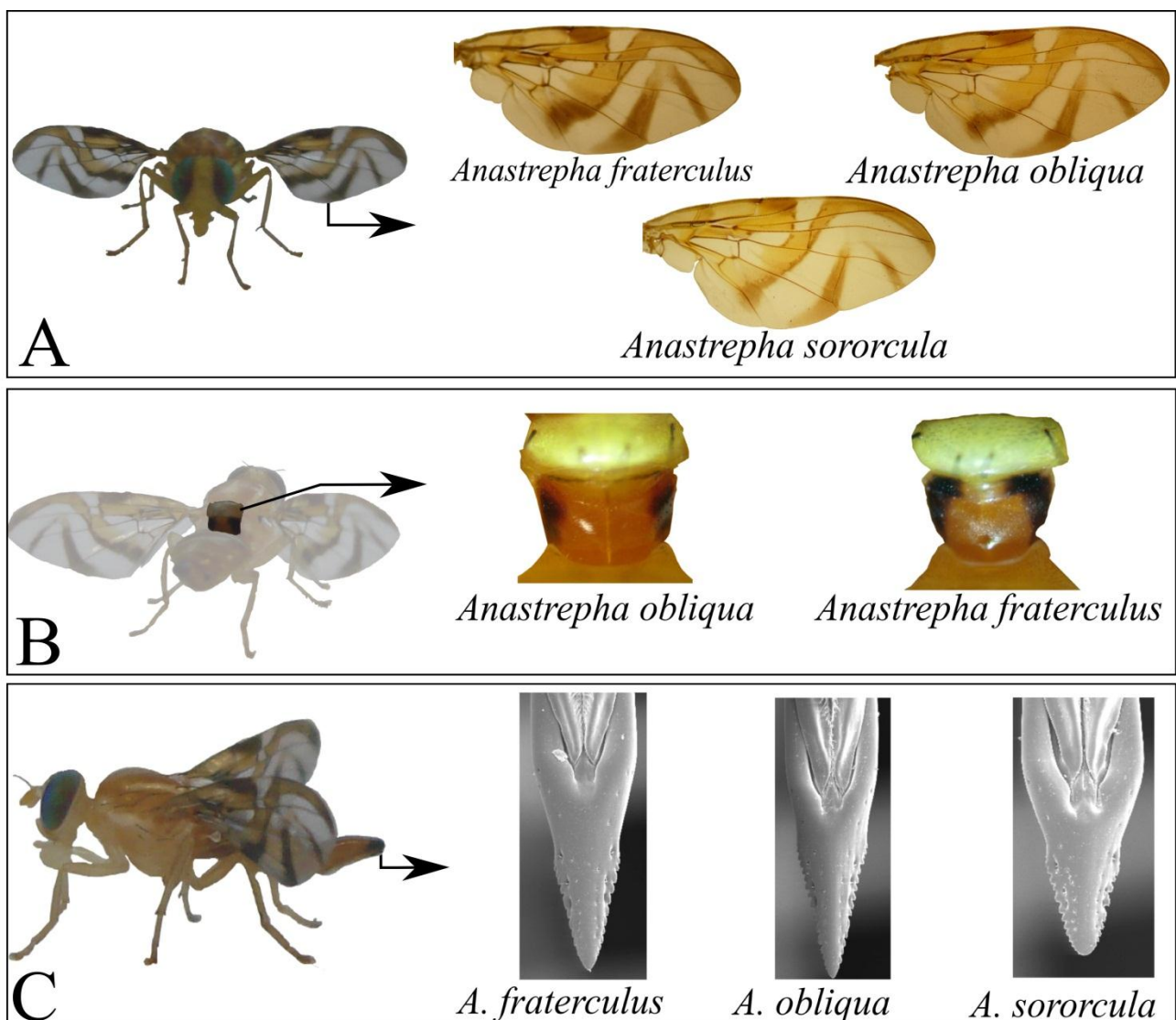
Os Tephritidae incluem aproximadamente 4500 espécies já descritas e muitas destas espécies compõem os principais gêneros de importância econômica *Anastrepha*, *Bactrocera*, *Ceratitis* (SEGURA et al, 2011) e *Rhagoletis* (ARAÚJO, 2011). Esta família não é uniformemente composta somente por fitófagos, abrangendo muitas espécies de saprófagos e ocasionalmente zoófagos, galhas e fitófagos não especializados (KORNEYEV, 1999b). Esta mesma subfamília possui a tribo Toxotrypanini que abrange três gêneros: *Anastrepha*, *Toxotrypana* e *Hexachaeta* (NORRBOM et al, 1999), e o gênero *Anastrepha* é o maior destes três gêneros (NORRBOM et al, 1999; SILVA e BARR, 2006).

O gênero *Anastrepha* é o que possui maior diversidade das Américas incluindo 212 espécies descritas (NORRBOM e KORYTKOWSKI, 2011) endêmicas das regiões do continente americano com sua distribuição cobrindo parte da América do Norte, América do Sul e muitas ilhas do Caribe (ALUJA, 1994). Estas espécies apresentam grande importância econômica por representarem importantes pragas da fruticultura (FACHOLI-BENDASSOLLI e UCHÔA-FERNANDES, 2006; NORRBOM et al, 1999; SEGURA et al, 2011; SELIVON et al, 2005a; SELIVON et al, 1999; SELIVON et al, 2005b; SELIVON e PERONDINI, 1999).

Dentro de *Anastrepha* o grupo *fraterculus* abrange 29 espécies, das quais *Anastrepha fraterculus* que forma um complexo de espécies crípticas, exibindo níveis distintos de isolamento reprodutivo pré e pós-zigóticos (DOS SANTOS et al, 2001; SELIVON et al, 1997; SELIVON et al, 2005a; SELIVON et al, 1999; SELIVON et al, 2005b; SELIVON et al, 2002). As espécies pertencentes ao grupo *fraterculus* possuem características morfológicas particulares, tais como listras sub-laterais do tipo xantina, cor pálida que muitas vezes não é evidente em espécies com coloração mais clara, mas são óbvias em espécies com coloração escura. O corpo possui coloração amarelada ou alaranjada e o sub-escutelo e/ou mediotergito podem apresentar listras marrom escuras na lateral que podem ter variações. Estas marcações variam intraspecificamente em algumas



espécies, mas não deixam de ser úteis taxonomicamente. As asas apresentam padrões similares sendo compostas por três bandas denominadas de Bandas C, S e V. Os padrões das asas de cerca de 95% das espécies de *Anastrepha* possuem alguma variação destes padrões ou são derivados destes. Essas três bandas podem ser separadas ou conectadas. Usualmente as bandas tocam se ao longo da veia  $R_{4+5}$  quando estas estão conectadas, embora ocasionalmente possam se conectar em outros lugares. Em algumas espécies partes da mesma banda podem ser ausentes especialmente o ápice e ou o braço distal da Banda V. Outros caracteres taxonômicos característicos das espécies pertencentes ao grupo *fraterculus* são a forma, denteção e o comprimento da ponta do acúleo, que em *Anastrepha* são variáveis; juntamente com o comprimento do acúleo. Estes últimos são os caracteres taxonômicos de maior importância entre estas espécies (NORRBOM et al, 1999). Na **Figura 1** são apresentadas algumas características utilizadas na identificação das *Anastrepha*. sp.

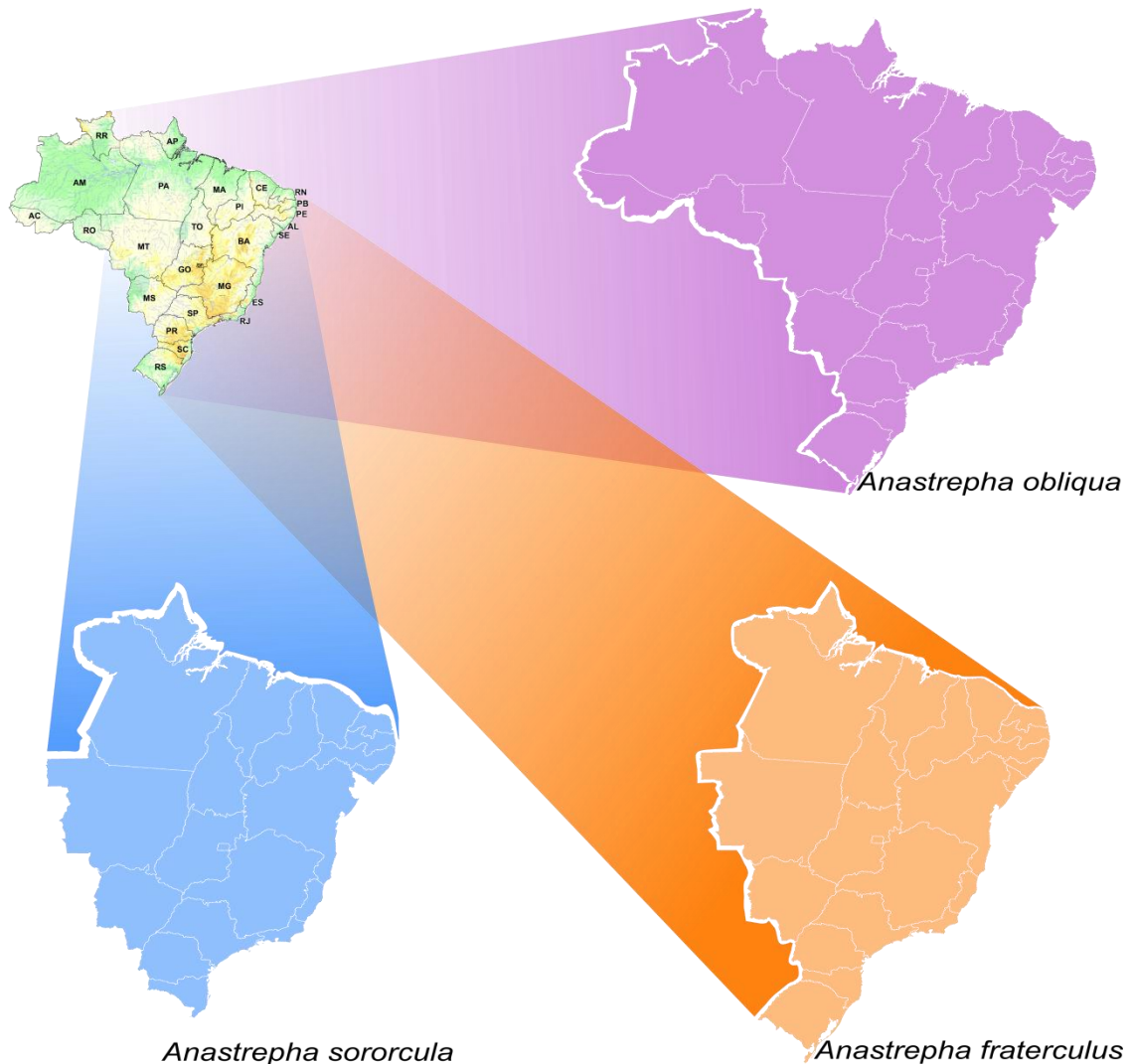


**Figura 1.** Caracteres morfológicos utilizados para identificação das moscas. Em **A**, temos o padrão das manchas em V nas asas. Em **B**, é mostrado o padrão de manchas no sub escutelo, em **C**, o padrão do ovipositor. As imagens das asas, tórax e ápice do ovipositor foram obtidos de ZUCCHI e URAMOTO (2007).



## 1.2 Biologia do Grupo *fraterculus*.

As Moscas-das-frutas apresentam ampla distribuição, abrangendo todos os estados brasileiros (ZUCCHI, 2008). Na **Figura 2**, destaca-se os locais de ocorrência das três espécies estudadas neste trabalho.



**Figura 2.** Distribuição das espécies estudadas no presente trabalho. Em **roxo** apresenta-se *Anastrepha obliqua* (Macquart); em **laranja** temos *Anastrepha fraterculus* (Wiedemann); e na região **Azul** evidencia-se *Anastrepha sororcula* (ZUCCHI). Figura criada a partir das ocorrências desses Toxotrypanini segundo ZUCCHI, 2008.

Há dimorfismo sexual em adultos, evidenciado pela presença do protuberante ovipositor em fêmeas (SIVINSKI e DODSON, 1992). A cópula ocorre geralmente em cima ou próximo dos frutos. O comportamento de cópulas das espécies tropicais ocorre frequentemente com a formação de um semicírculo ao redor da fêmea chamado de Leks, sendo estes leks usados como sítios de encontro para demonstrações de cortejo e acasalamento (ALUJA et al, 1999; CONDON e NORRBOM, 1999; SIVINSKI et al, 1999).



Os machos atingem a maturidade sexual entre 7 e 20 dias e as fêmeas entre 14 e 15 dias da emergência (CARVALHO et al, 1998; FACHOLI-BENDASSOLLI e UCHÔA-FERNANDES, 2006), sendo que a longevidade dos machos chega a 100 dias e das fêmeas a 105 dias aproximadamente com uma produção média de ovos/fêmea de 137 e com pico de oviposição entre 14 e 25 dias de idade (CARVALHO et al, 1998). O ciclo reprodutivo de *Anastrepha* dura entre 45 a 90 dias, podendo ser repetido várias vezes durante o ano (SELIVON, 1999).

Atualmente estão descritas 212 espécies de *Anastrepha* reconhecidas atualmente, sendo que 113 destas encontram-se no Brasil (ZUCCHI, 2000), com 52 hospedeiros conhecidos e 25 espécies que possuem apenas um hospedeiro. Já nos hospedeiros encontramos 38 famílias que são parasitadas por *Anastrepha*, e dentro destas, encontramos 14 espécies de *Anastrepha* infestando Myrtaceae, 16 para Sapotaceae, 11 para Anacardiaceae e 09 para Passifloraceae. Na relação parasita/hospedeiro, as espécies que apresentam maior número de hospedeiros associados são *A. fraterculus* (Wiedemann) (81) e *A. obliqua* (Macquart) (37) e *A. sororcula* ZUCCHI que não fica longe com 21 hospedeiros conhecidos. Das 81 espécies de hospedeiros que são infestados por *A. fraterculus* aproximadamente 40% são representantes da família Myrtaceae, enquanto que para *A. obliqua* esse número corresponde a 51% e para *A. sororcula* 71% desta mesma família de hospedeiro (ZUCCHI, 2008).

As larvas das moscas-das-frutas causam a destruição da fruta ao se alimentarem da polpa, assim como fêmeas adultas que ao fazerem orifícios na epiderme abrem uma porta de entrada para fungos causadores de podridão. Esses danos acabam causando o amadurecimento precoce dos frutos e, conseqüentemente uma queda abundante, o que faz diminuir a oferta de frutos para o mercado interno, geralmente resultando em aumento de preços do produto (DUARTE e MALAVASI, 1999). Sendo assim, moscas-das-frutas estão entre as principais pragas da fruticultura mundial (DÍAZ, 2010).

As espécies do grupo *fraterculus* apresentam pequenas diferenças em alguns dos seus marcadores, dentre os quais podemos destacar a morfologia do ápice do ovipositor, a morfologia do ovo, horário de cópula, padrão de coloração das asas e as manchas no escutelo (DOS SANTOS et al, 2001; SELIVON, 1999). Não há um único marcador que seja eficiente em diferenciar as espécies do grupo *fraterculus*, para tanto faz se necessário utilizar combinações de marcadores para efetivar a identificação das espécies. Contudo, existem evidências de que esses marcadores morfológicos e comportamentais não correspondam exatamente aos padrões de variação e divergência genética dessas moscas (MALAVASI e MORGANTE, 1982; MORGANTE et al, 1980). Um dos principais marcadores utilizados, o ápice do ovopositor, apresenta variação nas suas medidas (total e do ápice)



ao longo da distribuição geográfica da espécie e também em exemplares obtidos de um mesmo hospedeiro; portanto apresentando sobreposição interespecífica o que pode restringir sua aplicabilidade em certas circunstâncias. A plasticidade fenotípica dos marcadores morfológicos pode resultar na dificuldade em utilizá-los na separação de espécies muito próximas (ARAÚJO e ZUCCHI, 2006), dificultando a identificação de espécies, por exemplo, do grupo *fraterculus*, até mesmo com o uso de marcadores morfológicos (ZUCCHI, 1978), comportamentais (MORGANTE et al, 1983), alozímicos (MATIOLI et al, 1986; MORGANTE et al, 1980), cariotípicos (SOLFERINI e MORGANTE., 1987) e moleculares (Smith-Caldas, 2001; Sobrinho, 2009), sendo sugerido que *A. fraterculus* seja um complexo de espécies crípticas (SELIVON et al, 2005b; SOLFERINI e MORGANTE., 1987; STECK, 1991).

### 1.3 Busca por Marcadores Evolutivos

Muitos estudos moleculares sobre tefritídeos têm sido publicados recentemente, diversos deles focado sobre os gêneros *Anastrepha*, *Bactrocera*, *Ceratitis* e *Rhagoletis*. Estes estudos contribuem para o entendimento da biologia e evolução dos tefritídeos, provendo essenciais informações tanto para o manejo de pragas e programas de controle quanto para um melhor entendimento da filogenia, estrutura populacional e modelos de especiação destas moscas. Como a correta delimitação e identificação de espécies, ou complexos de espécies, é essencial para pesquisas básicas e aplicadas e tem profundas consequências práticas (SILVA e BARR, 2006), estes conhecimentos são de grande importância teórica e prática.

A compreensão do mecanismo pelo qual novas espécies são originadas, ou seja, o processo de especiação tem fascinado os biólogos desde antes de Charles Darwin publicar seu mais famoso livro intitulado Sobre a “Origem das Espécies por meio da Seleção Natural” (1859), há mais de 150 anos (KOCHER, 2004; OLIVER et al, 2009).

O entendimento da base genética do isolamento reprodutivo é um problema de importância crucial nos estudos sobre especiação (MACHOLÁN et al, 2011). Proteínas expressas em tecidos reprodutivos em geral apresentam maior divergência quando comparadas a proteínas expressas em tecidos não reprodutivos. Em *Drosophila*, encontra-se uma diversidade duas vezes maior em proteínas relacionadas a tecidos reprodutivos em relação a proteínas que não estão ligadas à reprodução (SWANSON e VACQUIER, 2002; SWANSON et al, 2004). Esta rápida divergência pode ser explicada pela evolução adaptativa, podendo ser importante para um melhor entendimento da nossa compreensão dos processos que permeiam a especiação, uma vez que as populações





tornem-se reprodutivamente isoladas (SWANSON e VACQUIER, 2002).

A rápida divergência acima apontada é condizente com a crescente evidência de que muitas características reprodutivas evoluem rapidamente durante os estágios iniciais da especiação. A rápida divergência desses genes que são essenciais ao desenvolvimento entre espécies irmãs, pode acrescentar outras particularidades em nosso entendimento do início da especiação (KULATHINAL et al, 2003).

#### 1.4 Bases para testes de neutralidade

Métodos filogenéticos são amplamente usados para estimar taxas evolutivas, detectando os rastros da seleção natural e usando essas informações para interpretar os dados genômicos. As alterações sofridas pelo códon são em muitos casos os recursos básicos usados nos principais modelos de substituição de aminoácidos. Nestes, a relação entre as taxas de substituições não-sinônimas ( $d_N$ ) pelas sinônimas ( $d_S$ ) ( $\omega = d_N/d_S$ ) mede a direção e magnitude da seleção com valores de  $\omega < 1$ ,  $= 1$  e  $> 1$  e, portanto, indicam seleção purificadora, neutra e positiva, respectivamente (YANG, 2007). A seleção positiva faz com que uma mutação vantajosa aumente rapidamente sua frequência em uma população até atingir a fixação. Na seleção negativa há a eliminação de mutações deletérias e em conjunto com as recorrentes mutações ocorre um acréscimo na proporção de mutações recentes em relação às mutações antigas; na seleção balanceadora, são mantidos, em um mesmo sítio, diferentes tipos de mutações (novas e antigas) em proporção equilibrada na população, diminuindo a proporção de mutações recentes em relação às mutações antigas (ASTHANA et al, 2005).

##### 1.4.1 Coalescência e Polimorfismos transpecíficos

Uma das principais características do DNA é sua capacidade de replicar-se, dando a esta molécula a capacidade de aumentar o seu número de cópias no futuro. Partindo deste princípio, se olharmos para diversas moléculas de DNA, hoje podemos inferir que estas derivaram de sucessivos processos reprodutivos bem sucedidos. Desta forma, se considerarmos o inverso destes eventos de replicação do DNA chegaremos a uma única cópia de DNA no passado que originou a todas as moléculas amostradas no presente e para este processo damos o nome de coalescência (KINGMAN, 1982). Este conceito não considera necessariamente mutações, mas somente a capacidade do DNA de replicar-se, mas a introdução de mutações permite a inferência da origem, segregação e extinção de novos alelos na população. Neste processo, vemos que quando uma cópia da forma ancestral sofre uma mutação para criar uma forma alternativa, as outras cópias a esta idênticas não mudam



simultaneamente, levando à coexistência na população de descendentes recentemente originados e diversas cópias idênticas a sua forma ancestral (TEMPLETON, 2006). Ao longo do tempo estes polimorfismos sofrerão a ação de forças evolutivas como a deriva genética ou a seleção natural e eventualmente serão fixados ou eliminados da população. Tal processo ocorrerá proporcionalmente ao tempo de divergência e será inversamente proporcional ao tamanho populacional, de forma que quanto maior este tempo e menor o  $N_{ef}$  (tamanho efetivo populacional de variância), mais provável que ocorra uma fixação ou perda de um polimorfismo (KIMURA, 1970). Por outro lado, quanto menos tempo houver entre divergências, ou maior o tamanho populacional, mais lento será este processo de fixação e mais polimorfismos ancestrais permanecerão segregando nas populações descendentes após a separação (ZOLET, 2009). Tal cenário define bem o potencial de existência de polimorfismos transespecíficos (polimorfismos ancestrais) que ocorrem quando alguns dos haplótipos (ou alelos ou inversões) encontrados em uma espécie são genealogicamente mais aparentados à linhagens haplotípicas encontradas em uma segunda espécie do que a outros haplótipos encontrados em sua própria espécie (TEMPLETON, 2006).

A existência de polimorfismos ancestrais cria disparidades entre árvores de genes e árvores de espécies e faz com que não seja esperada uma coerência completa entre o sinal filogenético de um gene e o que ele indica da separação entre as espécies. Diversos modelos têm sido criados para tentar se reconstruir as árvores de espécies a partir de árvores de genes distintas (MA et al, 2000; MADDISON, 1997; THAN e NAKHLEH, 2009) usando de metodologias bayesianas (LARGET et al, 2010) e da coalescência (THAN e NAKHLEH, 2009), mas tal processo continua potencialmente problemático. Ao considerar que algumas discordâncias podem ser estocásticas, estas árvores incongruentes muitas vezes refletem histórias evolutivas diferentes (MADDISON, 1997). Na tentativa de minimizar esses problemas, são empregadas metodologias bayesianas que irão tratar separadamente as árvores de espécie das de genes; dois programas destacam-se para estas análises, BEST (Estimativa Bayesiana de Árvore de espécie) que assume o modelo coalescente de múltiplas espécies e o BUCKy que usa similarmente uma abordagem Bayesiana para distinguir entre a incongruência e os erros da estimativa da árvore de gene (CHUNG e ANÉ, 2011). Ambos os programas apresentam resultados semelhantes e satisfatórios nas análises comparativas realizadas, mas o BUCKy destaca-se pela baixa demanda de tempo na obtenção dos resultados, contra o demasiado tempo gasto pelo BEST (CHUNG e ANÉ, 2011; CRANSTON et al, 2009), além do mais, o BUCKy obteve resultados satisfatórios para testes de adequação do modelo coalescente que provou ter baixo erro do Tipo I (CHUNG e ANÉ, 2011).



## 1.5 Genes

Problemas de incongruências em árvores de genes são provavelmente devido a imprecisões na reconstrução filogenética devido à falta de sítios informativos em sequências curtas de genes individuais, principalmente para curtas sequências de genes com baixa taxa evolutiva, que neste caso tem-se uma pronunciada incongruência na árvore de genes (RASMUSSEN e KELLIS, 2007). Por este motivo, atenção especial tem sido dada a proteínas relacionadas ao sexo que possuem, em geral, taxas de divergência maior, quando comparadas às proteínas que não estão relacionadas à reprodução. Uma gama de proteínas distintas tem sido identificada como tendo altas taxas evolutivas, mas de particular interesse são famílias gênicas que apresentam uma grande diversidade intraespecífica (entre cópias parálogas) e também entre espécies diferentes. Como exemplo destas famílias podemos mencionar as proteases e seus inibidores, as serpinas, as proteínas de cutícula e os receptores olfatórios. É possível que esta divergência em genes que atuam no comportamento e morfologia sexual seja provavelmente a primeira causa do isolamento sexual e início da diferenciação, sendo que o comportamento sexual frequentemente desenvolve-se rapidamente e é crítico para o isolamento reprodutivo (ELLISON et al, 2011).

Dentre as famílias gênicas potencialmente relevantes para o estudo da diferenciação podemos indicar as proteínas cuticulares. Diversas estruturas fibrosas são conhecidas como tendo arquitetura helicoidal. Tais estruturas incluem as cutículas de artrópodes, tendões de vertebrados, parede celular de plantas, entre outros (ICONOMIDOU et al, 1999). As proteínas cuticulares são importantes no reconhecimento de possíveis parceiros em insetos, através da quimiorrecepção de feromônios existentes na cutícula, estes hidrocarbonetos cuticulares apresentam várias diferenças quantitativas e qualitativas entre os sexos e podem desempenhar um papel importante no reconhecimento do parceiro (MING e LEWIS, 2010).

Larvas de *Drosophila melanogaster* possuem cutícula flexível com cinto ventral de dentículos esclerosados. No começo da metamorfose a cutícula do terceiro ínstar larval é transformada no pupário que protege e desenvolve a pupa em adultos. Durante a vida larval são expressos alguns genes que participam da formação da cutícula em adultos. Os genes *Lcp-c*, *-d*, *-e*, *-f* e *-g* são expressos durante a síntese de todas as três cutículas nos períodos de transição de um ínstar para o outro (CHARLES et al, 1997). O alto nível de conversão gênica existente em proteínas cuticulares faz com que estas proteínas sejam muito interessantes entre determinados genes de algumas espécies de insetos e novos estudos tem revelado uma notável diversidade entre estas proteínas (CORNMANN, 2009). Muitos tipos de proteínas cuticulares são encontrados em uma única espécie de inseto e suas características diversificadas entre os insetos. Isto pode se dar pela



duplicação de agrupamentos (*clusters*) de proteínas cuticulares que estão evoluindo independentemente entre as espécies de insetos (FUTAHASHI et al, 2008).

