



**Departamento de Genética e Evolução
Centro de Ciências Biológicas e da Saúde
Universidade Federal de São Carlos**



MARCOS ROSOLINO SZRAJER

**CARACTERIZAÇÃO E ANÁLISE DA REGIÃO CODIFICADORA DO GENE
FRUITLESS EM MOSCAS-DAS-FRUTAS (*Anastrepha obliqua*)**

São Carlos / 2011



**Departamento de Genética e Evolução
Centro de Ciências Biológicas e da Saúde
Universidade Federal de São Carlos**



Laboratório de Genética de Populações e Evolução

Título da Dissertação:

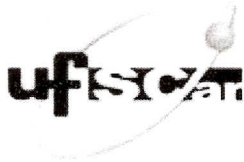
**CARACTERIZAÇÃO E ANÁLISE DA REGIÃO CODIFICADORA DO GENE
FRUITLESS EM MOSCAS-DAS-FRUTAS (*Anastrepha obliqua*)**

Dissertação apresentada ao Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular para obtenção do título de Mestre.

Aluno: Marcos Rosolino Szejzer

Orientador: Prof. Dr. Reinaldo Alves de Brito

São Carlos / 2011



UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Biológicas e da Saúde
Programa de Pós-Graduação em Genética Evolutiva e Biologia
Molecular

Relatório de Defesa de Dissertação
Candidato: Marcos Rosolino Szrajer

Aos 28/10/2011, às 14:00, realizou-se na Universidade Federal de São Carlos, nas formas e termos do Regimento Interno do Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular, a defesa de dissertação de mestrado sob o título: Caracterização e Análise da Região Codificadora do Gene Fruitless em Moscas-das-Frutas (*Anastrepha obliqua*), apresentado pelo candidato Marcos Rosolino Szrajer. Ao final dos trabalhos, a banca examinadora reuniu-se em sessão secreta para o julgamento, tendo os membros chegado ao seguinte resultado:

Participantes da Banca	Função	Instituição	Resultado
Prof. Dr. Reinaldo Otavio Alvarenga Alves de Brito	Presidente	UFSCar	<u> C </u>
Prof. Dr. Iran Malavazi	Titular	UFSCar	<u> B </u>
Profa. Dra. Zilá Luz Paulino Simões	Titular	USP	<u> B </u>

Resultado Final: B

Parecer da Comissão Julgadora*:

 APROVADO

Encerrada a sessão secreta, o presidente informou ao público presente o resultado. Nada mais havendo a tratar, a sessão foi encerrada e, para constar, eu, Ivanildes Regina de Menezes, representante do Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular, lavrei a presente ata, que assino com os demais membros da banca examinadora.

Prof. Dr. Reinaldo Otavio Alvarenga Alves de Brito

Prof. Dr. Iran Malavazi

Profa. Dra. Zilá Luz Paulino Simões

Representante do PPG: Ivanildes Regina de Menezes

() Houve () Não houve alteração no título da dissertação. O novo título passa a ser:

*Obs: Se o candidato for reprovado por algum dos membros, o preenchimento do parecer é obrigatório.

“Life is something that everyone should try at least once”.

Henry J. Tillman

“If we knew what it was we were doing,
it would not be called research, would it?”

Albert Einstein

“What would be your age
if you do not know how old you are?”

Cunfúcio

Dedico esta dissertação a todos aqueles que de alguma forma contribuíram para a formação do meu caráter como pesquisador e principalmente como pessoa, em especial meus pais Walter e Cláudia, meu irmão Álvaro, meus avós, tios, primos e amigos. Sou eternamente grato pela família que tenho e pela qual pude escolher.

AGRADECIMENTOS

Agradeço minha família pela criação, amor e apoio incondicional.

Todos os meus amigos por serem fundamentais para minha felicidade.

Meu orientador pela oportunidade, orientação, ensinamentos e amizade.

Meus companheiros de laboratório pelos momentos de trabalho e descontração.

À Universidade Federal de São Carlos, alicerce da minha formação acadêmica.

Ao Departamento de Genética e Evolução e seus pesquisadores pela infraestrutura e suporte.

Aos funcionários do Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular.

À Fundação de Amparo a Pesquisa do Estado de São Paulo pela viabilização dessa pesquisa.

1. RESUMO

As moscas-das-frutas (*Anastrepha sp*) são organismos endêmicos da região neotropical que apesar de serem semelhantes morfológicamente distinguem-se por variações comportamentais, reprodutivas e moleculares. Entre esses insetos podemos destacar o grupo *fraterculus* o qual é composto por um conjunto de espécies crípticas, entre elas *Anastrepha obliqua*, que representam ameaças à fruticultura nacional. O controle eficiente deste agrupamento exige a ampliação dos ainda escassos estudos referentes à biologia e história das populações cuja sistemática tem se mostrado bastante complexa. Trabalhos anteriores sugerem uma alta taxa evolutiva agindo sobre genes relacionados a aspectos reprodutivos e reconhecimento de parceiros devido as elevadas pressões seletivas envolvidas nessa relação, para tanto, escolhemos trabalhar com o gene *fruitless* por sua reconhecida importância na diferenciação sexual e comportamento de corte nos machos nos insetos dípteros. Uma vez que dispúnhamos de uma pequena região já amplificada deste gene em nosso laboratório, amplificamos as sequências flanqueadoras através da técnica de *RACE-PCR* que utilizou *primers* construídos e orientados para as extremidades 5' e 3'. Foram utilizadas bibliotecas de cDNAs produzidas a partir da extração de mRNAs de diferentes estágios de desenvolvimento desses organismos possibilitando a identificação de diferentes isoformas provenientes de *splincing alternativo* e caracterização estrutural da região codificadora (CDS) comum a ambos os sexos. Comparamos os dados gerados com as sequências correspondentes do gene *fruitless* em outras espécies de insetos identificando sua taxa de identidade nucleotídica e de aminoácidos as quais revelaram que os éxons que codificam domínios específicos como BTB e *Zinc Fingers* tendem a se manter mais conservados em relação às demais regiões e aos éxons da Região Conectora. Posteriormente, realizamos testes de seleção positiva conhecidos como *Relaxed Branch-site test* e *Strict Branch-site test*, os quais foram contrastados pelo *Likelihood Ratio Test* e possibilitaram a identificação de 25 novos sítios evoluindo sob seleção positiva, corroborando as descobertas de Sobrinho e de Brito (2010) e sugerindo a atuação dessas forças sobre regiões específicas deste importante gene ligado aos aspectos reprodutivos de diversos grupos principais de insetos.

ABSTRACT

The fruit flies (*Anastrepha sp*) are organisms endemic to the Neotropics that are similar morphologically but present some behavioral, reproductive and molecular variations. Among these insects we can highlight the *fraterculus group*, which consists of a set of cryptic species, between them *Anastrepha oblique*, that represents threats to national fruit culture. The effective control of this grouping requires the expansion of the few studies concerning to the biology and history of populations which has a very complex systematic. Previous works suggest a high rate of evolution of genes related to reproductive issues and recognition of partners due to the high selective pressures involved in that relationship, hence, we chose to work with the *fruitless* gene which is recognized for its role in sexual differentiation and male courtship behavior in dipteran insects. Since a small region of this gene was already been amplified in our laboratory, we amplified and described the flanking sequences by RACE-PCR technique which used *primers* constructed targeting the 5' and 3' ends. We used cDNA libraries produced from the extraction of mRNAs from different stages of development of these organisms enabling the identification of different isoforms arising from *alternative splicing* and structural characterization of the coding region (CDS) common to both sexes. We compared the data generated with the corresponding sequences of the *fruitless* gene in other insect species identifying the rate of similarity for nucleotides and amino acids (identity ratio) which revealed that the exons coding for specific domains such as *BTB* and *Zinc Fingers* tend to remain more conserved in relation to other regions and exons from *Connector Region*. Subsequently, we performed positive selection tests through the *Relaxed Branch-site test* and *Strict Branch-site test*, which were contrasted by *Likelihood Ratio Tests* (LRT) which identified 25 new sites evolving under positive selection, corroborating the findings of Sobrinho e de Brito (2010) and suggesting the presence of these forces on specific regions of this important gene linked to reproductive aspects of several major groups of insects.

KEY-WORDS: fruit flies; *Anastrepha sp*; *fraterculus group*; *fruitless*; RACE-PCR; gene characterization; identity ratio; positive selection.

LISTA DE FIGURAS

Figura 1: Exemplar de <i>Anastrepha</i> e caracteres morfológicos de diferenciação em <i>A. obliqua</i>	pág.12
Figura 2: Estrutura do gene <i>fruitless</i> em <i>D. melanogaster</i> e <i>A.gambiae</i>	pág.15
Figura 3: Arquitetura genômica dos transcritos de <i>fruitless</i> em drosofilídeos.....	pág.16
Figura 4: Arquitetura dos transcritos de <i>fruitless</i> em <i>Nasonia vitripennis</i>	pág.16
Figura 5: Sala climatizada destinada à manutenção das populações.....	pág.22
Figura 6: Expressão do gene <i>fruitless</i> ao longo do desenvolvimento dos drosofilídeos.....	pág.23
Figura 7: Resumo ilustrativo do procedimento do <i>Kit Marathon</i>	pág.27
Figura 8: Resumo ilustrativo dos procedimentos do <i>Kit First Choice RLM-RACE</i>	pág.28
Figura 9: Exemplo de placa contendo colônias recombinantes contendo fragmentos do <i>fruitless</i>	pág.31
Figura 10: Extração RNA total.....	pág.37
Figura 11: Isolamento mRNA.....	pág.37
Figura 12: Controle para diluição dos AL ds cDNAs.....	pág.37
Figura 13: Controle positivo da amplificação da região “C1” do gene <i>fruitless</i>	pág.37
Figura 14: Teste dos <i>primers</i> para 5’ e 3’ RACE.....	pág.37
Figura 15: Amplificação padronizada para ambas as extremidades.....	pág.37
Figura 16: Amplificação das colônias com <i>primers</i> M13 universais.....	pág.38
Figura 17: Exemplo de eletroferogramas provenientes do sequenciamento de colônias distintas.....	pág.40
Figura 18: Representação do gene <i>fruitless</i> identificada neste projeto para <i>A. obliqua</i>	pág.42
Figura 19: Exemplo de alinhamento no BioEdit para porção central do éxon ZnF_D.....	pág.47
Figura 20: Exemplo de alinhamento no BioEdit para porção central do éxon C3.....	pág.47
Figura 21: Árvore filogenética da região codificadora do gene <i>fruitless</i>	pág.48

LISTA DE TABELAS

Tabela 1: Taxa de similaridade nucleotídica (nível de identidade).....	pág.44
Tabela 2: Taxa de similaridade de aminoácidos (nível de identidade).....	pág.44
Tabela 3: Valores de <i>log likelihood</i> (Lnl) para os modelos de Branch-site M1a, MA e MA restrito....	pág.49
Tabela 4: <i>Likelihood Ratio Test</i> (LRT) e sítios sob seleção positiva por <i>Bayes Empirical Bayes</i> (BEB)..	pág.49

LISTA DE ANEXOS

Anexo 1: <i>Primers</i> e programas de amplificação utilizados nesse projeto.....	pág.60
Anexo 2: Alinhamento e identificação dos éxons da região codificadora do gene <i>fruitless</i>	pág.63
Anexo 3: Gráfico contendo os sítios sob seleção positiva para o gene <i>fruitless</i>	pág.67

LISTA DE ABREVIATURAS

- **AL DS cDNA (Adaptor-Ligated Double-Stranded CDNA):** cDNA de fita dupla ligado a adaptadores.
- **AP1 (Adaptor Primer 1):** Primer que se hibridiza ao adaptador do Kit Marathon.
- **BEB (Bayes Empirical Bayes):** Identifica a probabilidade de cada sítio de um gene estar evoluindo sob seleção positiva.
- **BTB (Bric-à-brac, Tramtrack and Broad complex transcription regulators):** Domínio característico do gene *fruitless* codificado por sequências presentes entre os éxons C1 e C2.
- **cDNA (Complementary DNA):** Fita complementar ao DNA.
- **CDS (CoDing Sequence):** Sequência codificadora de um gene.
- **CIP (Calf Intestine Alkaline Phosphatase):** Fosfatase Alcalina de Intestino de Bezerro.
- **dN/dS:** Taxa de substituições nucleotídicas não sinônimas por sinônimas.
- **g (gramas):** Unidade de medida de massa.
- **GSP (Gene Specific Primer):** Primers específicos para um gene.
- **Kb (Kilobase):** Unidade de comprimento equivalente a 1000 nucleotídeos (pb).
- **LRT (Likelihood Ratio Test):** Teste dos Índices de Máxima Verossimilhança que contrasta modelos evolutivos para identificação da ocorrência de seleção positiva em um gene.
- **M0 (Null Model):** Modelo nulo
- **MA (Model Alternative):** Modelo alternativo
- **mg (miligramas):** Milésima parte da unidade de massa Grama(g).
- **mL (mililitros):** Milésima parte da unidade de volume Litro(L).
- **mRNA (Messenger RiboNucleic Acid):** Ácido ribonucleico mensageiro.
- **PCR (Polymerase Chain Reaction):** Reação em cadeia pela polimerase.
- **pb (Pares de Base):** Unidade Nucleotídica.
- **RACE (Rapid Amplification of CDNA Ends):** Técnica de PCR baseada na amplificação rápida das extremidades do cDNA.
- **RT-PCR (Reverse Transcriptase-PCR):** Reação da Transcriptase reversa seguida de PCR.
- **Tampão TAE (Tampão Tris-Acetato-EDTA):** Solução tampão usada em eletroforese de ácidos nucleicos em géis de agarose.
- **Tampão TBE (Tampão Tris-Borato-EDTA):** Solução tampão usada em eletroforese de ácidos nucleicos em géis de agarose.
- **TAP (Tobacco Acid Pyrophosphatase):** Pirofosfatase Ácida de Tabaco
- **UTR (UnTranslated Region):** Região transcrita do mRNA que posteriormente não traduzidas.
- **ZnF (Zinc Fingers):** Domínio característico do gene *fruitless* codificado por sequências presentes normalmente no último éxon (extremidade 3') dos mRNAs transcritos.

SUMÁRIO

1. RESUMO.....	06
ABSTRACT.....	07
LISTA DE FIGURAS.....	08
LISTA DE TABELAS.....	09
LISTA DE ANEXOS.....	10
LISTA DE ABREVIATURAS.....	11
SUMÁRIO.....	12
2. INTRODUÇÃO.....	13
3. OBJETIVOS.....	21
4. MATERIAIS E MÉTODOS.....	22
4.1 Estratégia Amostral.....	22
4.2 Extração de RNA e Isolamento de RNA mensageiro.....	23
4.3 Precipitação e Concentração do Material.....	25
4.4 Síntese de cDNA e Ligação dos Adaptadores.....	26
4.5 Construção de <i>primers</i> e Reações de Amplificação (RACE-PCR).....	30
4.6 Clonagem e Sequenciamento.....	32
4.7 Análise das Sequências e Nível de Identidade.....	33
4.8 Reconstrução Filogenética e Testes de Seleção.....	34
5. RESULTADOS E DISCUSSÃO.....	36
5.1 Padronização dos Procedimentos.....	36
5.2 Padronização das Amplificações.....	38
5.3 Sequências obtidas e Alinhamentos.....	40
5.4 Organização Estrutural do gene <i>fruitless</i> em <i>Anastrepha</i>	42
5.5 Identidade entre as Sequências.....	44
5.6 Seleção Positiva sobre o gene <i>fruitless</i>	49
5.7 O gene <i>fruitless</i> como Marcador Molecular.....	51
6. CONCLUSÕES.....	53
7. REFERÊNCIAS BIBLIOGRÁFICAS.....	57
ANEXOS 1.....	62
ANEXOS 2.....	65
ANEXOS 3.....	69

2. INTRODUÇÃO

O gênero *Anastrepha* (Diptera, Tephritidae) oriundo da região Neotropical e encontrado no sul dos EUA, México, América Central e quase toda América do Sul (Malavasi e cols, 2000) representa uma das maiores ameaças à fruticultura nacional uma vez que são consideradas oligotróficas e depositam seus ovos nas culturas verdes ou maduras indistintamente, culminando no consumo do endocarpo da fruta por parte das larvas (Zucchi, 2000) e acelerando o processo de degradação da mesma pela exposição à fungos, bactérias e outros patógenos (Carvalho, 2005). O grupo *fraterculus* (Zucchi, 1978; Norrbom e Kim, 1988) faz parte desse gênero e é composto por um conjunto de espécies, algumas crípticas, de Moscas-das-Frutas que compartilham características fisiológicas, ecológicas e comportamentais, do qual fazem parte *A. fraterculus*, *A. obliqua*, *A. sororcula* e *A. zenildae*, nossos objetos de estudo.

Apesar da grande similaridade interespecífica, alguns marcadores morfológicos (Zucchi, 2000), comportamentais (Morgante e cols., 1980), reprodutivos (Selivon, 2000), alozímicos (Morgante e cols., 1980), cariotípicos (Solferini e Morgante, 1987, Selivon e cols., 2005) e de mtDNA (de Brito e Matioli, dados não publicados) já foram descritos na literatura. Usualmente em laboratório utilizamos caracteres morfológicos que agrupados fornecem uma classificação eficiente (Araujo e Zucchi, 2006) como as os padrões alares de nervuras, manchas presentes no sub-escutelo (Malavasi e Zucchi, 2000), a forma do ápice do ovipositor (para diferenciação de fêmeas) e a morfologia do ovo (Selivon, 2000) (Figura 1).

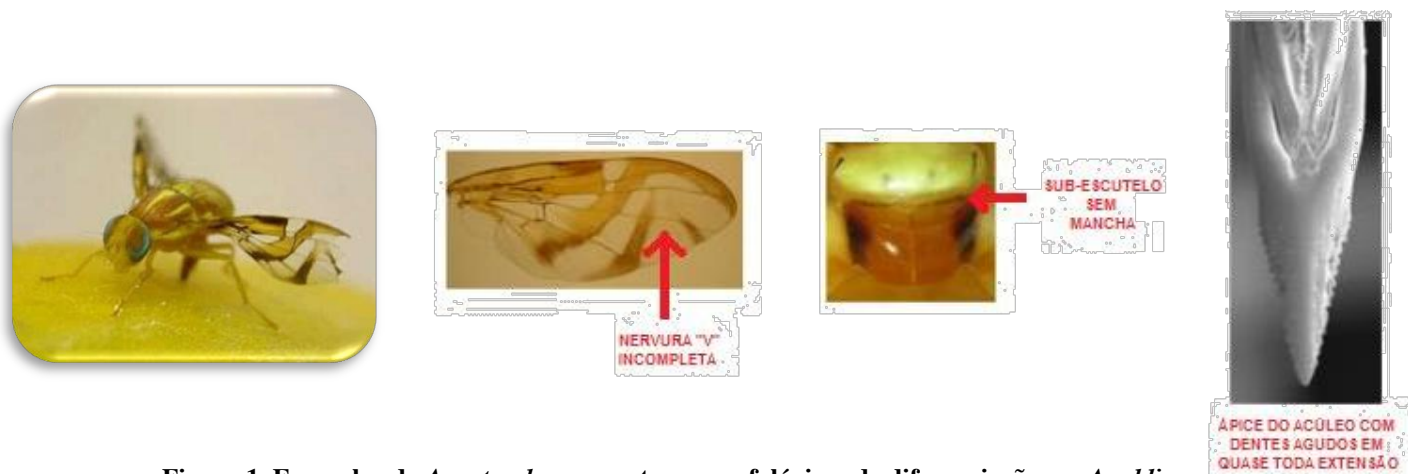


Figura 1. Exemplo de *Anastrepha* e caracteres morfológicos de diferenciação em *A. obliqua*.

