

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

Laboratório de Genética de Populações e Evolução

**Identificação de genes com alta taxa evolutiva em tecidos reprodutivos femininos de
moscas-das-frutas *Anastrepha obliqua***

Prof. Orientador: Dr. Reinaldo Alves de Brito

Aluna: Vanessa Regina Gonçalves

São Carlos

2010

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

Laboratório de Genética de Populações e Evolução

Identificação de genes com alta taxa evolutiva em tecidos reprodutivos femininos de moscas-das-frutas *Anastrepha obliqua*

Candidata: Vanessa Regina Gonçalves

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

SÃO CARLOS

2010

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

G635ig

Gonçalves, Vanessa Regina.

Identificação de genes com alta taxa evolutiva em tecidos reprodutivos femininos de moscas-das-frutas *Anastrepha obliqua* / Vanessa Regina Gonçalves. -- São Carlos : UFSCar, 2010.
102 f.

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

1. Genética e evolução. 2. Biblioteca de cDNA. 3. Anastrepha. 4. Seleção natural. I. Título.

CDD: 575 (20^a)

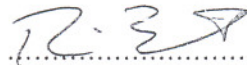
UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E EVOLUÇÃO

**IDENTIFICAÇÃO DE GENES COM ALTA TAXA EVOLUTIVA
EM TECIDOS REPRODUTIVOS FEMININOS DE MOSCAS-
DAS-FRUTAS ANASTREPHA OBLIQUA**

Dissertação de Mestrado de Vanessa Regina Gonçalves

Banca Examinadora


Prof. Dr. Reinaldo Alves de Brito



Profa. Dra. Vera Nisaka Solferini



Prof. Dr. Lúcio Antonio de Oliveira Campos



Dedico essa dissertação especialmente ao meu pai que sempre me apoiou em todas as minhas decisões e a todos que acreditaram em mim.

AGRADECIMENTOS:

Agradeço primeiramente a todos os professores e funcionários do departamento de genética e evolução, sem os quais não seria possível a realização deste trabalho.

A CAPES e a FAPESP pelo apoio financeiro.

Aos meus pais, pela confiança e incentivo em todos os momentos da trajetória de minha vida.

Aos amigos de laboratório de Genética de Populações e Evolução: Nelci, Gustavo, Bruno, Fernanda, Marcos, Daniel, André, Aline, Isabela, Carol, Vanessa e Victor pela amizade: de discussões de coisas importantes a muitos momentos de risos; e ao Ider pela amizade e pela grande disponibilidade de tempo e paciência para explicar e ensinar nos momentos de dúvida e desespero.

Ao professor Dr. Flávio e ao Wilson pela colaboração de uma importante parte do projeto, por disponibilizarem o laboratório e todos os equipamentos necessários.

Ao meu orientador, professor Dr. Reinaldo por ter me recebido tão bem em seu laboratório desde começo. Pela paciência, dedicação, disponibilidade de tempo, entusiasmo, confiança e incentivo. Agradeço por ter acreditado em mim, aprendi muito mais do que esperava e adorei tudo o que aprendi.

Termino este trabalho com a certeza de que escolhi o melhor lugar, os melhores colegas e o melhor orientador para realizá-lo.

A mente que se abre a uma nova idéia jamais voltará ao seu tamanho original.

Albert Einstein

Resumo

O gênero *Anastrepha* é o maior dentro da família Tephritidae (Trypetinae, Toxotrypanini). A construção de bibliotecas de cDNA é importante para a identificação de novos genes e no estabelecimento de marcadores genéticos em organismos não-modelos como do gênero *Anastrepha*. Dentre as proteínas envolvidas na reprodução podemos citar as pertencentes da cascata sexual e as que são responsáveis por formar a parede do ovo. Assim, este trabalho teve como objetivo construir uma biblioteca de cDNA a partir de um *pool* de tecidos reprodutivos femininos de *Anastrepha obliqua* para gerar ESTs, afim de buscar genes com alta taxa evolutiva, focando em proteínas coriônicas e vitelínicas. No Total foram seqüenciados 2304 clones obtidos a partir dos órgãos reprodutivos de fêmeas de moscas de *Anastrepha obliqua*. Ao todo foram gerados 310 contigs e 506 singlets que foram classificados em diferentes classes de proteínas. As proteínas coriônicas e o *Sxl* revelaram estar sob seleção positiva, assim como a vitelínica *Vm 26Aa'* que possivelmente foi gerada por uma duplicação gênica do *Vm 26Aa*, enquanto que o *Tra2* parece estar sob seleção purificadora. A inferência da estrutura secundária das proteínas estudadas revelou que os sítios sob seleção positiva estavam presentes do lado externo da membrana. As análises populacionais dos genes coriônicos, vitelínicos, *Sxl* e *Tra2* não revelaram uma separação entre as 3 espécies, apesar de em algumas situações ter ocorrido uma separação entre alguns dos haplótipos para uma determinada espécie. Com resultados obtidos, foi possível identificar novos genes, fazer o seqüenciamento completo para alguns genes coriônicos e vitelínicos em *pools* de outras espécies de *Anastrepha*, e também foi possível identificar que alguns destes genes estão sob pressão de seleção positiva.

Abstract

The genus *Anastrepha* is the largest in the family Tephritidae (Trypetinae, Toxotrypanini). The construction of cDNA libraries is important for identifying new genes and establishing genetic markers on non-model organisms such as the genus *Anastrepha*. Among the proteins involved in reproduction, we can cite those belonging to the cascade of sexual proteins and those responsible for forming the eggshell. The purpose of this study was to build a cDNA library from a pool of female reproductive tissues of *Anastrepha obliqua* to breed ESTs, to search for genes with a high evolutionary rate, focusing on chorionic and vitelline proteins. We sequenced 2304 clones obtained from the reproductive tissues of female flies *Anastrepha obliqua*. In total, 310 contigs and 506 singlets were generated which were classified into different protein classes. The chorionic proteins and *Sxl* revealed to be under positive selection as well as the vitelline *Vm 26Aa*, which was possibly generated by a gene duplication of *Vm 26aa*, while *Tra2* seems to be under purifying selection. The inference of the secondary structure of the proteins studied revealed that the sites under positive selection were present on the outside membrane. The population analysis of chorionic, vitelline, *Sxl* and *Tra2* revealed no separation between the three species, although in some situations separation between some of the haplotypes for a single species has occurred. With these results, it was possible to identify new genes, make the full sequencing for some vitelline and chorionic genes on pools of other species of *Anastrepha*, and we also identified that some of these genes are under positive selection pressure.

Sumário

RESUMO.....	6
1- INTRODUÇÃO.....	10
1.1- ESPECIAÇÃO.....	10
1.2 O GÊNERO ANASTREPHA.....	11
1.3- O APARELHO REPRODUTOR FEMININO.....	14
1.4- BIBLIOTECAS DE cDNA.....	15
1.5- GENES SOB SELEÇÃO POSITIVA.....	16
1.6- GENES DA CASCATA SEXUAL.....	17
1.7- PROTEÍNAS CORIÔNICAS E VITELÍNICAS.....	18
2- OBJETIVOS.....	20
2.1- OBJETIVO GERAL.....	20
2.2- OBJETIVOS ESPECÍFICOS.....	20
3- MATERIAIS E MÉTODOS.....	21
3.1.1- COLETA E MANUTENÇÃO DAS MOSCAS-DAS-FRUTAS.....	21
3.1.2- <i>Locais de coleta para construção da biblioteca.....</i>	<i>21</i>
3.2- CONSTRUÇÃO DA BIBLIOTECA DE cDNA.....	22
3.2.1- EXTRAÇÃO DE RNA TOTAL.....	22
3.2.2- <i>Extração de RNA mensageiro.....</i>	<i>23</i>
3.2.3- <i>Biblioteca de cDNA.....</i>	<i>23</i>
3.3- QUALIDADE DA BIBLIOTECA DE cDNA.....	24
3.4- <i>Seqüenciamento da Biblioteca de cDNA.....</i>	<i>25</i>
3.5- ANÁLISES DAS SEQÜÊNCIAS DE ESTS.....	25
3.6- CONSTRUÇÕES DE PRIMERS.....	26
3.1.3- <i>Locais de coleta para formação dos pools de DNA.....</i>	<i>27</i>
3.7- EXTRAÇÃO DE DNA.....	29
3.8 AMPLIFICAÇÃO POR PCR E PURIFICAÇÃO.....	30
3.9 QUANTIFICAÇÃO DOS PRODUTOS DE PCR.....	31
3.10- CLONAGEM.....	31
3.11- SEQÜENCIAMENTO DOS GENES <i>VM26AA</i> ; <i>VM26AA'</i> ; <i>Cps38</i> ; <i>Cps16</i> ; <i>TRA2</i> E <i>SXL</i>	33
3.12- TESTES DE DETECÇÃO DE SELEÇÃO POSITIVA.....	33
3.13- PREDIÇÃO DA ESTRUTURA PROTEICA.....	35
4- RESULTADOS E DISCUSSÃO.....	36
4.1- A BIBLIOTECA DE cDNA.....	36
4.1.1- <i>Integridade do RNA total.....</i>	<i>36</i>
4.1.2- <i>Integridade do RNA mensageiro.....</i>	<i>37</i>
4.2- ANÁLISE DO cDNA.....	38
4.2.2- <i>Classificação das Proteínas.....</i>	<i>42</i>
4.3- ANÁLISE DE SELEÇÃO PARA OS GENES CORIÔNICOS E DA CASCATA SEXUAL.....	44
4.4- ESTRUTURA SECUNDÁRIA DAS PROTEÍNAS QUE APRESENTARAM SELEÇÃO POSITIVA.....	55

4.5- AVALIAÇÃO DA UTILIDADE DOS GENES ESTUDADOS COMO MARCADORES DE IDENTIFICAÇÃO DE DIFERENTES ESPÉCIES NO GRUPO <i>FRATERCULUS</i>	63
5- CONCLUSÕES	76
6- REFERÊNCIAS	78

1- Introdução

1.1- Especiação

A causa mais comum de isolamento reprodutivo em animais sexuais – a impossibilidade de troca de genes entre dois organismos (WU & TING 2004) – são os obstáculos à fecundação entre as espécies (DOBZHANSKY 1937). A maioria dos pesquisadores considera que processos que levam ao isolamento reprodutivo são mais prováveis de levar à especiação, (BULTIN & RITCHIE 2001) e, por conseguinte, processos de especiação ocorreriam mais frequentemente em situações de ausência de fluxo gênico. Deriva daí a sugestão de que o modo mais comum de especiação esteja associado à diferenciação genética em populações isoladas – especiação alopátrica, na definição de Ernst Mayr. Isso não quer dizer que a especiação simpátrica não possa ocorrer, porém seria um processo mais raro (RITCHIE 2007), mas nem por isso menos importante. Assim, para a especiação acontecer não é necessário uma barreira genética que cesse o fluxo gênico, uma vez que o isolamento entre as espécies pode ocorrer com ou sem fluxo gênico. Muito embora todas as forças evolutivas possam causar diferenciação entre as populações e especiação, a maioria dos estudos relaciona eventos associados à seleção natural, a deriva genética e o balanço dessas duas forças (TEMPLETON 1989).

Tem sido longa a história de tentar identificar genes que causam o isolamento reprodutivo (COYNE & ORR 2004), e particularmente a tentativa de se identificar os genes envolvidos no processo de especiação. Após a ocorrência de isolamento, as espécies começam a se tornar divergentes devido ao acúmulo de diferenças genéticas ocorridas ao longo do tempo. Conforme o tempo de divergência entre duas espécies aumenta, torna-se mais difícil distinguir genes que são realmente responsáveis pelo isolamento reprodutivo de outros polimorfismos que surgiram após o processo de especiação. Assim, a identificação de genes que estão relacionados diretamente no processo de especiação é facilitada pelo estudo dos componentes genéticos das diferenças fenotípicas em espécies irmãs ou proximamente relacionadas (ZENG *et al.* 2000). Têm sido chamados (talvez erroneamente) de genes de especiação aqueles que contribuem para o isolamento reprodutivo. Esta definição inclui genes que causam isolamento devido a fatores fisiológicos, comportamentais ou até mesmo

ecológicos. Um híbrido, por exemplo, pode ser viável e fértil, porém, por exemplo, menos resistente ao frio (WU & TING 2004).

A maioria dos estudos buscando o isolamento de genes de especiação tem tentado identificar genes associados à produção de esterilidade ou inviabilidade masculina em híbridos de espécies distintas. Híbridos normalmente possuem comportamentos distintos, além de baixa viabilidade ou esterilidade gamética, e há casos em que há isolamento pós-zigótico devido ao comportamento disfuncional (Kawata & Yoshimura 2000).

1.2- O gênero *Anastrepha*

Dentro da família Tephritidae encontra-se gênero *Anastrepha* (Trypetinae, Toxotrypanini) (NORRBOM 2001). Este gênero é endêmico da região neotropical começando desde o sul dos EUA até a América do Sul, com exceção do Chile e do norte do deserto de Atacama, na fronteira com o Peru (MALAVASI *et al.* 2000). Atualmente são reconhecidas 197 espécies (NORRBOM *et al.* 2004), 95 das quais estão distribuídas no Brasil (ZUCCHI 2000). A família Tephritidae possui grande importância econômica, pois a maioria dessas espécies são responsáveis por danos a cultura de frutas carnosas, o que causa prejuízo econômico.

Entre os grupos de espécies de *Anastrepha* Schiner, destaca-se o grupo *fraterculus* que é composto por 29 espécies (NORRBOM *et al.* 2000). Algumas espécies deste grupo, dentre elas *A. obliqua*, *A. fraterculus* e *A. sororcula* possuem grande importância econômica (SOLFERINI & MORGANTE 1987; ZUCCHI 2000), embora os estudos nestas espécies sejam escassos. Apesar de estarem proximamente relacionada, as espécies do grupo *fraterculus* apresentam pequenas diferenças em alguns marcadores, dentre os quais podemos destacar denticção e o tamanho da extremidade do acúleo, as manchas presentes no subescutelo e o padrão de coloração e morfologia da asa (fig1) (SELIVON 2000; MALAVASI & ZUCCHI 2000). O formato do ápice do acúleo é o principal caráter para a identificação específica de *Anastrepha*, porém as medidas do acúleo (total e do ápice) dessas espécies variam ao longo da distribuição geográfica e também entre os exemplares obtidos em um mesmo hospedeiro (ARAUJO & ZUCCHI 2006). Embora não exista um único marcador que permita a diferenciação das espécies do grupo *fraterculus*, a análise de todas as características citadas em conjunto, permite na grande maioria dos casos, uma correta identificação da espécie.

As espécies deste grupo apresentam mais de um hospedeiro e muitas vezes são consideradas polípagas. As mais polípagas são *A. fraterculus* desenvolvendo-se em cerca de 67 hospedeiros e *A. obliqua* que se desenvolve em 27 (ZUCCHI 2000). Apesar desta polifagia, alguma especificidade pode ser encontrada. Em *A. fraterculus* que infesta cerca de 67 espécies, 24 delas são Myrtaceae, *A. sororcula* infesta cerca de 19 espécies, sendo 15 delas Myrtaceae, e *A. obliqua* que infesta 27 espécies, 16 pertencem a família Myrtaceae e 8 são Anacardiaceae (ZUCCHI 2000). Alguns autores têm considerado que em insetos fitófagos, o polimorfismo associado ao hospedeiro pode representar o primeiro estágio de especiação (MALAVASI & MORGANTE 1983; FEDER *et al.* 1998).

A possibilidade de cruzamento e de se obter híbridos viáveis em espécies do grupo *fraterculus*, (SANTOS *et al.* 2001; SELIVON *et al.* 2005) é importante para o estudo de genes que possam estar envolvidos na determinação de diferenças entre as espécies deste estudo. Esta obtenção de híbridos em laboratório demonstra que os mecanismos de isolamento reprodutivo ainda não estão definidos para o isolamento completo destas espécies, isto sugere que elas provavelmente se divergiram há pouco tempo, não havendo ainda muitas diferenças genéticas significativas.

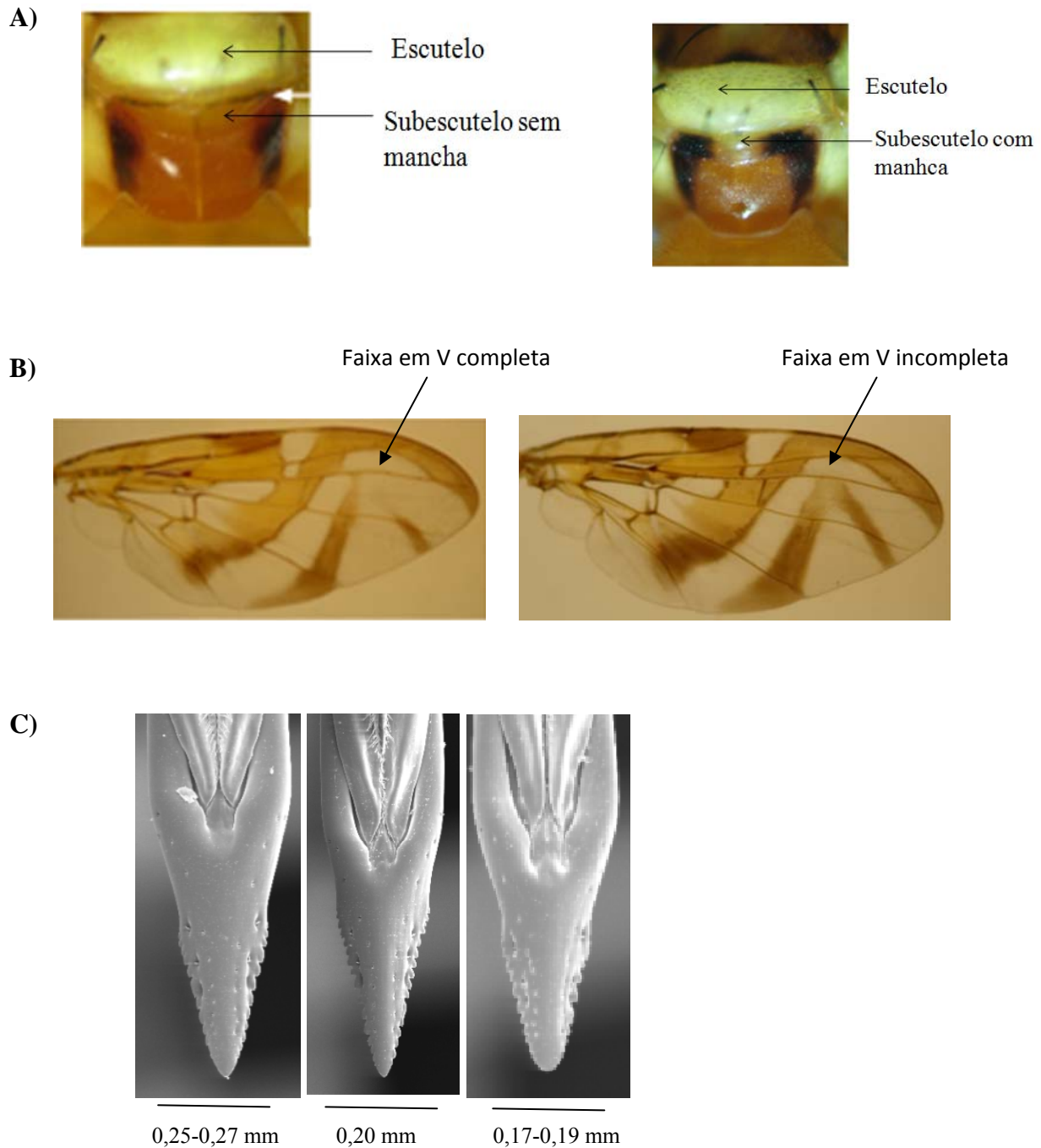


Fig.1- Caracteres usados na identificação das espécies. A- Tórax; B- asa; C- ápice do acúleo (Zucchi 2007). *A. fraterculus* possui mancha no subescutelo, faixa V geralmente completa, ápice do acúleo de 0,25 – 0,27 mm; *A. obliqua* não possui mancha no subescutelo, faixa V geralmente incompleta, ápice 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 de espessura.

1.3- O Aparelho reprodutor feminino

Os estudos dos órgãos reprodutivos devem abordar a herança e a variação dos padrões de morfológica por um lado e padrões de seleção fenotípica de outro (ARNQVIST 2000). A genitália animal evolui mais rapidamente e sua morfologia difere mesmo entre espécies proximamente relacionadas. Animais que possuem fecundação interna, a diversificação da genitália representa uma das mais surpreendentes formas de tendências evolutivas. (EBERHARD 1985). Em machos e fêmeas de *Drosófilas*, órgãos reprodutivos são em média duas vezes mais diferenciados entre as espécies do que caracteres não-reprodutivos (CIVETTA & SINGH 1995). A genitália masculina evolui para ser específica e única, de forma a adaptar-se adequadamente à genitália feminina (POMIANKOWSKI & MÜLLER 1995). Dessa forma, a divergência evolutiva dos traços, morfológicos e moleculares, que envolvem aspectos reprodutivos parece ser mais rápida, sendo o conflito sexual motor chave do processo de especiação (ARNQVIST *et al.* 2000).

A reprodução é o componente chave relacionado ao *fitness* de um organismo. Em nível molecular, a reprodução é um ato complexo de interações entre muitas proteínas (FINDLAY & SWANSON 2010). Genes que participam da reprodução também mostram normalmente sinais de evolução adaptativa. Mecanismos que poderiam explicar essa rápida evolução vista nessas proteínas seriam a seleção sexual, conflito sexual, imunidade e o reconhecimento dessas proteínas (SWANSON & VACQUIER 2002). Alguns estudos têm mostrado que genes codificadores de proteínas de glândulas acessórias evoluem diferentemente de outras classes de genes. Estes genes evidenciam uma alta divergência de aminoácidos quando comparados com os genes não reprodutivos em espécies estritamente relacionadas: *D. melanogaster* e *D. simulans* (SWANSON *et al.* 2001; MUELLER *et al.* 2005). Têm sido crescentes os estudos em evolução com proteínas relacionadas a caracteres sexuais e reprodução. Em machos há pesquisas detalhadas de proteínas ligadas com a reprodução, como por exemplo, as proteínas presentes no líquido seminal ejaculado pelos machos (GILLOTT 2003; CHAPMAN & DAVIES 2004; WOLFNER *et al.* 2005; CLARCK *et al.* 2006; PIOANI 2006; WONG *et al.* 2008), porém em fêmeas estes estudos ainda são escassos e as funções das proteínas ainda são pouco conhecidas (SWANSON *et al.* 2001; SWANSON *et al.* 2004; PANHUIS & SWANSON 2006). Partindo do fato de que muitas proteínas estão envolvidas em interações com outro sexo, de forma direta ou indireta, os

mecanismos evolutivos só serão mais bem compreendidos se seqüências gênicas de ambos os sexos forem estudadas.

Poucos são os genes que até o momento foram identificados como sendo genes de especiação. Alguns por terem altos níveis de seleção positiva, apesar de não apresentarem funções celulares e moleculares distintas, podem estar relacionados com o processo de especiação. Estes genes são potencialmente mais propensos a gerar isolamento reprodutivo e diferenciação (WU & TING 2004) e por isso o estudo dos mesmos tem sido enfatizado. Claramente, estudos genômicos em espécies que tenham divergido recentemente proporcionam uma grande oportunidade para se identificar genes que evoluem mais rapidamente entre as espécies (RITCHIE 2007).

1.4- Bibliotecas de cDNA

A construção de bibliotecas de cDNAs ou ESTs (Expressed Sequenced Tags) tem como objetivo inicial a identificação de um conjunto completo de genes que estão sendo expressos em um determinado tecido, podendo ser de grande importância no estabelecimento de novos marcadores genéticos, sejam eles microssatélites ou SNPs ([*single nucleotide polymorphism*](#)), que podem ser usados para o estabelecimento de mapas de ligação. 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 seqüenciados.