Por sua vez, a família gênica das proteases apresenta padrões similares de evolução com grande diferenciação pela geração de cópias ortólogas e parálogas. Proteases são enzimas que quebram as ligações peptídicas que unem os aminoácidos nas proteínas. As proteases séricas estão envolvidas em diversos processos biológicos (NEITZEL, 2010) como na digestão de alimentos, na penetração dos espermatozoides nos óvulos, nas infecções (NEITZEL, 2010; XUE e SETO, 2005), na coagulação do sangue e também no desenvolvimento de organismos (NEITZEL, 2010). Estas proteases apresentam um maior papel na reprodução em ambos os sexos, agindo nos espermatozoides (SCHILL, 1975), ou durante o processo de fertilização (CESARI et al, 2005). As cascatas de proteases séricas extracelulares estão envolvidas, nos vertebrados e invertebrados, na mediação rápida das relações fisiológicas ou sinais patológicos (LIGOXYGAKIS et al, 2003). Essas enzimas geram a ligação que ativa o padrão de sinalização Toll (proteína transmembrana responsável por recepção de estímulos externos) estabelecendo a polaridade dorsoventral em embriões de *Drosophila* (HASHIMOTO et al, 2003; LIGOXYGAKIS et al, 2003).

Proteases envolvidas na reprodução podem em alguns casos reduzir o polimorfismo ou a divergência, tendo sido sugerido para muitos loci a influência de uma seleção positiva direcional. Em contraste, proteases envolvidas em reprodução exibem notável diversidade intraespecífica gerada por seleção balanceadora, *splicing* alternativo, variação no número de cópias e na conversão gênica. Muitas vezes nestes casos estas proteases surgem de duplicações gênicas em linhagens específicas de proteases que evoluem para novas proteínas e sofrem mudanças regulatórias (KELLEHER et al, 2011). É possível que esteja acontecendo nestas proteases uma contínua competição espermática que sugere uma corrida armamentista entre os abundantes gametas de machos que visam fertilizar os limitados gametas femininos buscando controlar o processo de escolha (BIRKHEAD e MØLLER, 1998). Um exemplo desta alteração em proteases envolvidas no processo reprodutivo pode ser observado na dinâmica evolutiva de algumas proteases dos tratos reprodutivos femininos de *Drosophila mojavensis*. Tais proteases apresentam tanto seleção balanceadora quanto duplicação acelerada em alguns genes e apresentam um contraste com padrões de seleção direcional e escassez de duplicação de genes observados em *D. melanogaster* e estes dois congêneres também têm sistemas de acasalamento dramaticamente diferentes e níveis de promiscuidade feminina, bem como diferentes fisiologias reprodutivas. Estas características reprodutivas sugerem a permanência das assinaturas proteicas na história evolutiva envolvida em características reprodutivas e cria um quadro para possíveis comparações das dinâmicas das



proteínas reprodutivas ente organismos estreitamente relacionados (KELLEHER et al, 2011).

Inibidores de proteases séricas, que são também conhecidos como serpinas, pertencem à superfamília de proteínas que funcionam como inibidores de proteases. Elas foram primeiramente descritas como reguladoras da coagulação sanguínea, inflamação e ativação complementar em humanos. Aproximadamente 800 genes que codificam serpinas foram identificados em plantas, animais, vírus e procariotos. O genoma de *Drosophila melanogaster* contém 29 genes de serpinas, mas é provável que este número venha a aumentar quando contrastamos com os 211 genes de proteases séricas encontrados em *Drosophila* (REICHHART, 2005). Serpinas podem desempenhar funções diferentes das já mencionadas, inibidores de proteases, podendo desempenhar novas funções como a de armazenamento, ovalbumina (SERPINB2 e SERPINB1), no desenvolvimento de espermatozoide (SERPINA5) e genes supressores de tumor (maspin e SERPINC1) (SILVERMAN et al, 2001).

A maioria das serpinas inibidoras mostra uma conservação marcante dos sítios de clivagem das proteinases, indicando que estes inibidores não mudam facilmente sua especificidade dos alvos de proteinases. Apesar disso há uma irregularidade na distribuição dos ortólogos de *D. melanogaster* e muitas serpinas não são identificadas em ortólogos de espécies mais distantes, sendo que o equilíbrio entre as serpinas inibidoras e as proteinases que controladas por estas, formam uma forte restrição evolutiva, a maioria dos ortólogos em drosofilídeos das serpinas inibidoras mostram-se como não tendo redundantes funções genéticas. Três loci em *D. melanogaster* apresentam eventos de duplicações e divergência onde as restrições foram relaxadas, e cada complexo contém uma suposta serpina inibidora ancestral com ortólogos em toda linhagem de drosofilídeos. Da mesma forma que a evolução de super famílias de serpinas em drosofilídeos ocorre predominantemente por um mecanismo de duplicação de genes e diversificação (GARRETT et al, 2009), é possível que o mesmo processo seja comum a outras famílias de insetos.

Existem também genes envolvidos com tecidos reprodutivos que não apresentam taxas evolutivas rápidas e evoluem de forma mais conservada. Um exemplo de um potencial gene que se encaixa nesta categoria é o *Df31*, que codifica para uma fosfoproteína termoestável hidrofílica com 18kDa (CREVEL e COTTERILL, 1995; CREVEL et al, 2001; GUILLEBAULT e COTTERILL, 2007; SZOLLOSI et al, 2008) que pode catalisar a descondensação da cromatina do espermatozoide de *Xenopus* em um sistema em replicação (INOUE et al, 2011) e que pode atuar como uma chaperona de histonas (GUILLEBAULT e COTTERILL, 2007), ambas *in vitro*. A DF31 é detectada como um componente integral da cromatina presente ao mesmo nível que a histona H1,



sugerindo um papel estrutural para DF31 no dobramento da cromatina e também removendo proteínas específicas do espermatozoide. É provável que esta proteína tenha um papel na estrutura cromossomal agindo como uma proteína estrutural nos altos níveis de dobramento, maior do que dos nucleossomas (CREVEL et al, 2001; GUILLEBAULT e COTTERILL, 2007). A função na estrutura da cromatina é sugerida por observações após a injeção de RNA *antisense* contra RNAm de *Df31*, a falta de DF31 causa a desorganização do DNA no núcleo, além disso, DF31 pode interagir diretamente com histonas e são hábeis para atuar como um fator carregador de histonas junto ao DNA do espermatozóide e nos modelos de DNA nú (GUILLEBAULT e COTTERILL, 2007).

Uma das características importantes das proteínas relacionadas à reprodução é sua já mencionada taxa de divergência maior do que outras proteínas em geral (SWANSON e VACQUIER, 2002; SWANSON et al, 2004). Além desta característica, estas distintas proteínas podem criar BARReiras pré-zigóticas que ocorrem durante o processo de fertilização, sendo consequências de diferenças na forma em que as proteínas do espermatozoide e o oócito interagem, que em muitas vezes se manifestam por uma evolução mais rápida destas proteínas entre as espécies em simpatria do que em alopatria e uma possível consequência é estas proteínas estarem sob seleção direcional positiva (NYDAM e HARRISON, 2011). Neste estudo, buscamos investigar as taxas evolutivas de genes pertencentes a diversas classes evolutivas que têm sido associadas a padrões rápidos de evolução, como proteases, serpinas e proteínas cuticulares, contrastando os padrões de tais genes com um gene evolutivamente conservado. A escolha destas classes de genes é particularmente importante para o estudo de grupos que se diferenciaram recentemente como o grupo *fraterculus*, de forma que seu estudo pode auxiliar na investigação das diferenciações de espécies do grupo *fraterculus*, ao contribuir para um melhor entendimento de processos evolutivos como estimativas de fluxo gênico, estrutura populacional e pressões de seleção.

## 2 Objetivos

### 2.1 Objetivo Geral

Avaliar taxas de evolução molecular e o seu papel na identificação de taxas de fluxo gênico, diferenciação entre as espécies do grupo *fraterculus* e a construção de árvore de espécie para as espécies aqui estudadas.

### 2.2 Objetivos Específicos

- Averiguar possíveis desvios de neutralidade, e papel de seleção em sua diferenciação para cada um dos genes estudados,
- Avaliar diferenciação intra e interespecífica através do estudo dos genes investigados,
- Detectar taxas de migração entre os grupos e as espécies,
- Estimar árvores de genes e árvores de espécies.



### 3. Materiais e Métodos

#### 3.1 Coleta e Manutenção das Espécies

Os espécimes foram oriundos de coletas ou recebimento de frutas como manga, goiaba, seriguela, acerola, pitanga, carambola entre várias outras que possuem endocarpo carnoso. Quando houve coleta, esta se fez em áreas preferencialmente de reduzida intervenção humana coletando frutos diretamente do pé ou que estejam no chão há pouco tempo. No laboratório, as frutas foram depositadas em caixas de acrílico chamadas de gaiolas que seguem as seguintes dimensões 50 x 50 x 40 cm, forradas com vermiculita expandida para sua apropriada acomodação. As caixas de acrílico foram etiquetadas com localidade, dia da coleta, espécie (após identificação), e o número da geração desde a coleta da natureza.

No decorrer de uma semana a duas, as larvas contidas nas frutas vão para o substrato de vermiculita aonde irão empupar. As pupas permanecem enterradas na vermiculita até que esta seja peneirada para que proceda a separação das pupas do substrato. Então são acondicionadas em recipiente de vidro com pouco substrato para manter a umidade durante a metamorfose e resguardar a integridade da mosca até sua emergência para a vida adulta.

Até o terceiro dia após a eclosão, as jovens moscas foram separadas por espécie e sexo de acordo com caracteres morfológicos. As populações formadas a partir dessas moscas (10 machos e 10 fêmeas por gaiola) foram mantidas em recipiente contendo água e tira de papel com uma camada de alimento a base de proteína (60 ml de mel, 180 ml de hidrolisado de proteína de milho, 36 g de levedo de cerveja, 3 g de Sustagen sabor morango, 150 g de açúcar mascavo e 270 g de açúcar branco refinado) por 5 a 6 semanas. A partir da segunda semana foram adicionadas mangas à gaiola para que ocorra a oviposição, com 3 ou 4 trocas de frutas.

#### 3.2 Localidades Amostradas

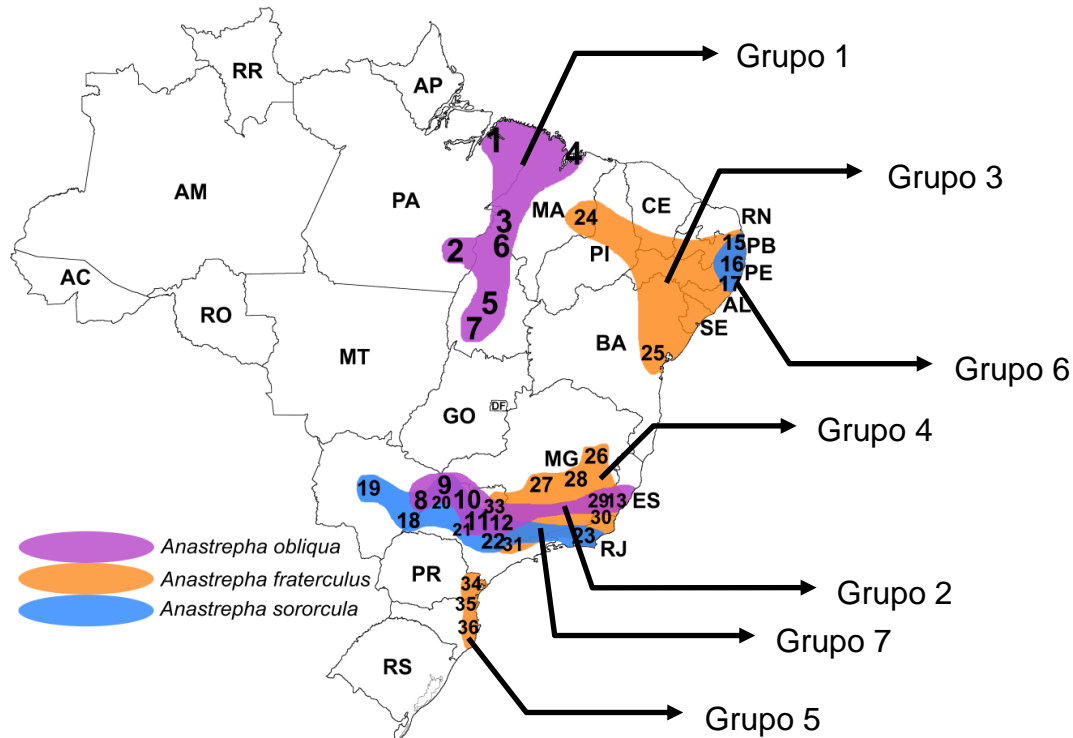
As amostras foram originárias de 38 localidades brasileiras (13 para *A. obliqua*, 16 para *A. fraterculus* e 9 para *A. sororcula*) que foram agrupadas em sete diferentes *pools* representando 3 diferentes regiões do Brasil, nordeste, sudeste e sul; estas divisões não representam as divisões oficiais no Brasil. As localidades (municípios), número de indivíduos por localidade, e espécies de cada *pool* (grupo) seguem abaixo na **Tabela 1**, e na **Figura 3** são indicado sobre o mapa do Brasil as localidades que compõem os grupos que são representadas no mapa.





**Tabela 1.** Localidades das coletas. As coletas foram separadas por grupos que representam as regiões constituintes dos pools.

<u>Anastrepha obliqua</u>					<u>Anastrepha fraterculus</u>				
<b>Grupo I - Nordeste</b>					<b>Grupo III - Nordeste</b>				
Município	UF	N	Latitude	Longitude	Município	UF	N	Latitude	Longitude
1 Belém	PA	1	-01°27'21"	-48°30'16"	24 Estreito	MA	2	-06°33'38"	-47°27'04"
2 Redenção	PA	2	-08°01'43"	-50°01'53"	15 João Pessoa	PB	2	-07°06'54"	-34°51'47"
3 Porto Franco	MA	2	-06°20'18"	-47°23'57"	17 Bonito	PE	2	-08°28'13"	-35°43'43"
4 São Luis	MA	2	-02°31'47"	-44°18'10"	25 Conceição Almeida	BA	1	-12°46'46"	-39°10'12"
5 Palmas	TO	2	-10°12'46"	-48°21'37"	<b>Grupo IV - Sudeste</b>				
6 Babaçulândia	TO	2	-07°12'17"	-47°45'25"	Município	UF	N	Latitude	Longitude
7 Gurupi	TO	2	-11°43'45"	-49°04'07"	26 Marilac	MG	2	-18°30'29"	-42°05'02"
<b>Grupo II - Sudeste</b>					27 Lagoa da Prata	MG	2	-20°01'21"	-45°32'37"
Município	UF	N	Latitude	Longitude	28 Itabira	MG	1	-19°37'09"	-43°13'37"
8 Três Lagoas	MS	2	-20°45'04"	-51°40'42"	29 Conceição do Castelo	ES	2	-20°22'06"	-41°14'38"
9 Estrela d'Oeste	SP	2	-20°17'16"	-50°24'03"	30 Vargem Alta	ES	1	-20°40'17"	-41°00'25"
10 S.J. Rio Preto	SP	2	-20°49'11"	-49°22'46"	31 Campinas	SP	2	-22°54'20"	-47°03'39"
11 Bariri	SP	2	-22°04'28"	-48°44'25"	32 Nova Odessa	SP	2	-22°46'39"	-47°17'45"
12 Cachoeira de Emas	SP	2	-21°59'46"	-47°25'33"	33 Ribeirão Preto	SP	2	-21°10'39"	-47°48'37"
13 Vitória	ES	2	-20°19'10"	-40°20'16"	23 Cabo Frio	RJ	2	-22°52'46"	-42°01'07"
<u>Anastrepha sororcula</u>					<b>Grupo V - Sul</b>				
<b>Grupo VI - Nordeste</b>					Município	UF	N	Latitude	Longitude
Município	UF	N	Latitude	Longitude	34 Curitiba	PR	2	-25°25'40"	-49°16'23"
15 João Pessoa	PB	2	-07°06'54"	-34°51'47"	35 Joinville	SC	2	-26°18'16"	-48°50'44"
16 Recife	PE	2	-08°03'14"	-34°52'52"	36 Florianópolis	SC	2	-27°35'48"	-48°32'57"
17 Bonito	PE	2	-08°28'13"	-35°43'43"	Grupos que formam os 7 pools				
<b>Grupo VII - Sudeste</b>					UF = União Federativa.				
Município	UF	N	Latitude	Longitude	NI = Número de Indivíduos.				
18 Bataguassu	MS	1	-21°42'51"	-52°25'20"					
19 Campo Grande	MS	2	-20°26'34"	-54°38'47"					
20 Estrela d'Oeste	SP	2	-20°17'16"	-50°24'03"					



**Figura 3.** Mapa do Brasil evidenciando os agrupamentos formados. Os números apontam os lugares de coletas, e a mancha delimita o agrupamento das localidades para cada espécie.



### 3.3 Dados Moleculares

Foram selecionados cinco genes das bibliotecas de cDNA dos aparelhos reprodutores feminino de *Anastrepha obliqua* Macquart (GONÇALVES et al, 2011) e masculina de *A. fraterculus* (Pascual et al, 2010) existentes no laboratório, para os quais foram desenhados conjuntos de *primers* para sua amplificação.

### 3.4 Construções de *Primers*

Para determinar as sequências dos *primers* aqui utilizados, as sequências dos genes aqui escolhidos obtidas dos contigs gerados das bibliotecas masculina e feminina de espécies de *Anastrepha* foram alinhadas no programa Clustal W (THOMPSON et al, 1994) implementado no BioEdit 7.0.9 (HALL, 1999) com outras espécies proximalmente relacionadas (outros Tephritidae, ou mesmo outros Diptera como Drosophilidae e Culicidae). Foram escolhidas regiões conservadas e preferencialmente fora da região codificadora para o desenho dos *primers*. Os conjuntos de *primers* foram desenhados seguindo critérios específicos como o tamanho entre 20 a 25 nucleotídeos, concentrações de citosina e guanina entre 50% e 60% evitando repetições de bases. Além disso, valores de  $\Delta G$  e as possibilidades de *Hairpin*, *Self-Dimer* e *Hetero-Dimer* foram calculados no programa *OligoAnalyzer 3.1* (<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>) visando reduzir a produção de artefatos nas reações de amplificação.

Os conjuntos de *primers* (**Tabela 2**) foram testados em diferentes condições de temperatura e concentração de Cloreto de Magnésio ( $MgCl_2$ ), para estipular as melhores condições de amplificação. Os reagentes foram utilizados nas seguintes concentrações:  $MgCl_2$ , 1,5 mM Tris-HCl pH = 9,0 10mM, KCl 50mM, BSA a 0.2 $\mu g/\mu L$ , dNTPs a 0.2mM cada, *Primers forward e reverse* a 0.2  $\mu M$ , Taq polimerase a 0.05 U/ $\mu L$  e DNA a 1 $\mu L/26\mu L$  de solução.

**Tabela 2.** Relação dos primers testados com suas respectivas temperaturas de hibridação (T) em graus Celsius (°C). F e R significam *forward* e *reverse* respectivamente.

Nome	Primer F (5' to 3')	Primer R (5' to 3')	T (°C)
<i>CG11912</i>	AATAATGAAGACTTTCATTGT	ATTTTCGTAATGCTTCAAAGAC	52
<i>Df31</i>	GAAACTCCAATTGTCGAGAAAGT	AACCTCTGATGATGCTTCGGAACC	58
<i>CG10031</i>	TGCTTCAGCATCGATCATCAC	CGTCAGTCAACGCATTCGCGC	64
<i>CG16713</i>	AAATGAAATTCTTCGCTGTAAT	TATCCTTTTATTCCACGCATT	55
<i>Lcp65Ac</i>	TCTTATTAAAGATGAAGTGTGC	TGTTAGCAAAGGAACGATCA	51



### 3.5 Extração de DNA

A extração do DNA foi feita através de protocolo que faz uso da digestão do material com proteinase K em tampão apropriado seguida de obtenção do DNA com Tiocianato de Guanidina (NELSON e KRAWETZ, 1992). A extração foi realizada individualmente para cada mosca o DNA extraído quantificado e posteriormente alíquotas deste extraído foram agrupadas de acordo com as localidade indicado na **Tabela 1** e sombreadas na **Figura 3**. Os sete *pools* (chamados no presente trabalho de grupos) foram montados com alíquotas de 20 µl por indivíduo para cada grupo.

### 3.6 Obtenção e preparação das sequências para o sequenciamento

As amplificações dos genes aqui estudados seguiram as concentrações padronizadas para os *primers* na técnica de reação em cadeia da polimerase. Para todos os genes foram realizados 35 ciclos nas temperaturas de hibridação determinadas para os *primers* no tempo de 30 segundos. As temperaturas e respectivos tempos padronizados para as demais etapas das reações foram: desnaturação a 94°C por 30 segundos, extensão 72°C por 55 segundos, e temperatura final de extensão a 72°C por 10 minutos.

Os produtos gerados nas amplificações de todos os genes em todas as etapas (amplificação pré-clonagem e pós-clonagem) foram purificados pelo protocolo de precipitação diferencial com Polietileno Glicol 8000 (LIS e SCHLEIF, 1975) e a qualidade e quantidade do material verificadas em géis de agarose 1% corados com brometo de etídio para visualização sob luz ultravioleta. O tamanho e concentração dos fragmentos amplificados foram conferidos por comparação com marcadores de peso molecular, *GeneRuler<sup>TM</sup> 1kb DNA Ladder* (Fermentas).

Os produtos amplificados foram individualizados através de clonagem bacteriana com o kit de clonagem InsTAclone (Fermentas), seguindo recomendações do fabricante. Esta técnica permitiu de forma simples e com uma única clonagem por grupo a individualização dos alelos e posteriormente uma avaliação global da variabilidade intra e interespecífica para cada um dos genes aqui estudados. Para realizar a clonagem utilizamos produtos amplificados e já purificados nas reações de ligação com o plasmídeo PTZ57R/T usado no kit acima citado, inserido em bactérias *DH5α* competentes. A transformação foi realizada pela metodologia do choque térmico de acordo com procedimentos padrão utilizando 100ng de plasmídeo por reação e plaqueados em placas de Petri com meio de cultura seletivo (com adição de ampicilina na concentração de 0,1µg/µL), e deixado em estufa de crescimento (na temperatura de 37°C) por 18 horas. Decorrido o tempo necessário para o crescimento bacteriano, colônias isoladas foram selecionadas e aquecidas a 90°C em 20 µl de água milliQ por 5 minutos e submetidas diretamente à reação de amplificação com



*primers* M13 (DALLAS-YANG et al, 1998) para detectar as colônias que continham os insetos.

### 3.7 Sequenciamento e processamento das sequências

As amostras amplificadas dos clones são novamente purificadas com PEG8000 conforme descrito acima e ressuspendidas em água ultrapura (26 mL) para serem enviadas à empresa Macrogen Inc. na Coreia do Sul ([www.macrogen.com](http://www.macrogen.com)), para o sequenciamento em sequenciadores ABI 3730. Os resultados foram disponibilizados eletronicamente no site da empresa.

A análise inicial do material obtido pelo sequenciamento foi feita no programa *Finch TV Version 1.4.0* disponível no site [www.geospiza.com](http://www.geospiza.com). Este programa possibilita uma rápida e eficiente análise das sequências quanto à qualidade de suas bases na visualização do cromatograma, propiciando a rejeição das sequências de baixa qualidade através da delimitação pelos *primers forward* e *reverse*, analisando somente as sequências entre os dois *primers*.

O alinhamento das sequências dos genes aqui obtidos com outras sequências similares (supostamente ortólogos) de moscas de grupos relacionados, por exemplo tefritídeos como *Rhagoletis pomonella*, e drosofilídeos como *Drosophila melanogaster* foi realizado com o programa Clustal W, para cada gene individualmente e os alinhamentos exportados para serem utilizados em outros programas. Com intuito de confirmar se as regiões amplificadas correspondem às classes de genes esperadas, foram feitas buscas sistemáticas nos bancos de dados NCBI e Flybase a partir tanto das sequências de DNA originais quanto de sua tradução em aminoácidos, o que foi realizado no site <http://insilico.ehu.es/translate/>.

### 3.8 Análises dos dados

#### 3.8.1 Testes de Neutralidade e Índice de Polimorfismo

Métodos que levam em consideração  $\omega$  (relação entre substituições sinônimas e não-sinônimas) para inferir padrões de seleção têm sido comumente usados para se inferir seleção positiva (BERLIN e SMITH, 2005; CASTRIC e VEKEMANS, 2007; CLARK e SWANSON, 2005; GALINDO et al, 2003). Tais métodos são úteis por permitir a identificação de regiões nos genes que apresentam mais alterações de aminoácidos do que seria esperado por acaso, indicando regiões que poderiam estar sob seleção positiva. Devemos, contudo, diferenciar seleção positiva direcional, quando há substituições que favorecem aminoácidos ou conjuntos de aminoácidos específicos em determinadas posições da seleção positiva diversificadora que apresenta substituições que favorecem quaisquer substituições nucleotídicas independente do aminoácido. Isto pode acontecer quando há seleção dependente de frequência negativa, por exemplo, em situações de



seleção sexual com favorecimento do macho raro, ou mesmo em outras situações nas quais há mudanças contínuas em resposta co-evolutiva ao ambiente, tais como caracterizadas pela relação parasita-hospedeiro através da corrida armamentista (KOSAKOVSKY-POND et al, 2007).

