

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

**Análise multilocus de parâmetros populacionais, evolução molecular e  
diferenciação em espécies de moscas-das-frutas do grupo *fraterculus*  
(Diptera, Tephritidae).**

Fernanda Fernandes

São Carlos  
2010

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

**Análise multilocus de parâmetros populacionais, evolução molecular e  
diferenciação em espécies de moscas-das-frutas do grupo *fraterculus*  
(Diptera, Tephritidae).**

**Fernanda Fernandes**

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Evolução 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**

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

F363am

Fernandes, Fernanda.

Análise multilocus de parâmetros populacionais, evolução molecular e diferenciação em espécies de moscas-das-frutas do grupo *fraterculus* (Diptera, Tephritidae) / Fernanda Fernandes. -- São Carlos : UFSCar, 2010.  
99 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2010.

1. Genética e evolução. 2. Coalescência. 3. Genética de populações. 4. Biblioteca de cDNA. 5. Anastrepha. I. Título.

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

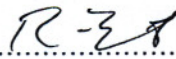
UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE  
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E EVOLUÇÃO

**ANÁLISE MULTILOCUS DE PARÂMETROS  
POPULACIONAIS, EVOLUÇÃO MOLECULAR E  
DIFERENCIAÇÃO EM ESPÉCIES DE MOSCAS-DAS-  
FRUTAS DO GRUPO *FRATERCULUS* (DIPTERA,  
TEPHRITIDAE).**

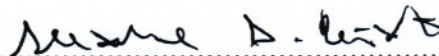
Dissertação de Mestrado de Fernanda Fernandes

Banca Examinadora

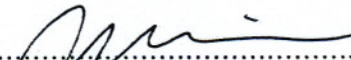
Prof. Dr. Reinaldo Alves de Brito

  
.....

Prof. Dr. Alexandre Afrânio Peixoto

  
.....

Prof. Dr. João Miguel de Barros Alexandrino

  
.....

"A ignorância gera confiança com mais frequência do que o conhecimento: são aqueles que sabem pouco, e não aqueles que sabem muito, que tão positivamente afirmam que esse ou aquele problema jamais será resolvido pela ciência."

Charles Darwin

## *Agradecimentos*

### **Agradeço...**

- À Deus, por tudo de bom que já aconteceu na minha vida e por me cercar de pessoas maravilhosas e de momentos inesquecíveis.
- Ao meu orientador Reinaldo pela dedicação e paciência durante a realização deste trabalho. E também por acreditar em mim e no meu potencial.
- Aos meus pais, pela educação, amor, e, sobretudo pelo apoio incondicional em todas as horas. E à minha irmã Flávia pela amizade e pela força todos os dias. Amo vocês!
- Aos meus avós, tios e primos pela torcida e preces para que eu obtivesse mais esta conquista na minha vida.
- À FAPESP, CNPq e CAPES pelo apoio financeiro que foi essencial para que esta pesquisa pudesse ser realizada.
- Aos meus amigos que sempre me deram força e sempre torceram para que tudo desse certo na minha vida!
- À Cí Camila, Cí Oi, Camillets, Carol Mizuno, Fabrício, Toshi, Bruninho pela amizade, carinho, cumplicidade e por terem- me aguentado durante todos esses anos, que não vou falar quantos, porque senão vou me sentir muito velha. Com certeza dividiremos ainda muitos momentos maravilhosos e inesquecíveis juntos.
- À Lelê, por ter paciência comigo, mesmo nos momentos de estresse, durante todos estes anos, principalmente pelos que fomos *roomates*, pela amizade, desabafos e sessões de filme e seriados SEMPRE!
- À Andréa pelo incentivo e torcida durante toda minha pesquisa.
- Ao Ider, por sempre me ajudar nas horas de desespero ensinando tudo o que ele sabia, pela paciência, pela tranquilidade e principalmente pela amizade, nunca me esquecerei disso!
- Ao Daniel, que me ajudou muito nas análises finais e mostrou ser um grande amigo. Muito Obrigada!
- A TODOS meus grandes amigos de laboratório Van Mieko, Nelci, Van Gonçalves, Marquinhos, Dé, Aline, Victor, Isabela, Bruno e ao Gú ( que este ano ficou mais em Rio Preto que aqui em Sanca) pela ajuda nos momentos difíceis, pelo apoio e pelas risadas e que com certeza ainda teremos muitas. Considero vocês todos da minha família!  
Muito Obrigada! É muito bom saber que eu tenho amigos como vocês!

## Resumo

O gênero *Anastrepha*, endêmico da região Neotropical, tem grande importância econômica por causar grandes prejuízos na produção de frutos. Neste trabalho, estudamos três espécies do grupo críptico *fraterculus*: *Anastrepha fraterculus*, *A. obliqua* e *A. sororcula*. Apesar de não haver um único marcador capaz de diferenciar essas espécies, alguns marcadores morfológicos e comportamentais, são capazes de distingui-las. Há evidências de que esses marcadores não correspondam aos padrões de variação e divergência genética dessas moscas. Investigamos o papel de diversos genes nucleares na especiação dessas moscas. Usamos uma biblioteca de cDNAs do aparelho reprodutivo de fêmeas de *A. fraterculus*, para a amplificação de genes expressos desse tecido. A partir de um quadro de divergência genética populacional, estimamos parâmetros demográficos relevantes que informaram a arquitetura genética da especiação neste grupo. Para estudar a separação dessas espécies, utilizamos árvores de genes como base e as comparamos com árvores de espécies. Foram amplificados 12 genes expressos na biblioteca de cDNAs. As análises de divergência populacional e nível de polimorfismo foram feitas no DNAsp e mostraram valores bem altos do nível de polimorfismo para a grande maioria dos loci e os testes de neutralidade foram significativos apenas para  $F_s$  de  $F_u$ . Esses resultados indicam uma possível expansão populacional das espécies estudadas. Estes dados foram corroborados pelos valores de  $q$  das análises realizadas pelo programa IMA2, que também mostrou uma taxa de migração alta entre a espécie mais distante na árvore para uma das outras duas. No entanto, as mesmas análises feitas com as espécies par a par não revelaram uma taxa de migração significativa entre as espécies, os valores de  $t$  e  $\theta$  foram próximos dos obtidos para as análises com as espécies juntas. Fizemos também testes para detectar a presença de seleção nos diferentes loci, usando o programa PAML e obtivemos indício de seleção positiva para três genes *CG7203*, *MLC* e *CG8064*; o gene *CG7009*, no entanto, mostrou evoluir por uma seleção purificadora branda, pois não rejeitamos somente o primeiro dos dois testes. As redes haplotípicas, feitas no TCS, mostraram o agrupamento dos haplótipos das espécies para cinco dos 12 genes *CG7009*, *CG8064*, *Elp*, *TCTP* e *Cyclophylin*. Para o *CG7009* notamos possíveis marcadores específicos para *A. sororcula*. Realizamos, pelo programa MESQUITE, análises de coalescência profunda comparando árvores de genes simuladas com árvores de genes reais para cada locus, colocando uma das espécies como grupo externo em cada modelo de árvore e

obtivemos resultados significativos apenas para os genes *Cyclophylin* e *TCTP*. A árvore obtida pela minimização das coalescências profundas mostrou que *A. sororcula* se encontra mais distante das outras duas espécies do gênero *Anastrepha*. Faz-se necessária, no entanto, uma investigação maior de diferentes genes, particularmente genes com taxas evolutivas mais altas e uma expansão das amostras destas espécies para considerar sua distribuição geográfica e auxiliar nos estudos da especiação deste grupo.

Palavras chave: Coalescência. Seleção positiva. Biblioteca de cDNA. *Anastrepha*



## *Abstract*

The genus *Anastrepha* which is endemic to the Neotropical region has big economic importance to cause great losses in fruit production. We studied three species of the cryptic group *fraterculus*: *Anastrepha fraterculus*, *A. obliqua* and *A. sororcula*. Although no single marker is able to distinguish these species, some morphological and behavioral markers are able to distinguish them. There is evidence that these markers do not match the patterns of genetic variation and divergence of these flies. We investigated the role of genes related or not to reproduction in the speciation of these flies. We used a cDNA library of the reproductive tissues of females of *A. fraterculus*, to isolate expressed genes from those tissues. From a framework of divergence population genetics, we estimated relevant demographic parameters that reported the genetic architecture of speciation in this group. To study the separation of these species, we used gene trees as a framework to be compared and to infer eventual species trees. We amplified 12 genes isolated from the cDNA library. The analysis of population divergence and levels of polymorphism were made in DNAsp and showed high levels of polymorphism for the vast majority of loci. Neutrality tests reveal that only  $F_s$  Fu was significantly negative for the combination of species here considered. These results indicate a possible evidence of population expansion on this group of species. These data were corroborated by the values of ancestral and current  $\theta$  estimated from the analysis undertaken on the software IMA2 program. These analysis also showed relevant levels of migration in some contrasts, but particularly among species which are more distant in the tree from the other two. However, the same analyses performed amongst species pairs fail to reveal high levels of migration rate among species, though  $t e \theta$  values were close to those estimated in the analysis with all three species together. We also tested for the presence of selection at different *loci*, using the program PAML and obtained evidence of positive selection for three genes *CG7203*, *CG8064* and *MLC*, whereas the gene *CG7009* showed evolution by a mild purifying selection. The haplotype networks showed the grouping of haplotypes of the species for five of the 12 genes *CG7009*, *CG8064*, *Elp*, *TCTP* and *Cyclophylin*. *CG7009* revealed possible species specific markers for *A. sororcula*. We carried out the deep coalescence analysis in the program MESQUITE, comparing gene trees simulated with real trees for each gene locus, placing the species as an outgroup for each tree model and obtained significant results, which would be compatible to simulated scenarios only for the genes

*Cyclophylin* and *TCTP*. The tree obtained by minimizing the deep coalescence estimated showed that *A. sororcula* is farther from the other two species of the genus *Anastrepha*. It is necessary, however, a larger investigation of different genes, particularly genes with higher evolutionary rates and an expansion of the samples of these species to consider geographical distribution and assist in studies of speciation of this group.

Keywords: Coalescence. cDNA library .Positive Selection.*Anastrepha*

## *Lista de Figuras*

- Figura 1**– Esquema de *Anastrepha* do grupo *fraterculus* com as regiões que auxiliam na identificação das espécies. **A**- Tórax; **B**- asa; **C**- ápice do acúleo (Zucchi 2007). *A. fraterculus* possui mancha no subescutelo, faixa V geralmente completa, ápice do acúleo de 0,25 – 0,27 mm; *A. obliqua* não possui mancha no subescutelo, faixa V geralmente incompleta, ápice do acúleo de 0,20 mm; *A. sororcula* possui mancha no subescutelo, faixa V geralmente incompleta, ápice do acúleo de 0,17 – 0,19 mm.....2
- Figura 2** - Ciclo de vida das moscas-das-frutas.....3
- Figura 3** Representação dos clones amplificados com o *primer* M13 do gene PCR 148 .....22
- Figura 4**- Eletroferograma mostrando polimorfismo no gene CG7203, causando também uma mudança de aminoácido de Valina para Isoleucina.....23
- Figura 5** – Sequenciamento do gene **Cyclophylin** exemplificando a presença de códon de terminação nas sequências de *A. sororcula* (**abaixo**) e *indel* nas sequências de *A. fraterculus* (**acima**).....28
- Figura 6** - Rede de haplótipos da região **CG7009** indicando as relações filogenéticas entre as 20 amostras de *A. fraterculus* (*frat*), 20 amostras de *A. obliqua* (*obl*) e 20 de *A. sororcula* (*sor*). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.21. Os haplótipos de *A. sororcula* estão conectados ao resto da árvore embora esteja acima dos limites da parcimônia, que neste caso é de nove passos mutacionais.....34
- Figura 7** - Rede de haplótipos da região **CG7203** indicando as relações filogenéticas entre as 20 amostras de *A. fraterculus* (*frat*), 20 amostras de *A. obliqua* (*obl*) e 20 de *A. sororcula* (*sor*). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.21.....35
- Figura 8**- Rede de haplótipos da região **CG8064** indicando as relações filogenéticas entre as 20 amostras de *A. fraterculus* (*frat*), 20 amostras de *A. obliqua* (*obl*) e 20 de *A. sororcula* (*sor*). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.21.....36
- Figura 9** - Rede de haplótipos da região **CG5325** indicando as relações filogenéticas entre as 20 amostras de *A. fraterculus* (*frat*), 15 amostras de *A. obliqua* (*obl*) e nove de *A. sororcula* (*sor*). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.....37
- Figura 10** - Rede de haplótipos da região **Elongation Protein** indicando as relações filogenéticas entre as 17 amostras de *A. fraterculus* (*frat*), 19 amostras de *A. obliqua* (*obl*) e 15 de *A. sororcula* (*sor*). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.....38
- Figura 11**- Rede de haplótipos da região **MLC** indicando as relações filogenéticas entre as 20 amostras de *A. fraterculus* (*frat*), 20 amostras de *A. obliqua* (*obl*) e 20 de *A. sororcula* (*sor*). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.....39

**Figura 12** - Rede de haplótipos da região **Porin** indicando as relações filogenéticas entre as 20 amostras de *A. fraterculus* (frat), 19 amostras de *A. obliqua* (obl) e 20 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.....40

**Figura 13**- Rede de haplótipos da região **TCTP** indicando as relações filogenéticas entre as 19 amostras de *A. fraterculus* (frat), 19 amostras de *A.obliqua* (obl) e 20 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.....40

**Figura 14**- Rede de haplótipos da região **Troponin C** indicando as relações filogenéticas entre as 20 amostras de *A. fraterculus* (frat), 20 amostras de *A. obliqua* (obl) e 20 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.....41

**Figura 15**- Rede de haplótipos da região **UQCRX** indicando as relações filogenéticas entre as amostras 18 de *A. fraterculus* (frat), 13 amostras de *A. obliqua* (obl) e 19 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.....42

**Figura 16**-Rede de haplótipos da região **Wings Up** indicando as relações filogenéticas entre as 13 amostras de *A. fraterculus* (frat), 16 amostras de *A. obliqua* (obl) e quatro de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.....42

**Figura 17**-Rede de haplótipos da região **Cyclophilyn** indicando as relações filogenéticas entre as 18 amostras de *A. fraterculus* (frat), dez amostras de *A. obliqua* (obl) e 21 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2. . Os haplótipos de *A. sororcula*, *A. fraterculus* e *A. obliqua* estão conectados ao resto da árvore embora esteja acima dos limites da parcimônia, que neste caso é de mais de dez passos mutacionais .....43

**Figura 18**- Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **K80**, mostrando as relações de parentesco entre moscas do grupo e grupos externos *Drosophila willistoni* (XP 002072331), *D. melanogaster* (NP 650947), *D. grimshawi* (XP 001995863), *D. mojavensis* (XP 001999852) e *D. virilis* (XP 002054908) para um fragmento do gene **CG7009**. Os dados das espécies de *Drosophila* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A.obliqua*); sor (*A.sororcula*). Teste de seleção se mostrou significativo para o *Relaxed Branch-site Test*.....44

**Figura 19**- Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **HKY85**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Rhagoletis pomonella* (EZ 127868), *Drosophila willistoni* (XP 002066731), *D. melanogaster* (NP 609143), *D. grimshawi* (XP 001993182), *D. mojavensis* (XP 002001922) e *D. virilis* (XP 002059236) para um fragmento do gene **CG7203**. Os dados das espécies de *Drosophila* e de *R. pomonella* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Teste de seleção se mostrou significativo para o *Strict Branch-site Test*.....45

**Figura 20-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **K80**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Drosophila willistoni* (XP 002072312), *D. melanogaster* (NP 650705), *D. grimshawi* (XP 001996274), *D. mojavensis* (XP 001999395) e *D. virilis* (XP 002054614) para um fragmento do gene **CG8064**. Os dados das espécies de *Drosophila* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Teste de seleção se mostrou significativo para o *Relaxed Branch-site Test* e para o *Strict Branch-site Test*.....45

**Figura 21-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **HKY85**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Drosophila willistoni* (XP 002069226), *D. melanogaster* (NP 723722), *D. grimshawi* (XP 001989171), *D. mojavensis* (XP 002001840) e *D. virilis* (XP 002052621) para um fragmento do gene **CG5325**. Os dados das espécies de *Drosophila* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Não foi significativo para nenhum contraste.....46

**Figura 22-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **K80**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Drosophila willistoni* (XP 002066814), *D. melanogaster* (NP 608834), *D. grimshawi* (XP 001989312), *D. mojavensis* (XP 002003521) e *D. virilis* (XP 002052755) para um fragmento do gene **Elongation Protein**. Os dados das espécies de *Drosophila* e de *R. pomonella* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Não foi significativo para nenhum contraste.....46

**Figura 23-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **TN93**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Rhagoletis pomonella* (EZ 137400), *Drosophila willistoni* (XP 002071798), *D. melanogaster* (NP 511049), *D. grimshawi* (XP 001992735), *D. mojavensis* (XP 002010859) e *D. virilis* (XP 002057174) para um fragmento do gene **MLC**. Os dados das espécies de *Drosophila* e de *R. pomonella* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Teste de seleção se mostrou significativo para o *Relaxed Branch-site Test* e para o *Strict Branch-site Test*.....47

**Figura 24-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **K80**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Rhagoletis pomonella* (EZ 124459), *Drosophila willistoni* (XP 002064648), *D. melanogaster* (NP 476813), *D. grimshawi* (XP 001993004), *D. mojavensis* (XP 002003241) e *D. virilis* (XP 002052597) para um fragmento do gene **Porin**. Os dados das espécies de *Drosophila* e de *R. pomonella* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Não foi significativo para nenhum contraste.....47

**Figura 25-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **K80**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Rhagoletis pomonella* (EZ 139567), *Drosophila willistoni* (XP 002073380), *D. melanogaster* (NP 650048), *D. grimshawi* (XP 001986483), *D. mojavensis* (XP 001998386) e *D. virilis* (XP002050135) para um fragmento do gene **TCTP**. Os dados das espécies de *Drosophila* e de *R. pomonella* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Não foi significativo para nenhum contraste.....48

**Figura 26-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **K80**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Rhagoletis pomonella* (EZ 140549), *Drosophila willistoni* (XP 002075414), *D. melanogaster* (NP 524122), *D. grimshawi* (XP 001987472), *D. mojavensis* (XP 002007253) e *D. virilis* (XP 002046615) para um fragmento do gene **Troponin C**. Os dados das espécies de *Drosophila* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Não foi significativo para nenhum contraste.....48

**Figura 27-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **TN93**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Drosophila willistoni* (XP 002065740), *D. melanogaster* (NP 648905), *D. grimshawi* (XP 001983443), *D. mojavensis* (XP 002012225) e *D. virilis* (XP 002046263) para um fragmento do gene **UQCRX**. Os dados das espécies de *Drosophila* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Não foi significativo para nenhum contraste.....49

**Figura 28-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **TN93**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Rhagoletis pomonella* (EZ 138366), *Drosophila willistoni* (XP 002071860), *D. melanogaster* (NP 728141), *D. grimshawi* (XP 001991571), *D. mojavensis* (XP 002010813) e *D. virilis* (XP 002046177) para um fragmento do gene **Wings Up**. Os dados das espécies de *Drosophila* e de *R. pomonella* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Não foi significativo para nenhum contraste.....49

**Figura 29** – Resultados obtidos pelo programa IMA2 para o modelo de árvore onde *A. fraterculus* se encontra mais separada de *A. obliqua* e *A. sorocula*.....55

**Figura 30** – Resultados obtidos pelo programa IMA2 para o modelo de árvore onde *A. obliqua* se encontra mais separada de *A. fraterculus* e *A. sorocula*.....56

**Figura 31** – Resultados obtidos pelo programa IMA2 para o modelo de árvore onde *A. sorocula* se encontra mais separada de *A. fraterculus* e *A. obliqua*.....56

**Figura 32** – Resultados obtidos pelo programa IMA2 para o modelo de árvore com as espécies *A. fraterculus* e *A. obliqua*.....58

**Figura 33** – Resultados obtidos pelo programa IMA2 para o modelo de árvore com as espécies *A. obliqua* e *A. sorocula*.....58

**Figura 34** – Resultados obtidos pelo programa IMA2 para o modelo de árvore com as espécies *A. fraterculus* e *A. sorocula*.....59

<b>Figura 35</b> – Valores de coalescência profunda para as árvores simuladas para o gene Elongation Protein no modelo de árvore onde <i>A. obliqua</i> está mais separada das outras duas espécies. A seta indica o valor da coalescência profunda da árvore de genes real para este <i>locus</i> . O intervalo de significância ( $p=0,05$ ) está delimitado entre as linhas vermelhas.....	62
<b>Figura 36</b> – Valores de coalescência profunda para as árvores simuladas para o gene Cyclo no modelo de árvore onde <i>A. obliqua</i> está mais separada das outras duas espécies. A seta indica o valor da coalescência profunda da árvore de genes real para este <i>locus</i> . O intervalo de significância ( $p=0,05$ ) está delimitado entre as linhas vermelhas.....	62
<b>Figura 37</b> – Árvore de espécies obtida pela minimização da coalescência profunda ( <i>minimizing deep coalescence</i> ) .....	63
<b>Figura A.1a</b> – Gráficos dos resultados de $m$ , $q$ e $t$ obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde <i>A. fraterculus</i> é a espécie mais distante na árvore.....	74
<b>Figura A.1b</b> – Gráficos dos resultados de $m$ , $q$ e $t$ obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde <i>A. fraterculus</i> é a espécie mais distante na árvore.....	75
<b>Figura A.1c</b> – Gráficos dos resultados de $m$ , $q$ e $t$ obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde <i>A. fraterculus</i> é a espécie mais distante na árvore.....	76
<b>Figura A.2a</b> – Gráficos dos resultados de $m$ , $q$ e $t$ obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde <i>A. obliqua</i> é a espécie mais distante na árvore.....	76
<b>Figura A.2b</b> – Gráficos dos resultados de $m$ , $q$ e $t$ obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde <i>A. obliqua</i> é a espécie mais distante na árvore.....	77
<b>Figura A.2c</b> – Gráficos dos resultados de $m$ , $q$ e $t$ obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde <i>A. obliqua</i> é a espécie mais distante na árvore.....	78
<b>Figura A.3a</b> – Gráficos dos resultados de $m$ , $q$ e $t$ obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde <i>A. sororcula</i> é a espécie mais distante na árvore.....	78
<b>Figura A.3b</b> – Gráficos dos resultados de $m$ , $q$ e $t$ obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde <i>A. sororcula</i> é a espécie mais distante na árvore.....	79
<b>Figura A.3c</b> – Gráficos dos resultados de $m$ , $q$ e $t$ obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde <i>A. sororcula</i> é a espécie mais distante na árvore.....	80
<b>Figura A.4</b> – Gráficos dos resultados de $m$ , $q$ e $t$ obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo de árvore com <i>A. fraterculus</i> e <i>A. obliqua</i> .....	81

<b>Figura A.5</b> – Gráficos dos resultados de $m$ , $q$ e $t$ obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo de árvore com <i>A.fartermculus</i> e <i>A.sororcula</i> .....	82
<b>Figura A.6</b> – Gráficos dos resultados de $m$ , $q$ e $t$ obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo de árvore com <i>A.obliqua</i> e <i>A.sororcula</i> .....	83
<b>Figura A.7</b> - Alinhamento do gene <b>CG7009</b> utilizado para análises de seleção positiva realizadas no PAML.....	84
<b>Figura A.8</b> Alinhamento do gene <b>CG7203</b> utilizado para análises de seleção positiva realizadas no PAML.....	85
<b>Figura A.9</b> Alinhamento do gene <b>CG8064</b> utilizado para análises de seleção positiva realizadas no PAML.....	86
<b>Figura A.10</b> Alinhamento do gene <b>CG5325</b> utilizado para análises de seleção positiva realizadas no PAML.....	87
<b>Figura A.11</b> Alinhamento do gene <b>Elongation Protein</b> utilizado para análises de seleção positiva realizadas no PAML.....	88
<b>Figura A.12</b> Alinhamento do gene <b>Myosin Light Chain</b> utilizado para análises de seleção positiva realizadas no PAML.....	89
<b>Figura A.13a</b> Alinhamento do gene <b>Porin</b> utilizado para análises de seleção positiva realizadas no PAML.....	90
<b>Figura A.13b</b> Alinhamento do gene <b>Porin</b> utilizado para análises de seleção positiva realizadas no PAML.....	91
<b>Figura A.14</b> Alinhamento do gene <b>Translationally Controlled Tumor Protein</b> utilizado para análises de seleção positiva realizadas no PAML.....	92
<b>Figura A.15</b> Alinhamento do gene <b>TroponinC</b> utilizado para análises de seleção positiva realizadas no PAML.....	93
<b>Figura A.16</b> Alinhamento do gene <b>Ubiquinol Cytochrome Reductase</b> utilizado para análises de seleção positiva realizadas no PAML.....	94
<b>Figura A.17a</b> Alinhamento do gene <b>WingsUp</b> utilizado para análises de seleção positiva realizadas no PAML.....	95
<b>Figura A.17b</b> Alinhamento do gene <b>WingsUp</b> utilizado para análises de seleção positiva realizadas no PAML.....	96
<b>Figura A.17c</b> Alinhamento do gene <b>WingsUp</b> utilizado para análises de seleção positiva realizadas no PAML.....	97



<b>Figura A.18a</b> Alinhamento do gene <b>Cyclophylin</b> utilizado para análises filogenéticas.....	98
<b>Figura A.18b</b> Alinhamento do gene <b>Cyclophylin</b> utilizado para análises filogenéticas.....	99

## *Lista de Tabelas*

<b>Tabela 1-</b> Regiões e indivíduos utilizados para os <i>pools</i> de DNA.....	11
<b>Tabela 2-</b> <i>Primers</i> testados e TM (°C) correspondentes.....	13
<b>Tabela 3-</b> <i>Primers</i> testados.....	21
<b>Tabela 4 -</b> Índices de polimorfismo e testes de neutralidade para cada gene com todas as espécies de <i>Anatrepha</i> .....	26
<b>Tabela 5 -</b> Índices de polimorfismo e testes de neutralidade de cada gene para <i>A.fraterculus</i> .....	26
<b>Tabela 6 -</b> Índices de polimorfismo e testes de neutralidade de cada gene para <i>A.obliqua</i> .....	27
<b>Tabela 7-</b> Índices de polimorfismo e testes de neutralidade de cada gene para <i>A.sororcula</i> .....	27
<b>Tabela 8 -</b> Índices de polimorfismo e testes de neutralidade de cada gene para a região <i>intrônica</i> .....	28
<b>Tabela 9 –</b> Sítios segregantes encontrados nas espécies e entre espécies por locus comparando <i>A. fraterculus</i> e <i>A. obliqua</i> .....	30
<b>Tabela 10 -</b> Sítios segregantes encontrados nas espécies e entre espécies por locus comparando <i>A. obliqua</i> e <i>A. sororcula</i> .....	30
<b>Tabela 11 -</b> Sítios segregantes encontrados nas espécies e entre espécies por locus comparando <i>A. fraterculus</i> e <i>A. sororcula</i> .....	31
<b>Tabela 12a –</b> Parâmetros estimados e valores de logaritmo da máxima verossimilhança para os modelos M1a, MA e MA restrito nos dados macroevolutivos.....	52
<b>Tabela 12b –</b> Parâmetros estimados e valores de logaritmo da máxima verossimilhança para os modelos M1a, MA e MA restrito nos dados macroevolutivos.....	53
<b>Tabela 13 –</b> Contraste dos modelos M1a, MA restrito e MA por LRT nos dados macroevolutivos.....	54
<b>Tabela 14 –</b> Tempo de divergência (milhões de anos) e Tamanho efetivo de cada população para cada modelo de árvore.....	59
<b>Tabela 15 –</b> Taxa de migração por gene por geração, na coalescência.....	60
<b>Tabela 16 –</b> Dados da mediana da coalescência profunda para as árvores simuladas para cada locus comparadas com a coalescência profunda das árvores de genes para cada um dos três modelos de árvores de espécie.....	61

## *Sumário*

<b>Resumo.....</b>	<b>II</b>
<b>Abstract.....</b>	<b>IV</b>
<b>1-Introdução.....</b>	<b>1</b>
1.1 O Gênero <i>Anastrepha</i> .....	1
1.2 Especiação.....	5
1.3 Aparelho e genes ligados à especiação.....	5
1.4 Bibliotecas de cDNA.....	6
1.5 Estudos sobre forças seletivas.....	7
<b>2-Objetivos.....</b>	<b>10</b>
<b>3-Material e Métodos.....</b>	<b>11</b>
3.1 Espécies estudadas.....	11
3.2 Coleta e Manutenção do Material.....	11
3.3 Procedimentos moleculares.....	11
3.4 Análises das sequências de DNA.....	16
<b>4-Resultados e Discussão.....</b>	<b>21</b>
4.1 Análises de Polimorfismo, Divergência e Testes de Neutralidade.....	23
4.2 Análise das redes de haplótipos.....	31
4.3 Testes de seleção.....	44
4.4 – Análises de Isolamento e Fluxo gênico.....	54
4.5 – Comparação das árvores de genes e de espécies.....	60
<b>5-Conclusão.....</b>	<b>64</b>
<b>6- Referências .....</b>	<b>66</b>
<b>7- ANEXOS.....</b>	<b>74</b>

## 1. Introdução

### 1.1 - O Gênero *Anastrepha*

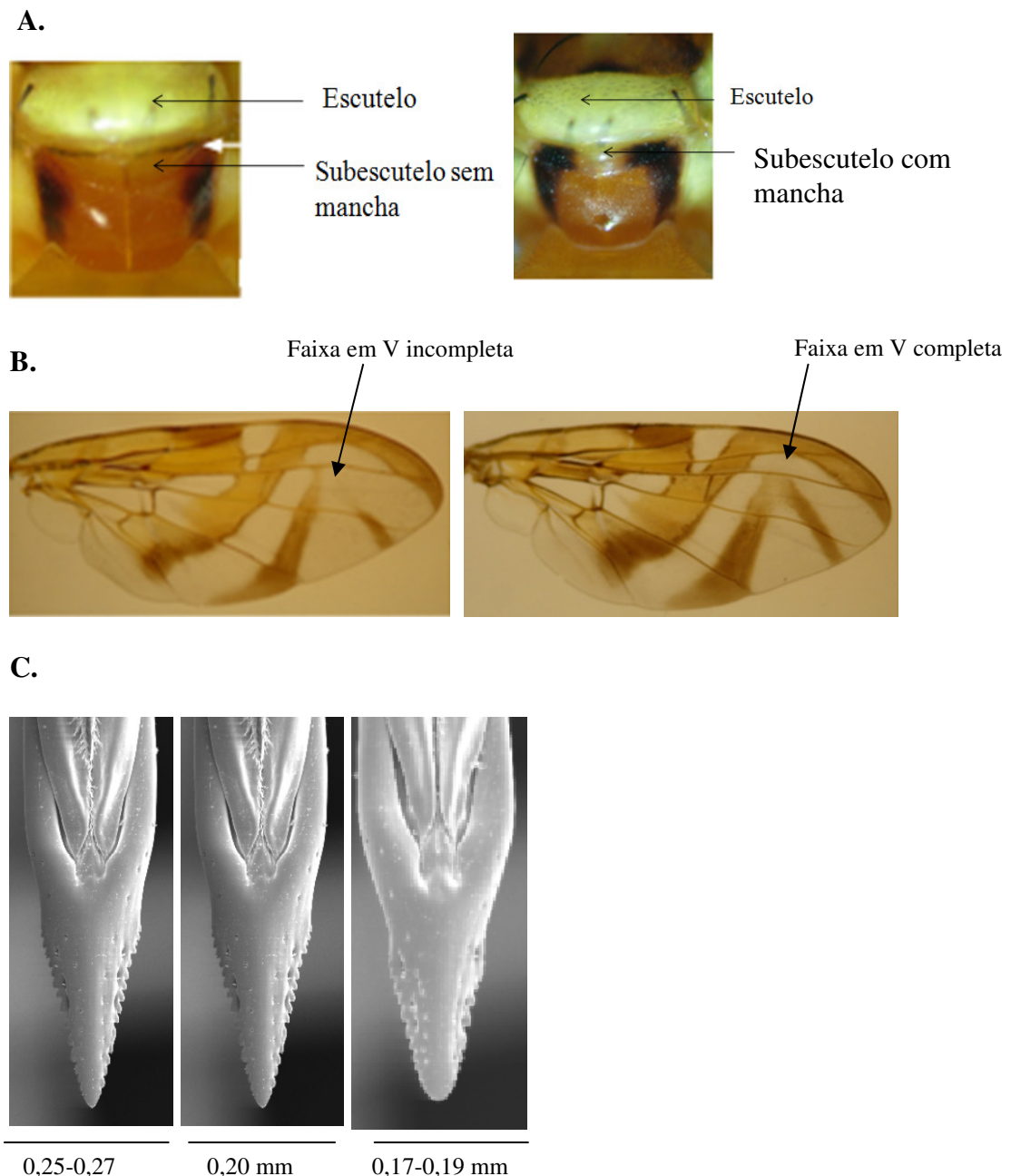
A família Tephritidae, que é endêmica à região Neotropical, é considerada de grande importância econômica por causar grandes prejuízos na produção de frutos. Esta família possui aproximadamente 500 gêneros, dentre eles, as moscas-das-frutas do gênero *Anastrepha* (Norrbom, 2004). O gênero *Anastrepha* é o maior dentro da família Tephritidae (Trypetinae, Toxotrypanini) (Norrbom, 2001), com 197 espécies reconhecidas atualmente, sendo que 94 destas se encontram no Brasil (Zucchi, 2000).

No Brasil, o gênero *Anastrepha* tem importância por apresentar espécies responsáveis por danos de grandes proporções às culturas de frutas carnosas, não só pelo desenvolvimento larval no endocarpo, mas também pelo processo mecânico da oviposição, que permite a invasão de outros organismos como fungos causadores de podridão, provocando o amadurecimento precoce dos frutos (Duarte & Malavasi, 2000). Dessa forma, as espécies do grupo *fraterculus* estão relacionadas com os maiores danos em cultivares por atacar frutos verdes e maduros indistintamente (Zucchi, 2000). A utilização de defensivos agrícolas para a eliminação destas moscas traz grandes prejuízos ambientais, o que estimula a procura por métodos alternativos para o seu controle. A presença de grande número de espécies crípticas de difícil distinção (Selivon, 2000) torna ainda mais difícil a elaboração de programas que visem o controle dessas pragas.

Neste trabalho, daremos enfoque a três espécies deste gênero: *Anastrepha fraterculus*, *Anastrepha obliqua* e *Anastrepha sororcula*. Tais espécies são importantes não apenas por sua ampla distribuição, mas também por suas preferências de frutos para a oviposição. Essas espécies estão incluídas no “complexo *fraterculus*”, composto por pelo menos 29 espécies (Norrbom *et al.*, 2000) e caracterizado por ser um complexo heterogêneo de espécies crípticas, de difícil distinção (Morgante *et al.*, 1980).

Existem algumas características, como a morfologia do ápice do ovipositor, a presença de manchas no subescutelo e o padrão de coloração das asas (Santos *et al.*, 2001; Selivon, 2000) que permitem a distinção entre *Anastrepha fraterculus*, *A. sororcula* e *A. obliqua*, objetos deste estudo (**Figura 1**). Em *A. fraterculus*, os cruzamentos ocorrem entre 7-13 horas da manhã, enquanto que em *A. sororcula*

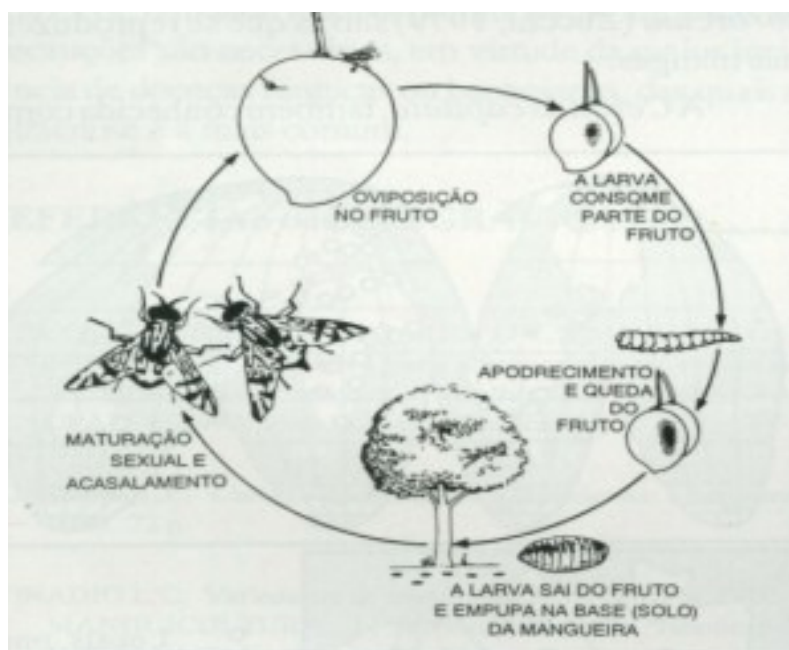
ocorrem entre 15-19 horas, não havendo, portanto sobreposição entre os períodos de cópula dessas duas espécies (Selivon, 2000). Já a espécie *A. obliqua* tem comportamento de corte geralmente, mas não exclusivamente, matutino (Heath *et al.*, 2000). Além disso, o isolamento temporal também pode auxiliar na identificação de espécies por contribuir com o isolamento reprodutivo (Malavasi & Morgante, 1982; Solferini & Morgante, 1987).



**Figura 1**– Esquema de *Anastrepha* do grupo *fraterculus* com as regiões que auxiliam na identificação das espécies. **A**- Tórax; **B**- asa; **C**- ápice do acúleo (Zucchi 2007). *A. fraterculus* possui mancha no subescutelo, faixa V geralmente completa, ápice do acúleo de 0,25 – 0,27 mm; *A. obliqua* não possui mancha no subescutelo, faixa V geralmente incompleta, ápice do acúleo de 0,20 mm; *A. sororcula* possui mancha no subescutelo, faixa V geralmente incompleta, ápice do acúleo de 0,17 – 0,19 mm.

As espécies deste grupo devem ser consideradas oligotróficas, pois embora exista certa especificidade, em geral elas não são restritas a apenas um hospedeiro (Solferini & Morgante, 1987). *A. fraterculus* infesta aproximadamente 67 espécies de plantas, 24 pertencentes à família *Myrtaceae*, *A. sororcula* infesta 19 espécies, sendo 15 delas *Myrtaceae*, e *A. obliqua* infesta 27 espécies, 16 *Myrtaceae* e oito *Anacardiaceae* (Zucchi, 2000). Alguns autores consideram que em insetos fitófagos, o polimorfismo associado ao hospedeiro pode representar o primeiro estágio de especiação (Malavasi & Morgante, 1982; Feder *et al.*, 1998).

O ciclo de vida das moscas do gênero *Anastrepha*, consiste na oviposição em frutos carnosos, os ovos eclodem em larvas que alimentam-se da polpa dos frutos provocando o apodrecimento do fruto que cai no solo. As larvas saem dos frutos enterram-se no solo onde ocorrerá o empupação das larvas. As pupas, entre oito e 14 dias, darão origem as moscas que se tornarão maduras sexualmente, após cinco a sete dias de nascidas, recomeçando o ciclo (Malavasi & Zucchi 2000) (**Figura 2**).



**Figura 2** - Ciclo de vida das moscas-das-frutas

Apesar de não haver um único marcador capaz de diferenciar as espécies do grupo *fraterculus*, o conjunto de características acima mencionadas é capaz, na grande maioria das vezes, de distinguir as espécies. Porém, há evidências de que esses marcadores morfológicos e comportamentais não correspondam aos padrões de variação e divergência genética dessas moscas (Morgante *et al.*, 1980; Malavasi & Morgante,

1982). Recentemente foi descrito que um dos principais caracteres utilizados para a identificação de espécies do grupo, o ápice do ovipositor, apresenta maior sobreposição interespecífica, e portanto, é um marcador menos confiável do que previamente se supunha (Araujo & Zucchi, 2006). Devido à alta plasticidade fenotípica destes marcadores morfológicos, que resulta na sobreposição interespecífica de diversas medidas quantitativas, há dificuldade em utilizá-los na separação de espécies muito próximas, principalmente as do grupo de espécies *fraterculus* (Malavasi & Zucchi, 2000; Araujo & Zucchi, 2006).

Dessa forma, um dos grandes problemas das moscas do grupo *fraterculus* é a difícil identificação das espécies, até mesmo com o uso de marcadores morfológicos (Zucchi, 1978), comportamentais (Morgante *et al.*, 1983), alozímicos (Morgante *et al.*, 1980; Matioli *et al.*, 1986), cariotípicos (Solferini & Morgante, 1987) e moleculares (Smith-Caldas *et al.*, 2001; Sobrinho, Jr, 2009), sendo sugerido que *A. fraterculus* seja de fato um complexo de espécies crípticas (Solferini & Morgante, 1987; Steck, 1991; Selivon *et al.*, 2005) e que é necessário um estudo mais detalhado acerca do status das espécies.

Na verdade, a dificuldade na identificação das espécies no grupo de moscas-das-frutas *A. fraterculus* oriunda de certas características que tornam este grupo um excelente modelo para o estudo dos processos envolvidos em especiação e diferenciação populacional. A existência de polimorfismos intraspecíficos bem como de cruzamentos viáveis entre diferentes espécies do grupo (Santos *et al.*, 2001; Selivon *et al.*, 2005) sugere que estas espécies do grupo *fraterculus* divergiram há pouco tempo e que os mecanismos de isolamento reprodutivo provavelmente não são totalmente eficientes para o isolamento destas espécies, possivelmente porque ainda não acumularam diferenças genéticas significativas em sistemas outros que não os envolvidos processo de especiação (Wu, 2001). Além disso, em certas condições no laboratório tais espécies são capazes de produzir híbridos viáveis quando cruzadas entre si (Henning & Matioli, 2006) o que facilita o estudo dos genes envolvidos na determinação das diferenças interespecíficas.

A identificação, isolamento e caracterização de marcadores genéticos que permitam a distinção entre as principais espécies do grupo *fraterculus* beneficiaria enormemente a elaboração de programas de manejo destas moscas. Além disso, o estudo referente à estrutura e distribuição populacional das espécies do grupo *fraterculus* permite o correto manejo das principais regiões afetadas por essa praga.

## 1.2 - Especiação

Muitos pesquisadores consideram que a especiação ocorra principalmente devido aos processos que causam isolamento reprodutivo, e esta barreira reprodutiva é na maioria das vezes, causada por impossibilidade de troca de material genético entre dois indivíduos (Wu & Ting 2004). Por este motivo, a maioria dos pesquisadores considera que processos que levam ao isolamento reprodutivo são mais prováveis de levar à especiação (Bultin & Ritchie, 2001). A especiação, portanto, ocorreria mais frequentemente em ocasiões onde não há fluxo gênico, tornando mais comum o processo de especiação alopátrica, que segundo Ernst Mayr consiste na diferenciação genética entre populações isolada. Porém, não se pode descartar a especiação simpátrica, que apesar de mais rara (Ritchie, 2007) pode também ser relevante.

Desta forma pode-se dizer que o processo de especiação não é um evento que dependa necessariamente de um isolamento genético para ocorrer, uma vez que não é um evento exclusivamente associado à ocorrência ou não de fluxo gênico. A diferenciação que ocorre entre populações e a especiação estão associadas, principalmente, à seleção, à deriva genética e ao balanço dessas duas forças (Templeton, 1989). Embora todas as outras forças evolutivas possam causar a diferenciação genética entre indivíduos e levar a especiação.

Há bastante tempo têm sido estudados genes que possivelmente causam o isolamento reprodutivo (Coyne & Orr, 2004), com o intuito de encontrar os genes envolvidos na especiação. As diferenças genéticas entre as espécies, após o isolamento, vão se acumulando com o passar do tempo. À medida que aumenta o tempo de divergência fica mais difícil identificar os genes que são realmente responsáveis pelo isolamento reprodutivo de outros polimorfismos que surgiram depois do processo de especiação.

Considerando-se a dificuldade em se encontrar marcadores que nos permitam identificar espécies do gênero *Anastrepha*, buscamos um entendimento maior das forças evolutivas envolvidas na diferenciação deste complexo de espécies estudando genes expressos no aparelho reprodutor das fêmeas de *A. fraterculus*.

## 1.3 - Aparelho Reprodutor e genes ligados à especiação.

Em animais que se reproduzem por fecundação interna, a divesificação da genitália representa uma forma de tendência evolutiva. Em geral, a evolução da



genitália animal é mais rápida e se difere em sua morfologia até mesmo entre espécies proximamente relacionadas (Eberhard, 1985). Podemos observar este fato em espécies de *Drosophila*, nas quais o aparelho reprodutivo de machos e fêmeas é aproximadamente duas vezes mais diferenciados entre as espécies que os caracteres não relacionados à reprodução (Civetta & Singh, 1995). Muitos pesquisadores sugerem que a evolução de caracteres morfológicos e moleculares relacionados aos aspectos reprodutivos acontece de forma mais rápida devido ao conflito sexual, que é a chave do processo de evolução (Arnqvist *et al.*, 2000). O conflito sexual consiste nas diferenças que existem nos custos e benefícios envolvidos na reprodução para machos e fêmeas (Watson *et al.*, 1998). Para exemplificar, podemos citar, por exemplo, diversas proteínas relacionadas à fertilização que foram identificadas como proteínas que evoluem rapidamente (Swanson *et al.* 2003). Como também, genes de autoincompatibilidade em plantas (Castric & Vekemans, 2007) e proteínas reprodutivas masculinas e femininas de *Drosophila* (Swanson *et al.* 2001; Panhuis & Swanson, 2006) têm mostrado o mesmo padrão de diversificação rápida. A maioria dos estudos moleculares, nesta área, têm se concentrado recentemente em genes envolvidos em fertilização ou interação entre machos e fêmeas e apenas poucos nos genes responsáveis pela diferenciação sexual.

Além disso, existem estudos que mostram que genes que estão associados à reprodução apresentam sinais de evolução adaptativa (Swanson & Vacquier, 2002) e que genes que codificam proteínas de glândulas acessórias apresentam uma evolução diferente de outras classes de genes, apresentando uma maior divergência de aminoácidos que genes não reprodutivos em espécies proximamente relacionadas (Swanson *et al.* , 2001; Mueller *et al.* , 2005).