Entretanto, nenhum desses estudos forneceu até o momento ao menos um marcador confiável que permita diferenciar as espécies do grupo *fraterculus*. Aliado a isso, a produção de híbridos viáveis, suas características reprodutivas e o fato de aparentemente terem divergido há pouco tempo favorecem a utilização desses “insetos-praga” como modelo para o entendimento de fatores envolvidos na especiação e diferenciação em grupos de espécies, o que promete fornecer importantes informações relativas à biologia evolutiva desse gênero (Resende e cols, 2009).

Baseados na grande variação genética encontradas entre as espécies do grupo *fraterculus* e a confirmação da presença de regiões evoluindo por seleção positiva em estudos realizados em nosso laboratório envolvendo os genes candidatos *doublesex* (Sobrinho e de Brito, 2006) e *fruitless* (Sobrinho e de Brito, 2010), buscamos nessa pesquisa obter e estudar em detalhes a estrutura codificante deste último, responsável pela diferenciação sexual dos machos e regulação do comportamento de corte, tornando possível acessar a maior parte dos éxons que compõem as diferentes regiões sujeitas às pressões de seleção positiva sobre esse gene como um todo, corroborando os estudos anteriores antes restritos a uma pequena porção do mesmo.

Em *Drosophila*, alguns genes participam ativamente da cascata de eventos diretamente responsável pela determinação sexual. Esse processo é controlado por um “interruptor” primário, *sex-lethal* (*sxl*). Os transcritos de *sxl* são processados diferentemente em cada sexo por meio de um *splicing* alternativo no mRNA. Em fêmeas, o produto do *sxl* (*sxlF*) regula o *splicing* diferencial de *transformer* (*tra*), que por sua vez, em conjunto com *transformer-2* (*tra-2*), direciona o *splicing* diferencial de *doublesex* (*dsx*), que está no fim desta cadeia e age em vários genes que determinam o fenótipo feminino (Sosnowski, 1994; Greenspan e Ferveur, 2000). Em machos, *sxl*, *tra* e *tra-2* estão inativos, e o mRNA do gene *dsx* produz uma proteína específica, DSX^M que ativa o gene *fruitless*, responsável pelo desenvolvimento dos atributos do fenótipo masculino (Burtis, 1989 e 1993; Taylor, 1994). Recentemente, estudos comprovaram essa conservação funcional da cascata também em *Anastrepha*

após verificar a indução de características parciais femininas em machos XY que tiveram as proteínas codificadas pelo gene *transformer* expressadas através do sistema GAL4-UAS (Ruiz e Sanchez, 2010).

Encontrado próximo ao final da hierarquia de determinação sexual, este gene é caracterizado como o primeiro desta cadeia a atuar especificamente no sistema nervoso central evidenciado por transcritos de *fru* em processos envolvidos nos aspectos neurais de processamento de informações quimiosensoriais ocorrendo no lobo antenal e neurônios cerebrais envolvidos no dimorfismo sexual por seus padrões quantitativos e de projeção distinta (Ryner, 1996; Kimura, 2005). É responsável também pelo desenvolvimento do par de músculos localizados no quinto segmento abdominal dos mesmos, o chamado “Músculos de Lawrence” (MOL) (Lawrence 1986), que ocorre principalmente pela mobilização de mioblastos para os miotúbulos dessa região (Taylor, 1995) e pela regulação do comportamento de corte realizado pelos machos (revisado por Hall, 1994).

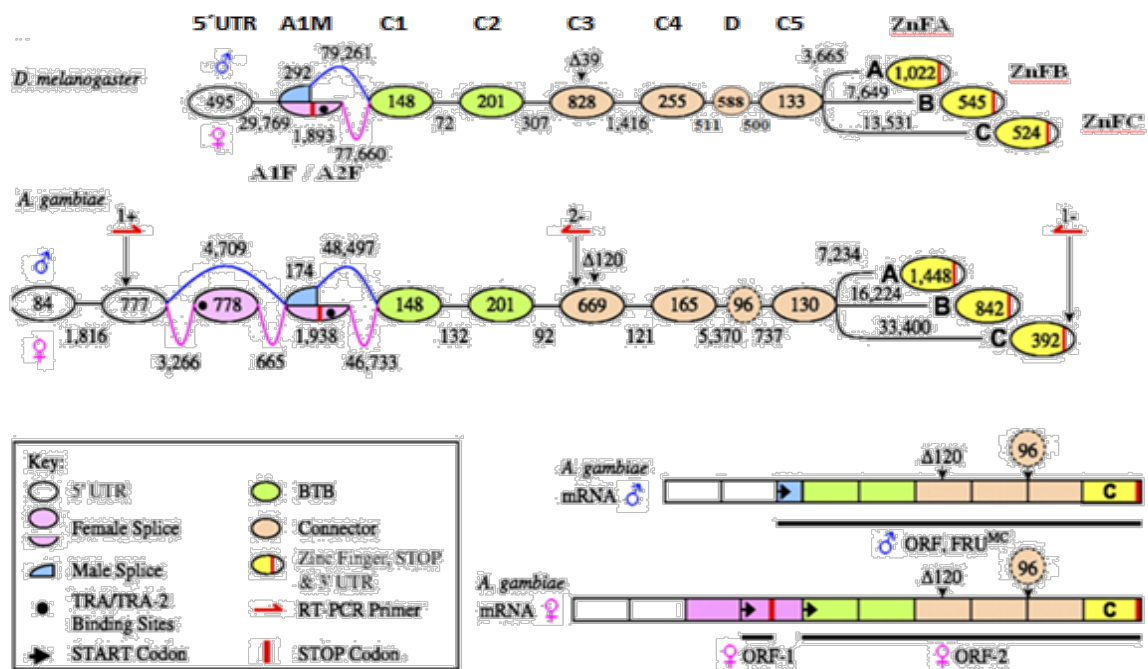


Fig.2 Estrutura do gene *fruitless* em *D. melanogaster* e *A.gambiae*. Os círculos indicam éxons, as linhas íntrons e os números suas extensões. Nomenclatura: A1M (Alternative Sex-Specifically Spliced Male) específico do macho, A1F e A2F (Alternative Sex-Specifically Spliced Female) indicam o 1º e o 2º éxons específicos da fêmea. Os éxons C1 e C2 (BTBs), C3,C4 e C5 (Conectores) estão marcados em verde e laranja, respectivamente, e são comuns em ambos os sexos. Éxons provenientes de *splicing* alternativo independente do sexo da extremidade 3' estão representados em amarelo (Adaptado de Gailey e cols., 2006).

É pouco provável que o gene *fruitless* não esteja desempenhando em *Anastrepha sp* função similar a que desempenha em *Drosophila* uma vez que todo o resto da cadeia de determinação sexual está, à exceção do gene *sxl* (Alvarez e cols, 2009). Outro indício dessa conservação remete ao ortólogo de *fru* também em *Anopheles gambiae*, no qual a isoforma da proteína *FRU* específica do macho, surge pela ativação de mecanismos conservados específicos ao sexo e *splicing alternativo* de éxons (Figura 2).

Um dos pontos interessantes dessa pesquisa está em averiguar se a grande variabilidade do gene *fruitless* e a recorrência do processo de *splicing* diferencial já descrito para outros grupos de insetos como *Drosophila heteroneura* (Davis e cols, 2000), *Drosophila melanogaster* (Demir e Dickson, 2005; Shirangi e McKeown, 2007), *Anopheles gambiae* (Gailey e cols, 2006) e *Nasonia vitripennis* (Bertossa, 2009), também se verificam em *Anastrepha*.

No caso dos drosofilídeos (Figura 3), diferentemente de *Nasonia* (Figura 4), o gene *fruitless* tem seu *splicing alternativo* dependente do sexo a partir do promotor P1, o qual é usado para interromper a tradução do transcrito da fêmea pela ligação do complexo TRA/TRA-2, enquanto que o produto masculino apresenta-se funcional (Usui-Aoki e cols, 2000). Nos transcritos específicos de cada sexo o éxon chamado de “S” aparece como primeiro domínio, enquanto que nos transcritos independentes do sexo do indivíduo o primeiro domínio é composto pelo domínio BTB. Este por sua vez, dependendo da espécie, pode ser composto por um ou dois éxons (C1 e C2). Da mesma forma, a região conectora subsequente será composta por dois ou três éxons (C3 e C4 ou C3, C4 e C5). O domínio de ligação ao DNA localizado na extremidade 3’ é composto por um de quatro éxons diferentes (A, B, C ou D), selecionados por *splicing diferencial*. Existem ainda as formas que independem do sexo e são expressas por três promotores distintos (P2, P3 e P4) a jusante do éxon inicial “S”, o que corrobora o padrão de expressão bastante complexo e variável de acordo com tecido e estágio de desenvolvimento (Anand e cols, 2001; Song e cols, 2002) (Figuras 3 e 4).

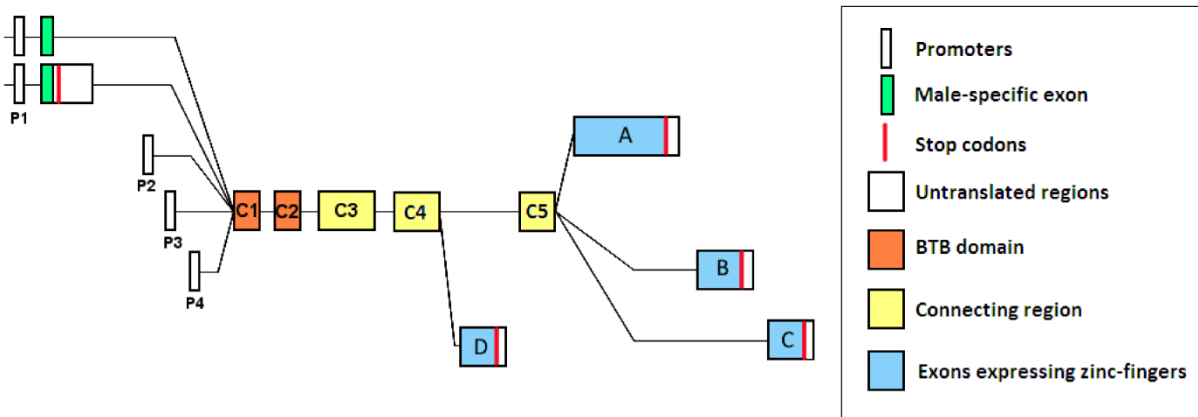


Fig.3 Arquitetura genômica dos transcritos de *fruitless* em drosofilídeos e grande parte das linhagens de insetos dípteros compostos por quatro regiões principais: domínios específicos aos sexos, região de dimerização (BTB), região conectora e domínio de ligação ao DNA que expressam Zinc-fingers (adaptado de Sobrinho e de Brito, 2010).

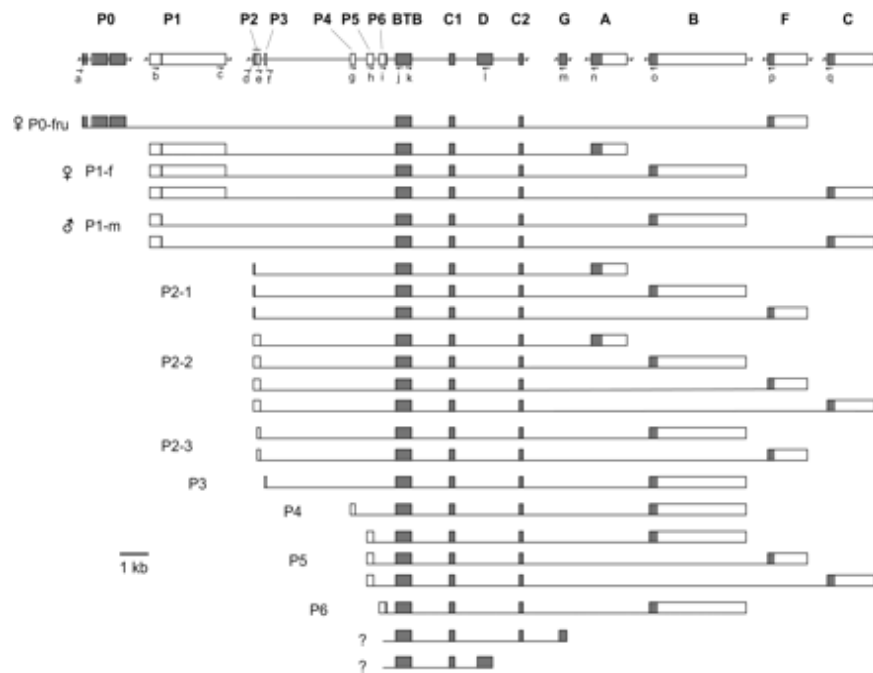


Fig.4 Arquitetura dos transcritos de *fruitless* em *Nasonia vitripennis*. No alto, as caixas representam os éxons em escala de comprimento e as linhas os íntrons. Abaixo, todos os transcritos encontrados em bibliotecas de cDNA. Cada isoforma, seja ela sexo específica (como as 6 primeiras) ou não, começa com um dos 7 éxons diferentes da extremidade 5' (P0–P6), seguidos da região referente ao BTB. Os éxons C1 e C2 (conectores e correspondentes aos éxons C3 e C5, respectivamente nas demais espécies), ligam o BTB a um dos 5 tipos diferenciais de éxons codificantes C2H2 ZincFinger (G, A, B, F e C). Amplificações realizadas por (RT-PCR) revelaram ainda 2 outros possíveis transcritos ZnF (D e G), os quais se iniciam diretamente no éxon BTB e são ainda pouco conhecidos (Adaptado de Bertosa e cols, 2009).

Essas características somadas acabam por dificultar o processo de identificação estrutural e a organização dos dados necessários para realização dos testes padrões de seleção positiva o que retarda o

processo de identificação de homologias para as novas sequências obtidas para *Anastrepha*, entretanto, a comparação destas sequências possibilita identificar pressões seletivas atuantes ao longo do tempo evolutivo sobre cada um dos sítios analisados, o que pode não somente confirmar, mas também expandir em grande quantidade os dados já encontrados em parte desse gene (Sobrinho e de Brito, 2010; Szrajner e de Brito, dados não publicados).

Visando obter tais dados, lançamos mão de PCRs convencionais aliadas a técnica de RACE-PCR. Esta última tem por finalidade permitir a obtenção da sequência completa do cDNA para mRNAs extraídos de nossas moscas nos quais apenas parte da sequência é obtida, através da amplificação separadamente das extremidades 5' e 3' flangeadoras ao gene com identificação parcial. Estas estratégias são geralmente referidas como 5'-RACE (*Rapid Amplification of cDNA Ends*) e 3'-RACE. A utilização dessa variação da técnica de PCR consiste basicamente na realização inicial de uma reação de amplificação tradicional que forneça sequências confiáveis que permitam a construção de oligonucleotídeos (*GSP-Gene Specific Primers* ou primers específicos para o gene) que hibridizam em ambas as extremidades dessa sequência e são orientados para as duas direções (5' e 3'). Tais *primers* são então utilizados em reações utilizando como molde mRNAs ou bibliotecas de cDNA, estendendo parcialmente os mRNAs partindo da região conhecida em direção à região final desconhecida, usando de *primers* que hibridizam a cauda poli(A) pré-existente na extremidade 3' ou a uma cauda de homopolímero “plugado” ou ainda a um *linker* na região terminal 5'. Os produtos de PCR gerados, que contém regiões terminais desconhecidas do cDNA, podem ser clonados e caracterizados usando as demais técnicas convencionais (Frohman e cols, 1988).

Em geral, a produção de clones parciais das extremidades 3' são mais simples em virtude da maioria dos mRNAs de eucariotos possuir uma cauda poli-A nesta extremidade. Isso facilita tanto a ancoragem dos adaptadores, quanto uma forma alternativa de amplificação que pode ser alcançada com o uso de *primers* poli-T, que se ligam à extremidade 3' dos cDNAs, e *primers* derivados do gene de

interesse. Neste caso, uma segunda rodada de ciclos de amplificações é recomendada usando *primers* aninhados para reduzir a produção de artefatos não-específicos.

Estratégias diferentes são usadas para gerar clones parciais da extremidade 5', que é de mais difícil obtenção. Neste caso, pode-se incorporar uma cauda poli(A) à extremidade 5' utilizando-se de uma desoxinucleotiltransferase terminal (TDT) e dATP. Alternativamente, pode-se optar por inserir sequências de DNA ligadoras (*linkers*) que são também incorporadas na região 5'. Em qualquer dos casos, *primers* complementares a estas sequências incorporadas são utilizadas em conjunto com o *primer específico* para o gene visando gerar o produto de PCR que contenha regiões desconhecidas do cDNA à montante. Da mesma forma que nas estratégias 3', uma segunda rodada de ciclos de PCR é feita usando *primers aninhados* tendo por finalidade aumentar o rendimento dos produtos específicos pretendidos (Frohman e Martin, 1989).

As informações geradas das RACE-PCR quando somadas às fornecidas pelas PCR tradicionais permitem o sequenciamento completo de genes de interesse, sendo particularmente relevantes em estratégias que se utilizam da amplificação de regiões conservadas em diferentes espécies para a criação de *primers degenerados* que são utilizados em PCR heterospecífica (Rose e cols, 1998; Fietto e cols, 2002). Nestes casos, é comum que parte significativa do gene, em geral a mais variável entre espécies distintas, não seja amplificada e permaneça desconhecida. As informações obtidas através destes procedimentos permitem dessa forma, não apenas a obtenção destas sequências desconhecidas, mas também o conhecimento das regiões que se estendem além das regiões codificadoras, os locais de iniciação do processo transcricional, bem como a determinação de regiões de atuação de possíveis *enhancers* (Scotto-Lavino e cols, 2006).

Utilizando como ponto de partida o fragmento sequenciado por Sobrinho e de Brito (2010), ou seja, a porção final do éxon "C2" (segundo éxon da Região BTB) e região inicial do éxon "C3" (primeiro éxon da Região Conectora) (Figura 3), correspondentes respectivamente aos éxons C2 e C3 de

drosophila sp (Gailey e cols, 2006), buscamos nessa dissertação amplificar as regiões flanqueadoras e unir esses dados descrevendo as regiões codificantes ainda desconhecidas de *fruitless* partindo de sua extremidades 5' e finalizando na extremidade 3' (CDS).