As bibliotecas feitas a partir de cDNAs são essenciais quando o que se deseja obter são genes que estão sendo expressos em um determinado organismo, tecido ou condição (CORTER *et al.* 1997). Os dados gerados por estas bibliotecas têm sido usados não apenas para a identificação de novos genes (SWANSON *et al.* 2004; KELLEHER *et al.* 2007; GOMULSKI *et al.* 2008), mas também para a produção de novos marcadores de forma a gerar um mapa de ligação nas espécies em estudo (BEEMAN & BROWN 1999) ou avaliar níveis de expressão em geral de forma a identificar genes que estejam sendo super ou sub expressos (KELLEHER *et al.* 2007). Posteriormente, ainda pode ser feito um estudo mais detalhado de alguns dos genes encontrados, para determinar sua expressão em diferentes tecidos e estágios da vida, caracterizando assim, a regulação e função destes genes (GOMULSKI *et al.* 2008). Se estas bibliotecas são feitas a partir de um *pool* representativo de indivíduos da população, as várias cópias obtidas de uma determinada região gênica permitem

também uma perscrutação dos níveis de polimorfismo para a mesma encontrados na população (ROETHELE *et al.* 2001; LORENZEN *et al.* 2005).

1.5- Genes Sob Seleção Positiva.

Conforme mencionado no item 2.3, diversas proteínas envolvidas na reprodução evoluem mais rapidamente (SWANSON *et al.* 2001; SWANSON & VACQUIER 2002; MUELLER *et al.* 2005), embora a identificação desta rápida evolução não seja um processo trivial (SOBRINHO & DE BRITO *subm.*). Há diversos testes atualmente estabelecidos para a identificação de desvios da neutralidade em uma molécula. Alguns destes testes são populacionais e lidam com o destino de polimorfismos (TAJIMA 1989; FU & LI 1993; FAY & WU 2000), embora a identificação de seleção positiva em dados populacionais seja ainda um procedimento complexo. Outros testes consideram contrastes nos quais já tenha havido tempo suficiente para a fixação da variação genética sinônima e não-sinônima. Em nível molecular, para medir a taxa de evolução dessas proteínas é calculado a razão da taxa de substituições não sinônimas (dN) para substituições sinônimas (dS) na região codificante (NIELSEN & YANG 1998; YANG 1998). Se esta razão (ω) for maior que um é um indicativo de que esteja ocorrendo seleção positiva. Diversas proteínas envolvidas em reconhecimento e determinação sexual têm revelado valores de dN/dS significativamente acima de 1 (SWANSON *et al.* 2001). Valores abaixo de 1 indicam que mutações não sinônimas podem estar tendo um efeito negativo no *fitness*, quando isto ocorre, chamamos de seleção purificadora (YANG 1998), e taxas próximas de 1 indicam neutralidade, quando não há mutações que causem danos funcionais. Assim, a comparação de substituições sinônimas e não sinônimas podem fornecer evidências de que uma proteína está evoluindo adaptativamente.

As estimativas da taxa de dN/dS calculadas por todo o comprimento de um gene raramente são maiores que 1, pois a maioria dos genes tende a ser conservado em sua totalidade, por estarem sob várias pressões seletivas. Por exemplo, as regiões essenciais de uma proteína podem estar submetidas à seleção purificadora, enquanto que uma região responsável por interagir com outra pode estar co-evoluindo de forma adaptativa (FINDLAY & SWANSON 2010). Tem sido observado que proteínas com uma taxa de dN/dS > 0,5 para o gene inteiro possuem um potencial para terem um conjunto de sítios específicos sob seleção positiva, embora seja necessário a análise de comparação com seqüências de genes de

espécies adicionais para concluir que de fato a seleção agiu neste sentido (FINDLAY & SWANSON 2010).

1.6- Genes da cascata sexual

Os genes da cascata sexual de *Drosophila melanogaster* determinam a diferenciação e o comportamento sexual, e também a fertilidade de fêmeas e machos (RIDDIFORD 1993, SÁNCHEZ *et al.* 2005, LEBO *et al.* 2009). A principal determinação do sexo é a relação do cromossomo X ao autossômico A, em fêmeas essa relação é igual a 1 e em machos igual a $\frac{1}{2}$. O primeiro gene desta cascata em *Drosophila*, *Sex lethal (Sxl)*, é responsável pela determinação do sexo controlando os genes *transformer (Tra)* e *transformer 2 (Tra2)* (SALZ & ERICKSON 2010)

Em *Drosophila melanogaster*, *Sxl* possui dois éxons iniciais que em fêmeas são eliminados por *splicing* alternativo na formação de mRNA. Um destes dois éxons que são eliminados possui um códon de parada, assim, para que ocorra a formação de uma proteína funcional é necessário que este éxon seja eliminado. Depois de formada a proteína ativa, é estabelecido um mecanismo de auto-regulação em que a própria proteína SXL regula seu *splicing* para garantir a eliminação do códon de parada. Em machos este *splicing* alternativo não ocorre, portanto não é produzida uma proteína funcional (POMIANKOWSKI & WILKINS 2004; SALZ & ERICKSON 2010).

O *Sxl* também é responsável pelo controle dos genes *transformer (Tra)* que atua em conjunto com o *transformer 2 (Tra2)*. O *Tra* assim como o *Sxl* também possui um éxon com códon de parada. Ele é transcrito em machos e fêmeas, porém somente em fêmeas o *Sxl* elimina o códon de parada, o que leva a codificação de uma proteína funcional. O *Tra* juntamente com a forma constitutiva *Tra2* são os genes que irão controlar o *splicing* sexo-específico do pré-mRNA do último gene da cascata, o *doublesex (Dsx)*. Este gene também é transcrito em ambos os sexos, mas em fêmeas o complexo *Tra* e *Tra2* dirige o *splicing* do pré-mRNA para a formação da proteína feminina (DSXf). Nos machos não há nenhuma proteína TRA ou TRA2, assim o *Dsx* segue seu *splicing* padrão dando origem a proteína masculina (DSXm) (LEBO *et al.* 2009). A proteína DSXf é que irá ativar o gene *Yolk*, cuja proteína é considerada uma vitelogenina importante para a nutrição do embrião durante seu estágio de desenvolvimento (LEBO *et al.* 2009).

Em espécies de outros gêneros como em *Ceratitis* e *Bactrocera*, embora o principal gene *Sxl* não esteja ativamente envolvido com o processo de determinação sexual, o resto da cascata sexual é bastante semelhante ao do gênero *Drosophila*, (PANE *et al.* 2002; LAGOS *et al.* 2007), em que somente as fêmeas produzem a proteína TRA na forma ativa. Em *Tephritídeos* do gênero *Ceratitis*, *Bactrocera* e *Anastrepha*, o gene *Tra* também possui um mecanismo de auto-regulação para determinação do sexo (PANE *et al.* 2002; LAGOS *et al.* 2007; RUIZ 2007; RUIZ 2010).

1.7- Proteínas coriônicas e vitelínicas

Diversas proteínas ligadas ao sexo parecem evoluir rapidamente, e isto tem grandes aplicações para estudos do processo de especiação. Esta rápida evolução é observada especialmente em proteínas que estão relacionadas à interação entre óvulo e espermatozóide (SWANSON & VACQUIER 1998). Algumas destas proteínas com alta taxa evolutiva são responsáveis pela formação da membrana do ovo. As vitelínicas, que estão localizadas no interior da membrana, como a *Vm 32E*, *Vm 34Ca*, *Vm 26 Aa* e *Vm 26Ab*; e as coriônicas, que se localizam NA PARTE EXTERNA, COMO *CPS 16*, *CPS15*, *CPS 18*, *CPS19*, *CPS 36 E CPS 38* (Parks & Spradling 1987). Estudos realizados em *Drosophila* têm indicado que as proteínas coriônicas como um grupo têm evoluído mais rapidamente do que as vitelínicas, particularmente as *Cps 16* e *Cps 15*, que são as que estão evoluindo mais rapidamente (JAGADEESHAN & SINGH 2007).

As proteínas vitelínicas formam uma membrana ao redor do ovo para protegê-lo de fatores externos. Estudos recentes em *Drosophila melanogaster* identificaram dois novos genes no genoma, além dos quatro citados acima, que são estruturalmente semelhantes aos genes vitelínicos já conhecidos. Um deles é o *Vm 26Ac* que está localizado entre o *Vm 26Aa* e *Vm 26 Ab*, e o outro é o *Vml* que também é expresso em fêmeas e não possui íntrons (ALATORTSEV 2005).

As proteínas coriônicas *Cps 36* e *Cps 38* estão localizadas no cromossomo X em *Drosophila melanogaster* e são expressas no início da formação do ovo (PARKS *et al.* 1986, PETER *et al.* 1993), enquanto em *Ceratitis capitata* as *Cps 36* e *Cps 38* são encontradas no cromossomo 5 (TOLIAS *et al.* 1990). Já as *Cps 16*, *Cps 19*, *Cps 15* e *Cps 18* são autossômicas, localizadas no cromossomo 3 em *Drosophila melanogaster* (PARKS & SPRADLING 1987), e são expressas nos estágios mais tardios da formação do ovo. As *Cps*

localizadas no cromossomo X são expressas entre os estágios 10 e 13, enquanto que as quatro *Cps* autossômicas são expressas nos estágios 13 e 14 (PARKS & SPRADLING 1987). Apesar da considerável diferença em suas seqüências, os genes *Cps 16*, *Cps 19*, *Cps 15* e *Cps 18*, são encontrados na mesma ordem tanto para *Drosophila melanogaster* quanto para *Ceratitis capitata* e possuem padrões similares de expressão (VLACHOU *et al.* 1997).

Neste trabalho buscamos a identificação de ESTs presentes em tecidos reprodutivos femininos, a fim de encontrar o maior número de proteínas vitelínicas e coriônicas investigando seu padrão evolutivo em 3 espécies do grupo *fraterculus*.

2- Objetivos

2.1- Objetivo Geral

Construir uma biblioteca de cDNAs a partir de um *pool* de tecidos reprodutivos femininos de *Anastrepha obliqua* para gerar ESTs (Expressed Sequenced Tags). Estes ESTs foram categorizados usando diferentes estratégias de classificação de proteínas de forma a isolar genes que sendo sujeitos à seleção positiva. O objetivo final foi a identificação de genes com alta taxa evolutiva neste gênero, focando em proteínas coriônicas que foram comparados com homólogos expressos em outras espécies do grupo fraterculus.

2.2- Objetivos específicos

1. Extrair RNAs mensageiros a partir de um pool de 100 órgãos reprodutivos femininos de moscas *Anastrepha obliqua* de localidades diferentes com aproximadamente 10 dias.
2. Construir uma biblioteca de cDNAs a partir dos mRNAs de fêmeas de *A. obliqua*;
3. Seqüenciar clones produzidos a partir dos cDNAs buscando produzir diversos ESTs (Expressed Sequenced Tags) de tecidos reprodutivos femininos;
4. Analisar os ESTs derivados de órgãos reprodutivos buscando identificar genes com maior variação em regiões codificadoras que possam estar sofrendo seleção positiva ou que apresentem maior variação intraspecífica.
5. Caracterizar molecularmente e fazer testes de seleção em alguns genes coriônicos e vitelínicos cujos ESTs foram encontrados na biblioteca de *A. obliqua*; dados estes que serão contrastados a de genes pertencentes a cascata sexual em *Drosophilas* (*Sex lethal* e *transformer 2*).

3- Materiais e Métodos

3.1.1- Coleta e manutenção das moscas-das-frutas

Fizemos a coleta das espécies do gênero *Anastrepha* de duas formas: pela coleta de adultos ou pela retirada de frutos da natureza em regiões de pouca interferência humana, sendo que esta segunda foi a mais comum. Posteriormente, levamos os frutos para o laboratório onde os colocamos e mantivemos em caixas acrílicas com vermiculita até a eclosão das pupas. Depois de aproximadamente 3 semanas peneiramos a vermiculita e transferimos as pupas para caixas acrílicas de 50x50x40 cm, após a eclosão, separamos as moscas por sexo. Uma semana depois, as moscas atingiram a maturidade sexual, e puderam ser então identificadas em nosso laboratório. Armazenamos algumas destas moscas em etanol 70% para posterior extração de DNA enquanto outras utilizamos em cruzamentos. Tratamos as moscas adultas sob uma dieta de 10ml de mel, 30ml de hidrolisado de proteína de milho, 6g de levedo de cerveja, 0,5g de Sustagem sabor morango, 25g de açúcar mascavo e 45g de açúcar branco.

3.1.2- Locais de coleta para construção da biblioteca

Fizemos as coletas em diferentes regiões do Brasil, para a construção da biblioteca de cDNA escolhemos de 30 a 40 indivíduos da espécie *Anastrepha obliqua* (fig 2) para 3 localidades distintas: Estrela do Oeste - SP; Bonito-PE; Bariri-SP, totalizando um *pool* de aproximadamente 100 indivíduos.



Fig.2- *Anastrepha obliqua*

3.2- Construção da biblioteca de cDNA

3.2.1- Extração de RNA total

Depois de aproximadamente 10 dias, que é o tempo da maturação sexual das moscas, retiramos o ovipositor e os órgãos reprodutivos (fig 3) de um *pool* de aproximadamente 100 fêmeas virgens da espécie *Anastrepha obliqua* e maceramos em Trizol (Invitrogen). Isolamos o RNA total através do protocolo de extração com trizol - Fenol/Tiocianato de Guanidina/Clorofórmio (Chomeczynski & Sacchi 1987). Quantificamos o RNA extraído no Biofotômetro Eppendorf pela relação de absorbância 260/280 nm. Testamos a integridade do RNA total por eletroforese em gel de agarose 1% corado com brometo de etídeo, visualizamos em luz ultravioleta, o que possibilitou a visualização das bandas do RNAs ribossômicos 18S e 28S.

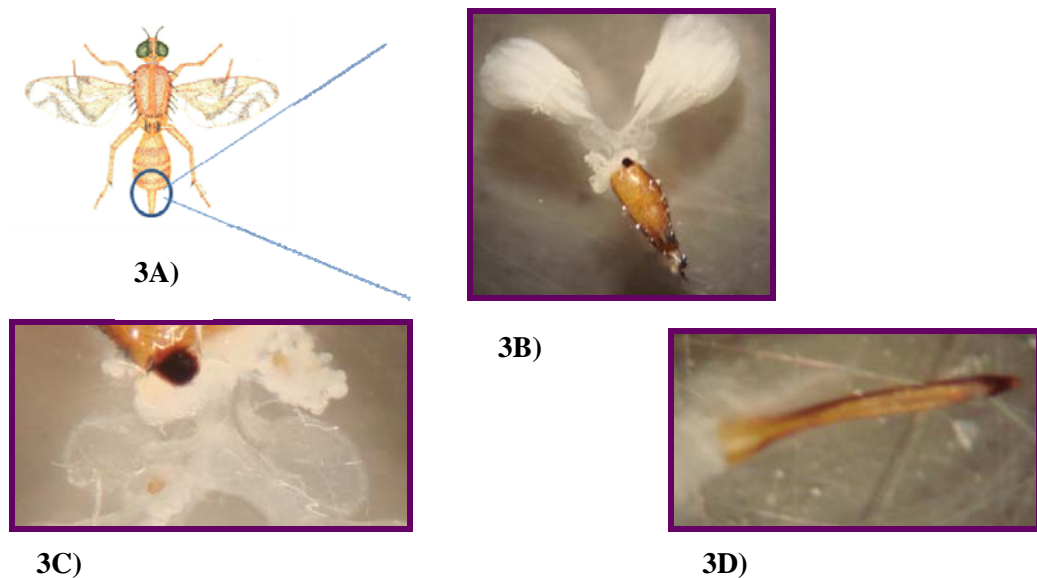


Fig.3 – Aparelho reprodutor feminino. O desenho 3A mostra a região onde se localiza o aparelho reprodutor. Na foto em 3B é possível observar os dois ovários, em 3C as glândulas acessórias e em 3D o acúleo.

3.2.2- Extração de RNA mensageiro

A partir do RNA total extraído, obtivemos o RNA mensageiro (mRNA) com o uso do kit PolyAtract mRNA Isolation Systems III (Promega). Este kit usa de oligos(dT) biotinilados ligados a microesferas magnéticas para atrair e se ligar a mRNAs que apresentem caudas poli(A); dessa forma, o uso de um suporte magnético permite o isolamento e purificação dos mRNAs que posteriormente ressuspendemos em 15 μ l de água MiliQ. Quantificamos o mRNA precipitado novamente através do Biofotômetro Eppendorf pela relação de absorvância 260/280 nm, e observamos sua integridade através da eletroforese em gel de agarose 1% corado com brometo de etídeo, visualizado em luz ultravioleta.

3.2.3- Biblioteca de cDNA

Construímos a biblioteca não normalizada de cDNA a partir de 2 μ g de mRNA, através do kit CloneMinerTM cDNA Library Construction KIT (Invitrogen), que constrói bibliotecas de cDNA sem o uso de enzimas de restrição, de acordo com as especificações do

fabricante. Na elaboração da biblioteca, ligamos adaptadores, *attB*, nas duas extremidades terminais do fragmento. A partir das fitas simples de mRNA isoladas construímos as primeiras fitas de cDNAs usando transcriptase reversa. Posteriormente adicionamos *E. coli* RNase para degradar os RNAs, e assim, foi construída uma segunda fita de cDNA usando desta vez, a primeira como molde. Fizemos o fracionamento diferencial dos cDNAs fita dupla resultantes em coluna Sephachryl S-500HR, o que permitiu a separação de frações de acordo com o tamanho médio dos mRNAs extraídos. Utilizamos as frações escolhidas para a clonagem que se baseia numa estratégia de recombinação sítio-específica para que os insertos sejam adequadamente introduzidos nos vetores. Usamos os clones para uma transformação em células competentes de *E. coli* ElectroMAX DH10B, resistente ao fago T1. Fizemos a seleção de clones em meio LB contendo Kanamicina, cuja resistência é conferida pelo plasmídeo pDONR 222, que também apresenta uma enzima DNA girase de *E. coli* que se expressa na ausência de um inserto e impede a replicação do plasmídeo, garantindo que apenas bactérias transformadas cresçam neste meio.

3.3- Qualidade da biblioteca de cDNA

Para verificar a qualidade da biblioteca construída, separamos inicialmente 24 clones, extraímos os plasmídeos bacterianos com o gene clonado através da técnica de mini preparação (miniprep). Esta técnica consiste em crescimento de colônias por 22 horas, quando então é feita uma lise alcalina e subsequente precipitação do DNA do plasmídeo com álcool isopropílico, cuja integridade testamos por eletroforese em gel de agarose 1% corado com brometo de etídeo, visualizado em luz ultravioleta.

A partir do plasmídeo extraído fizemos a digestão com a enzima de restrição BsrGI (5'...T/GTACA...3'; 3'...ACATG/T ...5'). Esta técnica trata-se de um simples corte na sequência correspondente ao DNA do plasmídeo para separação do fragmento inserido. Analisamos os resultados pelo tamanho dos fragmentos por eletroforese em gel de agarose 1% corado com brometo de etídeo, visualizado em luz ultravioleta.

Por fim, para verificar a integridade dos fragmentos inseridos no vetor, fizemos o uso da técnica de miniprep, e enviamos os produtos de outros 192 clones à empresa MacroGen, localizada em Geunchun-GU, Coréia do Sul, onde foram seqüenciados a partir do primer M13 *forward* a uma concentração de 5 µM.

3.4- Seqüenciamento da Biblioteca de cDNA

Os outros 2.112 clones foram seqüenciados pela empresa Beckman Coulter Genomics. Esta trata-se de uma empresa que presta serviços biotecnológicos localizada em Beverly-MA, EUA. Dentre os serviços oferecidos está o plaqueamento, extração de DNA e seqüenciamento de cDNAs gerados a partir de bibliotecas.

3.5- Análises das seqüências de ESTs

Agrupamos todos os ESTs seqüenciados em contigs através do software Desktop cDNA Annotation System (dCAS) (<http://exon.niaid.nih.gov>) (GUO *et al.* 2009). Este programa pode ser usado tanto para análise de uma biblioteca individual quanto para comparação de duas bibliotecas distintas. Para análise de uma biblioteca única primeiramente agrupamos as seqüências através do uso do programa NCBI Blast (<http://www.ncbi.nlm.nih.gov/>) e montamos contigs usando o *software* Cap 3 (HUANG *et al.* 1999). Analisamos então, estes contigs pelos *softwares* SignalP (EMANUELSSON *et al.* 2007) que prevê a presença e a localização de sítios de clivagem nas seqüências de aminoácidos a partir de diferentes organismos: procariotos gram-positivos e gram-negativos e eucariontes; TMHMM (KROGH *et al.* 2001) que prevê a topologia de proteínas; e pelo GenBlast, em que selecionamos todos os tipos de Blast database disponíveis: NR; GO; Mit-pla; RRNA; Conserved Domain Databases; and UniVec. Finalmente copilamos estes resultados em uma tabela de Excel.

Extraímos estas seqüências dos cromatogramas utilizando o programa Phred (EWING *et al.* 1998; EWING & GREEN 1998) que mede a qualidade das seqüências examinando os picos de cada base para atribuir uma pontuação de qualidade. Posteriormente, analisamos quanto à presença de vetores utilizando o programa Cross-Match, disponível no pacote Phred/Phrap. Consideramos válidas as seqüências com no mínimo 100 bases com Phred >20, sendo que Phred de média=20 representa 99% de chance da base estar correta. Investigamos as seqüências quanto a presença de uma fase de leitura aberta (Open Reading Frame) através do ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf.html>), que identifica todas as seqüências de aminoácidos possíveis dentro da seqüência de DNA dada.

3.5.2- Identificação das regiões sujeitas à seleção positiva

Para a identificação das regiões sujeitas à seleção positiva fizemos inicialmente a comparação das seqüências de alguns contigs aqui obtidos com as seqüências de outros *Tephritidae* além de outros insetos do gênero *Drosophila* (EADS *et al.* 2007). Calculamos a taxa de dN/dS (relação entre substituições não-sinônimas e sinônimas) pelo programa K-estimator (COMERON 1999) (<http://en.bio-soft.net/format/KEstimator.html>). Selecionamos regiões que apresentaram alta taxa evolutiva entre famílias, particularmente em sítios codificadores de aminoácido para maior investigação.

3.6- Construções de primers

Desenvolvemos primers para diversos contigs encontrados em *A. obliqua* que pareciam corresponder a genes coriônicos e vitelínicos, que poderiam estar sob seleção positiva, indicado pela taxa de dN/dS, e para outros genes de interesse. Para estes contigs encontrados na biblioteca de cDNAs de *A. obliqua*, desenvolvemos *primers* para a amplificação de *pools* de DNA para espécie *A. obliqua* e para outras 2 espécies do gênero *Anastrepha*: *A. sororcula* e *A. fraterculus*.

Assim, desenvolvemos *primers* para os genes vitelínicos que encontramos com a seqüência completa na biblioteca de cDNA: *Vm 26Aa*, *Vm 26Ab*; *Vm 26Aa'*; e para os genes coriônicos que também possuíam seqüência completa: *Chorion s38 (Cps38)*, *Chorion s16 (Cps16)*, *Chorion s18 (Cps18)*, *Chorion s19 (Cps19)*; afim de caracterizar estes genes completos para as 3 espécies de *Anastrepha*. Também construímos *primers* para os genes: *Transformer 2 (Tra2)* e *Sex lethal (Sxl)*, mas para estes não foi com o objetivo de obter o gene completo.

Construímos estes *primers* através do alinhamento das seqüências de *A. obliqua* com as do gênero *Ceratitis*; *Bactrocera* ou quando não era encontrado nenhum gênero da família *Tephritidae*, utilizamos espécies do gênero *Drosophila*. O alinhamento dos aminoácidos fizemos no programa BioEdit (HALL 1999), pelo uso do *software* Clustal W (HIGGINS *et al.* 1994), este *software* permite alinhar seqüências mesmo quando o tamanho entre elas são diferentes, relacionado a processos de inserções ou deleções. Assim pudemos identificar regiões conservadas para a construção dos primers e também identificamos regiões de possíveis íntrons. Todas as seqüências de outras espécies utilizadas foram obtidas no GenBank (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), ou no Flybase (<http://www.flybase.org>). Desenhamos os *primers* logo no começo, ou preferencialmente antes, da região codificadora

(*primer forward*) e no final da região codificadora (*primer reverse*). Depois de escolhermos uma região conservada foi necessário verificar a estabilidade das estruturas destes primers, para isto, usamos o programa *OligoAnalyzer3.0* (<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx/oligocalc.asp>).