Estimativas de  $\omega$  globais e para diferentes ramos da filogenia, foram obtidas por máxima verossimilhança usando os softwares PAML (YANG, 2007) e o Hyphy (ver. 2.1.0) (KOSAKOVSKY-POND et al, 2011). Árvores de máxima verossimilhança foram inferidas no programa PhyML v. 3.0, usando modelos evolutivos previamente inferidos no jModelTest v.0.1.1 para cada gene com oito sequências, sendo 4 correspondentes ao grupo *Anastrepha*, uma de *Rhagoletis pomonella* e diferentes espécies de *Drosophila*, (*D. melanogaster*, *D. mojavensis* e *D. grimshawi* para os genes *CG16713*, *CG11912*, *CG10031*; *D. melanogaster*, *D. mojavensis* e *D. yakuba* para o gene *Df31*; *D. melanogaster*, *D. willistoni* e *D. virilis* para o gene *Lcp65Ac*). Estes testes utilizam apenas um espécime de cada espécie de *Anastrepha* porque estes modelos evolutivos lidam com estimativas de  $\omega$  que são estimativas que têm maior confiança quando estimadas em ramos evolutivos maiores e sofrem interferência de polimorfismos ainda segregando nas populações, de forma que a inclusão de outros espécimes de *Anastrepha* não apenas reduz os graus de liberdade para as estimativas aqui feitas como podem levar a interpretações errôneas (KRYAZHIMSKIY e PLOTKIN, 2008). Para cada gene, a árvore mais verossímil foi utilizada no modelo M0 do CODEML (pacote PAML), e seu resultado utilizado nos modelos subsequentes M1a (modelo Nulo), MA (modelo alternativo) e MAnull (modelo nulo). Para esta análise, cada árvore acima mencionada foi dividida nos modelos MA (modelo alternativo) e MAnull (modelo nulo) em dois ramos, *foreground* e *background*, enquanto no modelo M1a a árvore não foi subdividida. Espécies de *Anastrepha* foram consideradas o ramo *foreground* no qual foi testada a ocorrência de seleção positiva e o *background* destinado ao restante da árvore, considerando três diferentes valores para  $\omega$  que são:

$$\omega_0 \rightarrow 0 < \omega < 1, \omega_1 \rightarrow \omega = 1 \text{ e } \omega_2 \rightarrow \omega > 1,$$

que foram usados nos três modelos. No modelo M1a a árvore é considerada em sua totalidade e os sítios têm liberdade para variar entre si, com o  $\omega$  podendo assumir valores correspondentes ao  $\omega_0$ ,  $\omega_1$ . No modelo MA, a árvore é separada em *foreground* e *background* e aqui os sítios também podem variar entre si. No ramo *foreground*,  $\omega$  assume os três valores possíveis  $\omega_0$ ,  $\omega_1$  e  $\omega_2$ . No ramo *background*, as taxas dos sítios podem variar livremente entre si, desde que  $\omega$  assumam valores entre  $\omega_0$  e  $\omega_1$ . No modelo MAnull, a árvore também é dividida em *foreground* e *background*. No ramo *foreground*, as taxas dos sítios podem variar entre si e  $\omega$  assume os três valores possíveis  $\omega_0$ ,  $\omega_1$  e  $\omega_2$  como no modelo MA, porém com uma diferença importante, o valor do  $\omega_2$  ( $\omega > 1$ ) é alterado



ficando fixado em 1 ( $\omega=1$ ). No ramo *background* os sítios podem variar entre si, com o  $\omega$  assumindo valores correspondentes a  $\omega_0$  e  $\omega_1$ . Os modelos então são contrastados em dois diferentes testes, o *relaxed branch-site test*, que usa os modelos MA X M1a e o *strict branch-site test* que utiliza os modelos MA X MAnull.

As metodologias implementadas no CODEML podem ser usadas para se detectar seleção positiva agindo em uma linhagem particular (*foreground*) da árvore filogenética. O *strict branch-site test* é útil para detectar seleção positiva após duplicação gênica, em que uma cópia duplicada pode ter adquirido uma nova função e pode ter assim evoluído a taxas aceleradas. O modelo de sítios trata a taxa  $\omega$  para qualquer sítio (códon) no gene como uma variável aleatória para a distribuição estatística, assim permitindo  $\omega$  variar entre códons. Seleção positiva é definida como presença de códons em que  $\omega > 1$ . A LRT (testes de razão de verossimilhanças) foi estimada para comparar o modelo nulo que não permite quaisquer códons com  $\omega > 1$ , contra um modelo mais geral que o permite (utilizando o  $\chi^2$  com os valores críticos de 3,84 e 5,99) (YANG, 2007). Para tanto, foi utilizado o programa CODEML implementado no pacote PAML para detecção de sinais de seleção positiva.

Recentemente, a amplitude da aplicabilidade do modelo do *branch site test* implementada no PAML foi questionada (KOSAKOVSKY-POND et al, 2011) em função de potenciais erros tipo 1 e tipo 2, de forma que utilizamos também análises de seleção direcional e diversificadora de acordo com KOSAKOVSKY-POND et al, 2011) no programa Hyphy ver. 2.1.0 KOSAKOVSKY-POND et al, 2005. Este programa também realiza análise para seleção positiva, porém utiliza de metodologias distintas para sua identificação. O Hyphy apresenta diferenças em relação ao modelo *strict branch-site test* (YANG, 2007), uma vez que as taxas de substituições podem variar de ramo em ramo e de sítio para sítio e esta incorporação é realizada através de efeitos aleatórios (Probabilidade de efeitos aleatórios – REL *branch-site*) (KOSAKOVSKY-POND et al, 2011). A análise implementada no Hyphy permite  $\omega$  variar livremente entre os ramos da árvore, que não precisa estar separada nos ramos *foreground* e *background*, assim testando cada ramo individualmente para seleção positiva. Outros modelos utilizados são os programas MEDS e FEEDS. O programa MEDS (modelo episódico de seleção direcional) permite diferenciar seleção positiva direcional da diversificadora atuando nos ramos da árvore filogenética. Por sua vez, o programa FEEDS tem maior poder em detectar seleção positiva direcional. No MEDS dividimos a árvore em ramos *foreground* e *background* para que seja detectada seleção direcional ao longo do ramo *foreground*, enquanto no ramo *background* assume-se que não tenha havido seleção direcional (mas não necessariamente neutra), pois esta divisão permite que ambos os ramos assumam uma



única taxa de substituição sinônima, mas para cada ramo é permitida suas próprias taxas de substituições não sinônimas. O FEEDS irá permitir que a taxa de substituição sinônima varie em ambos os ramos. Pressões seletivas por muitas vezes não são constantes sobre toda a filogenia, mas sim, restrita a subconjuntos dos ramos: isto é episódico. E a seleção direcional atuará favorecendo substituições e um ou mais aminoácidos alvos nos ramos que pode ser contrastada com a seleção diversificadora, em que substituições nucleotídicas que mudam os aminoácidos são favorecidas independentemente de qual aminoácido os substituem em uma determinada linhagem. Portanto, seleção diversificadora aproxima de maneira contínua mudanças co-evolutivas ambientais, tipificado pela co-evolução da corrida armamentista dos patógenos-hospedeiros. (KOSAKOVSKY-POND et al, 2007)

Os testes de neutralidade e polimorfismo populacionais foram realizados no programa DnaSP ver.5.10.01 (LIBRADO e ROZAS, 2009). Os seguintes testes foram realizados neste programa, teste D de Tajima, D e F de Fu e Li, H de Fay e Wu, diversidade haplotípica, diversidade nucleotídica e theta-W ( $\Theta_w$ ). Assim os valores positivos significativos do teste D de Tajima, que se baseia no número de sítios segregantes amostrados e no número médio de diferenças nucleotídicas entre cada par de sequências, indicam um excesso de alelos com frequência intermediária, quando comparadas com as frequências esperadas em neutralidade, com isso, tem-se um sinal de seleção balanceadora. Já valores negativos de D de Tajima correspondem a um excesso de alelos com baixa frequência dando sinal de seleção purificadora. Os valores positivos de D de Tajima também podem ocorrer por subdivisão populacional, enquanto valores negativos podem ser resultantes de contração seguida de expansão populacional (TAJIMA, 1989a; TAJIMA, 1989b). O teste H de Fay e Wu que usa grupo externo leva em consideração o número médio de diferenças nucleotídicas entre os pares de sequências formados pelo grupo externo e pelas sequências da população em estudo e a frequência dessas variantes na população (FAY e Wu, 2000), sendo negativo caso haja excesso de mutações abundantes (sinal de seleção positiva) e positivo caso haja excesso de mutações intermediárias dando sinal de seleção balanceadora (ZENG et al, 2006). O teste D de Fu e Li é baseado na diferença entre  $\theta$  estimado a partir do número total de divergências ( $\theta_d$ ) e estimado a partir do número total de polimorfismos ( $\theta_p$ ). O Teste F de Fu e Li são baseados na diferença entre o valor de  $\theta$  estimado a partir do número total de divergências ( $\theta_d$ ) e um a partir da média de diferenças nucleotídicas entre duas sequências do grupo interno ( $\theta_\pi$ ). Sendo  $\theta_d$  uma estimativa das mutações antigas,  $\theta_p$  e  $\theta_\pi$  das mutações recentes. Valores de D e F de Fu e Li serão negativos quando a proporção de mutações recentes forem maiores e serão positivas quando forem menores (FU e LI, 1993).



### 3.8.2 Análise Populacional

O *Fst* par-a-par para cada um dos genes aqui estudados e os valores médios para todos os genes foram calculados no software DnaSP e representados graficamente mostrando os índices médios de *Fst* estimados tanto para as três espécies quanto para os sete grupos.

O programa Arlequin ver 3.5.1.3 foi utilizado para análise de variância molecular (AMOVA) (EXCOFFIER e LISCHER, 2010). Esta análise produz estimativas de componentes de variância de forma análoga à estatística F, refletindo a correlação da diversidade haplotípica em diferentes níveis de subdivisão hierárquica. Para estas subdivisões são estimados os índices de  $\Phi_{st}$ ,  $\Phi_{ct}$ , e  $\Phi_{sc}$ . O  $\Phi_{st}$  se refere à correlação de haplótipos aleatórios dentro da população, em relação aos pares aleatórios de haplótipos extraídos de toda espécie, o  $\Phi_{ct}$  como a correlação dos haplótipos aleatórios dentro do grupo de populações, em relação aos pares de haplótipos aleatórios extraídos de toda espécie e o  $\Phi_{sc}$  como a correlação de diversidade molecular dos haplótipos aleatórios dentro da população, em relação aos pares de haplótipos aleatórios extraídos da região (EXCOFFIER et al, 1992). A análise de variância molecular foi realizada considerando-se estruturas distintas, buscando avaliar o quanto da variação genética é explicado pelas diferenças entre as localidades e entre as espécies.

Inicialmente, investigamos a partição da variação genética por regiões e espécies. Para esta análise, foi montada uma estrutura com um único agrupamento com todos os grupos formados (Entre Grupos). Esta análise não distingue se o efeito se deve ao fato de serem espécies distintas, ou por estarem em regiões distintas. Assim, uma segunda AMOVA considerou uma estrutura com três agrupamentos separando *A. obliqua*, *A. fraterculus* e *A. sororcula* para avaliar o quanto da variação seria explicada tanto pela diferença entre espécies ou pelas diferenças entre indivíduos de uma mesma espécie mas de grupos diferentes (Espécies Agrupadas). Buscando separar o efeito geográfico, realizamos outra análise que considerou apenas o efeito das separações entre as regiões geográficas, independente dos indivíduos pertencerem ou não à mesma espécie, separando a região nordeste e norte da região sudeste e da região sul (Entre Regiões). Finalmente, foi realizada uma análise em que apenas a separação das espécies foi considerada, sem se considerar as diferenças geográficas entre as mesmas (Por Espécie).

### 3.8.3 Rede Haplotípica

Uma rede haplotípica permite a incorporação de informações genealógicas associadas com níveis de divergência para a inferência das relações entre os haplótipos. Tais redes haplotípicas





podem ser criadas com dados de sítios de restrição ou com sequências de nucleotídeos, e são particularmente úteis para inferir genealogias quando a divergência entre os haplótipos é baixa. As relações genealógicas entre as sequências obtidas para os genes aqui estudados foram feitas por parcimônia estatística, respeitando os limites de 95% da parcimônia, usando o algoritmo desenvolvido por TEMPLETON et al. (1992) e implementado no programa TCS v. 1.21 (CLEMENT et al, 2000) (disponível no site: <http://darwin.uvigo.es/software/tcs.html>). Esta análise foi realizada com a região codificante dos genes que possuem *íntrons*, ou seja, *CG16713*, *Lcp65Ac*, *Df31*, e *CG10031*.

### 3.8.4 Análise de Migração

Para a obtenção das taxas de migração foi utilizado o programa Migrate-n (BEERLI e PALCZEWSKI, 2010) que estima as taxas de migrações passadas entre  $n$  populações, assumindo o modelo de matriz de migração com migração assimétrica e em diferentes tamanhos das subpopulações, utilizando a metodologia de máxima verossimilhança para estimar todos os parâmetros em conjunto. Foram realizadas 20 cadeias curtas, amostradas 50000 árvores, destas 500 foram gravadas. Foram feitas 4 cadeias longas nas quais foram amostradas 500000 árvores que tiveram 5000 gravadas, e os valores de theta e migração foram estimados por *Fst*. O modelo de migração é por matriz com theta variável, com taxas de mutação constantes em todos os loci. Este programa está disponível em <http://popgen.sc.fsu.edu/Migrate/Migrate-n.html>.

### 3.8.5 Árvore de Espécies

BUCKy é um programa para combinar os sinais filogenéticos de vários loci ortólogos com potencial conflito em suas árvores filogenéticas. A principal árvore de concordância estimada resume o sinal vertical filogenético compartilhado pela maior proporção de loci, enquanto estima fatores de concordância (FC) e fornece informações sobre o sinal horizontal. BUCKy toma como entrada os arquivos de topologia de árvores geradas pela análise Bayesiana de cada locus individualmente pelo programa MrBayes v. 3.1.2. Neste caso, este programa gerou um total de 2 milhões de árvores que foram amostradas a cada 100. O programa mbsum sumariza as árvores obtidas pelo MrBayes para um valor desejado criando o output necessário para análise bayesiana no programa BUCKy. Vinte mil árvores amostradas no MrBayes foram sumarizadas para 5 mil para cada gene. O BUCKy a partir destes *outputs* realiza a análise bayesiana que consiste da amostra de árvores de genes de sua distribuição conjunta a partir da qual os fatores de confiança são estimados com intervalos de confiança. Finalmente, esses fatores de concordância são usados para produzir o *output* principal, uma árvore de concordância da população (LARGET et al, 2010). A árvore de



espécie obtida pelo BUCKy com os 5 genes contou com 24.957 diferentes topologias de árvores. Uma recente comparação entre duas metodologias de análises bayesianas de inferência de árvores, demonstrou que o BUCKy além de ser mais rápido na análise (contando desde da primeira análise realizada no MrBayes até a obtenção da árvore de concordância pelo BUCKy propriamente dito) do que o BEST (*Bayesian estimation of species trees*), também é eficiente tendo demonstrado na sua adequação ao modelo coalescente provado ter baixa taxa de erro Tipo I (CHUNG e ANÉ, 2011).

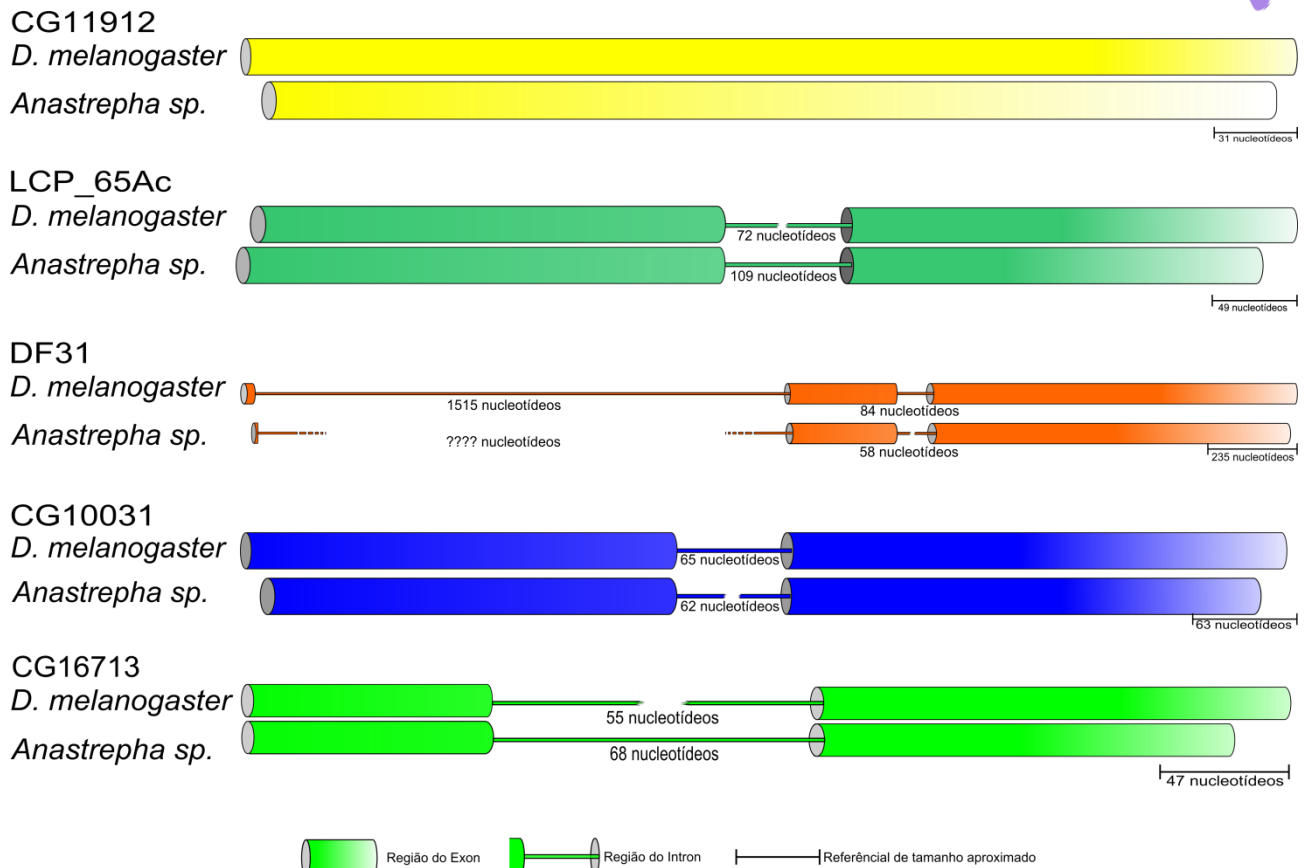


## 4 Resultados

### 4.1 Primers e Alinhamento

Na **Tabela 1** são mostrados os cinco conjuntos de *primers* desenhados. As ampliações com estes *primers* foram bem sucedidas e resultaram em produtos amplificados que correspondiam às regiões esperadas. A clonagem destes produtos atingiu em média 85% de colônias positivas que geraram 345 sequências com qualidade para serem utilizadas. Os alinhamentos dos genes *CG10031*, *Df31*, *Lcp65Ac*, *CG11912* e *CG16713* estão apresentados no **Anexo I**.

Os fragmentos sequenciados não abrangem necessariamente toda a extensão dos genes estudados, como é demonstrado graficamente na **Figura 4** em função da necessidade de adequação do posicionamento dos *primers*. Para o gene similar a *CG10031*, uma serpina, foram sequenciados 451 nucleotídeos com um *intron* de 65 nucleotídeos. Isso corresponde a todos os 78 aminoácidos inferidos para a ORF de *A. obliqua*, em contraste aos 84 observados em *D. melanogaster*. O gene similar a *CG11912*, uma protease com atividade endopeptidase tipo serina, apresentou 837 nucleotídeos e não apresentou *intron*, correspondendo a todos os 268 aminoácidos inferidos da ORF de *A. obliqua*, em contraste aos 270 observados em *D. melanogaster*. O gene *Df31*, *Sperm Decondensation Factor* (Fator de descondensação de espermatozoide) teve 471 nucleotídeos sequenciados, com um *intron* de 58 nucleotídeos. O primer *forward* deste gene foi posicionado após o 1º *intron* em função do tamanho do mesmo em *Drosophila melanogaster* ser maior do que 1 kb, o que dificultaria sua amplificação genômica, e também que o 1º *éxon* codifica para apenas 7 aminoácidos em *Anastrepha obliqua* (9 em *D. melanogaster*). Dessa forma, o produto amplificado correspondeu a 136 aminoácidos sequenciados de um total de 141 inferidos do contig de *A. obliqua*. A proteína de cutícula *Lcp65Ac* apresenta 594 nucleotídeos com um *intron* de 236 nucleotídeos localizado próximo ao centro do gene, correspondendo a todos os 104 aminoácidos inferidos para a ORF de *A. obliqua* em comparação com os 109 observados em *D. melanogaster* e a serpina *CG16713* possui 324 nucleotídeos com *intron* de 60 nucleotídeos na posição proximal, correspondendo a todos os 84 aminoácidos inferidos para a ORF de *A. obliqua* em comparação com os 82 observados em *D. melanogaster*. A **Figura 4** mostra graficamente os genes usados neste trabalho, alinhados com seus supostamente homólogos em *Drosophila melanogaster* indicando seus *introns* e a região correspondente sequenciada.



**Figura 4.** Regiões amplificadas, evidenciando as regiões dos *introns* e exons de cada gene e o tamanho sequenciado em *Anastrepha sp.* em comparação à sequências em *Drosophila melanogaster*.

## 4.2 Análises

### 4.2.1 Testes de Neutralidade e Índice de Polimorfismo

As análises para os testes de seleção positiva foram realizadas nos programas PAML (CODEML) e PhyML, os testes para seleção diversificadora e direcional no programa PhyML mostrados nas Tabelas 3 e 4, o teste neutralidade e índice de polimorfismo, realizado no programa DnaSP, e mostrados na **Tabela 5**. As árvores geradas no PhyML para cada gene evidenciando os ramos *foreground* e *background* que foram usadas no PAML estão expostas para cada gene que podem ser vistas no **Anexo 2** de acordo com os seguintes modelos evolutivos, inferidos pelo programa jModelTest ver 1.1: K80+G (-lnL = 1027,9084 e gamma = 0,291) para *CG16713*, TrNef+I+G (-lnL = 4037,8338 e gamma = 0,29) para *CG11912*, TPM3+g (-lnL = 1379,1533 e gamma = 0,509) para *CG10031*, TPM3uf+G (-lnL = 1024,9902 e gamma = 0,5860) para *Df31* e TIM3+G (-lnL = 2143,4574 e gamma = 1,4900) para *Lcp65Ac*.



**Tabela 3.** Testes de seleção positiva realizados no PAML (CODEML).

CG10031					Lcp65Ac								
Model	LnL (Log Likelihood)	Contraste entre os Modelos			Model	LnL (Log Likelihood)	Contraste entre os Modelos						
1-) Branch-site test 1 (relaxed branch-site test)		Contraste	LRT	DF	Prob $X^2$	1-) Branch-site test 1 (relaxed branch-site test)		Contraste	LRT	DF	Prob $X^2$		
M1a	-1094.678	RBST - 1	MA x M1a	16.91	2	0.000213***	M1a	-1144.706	RBST - 1	MA x M1a	4.787	2	0.091297
MA	-1086.223	SBST - 2	MA x MA null	2.438	1	0.118432	MA	-1142.313	SBST - 2	MA x MA null	2.995	1	0.083531
2) Branch-site test 2 (strict branch-site test)							2) Branch-site test 2 (strict branch-site test)						
MA null	-1087.442						MA null	-1143.810					
MA	-1086.223						MA	-1142.313					

CG11912					Df31								
Model	LnL (Log Likelihood)	Contraste entre os Modelos			Model	LnL (Log Likelihood)	Contraste entre os Modelos						
1-) Branch-site test 1 (relaxed branch-site test)		Contraste	LRT	DF	Prob $X^2$	1-) Branch-site test 1 (relaxed branch-site test)		Contraste	LRT	DF	Prob $X^2$		
M1a	-3789.466	RBST - 1	MA x M1a	26.51	2	0.000002***	M1a	-1657.099	RBST - 1	MA x M1a	4.879	2	0.087189
MA	-3776.209	SBST - 2	MA x MA null	7.772	1	0.005305**	MA	-1654.660	SBST - 2	MA x MA null	2.983	1	0.084159
2) Branch-site test 2 (strict branch-site test)							2) Branch-site test 2 (strict branch-site test)						
MA null	-3780.095						MA null	-1656.151					
MA	-3776.209						MA	-1654.660					

CG16713						
Model	LnL (Log Likelihood)	Contraste entre os Modelos				
1-) Branch-site test 1 (relaxed branch-site test)		Contraste	LRT	DF	Prob $X^2$	
M1a	-1237.139	RBST - 1	MA x M1a	5.19	2	0.074634
MA	-1234.544	SBST - 2	MA x MA null	0.283	1	0.594654
2) Branch-site test 2 (strict branch-site test)						
MA null	-1234.685					
MA	-1234.544					

Legenda	
Prob $X^2$	Probabilidade em uma distribuição de qui-quadrado
DF	Grau de liberdade
M1a	Modelo Nulo - ( $\omega_0 < 1, \omega_1 = 1$ )
MA	Modelo Alternativo - ( $\omega_0 < 1, \omega_1 = 1, \omega_2 > 1$ )
MA null	Modelo Nulo - ( $\omega_0 < 1, \omega_1 = 1$ e o $\omega_2$ fixado em 1)
LRT	Likelihood ratio tests
RBST	relaxed branch-site test
SBST	strict branch-site test

\*p < 0,05, \*\*p < 0,01 e p\*\*\* < 0.001

**Tabela 4.** Testes de seleção positiva realizados no software Hyphy

Gene	MEDS	FEEDS	branch-site REL
CG10031	0 sítios	9 sítios (21, 23, 30, 36, 48, 49, 62, 77, 80)	0,00597
CG11912	0 sítios	9 sítios (3, 30, 63, 82, 139, 143, 178, 200, 202)	0,0004
Df31	1 sítio (164)	3 sítios (72, 91, 143)	> 0,05
Lcp65Ac	1 sítio (78)	0 sítios	> 0,05
CG16713	0 sítios	1 sítio (42)	> 0,05



**Tabela 5.** Testes de neutralidade e índices de polimorfismos para cada gene.

Índice de polimorfismo e teste de neutralidade									
CG16713									
Região	N	S	Hd	$\pi$	$\Theta_w$	Fu&Li's D	Fu&Li's F	Fay&Wu's H	Tajima's D
<i>Anastrepha</i>	66	24	0.941	0.0252	0.0771	-1.7959	-1.985	-5.4918	-1.551
<i>A.obli</i>	18	11	0.993	0.0145	0.0259	-1.1107	1.5642	-4.0784	-1.464
<i>A.frat</i>	28	14	0.984	0.0221	0.0299	-0.0948	-0.173	-3.7407	-0.545
<i>A.soro</i>	20	7	0.958	0.0142	0.0186	1.2295	0.9956	-3.7158	-0.552
<i>O Grupo 1</i>	10	7	0.911	0.012	0.017	-1.274	-1.519	-0.622	-1.383
<i>O Grupo 2</i>	8	5	0.929	0.013	0.013	0.471	0.280	-1.857	0.0005
<i>F Grupo 3</i>	10	5	0.933	0.017	0.015	0.592	0.821	-0.267	0.981
<i>F Grupo 4</i>	9	7	0.972	0.020	0.018	-0.121	0.038	0.611	0.639
<i>F Grupo 5</i>	9	6	0.944	0.013	0.015	0.636	0.414	-1.778	-0.740
<i>S Grupo 6</i>	10	2	0.911	0.010	0.010	0.989	1.278	0.267	1.439
<i>S Grupo 7</i>	10	6	0.911	0.016	0.018	1.272	1.154	-2.222	-0.236