3. OBJETIVOS

Objetivo Geral:

Identificar a estrutura da região codificadora (CDS) do gene *fruitless* a partir do sequenciamento de cDNA amplificados através de RACE-PCR em espécies do grupo *fraterculus*, fornecendo subsídio para investigar o potencial deste gene como marcador em estudos da diferenciação e especiação deste grupo de moscas-das-frutas.

Objetivos Específicos:

1. Padronização e otimização das extrações de RNAs advindas de diferentes fases do desenvolvimento.
2. Estabelecimento de *primers* que amplifiquem as regiões à montante e à jusante da região conhecida.
3. Testar a aplicabilidade da técnica de RACE-PCR para a obtenção de sequências completas deste gene.
4. Verificar as sequências obtidas visando identificar similaridades destas com espécies próximas.
5. Organizar uma estrutura linear da região codificadora contendo as novas regiões descobertas.
6. Investigar os transcritos do gene *fruitless* expressos em *Anastrepha sp.*
7. Comparar a estrutura completa do gene *fruitless* e cada um de seus éxons separadamente com o seu correspondente em espécies relacionadas, avaliando sua taxa de similaridade interespecífica (identidade nucleotídica e de aminoácidos).
8. Testar se o gene *fruitless* tem evoluído de maneira neutra ou sob seleção positiva em *Anastrepha* analisando longos períodos evolutivos (*Long-term Evolutionary Analysis*) baseados em variações ao nível protéico, identificando eventuais sítios sujeitos a pressões de seleção.

4. MATERIAIS E MÉTODOS

4.1. Estratégia Amostral

A coleta das populações de *Anastrepha sp* na natureza foram realizadas pela coleta de frutos *in natura* de diversas regiões do Brasil, ocorrendo principalmente em localidades com alguma interferência antrópica e dando-se prioridade a mangas e goiabas maduras. Tanto os frutos recolhidos quanto as moscas originárias destes, foram mantidos em “gaiolas” distintas em ambiente com temperatura e iluminação controladas (Figura 5) durante todo o processo de eclosão, maturação sexual e acasalamento. As populações foram mantidas sob dieta de 10 ml de mel, 30 ml de hidrolisado de proteína de milho, 0,5 ml de Levenil, 6g de levedo de cerveja, 0,5g de Sustagen, 25g de açúcar mascavo e 45 g de açúcar branco, utilizando frutas (usualmente goiabas) tratadas com Nipagin (Metilparabeno: conservante farmacêutico) como substrato para oviposição.

Para este estudo foram utilizadas tanto populações já existentes em nosso laboratório quanto novas populações, já identificadas, originárias de diversos pontos do Brasil derivadas de frutas das novas coletas e mantidas no Laboratório de Genética de Populações e Evolução, realizadas no período a que se refere esse estudo, entre elas: Araras (SP), Estrela-do-Oeste (SP), São Carlos (SP), Ribeirão Preto (SP), Itabira (MG), Uberlândia (MG), Búzios (RJ), Araguatins (TO), Goiânia (GO), Porto Alegre (RS), entre outras. As frutas foram coletadas e transportadas até nosso laboratório, localizado no Departamento de Genética e Evolução (DGE) da UFSCar, embrulhadas em jornal e posteriormente acondicionadas em caixas plásticas com vermiculita até que as pupas emergissem e fossem transferidas para caixas acrílicas de 50 x 50 x 40 cm. As moscas nascidas foram separadas por sexo e identificadas até 24 horas após a eclosão.

Decorrente da elevada disponibilidade de populações obtidas em relação às demais e os bons resultados iniciais obtidos, *A. obliqua* foi primordialmente escolhida para a maioria dos experimentos aos quais faziam parte das populações de Estrela-do-Oeste (SP), São Luís (MA) e Passos (MG),

contribuindo com três exemplares jovens e vivos cada uma, sendo eles do mesmo sexo e espécie para composição do *pool* de indivíduos (9 moscas). Estes foram sacrificados apenas no momento da extração do material genético. Em alguns experimentos de confirmação utilizamos *A. fraterculus*, enquanto que *A. sororcula* e *A. zenilidae*, por sua restrição populacional, não chegaram a ser testadas.



Figura 5. Sala climatizada de nosso laboratório destinada à manutenção das populações.

A formação de *pools* se fez necessária em decorrência do grande aporte inicial de mRNA exigido pelo *Kit Marathon* (Clontech), escolhido para iniciar os experimentos de RACE-PCR. Em uma segunda etapa passamos a utilizar o *Kit First Choice RLM-RACE* (Ambion), que possibilitou que as extrações subsequentes fossem realizadas a partir de uma única mosca adulta, uma única pupa e uma única larva.

4.2. Extração de RNA e Isolamento de RNA mensageiro

Como mencionado, escolhemos moscas da espécie *A. obliqua* para formar o primeiro *pool* de machos já nascidos uma vez que *pools* diferentes para cada sexo são importantes pela comprovada existência de *splicings* alternativos e por apresentarem níveis de expressão distintos dependendo do sexo

do indivíduo e estágio dos indivíduos. Para essa escolha optamos por nos basear em dados dos momentos de maior expressão desse gene disponíveis para machos e fêmeas de *Drosophila* (Figura 6).

Sendo assim, foram utilizadas moscas recém-nascidas ou ainda jovens (entre zero e dez dias de vida). Para as amostras individualizadas que não formariam *pools* (*Kit First Choice RLM-RACE*) estas moscas também foram escolhidas respeitando essa característica, enquanto que também foram selecionadas nesse segundo momento larvas e pupas (fase de metamorfose) em diversos períodos.

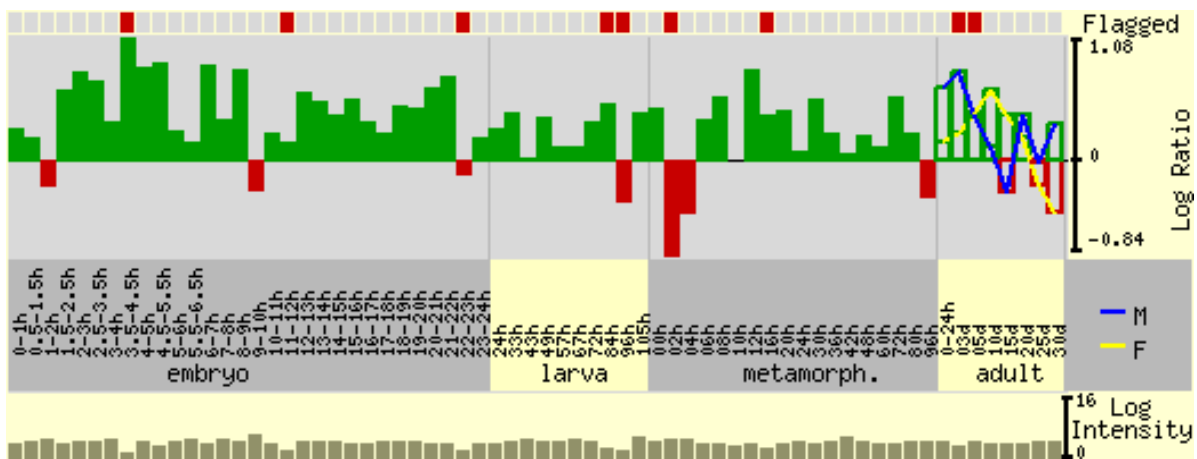


Figura 6. Expressão do gene *fruitless* ao longo do desenvolvimento de machos e fêmeas de *Drosophila* (fonte: FlyBase)

As moscas recém-nascidas escolhidas foram mantidas em gaiolas apenas com água *overnight* visando reduzir resíduos alimentares nas amostras, tendo seu RNA total extraído e isolado através da maceração de três moscas inteiras em cada tubo utilizando nitrogênio líquido com posterior extração com Trizol (Invitrogen) - uma combinação de Fenol/Tiocianato de Guanidina/Clorofórmio que deriva da melhoria do método de isolamento de RNA em um único passo (Chomczynski e Sacchi, 1987) - de acordo com especificações do fabricante. Já no caso das larvas e pupas, estas foram selecionadas, “lavadas” em solução salina para retirada de impurezas e submetidas ao mesmo procedimento. O Trizol tem a finalidade de manter a integridade do RNA enquanto rompe as células e dissolve os componentes celulares. Seu protocolo passou por diversos testes e ajustes tendo em vista o melhor aproveitamento do material utilizado destacando entre elas: a realização de todo o procedimento dentro da câmara de fluxo

laminar; adição das moscas ainda vivas ao nitrogênio líquido, seguido da aplicação do TRIZOL o mais rápido possível antes da maceração; retirada cuidadosa dos resíduos de etanol com o auxílio de uma pipeta e ponteira com filtro antes de proceder a ressuspensão do material em água estéril.

Todas as quantificações e verificações de integridade dos produtos ao longo de todos os procedimentos desse projeto foram realizada tanto em gel de agarose (EtBr 1,2%) quanto no aparelho NanoDrop 1000 (Thermo Scientific) generosamente disponibilizado pelo Prof. Dr. Flávio Henrique da Silva de nosso departamento. Sendo assim, a qualidade do RNA total extraído foi avaliada pela relação de absorbância 260/280 nm neste aparelho e sua integridade foi testada por eletroforese em gel de agarose não-desnaturante 1% com tampão TAE livre de RNase, tendo sido corado com brometo de etídeo e visualizado em luz ultravioleta, que permite a distinção das bandas do RNAs ribossômicos 18S e 28S. Nessas condições o RNA não migra precisamente de acordo com seu tamanho, entretanto é possível visualizar alguma separação de RNA sendo possível identificar inclusive o RNA ribossomal. Todas as alíquotas de RNA total extraídas agrupadas em *pools* foram utilizadas posteriormente para o isolamento do RNA mensageiro (mRNA) através do uso do *kit* PolyATract mRNA Isolation Systems III (Promega), bem como as amostras individuais. Este *kit* usa de oligos(dT) biotinizados ligados a microesferas 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 foram posteriormente ressuspensos em 100 µl de água MiliQ. A qualidade do mRNA precipitado foi novamente observada através da eletroforese em gel de agarose e quantificado tanto no NanoDrop quanto em Biofotômetro Eppendorf, localizados nos laboratórios dos professores Dr. Flávio Henrique da Silva e Dr. Pedro Galetti, ambos de nosso departamento.

4.3. Precipitação e Concentração do material

Após obtenção de uma quantidade elevada de RNAm procedemos para a realização das reações de precipitação e concentração, quando necessário (no caso do *kit Marathon*), com Acetato de Amônio e Isopropanol (modificado de Sambrook e Russell, 2001), seguida por duas lavagens com etanol 80% e rápida secagem a temperatura ambiente. A metodologia de precipitação e concentração também teve de ser ajustada uma vez que os primeiros testes seguindo o protocolo original mostraram uma perda de aproximadamente 80% de RNA mensageiro (apenas 0,7 µg concentrados a partir de 4 µg iniciais). Foram feitos alguns experimentos visando à otimização dessas reações e observamos que as seguintes alterações produziram melhores resultados: reduzir o tempo de “overnight” a - 20°C para que ocorra a ressuspensão do RNA mensageiro recém-isolado (no máximo 12 horas) antes de concentrá-lo, acelerar as secagens utilizando cuidadosamente a pipeta com uma ponteira que possua filtro para retirar manualmente os resíduos de acetato/isopropanol e etanol, aumentar o tempo de centrifugação utilizada e manter por um maior tempo (2-3 horas) o material na geladeira a 4°C para uma boa ressuspensão em água livre de RNAses antes de iniciar sua transformação em cDNA.

No caso das amostras a serem utilizadas no *Kit First Choice RLM-RACE*, essa etapa não foi necessária por sua eficácia satisfatória com uma menor quantidade de material.

4.4. Síntese de cDNA e Ligação dos Adaptadores

A produção do cDNA utilizando o *kit marathon* foi feita a partir de 1 µg de RNAm (poly A⁺) concentrado em até 4µl de solução, em que sua primeira fita é sintetizada combinando-se a ação da AMV transcriptase reversa e *primers* específicos para síntese de cDNA, seguindo-se o método proposto por Borson et al. (1992) e Chenchik et al. (1995). Já a síntese da fita complementar é realizada de acordo com o método de Gubler e Hoffmann (1983) que utiliza uma solução contendo DNA Polimerase I e DNA Ligase isoladas de *E.coli*, somadas a RNase H, responsável também pela degradação do RNA remanescente da etapa anterior. Utilizando o *blunt end* ou “final-reto” gerado pela ação da T4 DNA

Ligase, o DS cDNA (Double-stranded cDNA) é ligado aos adaptadores do *kit* Marathon (Clontech) que apresentam uma dupla-fita parcial e fosforilações em suas extremidades 5' que facilitam sua ligação aos *blunt ends* de ambos os lados do cDNA.

Para confirmar a obtenção de DS cDNAs foi realizada uma PCR tradicional que amplifica um fragmento conhecido do gene que estava sendo expresso no tecido extraído. Em nosso caso, utilizamos *primers* capazes de amplificar um fragmento de aproximadamente 850 pb referente a região conectora do gene em questão (Sobrinho e de Brito,2010). Já a constatação da ligação destes adaptadores ao DS cDNA, ou seja, a existência de AL DS cDNA (Adaptor-Ligated Double-Stranded cDNA), é geralmente feita pela amplificação do produto ligado utilizando-se um *primer* interno, específico ao gene (GSP), e outro específico do adaptador(AP1 *primer*). A padronização desta amplificação foi feita utilizando-se de um controle positivo também composto por RNAm (poly A+) proveniente de material placentário humano e *primers* específicos para este gene fornecidos juntamente com o *kit*. Estas amplificações são então comparadas para estabelecer a concentração relativa do pool de AL DS cDNAs para que possamos fazer a diluição adequada baseando-se no padrão esperado para o controle positivo. A diluição padrão para o controle fornecido pelo *kit* é de 250 µl de Tampão Tricina-EDTA para cada µl de AL DS cDNA, dessa forma se a amostra for de comprimento e nitidez semelhante ou superior deve-se usar a mesma diluição, caso contrário (seja inferior ao controle) deve-se diluir em uma quantidade proporcionalmente menor desse tampão. Em nosso caso testamos ainda diluições distintas de 50, 100 e 250 µl. Estas diluições servem tanto para utilização para 5' quanto 3'RACE, dependendo do par de *primers* utilizado na reação (Figura 7).

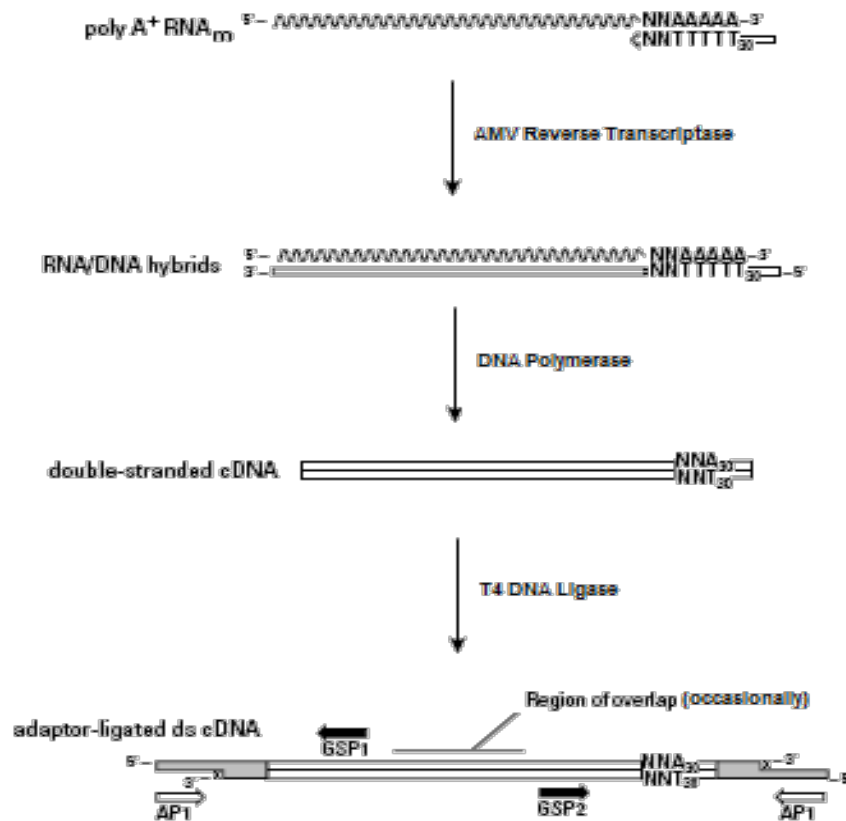


Fig.7 Resumo ilustrativo do procedimento do *Kit Marathon*

Para o *Kit First Choice RLM-RACE*, as reações de produção de cDNA e a ligação de adaptadores para posteriores 5'RACE e 3'RACE utilizam metodologias diferentes e requerem quantidades também distintas de material para a síntese de cDNA e ligação de adaptadores (Figura 8).

Foram utilizados 250ng de mRNA (poly A⁺) para a reação 5' a qual consiste da utilização de um reagente denominada CIP (Calf Intestine Alkaline Phosphatase) responsável por remover fosfatos 5' livres presentes das moléculas de rRNAs, tRNAs, mRNAs degradados e e/ou DNA genômicos contaminantes. As estruturas CAP intactas encontradas nas extremidades 5' dos mRNAs de boa qualidade não são afetadas. A seguir o material é submetido a um tratamento com TAP (Tobacco Acid Pyrophosphatase) para remover a estrutura de CAP de fitas completas de mRNA, deixando um 5'-monofosfato. Um oligonucleotídeo adaptador de RNA com 45pb é ligada à população de mRNAs com

uma T4 RNA ligase. O adaptador não pode ligar-se a RNAs desfosforilados porque nessas moléculas o 5'-fosfato necessárias para a ligação está ausente. Durante a ligação, a maior parte das fitas completas de mRNA apresenta-se sem CAP e acabam por adquirir o adaptador em sua extremidade 5'. Nesta etapa, somos capazes de realizar uma transcrição reversa seguida de uma 5'RACE-PCR com *primers* internos específicos sendo combinados ao *primer* complementar ao adaptador e orientado no sentido 3'.

Para produzir o cDNA a ser usado nas reações da extremidade 3', uma menor quantidade de mRNA(poly A⁺) é necessária, tendo sido utilizados nesses casos 50ng mRNA(poly A⁺) para início da síntese da primeira fita de cDNA. Esta é feita utilizando juntamente com a transcriptase reversa um adaptador 3'RACE fornecido com o *kit*. Em seguida o material está pronto para realização de uma 3'RACE-PCR.