Apesar do crescimento das pesquisas em evolução com proteínas relacionadas a caracteres sexuais e reprodução, principalmente em machos, ainda há poucos estudos com fêmeas ( Swanson *et al.* , 2004; Panhuis & Swanson, 2006). Portanto, é preciso que sejam estudadas as sequências gênicas de ambos os sexos para uma melhor compreensão dos mecanismos evolutivos, visto que muitas proteínas estão envolvidas direta ou indiretamente em interações com o outro sexo.

#### **1.4 - Bibliotecas de cDNA**

Quando buscamos a identificação de genes expressos em um determinado tecido ou condição é necessária a utilização de bibliotecas construídas a partir de cDNAs ou ESTs (*Expressed Sequenced Tags*) (Cortner & Woude, 1997). Os resultados obtidos

através destas bibliotecas têm sido usados para a identificação de novos genes (Swanson *et al.*, 2004; Kelleher *et al.*, 2007; Gomulski *et al.*, 2008), desenvolvimento de novos marcadores, com o objetivo de criar um mapa de ligação para as espécies estudadas (Beeman & Brown, 1999) e também avaliar níveis de expressão para identificar genes que estejam sendo super ou sub expressos (Kelleher *et al.*, 2007). Além disso, estudos com alguns genes para identificar sua expressão em diferentes tecidos e estágios de vida, com o objetivo de caracterizar sua regulação e função, podem ser feitos posteriormente (Gomulski *et al.*, 2008).

Uma vez que novas tecnologias relativamente rápidas têm facilitado o processo de produção de bibliotecas cDNA, tais bibliotecas têm se mostrado uma das ferramentas mais eficientes para a identificação de genes em espécies sem genomas sequenciados.

A estratégia mais utilizada é a construção de bibliotecas de cDNA a partir de um *pool* de indivíduos, o que permite um isolamento rápido de um grande número de marcadores genéticos a partir do sequenciamento de ESTs (Expressed Sequenced Tags). Como estas bibliotecas são geradas a partir de um pool de indivíduos, as várias cópias obtidas para uma determinada região gênica fornecem já uma primeira avaliação do polimorfismo encontrado na população (Roethele *et al.*, 2001; Lorenzen *et al.*, 2005).

Com o intuito de investigar o papel de genes ligados ou não à reprodução na especiação das moscas das frutas do gênero *Anastrepha*, este trabalho buscou a amplificação, a partir de uma biblioteca genômica de cDNAs, desenvolvida no Laboratório de Genética de Populações e Evolução na Universidade Federal de São Carlos (Kaminski, 2007), de genes expressos no aparelho reprodutivo de fêmeas de *Anastrepha fraterculus* com o objetivo de avaliar inicialmente o eventual papel de cada um na diferenciação deste complexo, bem como investigar mais detalhadamente a dinâmica das forças evolutivas envolvidas com os padrões de evolução molecular destes genes.

### **1.5 - Estudos sobre forças seletivas, evolução molecular e parâmetros populacionais.**

Os métodos mais comuns para se investigar a evolução adaptativa molecular consistem em se basear nas expectativas neutras do espectro de frequências de haplótipos que fornecem uma idéia geral das forças seletivas que estão agindo sobre a região de DNA estudada usando parâmetros populacionais (Tajima, 1989a; McDonald & Kreitman, 1991; Fu & Li, 1993). No entanto, estes métodos podem ser influenciados

por mudanças demográficas e sua interpretação é raramente inequívoca (Tajima, 1989a; Yang, 2002). Por isso, métodos que levam em consideração a razão entre mudanças não sinônimas e sinônimas ( $\omega$ ), ou seja, não enviesados, para se inferir padrões de seleção positivas, tem sido comumente usados para se investigar mudanças positivamente selecionadas (Galindo *et al.*, 2003; Clark & Swanson, 2005; Castric & Vekemans, 2007).

Os desafios associado à inferência das relações evolutivas entre espécies ou populações que divergiram recentemente consiste na triagem incompleta (*lineage sorting*) das linhagens. É importante lembrarmos que a delimitação das espécies pode ser equivocada se, por discordância, as linhagens de genes dentro de uma espécie coalescem abaixo da divergência das espécies (ou seja, abaixo de um evento de especiação), também conhecida como problema de discordância entre árvore de genes e árvore de espécies (Maddison, 1997).

No entanto, é também um equívoco acreditar que as árvores de genes não fornecem informações sobre as espécies. De fato, trabalhos recentes demonstram que genealogias de genes fornecem informações importantes sobre a história da separação de espécies (ou seja, a árvore de espécies), apesar da incompleta triagem das linhagens em uma árvore de gene (Maddison & Knowles, 2006; Knowles & Carstens, 2007). Estes estudos nos mostram que podemos delimitar linhagens de espécies mesmo que a monofilia recíproca ainda não tenha sido seja atingida.

Populações que ainda trocam migrantes podem ainda assim se diferenciar até ao ponto de especiação. Nesse caso o resultado final dependerá do balanço de forças evolutivas diversificadoras e homogeneizadoras. A deriva genética aumentará divergência, enquanto fluxo gênico retardará. A seleção natural pode homogeneizar populações quando *selective sweeps* são transportados por migração entre populações e pode aumentar divergência por meio de *selective sweeps* específicos de uma dada população (Hey e Nielsen , 2004). Utilizamos então nesse trabalho, um modelo de isolamento com migração que esclarece o balanço da deriva genética e migração na diferenciação das espécies desse grupo (Hey & Nielsen 2004; Hey & Nielsen 2007).

Em seguida, aplicamos um método que usa a coalescência para estimar as probabilidades de árvores de genes (Degnan & Salter, 2005) dado uma árvore de espécies, levando em conta eventos demográficos que foram significativos para o isolamento reprodutivo, para avaliar a probabilidade de divisão da linhagem (Hudson & Coyne, 2002). A mesma teoria da coalescência serve de base para estimar a

probabilidade de monofilia recíproca nas árvores gene (Rosenberg, 2003; Hudson & Turelli, 2003). No entanto, a abordagem aqui proposta pode ser aplicada a espécies para as quais não houve tempo suficiente para eliminarmos o polimorfismo ancestral por deriva genética. Focalizamos a atenção sobre a relação entre árvores de genes e divergência das linhagens das espécies ao invés de igualar as árvores gene com a história da espécie (Maddison, 1997).

Compreender a natureza, o tempo e o contexto geográfico dos acontecimentos históricos e dos processos populacionais responsáveis pela distribuição espacial da diversidade genética é muito relevante para abordar questões relacionadas à especiação, seleção e aplicação desses conhecimentos para o controle e conservação das espécies. A partir de um quadro de divergência genética populacional (Hey & Nielsen 2004; Hey & Nielsen 2007), estimamos parâmetros demográficos relevantes buscando esclarecer a arquitetura genética da diferenciação entre espécies deste grupo. Tal estudo foca em diversos genes nucleares que são investigados para padrões de evolução molecular e inferências filogenéticas buscando comparar suas árvores de genes para estudar a separação das espécies do grupo *fraterculus*.

## 2. *Objetivos*

### **Objetivo Geral**

- Analisar a diversidade populacional e dados de evolução molecular de 12 genes nucleares buscando investigar se ocorre ou não a separação das três espécies das moscas-das-frutas do grupo *fraterculus*, bem como avaliar o papel de cada um dos genes aqui estudados nesta diferenciação.

### **Objetivos Específicos**

- Amplificar sequências de diversos genes de uma biblioteca genômica de cDNAs expressos em aparelho reprodutivo de fêmeas de *Anastrepha fraterculus*.
- Estudar o nível de variabilidade genética intra e interespecífica das regiões amplificadas dos genes estudados neste trabalho.
- Testar a hipótese destes genes estarem evoluindo devido à ação de seleção positiva ou devido a um relaxamento de restrição seletiva.
- Verificar o potencial uso de algumas destas regiões como marcadores para diferenciar as espécies *A. fraterculus*, *A. obliqua* e *A. sororcula*.
- Estimar parâmetros populacionais e inferir a árvore de espécies usando dados das regiões gênicas aqui estudadas utilizando estratégias bayesianas de máxima verossimilhança e coalescência.

### 3. Material e Métodos

#### 3.1 - Espécies estudadas

As moscas-das-frutas do grupo *fraterculus* pertencem à família Tephritidae, que possui aproximadamente 500 gêneros, dentre eles, o gênero *Anastrepha* (Norrbom, 2004). Dentro deste gênero se encontram as três espécies estudadas neste trabalho: *A. fraterculus*, *A. obliqua* e *A. sororcula* de diversas localidades do Brasil (**Tabela 1**).

#### 3.2 – Coleta e manutenção do material

Acondicionamos os frutos provenientes de diversas regiões do Brasil em caixas plásticas com vermiculita até que houvesse a empupação das larvas e a emergência das moscas. Alguns dias após o nascimento, identificamos morfológicamente as espécies de *Anastrepha* e estocamos em etanol 70% no freezer ou mantivemos, algumas populações, em laboratório, sob dieta artificial constituída de 10 ml de mel, 30 ml de hidrolisado de proteína de milho, 0,5 ml de Levenil, 6g de levedo de cerveja, 0,5g de Sustagen, 25g de açúcar mascavo e 45g de açúcar branco até o momento de sua utilização.

#### 3.3 - Procedimentos moleculares

##### Extração de DNA

Extraímos o DNA com Tiocianato de Guanidina, fazendo a digestão enzimática prévia de proteínas com proteinase K em tampão apropriado (Nelson & Krawetz, 1992). Fizemos, então, um *pool* de DNAs para cada uma das espécies estudadas, com 5 indivíduos de cinco localidades distintas (**Tabela 1**).

**Tabela 1-** Regiões e indivíduos utilizados para os *pools* de DNA.

	Estado	Nº de indivíduos/local	Geração
<u><i>Pool A. fraterculus</i></u>		20	
Curitiba	PR	5	F4
Florianópolis	SC	5	F2
Joinville	SC	5	F3
João Pessoa	PB	5	F1
<u><i>Pool de A. obliqua</i></u>		20	
Bonito	PE	5	F4
São Carlos	SP	5	F5
Bariri	SP	5	F5
Araguaina	TO	5	F0
<u><i>Pool de A. sororcula</i></u>		21	
Guarapari	ES	9	F1
Rio das Ostras	RJ	6	F0
Búzios	RJ	6	F1

### **Construção dos primers**

Para o isolamento e análise de genes distintos no genoma de *Anastrepha*, utilizamos 27 conjuntos de *primers*, que foram desenvolvidos a partir do alinhamento manual no programa BioEdit (Hall, 1999), de diversas sequências, obtidas de uma biblioteca de cDNAs de genes do aparelho reprodutivo feminino de *Anastrepha fraterculus* (Kaminski, 2007), com o de outras espécies de insetos próximas (outros Tephritidae ou mesmo outros Diptera como Drosophilidae e Culicidae). Fizemos estes alinhamentos utilizando, também, as sequências de aminoácidos correspondentes a cada sequência de nucleotídeos usada no desenvolvimento de cada um dos *primers*. O alinhamento dos aminoácidos também fizemos no programa BioEdit, pelo uso do *software* Clustal W (Higgins *et al.* 1994), este *software* permite alinhar sequências mesmo quando o tamanho entre elas são diferentes, relacionado a processos de inserções ou deleções. Então, construímos esses *primers* selecionando regiões mais conservadas para os genes aqui estudados. Para tanto, escolhemos regiões de aproximadamente 20 bases, com quantidade de citosina e guanina entre 50% e 60%, evitando repetições de bases.

Analizamos as temperaturas de desnaturação, valores de  $\Delta G$  e possibilidade de formação de estruturas secundárias, tanto em hibridações entre os mesmos *primers* como entre os diferentes do conjunto no *software* *OligoAnalyser 3.0* ([www.idtdna.com/analyser/Default.aspx/oligocalc.asp](http://www.idtdna.com/analyser/Default.aspx/oligocalc.asp)). Dessa forma, foi verificada a estabilidade das estruturas com o objetivo de tentar reduzir a produção de artefatos nas reações de amplificação.

### **Amplificação por PCR e purificação do produto**

Testamos cada conjunto de *primers* desenvolvidos (**Tabela 2**) em diferentes temperaturas e usamos diversas concentrações de reagentes para estipular as condições ótimas de amplificação. Fizemos testes utilizando diferentes concentrações de Cloreto de Magnésio ( $MgCl_2$ ) - (1,0mM, 1,5mM e 2,5mM) - e alteramos também as concentrações dos *primers*, 0,2mM e 0,4mM. Nas reações de amplificação utilizamos uma mistura de *Taq polimerase* e 1/100 U de *Pfu polimerase*, que reduz a taxa de incorporação de erros e aumenta a fidelidade na PCR (Cline *et al.*, 1996). Fizemos as reações de amplificação utilizando os reagentes nas seguintes concentrações: Água, Tampão ( $MgCl_2$  1,0 mM, Tris-HCl pH = 9,0 10mM, KCl 50mM), BSA (0,2 $\mu g/\mu L$ ),

dNTPs (0,2 $\mu$ M), *primer forward* (0,2mM), *primer reverse* (0,2mM), Taq polimerase(0,05 U/ $\mu$ L) e DNA(1 $\mu$ L/26  $\mu$ L de solução).

Corremos os produtos das reações de amplificação em géis de agarose 1%, adicionados de brometo de etídio, que permite a visualização em luz ultravioleta para verificar a eficiência dos *primers* e estabelecer as melhores temperaturas e condições de *primers*. Utilizamos marcadores de peso molecular para confirmar se o tamanho do fragmento amplificado correspondia ao tamanho esperado.

Uma vez estabelecidas as condições de amplificação, os produtos de PCR obtidos de *pools* de indivíduos de cada espécie foram purificados através do protocolo de precipitação diferencial com PEG (Polietileno Glicol) 8000 (Lis & Schleif, 1975).

**Tabela 2-** *Primers* testados e temperaturas de *melting* (Tm) (°C) correspondentes.

<i>Loci</i>	<b>Primer F (5'-3')</b>	<b>Primer R (5'-3')</b>	<b>TM (°C)</b>
<b>TCTP</b>	ATGAAGATCTACAAGGATATCATC	TTCACACATGGCTACCATACCAT	50
<b>Cyclophylin</b>	CTCCCCGTGTATTTTCGAT	TCCAGTACCGTTGTGATTCGT	52,9
<b>Troponin C</b>	GGATCTAACACCTGAGCAAATTG	TTCATCCACTTCTTCGATGAG	52,1
<b>UQCRX</b>	ATGGCATGCAACGCGAATAAAG	TACACGGGTGATTGGCAGGGA	56,8
<b>Elongation Protein</b>	TGTCATAGTTATCGGTGAAATCATA	ACGCAGTCATACACGGGATAC	51,8
<b>Porin</b>	AAGACCAATACCCCTCTGGT	AAGACCAATACCCCTCTGGT	56,2
<b>MLC</b>	AATTAGCTGAATTCCAAGAAGC	CGTAGTTAAAAGATGTCTCAAT	49
<b>CG8064</b>	CTGGAGTGCTTGGAGATTTGTAA	CATGCACATTCATACCAATTGTAT	52
<b>Wings up</b>	GCTAAACAGGCCGAGATTGAGC	CTTGCGCTTAAACAACATACTCAAGG	56,4
<b>CG7009</b>	GAGACATATACTATCGATTGGCC	ACCAGTTACATCAGGTGCACCATC	52,5
<b>CG5325</b>	ATGGCAGAAAATAAACAGAACAAAG	GTGGTAATTCACCGCTGGGTGT	52,2
<b>CG7203</b>	ATGAAGTTCGCCGTTCCGTAGTC	GTGAATTTGGAACCAGCACCAGGC	61,3
<b>sniffer</b>	ATGAATTCTACTAATTACGGG	AAGATCAATGCTCATGGACTT	49
<b>Copper Transporter</b>	ATGATTATGGTGTTCCATGGTGG	CAGTTGTTGACGTGAAAATTTCA	52,4
<b>C-type lectin</b>	AGAGCCACCACAACCACATCTACG	CGCAAATACGTCCGGCAGTCCAAATG	60,9
<b>Nimrod</b>	ACGGCGGAGTTCATCAATAAGA	CGCCGTGCAAATGCCATT	56,2
<b><math>\alpha</math>-trypsin</b>	CGTACTGGGCGCATCTGTACCTGAGG	CCTCAGGTACAGATGCGCCAGTACG	64,2
<b>Ubiquitin</b>	ATGCAAATTTTCGTAAAAACCC	GGCCATGAAAACACCAGCCCC	51
<b>Vitelline</b>	GCGCCGCCATGCCAAGAA	TCCGCTACTGCTACCCGAT	58
<b>CG 3751</b>	AAGGACACTGGTTTGTGGGGTATC	CGATGAGCGGATTCTGTTGTATTCAGG	58,9
<b>CG 10306</b>	ATGTCGCACCTCGTGAAATGG	ATGCTGAACTTCGACATCTCACGTAC	58,1
<b>CG9894</b>	ATGTCAGCAGCCACAGAAC	CKGARTCWSRCTRCGCGGC	50,2
<b>CG1822</b>	ATGGATCCATATACAGTGTG	GCGAGCAGATTGTAAAGC	48
<b>CG6503</b>	TGGCACCATCGGTATTAGCCTCAA	CATGAAAGTGCTGTTCTTCTCAC	56,2
<b>CG6878</b>	GCATTCAACCGCGAGCTGCATAAA	TAGCGGAGTTTATGGCCAAGCACA	60,6
<b>CG3153</b>	TTCCGTCCAGTACCGTTGTGATTC	TGGCCCTAGTTAAAGCGTTGGTGA	58,8
<b>CG30273</b>	TTAGCAACGTATGCCTGTCCCGAT	TGCGCAATGACCTACAGTCTGTGT	60,3



Para determinar a pureza e concentração dos fragmentos obtidos, realizamos uma nova eletroforese em gel de agarose 1% com brometo de etídio. Esta concentração foi também estabelecida através de *Dot Blots* (quantificação em gota), comparando a intensidade da fluorescência (proporcional a concentração de DNA aplicado) de cada amostra com padrões conhecidos (marcadores).

As regiões do genoma testadas neste trabalho, nas moscas das frutas do grupo *fraterculus* codificam para as seguintes proteínas e apresentam as seguintes funções:

- **CG7009**, da família das histidinas, responsável pela metilação do rRNA;
- **Troponin C**, família das troponinas, e que detecta os níveis de cálcio celular promovendo contração muscular, em *Drosophila* esta proteína é codificada por três genes, sendo que, um deles parece ser específico para os músculos de vôo;
- **Myosin Light Chain (MLC)**, família das miosinas, e é importante no mecanismo de contração muscular, a partir da entrada de cálcio ( $\text{Ca}^{++}$ ) pelo retículo sarcoplasmático ou pelo espaço extracelular;
- **Wings Up**, responsável pelo posicionamento vertical das asas de *Drosophila*, os mutantes para este gene não podem voar e também não são capazes de bater as asas durante o comportamento de corte;
- **Porin**, proteína da família das porinas, presente na mitocôndria onde regula o metabolismo energético, a apoptose e o transporte de metabólitos entre o citosol e a mitocôndria. As isoformas Porin 1 e Porin 2 encontram-se presentes nos tecidos reprodutivos de *Drosophila*, sendo que a Porin 2 é encontrada em maior quantidade nos testículos do macho, que nos ovários das fêmeas. Além disso, esta é uma proteína que, em *Drosophila*, pode ter influência na formação dos olhos;
- **Ubiquinol Cytochrome Reductase (UQCRX)**, que está presente na cadeia respiratória mitocondrial;
- **CG5325**, que participa do desenvolvimento do sistema nervoso;
- Família de proteínas **Translationally Controlled Tumor Protein (TCTP)**, relacionada com o desenvolvimento embrionário e participa do controle do ciclo celular (Hinojosa-Moya *et al.*, 2008).
- **Cyclophilin**, que é constituinte das moléculas do sistema imune. A expressão de proteínas do sistema imune em órgãos reprodutivos já foi indicada por

outros autores (Schmid-Hempel, 2005) e pode estar associada tanto a uma barreira de proteção contra bactérias e fungos quanto a um sistema de defesa das fêmeas contra not-selfs como por exemplo, o esperma do macho, que pode ser considerado pela fêmea como um organismo intruso dentro de seu sistema ou mesmo contra doenças sexualmente transmissíveis;

- **Carbonyl Reductase Sniffer**, confere proteção contra os efeitos deletérios da oxidação nos neurônios;
- **Nimrod e  $\alpha$ -Trypsin**, envolvidos no processo biológico da proteólise;
- **Vitelline**, faz parte da família de proteínas vitelínicas, constituinte da membrana vitelínica, participando, então, da oogênese;
- **Ubiquitin**, desempenha função na regulação das proteínas. Participando biogênese dos ribossomos, na modulação de receptores celulares, na expressão gênica, na reparação do DNA, na resposta ao stress, na morfogênese neuronal, além de desempenhar funções no controle dos ritmos circadianos;
- **C-type Lectin**, família da lectinas, e participa do processo metabólico da quitina;
- **Copper Transporter**, proteína importante para o transporte de cobre para as células, estudos mostram a participação do cobre no crescimento e desenvolvimento embrionário de *Drosophila melanogaster* (Zhou *et al*,2003).
- **Elongation Protein ( Elp)** e **CG10306**, desempenhando papéis na tradução de RNA, sendo que, a Elongation Protein está envolvida na fase de alongamento e a proteína **CG10306** na fase de iniciação.
- **CG1822**, que participa da segregação cromossômica na meiose de células femininas.
- E alguns ainda com função desconhecida **CG7203**, **CG8064**, **CG3153**, **CG30273**, **CG6878**, **CG9894** e **CG6503**.

### Clonagem e Sequenciamento

Uma vez que os produtos foram amplificados de *pools* de indivíduos de cada espécie, é necessário fazer a clonagem destes produtos para a identificação de fragmentos de sequências distintas em cada espécie. Tal estratégia permite-nos fazer de forma simples e com apenas uma clonagem, uma avaliação global da variabilidade intra e interespecífica para cada um dos genes aqui investigados. Utilizamos os produtos de PCR purificados em reações de ligação com plasmídeos PTZ57R/T usando o Kit de

clonagem InsTAclone (Fermentas). Utilizamos o produto desta ligação na transformação de bactérias competentes da cepa DH5 $\alpha$ . Realizamos a transformação utilizando a metodologia de choque térmico de acordo com procedimentos padrão utilizando 100ng de plasmídeo por reação. Incubamos os tubos em gelo por 30 minutos e depois em banho-maria a 42°C por 2 minutos e novamente em gelo por 5 minutos. Adicionamos aos tubos 500 $\mu$ L de meio LB líquido e deixamos as bactérias crescerem por 1 hora em banho-maria a 37°C. Dividimos em 4 placas de Petri, com meio de cultura seletivo (com adição de ampicilina-concentração:0,1 $\mu$ g/ $\mu$ L), os 500  $\mu$ L de meio LB líquido com as colônias recombinantes já crescidas e incubamos as placas a 37°C por até 18 horas.

Amostramos ao menos 20 colônias por clonagem e avaliamos a presença do inserto nas colônias recombinantes por PCR direta de colônia (Dallas-Yang *et al.*, 1998) com os *primers* M13 *forward* e M13 *reverse* que amplificam regiões flanqueadoras da região dos insertos no vetor.

Purificamos os insertos que apresentaram o tamanho de fragmento esperado novamente através de precipitação diferencial com PEG (Polietileno Glicol) 8000 (Lis e Schleif, 1975) e ressuspendemos os produtos em 26  $\mu$ l de água ultrapura os deixando prontos para serem sequenciados. Enviamos nossas amostras para a empresa MAGROGEN INC. na Coréia do Sul, onde tais fragmentos foram sequenciados em um sequenciador ABI 3730 e os resultados disponibilizados eletronicamente.

### 3.4 - Análises das sequências de DNA

Utilizamos o software Chromas 2.31 (<http://en.bio-soft.net/dna/chromas.html>) que permite a visualização dos eletroferogramas e inspeção visual da qualidade do sequenciamento e, em seguida, a ferramenta BLAST no GENBANK (<http://www.ncbi.nlm.nih.gov>), para conferir se as sequências correspondiam a nossa região de interesse, e as alinhamos no programa BioEdit (Hall,1999) através do uso do software Clustal W (Thompson *et al.*, 1994)

Através do programa DNAsp v.5.00.03(Rozas *et al.*, 2003) investigamos o nível de polimorfismo de cada uma das regiões amplificadas pela estimativa do número de sítios polimórficos e do índice de diversidade de nucleotídeos ( $\pi$ ) (Nei,1987), através da comparação par a par do número de nucleotídeos diferentes por sítio entre todas as sequências estudadas e a média entre todas as combinações dentro das espécies estudadas. Realizamos essas análises utilizando as sequências das três espécies juntas e

separadas, além fazermos as análises da região intrônica dos *loci* que amplificaram tanto *intron* como *exon*. Pelo DNAsp também realizamos testes de neutralidade seletiva como D de Tajima (Tajima, 1989b), F e D de Fu e Li (Fu & Li, 1993), Fs de Fu (Fu, 1997) e H de Fay&Wu (Fay & Wu, 2000). Para os testes de neutralidade, quando um grupo externo era necessário, usamos sequências obtidas no GENBANK (<http://www.ncbi.nlm.nih.gov>) de *Drosophila melanogaster*, *D. virilis*, *D. willistoni*, *D. grimshawi* e *D. mojavensis* e quando disponíveis sequências de Tephritidae como *Rhagoletis pomonella*. Em função do número de testes realizados utilizamos a correção de Bonferroni para testes múltiplos. E para compararmos os resultados das análises das regiões de *exon* e das regiões intrônicas, dos *loci* que as possuíam, realizamos o Teste Z.

Como a presença de recombinação nas sequências das regiões estudadas interfere tanto nas inferências filogenéticas quanto na capacidade de detecção de seleção, utilizamos para detectar eventos de recombinação o programa RDP 3 Beta 42 (<http://www.bioinf.manchester.ac.uk/recombination/programs.html>). Este *software* compara as sequências três a três, determinando duas mais proximamente relacionadas. Através dessa comparação, é calculada uma porcentagem de identidade nas diferentes posições das sequências, esperando-se encontrar maior proximidade entre as duas sequências que apresentam maior relação. Se em uma determinada região encontrarmos uma maior porcentagem de identidade entre duas sequências que apresentam menor relação é possível que esteja ocorrendo, nessa região, recombinação entre essas duas sequências.

Para a obtenção das redes de haplótipos que representam as relações genealógicas entre as sequências obtidas para as espécies estudadas, utilizamos o programa computacional TCS versão 1.21 (Clement *et al.*, 2000). O TCS aplica a estatística de parcimônia pelo algoritmo desenvolvido por Templeton (1992). Os “*loops*” presentes na rede foram resolvidos utilizando critérios baseados na teoria da coalescência (Crandall & Templeton, 1993) e podem ser resumidos da seguinte maneira: haplótipos são mais prováveis de estarem conectados com haplótipos de maior frequência do que com haplótipos raros e, haplótipos são mais prováveis de estarem conectados com haplótipos do interior do que com haplótipos das extremidades.

Para estimarmos o modelo evolutivo, de cada uma das regiões estudadas, utilizamos o modelo de comparação MODELTEST (Posada & Crandall, 1998) através do programa HYPHY\_win32 (Pond *et al.*, 2005).

Para a análise filogenética usamos, novamente, as sequências de *Drosophila melanogaster*, *D. mojavensis*, *D. virilis*, *D. willistoni*, *D. grimshawi* e *Rhagoletis pomonella* obtidas do GenBank, como grupos externos. Alinhamos as sequências de nucleotídeos e usamos o programa MEGA 4 (Tamura *et al.*, 2007) para inferir relações filogenéticas através de Neighbor-Joining e as utilizamos como base para inferir as árvores filogenéticas por Máxima Verossimilhança com as distâncias baseadas em mudanças nucleotídicas por sítio no *software* PhyML v.3.0 (Guindon & Gascuel, 2003).

Realizamos os testes de seleção positiva em dados macroevolutivos através de uma metodologia que utiliza o CODEML pelo programa PAML ver. 4 (Yang, 2007). O PAML ver. 4 nos permitiu investigar as razões entre as taxas de substituição sinônima/não-sinônima (dN/dS, também denominado  $\omega$ ). Quando as substituições não-sinônimas têm taxas de fixação mais altas do que as mutações sinônimas, há a indicação de seleção positiva ( $\omega > 1$ ). Por outro lado,  $\omega = 1$  indica neutralidade, enquanto um  $\omega < 1$  indica seleção purificadora.

Como as análises no CODEML (PAML) exigem uma árvore filogenética por Máxima Verossimilhança com as distâncias entre os ramos baseadas em substituições nucleotídicas por códon, rodamos o modelo M0 no PAML utilizando a árvore de máxima verossimilhança obtida no programa PhyML v. 3.0 (para obtermos a árvore necessária para rodar os outros modelos. Para a realização do *Relaxed Branch-site test* (Teste relaxado de sítios e ramos) e *Strict Branch-site test* (Teste Estrito de sítios e ramos) (Yang & Nielsen 2002; Zhang *et al.* 2005) consideramos o ramo das *Anastrepha* como o ramo *foreground*, no qual testamos a ocorrência de seleção positiva, e depois o contrastamos com o ramo que denominamos de *background*, contendo as espécies de *Drosophila* e, eventualmente, *Rhagoletis*. Através de um Teste de Razão de Verossimilhanças (LTR – *log-likelihood ratio test*) contrastamos o modelo alternativo, que é para ambos os testes o modelo MA, no qual apenas o ramo *foreground* pode conter sítios com  $\omega > 1$ , com os modelos nulos. O primeiro teste consiste em contrastar o modelo MA contra o modelo M1A (modelo nulo), que admite a mesma taxa evolutiva para todos os sítios e ramos, com variação de  $\omega$ , para todos os sítios, entre 0 e 1. Para o segundo teste, o modelo nulo (modelo MA restrito), no qual os valores de  $\omega$  para cada sítio são fixados a um valor máximo de 1, é contrastado contra o modelo MA. Para o Teste de Razão de Verossimilhança (LTR) para o *Relaxed Branch-site test* usamos um grau de liberdade igual a 1 e para o *Strict Branch-site test* o grau de liberdade que

utilizamos para os cálculos foi 2. Pelas análises do PAML obtivemos, também, os resultados das análise Bayesiana para detecção de sítios com sinal de seleção positiva Bayes Empirical Bayes (BEB) (Yang *et al*, 2005).

Para obtermos estimativas bayesianas de máxima verossimilhança de tamanho populacional, tempo de divergência ( $t = t\mu_i$ ) e taxas de mutação e migração usamos o programa IMa2 (Hey, 2009), que utiliza um método de coalescência Bayesiana, que aproxima a integração das genealogias possíveis, através de uma simulação da Cadeia de Markov Monte Carlo (*Markov Chain Monte Carlo* – MCMC). Este teste assume que os genes estejam evoluindo por seleção neutra. Ao contrário dos programas IM (Hey e Nielsen, 2004) e IMa (Nielsen & Wakeley, 2001; Hey & Nielsen, 2004; Hey & Nielsen, 2007) , com o IMa2 podemos realizar as análises para mais de duas populações, para isto é necessário que o usuário especifique uma árvore filogenética de máxima verossimilhança enraizada, que contenha também a informação das distâncias entre os ramos.

Para as análises no IMa2 usamos o modelo de Hasegawa-Kishino-Yano (HKY), com *burn-in* de 1.500.000 genealogias e amostragem a cada 10.000 genealogias. Para termos uma estimativa do tamanho efetivo populacional por indivíduos para cada espécie ( $N_e$ ) e o tempo de separação da população em anos, calculamos  $\mu$  comparando as espécies de *Drosophila* entre si, para as quais já sabíamos o tempo de divergência, (Tamura *et al*, 2004) e geramos um valor médio para cada um dos genes aqui utilizados chegando à um tempo de divergência estimado por locus para as espécies estudadas. Utilizamos para as análises do IMa2 as três genealogias possíveis em que as espécies são consideradas juntas, além de também termos realizado as análises utilizando as espécies duas a duas..

Fizemos a comparação entre a árvore de espécies e a árvore de genes usando o programa MESQUITE versão 2.5 (Maddison & Maddison, 2008). Especificamente, avaliamos o ajuste das árvores dos genes em três possíveis árvores enraizadas para as espécies estudadas, considerando cada uma das espécies aqui estudadas como sendo externa às outras duas, ou seja, testamos três topologias distintas considerando *A. fraterculus*, *A. obliqua* e *A. sororcula* como sendo a espécie mais distante das outras duas espécies.

Realizamos as análises usando o método de minimização da coalescência profunda (*deep coalescence*) (Maddison, 1997) implementado pelo programa

MESQUITE. Este método consiste em conciliar a árvore de genes dentro da árvore de espécies de modo a minimizar a número de coalescências profundas pelo critério da parcimônia, a "melhor" árvore de espécies inferida é a que requer o menor número de coalescências profundas.

Também fizemos análises comparando a coalescência profunda das árvores de genes reais com 100 árvores de genes simuladas para cada um dos *loci* estudados nos três modelos de árvores de espécies, nas quais em cada modelo uma das espécies se encontra mais separada das outras duas. Usando o programa MS (Hudson, 2002) foram simuladas 100 árvores com 3 populações isoladas e sem migração, evoluindo segundo modelo neutro de Wright-Fisher, assumindo modelo de mutação de sítios infinitos. Foram simuladas árvores representando as três topologias possíveis para as três espécies do estudo (*A. fraterculus*(*A. obliqua*, *A. sororcula*); (*A. obliqua* (*A. fraterculus*, *A. sororcula*) e (*A. sororcula*(*A. obliqua*, *A. fraterculus*,). Os tempos de divergência entre as espécies (em unidades de coalescência) foram determinados de acordo com os tempos e tamanhos populacionais obtidos no programa Ima2 para as respectivas topologias.

Após a obtenção das 100 simulações de árvores para cada topologia, transformamos essas árvores em sequências de genes evoluindo de acordo com modelo de substituição nucleotídica específico para cada um dos 12 genes, que foram HKY ou REV, conforme determinado no programa MODELSTEST versão 3.0 (Posada & Crandall, 1998) e tamanho do segmento de DNA sequenciado, no programa Seq-Gen versão 2.0 (Strope *et al*, 2009). Estes dados gerados no Seq-gen foram rodados no programa PhyML 3.0 (Guindon & Gascuel, 2003). para gerar as melhores topologias por máxima verossimilhança, usando os modelos de substituição de nucleotídeos para cada gene, HKY85 ou GTR, de acordo com os resultados do MODELTEST . Estas topologias foram então utilizadas no MESQUITE para compararmos seus valores de coalescência profunda com as árvores de genes originalmente obtidas. Este procedimento busca criar uma distribuição de valores de coalescência profunda considerando um determinado modelo evolutivo (por exemplo, uma separação de *A. fraterculus* antes de *A. sororcula* e *A. obliqua*, sem migração entre elas). Dessa forma, os valores originalmente obtidos para cada uma de nossas árvores quando contrastados com esta distribuição, podem ser considerados como teste de adequação desta topologia ao modelo evolutivo específico sendo testado.

#### 4. Resultados e Discussão

Testamos os 27 conjuntos nas três espécies estudadas utilizando o *pool* de DNA de indivíduos de regiões diferentes do Brasil. Após diversos testes de temperatura e concentração de *primers*, os *primers* **CG7009**, **CG7203**, **CG8064**, **CG5325**, **Elongation Protein**, **Cyclophilin**, **Porin**, **Myosin Light Chain**, **Translationally Controlled Tumor Protein**, **Ubiquinol Cytochrome Reductase**, **Wings Up**, **Troponin C**, **CG6503**, **Ubiquitin** e **Copper Transporter** amplificaram ao menos uma banda consistentemente (**Tabela 3**).

**Tabela 3-** *Primers* testados.

<i>Loci</i>	Tamanho esperado (pb)	TM (°C)	T hibridação (°C)	[Mg <sup>2+</sup> ] (Mm)
<b>TCTP</b>	396	50	45	2,5
<b>Cyclophilin</b>	175	52,9	48	2,5
<b>Troponin C</b>	200	52,1	46	2,5
<b>UBCRX</b>	260	56,8	49	2,5
<b>Elongation Protein</b>	366	51,8	47	1,5
<b>Porin</b>	441	56,2	51	2,5
<b>MLC</b>	313	49	44	2,5
<b>CG8064</b>	339	52	48	1,5
<b>Wings up</b>	920	56,4	51	1,5
<b>CG7009</b>	376	52,5	54	2,5
<b>CG5325</b>	381	52,2	47	1,5
<b>CG7203</b>	393	61,3	55	2,5
<b>Copper Transporter</b>	200	52,4	45	2,5
<b>Ubiquitina</b>	400	51	45	2,5
<b>CG6503</b>	450	56,2	48	1,5
<b>Nimrod*</b>	250	56,2	-	-
<b>α-trypsin*</b>	450	64,2	-	-
<b>C-type lectin*</b>	300	60,9	-	-
<b>Vitelline*</b>	150	58	-	-
<b>CG 3751*</b>	370	58,9	-	-
<b>CG 10306*</b>	210	58,1	-	-
<b>CG9894*</b>	310	50,2	-	-
<b>CG1822*</b>	360	48	-	-
<b>Sniffer*</b>	560	49	-	-
<b>CG6878*</b>	300	60,6	-	-
<b>CG3153*</b>	375	58,8	-	-
<b>CG30273*</b>	380	60,3	-	-

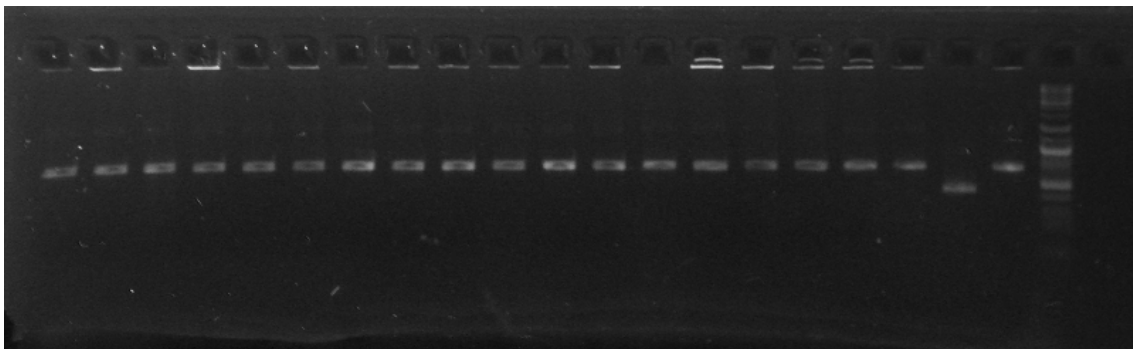
\* não amplificou

Após repetições, alguns destes *primers*, **CG6503**, **Ubiquitin** e **Copper Transporter**, começaram a amplificar inconsistentemente e não foram estudados. Os



*primers* que amplificavam fragmentos dos genes **CG7009**, **CG7203**, **CG8064**, **CG5325**, **Elongation Protein**, **Cyclophilin**, **Porin**, **Myosin Light Chain**, **Translationally Controlled Tumor Protein**, **Ubiquinol Cytochrome Reductase**, **Wings Up** e **Troponin C** foram amplificados em *pools* de cada uma das espécies, *A. fraterculus*, *A. obliqua* e *A. sororcula* e seus produtos clonados e sequenciados.

Ao menos vinte colônias produzidas foram amplificadas com primers M13 *forward* e *reverse* (Exemplo que pode ser conferido na **Figura 3**) e presença de inserto no tamanho esperado foi verificada pela corrida em gel de agarose 1%, utilizando um marcador de peso molecular (**Tabela 3**).



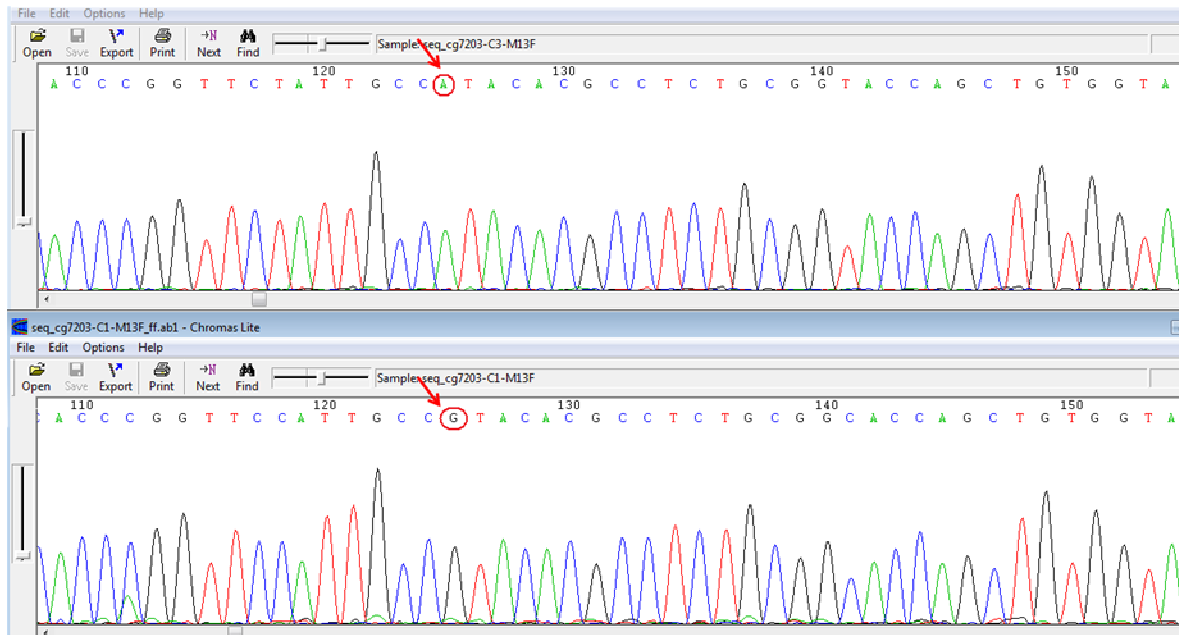
**Figura. 3** Amplificação dos clones, com o *primer* M13, do gene **Porin**, sendo que a banda de tamanho menor representa uma colônia negativa.

Determinamos os tamanhos dos produtos e quantificamos as amostras que possuíam bandas do tamanho esperado para o conjunto de *primers* (**Tabela 3**). As amostras quantificadas tinham entre 50 e 100 ng/ $\mu$ l. Este material foi purificado e enviado para ser sequenciado na empresa MACROGEN INC. na Coréia do Sul.

A qualidade das sequências *forward* e *reverse* obtidas (exemplo que pode ser conferido na **Figura 4**) foi verificada no programa Chromas 2.31 (<http://en.biosoft.net/dna/chromas.html>) e usadas para a criação de sequências consenso. Estas sequências foram, então, conferidas no GENBANK (<http://www.ncbi.nlm.nih.gov>) através da ferramenta BLAST e uma vez que correspondessem às regiões amplificadas esperadas eram alinhadas no programa Bioedit (<http://www.mbio.ncsu.edu/BioEdit/BioEdit.html>), utilizando a ferramenta CLUSTAL W.

A partir do alinhamento das sequências *consenso*, conseguimos identificar polimorfismos nas regiões amplificadas de cada um dos 12 genes tanto dentro quanto entre as três espécies de *Anastrepha* estudadas neste trabalho. A **Figura 4** representa um

dos polimorfismos presentes na região amplificada do gene **CG7203**, onde a mudança de base levou a uma substituição de aminoácido.



**Figura 4-** Eletroferograma mostrando um polimorfismo no gene CG7203, causando também uma mudança de aminoácido de Valina para Isoleucina.

Como a presença de recombinação interfere nas análises filogenéticas, porque mistura sinais históricos de parentesco, utilizamos o programa RDP 3 Beta 42 para detectar a presença de regiões recombinantes, calculando uma porcentagem de identidade nas diferentes posições das sequências, esperando-se encontrar maior proximidade entre as duas sequências que apresentam maior relação, e não encontramos nenhum sinal de recombinação para as regiões amplificadas de cada um dos genes estudados neste trabalho.

#### **4.1 - Análises de Polimorfismos, Divergência e Testes de Neutralidade.**

Através do programa DnaSP v. 5.00.03 (Rozas, *et al.*, 2003) analisamos o nível de polimorfismo e divergência para cada região gênica aqui estudada para *A. fraterculus*, *A. obliqua*, *A. sororcula* juntas e separada por espécie. A porcentagem média de nucleotídeos diferentes por sítio entre sequências define um parâmetro importante de diversidade molecular, a diversidade nucleotídica. Este parâmetro, bem como o número de sítios polimórficos e os valores de *theta* (indicativo da quantidade de polimorfismos mantido na população pelo balanço entre deriva genética e mutação) por sequência foram utilizados para determinar o grau de polimorfismo dentro e entre as espécies.

A **Tabela 4** indica os índices de polimorfismo e testes de neutralidade para cada região estudada considerando o gênero *Anastrepha* como um todo e as **Tabelas 5, 6, 7** mostram as mesmas análises feitas para as espécies separadas, respectivamente, *A.fraterculus*, *A. obliqua* e *A. sororcula*.

Para as análises realizadas com as três espécies de *Anastrepha* juntas (**Tabela 4**), os resultados dos índices de polimorfismo mostram que, a maioria das regiões gênicas amplificadas neste trabalho apresenta altos valores de *theta*, índice de diversidade nucleotídica ( $\pi$ ) e número de sítios polimórficos (S), por exemplo, a região amplificada pelo *primer* **UQCRX** ( $\pi = 0,034$ ), quando comparados com outras espécies estudadas (por exemplo, apresentam valores compatíveis ou mais altos a dados mitocondriais (ver Powell, 1997), que geralmente são mais altos do que regiões nucleares. Para o gênero *Anastrepha* como um todo, os testes D de Tajima e D e F de Fu e Li não apresentaram desvios significantes da neutralidade para nenhum dos genes aqui estudados após a correção de Bonferroni. O teste de H de Fay e Wu foi significativo para as regiões amplificadas pelos *primers* **CG7009**, **CG5325** e **MLC**, enquanto o Fs de Fu apresentou resultados significativos para a maioria dos genes aqui estudados, com exceção para os *loci* **Wings Up** e **Cyclophylin** (**Tabela 4**).