Índice de polimorfismo e teste de neutralidade									
LCP_65Ac									
Região	N	S	Hd	$\pi$	$\Theta_w$	Fu&Li's D	Fu&Li's F	Fay&Wu's H	Tajima's D
<i>Anastrepha</i>	66	71	0.977	0.027	0.027	-3.4775*	-3.6656*	-27.9403*	-2.18373*
<i>A.obli</i>	18	21	0.817	0.0072	0.0155	-1.8106	-2.2908	-3.8431	-1.987
<i>A.frat</i>	28	38	1.000	0.0139	0.0242	-1.9538	-2.3448	-11.418	-1.672
<i>A.soro</i>	20	24	0.979	0.013	0.0182	-0.5688	-0.8363	7.6211	-0.775
<i>O Grupo 1</i>	10	15	0.867	0.009	0.013	-1.137	-1.497	-1.600	-1.459
<i>O Grupo 2</i>	8	9	0.786	0.006	0.009	-1.745	-2.048	-0.429	-1.723
<i>F Grupo 3</i>	10	14	1.000	0.010	0.011	-0.410	-0.588	-2.667	-0.385
<i>F Grupo 4</i>	9	17	1.000	0.011	0.014	-1.231	-1.520	-1.222	-1.236
<i>F Grupo 5</i>	9	13	0.972	0.011	0.011	0.306	0.089	-5.806	-0.172
<i>S Grupo 6</i>	10	11	0.933	0.011	0.011	-0.094	-0.215	-2.667	0.026
<i>S Grupo 7</i>	10	20	0.978	0.015	0.017	-0.303	-0.514	-4.267	-0.637

Índice de polimorfismo e teste de neutralidade									
DF 31									
Região	N	S	Hd	$\pi$	$\Theta_w$	Fu&Li's D	Fu&Li's F	Fay&Wu's H	Tajima's D
<i>Anastrepha</i>	66	39	0.905	0.0083	0.0349	-5.1671*	-4.9147*	-6.2956	-2.40811*
<i>A.obli</i>	18	18	0.935	0.0083	0.013	-1.9473	-2.152	-3.2157	-1.402
<i>A.frat</i>	28	28	0.966	0.0073	0.010	-3.7039**	-3.7989**	0.7566	-2.085
<i>A.soro</i>	20	11	0.958	0.0074	0.0077	-2.8171	-2.9806	-0.9474	-1.536
<i>O Grupo 1</i>	10	13	0.933	0.008	0.011	-2.020	-2.427	-3.022	-1.976
<i>O Grupo 2</i>	8	10	0.964	0.010	0.010	-0.063	-0.014	-0.357	0.138
<i>F Grupo 3</i>	10	8	0.967	0.008	0.008	-1.738	-2.029	1.067	-1.626
<i>F Grupo 4</i>	9	5	0.989	0.008	0.005	-1.636	-1.729	0.639	-0.910
<i>F Grupo 5</i>	9	4	0.906	0.007	0.004	-2.031	-2.243	0.583	-0.689
<i>S Grupo 6</i>	10	8	0.944	0.007	0.007	-1.530	-1.842	-0.444	-1.127
<i>S Grupo 7</i>	10	4	0.844	0.006	0.004	-1.427	-1.544	-0.089	-0.94297

Índice de polimorfismo e teste de neutralidade									
CG11912									
Região	N	S	Hd	$\pi$	$\Theta_w$	Fu&Li's D	Fu&Li's F	Fay&Wu's H	Tajima's D
<i>Anastrepha</i>	66	137	0.999	0.02009	0.0401	-1.7982	-2.1957	-64.4625*	-1.733
<i>A.obli</i>	18	18	1.000	0.01266	0.0172	-0.1291	-0.5474	-15.3725	-1.081
<i>A.frat</i>	28	84	0.997	0.01902	0.0307	-0.9949	-1.4753	-33.5714	-1.469
<i>A.soro</i>	20	56	0.995	0.01848	0.0212	-0.299	-0.4536	-17.8526	-0.522
<i>O Grupo 1</i>	10	32	1.000	0.013	0.015	-0.054	-0.293	-7.022	-0.577
<i>O Grupo 2</i>	8	26	1.000	0.012	0.013	0.873	0.630	-7.500	-0.672
<i>F Grupo 3</i>	10	37	1.000	0.019	0.019	0.561	0.454	-3.644	0.036
<i>F Grupo 4</i>	9	42	1.000	0.017	0.021	-0.318	-0.733	-14.556	-1.102
<i>F Grupo 5</i>	9	35	0.972	0.017	0.018	-0.400	-0.516	-7.250	-0.210
<i>S Grupo 6</i>	10	36	0.978	0.014	0.017	-0.559	-0.891	-8.533	-0.766
<i>S Grupo 7</i>	10	26	1.000	0.012	0.012	0.673	0.465	-11.822	-0.318

Índice de polimorfismo e teste de neutralidade									
CG10031									
Região	N	S	Hd	$\pi$	$\Theta_w$	Fu&Li's D	Fu&Li's F	Fay&Wu's H	Tajima's D
<i>Anastrepha</i>	66	61	0.984	0.0254	0.0432	-2.9953*	-2.8642*	-9.1991	-1.394
<i>A.obli</i>	18	22	0.98	0.0147	0.0217	-0.8056	-1.165	-7.6863	-1.281
<i>A.frat</i>	28	32	0.979	0.01894	0.0256	-1.3426	-1.4587	-2.9577	-0.958
<i>A.soro</i>	20	28	0.900	0.02453	0.0246	0.0124	0.0172	-1.5263	-0.009
<i>O Grupo 1</i>	10	12	0.978	0.010	0.013	-1.100	-1.277	-1.511	-0.944
<i>O Grupo 2</i>	8	16	1.000	0.019	0.022	-0.283	-0.453	-3.000	-0.623
<i>F Grupo 3</i>	10	16	1.000	0.015	0.018	-0.423	-0.606	-1.511	-0.700
<i>F Grupo 4</i>	9	13	1.000	0.015	0.015	-0.079	-0.054	-0.583	0.050
<i>F Grupo 5</i>	9	11	0.961	0.011	0.013	-0.337	-0.555	-1.556	-0.507
<i>S Grupo 6</i>	10	14	0.911	0.016	0.015	0.596	0.554	0.711	0.089
<i>S Grupo 7</i>	10	14	0.667	0.010	0.015	-1.100	-1.502	-5.689	-1.557

Legenda			
Regiões	Espécies que as Compõem	Siglas e Símbolos	
<i>O Grupo 1</i>	<i>Anastrepha obliqua (A.obli)</i>	N	Número de seqüências
<i>O Grupo 2</i>	<i>Anastrepha obliqua (A.obli)</i>	S	Número de sítios polimórficos (segregando)
<i>F Grupo 3</i>	<i>Anastrepha fraterculus (A. frat)</i>	Hd	Diversidade Haplotípica
<i>F Grupo 4</i>	<i>Anastrepha fraterculus (A. frat)</i>	$\pi$	Diversidade Nucleotídica
<i>F Grupo 5</i>	<i>Anastrepha fraterculus (A. frat)</i>	$\Theta_w$	Theta-W
<i>S Grupo 6</i>	<i>Anastrepha sororcula (A.soro)</i>	p	p-value
<i>S Grupo 7</i>	<i>Anastrepha sororcula (A.soro)</i>		

\*P < 0.05 considerando-se significância corrigida para Bonferroni (equivalente a P < 0,01\*, 0,003\*\* e 0,001\*\*\*)

\*\*Testes realizados com grupo externo feito com *Rhagoletis pomonella*



### 4.2.2 Análise populacional

Os testes de Variância Molecular (AMOVA) são apresentados na **Tabela 6**. Na primeira análise em que todos os grupos são combinados em um único agrupamento, apenas o gene *CG16713* não apresentou valores significativos para a separação entre grupos. A segunda AMOVA, que avalia se a geografia responde por parte da diversidade encontrada entre as regiões, independentemente da influência das diferenças entre as espécies, não indicou valores significativos para a separação entre regiões, embora como a primeira AMOVA tenha encontrado valores significativos para a separação entre grupos para todos os genes com exceção do *CG16713*. Espécies agrupadas, que considera o quanto da variabilidade genética é explicada pela separação dentro das espécies (entre os grupos de cada espécie), revela apenas um gene com parte da variação sendo explicada pela separação entre espécies (*CG11912*) e por fim, a análise por espécie que traz para como resultados a diversidade genética sendo explicada muito mais pela diversidade dentro das espécies do que entre elas. As análises dos índices de *Fst* par-a-par entre todos os grupos estão representadas na **Figura 12**, que revelam a baixa diversidade encontrada em *A. obliqua* e o potencial fluxo gênico entre as populações do nordeste de *A. fraterculus* e *A. sororcula*.

**Tabela 6.** Resultados das análises da AMOVA, para todos os genes, em quatro análises distintas, Entre Grupos, por espécie e região, Por Região, Por Espécie. Os valores em azul são os valores que apresentam significância para  $P < 0,05$ .

	Entre grupos					
	<i>Lcp65Ac</i>	<i>Df31</i>	<i>CG10031</i>	<i>CG11912</i>	<i>CG16713</i>	
Entre os Grupos	43,22***	35,8*	87,04***	56,92***	11,65	φCT
Dentro dos Grupos	56,78***	64,2*	12,96***	43,08***	88,35	φST
	Entre Regiões					
	<i>Lcp65Ac</i>	<i>Df31</i>	<i>CG10031</i>	<i>CG11912</i>	<i>CG16713</i>	
Entre Regiões	20,01	-76,82	-11,61	-24,08	27,29	φCT
Entre grupos, Dentro das regiões	26,52***	98,61**	98,22***	78,04***	-8,78	φSC
Dentro dos Grupos	53,47***	78,22*	13,39***	46,04***	81,50	φST
	Espécies Agrupadas					
	<i>Lcp65Ac</i>	<i>Df31</i>	<i>CG10031</i>	<i>CG11912</i>	<i>CG16713</i>	
Entre Espécie	-25,41	27,71	24,77	28,8*	-44,57	φCT
Entre grupos, Dentro das Espécies.	65,24***	12,37	63,03***	31,08***	46,96	φSC
Dentro dos Grupos	60,17***	59,92*	12,2***	40,12***	97,62	φST
	Por espécie					
	<i>Lcp65Ac</i>	<i>Df31</i>	<i>CG10031</i>	<i>CG11912</i>	<i>CG16713</i>	
Entre Espécie	19,33***	11,09***	31,51***	21,16***	2,03***	φCT
Dentro das Espécies	80,67***	88,91***	68,49***	78,84***	97,97***	φST

\* $p < 0,05$ , \*\* $p < 0,01$  e  $p^{***} < 0,001$

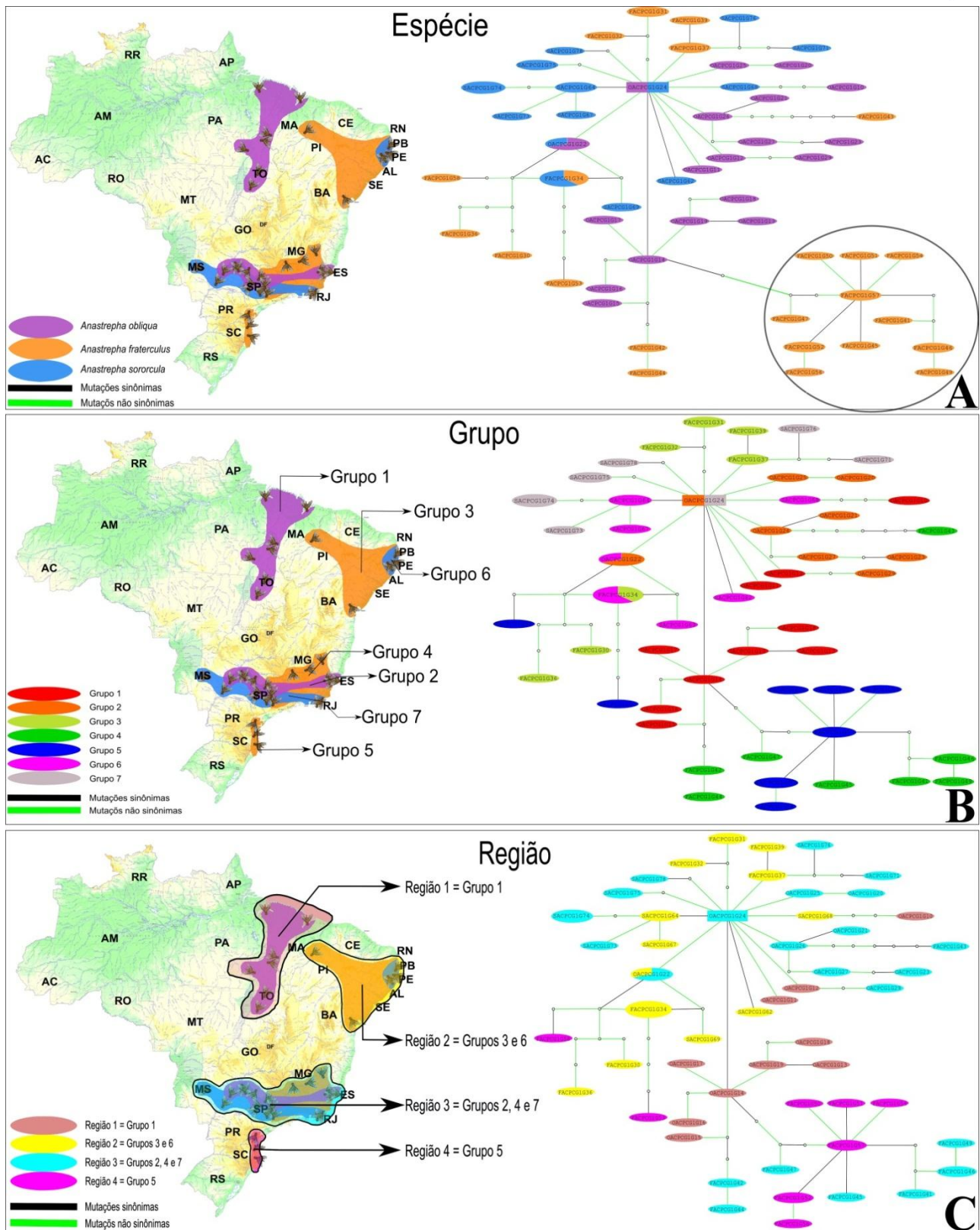


### 4.2.3 Rede Haplótica, análise de migração e árvore de espécies

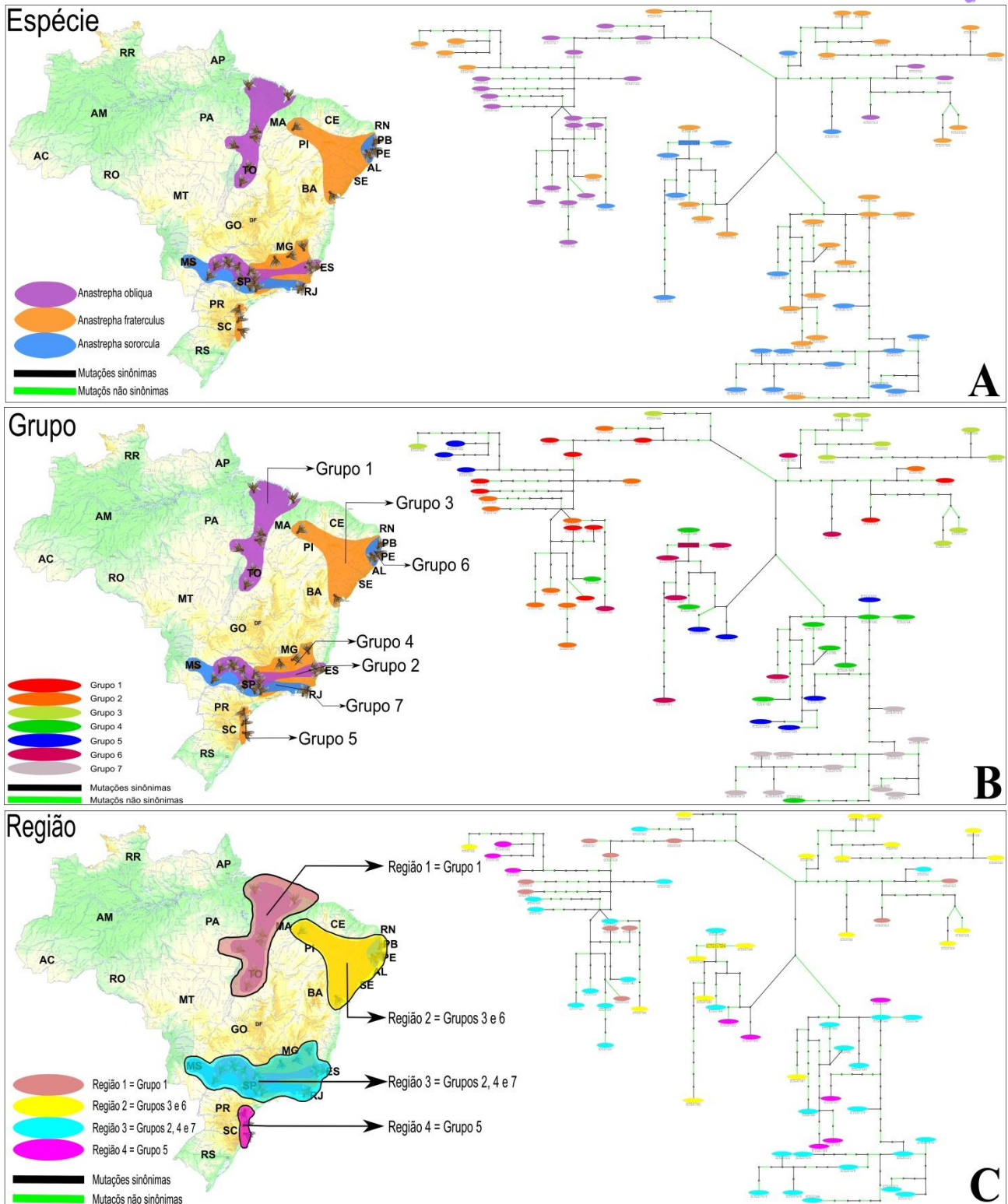
As redes haplotípicas das regiões codificantes para cada um dos genes analisados são apresentadas nas **Figuras 5, 6, 7, 8 e 9** e revelam que em geral tais genes não definem linhagens para as espécies aqui estudadas. No intuito de facilitar a visualização destas relações, as redes de haplótipos são representadas diferentemente para ressaltar possíveis relações entre os grupos, regiões e espécies.

A **Figura 10** mostra a árvore de espécie originada pela análise conjunta dos 5 genes aqui estudados realizada no programa BUCKy, construída pela congruência de ramos das 24.957 diferentes topologias obtidas que revela que, com exceção de populações do Nordeste de *A. fraterculus* e *A. sororcula*, as espécies são definidas por linhagens bem estabelecidas. As estimativas de migração (M) realizadas no programa Migrate, analisando as relações entre as espécies e os diferentes grupos, estão indicadas na **Figura 11**.

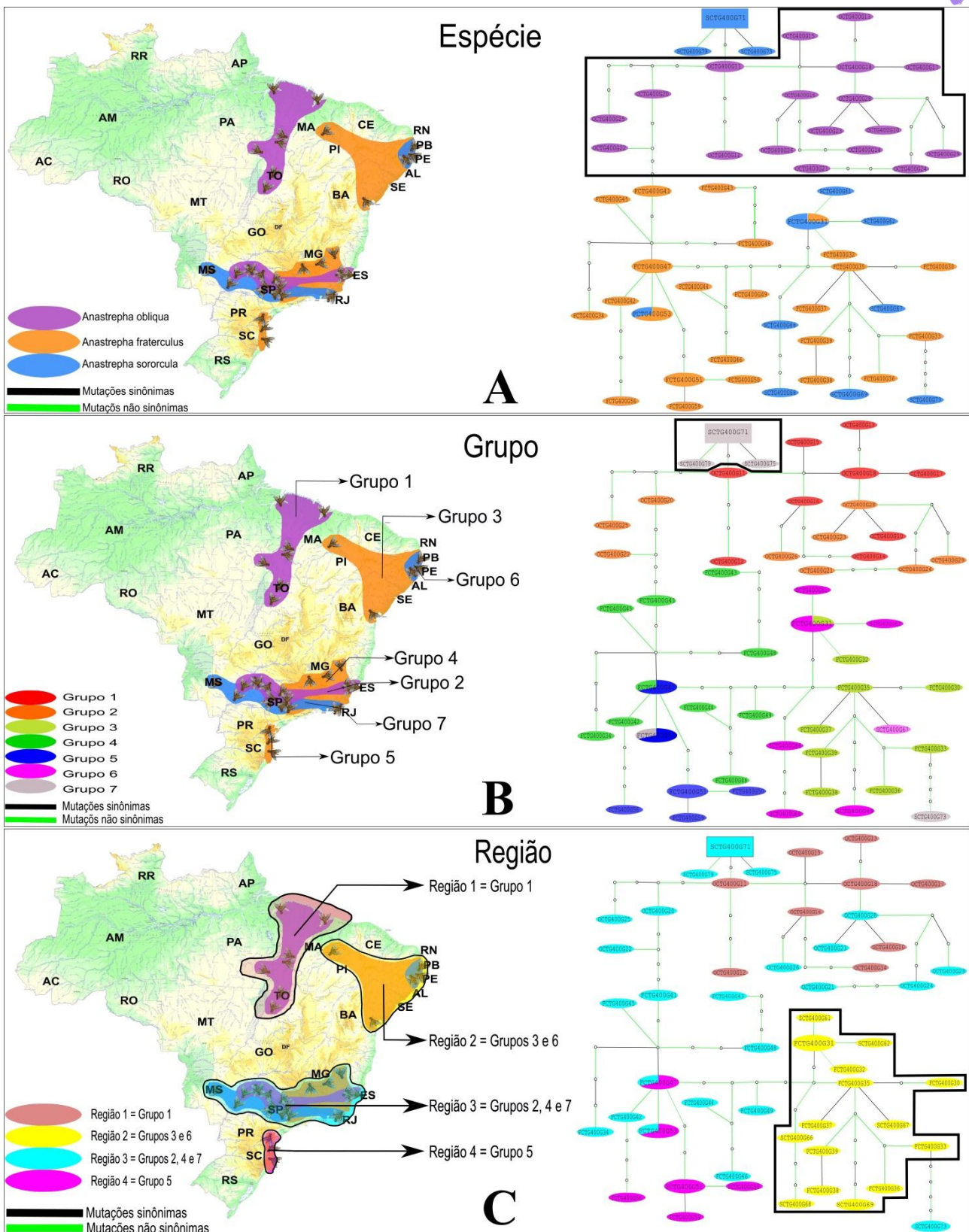




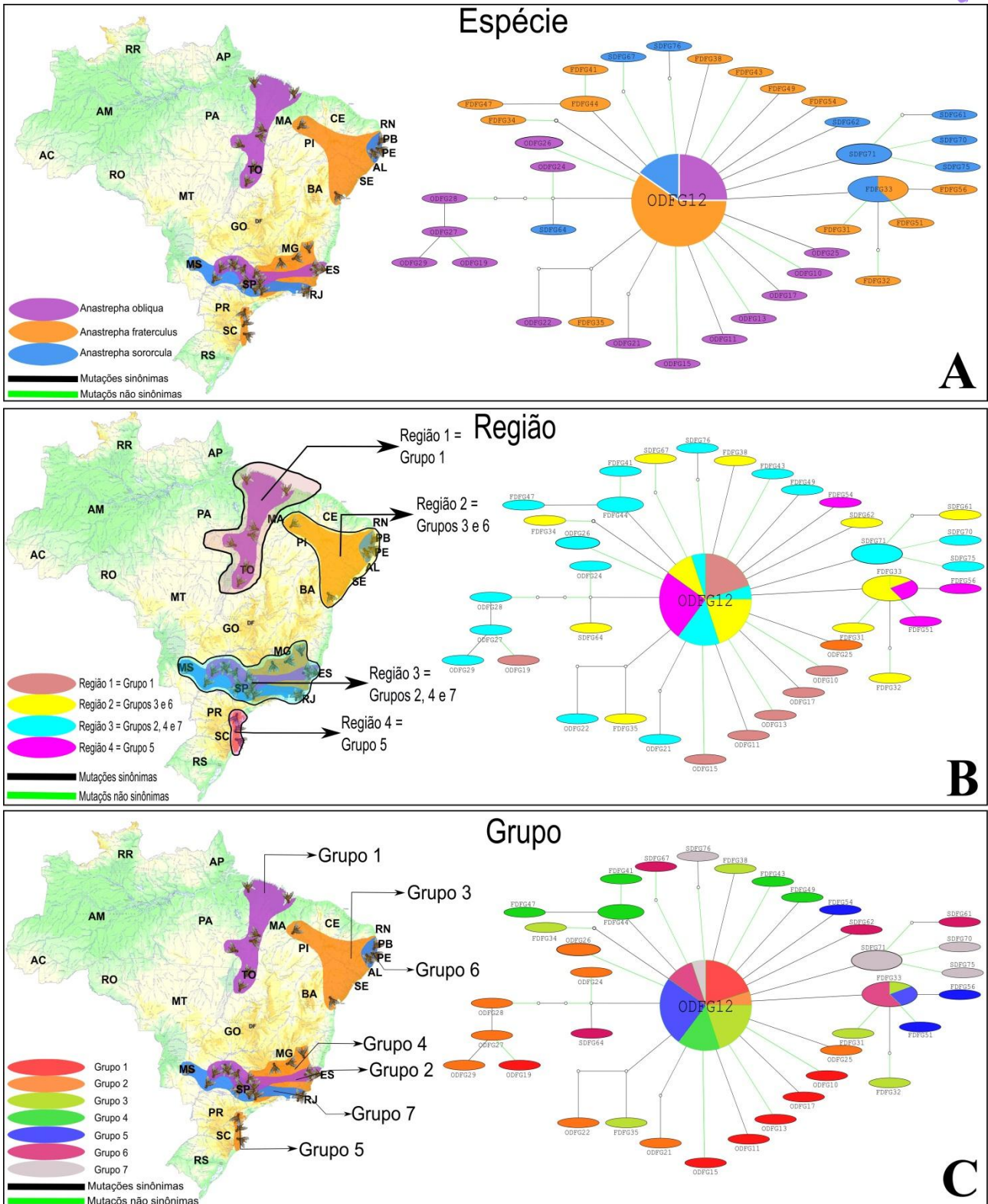
**Figura 5.** Rede haplotípica para a região codificante do gene *CG16713*, indicando as relações filogenéticas entre as três espécies estudadas neste trabalho. Para esta análise foram usadas 20 seqüências para *A. obliqua*, 29 para *A. fraterculus* e 20 para *A. sororcula*. A rede haplotípica **A** ressalta em cores distintas as diferentes espécies, enquanto **B** realça os distintos grupos e espécies e **C** ressalta as amostras de diferentes regiões. Os pequenos círculos vazios são os haplótipos inferidos pelo programa TCS mas não observados na amostragem.