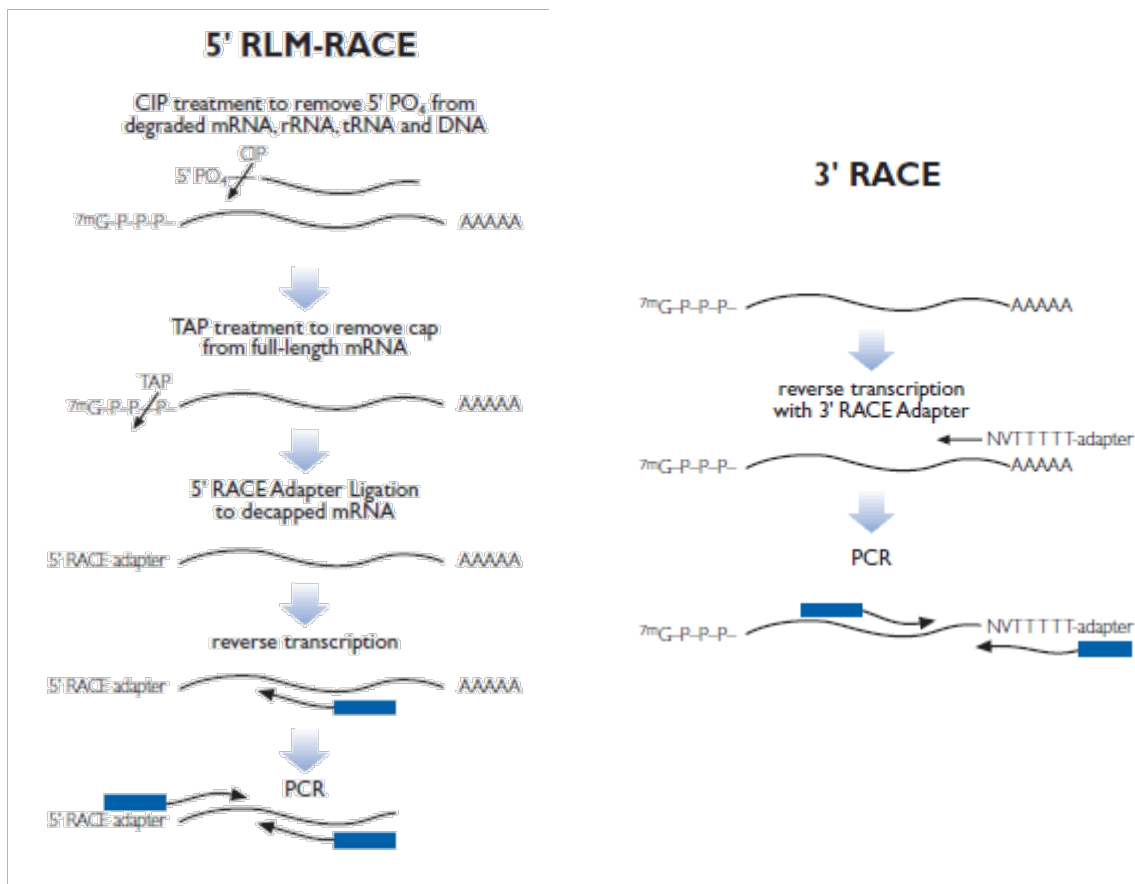


Fig.8 Resumo ilustrativo dos dois procedimentos distintos para o *Kit First Choice RLM-RACE*.

4.5. Construção de *primers* e Reações de Amplificação (RACE-PCR)

Baseando-se em 100 sequências (Sobrinho Jr. e de Brito, 2010) obtidas para diferentes espécies de *Anastrepha sp* de 850 pares de bases relativos a região codificadora comum aos sexos do gene *fruitless*, construímos um conjunto de *primers* através do software *Primer3* (Rozen e Skaletsky, 2000) e da ferramenta *OligoAnalyser 3.0* disponível no site (www.idtdna.com/analyser/Default.aspx/oligocalc.asp) que nos permitiu a amplificação das regiões 5' e 3' desse fragmento. Analisamos as estimativas das temperaturas de hibridização, valores de ΔG , possibilidade de formação de estruturas secundárias, chances de formação de dímeros tanto entre os mesmos *primers* (*selfdimer*) como entre estes e os *primers* do kit utilizado (*heterodimer*). Dessa forma, reduzimos a produção de artefatos nas reações de amplificação. O elevado número de sequências de alta qualidade revelou conservação interespecífica de algumas regiões proporcionando *primers* bastante específicos com poucas degenerações devido as suas altas afinidades às regiões-alvo. Desenhamos um par de *primers* para cada uma das extremidades do fragmento conhecido buscando tanto a amplificação inicial quanto a amplificação aninhada do produto final (*nested PCR*), que é mais precisa e diminui bastante a chance de amplificações espúrias não associadas ao gene que está sendo isolado. Porém, após algumas falhas e impossibilidade de realizar *PCRs-aninhadas* em algumas amostras, identificamos a necessidade de um maior número de *primers* os quais foram construídos e testados em diferentes temperaturas, utilizamos diversas concentrações de reagentes e programas distintos para estipular as condições ótimas para a amplificação (Ver **Anexo 1** para lista de *primers* e programas utilizados). Duas bibliotecas parciais de cDNA, construídas *a priori* por pesquisadoras de nosso laboratório, também foram utilizadas em alguns experimentos afim de testar a amplificação de regiões do gene *fruitless* eventualmente expressas nestas. Estas foram obtidas através de mRNAs expressos em *pools* de aproximadamente 100 tecidos reprodutivos dissecados de machos e fêmeas de *Anastrepha*, ambas utilizando o *Kit CloneMiner* (Invitrogen) seguindo os procedimentos que

evitam contaminações do material e as demais recomendações do fabricante (Pascual e cols, 2010; Gonçalves e cols. *subm*, 2011).

Submetemos todos os produtos de PCR à eletroforese em géis de agarose 1%, corados com brometo de etídeo e visualizados em luz ultravioleta para verificar a eficiência da reação, e assim estabelecemos as melhores temperaturas e condições de *primers* e dNTPs na amplificação das regiões de interesse. Marcadores de peso moleculares foram utilizados como parâmetro de comprimento para os fragmentos amplificados. Foram selecionadas amostras que abrangessem todos os comprimentos obtidos, entretanto um maior número de fragmentos de comprimento próximo e superiores a 1kb foram escolhidos por dispormos de registros evidenciando comprimentos iguais e superiores a estes para o gene *fruitless* em outras espécies de insetos, como: *Nasonia sp*, *Anopheles sp*, *Apis sp*, *Drosophila sp*, entre outras.

Uma vez estabelecidas boas condições de amplificação, fizemos reações com volume total de 50µl fornecendo material amplificado suficiente para purificação e clonagem. Todos os produtos de RACE-PCR provenientes da utilização de ambos os *kits* foram precipitados com uma solução de PEG 8000 (20% de Polietilenoglicol 8000 e 2,5M NaCl) (Lis e Schleif, 1975). Este procedimento garante a precipitação do produto amplificado e, uma vez que não precipita eficientemente produtos com menos de 200 pares de bases, serve para eliminar artefatos da reação de amplificação como dímeros de *primer* e nucleotídeos não incorporados. O produto purificado é novamente submetido à eletroforese em gel de agarose 1% e corado com brometo de etídeo a fim de se determinar a pureza e concentração dos fragmentos obtidos. Esta concentração é também estabelecida através de Dot Blot (processo de quantificação em gota), comparando a intensidade da fluorescência (proporcional a concentração de cDNA aplicado) de cada amostra a padrões conhecidos (marcadores de 10, 20, 30, 40 e 50 ng).

4.6. Clonagem e Sequenciamento

Todas as amostras purificadas foram submetidas aos procedimentos de clonagem usando o *Kit* de Clonagem InsTAclone (Fermentas), através de ligação com plasmídeos do vetor PTZ57R/T e subsequentemente transformação pelo método do choque térmico com bactérias competentes (cepa DH5 α), que apresentam resistência à ampicilina, de acordo com recomendações do fabricante. O *kit* de clonagem T&A Cloning *Kit* (Real Biotech Corporation) também foi utilizado, mas produziu pequeno número de colônias positivas necessárias.

Plaqueamos as colônias recombinantes em meio de cultura LB seletivo (contendo ampicilina (100 μ g/ μ l) e coletamos colônias individuais após um tempo de aproximadamente 18h (Figura 9). Replaqueamos as colônias amostradas em novas placas seletivas para uso futuro. Amostramos aproximadamente 12 colônias para cada reação de ligação, as quais continham uma variedade de fragmentos diferentes de AL DS cDNA amplificado, tanto para 5'RACE quanto para 3'RACE realizados separadamente.



Figura.9 Exemplo de placa contendo colônias recombinantes para fragmentos à jusante (3'RACE) da região conhecida do gene *fruitless* provenientes da reação de ligação e transformação crescidas em estufa após 18h.

Colocamos estas colônias em água ultra pura por 5 minutos à 85°C para lise das células e posterior liberação dos plasmídeos, e amplificamos este material com *primers* M13 forward e reverse (0,2mM), que ligam-se próximo ao sítio de clonagem do plasmídeo, para revelar as colônias que incorporaram o inserto. Dessa forma, diferenciamos as colônias que incorporaram o inserto dos falsos positivos. Os produtos de PCR nos comprimentos escolhidos foram novamente purificados através da precipitação com PEG 8000 e ressuspendidos em 15 μ l de água ultrapura.

Os produtos amplificados destas clonagens foram purificados, quantificados e concentrados na relação exigida pela empresa Macrogen na Coréia do Sul (aproximadamente 25ng/μl), responsável pelos sequenciamentos conduzidos sob as condições de ciclagem do BigDye™ Terminator e utilizando o sequenciador automático ABI3730XL (Applied Biosystem).

4.7 Análises das Sequências e Nível de Identidade

As sequências obtidas para *A. obliqua* foram analisadas com o software Chromas 2.31 (<http://en.bio-soft.net/dna/chromas.html>) que permite a visualização dos eletroferogramas e a suas qualidades. Ambas as fitas foram sequenciadas, e o alinhamento de ambos os eletroferogramas permite a obtenção de uma sequência consenso para o clone em questão. A confirmação da similaridade desses fragmentos com a região de interesse foi feita utilizando a ferramenta *Basic Local Alignment Search Tool* para busca por similaridade nucleotídica, protéica e utilizando nucleotídeos transcritos para busca de proteínas (*BlastN*, *BlastP* e *BlastX*, respectivamente) no GenBank (<http://ncbi.nlm.nih.gov>).

Estas sequências são extraídas dos cromatogramas utilizando o programa Phred (Ewing e Green, 1998) que mede a qualidade das sequências examinando os picos de cada base para atribuir uma pontuação de qualidade. Posteriormente, são analisadas quanto à presença de vetores utilizando o programa Cross-Match, disponível no pacote Phred/Phrap. A qual considera válidas as sequências com no mínimo 100 bases com Phred >20, sendo que Phred de média = 20 representa 99% de chance de a base estar correta. Posteriormente as sequências foram investigadas para 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 sequências de aminoácidos possíveis dentro da sequência de DNA fornecida.

As sequências foram transferidas para o software BioEdit (<http://www.mbio.ncsu.edu/BioEdit/BioEdit.html>) para alinhamento com as demais espécies. Todas as sequências, inclusive aquelas que apresentaram baixa ou nenhuma similaridade com o gene *fruitless*,

foram conferidas manualmente visando maior confiabilidade nos alinhamentos e evitar a perda de sequências informativas caso estas fossem muito distintas das existentes nos bancos de dados.

As taxas de similaridade entre as sequências do gene *fruitless* produzidas neste trabalho para *Anastrepha* e as depositadas em bancos de dados (GenBank e EMBL principalmente) para as espécies próximas foram realizadas utilizando os alinhamentos produzidos e inseridos no programa *DnaSP versão 5* (Rozas, 2009).

4.8 Reconstrução Filogenética e Testes de Seleção

Os testes de seleção foram realizados no programa *PAML ver.4* (Yang, 2007), que utiliza tanto o alinhamento das sequências entre as diferentes espécies, quanto uma árvore filogenética de máxima verossimilhança gerada no programa *PHYML ver.3.0* (Guindon e Gascuel, 2003).

Para a reconstrução da árvore filogenética um modelo otimizado de substituições nucleotídicas foi determinado pelo critério de informação de Akaike (AIC) utilizando o *MODELTEST ver.3.7* (Posada e Crandall, 1998) implementado no pacote *HyPhy ver.0.95 beta* (Pond e cols, 2005). A árvore foi estimada por máxima verossimilhança sob o modelo de substituições de nucleotídeos TIM+G utilizando a sequência completa do gene *fruitless* estabelecida ao longo desse projeto para *Anastrepha* (dados não publicados) em conjunto com as sequências de cada espécie de drosofilídeos, sendo elas: *D. melanogaster* – [D84437.1], *D. heteroneura* – [AF051668.1] e *D. silvestris* – [AF051665.1].

Buscando investigar as pressões seletivas que modelaram a evolução de *fruitless*, realizamos e contrastamos os resultados do testes de seleção “*Relaxed Branch-site test*” (Zhang e cols, 2005) e “*Strict Branch-site test*” (Yang and Nielsen, 2002) que utilizam o *software CODEML*, presente no pacote *PAML*, para simular diferentes cenários evolutivos. Ambos os testes utilizam modelos nulos e alternativos para calcular os índices decorrentes das taxas de substituições não-sinônimas/sinônimas ($dN/dS = \omega$) que são estimados visando identificar pressões seletivas. Qualquer valor $\omega > 1$ em um sítio

específico indica ocorrência de seleção positiva, uma vez que substituições não-sinônimas teriam uma maior probabilidade de fixação do que mutações sinônimas devido à vantagem seletiva. Por outro lado, $\omega < 1$ indicam ocorrência de seleção purificadora, causados provavelmente restrições seletivas na posição do códon. Em ambos os testes determinamos também a topologia da árvore na qual o gênero *Anastrepha* representa o ramo *foreground*, no qual a seleção positiva é testada, em detrimento das demais linhagens utilizadas que compõem o ramo *background*.

Em ambos os testes, a hipótese alternativa (MA) permite que os códons do ramo *foreground* assumam valores superiores a 1 ($\omega > 1$), enquanto que os códons do ramo *background* são limitados $\omega \leq 1$. Já a hipótese nula, apesar de não permitir seleção positiva em nenhum dos dois modelos, existe uma pequena variação entre os testes, sendo que no teste “*Relaxed*” os sítios podem variar seus valores de ω entre 0 e 1, diferentemente do teste “*Strict*” que fixa em $\omega = 1$ todos aqueles sítios que apresentem sinais $\omega > 1$ (ou seja, é forçado a evoluir de maneira neutra).

Dessa forma, podemos identificar a ocorrência de seleção positiva através de um Teste dos Índices de Máxima Verossimilhança (ou *LRT – Likelihood Ratio Test*), que transforma o log da razão destes testes em um resultado que após submetido a um teste de distribuição de qui-quadrado (X^2), pode ser considerado significativo ou não, ou seja, aceitando ou rejeitando a hipótese que sugere ocorrência de seleção positiva (Zhang e cols, 2005) nos sítios sujeitos a seleção do modelo alternativo identificados pelo método *Bayes Empirical Bayes* (Yang e cols, 2005).

Tanto as taxas de identidade quanto as análises de seleção positiva fornecem informações valiosas a respeito da diferenciação não somente de todo o gene em questão mas também de regiões específicas onde observamos a variação das taxas evolutivas entre as espécies comparadas ao longo do tempo evolutivo (Yang e Swason, 2002; Carter e Drouin, 2009).

5. RESULTADOS E DISCUSSÃO

5.1. Padronização dos Procedimentos

As adaptações realizadas aos protocolos existentes representaram um aumento significativo na eficiência de tais procedimentos e foram essenciais para a obtenção da grande quantidade de material necessária para a prática das RACE-PCRs com o *kit Marathon* e possibilitou a obtenção de RNA total de apenas um indivíduo (uma única larva, pupa ou mosca adulta) para sua utilização no *kit First Choice RLM-RACE*. Corroboramos tais afirmações através dos ganhos relativos a cada uma das etapas que antecedem a síntese do cDNA uma vez que o rendimento médio foi ampliado em mais de 100% na extração do RNA total com TRIZOL (Figura 10), o que corresponde a aproximadamente 2000ng/μl em 10μl de solução e atingindo um máximo de 3058ng/μl em 10μl de solução em uma das extrações feita de uma única larva. Na etapa subsequente notou-se o isolamento de 15% a mais de RNA mensageiro quando realizamos pequenas mudanças e o procedimento sendo realizado no mesmo dia da extração. Além do acoplamento entre os dois procedimentos anteriores, foi observado que as modificações adotadas sobre a própria metodologia de isolamento do mensageiro através das esferas magnéticas (*beads*) proporcionaram um aumento significativo na quantidade de RNA mensageiro recapturado (Figura 11) passando de 1,57% para 2,2% do RNA total em comparação ao *kit PolyATtract® System 1000* (Promega) também testado, o que representa um crescimento de 71,36% em eficiência.

Um dos grandes empecilhos observados nesse projeto foi a grande perda de material durante a etapa de concentração na qual cerca de 80% de todo o material era perdido nos primeiros experimentos seguindo-se os protocolos tradicionais. Depois de testes e consultas com outros pesquisadores, novas condições foram estabelecidas e a quantidade de material máxima utilizada de uma só vez foi reduzida, recuperando em torno de 50% do material após a realização desse procedimento. Dessa forma, o aporte inicial de RNA mensageiro concentrado exigido para dar início a síntese dos AL DS cDNAs pôde ser respeitado (1μg de material diluído em 1-4 μl de solução) e suas etapas posteriores puderam ser

realizadas de acordo com as especificações do fabricante do *kit Marathon*. Vale lembrar que esta etapa pôde ser evitada na utilização do *kit First Choice* graças às otimizações adotadas e o mesmo necessitar de 4x menos material inicial (0,25µg de mRNA) para produzir resultados satisfatórios.

Em seguida foi sintetizado o AL DS cDNA (cDNA de dupla-fita ligado a adaptadores) seguindo-se as orientações do fabricante do *kit Marathon*. Portanto, foram utilizados 4 µl do nosso material concentrado contendo aproximadamente 1,1 µg. A comparação entre nossa amostra de AL DS cDNA e o controle de material placentário humano fornecido pelo *kit* forneceu uma estimativa da diluição a ser utilizada antes da realização das RACE-PCRs. O controle foi 3-4 vezes mais nítido que nossa amostra e optamos por realizar três diluições: 25, 100 e 250µl de tampão com tricina-EDTA (Figura 12).

Antes da realização das RACE-PCRs e visando confirmar a presença de AL DS cDNAs (controle-positivo) esperados nessa etapa do processo, foram realizadas PCRs com *primers* para amplificação da região conhecida do gene *fruitless* com as três diluições. Foi obtida uma banda de aproximadamente 850 pb nos três casos, o que evidencia não apenas a expressão do gene em questão no material extraído, mas também o sucesso do isolamento de seus produtos na produção de AL DS cDNA (Figura 13). Este teste também foi realizado após obtenção dos cDNA ligados a adaptadores no *kit First Choice* e não consta do protocolo de nenhum dos dois *kits*, mas se mostrou como uma eficiente forma de assegurar a não degradação do material de interesse após essa fase trabalhosa de inclusão dos *linkers*.

A seguir apresentamos as imagens dos géis de agarose 1% fotografados ao longo das etapas dessa metodologia utilizando os protocolos adaptados.