Quando fizemos os testes utilizando apenas as sequências de *A. fraterculus* (**Tabela 5**) os índices de polimorfismos continuaram altos, como, por exemplo, o índice de diversidade nucleotídica ( $\pi$ ) para o gene **UQCRX** ( $\pi = 0,03$ ), maiores que os valores encontrados para dados mitocondriais de *Drosophila buzzatii* ( $\pi$  entre 0,00039 e 0,0082) (de Brito *et al.*, 2002). Para estes testes *A. fraterculus* foi a que apresentou os maiores valores de polimorfismos quando comparada com *A. obliqua* e *A. sororcula* analisadas separadamente. Os valores de F e D de Fu e Li e D de Tajima não apresentaram valores significantes para nenhuma das regiões gênicas amplificadas. O teste de H de Fay e Wu para *A. fraterculus*, não apresentou resultados significativos para nenhum dos genes e os valores para Fs de Fu foram significativos para os genes **CG7009**, **TCTP**, **Troponin C**, **Porin** e **ELp**.

Pelas análises feitas com as sequências de *A. obliqua* (**Tabela 6**) podemos observar um índice de diversidade nucleotídica ( $\pi$ ) alto, por exemplo, como o encontrado para o gene **UQCRX** ( $\pi = 0,056$ ), maiores por exemplo, que os valores encontrados para *Gryllus rubens* e *G. texensis* (Gray *et al.*, 2008). Não encontramos resultados significativos para o Teste de F e D de Fu e Li e para o teste D de Tajima. O

teste de H de Fay e Wu para *A. obliqua*, foi significativo apenas para a região gênica amplificada pelo *primer* **CG7009**. Os valores para o teste  $F_s$  de  $F_u$  foram significativos para a maioria dos genes com exceção dos *loci* **Wings Up** e **Cyclophylin** (Tabela 6). Apesar dos valores de  $F_s$  de  $F_u$ , para *A. fraterculus* e para o gênero *Anastrepha*, se mostrarem significativos para as mesmas regiões gênicas, os resultados, quando agrupamos as três espécies, apresentaram valores mais extremos de desvio de neutralidade.

Para as sequências de *A. sororcula*, os índices de diversidade nucleotídica ( $\pi$ ) foram altos para algumas regiões gênicas, por exemplo, para o gene **UQCRX** ( $\pi = 0,019$ ), não encontramos, para nenhuma das regiões amplificadas, valores significativos para os testes de F e D de  $F_u$  e Li e para o teste D de Tajima. O teste de H de Fay e Wu para *A. sororcula*, não foram significativos para nenhum dos genes e o teste de  $F_s$  de  $F_u$  foi significativo apenas para as regiões amplificadas pelos *primers* **CG7009**, **CG8064**, **MLC** e **Porin** (Tabela 7).

A Tabela 8 apresenta os testes de neutralidade e níveis de polimorfismo para as regiões intrônicas dos genes amplificados, considerando o grupo *fraterculus* como um todo para a análise. Estes testes nos mostram níveis de polimorfismo mais altos e testes de neutralidade menos significativos para a região do *íntron* quando comparados com a região *exônica*. O nível de polimorfismo das regiões exônicas é bem mais baixo do que regiões não codificadoras. Por exemplo, na região amplificada pelo *primer* **Elongation Protein** encontramos 26 substituições em cerca de 300 pares de bases presentes em um *exon*, contra 24 em cerca de 60 pares de bases intrônicas. Apenas as regiões amplificadas pelos genes **CG7009** e **CG5325** apresentaram uma maior quantidade de alterações na região do *éxon*. Os valores de  $\pi$ , no entanto, não apresentaram diferença significativa entre *exon* e *íntron* pelo Teste Z feito para cada um dos *locus*. O Teste D de Tajima não foi significativo para nenhum gene. Os resultados para o teste de  $F_s$  de  $F_u$  foram significativos para os genes **CG7009**, **CG5325**, **UQCRX** e **Wings Up** (Tabela 8), embora deva se ressaltar que os outros genes apresentaram íntrons com tamanho reduzido, o que diminui o grau de liberdade dos testes aqui realizados.

O teste de  $F_s$  de  $F_u$  contrasta a probabilidade de se encontrar um número igual ou superior de alelos em comparação com o esperado em uma população estacionária (Fu, 1997) e apresenta maior poder estatístico para detectar expansão populacional do que outros aqui testados (Ramos-Onsins e Rozas, 2002). Como o padrão geral dos testes de neutralidade aqui realizados foi a detecção de valores significativos de  $F_s$  para todos

os genes investigados, inclusive em seus íntrons, isto indicaria a existência de uma expansão populacional afetando as populações estudadas, mais particularmente em *A. obliqua* que também apresentou valores significativos de  $F_s$  para todos seus genes, quando analisada individualmente, o que não ocorreu com as outras espécies, nas quais apenas parte dos genes indicou testes significativamente negativos para  $F_s$ .

**Tabela 4** - Índices de polimorfismo e testes de neutralidade para cada gene com todas as espécies de *Anastrepha*.

Locus	N	Tamanho (pBase)	$\pi$	S	$\theta$	Fu &Li's D <sup>#</sup>	Fu &Li's F <sup>#</sup>	Fay&Wu's H <sup>#</sup>	Fu's $F_s$	Tajima's D
CG7009	60	321	0,024	49	0,025	-1,62	-1,77	-24,64*	-15,97*	-0,97
CG7203	60	285	0,019	46	0,019	-2,37	-2,57	-8,27	-37,68*	-1,70
CG8064	60	339	0,010	36	0,010	-3,14	-3,19	-5,55	-35,89*	-1,93
CG5325	44	220	0,019	38	0,019	-1,65	-2,13	-10,98*	-23,58*	-1,88
MLC	60	266	0,014	39	0,014	-2,66	-2,93	-14,99*	-43,41*	-2,02
TCTP	58	396	0,009	40	0,009	-5,74	-5,24	-1,07	-26,91*	-2,02
Troponin C	60	141	0,006	15	0,006	-3,02	-3,32	-4,68	-18,79*	-2,33
UQCRX	57	132	0,034	50	0,035	-3,20	-3,54	-8,92	-28,76*	-2,36
Porin	59	345	0,006	24	0,006	-2,90	-3,06	-4,51	-14,88*	-1,97
Wings Up	37	290	0,022	25	0,022	-4,15	-4,23	1,81	-3,48	-2,37
Elp	51	300	0,017	26	0,017	-2,33	-2,50	-3,89	-24,90*	-1,79
Cyclophylin	49	175	0,094	49	0,110	-0,62	-0,60	0,84	-5,95	-0,31

\* $P < 0,05$  considerando-se a significância corrigida para Bonferroni (equivalente a  $P < 0,004$ )\*.

N = número de sequências utilizadas para a análise.

# Testes feitos utilizando como grupo externo *D. melanogaster*, *D. virilis*, *D. willistoni*, *D. mojavensis* e *D. grimshawi* ou *Rhagoletis pomonella*.

**Tabela 5** - Índices de polimorfismo e testes de neutralidade de cada gene para *A. fraterculus*.

Locus	N	$\pi$	S	$\theta$	Fu &Li's D <sup>#</sup>	Fu &Li's F <sup>#</sup>	Fay&Wu's H <sup>#</sup>	Fu's $F_s$	Tajima's D
CG7009	20	0,01	17	0,009	-1,24	-1,59	-7,55	-6,63*	-1,36
CG7203	20	0,015	19	0,015	0,03	-0,25	-4,45	-4,64	-0,78
CG8064	20	0,011	17	0,011	-0,31	-0,67	-4,68	-4,35	-0,99
CG5325	20	0,008	13	0,008	-0,88	-1,36	-3,41	-3,72	-1,82
MLC	20	0,016	16	0,016	-0,95	-0,96	-2,71	-4,77	-0,5
TCTP	19	0,009	20	0,009	-3,14	-3,16	0,94	-7,99*	-1,51
Troponin C	20	0,008	7	0,008	-1,42	-1,79	-2,51	-5,99*	-1,66
UQCRX	20	0,03	30	0,031	-2,79	-3,19	-5,11	-2,85	-2,23
Porin	20	0,003	8	0,003	-1,42	-1,83	-4,45	-5,12*	-1,91
Wings Up	14	0,013	17	0,013	-3,79	-4,03	2,11	0,56	-2,25
Elp	18	0,021	19	0,021	-0,87	-1,24	-2,8	-13,23*	-1,54
Cyclophylin	18	0,07	23	0,077	0,84	0,95	1,49	1,37	0,62

\* $P < 0,05$  considerando-se a significância corrigida para Bonferroni (equivalente a  $P < 0,004$ )\*.

N = número de sequências utilizadas para a análise

# Testes feitos utilizando como grupo externo *D. melanogaster*, *D. virilis*, *D. willistoni*, *D. mojavensis* e *D. grimshawi* ou *Rhagoletis pomonella*.

**Tabela 6 - Índices de polimorfismo e testes de neutralidade de cada gene para *A. obliqua*.**

<i>Locus</i>	N	$\pi$	S	$\theta$	Fu & Li's D <sup>#</sup>	Fu & Li's F <sup>#</sup>	Fay&Wu's H <sup>#</sup>	Fu's Fs	Tajima's D
CG7009	20	0,006	16	0,006	-0,15	-0,85	-13,85*	-7,76*	-2,06
CG7203	20	0,02	27	0,021	-1,29	-1,41	0,08	-14,31*	-1,06
CG8064	20	0,01	14	0,01	-1,42	-1,30	-1,07	-7,87*	-0,63
CG5325	15	0,022	23	0,023	-1,46	-1,75	-3,46	-12,34*	-1,33
MLC	20	0,012	18	0,011	-1,53	-1,82	-3,79	-12,6*	-1,5
TCTP	19	0,005	12	0,005	-2,14	-2,33	-0,55	-6,21*	-1,46
Troponin C	20	0,008	7	0,007	-0,72	-1,14	-0,73	-4,77*	-1,52
UQCRX	18	0,056	35	0,06	-1,85	-2,23	-3,19	-10,47*	-1,69
Porin	19	0,007	12	0,007	-1,43	-1,75	-4,37	-7,13*	-1,17
Wings Up	18	0,028	17	0,029	-2,85	-3,09	2,17	0,27	-1,94
Elp	19	0,013	20	0,013	-0,98	-1,32	-3,24	-6,94*	-1,4
Cyclophylin	10	0,092	29	0,104	0,80	1,01	5,07	1,38	0,81

\*P< 0,05 considerando-se a significância corrigida para Bonferroni (equivalente a P< 0,004) \*.

N = número de sequências utilizadas para a análise

# Testes feitos utilizando como grupo externo *D. melanogaster*, *D. virilis*, *D. willistoni*, *D. mojavensis* e *D. grimshawi* ou *Rhagoletis pomonella*.

**Tabela 7- Índices de polimorfismo e testes de neutralidade de cada gene para *A. sororcula*.**

<i>Locus</i>	N	$\pi$	S	$\theta$	Fu & Li's D <sup>#</sup>	Fu & Li's F <sup>#</sup>	Fay&Wu's H	Fu's Fs	Tajima's D
CG7009	20	0,006	15	0,006	-2,56	-2,88	-4,86	-8,21*	-2,05
CG7203	20	0,011	11	0,011	-1,72	-1,86	-0,04	-4,78	-0,65
CG8064	20	0,004	12	0,003	-3,02	-3,30	-1,14	-7,47*	-2,31
CG5325	09	0,021	12	0,021	0,34	0,29	-1,64	-3,03	0,1
MLC	20	0,012	15	0,012	-0,15	-0,54	-5,62	-6,44*	-0,95
TCTP	20	0,003	10	0,003	-3,20	-3,43	0,93	-4,28	-2,03
Troponin C	20	0,002	3	0,002	-2,58	-2,75	0,28	-1,14	-1,72
UQCRX	19	0,019	13	0,019	-0,69	-1,11	-2,58	-3,97	-1,32
Porin	20	0,008	17	0,008	-2,76	-2,94	-0,13	-9,21*	-1,73
Wings Up	05	0,054	33	0,058	-2,13	-2,30	7,8	0,04	-1,15
Elp	15	0,007	8	0,007	-1,17	-1,25	0,62	-3,03	-0,51
Cyclophylin	21	0,021	22	0,021	-1,92	-2,22	-4,99	-6,6	-1,63

\*P< 0,05 considerando-se a significância corrigida para Bonferroni (equivalente a P< 0,004) \*.

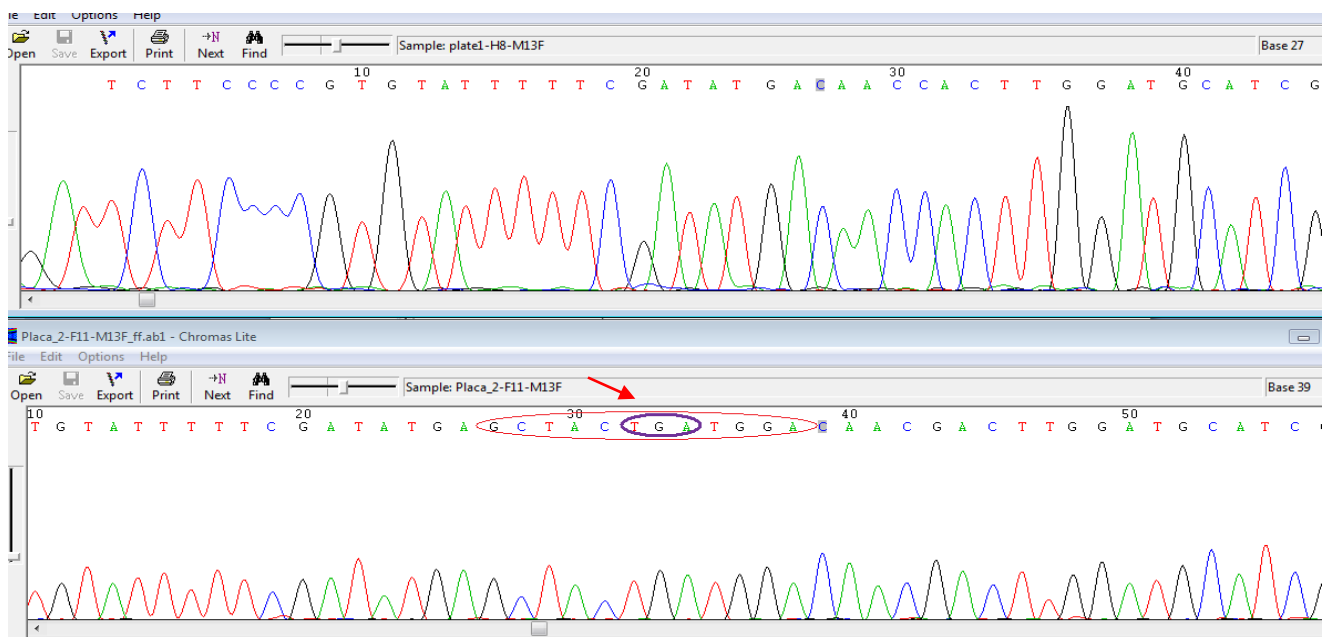
N = número de sequências utilizadas para a análise # Testes feitos utilizando como grupo externo *D. melanogaster*, *D. virilis*, *D. willistoni*, *D. mojavensis* e *D. grimshawi* ou *Rhagoletis pomonella*.

**Tabela 8** - Índices de polimorfismo e testes de neutralidade de cada gene para a região intrônica.

Locus	Tamanho(pb)	$\pi$	S	$\theta$	Fu's Fs	Tajima's D
CG7009	55	0,023	6	0,023	-2,84	-0,75
CG5325	159	0,01	21	0,01	-16,85*	-2,29
UQCRX	133	0,048	53	0,051	-28,11*	-1,85
Porin	96	0,034	22	0,036	-4,92	-1,04
Wings Up	634	0,025	75	0,026	-23,42*	-1,86
Elp	62	0,079	24	0,089	-5,28	-0,46

\*P < 0,05 considerando-se a significância corrigida para Bonferroni (equivalente a P < 0,008) \*.  
N = número de sequências utilizadas para a análise

O alinhamento das sequências obtidas para a região amplificada do gene **Cyclophylin** das três espécies de *Anastrepha* do grupo *fraterculus* revelou a presença de alguns códons de terminação e de alguns indels que mudam a fase da leitura (**Figura 5**). Tais resultados sugerem a possibilidade desta região amplificada corresponder a um *pseudogene*, que são regiões do genoma que apresentam sequência similar ao gene funcional, por terem se originado de duplicações recentes destes genes, porém com algumas mudanças de nucleotídeos que impedem sua expressão. Os pseudogenes seguem em geral um modelo de evolução neutra, porque na grande maioria das vezes qualquer nova mutação nestes genes é determinada por deriva genética sob a hipótese de mutação neutra (Li *et al.*, 1981).



**Figura 5** – Sequenciamento do gene **Cyclophylin** exemplificando a presença de códon de terminação nas sequências de *A. sororcula* (abaixo) e indel nas sequências de *A. fraterculus* (acima).

Analizamos também o número de sítios polimórficos existentes nas espécies e entre as espécies, comparando em *A. fraterculus* e *A. obliqua* (**Tabela 9**); *A. obliqua* e *A. sororcula* (**Tabela 10**) e *A. fraterculus* e *A. sororcula* (**Tabela 11**), o número de polimorfismos específicos para cada espécie, comum a ambas ou fixados entre as mesmas.

A comparação entre *A. fraterculus* e *A. obliqua* revelou que a média de sítios polimórficos exclusivos para cada uma, respectivamente, foi de 12 e 15 sítios. O *locus* que apresentou o maior número de sítios polimórficos compartilhados entre as espécies foi o *pseudogene Cyclophylin* com 33 sítios compartilhados, sendo que não observamos a presença de sítios polimórficos fixados (**Tabela 9**).

Quando comparamos *A. obliqua* e *A. sororcula* tivemos como resultado uma média de sítios polimórficos exclusivos para cada uma das espécies, respectivamente, de 17 e oito sítios. Observamos na região amplificada pelo *primer Porin* 8 sítios polimórficos compartilhados entre as espécies, e este foi o *locus* que apresentou o maior valor nesta comparação. O gene **CG7009** apresentou 11 sítios polimórficos fixados, sendo este o maior valor dentre as três comparações feitas para este tipo de substituição (**Tabela 10**).

A comparação entre as espécies *A. fraterculus* e *A. sororcula* nos mostrou, respectivamente, uma média de 16 e 11 sítios polimórficos que ocorrem exclusivamente para cada uma das espécies. O *locus* que apresentou o maior número de sítios polimórficos compartilhados entre as espécies foi o **UQCRX** e os genes **CG7009**, **TCTP** e **Cyclophylin** apresentaram sítios polimórficos fixados, respectivamente, sete, uma e cinco substituições (**Tabela 11**).

Estes dados em conjunto sugerem uma maior proximidade evolutiva entre *A. fraterculus* e *A. obliqua*, revelada pela ausência de polimorfismos específicos para espécie que distingam tais espécies e uma maior presença de polimorfismos compartilhados. Por outro lado, *A. sororcula* apresentou diferenças fixadas com as outras duas espécies para três genes (**CG7009**, **TCTP** e **Cyclophylin**). Estes resultados devem ser considerados cuidadosamente, uma vez que também podem ser reflexo da amostragem geográfica de *A. sororcula* ser bem mais circunscrita do que a das outras duas espécies, o que facilita o encontro de polimorfismos que embora fixados na região amostrada não o seriam no resto da distribuição não amostrada da espécie. Deve-se



ressaltar ainda a possibilidade de *A. fraterculus* apresentar espécies crípticas que podem ter sido aqui amostradas e consideradas como uma única espécie.

**Tabela 9** – Sítios segregantes encontrados nas espécies e entre espécies por locus comparando *A. fraterculus* e *A. obliqua*.

<b><i>A. fraterculus</i> (1) x <i>A. obliqua</i> (2)</b>				
<b>Locus</b>	<b>Polimórfico para (1)</b>	<b>Polimórfico para (2)</b>	<b>Polimórfico para ambas as espécies</b>	<b>Diferenças fixadas</b>
<b>CG7009</b>	15	14	2	0
<b>CG7203</b>	15	24	4	0
<b>CG8064</b>	13	10	4	0
<b>CG5325</b>	7	16	6	0
<b>Elp</b>	14	5	6	0
<b>MLC</b>	14	15	3	0
<b>Porin</b>	6	9	2	0
<b>Troponin C</b>	7	6	1	0
<b>TCTP</b>	18	10	2	0
<b>Wings Up</b>	4	17	1	0
<b>UQCRX</b>	19	32	9	0
<b>Cyclophylin</b>	12	14	13	0

**Tabela 10** - Sítios segregantes encontrados nas espécies e entre espécies por locus comparando *A. obliqua* e *A. sororcula*.

<b><i>A. obliqua</i>(2)x <i>A. sororcula</i>(3)</b>				
<b>Locus</b>	<b>Polimórfico para (2)</b>	<b>Polimórfico para (3)</b>	<b>Polimórfico para ambas as espécies</b>	<b>Diferenças fixadas</b>
<b>CG7009</b>	14	13	2	11
<b>CG7203</b>	23	9	5	0
<b>CG8064</b>	13	13	1	0
<b>CG5325</b>	20	8	2	0
<b>Elp</b>	20	6	2	0
<b>MLC</b>	16	13	2	0
<b>Porin</b>	3	10	8	0
<b>Troponin C</b>	7	3	0	0
<b>TCTP</b>	12	10	0	2
<b>Wings Up</b>	16	5	2	0
<b>UQCRX</b>	33	5	7	0
<b>Cyclophylin</b>	26	12	1	3

**Tabela 11** – Sítios segregantes encontrados nas espécies e entre espécies por locus comparando *A. fraterculus* e *A. sororcula*.

<i>A. fraterculus</i> (1) x <i>A. sororcula</i> (3)				
Locus	Polimórfico para (1)	Polimórfico para (3)	Polimórfico para ambas as espécies	Diferenças fixadas
CG7009	16	14	1	7
CG7203	17	12	2	0
CG8064	16	13	1	0
CG5325	12	10	1	0
Elp	19	3	1	0
MLC	15	13	2	0
Porin	7	17	1	0
Troponin C	8	3	0	0
TCTP	20	10	0	1
Wings Up	13	17	3	0
UQCRX	25	9	4	0
Cyclophylin	24	12	1	5

#### 4.2 - Análises das redes de haplótipos

As redes de haplótipos inferidas no TCS 1.21 (Clemente *et al.*, 2000) para cada uma das regiões gênicas (**Figura 6** a **Figura17**) não foram significativamente diferentes das árvores filogenéticas de máxima verossimilhança inferidas no PHYML v. 3.0 (**dados não apresentados**) também feitas para cada um dos 12 genes. A relação entre os haplótipos estimada pelo programa TCS é feita através dos métodos de parcimônia estatística, que considera a distância entre os haplótipos e os limites da parcimônia para criar redes que conectam haplótipos de forma a minimizar a matriz de distância. Para esta análise, é relevante considerar que o padrão normalmente esperado em regiões que estão evoluindo neutramente de acordo com a coalescência, é que mutações mais antigas tendem a estar presentes em mais cópias na população, e conseqüentemente mais internas na rede de haplótipos, enquanto mutações mais novas tendem a estar mais no extremo da rede (Castelloe & Templeton, 1994).

Nas redes geradas pelo programa TCS os haplótipos são representados por elipses, sendo que o tamanho das elipses é proporcional ao número de indivíduos encontrados para cada haplótipo. Os círculos menores, que ligam os haplótipos identificados, correspondem aos haplótipos inferidos não amostrados e são classificados

como intermediários, e os haplótipos representados por retângulos correspondem aos haplótipos ancestrais inferidos por coalescência.

As redes haplotípicas de cada gene revelam que não conseguimos separar as espécies *A. fraterculus*, *A. obliqua* e *A. sororcula* para a grande maioria das regiões gênicas amplificadas neste trabalho. Dos 12 genes para os quais foram feitas as redes de haplótipos (**Figura 6 a Figura 17**), cinco deles, **CG7009 (Figura 6)**, **CG8064 (Figura 8)**, **Elp (Figura 10)**, **TCTP (Figura 13)** e **Cyclophylin (Figura 17)**, apresentaram uma separação dos haplótipos de *A. sororcula* dos outros dois haplótipos, de *A. fraterculus* e *A. obliqua*. Além disso, as redes feitas utilizando os genes **CG8064 (Figura 8)**, **Elp (Figura 10)**, **TCTP (Figura 13)** e **Cyclophylin (Figura 17)** mostram, inclusive, um agrupamento dos haplótipos de *A. fraterculus* e *A. obliqua*, sendo que, observamos uma separação com muitos passos mutacionais dos haplótipos agrupados para a rede do gene **Cyclophylin (Figura 17)**. Dessas redes, nas quais houve algum agrupamento dos haplótipos das espécies, a maioria apresentou uma proporção grande de mutações não-sinônimas, com exceção do gene **Elp** que mostrou uma quantidade maior de mutações sinônimas quando comparadas com as mutações não-sinônimas.

Dentre as cinco redes de haplótipos que separaram *A. sororcula* das outras duas espécies do gênero *Anastrepha*, a rede feita com as sequências dos genes **CG7009 (Figura 6)** e **Cyclophylin (Figura 17)** foram as que separaram de forma mais clara aquela espécie, mostrando que os haplótipos de *A. sororcula*, nestas redes, encontram-se separados, respectivamente por, nove e mais de dez passos mutacionais dos outros haplótipos. É importante observar que todas as nove mutações que levam à separação dos haplótipos de *A. sororcula*, na rede feita com as sequências do gene **CG7009 (Figura 6)** são sinônimas. A ausência de polimorfismos específicos de espécie para os outros genes é facilmente visualizada na maioria das árvores de haplótipos inferidas. As conexões entre os haplótipos de *A. sororcula* no gene **CG7009** e entre os haplótipos das três espécies de *Anastrepha* do grupo *fraterculus* no gene **Cyclophylin** violam os limites da parcimônia que são premissa para a utilização do TCS, mas para efeito de visualização mudamos o limite da parcimônia, de forma que os haplótipos nas duas redes ficassem conectados.

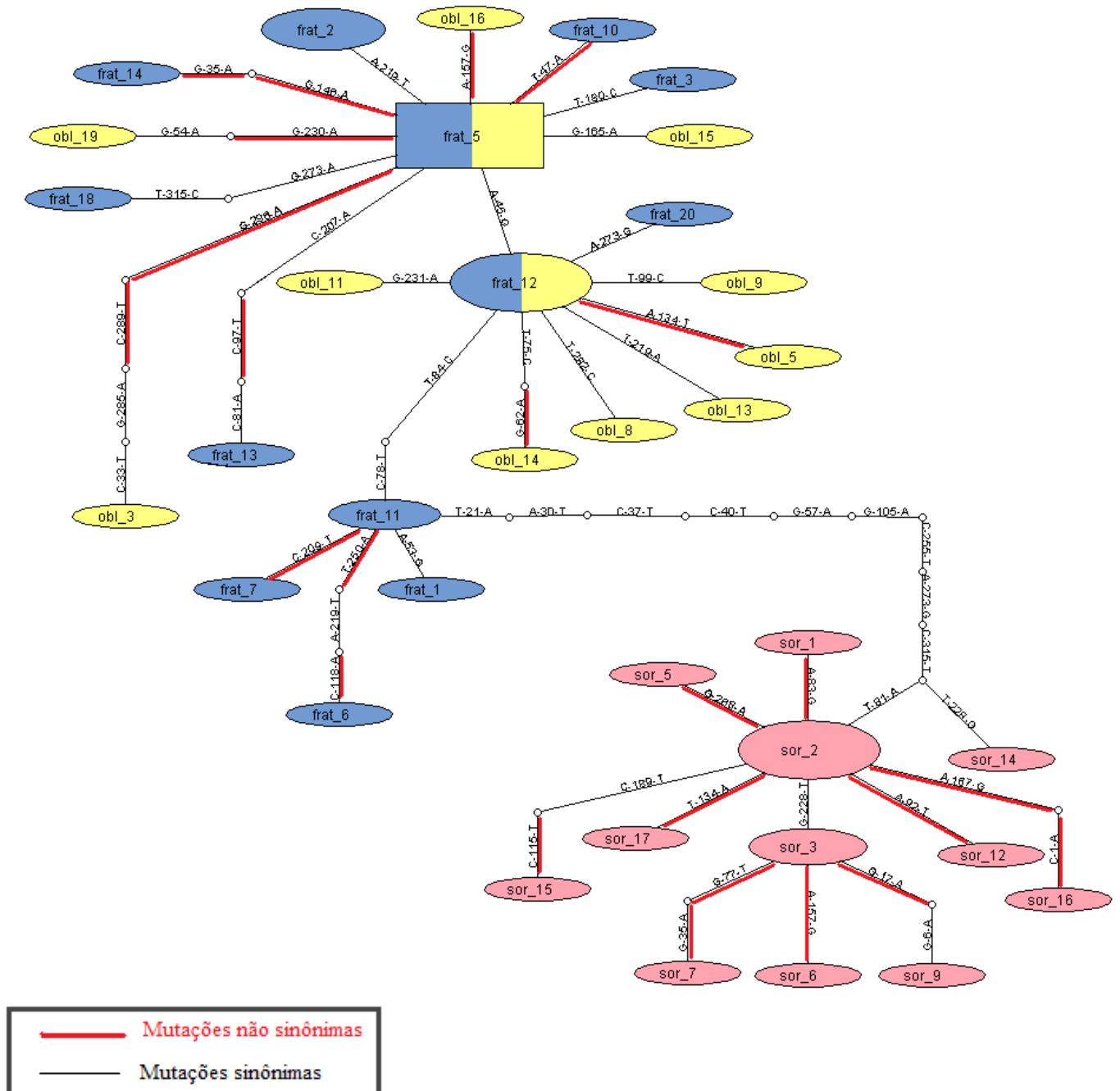
Uma grande parte das redes de haplótipos apresentou diversas mutações não-sinônimas, principalmente para as redes feitas para os genes **CG7009 (Figura 6)**, **CG8064 (Figura 8)**, **CG5325 (Figura 9)**, **TCTP (Figura 13)**, **UQCRX (Figura 15)** e

**Wings up (Figura 16)**, onde houve um predomínio das mutações que levam a mudança de aminoácidos sobre as mutações sinônimas, que não levam a mudança de aminoácido, com exceção do **CG7009**, que apesar uma grande quantidade de mutações que levam à mudança de aminoácido, estas não superou a quantidade de mutações sinônimas. Desconsideramos aqui as mudanças em **Cyclophylin (Figura 17)** uma vez que consideramos este como um pseudogene.

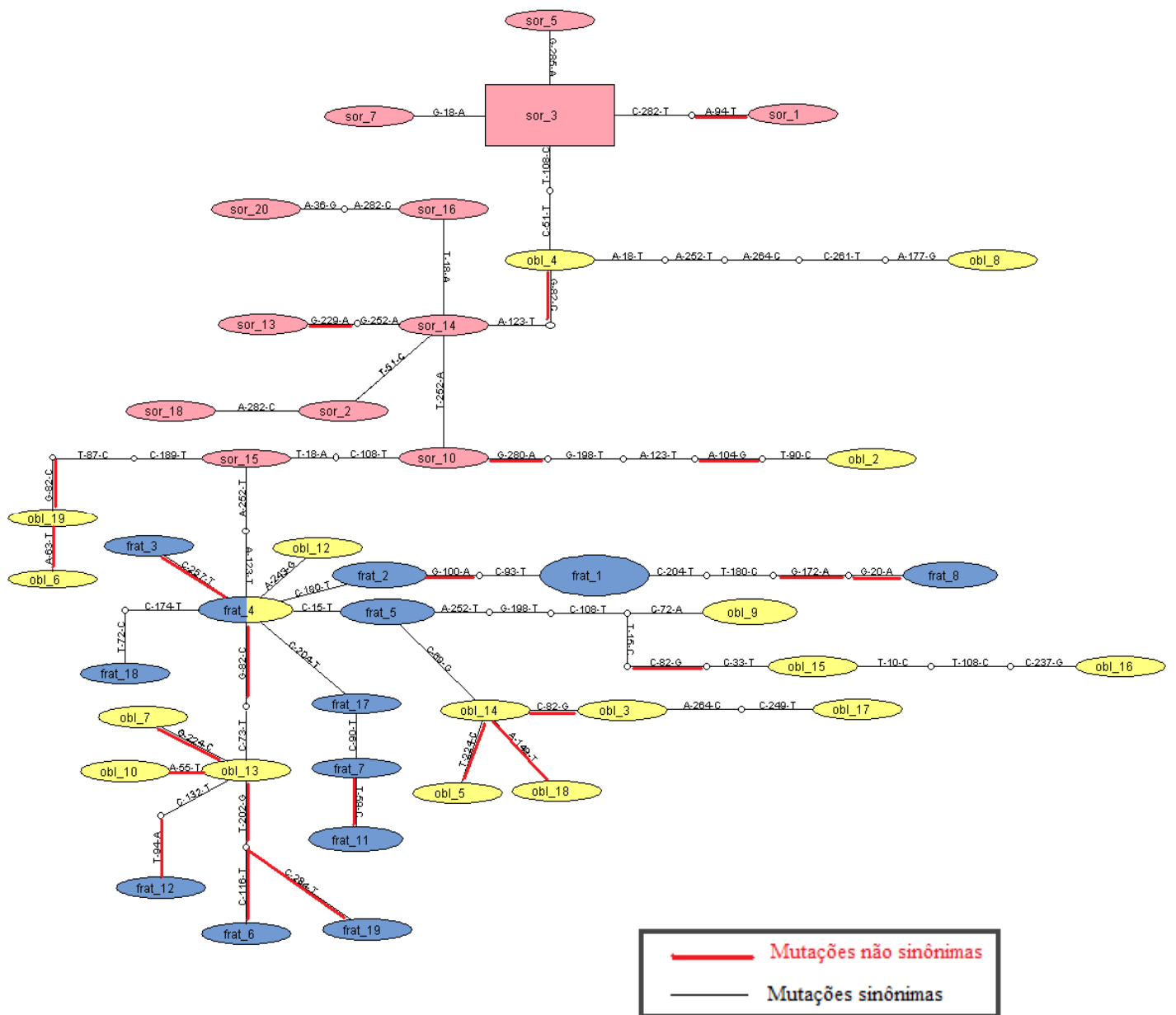
É importante ressaltar que a separação de *A. sororcula* na rede haplotípica feita com as sequências do **CG7009** indica que isso pode ser reflexo de seu papel na diferenciação. No entanto, considerando-se que apenas mudanças sinônimas foram encontradas no ramo que separa as espécies pode ser também considerar que a mudança esteja próxima deste gene e a diferenciação aqui teria ocorrido por carona (*hitchhiking*) ou mesmo que esta diferenciação seja apenas um reflexo da estocasticidade do processo evolutivo, em que algumas regiões do genoma levam tempos diferentes para coalescer. Apenas estudos mais detalhados desta região permitirão uma melhor resposta a esta questão.

Considerando que ao menos para *A. fraterculus* há evidências de espécies crípticas e que a determinação de espécies neste grupo é sempre um processo não trivial, pode ser que uma parcela da diferenciação encontrada reflita de fato segregações encontradas na natureza. Apenas um estudo mais aprofundado com coletas adequadas e análises filogeográficas nos permitiria investigar de fato esta hipótese.

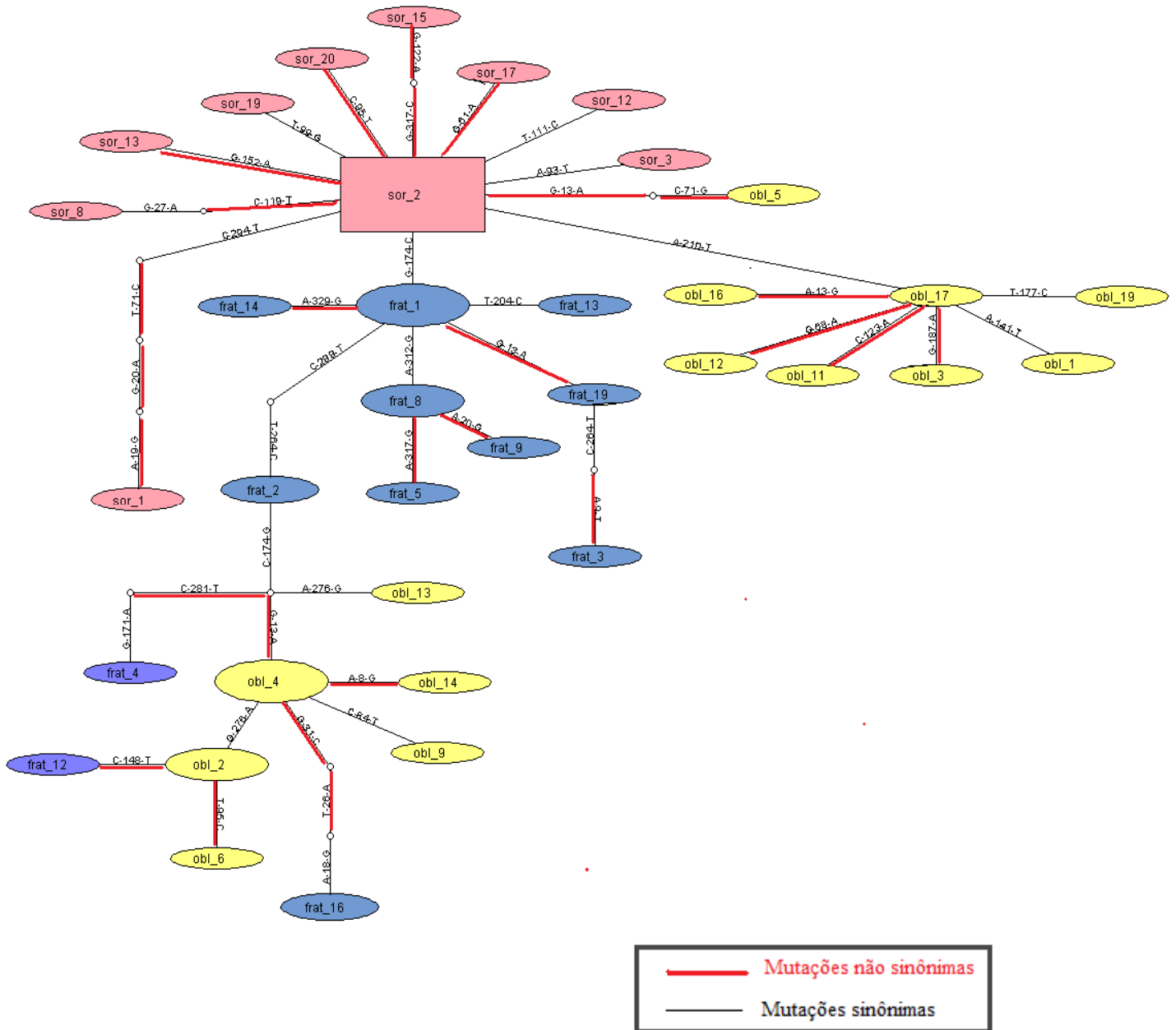
Outras regiões estudadas para este grupo de espécies apresentaram características similares. Dados mitocondriais (Smith-Caldas *et al.*, 2001) e dados obtidos no laboratório para genes coriônicos e vitelínicos (Gonçalves, 2010) e dos genes *doublesex* e *fruitless* (Sobrinho Jr., 2009) também revelaram em geral uma ausência de marcadores específicos de espécie para este grupo. Este padrão pode ser indicação de uma divergência recente entre as espécies do grupo *fraterculus* aqui estudadas, ou pode refletir existência ainda de fluxo gênico na natureza ou mesmo um problema maior taxonômico na identificação de espécies do grupo. Estas questões serão discutidas com os resultados das outras análises realizadas neste trabalho nos itens abaixo.



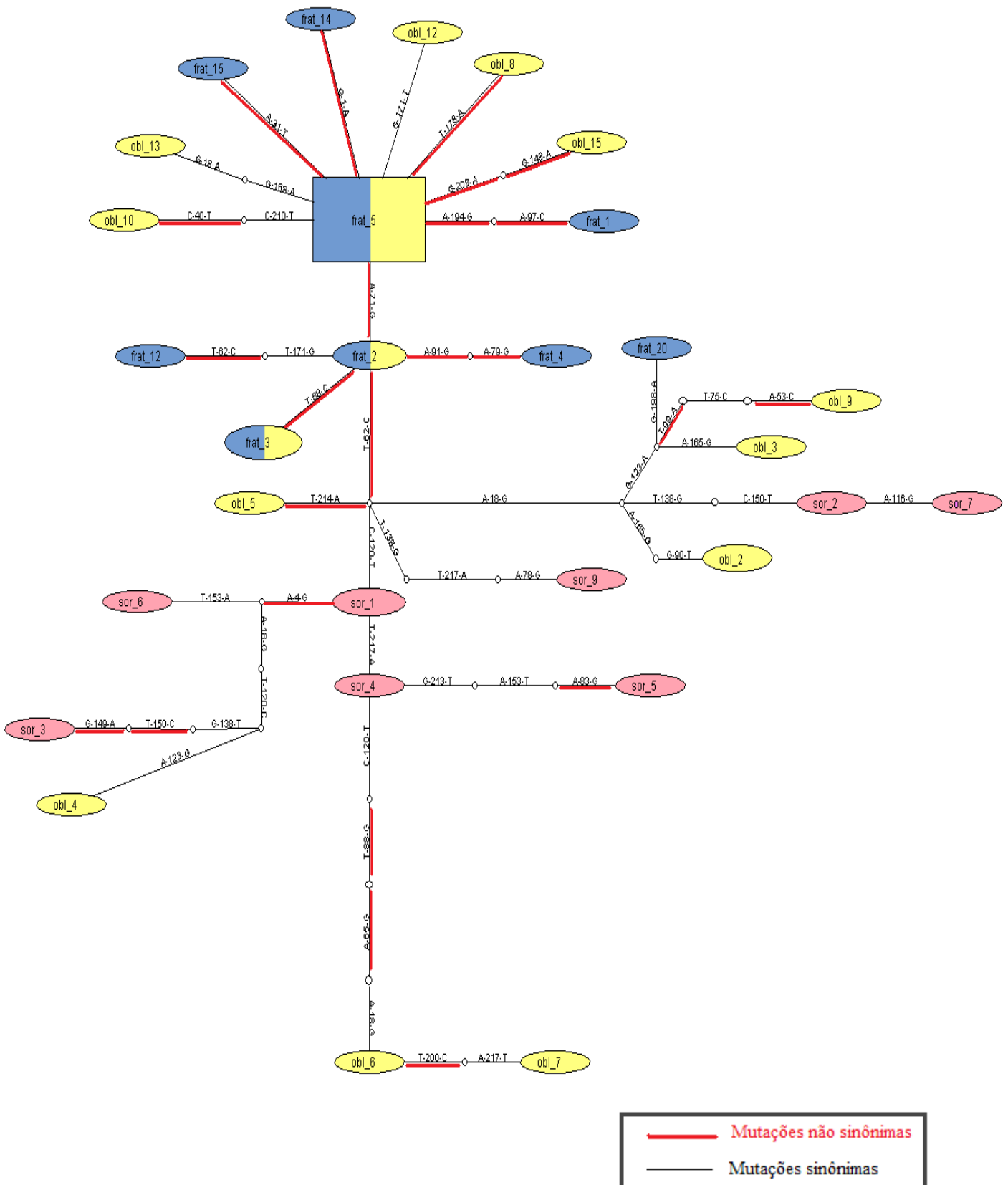
**Figura 6** - Rede de haplótipos da região **CG7009** indicando as relações filogenéticas entre as 20 amostras de *A. fraterculus* (frat), 20 amostras de *A. obliqua* (obl) e 20 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.21. Os haplótipos de *A. sororcula* estão conectados ao resto da árvore embora esteja acima dos limites da parcimônia, que neste caso é de nove passos mutacionais.



**Figura 7**-Rede de haplótipos da região **CG703** indicando as relações filogenéticas entre as 20 amostras de *A. fraterculus* (frat), 20 amostras de *A. obliqua* (obl) e 20 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.21.

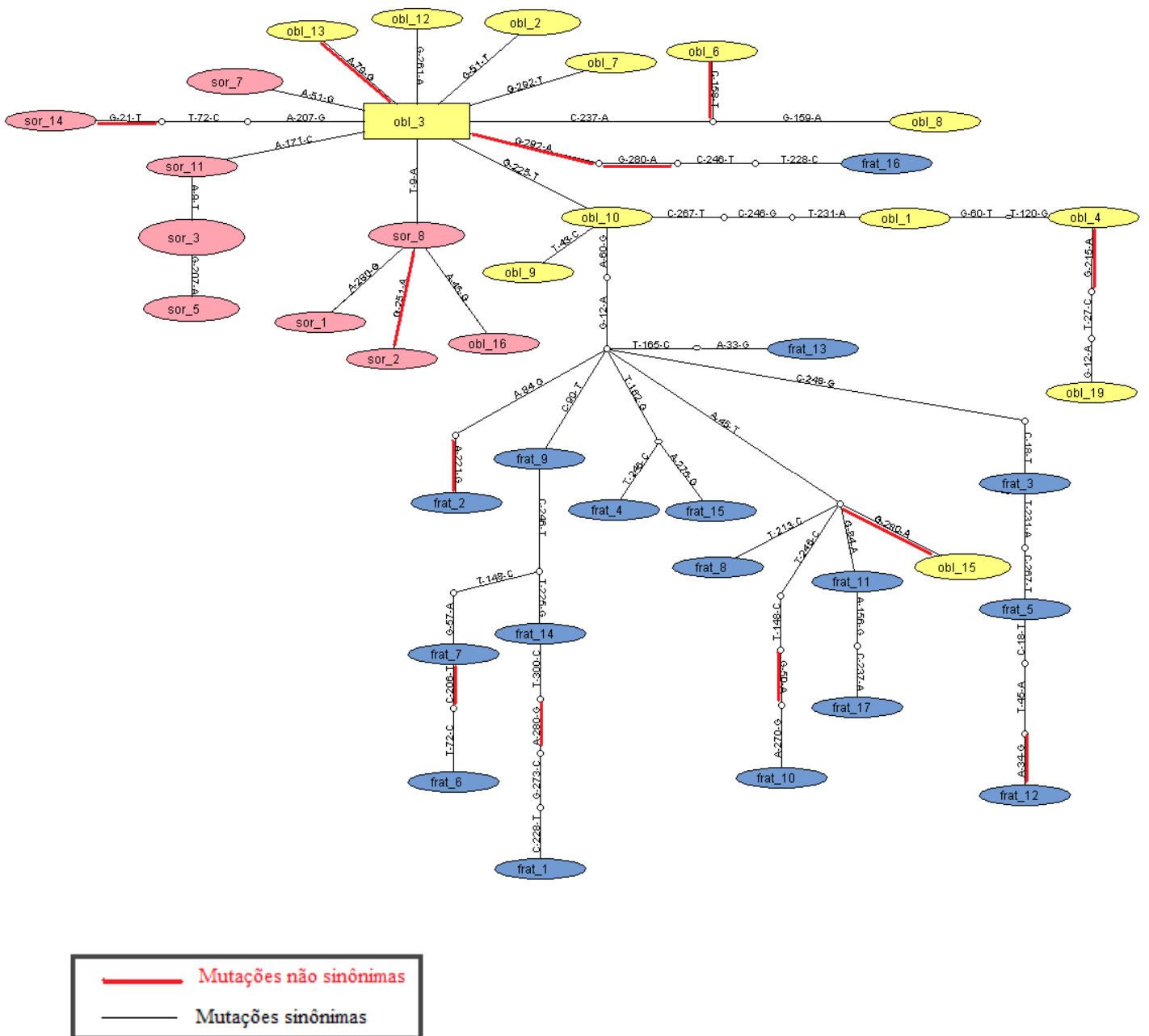


**Figura 8** - Rede de haplótipos da região **CG8064** indicando as relações filogenéticas entre as 20 amostras de *A. fraterculus* (frat), 20 amostras de *A. obliqua* (obl) e 20 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.21.