Neste programa analisamos temperaturas de desnaturação, quantidade de bases C e G, valores de delta G que avaliam a possibilidade de formação de hairpin e formações de estruturas secundárias, tanto em ligações entre os mesmos tipos de primers (*forward-forward* ou *reverse-reverse*) como entre os diferentes (*forward-reverse*). Analisando todos estes componentes garantimos a estabilidade das estruturas, assim diminuimos a produção de artefatos durante as reações de ampliações.

Abaixo encontra-se uma tabela com as seqüências de cada um dos *primers* gerados (*forward* e *reverse*), para cada um dos genes:

Gene	Primer <i>Forward</i>	Primer <i>Reverse</i>
<i>Cps38</i>	caaggcgtgaacaagttccg	tcttagtattgtgaagctgatgg
<i>Cps16</i>	cgtaaaacaacgcattagaatat	tcagattgccacaacaccgat
<i>Cps18</i>	atttacgcaaatgaatcg	ttcgtagtagccagctgc
<i>Cps19</i>	aactacttcacttcacgtc	gacagattaatagctgctgga
<i>Vm26Aa</i>	caaaactcactatgaaattctcg	cttccgctcgattccatgta
<i>Vm26Ab</i>	tgaaaRtcttcatcactgtcac	aaatcctaatgcgatgtctatg
<i>Vm26 Aa'</i>	cagaaacttategcaaaagatg	attgccattcatccttgcc
<i>Tra2</i>	ggcattctagagaaaaagt	cttgattcgttcgattgggc
<i>Sxl</i>	aatYtaccRcgYacNatNaccga	tttgYgcatYgaaatagggatg

Tabela 1: Sequências dos conjuntos de primers *Forward* e *Reverse* usados para amplificação.

3.1.3- Locais de coleta para formação dos pools de DNA.

Para a amplificação dos genes vitelínicos: *Vm 26Aa*, *Vm 26Ab*, *Vm 26Aa'*; dos genes coriônicos: *Cp38*, *Cps19*, *Cps18*, *Cp16*; e também dos genes: *Tra2* e *Sxl*, usamos um *pool* de DNA para as 3 espécies presentes no laboratório: *A. fraterculus*; *A. sororcula* e *A. obliqua*. Formamos o *pool* de cada espécie a partir de diferentes regiões do Brasil, conforme mostra a tabela abaixo:

Pontos de coleta	Nº total de indivíduos	Nº indivíduos/local
<i>Pool fraterculus</i>		
Curitiba PR		5
Florianópolis SC	20	5
Joinville SC		5
João Pessoa PA		5
<i>Pool obliqua</i>		
Bonito PE		5
São Carlos SP	20	5
Bariri SP		5
Araguaina TO		5
<i>Pool sororcula</i>		
Guarapari ES		9
Rio das Ostras RJ	21	6
Buzios RJ		6

Tabela 2: Número de indivíduos representados no *pool* e suas respectivas localidades.

3.7- Extração de DNA

Extraímos DNA de 5 localidades diferentes para cada uma das 3 espécies *A. obliqua*, *A. fraterculus* e *A. sororcula*. Primeiramente perfuramos os corpos dos insetos e colocamos em tubos do tipo eppendorf contendo solução de proteinase K (20mg/ml); 100mM de NaCl; 100mM de Tris-Cl pH 8; 25mM de EDTA pH 8 e 0,5% de SDS, incubamos os tubos a 37°C de 2 a 3 horas. Após este período, adicionamos RNase (4mg/ml), e novamente incubamos a

solução a 37°C de 15 a 60 minutos. Posteriormente centrifugamos a solução a 300 rpm por 1 minuto para que fossem retiradas as impurezas insolúveis, transferimos o sobrenadante para outro tubo contendo uma solução de 4M de tiocianato de guanidina e 0,1M de Tris-HCl pH 7,5, homogeneizamos então, a solução em vórtex por 5 segundos e centrifugamos a 1300 rpm por 10 minutos para que as proteínas fossem precipitadas. Após esta centrifugação, retiramos o sobrenadante e colocamos em outro tubo contendo isopropanol 100%, novamente centrifugamos esta solução para a precipitação do DNA. Descartamos o sobrenadante e lavamos o precipitado formado de DNA por duas vezes com etanol 70%, invertemos os tubos gentilmente por várias vezes, em seguida centrifugamos a 1300 rpm por 5 minutos e descartamos o sobrenadante. Após a última lavagem os tubos ficaram abertos para que o etanol pudesse ser evaporado por completo e por fim ressuspendemos o precipitado em 200 µl de água ultrapura.

3.8 Amplificação por PCR e Purificação

Usamos os *primers* elaborados acima para amplificar *pools* de DNA de *A. obliqua* e também de outras espécies do grupo *fraterculus* (*A. fraterculus* e *A. sororcula*). Os reagentes que utilizamos e suas concentrações foram: tampão 10X [100mM Tris-Cl pH 9; 500mM KCl; 25mM MgCl₂; 0,1-0,2mg/ml BSA (0,1µg/µl)]; dNTPs 0,2mM; *primers forward* e *reverse* 0,2mM e *Taq* polimerase 0,5U. Analisamos a amplificação em gel de agarose 1% corado com brometo de etídeo, visualizado em luz ultravioleta.

Purificamos os produtos de PCR dos *primers* que amplificaram a região de tamanho esperado utilizando o protocolo de precipitação diferenciada com PEG 8000 (Polietilenoglicol) (LIS & SCHLEIF 1975). Neste, o produto amplificado é precipitado eliminando artefatos como dímeros de *primers* e nucleotídeos não incorporados, uma vez que o Polietilenoglicol não é eficiente para precipitar produtos com menos de 200 pares de bases. Então ressuspendemos o precipitado em água ultrapura. Analisamos o resultado também em gel de agarose 1% corado com brometo de etídeo.

Abaixo encontra-se uma tabela com o programa padrão utilizado na reação de amplificação pela técnica de PCR para todos os *primers* sintetizados:

Temperatura	Tempo	Número de ciclos
95°C	60s	1x
94°C	60s	3x
56°C	30s	
72°C	75s	
94°C	60s	30x
53°C	30s	
72°C	75s	
72°C	10.00 min	1x

Tabela3: Programa padrão utilizado na reação de amplificação para todos os *primers*, são mostrados o tempo de cada temperatura e o número de ciclos.

3.9 Quantificação dos produtos de PCR

Para a quantificação dos produtos obtidos na purificação, conforme descrito acima submetemos novamente cada amostra à eletroforese em gel de agarose 1% corado com brometo de etídeo, juntamente com um marcador de peso molecular. Fizemos a quantificação através da comparação das intensidades da coloração das bandas das amostras com as bandas do marcador.

3.10- Clonagem

Posteriormente, clonamos o produto amplificado e purificado, usando o Kit de clonagem InsTAclone (Fermentas), cujo vetor apresenta resistência a ampicilina. Fizemos a transformação de bactérias competentes da cepa DH5 α através do método de choque térmico, utilizando os produtos da ligação feita com plasmídeos PTZ57R/T.

Após a transformação, plaqueamos as colônias recombinantes presentes no meio de cultura LB seletivo (com concentração de ampicilina de 100 $\mu\text{g}/\mu\text{l}$), depois de aproximadamente 18h coletamos as colônias individuais. Colocamos cada colônia em água ultrapura que fervemos a 90°C por 5 minutos para lise das células e liberação dos plasmídeos. Amplificamos então, este material com primers M13 *forward* e *reverse* (0,2 μM) que se ligaram próximo ao sítio de clonagem do plasmídeo. Submetemos as amostras que se mostraram positivas para o inserto à precipitação diferencial com PEG (Polietileno Glicol)

8000 (LIS & SCHLEIF 1975), conforme descrito acima e ressuspendemos em 20 μ l de água ultrapura.

Abaixo encontra-se uma tabela com o programa utilizado na reação de amplificação para os *primers* M13 *forward* e *reverse*:

Temperatura	Tempo	Número de ciclos
95°C	60s	1x
94°C	60s	3x
60°C	30s	
72°C	75s	
94°C	60s	28x
57°C	30s	
72°C	75s	
72°C	10.00 min	1x

Tabela4: Programa padrão utilizado na reação de amplificação para o *primers* M13, são mostrados o tempo de cada temperatura e o número de ciclos.

3.11- Seqüenciamento dos genes *Vm26Aa*; *Vm26Aa'*; *Cps38*; *Cps16*; *Tra2* e *Sxl*

Enviamos as amostras depois de purificadas e quantificadas a 50ng/μl, para seqüenciamento pela empresa MacroGen localizada em Geunchun-GU, Coréia do Sul. Novamente utilizamos o programa ChromasPro 2.31 (<http://en.bio-soft.net/dna/chromas.html>) para visualização dos eletroferogramas. Dessa forma, pudemos obter as seqüências do *pool* dos genes amplificados *Vm26Aa*; *Vm26Aa'*; *Cps38*; *Cps16*; *Tra2* e *Sxl* para 3 espécies do grupo *fraterculus*.

3.12- Testes de detecção de seleção positiva

Para todos os genes estudados que foram amplificados e seqüenciados, fizemos um teste de seleção positiva. Para esta análise utilizamos uma seqüência de cada espécie, ou seja, uma seqüência de *A. fraterculus*, uma de *A. obliqua* e uma de *A. sororcula*, para todos os genes, exceto para o gene *Cps38* em que utilizamos 3 seqüências de cada espécie, devido a uma grande divergência intra-específica nas seqüências. Além destas, também usamos seqüências de outras espécies do gênero *Anastrepha*, *Ceratitis*, *Bactrocera*, *Musca* e *Drosophila*, quando estavam disponíveis no GenBank (<http://www.ncbi.nlm.nih.gov>). Novamente, alinhamos todas as seqüências de aminoácidos no programa BioEdit (HALL 1999), pelo uso do *software* Clustal W (HIGGINS *et al.* 1994).

Para a realização deste teste, primeiramente construímos uma árvore filogenética usando o programa PhyML 3.0 (dados não mostrados) que estima árvores por máxima

verossimilhança a partir de um modelo escolhido. Este modelo pode ser facilmente obtido através do programa ModelTest (<http://darwin.uvigo.es/software/modeltest.html>), que faz a seleção do modelo de substituição de nucleotídeos que melhor se ajusta aos dados testando 56 modelos diferentes.

A partir da árvore obtida pelo PhyML 3.0, utilizamos o programa PAML que faz uso do CODEML (YANG 2007) para testar seleção positiva. Fizemos dois tipos de testes: *Relaxed Branch-site test* e *Strict Branch-site test* (YANG & NIELSEN 2002; ZHANG *et al.* 2005). Em ambos os testes escolhemos um ramo chamado de *foreground*, foi neste ramo em que foi testada a ocorrência de seleção positiva, os demais ramos de gêneros de outras espécies foi chamado de *background*. Neste trabalho fizemos duas análises para cada um dos dois testes. Uma considerando o ramo que isola o gênero *Anastrepha* como *foreground* e outra considerando o ramo que isola a família *Tephritidae* inteira como *foreground*. Para que fossem realizados ambos os testes primeiramente foi preciso testar o modelo M0, este modelo fixou um único ω (razão entre substituições não sinônimas por substituições sinônimas) para todos os ramos e fixou um único ω para todos os códons. Foi necessário primeiramente testar este modelo para que construíssemos uma nova árvore filogenética com a distância de mudança nucleotídica por códon, foi a partir dela que testamos os outros modelos.

No primeiro teste, *Relaxed Branch-site test*, contrastamos o modelo M1a (nulo) com o modelo MA. O modelo M1a assume que a taxa evolutiva é a mesma para todos os ramos e sítios, os valores de ω só podem variar entre 0 e 1. Já no modelo MA separamos os ramos em *foreground* e *background*, e o ramo *foreground* pode possuir sítios com $\omega > 1$. Fizemos o contraste destes dois modelos através de um teste chamado de *log-likelihood ratio test* (LRT) (YANG *et al.* 2005; ZHANG *et al.* 2005).

No segundo teste, *Strict Branch-site test*, contrastamos o modelo MA nulo com o modelo MA. No modelo MA nulo, também separamos os ramos em *foreground* e *background*, porém assumimos que todos os sítios com valores de $\omega > 1$ estivessem fixados a 1. Fizemos o contraste destes dois modelos também através do *log-likelihood ratio test* (LRT) (ZHANG *et al.* 2005). No teste de seleção positiva calculado pelo modelo MA, também pudemos realizar outra análise: a *Bayes Empirical Bayes* (BEB) (YANG *et al.* 2005) que é baseada nos mesmos parâmetros do modelo MA. Através dela, quando LRT foi significativa, pudemos identificar quais sítios obtiveram probabilidade maior que 50%, 95% e 99%, de estarem sob seleção positiva.

3.13- Predição da estrutura proteica.

Fizemos as predições das estruturas proteicas para os genes que mostram estar sob seleção positiva, pelo programa MEMSAT 3 disponível no site da UCL (University College London) (<http://bioinf4.cs.ucl.ac.uk:3000/psipred>). Este programa é de fácil acesso, bastando colocar a seqüência de aminoácidos de uma determinada proteína que ele prevê a estrutura secundária e a topologia da membrana baseando-se em alguns modelos (JONES *et al* 1994; JONES 2007; NUGENT & JONES 2009). Fizemos os testes através de um conjunto de tabelas estatísticas de LRT. Estas tabelas estatísticas mostram tendências definidas de determinados aminoácidos de estarem localizados no interior, no meio ou fora da membrana celular.

3- Resultados e Discussão

No total sequenciamos a partir da região 5', 2304 clones obtidos a partir dos órgãos reprodutivos de fêmeas de moscas de *Anastrepha obliqua*. A tabela de análise destes ESTs do programa dCAS produziu 310 contigs e 506 singlets (Anexo) que classificamos em 8 classes diferentes de proteínas de acordo com o GO e em 13 classes de acordo com o KOG.

4.1- A biblioteca de cDNA.

4.1.1- Integridade do RNA total.

Para se iniciar a construção da biblioteca de cDNAs primeiramente foi necessário extrair o RNA total e sua integridade analisada em gel de agarose 1% (fig.1). A qualidade do RNA total foi comprovada pelas duas bandas visualizadas acima na fig.1, que indicam os RNAs ribossômicos (rRNA) íntegros. Abaixo encontram-se os RNAs transportadores e como os RNAs mensageiros são de diferentes tamanhos eles correspondem ao arraste na vertical.

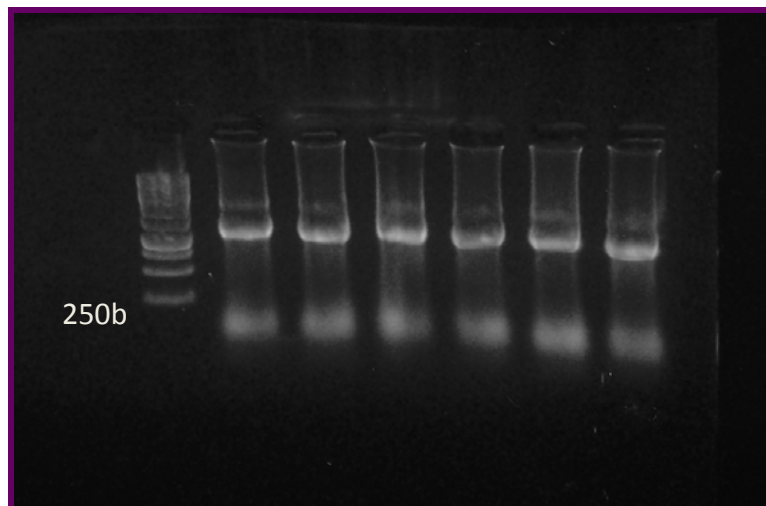


Fig. 4- RNA total mostrado em foto de gel de agarose 1% corado com brometo de etídeo. Na primeira coluna da esquerda para direita está o marcador. Cada banda corresponde a 250b. Nas demais colunas estão seis amostras. Acima se encontram duas bandas de rRNA 28s e 18s respectivamente.

4.1.2- Integridade do RNA mensageiro.

Após a extração e quantificação do RNA total, extraímos o mRNA usando o kit PolyAtract mRNA Isolation Systems III (Promega), e novamente avaliamos sua integridade em gel de agarose 1% (fig.2). As bandas começam próximo de 3 kb e terminam em 250b, o que demonstra sua integridade. As bandas mais intensas correspondem aos ESTs mais expressos.

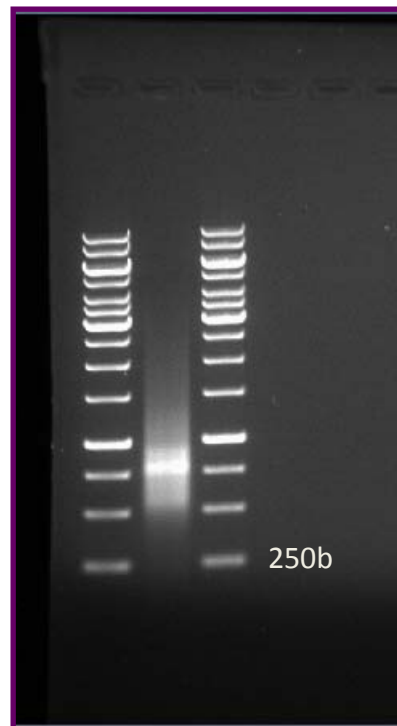


Fig. 5- RNA mensageiro mostrado em foto de gel de agarose 1% corado com brometo de etídeo. Na primeira e última coluna está o marcador. Cada banda corresponde a 250b. Na coluna do meio está o mRNA.

4.1.3- Tamanho dos ESTs

Depois de verificada a integridade do mRNA, o quantificamos e utilizamos para a construção da biblioteca de cDNA de acordo com instruções do fabricante do kit CloneMiner (Invitrogen). Para calcular a titulação da biblioteca foi necessário fazer várias diluições das bactérias transformadas, fizemos diluições de 10^{-1} , 10^{-2} , 10^{-3} , e 10^{-4} . Após os cálculos, vimos que a titulação da biblioteca foi de $1,5 \times 10^8$.

Para saber a média do tamanho dos ESTs que inserimos nos vetores, fizemos um tratamento com a enzima de restrição BsrGI que corta o inserto do vetor. Ao todo digerimos

24 clones obtidos através de transformação bacteriana, sendo que todos os 24 clones continham o inserto, a média do tamanho do fragmento inserido foi de 1,1 kb (fig.6).

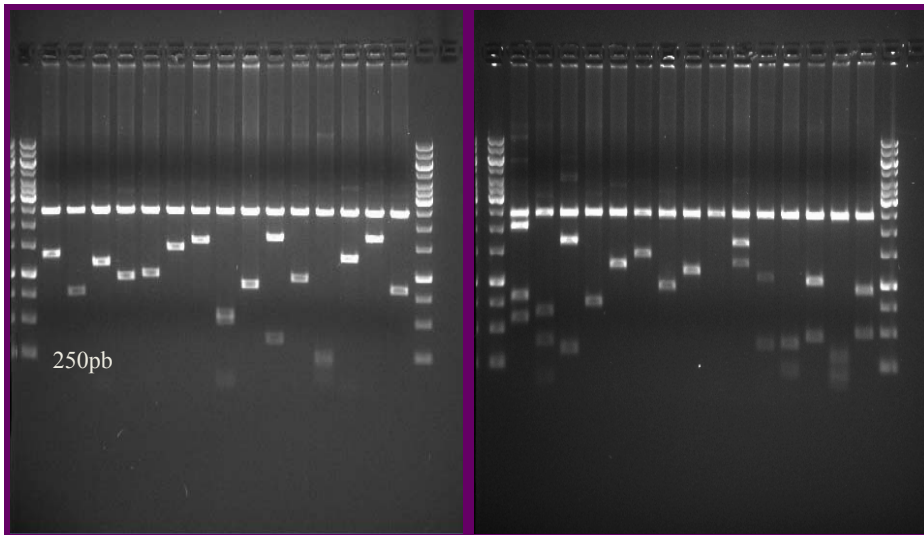


Fig. 6- Tratamento com enzima de restrição BsrGI mostrado em gel de agarose 1% corado com brometo de etídeo. Foram fragmentados 24 clones com média de 1,1kb. Nas colunas das extremidades está o marcador, cada banda corresponde a 250pb. A banda de 2,5 kb corresponde ao vetor.

No total sequenciamos 2308 clones ao acaso, sendo que destes 2101 produziram seqüências com qualidade e os utilizamos nas análises, isto é, seqüências com no mínimo 100pb e com Phred >20. Analizamos as seqüências obtidas e agrupamos em contigs através do *Software Desktop cDNA Annotation System (dCAS)* (GUO *et al.* 2009). O programa Phred acoplado ao dCAS (Ewing & Green 1998; Ewing *et al* 1998) atribuiu uma pontuação de qualidade das seqüências, sendo que eliminamos as de baixa qualidade, e o vetor retiramos pelo Cross-Match. O *software* dCAS utilizou então os *softwares* CAP3 (HUANG *et al.* 1999) e BLAST para produzir contigs que compilamos em uma tabela Excel. Esta análise produziu 309 contigs e 506 singlets, que são os ESTs restantes que não foram redundantes. Deste conjunto de contigs e singlets, 25 apresentaram uma taxa média de identidade com *Tephritidae* de 84,04%, e 485 contigs/singlets apresentaram taxa média de identidade com *Drosophila* igual a 71,38%. Os outros 305 contigs/singlets não foram alinhados com nenhuma seqüência do GenBank ou foram alinhados com seqüências de famílias muito distantes da família *Tephritidae*.

Apesar deste conjunto de contigs e singlets somarem um total de 815 sequências, o que parece ser um número alto para uma biblioteca não normalizada, estes dados são semelhantes aos de Kelleher *et al* (2007) que encontrou um número um pouco menor. A partir de tecidos reprodutivos de fêmeas virgens e grávidas e do mesmo número de ESTs aos obtidos neste trabalho, foi obtido um total de 649 contigs/singlets. Em outro trabalho, a partir de tecidos reprodutivos de fêmeas grávidas de *Glossina morsitans morsitans*, de um total de 3048 ESTs utilizados, foram formados 1391 contigs/singlets (ATTARDO *et al.* 2010).

O comprimento médio das seqüências foi de 644pb, isto indica que geramos aproximadamente 1,4 megabases nesta análise desta biblioteca. Estes valores foram semelhantes aos encontrados para uma biblioteca de embrião, em média 700 pb, e para biblioteca de cabeça, cerca de 723pb, em *Ceratitis capitata* (GOMULSKI *et al.* 2008). Entretanto na biblioteca de tecidos reprodutivos de *Glossina morsitans morsitans*, esta média foi um pouco maior, o comprimento médio das seqüências foi de 1125pb (ATTARDO *et al.* 2010). Apenas 68 contigs chegaram à polyA, entretanto isto não significa que obtivemos apenas 68 contigs completos, pois muitos dos contigs analisados apesar de não apresentarem a cauda polyA estavam inteiros.

Um dos fatores que determinam a qualidade de uma biblioteca é o número de seqüências completas obtidas para que se possa saber também a seqüência completa correspondente aos aminoácidos da proteína codificada. Assim a qualidade depende da integridade do mRNA e conseqüentemente da integridade do cDNA gerado. Abaixo encontra-se um gráfico (fig.4) mostrando que 420 contigs/singlets alinham com seqüências do GenBank logo nos primeiros 10 nucleotídeos, isto sugere que possuem o códon de início.