Fig.10 RNA total

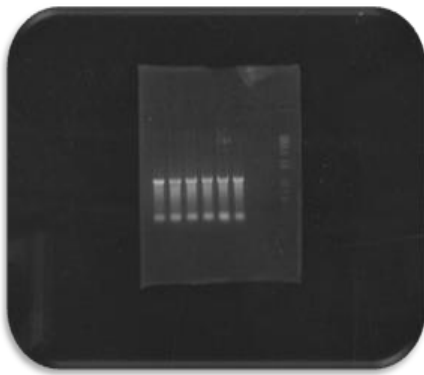


Fig.11 mRNA

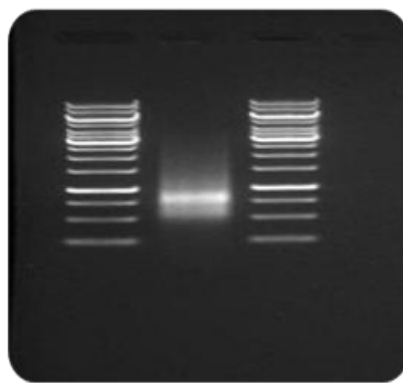


Fig.12 Controle AL ds cDNA (diluição)

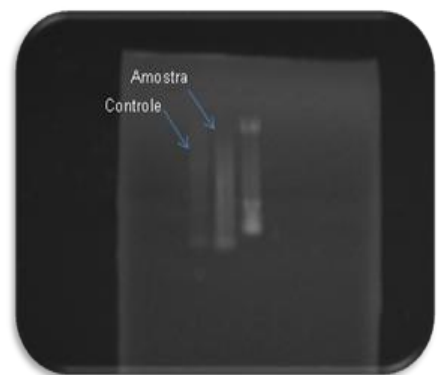
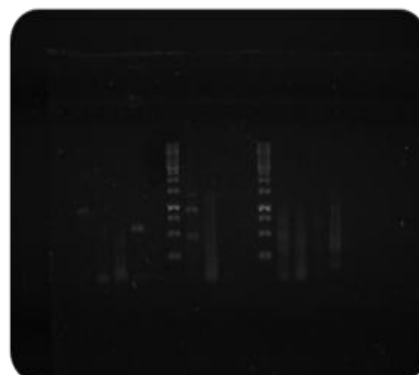
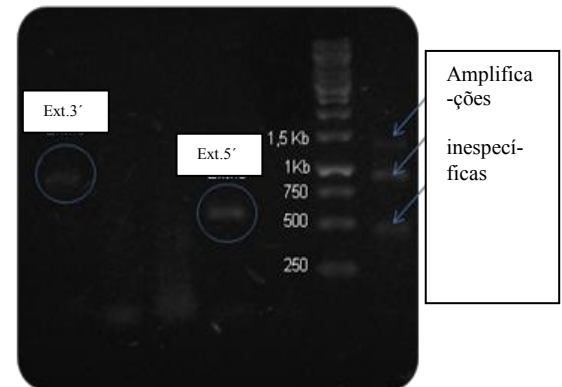
Fig.13 Controle *fruitless* AL ds cDNAFig.14 Teste dos *primers*Fig. 15 3'RACE Aprox. 950 pb
5'RACE Aprox. 590 pb
GeneRuler 1kb Ladder (Fermentas)

Fig.10 Seis amostras com aproximadamente 2000 ng RNA total.

Fig.11 mRNA mensageiro (centro) isolado pelas *beads*, o que representa 2,2% do RNA total.

Fig.12 Comparação entre a amostra de AL DS cDNA obtida (esquerda) e o controle de material placentário humano (centro). Marcador de peso molecular (direita) utilizado não foi satisfatório.

Fig.13 Controle positivo através da amplificação da região "C1" do gene *fruitless* utilizando *primers* já padronizados usando como amostra AL DS cDNA nas diluições 1/100 (esquerda) e 1/250 (direita)

Fig.14 Teste dos *primers* para 5' e 3' RACE em diferentes condições e programas.

Fig.15 Amplificação padronizada para ambas as extremidades.

Obs: Cada linha resulta da aplicação de 1 μ l de amostra. Utilizamos como parâmetro de peso molecular o marcador GeneRuler 1kb Ladder (Fermentas).

5.2. Padronização das Amplificações

Como descrito na metodologia, diversas *RACE-PCRs* (Figura 14 e 15) utilizando combinações diferentes de *primers* foram realizados seguidas por *nested-PCRs* com *primers* mais internos visando uma maior confiabilidade dos fragmentos amplificados. Em poucos casos, em que os *primers* escolhidos eram mais externos aos demais, essa segunda rodada de amplificações não pôde ser realizada, sendo assim, a validação nesses casos se deu pelo sequenciamento direto dos fragmentos obtidos.

Os fragmentos de comprimentos distintos obtidos foram clonados, reamplificados com *primers* M13 *forward e reverse* (Figura 16) e sequenciados visando encontrar aqueles correspondentes ao gene *fruitless* uma vez que seu comprimento é desconhecido para nossa espécie. Outro fator importante para seleção de uma ampla quantidade de fragmentos encontrados está na existência de diversas isoformas que apresentam provavelmente comprimentos variáveis.

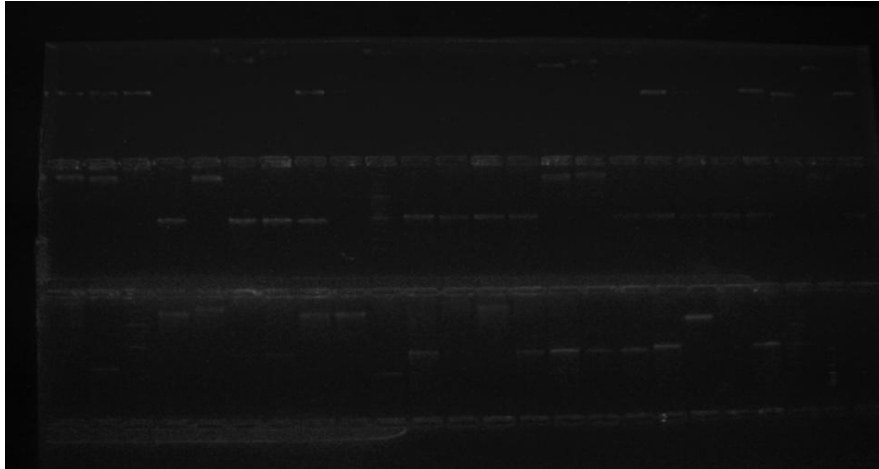


Fig.16 Exemplo de gel de agarose 1% contendo vários fragmentos de tamanhos distintos posteriores a amplificação com *primers* M13 universais. Para fins de comparação foram escolhidos dois fragmentos de cada um dos comprimentos obtidos e submetidos ao sequenciamento.

De porte destes sequenciamentos pudemos identificar fragmentos similares àqueles expressos pelo gene *fruitless* em outras espécies, facilitando a caracterização parcial deste e a identificação das possíveis isoformas por nós isoladas.

Através dos resultados das amostras e dos controles realizados podemos inferir que o protocolo do *kit Marathon* apesar de bastante trabalhoso e de alto investimento financeiro, não apresentou grandes dificuldades para sua realização. Contudo, sua eficiência na síntese de cDNAs ligados aos adaptadores artificiais para o gene *fruitless* ainda não está totalmente comprovada, uma vez que sua amplificação não atingiu o nível de especificidade que se aguardava (aproximadamente 50% de colônias falso-positivas) e não logramos sucesso completo na amplificação da região 5'. Entretanto, importantes sequências

pretendidas para o final 3' foram obtidas e confirmam sua alta variabilidade de formas e complexidade estrutural, seguindo o padrão já descrito e revisado para outros insetos (Salvemini e cols, 2010).

Já o *kit First Choice* também foi capaz de amplificar eficientemente os produtos desejados, apresentando a característica de necessitar de menos material para sua realização, o que lhe confere uma vantagem importante uma vez que uma grande quantidade de mRNAs de boa qualidade demonstrou-se de difícil obtenção para nossos organismos. As sequências provenientes da utilização desse *kit* também se referem majoritariamente a extremidade 3' e possibilitaram a identificação de variantes produzidas pelo mecanismo de *splicing alternativo* especialmente das fases larvais e de pupa.

Uma vez que obtivemos uma quantidade reduzida de sequências informativas a respeito da extremidade 5', fomos obrigados a lançar mão de um método de obtenção alternativo que consistiu na amplificação das sequências-alvo utilizando *primers universais M13* em conjunto com *primers* específicos para o gene *fruitless* orientados para extremidade 5' utilizando como sequências molde duas bibliotecas de cDNA distintas provenientes de tecidos reprodutores de machos (Pascual e cols, 2010) e de fêmeas (Gonçalves e cols.*subm.*, 2011).

5.3 Sequências Obtidas e Alinhamentos

Ao total foram sequenciados 207 fragmentos, os quais após tratados e conferidos no GENBANK (através da ferramenta *BLAST*), revelaram 114 sequências informativas referentes a diversas regiões genômicas de *Anastrepha sp* pela similaridade com *Drosophila* e demais insetos dípteros, a grande maioria delas ainda não descrita para esse gênero. Destas, três sequências foram identificadas como sendo referentes à extremidade 5' do gene *fruitless*, enquanto que outras 45 sequências dizem respeito à extremidade 3'. Depois de analisadas individualmente fomos capazes de estabelecer a correspondência destes fragmentos com a região específica do gene *fruitless*, entre elas: uma sequência de 210pb (relativa à uma região a montante gene *fruitless* na qual encontram-se os éxons BTB1 e BTB2); duas sequências de 299pb (que se inicia na região conhecida do éxon C3, passa pelo éxon C4 e termina no

início de D em *Drosophila silvestris*); duas sequências praticamente idênticas de 861pb (que englobam o final do éxon C3, passando pelo éxon C4 até a região mediana do éxon D) que diferem em apenas uma substituição sinônima (Figura 17); uma sequência de 544pb (relativo à amplificação da porção final do éxon D, último presente na isoforma D); duas sequências de 531pb (referentes a porção mediana e final do éxon D, que sobrepõem em suas extremidades as regiões iniciais e finais do mesmo éxon já mencionadas, possibilitando sua completa identificação); duas sequências de 762pb (referentes à região já conhecida do BTB2 e C3 em *Anastrepha sp*); uma sequência 556pb (que se inicia no final do éxon C5 e termina no região intermediária do éxon ZincFinger do “tipo C”); uma sequência de 809pb (que se inicia no final do éxon C4, apresenta C5 e finaliza no stop códon do éxon ZincFinger do “tipo C”); outras 34 sequências analisadas, variando entre 423pb e 856pb, referiam-se a fragmentos incompletos ou de menor extensão da extremidade 3’ (regiões dos éxons C4, D, C5 e domínios Zinc-Fingers).

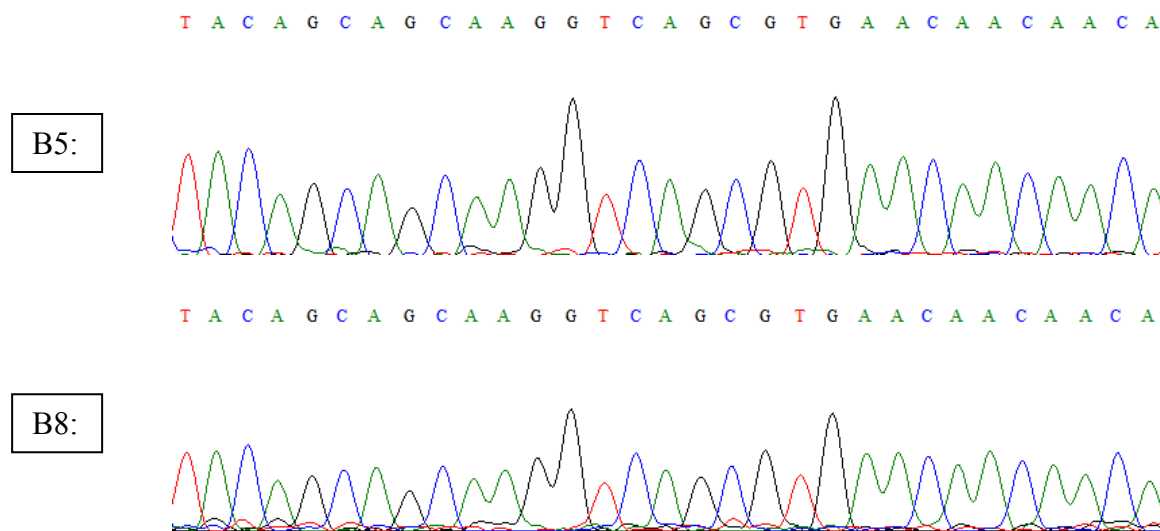


Fig. 17: exemplos referentes a um mesmo trecho (520-550pb) dos sequenciamentos de duas colônias distintas da extremidade 3’ do gene *fruitless*.

Observando-se o total de fragmentos obtidos e aqueles realmente dizem respeito às regiões pretendidas nesse projeto observamos uma baixa especificidade inerente à utilização desta técnica, mesmo utilizando re-amplificações com *primers* aninhados (*nested-primers*), isso se deve

provavelmente a primeira rodada das RACE-PCRs utilizar apenas um *primer específico para o gene* ao invés de dois como normalmente ocorre em PCRs tradicionais, elevando a chance de hibridizações e subsequentes ampliações inespecíficas.

As sequências informativas deste trabalho apresentaram altas taxas de similaridade nucleotídica e protéica com CDS do gene *fruitless* de diversas espécies de *Drosophila* e outros dípteros, os quais já têm seu genoma completamente sequenciado. Quando buscamos traduzir estes nucleotídeos, vimos que os produtos referem-se a uma única ORF (*Open Reading Frame*) indicando que se tratam de fato de mRNAs funcionais únicos.

5.4 Organização Estrutural do gene *fruitless* em *Anastrepha*

O alinhamento das proteínas preditas com a de outras espécies de dípteros revelou que nossa estratégia de fato conseguiu identificar tanto éxons a jusante quanto a montante da região conhecida. De posse desses dados e sobrepondo regiões similares entre as sequências obtidas separadamente fomos capazes de organizar estruturalmente a região codificadora do gene *fruitless* para *Anastrepha*, que se estende desde os domínios BTB que agem como moduladores de interações proteína-proteína e se enquadram em uma variedade de classes funcionais já relatadas (Privé, 2005), até os *motifs* terminais que expressam sítios de ligação ao DNA do tipo Zinc-Fingers (Gailey, 2006), sendo elas: a região 5' representada pelos dois éxons BTB1 (chamado C1) e BTB2 (chamado C2) (*Bric-à-brac*, *Tramtrack* and *Broad complex transcription regulators*), um íntron de 79 pb conhecido graças a amplificação de uma porção de DNA deste gene entre os éxons BTB2 e C3 (Sobrinho e de Brito, 2010), os três éxons da região conectora C3, C4 e C5, o éxon ZnF_D (estruturalmente localizado entre C4 e C5) que codifica para um Zinc-Finger de expressão limitada em *Drosophila* compondo apenas uma das 14 isoformas já descritas para esse gênero; e por fim o éxon ZnF_C à jusante do éxon conector C5, similar aquele

encontrado nos demais dípteros e característico do final das proteínas produzidas por *fruitless* (Figura 18 e Anexo 2).

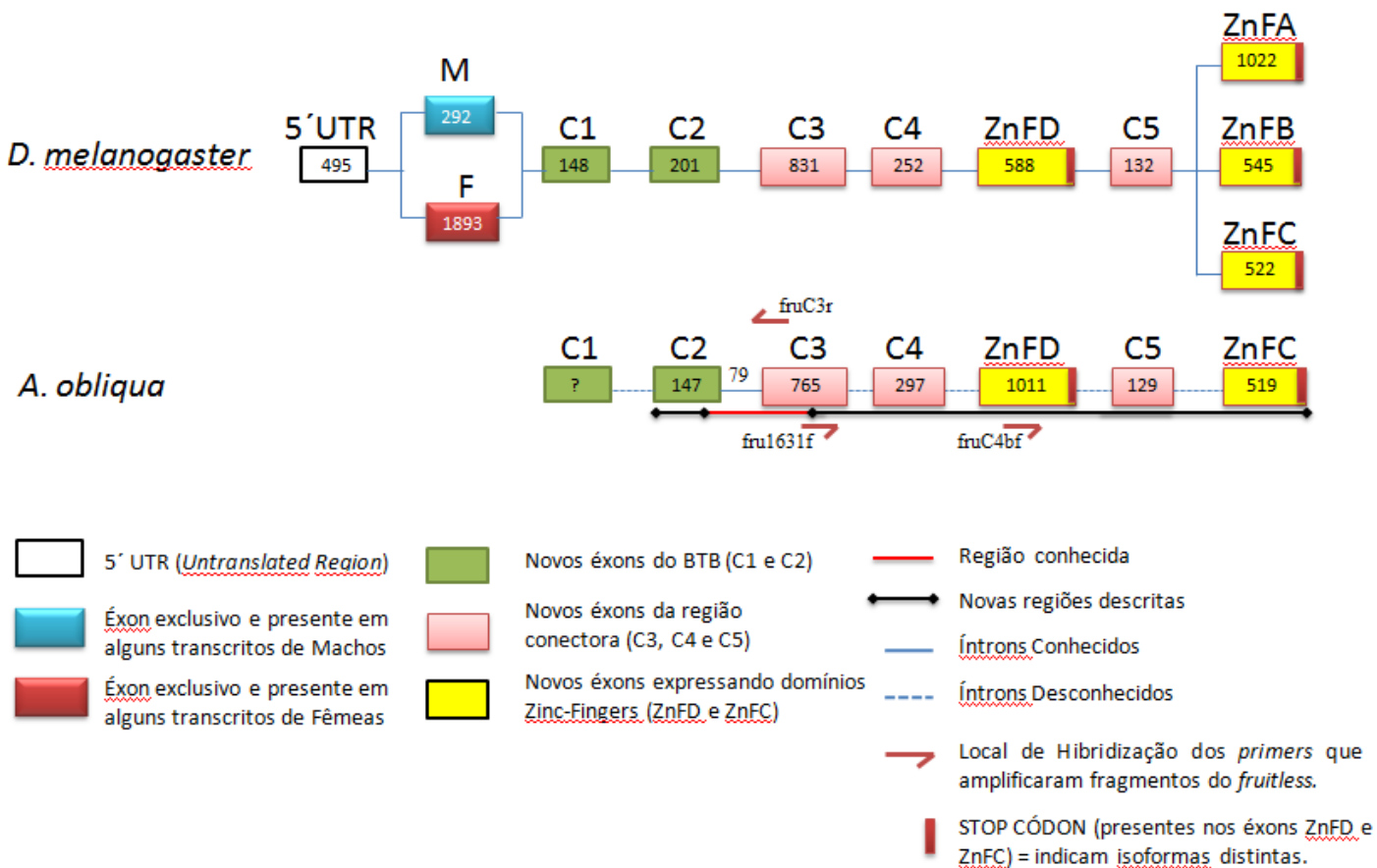


Fig.18 Representação esquemática do gene *fruitless* identificada neste projeto para *A. obliqua* em contraste com *Drosophila melanogaster*.