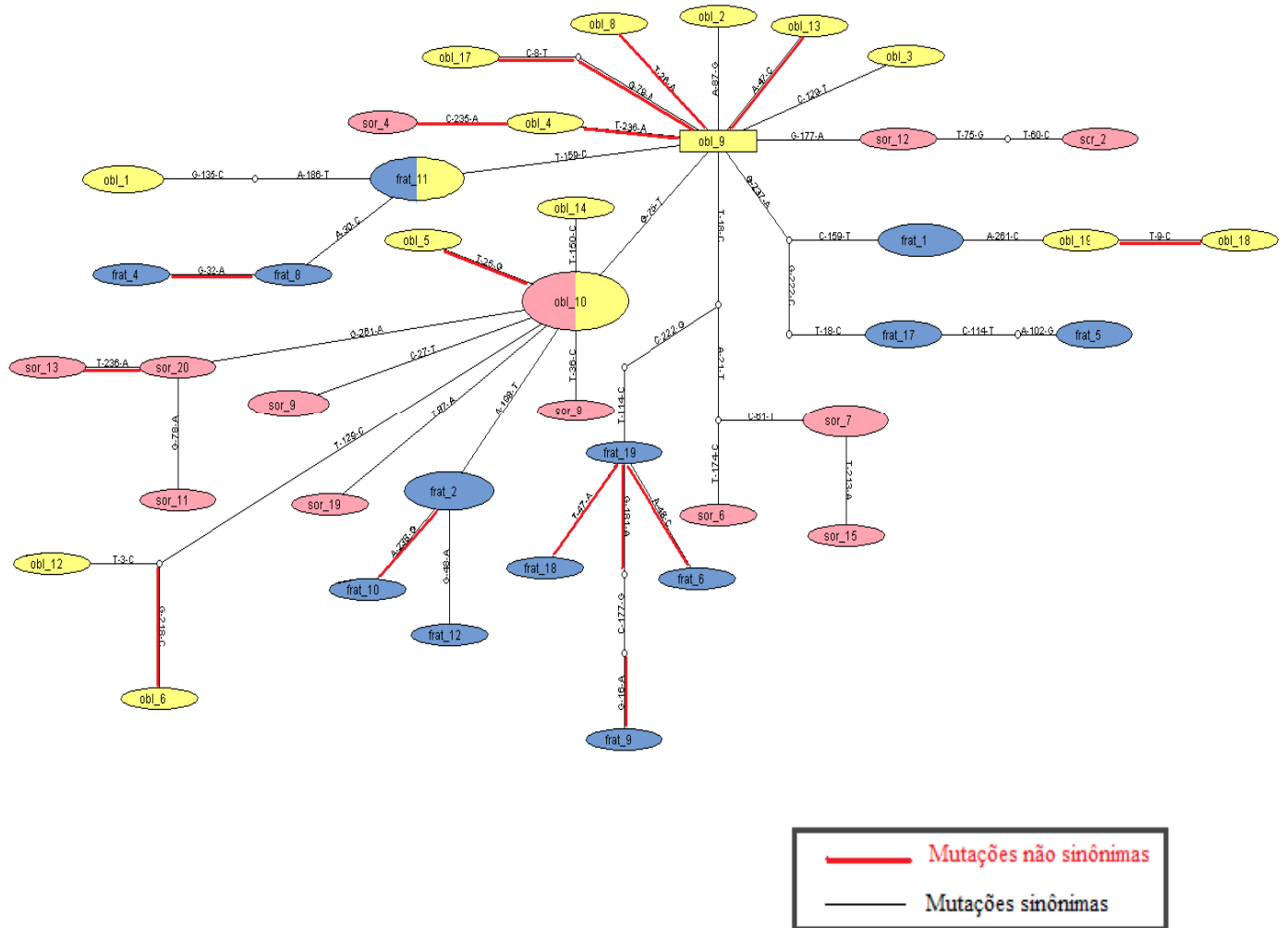


**Figura 9** - Rede de haplótipos da região **CG5325** indicando as relações filogenéticas entre as 20 amostras de *A. fraterculus* (frat), 15 amostras de *A. obliqua* (obl) e nove de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.

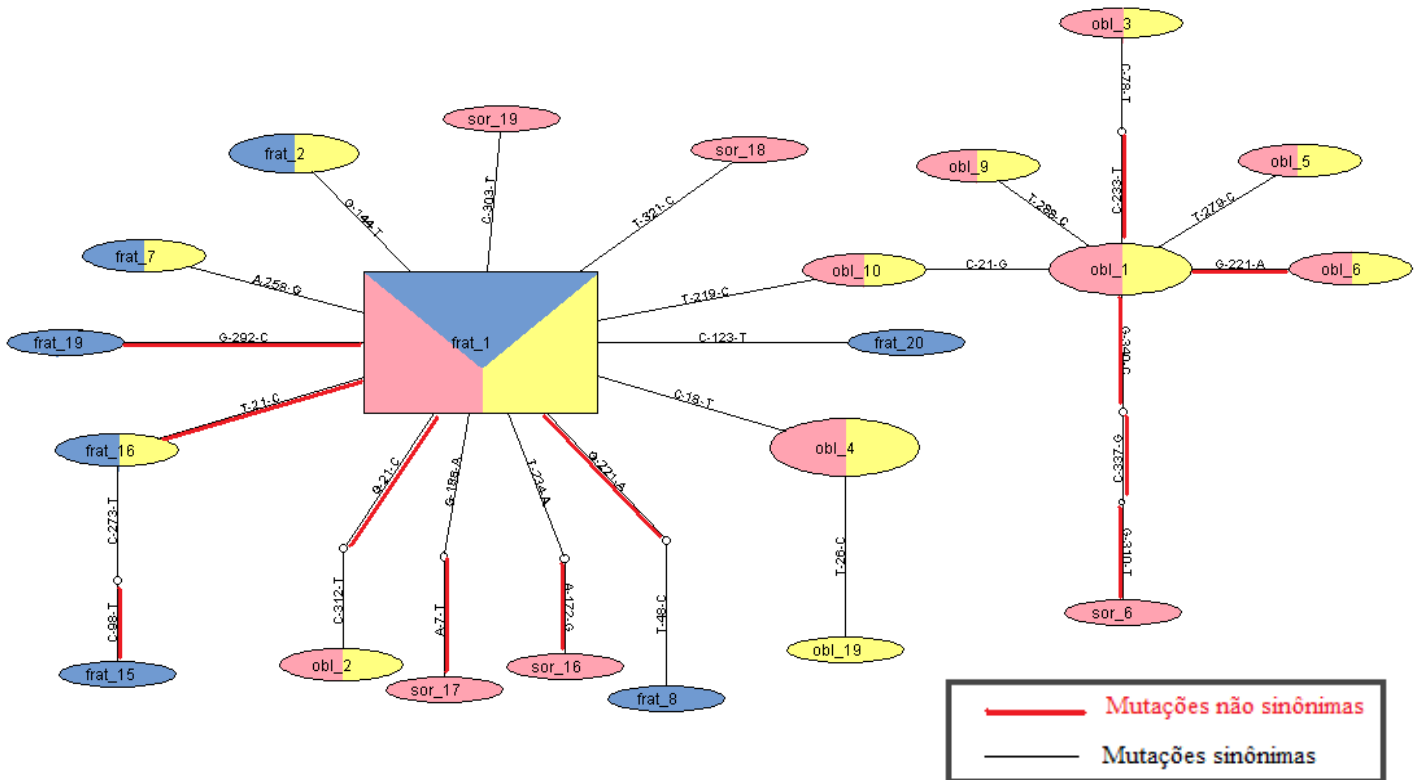




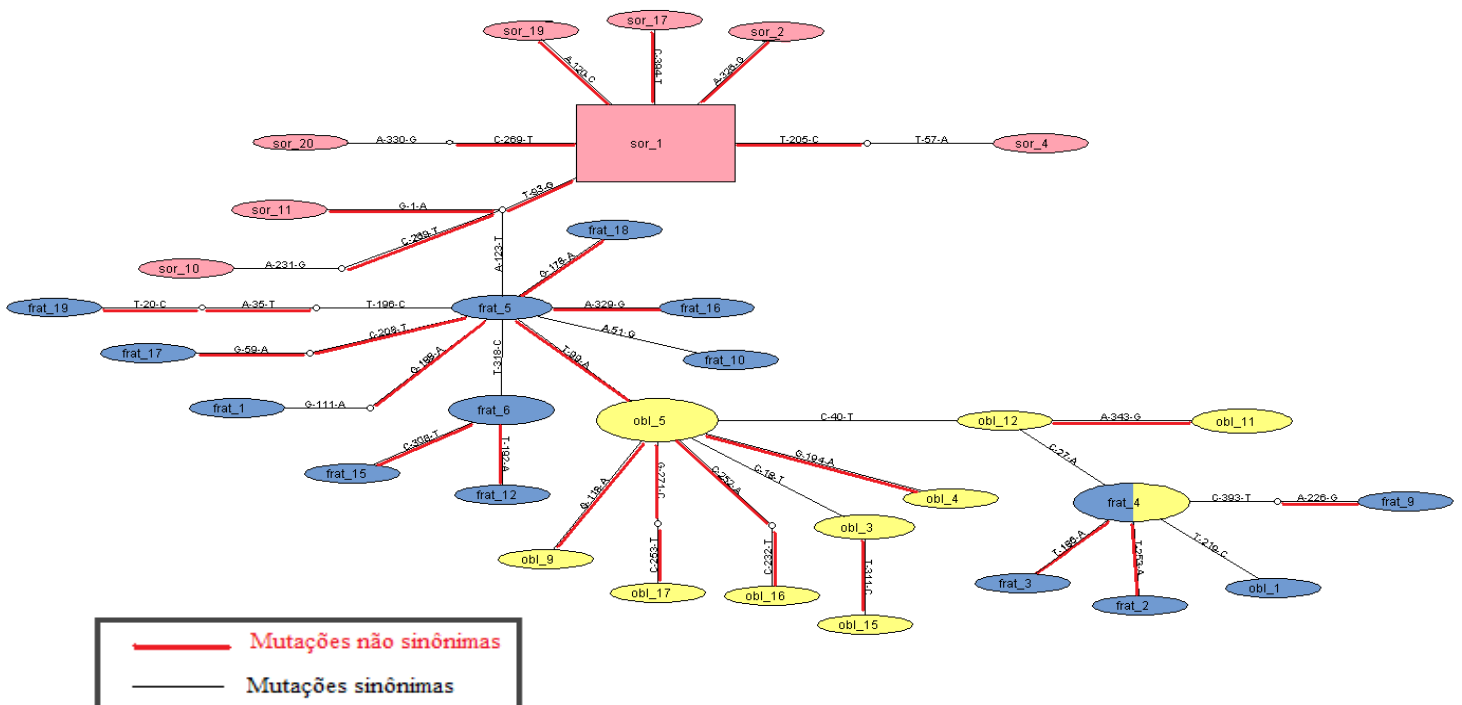
**Figura 10** - Rede de haplótipos da região **Elongation Protein** indicando as relações filogenéticas entre as 17 amostras de *A. fraterculus* (frat), 19 amostras de *A. obliqua* (obl) e 15 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.



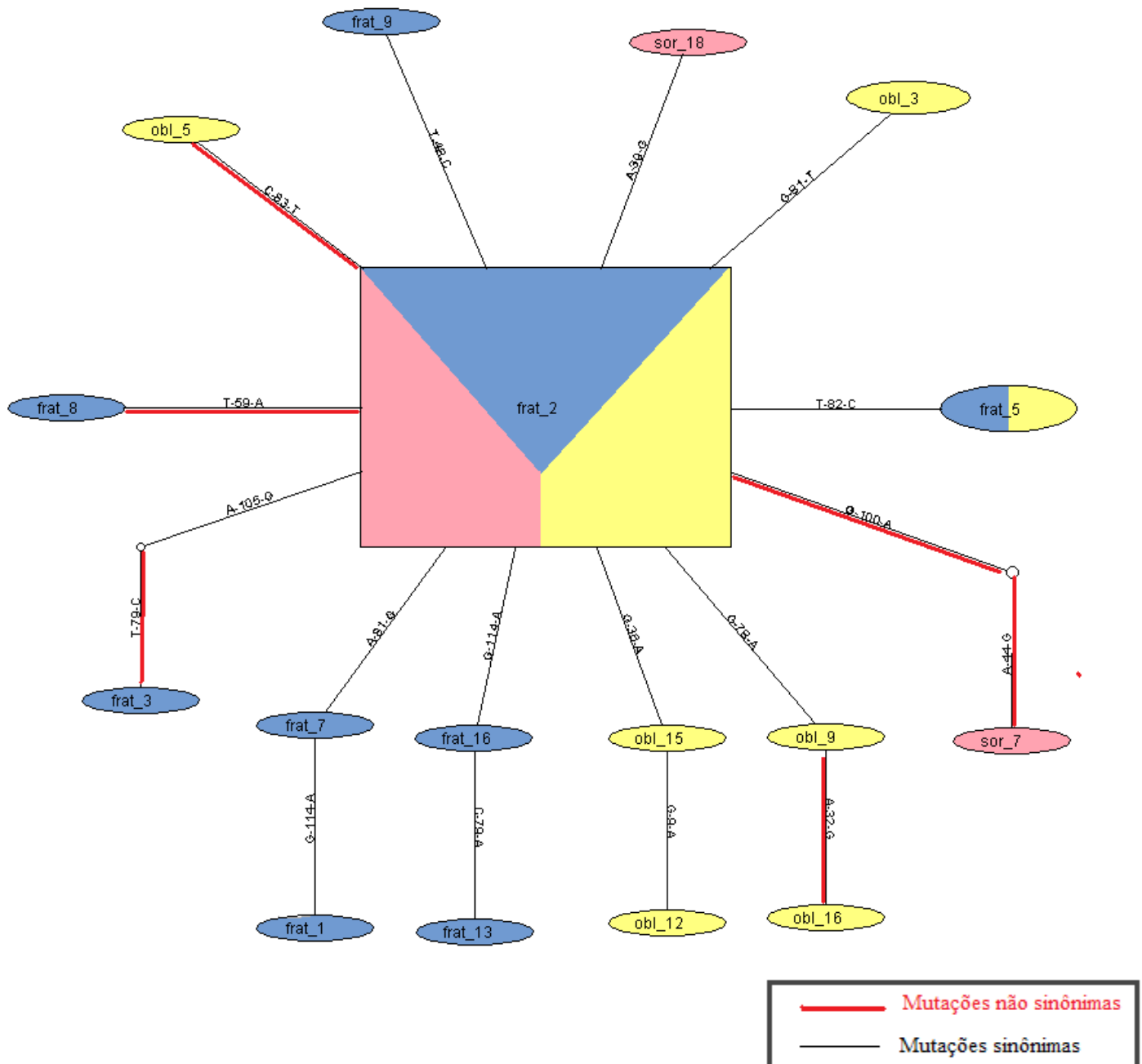
**Figura 11** - Rede de haplótipos da região **MLC** indicando as relações filogenéticas entre as 20 amostras de *A. fraterculus* (frat), 20 amostras de *A. obliqua* (obl) e 20 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.



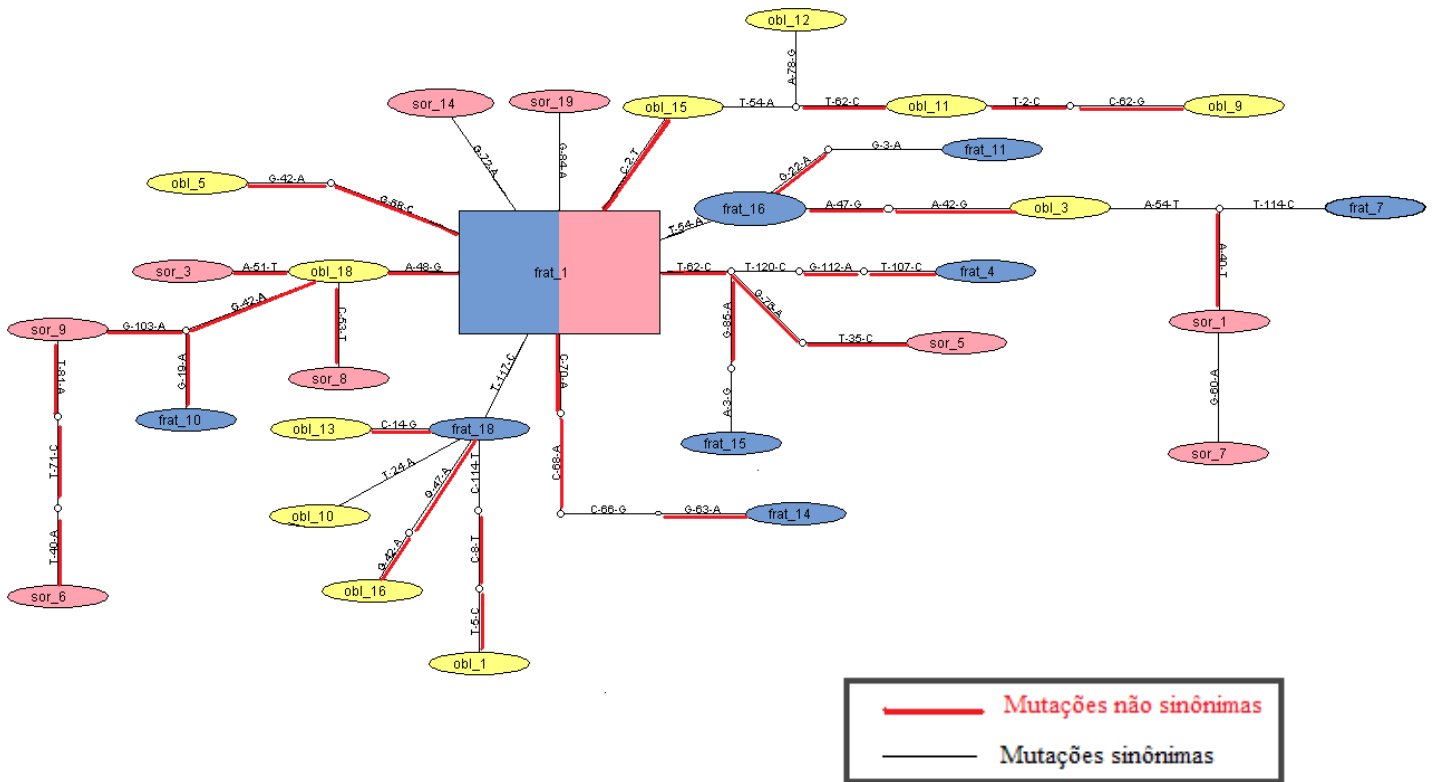
**Figura 12** - Rede de haplótipos da região **Porin** indicando as relações filogenéticas entre as 20 amostras de *A. fraterculus* (frat), 19 amostras de *A. obliqua* (obl) e 20 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.



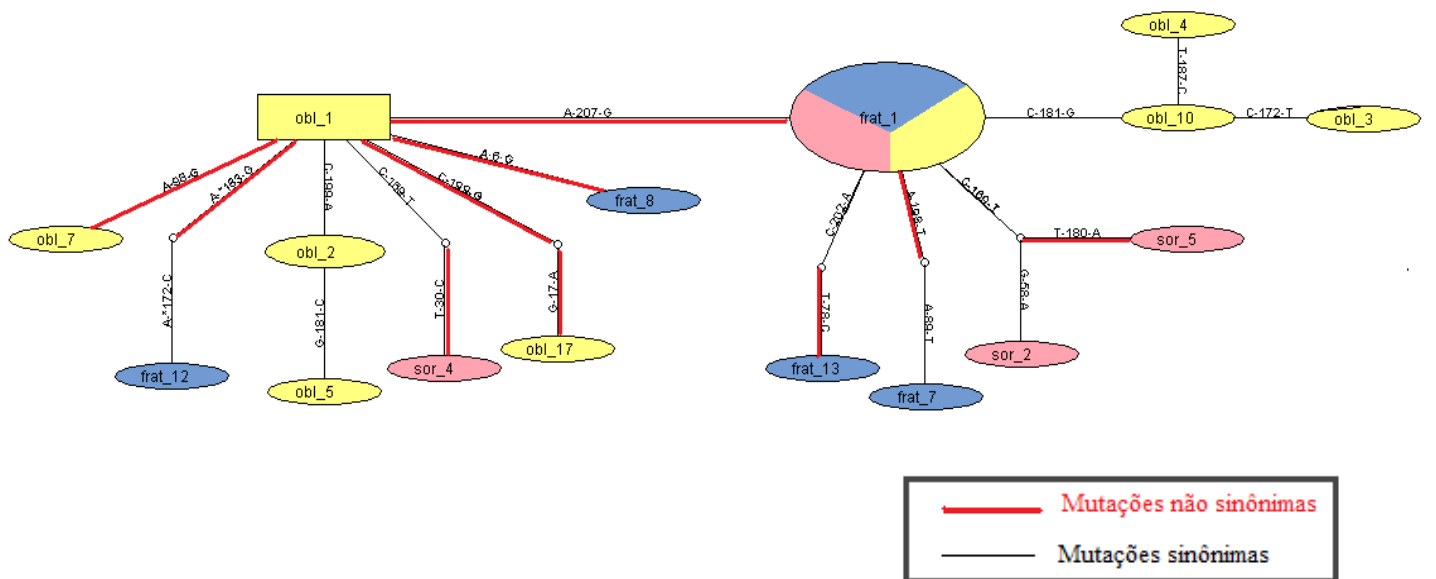
**Figura 13**-Rede de haplótipos da região **TCTP** indicando as relações filogenéticas entre as 19 amostras de *A. fraterculus* (frat), 19 amostras de *A. obliqua* (obl) e 20 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.



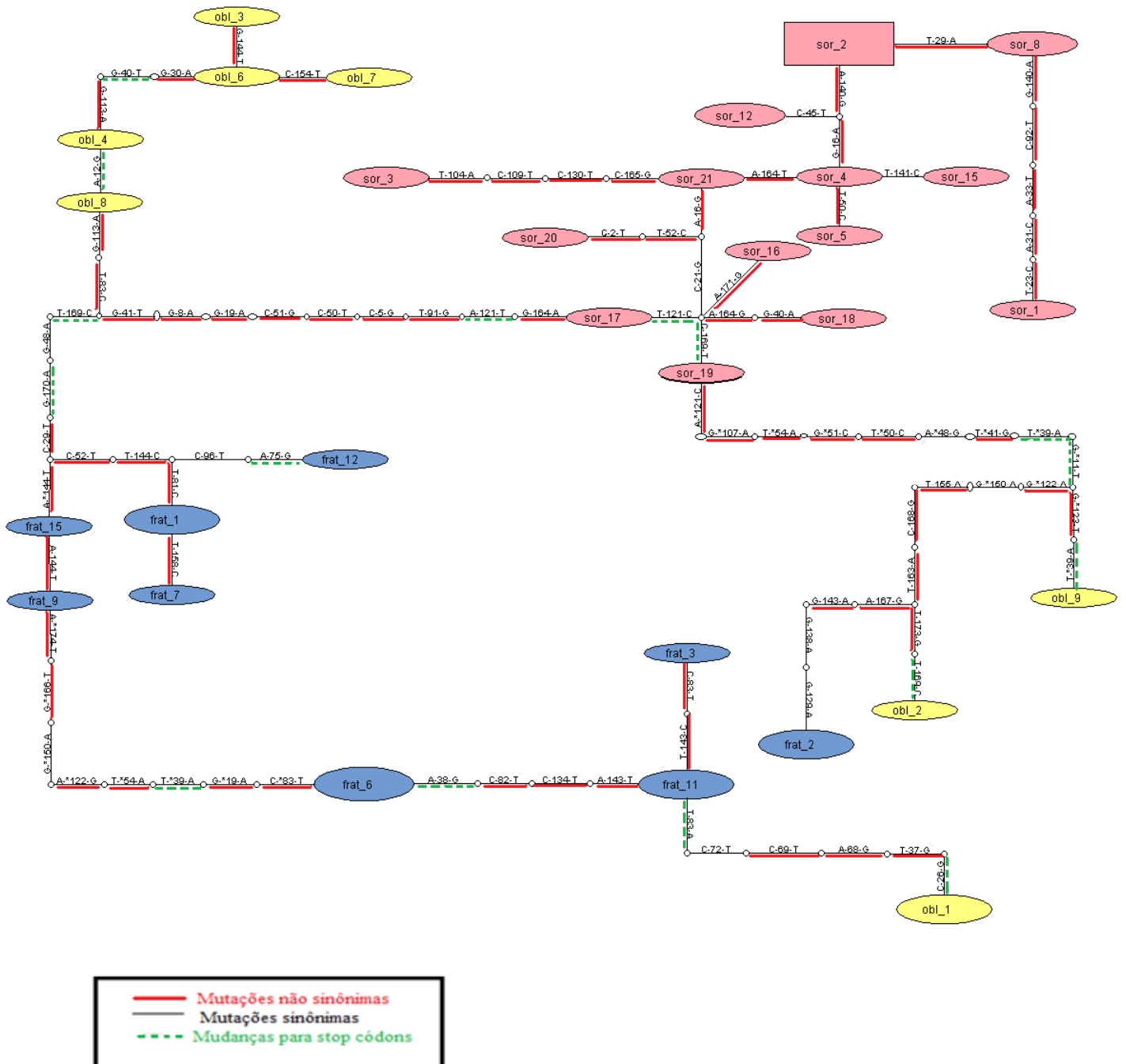
**Figura 14**- Rede de haplótipos da região **Troponin C** indicando as relações filogenéticas entre as 20 amostras de *A. fraterculus* (frat), 20 amostras de *A. obliqua* (obl) e 20 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.



**Figura 15**-Rede de haplótipos da região **UQCRX** indicando as relações filogenéticas entre as amostras 18 de *A. fraterculus* (frat), 13 amostras de *A. obliqua* (obl) e 19 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.



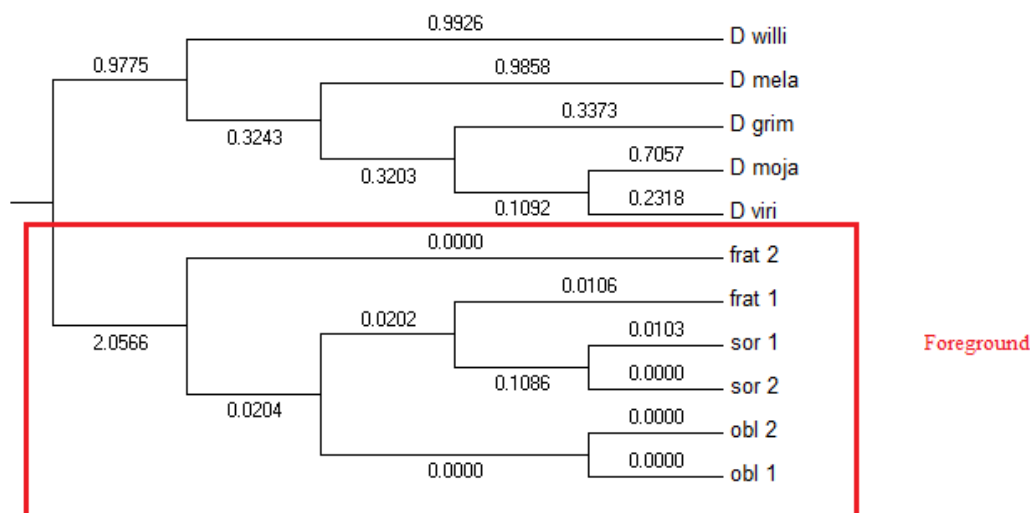
**Figura 16**-Rede de haplótipos da região **Wings Up** indicando as relações filogenéticas entre as 13 amostras de *A. fraterculus* (frat), 16 amostras de *A. obliqua* (obl) e quatro de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2.



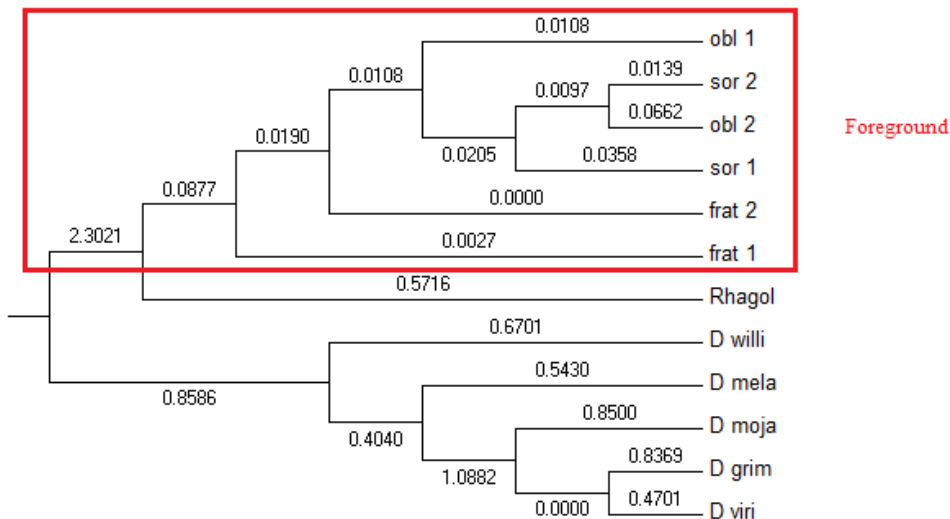
**Figura 17**-Rede de haplótipos da região **Cyclophylin** indicando as relações filogenéticas entre as 18 amostras de *A. fraterculus* (frat), dez amostras de *A. obliqua* (obl) e 21 de *A. sororcula* (sor). Os círculos vazios correspondem a haplótipos inferidos pelo programa TCS 1.2. Os haplótipos de *A. sororcula*, *A. fraterculus* e *A. obliqua* estão conectados ao resto da árvore embora esteja acima dos limites da parcimônia, que neste caso é de mais de dez passos mutacionais.

### 4.3 - Testes de seleção.

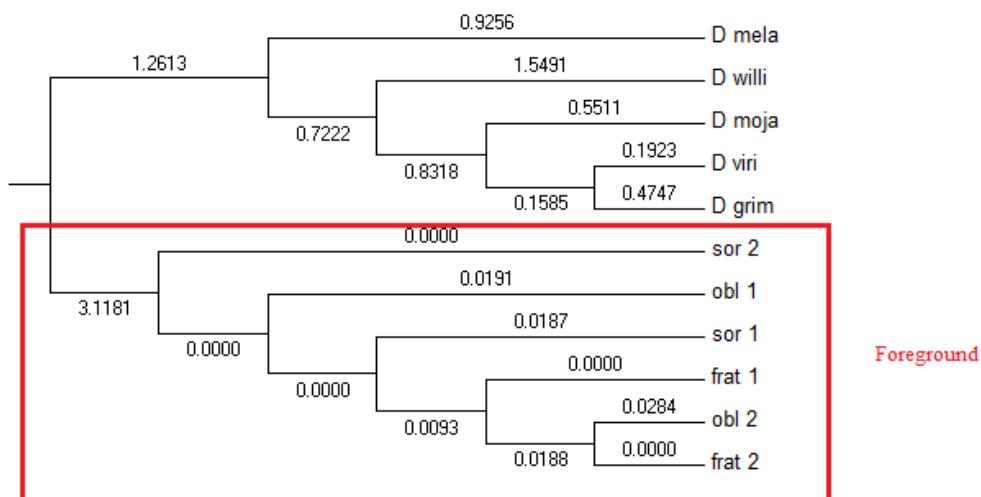
Para investigar padrões de seleção positiva que poderiam estar agindo nos genes aqui estudados, escolhemos aleatoriamente duas sequências de cada uma das espécies e as contrastamos com sequências de Tephritidae do gênero *Rhagoletis* (quando disponível) e algumas espécies que representam uma parcela da diversidade de *Drosophila*. As árvores filogenéticas por máxima verossimilhança geradas pelo programa PhyML v. 3.0, para cada região gênica (**Figuras 18 a 28**), evidenciam os ramos *foreground* e *background* contrastados na análise. Para este tipo de análise de dados macroevolutivos são considerados para o teste de heterogeneidade os valores de  $\omega$  entre ramos e códons.



**Figura 18-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **K80**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Drosophila willistoni* (XP 002072331), *D. melanogaster* (NP 650947), *D. grimshawi* (XP 001995863), *D. mojavensis* (XP 001999852) e *D. virilis* (XP 002054908) para um fragmento do gene **CG7009**. Os dados das espécies de *Drosophila* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sororcula*). Teste de seleção se mostrou significativo para o *Relaxed Branch-site Test*.



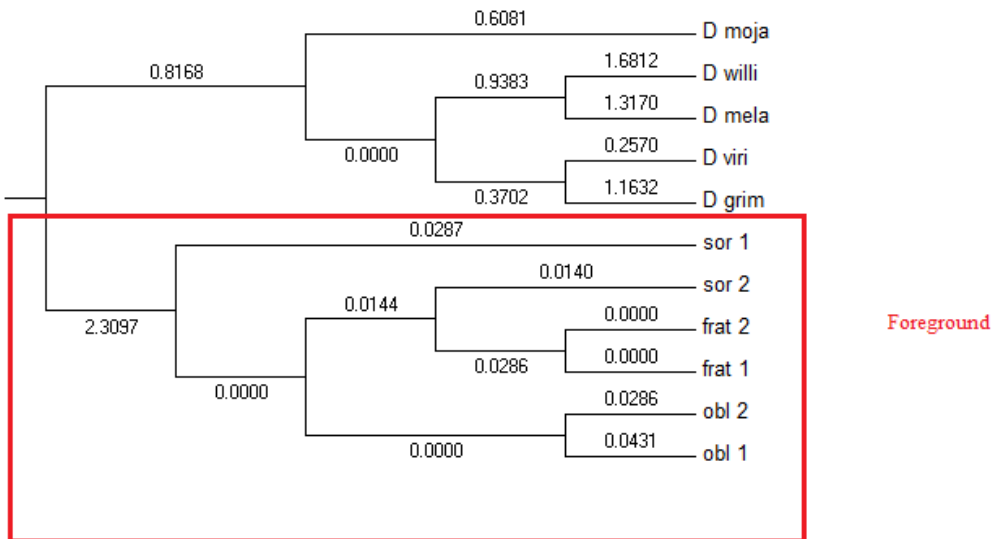
**Figura 19-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **HKY85**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Rhagoletis pomonella* (EZ 127868), *Drosophila willistoni* (XP 002066731), *D. melanogaster* (NP 609143), *D. grimshawi* (XP 001993182), *D. mojavensis* (XP 002001922) e *D. virilis* (XP 002059236) para um fragmento do gene **CG7203**. Os dados das espécies de *Drosophila* e de *R. pomonella* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Teste de seleção se mostrou significativo para o *Strict Branch-site Test*.



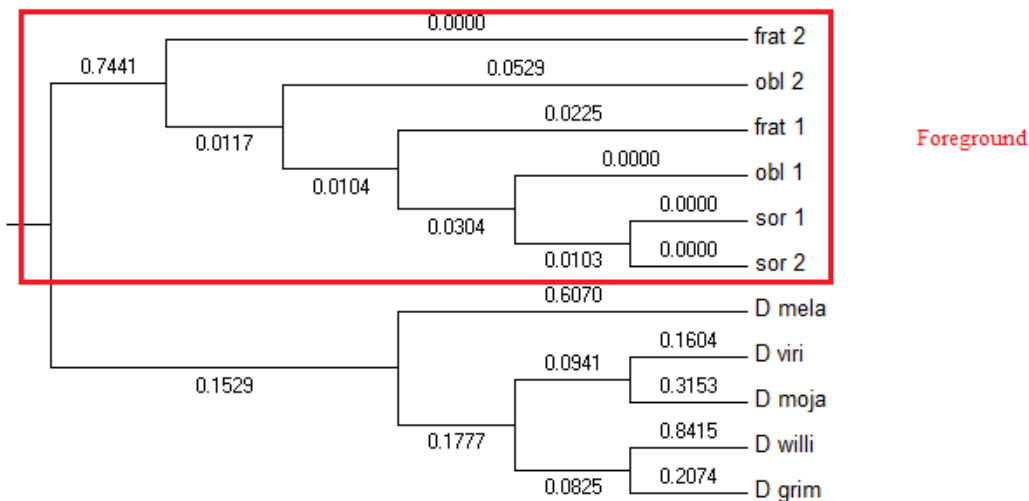
**Figura 20-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **K80**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Drosophila willistoni* (XP 002072312), *D. melanogaster* (NP 650705), *D. grimshawi* (XP 001996274), *D. mojavensis* (XP 001999395) e *D. virilis* (XP 002054614) para um fragmento do gene **CG8064**. Os dados das espécies de



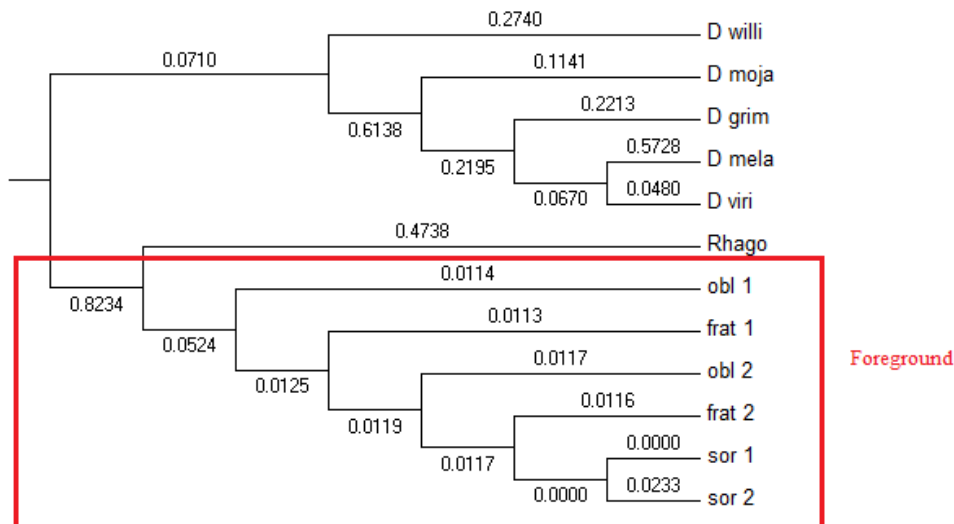
*Drosophila* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Teste de seleção se mostrou significativo para o *Relaxed Branch-site Test* e para o *Strict Branch-site Test*.



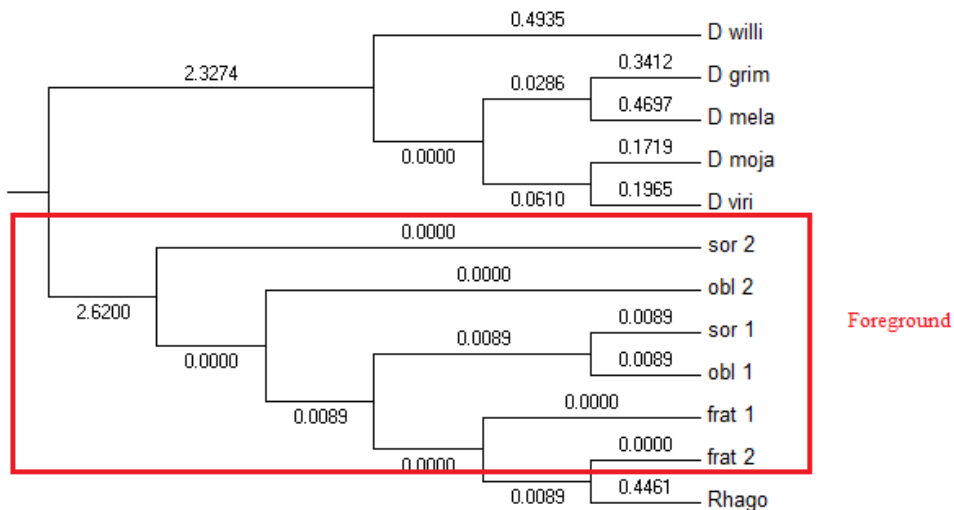
**Figura 21-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **HKY85**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Drosophila willistoni* (XP 002069226), *D. melanogaster* (NP 723722), *D. grimshawi* (XP 001989171), *D. mojavensis* (XP 002001840) e *D. virilis* (XP 002052621) para um fragmento do gene **CG5325**. Os dados das espécies de *Drosophila* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Não foi significativo para nenhum contraste.



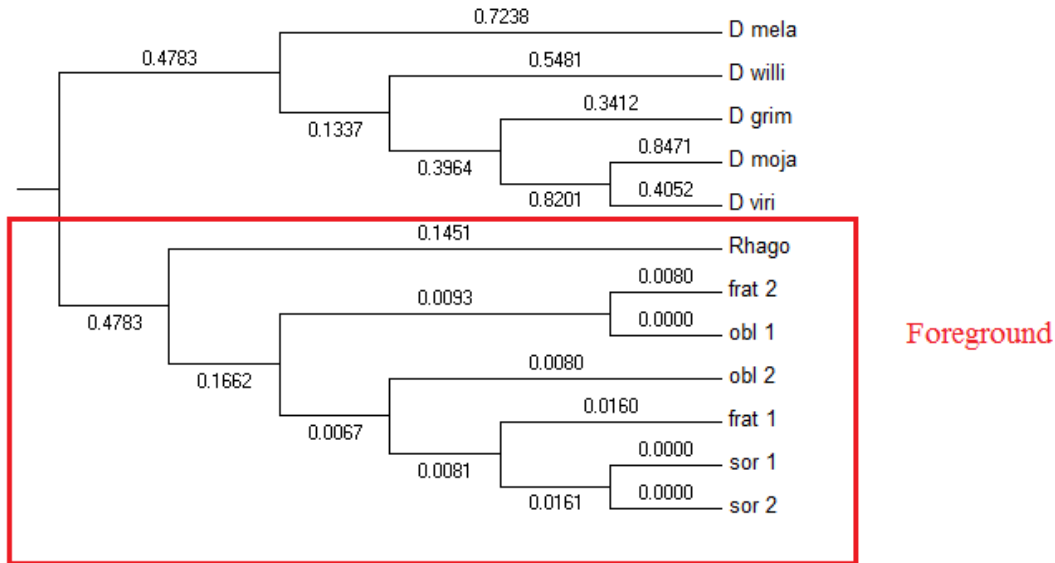
**Figura 22-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **K80**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Drosophila willistoni* (XP 002066814), *D. melanogaster* (NP 608834), *D. grimshawi* (XP 001989312), *D. mojavensis* (XP 002003521) e *D. virilis* (XP 002052755) para um fragmento do gene **Elongation Protein**. Os dados das espécies de *Drosophila* e de *R. pomonella* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Não foi significativo para nenhum contraste.



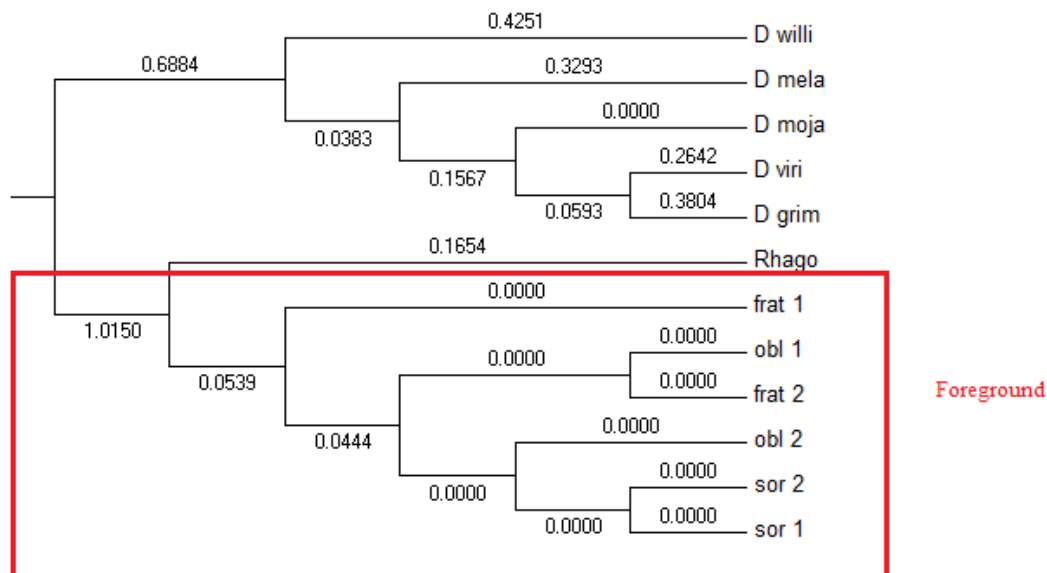
**Figura 23-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **TN93**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Rhagoletis pomonella* (EZ 137400), *Drosophila willistoni* (XP 002071798), *D. melanogaster* (NP 511049), *D. grimshawi* (XP 001992735), *D. mojavensis* (XP 002010859) e *D. virilis* (XP 002057174) para um fragmento do gene **MLC**. Os dados das espécies de *Drosophila* e de *R. pomonella* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Teste de seleção se mostrou significativo para o *Relaxed Branch-site Test* e para o *Strict Branch-site Test*.