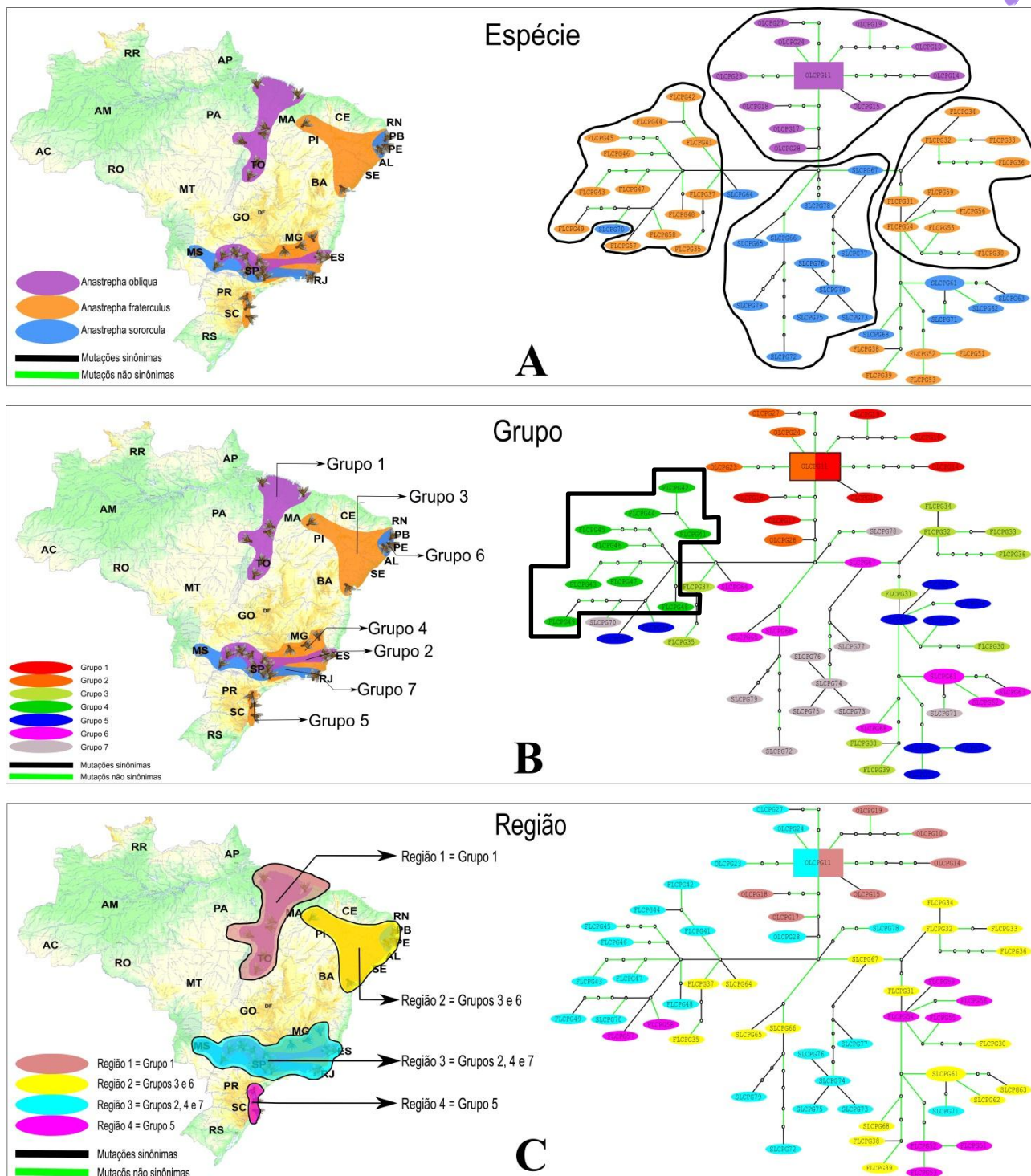
**Figura 6.** Rede haplotípica para a região codificante do gene *CG11912* indicando as relações filogenéticas entre as três espécies estudadas neste trabalho. Para esta análise foram usadas 20 seqüências para *A. obliqua*, 30 para *A. fraterculus* e 20 para *A. sororcula*. A rede haplotípica **A** ressalta em cores distintas as diferentes espécies, enquanto **B** realça os distintos grupos e espécies e **C** ressalta as amostras de diferentes regiões. Os pequenos círculos vazios são os haplótipos inferidos pelo programa TCS mas não observados na amostragem.



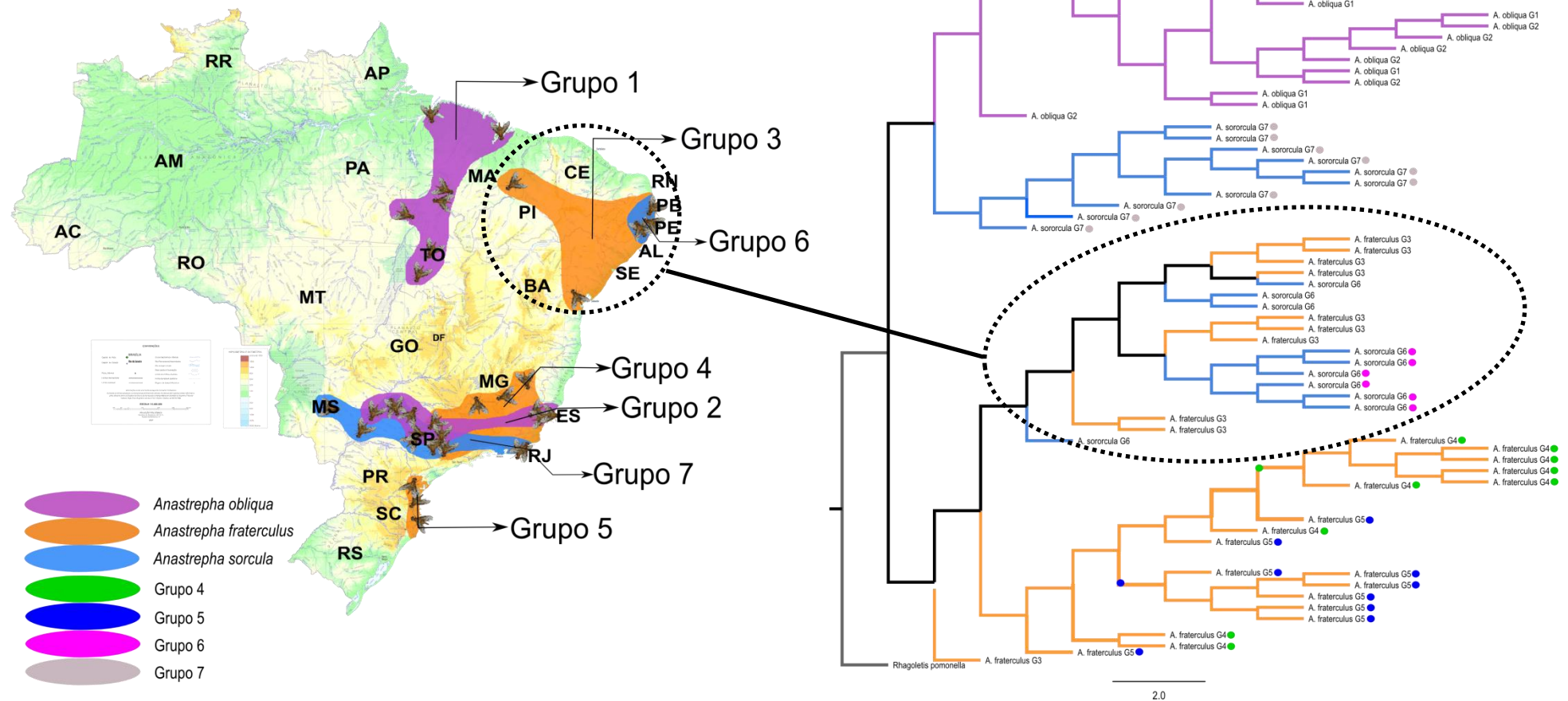
**Figura 7.** Rede haplotípica para a região codificante do gene *CG10031*, indicando as relações filogenéticas entre as três espécies estudadas neste trabalho. Para esta análise foram usadas 20 seqüências para *A. obliqua*, 30 para *A. fraterculus* e 20 para *A. sororcula*. A rede haplotípica **A** ressalta em cores distintas as diferentes espécies, enquanto **B** realça os distintos grupos e espécies e **C** ressalta as amostras de diferentes regiões. Os pequenos círculos vazios são os haplótipos inferidos pelo programa TCS mas não observados na amostragem.



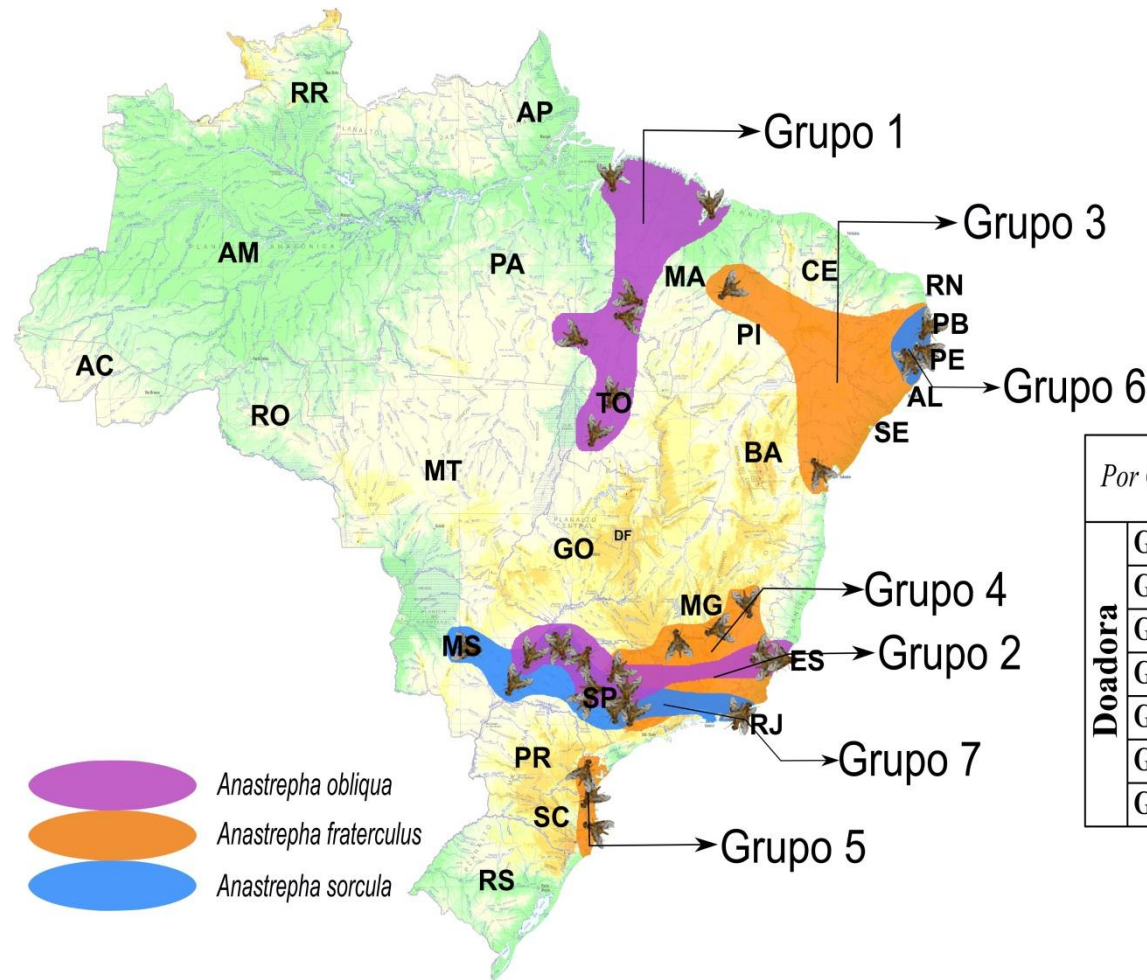
**Figura 8.** Rede haplotípica para a região codificante do gene *Df31*, indicando as relações filogenéticas entre as três espécies estudadas neste trabalho. Para esta análise foram usadas 19 seqüências para *A. obliqua*, 30 para *A. fraterculus* e 20 para *A. sororcula*. A rede haplotípica **A** ressalta em cores distintas as diferentes espécies, enquanto **B** realça os distintos grupos e espécies e **C** ressalta as amostras de diferentes regiões. Os pequenos círculos vazios são os haplótipos inferidos pelo programa TCS mas não observados na amostragem.



**Figura 9.** Rede haplotípica para a região codificante do gene *LCP65Ac* indicando as relações filogenéticas entre as três espécies estudadas neste trabalho. Para esta análise foram usadas 19 seqüências para *A. obliqua*, 18 para *A. fraterculus* e 20 para *A. sororcula*. A rede haplotípica **A** ressalta em cores distintas as diferentes espécies, enquanto **B** realça os distintos grupos e espécies e **C** ressalta as amostras de diferentes regiões. Os pequenos círculos vazios são os haplótipos inferidos pelo programa TCS mas não observados na amostragem.



**Figura 10.** Árvore de espécies gerada pelo programa BUCKy vs 1.4.0. Os ramos coloridos destacam as espécies através das cores representativas e as elipses destacam os grupos com alguma congruência, a linha tracejada indica as relações filogenéticas entre os dois grupos que apresentam sobreposição para região nordeste.

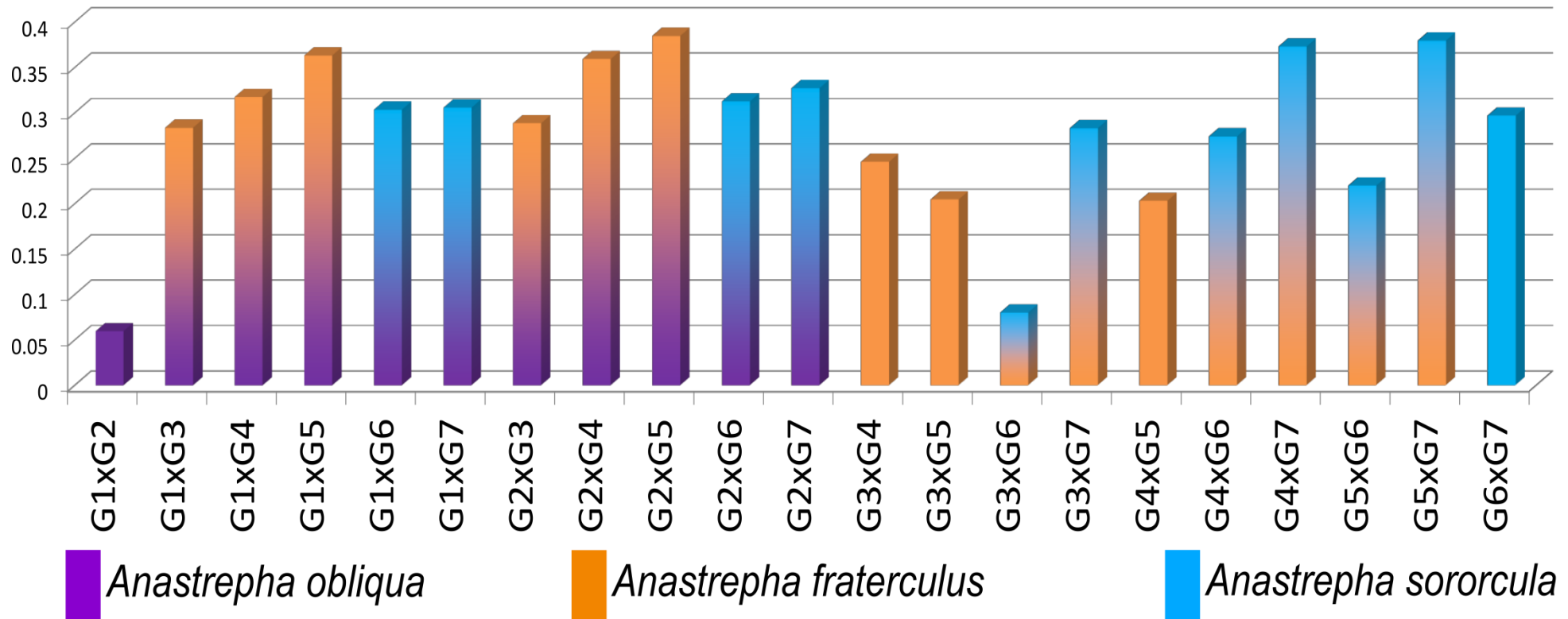


Por Espécie		Receptora		
		<i>A. obliqua</i>	<i>A. fraterculus</i>	<i>A. sorocula</i>
Doadora	<i>A. obliqua</i>		63.94	56.75
	<i>A. fraterculus</i>	2.32		52.75
	<i>A. sorocula</i>	4.28	46.07	

Por Grupo		Receptora						
		Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7
Doadora	Grupo 1		483.30	17.49	76.26	8.48	0	0
	Grupo 2	89.96		0	0	0	161.13	47.36
	Grupo 3	25.57	0		35.80	51.14	66.47	5.12
	Grupo 4	0	0	22.62		149.26	31.91	0
	Grupo 5	0	0	15.83	74.86		0	20.21
	Grupo 6	10.20	0	0	102.31	51.17		48.19
	Grupo 7	20.55	0	68.62	47.93	6.85	68.5032	

**Figura 11.** Estimativas de fluxo gênico inferidas por máxima verossimilhança. Neste resultado são mostradas duas tabelas com os valores de M para migração, sendo a primeira referente a migração entre as Espécies e a segunda referente a migração entre os Grupos. Doadora refere-se à população que cede migrantes, enquanto Receptora é a população que recebe migrantes.

# Fst - Média com todos os genes



**Figura 12.** *Fst* par-a-par médio para todos os genes. A cor roxa na base representa *A. obliqua*, laranja representa *A. fraterculus* e azul esta representando *A. sororcula*.





## 5. Discussão

Neste trabalho buscamos analisar as relações existentes entre três espécies do grupo *fraterculus* para o gênero *Anastrepha*, *A. obliqua*, *A. fraterculus* e *A. sororcula*. As análises feitas aqui visam não apenas um melhor entendimento dos padrões evolutivos destas espécies, suas possíveis relações e estruturas populacionais, mas também uma investigação da utilidade de certos genes com taxas evolutivas mais altas para tais inferências. Como diversas evidências têm indicado que genes expressos nos tecidos reprodutivos evoluem sob taxas mais altas do que quando comparados com genes que não participam dos processos reprodutivos (SWANSON e VACQUIER, 2002; SWANSON et al, 2004), os padrões evolutivos mais rápidos destes genes poderiam ser utilizados como indicadores mais efetivos para distinguir entre espécies proximamente relacionadas. Partindo deste princípio, estudamos um gene considerado evolutivamente conservado e quatro genes que codificam para proteínas potencialmente com rápidas taxas evolutivas: uma protease, duas serpinas e uma proteína cuticular. Esta estratégia buscou investigar se genes potencialmente sujeitos a taxas evolutivas mais altas seriam mais informativos sobre a história evolutiva de moscas das frutas do grupo *fraterculus*.

O primeiro aspecto a ser avaliado é o papel da seleção na diferenciação dos genes aqui estudados. Isto foi avaliado de formas diferentes considerando tanto aspectos macroevolutivos, ao considerar a diferenciação do grupo *fraterculus* com espécies mais distantemente relacionadas, quanto padrões microevolutivos, ao considerar a variação inter e intraspecífica no grupo *fraterculus*.

Quando analisamos os padrões macroevolutivos para estes genes, encontramos um forte sinal de seleção positiva para a protease *CG11912*, detectado tanto no PAML (*strict branch-site test*,  $p = 0,005305$ ) quanto no Hyphy (*branch-site REL*,  $p = 0,0004$ ). Interessantemente, o teste FEEDS indicou nove sítios neste gene sujeitos à seleção positiva diversificadora e nenhum sujeito à seleção direcional. Já a serpina *CG10031* apresenta sinais de restrições seletivas relaxadas (ou seja, uma evolução próxima à neutralidade) no PAML (*relaxed branch-site test*,  $p = 0,000213$ ), mas apresentando evidências de seleção positiva no Hyphy (*branch-site REL*,  $p = 0,00597$ ) e novamente 9 sítios sob seleção positiva diversificadora e nenhum sob seleção direcional. A diferença nos resultados encontrados para as análises realizadas no PAML e no HYPHY pode ser devido ao maior poder deste último em lidar com heterogeneidades das taxas evolutivas na comparação de ramos *foreground* e *background*. Por este motivo, e por termos encontrado este grande número de sítios sujeitos à seleção diversificadora, consideramos que estes dois genes apresentam evidências significativas de seleção positiva diversificadora. A análise para os demais genes no PAML apresentou valores quase significativos para seleção positiva ( $p \sim 0,08$ ), apesar de não apresentar tais



resultados no HYPHY. É possível que estes resultados sejam apenas indicação de ausência de seleção positiva de fato, mas estudos mais detalhados nestes genes são necessários até que possamos resolver melhor esta questão, particularmente porque testes de MEDS e FEEDS indicaram um número reduzido de sítios sob seleção positiva direcional para os genes *Df31* (1 sítio) e *Lcp65Ac* (1 sítio) e seleção positiva diversificadora para o gene *Df31* (3 sítios) e *CG16713* (1 sítio). Outros estudos têm revelado importantes genes da cadeia sexual que estão sujeitos a uma combinação de pressões de seleção opostas em regiões distintas do gene, seleção purificadora em regiões mais conservadas, e seleção positiva em outras regiões (SOBRINHO e DE BRITO, 2010; *subm*). Nestes casos, o processo de identificação das forças evolutivas agindo nestes genes é mais complexo e muitas vezes é apenas capaz de detectar o sinal mais forte de seleção, muitas vezes a purificadora, uma vez que a seleção direcional ocorre de forma mais restrita, tanto no tempo quanto no gene.

Uma vez que alguns dos genes aqui estudados apresentaram sinais de seleção positiva quando comparados com sequências de drosofilídeos, buscamos investigar quais níveis de diversidade, padrões de seleção e sinais filogenéticos tais genes ofereceriam para a diferenciação inter e intraespecífica no grupo *fraterculus*.

Ao considerar os níveis de diversidade apresentados pelos genes aqui estudados, consideramos três níveis hierárquicos. O primeiro considerou a variação regional, em que amostras foram coletadas de localidades diferentes mas geograficamente próximas. Consideramos também a variabilidade apresentada para cada espécie como um todo e finalmente consideramos toda a linhagem *Anastrepha*, neste caso o grupo *fraterculus*.

O gene *Df31* apresentou os valores mais baixos para a diversidade nucleotídica ( $\pi$ ) e também para  $\theta_w$  quando consideramos as espécies individualmente. Tais valores são compatíveis com o papel deste gene que codifica para uma chaperona, que potencialmente apresenta interações *in vitro* com diversas histonas, e evidências fortes para uma função estrutural no dobramento da cromatina e associações nos mecanismos que estão associados a este processo (GUILLEBAULT e COTTERILL, 2007), sugerindo um padrão de seleção purificadora que foi indicado pelos valores significativamente negativos para o gênero nos testes D de Tajima e D e F de Fu e Li. Apenas a espécie *A. fraterculus* apresentou desvios significativo de neutralidade para este gene para D e F de Fu e Li, o que pode se dever a um maior tamanho amostral para esta espécie. É também possível que outras forças evolutivas estejam causando tal desvio, embora uma expansão populacional, o outro evento mais provável de tê-lo causado, não seja muito provável por não ter sido detectada nos outros genes aqui estudados. A rede de haplótipos do gene *Df31* também apresenta padrão de rede



haplotípica em forma de estrela compatível com evolução conservada e seleção purificadora eliminando variação genética. Tal padrão, que também pode ser explicada por uma expansão populacional, é função da seleção purificadora eliminando variação genética deletéria e mantendo ramos curtos sem muito sinal filogenético, de forma que há pouca diferenciação entre espécies e linhagens e pouco sinal filogenético entre as espécies, como pôde ser observado.

O gene que codifica para a proteína cuticular *Lcp65Ac* também apresenta valores reduzidos de  $\theta_w$ , sendo os mais baixos para os genes estudados. Interessantemente, os valores de  $\pi$  para as espécies individualmente foram baixos, mas o valor global para o grupo *fraterculus* foi o mais alto de todos os genes. Como o *Df31*, o gene *Lcp65Ac* apresentou desvios significantes da neutralidade, para todos os testes realizados para o grupo *fraterculus*. No entanto, a rede haplotípica para este gene é bastante diferente da encontrada para *Df31*, apresentando diversos haplótipos que definem grupos aparentemente bem distintos dentro das espécies, alguns dos quais definidos por mutações não-sinônimas. Na verdade, este gene separa as três espécies estudadas, porém subdivide *A. fraterculus* basicamente em dois agrupamentos distintos. Considerando a diferenciação entre as espécies aqui encontrada e a segregação da diversidade, é possível que o desvio de neutralidade encontrado reflita esta heterogeneidade observada para o grupo *fraterculus* mais do que um papel purificador da seleção. Muito embora diversas proteínas de cutícula apresentem padrões compatíveis com ação de seleção positiva (CORNMANN et al, 2008), isso não acontece para todas, uma vez que diversas destas proteínas são expressas em larvas e pupas e não estão, necessariamente, envolvidas com padrões de reconhecimento específico para a espécie (ICONOMIDOU et al, 1999). Não sabemos exatamente qual o papel deste gene e seu padrão de expressão em *Anastrepha* a não ser que apenas uma cópia deste gene foi obtida da biblioteca feminina de *A. obliqua*, contudo, a observação desta segregação de polimorfismos entre as espécies indica um gene promissor para estudos populacionais mais detalhados.