Temos confiança que essas sequências referem-se às regiões codificadora do gene *fruitless* em questão por representarem *ORFs*, terem alta qualidade de sequenciamento (sequenciamentos médios de 750 nucleotídeos com boa resolução, tendo alcançado 861pb), apresentarem alta similaridade das regiões iniciais e terminais dos fragmentos com espécies próximas, sobreposição de regiões semelhantes obtidas em fragmentos distintos, ausência de códons de terminação e presença da cauda poli-A nos eletroferogramas amplificados com *primers* M13rev.

Apesar de algumas amostras não possibilitarem seu sequenciamento reverso em função da existência da cauda poli-A, que promove alterações no sequenciamento em função de “gagueira”

(*stutter*) ou escorregões da enzima que inviabilizam a obtenção de sequências de qualidade após tal cauda, conseguimos obter novos fragmentos que puderam ser sequenciados completamente até as regiões terminais dos éxons que apresentavam códons de terminação (*Stop códons*) através da amplificação desses clones por *gene-walking*.

De posse dessas informações, identificamos com segurança a extremidade 5' quase que em sua totalidade, restando apenas a região 5'UTR, os eventuais promotores, os éxons específicos de cada sexo, caso existam, e o primeiro éxon do domínio BTB. Temos confiança também que a extremidade 3' está totalmente caracterizada ao menos para as transcritos terminados nos éxons que codificam os Zinc-Fingers dos tipos D e C. Dessa forma, este projeto contribuiu significativamente para expansão do banco de dados a respeito da região codificadora do gene *fruitless* em *Anastrepha*, tornando conhecidas diversas regiões, transcritos e sequências que podem servir como locais confiáveis para hibridização de *primers específicos* visando a obtenção de outras isoformas em estudos posteriores.

5.5 Identidade entre as Sequências

Depois de estabelecida a sequência linear de organização para *Anastrepha* calculamos a taxa de similaridade nucleotídica e de aminoácidos codificados (incluindo os *indels*; sequência de bases presentes em apenas uma das espécies) entre este e os demais grupos aparentados, para as quais o gene em questão já se encontra parcialmente (como em *Bactrocera* e *Ceratitis*) ou completamente descrito, produzindo os dados encontrados nas Tabelas 1 e 2, respectivamente.

Tabela 1. Taxa de similaridade nucleotídica global (nível de identidade) e por éxons entre a sequência do gene *fruitless* identificada para *Anastrepha* em comparação com seu correspondente em outros insetos.

Regiões e éxons do gene <i>fruitless</i>		Região BTB	Região Conectora			Zinc Fingers		TOTAL
Espécies (Família)		C2	C3	C4	C5	ZnF_D	ZnF_C	CDS
A	<i>Bactrocera cucurbitae</i> (Tephritidae)	92,59%	--	--	--	--	--	92,59%
	<i>Ceratitis capitata</i> (Tephritidae)	95,37%	37,09%	--	--	--	--	66,23%
	<i>D. melanogaster</i> (Drosophilidae)	86,11%	46,23%	54,33%	60,60%	31,66%	54,54%	55,57%
	<i>Anopheles gambiae</i> (Culicidae)	65,74%	33,59%	26,51%	47,36%	17,90%	44,42%	39,25%
B	<i>Bombyx mori</i> (Bombycidae)	67,59%	32,50%	34,66%	44,27%	14,13%	33,97%	37,85%
C	<i>Tribolium castaneum</i> (Tenebrionidae)	70,37%	14,34%	25,16%	50,75%	20,71%	21,19%	33,75%
D	<i>Nasonia vitripennis</i> (Pteromalidae)	68,51%	22,94%	*	38,93%	15,92%	28,70%	34,89%
	<i>Apis mellifera</i> (Apidae)	67,59%	23,28%	*	44,69%	32,06%	28,91%	39,30%
Taxa média de Identidade por éxon		76,73%	29,99%	35,16%	47,76%	22,06%	35,28%	

Tabela 2. Taxa de similaridade global entre os aminoácidos codificados pelas sequências comparadas (nível de identidade) e por éxons para o gene *fruitless* identificada para *Anastrepha* e seu correspondente em outros insetos.

Regiões e éxons do gene <i>fruitless</i>		Região BTB	Região Conectora			Zinc Fingers		TOTAL
Espécies (Família)		C2	C3	C4	C5	ZnF_D	ZnF_C	CDS
A	<i>Bactrocera cucurbitae</i> (Tephritidae)	97,22%	--	--	--	--	--	97,22%
	<i>Ceratitis capitata</i> (Tephritidae)	97,22%	41,76%	--	--	--	--	69,49%
	<i>D. melanogaster</i> (Drosophilidae)	97,22%	47,91%	59,00%	61,36%	33,92%	66,66%	61,00%
	<i>Anopheles gambiae</i> (Culicidae)	69,44%	18,03%	14,14%	20,00%	13,35%	41,66%	29,43%
B	<i>Bombyx mori</i> (Bombycidae)	83,33%	10,58%	10,48%	13,95%	8,03%	34,30%	26,77%
C	<i>Tribolium castaneum</i> (Tenebrionidae)	75,00%	14,50%	15,00%	32,55%	19,04%	23,83%	29,98%
D	<i>Nasonia vitripennis</i> (Pteromalidae)	69,44%	12,54%	*	18,18%	16,36%	33,13%	28,93%
	<i>Apis mellifera</i> (Apidae)	66,66%	11,37%	*	22,72%	19,28%	30,93%	30,19%
Taxa média de Identidade por éxon		81,94%	22,38%	24,65%	28,12%	18,33%	38,41%	

Em ambas as tabelas (1 e 2): (--) indicam cálculos não realizados pela ausência de dados para as espécies relacionadas. (*) indicam a não existência de determinados éxons para as espécies. As letras indicam as ordens as quais pertence cada gênero: (A) Diptera, (B) Lepidoptera, (C) Coleoptera e (D) Hymenoptera.

Esses dados confirmam nossas expectativas a respeito da região codificadora do BTB ser a mais conservada desse gene entre todos os grupos comparados, mantendo valores acima de 65% de similaridade mesmo para os grupos menos relacionados (*p.ex. Apis mellifera* = 66,66%). Quanto a essa região, destacamos principalmente sua conservação entre os tefritídeos nos quais obtivemos taxas de identidade acima de 92% levando-se em conta os nucleotídeos e superiores a 97% quando considerados os aminoácidos codificados pelas sequências.

Outros éxons que apresentaram regiões internas com significativa conservação em detrimento de outras mais variáveis foram os éxons C5 e ZnF_C, que apesar de não superarem 60% de conservação em nenhuma situação, são compostos basicamente pelos mesmo aminoácidos em suas porções medianas e terminais (extremidade 3') em praticamente todas as espécies. Ao menos para o éxon ZnF_C isso pode ser explicado por essas porções coincidirem com os aminoácidos que codificam para os importantes domínios do tipo *ZincFinger* característicos dessa região terminal de alguns transcritos (Gailey e cols, 2006).

Entre os demais éxons com menores taxas de similaridade aquele que mais merece destaque sem dúvida é o ZnF_D, raramente expresso em drosofilídeos. Sua estrutura em *Anastrepha* só pôde ser identificada com segurança após a sobreposição de uma série de fragmentos amplificados em sequência iniciando-se no éxon anterior (éxon C4) e terminando no éxon posterior (éxon C5), mais conservados, que vieram a preencher as divergências deixadas pelo processo evolutivo. Além de o éxon ter praticamente o dobro do comprimento do observado em *Drosophila melanogaster* (588 pb contra 1011 em *Anastrepha*), identificamos uma grande similaridade na porção medial complementada por uma espécie de “fragmentação de alinhamento” ao longo do restante do mesmo, no qual pequenas sequências não contendo mais do que 15 aminoácidos foram alinhadas entre extensas regiões apresentando *indel* (Anexo 2).

Uma das hipóteses elaboradas por Bertossa (2009) para explicar as grandes variações encontradas entre as regiões que codificam *zinc fingers* e outras regiões como apontadas nesse e em demais trabalho como a ausência de *zinc fingers* do tipo A e F em *Bombyx mori*; do tipo F em drosofilídeos; de éxons específicos para cada sexo em drosofilídeos havaianos; e da fusão de éxons da região BTB e conectora em *Nasonia*, sugere que a história evolutiva do gene *fruitless* entre os insetos se deve a existência de apenas dois ou três C₂H₂ *zinc fingers* basais no ancestral anterior a irradiação dos holometábolos que deram origem um deles aos tipos A e C, e o outro aos tipos B e F, tendo estes pares funções redundantes herdadas do ancestral comum e sendo suscetíveis a evolução multifuncional no decorrer do período evolutivo, tendo assim se diferenciado, gerado novos domínios e possivelmente adquirido novas funções. Para testar essa hipótese, pode-se expressar ou inativar, através de *knock-out*, os *zinc fingers* teoricamente mais derivados (*p.ex.* tipo F) ausentes ou presentes em espécies de interesse, visualizando seus eventuais efeitos fenótipos.

Pelo fato de considerarmos os *indels*, presentes entre todos os alinhamentos realizados, esperamos que a porcentagem de identidade entre as sequências tenham seus valores subestimados, uma vez que o programa *DnaSP* considera como ausência total de similaridade estas situações, ou seja, regiões onde o alinhamento apresenta um “GAP” em qualquer uma das sequências comparadas, o *software* atribui uma diferença entre ambas. Se considerarmos, por exemplo, apenas as regiões com correspondência mesmo entre espécies mais distantes entre *Anastrepha* e *Nasonia* e para um éxon altamente variável como o ZnF_D citado acima encontraremos 64,70% de identidade de aminoácidos, bastante diferentes dos 15,92% obtidos na análise inicial (Figura 19).

Outro exemplo pode ser obtido contrastando o éxon C3 entre *Anastrepha* e *Ceratitis*, no qual a análise inicial apontou 37,09% de similaridade nucleotídica, porém, após a parte final do éxon ter sido desconsiderada da análise por ser uma região exclusiva de *Anastrepha* e sem similaridade com as demais espécies, essa identidade atingiu 81,95% (Figura 20).

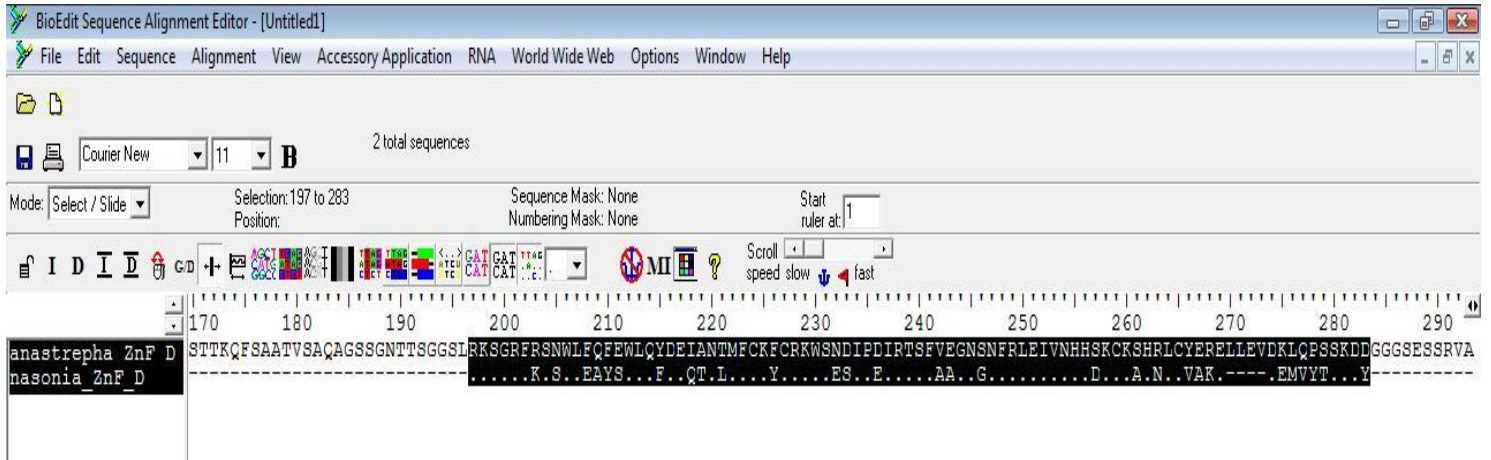


Fig.19 Imagem do alinhamento realizado no programa BioEdit para simular a taxa de similaridade entre os aminoácidos codificados pelas seqüências da porção central do éxon ZnF_D entre *Anastrepha* e *Nasonia*, desconsiderando-se as regiões de *indel*. A área em destaque escuro representa a região utilizada para nova análise. Os traços (--) representam os *indels* e os pontos (.) representam aminoácidos iguais entre as seqüências.

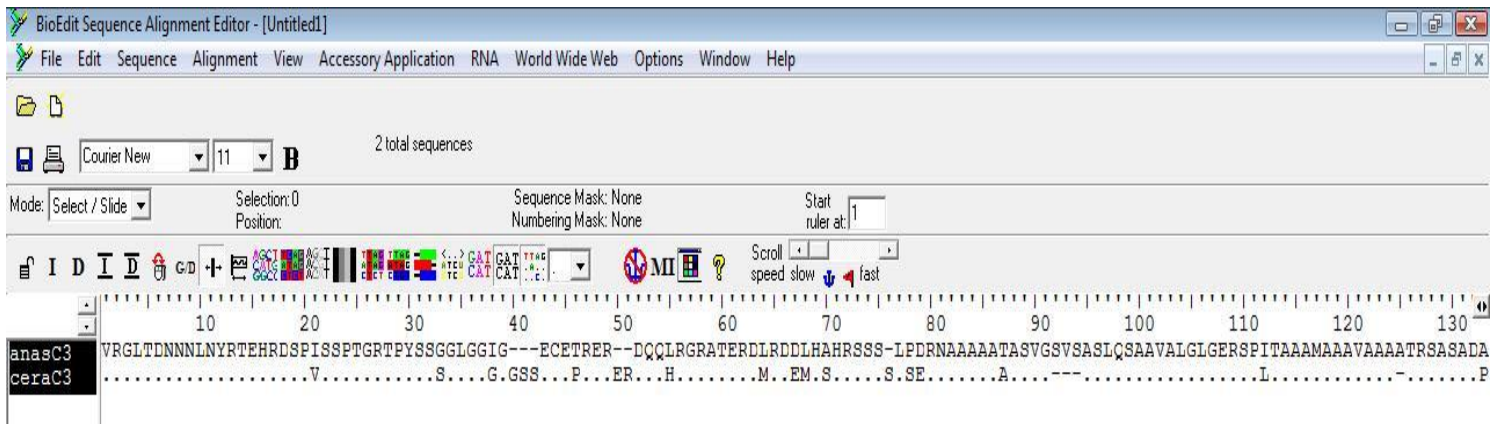


Fig.20 Imagem do alinhamento realizado no programa BioEdit para simular a taxa de similaridade entre os aminoácidos codificados pelas seqüências da porção central do éxon C3 entre *Anastrepha* e *Ceratitis*, desconsiderando-se as regiões de *indel*. A área em destaque escuro representa a região utilizada para nova análise. Os traços (--) representam os *indels* e os pontos (.) representam aminoácidos iguais entre as seqüências.

Ainda que com essa ressalva que requer uma avaliação mais minuciosa de cada uma das regiões comparadas nessa análise, foram obtidos dados valiosos a respeito de diversas regiões específicas do gene antes desconhecidas, tanto aquelas que mantiveram uma maior conservação como as que passaram por grandes diferenciações ao longo do tempo evolutivo, possibilitando a identificação principalmente

de aminoácidos distintos que vem sendo codificados pelos organismos em questão, o que acaba por nos fornecer indicadores interessantes sobre possíveis alterações na estrutura das proteínas produzidas que podem levar a alterações estruturais e funcionais singulares entre as espécies.

5.6 Seleção Positiva sobre o gene *fruitless*

A árvore filogenética (Figura 21) gerada no programa *PHYML* foi utilizada juntamente com as sequências provenientes da região codificadora obtida para os estudos evolutivos do gene *fruitless*. A região codificadora totalizou 2868 pb, se iniciando no segundo éxon da região BTB (éxon C2) e terminando no códon de terminação do éxon conhecido como ZnF_C.

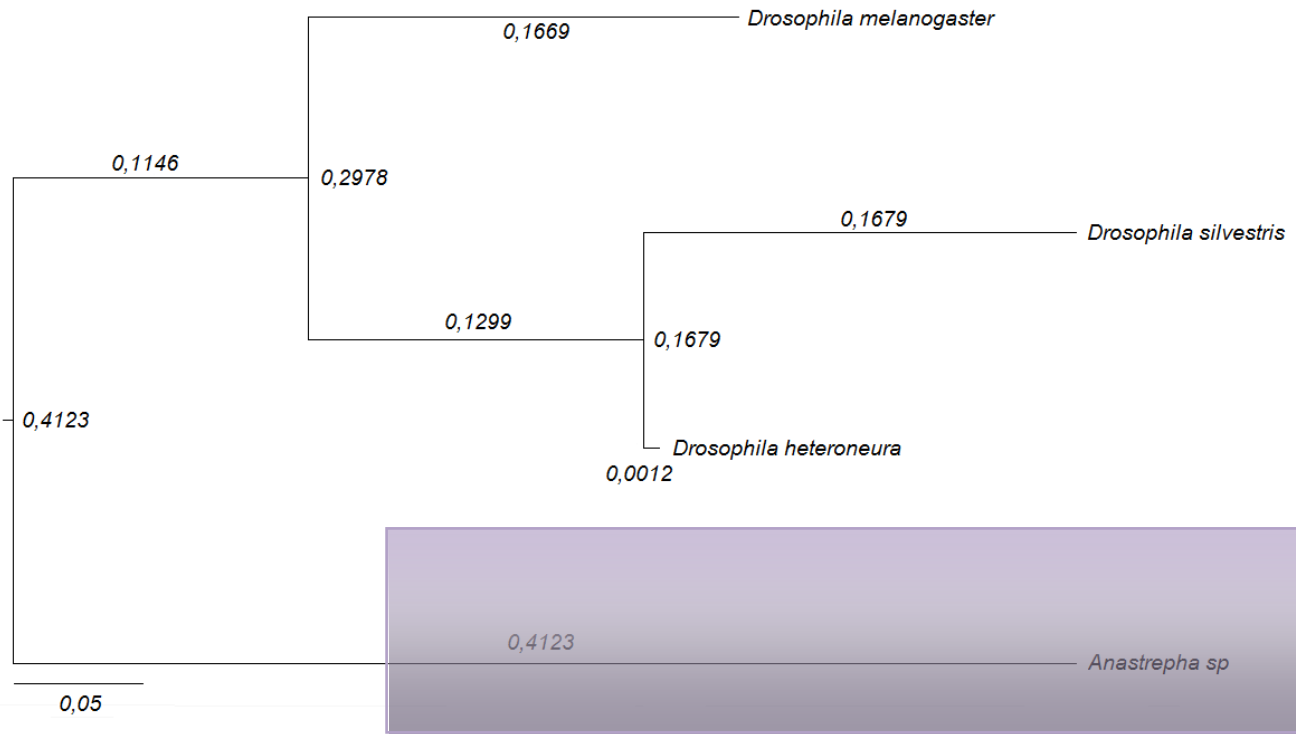


Fig.21 Árvore filogenética estimada por máxima verossimilhança da região codificadora do gene *fruitless* utilizada para detecção de seleção. O ramo *foreground* está destacado em cinza. A distância genética está expressa em substituições de nucleotídeo por códon.