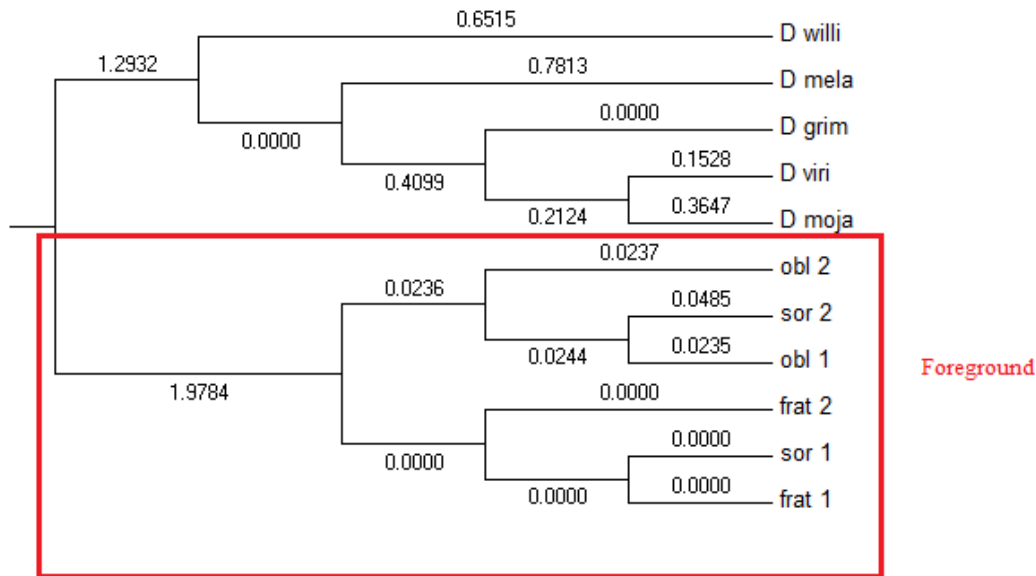
**Figura 24-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **K80**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Rhagoletis pomonella* (EZ 124459), *Drosophila willistoni* (XP 002064648), *D. melanogaster* (NP 476813), *D. grimshawi* (XP 001993004), *D. mojavensis* (XP 002003241) e *D. virilis* (XP 002052597) para um fragmento do gene **Porin**. Os dados das espécies de *Drosophila* e de *R. pomonella* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Não foi significativo para nenhum contrast



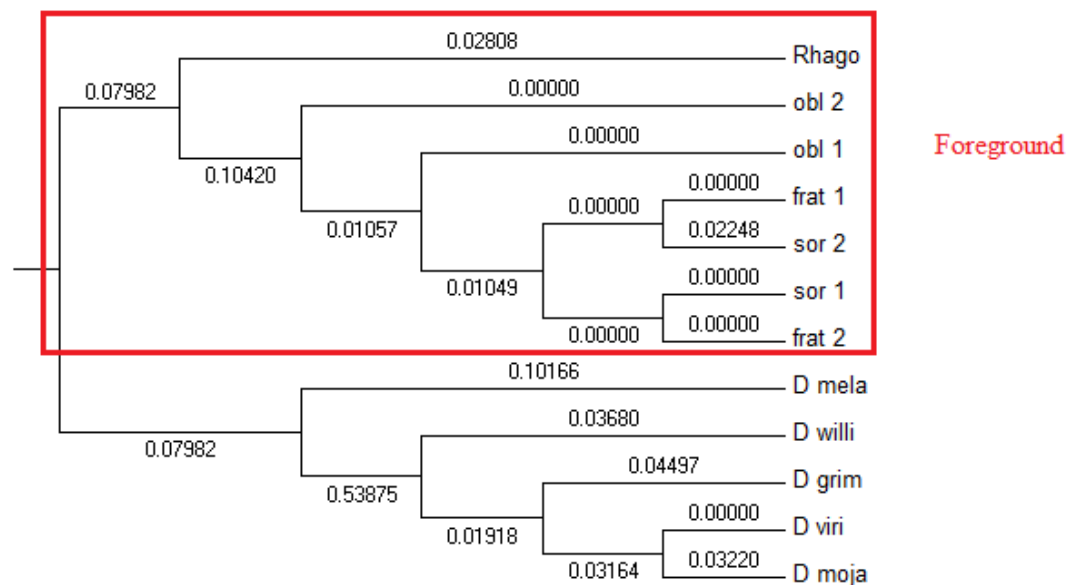
**Figura 25-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **K80**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Rhagoletis pomonella* (EZ 139567), *Drosophila willistoni* (XP 002073380), *D. melanogaster* (NP 650048), *D. grimshawi* (XP 001986483), *D. mojavensis* (XP 001998386) e *D. virilis* (XP002050135) para um fragmento do gene **TCTP**. Os dados das espécies de *Drosophila* e de *R. pomonella* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Não foi significativo para nenhum contraste.



**Figura 26-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **K80**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Rhagoletis pomonella* (EZ 140549), *Drosophila willistoni* (XP 002075414), *D. melanogaster* (NP 524122), *D. grimshawi* (XP 001987472), *D. mojavensis* (XP 002007253) e *D. virilis* (XP 002046615) para um fragmento do gene **Troponin C**. Os dados das espécies de *Drosophila* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Não foi significativo para nenhum contraste.



**Figura 27-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **TN93**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Drosophila willistoni* (XP 002065740), *D. melanogaster* (NP 648905), *D. grimshawi* (XP 001983443), *D. mojavensis* (XP 002012225) e *D. virilis* (XP 002046263) para um fragmento do gene **UQCRX**. Os dados das espécies de *Drosophila* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Não foi significativo para nenhum contraste.



**Figura 28-** Árvore filogenética, inferida por máxima verossimilhança, utilizando o modelo evolutivo **TN93**, mostrando as relações de parentesco entre moscas do grupo *fraterculus* e grupos externos *Rhagoletis pomonella* (EZ 138366), *Drosophila willistoni* (XP 002071860), *D. melanogaster* (NP 728141), *D. grimshawi* (XP 001991571), *D. mojavensis* (XP 002010813) e *D. virilis* (XP 002046177) para um fragmento do gene **Wings Up**. Os dados das espécies de *Drosophila* e de *R. pomonella* foram obtidos do GenBank. frat. (*A. fraterculus*); obl (*A. obliqua*); sor (*A. sorocula*). Não foi significativo para nenhum contraste.

Através das análises feitas pelo PAML (Yang 2007) investigamos, pelo contraste da razão das taxas de substituições não-sinônimas por sinônimas ( $\omega$ ) (Yang *et al.* 2005; Zhang *et al.* 2005) a presença de seleção positiva nas regiões gênicas amplificadas neste trabalho, com exceção do pseudogene **Cyclophylin**, uma vez que este teste apenas pode ser feito em regiões codificadoras. Os parâmetros dos modelos nulos M1a e MA restrito e do modelo alternativo MA estão apresentados na **Tabela 12**. O teste que contrasta os modelos M1A-MA (*Relaxed Branch-site Test*) rejeitou o modelo nulo de restrição seletiva ( $0 < \omega \leq 1$ ) para os genes **MLC**, **CG8064**, e **CG7009** indicando que os ramos *background* e *foreground* divergiram nesta região por meio de uma restrição seletiva mais branda ou por seleção positiva (**Tabela 13**). Para discriminar as duas hipóteses, contrastamos os modelos MA restrito e MA (*Strict Branch-site Test*), para cada um dos genes. Para os genes **MLC** e **CG8064** (**Tabela 13**) rejeitamos os contrastes do modelo nulo, e a partir destes resultados podemos inferir que, provavelmente, estes genes evoluíram por meio de uma seleção positiva. Entretanto, o contraste do modelo nulo não foi rejeitado para o gene **CG7009** (**Tabela 13**) nos mostrando que este gene parece estar evoluindo por meio de uma restrição seletiva mais branda.

Para o gene **CG7203** (**Tabela 13**) não rejeitamos o contraste do modelo nulo para o primeiro teste (*Relaxed Branch-site Test*), no entanto, conseguimos rejeitá-lo para o segundo teste (*Strict Branch-site Test*). Como em teoria o primeiro teste seria o mais abrangente, como ele não rejeitou o contraste do modelo nulo, consideramos que esta região gênica analisada não poderia estar evoluindo por seleção positiva, embora não entendamos em que situações o teste mais estrito daria desvios significativos de neutralidade quanto o primeiro não o tenha feito.

Por meio da análise Bayesiana, Bayes Empirical Bayes (BEB), estimamos, para cada região gênica, sítios com probabilidade superior a 0,95 ou 0,99 de possuírem  $w > 1$ , sendo que, para o gene **MLC** foi estimado um sítio, um sítio para a região do gene **CG7203**, um sítio para o gene **CG7009**, todos com probabilidade maior que 95% de apresentarem seleção positiva e sete sítios para o **CG8064**, em que alguns apresentaram probabilidade maior que 95% e outros uma probabilidade maior que 99% de possuírem seleção positiva (**Tabela 13**). Para as outras regiões gênicas amplificadas não conseguimos rejeitar nenhum dos contrastes (**Tabela 13**).

Para os genes nos quais detectamos seleção positiva, observamos um grande número de mutações não-sinônimas, mas não foram os únicos para os quais houve uma maior proporção de mutações não silentes com relação às mutações sinônimas. Nestes

casos, em geral, mas mutações não-sinônimas estão mais distribuídas nas partes externas das redes de haplótipos (tips) do que nos interiores, o que é sugestão destas mudanças serem mais recentes do que as sinônimas (Casteloe & Templeton, 1994). Para as regiões com sinal de seleção positiva vimos também que estas apresentavam os maiores valores para o Teste de  $F_s$  de  $F_u$  feito para cada locus com as três espécies do gênero *Anastrepha* juntas.

**Tabela 12a**—Parâmetros estimados e valores de logaritmo da máxima verossimilhança (LNL) para os modelos M1a, MA e MA restrito nos dados macroevolutivos.

	<b>Modelo</b>	<b>Parâmetros</b>	<b>LNL (Log likelihood)</b>
<b>CG7009</b>	M1A	$p_0 = 0,978$ $p_1 = 0,022$ $\omega_0 = 0,045$ $\omega_1 = 1,000$	-1449,915
	MA	$p_0 = 0,867$ $p_1 = 0,014$ $(p_2+p_3) = 0,119$ $\omega_0 = 0,044$ $\omega_2 = 42,57$	-1444,684
	MA restrito	$p_0 = 0,85$ $p_1 = 0,014$ $(p_2+p_3) = 0,134$ $\omega_0 = 0,042$ $\omega_2 = 1,000$	-1445,099
<b>CG7203</b>	M1A	$p_0 = 0,567$ $p_1 = 0,432$ $\omega_0 = 0,042$ $\omega_1 = 1,000$	-1855,649
	MA	$p_0 = 0,663$ $p_1 = 0,155$ $(p_2+p_3) = 0,181$ $\omega_0 = 0,106$ $\omega_2 = 183,678$	-1853,147
	MA restrito	$p_0 = 0,519$ $p_1 = 0,380$ $(p_2+p_3) = 0,099$ $\omega_0 = 0,036$ $\omega_2 = 1,000$	-1855,158
<b>CG5325</b>	M1A	$p_0 = 0,704$ $p_1 = 0,295$ $\omega_0 = 0,042$ $\omega_1 = 1,000$	-1221,514
	MA	$p_0 = 0,604$ $p_1 = 0,260$ $(p_2+p_3) = 0,134$ $\omega_0 = 0,037$ $\omega_2 = 1,000$	-1220,669
	MA restrito	$p_0 = 0,604$ $p_1 = 0,260$ $(p_2+p_3) = 0,134$ $\omega_0 = 0,037$ $\omega_2 = 1,000$	-1220,669
<b>CG8064</b>	M1A	$p_0 = 0,915$ $p_1 = 0,085$ $\omega_0 = 0,031$ $\omega_1 = 1,000$	-1549,872
	MA	$p_0 = 0,748$ $p_1 = 0,065$ $(p_2+p_3) = 0,186$ $\omega_0 = 0,031$ $\omega_2 = 3,202$	-1537,105
	MA restrito	$p_0 = 0,736$ $p_1 = 0,066$ $(p_2+p_3) = 0,197$ $\omega_0 = 0,029$ $\omega_2 = 1,000$	-1538,499
<b>MLC</b>	M1A	$p_0 = 0,990$ $p_1 = 0,009$ $\omega_0 = 0,004$ $\omega_1 = 1,000$	-908,051
	MA	$p_0 = 0,981$ $p_1 = 0$ $(p_2+p_3) = 0,018$ $\omega_0 = 0,003$ $\omega_2 = 16,104$	-900,917
	MA restrito	$p_0 = 0,975$ $p_1 = 0$ $(p_2+p_3) = 0,025$ $\omega_0 = 0,002$ $\omega_2 = 1,000$	-902,466
<b>Porin</b>	M1A	$p_0 = 0,99999$ $p_1 = 0,00001$ $\omega_0 = 0,009$ $\omega_1 = 1,000$	-1217,381
	MA	$p_0 = 0$ $p_1 = 0$ $(p_2+p_3) = 1,000$ $\omega_0 = 0,011$ $\omega_2 = 999,000$	-1216,846
	MA restrito	$p_0 = 0$ $p_1 = 0$ $(p_2+p_3) = 1,000$ $\omega_0 = 0,011$ $\omega_2 = 1,000$	-1216,869

Lnl = logaritmo da máxima verossimilhança;  $w_0$  = valores de dN/dS para sítios com  $0 < w < 1$ ;  $w_1$  = valores de dN/dS fixados a 1;  $w_2$  = valores de dN/dS para sítios  $w > 1$  e correspondem apenas aos sítios no ramo *foreground*.

**Tabela 12b** –Parâmetros estimados e valores de logaritmo da máxima verossimilhança (LNL) para os modelos M1a, MA e MA restrito nos dados macroevolutivos.

	<b>Modelo</b>	<b>Parâmetros</b>	<b>LNL (Log likelihood)</b>
<b>Elp</b>	M1A	$p_0 = 0,979$ $p_1 = 0,020$ $\omega_0 = 0,010$ $\omega_1 = 1,000$	-1070,543
	MA	$p_0 = 0,905$ $p_1 = 0,018$ $(p_2+p_3) = 0,076$ $\omega_0 = 0,008$ $\omega_2 = 1,000$	-1070,072
	MA restrito	$p_0 = 0,905$ $p_1 = 0,018$ $(p_2+p_3) = 0,076$ $\omega_0 = 0,008$ $\omega_2 = 1,000$	-1070,072
<b>Troponin C</b>	M1A	$p_0 = 0,957$ $p_1 = 0,043$ $\omega_0 = 0$ $\omega_1 = 1,000$	-416,661
	MA	$p_0 = 0,879$ $p_1 = 0,039$ $(p_2+p_3) = 0,081$ $\omega_0 = 0$ $\omega_2 = 1,252$	-416,661
	MA restrito	$p_0 = 0,876$ $p_1 = 0,039$ $(p_2+p_3) = 0,084$ $\omega_0 = 0$ $\omega_2 = 1,000$	-416,661
<b>UQCRX</b>	M1A	$p_0 = 0,99999$ $p_1 = 0,00001$ $\omega_0 = 0,089$ $\omega_1 = 1,000$	-564,245
	MA	$p_0 = 0,494$ $p_1 = 0,066$ $(p_2+p_3) = 0,439$ $\omega_0 = 0,049$ $\omega_2 = 1,000$	-562,604
	MA restrito	$p_0 = 0,494$ $p_1 = 0,066$ $(p_2+p_3) = 0,439$ $\omega_0 = 0,049$ $\omega_2 = 1,000$	-562,604
<b>Wings Up</b>	M1A	$p_0 = 0,837$ $p_1 = 0,162$ $\omega_0 = 0,010$ $\omega_1 = 1,000$	-766,105
	MA	$p_0 = 0,837$ $p_1 = 0,162$ $(p_2+p_3) = 0$ $\omega_0 = 0,010$ $\omega_2 = 1,000$	-766,105
	MA restrito	$p_0 = 0,837$ $p_1 = 0,162$ $(p_2+p_3) = 0$ $\omega_0 = 0,010$ $\omega_2 = 1,000$	-766,105
<b>TCTP</b>	M1A	$p_0 = 0,934$ $p_1 = 0,065$ $\omega_0 = 0,043$ $\omega_1 = 1,000$	-1779,409
	MA	$p_0 = 0,831$ $p_1 = 0,065$ $(p_2+p_3) = 0,111$ $\omega_0 = 0,043$ $\omega_2 = 1,000$	-1779,409
	MA restrito	$p_0 = 0,933$ $p_1 = 0,065$ $(p_2+p_3) = 0,002$ $\omega_0 = 0,043$ $\omega_2 = 1,000$	-1779,409

Lnl = logaritmo da máxima verossimilhança;  $w_0$  = valores de dN/dS para sítios com  $0 < w < 1$ ;  $w_1$  = valores de dN/dS fixados a 1;  $w_2$  = valores de dN/dS para sítios  $w > 1$  e correspondem apenas aos sítios no ramo *foreground*.



**Tabela 13** – Contraste dos modelos M1a, MA restrito e MA por LRT nos dados macroevolutivos.

	<b>Contraste</b>	<b>LTR</b>	<b>Sítios com seleção Positiva</b>
<b>CG7009</b>	MAxM1A	<b>10,463*</b>	54 <sup>•</sup>
	MAxMAnull	0,831	
<b>CG7203</b>	MAxM1A	5,00	22 <sup>•</sup>
	MAxMAnull	<b>4,02*</b>	
<b>CG5325</b>	MAxM1A	1,69	-
	MAxMAnull	0	
<b>CG8064</b>	MAxM1A	<b>25,534**</b>	64 <sup>•</sup> , 65 <sup>••</sup> , 67 <sup>•</sup> , 75 <sup>•</sup> , 89 <sup>•</sup> , 100 <sup>••</sup> , 106 <sup>•</sup>
	MAxMAnull	<b>2,79*</b>	
<b>MLC</b>	MAxM1A	<b>14,268*</b>	47 <sup>•</sup>
	MAxMAnull	<b>3,098*</b>	
<b>Porin</b>	MAxM1A	1,07	-
	MAxMAnull	0,047	
<b>Elp</b>	MAxM1A	0,941	-
	MAxMAnull	0	
<b>Troponin C</b>	MAxM1A	0	-
	MAxMAnull	0	
<b>UQCRX</b>	MAxM1A	3,281	-
	MAxMAnull	0	
<b>Wings Up</b>	MAxM1A	0	-
	MAxMAnull	0	
<b>TCTP</b>	MAxM1A	0,047	-
	MAxMAnull	0	

LTR ( *log-likelihood ratio test* )**Relaxed Branch-site test**  $p < 0,01^*$ ;  $p < 10^{-8}^{**}$ **Strict Branch-site test** (valores críticos) 2.71 (alfa = 0.05)\*; 5.41 (alfa = 0.01)\*\*

• se a probabilidade for posterior a 95%, e •• se a probabilidade for posterior a 99%.

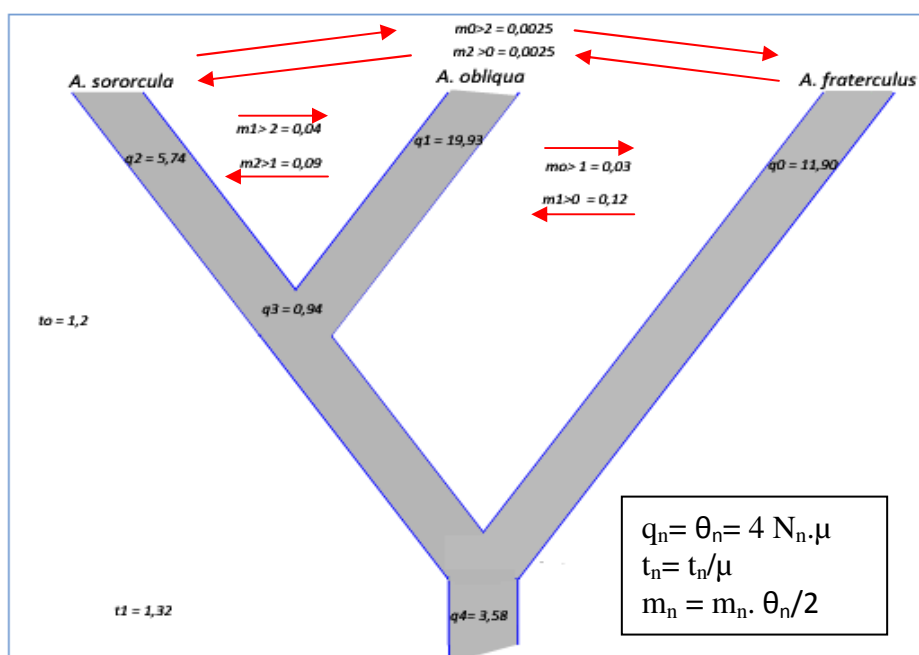
#### 4.4 – Análises de Isolamento e Fluxo gênico

Padrões históricos de isolamento populacional e fluxo gênico foram investigados através do programa IMA2. A investigação da separação entre três ou mais espécies no IMA2 requer uma topologia da árvore filogenética das espécies sendo investigadas. As três topologias possíveis entre *A. fraterculus*, *A. obliqua* e *A. sororcula* foram consideradas para análises distintas no IMA2 para fornecer estimativas de máxima verossimilhança do tamanho populacional, tempo de divergência ( $t = t\mu_i$ ) e taxas de

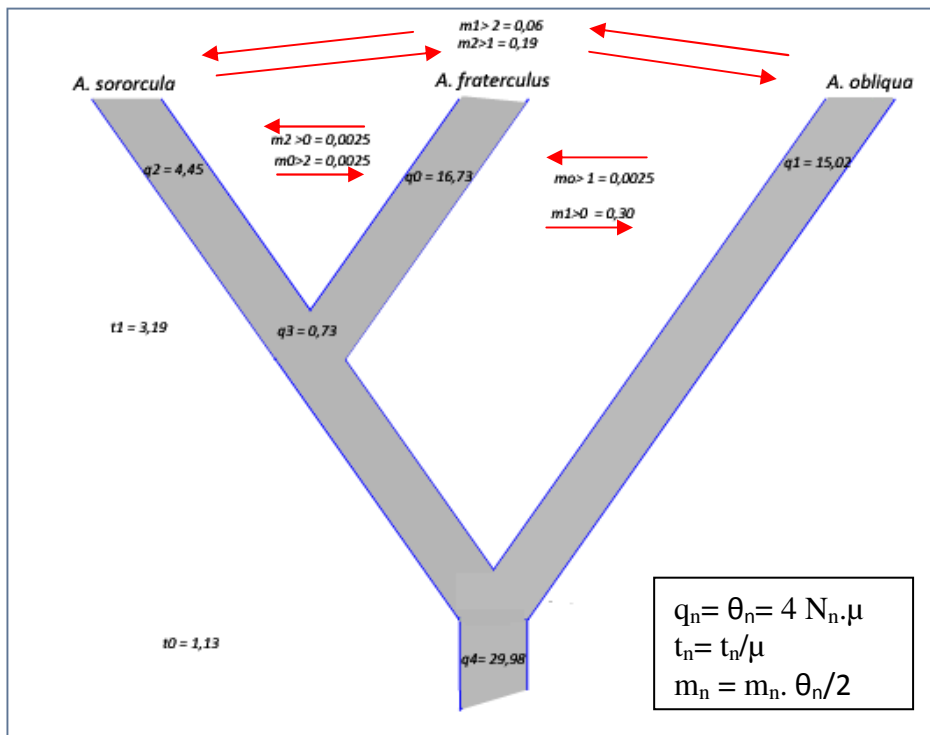
mutação ( $\mu$ ) e migração ( $m_n \cdot q_n / 2$ ) (Tabelas 14 e 15). A média da taxa de mutação obtida no IMA 2 e utilizada nos cálculos foi 0,0000007891.

As Figuras 29, 30 e 31 mostram os resultados obtidos pelas análises do programa IMA2 para os três modelos de árvore utilizando as três espécies de moscas das frutas do grupo *fraterculus* nas quais, respectivamente, as espécies *A. fraterculus*, *A. obliqua* e *A. sororcula* foram consideradas como sendo a espécie mais distante das outras duas em cada modelo.

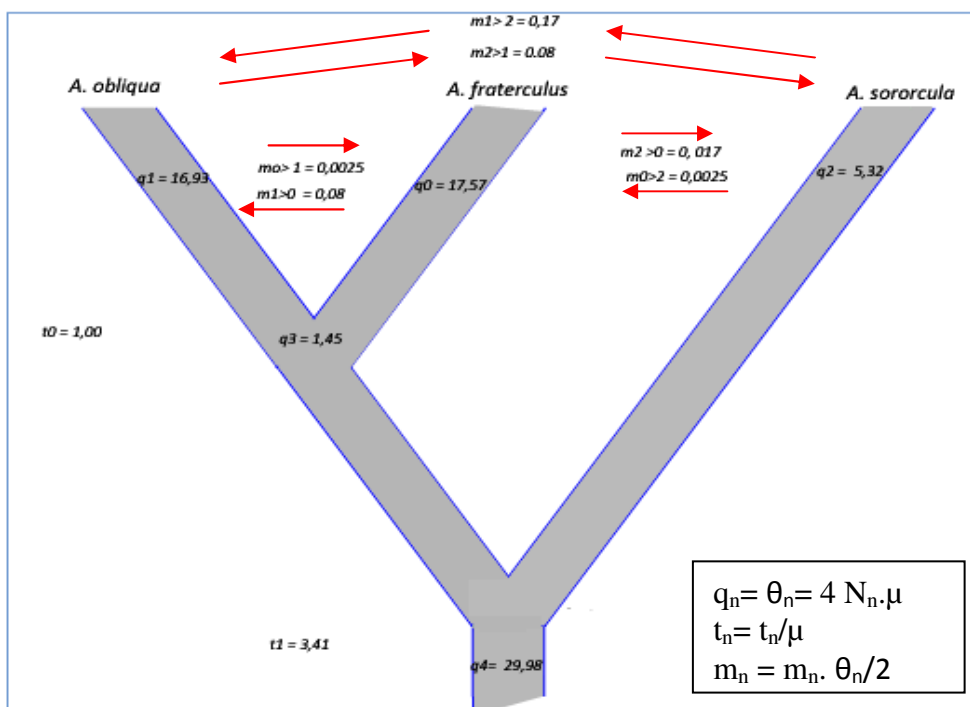
Podemos observar que os valores de tamanho populacional das três espécies e tempo de divergência são bem próximos para os três modelos de árvore utilizando as três espécies juntas. Os valores de migração por geração apresentaram-se baixos na maior parte das vezes, com exceção das migrações de *A. fraterculus* para *A. obliqua* nos dois primeiros modelos e de *A. sororcula* para *A. obliqua* no último modelo com as três espécies juntas. Esses resultados nos mostram que nos três modelos sempre ocorre uma migração significativa entre uma das espécies próximas na árvore com a espécie que este mais separada. Indicando que talvez esse não seja o melhor modelo de árvore para estas espécies. No entanto, apesar de termos obtido para os modelos analisados baixos valores de ESS (*Effective Sample Size*), desses três modelos, o que apresentou os maiores valores de ESS foi o com *A. sororcula* mais separada das outras duas. Os valores de  $q$  quando analisados com os do teste de Fs de Fu (Tabelas 4, 5, 6, 7 e 8) corroboram a hipótese de que essas espécies estão, possivelmente, em expansão populacional.



**Figura 29** – Resultados obtidos pelo programa IMA2 para o modelo de árvore onde *A. fraterculus* se encontra mais separada de *A. obliqua* e *A. sororcula*.

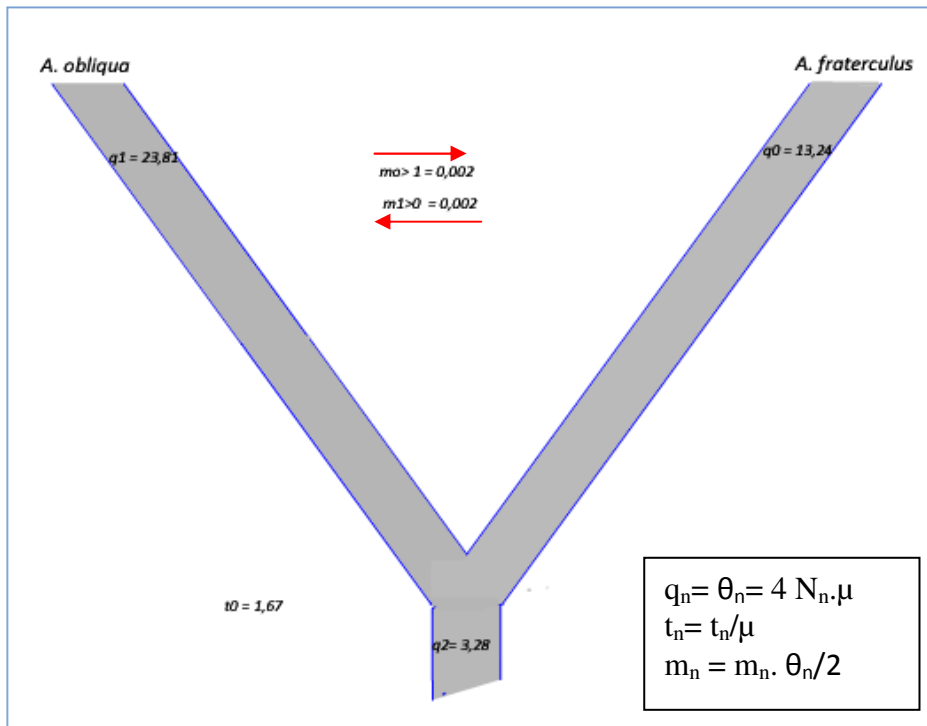


**Figura 30** – Resultados obtidos pelo programa IMA2 para o modelo de árvore onde *A. obliqua* se encontra mais separada de *A. fraterculus* e *A. sororcula*.

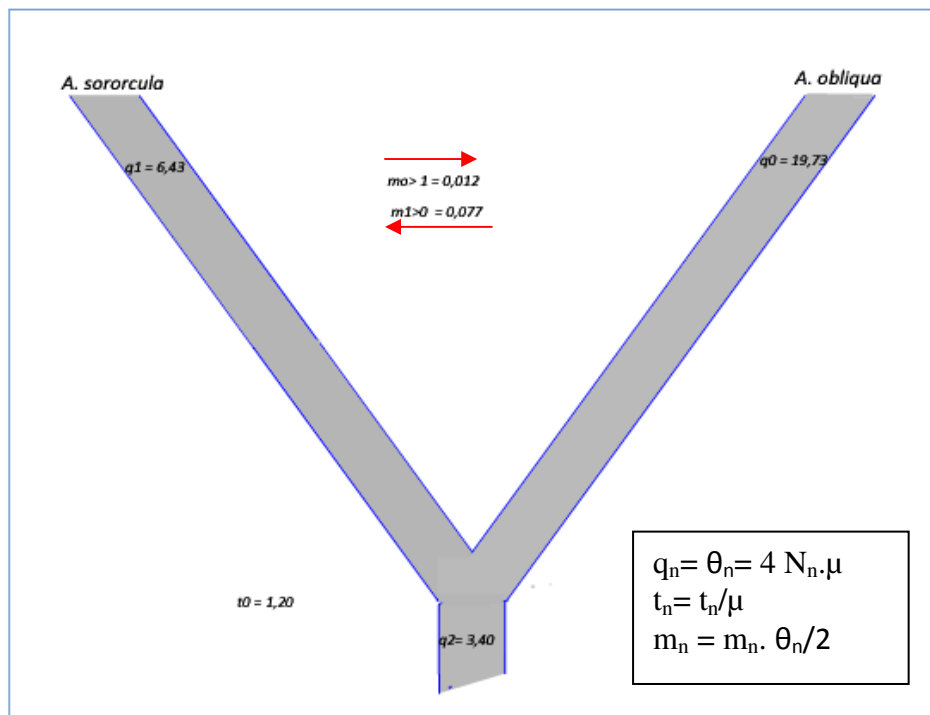


**Figura 31** – Resultados obtidos pelo programa IMA2 para o modelo de árvore onde *A. sororcula* se encontra mais separada de *A. fraterculus* e *A. obliqua*.

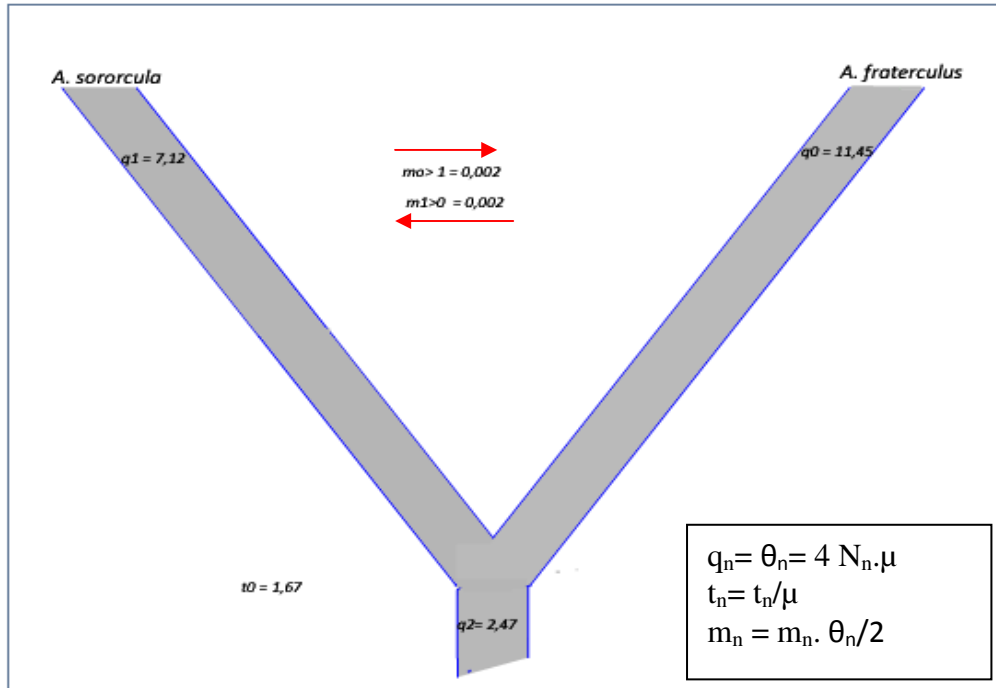
Uma vez que os graus de liberdade para cálculo concomitante de *theta*, tamanho efetivo e tempo de divergência no IMA2 quando temos mais de duas populações aumenta sobremaneira, realizamos estas estimativas também par-a-par buscando contrastar os valores obtidos nestas estimativas distintas. As **Figuras 32, 33 e 34** mostram, respectivamente, os resultados das análises feitas com *A. fraterculus* e *A. obliqua*; *A. obliqua* e *A. sororcula* e *A. fraterculus* e *A. sororcula*. Os valores de tamanho populacional, tempo de divergência ( $t = t\mu_i$ ) e migração ( $m_n \cdot q_n / 2$ ) (**Tabelas 14 e 15**) foram feitos considerando-se a média da taxa de mutação por ano por região de 0,0000007891, que é a média harmônica dos valores individuais calculados para cada um dos genes aqui estudados. Nas análises par a par os valores de tempo de divergência e tamanho populacional foram próximos para os três modelos, no entanto, foram, no geral, mais elevados que os resultados obtidos nas análises utilizando as três espécies juntas. É interessante contrastar os valores de migração por geração estimados para modelos com três espécies e espécies par-a-par. Enquanto as estimativas foram baixas para todos os modelos par-a-par (valores de M em geral inferiores a 0.25), algumas das estimativas no modelo com o trio de espécies apresentou valores altos (mesmo M=2.25), revelando fluxo gênico relevante e talvez suficiente para, em modelos na ausência de seleção, impedir a diferenciação. No entanto, em todos estes casos, os valores significativos de M coincidem com a migração entre o grupo externo e uma das espécies do grupo interno. Por exemplo, no modelo 1, M=1,19 entre *A. fraterculus* (externa) e *A. obliqua*. No modelo 2, M=2,25 entre *A. obliqua* (externa) e *A. fraterculus* e no modelo 3, M=1,43 entre *A. sororcula* (externa) e *A. obliqua*. Embora isso possa ser uma coincidência, ou reflexo de valores ainda baixos de número de corridas para estimativas confiáveis de parâmetros populacionais é muito mais provável que isso seja reflexo da inadequação dos modelos de árvores para representar o padrão de diversidade nuclear aqui estimado. Dessa forma, taxas de fluxo gênico inferidos podem ser reflexo de topologias errôneas das árvores trazendo mais divergência do que deveria acontecer e tornando segregação de polimorfismos ancestrais em inferências de fluxo gênico recorrente. Uma indicação desta possibilidade são as estimativas par-a-par destes mesmos parâmetros, que revelaram valores bem mais baixos de fluxo gênico entre as espécies apesar das estimativas de tamanho efetivo tenham ficado em valores não muito diferenciados. As estimativas de *t* também sofrem com estas estimativas distintas entre os modelos com trio ou duo de espécies, uma vez que a diferenciação nas estimativas de trio ocorreu há aproximadamente o dobro do tempo das estimativas de duo.



**Figura 32** – Resultados obtidos pelo programa IMA2 para o modelo de árvore com as espécies *A.fraterculus* e *A. obliqua*.



**Figura 33** – Resultados obtidos pelo programa IMA2 para o modelo de árvore com as espécies *A.obliqua* e *A. sororcula*.



**Figura 34** – Resultados obtidos pelo programa IMA2 para o modelo de árvore com as espécies *A.fraterculus* e *A. sororcula*.

**Tabela 14** – Tempo de divergência (em milhões de anos) e tamanho efetivo de cada população (em milhões) para cada modelo de árvore.

	t0	t1	Nef (1)	Nef (2)	Nef (3)
<b>Modelo 1</b>	1,5	1,7	3,7	6,3	1,8
<b>Modelo 2</b>	1,49	4	5,3	4,7	1,4
<b>Modelo 3</b>	1,27	4,3	5,5	5,3	1,7
<b>Modelo 4</b>	2,1	-	4,2	7,5	-
<b>Modelo 5</b>	1,5	-	-	3,6	2,4
<b>Modelo 6</b>	2,1	-	6,2	-	2

**Modelo 1** – *A. fraterculus* separada de *A. obliqua* e *A. sororcula*.

**Modelo 2** – *A. obliqua* separada de *A. fraterculus* e *A. sororcula*.

**Modelo 3** – *A. sororcula* separada de *A.obliqua* e *A. sororcula*.

**Modelo 4** – *A.fraterculus* e *A. obliqua*.

**Modelo 5** - *A.obliqua* e *A. sororcula*.

**Modelo 6** – *A.fraterculus* e *A. sororcula*.

**Nef (1)** – *A. fraterculus*.

**Nef (2)** – *A. obliqua*.

**Nef (3)** – *A.sororcula*.

**Tabela 15** – Taxa de migração por gene por geração, na coalescência.

	<b>m0&gt;1</b>	<b>m0&gt;2</b>	<b>m1&gt;0</b>	<b>m1&gt;2</b>	<b>m2&gt;0</b>	<b>m2&gt;1</b>
Modelo 1	0,18	0,01	1,19*	0,40	0,007	0,26
Modelo 2	0,02	0,02	2,25*	0,45	0,005	0,40
Modelo 3	0,02	0,02	0,67	1,43*	0,045	0,21
Modelo 4	0,01	-	0,03	-	-	-
Modelo 5	0,12	-	0,25	-	-	-
Modelo 6	0,007	-	0,01	-	-	-

**Modelo 1** – *A. fraterculus* separada de *A. obliqua* e *A. sororcula*.

**Modelo 2** – *A. obliqua* separada de *A. fraterculus* e *A. sororcula*.

**Modelo 3** – *A. sororcula* separada de *A. obliqua* e *A. sororcula*.

**Modelo 4** – *A. fraterculus* e *A. obliqua*.

**Modelo 5** – *A. obliqua* e *A. sororcula*.

**Modelo 6** – *A. fraterculus* e *A. sororcula*.

\*valores significativos.

#### 4.5 – Comparação das árvores de genes e de espécies.

Utilizamos o programa MS para obtermos 100 árvores simuladas para cada *locus* e modelo de árvore de espécie, usando os dados de  $t$  e  $\theta$  obtidos no programa IMA2 para estimarmos as distâncias entre os ramos para cada modelo. O MS, então, nos forneceu as simulações de árvores que transformamos em sequências de nucleotídeos pelo programa seq-gen utilizando os modelos evolutivos estimados para aquele *locus* específicos. Estas seqüências simuladas foram então rodadas no PhyML para gerar as melhores topologias inferidas por máxima verossimilhança para serem usadas no MESQUITE. Esta estratégia busca gerar uma distribuição dos valores de coalescência profunda, computada a estocasticidade do processo evolutivo, de acordo com um determinado padrão evolutivo, de forma que os valores encontrados nas árvores podem ser contrastados com estas estimativas para fornecer um teste de adequação dos mesmos aos padrões evolutivos usados para geral tal distribuição. Comparamos os valores de coalescência profunda das árvores simuladas de cada *locus* para cada um dos modelos com as árvores de genes “reais” para cada *locus*, consideramos significativos os valores de coalescência profunda das árvores de genes “reais” que se encaixavam no gráfico que mostra a coalescência profunda para cada uma das árvores simuladas, utilizamos uma significância de 0,05 e obtivemos poucos valores significativos (**Tabela 16**). Apenas os genes **TCTP**, para a árvore em que *A. fraterculus* é mais distante das outras duas espécies e **Cyclophylin**, em que os valores de coalescência profunda se mostraram

significativos quando comparamos as árvores simuladas com as árvores de genes para os três modelos de árvore.

Podemos observar como exemplo, o gráfico que mostra os valores de coalescência profunda para as 100 árvores simuladas pelo programa MS, para o gene **Elongation Protein (Figura 35)**. Neste caso, o valor de coalescência profunda obtida para a árvore de espécie não se mostrou significativo quando comparado com os valores de coalescência profunda das árvores simuladas para este gene no modelo de árvore onde *A. obliqua* está mais distante das outras espécies do gênero *Anastrepha*. Em contrapartida, o gráfico da **Figura 36** mostra o resultado significativo para a comparação da coalescência profunda das árvores simuladas para o gene **Cyclophylin** com sua respectiva árvore de genes para o modelo de árvore em que *A. obliqua* se encontra mais distante de *A. fraterculus* e *A. sororcula*. Na **Tabela 16** podemos observar que os valores de coalescência profunda para as árvores simuladas foram em geral maiores nas análises utilizando o modelo 1, em que *A. fraterculus* se encontra mais separada das outras duas espécies.

**Tabela 16** – Dados da mediana da coalescência profunda para as árvores simuladas para cada locus comparadas com a coalescência profunda das árvores de genes para cada um dos três modelos de árvores de espécie.

<b>Locus</b>	<b>Modelo 1</b>	<b>Modelo 2</b>	<b>Modelo 3</b>
CG7009	17/55	12/55	13/55
CG7203	18/29	13/20	14/29
CG8064	18/26	14/26	13/26
CG5325	17/41	14/ 41	13/41
Elp	18/32	16/32	15/32
Cyclophylin	18/13*	14/13*	13/13*
TCTP	67/41*	12/41	13/41
Troponin C	19/71	13/71	14/71
MLC	17/51	13/51	13/51
Porin	17/74	13/74	12/74
UQCRX	19/63	14/63	13/63
Wings Up	17/32	12/32	12/32

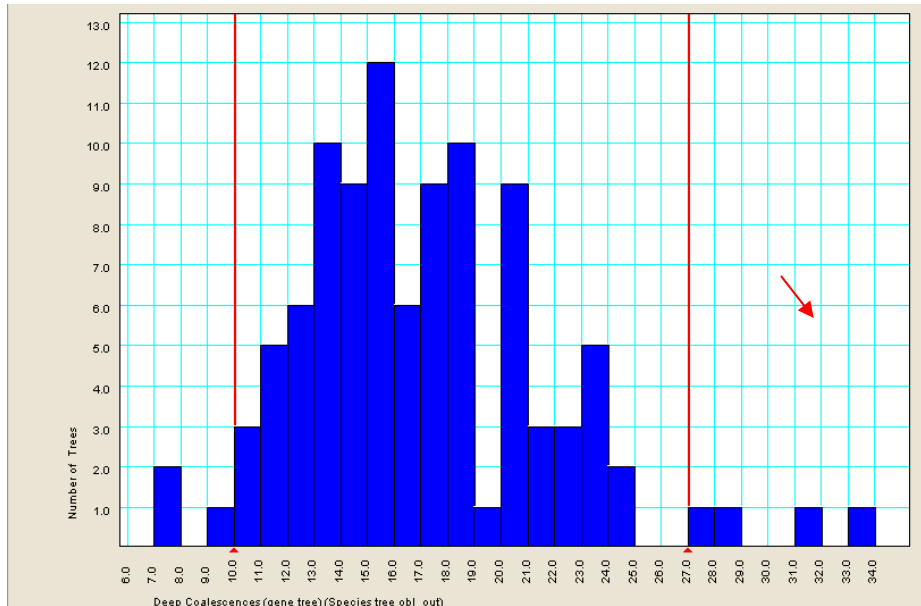
**Modelo 1** – *A. fraterculus* separada de *A. obliqua* e *A. sororcula*.

**Modelo 2** – *A. obliqua* separada de *A. fraterculus* e *A. sororcula*.

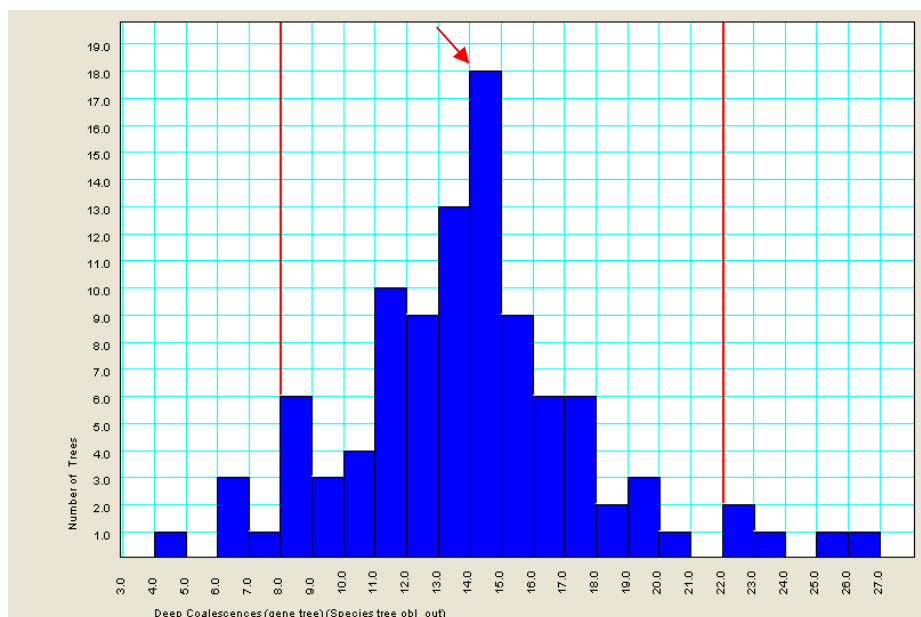
**Modelo 3** – *A. sororcula* separada de *A.obliqua* e *A. sororcula*.