O gene *CG10031* apresenta altos valores de  $\theta_w$  para o grupo *fraterculus* bem como para cada espécie individualmente. Os testes de desvio de neutralidade revelaram desvios negativos significativos apenas nos testes D e F de Fu & Li. Estes testes são mais sensíveis para se detectar seleção de fundo (*background selection*), que ocorre quando sítios ao longo da molécula são eliminados por estarem associados a outros sítios que de fato estão sofrendo seleção purificadora. É possível que esta seleção de fundo seja reflexo da seleção positiva diversificadora encontrada para diversos sítios neste gene no HYPHY. A rede de haplótipos, contudo, embora não seja capaz de revelar linhagens distintas para *A. fraterculus* e *A. sororcula*, separa quase totalmente *A. obliqua* destas duas primeiras. Na verdade, o padrão imbricado encontrado para as duas primeiras espécies



ocorre principalmente com amostras destas espécies do nordeste sugerindo ou uma região de possível hibridização entre *A. fraterculus* e *A. sororcula* ou uma maior inabilidade em discriminar os espécimes desta região. Na verdade, *A. sororcula* do sudeste se agrupa, na maioria dos casos, com *A. obliqua* sugerindo até potencialmente uma diferenciação. Uma amostragem mais detalhada para estas regiões se faz necessária, mas de qualquer forma é um gene bastante interessante, uma vez que esta serpina foi encontrada também na biblioteca de tecidos reprodutivos femininos e é expressa preferencialmente no corpo gorduroso, coração e espermateca em *D. melanogaster*. Dessa forma, é possível que este gene esteja de fato diretamente sujeito à pressão de seleção sexual, o que deve ser confirmado por experimentos de expressão em tempo real.

O padrão encontrado para os genes *CG16713* e *CG11912* foi similar, uma vez que valores de  $\theta_w$  são relativamente altos enquanto os de  $\pi$  não se destacam dos outros encontrados. Da mesma forma, as árvores haplotípicas para estes genes apresentam grande variação intra e interespecífica, porém não se diferenciam bem em linhagens seja por espécie ou por regiões, revelando grande segregação de polimorfismos ancestrais. Há apenas um agrupamento de *A. sororcula* do sudeste para o gene *CG11912*. Enquanto para o primeiro gene não houve nenhum teste de neutralidade que atingisse valores de significância, para *CG11912* houve um desvio significativo de neutralidade para o teste H de Fay e Wu. Como este teste é mais sensível do que os outros aqui realizados para se detectar variações em frequências alélicas intermediárias (FU e LI, 1993), é mais sensível para se investigar potencial seleção positiva, o que já havia sido revelado pelos testes realizados no HYPHY.

Os testes de desvio da neutralidade utilizados neste trabalho utilizam-se de metodologias diferentes para detectar se os genes estão evoluindo neutralmente ou se ocorre desvio da neutralidade. Quando há concordância entre os testes como ocorre com os genes *CG11912* (PAML e H de Fay & Wu), *Lcp65Ac* (D, F de Fu & Li; H de Fay & Wu e D de Tajima) e *Df31* (D, F de Fu & Li e D de Tajima), aumenta a confiança na veracidade dos resultados do que pode estar ocorrendo com as espécies. Neste trabalho encontramos que os resultados significativos obtidos desvios de neutralidade são quase exclusivamente para o grupo *fraterculus* o que pode indicar que a presença de polimorfismo ancestral ainda segregando (ou subestrutura populacional) limita o poder destes testes ou mesmo que apenas na análise conjunta atingimos grau de liberdade suficiente para resgatar tal poder.

A análise populacional da variação nestes genes indica que apesar de diversos haplótipos apresentarem dispersão por várias regiões sem associação quer com regiões geográficas, quer com limites específicos, uma parcela relevante da variação genética apresentou tal associação. Quando



quantificamos a parcela da variação genética total que seria explicada pela separação entre os grupos, que são na verdade a separação das espécies por diferentes regiões geográficas, quase todos os genes apresentaram uma parcela significativa da variação, que variou, quando significativa, entre 35,8% para o *Df31* e 87,4% para o *CG10031*, explicada pela separação entre regiões nas espécies, indicando que há sinal informativo sobre a separação entre as espécies e ainda sobre a variação intraespecífica. Contudo, se considerarmos apenas esta variação geográfica, desconsiderando a qual espécie os indivíduos pertencem, este valor se reduz drasticamente e nenhum dos genes apresentou variação explicada pelas separações geográficas desconsiderando as espécies. Por outro lado, se incluirmos na análise um nível a mais que indicaria a separação entre as espécies, apenas um dos genes, *CG11912*, apresenta uma parcela significativa da variação (embora não muito alta, 28,8%) explicada pela separação das espécies, considerando também a separação entre as regiões. A maior parte da variação continua alocada à separação das regiões neste caso no contexto das espécies, mas isso não é surpreendente já que apenas para *A. fraterculus* há mais do que duas regiões por espécie, confundindo o sinal da separação entre as regiões e entre as espécies. Por este motivo incluimos mais um contraste, considerando apenas o quanto a variação genética é explicada pelas diferenças entre as espécies. Neste caso, todos os genes apresentaram valores significativos, mas que variam de 2% (para *CG16713*) a 31% (para *CG10031*). Isso indica que existe de fato uma parcela da variação genética que é explicada pelas diferenças entre as espécies, mas esse valor não é muito alto, e mais da variação é explicada pelos agrupamentos regionais (em se considerando as espécies). Isso pode refletir uma heterogeneidade intraespecífica ou mesmo que há questões de definições específicas que necessitam reconsideração. Em suma, as análises de AMOVA indicam que há sinal significativo da variação genética explicada pela diferença entre as espécies, e também entre as regiões distintas na mesma, sugerindo que esta variação pode estar trazendo informação relevante sobre o processo de diferenciação deste grupo. Contudo, parte da variação genética não é explicada pela separação entre as espécies, e devemos investigar se isso pode ser devido ao fluxo gênico entre as populações e espécies, ou à segregação de polimorfismos ancestrais.

Para avaliar os níveis de fluxo gênico entre as regiões e espécies consideramos inicialmente um modelo simples de estimativa de *Fst* par-a-par que revelam uma baixa divergência entre populações de *A. obliqua* entre si, mas valores altos para quase todos os outros contrastes. Estes dados indicam que as populações de *A. obliqua* das regiões nordeste e sudeste comportam-se como se fosse uma única grande população apresentando valores de *Fst* entre três (*A. fraterculus*) e quatro (*A. sororcula*) vezes menor do que as divergências intraspecíficas de outras espécies do grupo. Por outro lado, todos os outros contrastes incluindo as três espécies das diversas regiões amostradas



apresentaram valores de *Fst* superiores a 0,1 com exceção do contraste entre populações do nordeste para *A. fraterculus* e *A. sororcula*. O baixo valor de *Fst* (0,07) encontrado para estas populações sugere a existência ou de polimorfismos ancestrais ainda segregando nestas populações, potencial fluxo gênico ou mesmo a existência de espécies crípticas que podem não ter sido corretamente identificadas em nosso estudo. Este padrão é compatível com o que vinha sendo observado nas redes haplotípicas, mas seria importante uma melhor quantificação do fluxo gênico entre estas regiões e espécies, que foi feita por máxima verossimilhança e corroborou em parte estes resultados.

Os resultados da análise de fluxo gênico indicam alguns aspectos interessantes. O primeiro deles é que há de fato, muita inferência de fluxo gênico, tanto entre as espécies, quanto entre as regiões. Na análise em que apenas consideramos a separação por espécies fica claro que *A. obliqua* se apresenta mais isolada do que as outras, mas mesmo assim os valores de fluxo gênico de outras espécies do grupo *fraterculus* para *A. obliqua* gira entre 2 e 4, valores teoricamente mais do que suficientes para levar à panmixia, caso estes genes estivessem evoluindo neutramente. Todos os outros contrastes apresentam valores significativamente maiores do que 4 e indicam um alto potencial de compartilhamento de variação genética. Padrões similares foram encontrados quando consideramos uma análise por grupo, na qual fica evidente o maior isolamento entre *A. obliqua* e as outras espécies aqui estudadas. O grupo 2, *A. obliqua* do sudeste, mostra-se bastante isolado e apresentando fluxo gênico unicamente com *A. sororcula* não recebendo migrantes de nenhuma espécie, exceto de sua própria espécie. O grupo 1, *A. obliqua* do Norte/Centro Oeste, por sua vez não envia migrantes para *A. sororcula*, mas recebe migrantes de *A. sororcula* de ambos os grupos e também recebe migrantes de *A. fraterculus* pelo grupo 3; *A. sororcula* só não envia migrantes para o grupo 3 (*A. fraterculus*), além do já mencionado grupo 2 (*A. obliqua*). *A. sororcula* como espécie receptora não recebe migrantes do grupo 1 (*A. obliqua*) e o grupo 6 não recebe migrantes do grupo 5 (*A. fraterculus*), o grupo 7 não recebe migrantes do grupo 4 (*A. fraterculus*). É importante ressaltar que estas inferências de potencial fluxo gênico são influenciadas pela existência de fluxo gênico ancestral, de forma que estes valores podem estar grandemente inflados caso haja a manutenção de polimorfismos transespecíficos nestas espécies. Uma característica interessante nesta análise ocorre entre *A. fraterculus* e *A. sororcula* do nordeste (grupos 3 e 6 respectivamente), mesmo que nós tenhamos detectado uma região que apresenta relativa coesão para estas espécies, como já mostrado pela rede de haplótipo (Figura 6C) e também mostrado na árvore de espécie (Figura 9), a migração entre estas espécies apresenta padrões desiguais, fazendo com que *A. sororcula* (grupo 6) comporte-se como uma espécie receptora de migrantes e *A. fraterculus* como



doadora de migrantes para este caso do nordeste. E este canal com fluxo de migrantes entre *A. fraterculus* e *A. sororcula* pode ser o principal responsável pelos resultados que apontam um alto relacionamento destas espécies que pode refletir facilidades diferentes de cruzamento intraspecífico ou mesmo de sobrevivência do eventual híbrido na natureza. Não obstante, estes resultados merecem uma investigação mais profunda buscando eliminar potenciais espécies crípticas e erros de identificação buscando confirmar tal padrão.

Após observar os resultados das redes de haplótipos e da variância molecular fica a dúvida de como está a relação filogenética entre estas espécies. A questão mais importante que se impõe é se toda a variação que encontramos revela a existência de troca de material genético entre diferentes linhagens, o que indicaria que a definição das espécies neste grupo mereceria reconsideração ou se estamos basicamente lidando com a segregação de polimorfismos ancestrais e as espécies seriam de fato linhagens evolutivas independentes que não estariam, na maior parte das vezes, trocando material genético. E uma tentativa de inferir esta relação foi através do programa BUCKy que gerou uma árvore de espécie por concordância de ramos com resultados interessantes. O BUCKy utiliza em suas análises da abordagem bayesiana de concordância (BCA) que trata de um método alternativo que irá integrar incertezas sobre a árvore de genes e não irá criar qualquer suposição particular sobre os motivos de discordância. O BCA gera um agrupamento não paramétrico de genes com o compartilhamento de informações cruzadas dos genes. Com isso o BUCKy torna-se um programa para combinar vários loci ortólogos com potencial conflito entre suas árvores filogenéticas. As estimativas de concordância das árvores primárias resumem o sinal filogenético vertical da concordância da maior proporção de loci. (LARGET et al, 2010). Isso não quer dizer que este sinal seja o único existente nos genes investigados, mas será o principal.

Uma característica interessante encontra-se em *A. fraterculus*, que apresenta maior estruturação populacional do que *A. obliqua*, geograficamente mais homogênea. O BUCKy indica que *A. obliqua* e *A. sororcula* formam uma linhagem separada de *A. fraterculus* e *A. sororcula* do nordeste. Este último ramo é interessante e pode ser explicado pela possibilidade da amostragem destas referidas espécies (grupo 3 e 6) ter sido feita em uma região de coexistência e de possível hibridização. Se de fato este padrão é reflexo de hibridização, então a análise do BUCKy é interessante uma vez que mostra que as espécies aqui estudadas formariam linhagens bem definidas com exceção da região em que hibridizam. Por outro lado, caso esta hibridização não se confirme, então teríamos indicação (embora sem grande suporte estatístico, dado os baixos valores de suporte para estes ramos definidos em *A. fraterculus* e *A. sororcula* e baixo suporte no estudo individual de cada gene, de diversas linhagens para estas duas espécies, o que corroboraria estudos anteriores que



afirmam que a espécie *A. fraterculus* não seria monofilética (SILVA e BARR, 2006). No entanto, a análise conjunta dos dados dos genes aqui estudados sugere ser mais provável que esteja havendo algum processo de troca de material genético no nordeste, que é sugerida pelas estimativas de fluxo gênico aqui inferidas.

Os resultados apresentados neste trabalho mostram pontos específicos para estes organismos. *A. obliqua* parece comportar-se como uma espécie de maior coesão e algo diferenciada em relação às outras duas aqui estudadas. A sugestiva coesão de *A. obliqua* contrasta com os resultados de SMITH-CALDAS et al (2001) que sugeriram para esta espécie, a partir de dados de COI, um potencial complexo de espécies. Por sua vez, *A. fraterculus* apresenta maior variabilidade genética, o que já foi indicado por dados de isozimas (STECK, 1991) e moleculares (SMITH-CALDAS, et al, 2001) e apresenta-se como uma espécie possivelmente subestruturada o que pode ser reflexo tanto desta variabilidade populacional quanto da existência de espécies crípticas já sugeridas por outro autores (SELIVON et al, 2002 SELIVON et al, 1997; SELIVON et al, 2005b STECK, 1991). Finalmente, *A. fraterculus* aparece na árvore de espécies filogeneticamente mais próxima à *A. sororcula* a quem é, na verdade, morfologicamente mais similar. No entanto, em virtude da potencial hibridização, ou retenção de polimorfismos ancestrais encontrados para as populações do nordeste para estas espécies, tal proximidade pode ser reflexo destes processos. Por sua vez, *A. sororcula* expõe claramente uma forte coesão no sudeste, que se mostra como uma linhagem coesa, mas o mesmo não pode ser afirmado para *A. sororcula* do nordeste que apresenta um grande compartilhamento de haplótipos com *A. fraterculus* do nordeste, sugerindo que possa estar ocorrendo hibridização. Embora a amostragem aqui realizada tenha sido razoavelmente extensa, não chega perto da amplitude da distribuição das espécies e muito menos consegue representar uma amostra significativa por região. Assim, estudos com maior com maior amostragem poderão fornecer uma melhor evidência das possíveis relações existentes entre estas espécies e quem sabe melhor considerar as diversidades regionais e polimorfismos ancestrais.





## 6. Conclusões

Proteínas relacionadas a tecidos reprodutivos tendem a ter taxas evolutivas mais altas do que proteínas de tecidos não reprodutivos. Neste trabalho encontramos dois dos genes sob seleção positiva e grande variabilidade populacional para todos os genes de tecidos reprodutivos, com exceção do *Df31*, que embora também marginalmente relacionado a tecidos reprodutivos tem função molecularmente mais conservada de chaperona associado a histonas. Outro ponto evidente dos resultados apresentados é a necessidade do uso de vários genes para detecção da história evolutiva, uma vez que genes diferentes podem apresentar histórias diferentes dependendo das taxas evolutivas e dos ambientes vivenciados por estes genes e também, o grau de polimorfismo ancestral que eles apresentam e de sua manutenção na espécie pelas migrações encontradas. A menos que estejamos lidando com os genes diretamente envolvidos com o processo de diferenciação das espécies, os chamados genes de especiação (ORR et al, 2004), os sinais filogenéticos serão em parte “embaralhados” pelos processos estocásticos o que diminui o potencial de inferência a menos que diversos genes sejam considerados.

O uso de cinco genes envolvidos em processos reprodutivos trouxe maiores esclarecimentos para o entendimento das possíveis relações entre estas espécies. Dentre estes, os genes *CG10031* e *CG11912* apresentaram sinais de seleção positiva diversificadora, outros apresentaram sinais mais fracos de seleção direcional (*Df31* e *Lcp65Ac*) e diversificadora (*Df31*, *CG16713*, *CG11912* e *CG10031*). Embora o estudo da variação individual dos genes aqui investigados não indique nenhum gene cuja variação esteja completamente associada às diferenças entre as espécies do grupo *fraterculus*, trouxeram informações relevantes para tal estudo. Alguns dos genes apresentaram sinais informativos tanto para diferenciações entre espécies, ou mesmo para identificação de variação regional.

A análise conjunta destes genes nos permite indicar potenciais linhagens evolutivas que, no geral, são coerentes com a classificação das espécies aqui estudadas. Tal análise sugere coesão de *A. obliqua*, que se apresenta com menor diversidade em relação às outras espécies aqui estudadas. Por sua vez, *A. sororcula* apresenta uma linhagem bem definida nas populações do sudeste, enquanto as populações do nordeste apresentam grande sobreposição com espécimes de *Anastrepha fraterculus* também do nordeste, que pode ser resultado de troca de material genético atual ou muito recente. Se desconsiderarmos as populações do nordeste tanto para *A. fraterculus* como para *A. sororcula*, estas espécies se apresentam como potenciais linhagens evolutivas distintas. De outra forma indicaria que *A. fraterculus*, e mesmo *A. sororcula*, seria um complexo de espécies crípticas, corroborando sugestões de estudos anteriores.



A diversidade encontrada para a maioria destes genes, tanto regionalmente quanto entre populações e até espécies sugere que estes genes são marcadores interessantes para o entendimento das espécies aqui estudadas, bem como para o estudo das relações entre estas e outras do grupo *fraterculus*. No entanto, novos estudos se fazem necessários para um melhor entendimento da dinâmica destas populações, principalmente na região nordeste, a fim de confirmar a potencial hibridização entre *A. fraterculus* e *A. sororcula*. Para tanto novas análises que ampliarão as populações analisadas poderão fornecer melhores indicativos dos processos evolutivos envolvidos na diferenciação deste grupo de espécies.

## 7. Bibliografia

- ALUJA, M. 1994. **Bionomics and Management of *Anastrepha***. *Annual Review of Entomology* **39**: 155-178.
- ALUJA, M; et al. 1999. Behavior of flies in the genus *Anastrepha* (Trypetinae: Toxotrypanini). In: ALUJA, M e NORRBOM, AL **Fruit Flies (Tephritidae): phylogeny and evolution of behavior**. pp. 375- 406.CRC Press. Boca Raton, Florida.
- ARAÚJO, AAR. 2011. **Moscas-das-frutas (Diptera: Tephritidae) e seus parasitóides em frutíferas nativas no estado do Piauí, Brasil**. Dissertação (Mestrado em Agronomia/ Produção Vegetal) - Universidade Federal do Piauí.
- Araujo, EL; ZUCCHI, RA. 2006. **Medidas do acúleo na caracterização de cinco espécies de *Anastrepha* do grupo *fraterculus* (Diptera: Tephritidae)**. *Neotropical Entomology* **35**: 329-337.
- BEERLI, P; PALCZEWSKI, M. 2010. **Unified framework to evaluate panmixia and migration direction among multiple sampling locations**. *Genetics* **185**: 313-326.
- BERLIN, S; SMITH, NGC. 2005. **Testing for adaptive evolution of the female reproductive protein ZPC in mammals, birds and fishes reveals problems with the M7-M8 likelihood ratio test**. *BMC Evolutionary Biology* **5**: 65.-
- BIRKHEAD, TR; MØLLER, AP 1998. **Sperm Competition and Sexual Selection**. 826 pp. Elsevier Science & Technology.
- CARVALHO, RS; NASCIMENTO, AS; FERNANDES, ÊB. 1998. **Dados biológicos de *Anastrepha obliqua* Macquart (Diptera: Tephritidae) em manga**. *Anais Da Sociedade Entomológica Do Brasil* **27**: 469-472.
- CASTRIC, V; VEKEMANS, X. 2007. **Evolution under strong balancing selection: how many codons determine specificity at the female self-incompatibility gene SRK in Brassicaceae?** *BMC Evolutionary Biology* **7**: 132.-
- CESARI, A; KATUNAR, MR; MONCLUS, MA; VINCENTI, A; de ROSAS, JC; FORNÉS, MW. 2005. **Serine protease activity, bovine sperm protease, 66 kDa (BSp66), is present in hamster sperm and is involved in sperm-zona interaction**. *Reproduction (Cambridge, England)* **129**: 291-
- CHARLES , JP; CHIHARA, C; NEJAD, S; RIDDIFORD, LM. 1997. **A cluster of cuticle protein genes of *Drosophila melanogaster* at 65A: sequence, structure and evolution**. *Genetics* **147**: 1213-1224.
- CHUNG, Y; ANÉ, C. 2011. **Comparing two Bayesian methods for gene tree/species tree reconstruction: simulations with incomplete lineage sorting and horizontal gene transfer**. *Systematic Biology* **60**: 261-
- CLARK, NL; SWANSON, WJ. 2005. **Pervasive adaptive evolution in primate seminal proteins**. *PLoS Genetics* **1**: 35-
- Clement, M; POSADA, D; CRANDALL, KA. 2000. **TCS: a computer program to estimate gene genealogies**. *Molecular Ecology* **9**: 1657-
- CONDON, MA; NORRBOM, AL. 1999. Behavior of Flies in the Genus *Blepharoneura* (Blepharoneurinae). In: ALUJA, M e NORRBOM, AL **Fruit Flies (Tephritidae): phylogeny and evolution of behavior**. pp. 375- 406.CRC Press. Boca Raton, Florida.
- CORNMAN, RS. 2009. **Molecular evolution of *Drosophila* cuticular protein genes**. *PloS One* **4**: e8345-
- CORNMAN, RS; TOGAWA, T; DUNN, WA; HE, N; EMMONS, AC; WILLIS, JH. 2008. **Annotation and analysis of a large cuticular protein family with the R&R Consensus in *Anopheles gambiae***. *BMC Genomics* **9**: 22.
- CRANSTON, KA; HURWITZ, B; WARE, D; STEIN, L; WING, RA. 2009. **Species trees from highly incongruent gene trees in rice**. *Systematic Biology* **58**: 489-500.
- CREVEL, G; COTTERILL, S. 1995. **DF 31, a sperm decondensation factor from *Drosophila melanogaster***:

**purification and characterization.** *The EMBO Journal* **14**: 1717.

CREVEL, G; HUIKESHOVEN, H; COTTERILL, S. 2001. **Df31 is a novel nuclear protein involved in chromatin structure in *Drosophila melanogaster*.** *Journal of Cell Science* **114**: 47.

DALLAS-YANG, Q; JIANG, G; SLADEK, FM. 1998. **Avoiding false positives in colony PCR.** *BioTechniques* **24**: 580.

DÍAZ, ORAR. 2010. **Espécies de moscas-das-frutas do gênero *Anastrepha* Schiner, 1868 (DIPTERA:TEPHRITIDAE) coletadas nos estados de Concepción e Misiones, Paraguai.** Dissertação (Mestrado em Programa de Pós-Graduação) - Escola Superior de Agricultura Luiz de Queiroz/USP.

DOS SANTOS, P; URAMOTO, K; MATIOLI, SR. 2001. **Experimental Hybridization Among *Anastrepha* Species (Diptera: Tephritidae): Production and Morphological Characterization of F 1 Hybrids.** *Annals of the Entomological Society of America* **94**: 725.

DUARTE, AL; MALAVASI, A. 1999. **Tratamentos Quarentenários.** In: MALAVASI, A e ZUCCHI, RA. **Moscas das Frutas de importância Econômica no Brasil. Conhecimento Básico e Aplicado.** pp. 233. Holos, Editora Ltda-ME. Ribeirão Preto.

ELLISON, CK; WILEY, C; SHAW, KL. 2011. **The genetics of speciation: genes of small effect underlie sexual isolation in the Hawaiian cricket *Laupala*.** *Journal of Evolutionary Biology* **24**: 1110

EXCOFFIER, L; SMOUSE, PE; QUATTRO, JM. 1992. **Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data.** *Genetics* **131**: 479-

EXCOFFIER, L; LISCHER, HEL. 2010. **Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows.** *Molecular Ecology Resources* **10**: 564.

FACHOLI-BENDASSOLLI, MCN; UCHÔA-FERNANDES, MA. 2006. **Comportamento sexual de *Anastrepha sororcula* ZUCCHI (Diptera, Tephritidae) em laboratório.** *Revista Brasileira De Entomologia* **50**: 412.

FAY, JC; WU, CI. 2000. **Hitchhiking under positive Darwinian selection.** *Genetics* **155**: 1405

FU, YX; LI, WH. 1993. **Statistical tests of neutrality of mutations.** *Genetics* **133**: 693

FUTAHASHI, R; OKAMOTO, S; KAWASAKI, H; ZHONG, Y-S; IWANAGA, M; MITA, K; FUJIWARA, H. 2008. **Genome-wide identification of cuticular protein genes in the silkworm, *Bombyx mori*.** *Insect Biochemistry and Molecular Biology* **38**: 1138

GALINDO, BE; VACQUIER, VD; SWANSON, WJ. 2003. **Positive selection in the egg receptor for abalone sperm lysin.** *Proceedings of the National Academy of Sciences of the United States of America* **100**: 4639

GARRETT, M; FULLAONDO, A; TROXLER, L; MICKLEM, G; GUBB, D. 2009. **Identification and analysis of serpin-family genes by homology and synteny across the 12 sequenced Drosophilid genomes.** *BMC Genomics* **10**: 489.