A Tabela 3 mostra tanto os parâmetros inferidos para os modelos nulos M1a e MA restrito, quanto para os modelos MA alternativos.

Modelo	Parâmetros			Lnl
M1a ($\omega_1 = 1$, fixado)	$\omega_0 = 0,075$	$\omega_1 = 1.000$		
<i>Relaxed Brach-site test nulo</i>	$p_0 = 0,487$	$p_1 = 0,513$		-10071,980644
MA restrito ($\omega_2 = 1$, fixado)	$\omega_0 = 0,052$	$\omega_2 = 1.000$		
<i>Strict branch-site test nulo</i>	$p_0 = 0,407$	$p_1 = 0,298$	$(p_2+p_3) = 0,295$	-10049,419875
MA	$\omega_0 = 0,066$	$\omega_2 = 186,015$		
Modelo alternativo	$p_0 = 0,423$	$p_1 = 0,295$	$(p_2+p_3) = 0,282$	-10005,583645

Tabela 3. Parâmetros estimados e valores de *log likelihood* (Lnl) para os modelos M1a, MA e MA restrito. ω_0 = valores de dN/dS para sítios com $0 < \omega < 1$; ω_1 = valores de dN/dS fixados em 1; ω_2 = valores de dN/dS para sítios com $\omega > 1$, os quais correspondem somente aos sítios do ramo *foreground* (*Anastrepha*). Os parâmetros (p_2+p_3) representam a proporção total de sítios sob seleção positiva.

O *relaxed branch-site test* rejeitou o modelo nulo (M1a) com restrições seletivas ($0 < \omega \leq 1$) indicando que o ramo *foreground* (Figura 21) divergiu na região analisada por limitações seletivas relaxadas ou por seleção positiva (Tabela 4). Contrastando o modelo MA restrito (null MA, ou MA nulo) com o modelo MA alternativo do *strict branch-site test*, pudemos discriminar entre as duas hipóteses e novamente o modelo nulo foi rejeitado (Tabela 4) em favor da hipótese de diferenciação por seleção positiva. Sendo assim, através do *Bayes Empirical Bayes* pudemos identificar 25 sítios evoluindo sob seleção positiva com probabilidade superior a 95% (Tabela 4) e criar um gráfico contendo todos os sítios para os quais essa probabilidade superou 50% (Anexo 3).

Teste	Contraste	LRT	GL	Probabilidade χ^2	Sítios sob Seleção Positiva
<i>Relaxed Brach-site test</i>	M1a x MA	133	2	$p < 10^{-8}$	20,21,23,24,29,31,92,122,147,262,405,429,473,488
<i>Strict Brach-site test</i>	Restricted MA x MA	87,7	1	$p < 10^{-10}$	489,490,495,502,507,521,536,703,768,785,927.

Tabela 4. Comparação entre os modelos nulos e alternativos por LRT (*Likelihood Ratio Test*) e estimativa de sítios evoluindo sob seleção positiva com por BEB (*Bayes Empirical Bayes*) com confiabilidade superior 0.95. GL: graus de liberdade.

Esse gráfico facilitou a identificação visual de regiões com tendência elevada a apresentar sítios provavelmente sob efeito de seleção. O contraste destes com suas posições reais na estrutura do gene *fruitless* revelaram maiores aglomerações de sítios na região inicial do éxon C2, porção 5' do éxon C3, terço inicial e final do ZnF_D, região de transição entre os éxons C5/ZnF_C e extremidade 3' do ZnF_C (Anexo 3). Essas observações evidenciam regiões específicas do gene *fruitless* com maiores tendências a apresentarem mutações não-sinônimas ao invés de sinônimas, que podem ser explicadas, ao menos em parte, devido à seleção positiva ocorrendo preferencialmente nas regiões que não codificam domínios essenciais característicos desse gene.

5.7 O gene *fruitless* como marcador molecular

Além da disponibilidade de informações geradas por trabalhos anteriores realizados em nosso laboratório, interesse científico e curiosidade pessoal, outra prerrogativa para escolha desse gene foi a investigação do potencial de genes que afetam características reprodutivas para diferenciação molecular entre espécies com divergência recente e grande tamanho efetivo populacional, como é o caso do gênero *Anastrepha*.

Estudos *multilocus* indicam a majoritária retenção de polimorfismos ancestrais em grande parte de genomas analisados entre espécies proximamente relacionadas (Smith-Caldas e cols, 2001; Morgante e cols, 1980), no entanto alguns estudos realizados com genes relacionados à produção de feromônios (Dopman e cols, 2005), proteínas do sêmen (Andrés e cols, 1983; Maroja e cols, 2009) ou função espermatogênica (Ting e cols, 2000) tem apresentado variações exclusivas em regiões sob seleção positiva. Nesses casos, os polimorfismos ancestrais tendem a ser eliminados de maneira acelerada concomitantemente a fixação dessas regiões selecionadas contendo variações exclusivas em cada grupo ou espécie (Ting e cols, 2000), aumentando a taxa evolutiva e possivelmente afetando de alguma forma a ocorrência de isolamento reprodutivo. Tomando por base nossos resultados que demonstraram novos

sítios sob seleção, não refutamos a hipótese de que o gene *fruitless* possa vir a ser utilizado futuramente como um marcador molecular eficiente que ajude na elucidação do processo de especiação entre os insetos, em particular para nosso complexo de espécies de interesse chamado de *grupo fraterculus*, entretanto novas análises com foco populacional que envolvam toda a ampla região descrita nesse trabalho serão necessárias para uma avaliação precisa.

6. CONCLUSÕES

Os resultados obtidos aliados aos estudos desse importante gene de regulação do comportamento sexual revelam a conservação estrutural dos importantes domínios e manutenção da variabilidade e complexidade de transcritos produzidos por esse gene, dando suporte à hipótese de que *fruitless* é um gene de origem antiga que tem conservado suas funções nas diferentes espécies de insetos (Bertossa e cols, 2009; Ruiz e Sanchez, 2010), permanecendo praticamente inalterada a regulação de sua expressão baseada na estratégia de *splicing diferencial* nos últimos 350 milhões de anos (Salvemini e cols, 2010), podendo inclusive ser considerado o gene prototípico do comportamento sexual masculino entre os insetos dípteros dada a sua conservação estrutural no contexto evolutivo transcorrido ao longo de aproximadamente 250 milhões de anos que separam *Drosophila* e *Anopheles* (Gailey e cols, 2006). Entretanto, divergências consideráveis quanto ao nível de conservação e composição de cada um dos éxons são esperadas e realmente ocorreram em nossos dados, o que explica as diferenças espécie específicas e pode ajudar no entendimento da diversidade de comportamentos sexuais observada entre os insetos (Bertossa e cols, 2009).

De posse desses dados fomos capazes de descrever com detalhes a maior parte da região codificadora desse importante gene responsável pelo comportamento sexual da espécie em questão, principalmente no que diz respeito aos éxons expressos independentemente do sexo ao longo dos diferentes estágios de desenvolvimento. A não amplificação dos éxons específicos dos machos e das fêmeas acaba por proporcionar novas oportunidades para estudos subsequentes uma vez que a não obtenção dessa região pode ser explicada tanto por artefatos metodológicos como pela própria ausência da mesma em *Anastrepha sp*, fato este já identificado como recorrente em outros drosofilídeos e bastante caracterizado para duas espécies de drosofilídeos havaianos (*D.heteroneura* e *D.silvestris*), os quais não apresentam *splicing sexo-específico* e tem todos os seus transcritos iniciados diretamente no primeiro éxon C1 da região do BTB, o que hipoteticamente leva a uma mudança comportamental pouco

severa incapaz de prevenir o acasalamento e pode ter desencadeado o comportamento agressivo observado entre os machos dessas espécies (Davies e cols. 2000a e 2000b). Vale destacar que até a presente data, as únicas espécies dentre as que foram citadas neste trabalho que se tem confirmação do mecanismo de *splicing diferencial* para cada um dos sexos são: *Drosophila melanogaster*, *Anopheles gambiae*, *Nasonia vitripennis* e *Apis mellifera*.

Fomos capazes de identificar também diferenças significativas com relação aos comprimentos e composição dos éxons como é o caso do éxon D, relatado como raramente expresso e de comprimento bastante reduzido nos drosofilídeos, evidenciando sua conservação principalmente na região central enquanto que a porção terminal revela um número maior de sítios sob seleção. Além disso, confirmamos a presença de diversas isoformas que ainda que apresentem sequências diferentes seguem o padrão de expressão desse gene com uma forte tendência a conservação de estrutura primária e algumas vezes secundária na composição das regiões BTB, Conectora e que codificam os domínios *Zinc Finger* (Davies e cols, 2000).

Todos estes dados proporcionam uma caracterização pioneira a cerca da estrutura e mecanismos de expressão a respeito do gene *fruitless* e são de grande valia para a continuidade de estudos evolutivos, abrindo caminho para o esclarecimento de sua complexa conformação e formas de expressão, possibilitando a identificação de novas regiões não só em *A. obliqua*, mas também nas demais espécies próximas. Essa gama de novas informações auxilia no entendimento da biologia evolutivos deste grupo e possibilita análises mais complexas sobre o padrão de evolução molecular do gene *fruitless*.

Pudemos constatar que o gene *fruitless* realmente tem uma organização genômica bastante complexa quando comparado aos demais genes principais da cascata de determinação sexual (como o *doublesex* e o *transformer*), expressando um elevado número de isoformas distintas que provavelmente requerem uma minuciosa combinação para o desenvolvimento correto de certos circuitos neuronais, o controle dos aspectos de corte e o próprio ato de cópula. Sendo assim, é esperado que importantes

domínios tenham mantido conservado o cerne de suas estruturas funcionais, enquanto que regiões de transição ou contendo aminoácidos menos importantes para funcionalidade das proteínas produzidas pela expressão desse gene tenham sofrido menores restrições para se modificar.

A análise da identidade da região codificadora do gene *fruitless* revelou que a região BTB exibe a maior taxa de similaridade tanto nucleotídica quanto de aminoácidos entre todas as sequências dos insetos comparados; situação oposta ao éxon ZnF_D que apresentou a menor. Observamos também que, salvo poucas exceções, ocorre uma diminuição da taxa de identidade deste gene quando comparamos grupos filogeneticamente mais distantes, o que sugere uma separação mais recente, ou seja, menor tempo evolutivo hábil para alterações grosseiras, entre os grupos mais aparentados.

As comparações das sequências do gene *fruitless* de insetos provenientes de linhagens evolutivas bastante distintas mostram que apesar da alta conservação evolutiva do BTB e domínios de ligação do DNA, toda a região conectora apresenta elevada divergências entre as espécies estudadas (Gailey e cols, 2006; Bertossa e cols, 2009). Essa divergência corrobora os achados de Sobrinho e de Brito (2010) e pode ser decorrente de limitações seletivas relaxadas e/ou forças de seleção positiva que atuam sobre porções específicas do gene, sendo essas destacadas pela análise *BEB* (Tabela 4).

Tanto os testes de similaridade quanto os testes de seleção ao nível de espécies demonstraram que ocorreu diferenciação entre o *background* (Drosophilidae) e o *foreground* (Tephritidae) especialmente na região conectora em detrimento dos demais éxons desse gene, o que pode sugerir uma ausência de função primordial codificada por essa região ou ainda a presença de informações singulares de cada espécie. Essas conclusões se baseiam não somente na comparação entre *taxa* distantes apresentarem divergências elevadas decorrentes da longa história evolutiva transcorrida, mas também por episódios de evolução adaptativa observados entre linhagens de separação recente, o que sugere que as características sexuais influenciadas por esse gene, como a diferenciação sexual o comportamento de corte dos machos, estejam envolvidos de alguma forma nesse processo evolutivo.

Os dados gerados nesse trabalho não somente corroboram as inferências propostas por Sobrinho e de Brito (2010) como evidenciam a ocorrência de seleção positiva atuando sobre pontos específicos ao longo de todo o gene *fruitless* durante a diferenciação dos insetos. Foram encontrados nesse estudo 25 sítios sob efeito da seleção positiva, que somados aos 15 propostos pelos estudos precursores validam a hipótese da alta taxa evolutiva observada para regiões individuais deste gene.

A aplicabilidade de técnica de RACE-PCR para obtenção das sequências pretendidas não está totalmente comprovada uma vez que ambos os *kits* utilizados apresentaram grandes quantidades de amplificações espúrias mesmo utilizando posteriormente *primers* aninhados (*nested-PCR*), provavelmente isso se deve à falta de especificidade da reação de amplificação que é reflexo da utilização de apenas um *primer gene-específico*. Dessa forma, não nos surpreendemos com tais resultados e tivemos que lançar-mão de técnicas complementares em vista de alcançar os objetivos propostos inicialmente.

Através dessa pesquisa, esperamos ter elucidado grandemente a estrutura codificadora desse importante gene candidato, determinando regiões potencialmente importantes para o processo de diferenciação do grupo *fraterculus*, viabilizando a adequação do mesmo em um modelo evolutivo a ser considerado para futuras inferências filogenéticas, filogeográficas e de seleção positiva com abordagens populacionais, averiguando de maneira mais robusta seu potencial como marcador molecular para essas espécies.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- ALVAREZ, M.; RUIZ M. F.; SÁNCHEZ, L. Effect of the gene doublesex of *Anastrepha* on the somatic sexual development of *Drosophila*. **PLoS One**, v. 4, n. 4, 2009.
- ANAND, A. et al. Molecular genetic dissection of the sex-specific and vital functions of the *Drosophila melanogaster* sex determination gene fruitless. **Genetics**, v. 158, p. 1569-1595, 2001.
- ANDRÉS, J. A.; MAROJA, L. S.; HARRISON, R. G. Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. **Proceedings of the Royal Society B: Biological Sciences**, v. 275, p. 1975-1983, 2008.
- ARAUJO, E. L.; ZUCCHI, R. A. Medidas do acúleo na caracterização de cinco espécies de *Anastrepha* do grupo *fraterculus* (Diptera: Tephritidae). **Neotropical Entomology**, v. 35, p. 1-9, 2006.
- BERTOSSA, R. C.; VAN DE ZANDE, L.; BEUKEBOOM, L. W. The fruitless gene in *nasonia* displays complex sex-specific splicing and contains new zinc finger domains. **Molecular Biology and Evolution**, v. 26, p. 1557-1569, 2009.
- BORSON, N. D.; SALO, W. L.; DREWES, L. R. A lock-docking oligo(dT) primer for 5' and 3' race PCR. **PCR Methods and Applications**, v. 2, p. 144-148, 1992.
- BURTIS, K. C.; BAKER, B. S. *Drosophila* doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. **Cell**, v. 56, p. 997-1010, 1989.
- BURTIS, K. C. The regulation of sex determination and sexually dimorphic differentiation in *Drosophila*. **Current Opinion in Cell Biology**, v. 5, p. 1006-16, 1993.
- CHENCHIK, A.; MOQADAM, F.; SIEBERT, P. Marathon cDNA amplification: a new method for cloning full-length cDNAs. **Clontechiques**, v. 1, p. 5-8, jan. 1995.
- CHOMCZYNSKI, P.; SACCHI, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. **Analytical Biochemistry**, v. 162, p. 156-159, 1987.
- CARTER, R.; DROUIN, G. Structural differentiation of the three eukaryotic RNA polymerases. **Genomics**, v. 94, p. 388-396, 2009.
- CARVALHO, R. S. Avaliação das liberações inundativas do parasitóide exótico *Diachasmimorpha longicaudata* (Ashmead) (Hymenoptera: Braconidae) em pomar diversificado em Conceição do Almeida, BA. **Neotropical Entomology**, v. 34, p. 799-805, 2005.
- DAVIS, T.; KURIHARA, J.; YAMAMOTO, D. Genomic organisation and characterisation of the neural sex determination gene fruitless (*fru*) in the hawaiian species *Drosophila heteroneura*. **Genetics**, v. 246, p. 143-149, 2000 (a).

DAVIS, T.; KURIHARA, J.; YAMAMOTO, D. Genomic organisation of the neural sex determination gene fruitless (*fru*) in the Hawaiian species *Drosophila silvestris* and the conservation of the *fru* BTB protein-protein-binding domain throughout evolution. **Hereditas**, v. 132, p. 67–78, 2000 (b).

DEMIR, E.; DICKSON, B. J. Fruitless splicing specifies male courtship behavior in *Drosophila*. **Cell**, v. 121, p. 785–794, 2005.

DOPMAN, E. B. et al. Consequences of reproductive barriers for genealogical discordance in the European corn borer. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, p. 14706-14711, 2005.

EWING, B.; GREEN, P. Base-calling of automated sequencer traces using Phred.II. error probabilities. **Genome Research**, v. 8, p. 186-194, 1998.

FIETTO, J. L.; DEMARCO, R.; VERJOVSKI-ALMEIDA, S. Use of degenerate primers and touchdown PCR for construction of cDNA libraries. **Biotechniques**, v. 32, n. 6, p. 1404-8, jun. 2002.

FROHMAN, M. A.; DUSH, M. K.; MARTIN, G. R. Rapid production of full-length cDNAs from rare transcripts by amplification using a single gene-specific oligonucleotide primer. **Proceedings of the National Academy of Sciences of the United States of America**, v. 85, p. 8998–9002, 1988.

FROHMAN, M. A.; MARTIN, G. R. Rapid amplification of cDNA ends using nested primers. **Technique 1**, p. 165–173, 1989.

GAILEY, D. A. Functional conservation of the fruitless male sex-determination gene across 250 myr of insect evolution. **Molecular Biology and Evolution**, v. 23, p. 633-643, 2006.

GONÇALVES, V. R. et al. Transcriptome analysis of female reproductive tissues and molecular evolutionary studies of eggshell proteins in *Anastrepha fraterculus* fruit flies. Submetido em **Insect Biochemistry and Molecular Biology**, 2011.

GUBLER, U.; HOFFMAN, B. J. A simple and very efficient method for generating complimentary DNA libraries. **Genetics**, v. 25, p. 263–269, 1983.

GUINDON, S.; GASCUEL, O. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. **Systematic Biology**, v. 52, p. 696-704, 2003.

GREENSPAN, R. J.; FERVEUR, J. F. Courtship in *Drosophila*. **Annual Review of Genetics**, v. 34, p. 205-232, 2000.

HALL, J. C. The mating of the fly. **Science**, v. 264, p. 1702-44, 1994.

