

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CURSO DE BACHARELADO EM CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE GENÉTICA E EVOLUÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA EVOLUTIVA E
BIOLOGIA MOLECULAR

MARIANA DE CÁSSIA BISIO

**EVOLUÇÃO MOLECULAR E CARACTERIZAÇÃO DO GENE
DEFECTIVE CHORION 1 EM *ANASTREPHA FRATERCULUS*
E *ANASTREPHA OBLIQUA* (DIPTERA, TEPHRITIDAE)**

SÃO CARLOS – SP

JUNHO/2023

MARIANA DE CÁSSIA BISIO

**Evolução molecular e caracterização do gene *defective chorion-1*
em *Anastrepha fraterculus* e *Anastrepha obliqua* (Diptera,
Tephritidae)**

Dissertação apresentada ao Programa de Pós-graduação em Genética Evolutiva e Biologia Molecular da Universidade Federal de São Carlos, *campus* São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Genética e Evolução.

Orientador: Prof. Dr. Reinaldo Alves de Brito

São Carlos - SP

Junho/2023



Folha de Aprovação

Defesa de Dissertação de Mestrado da candidata Mariana de Cássia Bisio, realizada em 29/06/2023.

Comissão Julgadora:

Prof. Dr. Reinaldo Otavio Alvarenga Alves de Brito (UFSCar)

Profa. Dra. Maura Helena Manfrin (USP)

Profa. Dra. Mariana Freitas Nery (UNICAMP)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular.

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer a toda minha família pelo apoio incondicional e por toda a ajuda durante a minha caminhada profissional, principalmente à minha irmã Vanessa e aos meus sobrinhos Luís Otávio e João Henrique. Agradeço especialmente aos meus pais e parceiros de vida, Auro e Maria de Lourdes, pelos conselhos, ensinamentos e companheirismo, e pelo auxílio na superação dos obstáculos da vida sem abandonar minhas origens e meus princípios.

Ao meu orientador Prof. Dr. Reinaldo Alves de Brito, por toda ajuda, orientação e paciência durante todo esse tempo de convivência e trabalho, e por manter as portas do seu laboratório sempre abertas desde o início.

À minha amiga e colega de trabalho Samira Chahad-Ehlers, por todo apoio e suporte durante todo o meu período de mestrado. Obrigada por compartilhar suas experiências e por me ajudar a ser uma profissional melhor e a evoluir a cada dia.

À profª Drª Andrea Peripato, por toda parceria, paciência e ensinamentos na área da educação, e por me proporcionar experiências profissionais novas. Foi um imenso prazer trabalhar com você.

Ao meu namorado Rogério, por todo o companheirismo e apoio durante esses tempos difíceis de pandemia que enfrentamos, e por me ajudar a transformá-los em momentos de muita alegria, acolhimento e construção pessoal.

Aos meus amigos e colegas de trabalho do Laboratório de Genética de Populações e Evolução Maria Isabel, Beatriz, Bruna e Bruno, por toda a ajuda e apoio durante todas as etapas do meu trabalho, pelas risadas e conversas, nada disso seria possível sem vocês.

Aos meus grandes amigos e irmãos que a vida me deu, por me apoiarem em meus melhores e piores momentos, e por tudo o que compartilhamos juntos.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 88887.497759/2020-00.