N de Contigs/Singlets

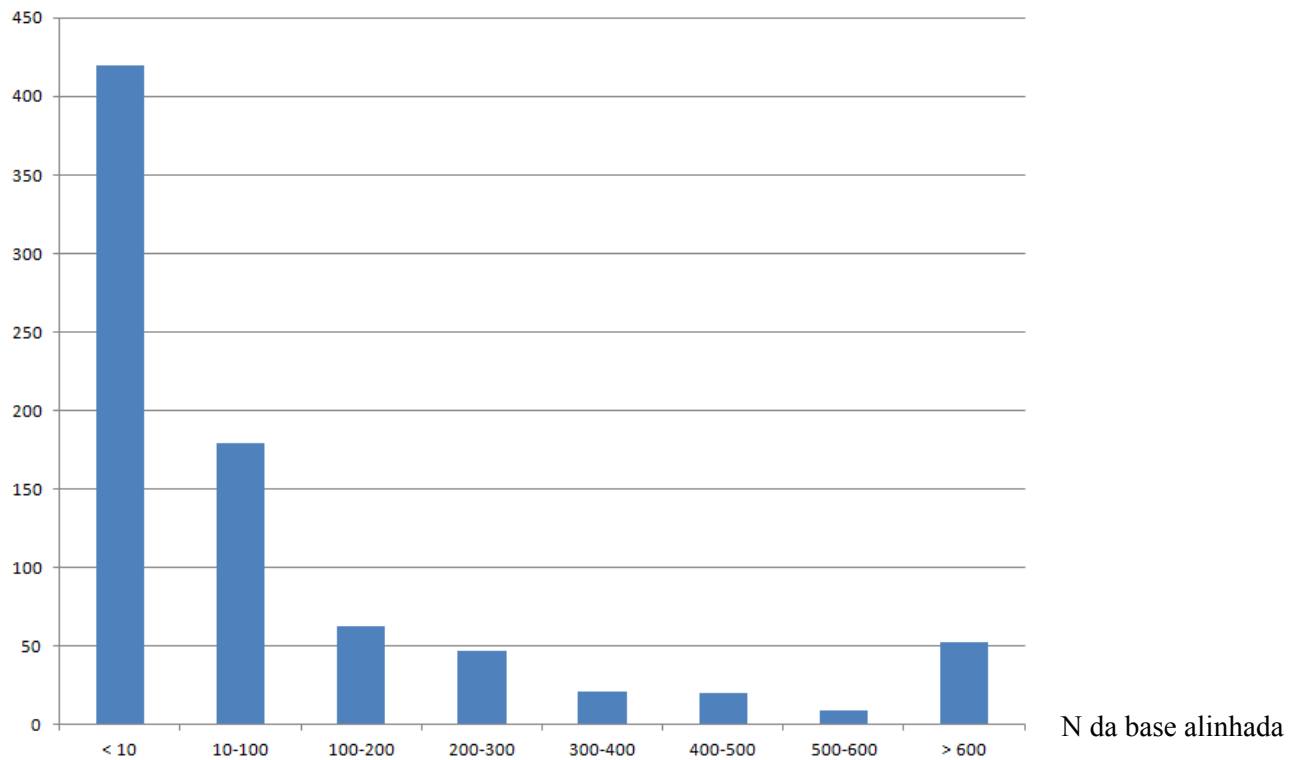


Fig.7- Gráfico evidenciando a integridade do cDNA. No eixo X está o número da base em que a seqüência da biblioteca foi alinhada com a seqüência do GenBank. No eixo Y está o número de contigs/singlets. Foram consideradas as seqüências com pelo menos 100 nucleotídeos.

O maior número de ESTs em um único contig foi o gene *yolk* com 113 cópias e o segundo contig mais expresso foi o *16s ribosomal* com 61 cópias, mas apenas 30 contigs tiveram número maior ou igual a 10 ESTs, estes resultados são semelhantes aos encontrados na biblioteca de cDNA de embrião e cabeça de *Ceratitits capitata* (GOMULSKI *et al.* 2008). Na biblioteca de embrião, de 11512 seqüências, somente 54 contigs tinham mais de 10 ESTs, e na biblioteca de cabeça apenas 33 contigs num total de 12518 seqüências (GOMULSKI *et al.* 2008), entretanto ambas as bibliotecas foram normalizadas.

O gene mais expresso, *yolk*, está localizado no cromossomo X em *Drosophila melanogaster*, é ativado pelo último gene da cascata sexual o *doublesex*, através de sua proteína ativa DSXf (ALVARES *et al.* 2009). O *yolk* é um gene que codifica uma proteína do ovo conhecida como uma vitelogenina, cuja função está relacionada ao desenvolvimento embrionário, na formação do produto nutritivo do embrião (LEBO *et al.* 2009). Sua proteína é expressa em dois locais distintos, nos tecidos gordurosos do corpo adulto de fêmeas, onde posteriormente é levada para os ovários, e principalmente nas células do folículo ovariano (BEBAS *et al.* 2008).

Neste trabalho encontramos 49 tipos de proteínas ribossomais, que normalmente são altamente conservadas (fig. 8). Também encontramos algumas proteases importantes para estudos evolutivos, ao todo foram 11 tipos de serinas, 8 cisteínas e 5 catepsinas. Proteases digestivas encontradas no aparelho reprodutor feminino de *Drosophila arizonae* estão associadas a recentes duplicações gênicas, com rápida aquisição de nova função. Foi verificado que estas duplicações das espécies do grupo *Drosophila repleta* estão sob forte pressão de seleção positiva (KELLEHER *et al.* 2007). Estudos filogenéticos com outro grupo de proteases, feitos com sequências disponíveis no GenBank, evidenciam que a família das serinas proteases evoluíram por diversos eventos de duplicação gênica como um resultado da seleção positiva (ANDRÉS *et al.* 2006). Também há vários estudos com as cisteínas e catepsinas envolvidas na reprodução, que mostram eventos de duplicações e seleção positiva (CIVETTA 2003, CHUONG *et al.* 2010).

O gráfico abaixo (Fig.8) apresenta os 20 contigs mais expressos, sua porcentagem de identidade com a espécie mais próxima que possui sequência no GenBank e sua taxa de substituição não-sinônima por sinônima (dN/dS), calculada em relação à *Drosophila melanogaster*, que escolhemos por ser um organismo modelo para o qual haviam sequências disponíveis no GenBank para comparações. Fizemos o cálculo na região completa dos ESTs e valores maiores podem ser um indicativo de que esteja ocorrendo seleção positiva. Alguns como o *envelope protein* e *Chorion s19* apresentaram altos valores de dN/dS, enquanto a maioria apresentou valores bem baixos, como já era esperado para genes conservados, como as diversas proteínas ribossômicas encontradas (ex: *60S Ribosomal protein L44*, *40S Ribosomal protein S4*, *16S ribosomal*).

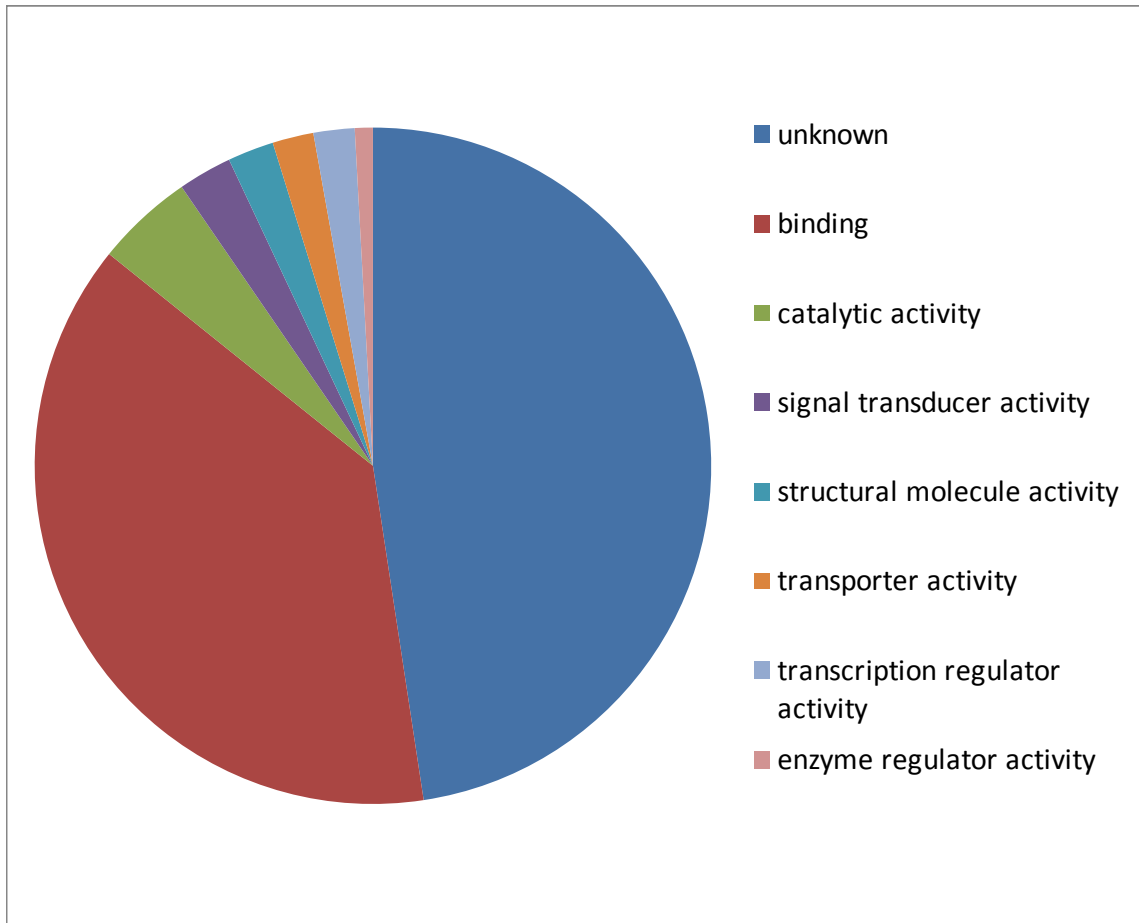


Fig 9- Classificação das proteínas de acordo com sua função de acordo com o GO.

Na classificação feita de acordo com o KOG formamos 13 classes diferentes de proteínas, sendo estas um pouco mais específicas e completas do que as formadas pelo GO. Como trata-se de uma classificação mais completa, obtivemos um número de classes protéicas muito grande, muito superior as 13 que foram mostradas. Assim, consideramos classes que tivessem pelo menos 20 contigs dentro dela, e as demais classes foram agrupadas como “Outras”, desta forma, como já era o esperado, a maior classe de proteínas foi esta que agrupa várias classes com 25%.

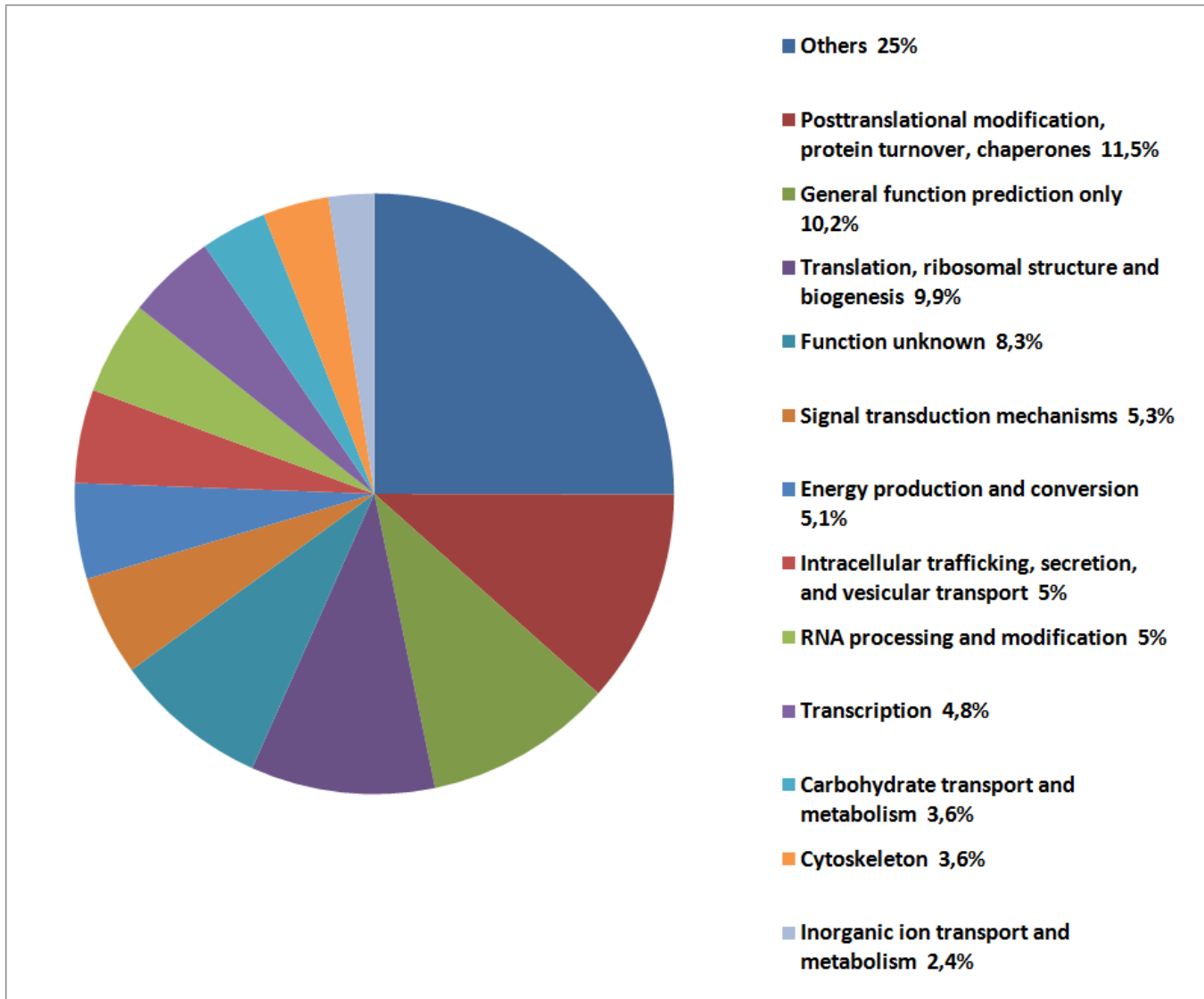


Fig 10- Classificação das proteínas de acordo com sua função de acordo com o KOG.

4.3- Análise de seleção para os genes coriônicos e da cascata sexual

Duas das proteínas coriônicas estão entre os 20 genes mais expressos (Fig. 5). Além disso, observamos que a *Cps 19* possui um valor médio de dN/dS igual a 0,8 para toda a extensão do gene, que foi o mais alto de todas as proteínas presentes na Fig. 8. Por ser superior a 0,5, indica um potencial para este gene ter um conjunto de sítios específicos sob seleção positiva, uma vez que este valor é medido para toda a extensão do gene, indicando que mesmo que existam regiões conservadas, no geral há grande diferenciação entre as seqüências (FINDLAY & SWANSON 2010; KELLEHER *et al.* 2007). Isto nos mostrou que as proteínas coriônicas de fato poderiam estar evoluindo mais rapidamente sob pressão de seleção positiva. Por outro lado, a *Cps 38* apresentou um dos valores mais baixos das

proteínas avaliadas na Fig. 5, indicando que podem haver dinâmicas diferentes entre proteínas coriônicas distintas. Por este motivo, neste trabalho fizemos *primers* para buscar amplificar todas as proteínas coriônicas e vitelínicas completas encontradas na biblioteca de cDNAs e dois outros genes pertencentes à cascata sexual para uma análise mais aprofundada da sua variabilidade intra e interespecífica no grupo *fraterculus*. Apesar de diversas tentativas, apenas foram amplificados eficientemente duas proteínas coriônicas: *Cps38* e *Cps16*; e outras duas vitelínicas: *Vm26Aa* e *Vm26Aa'*; sendo que estas últimas são provavelmente duplicações recentes da *Vm26Aa* encontrada em outros *Dipteras*. Além destes, os genes *Tra2* e *Sxl* também foram amplificados. Os genes coriônicos e vitelínicos foram seqüenciados por completo, já os da cascata sexual apenas uma região foi seqüenciada devido ao seu grande comprimento, e uma vez que o intuito neste caso era de fornecer bases de comparação com dados gerados das proteínas vitelínicas e coriônicas e não sua descrição completa.

Após a amplificação, clonagem e seqüenciamento do *pool* de cada uma das espécies aqui investigadas, alinhamos as seqüências dos genes de *Anastrepha* entre si e com outros *Diptera* filogeneticamente próximos encontrados no GenBank. Uma primeira análise de seleção foi utilizada pelo contraste da razão dN/dS (YANG & NIELSEN 2002; ZHANG *et al.* 2005) nas seqüências como um todo e também nos ramos que separam *Drosophilidae* de *Tephritidae*. Como esta análise é inadequada para a investigação de ramos evolutivos curtos, em função da confusão entre polimorfismos ainda segregando e mutações já fixadas (KRYAZHIMSKIY & PLOTKIN 2008), apenas uma seqüência de cada espécie foi aleatoriamente amostrada e considerada para estes testes.

Descrevemos a seguir as árvores filogenéticas inferidas por máxima verossimilhança (Figs. 11 a 16) para os genes coriônicos, vitelínicos e os pertencentes à cascata sexual, criadas pelo modelo M0, que fixa um único ω para todos os ramos e códons. Em função da similaridade dos dois genes vitelínicos, também construímos uma árvore em que os mesmos são considerados conjuntamente, alinhando ambos com *Vm26Aa* de espécies do gênero *Drosophila*. A taxa de ω fixo para cada um destes genes foi de: *Vm26Aa*= 0,19; *Vm26Aa'*= 0,16; *Vm26Aa* e *Vm26Aa'*= 0,23; *Cps16*= 0,3; *Cps38*= 0,16; *Sxl* = 0,04; *Tra2*= 0,1

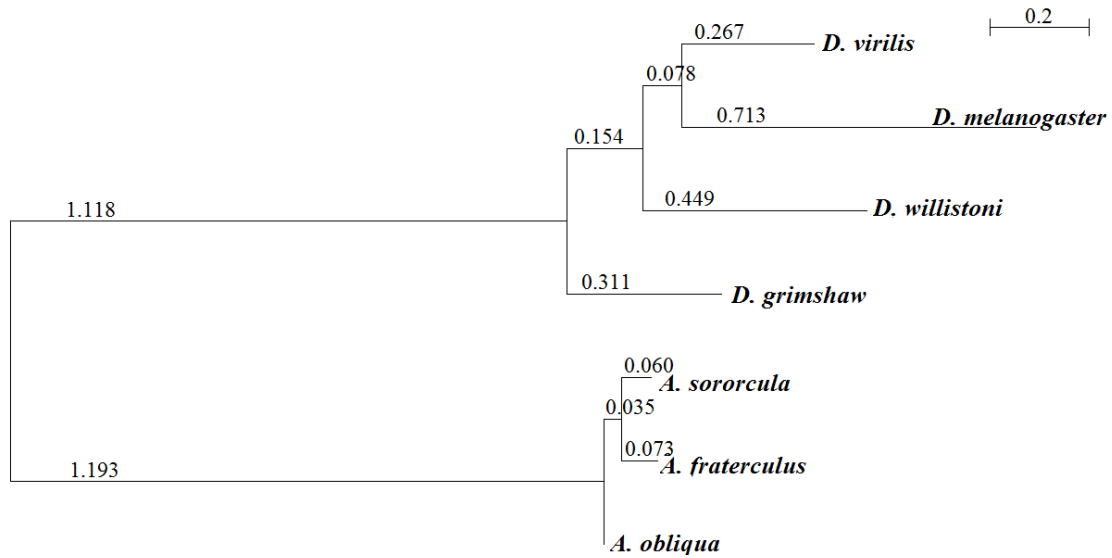


Fig.11 – Árvore filogenética de Máxima Verossimilhança estimada pelo modelo M0 para a *Vm26Aa*. Os tamanhos dos ramos foram estimados por substituições nucleotídicas por códon.

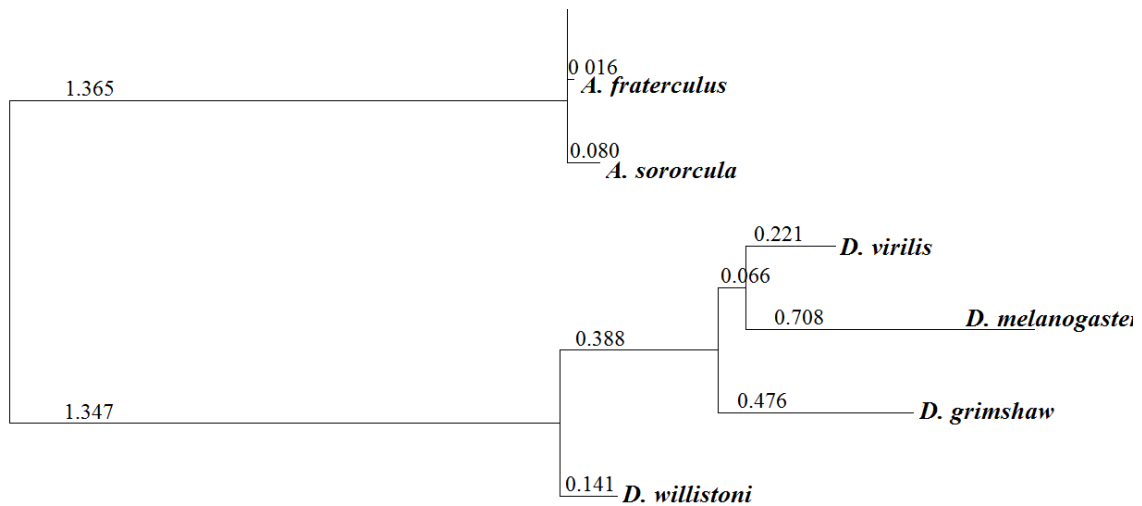


Fig.12 – Árvore filogenética de Máxima Verossimilhança estimada pelo modelo M0 para a *Vm26Aa'*. Os tamanhos dos ramos foram estimados por substituições nucleotídicas por códon.

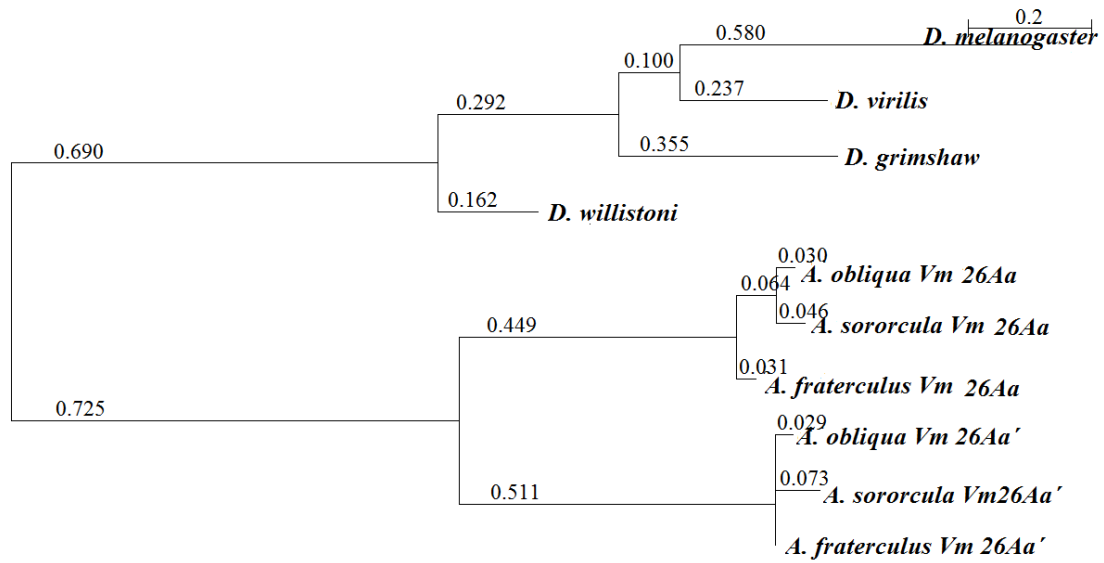


Fig.13 – Árvore filogenética de Máxima Verossimilhança estimada pelo modelo M0 para a análise conjunta de *Vm26Aa* e *Vm26Aa'*. Os tamanhos dos ramos foram estimados por substituições nucleotídicas por códon.