GONÇALVES, VR; Jr, ISS; Jr, WM; SILVA, FH; BRITO, RAd. 2011. **Transcriptome analysis of female reproductive tissues and molecular evolutionary studies of eggshell proteins in *Anastrepha fraterculus* fruit flies.** *Subm to Insect Biochem Mol. Biol*

GUILLEBAULT, D; COTTERILL, S. 2007. **The *Drosophila* Df31 protein interacts with histone H3 tails and promotes chromatin bridging in vitro.** *Journal of Molecular Biology* **373**: 903-

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

HASHIMOTO, C; KIM, DR; WEISS, LA; MILLER, JW; MORISATO, D. 2003. **Spatial Regulation of Developmental Signaling by a Serpin.** *Developmental Cell* **5**: 945

- ICONOMIDOU, VA; WILLIS, JH; HAMODRAKAS, SJ. 1999. **Is beta-pleated sheet the molecular conformation which dictates formation of helicoidal cuticle?** *Insect Biochemistry and Molecular Biology* **29**: 285.
- INOUE, A; OGUSHI, S; SAITOU, M; SUZUKI, MG; AOKI, F. 2011. **Involvement of mouse nucleoplasmin 2 in the decondensation of sperm chromatin after fertilization.** *Biology of Reproduction* **85**: 70.
- KELLEHER, ES; CLARK, NL; MARKOW, TA. 2011. **Diversity-enhancing selection acts on a female reproductive protease family in four subspecies of *Drosophila mojavensis*.** *Genetics* **187**:
- KIMURA, M. 1970. **The length of time required for a selectively neutral mutant to reach fixation through random frequency drift in a finite population.** *Genetical Research* **15**: 131-
- KINGMAN, J.F.C. 1982 **The coalescent. Stochastic Process.** *Appl.* **13**: 235-248.
- KOCHER, TD. 2004. **Adaptive evolution and explosive speciation: the cichlid fish model.** *Nature Reviews. Genetics* **5**: 288.
- KORNEYEV, VA. 1999a. Phylogenetic Relationship among the families of the Superfamily Tephritoidea. In: ALUJA, M e NORRBOM, AL **Fruit Fies (Tephritidae): phylogeny and evolution of behavior.** pp. 375- 406.CRC Press. Boca Raton, Florida.
- KORNEYEV, VA. 1999b. Phylogenetic Relationships among Higher Groups of Tephritidea. In: ALUJA, M e NORRBOM, AL. **Fruit Fies (Tephritidae): phylogeny and evolution of behavior.** pp. 375- 406.CRC Press. Boca Raton, Florida.
- KOSAKOVSKY-POND, SL; et al. 2012. Evolution of viral genomes - Interplay between selection, recombination and other forces. In: Anisimova, Maria **Evolutionary Genomics** pp.560. Statistical and Computational Methods, Volume 2 Series: Methods in Molecular Biology, Vol. 856
- KOSAKOVSKY-POND, SL; Frost, SDW; Muse, SV. 2005. **HyPhy: hypothesis testing using phylogenies.** *Bioinformatics (Oxford, England)* **21**: 676
- KOSAKOVSKY-POND, SL; Murrell, B; Fourment, M; Frost, SDW; Delport, W; Scheffler, K. 2011. **A random effects branch-site model for detecting episodic diversifying selection.** *Molecular Biology and Evolution* **28**: 3033
- KOSAKOVSKY-POND, SL; Poon, AFY; Frost, SDW. 2007. **Evolution of viral genomes - Interplay between selection, recombination and other forces.** *Evolutionary Genomics: Statistical and Computational Methods* **44**.
- KRYAZHIMSKIY, S; PLOTKIN, JB. 2008. **The population genetics of dN/dS.** *PLoS Genetics* **4**: e1000304
- KULATHINAL, RJ; Skwarek, L; Morton, RA; Singh, RS. 2003. **Rapid evolution of the sex-determining gene, transformer: structural diversity and rate heterogeneity among sibling species of *Drosophila*.** *Molecular Biology and Evolution* **20**: 441
- LARGET, BR; Kotha, SK; Dewey, CN; ANÉ, C. 2010. **BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis.** *Bioinformatics (Oxford, England)* **26**: 2910
- LIBRADO, P; ROZAS, J. 2009. **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.** *Bioinformatics (Oxford, England)* **25**: 3023
- LIGOXYGAKIS, P; Roth, S; REICHHART, J-M. 2003. **A serpin regulates dorsal-ventral axis formation in the *Drosophila* embryo.** *Current Biology : CB* **13**: 145
- Lis, JT; SCHLEIF, R. 1975. **Size fractionation of double-stranded DNA by precipitation with polyethylene glycol.** *Nucleic Acids Research* **2**: 2097
- MA, B; LI, M; ZHANG, L. 2000. **From Gene Trees to Species Trees.** *SIAM Journal on Computing* **30**: 729-752.

- MACHOLÁN, M; BAIRD, SJE; DUFKOVÁ, P; MUNCLINGER, P; BÍMOVÁ, BV; PIÁLEK, J. 2011. **Assessing multilocus introgression patterns: a case study on the mouse X chromosome in central Europe.** *Evolution; International Journal of Organic Evolution* **65**: 196
- MADDISON, WP. 1997. **GENE TREES IN SPECIES TREES.** *Syst. Biol* **46**: 523-536.
- MALAVASI, A; MORGANTE, JS. 1982. **Genetic variation in natural population of *Anastrepha* (Diptera: Tephritidae).** *Rev. Bras. Genet.* **5**: 263-278.
- MATIOLI, SR; MORGANTE, JS; MALAVASI, A. 1986. **Genetical and biochemical comparisons of alcohol dehydrogenase isozymes from *Anastrepha fraterculus* and *A. obliqua* (Diptera: Tephritidae): Evidence for gene duplication.** *Biochemical Genetics* **24**: 13-24.
- MING, Q-L; LEWIS, SM. 2010. **Mate Recognition and Sex Differences in Cuticular Hydrocarbons of the Diurnal Firefly *Ellychnia corrusca* (Coleoptera: Lampyridae).** *Annals of the Entomological Society of America* **103**: 128-133.
- MORGANTE, JS; MALAVASI, A; BUSH, GL. 1980. **Biochemical Systematics and Evolutionary Relationships of Neotropical *Anastrepha*.** *Annals of the Entomological Society of America* **73**: 622-630.
- MORGANTE, JS; MALAVASI, A; PROKOPY, RJ. 1983. **Mating Behavior of Wild *Anastrepha fraterculus* (Diptera: Tephritidae) on a Caged Host Tree.** *The Florida Entomologist* **66**: 234
- NEITZEL, JJ. 2010. **Enzyme Catalysis: The Serine Proteases.** *Nature Education* **3**: 21
- NELSON, JE; KRAWETZ, SA. 1992. **Purification of cloned and genomic DNA by guanidine thiocyanate/isobutyl alcohol fractionation.** *Analytical Biochemistry* **207**: 197
- NORRBOM, Allen L.; KORYTKOWSKI, CHESLAVO A. 2011 **New species of and taxonomic notes on *Anastrepha* (Diptera: Tephritidae).** *Zootaxa* **2740**: 1-23.
- NORRBOM, AL; ZUCCHI, RA; HERNÁNDEZ-ORTIZ, V. 1999. **Phylogeny of the Genera *Anastrepha* and *Toxotrypana* (Trypetinae: Toxotrypanini) Based on Morphology.** In: ALUJA, M e NORRBOM, AL. **Fruit Fies (Tephritidae): phylogeny and evolution of behavior.** pp. 375- 406. CRC Press. Boca Raton, Florida.
- NYDAM, ML; HARRISON, RG. 2011. **Reproductive protein evolution in two cryptic species of marine chordate.** *BMC Evolutionary Biology* **11**: 18
- OLIVER, PL; GOODSTADT, L; BAYES, JJ; BIRTLE, Z; ROACH, KC; PHADNIS, N; BEATSON, SA; LUNTER, G; MALIK, HS; PONTING, CP. 2009. **Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa.** *PLoS Genetics* **5**: e1000753.
- RASMUSSEN, MD; KELLIS, M. 2007. **Accurate gene-tree reconstruction by learning gene- and species-specific substitution rates across multiple complete genomes.** *Genome Research* **17**: 1932-
- REICHHART, J-M. 2005. **Tip of another iceberg: *Drosophila* serpins.** *Trends in Cell Biology* **15**: 659-
- SCHILL, WB. 1975. **Significance of proteases of the male genital tract for reproduction.** *Die Medizinische Welt* **26**: 1450
- SEGURA, DF; VERA, MT; RULL, J; WORNOPYORN, V; ISLAM, A; ROBINSON, AS. 2011. **Assortative mating among *Anastrepha fraterculus* (Diptera: Tephritidae) hybrids as a possible route to radiation of the *fraterculus* cryptic species complex.** *Biological Journal of the Linnean Society* **102**: 346-
- SELIVON, D; MORGANTE, JS; PERONDINI, ALP. 1997. **Egg size, yolk mass extrusion and hatching behavior in two cryptic species of *Anastrepha fraterculus* (Wiedemann) (Diptera, Tephritidae).** *Brazilian Journal of Genetics* **20**:

- SELIVON, D; PERONDINI, ALP; MORGANTE, JS. 2005a. **A Genetic–Morphological Characterization of Two Cryptic Species of the *Anastrepha fraterculus* Complex (Diptera: Tephritidae)**. *Annals of the Entomological Society of America* **98**: 367-
- SELIVON, D; PERONDINI, ALP; MORGANTE, JS. 1999. **Haldane's rule and other aspects of reproductive isolation observed in the *Anastrepha fraterculus* complex (Diptera: Tephritidae)**. *Genetics and Molecular Biology* **22**:
- SELIVON, D; PERONDINI, ALP; ROCHA, LS. 2005b. **Karyotype characterization of *Anastrepha* fruit flies (Diptera: Tephritidae)**. *Neotropical Entomology* **34**: 273-
- SELIVON, D. 1999. Biologia e Padrões de especiação. In: MALAVASI, A e ZUCCHI, RA. **Moscas das Frutas de importância Econômica no Brasil. Conhecimento Básico e Aplicado**. pp. 233. Holos, Editora Ltda-ME. Ribeirão Preto.
- SELIVON, D; PERONDINI, ALP. 1999. Morfologia dos Ovos de Moscas-das-Frutas do Gênero *Anastrepha*. In: MALAVASI, A e ZUCCHI, RA. **Moscas das Frutas de importância Econômica no Brasil. Conhecimento Básico e Aplicado**. pp. 233. Holos, Editora Ltda-ME. Ribeirão Preto.
- SELIVON, DENISE., VRETOS, C., FONTES, L., e PERONDINI, A. L. P. 2002. **New variant forms in the *Anastrepha fraterculus* complex (Diptera: Tephritidae)**. Proceedings of 6th International Fruit Fly Symposium V6. p. 253
- SILVA, JANISETE G. and BARR, NORMAN B. 2006. Recent Advances in Molecular Systematics of *Anastrepha* Schiner Salvador, In: **Proceedings of the 7th International Symposium on Fruit Flies of Economic Importance**. p. 13-28. Salvador, Brazil. SBPC, Press Color
- SILVERMAN, GA; BIRD, PI; CARRELL, RW; CHURCH, FC; COUGHLIN, PB; GETTINS, PG; IRVING, JA; LOMAS, DA; LUKE, CJ; MOYER, RW; PEMBERTON, PA; REMOLD-O'DONNELL, E; SALVESEN, GS; TRAVIS, J; WHISSTOCK, JC. 2001. **The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature**. *The Journal of Biological Chemistry* **276**:
- SIVINSKI, J; ALUJA, M; DODSON, G; FREIDBERG, A; HEADRICK, D; KANESHIRO, K; LANDOLT, P. 1999. Topics in the Evolution of Sexual Behavior in the Tephritidae. In: ALUJA, M e NORRBOM, AL. **Fruit Flies (Tephritidae): phylogeny and evolution of behavior**. pp. 375- 406. CRC Press. Boca Raton, Florida.
- SIVINSKI, JM; DODSON, G. 1992. **Sexual dimorphism in *Anastrepha suspensa* (Loew) and other tephritid fruit flies (Diptera: Tephritidae): Possible roles of developmental rate, fecundity, and dispersal**. *Journal of Insect Behavior* **5**: 491.
- SMITH-CALDAS, MRB; MCPHERON, BA; SILVA, JG; ZUCCHI, RA. 2001. **Phylogenetic Relationships Among Species of the *fraterculus* Group (*Anastrepha*: Diptera: Tephritidae) Inferred from DNA Sequences of Mitochondrial Cytochrome Oxidase I**. *Neotropical Entomology* **30**:
- SOBRINHO, IdSJ. 2009. **Evolução molecular dos genes doublesex e fruitless em moscas-das-frutas do grupo *Anastrepha fraterculus* (Diptera, Tephritidae)**. Tese (Doutorado em Genética e Evolução) - Universidade Federal de São Carlos.
- SOBRINHO, IS; DE BRITO, RA. 2010. **Evidence for positive selection in the gene fruitless in *Anastrepha* fruit flies**. *BMC Evolutionary Biology* **10**: 293.
- SOLFERINI, VN; MORGANTE., JS. 1987. **Karyotype study of eight species of *Anastrepha* (Diptera: Tephritidae)**. *Caryologia* **40**: 229-241.
- STECK, GJ. 1991. **Biochemical Systematics and Population Genetic Structure of *Anastrepha fraterculus* and Related Species (Diptera: Tephritidae)**. *Annals of the Entomological Society of America* **84**: 10-28.
- SWANSON, WJ; VACQUIER, VD. 2002. **The rapid evolution of reproductive proteins**. *Nature Reviews. Genetics* **3**: 137

- SWANSON, WJ; WONG, A; WOLFNER, MF; AQUADRO, CF. 2004. **Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection.** *Genetics* **168**: 1457
- SZOLLOSI, E; BOKOR, M; BODOR, A; PERCZEL, A; KLEMENT, E; MEDZIHRADESKY, KF; TOMPA, K; TOMPA, P. 2008. **Intrinsic structural disorder of DF31, a *Drosophila* protein of chromatin decondensation and remodeling activities.** *Journal of Proteome Research* **7**: :2291
- TAJIMA, F. 1989a. **Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism.** *Genetics* **123**: 585-595.
- TAJIMA, F. 1989b. **The Effect of Change in Population Size on DNA Polymorphism.** *Genetics* **123**: 597-601.
- TEMPLETON, AR 2006. *Population Genetics and Microevolutionary Theory*. 720 pp. John Wiley & Sons, Inc. Hoboken, NJ, USA.
- THAN, C; NAKHLEH, L. 2009. **Species tree inference by minimizing deep coalescences.** *PLoS Computational Biology* **5**: e1000501-
- THOMPSON, JD; HIGGINS, DG; GIBSON, TJ. 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-
- XUE, F; SETO, CT. 2005. **Selective inhibitors of the serine protease plasmin: probing the S3 and S3' subsites using a combinatorial library.** *Journal of Medicinal Chemistry* **48**: 6908-
- YANG, Z. 2007. **PAML 4: phylogenetic analysis by maximum likelihood.** *Molecular Biology and Evolution* **24**: 1586-
- ZENG, K; FU, Y-X; SHI, S; WU, C-I. 2006. **Statistical tests for detecting positive selection by utilizing high-frequency variants.** *Genetics* **174**: 1431-
- ZOLET, ACT. 2009. **Filogeografia e sistemática molecular de *Schizolobium parahyba* (Vell.) Blake (Guapuruvu) através do sequenciamento de regiões cloroplásticas e nucleares.** Tese (Doutorado em Genética e Biologia Molecular) - Universidade Federal do Rio Grande do Sul.
- ZUCCHI, RA. 1978. **Taxonomia das espécies de *Anastrepha* Schiner, 1868 (Diptera, Tephritidae) assinaladas no Brasil.** Dissertation, Universidade de São Paulo, Piracicaba. vi + 105 p. [monograph of 77 spp. in Brazil]
- ZUCCHI, R.A. (2000). Taxonomia. In: MALAVASI, A e ZUCCHI, RA. **Moscas das Frutas de importância Econômica no Brasil. Conhecimento Básico e Aplicado.** pp. 13- 28. Holos, Editora Ltda-ME. Ribeirão Preto.
- ZUCCHI, R.A. 2008. **Fruit flies in Brazil - *Anastrepha* species their host plants and parasitoids.** Available in: [www.lea.esalq.usp.br/Anastrepha/](http://www.lea.esalq.usp.br/Anastrepha/), updated on August 10, 2011. Accessed on 10/11/2011



8. Anexos

**ANEXO 1:**

**Alinhamento do gene *Cg16713***

Alinhamento do gene amplificado com o *Primer CG16713* que confirma sua identidade com o representante de *Drosophila melanogaster*.

**CG16713 - *D. melanogaster*.**

	5' UTR	Éxon 1	Íntron 1
	0 1 2 3 4 0	0 6 7 8 9 1	1 2 3 4 1 6 7 8 9 2
	0 0 0 0 0 5	0 0 0 0 0 0	0 0 0 0 0 5 0 0 0 0 0
contig_1397	agttgcaaaaaattttattcaacaaagatttttaggtaattcttttaaaaaatgaattcttcgctgtaattttggctgtttc	M K F F A V I L A V F - A L I G C S F - A L K D	-----gcgcttatcgctgctcgttc-----gcgttaaggatg-----
Aobliq	-----atgaaattcttcgctgtaattttggctgtttc-----gcgcttatcgctgctcattc-----gccttaaggatgg-----	M K F F A V I L A V F - A L I G C S F - A L K D	-----taagtttgagagaactaaagactaagtgcccactgtaatttatt-cttttcata-----tt
Afrat	-----atgaaattcttcgctgtaattttggctgtttc-----gcgcttatcgctgctcattc-----gccttaaggatgg-----	M K F F A V I L A V F - A L I G C S F - A L K D	-----taagtttgagagaactaaagactaagtgcccactgtaatttatt-attttcata-----tt
Asororc	-----atgaaattcttcgctgtaattttggctgtttc-----gcgcttatcgctgctcattc-----gcgttaaggatgg-----	M K F F A V I L A V F - A L I G C S F - A L K D	-----taagtttgagagaactaaagactaagtgcccctgtaatttatt-attttcata-----tt
Dmelano	-----atgaaactgttg-----attttggttcttcgctgttttggcccaacgcttg-----gcctgaaaaatgg-----	M K I L - I L V F V L V V S I S S C F - A L K N	-----ttggttc-----ctaactgttca-taactcaaatggtttgta-atctg-aaa-----tt
Dmojav	-----atgaagattctg-----gtgctgttcgtagtactcgttctcaatagcagctgttt-----gcgctcaaaaaatggtgaaactcattt	M K F F - V L L V L A L F V S C L A A P K N	-----taagtttgagagaactaaagactaagtgcccactgtaatttatt-attttcata-----tt
Dgrimsh	-----atgaagtttttc-----gtgtgctgtgtgctcgtcgtcccttttgcagctgccttgggcccccaaaaatgg-----	M K F L - A I F L M L A A F V A T S W - A L K H	-----tggtgtgtgattttttggaat-gccttacatagaattatttataaactcgata-----tggttc
Dwillist	-----atgaagtctctt-----gctatatttttaagtgtggtgccttttggcaactctcgtg-----gcctgaaacatcg-----	M K F L - A I F L M L A A F V A T S W - A L K H	-----tatgtggcaattatgtgat-gttttacctctttttcttac-----attctggcactaacgtt-----

**Íntro**

**Éxon 2**

	2 0 1 2 3 4 2 5 6 7 8 9 0 3 1 2 3 4 3 5 6 7 8 9 0 4
	1 0
contig_1397	-----cgttttgcggtcaagagcattcgccaatggcattggcattccagctgtgcagcttttggcctcctggacctcgtatgccagcgccaataagtgctggagttcgtttatggcgctgtgtggcaatgacaatcgtttcgatcagaacaggcttgcaggcaaaaatcggtggaataaaaggata
Aobliq	-----tttcgcatagccgttttgggtcaagagcattcgccaatggcattggcattccagctgtgcagcttttggcctcctggacctcgtatgccagcgccaataagtgctggagttcgtttatggcgctgtgtggcaatgacaatcgtttcgatcagaacaggcttgcaggcaaaaatcggtggaataaaaggata
Afrat	-----tttcgcaatagccgttttgcggtcaagagcattcgccaatggcattggcattccagctgtgcagcttttggcctcctggacctcgtatgccagcgccaataagtgctggagttcgtttatggcgctgtgtggcaatgacaatcgtttcgatcagaacaggcttgcaggcaaaaatcggtggaataaaaggata
Asororc	-----tttcgcaatagccgttttgcggtcaagagcattcgccaatggcattggcattccagctgtgcagcttttggcctcctggacctcgtatgccagcgccaataagtgctggagttcgtttatggcgctgtgtggcaatgacaatcgtttcgatcagaacaggcttgcaggcaaaaatcggtggaataaaaggata
Dmelano	-----tttc-----cagcaactctgtggtctaccctaccctaaacggagatgagcagaatactctggaggcctataaccagctgttctcctagcagcgatcgaaacagtagctgcgcaaatctatctacggaggctgcggaggcaatacaatagatttaattcgagggaatctgtgagacaagtgtttgcaataa
Dmojav	-----tttgcattagaacctttgcggtctgccaactctgcagatggtgctggggccaactctgacgctcactttccaagctggagatatactgtggcgctaatgaatgctgtaactcaatctatgagcgctgtggaggcaatgacaatcgtttcgacaccaaaagaaactctgtgaggaagaagtcttggaaataa
Dgrimsh	-----tttgatcaagtgagtgagctgagagcattcctcaatggcagatgagcgtatttctgtgagcatatatacccaactgacctctgcactctgcagccaagaaatgctccagtttatctatgagcgctgtggagggaatgccaatcgtttgaaaccaagaagtctgtgagagagatcctcaattga
Dwillist	-----taacgagcactgtggtctctctcaactcaactgaacggcagtgagcatttcttgcagagcttatatacccaactggtcctctcgtat-----ggtcaaggatgctggaattcttatggcgatgaggtggcaatgccaatcgttttggcactaaagagcagtgaggccaagtgttg-----taa-----

**3' UTR**

	4 0 1 2 3 4 5 6 7 8 9 0 5 1 5
	1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
contig_1397	ttttgtgtgcccgaatgcatatgaaactattttgcactttacatattgagtgatgtgattctctatatggatataaaataaattgaagagctctgctgtgatttcaaaataaaa



### Alinhamento do gene *Df31*

Alinhamento do gene amplificado com o *Primer DF31* confirmando sua identidade com o gene **CG2207** de *Drosophila melanogaster*. **CG2207 (Df31) - D. melanogaster**,

#### 5' UTR

```

0      1      2      3      4      5      6      7      8      9      0      1      2      3      4      5      6      7      8      9      0
1      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0
contig_573 gttgcttctacggttcggcagctgttcttcttaattttgttcgctcaagtaagtcattctttcataaaaaaagttggttaattttgagtttttgcatagcagatttaattagcataactcatttctcaactcctgtgttttaattcataaaacgtaaaaaaacgtagttcatttgagcttaattcattcct
  
```

5' UTR	Éxon 1	Intron 1	Éxon 2
--------	--------	----------	--------

```

0      1      2      3      4      5      6      7      8      9      0      1      2      3      4      5      6      7      8      9      0
1      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0
contig_573 gctcgtagagaaaagaactactgacaaaaatggccgatgtt---gaaaagaaa---... 1.5kb ... E T P I V E K V A A E E V E A Q K E K E V A - E E V A P - - - - T A E N G S A D S A S
Aobliq -----gaaactccaattgctgagaagaatcgctgctgaagaagtagaagcacaaaagaaaaagagttgca---gaggagtagcacca-----accgccgaaaaatgtagtctgacagtgatcc
Afrat -----gaaactccaattgctgagaagaatcgctgctgaagaagtagaagcacaaaagaaaaagagttgca---gaggagtagcacca-----gcccgcgaaaaatgtagtctgacagtgatcc
Asororc -----gaaactccaattgctgagaagaatcgctgctgaagaagtagaagcacaaaagaaaaagagttgca---gaggagtagcacccg-----gcccgcgaaaaatgtagtctgacagtgatcc
Dmojav -----M A D V A E K K N -----gaaactccaattgctgagaagaatcgctgctgaagaagtagaagcacaaaagaaaaagagttgca---gaggagtagcacca-----gcccgcgaaaaatgtagtctgacagtgatcc
Dmelano -----atggctgatggtgctgagaaaaagaa... ? ...cgagacccccgtggtgagaagaattgcccgtgagaaggtgtagctgtgaaagaagtagctgctgcccagcagtgagcagcagagacaccgaccaccacagagaacggtgctgcccagagtagcagc
Dyakuba -----M A D V A D Q K N -----gaaactccaattgctgagaagaatcgctgctgaagaagtagaagcacaaaagaaaaagagttgca---gaggagtagcacccg-----gcccgcgaaaaatgtagtctgacagtgatcc
  
```

#### Éxon 2

```

4      5      6      7      8      9      0      1      2      3      4      5      6      7      8      9      0
0      1      2      3      4      5      6      7      8      9      0      1      2      3      4      5      6      7      8      9      0
1      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0
contig_573 - - - - - A K E N G D - - - - - A A E V A A - E S T T E - - - - T P - - - - - K E A - S A E A E A A A P A E E - N G K D D E
Aobliq c-----gccaagaagaaatggtgac-----gctgccaagtagccgct---gaaactacaactgaa---acgcc-----aaagaagca---tctgctgaggtgaaagctgcccctcctgccgaagaa---aatggcaagtagcagaag
Afrat c-----gccaagaagaaatggtgac-----gctgccaagtagccgct---gaaactacaactgaa---acgcc-----aaagaagca---tctgctgaggtgaaagctgcccctcctgccgaagaa---aatggcaagtagcagaag
Asororc c-----gccaagaagaaatggtgac-----gctgccaagtagccgct---gaaactacaactgaa---acgcc-----aaagaagca---tctgctgaggtgaaagctgcccctcctgccgaagaa---aatggcaagtagcagaag
Dmojav c-----gccaagaagaaatggtgac-----gctgccaagtagccgct---gaaactacaactgaa---acgcc-----aaagaagca---tctgctgaggtgaaagctgcccctcctgccgaagaa---aatggcaagtagcagaag
Dmelano c-----gtagccaagtagcagaagc-----gccaagtagcagaagc-----gccaagtagcagaagc-----gccaagtagcagaagc-----gccaagtagcagaagc-----gccaagtagcagaagc
Dyakuba c-----gtaaccaagtagcagaagc-----gccaagtagcagaagc-----gccaagtagcagaagc-----gccaagtagcagaagc-----gccaagtagcagaagc
  
```

Éxon 2	Intron 1	Éxon 3
--------	----------	--------

```

6      7      8      9      0      1      2      3      4      5      6      7      8      9      0
0      1      2      3      4      5      6      7      8      9      0      1      2      3      4      5      6      7      8      9      0
1      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0
contig_573 A - - - - E S - - - E T T ----- D S A P A E A V K R K V A E D E
Aobliq cg-----gaaagc-----gaaacaactg-----attcagcaccgcagaggtctaaagagaaaaatgtagtgaagcagaat
Afrat cg-----gaaagc-----gaaacaactg-----gt-----aagt-aaaacttgat-----actatatactactt-aaaatgttctggaactaat-----tgaaatcaatcca-----ttcagactcagcaccgcagaggtctaaagagaaaaatgtagtgaagcagaat
Asororc cg-----gaaagc-----gaaacaactg-----gt-----aagt-aaaacttgat-----actatatacaactc-aaaatgttctggaactaat-----tgaaatcaatccta-----ttcagactcagcaccgcagaggtctaaagagaaaaatgtagtgaagcagaat
Dmojav A V E N G A D D - - A E A A N G D S K -----cggttgagaatggcgcgatgat-----gcagagcagctaacggtgactcaaaaag---gtaaatgtgaa---agaataaacataacctgaaat---gacatttaattttg---tattttcttttctatttttgggtccccatttttag---atgccctccctgagagctgaaagcgaaggtgtagtgaagcagc
Dmelano A V L P A V E N G S E V T N G D S T -----aagt-aaaacttgat-----actatatactactt-aaaatgttctggaactaat-----tgaaatcaatcca-----ttcagactcagcaccgcagaggtctaaagagaaaaatgtagtgaagcagc
Dyakuba cagtttctactctgctgtagaagtggtcctgctgagggcactaacggtgactcaaacag---gtatg---ta-----gatatacctgctgagaat---ccagcttacggtcagttctctattatcc-tttttactctggccctacag---acgctcccgccattgaggtctaaagcgaaggtgtagtgaagcagc
  