\*valores significativos.



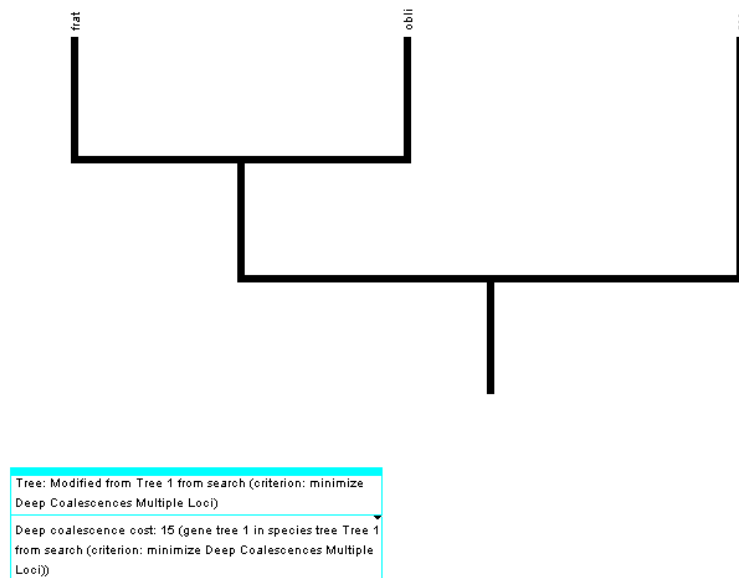


**Figura 35** – Valores de coalescência profunda para as árvores simuladas para o gene Elongation Protein no modelo de árvore onde *A. obliqua* está mais separada das outras duas espécies. A seta indica o valor da coalescência profunda da árvore de genes real para este locus. O intervalo de significância ( $p=0,05$ ) está delimitado entre as linhas vermelhas.



**Figura 36** – Valores de coalescência profunda para as árvores simuladas para o gene Cyclo no modelo de árvore onde *A. obliqua* está mais separada das outras duas espécies. A seta indica o valor da coalescência profunda da árvore de genes real para este locus. O intervalo de significância ( $p=0,05$ ) está delimitado entre as linhas vermelhas.

As topologias geradas por cada um dos genes individuais podem também ser utilizadas para avaliar o grau de concordância entre os sinais destes genes para explicar os processos evolutivos subjacentes. Dessa forma, consideramos cada gene como indicador independente do processo evolutivo e buscamos no programa MESQUITE uma árvore de espécies que minimize o número de coalescências profundas considerando-se todos estes genes em conjunto. A árvore obtida, através da minimização das coalescências profundas, mostrou que *A. sororcula* se encontra mais distante das outras duas espécies do gênero *Anastrepha* (**Figura 37**).



**Figura 37** – Árvore de espécies obtida pela minimização da coalescência profunda (*minimizing deep coalescence*).

## 5. Conclusões

As análises, no geral, indicam que as três espécies estudadas são realmente linhagens distintas. As análises de polimorfismos mostram valores bem altos para as três espécies, analisadas juntas e separadamente. No entanto, os valores para os testes de neutralidade, em geral, não foram significativos para a maioria das regiões gênicas amplificadas, com exceção dos valores de  $F_s$  e  $F_u$  que mostraram significância para a maioria dos *loci*, nos indicando uma provável expansão populacional, hipótese corroborada pelos valores de *theta* obtidos nas análises realizadas pelo programa IMA2.

Ao fazermos as redes de haplótipos para cada gene observamos que a espécie *A. sororcula* é a que mais se difere das outras duas espécies e que para alguns genes ocorre um agrupamento dos haplótipos das mesmas espécies, além disso, o gene **CG7009** e **Cyclophylin** nos mostraram a presença de polimorfismos específicos para a espécie *A. sororcula*, separando os haplótipos desta espécie dos haplótipos das outras duas. Considerando-se que o **CG7009** nas análises de seleção mostrou evidências de seleção positiva, é importante investigá-lo mais a fundo como potencial marcador genético da espécie *A. sororcula*. Pelos testes de seleção detectamos também, a provável evolução por seleção positiva em mais um dos 12 *loci* analisados.

As análises do IMA2 utilizando-se do trio de espécies juntas mostrou-se ainda pouco significativa devido ao fato de termos corrido um número reduzido de gerações. No entanto, as análises par a par das espécies mostrou valores altos de ESS (*Effective Sample Size*), indicando uma alta mistura dos dados, fazendo com que estes resultados sejam bem mais significativos.

A análise de minimização da coalescência, feita pelo MESQUITE, indicou *A. sororcula* como sendo a espécie mais separada das outras duas. Estes dados são concordantes com a análise de polimorfismos fixos e compartilhados entre as espécies, que revelou que *A. sororcula* apresenta mais polimorfismos fixados do que as outras duas espécies, que por sua vez apresentam mais polimorfismos compartilhados. No entanto, a comparação das coalescências profundas das árvores simuladas para cada uma das regiões amplificadas, nas três topologias testadas, revelou alta discordância mostrando que os modelos testados não são os mais adequados para explicar os dados obtidos. Isso pode ter ocorrido pelos valores dos ramos que utilizamos para a construção das árvores simuladas estarem sub ou super estimados, mas também pode ser porque o próprio modelo não está considerando todos os processos evolutivos ocorridos. Todas

as topologias testadas no IMA2 indicaram valores positivos de fluxo gênico entre as espécies. Em alguns casos estes valores de fluxo gênico foram acima de 1, valor considerado suficiente para impedir a diferenciação entre linhagens. Contudo, estes valores foram sempre observados entre a espécie considerada externa e alguma do grupon interno, o que pode sugerir que topologias dicotômicas podem não ser as mais adequadas para explicar diferenciação deste grupo de espécies. Embora os resultados das análises par a par das espécies, para os quais obtivemos uma maior confiança por necessitarem de menores graus de liberdade para inferência, mostram baixos valores de migração entre as espécies, não podemos rejeitar a existência de algum fluxo gênico entre estas espécies na natureza. Apesar de não termos conhecimento de zonas híbridas ou híbridos entre estas espécies na natureza, há evidências de formação de híbridos entre estas espécies em laboratório (Henning e Matioli, 2006). Isso pode indicar de fato que o processo de diferenciação das espécies deste grupo ainda seja insipiente e pode reforçar o papel de certos genes específicos na diferenciação destas espécies. Alguns estudos têm revelado que o genoma pode ser “poroso” ao longo da distribuição da espécie de forma que alguns marcadores que não estão diretamente envolvidos com processo de especiação podem segregar entre as espécies e tendem a seguir mais padrão geográfico do que específico, enquanto outros não (Kane *et al.*, 2009), o que seria uma violação dos modelos tradicionais alopátricos de especiação. Outras espécies de Tephritidae estão associados a processos de especiação simpátricos em que há papel relevante de adaptação a hospedeiros e seleção sexual (Feder *et al.*, 2005), o que pode ser padrão similar ao aqui encontrado. Uma investigação maior de diferentes genes, particularmente genes com taxas evolutivas mais altas e uma expansão das amostras destas espécies de forma a considerar sua distribuição geográfica pode nos auxiliar a resolver tal questão.

## 6. Referências

- ARAÚJO, E.L. ; ZUCCHI, R.A. (2006). Medidas do acúleo na caracterização de cinco espécies de *Anastrepha* do grupo *fraterculus* (Diptera: Tephritidae). *Neotropical Entomology* 35: 1-9.
- ARNQVIST, G.; EDVARDSSON, M.; FRIBERG, U. ; NILSSON, T. (2000). Sexual conflict promotes speciation in Insects. *Proc. Nat. Acad. Sci. USA*. 97: 10460-10464.
- BEEMAN, R.W.; BROWN, S.J. (1999). RAPD-based genetic linkage maps of *Tribolium castaneum*. *Genetics* 153, 333–338.
- BUTLIN, R. ; RITCHIE, M.G. (2001). Searching for speciation genes. *Nature*. **412**: 31-33.
- CASTELLOE, J.M. ; TEMPLETON, A.R. (1994). Root Probabilities for Intraspecific Gene Trees under Neutral Coalescent Theory. *Mol. Phylogenet. Evol.* **3**: 102--113.
- 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.
- CIVETTA, A. ; SINGH, R.S. (1995). High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species. *J. Mol. Evol.* 41: 1085–1095.
- CLARK, N. L. ; SWANSON, W. J. (2005). Pervasive adaptive evolution in primate seminal proteins. *PLoS Genetics*, 1: e35.
- CLEMENT, M.; POSADA, D. ; CRANDALL, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, 9: 1657-1659.
- CLINE, J. ; BRAMAN, J. C. ; HOGREFE, H. H. (1996). PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research*, 24: 3546-3551.
- CORTNER, J.; WOUDE G.F.V. (1997).Essentials of molecular biology – cDNA libraries. In: Cancer: Principles and practice of oncology. DeVita V.T. Jr., Hellman S. and Rosenberg S.A. (ed), 5th ed, Lippincott – Raven Publishers, Philadelphia.
- COYNE, J.A.; ORR, H.A. (2004). *Speciation*. Sunderland, MA: Sinauer. 545 pp.
- CRANDALL, K.A.; TEMPLETON, A.R. (1993). Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*. v. 134, n. 3, p. 959–969.
- DALLAS-YANG, Q.; JIANG, G. ; SLADEK, F.M. (1998). Avoiding false positives in colony PCR. *BioTechniques* **24**: 580-582.

- DE BRITO, R.A.; MANFRIN, M.H. ; SENE, F.M. (2002). Mitochondrial DNA phylogeography of Brazilian populations of *Drosophila buzzatii*. *Genetics and Molecular Biology*, 25, 2, 161-171.
- DEGNAN, J. H. ; SALTER, L.A. (2005). Gene tree distributions under the coalescent process. *Evolution* 59:24–37.
- DUARTE, A. L.; MALAVASI, A. (2000) Tratamentos quarentenários. In: MALAVASI, A.; ZUCCHI, R. A. (Ed.). Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado. Ribeirão Preto: FAPESP-Holos. p. 187-192.
- EBERHARD, W. G. (1985). Sexual Selection and the Evolution of Animal Genitalia. Cambridge, Mass.: Harvard University Press.
- FAY, J. C. ; WU, C.-I. (2000). Hitchhiking Under Positive Darwinian Selection. *Genetics*, 155: 1405-1413.
- FEDER, J. L.; BERLOCHER, S. H. ; OPP, S. B. (1998). Sympatric host race formation and speciation in *Rhagoletis* (Diptera: Tephritidae): a tale of two species for Charles D., pp. 408–441 in *Genetic Structure in Natural Insect Populations: Effects of Host Plants and Life History*, edited by S. MOPPER and S. STRAUSS. Chapman & Hall, New York.
- FEDER, J.L.; XIE, X.; RULL, J.; VELEZ, S.; FORBES, A.; LEUNG, B.; DAMBROSKI, H.; FILCHAK, K.E.; ALUJA, M. (2005). Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*. *PNAS* 102: 6573-6580. Fu, Y.-X. e Li, W.-H. (1993). Statistical tests of neutrality of mutations. *Genetics*, 133: 693-709.
- FU, X.-Y. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915-925.
- FU, Y.-X. ; LI, W.-H. (1993). Statistical tests of neutrality of mutations. *Genetics*, 133: 693-709.
- GALINDO, B. E., VACQUIER, V. D. ; SWANSON, W. J. (2003). Positive selection in the egg receptor for abalone sperm lysin. *Proceedings of the National Academy of Sciences*, 100: 4639-4643.
- GOMULSKI, L.M.; DIMOPOULOS G.; XI, Z.; SOARES, M.B., BONALDO, M.F.; MALACRIDA, A.R. ; GASPERI, G. (2008) Gene discovery in an invasive tephritid model pest species, the Mediterranean fruit fly, *Ceratitis capitata*. *BMC Genomics* 9: 243.
- GONÇALVES, V.R. (2010). Identificação de genes com alta taxa evolutiva em tecidos reprodutivos femininos de moscas-das-frutas *Anastrepha obliqua*. Tese (Mestrado em Genética e Evolução) – Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, São Paulo.
- GRAY, D.A.; HUANG, H.; KNOWLES, L. L. (2008). Molecular evidence of a peripatric origin for two sympatric species of field crickets (*Gryllus rubens* and *G. texensis*) revealed from coalescent simulations and population genetic tests. *Mol. Ecol.* 17, 3836–3855.

- GUINDON, S. ; GASCUEL, O. (2003). A simple, fast and accurate algorithm to estimate large phylogenies by max-mum likelihood *Systematic Biology* 52(5):696-704.
- HALL, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41:95-98.
- HEATH, R. R.; LANDOLT, P.J.; ROBACKER, D.C.; DUEBEN, B.D. ; EPSKY,N. (2000). Sexual pheromones of tephritid flies: clues to unravel phylogeny and behavior. Aluja, M. e Norrbom, A.L., editors. *Fruit flies (Tephritidae): phylogeny and evolution of behavior.* 793–809.CRC. Boca Raton, FL.
- HENNING, F. ; MATIOLI, S.R. (2006). Mating time of the West Indian fruit fly *Anastrepha obliqua* (Macquart) (Diptera: Tephritidae) under laboratory conditions. *Neotropical Entomology* 35: 145-148.
- HEY, J. ; NIELSEN, R. (2004). Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167:747-760.
- HEY, J. ; NIELSEN, R. (2007). Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *PNAS* 104:2785–2790.
- HEY, J.(2009) Isolation with migreation models for more than two populations. *Mol. Biol. Evol.*, doi:10.1093/molbev/msp296.
- HINOJOSA-MOYA, J.; XOCONOSTLE-CAZARES, B.; PIEDRA-IBARRA, E.;MENDEZ-TENORIO, A.; LUCAS, W. J. ; RUIZ-MEDRANO, R. (2008). Phylogenetic and structural analysis of translationally controlled tumor proteins. *Journal of molecularevolution*, v. 66, p. 472-483.
- HUDSON, R. (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, 18, 337–338
- HUDSON, R. R. ; COYNE, J.A. (2002). Mathematical consequences of the genealogical species concept. *Evolution* 56:1557–1565.
- HUDSON, R. R. ; TURELLI, M. (2003). Stochasticity overrules the “three-times” rule: Genetic drift, genetic draft, and coalescence times for nuclear *loci* versus mitochondrial DNA. *Evolution* 57:182–190.
- KAMINSKI, A.C.F.(2007). Identificação de genes expressos no aparelho reprodutivo de fêmeas de moscas-das-frutas *Anastrepha fraterculus*. Monografia – Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, São Paulo.
- KANE, N.C.; KING, M.G.; BARKER, M.S.; RADUSKI, A.; KARRENBERG, S.; YATABE, Y.; KNAPP, S.J.; RIESEBERG, L.H. (2009). Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution* 63: 2061-2075.

- KELLEHER E.S.; SWANSON, W.J.; MARKOW, T.A. (2007). Gene duplication and adaptive evolution of digestive proteases in *Drosophila arizonae* female reproductive tracts. *PLoS Genet.* 3:e148.
- KNOWLES, L. L. ; CARSTENS, B.C. (2007). Inferring a population divergence model for statistical-phylogeographic tests in montane grasshoppers. *Evolution* 61:477–493.
- LI, W.-H.; GOJOBORI, T.; NEI, M. (1981). Pseudogenes as a paradigm of neutral evolution *Nature* 292:237-239.
- LIS, J.T. ; SCHLEIF, R. (1975). Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. *Nucleic Acids Res* 2: 383-389.
- LORENZEN, M.D.; DOYUNGAN, Z.; SAVARD, J.; SNOW, K.; CRUMLY, L.R.; SHIPPY, T.D.; STUART, J.J.; BROWN, S.J.; BEEMAN, R.W. (2005). Genetic Linkage Maps of the Red Flour Beetle, *Tribolium castaneum*, Based on Bacterial Artificial Chromosomes and Expressed Sequence Tags. *Genetics* 170: 741-747.
- MADDISON, W. P. (1997). Gene trees in species trees. *Syst. Biol.* 46:523–536.
- MADDISON, W. P. ; KNOWLES, L.L. (2006). Inferring phylogeny despite incomplete lineage sorting, *Syst. Biol.* 55:21–30.
- MADDISON, W.P. ; MADDISON, D.R. (2008). Mesquite: a modular system for evolutionary analysis. Version 2.5, build j55.
- MALAVASI, A. ; MORGANTE, J.S. (1982). Genetic variation in natural populations of *Anastrepha* (Diptera: Tephritidae). *Rev. Bras. Genet.* 5: 263-278.
- MALAVASI, A. ; ZUCCHI, R.A. (2000). *Moscas-das-frutas de importância econômica no Brasil: Conhecimento básico e aplicado*. First ed. Ribeirão Preto: Holos Editora
- MATIOLI, S.R.; MORGANTE, J.S. ; MALAVASI, A. (1986). Genetical and biochemical comparisons of alcohol dehydrogenase isozymes from *Anastrepha fraterculus* and *A. obliqua*: evidence for gene duplication. *Biochem. Genet.* 24: 13-24.
- MCDONALD, J. H. ; KREITMAN, M. (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, 351: 652-654.
- MORGANTE, J.S.; MALAVASI, A. ; BUSH, G.L. (1980). Biochemical systematics and evolutionary relationships of Neotropical *Anastrepha*. *Ann. Entomol. Soc. Am.* 73: 622-630.
- MORGANTE, J.S.; MALAVASI, A. ; PROKOPY, R.J. (1983). Mating behavior of wild *Anastrepha fraterculus* flies (Diptera, Tephritidae) on a caged host tree. *Florida Entomol.* 66: 234-241.



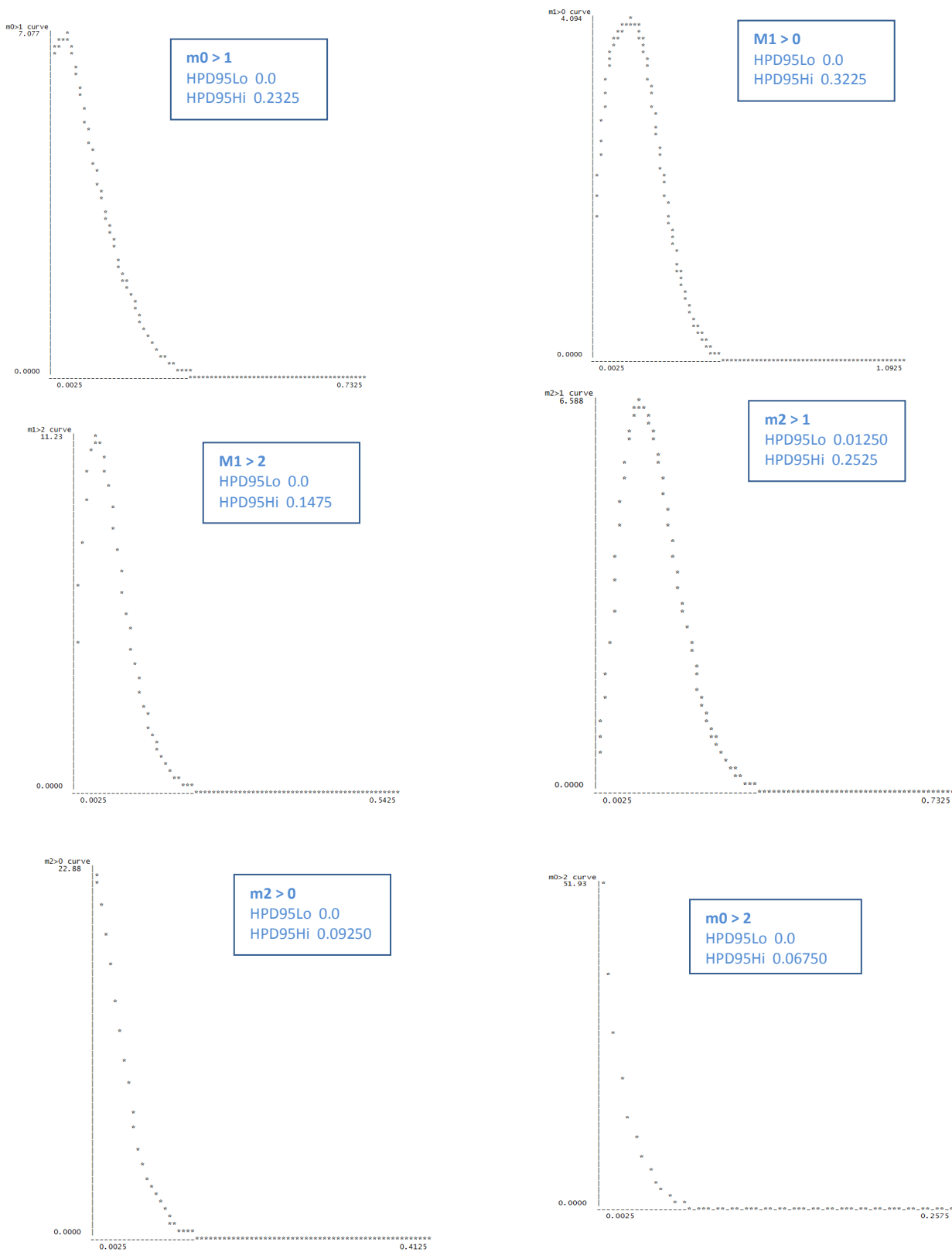
- MUELLER, J.L.; RAM, K.R.; MCGRAW, L.A.; BLOCH QAZI, M.C.; SIGGIA, E.D.; CLARK, A.G.; AQUADRO, C.F. ; WOLFNER, M.F. (2005). Cross-Species Comparison of *Drosophila* Male Accessory Gland Protein Genes. *Genetics* **171**: 131-143.
- NEI, M. (1987). Molecular Evolutionary Genetics. *New York: Columbia Univ. Press*
- Nelson, J.E. e Krawetz, S.A. (1992). Purification of Cloned and Genomic Dna by Guanidine Thiocyanate Isobutyl Alcohol Fractionation. *Anal Biochem* **207**: 197-201.
- NELSON, J.E.; KRAWETZ, S.A. (1992). Purification of Cloned and Genomic DNA by Guanidine Thiocyanate Isobutyl Alcohol Fractionation. *Anal Biochem* **207**:197-201.
- NIELSEN, R.; WAKELEY, J. (2001). Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**: 885-896.
- NORRBOM, A.L.; ZUCCHI, R.A. ; HERNÁNDEZ-ORTIZ, V. (2000). Phylogeny of the genera *Anastrepha* and *Toxotrypana* (Trypetinae: Toxotrypaninae) based on morphology. p. 299-342. In M. Aluja A.L. Norrbom (eds.), Fruit Flies (Tephritidae): Phylogeny and Evolution of Behavior. Boca Ratón, CRC Press, 944p.
- NORRBOM, A.L. (2001). Tephritidae classification table. <http://www.sel.barc.usda.gov/diptera/tephriti/Tephclas.htm>. (27 junho2001)
- NORRBOM, A. L. (2004). Host plant database for *Anastrepha* and *Toxotrypana* (Diptera: Tephritidae: Toxotrypanini). Diptera Data Dissemination Disk (CD-ROM) 2: 5,984 records.
- PANHUIS, T.M. ; SWANSON, W.J. (2006). Molecular Evolution and Population Genetic Analysis of Candidate Female Reproductive Genes in *Drosophila*. *Genetics* genetics.
- POND, S.L.K.; FROST, S.D.W. ; MUSE, S.V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21** (5), 676–679.
- POSADA, D. ; CRANDALL, K.A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* **14** (9): 817-818.
- POWELL, J.R. (1997). Progress and Prospects in Evolutionary Biology: The *Drosophila* Model. *Oxford University Press*. Pp 381-395.
- RAMOS-ONSINS, S.E. ; ROZAS, J. (2002). Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol.* **19**: 2092-2100.
- RITCHIE, M.G. ( 2007). Sexual Selection and Speciation . *Ann. Rev. Ecol. Evol. Syst.* **38**:79-102.
- ROETHELE, J.B.; ROMERO-SEVERSON, J. ; FEDER, J.L. (2001). Evidence for Broad-Scale Conservation of Linkage Map Relationships Between *Rhagoletis pomonella* (Diptera: Tephritidae) and *Drosophila melanogaster* (Diptera: Drosophilidae). *Ann Entomol Soc Am* **94**: 936-947.

- ROSENBERG, N. A. (2003). The shapes of neutral gene genealogies in two species: Probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution* 57:1465–1477.
- ROZAS, J.; SÁNCHEZ-DELBARRIO, J. C.; MESSEGUER, X. ; ROZAS, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 19: 2496-2497.
- SANTOS, P.D.; URAMOTO, K. ; MATIOLI, S.R. (2001). Experimental hybridization among *Anastrepha* species (Diptera: Tephritidae): Production and morphological characterization of F<sub>1</sub> hybrids. *Ann. Entomol. Soc. Am.* **94**: 717-725.
- SCHMID-HEMPEL, P. (2005). Evolutionary ecology of insect immune defenses. *Annual Review of Entomology* 50: 529-51.
- SELIVON, D. (2000). Biologia e padrões de especiação. In: *Moscas-das-frutas de importância econômica no Brasil Conhecimento Básico e Aplicado* (A Malavasi and RA Zucchi). pp. 25- 28. Holos Editora. Ribeirão Preto, SP.
- SELIVON, D.; PERONDINI, A.L.P.; MORGANTE, J.S. (2005). A Genetic-Morphological Characterization of Two Cryptic Species of the *Anastrepha fraterculus* Complex (Diptera: Tephritidae). *Ann. Entomol. Soc. Am.* **98**: 367-381.
- SMITH-CALDAS, M.R.B.; MCPHERON, B.A.; SILVA, J.G. ; ZUCCHI, R.A. (2001). Phylogenetic relationship among species of the *fraterculus* group (*Anastrepha*: Diptera: Tephritidae) inferred from DNA sequences of mitochondrial cytochrome oxidase I gene. *Neotropical Entomology* **30**: 565-573.
- SOBRINHO-JR, I.S. (2009). Evolução molecular dos genes doublesex e fruitless em moscas-das-frutas do grupo *Anastrepha fraterculus* (Diptera, Tephritidae). Tese (Doutorado em Genética e Evolução) – Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, São Paulo.
- SOLFERINI, V.N. ; MORGANTE, J.S. (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). *Ann. Entomol. Soc. Am.* **84**: 10-28.
- STROPE, C.L., ABEL, K., SCOTT, S.D.; MORIYAMA, E.N. (2009) Biological sequence simulation for testing complex evolutionary hypotheses: indel-Seq-Gen version 2.0. *Mol. Biol. Evol* 26: 2581-2593.
- SWANSON, W.J.; CLARK, A.G.; WALDRIP-DAIL, H.M. ; WOLFNER, M.F. ; AQUADRO, C.F.(2001). Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc Natl Acad Sci USA.* 98:7375–7379.

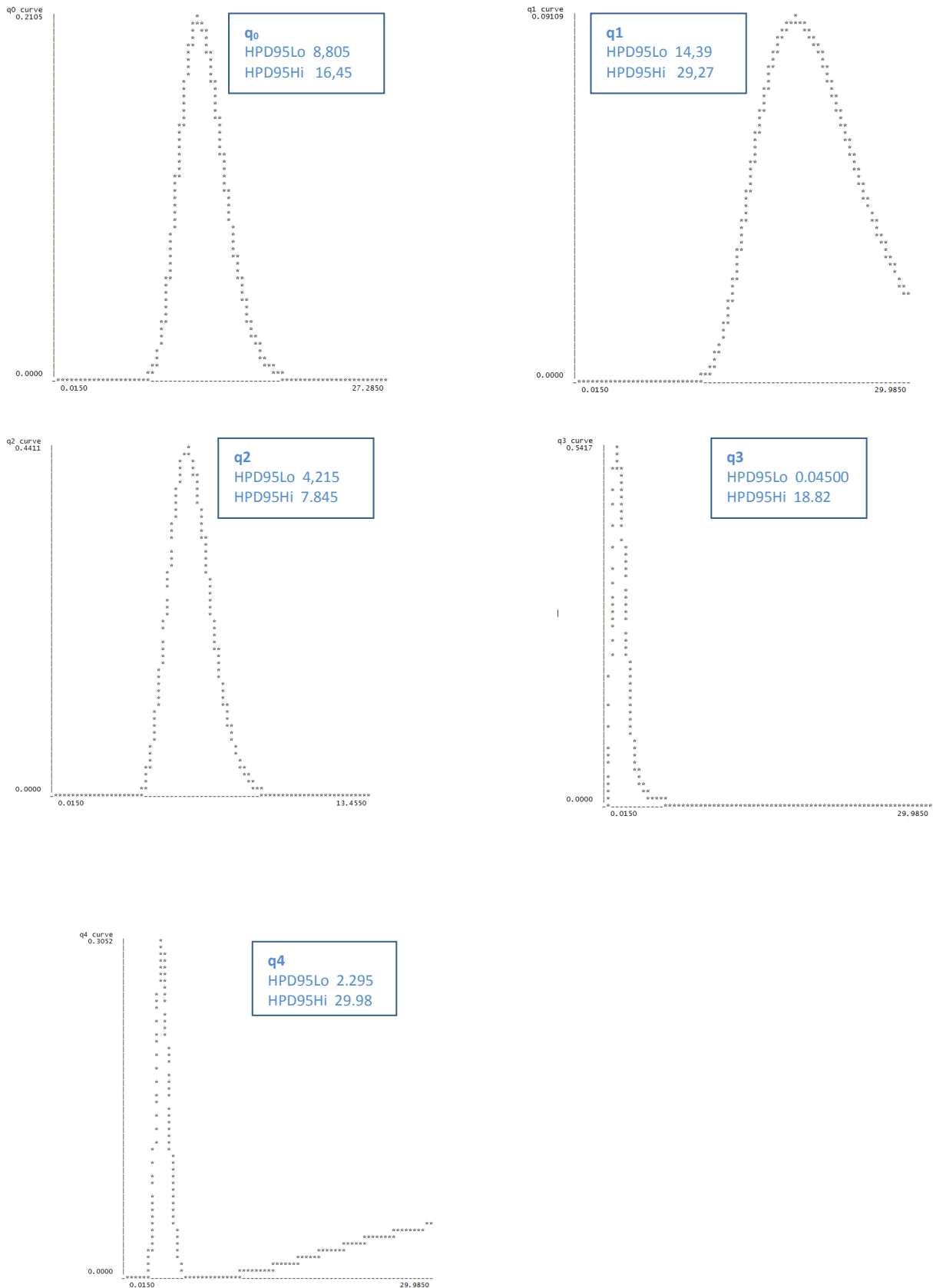
- SWANSON, W.J. ; VACQUIER, V.D. (2002). The rapid evolution of reproductive proteins. *Nat. Rev Genet.* 3:137–144.
- SWANSON, W. J.; NIELSEN, R. ; YANG, Q. (2003). Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol*, 20: 18-20.
- SWANSON, W. J.; WONG, A.; WOLFNER, M.F. ; AQUADRO, C.F. (2004). Evolutionary Expressed Sequence Tag Analysis of *Drosophila* Female Reproductive Tracts Identifies Genes Subjected to Positive Selection. *Genetics*, 168: 1457-1465.
- 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.
- TAMURA, K.; SUBRAMANIAN, S. ; KUMAR, S. (2004). Temporal Patterns of Fruit Fly (*Drosophila*) Evolution Revealed by Mutation Clocks. *Mol. Biol. Evol.* 21(1):36-44.
- TAMURA, K.; DUDLEY, J.; NEI, M. ; KUMAR, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**: 1596-1599.
- TEMPLETON, A.R. (1989). The meaning of species and speciation: A genetic perspective. In: *Speciation and its Consequences* (D Otte and JA Endler). pp. 3-27. Sinauer. Sunderland, MA.
- TEMPLETON, A.R. (1992). Human origins and analysis of mitochondrial DNA sequences. *Science*, v. 255, n. 5045, p. 737.
- THOMPSON, J.D.; HIGGINS, D.G.; GIBSON, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matrix choice. *Nucl. Acids Res.* 22:4673-4680.
- WATSON, P.J.; ARNQVIST, G. ; STALLMANN, R.R. (1998). Sexual conflict and the energetic costs of mating and mate choice in water striders. *The American Naturalist*, 151: 46–58.
- WU, C-I. (2001). The genic view of the process of speciation. *J. Evol. Biol.* 14 851-865.
- WU, C-I ; TING, C.T. (2004). Genes and speciation. *Nat. Genet.* **5**: 114-122.
- YANG, Z. (2002). Inference of selection from multiple species alignments. *Current Opinion in Genetics & Development*, 12: 688-694.
- YANG, Z. ; NIELSEN, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol*, 19: 908-917.

- YANG, Z.; WONG, W.S.W. ; NIELSEN, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection, *Mol. Biol. Evol.* **22** (2005), pp. 1107–1118.
- YANG, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24: 1586-1591.
- ZHANG, J.; NIELSEN, R. E YANG, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, 22: 2472-2479.
- ZHOU, H.; CADIGAN, K.M. E THIELE, D.J. (2003). A Copper-regulated Transporter Required for Copper Acquisition, Pigmentation, and Specific Stages of Development in *Drosophila melanogaster* *J. Biol. Chem.* 278: 48210-48218.
- ZUCCHI, R. A. (1978). Taxonomia das espécies de *Anastrepha* Schiner, 1868 (Diptera: Tephritidae) assinaladas no Brasil. PhD Dissertation, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba.
- ZUCCHI, R.A. (2000). Taxonomia. In: *Moscas-das-frutas de importância econômica no Brasil Conhecimento Básico e Aplicado* (A Malavasi and RA Zucchi). pp. 13- 28. Holos Editora. Ribeirão Preto, SP.

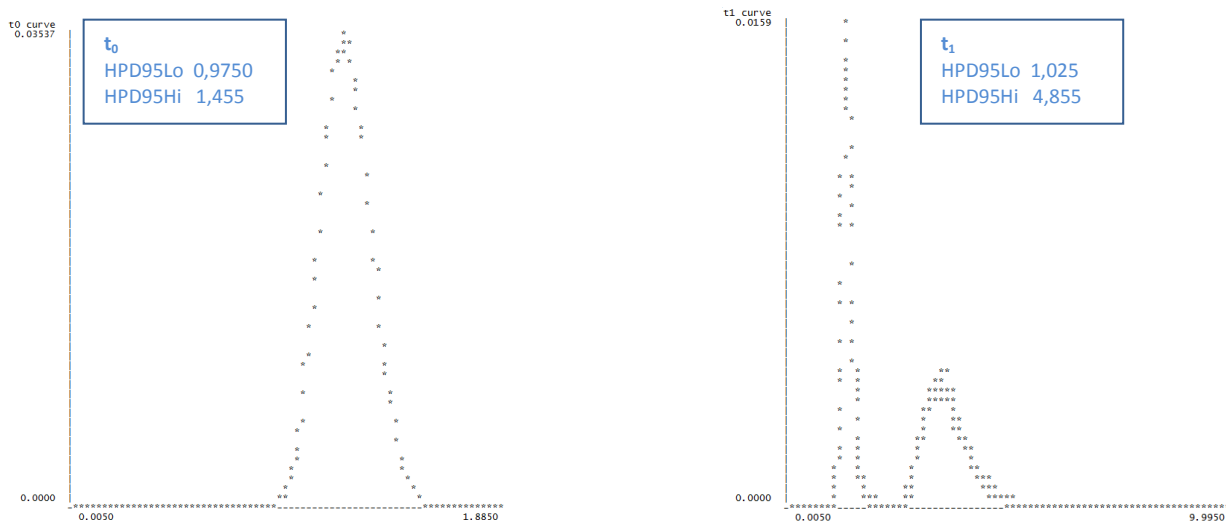
## 7. ANEXOS



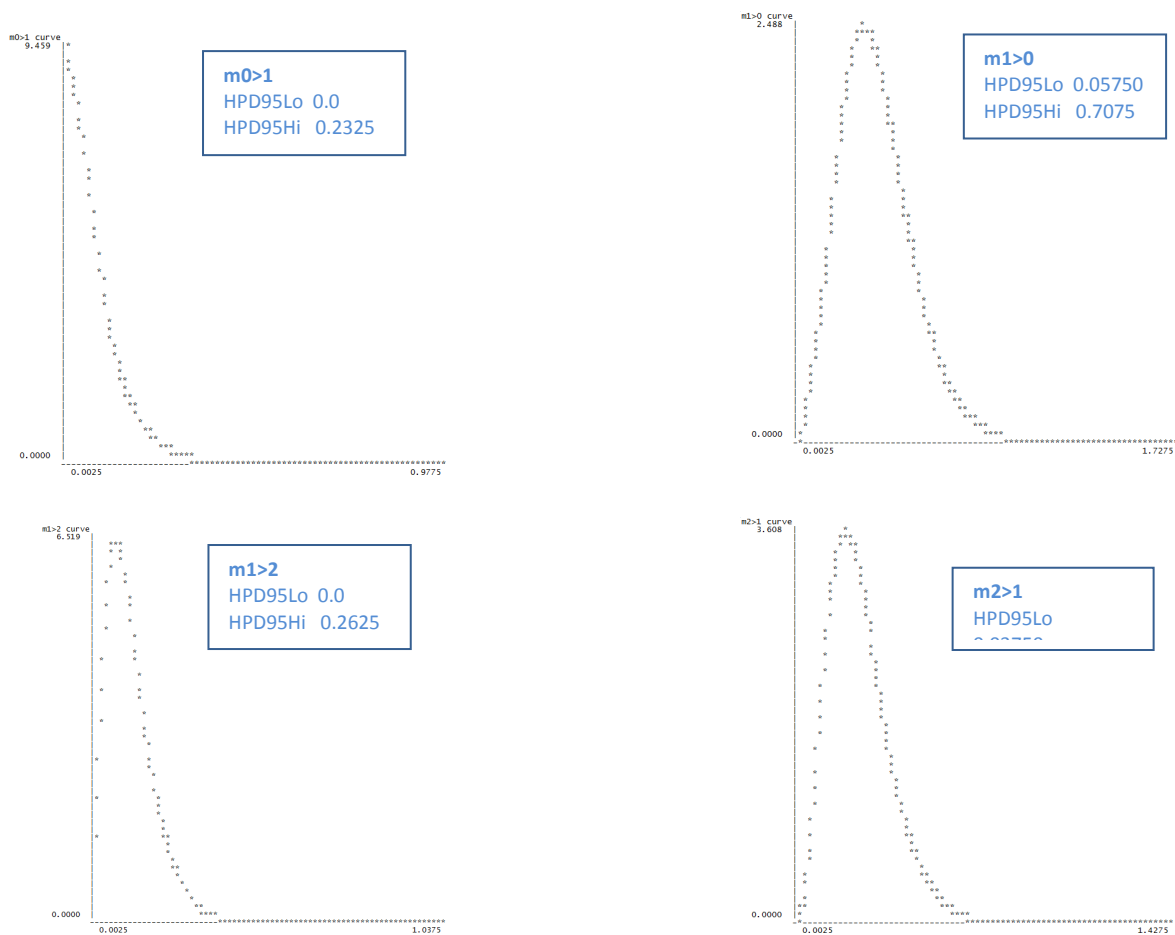
**Figura A.1a** – Gráficos dos resultados de  $m$ ,  $q$  e  $t$  obtidos no programa IMa2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde *A. fraterculus* é a espécie mais distante na árvore.



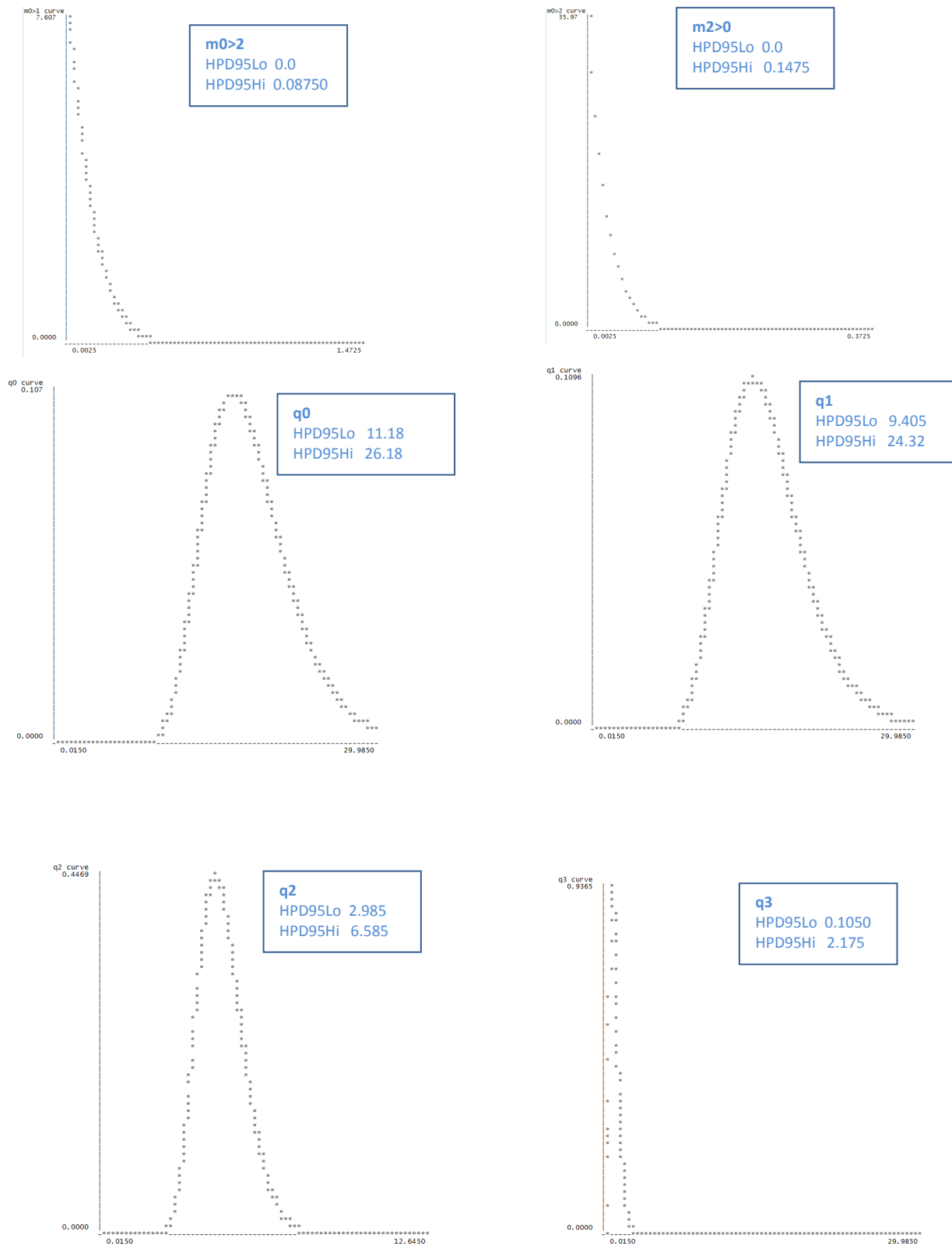
**Figura A.1b** – Gráficos dos resultados de  $m$ ,  $q$  e  $t$  obtidos no programa IMa2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde *A. fraterculus* é a espécie mais distante na árvore.



**Figura A.1c** – Gráficos dos resultados de  $m$ ,  $q$  e  $t$  obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde *A. fraterculus* é a espécie mais distante na árvore.

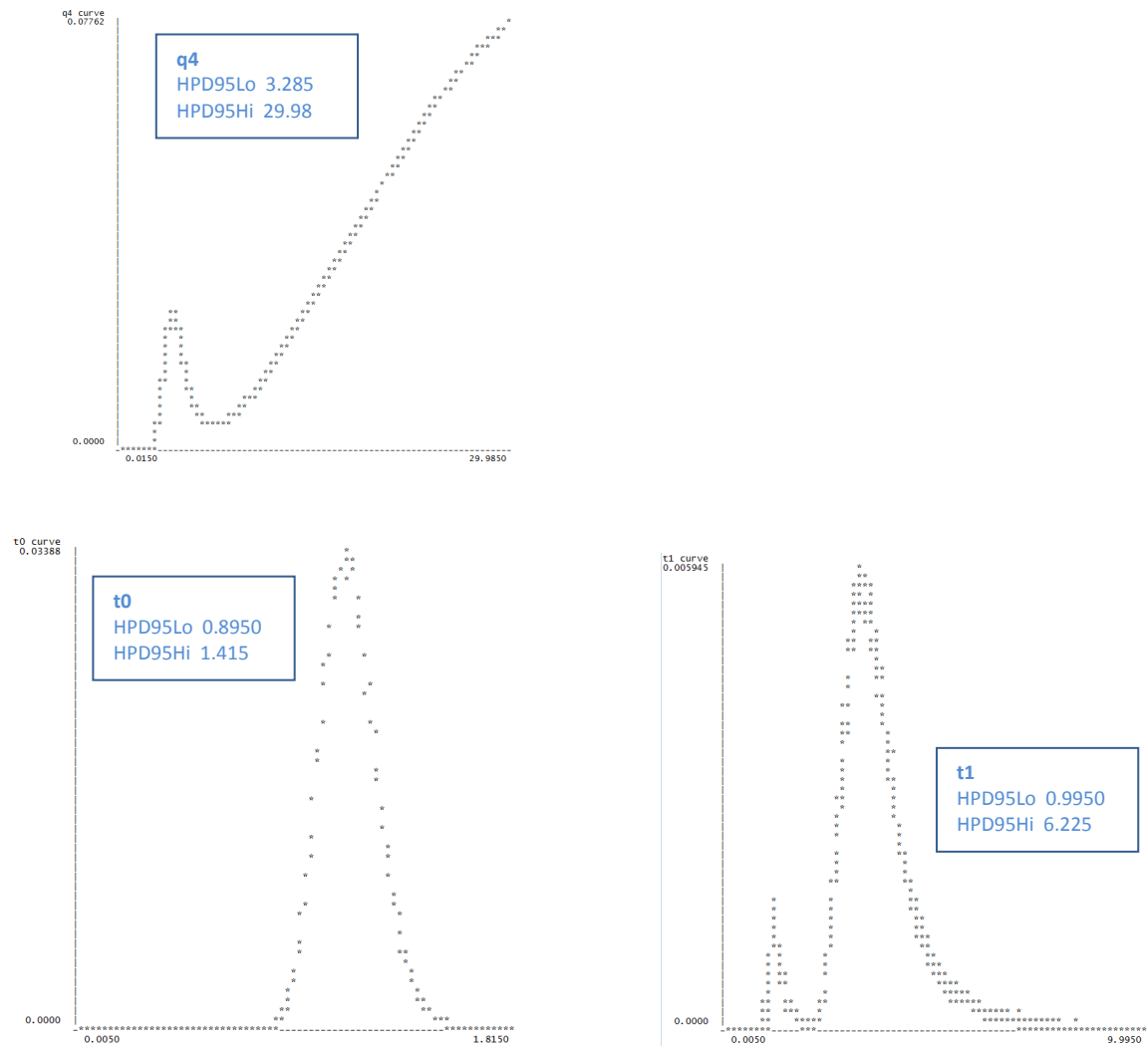


**Figura A.2a** – Gráficos dos resultados de  $m$ ,  $q$  e  $t$  obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde *A. obliqua* é a espécie mais distante na árvore.