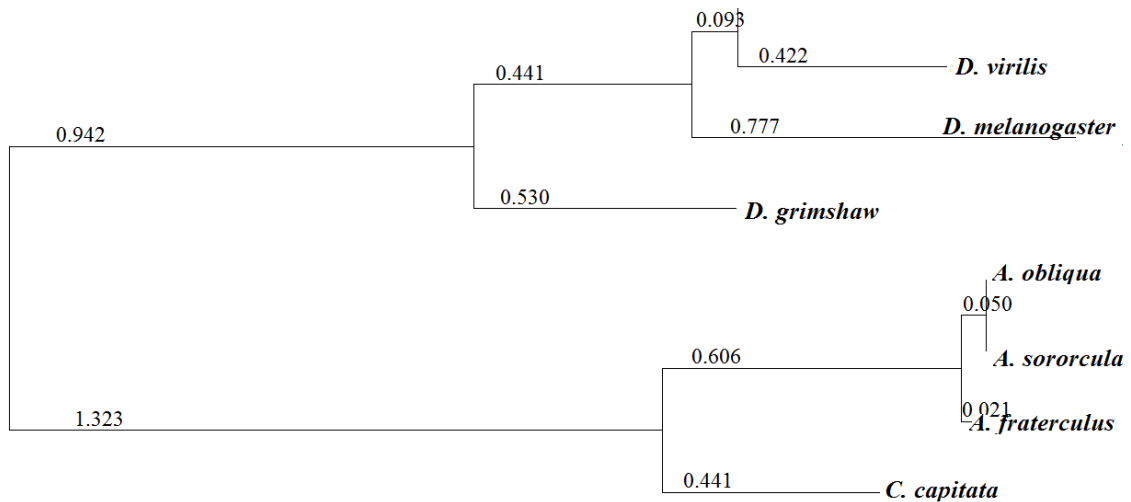


Fig.14 – Árvore filogenética de Máxima Verossimilhança estimada pelo modelo M0 para *Cps 16*. Os tamanhos dos ramos foram estimados por substituições nucleotídicas por códon.

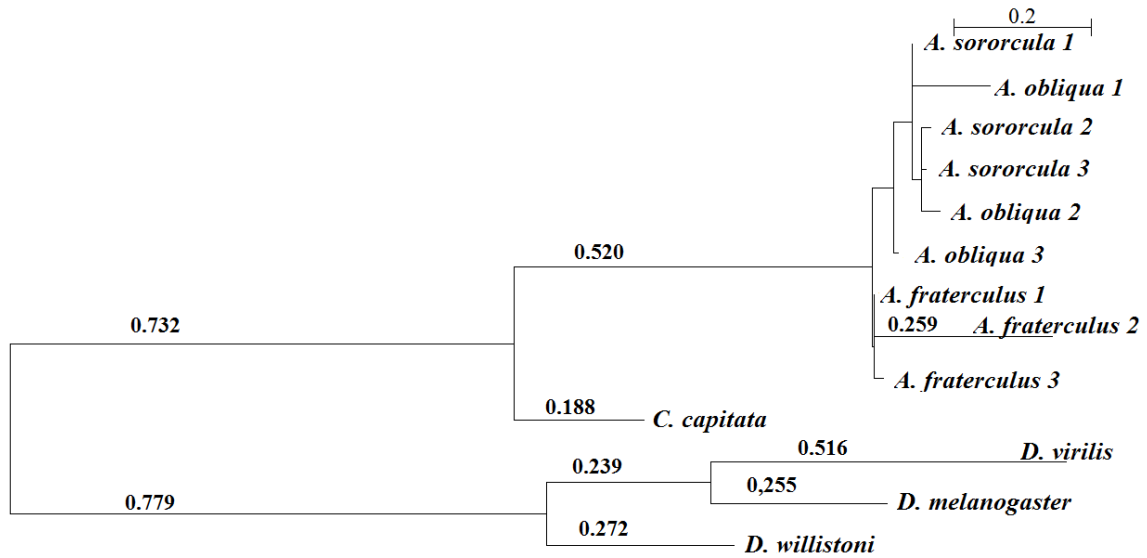


Fig.15 – Árvore filogenética de Máxima Verossimilhança, estimada pelo modelo M0 para *Cps 38*. Os tamanhos dos ramos foram estimados por substituições nucleotídicas por códon.

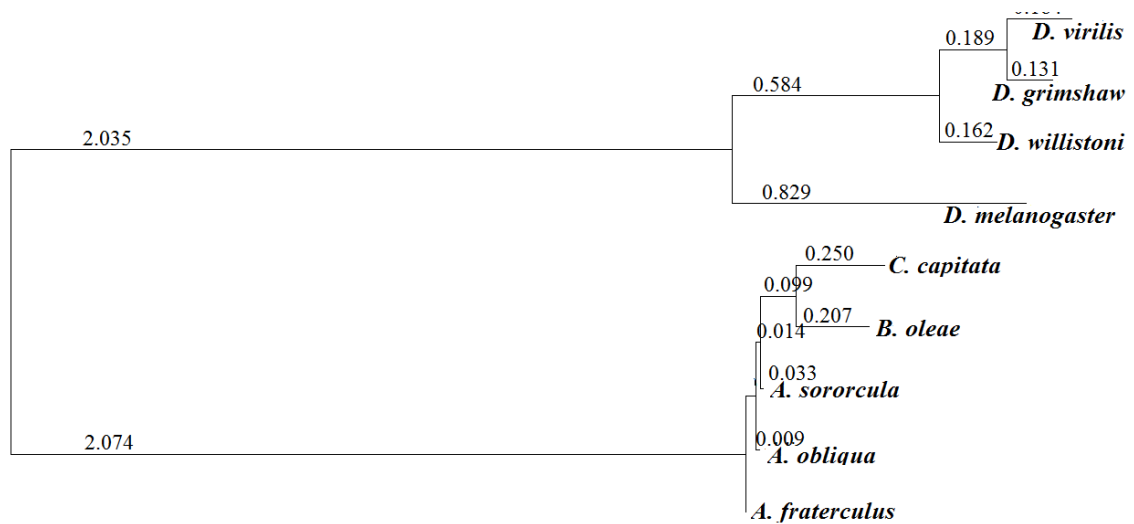


Fig.16 – Árvore filogenética de Máxima Verossimilhança, estimada pelo modelo M0 para *Sxl*. Os tamanhos dos ramos foram estimados por substituições nucleotídicas por códon.

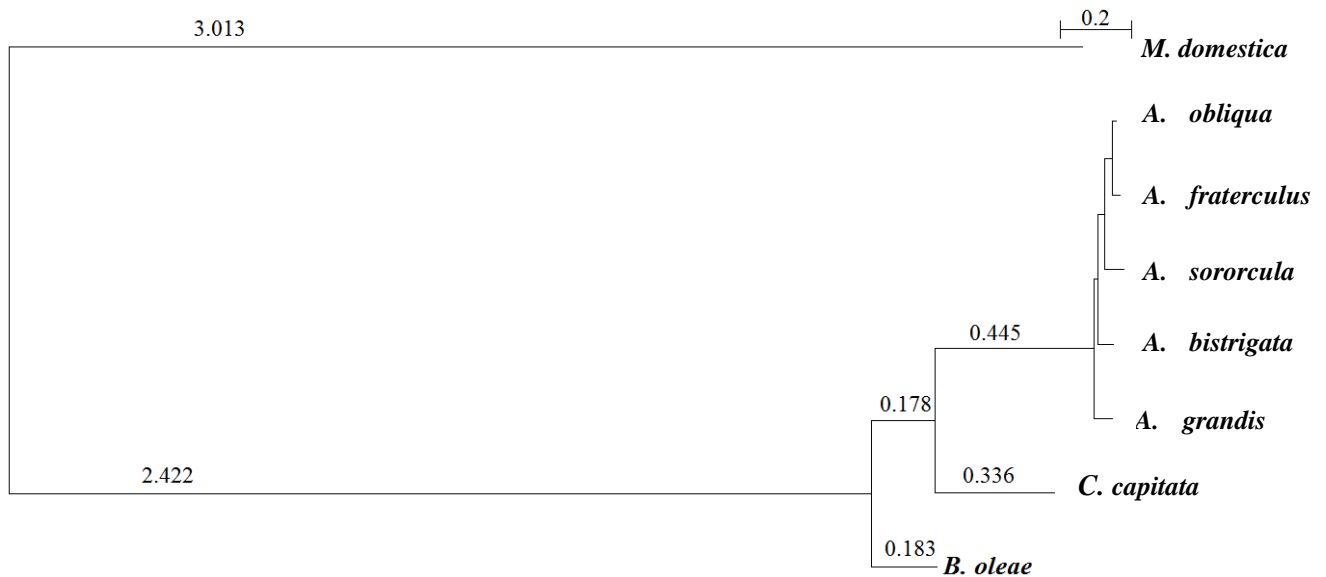


Fig.17 – Árvore filogenética de Máxima Verossimilhança, estimada pelo modelo M0 para *Tra2*. Os tamanhos dos ramos foram estimados por substituições nucleotídicas por códon.

Para cada um dos genes coriônicos, vitelínicos e pertencentes à cascata sexual fizemos o teste de seleção positiva, através dos diferentes modelos evolutivos presentes no PAML. Os modelos testados foram o M1a, o MA restrito (modelo nulo) com $\omega_2=1$, fixado e o modelo MA alternativo. Quando possível, fizemos dois tipos de análises: uma considerando como ramo *foreground* a família *Tephritidae* e outra considerando o gênero *Anastrepha* como *foreground*. Para os genes vitelínicos *Vm26Aa* e *Vm26Aa'* também fizemos uma análise em conjunto em que os ramos considerados como *foreground* foram o de *Anastrepha* que incluiu ambos os genes, e posteriormente cada gene separadamente. As tabelas em anexo resumem os contrastes relevantes inferidos dos parâmetros estimados para cada um dos genes para cada modelo testado.

Após esta primeira análise, fizemos uma comparação entre os modelos nulos (M1a e MA restrito) e o modelo MA alternativo. As *Cps16* e *Cps38* e o *Sxl* mostram estar sob seleção positiva para ambos os testes (com e sem *gaps*) e para ambos os ramos. No caso do *Sxl* não foi separado um ramo para as *Anastrepha*, por estas estarem filogeneticamente muito próximas aos demais *Tephritidae*, como é possível observar na Fig. 16.

Genes	MA x M1a (LRT)	MA restrito x MA (LRT)	Sítios selecionados positivamente
Tra2 (<i>Tephritidae</i>)	21,38881 ***	1,809956**	-
Tra2 (<i>Anastrepha</i>)	0*	0**	-
Sex lethal (<i>Tephritidae</i>)	47,25089*	10,82756**	12 ^{••} ; 74 ^{••} ; 84 ^{••} ; 85 [•] ; 86 [•] ; 90 ^{••} ; 91 [•] ; 92 [•] ; 101 ^{••} ; 102 ^{••} ; 109 ^{••} ; 122 ^{••}
Vm26Aa (<i>Anastrepha</i>)	3,283314*	0,536582****	-
Vm26Aa' (<i>Tephritidae</i>)	8,519784**	4,433312****	-
Vm26Aa e Vm26Aa' (<i>Anastrepha</i>)	0*	0***	-
Vm26Aa e Vm26Aa' (alternativo para Vm26Aa de <i>Anastrepha</i>)	3,276462	0,318158****	-
Vm26Aa e Vm26Aa' (alternativo para Vm26Aa' de <i>Anastrepha</i>)	6,330526**	2,576988****	77 [•]
Cps16 (<i>Tephritidae</i>)	21,0861**	3,4111****	61 [•] ; 65 [•] ; 89 [•] ; 96 [•] ; 99 [•] ; 109 [•] ; 114 [•] ; 143 [•]
Cps16 (<i>Anastrepha</i>)	8,204998**	2,72656****	132 [•] ;
Cps38 (<i>Tephritidae</i>)	45,06276****	16,69391****	9 [•] ; 13 [•] ; 14 ^{••} ; 19 [•] ; 22 [•] ; 76 [•] ; 87 ^{••} ; 90 [•] ; 91 ^{••}
Cps38 (<i>Anastrepha</i>)	9,73977**	7,884736****	187 [•]

Tabela 5– Comparação por LRT considerando os *gaps*, dos modelos evolutivos M1a x MA e MA restrito x MA, e os sítios selecionados positivamente feitos por análise de *Bayes Empirical Bayes* para linhagem *foreground* ($w > 1$). * $P < 1$, *x $P < 0,05$ x** $P < 0,01$ *x** $P < 0,001$. • se a probabilidade for posterior a 95%, e •• se a probabilidade for posterior a 99%.

Genes	MA x M1a	MA restrito x MA	Sítios selecionados positivamente
Tra2 (<i>Tephritidae</i>)	21,38881 ^{***} *	1,809956 ^{***}	-
Tra2 (<i>Anastrepha</i>)	0 [*]	0 ^{***}	-
Sex lethal (<i>Tephritidae</i>)	40,4745 ^{****}	8,532582 ^{***}	12 ^{••} ; 74 ^{••} ; 84 ^{••} ; 88 ^{••} ; 90 [•] ; 91 ^{••} ; 92 ^{••} ; 99 [•] ; 109 ^{••}
Vm26Aa (<i>Tephritidae</i>)	3,839252 [*]	0,525142 ^{***}	-
Vm26Aa' (<i>Anastrepha</i>)	2,884894 [*]	1,438876 ^{***}	-
Vm26Aa e Vm26Aa' (<i>Anastrepha</i>)	0 [*]	0 ^{***}	-
Vm26Aa e Vm26Aa' (alternativo para Vm26Aa de <i>Anastrepha</i>)	3,997508 [*]	0,574742 ^{***}	-
Vm26Aa e Vm26Aa' (alternativo para Vm26Aa' de <i>Anastrepha</i>)	5,84035 ^{**}	1,287932 ^{***}	-
Cps16 (<i>Tephritidae</i>)	19,76072 ^{***} *	3,091196 ^{***}	33 [•] ; 37 [•] ; 60 [•] ; 67 [•] ; 70 [•] ; 80 [•] ; 85 [•] ; 114 [•]
Cps16 (<i>Anastrepha</i>)	8,269518 ^{***}	4,089256 ^{***}	-
Cps38 (<i>Tephritidae</i>)	31,58422 ^{***} *	8,669284 ^{***}	8 [•] ; 12 [•] ; 13 ^{••} ; 18 [•] ; 21 [•] ; 48 [•] ; 57 ^{••} ; 59 [•]
Cps38 (<i>Anastrepha</i>)	5,439934 ^{**}	4,759078 ^{***}	150 [•]

Tabela 6– Comparação por LRT não considerando os *gaps*, dos modelos evolutivos M1a x MA e MA restrito x MA, e os sítios selecionados positivamente feitos por análise de *Bayes Empirical Bayes* para linhagem *foreground* ($w > 1$). * $P < 1$, ** $P < 0,1$, *** $P < 0,01$, **** $P < 0,001$. • se a probabilidade for posterior a 95%, e •• se a probabilidade for posterior a 99%.

A região estudada do gene *Sxl* apresentou seleção positiva pela análise de LRT dos modelos MA restrito e MA alternativo e pela análise de BEB, que indica quais sítios possuem maior probabilidade de estarem sob seleção positiva. Como o gênero *Anastrepha* apresenta muito pouca diferenciação em relação aos demais *Teprítídeos*, evidenciada pelos ramos curtos entre espécies desta família para este gene, apenas a família *Tephritidae* foi considerada como *foreground*. Foram 12 os sítios que apresentaram estar sob seleção positiva (Tab.5); por apresentarem probabilidade superior a 95%, embora esta análise tenha sido feita considerando os *gaps* dos alinhamentos. Quando estes não foram considerados, 9 sítios estavam sob seleção positiva (Tab.6). Não existem ainda estudos que relatem seleção positiva no gene *Sxl*, mas alguns genes que pertencem à cascata sexual já foram estudados como o *Dsx*, *Tra* e *Tra2* (ALVAREZ *et al.* 2009; RUIZ *et al.* 2007; RUIZ *et al.* 2010) e eles têm se mostrado conservados em *Anastrepha*. Entretanto apesar de não sabermos exatamente qual função *Sxl* desempenha em *Teprítídeos* e outros *Dipteros*, é conhecido que ele não pertence à cascata sexual (SIERA & CLINE 2008), de forma que não deixa de ser interessante que tenhamos encontrado evidência de seleção positiva.

O outro gene da cascata estudado foi o *Tra2*, cujo resultado da análise foi igual independentemente de considerarmos ou não os *gaps*. Quando os tefritídeos foram considerados como *foreground*, não detectamos seleção positiva através da comparação dos modelos MA restrito e MA alternativo. Também não encontramos forte seleção purificadora ao compararmos os modelos M1a com o modelo MA alternativo, entretanto quando o gênero *Anastrepha* foi considerado como *foreground*, a comparação destes dois modelos indicou que esta região gênica estudada parece estar sob forte pressão de seleção purificadora em *Anastrepha*.

A análise de LRT dos modelos MA restrito e MA alternativo revelou que os genes coriônicos estão sob seleção positiva tanto para a análise em que consideramos os *gaps*, quanto quando não os consideramos. Ao fazermos o branch-site test, detectamos que a seleção positiva foi encontrada no ramo *foreground* tanto para os *Teprítídeos* quanto para o ramo das *Anastrepha*. Para a *Cps 16* encontramos 8 sítios que possuem probabilidade superior de 95% de estarem sob seleção positiva considerando *Teprítídeos* como *foreground* enquanto quando consideramos *Anastrepha* apenas um sítio alcançou probabilidade maior que 95%, isto considerando-se *gaps*. Quando os *gaps* não foram considerados, o número de sítios sob seleção com probabilidade superior a 95 % continuou o mesmo para os *Teprítídeos*, mas para

Anastrepha não houve nenhum sítio sob seleção positiva. Estes resultados indicam que independente dos *gaps* os *Tefritídeos* de maneira geral apresentam mais sinais de sítios sob seleção positiva em relação à *Drosophila*, do que quando consideramos apenas o gênero *Anastrepha*. Isso é de se esperar considerando que o ramo que separa *Tephritidae* de *Drosophilidae* é bem maior do que o ramo que separa *Anastrepha* de outros *Tephritídeos*, e como relatado anteriormente, o poder estatístico deste teste de seleção é diretamente proporcional ao tempo evolutivo e número de mudanças evolutivas entre os contrastes realizados (KRYAZHIMSKIY & PLOTKIN 2008) e pode inclusive produzir resultados espúrios caso a análise considere polimorfismos com sendo substituições fixadas entre linhagens.

Na análise de BEB para a *Cps 38* os sítios sob seleção positiva quando os *gaps* não foram considerados foram 9, um a mais que a *Cps 16*, sendo que alguns deles possuíam probabilidade de seleção maior que 95% e outros maior que 99%, considerando tefritídeos como *foreground*. Quando as *Anastrephas* foram consideradas como *foreground* houve apenas um sítio, o 187, com alta probabilidade de seleção. Quando não consideramos os *gaps* nas análises, os resultados também foram semelhantes, 8 sítios estavam sob seleção para os *Tefritídeos* e apenas o sítio 150 para *Anastrepha*.

Encontramos resultados distintos do que vimos com os genes coriônicos quando analisamos os genes vitelínicos. O gene *Vm 26Aa* não apresentou sinal de seleção positiva para análise de LRT quando analisado separadamente e quando analisado em conjunto com o *Vm26 Aa'* para ambos os testes (com e sem *gaps*). Entretanto quando analisamos o modelo M1a com o modelo MA alternativo notamos que este gene também não está sob forte pressão de seleção purificadora.

Na atual investigação desta biblioteca não encontramos nenhum dos dois genes vitelínicos, *Vm26Ac* e *Vml*, descritos para *Drosophila* (ALATORTSEV 2005). Entretanto, encontramos um novo gene, o *Vm 26Aa'* que possivelmente foi gerado por uma duplicação gênica do *Vm 26Aa* na família *Tephritidae* ou por uma duplicação exclusiva do gênero *Anastrepha*. A distância genética entre as duplicações *Vm26 Aa* faz sugerir uma diferenciação equivalente à encontrada entre *D. willistoni* e outras espécies de *Drosophila* (Fig. 13), que possivelmente ocorreu há cerca de 36,3 milhões de anos (RUSSO *et al.* 1995). Assumindo taxas evolutivas equivalentes para estas linhagens, isso indica que a duplicação de *Vm26Aa* precedeu a diferenciação dos *Tefritídeos* aqui estudados, contudo a seleção positiva encontrada em um destes genes duplicados pode ter modificado as taxas evolutivas após a

duplicação, de forma que tal questão poderá apenas ser esclarecida pelo estudo de outras espécies de *Teprítídeos*.

Uma das importantes formas de se obter novos genes com novas funções é através da duplicação gênica. Diversos estudos recentes têm indicado o papel importante da duplicação na produção de variação genética (CIVETTA 2003, ANDRÉS *et al.* 2006, KELLEHER *et al.* 2007), papel este ainda reforçado em proteínas envolvidas na reprodução, como alguns estudos que evidenciaram algumas proteases que adquiriram nova função (KELLEHER *et al.* 2007) e o próprio gene da especiação *Odysseus* que também é uma duplicação de um retroposon que sofreu rápida diferenciação em espécies de *Drosophila* do grupo *melanogaster* estando envolvido com o isolamento reprodutivo de *D. simulans* e *D. mauritiana* (TING *et al.* 2004). Entretanto para que uma duplicação tenha uma contribuição adaptativa é necessário que ocorram mutações adaptativas. Porém, mutações podem aperfeiçoar uma função e comprometer outras, este conflito adaptativo normalmente é resolvido por uma duplicação gênica seguida de subfunções e seleção positiva (HITTINGER & CARROLL 2007). Tal padrão é facilitado em eventos de duplicação uma vez que estas novas mutações que eventualmente ocorrem não comprometem o funcionamento normal da proteína original, de forma que tais mutações têm chance de não serem deletérias, e mesmo efetivamente neutras, quando não vantajosas.

Detectamos seleção positiva no gene *Vm26Aa'* quando fizemos sua análise separadamente com *gaps*, quando desconsideramos os *gaps* e quando fizemos a análise em conjunto este sinal de seleção foi perdido. Caso este sinal de seleção se confirme através de análises mais detalhadas, diferente do que ocorreu com o *Vm 26Aa*, é possível que a proteína produzida por esta suposta duplicação esteja exercendo uma nova função celular. A diferença encontrada entre análises com e sem *gaps* reflete limitações de testes de seleção positiva como estes que não são eficientes em lidar com *indels* e sua evolução.

Os testes feitos para investigação de seleção no gene *Tra2* além de não apresentar seleção positiva, revelou estar possivelmente sob seleção purificadora. Ruiz *et al.* (2010) em seu trabalho também com espécies do gênero *Anastrepha*, também encontrou seleção purificadora quando analisou o *Tra2* por todo o seu comprimento. Este tipo de seleção pode estar ocorrendo, não apenas por *Tra2* ser um gene conservado da cascata sexual em insetos em geral, como *Drosophilidae* e *Tephritidae*, mas também porque a região aqui amplificada faz parte de algumas regiões que se mostraram altamente conservadas, possivelmente mantendo algumas características estruturais da proteína que são essenciais na sua função.