RESUMO

A família Tephritidae compreende muitas espécies de moscas-das-frutas com larvas predominantemente frugívoras, diversas das quais de grande importância econômica. O grupo *fraterculus* é o mais importante grupo de espécies, pois inclui algumas das espécies economicamente mais significativas que causam prejuízos à fruticultura no Brasil e no mundo. Algumas espécies desse grupo são espécies taxonomicamente relacionadas que infestam diversos frutos hospedeiros, possuem ampla distribuição geográfica e mostram considerável similaridade e proximidade evolutiva, o que dificulta sua identificação. Portanto, a investigação de genes com altas taxas evolutivas, como o *defective chorion 1 (dec-1)*, pode ser relevante para uma melhor compreensão da filogenia do grupo, bem como das forças evolutivas envolvidas em sua diferenciação. O gene *dec-1* está envolvido com a produção e modificação de proteínas específicas que participam da formação do córion do ovo durante a oogênese e seu estudo pode aumentar nosso conhecimento sobre a diferenciação das espécies de *Anastrepha*. Este trabalho objetiva investigar a estrutura molecular e os padrões evolutivos de *dec-1* para entender o processo de diferenciação de espécies do gênero *Anastrepha* e entre espécies de Tephritidae. Um fragmento específico do *dec-1* de duas espécies de *Anastrepha* (*A. fraterculus* e *A. obliqua*) foi amplificado a partir de amostras coletadas na natureza e comparado com dados de sequências de transcriptomas de última geração disponíveis. A PCR convencional foi realizada com *primers* específicos do *dec-1* e os fragmentos foram submetidos ao sequenciamento Sanger. Para a predição da estrutura, a sequência do gene *dec-1* do genoma de *Anastrepha ludens* foi usada como modelo. Um conjunto de supostos ortólogos para este gene em outras moscas proximamente relacionadas a essas espécies foi recuperado do NCBI para testar assinaturas de seleção natural em clado e ramo específico com os softwares PAML e HyPhy. A fim de caracterizar o desenvolvimento ovariano, ovários de fêmeas da espécie *A. fraterculus* foram coletados em diferentes idades, medidos e submetidos a análises estatísticas. A caracterização ovariana mostrou que a coriogênese provavelmente ocorre em moscas de *A. fraterculus* dos dias 10 ao 14 da vida adulta, período no qual o produto da transcrição do gene *dec-1* seria clivado em isoformas e derivados proteicos. Os dados coletados dos ovários mostraram ser informativos para fundamentar e guiar estudos acerca da caracterização dos desenvolvimentos ovariano e reprodutivo de populações desta espécie. Uma combinação de dados do transcriptoma com procedimentos clássicos de sequenciamento permitiu identificar três transcritos alternativos em *A. fraterculus*

e *A. obliqua*, que corroboram o padrão encontrado em outros tefritídeos e recupera três dos seis transcritos conhecidos de *Drosophila melanogaster*. No geral, os resultados são consistentes com ação predominante de seleção purificadora tanto na estrutura do gene quanto nos ramos estudados. Apesar das evidências de alto grau de conservação, alguns códons e ramificações mostraram sinais significativos de seleção positiva, observados em níveis hierárquicos distintos de separação entre moscas de cada uma das espécies estudadas, e também de sua separação com outros tefritídeos e dípteros. As análises de algumas alterações selecionadas positivamente revelaram que as substituições de aminoácidos levaram a sete alterações radicais em suas propriedades físico-químicas. Esses resultados podem indicar que a seleção positiva afetou essas regiões durante um tempo evolutivo específico e foi fixada por conferir alguma vantagem adaptativa. Assim, este gene demonstrou ter um possível papel no processo de diferenciação em *Anastrepha*.