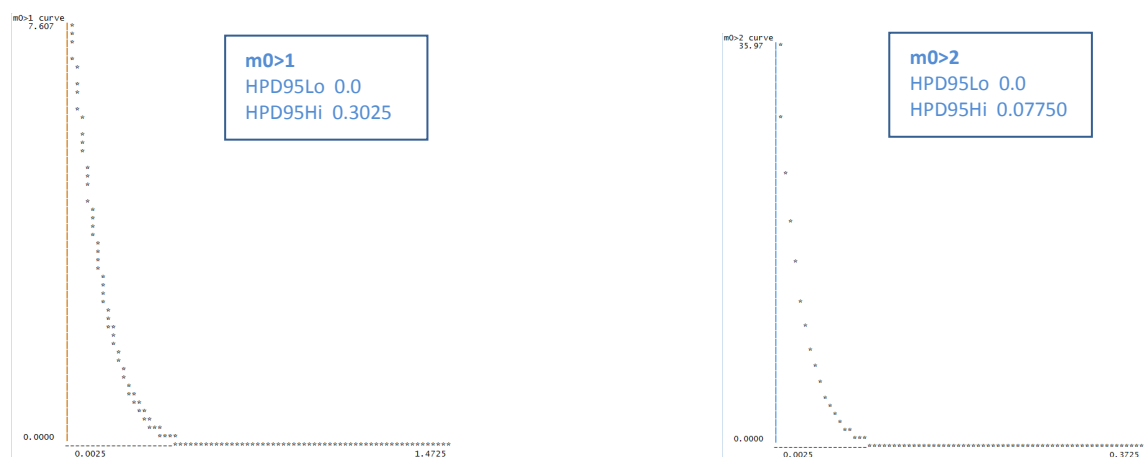


**Figura A.2b** – Gráficos dos resultados de  $m$ ,  $q$  e  $t$  obtidos no programa IMa2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde *A. obliqua* é a espécie mais distante na árvore.

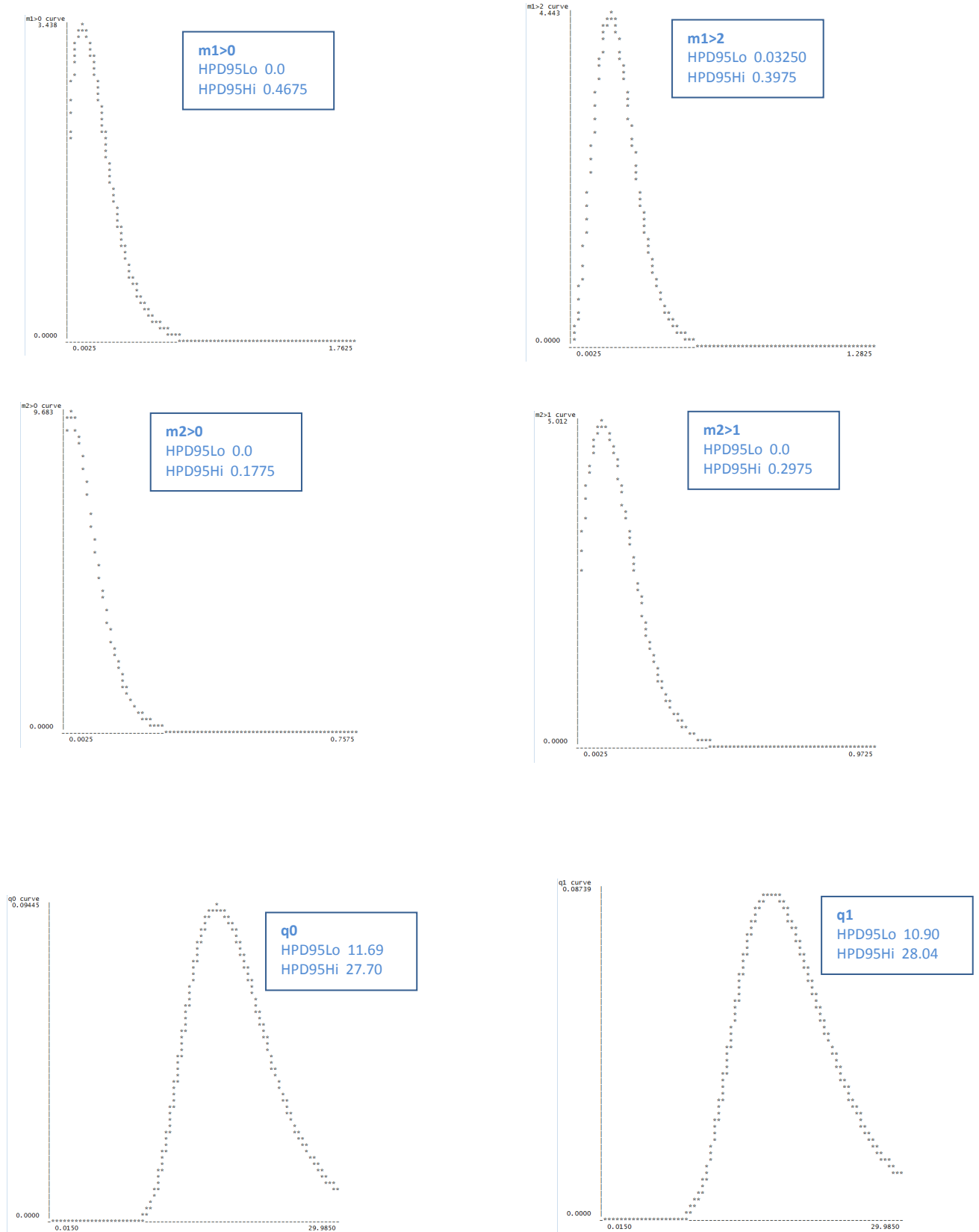




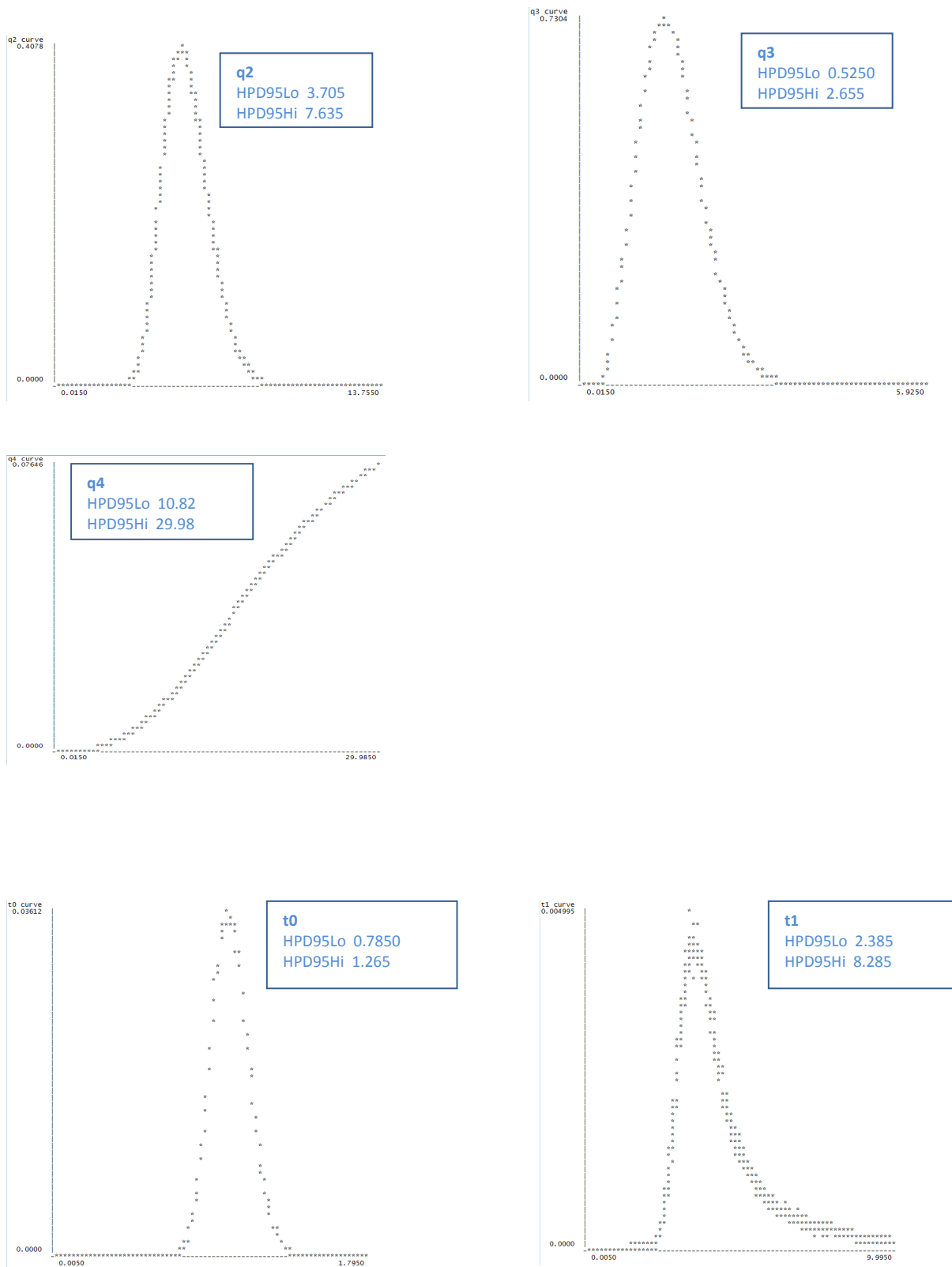
**Figura A.2c** – Gráficos dos resultados de  $m$ ,  $q$  e  $t$  obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde *A. obliqua* é a espécie mais distante na árvore.



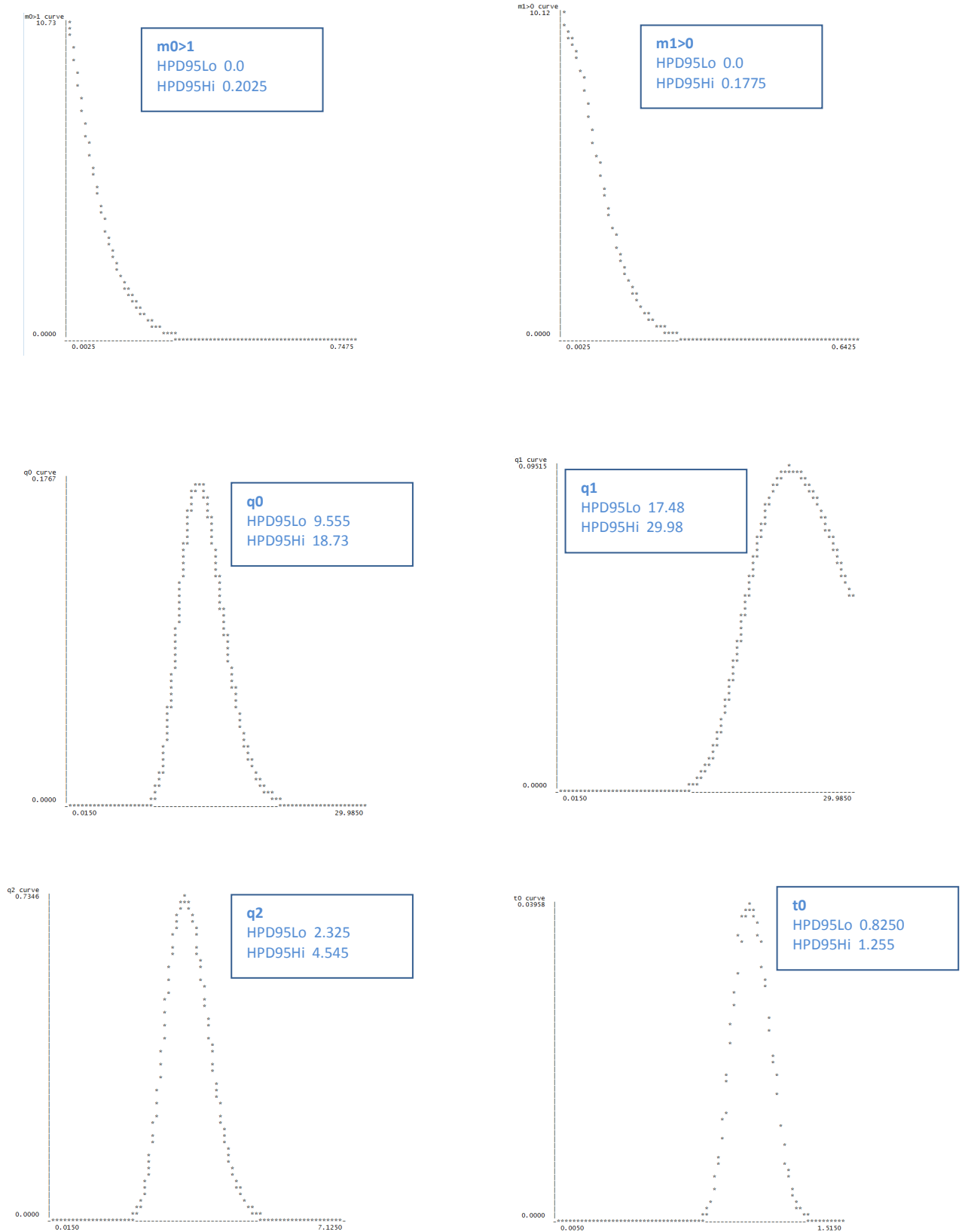
**Figura A.3a** – Gráficos dos resultados de  $m$ ,  $q$  e  $t$  obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde *A. sororcula* é a espécie mais distante na árvore.



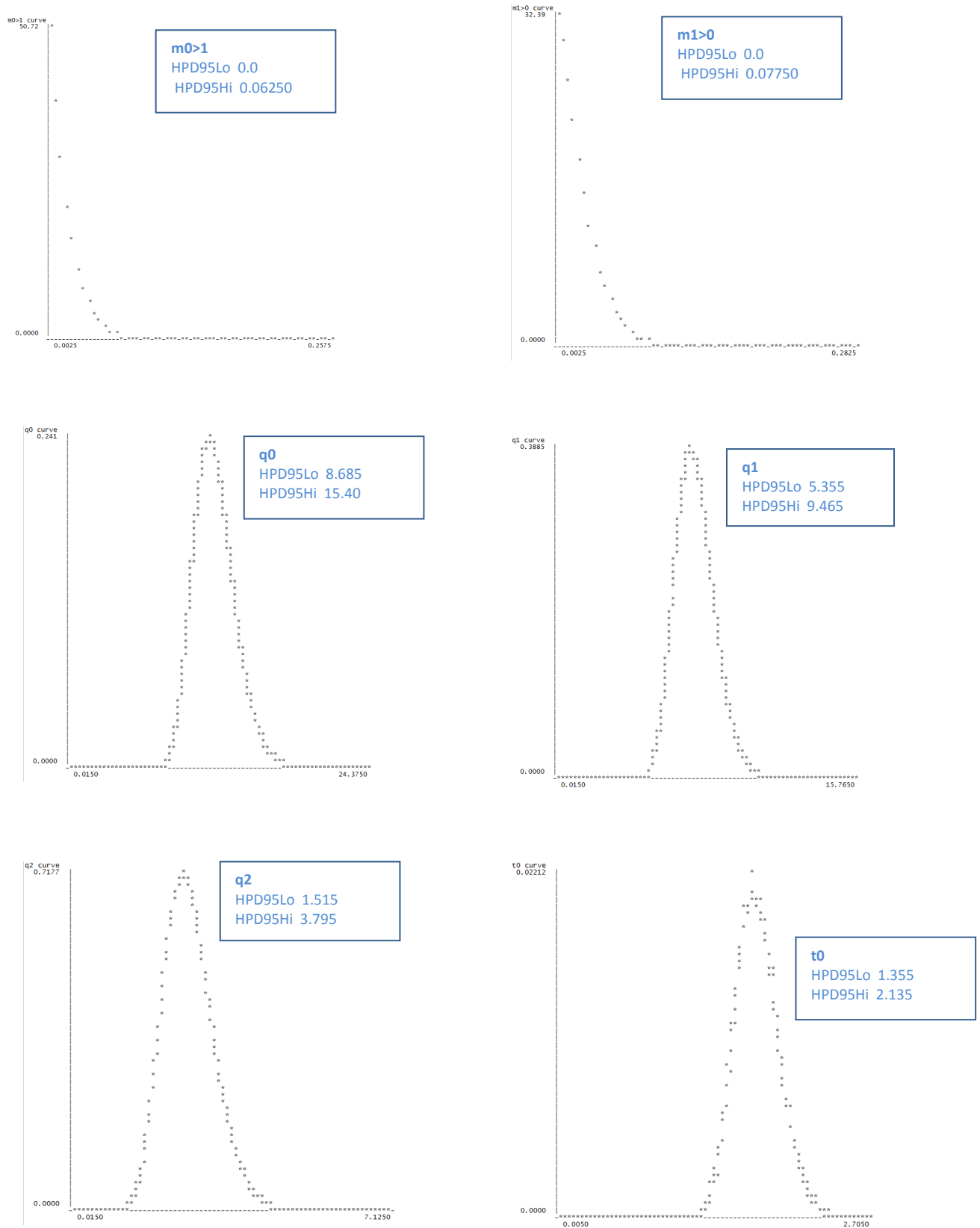
**Figura A.3b** – Gráficos dos resultados de  $m$ ,  $q$  e  $t$  obtidos no programa IMa2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde *A. sororcula* é a espécie mais distante na árvore.



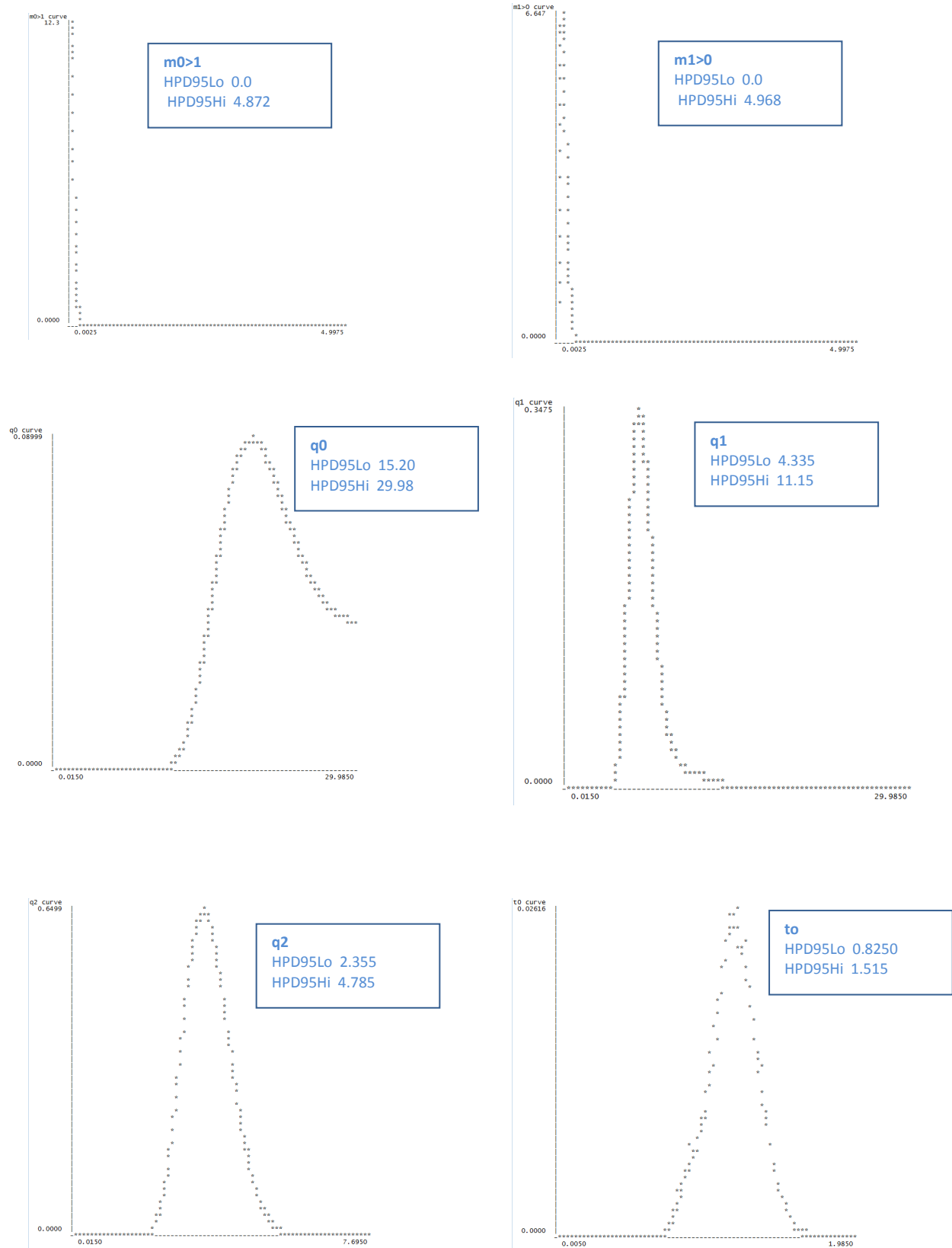
**Figura A.3c** – Gráficos dos resultados de  $m$ ,  $q$  e  $t$  obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo onde *A. sororcula* é a espécie mais distante na árvore.



**Figura A.4** – Gráficos dos resultados de  $m$ ,  $q$  e  $t$  obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo de árvore com *A.farterculus* e *A.obliqua*.



**Figura A.5** – Gráficos dos resultados de  $m$ ,  $q$  e  $t$  obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo de árvore *cA.farterculus* e *A.sororcula*.



**Figura A.6** – Gráficos dos resultados de  $m$ ,  $q$  e  $t$  obtidos no programa IMA2, com seus respectivos intervalos de confiança, para as análises realizadas com o modelo de árvore com *A. obliqua* e *A.sororcula*.

```

0          1          2          3          4          5          6          7          8          9          10         11         12         13         14         15         16         17         18         19         20
D.melanogaster  K E Q G W R A R S A F K L L Q A D E T F Q L L E G L T R A V D L C A A P G S W S Q V L A K R L Y E P L P P E E R E K V K I I A V D L Q
AAGGAGCAGGGATGGCGTGCACAAAGTGGCTTCAAACCTGCTCAGCGGGATGAGACCTTCAGTTGCTAGAGGCTCACCAGGCGCGTGGACCTTTCGCGCGCTCCTCGCAGCTGGTACAGGTCGTGGCCAAACGCTTGTACGAGCCTCTGCCCCAGAGGAACGGGAAAGGTAAGATCATCGCCGTAGACCTGCA
D.mojavensis   K E Q G W R A R S A F K L L Q A D E T F N L L E G L T R A V D L C A A P G G W S Q V L S K R M Y E P R S E K D R E Q V K I I A V D M Q
AAGGAAACAGGGATGGCGAGCCAGAAGTGGCTCAAACCTACTACAAGCAGATGAAACGTTTAAATTTGTGGAGGGCTTAACACGTCGCGTGGACTTATCGCCAGCACCTGGAGGATGGTACAGGTTTCTCAAGAGCTATGTACGAACCCGGATCGGAGAAAGCCGGGACAGTAAAGATAATGCTGTGGACATGCA
D.willistoni    K Q Q G W R A R S A F K L L Q A E E T F H L L D G V T R A V D L C A A P G S W S Q V L S K H M Y E P R T P E E R E Q V K I I A V D M Q
AAACAACAAGCTGGAGAGCTAGGATGCTTAAATTTGCTCAAGCAGAAGAAACGTTTCAATTTGCTGGATGGCTCACACGAGCCGTGGATCTTTGTGCGGCTCCTGGAAGCTGGTCCAGGTTATTCAAAACATATGTACGAGCCGCAACCCCTGAAGAACGGGAAACAGTAAATAATGCTGTGGATATGCA
D_virilis       K E Q G W R A R S A F K L L Q A D E T F N L L E G V T R A V D L C A A P G G W S Q V L S K R L Y E P R T P E E R E Q V K I I A V D M Q
AAGGAGCAGGGATGGCGCCAGAAGTGCATTCAAACCTGCTCAAGCCGACGAAACGTTTAAATTTACTCGAGGGTGAACACGCTGCTGVDGACTTATCGCCTGCCCGGGTGGTCCAGGTCGTCCAAAGCTTTGTACGAACCCGCGCACCCCGAGGAACGGGACAGGTGAAGATCATCGCTGTGGATATGCA
D.grimshawi     K E Q G W R A R S A F K L L Q A D E T F N L L E G V E R A V D L C A A P G G W S Q V L S R R M Y E P R T V E E R E K V K I I A V D T Q
AAGGAAACAGGGATGGCGCCAGAAGTGCCTCAAACCTGCTACAAGCTGATGAAACGTTTAAATTTATGGAGGGTGTGCAAGCGGCGGTGATTTATGCTGCTGCACTGGCGGCTGGTCCAGGTCGTATCAAGCGCATGTACGAGCCGCAACTGTCCAGGAACCGAGAAAGTGAAGATCATCGCAGTGGATATGCA
A.fraterculus_1 K E E G W R A R S A F K L L Q I D E K F H I L K D V T R A V D L C A A P G S W T Q V L S K R L Y E - - N R S N N E E V K I I A V D L Q
AAAGAAGAAGCTGGCGAGCTCGGAGTGCATTAACTGCTGCAGATCGATGAAAGTTCACATTTTAAAGATGTACACAGTCGCGTAGACTTATGCTGCTGCGCTGGTAGCTGCAAGTGCCTTCCAAAGCCCTTTATGAG-----AATCGTCCAACATGAGGAGGTAATAATAATTCGCGTAGATTTGCA
A.fraterculus_2 K E E G W R A R S A F K L L Q I D E K F H I L K D V T R A V D L C A A P G S W T Q V L S K R L Y E - - N R S N N E E V K I I A V D L Q
AAAGAAGAAGCTGGCGAGCTCGGAGTGCATTAACTGCTGCAGATCGATGAAAGTTCACATTTTAAAGATGTACACAGCGCGTAGACTTATGCTGCTGCGCTGGTAGCTGCAAGTGCCTTCCAAAGCCCTTTATGAG-----AATCGTCCAACATGAGGAGGTAATAATAATTCGCGTAGATTTGCA
A.obliqua_1     K E E G W R A R S A F K L L Q I D E K F H I L K D V T R A V D L C A A P G S W T Q V L S K R L Y E - - N R S N N E E V K I I A V D L Q
AAAGAAGAAGCTGGCGAGCTCGGAGTGCATTAACTGCTGCAGATCGATGAAAGTTCACATTTTAAAGATGTACACAGCGCGTAGACTTATGCTGCTGCGCTGGTAGCTGCAAGTGCCTTCCAAAGCCCTTTATGAG-----AATCGTCCAACATGAGGAGGTAATAATAATTCGCGTAGATTTGCA
A.obliqua_2     K E E G W R A R S A F K L L Q I D E K F H I L K D V T R A V D L C A A P G S W T Q V L S K R L Y E - - N R S N N E E V K I I A V D L Q
AAAGAAGAAGCTGGCGAGCTCGGAGTGCATTAACTGCTGCAGATCGATGAAAGTTCACATTTTAAAGATGTACACAGCGCGTAGACTTATGCTGCTGCGCTGGTAGCTGCAAGTGCCTTCCAAAGCCCTTTATGAG-----AATCGTCCAACATGAGGAGGTAATAATAATTCGCGTAGATTTGCA
A.sororcula_1   K E E G W R A R S A F K L L Q I D E K F H I L K D V T H A V D L C A A P G S W T Q V L S K R L Y E - - N R S N N E E V K I I A V D L Q
AAAGAAGAAGCTGGCGAGCAGGAGTGCCTTAAATTTGCTGAGATCGATGAAAGTTCACATTTTAAAGATGTACACTATGCTGAGACTTATGCTGCTGCACTGGTAGCTGCAAGTGCCTTCCAAAGCCCTTTATGAG-----AATCGTCCAACATGAGGAGGTAATAATAATTCGCGTAGATTTGCA
A.sororcula_2   K E E G W R A R S A F K L L Q I D E K F H I L K D V T R A V D L C A A P G S W T Q V L S K R L Y E - - N R S N N E E V K I I A V D L Q
AAAGAAGAAGCTGGCGAGCAGGAGTGCCTTAAATTTGCTGAGATCGATGAAAGTTCACATTTTAAAGATGTACACTGCTGCGTAGACTTATGCTGCTGCACTGGTAGCTGCAAGTGCCTTCCAAAGCCCTTTATGAG-----AATCGTCCAACATGAGGAGGTAATAATAATTCGCGTAGATTTGCA

```

```

2          3          4          5          6          7          8          9          10         11         12         13         14         15         16         17         18         19         20
D.melanogaster  G M A P I E G V K Q L R A D I S K E S T A E A I I E F F G G E - - K A Q I V V S
GGGAATGGCTCCGATCGAAGGAGTCAAACAACCTGGCGCGGACATCAGCAAGGATGCGACAGCGAGGCCATAATGAGTTCCTTGGCGCGGAG-----AAGGCACAGATTTGTTGTTAGC
D.mojavensis   G M A P I D G V T Q L R A D I T K E S T A E E I I K F F D G K - - K A E L V V T
GGGTATGGCACCATCGATGGCGTACTCAATTCGGCGCGACATTAAGGAGTCAACCGCAGAGGAAATATTAATTTTCGATGGCAAA-----AAGCGCAACTGGTGTGCTCACT
D.willistoni    G M A P I D G V K Q L R A D I T K E S T A K E I I E Y F G G Q - - K A Q L V I S
GGGATGGCACCATCGATGGAGTCAAGCAACTGAGGGCGGATATTACCAAAGAGTCCACTGCCAAGAGATATATGATACTTCGGGGGTCAA-----AAGCACAAITGGTATCTCA
D_virilis       G M A P I D G V T Q L R E D I T K E E T A E A I I K F F D G K - - K A Q L V V S
GGGTATGGCACCATTGATGGTGCACACAGTTCGGCGAAGATATTACCAAGGAGGAAACAGCGGAGGCAATTAATTAANGTTTTTGTATGGCAAA-----AAGCGCAACTGGTGTGCTCAGC
D.grimshawi     G M A P I D G V T Q L R A D I T K E S T A D A I I E F F G G K - - K A Q L V V S
GGGATGGCACCATTGATGGTGCACACAGTTCGGCGCGCACTTACCAAAGAGTCAACAGCAGATGCAATTAATTAANGTTTTTGTATGGCAAA-----AAGCGCAACTGGTGTGCTCAGC
A.fraterculus_1 A M A P L P G V T Q L Q G D I T K L S T A Q A I I E H F G G E S Q K A Q L V I C
AGCGATGGCACCATTGCCTGGAGTACGCAACTTCAAGGCGATATAACGAAACTAAGTACCCTCAGGCTATAATGAACATTTTGGTGGAGAATCGCAAAAGGCCAGTTGGTTATTGTC
A.fraterculus_2 A M A P L P G V T Q L Q G D I T K L S T A Q A I I E H F G G E S Q K A Q L V I C
AGCGATGGCACCATTGCCTGGAGTACGCAACTTCAAGGCGATATAACGAAACTAAGTACCCTCAGGCTATAATGAACATTTTGGTGGAGAATCGCAAAAGGCCAGTTGGTTATTGTC
A.obliqua_1     A M A P L P G V T Q L Q G D I T K L S T A Q A I I E H F G G E S Q K A Q L V I C
AGCGATGGCACCATTGCCTGGAGTACGCAACTTCAAGGCGATATAACGAAACTAAGTACCCTCAGGCTATAATGAACATTTTGGTGGAGAATCGCAAAAGGCCAGTTGGTTATTGTC
A.obliqua_2     A M A P L P G V T Q L Q G D I T K L S T A Q A I I E H F G G E S Q K A Q L V I C
AGCGATGGCACCATTGCCTGGAGTACGCAACTTCAAGGCGATATAACGAAACTAAGTACCCTCAGGCTATAATGAACATTTTGGTGGAGAATCGCAAAAGGCCAGTTGGTTATTGTC
A.sororcula_1   A M A P L P G V T Q L Q G D I T K L S T A Q A I I E H F G G E S Q K A Q L V I C
AGCGATGGCACCATTGCCTGGAGTACGCAACTTCAAGGCGATATAACGAAACTAAGTACCCTCAGGCTATAATGAACATTTTGGTGGAGAATCGCAAAAGGCCAGTTGGTTATTGTC
A.sororcula_2   A M A P L P G V T Q L Q G D I T K L S T A Q A I I E H F G G E S Q K A Q L V I C
AGCGATGGCACCATTGCCTGGAGTACGCAACTTCAAGGCGATATAACGAAACTAAGTACCCTCAGGCTATAATGAACATTTTGGTGGAGAATCGCAAAAGGCCAGTTGGTTATTGTC

```

Figura A.7 Alinhamento do gene *CG7009* utilizado para análises de seleção positiva realizadas no PAML.

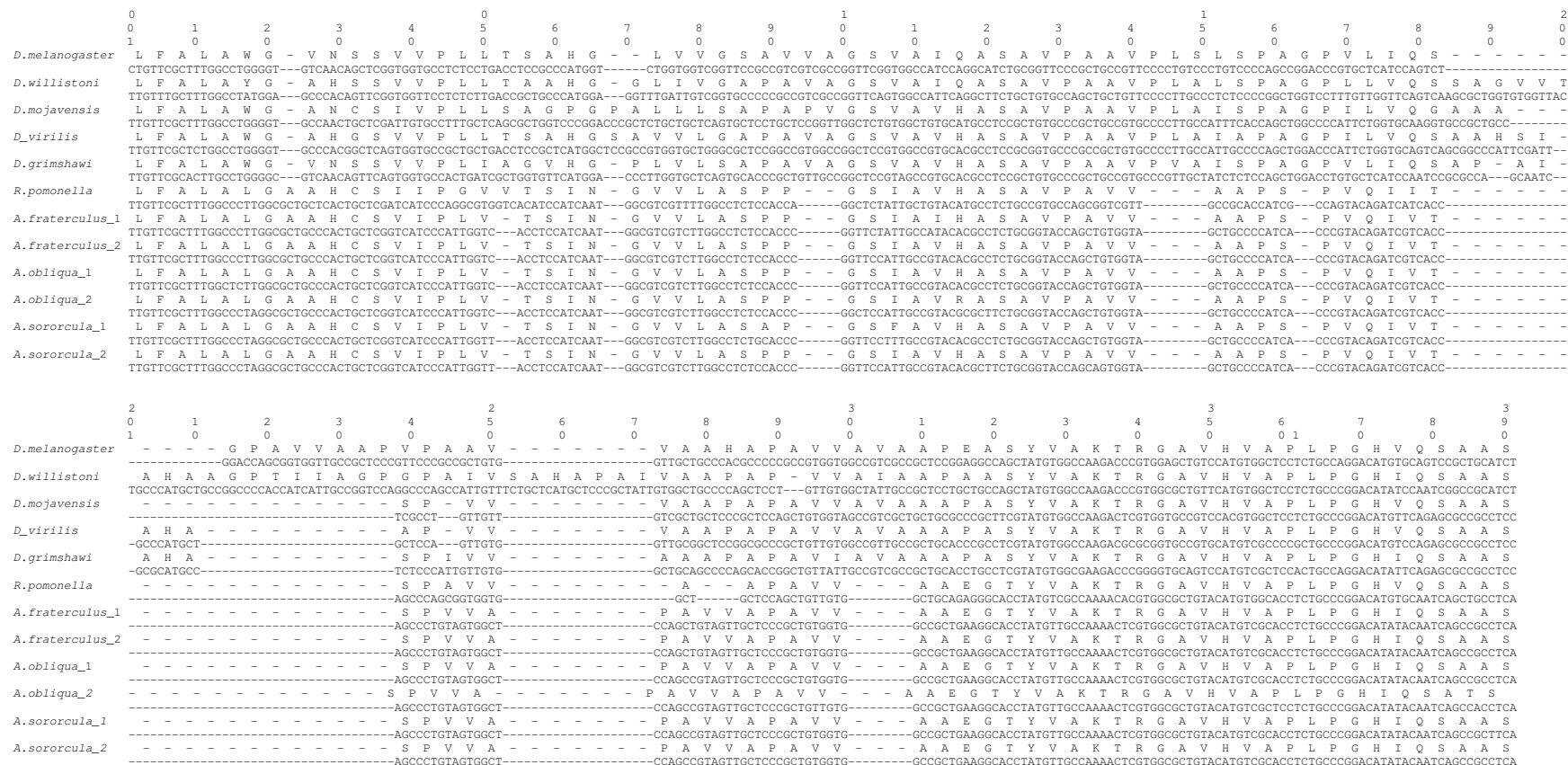


Figura A.8 Alinhamento do gene *CG7203* utilizado para análises de seleção positiva realizadas no PAML.



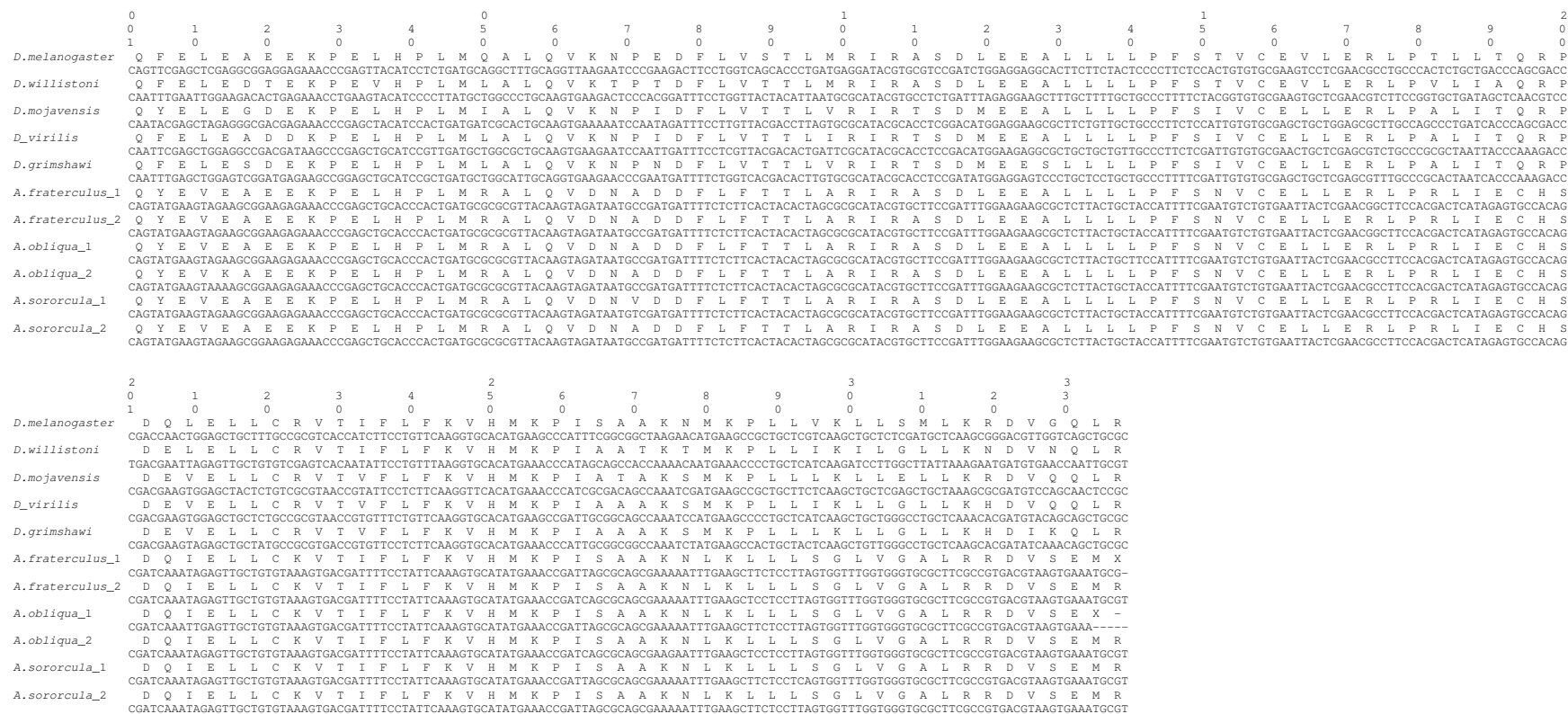


Figura A.9 Alinhamento do gene *CG8064* utilizado para análises de seleção positiva realizadas no PAML.

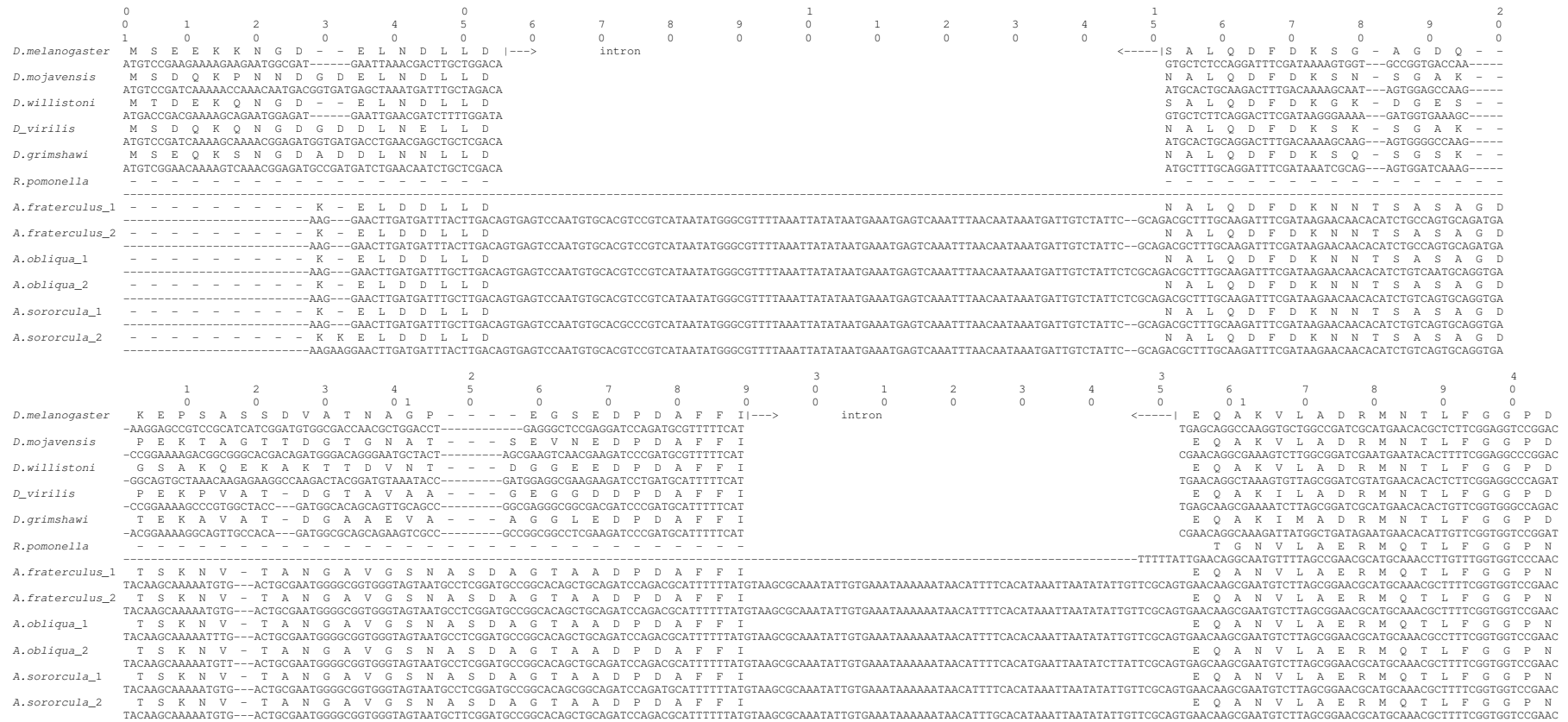


Figura A.10 Alinhamento do gene CG5325 utilizado para análises de seleção positiva realizadas no PAML.