```



Alinhamento do gene *Df31 cont.*

éxon 3

3' UTR

	8				8				9					9				9																				
	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	9	9																			
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																			
	S	K	P	E	E	S	T	A	A	T	P	E	K	K	A	K	L	D	E	P	V	-	K	E	D	T	Q	N	G	S	E	A	S	E	V	A	A	*
contig_573	caaaccagaagaatccactgcagcacaacacctgagaaaaagccaagctagacgaaccagtt---aaagaagacacccaaaatggtccgaagcatcagaggttgcggcttaacttagatctaatacttaactattcaaatatcagtaacaaaaaa																																					
	S	K	P	E	E	S	T	A	A	T	P	E	K	K	A	K	L	D	E	P	V	-	K	E	D	T	Q	N	G	S	E	A	S	E	V	-	-	
Aobliq	aaaaccagaagaatccactgcagcaacacctgagaaaaagccaagctagacgaaccagtt---aaagaagacacccaaaatggtccgaagcatcagaggtt-----																																					
	S	K	P	E	E	S	T	A	A	T	P	E	K	K	A	K	L	D	E	P	V	-	K	E	D	T	Q	N	G	S	E	A	S	E	V	-	-	
Afrat	aaaaccagaagaatccactgcagcaacacctgagaaaaagccaagctagacgaaccagtt---aaagaagacacccaaaatggtccgaagcatcagaggtt-----																																					
	S	K	P	E	E	S	T	A	A	T	P	E	K	K	A	K	L	D	E	P	V	-	K	E	D	T	Q	N	G	S	E	A	S	E	V	-	-	
Asororc	aaaaccagaagaatccactgcagcaacacctgagaaaaagccaagctagacgaaccagtt---aaagaagacacccaaaatggtccgaagcatcagaggtt-----																																					
	A	K	A	D	E	V	V	A	-	T	P	E	K	K	A	K	L	D	E	A	S	S	K	D	E	V	Q	N	G	A	E	A	S	E	V	A	A	*
Dmojav	ctaaagccgatgaggtcggtgcc---acgccggagaaagggcgaagctggacgagccagctcaaaggatgaagttcaaaatggtgccgaagccagcgaggtggctgcttaa-----																																					
	A	K	A	D	E	A	V	A	---	T	P	E	K	K	A	K	L	D	E	A	S	T	K	D	E	V	Q	N	G	A	E	A	S	E	V	A	A	*
Dmelano	ccaagccagatgagccgctgcc---acgccggagaaagggcgaagctggacgagccagcacaaggatgaggttcagaatggggccgaggtagcgaagtggccgcctaa-----																																					
	A	K	A	D	E	S	V	A	-	T	P	E	K	K	A	K	L	D	E	S	S	T	K	D	E	V	Q	N	G	A	E	A	S	E	V	A	A	*
Dyakuba	ccaagccgacgagctctgctgcc---acgccggagaaagggcgaagctggacgagccagcacaaggatgaggtgcagaacggggccgaggtagcgaagtggccgcctaa-----																																					

**Alinhamento do gene *Lpc65Ac***  
 Alinhamento do gene amplificado com o *Primer LCP65Ac* confirmando sua identidade com o gene **CG6956 (*Lcp65Ac*)** de *Drosophila melanogaster*. **CG6956 - *D. melanogaster*,**

<b>5' UTR</b>																			
0											1						1		2
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

contig\_1127 tccaacctaaccaatcaaccaaaccaatcaaccgccatcgccagcagccatccaacctccaactactaaaaacaaacaaaaccctgattccacgtagaaaacctctccctcgaaacgaacaccccgtattgaagcgcggtgcttggctgtgtgctatagaaaagaaggaaggtgccatttcactaa

5' UTR	Éxon 1	Íntron 1
2	2	3
0	1	2
1	0	0
M K C - - A I L F V V V C I A S A M A A P A P - Q Q E V Q V L R Y D S D V E P E G Y K F L		
contig_1127	gaagattacgcaatctttaaagatgaagtgt----gctatatttattggtggttgcctcgcctccgcatggctgcgcagcaact---caacaagaagtacaggttttacctgtacgattcagcgtggagcctgaggatataaaattct	
Aobliq	M K C - - A I L F V V V C I A S A M A A P A P - Q Q E V Q V L R Y D S D V E P E G Y K F L	
Afrater	-----tctttaaagatgaagtgt----gctatatttattggtggttgcctcgcctccgcatggctgcgcagcaact---caacaagaagtacaggttttgcgtacgattcagcgtggagcctgaggatataaaatttctaagtgaaataaaatgtacatacgaagtgtgtatgaag	
Asororc	M K C - - A I L F V V V C I A S A M A A P A P - Q Q E V Q V L R Y D S D V E P E G Y K F L	
Dmelano	-----tctttaaagatgaagtgt----gctatatttattggtggttgcctcgcctccgcatggctgcgcagcaact---caacaagaagtacaggttttgcgtacgattcagcgtggagcctgaggatataaaatttctaagtgaaataaaatgtacatacgaagtgtgtatgaag	
Dwillis	M K C T V A I V F T A L - F A V V L A A P A P - D A D T Q I L R L E S D V Q P E G Y N F A	
Dvirili	-----atgaagtgcacagtgccatcgctccaccgcctc---ttgcgctcgttctgctgcccggcctccc---gatcgcgatccagcagctcgcgctccagcgtggagcctcacaacttgcgtatg	

-----atgaagtgcacagtgccatcgctccaccgcctc---ttgcgctcgttctgctgcccggcctccc---gatcgcgatccagcagctcgcgctccagcgtggagcctcacaacttgcgtatg  
 M K Y T I A I V F A C L - L A V V L A A P A - - D N T A E I L R L E S D V Q P E G Y K F A  
 -----atgaagtgcacagtgccatcgctccaccgcctc---ttgcgctcgttctgctgcccggcctccc---gatcgcgatccagcagctcgcgctccagcgtggagcctcacaacttgcgtatg  
 M K Y T I A I V L V A F - L G A A V L A A P L D D S K T A Q I L R L D S D V Q P N G Y K F G  
 -----atgaagtgcacagtgcttattggtgctcattc---ctgcccgcctcgtgcccgcagcctcgatgacagcagcagcagcagcctcgctcgcctcgatcgctccagcccaacgcgctacaagtgcggtaagc-----

	Íntron 1	Éxon 1
	0	6
4	5	0
0	1	2
1	0	0
contig_1127	gtgtgaactcgcgtgtatgcgaattgccaggtttcgatggctccattcacttgagagtggtcacgtcacaggtcaagatgcaa---taatgatttgcattaaacactaaacagtcataacatttttgcctcttcttctgctcatatttcttcgctctctgatgctccgcataacctacataccgaatagcaggt	
Aobliq	gtgtgaactcgcgtgtatgcgaattgccaggtttcgatggctccattcacttgagagtggtcacgtcacaggtcaagatgcaa---taatgatttgcattaaacactaaacagtcataacatttttgcctcttcttctgctcatatttcttcgctctctgatgctccgcataacctacataccgaatagcaggt	
Afrater	-----gtatcaaacctccgctcagcctcggtgcgacgattccaagggatataaa-----taccacttctccacagtt	
Asororc	-----tggattact---gtttact--gccaacattcaaaaacgcttattgaaataatagct-----tctgattgattgacatagctt	
Dmelano	-----gtatcaaacctccgctcagcctcggtgcgacgattccaagggatataaa-----taccacttctccacagtt	
Dwillis	-----tggattact---gtttact--gccaacattcaaaaacgcttattgaaataatagct-----tctgattgattgacatagctt	
Dvirili	-----taaca--gcacagcgcctgagagagagagagagagacgcgcaa---taactg-----agcgtatctcctcagcagttg	

<b>Éxon 2</b>																																
6																									7							8
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0		
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
contig_1127	E T S D G K L H E E E G K L K D V G T E Q E A I V V R G S F S Y V G D D G Q T Y T V N Y I A D E N G F Q P E G A H L P R V - - - - *																															
Aobliq	ggagactagcgatggcaactgcatgaagaggaaggcaaatgaaggaagcctggacagcaagagggcgatgtatgctgctggtcattctcctatgctggcagatggtcaaacctaacagctcaactatagccgatgagaatggtttccaacggagggcgcacattgccacggtt-----taag																															
Afrater	E T S D G K L H E E E G K L K D V G T E Q E A I V V R G S F S Y V G D D G Q T Y T V N Y I A D E N G F Q P E G A H L P R V - - - - *																															
Asororc	ggagactagcgatggcaactgcatgaagaggaaggcaaatgaaggaagcctggacagcaagagggcgatgtatgctgctggtcattctcctatgctggcagatggtcaaacctaacagctcaactatagccgatgagaatggtttccaacggagggcgcacattgccacggtt-----taag																															
Dmelano	E T S D G K L H E E E G K L K D V G T E Q E A I V V R G S F S Y V G D D G Q T Y T V N Y I A D E N G F Q P E G A H L P R V - - - - *																															
Dwillis	ggagactagcgatggcaactgcatgaagaggaaggcaaatgaaggaagcctggacagcaagagggcgatgtatgctgctggtcattctcctatgctggcagatggtcaaacctaacagctcaactatagccgatgagaatggtttccaacggagggcgcacattgccacggtt-----taag																															
Dvirili	E T S D G Q K H D E E G S L T N P G A E N E S I A V R G S Y S F T A D D G Q V Y T V N Y V A D E N G F Q P E G A H L P N V P I G N *																															
	ggaaactagcagcggcaagaccatcaggaaggaagtcgaattgaaggaacatcggtaccggaagcgaattgctcctgctcactactccttcttgcgcatgatggaacagcctcaactatgctgagagaagcgttccaactcaggggtgctcaattgcccaattgcccaactaa																															
	E T S D G Q K H D E E G S L T N P G A E N E S I A V R G S Y S F T A D D G Q V Y T V N Y V A D E N G F Q P E G A H L P N V P I G N *																															
	ggagaccagcgatggaacagaacgcagcaggaagggcgactgaccatcctggtgctgagaagaatcgatcgtctgcgcgctcctactccttaccgccaacgcagtcgccaagtgtatccgctgaactatgctggcagatgagaacggctccaacggagggcgccactgccaagtgtgccattgccaactaa																															



Alinhamento do gene *Lpc65Ac cont.*

3' UTR

	8	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
contig_1127	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
Aobliq	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Afrater	-----																			
Asororc	-----																			
Dmelano	-----																			
Dwillis	-----																			
Dvirili	-----																			



Dwillis ctccaactcgaatgcatgccccctgatggtgagtcaccctgacgacctcctgctcccaacaattctttcgcccaacagtgatggtgctctttgctgggggacgcgacaacgctggcaagtggccgcacaactcgaaactcgagaaactggatgctgggacattattgggtacaatgattgctgcagctgcatgccaag  
 L D L T A V A R D G S L F V A A I A L P T L Q L Q A I D D G V L F G W G R D N S G A L A K K L Q T L N V T V L D F P T C K A E L P S  
 cctagatctgagcgttctgctgatggtgagtcaccctgacgacctcctgctcccaacaattctttcgcccaacagtgatggtgctctttgctgggggacgcgacaacgctggcaagtggccgcacaactcgaaactcgagaaactggatgctgggacattattgggtacaatgattgctgcagctgcatgccaag  
 F D I N S L A R N - A P V A T I A V A S K G V A T T S N G V L V S W S R D N S G S L P D K L Q K L D V D I I D Y K T C L D N I P K  
 Dmojav attcagatcaatagcttgcccgcaac----gcgctgctgccaccactcgtggtgctccaaagcagtagccacgaccggcaatggtgctcctcagcgtgggtgcgcacaactcgtgctgcccgcacaactcgagaaactggatgctggatcattgactacaagcgtgctggataacctgctcaag

### Alinhamento do gene *CG11912 cont.*

#### Éxon 1

	6					6						7							7						8	
	0	1	2	3	4	5	6	7	8	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	F	A	P	I	K	D	V	N	I	C	S	Y	N	A	G	T	D	G	A	C	N	G	D	S	G	G
contig_418	t	c	g	c	c	c	c	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
Aobliq	F	A	P	I	K	D	V	N	I	C	S	Y	N	A	G	T	D	G	A	C	N	G	D	S	G	G
Afrater	t	c	g	c	c	c	c	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
Asororc	F	A	P	I	K	D	V	N	I	C	S	Y	N	A	G	T	D	G	A	C	N	G	D	S	G	G
Rhagol	t	c	g	c	c	c	c	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
dmelano	N	N	S	L	A	E	T	N	V	C	T	H	T	P	G	K	A	D	G	S	C	N	G	D	S	G
Dvirili	a	c	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
Dgrimsh	G	N	K	L	D	Q	T	N	I	C	T	Y	T	A	G	T	D	G	A	C	N	G	D	S	G	G
Dwillis	T	S	L	E	C	N	I	C	T	H	N	A	G	T	A	D	G	A	C	N	G	D	S	G	G	L
Dmojav	g	c	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	

#### Éxon 1

#### 5' UTR

	8																								
	0	1	2	3	4	5	6	7	8	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
contig_418	K	W	I	T	E	K	I	R	-	N	*														
Aobliq	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
Afrater	K	W	I	T	E	K	I	R	-	N	*														
Asororc	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
Rhagol	K	W	I	T	E	K	I	R	-	N	*														
dmelano	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
Dvirili	P	W	I	D	E	N	R	K	A	K															
Dgrimsh	c	c	t	t	g	g	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	
Dwillis	D	W	I	N	E	H	S	N	D	-	*														
Dmojav	g	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	

### Alinhamento do gene *CG10031*

Alinhamento do gene amplificado com o *Primer CG10031* confirmando sua identidade com o gene *CG10031* de *Drosophila melanogaster*.

5' UTR					Éxon 1																								
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25				
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
contig_1062	M K C C - - I L F V F G I I A L L - - - L C S T V Q A V D K S A - - - - -																												
Aobliq	M K C - - I L F V F G I I A L L - - - L C T T V Q A V D K S G - - - - -																												
Afrater	M K C - - I L F V F S I I A L L - - - L C T T V Q A V D K S G - - - - -																												
Asororc	M K C - - I L F V F S I I A L L - - - L C T T V Q A V D K S G - - - - -																												
Dmelano	M L L R - - M T Q F L C L A A L V L L S L M Q L T Q A V A T P N N N N N N N N N I K P K A V V Q V Q P G N P S G																												
Dmojav	M K R V L G M T Q F L C L A A I V L L G L T Q L A E A V T A D - - - - -																												
Dvirili	M K R T T R M T Q F L W L A A I V L L C L T Q M T Q A V T T P D - - - - -																												
Dgrimsh	M T Q F C I V V I L L I C L T Q F T G A E T T P D - - - - -																												

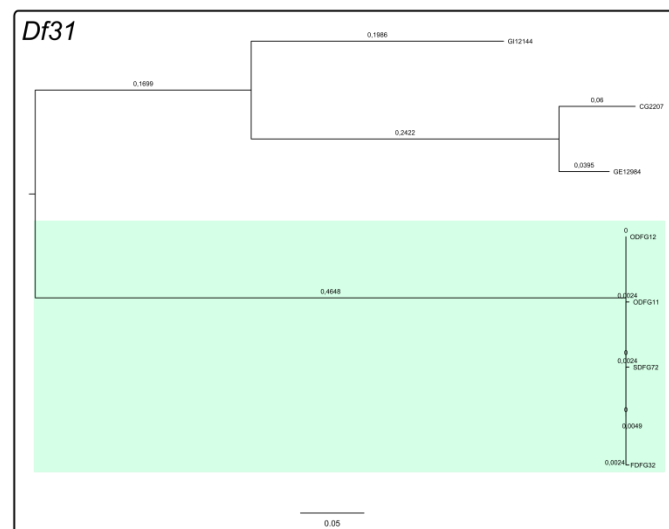
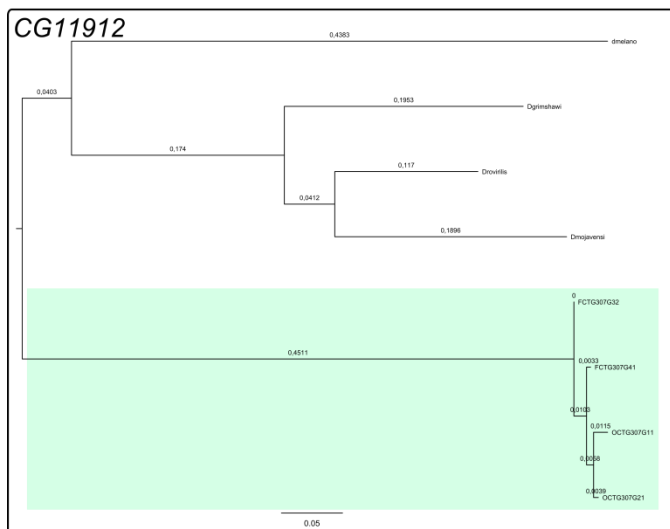
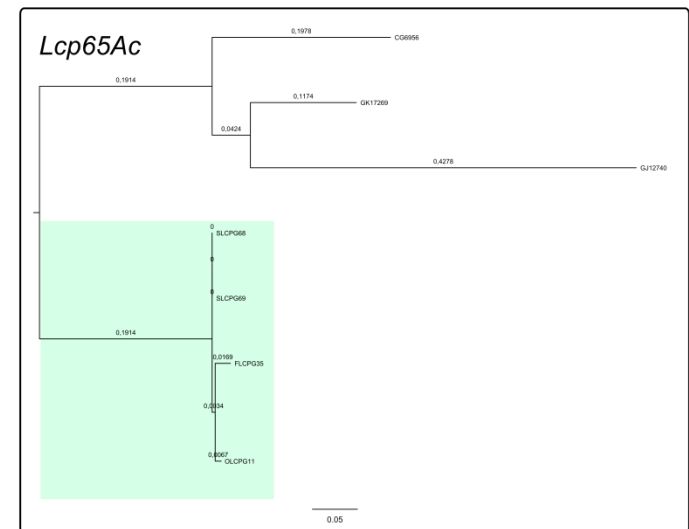
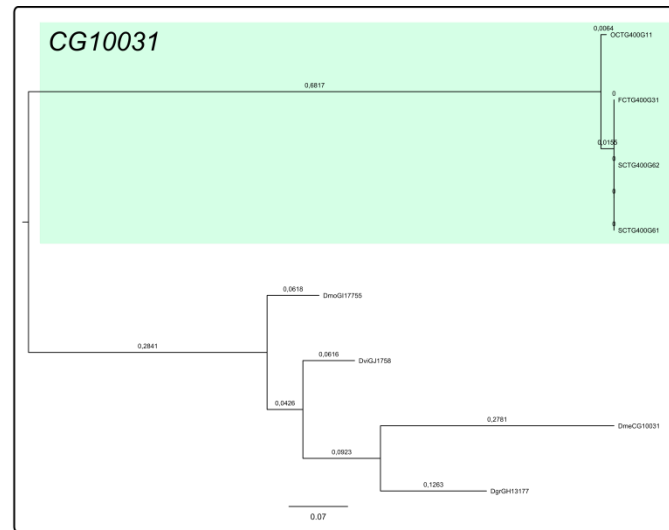
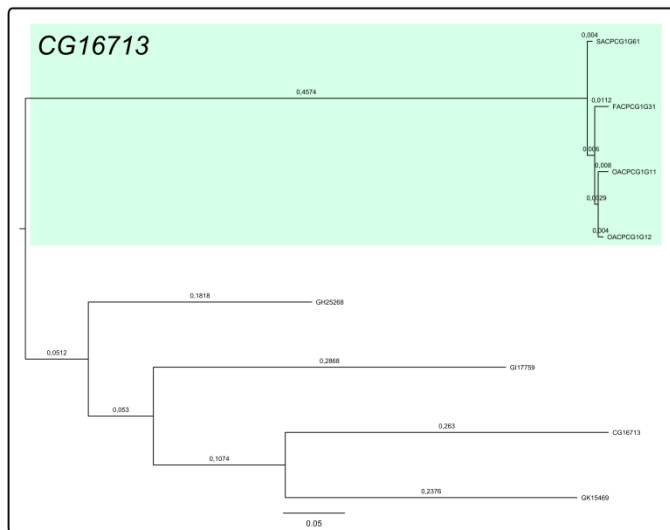
Éxon 1					Íntron 1															Éxon 2					
2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
contig_1062	aagcgcaacaggcgtgaagagtctaattaagcct----aactta-actagcaaaaatcaagttttattctgctttctaattcattgtctttcactttttctctgatcagc-----tgctcacagccacgcgaaatcggtccctgc-----																								
Aobliq	aagcgcaacaggcgtgaagagtctaattaagcct----aactta-actagcaaaaatcaagttttattctgctttctaattcattgtctttcactttttctctgatcagc-----tgctcacagccacgcgaaatcggtccctgc-----																								
Afrater	aagcgcaacaggcgtgaagagtctaattaagcct----aactta-actagcaaaaatcaagttttattctgctttctaattcattgtctttcactttttctctgatcagc-----tgctcacagccacgcgaaatcggtccctgc-----																								
Asororc	aagcgcaacaggcgtgaagagtctaattaagcct----aactta-actagcaaaaatcaagttttattctgctttctaattcattgtctttcactttttctctgatcagc-----tgctcacagccacgcgaaatcggtccctgc-----																								
Dmelano	N K P A V A T T T I K P Q R L V P E D F L Y G G C K G N D N R F D T K E E C E K T C L - - - * acaagcctgcagttgccacaactacaatcaaacacagagactgtaccaggtatgtttattataaactgatcgaagaactttatggacctttatataatcccttaatttga-----gatcccaatgctgcagccttggaactgttgcctat-----																								
Dmojav	R K A D P K F F Y N S E K K A C E D F L Y G G C K G N D N R F D T K E E C E K T C L - - - * ccgcaagccgaccgaaattcttcaacaatagcgagaaggcttgcgaggacttctctatggaggtgcaaggcgaatgataatcgcttcgataccaagaggcaatcgagaagacttgcct-----																								
Dvirili	R K A D P K F F Y N S E K K A C E D F L Y G G C K G N D N R F D T K E E C E K T C L - - - * ccgcaagccgaccgaaattcttcaacaatagcgagaaggcttgcgaggacttctctatggaggtgcaaggcgaatgataatcgcttcgataccaagaggcaatcgagaagacttgcct-----																								
Dgrimsh	R M S L E R F Y Y N K D S K A C E T F K Y G G C R G N D N R W G F R O T C E E A C I P K K * ccgatgagcctgagaaggttctactacaacaaggattcaaaagcttgcgagacattcaagtagcggagattgcgaggcaatgacaactgttggggattcgcgcaaaatgtgaggagcttcttatcccaaaaagtga-----																								

Éxon 2																									3' UTR				
4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25				
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
contig_1062	R K A D P K F F Y N S E K K A C E D F L Y G G C K G N D N R F D T K E E C E K T C L - - - * cgcaagccgaccgaaattcttcaacaatagcgagaaggcttgcgaggacttctctatggaggtgcaaggcgaatgataatcgcttcgataccaagaggcaatcgagaagacttgcct-----																												
Aobliq	R K A D P K F F Y N S E K K A C E D F L Y G G C K G N D N R F D T K E E C E K T C L - - - * ccgcaagccgaccgaaattcttcaacaatagcgagaaggcttgcgaggacttctctatggaggtgcaaggcgaatgataatcgcttcgataccaagaggcaatcgagaagacttgcct-----																												
Afrater	R K A D P K F F Y N S E K K A C E D F L Y G G C K G N D N R F D T K E E C E K T C L - - - * ccgcaagccgaccgaaattcttcaacaatagcgagaaggcttgcgaggacttctctatggaggtgcaaggcgaatgataatcgcttcgataccaagaggcaatcgagaagacttgcct-----																												
Asororc	R M S L E R F Y Y N K D S K A C E T F K Y G G C R G N D N R W G F R O T C E E A C I P K K * ccgatgagcctgagaaggttctactacaacaaggattcaaaagcttgcgagacattcaagtagcggagattgcgaggcaatgacaactgttggggattcgcgcaaaatgtgaggagcttcttatcccaaaaagtga-----																												
Dmelano	R M S L E R Y Y Y N K E K N S C E T F K Y G G C R G N D N R W G F R O T C E E A C L - - - tcgatgagcctggagcgtattactacaataagcgagaaggcttgcgagacattcaagtagcggagattgcgaggcaatgacaactgttggggattcgcgcaaaatgtgaggagcttcttatcccaaaaagtga-----																												
Dmojav	R M S L E R Y Y Y N K E A N A C Q T F K Y G G C K G N D N R W G F R O T C E E A C L - - - tcgaaatgagcctggagcgtattactacaataagcgagaaggcttgcgagacattcaaataggcggctgtaaggcgaatgacaactgttggggattcgcgcaaaatgtgaggagcttcttatcccaaaaagtga-----																												
Dvirili	R M K L E R Y Y Y N K D K N A C E T F N F G G C R G N D N R W G F R O T C E D A C L - - - tcgatgaaactggagcgtattactacaacaaggataagaatgcgtgtgagacattcaatttggcggctgcgaggcaatgacaactgttggggattcgcgcaaaatgtgaggagcttcttatcccaaaaagtga-----																												

3' UTR																									
6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
contig_1062	taattgtgttttggtaaatatgttgaactataattatataatattgttgatttaattgttctcaatttatgttgggtggtgoatactacgttgcgagcggctgtattccaogtgatacggaaataaaagtaagaataatgaaag																								







**Anexo 2.** Árvores obtidas pelo programa PhyML v. 3.0. por máxima verossimilhança. Estas árvores foram utilizadas no CODEML (PAML) para os genes *CG16713*, *CG10031*, *CG11912*, *Df31* e *Lcp65Ac*. Em Verde destaca-se o ramo *foreground* das *Anastrephas*, e o restante da árvore considera-se como o *background*.