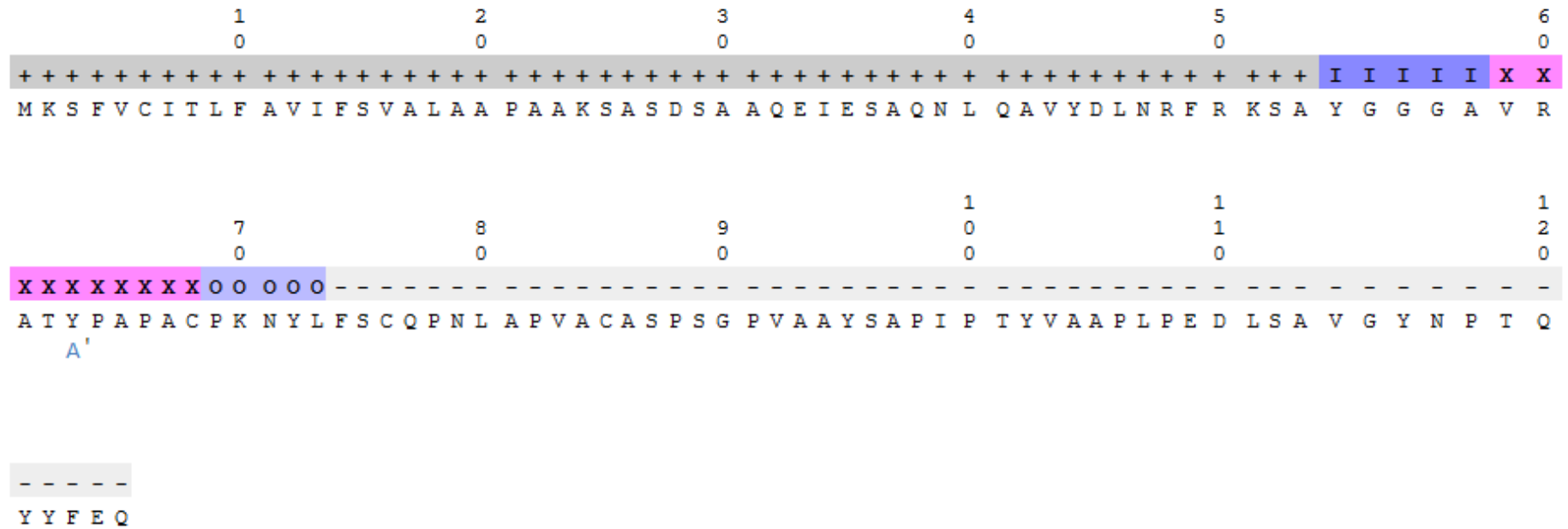
Ambos os genes coriônicos demonstraram estar evoluindo de forma semelhante, com sítios sob forte seleção positiva principalmente quando consideramos *Tephritidae* como *foreground*. Este sinal de seleção positiva em genes coriônicos já foi encontrado antes em *Drosophila* (JAGADEESHAN & SINGH 2007) com análises semelhantes à deste trabalho. A seleção positiva foi encontrada principalmente nas *Cps 15* e *Cps 16*, mas também foi encontrada nas *Cps 38* e *Cps 18*, o único gene coriônico que não apresentou seleção positiva foi a *Cps36*. Ainda neste mesmo trabalho, foi visto que os genes vitelínicos não apresentaram seleção positiva, assim como foi visto neste estudo. Entretanto quando consideramos os *gaps* nas análises, encontramos seleção positiva para o *Vm 26Aa*'.

As proteínas vitelínicas localizam-se na camada vitelínica que é a região mais interna do ovo, enquanto que as coriônicas estão localizadas em camadas superiores, no córion (PARKS & SPRADLING 1987), de forma que é possível que a seleção positiva encontrada nas proteínas coriônicas reflita o fato de estarem mais em contato com o meio externo do ovo. Ou seja, as proteínas coriônicas podem estar relacionadas com a interação espécie-específica de óvulo e espermatozóide durante a fertilização, uma vez que a proteína lisina presente no esperma precisa criar um buraco no envoltório do ovo (SWANSON & VACQUIER 1998), e/ou esta divergência indica que a evolução dos ovos foi influenciada principalmente por fatores ecológicos, através da diversificação do comportamento da oviposição (JAGADEESHAN & SINGH 2007).

Outro fator que explicaria estes sinais de seleção positiva nas duas proteínas coriônicas estudadas, é que tanto a membrana vitelínica como a coriônica são estabilizadas por *cross-linking* covalente, o que torna tanto as proteínas vitelínicas como as coriônicas insolúveis. As proteínas da membrana vitelínica *Vm26Aa* e *Vm26Ab* tornam-se completamente insolúveis no estágio 12, enquanto que as proteínas da membrana coriônica se mantêm solúveis até o estágio 14 (PASCUCCI *et al*, 1996), o que permitiria um tempo maior para ocorrer mais alterações até elas se tornarem estáveis.

4.4- Estrutura secundária das proteínas que apresentaram seleção positiva.

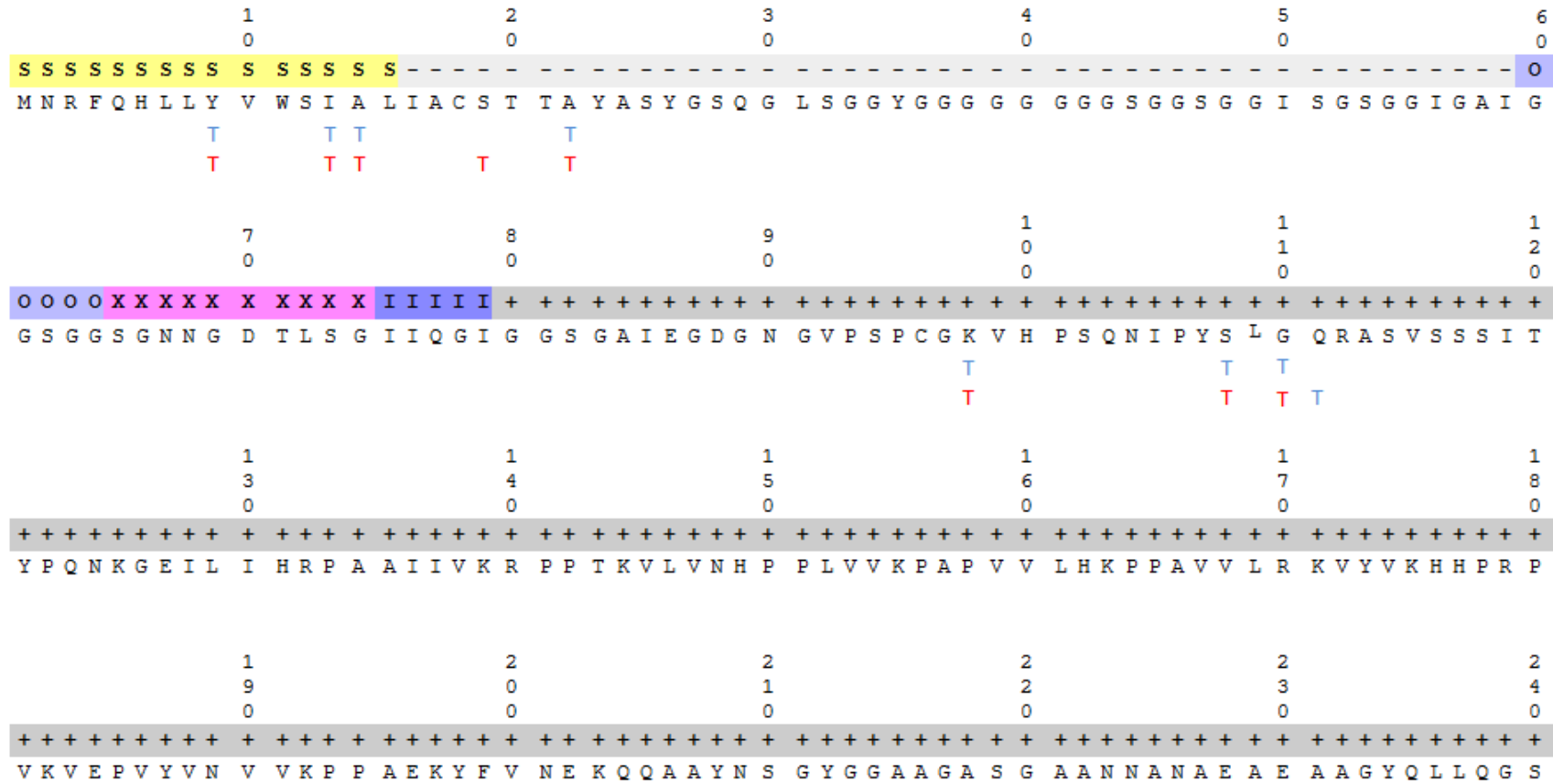
No intuito de entender melhor os padrões de seleção encontrados para as proteínas estudadas, utilizamos de *softwares* para inferir suas estruturas secundárias ressaltando nas figuras abaixo os sítios que apresentaram seleção positiva pela análise de BEB.



T= foreground= Tephritidae- considerando os gaps
 T= foreground= Tephritidae- não considerando os gaps
 A= foreground= Anastrepha- considerando os gaps
 A= foreground= Anastrepha- não considerando os gaps

Possible N-terminal signal peptide S
 Outside helix cap O
 Inside helix cap I
 Outside loop -
 Inside loop Central +
 Transmembrane helix segment X

Fig.19- Estrutura secundária inferida para o gene *Vm26Aa*'.



A
B

Fig.21- Estrutura secundária inferida para o gene *Chorion s38*.

Como demonstrado acima, os sítios que apresentaram seleção positiva foram praticamente os mesmos independente da análise, considerando ou não os *gaps*. Porém os sítios que apresentaram seleção quando *Tephritidae* foi considerado como *foreground* não foram os mesmos quando o gênero *Anastrepha* foi considerado. Assim, para análise de BEB, praticamente não houve diferenciação entre os testes realizados, mas para o cálculo de LRT a diferença entre os valores foi bastante significativa, sendo que quando os *gaps* foram considerados havia uma tendência de um valor mais alto de LRT.

A *Cps38* apresentou sítios com sinais de seleção positiva principalmente na região de *Possible N-terminal signal peptide* e na região de *Inside loop Central*. A *Cps36* quase não possui região de *Outside loop*, sendo que os sítios sob seleção foram na região de *Inside loop Central*. Já o *Sxl* a região que mais apresentou sinais de seleção positiva foi numa repetição de glicinas em *Outside loop*.

A região transmembrana, de *Outside helix cap* e de *Inside helix cap* manteve-se conservada em ambos os genes coriônicos, assim como no *Sxl*, o que sugere uma maior conservação na região intermembrana. Interessantemente, o único sítio com sinal de seleção positiva em *Vm26Aa'* foi justamente na região de transmembrana, isto talvez possa ser uma mutação que esteja levando a uma importante alteração de função.

Não sabemos ao certo onde as proteínas coriônicas aqui estudadas são encontradas na membrana coriônica, mas estudos em *Drosophila* revelam que a *Cps15* e a *Cps16* estão localizadas na região de micrópila, onde entra o espermatozóide, e ambas as proteínas coriônicas interagem com as do esperma (JAGADEESHAN & SINGH 2007). Interessantemente as regiões que mais apresentaram seleção positiva foram as que os sítios estavam presentes do lado externo da membrana. Isto da idéia de que possivelmente esta seleção positiva seja devido a interações óvulo e espermatozóide durante a fertilização, e/ ou devido a fatores ecológicos, por estas proteínas estarem provavelmente expostas a um ambiente externo (SWANSON & VACQUIER 1998, JAGADEESHAN & SINGH 2007).

4.5- Avaliação da utilidade dos genes estudados como marcadores de identificação de diferentes espécies no grupo *fraterculus*.

Uma vez que identificamos sinais de seleção positiva em alguns dos genes aqui estudados, investigamos qual a sua capacidade em diferenciar espécimes pertencentes a diferentes espécies no grupo *fraterculus*. Para isto, sequenciamos 10 cópias de cada gene

derivados das espécies *A. fraterculus*, *A. obliqua* e *A. sororcula*. Estas sequências foram utilizadas para a criação de árvores de haplótipos que permitem-nos avaliar não apenas o grau de variabilidade encontrado nestas regiões, como também o quanto desta variabilidade é intra ou interespecífica. A segregação da variabilidade para cada um dos genes está representada nas figuras abaixo (Figs. 24-29), criadas no programa TCS v. 1.21 (POSADA *et al.* 2001).

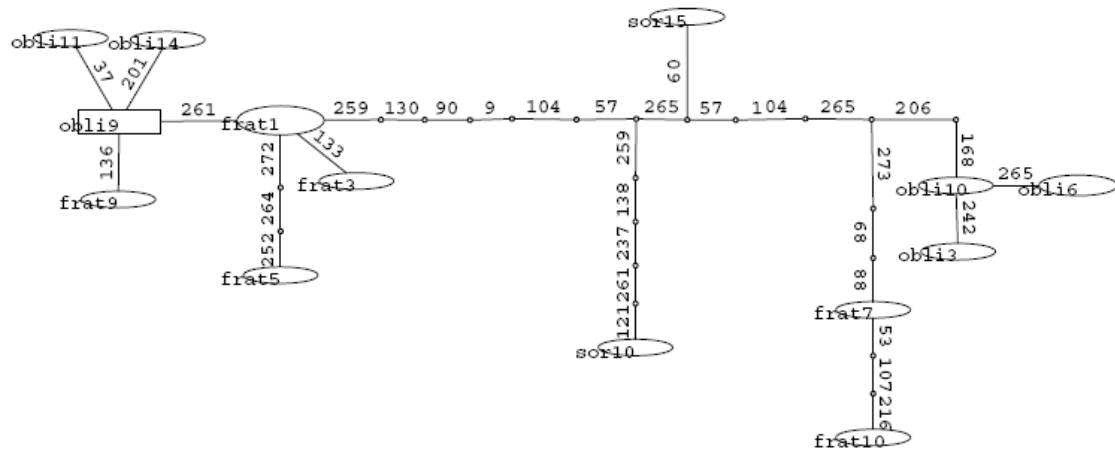


Fig.24 – Rede de haplótipos de sequências de *Anastrepha fraterculus* (fratN), *A. obliqua* (obliN) e *A. sororcula* (sorN) para o gene *Vm 26Aa* estimada por parcimônia estatística. Número em cada ramo indica sítio modificado entre haplótipos representados por elipses.

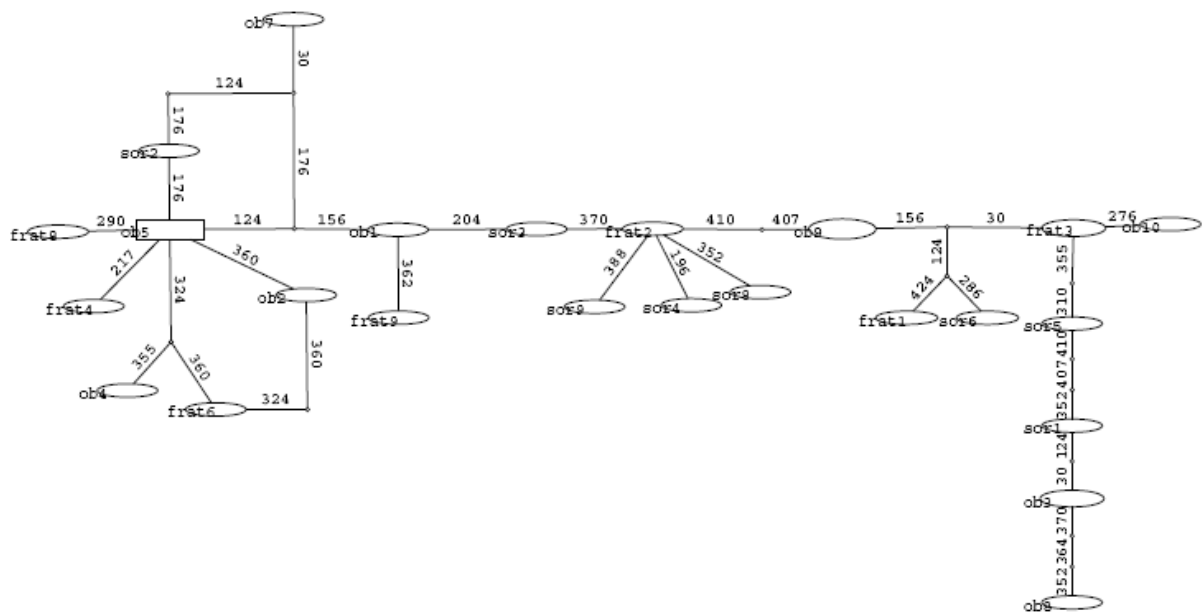


Fig.28 – Rede de haplótipos de seqüências de *Anastrepha fraterculus* (fratN), *A. obliqua* (obliN) e *A. sororcula* (sorN) para o gene *Sxl* estimada por parcimônia estatística. Número em cada ramo indica sítio modificado entre haplótipos representados por elipses.

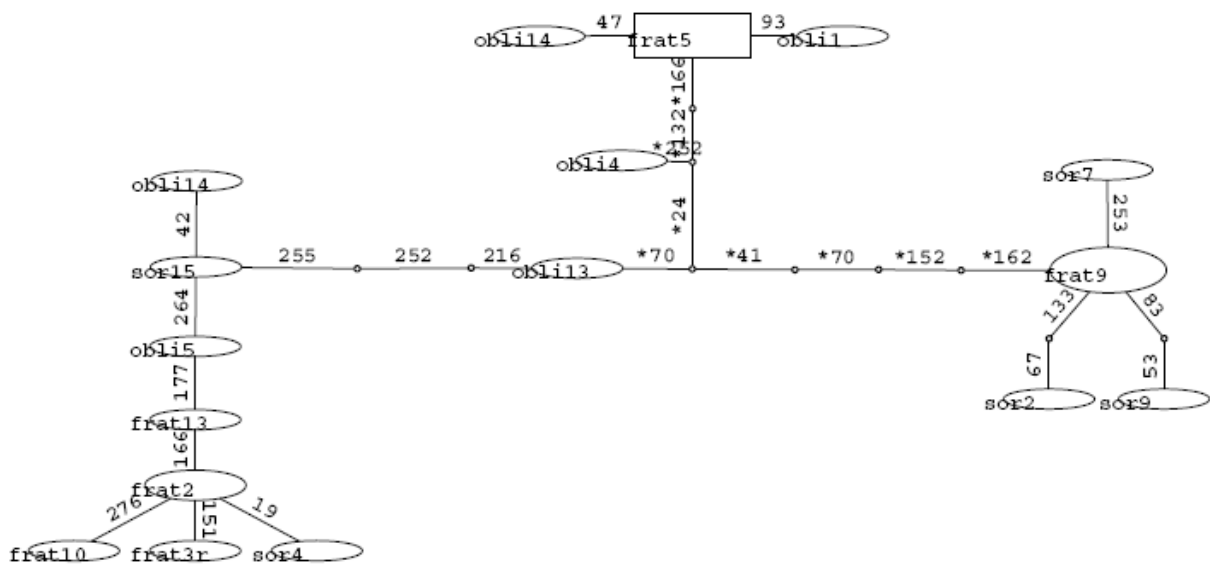


Fig.29 – Rede de haplótipos de seqüências de *Anastrepha fraterculus* (fratN), *A. obliqua* (obliN) e *A. sororcula* (sorN) para o gene *Tra2* estimada por parcimônia estatística. Número em cada ramo indica sítio modificado entre haplótipos representados por elipses.

A análise das sequências obtidas para cada espécie em geral indicou existência de significativa variação intraespecífica nos genes estudados. Encontramos uma associação em alguns casos entre o nível de variação intra e interespecífica e a detecção de seleção positiva em estudos de dN/dS. Embora a dinâmica populacional de polimorfismos intraspecíficos seja distinta da dinâmica de mutações fixadas, ambas respondem às mesmas pressões de seleção. As principais diferenças na resposta à seleção dizem respeito à extensão da duração da seleção e ao efeito estocástico do processo evolutivo (KRYAZHIMSKIY & PLOTKIN 2008). Em função de tal estocasticidade, apenas um tempo evolutivo grande permitiria graus de liberdade grandes o suficiente para que pudessemos detectar evidências de seleção positiva. Por este motivo, estes testes devem ser normalmente realizados apenas entre espécies que se diferenciaram há bastante tempo, como os contrastes realizados no item 4.3. Contudo, a detecção de seleção positiva em dados populacionais é mais complexa e apesar de existirem alguns testes sugeridos não existe ainda uma metodologia comprovada que permita diferenciar processos populacionais dos processos seletivos subjacentes (SOBRINHO & DE BRITO *subm.*). Como os dados intraespecíficos aqui obtidos para cada gene não nos permitem avaliar parâmetros filogeográficos, por ser uma amostra agrupada de diversas regiões do Brasil, utilizamos os dados de variação populacional, apenas para a identificação eventual dos processos de segregação de polimorfismo ancestral.

Entretanto, estas redes de haplótipos foram feitas apenas com as regiões codificadoras dos genes aqui estudados, de forma que foram eliminadas as regiões 5' e os íntrons eventualmente presentes. Esta estratégia se justifica uma vez que estamos interessados no processo evolutivo do gene, mas um contraste com a variação neutra poderia ser interessante para revelar melhor os padrões evolutivos nesta região.

Uma análise global das sequências obtidas não revelou marcadores específicos para diferentes espécies de *Anastrepha* embora tenha revelado padrões interessantes de variação que devem ser destacados em particular. 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). Isto estabelece um contraste entre mutações internas e de ponta utilizadas na genética de populações que indica algo sobre a dinâmica evolutiva.

Embora o gene *Vm 26Aa* não tenha revelado sinais de seleção positiva, observamos que uma parte significativa da variação presente internamente na árvore de haplótipos promove alterações de aminoácidos. Na verdade, 5 das 14 mutações que promovem alteração de aminoácidos são internas e quatro destas estão localizadas no ramo principal que distingue basicamente três grupos de haplótipos distintos. Estes três grupos não nos permitem diferenciar as três espécies, embora haja certo agrupamento de forma que haplótipos próximos tendam a ser da mesma espécie.

A análise de *Vm26Aa'* embora revele topologia similar à de *Vm26Aa*, indica um padrão distinto de relação entre as espécies. Esta região apresentou uma quantidade bem maior de substituições não-sinônimas do que *Vm 26Aa*, uma vez que esta molécula, embora aproximadamente tenha o mesmo tamanho daquela, apresenta 31 substituições não-sinônimas. Contudo, o padrão mutacional foi mais disperso na árvore, e não tão concentrado em mudanças no interior da rede de haplótipos, em oposição a mudanças na ponta da árvore. Por outro lado, vemos que dos três grupos claramente definidos, um é composto de espécies de *A. fraterculus*, o outro de *A. obliqua* e um terceiro que apresenta tanto uma mescla de ambas quanto um ramo mais comprido que apresenta grande parte das *A. sororcula*. Apesar destes grupos apresentarem apenas espécimes destas espécies, isso não significa que tais grupos abranjam todas os espécimes para cada uma destas espécies, embora encontremos uma segregação maior do que a encontrada em *Vm 26Aa*.

O padrão encontrado para *Cps38* é distinto dos anteriores, uma vez que não observamos agrupamentos organizados entre haplótipos distintos, mas sim uma quantidade razoável de haplótipos bastante distintos entre si, ou seja, separados por ramos longos. Há inclusive alguns destes ramos que violam os limites da parcimônia (isto é, apresentam um número tão grande de mudanças que há probabilidade razoável de haver homoplasia neste ramo), como é o caso da conexão de obl21 e obl18 com o resto da rede, o que indica grande variação entre haplótipos existentes. Este padrão também costuma ser encontrado quando há recombinações entre as moléculas investigadas, o que viola uma premissa essencial da inferência filogenética. Apesar de testes de recombinação realizados no software RDP não terem detectado evidências significativas de recombinação (dados não apresentados), estes testes não são muito eficazes quando há pouca diferenciação entre moléculas recombinantes e pode não detectar tais recombinações. Contudo, uma observação mais detalhada nos revela que boa parte da variação encontrada está diferenciada entre as espécies. O lado direito da rede de haplótipos é completamente constituído *A. fraterculus* enquanto o lado esquerdo

possui basicamente *A. obliqua* e *A. sororcula*. Embora existam no meio ainda uma pouca diferenciação entre estas espécies, no geral o padrão entre elas está mais definido. Apesar desta definição, boa parte das mutações que separa tais ramos é sinônima. Apesar disto, existe grande quantidade de variação não-sinônima, uma vez que 42 das 92 substituições levam à alteração de aminoácidos, e 21 destas encontra-se em ramos internos da rede. Além disso, verificamos neste gene uma grande variação em *indels*, de forma que existem quatro posições distintas na proteína em que há diferenças, inserções/deleções de aminoácidos, em geral de mais de um ao mesmo tempo. Por exemplo, em duas posições há haplótipos que diferem de 3 aminoácidos entre si, e existe ainda uma outra região em que os haplótipos diferem de 6 ou mais aminoácidos. Estas mudanças ocorrem principalmente em uma região rica em glicina na qual foram encontrados sítios com evidências de seleção positiva (Fig. 22), o que pode também refletir a existência de um padrão distinto de variação nesta região. Interessantemente, a região que apresentou maior proporção de mutações não-sinônimas nas análises populacionais foi entre as posições 40 e 60, outra região rica em glicina próxima ao início da proteína.