KIMURA, K. I. et al. Fruitless specifies sexually dimorphic neural circuitry in the *Drosophila* brain. **Nature**, v. 438, n. 10, 2005.

LAWRENCE, P.; JOHNSTON, P. The muscle pattern of a segment of *Drosophila* may be determined by neurons and not by contributing myoblasts. **Cell**, v. 45, p. 505-513, 1986.

- LIS, J. T.; SCHLEIF, R. Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. **Nucleic Acids Research**, v. 2, p. 383-389, 1975.
- MALAVASI, A.; ZUCCHI, R. A. **Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado**. Ribeirão Preto: Holos, 2000. 324 p.
- MALAVASI, A.; ZUCCHI, R. A.; SUGAYAMA, R. L. Biogeografia, In: MALAVASI, A.; ZUCCHI, R. A. **Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado**. Ribeirão Preto, Holos, 2000. p. 93-98.
- MAROJA, L. S.; ANDRÉS, J. A.; HARRISON, R. G. Genealogical discordance and patterns of introgression and selection across a cricket hybrid zone. **Evolution**, V. 63, p. 2999-3015, 2009.
- MORGANTE, J. S.; MALAVASI, A.; BUSH, G. L. Biochemical systematics and evolutionary relationships of neotropical anastrepha. **Annals of the Entomological Society of America (USA)**, v. 73, p. 622-630, 1980.
- NORRBOM, A. L.; KIM, K. C. A list of the reported host plants of the species of anastrepha (Diptera: Tephritidae). **United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarentine**, 1988. 114 p.
- PASCUAL, A. N. T. et al. Expressed sequences tags de tecidos reprodutivos de nachos de *Anastrepha fraterculus* (Tephritidae). In: XIV Congreso Latinoamericano de Genética (ALAG), 2010, Viña del mar, Chile.
- POND, S. L. K.; FROST, S. D. W.; MUSE, S. V. HyPhy: hypothesis testing using phylogenies. **Bioinformatics**, v. 21, p. 676-679, 2005.
- POSADA, D.; CRANDALL, K. A. Modeltest: testing the model of DNA substitution. **Bioinformatics**, v. 14, p. 817-818, 1998.
- PRIVÉ, G. G. et al. Sequence and structural analysis of BTB domain proteins. **Genome Biology**, v. 6, 2005.
- RESENDE, A. L. et al. Expressed sequence tags from the reproductive tract of *Anastrepha fraterculus*. In: 3rd Annual Arthropod Genomic Symposium, 2009, Kansas City. **KS**, 2009.
- RYNER, L. C. et al. Control of male sexual behavior and sexual orientation in *Drosophila* by the fruitless gene. **Cell**, v. 87, p. 1079-1089, 1996.
- ROSE, T. M. et al. Consensus-degenerate hybrid oligonucleotide *primers* for amplification of distantly related sequences. **Nucleic Acids Research**, v. 26, p. 1628-1635, 1998.
- ROZAS J.; LIBRADO, P. DNASP v5: a software for comprehensive analysis of DNA polymorphism data. **Bioinformatics**, v. 25, p. 1451-1452, 2009.

ROZEN, S.; SKALETZKY, H. J. Primer3 on the WWW for general users and for biologist programmers. In: KRAWETZ, S. A.; MISENER, S. **Bioinformatics Methods and Protocols: methods in molecular biology**. Totowa: Humana Press, 2000. p. 365- 386.

RUIZ, M. F.; SANCHEZ, L. Effect of the gene transformer of *Anastrepha* on the somatic sexual development of *Drosophila*. **International Journal of Developmental Biology**, v. 54, n. 4, p. 627-633, 2010.

SALVEMINI, M.; POLITO, C.; SACCONI, G. Fruitless alternative splicing and sex behaviour in insects: an ancient and unforgettable love story? **Journal of Genetics**, v. 89, p. 287–299, 2010.

SAMBROOK, J.; RUSSELL. **Molecular cloning: a laboratory manual**. New York: Cold Spring Harbor Laboratory Press, 2001. 2344 p.

SCOTTO–LAVINO, E.; DU, G.; FROHMAN, M. A. 5' end cDNA amplification using new RACE. **Nature Protocols**, 2006.

SELIVON, D.; PERONDINI, A. L. P.; MORGANTE, J. S. A genetic-morphological characterization of two cryptic species of the *Anastrepha fraterculus* complex (Diptera: Tephritidae). **Annals of the Entomological Society of America**, v. 98, p. 367-381, 2005.

SELIVON, D. Biologia e padrões de especiação. In: MALAVASI, A.; ZUCCHI, R. A. **Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado**. Ribeirão Preto: Holos, 2000. p. 25- 28.

SHIRANGI, T. R.; MCKEOWN, M. Sex in flies: what mind body dichotomy. **Developmental Biology**, v. 306, p.10-19, 2007.

SMITH-CALDAS, M. R. B. et al. Phylogenetic relationships among species of the fraterculus group (*Anastrepha*: Diptera: Tephritidae) inferred from DNA sequences of mitochondrial cytochrome oxidase I. **Neotropical Entomology**, v. 30, p. 565-573, 2001.

SOBRINHO JUNIOR, I. S.; BRITO, R. A. Marcadores moleculares para estudos populacionais de moscas-das-frutas do complexo fraterculus. In: 52º Congresso Brasileiro de Genética, 2006, Foz do Iguaçu.

SOBRINHO JUNIOR, I. S.; BRITO, R. A. Evidence for positive selection in the gene fruitless in *Anastrepha* fruit flies. **BMC Evolutionary Biology**, v. 10 p. 293, 2010.

SOLFERINI, V. N.; MORGANTE, J. S. Karyotype study of eight species of *Anastrepha* (Diptera, Tephritidae). **Caryologia**, v. 40, p. 229-241, 1987.

SONG, H. J. et al. The fruitless gene is required for the proper formation of axonal tracts in the embryonic central nervous system of *Drosophila*. **Genetics**, v. 162, p. 1703-1724, 2002.

SOSNOWSKI, B. A. et al. Multiple portions of a small region of the *Drosophila* transformer gene are required for efficient in vivo sex-specific regulated RNA splicing and in vitro sex-lethal binding. **Developmental Biology**, v. 161, p. 302-12, 1994.

TAYLOR, B. J. et al. Behavioral and neurobiological implications of sex-determining factors in *Drosophila*. **Developmental Genetics**, v. 15, p. 275-296, 1994.

TAYLOR, B. J.; KNITTEL, L. M. Sex-specific differentiation of a male-specific abdominal muscle, the muscle of Lawrence, is abnormal in hydroxyurea-treated and in fruitless male flies. **Development**, v. 121, p. 3079-88, 1995.

TING, C. T.; TSAUR, S. C.; WU, C. I. The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, p. 5313-5316, 2000.

USUI-AOKI, K. et al. Formation of the male-specific muscle in female *Drosophila* by ectopic fruitless expression. **Nature Cell Biology**, v. 2, p. 500-506, 2000.

YANG, Z.; SWANSON, W. J. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. **Molecular Biology and Evolution**, v. 19, n. 1, p. 49-57, 2002.

YANG, Z.; NIELSEN, R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. **Molecular Biology and Evolution**, v. 19, p. 908-917, 2002.

YANG, Z.; WONG, W. S. W.; NIELSEN, R. Bayes empirical Bayes inference of amino acid sites under positive selection. **Molecular Biology and Evolution**, v. 22, p. 1107-1118, 2005.

YANG, Z. PAML 4: phylogenetic analysis by maximum likelihood. **Molecular Biology and Evolution**, v. 24, p. 1586-1591, 2007.

ZHANG, J.; NIELSEN, R.; YANG, Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. **Molecular Biology and Evolution**, v. 22, p. 2472-2479, 2005.

ZUCCHI, R. A. **Taxonomia das espécies de *Anastrepha Schiner, 1868 (Diptera: Tephritidae) assinaladas no Brasil***. 1978. Tese (Doutorado em Entomologia) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1978.

ZUCCHI, R. A. Espécies de *Anastrepha*, sinónimas, plantas hospedeiras e parasitóides. In: MALAVASI, A.; ZUCCHI, R. A. **Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado**. Ribeirão Preto: Holos, 2000. p. 41-48.

ANEXO 1

Os *primers* em **vermelho** foram os que apresentaram melhores resultados de amplificação inicialmente e por isso foram escolhidos para serem utilizados na etapa inicial do trabalho (Quadro 1).

<i>RACE</i>	<i>Primer</i>	<i>Seqüência (5' - 3')</i>	<i>Tm</i>	<i>Ta</i>
5'	fru783r	taRttgaggttgttgttccgtgagacc	64.7° C	68.0° C
5'	fru837r	gtggcgaggatatggcgagtc	68.3° C	-
3'	fru1568f	gccggtggatcatataggtagtaataagcg	63.1° C	-
3'	fru1631f	tcgttgacctcgacacataacgcaaagtga	64.8° C	68.0° C

Quadro 1. Par de *Primers* específicos para o gene e suas condições otimizadas para RACE-PCR, onde “R” indica uma degeneração relativa a uma Adenina ou Guanina.

A) Novos *Primers* construídos para extensão da extremidade 3':

fru3100f	5' GGT CTC AGT CAA GCG CTT AG 3'	fru3002f	5' AGC ACG TCC GAA TCG GAG GA 3'
fru3611r	5' TCATCGTATTGCAGCCATTCAAATTG 3'	fru3739r	5' TGT GAT GAT TGA CAA TCT CCA GCC G 3'
fru44af	5' GGAGGACGCGATCATCACAGGTGTTGGC 3'	fru44bf	5' GGCGTGGGTAGCGGCAGTGTTAGTGCG '
fru3C3C4f	5' GGC AAA GGT ACA TGG GTC TCA CG 3'	fru3C3C4r*	5' AGT CGT GAT GAA CTC ATG CA 3'
BTB2f	5' ATG GAC CAG CAA TTC TGC TTA C 3'	fru33bf	5' CGC GAA CAC CGC CAA ATA CAA TGC 3'
fru33af	5'CAA CGA CCG CTA ACG CCT GCA GAC GC 3'	fru 1-	5' GTGTTACAGTGCGGCGTTAGC 3'

B) Novos *Primers* construídos para extensão da extremidade 5':

fru 2-	5' TCCTCCGCCGCCAGATAGT 3'	fru33r	5'TGTAGACTCTCGGCTGTTTTGAGGAACAT 3'
fru 1+	5' CGCCTGTGACACCAGACCA 3'	fru3C3C4r*	5' AGT CGT GAT GAA CTC ATG CA 3'

* O primer fru3C3C4r foi utilizado tanto para a amplificação da região intermediária entre os éxons C3 e C4, quanto para posterior amplificação de fragmentos localizados à montante (5') da região conhecida do gene *fruitless*.

Tanto os programas sugeridos pelos protocolos dos fabricantes dos *kits* utilizados, quanto variantes destes foram testados. Abaixo se encontram aqueles que produziram os melhores resultados:

PROGRAMA 1: *Kit Marathon* (Termociclador Hot-Lid): 5'RACE e 3'RACE

T = 94°C por 0:01:00

5 ciclos:

T = 94°C por 0:00:30

T = 72°C por 0:04:00

5 ciclos:

T = 94°C por 0:00:30

T = 70°C por 0:04:00

30 ciclos:

T = 94°C por 0:00:30

T = 72°C por 0:04:00

Manter 4°C

Fim

PROGRAMA 2: *Kit First Choice RLM-RACE* (Termociclador Hot-Lid): 5'RACE

T = 94°C por 0:03:00

35 ciclos:

T = 94°C por 0:00:30

Gradiente T = 55-65°C por 0:00:30

T = 72°C por 0:02:15 ou 0:04:30

1 ciclo:

T = 72°C por 0:07:00

Manter 4°C

Fim

PROGRAMA 3: *Kit First Choice RLM-RACE* (Termociclador Hot-Lid): 3'RACE

T = 94°C por 0:03:00

35 ciclos:

T = 94°C por 0:00:30

Gradiente T = 55-65°C por 0:00:30

T = 72°C por 0:02:15

1 ciclo:

T = 72°C por 0:07:00

Manter 4°C

Fim

ANEXO 2:

Alinhamento e identificação dos éxons das regiões já conhecidas entre *Drosophila melanogaster*, *Anastrepha obliqua*, *Drosophila silvestris*, *Bactrocera sp* e *Ceratitis sp*. Os primers utilizados e seus locais de ancoramento estão coloridos.

BTB (C1)

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
Dmelanogaster
Dmelanogaster
Aobliqua
AobliquaFB26ac.2
Dsilvestris
Dsilvestris
Bactrocera
Bactrocera
Ceratitis
Ceratitis

BTB2 (C2)

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
Dmelanogaster
Dmelanogaster
Aobliqua
AobliquaFB26ac.2
Dsilvestris
Dsilvestris
Bactrocera
Bactrocera
Ceratitis
Ceratitis

CONECTORA 1 (C3)

4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
Dmelanogaster
Dmelanogaster
Aobliqua
AobliquaFB26ac.2
Dsilvestris
Dsilvestris
Bactrocera
Bactrocera
Ceratitis
Ceratitis

CONECTORA 1 (C3)

6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
Dmelanogaster
Dmelanogaster
Aobliqua
AobliquaFB26ac.2
Dsilvestris
Dsilvestris
Bactrocera
Bactrocera
Ceratitis
Ceratitis

CONECTORA 2 (C4)

ZnF D

Sequence alignment for CONECTORA 2 (C4) and ZnF D. Includes species: Dmelanogaster, Aobliqua, AobliquaB8, AobliquaB5, AobliquaRaro (isof), Dsilvestris, Dsilvestris. Features fruC3C4r and 511bp intron.

ZnF D

Sequence alignment for ZnF D. Includes species: Dmelanogaster, Dmelanogaster, Aobliqua. Includes motif fruC3C4r.

ZnF D

Sequence alignment for ZnF D. Includes species: Dmelanogaster, Dmelanogaster, Aobliqua.

ZnF D

Sequence alignment for ZnF D. Includes species: Dmelanogaster, Dmelanogaster, Aobliqua.

ZnF D

Sequence alignment for ZnF D. Includes species: Dmelanogaster, Dmelanogaster, Aobliqua.

ZnF D

```

2
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Dmelanogaster P E I I T I N V G K N S A *
Dmelanogaster -----cccgaaattaccacataaatgtggcgaagaacagcgcatag
Aobliqua S |G| G G K R S G M A M D N E L G A T L R E T M T K R N N T P E V I T I T V G E D S T *
AGTGGTGGTGGTAAAGCAGTGGCATGGCAATGGATAACGAACCTTGGCCCACTTTACGTGAGACAATGACAAAACGCAACAACACACCTGAAGTGATCACTATCACTGTGGGAGGATAGCACGTGA

```

Conectora 3 (C5)

```

2
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Dmelanogaster S A A L K L H A E D M S T L L T Q H A L Q A A D A R D E H N D A K Q L Q L D Q T D N I D <| 3.6kb intron |>
Dmelanogaster agcgcgctctaaagctgcacgcggaggatgtcaacgctgctcacgcagcatgctttgcaagcagcagatgcgcgggacgcagcacaacgcgcgcaaacagctgcagctggaccagacggacaatcgac <| 3.6kb intron |>
Aobliqua P G A L K L H P E E M S Q L L S T H G L Q Q - D G R E D D Q N D G K Q L P L D Q S D N I D <| 3.6kb intron |>
Aobliqua cccggtgtttaaagctacatcccgaggaatgtcccaattattatcaacacatggcctgcagcaa - gatggtcgagaggtatcagaatgatggcaagcagctaccattggatcagctggacaacattgat <| 3.6kb intron |>
Dsilvestris T A A L K M H A E D M S T L L T Q H A L Q A A D A R D E H N D A K Q L P L D P T D N I D <| 3.6kb intron |>
Dsilvestris accgcgctctgaaaatgcacgcggaggatgtcaacgctgttaacccaacacgcactgcaggtgcgggatgagcagcaatgatgccaagctggccactggatccgacggacaacattgac <| 3.6kb intron |>

```

ZincFinger C (ZnFC)

```

3
4 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56
1 - S S A R H H L S - T P L S T S S S A S P P P P P P F G M H L S A A L K R E Y H P L H Y M A A G N G H N G P S A L G Y G N Q G S G N A P
Dmelanogaster ---agcagcggcccgccaccctgtcg - accccctgtgacactgtgctggcctgcccggcgcccccctttcgggatgcacctgtcggcgccctgaaacgcgagtgaccatccctgcaactataggcgcggcgaacgtcacaacggcccatggcgcttggatggcaatcagggatcgggcaatgcgccc
Aobliqua G Q V G R H H L S G T P L S T P S S C S P P P P P P P P H A F T V H L S A A L K N E Y H P L D Y I R A G G C S C G G N H N T N G G Q A
gggcaagtggagctcatcctctatccggcacaccactttcgaccacatcatcctgttaccacgcgcaccaccaccacgcgcagcgtttaccgctgactgagtgagcgcgcaaaacgaatatcatccattggactatatacagagcggcggtggcagcggcggtggcaatcacaacactaacgcggcgaagca

```

ZincFinger C (ZnFC)

```

Dmelanogaster N S A G G A G S V A G G V G A G G G A G G A T G A A G H N S H H T M S Y H N M F T P S R D P G T M W R C R S C G K E V T N R W H H F H
aatagtccggaggagctggatcgggttgcggcggaagtggagccggcggagcggcggagcgaactggagcagctggccataatctcgatcacaacctgtcgtaccacaacatgttcaacgcgtcccgcgatccgggcaccatgtggcggcggcctcctgcggcaaggaggtgaccaatcgctggcaccactttcac
Aobliqua S H L Y N Q S A N Q G G L - - G N G Q G G N S - A I V H N S H H Q M S Y H N M F T Q S R E P G T M W R C R S C G K E V T N R W H H F H
tcacatctatacaatcaatcggcgaacaaaggcggtcta-----ggtaacgctcaaggaggtaatagc---gccattgtacacaattcacatcatcaaatgtcctatcacaatattgttacacaactcaacgcaacggcgcccaaatgtggcgttgtcgctcctgcggcaaggaggttaccatcgatggcatcactttcat

```

ZincFinger C (ZnFC)

```

Dmelanogaster S H T A Q R S M C P Y C P A T Y S R I D T L R S H L R V K H P D R L L K L N S S I *
tccacaccccgccagcggctccatcctactgcccggccacctacagcaggtatgatcagctgctccatttgcgggtaagcatccggatcgcctgctcaagctgaactcatccatttaa
Aobliqua S H T A Q R S L C P Y C P A T Y S R I D T L R S H L R V K H P E R L L K L S P T L *
tcacacaccccgcccaagcttcaactatgccctactgtccggccacatacagctcgcatacattgctcaccactcagggtaaacatcccgaagattgctcaaatgagtccaaccttgtaa

```

ANEXO 3: Gráfico contendo os sítios relativos ao gene *fruitless* que apresentaram mais de 0.5 de probabilidade de estarem evoluindo sob seleção positiva ($\omega > 1$). A linha vermelha delimita os sítios para os quais a confiabilidade superou 0.95.