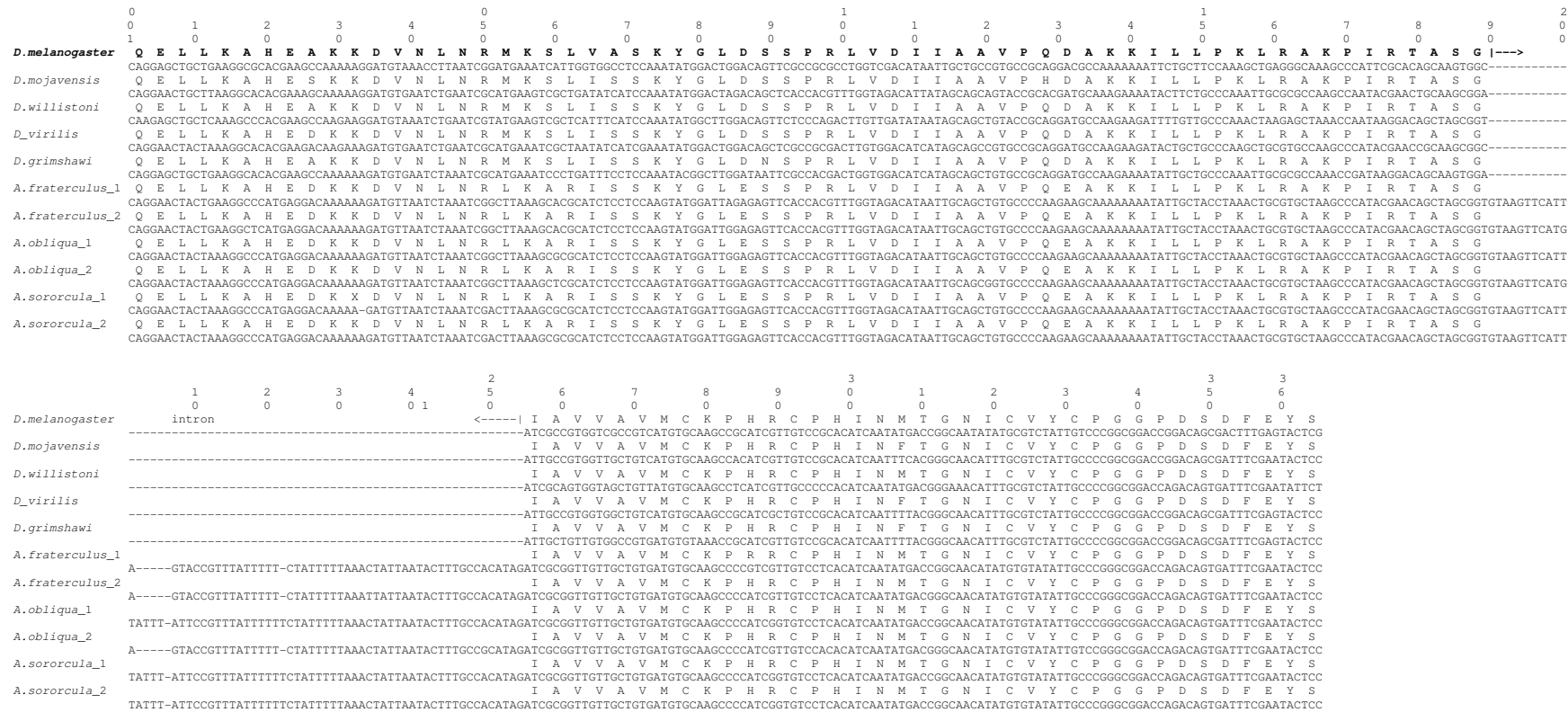


Figura A.11 Alinhamento do gene *Elongation Protein* utilizado para análises de seleção positiva realizadas no PAML.

```

0          1          2          3          4          5          6          7          8          9          10         11         12         13         14         15         16         17         18         19         20
0          0          0          0          0          0          0          0          0          0          0          0          0          0          0          0          0          0          0          0          0
D.melanogaster  F N L F D N R G D G K I Q L S Q V G E C L R A L G Q N P T E S D V K K C T H Q L K P D E R I S F E V F L P I Y Q A I S K A R S G D T A
TTCAACCTTTCGACAACCGCGGAGATGGCAAAATTCAGTTCAGCCAGGTGGGCGAGTCTGAGGGCCCTGGGCCAGAACCCACCGAGTGGGATGCAAGAAAGTGCACCCACCGAGTGAAGCCCGAGCGAGCGATCTCGTTTCGAGGTGTTCTGCCCATCTATCAGGCCATCTCAAGGCCCGCTCGGGGATACGGC
D.mojavensis   F N L F D N R G D G K I Q L S Q V G E C L R A L G Q N P T E S D V K K C T H Q L K P D E R I S F E V F L P I Y Q A I S K A R S G D T A
TTCAACCTTTCGATAATCGCGCGATGGCAAAATTCAGTTCAGCCAGGTGGGCGAATGCTTCGGTGCATTCGGCCAGAAATCCACCGAATCGGATGTGAAGAAAGTGCACCCACCGAGTGAAGCCCGAGCGAGCGATCTCGTTTCGAGGTCTTTCTGCCCATCTATCAGGCCATCTCAAGGCCCGCTCGGGGATACGGC
D_virilis      F N L F D N R G D G K I Q L S Q V G E C L R A L G Q N P T E S D V K K C T H Q L K P D E R I S F E V F L P I Y Q A I S K A R S G D T A
TTCAATCTGTTTGACAATCGCGCGATGGCAAGATACAGTTCAGCCAGGTGGGCGAGTGCCTGGGCCACTGGGCCAGAAATCCACCGAATCGGATGTGAAGAAAGTGCACCCACCGAGTGAAGCCCGAGCGAGCGATCTCGTTTCGAGGTCTTTCTGCCCATCTATCAGGCCATCTCAAGGCCCGCTCGGGGATACGGC
D.willistoni   F N L F D N R G D G K I Q L S Q V G E C L R A L G Q N P T E S D V K K C T H Q L K S D E R I S F E V F L P I Y Q A I S K A R S G D T A
TTCAATTTGTCGACAATCGAGGTGATGGCAAAATCCAATTCAGCCAGTGGGCGAATGCTTCGGTGCCTTCGGGCCAAATCCCTACCGAATCGGACGTGAAGAAATGCACACCAATGAAATCGGATGAGAGGATCTCGTTTGAAGTCTTTTCTGCCCATCTATCAGGCCATCTCAAGGCCCGCTCGGGGATACGGC
D.grimshawi   F N L F D N R G D G K I Q L G Q V G E C L R A L G Q N P T E S D V K K C T H Q L K P D E R I S F E V F L P I Y Q A I S K A R S G D T A
TTCAATCTGTTTCGACAATCGCGCGATGGCAAGATACAGTTCAGCCAGGTGGGCGAGTGTCTGGCGCCTTCGGGACAGAAATCCACCGAATCGGATGTGAAGAAAGTGCACACCTACCGAATGAAAGTGAAGCCCGAGCGAGCGATCTCGTTTCGAGGTCTTTCTGCCCATCTACCAGCGATCTCAAGGCCCGCTCGGGGATACGGC
R.pomonella   F N L F D N R G D G K I Q L S Q V G E C L R A L G Q N P T E S D V K K C T H Q L K P D E R I S F E V F L P I Y Q A I S K A R S G D T A
TTCAACCTTTCGACAATCGCGGTGATGGTAAAATCCAATTCAGTTCAGTTCAGTGGGCGAGTGTTCGGTGCATTCGGTTCAGAAATCCCACTGAAATCGGATGTGAAGAAATGCACCTACCAATGAAAGTGAAGCCCGAGCGAGCGATCTCGTTTCGAGGTCTTTCTACCCTATCTATCAGGCCATCTCAAGGCCCGCTCGGGGATACGGC
A.fraterculus_1 F N L F D N R G D G K I Q L S Q V G E C L R A L G Q N P T E S D V K K C T H Q L K P D E R I S F E V F L P I Y Q A I S K A R S G D T A
TTCAATCTTTTCGACAATCGAGGTGATGGCAAAATCCAATTCAGTTCAGTTCAGTGTGAGTGTCTCGGTGCATTCAGTTCAGAAATCCCAAGAAATCGGATGTGAAGAAATGCACCTACCAATGAAAGTGAAGCCCGAGCGAGCGATCTCGTTTCGAGGTCTTTCTACCCTATCTATCAGGCCATCTCAAGGCCCGCTCGGGGATACGGC
A.fraterculus_2 F N L F D N R G D G K I Q L S Q V G E C L R A L G Q N P T E S D V K K C T H Q L K P D E R I S F E V F L P I Y Q A I S K A R S G D T A
TTCAATCTTTTCGACAATCGAGGTGATGGCAAAATCCAATTCAGTTCAGTTCAGTGTGAGTGTCTCGGTGCATTCAGTTCAGAAATCCCAAGAAATCGGATGTGAAGAAATGCACCTACCAATGAAAGTGAAGCCCGAGCGAGCGATCTCGTTTCGAGGTCTTTCTACCCTATCTATCAGGCCATCTCAAGGCCCGCTCGGGGATACGGC
A.obliqua_1   F N L F D N R G D G K I Q L S Q V G E C L R A L G Q N P T E S D V K K C T H Q L K P D E R I S F E V F L P I Y Q A I S K A R S G D T A
TTCAATCTTTTCGACAATCGAGGTGATGGCAAAATCCAATTCAGTTCAGTTCAGTGTGAGTGTCTCGGTGCATTCAGTTCAGAAATCCCAAGAAATCGGATGTGAAGAAATGCACCTACCAATGAAAGTGAAGCCCGAGCGAGCGATCTCGTTTCGAGGTCTTTCTACCCTATCTATCAGGCCATCTCAAGGCCCGCTCGGGGATACGGC
A.obliqua_2   F N L F D N R G D G K I Q L S Q V G E C L R A L G Q N P T E S D V K K C T H Q L K P D E R I S F E V F L P I Y Q A I S K A R S G D T A
TTCAATCTTTTCGACAATCGAGGTGATGGCAAAATCCAATTCAGTTCAGTTCAGTGTGAGTGTCTCGGTGCATTCAGTTCAGAAATCCCAAGAAATCGGATGTGAAGAAATGCACCTACCAATGAAAGTGAAGCCCGAGCGAGCGATCTCGTTTCGAGGTCTTTCTACCCTATCTATCAGGCCATCTCAAGGCCCGCTCGGGGATACGGC
A.sororcula_1 F N L F D N R G D G K I Q L S Q V G E C L R A L G Q N P T E S D V K K C T H Q L K P D E R I S F E V F L P I Y Q A I S K A R S G D T A
TTCAATCTTTTCGACAATCGAGGTGATGGCAAAATCCAATTCAGTTCAGTTCAGTGTGAGTGTCTCGGTGCATTCAGTTCAGAAATCCCAAGAAATCGGATGTGAAGAAATGCACCTACCAATGAAAGTGAAGCCCGAGCGAGCGATCTCGTTTCGAGGTCTTTCTACCCTATCTATCAGGCCATCTCAAGGCCCGCTCGGGGATACGGC
A.sororcula_2 F N L F D N R G D G K I Q L S Q V G E C L R A L G Q N P T E S D V K K C T H Q L K P D E R I S F E V F L P I Y Q A I S K A R S G D T A
TTCAATCTTTTCGACAATCGAGGTGATGGCAAAATCCAATTCAGTTCAGTTCAGTGTGAGTGTCTCGGTGCATTCAGTTCAGAAATCCCAAGAAATCGGATGTGAAGAAATGCACCTACCAATGAAAGTGAAGCCCGAGCGAGCGATCTCGTTTCGAGGTCTTTCTACCCTATCTATCAGGCCATCTCAAGGCCCGCTCGGGGATACGGC
2          2
0          1          2          3          4          5          6
1          0          0          0          0          0          0
D.melanogaster  D D F I E G L R H F D K D A S G Y I S S A
TGATGATTTTCATCGAAGGACTTGGGTCACTTCGACAAGGACCGCGGGTTATATCTCCTCGCC
D.mojavensis   D D F I E G L R H F D K D A S G Y I S S A
CGACGATTTTCATTCAGGGCTTGGCCATTCGATAAGGATGCTAGTGGCTACATATCGTCCGT
D_virilis      D D F I E G L R H F D K D A S G Y I S S A
CGATGATTTTCATTCAGGGATTCGGTCACTTCGATAAGGATGCCAGCGCTATATATCATCGGT
D_willistoni   D D F I E G L R H F D K D A S G Y I S S A
CGATGATTTTCATTCAGGGTTCGGCCATTCGATAAGGATGCCAGCGGTTCATATCTCCTCGCC
D.grimshawi   D D F I E G L R H F D K D A S G Y I S S A
CGATGATTTTCATTCAGGGCTTGGCCATTCGATAAGGATGCCAGCGGTTCATATCTCCTCGCT
R.pomonella   D D F I E G L R H F D K D A S G F I S S A
CGATGATTTTCATTCAGGGTTTGGGTCACTTCGACAAGATGCCAGCGGTTTCATTCATCAGCT
A.fraterculus_1 D D F I E G L R H F D K D A S G F I S S A
TGATGATTTTCATTCAGGGTTCGGTCACTTCGACAAGATGCCAGCGGTTTCATTCCTCAGCC
A.fraterculus_2 D D F I E G L R H F D K D A S G F I S S A
TGATGATTTTCATTCAGGGTTCGGTCACTTCGACAAGATGCCAGCGGTTTCATTCCTCAGCC
A.obliqua_1   D D F I E G L R H F D K D A S G F I S S A
TGATGATTTTCATTCAGGGTTCGGTCACTTCGACAAGATGCCAGCGGTTTCATTCCTCAGCC
A.obliqua_2   D D F I E G L R H F D K D A S G F I S S A
TGATGATTTTCATTCAGGGTTCGGTCACTTCGACAAGATGCCAGCGGTTTCATTCCTCAGCC
A.sororcula_1 D D F I E G L R H F D K D A S G F I S S A
TGATGATTTTCATTCAGGGTTCGGTCACTTCGACAAGATGCCAGCGGTTTCATTCCTCAGCC
A.sororcula_2 D D F I E G L R H F D K D A S G F I S S A
TGATGATTTTCATTCAGGGTTCGGTCACTTCGACAAGATGCCAGCGGTTTCATTCCTCAGCC

```

Figura A.12 Alinhamento do gene *Myosin Light Chain* utilizado para análises de seleção positiva realizadas no PAML.

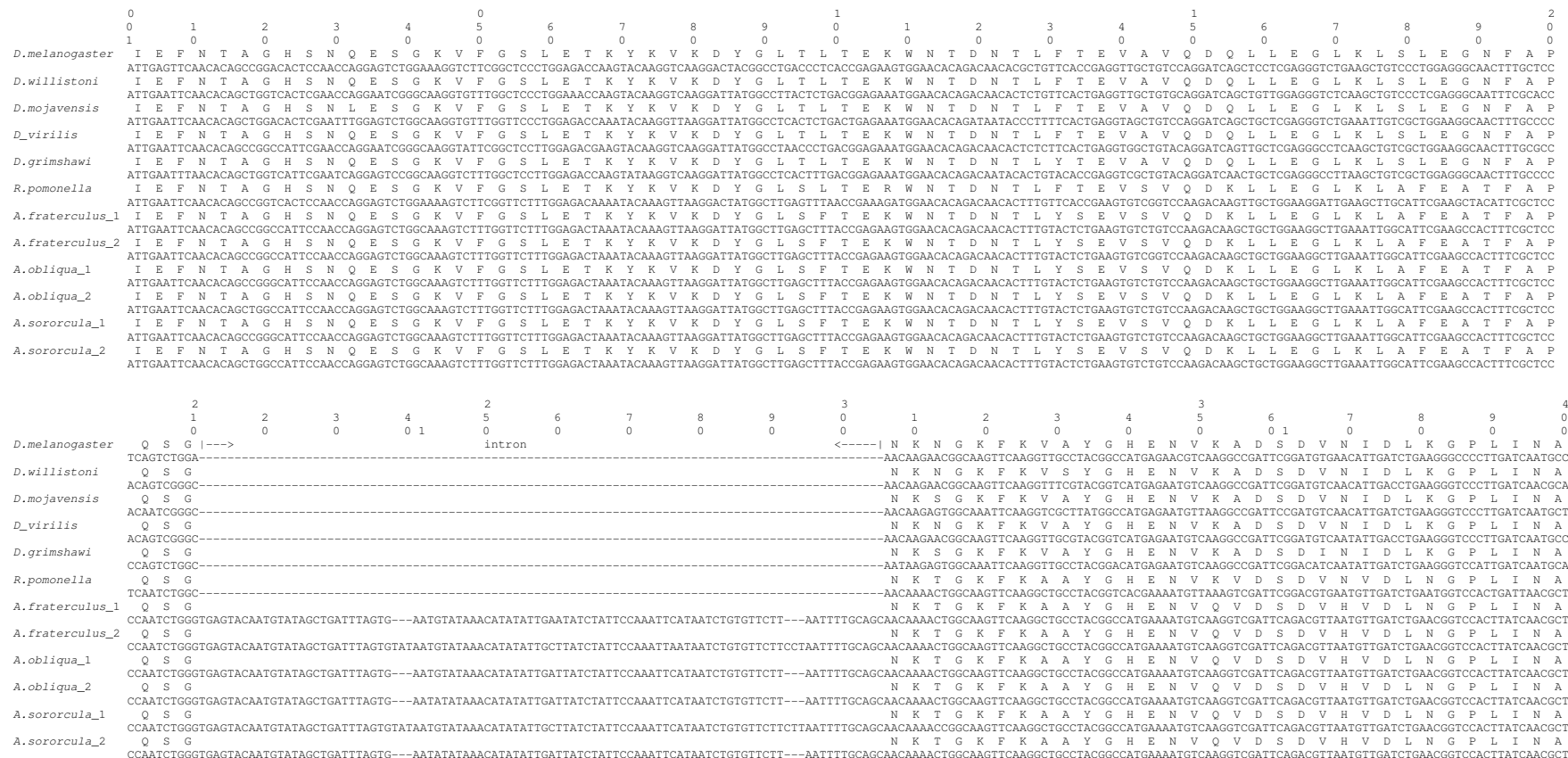


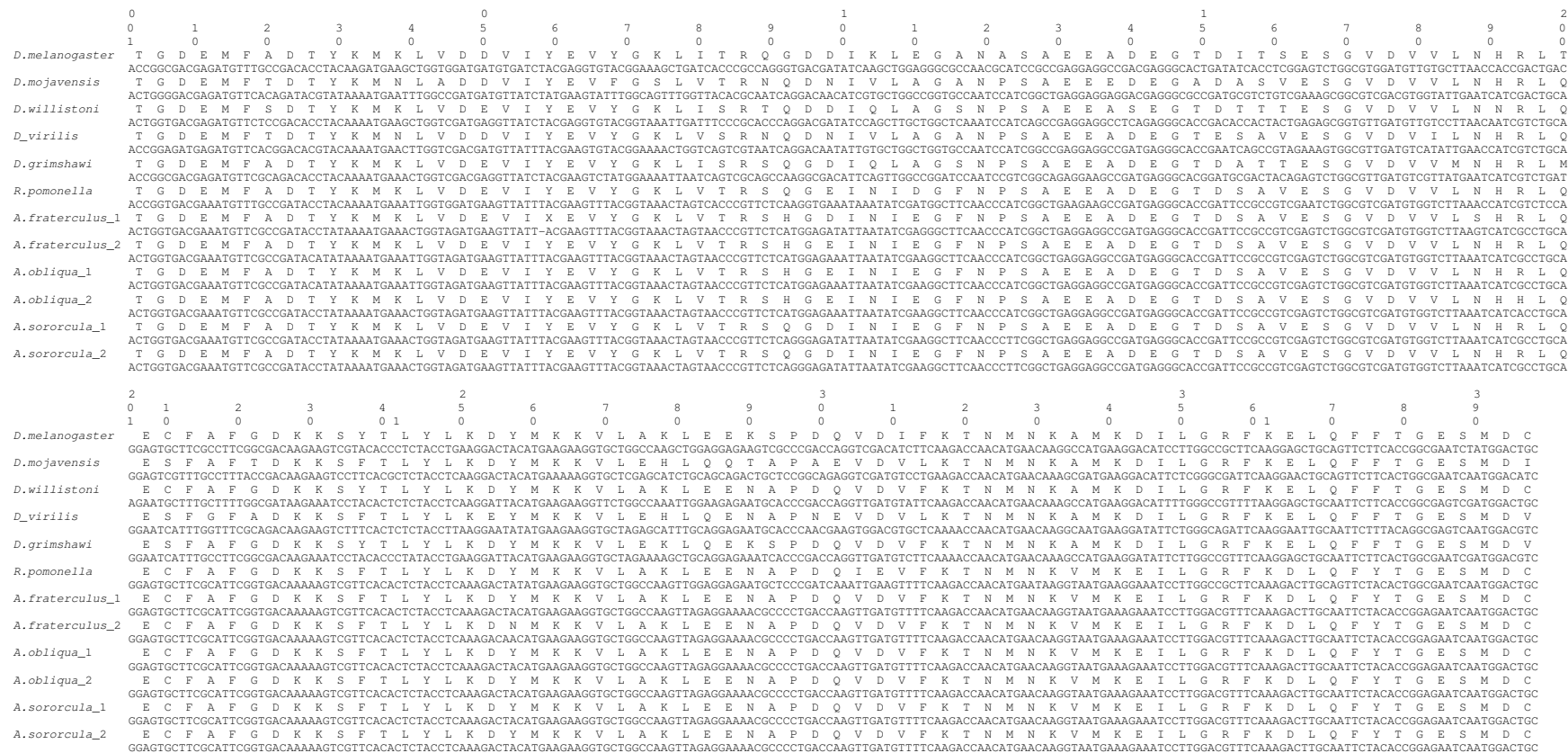
Figura A.13a Alinhamento do gene *Porin* utilizado para análises de seleção positiva realizadas no PAML.

```

          1       2       3       4
          0       0       0       0
D.melanogaster  S A V L G Y Q G W L A G Y
                TCTGCCGTGGCTGGCTACCAAGGATGGTTGGCCGGCTAC
D.willistoni   S A V L G Y Q G W L A G Y
                TCTGCTGTGCTGGGCTACCAAGGCTGGTTGGCTGGCTAT
D.mojavensis  S A V L G Y Q G W L A G Y
                TCGGCTGTTTTGGGTTATCAGGGCTGGTTGGCTGGCTAC
D_virilis     S A V L G Y Q G W L A G Y
                TCTGCCGTGGCTGGGCTACCAAGGCTGGTTGGCCGGTTAT
D.grimshawi   S A V L G Y Q G W L A G Y
                TCTGCTGTGCTGGCTACCAAGGCTGGTTGGCCGGCTAT
R.pomonella   S A V L G Y E G W L A G Y
                TCAGCTGTTTTGGGTTATGAAGGCTGGTTGGCTGGTTAC
A.fraterculus_1 S A V L G Y E G W L A G Y
                TCAGCTGTTTTGGGTTATGAGGTTGGTTAGCCGGTTAC
A.fraterculus_2 S A V L G Y E G W L A G Y
                TCAGCTGTTTTGGGTTATGAGGTTGGTTAGCCGGTTAC
A.obliqua_1   S A V L G Y E G W L A G Y
                TCAGCCGTTTTGGGTTATGAGGTTGGTTAGCCGGTTAC
A.obliqua_2   S A V L G Y E G W L A G Y
                TCAGCTGTTTTGGGTTATGAGGTTGGTTAGCCGGTTAC
A.sororcula_1 S A V L G Y E G W L A G Y
                TCAGCTGTTTTGGGTTATGAGGTTGGTTAGCCGGTTAC
A.sororcula_2 S A V L G Y E G W L A G Y
                TCAGCTGTTTTGGGTTATGAGGTTGGTTAGCCGGTTAC

```

**Figura A.13b** Alinhamento do gene *Porin* utilizado para análises de seleção positiva realizadas no PAML.



**Figura A.14** Alinhamento do gene *Translational Control Tumor Protein* utilizado para análises de seleção positiva realizadas no PAML.

```

0          1          2          3          4          5          6          7          8          9          10         11         12         13         14
1          0          0          0          0          0          0          0          0          0          0          0          0          0
D.melanogaster V L Q K A F N S F D H Q K T G S I P T E M V A D I L R L M G Q P F D K K I L E E L I E E V D E
GTGCTCCAGAAGGCGTTCAACAGCTTCGATCACCAGAAAGACTGGCTCCATCCCCACCGAGATGGTCGGCGACATACTGCGCCTGATGGGTGAGCCCTTCGACAAAGAAATCCTGGAGGAACTGATCGAGGAGGTCGATGAG
D.willistoni V L Q K A F N S F D H Q K T G S I P T E M V A D I L R L M G Q P F D K K I L E E L I E E V D E
GTGCTCCAGAAGGCGTTCAACAGCTTCGATCACCAGAAAGACTGGCAGCATCCCCACCGAGATGGTGTGATATCCTGCGCCTTATGGGTCAACCCCTTCGATAAGAAAATCCTGGAGGAACTGATCGAGGAGGTCGACGAA
D.mojavensis V L Q K A F N S F D H Q K T G S I P T E M V A D I L R L M G Q P F D K K I L E E L I E E V D E
GTGCTCCAGAAGGCGTTCAACAGCTTCGATCACCAGAAAGACTGGCAGCATCCCCACCGAAATGGTGGCGATATTCGCGGTCTGATGGGTGAGCCCTTCGACAAAGAAATCCTGGAGGAACTGATCGAAGAGGTCGATGAG
D.virilis V L Q K A F N S F D H Q K T G S I P T E M V A D I L R L M G Q P F D K K I L E E L I E E V D E
GTGCTCCAGAAGGCGTTCAACAGCTTCGATCACCAGAAAGACTGGCAGCATCCCCACCGAAATGGTGGCGATATTCGCGGTCTGATGGGTGAGCCCTTCGACAAAGAAATCCTGGAGGAACTGATCGAGGAGGTCGACGAA
D.grimshawi V L Q K A F N S F D H Q K T G S I P T E M V A D I L R L M G Q P F D K K I L E E L I D E V D E
GTGCTCCAGAAGGCGTTCAACAGCTTCGATCACCAGAAAGACTGGCAGCATCCCCACCGAAATGGTGGCGATATTCGCGGTCTGATGGGTGAGCCCTTCGACAAAGAAATCCTGGAGGAACTGATCGATGAGGTCGATGAG
R.pomonella V L Q K A F N S F D H Q K T G S I P T E M V A D I L R L M G Q P F D K K I L E E L I D E V D E
GTGCTCCAGAAGGCGTTTAAATAGCTTCGATCACCAGAAAGACTGGTGTAGTATCCCTACGAAATGGTGGCGGATATTCGCGGTGATGGGCCAACCCCTTCGATAAGAAAATTTGGAGGAACTGATCGATGAGGTCGATGAA
A.fraterculus_1 V L Q K A F N S F D H Q K T G S I P T E M V A D I L R L M G Q P F D K K I L E E L I E E V D E
GTGCTCCAGAAGGCGTTTAAATAGCTTCGATCACCAGAAAGACTGGTGTAGTATCCCTACGAAATGGTGGCGGATATTCGCGGTGATGGGCCAACCCCTTCGATAAGAAAATTTGGAGGAACTGATCGAAGAGGTCGATGAA
A.fraterculus_2 V L Q K A F N S F D H Q K T G S I P T E M V A D I L R L M G Q P F D K K I L E E L I E E V D E
GTGCTCCAGAAGGCGTTTAAATAGCTTCGATCACCAGAAAGACTGGTGTAGTATCCCTACGAAATGGTGGCGGATATTCGCGGTGATGGGCCAACCCCTTCGATAAGAAAATTTGGAGGAACTGATCGAAGAGGTCGATGAA
A.obliqua_1 V L Q K A F N S F D H Q K T G S I P T E M V A D I L R L M G Q P F D K K I L E E L I E E V D E
GTGCTCCAGAAGGCGTTTAAATAGCTTCGATCACCAGAAAGACTGGTGTAGTATCCCTACGAAATGGTGGCGGATATTCGCGGTGATGGGCCAACCCCTTCGATAAGAAAATTTGGAGGAACTGATCGAAGAGGTCGATGAA
A.obliqua_2 V L Q K A F N S F D H Q K T G S I P T E M V A D I L R L M G Q P F D K K I L E E L I E E V D E
GTGCTCCAGAAGGCGTTTAAATAGCTTCGATCACCAGAAAGACTGGTGTAGTATCCCTACGAAATGGTGGCGGATATTCGCGGTGATGGGCCAACCCCTTCGATAAGAAAATTTGGAGGAACTGATCGAAGAGGTCGATGAA
A.sororcula_1 V L Q K A F N S F D H Q K T G S I P T E M V A D I L R L M G Q P F D K K I L E E L I E E V D E
GTGCTCCAGAAGGCGTTTAAATAGCTTCGATCACCAGAAAGACTGGTGTAGTATCCCTACGAAATGGTGGCGGATATTCGCGGTGATGGGCCAACCCCTTCGATAAGAAAATTTGGAGGAACTGATCGAAGAGGTCGATGAA
A.sororcula_2 V L Q K A F N S F D H Q K T G S I P T E M V A D I L R L M G Q P F D K K I L E E L I E E V D E
GTGCTCCAGAAGGCGTTTAAATAGCTTCGATCACCAGAAAGACTGGTGTAGTATCCCTACGAAATGGTGGCGGATATTCGCGGTGATGGGCCAACCCCTTCGATAAGAAAATTTGGAGGAACTGATCGAAGAGGTCGATGAA

```

**Figura A.15** Alinhamento do gene *TroponinC* utilizado para análises de seleção positiva realizadas no PAML.



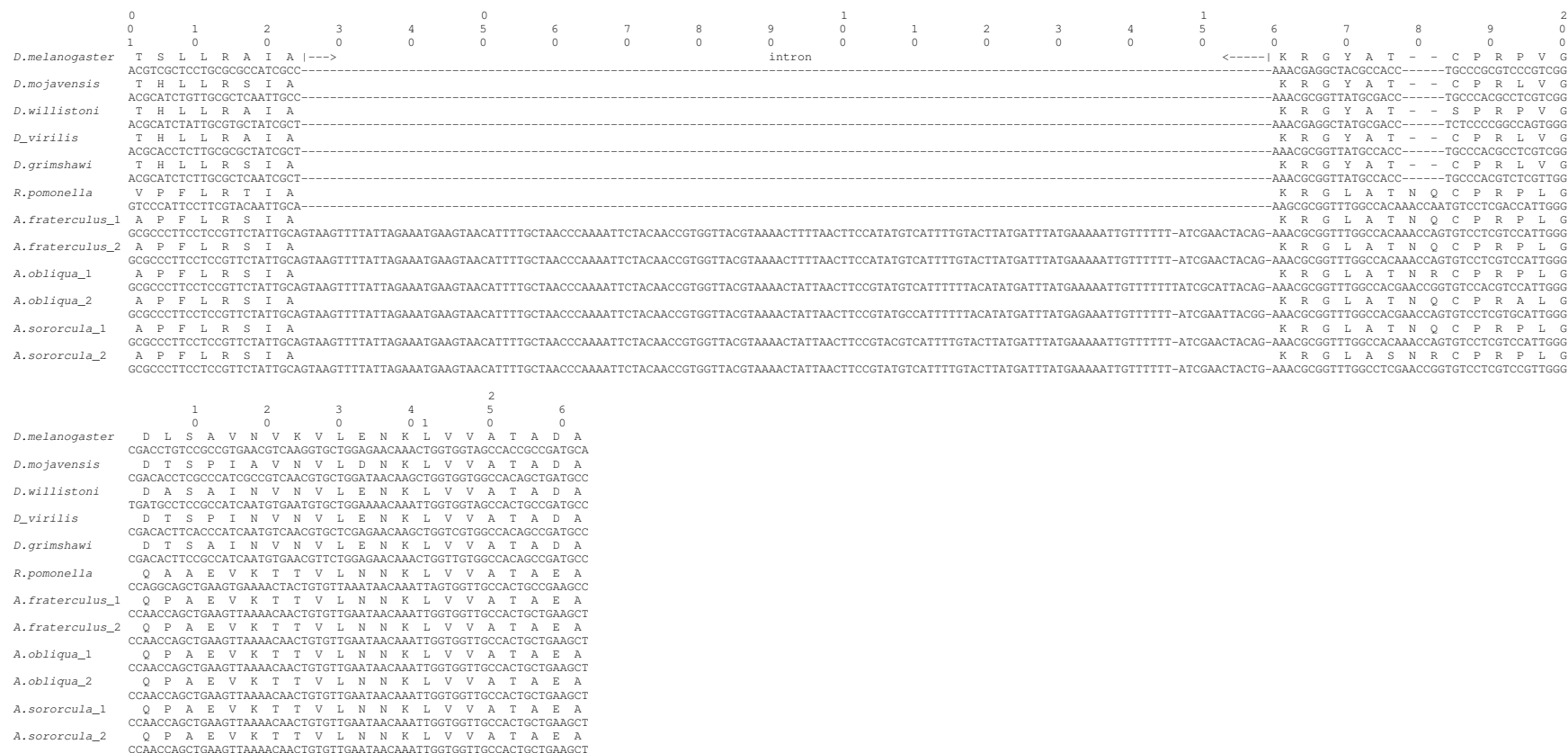


Figura A.16 Alinhamento do gene *Ubiquinol Cytochrome Reductase* utilizado para análises de seleção positiva realizadas no PAML.

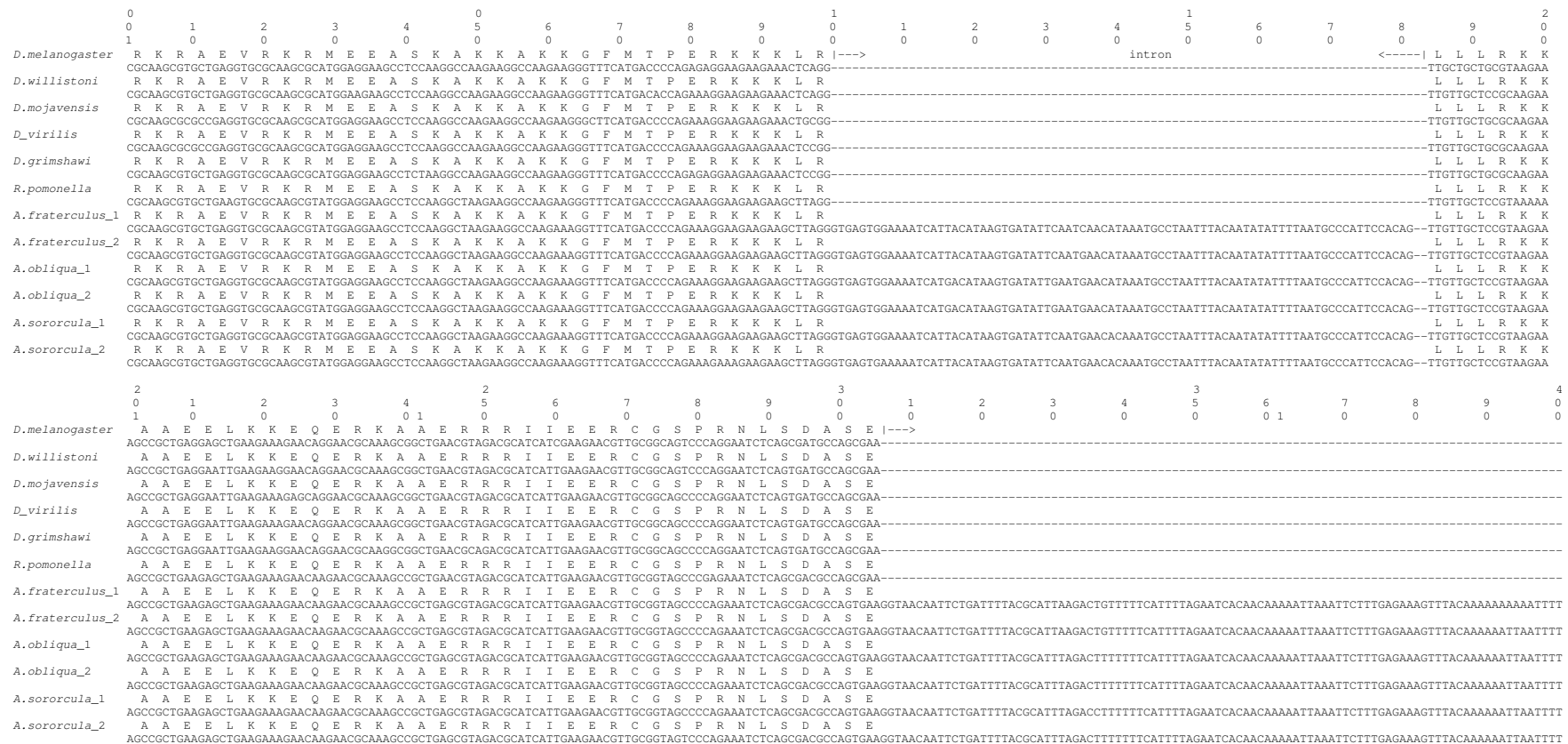


Figura A.17a Alinhamento do gene *WingsUp* utilizado para análises de seleção positiva realizadas no PAML.

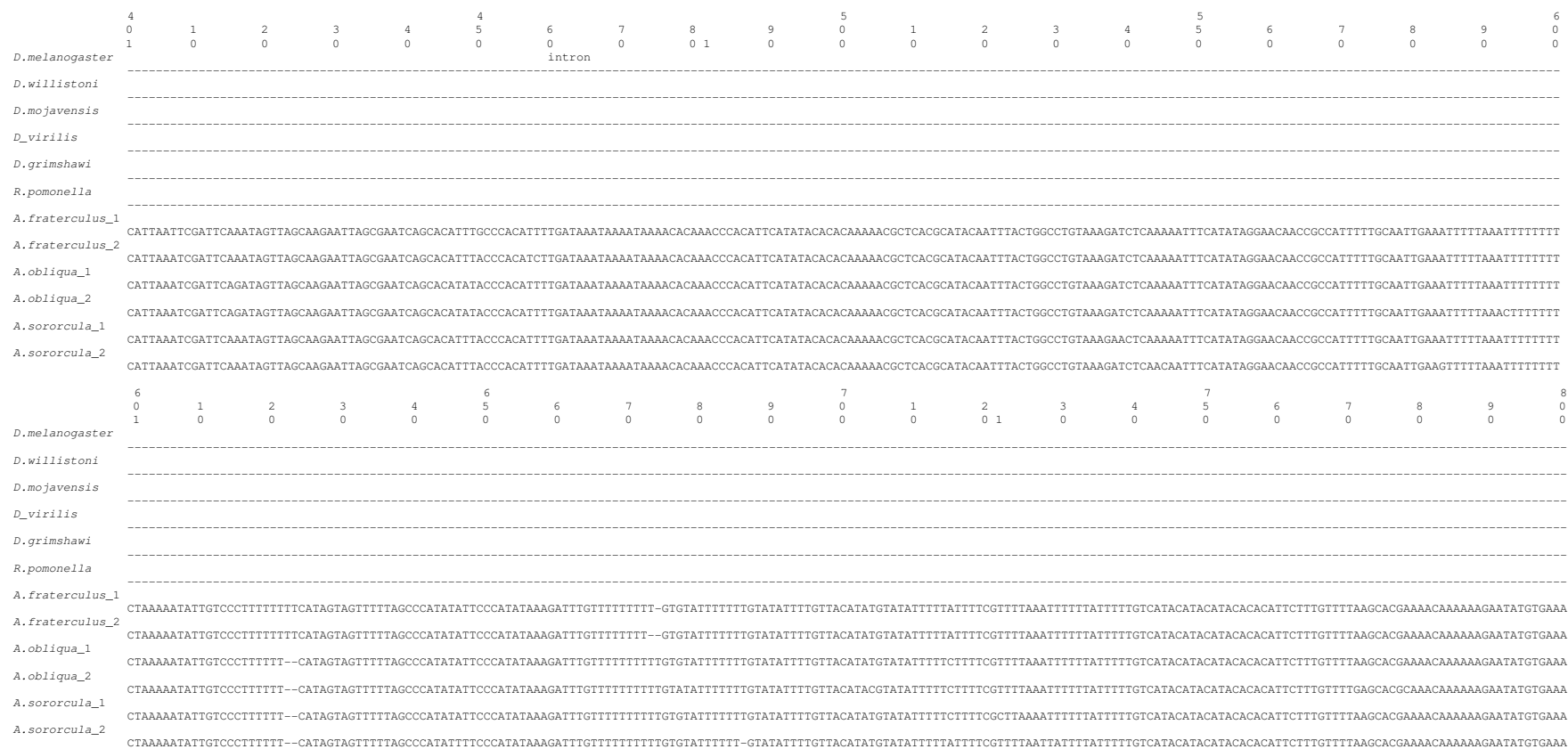


Figura A.17b Alinhamento do gene *WingsUp* utilizado para análises de seleção positiva realizadas no PAML.

```

      8           8           9
      0         1         2         3         4         5         6         7         8         9         0         1         2
      1         0         0         0         0 1         0         0         0         0         0         0         0
D.melanogaster <-----| D T L K S L I K Q H Y D R I N K L E D Q K Y
                G A C A C A C T C A A A T C T G A T C A A G C A A C A C T A T G A C A G G A T T A A T A A A T T G G A G G A C C A G A A A T A T G A
D.willistoni    G E L Q E I C E E Y Y E R M Y I C E G Q K W
                G C C G A A T T G C A A G A G A T T G C G A A G A G T A T T A C G A G C G T A T G T A T A T T T G T G A A G G C C A G A A A T G G G A
D.mojavensis   G E L Q E I C E E Y Y E R M Y I C E G Q K W
                G C C G A A T T G C A A G A G A T T G C G A A G A G T A T T A C G A G C G T A T G T A T A T T T G T G A A G G C C A G A A A T G G G A
D_virilis      G E L Q E I C E E Y Y E R M Y I C E G Q K W
                G C C G A A T T G C A A G A G A T T G C G A A G A G T A T T A C G A G C G T A T G T A T A T T T G T G A A G G C C A G A A A T G G G A
D.grimshawi    G E L Q E I C E E Y Y E R M Y I C E G Q K W
                G C C G A A T T G C A A G A G A T T G C G A A G A G T A T T A C G A G C G T A T G T A T A T T T G T G A A G G C C A G A A A T G G G A
R.pomonella    E T L K N I I K Q H Y D R I N K L E D Q K Y
                G A A C A C T T A A A A T A T A A T C A A G C A A C A C T A T G A C A G G A T T A A T A A A T T G G A G G A C C A G A A A T A T G A
A.fraterculus_1 T T G T T A A T T T T G T T A A A A A C G C A A T T C T T C A C T T C C A C T T A T C T G C A - G A C A C A C T T A G G A G T T T A A T C A A G C A A C A C T A T G A A A G G A T T T G T A A A T T G G A G G A C C A G A A A T A T G A
A.fraterculus_2 D T L R S L I K Q H Y E R I C K L E D Q K Y
                D T L K S L I K Q H Y E R I C K L E D Q K Y
A.obliqua_1    T T G T T A A T T T T G T T A A A A A C G C A A T T C T T C A C T T C C A C T T A T C T G C A - G A C A C A C T T A A G A G T T T A A T C A A G C A A C A C T A T G A A A G G A T T T G T A A A T T G G A G G A C C A G A A A T A T G A
A.obliqua_2    E T L K S L I K Q H Y E R I C K L E D Q K Y
                G A A C A C T T A A A A T A T A A T C A A G C A A C A C T A T G A C A G G A T T A A T A A A T T G G A G G A C C A G A A A T A T G A
A.sororcula_1 D T L R S L I K Q H Y E R I C K L E D Q K Y
                G A A C A C T T A A G A G T T T A A T C A A G C A A C A C T A T G A A A G G A T T T G T A A A T T G G A G G A C C A G A A A T A T G A
A.sororcula_2 D T L R S L I K Q H Y E R I C K L E D Q K Y
                D T L R S L I K Q H Y E R I C K L E D Q K Y
T T G T T A A T T T T G T T A A A A A C G C A A T T C T T C A C T T C C A C T T A T C T G C A - G A C A C A C T T A G G A G T T T A A T C A A G C A A C A C T A T G A A A G G A T T T G T A A A T T G G A G G A C C A G A A A T A T G A

```

Figura A.17c Alinhamento do gene *WingsUp* utilizado para análises de seleção positiva realizadas no PAML.

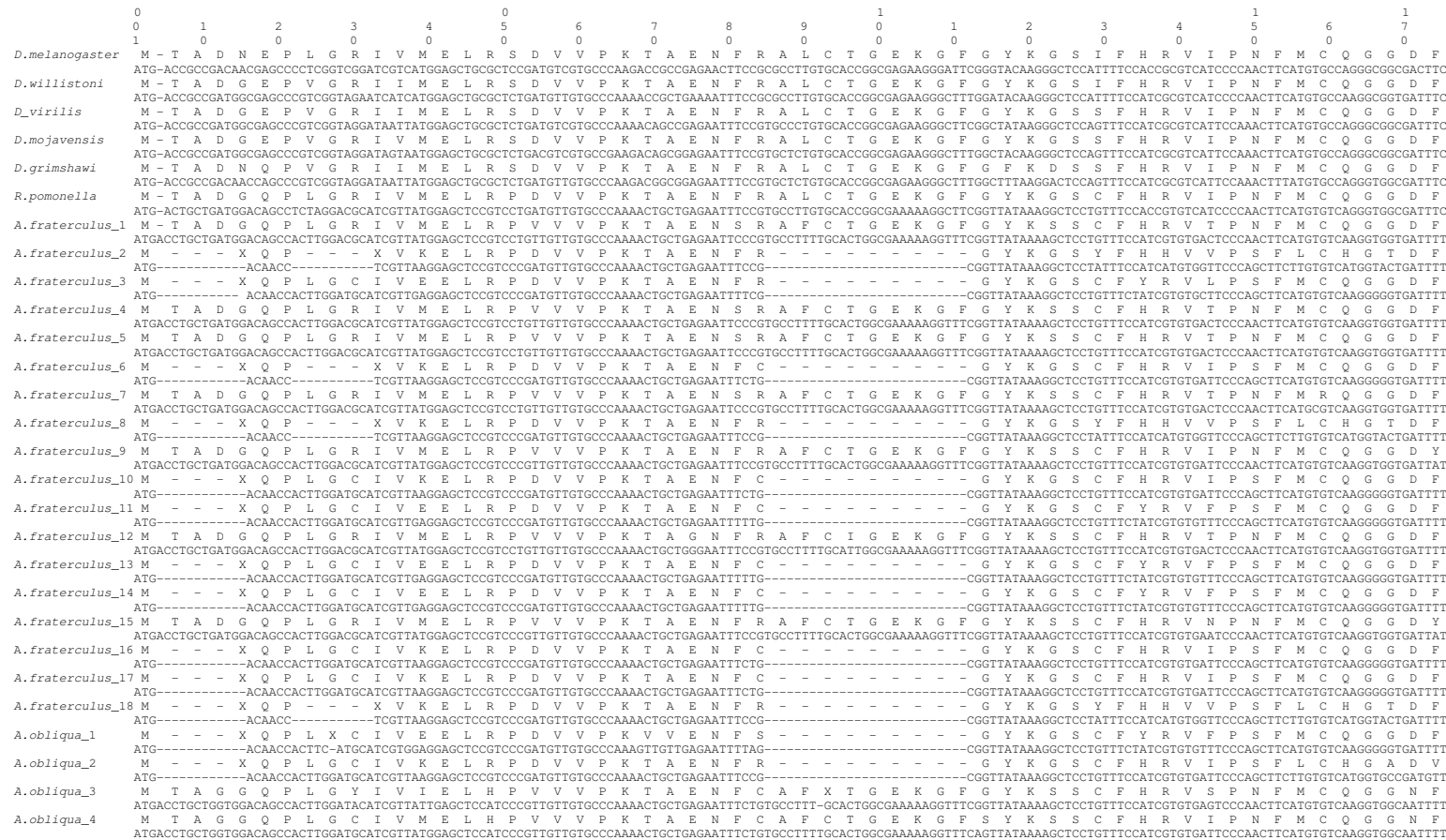


Figura A.18a Alinhamento do gene *Cyclophylin* utilizado para análises filogenéticas.