O gene *Cps16* apresentou indícios de seleção positiva no PAML, particularmente entre as posições 51 e 105. Uma avaliação da rede haplotípica produzida para este gene (Fig. 27) revela que embora haja variação genética intra e interespecífica, o padrão da rede de haplótipos é similar ao encontrado para *Cps38*, em que não há o claro agrupamento de haplótipos em grupos menores. Encontramos apenas do lado esquerdo da rede um agrupamento que contém praticamente todos os haplótipos de *A. sororcula* (apenas um não está neste grupo), separada por três substituições, sendo duas não-sinônimas que estão próximas e localizadas logo após o início do *loop* externo, enquanto os outros agrupamentos contêm haplótipos das duas outras espécies. 17 das 26 mutações não sinônimas observadas na análise populacional, encontra-se nesta região próxima ao início da proteína (até o aminoácido na posição 50), o que destoa da região na qual foi detectada seleção positiva revelada anteriormente (Fig. 19).

Os resultados de *Sxl* e *Tra2* foram considerados conjuntamente por terem apresentado padrões similares. Em ambos foi encontrada uma baixa quantidade variação haplotípica o que reflete em redes em que os haplótipos estão separados por um número menor de passos evolutivos. Da mesma forma que outros genes, não há uma distinção dos haplótipos encontrados em espécies distintas, mas aqui, possivelmente em função da menor variabilidade genética, encontramos poucos agrupamentos e em nenhum destes há indicações de um certo

alinhamento ao padrão específico. Enquanto *Sxl* apresentou 10 substituições não-sinônimas, sendo apenas uma compartilhada por mais de um haplótipo, o *Tra2* apresentou apenas 10 substituições não-sinônimas e destas apenas três compartilhadas. Estes valores destoam dos encontrados nos outros genes aqui estudados e revelam maior conservação nas substituições encontradas. A baixa variação do *Sxl* pode refletir a seleção positiva detectada no item 4.3, e pode indicar que esta seleção ocorreu recentemente de forma que não houve tempo para recuperação da variação genética desta região.

A análise conjunta dos genes aqui estudados sugere que os aminoácidos que estão sob ação da seleção positiva no contraste entre *Tephritidae* de *Drosophilidae* não são normalmente os mesmos que apresentam maior variação não silente intra e interespecificamente. Em algumas situações as regiões protéicas estão próximas, como em *Cps38*, mas em outros, como *Cps16* e *Sxl*, regiões distintas estão associadas, o que pode indicar que sejam de fato processos distintos que estiveram envolvidos no processo global de diferenciação destas distantes linhagens evolutivas e na diferenciação de espécies de *Anastrepha*. Por outro lado, pode ser também que a seleção positiva identificada no item 4.3 tenha eliminado a variação genética nestes sítios, enquanto que a variação ainda segregante nos outros sítios, não nos permita inferir um processo seletivo pela ausência ainda da fixação de tais polimorfismos.

Os resultados das análises populacionais destes genes indicam que, no geral, não há separação entre as espécies de forma a nos permitir ter marcadores específicos de espécies definitivos. Entretanto há em diversas situações, uma clara separação entre alguns dos haplótipos de uma espécie, de forma a sugerir alguma segregação dos polimorfismos ancestrais. Estes dados são mais evidentes em genes que apresentaram uma taxa evolutiva maior e ainda mais com sinais de seleção positiva, do que em outros genes. Isto pode ser evidência de que estes genes estejam de fato envolvidos no processo de diferenciação. Se for este o caso, como explicar que eles não sejam de fato específicos para a espécie? Uma limitação de nossas análises diz respeito à localização geográfica dos espécimes coletados de cada espécie. Como fizemos um pool *para* cada espécie, com espécimes de localidades distintas, não há como buscar uma associação entre localidades específicas e certos agrupamentos aqui encontrados. 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.

Por outro lado, consideramos mais provável que estes genes não estejam diretamente envolvidos no processo de diferenciação, mas sim, por terem uma taxa evolutiva maior, melhor representem o rápido processo evolutivo de diferenciação das moscas deste grupo. Desta forma, não esperamos uma congruência total entre os grupos formados pela análise aqui realizada e as espécies, mas uma associação geral, como a que foi encontrada. Uma análise de coalescência realizada com diversos genes pode muito mais eficientemente permitir a identificação de linhagens evolutivas e mesmo testar a existência de tais linhagens com a existência de polimorfismos ancestrais (CARSTENS & KNOWLES 2007, KUBATKO *et al.* 2009). Como tais estudos se beneficiam de um número maior de genes para sua melhor resolução estamos considerando em outra publicação cerca de 20 genes distintos para tal análise (Fernanda Fernandes, Vanessa R Gonçalves, Iderval da Silva Sobrinho, Reinaldo Alves de Brito, dados não publicados).

4- Conclusões

Através do seqüenciamento dos clones obtidos na biblioteca de cDNA de *Anastrepha obliqua* foi possível identificar um grande número de diferentes proteínas. Assim, o *screening* desta biblioteca ainda indica uma baixa saturação, com poucos contigs que foram mais representados. Entretanto, os resultados obtidos foram de grande importância para a identificação de novos genes, além de suas funções e processos biológicos envolvidos, tais como: determinação do sexo, diferenciação sexual, reprodução e muitos dos ESTs encontrados ainda não são conhecidos.

Desta biblioteca, focamos em estudar genes que estivessem ligados a ao processo de reprodução e diferenciação sexual, de forma que buscamos estudar com mais detalhes alguns genes coriônicos, vitelínicos, o gene *Sxl* e *Tra2*. Destes, os dois genes coriônicos e o *Sxl* revelaram estar sob seleção positiva de acordo com o LRT e também com a análise de BEB foi possível identificar vários sítios sob seleção. Dos dois genes vitelínicos apenas o Vm26Aa', demonstrou sinal de seleção positiva, este sinal possivelmente reflexo de uma duplicação recente que pode ter ocorrido e levado a uma nova função gênica. O *Sxl* foi outro gene que assim como os coriônicos demonstrou estar sob seleção positiva. Já o *Tra2*, embora também envolvido na diferenciação sexual, revelou estar sob seleção purificadora, possivelmente reflexo do papel central deste gene na cascata sexual.

As estruturas secundárias das proteínas inferidas revelaram que, de maneira geral, as regiões que mais sujeitas à seleção positiva estavam presentes do lado externo da membrana coriônica mantendo a possibilidade de que tal seleção seja reflexo de interações entre óvulo e espermatozóide durante a fertilização, e/ ou devido a fatores ecológicos, por estas proteínas estarem provavelmente expostas a um ambiente (quase) externo.

As análises populacionais dos genes coriônicos, vitelínicos, *Sxl* e *Tra2* não revelaram marcadores específicos para a espécie, ou seja, embora a variação nestes genes em algumas situações ter conseguido separar conjuntos de haplótipos que estavam preferencialmente em uma espécie, não houve nenhum marcador que fosse determinante para a identificação de uma espécie. Como fizemos um *pool* para cada espécie, com espécimes de localidades distintas, não foi possível identificar se estes agrupamentos observados seguem algum padrão filogeográfico específico, o que deve ser buscado futuramente.

Finalmente, os genes aqui estudados fornecem importantes informações sobre estudos de especiação neste grupo. O fato de termos identificado genes sob seleção positiva, com taxas evolutivas mais altas, permitiu uma observação mais detalhada do processo evolutivo nestas espécies e confirmou a recente divergência deste grupo de moscas. Embora neste estudo tenhamos conseguido identificar diversos sítios sob seleção positiva, tais análises filogeográficas, assim como uma melhor elucidação das estruturas secundárias e terciárias das proteínas dos genes estudados, serão necessários para melhor entendermos as forças envolvidas na seleção que está possivelmente ocorrendo, para que possamos avaliar se alguns dos genes aqui identificados sob seleção positiva estão direta, ou indiretamente envolvidos no processo de especiação deste grupo de moscas.

5- Referências

- Alatortsev, VE. 2005. New genes for vitelline membrane proteins in *Drosophila*. *Molecular Biology*. **40**: 330-332.
- Alvarez, M; Ruiz, MF; Sanchez, L. 2009. Effect of the Gene doublesex of *Anastrepha* on the Somatic Development of *Drosophila*. *PLoS ONE* **4(4)**: e5141.
- Andrés J; Escobar G; Arenas AF; Gómez-Marin JE. 2006. Molecular Evolution of Serine/Arginine Splicing Factors Family (SR) by Positive Selection. *In Silico Biology*. **6**: 347-350.
- Araujo, EL; Zucchi, RA. 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.
- Attardo, MG; Ribeiro, JMC; Wu, Y; Berriman, M; Aksoy, S. 2010. Transcriptome analysis of reproductive tissue and intrauterine developmental stages of the tsetse fly (*Glossina morsitans morsitans*). *BMC Genomics*. **11**:160.
- Beeman, RW & Brown, SJ. 1999. RAPD-Based Genetic Linkage Maps of *Tribolium castaneum*. *Genetics*. **153**: 333–338.
- Brooks, ML. 1999. Effects of Protective Fencing on Birds, Lizards and Black-Tailed Hares in the Western Mojave Desert. *Environmental Management* **23**:387-400.
- Butlin, R. e Ritchie, MG. 2001. Searching for speciation genes. *Nature*. **412**: 31-33.
- Carstens, BC. e Knowles, LL. 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus grasshoppers*. *Syst. Biol.* **56**: 1-12.

- Castelloe, JM. e Templeton, AR. 1994. Root Probabilities for Intraspecific Gene Trees under Neutral Coalescent Theory. *Mol. Phylogenet. Evol.* **3**: 102--113.
- Civetta A. & Singh RS. 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.
- Civetta A. 2003. Positive Selection Within Sperm-Egg Adhesion Domains of Fertilin. An ADAM Gene with a Potential Role in Fertilization. *Mol. Biol. Evol.* **20**: 21–29
- Clark NL, Aagaard JE, Swanson WJ. 2006. Evolution of reproductive proteins from animals and plants. *Reproduction.* **131**: 11–22.
- Chapman T, Davies SJ. 2004. Functions and analysis of the seminal fluid proteins of male *Drosophila melanogaster* fruit flies. *Peptides.* **25**: 477–1490.
- Chomczynski & Sacchi N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162**: 156–159.
- Chuong EB, Tong W, Hoekstra HE 2010. Maternal-Fetal Conflict: Rapidly Evolving Proteins in the Rodent Placenta. *Mol. Biol. Evol.* **27**:1221-1225.
- Comeron, JM. 1999. K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. *Bioinformatics.* **15**: 763-4.
- Cortner, J.; Woude G.F.V. Essentials of molecular biology – cDNA libraries. In: Cancer: Principles and practice of oncology. DeVita V.T. Jr., Hellman S. and Rosemberg S.A. (ed), 5th ed, Lippincott – Raven Publishers, Philadelphia, 1997. Coyne JA, Orr HA. 2004. *Speciation*. Sunderland, MA: Sinauer. 545 pp.
- Dobzhansky T. 1937. Genetics and the Origin of Species. New York: Columbia Univ. Press.
- Eads, B; Colbourne, J; Bohuski, E; Andrews, J. 2007. Profiling sex-biased gene expression during parthenogenetic reproduction in *Daphnia pulex*. *BMC Genomics* **8**: 464-

- Eberhard, W. G. 1985. *Sexual Selection and the Evolution of Animal Genitalia*. Cambridge, Mass.: Harvard University Press.
- Emanuelsson, O; Brunak, S; von Heijne, G; Nielsen, H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protoc*: **2**: 953-971.
- Ewing, B; Green, P. 1998. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Res*. **8**: 186-194.
- Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Res*. **8**: 175-185.
- Feder, JL; Xie, X; Rull, J; Velez, S; Forbes, A; Leung, B; Dambroski, H; Filchak, KE; Aluja, M. 2005. Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*. *PNAS*. **102**: 6573-6580.
- Ferreira, ME; Grattapaglia, D. *Introdução ao uso de marcadores moleculares em análise genética*. 3. ed. Brasília: EMBRAPA-CENARGEN, 1998. 220p.
- Findlay, GD; Swanson WJ. 2010. Proteomics enhances evolutionary and functional analysis of reproductive proteins. *BioEssays*. **32**: 26-36.
- Gillott C. 2003. Male accessory gland secretions: modulators of female reproductive physiology and behavior. *Annu Rev Entomol*. **48**: 163-184.
- Guo, Y; Ribeiro, JMC, Anderson, JM; Bour, S. 2009. dCAS: a desktop application for cDNA sequence annotation. *Bioinformatics*. **25**: 1195-1196.
- Greenberg AJ, Moran JR, Coyne JA, Wu CI. 2003. Ecological adaptation during incipient speciation revealed by precise gene replacement. *Science*. **302**: 1754-57.
- Higgins DG. 1994. CLUSTAL V: multiple alignment of DNA and protein sequences. *Methods Mol Biol*. **25**: 307-318.
- Hittinger CT & Carroll SB. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature*. **449**: 677-682.

- Hofmann, MA; Brian, DA. 1993. Sequencing DNA Amplified Directly from a Bacterial Colony. **15**: 205- 210.
- Huang, X., Madan, A. 1999. CAP3: A DNA sequence assembly program. *Genome Res.* **9**: 868-77
- Jones DT, Taylor WR, Thornton JM. 1994. A Model Recognition Approach to the Prediction of All-Helical Membrane Protein Structure and Topology. *Biochemistry.* **33**: 3038-3049.
- Jones DT. 2007. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics.* **23**: 538-544
- Kawata M, Yoshimura J. 2000. Speciation by sexual selection in hybridizing populations without viability selection. *Evol. Ecol. Res.* **2**: 897–909.
- Kelleher, ES, Swanson, WJ, Markow, TA. 2007. Gene Duplication and Adaptive Evolution of Digestive Proteases in *Drosophila arizonae* Female Reproductive Tracts. *PLoS Genet.* **3**: 1541-1549.
- Knipling, EF. 1955. Possibilities of insect control or eradication through the use of sexually sterile males J. *Econ.Entomol.* **48**: 459–462.
- Krafsur, ES. 1998. Sterile insect technique for suppressing and eradicating insect populations: 55 years and counting J. *Agric. Entomol.* **15**: 313–317.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* **3**: 567-80.
- Kryazhimskiy S, Plotkin JB. 2008: The Population Genetics of dN/dS. *PLoS Genet.* **4**:100-304
- Komitopoulou, K; Christophides, GK; Kalosaka, K; Chrysanthis, G; Theodoraki, MA; Savakis, C; Zacharopoulou, A; Mintzas, AC. 2004. Medfly promoters relevant to the sterile insect technique. *Insect Biochem Mol Biol.* **34**: 149-157.

- Kovaleski, ARL. Sugayama & A. Malavasi. 2000. Controle químico em macieiras, p. 135-141. In: A. Malavasi & R. A. Zucchi (Ed.), *Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado*. Ribeirão Preto, Holos, 325 p.
- Kubatko L., Carstens, BC., Knowles LL. 2009. STEM: Species tree estimation using Maximum likelihood for gene trees under coalescence. *Bioinformatics*. **25**: 971-973.
- Lagos D, Koukidou M, Savakis C, Komitopoulou K. 2007. The *transformer* gene in *Bactrocera oleae*: the genetic switch that determines its sex fate. *Insect Mol Biol*. **16**: 221–230.
- Lis, JT; Schleif, R. 1975. Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. *Nucleic Acids Res* **2**: 383-389.
- Lorenzen, MD; Doyungan, Z; Savard, J; Snow, K; Crumly, LR; Shippy, TD; Stuart, JJ; Brown, SJ; Beeman, RW. 2005. Genetic Linkage Maps of the Red Flour Beetle, *Tribolium castaneum*, Based on Bacterial Artificial Chromosomes and Expressed Sequence Tags. *Genetics*. **170**: 741-747.
- Malavasi, A. e R. Zucchi, A. 2000. *Moscas-das-frutas de importância econômica no Brasil: Conhecimento básico e aplicado*. First ed. Ribeirão Preto: Holos Editora
- Morgante, JS; Selivon, D; Solferini, VN; Matioli, SR. 1993. Evolutionary patterns in specialist and generalist species of *Anastrepha*. In: *Fruit flies: Biology and Management* (M Aluja and P Liedo). pp. 15- 20. Springer-Verlag. New, York.
- Mueller, JL; Ram, KR; McGraw, LA; Bloch Qazi, MC; Siggia, ED; Clark, AG; Aquadro, CF; Wolfner, MF. 2005. Cross-Species Comparison of *Drosophila* Male Accessory Gland Protein Genes. *Genetics*. **171**: 131-143.
- Nielsen, R; Yang, Z. 1998. Likelihood Models for Detecting Positively Selected Amino Acid Sites and Applications to the HIV-1 Envelope Gene. *Genetics*. **148**: 929-936.
- Norrbom, A L. 2001. Tephritidae classification table. <http://www.sel.barc.usda.gov/diptera/tephriti/Tephclas.htm>. (27 junho2001)

- Norrbom, A L; R A; Zucchi & Hernández-Ortiz V. 1999. Phylogeny of the genera *Anastrepha* and *Toxotrypana* (Trypetinae: Toxotripanini) based on morphology, p. 299-342. In: A. L. NORRBOM & M. ALUJA (eds.). Fruit flies (Tephritidae): phylogeny and evolution of behavior. Boca Raton, *CRC Press*, xviii+963 p.
- Norrbom, A.L., R.A. Zucchi & V. Hernández-Ortiz. 2000. Phylogeny of the genera *Anastrepha* and *Toxotrypana* (Trypetinae: Toxotripanini) based on morphology. In M. Aluja & A.L. Norrbom (eds.), Fruit flies (Tephritidae): Phylogeny and evolution of behavior. Boca Raton , *CRC Press*, 944p.
- Nugent, T. & Jones, DT. 2009. Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics*. **10**:159
- Parks S, Wakimoto B, Spradling AC. 1986. Replication and ex-pression of an X-linked cluster of *Drosophila* chorion genes. *Dev Biol*. **117**:294–305.
- Parks S & Spradling A 1987. Spatially regulated expression of chorion genes during *Drosophila* oogenesis. *Dev. Biol*. **91**: 325– 336.
- Pane A, Salvemini M, Bovi PD, Polito C, Saccone G.2002. The *transformer* gene in *Ceratitidis capitata* provides a genetic basis for selecting and remembering the sexual fate. *Development* **129**: 3715–3725.
- Panhuis, TM; Swanson, WJ. 2006. Molecular Evolution and Population Genetic Analysis of Candidate Female Reproductive Genes in *Drosophila*. *Genetics*. **173**: 2039-2047 -
- Parkinson, J; Anthony, A; Wasmuth, J; Schmid, R; Hedley, A; Blaxter, M. 2004. PartiGene--constructing partial genomes. *Comput. Appl. Biosci*. **20**: 1398-1404.
- Perez DE, Wu CI. 1995. Further characterization of the Odysseus locus of hybrid sterility in *Drosophila*. One gene is not enough. *Genetics*. **140**:201–106.
- Poiani A. 2006. Complexity of seminal fluid: a review. *Behav Ecol Sociobiol*. **60**:289-310.

- Pomiankowski, A; Muller, AP. 1995. A resolution of the lek paradox. *Proc. R. Soc. Lond B* **260**: 21-29.
- Pomiankowski A, Nothiger R, Wilkins A. 2004. The Evolution of the *Drosophila* Sex-Determination Pathway. *Genetics*. **166**:1761-1773.
- Ritchie, MG. 2007. Sexual Selection and Speciation Rev. *Ecol. Evol. Syst.* **38**:79-102
- Roethele, JB; Romero-Severson, J; Feder, JL. 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.
- Ronchi-Teles, B. 2000. Ocorrência e flutuação populacional de espécies de moscas-das-frutas e parasitóides com ênfase para o gênero *Anastrepha* (Diptera: Tephritidae) na Amazônia brasileira. Tese de Doutorado. Manaus, Instituto Nacional de Pesquisas da Amazônia, Fundação Universidade do Amazonas, 156p.
- Rozen, S; Skaletsky, HJ. 2000. Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (SA Krawetz and S Misener). pp. 365- 386.Humana Press. Totowa, NJ.
- Ruiz, M. F., Eirín-López, J. M., Stefani, R. N., Perondini, A. L. P., Selivon, D. e Sánchez, L. 2007. The gene *doublesex* of *Anastrepha* fruit flies (Diptera, Tephritidae) and its evolution in insects. *Development Genes and Evolution*, **217**: 675-731.
- Russo CA, Takezaki N, Nei M. 1995. Molecular phylogeny and divergence times of drosophilid species. *Molecular Biology and Evolution*. **12**: 391-404.
- Saccone, G; Pane, A; Simone, A; Salvemini, M; Milano, A; Annunziata, L; Mauro, U; Polito, L. 2007. New Sexing Strains for Mediterranean Fruit Fly *Ceratitidis capitata* : Transforming Females into Males. 95- 102.
- Salz HK & Erickson JW. 2010. Sex determination in *Drosophila*: The view from the top. *Landes Bioscience*. **4(1)**: 60–70.

- Santos, Pd; Uramoto, K; Matioli, SR. 2001. Experimental hybridization among *Anastrepha* species (Diptera: Tephritidae): Production and morphological characterization of F₁ hybrids. *Ann Entomol Soc Am* **94**: 717-725.
- 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, ALP; Morgante, JS. 2005. A Genetic-Morphological Characterization of Two Cryptic Species of the *Anastrepha fraterculus* Complex (Diptera: Tephritidae). *Ann Entomol Soc Am*. **98**: 367-381.
- Siera , SG & Cline, TW. 2008. Sexual Back Talk With Evolutionary Implications: Stimulation of the *Drosophila* Sex-Determination Gene Sex-lethal by Its Target transformer. *Genetics*. **180**: 1963–1981.
- Sobrinho IJ & de Brito RA *subm.* Evidence of positive selection in fruitless *subm.* to BMC Evol. Biol.
- Solferini, VN; Morgante, JS. 1987. Karyotype study of eight species of *Anastrepha* (Diptera, Tephritidae). *Caryologia* **40**: 229-241.
- Steck, GJ. 1991. Biochemical systematics and population genetic structure of *Anastrepha fraterculus* and related species (Diptera: Tephritidae). *Ann. Entomol. Soc. Am.* **84**: 10-28.
- Steinke, D; Salzburger, W; Meyer, A. 2004. EverEST - A phylogenomic EST database approach. *Phyloinformatics*. **6**: 1-4.
- Steinke, D; Salzburger, W; Meyer, A. 2006. Novel Relationships Among Ten Fish Model Species Revealed Based on a Phylogenomic Analysis Using ESTs. *J Mol Evol*. **62**: 772-784.
- Sun S, Ting CT, Wu CI. 2004. The normal function of a speciation gene, *Odysseus*, and its hybrid sterility effect. *Science*. **305**: 81–83.

- Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc Natl Acad Sci USA*. **98**: 7375–7379.
- Swanson WJ, Vacquier VD. 1998. Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. *Science*. **281**: 710–712.
- Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. *Nat. Rev Genet*. **3**: 137–144.
- Swanson, WJ; Wong, A; Wolfner, MF; Aquadro, CF. 2004. Evolutionary Expressed Sequence Tag Analysis of *Drosophila* Female Reproductive Tracts Identifies Genes Subjected to Positive Selection. *Genetics*. **168**: 1457-1465.
- Templeton, AR. 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.
- Terai Y, Seehausen O, Sasaki T, Takahashi K, Mizoiri S, et al. 2006. Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS Biol*. **4**:e433
- Ting CT, Tsaor SC, Wu ML, Wu CI. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**:1501–04.
- Tolias, PP., Konsolaki, M., Komitopoulou, K., and Kafatos, FC. 1990. The chorion genes of the medfly, *Ceratitis capitata*. II. Characterization of three novel cDNA clones obtained by differential screening of an ovarian library. *Dev. Biol*. **140**: 105–112.
- Trougakos, I. P., Lamnissou, K., Margaritis, L. H. 1999. Biochemical and Immunocytochemical Analysis of Vitellogenesis in the Olive Fruit Fly *Dacus (Bactrocera) Oleae* (Diptera: Tephritidae), *Cell Biol. Inter*. **23**: 417-429.
- Tsaor, SC; Wu, CI. 1997. Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of *Drosophila*. *Mol. Biol. Evol*. **14**: 544-549.

- Uramoto, K; Walder, MM; Zucchi RA. 2004. Biodiversidade de moscas-das-frutas do gênero *Anastrepha* (Diptera, Tephritidae) no campus da ESALQ-USP, Piracicaba, São Paulo. *Revista Brasileira de Entomologia*. **48**: 409-414
- Vlachou, D; Konsolalti, M; Tolias, PP; Kafatoss, FC; Komitopoulou, K. 1997. The Autosomal Chorion Locus of the Medfly *Ceratitidis capitata*. I. Conserved Synteny, Amplification and Tissue Specificity but Sequence Divergence and Altered Temporal Regulation. *Genetics*. **147**: 1829-1842.
- Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**: 568-573.
- Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**: 1586-1591.
- West-Eberhard, MJ. 1985. Alternative adaptations, speciation, and phylogeny (A Review). *Proc. Natl. Acad. Sci. USA*. **83**: 1388-1392.
- WilliansI JGK.; Kubelic AR.; Livak, KJ.; Rafalski JA.; Tingey SV. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, **18**: 6531-6535.
- Wolfner MF, Heifetz Y, Applebaum SW. 2005. Gonadal glands and their gene products. In: Gilbert LI, Iatrou K, Gill SS, editors. *Comprehensive molecular insect science: reproduction and development*. Oxford: Elsevier Ltd. p.179–212.
- Wong, A; Turchin, MC; Wolfner, MF; Aquadro, CF. 2008a. Evidence for Positive Selection on *Drosophila melanogaster* Seminal Fluid Protease Homologs. *Mol Biol Evol.* **25**: 497-506.
- Wu,C.I e Ting, C.T. 2004. Genes and speciation. *Nat. Genet.* **5**: 114-122.
- Zacharopoulou, AM; Frisardi, C; Savakisa, S; Robinson, P. 1992. The genome of the mediterranean fruitfly *ceratitidis capitata*: localization of molecular markers by *in situ* hybridization to salivary gland polytene chromosomes. *Chromosoma*.**101**: 448-455.

- Zeng, ZB; Liu, J; Stam, LF; Kao, CH; Mercer, JM; Laurie, CC. 2000. Genetic Architecture of a Morphological Shape Difference Between Two *Drosophila* Species. *Genetics*. **154**: 299-310.
- 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.
- Zhang, J. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*. **24**: 1586-1591.
- Zucchi, RA. 2000. Espécies de *Anastrepha*, sinonímias, plantas hospedeiras e parasitóides . *In: Moscas-das-frutas de importância econômica no Brasil Conhecimento Básico e Aplicado* (A Malavasi and RA Zucchi). pp. 41- 48.Holos Editora . Ribeirão Preto, SP.

ANEXO A- Modelos Evolutivos

Modelo	Parâmetros	LnL
M1a (modelo nulo)	$\omega_0 = 0,054$; $\omega_1 = 1,000$ $p_0 = 0,622$; $p_1 = 0,378$	-1950,671956
MA restrito (modelo nulo para os <i>Tephritidae</i>)	$\omega_0 = 0,044$; $\omega_2 = 1,000$ $p_0 = 0,427$; $p_1 = 0,258$; $(p_{2a}+p_{2b}) = 0,3144$	-1942,337195
MA restrito (modelo nulo para as <i>Anastrepha</i>)	$\omega_0 = 0,044$; $\omega_2 = 1,000$ $p_0 = 0,552$; $p_1 = 0,327$; $(p_{2a}+p_{2b}) = 0,124$	-1946,537197
MA (modelo alternativo para os <i>Tephritidae</i>)	$\omega_0 = 0,049$; $\omega_2 = 4,743$ $p_0 = 0,446$; $p_1 = 0,262$; $(p_{2a}+p_{2b}) = 0,292$	-1940,791597
MA (modelo alternativo para as <i>Anastrepha</i>)	$\omega_0 = 0,043$; $\omega_2 = 999,000$ $p_0 = 0,584$; $p_1 = 0,343$; $(p_{2a}+p_{2b}) = 0,072$	-1948,581825

Tabela A.1 - Modelos evolutivos para o gene *Cps 16* não considerando os *gaps*. LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS de $0 < \omega < 1$; ω_1 = dN/dS fixado a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ que correspondem apenas a sítios do ramo *foreground*.

Modelo	Parâmetros	LnL
M1a (modelo nulo)	$\omega_0 = 0,055$; $\omega_1 = 1,000$ $p_0 = 0,573$; $p_1 = 0,427$	-2269,680752
MA restrito (modelo nulo para os <i>Tephritidae</i>)	$\omega_0 = 0,046$; $\omega_2 = 1,000$ $p_0 = 0,382$; $p_1 = 0,278$; $(p_{2a}+p_{2b}) = 0,339$	-2260,843252
MA restrito (modelo nulo para as <i>Anastrepha</i>)	$\omega_0 = 0,042$; $\omega_2 = 1,000$ $p_0 = 0,482$; $p_1 = 0,339$; $(p_{2a}+p_{2b}) = 0,17821$	-2266,941533
MA (modelo alternativo para os <i>Tephritidae</i>)	$\omega_0 = 0,051$; $\omega_2 = 5,668$ $p_0 = 0,398$; $p_1 = 0,279$; $(p_{2a}+p_{2b}) = 0,323$	-2259,137702
MA (modelo alternativo para as <i>Anastrepha</i>)	$\omega_0 = 0,044$; $\omega_2 = 26,300$ $p_0 = 0,531$; $p_1 = 0,376$; $(p_{2a}+p_{2b}) = 0,093$	-2265,578253

Tabela A.2- Modelos evolutivos para o gene *Cps16* considerando os *gaps*. LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS de $0 < \omega < 1$; ω_1 = dN/dS fixado a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ que correspondem apenas a sítios do ramo *foreground*.

Modelo	Parâmetros	LnL
M1a (modelo nulo)	$\omega_0 = 0,034$; $\omega_1 = 1,000$ $p_0 = 0,946$; $p_1 = 0,054$	-1547,030898
MA restrito (modelo nulo para os <i>Tephritidae</i>)	$\omega_0 = 0,025$; $\omega_2 = 1,000$ $p_0 = 0,771$; $p_1 = 0,029$; $(p_{2a}+p_{2b}) = 0,20031$	-1528,819230
MA (modelo alternativo para os <i>Tephritidae</i>)	$\omega_0 = 0,030$; $\omega_2 = 142,098$ $p_0 = 0,746$; $p_1 = 0,017$; $(p_{2a}+p_{2b}) = 0,23648$	-1523,405451

Tabela A.3- Modelos evolutivos para o gene *Sex lethal* considerando os *gaps*. LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS de $0 < \omega < 1$; ω_1 = dN/dS fixado a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ que correspondem apenas a sítios do ramo *foreground*.

Modelo	Parâmetros	LnL
M1a (modelo nulo)	$\omega_0 = 0,033$; $\omega_1 = 1,000$ $p_0 = 0,949$; $p_1 = 0,051$	-1437,890372
MA restrito (modelo nulo para os <i>Tephritidae</i>)	$\omega_0 = 0,026$; $\omega_2 = 1,000$ $p_0 = 0,777$; $p_1 = 0,032$; $(p_{2a}+p_{2b}) = 0,189$	-1421,919415
MA (modelo alternativo para os <i>Tephritidae</i>)	$\omega_0 = 0,030$; $\omega_2 = 128,153$ $p_0 = 0,762$; $p_1 = 0,024$; $(p_{2a}+p_{2b}) = 0,214$	-1417,653124

Tabela A.4- Modelos evolutivos para o gene *Sex lethal* não considerando os *gaps*. LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS de $0 < \omega < 1$; ω_1 = dN/dS fixado a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ que correspondem apenas a sítios do ramo *foreground*.

Modelo	Parâmetros	LnL
M1a (modelo nulo)	$\omega_0 = 0,076$; $\omega_1 = 1,000$ $p_0 = 0,934$; $p_1 = 0,066$	-1213,514210
MA restrito (modelo nulo para os <i>Tephritidae</i>)	$\omega_0 = 0,052$; $\omega_2 = 1,000$ $p_0 = 0,587$; $p_1 = 0,046$; $(p_{2a}+p_{2b}) = 0,366$	-1203,423224
MA restrito (modelo nulo para as <i>Anastrepha</i>)	$\omega_0 = 0,076$; $\omega_2 = 1,000$ $p_0 = 0,934$; $p_1 = 0,066$; $(p_{2a}+p_{2b}) = 0,000$	-1213,514210
MA (modelo alternativo para os <i>Tephritidae</i>)	$\omega_0 = 0,057$; $\omega_2 = 999,000$ $p_0 = 0,579$; $p_1 = 0,044$; $(p_{2a}+p_{2b}) = 0,37727$	-1202,475546
MA (modelo alternativo para as <i>Anastrepha</i>)	$\omega_0 = 0,076$; $\omega_2 = 1,000$ $p_0 = 0,934$; $p_1 = 0,066$; $(p_{2a}+p_{2b}) = 0,000$	-1213,514210

Tabela A.5- Modelos evolutivos para o gene *Tra2* considerando os *gaps*. LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS de $0 < \omega < 1$; ω_1 = dN/dS fixado a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ que correspondem apenas a sítios do ramo *foreground*.

Modelo	Parâmetros	LnL
M1a (modelo nulo)	$\omega_0 = 0,075$; $\omega_1 = 1,000$ $p_0 = 0,936$; $p_1 = 0,064$	-1134,297731
MA restrito (modelo nulo para os <i>Tephritidae</i>)	$\omega_0 = 0,052$; $\omega_2 = 1,000$ $p_0 = 0,598$; $p_1 = 0,047$; $(p_{2a} + p_{2b}) = 0,356$	-1124,508306
MA restrito (modelo nulo para as <i>Anastrepha</i>)	$\omega_0 = 0,075$; $\omega_2 = 1,000$ $p_0 = 0,936$; $p_1 = 0,064$; $(p_{2a} + p_{2b}) = 0,000$	-1134,297731
MA (modelo alternativo para os <i>Tephritidae</i>)	$\omega_0 = 0,057$; $\omega_2 = 999,000$ $p_0 = 0,589$; $p_1 = 0,044$; $(p_{2a} + p_{2b}) = 0,367$	-1123,603328
MA (modelo alternativo para as <i>Anastrepha</i>)	$\omega_0 = 0,075$; $\omega_2 = 1,000$ $p_0 = 0,936$; $p_1 = 0,064$; $(p_{2a} + p_{2b}) = 0,000$	-1134,297731

Tabela A.6- Modelos evolutivos para o gene *Tra2* não considerando os *gaps*. LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS de $0 < \omega < 1$; ω_1 = dN/dS fixado a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ que correspondem apenas a sítios do ramo *foreground*.

Modelo	Parâmetros	
M1a (modelo nulo)	$\omega_0 = 0,090$; $\omega_1 = 1,000$ $p_0 = 0,727$; $p_1 = 0,273$	-1276,682843
MA restrito (modelo nulo para os <i>Tephritidae</i>)	$\omega_0 = 0,076$; $\omega_2 = 1,000$ $p_0 = 0,634$; $p_1 = 0,232$; $(p_{2a} + p_{2b}) = 0,13431$	-1275,309477
MA (modelo alternativo para os <i>Tephritidae</i>)	$\omega_0 = 0,077$; $\omega_2 = 999,00000$ $p_0 = 0,629$; $p_1 = 0,224$; $(p_{2a} + p_{2b}) = 0,148$	-1275,041186

Tabela A.7- modelos evolutivos para o gene *Vm 26Aa* considerando os *gaps*. LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS de $0 < \omega < 1$; ω_1 = dN/dS fixado a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ que correspondem apenas a sítios do ramo *foreground*.

M1a (modelo nulo)	$\omega_0 = 0,078$; $\omega_1 = 1,000$ $p_0 = 0,756$; $p_1 = 0,244$	-964,372374
MA restrito (modelo nulo para os <i>Tephritidae</i>)	$\omega_0 = 0,062$; $\omega_2 = 1,000$ $p_0 = 0,634$; $p_1 = 0,196$; $(p_{2a} + p_{2b}) = 0,169$	-962,715319
MA (modelo alternativo para os <i>Tephritidae</i>)	$\omega_0 = 0,063$; $\omega_2 = 999,000$ $p_0 = 0,633$; $p_1 = 0,191$; $(p_{2a} + p_{2b}) = 0,176$	-962,452748

Tabela A.8- Modelos evolutivos para o gene *Vm 26Aa* não considerando os *gaps*. LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS de $0 < \omega < 1$; ω_1 = dN/dS fixado a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ que correspondem apenas a sítios do ramo *foreground*.

M1a (modelo nulo)	$\omega_0 = 0,080$; $\omega_1 = 1,000$ $p_0 = 0,715$; $p_1 = 0,285$	-1409,441859
MA restrito (modelo nulo para os <i>Tephritidae</i>)	$\omega_0 = 0,077$; $\omega_2 = 1,000$ $p_0 = 0,641$; $p_1 = 0,194$; $(p_{2a} + p_{2b}) = 0,16447$	-1407,398623
MA (modelo alternativo para os <i>Tephritidae</i>)	$\omega_0 = 0,081$; $\omega_2 = 999,00000$ $p_0 = 0,623$; $p_1 = 0,170$; $(p_{2a} + p_{2b}) = 0,20716$	-1405,181967

Tabela A.9- Modelos evolutivos para o gene *Vm 26Aa'* considerando os *gaps*. LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS de $0 < \omega < 1$; ω_1 = dN/dS fixado a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ que correspondem apenas a sítios do ramo *foreground*.

M1a (modelo nulo)	$\omega_0 = 0,080$; $\omega_1 = 1,000$ $p_0 = 0,820$; $p_1 = 0,180$	-1065,746426
MA restrito (modelo nulo para os <i>Tephritidae</i>)	$\omega_0 = 0,076$; $\omega_2 = 1,000$ $p_0 = 0,765$; $p_1 = 0,1479$; $(p_{2a} + p_{2b}) = 0,087$	-1065,023417
MA (modelo alternativo para os <i>Tephritidae</i>)	$\omega_0 = 0,077$; $\omega_2 = 999,000$ $p_0 = 0,751$; $p_1 = 0,138$; $(p_{2a} + p_{2b}) = 0,11109$	-1064,303979

Tabela A.10- Modelos evolutivos para o gene *Vm 26Aa'* não considerando os *gaps*. LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS de $0 < \omega < 1$; ω_1 = dN/dS fixado a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ que correspondem apenas a sítios do ramo *foreground*.

Modelo	Parâmetros	LnL
M1a (modelo nulo)	$\omega_0 = 0,110$; $\omega_1 = 1,000$ $p_0 = 0,648$; $p_1 = 0,353$	-1621,408833
MA restrito (modelo nulo para as <i>Anastrepha</i>)	$\omega_0 = 0,110$; $\omega_2 = 1,000$ $p_0 = 0,62368$; $p_1 = 0,33934$; $(p_{2a} + p_{2b}) = 0,03698$	-1621,408835
MA restrito (modelo nulo para o gene <i>Vm26Aa</i> de <i>Anastrepha</i>)	$\omega_0 = 0,095$; $\omega_2 = 1,000$ $p_0 = 0,383$; $p_1 = 0,200$; $(p_{2a} + p_{2b}) = 0,416$	-1619,929681
MA restrito (modelo nulo para o gene <i>Vm26Aa'</i> de <i>Anastrepha</i>)	$\omega_0 = 0,093$; $\omega_2 = 1,000$ $p_0 = 0,462$; $p_1 = 0,204$; $(p_{2a} + p_{2b}) = 0,334$	-1619,532064
MA (modelo alternativo para as <i>Anastrepha</i>)	$\omega_0 = 0,110$; $\omega_2 = 1,000$ $p_0 = 0,595$; $p_1 = 0,324$; $(p_{2a} + p_{2b}) = 0,082$	-1621,408836
MA (modelo alternativo para o gene <i>Vm26Aa</i> de <i>Anastrepha</i>)	$\omega_0 = 0,101$; $\omega_2 = 4,559$ $p_0 = 0,564$; $p_1 = 0,265$; $(p_{2a} + p_{2b}) = 0,171$	-1619,770602
MA (modelo alternativo para o gene <i>Vm26Aa'</i> de <i>Anastrepha</i>)	$\omega_0 = 0,109$; $\omega_2 = 7,537$ $p_0 = 0,583$; $p_1 = 0,210$; $(p_{2a} + p_{2b}) = 0,207$	-1618,243570

Tabela A.11- Modelos evolutivos para o gene *Vm 26Aa* junto com *Vm 26Aa'* considerando os *gaps*. LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS de $0 < \omega < 1$; ω_1 = dN/dS fixado a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ que correspondem apenas a sítios do ramo *foreground*.

Modelo	Parâmetros	LnL
M1a (modelo nulo)	$\omega_0 = 0,106$; $\omega_1 = 1,000$ $p_0 = 0,768$; $p_1 = 0,232$	-1081,389292
MA restrito (modelo nulo para as <i>Anastrepha</i>)	$\omega_0 = 0,107$; $\omega_2 = 1,000$ $p_0 = 0,764$; $p_1 = 0,231$; $(p_{2a}+p_{2b}) = 0,004$	-1081,389292
MA restrito (modelo nulo para o gene <i>Vm26Aa</i> de <i>Anastrepha</i>)	$\omega_0 = 0,090$; $\omega_2 = 1,000$ $p_0 = 0,475$; $p_1 = 0,138$; $(p_{2a}+p_{2b}) = 0,386$	-1079,677909
MA restrito (modelo nulo para o gene <i>Vm26Aa'</i> de <i>Anastrepha</i>)	$\omega_0 = 0,093$; $\omega_2 = 1,000$ $p_0 = 0,509$; $p_1 = 0,134$; $(p_{2a}+p_{2b}) = 0,358$	-1079,113083
MA (modelo alternativo para as <i>Anastrepha</i>)	$\omega_0 = 0,106$; $\omega_2 = 1,000$ $p_0 = 0,658$; $p_1 = 0,199$; $(p_{2a}+p_{2b}) = 0,144$	-1081,389301
MA (modelo alternativo para o gene <i>Vm26Aa</i> de <i>Anastrepha</i>)	$\omega_0 = 0,092$; $\omega_2 = 4,557$ $p_0 = 0,651$; $p_1 = 0,176$; $(p_{2a}+p_{2b}) = 0,17331$	-1079,390538
MA (modelo alternativo para o gene <i>Vm26Aa'</i> de <i>Anastrepha</i>)	$\omega_0 = 0,094$; $\omega_2 = 5,492$ $p_0 = 0,675$; $p_1 = 0,169$; $(p_{2a}+p_{2b}) = 0,156$	-1078,469117

Tabela A.12- Modelos evolutivos para o gene *Vm 26Aa* junto com *Vm 26Aa'* não considerando os *gaps*. LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS de $0 < \omega < 1$; ω_1 = dN/dS fixado a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ que correspondem apenas a sítios do ramo *foreground*.

Modelo	Parâmetros	LnL
M1a (modelo nulo)	$\omega_0 = 0,089$; $\omega_1 = 1,000$ $p_0 = 0,740$; $p_1 = 0,260$	-2616,154119
MA restrito (modelo nulo para os <i>Tephritidae</i>)	$\omega_0 = 0,053$; $\omega_2 = 1,000$ $p_0 = 0,543$; $p_1 = 0,177$; $(p_{2a}+p_{2b}) = 0,281$	-2601,969691
MA restrito (modelo nulo para as <i>Anastrepha</i>)	$\omega_0 = 0,083$; $\omega_2 = 1,000$ $p_0 = 0,715$; $p_1 = 0,250$; $(p_{2a}+p_{2b}) = 0,035$	-2615,226602
MA (modelo alternativo para os <i>Tephritidae</i>)	$\omega_0 = 0,070$; $\omega_2 = 181,901$ $p_0 = 0,599$; $p_1 = 0,154$; $(p_{2a}+p_{2b}) = 0,24715$	-2593,622737
MA (modelo alternativo para as <i>Anastrepha</i>)	$\omega_0 = 0,084$; $\omega_2 = 998,92198$ $p_0 = 0,732$; $p_1 = 0,235$; $(p_{2a}+p_{2b}) = 0,033$	-2611,284234

Tabela A.13- Modelos evolutivos para o gene *Cps 38* não considerando os *gaps*. LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS de $0 < \omega < 1$; ω_1 = dN/dS fixado a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ que correspondem apenas a sítios do ramo *foreground*.

Modelo	Parâmetros	LnL
M1a (modelo nulo)	$\omega_0 = 0,096$; $\omega_1 = 1,000$ $p_0 = 0,772$; $p_1 = 0,228$	-2203,596636
MA restrito (modelo nulo para os <i>Tephritidae</i>)	$\omega_0 = 0,052$; $\omega_2 = 1,000$ $p_0 = 0,552$; $p_1 = 0,181$; $(p_{2a}+p_{2b}) = 0,26771$	-2192,139169
MA restrito (modelo nulo para as <i>Anastrepha</i>)	$\omega_0 = 0,093$; $\omega_2 = 1,000$ $p_0 = 0,760$; $p_1 = 0,220$; $(p_{2a}+p_{2b}) = 0,020$	-2203,256208
MA (modelo alternativo para os <i>Tephritidae</i>)	$\omega_0 = 0,066$; $\omega_2 = 34,362$ $p_0 = 0,622$; $p_1 = 0,173$; $(p_{2a}+p_{2b}) = 0,206$	-2187,804527
MA (modelo alternativo para as <i>Anastrepha</i>)	$\omega_0 = 0,093$; $\omega_2 = 998,98891$ $p_0 = 0,767$; $p_1 = 0,216$; $(p_{2a}+p_{2b}) = 0,017$	-2200,876669

Tabela A.14- Modelos evolutivos para o gene *Cps 38* não considerando os *gaps*. LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS de $0 < \omega < 1$; ω_1 = dN/dS fixado a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ que correspondem apenas a sítios do ramo *foreground*.

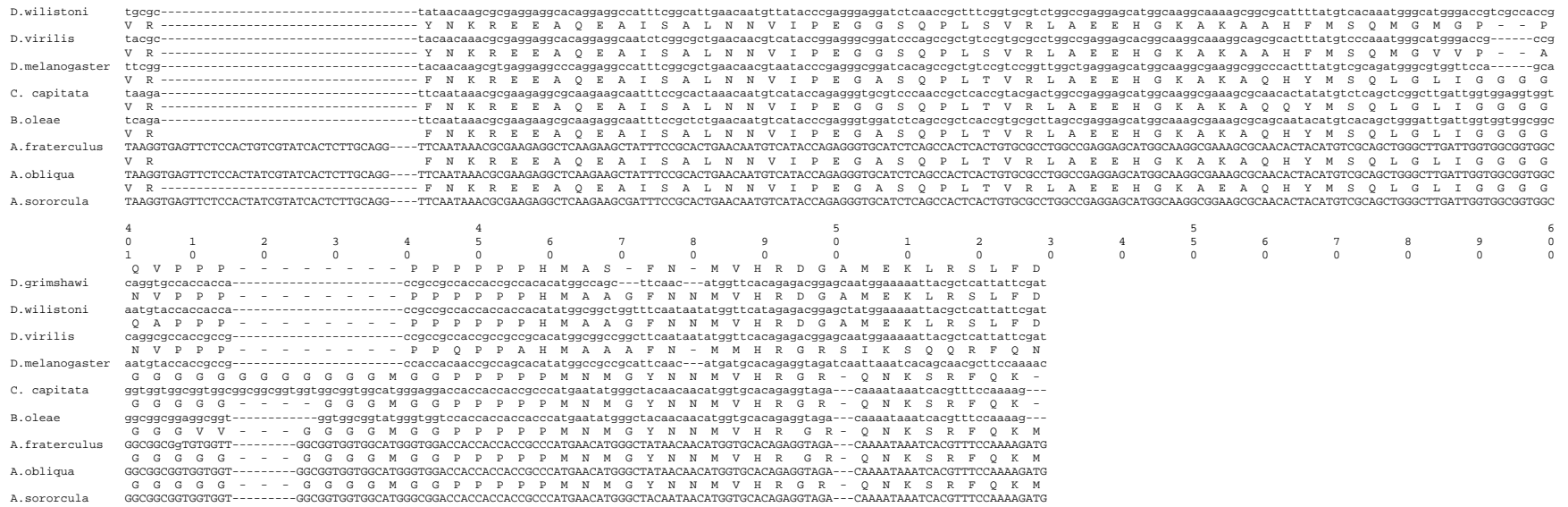


Figura A.5 Alinhamento do gene *sx/* utilizado para análises de seleção positiva realizadas no PAML.