Palavras-chave: *dec-1*; *Anastrepha*; caracterização gênica; diferenciação

ABSTRACT

The Tephritidae family comprises many fruit fly species with predominance of frugivorous larvae, several of which of great economic importance. The *fraterculus* group is the most important species group, because it includes some of the most economically important species that cause damage to fruit production in Brazil and worldwide. Some species in this group are closely related, infest several host fruits, have a wide geographical distribution, and show considerable similarity and evolutionary proximity, which renders their identification difficult. Therefore, the investigation of genes with high evolutionary rates, such as the *defective chorion 1* (*dec-1*), may be relevant for a better understanding of their phylogeny and the underlying differentiation forces involved. The gene *dec-1* is involved with the production and modification of specific proteins that participate in the egg chorion formation during oogenesis. Evidence about this gene may bring relevant information on the differentiation of *Anastrepha* species. This work aims to investigate the molecular structure and the evolutionary patterns of *dec-1* in order to understand the differentiation process of species in the *Anastrepha* genus, as well as among Tephritidae. A specific fragment of the *dec-1* from two species of *Anastrepha* (*A. fraterculus* and *A. obliqua*) was amplified from samples collected in nature and compared to transcriptome sequences from next-generation data available in our laboratory. Conventional PCR was performed with specific *dec-1* primers and fragments were submitted to Sanger sequencing. For structure prediction, *dec-1* gene sequence of *Anastrepha ludens* was used as model. A group of orthologous sequences for this gene from other related flies was retrieved from NCBI and used to test for signatures of natural selection using clade and branch-specific methods with PAML and HyPhy software. In order to characterize the ovarian development, female ovaries of the species *A. fraterculus* were collected at different developmental periods, measured and submitted to statistical analysis. The ovarian characterization showed that choriogenesis probably occurs in *A. fraterculus* flies from 10 to 14 days of adult life, a period in which *dec-1* gene products would be cleaved into isoforms and protein derivatives. The data collected from the ovaries proved to be informative to base and guide studies about ovarian characterization and reproductive development of *A. fraterculus* populations. A combination of transcriptome data

with classical sequencing procedures allowed the identification of three alternative transcripts in both *A. fraterculus* and *A. obliqua*, corroborating the pattern found in other tephritids and recovering three of the six known *D. melanogaster* transcripts. Overall, there is evidence of purifying selection having a preponderant role both, for gene structure as well as along evolutionary branches. Despite evidence of a significant evolutionary conservation, some codons and branches under positive selection were detected, such as in the branch that separates *A.obliqua* flies, as well as on the separation of *A. fraterculus* into three clusters, and on the separation of Tephritidae from other Diptera. Analyses on some positively selected changes revealed seven non-synonymous changes that led to radical amino acid replacements which altered physicochemical properties. These results indicate that positive selection affected these regions during a specific evolutionary time and were fixed in different lineages, perhaps because it conferred some adaptive advantage. Thus, this gene showed some positively selected changes that seem to have been important to differentiate species in the group, suggesting that it might have a possible role in their differentiation process.

Key words: *dec-1*; *Anastrepha*; gene characterization; differentiation

LISTA DE FIGURAS

- Figura 1:** Comparação entre as estruturas do córion encontradas para *Drosophila melanogaster* (a) e para *Anastrepha sororcula* (b). Modificado de Figueiredo e colaboradores (2016). 16
- Figura 2:** Esquema das etapas do processo de oogênese. Em verde está representada a etapa desse processo na qual ocorre a formação da casca do ovo, e em laranja, a etapa na qual o gene *dec-1* é transcrito, de acordo com o proposto para *Drosophila melanogaster*. 18
- Figura 3:** Estrutura do gene *dec-1* em *Drosophila melanogaster*, em *Rhagoletis zephyria* e a produção de três isoformas na mosca-das-frutas *Anastrepha fraterculus*. Os seis éxons detectados são representados proporcionalmente por cores diferentes. O éxon 5 (em amarelo) mostra os sítios de *splicing* internos (setas). As posições dos íntrons (linhas pretas) são previsões baseadas no genoma de *Anastrepha ludens*. UTR, região não traduzida: 5' e 3'; códon de iniciação (ATG); códon de terminação (*). 36
Nenhuma entrada de índice de ilustrações foi encontrada.
- Figura 4:** Árvores filogenéticas dos grupos “*Anastrepha*”, “*Tephritidae*” e “*Cyclorrhapha*”. Os ramos em rosa foram submetidos a análise de seleção de ramos. 41
- Figura 5:** Análise gráfica do *Bayes Empirical Bayes (BEB)* mostrando probabilidades posteriores de sítios sob seleção purificadora ($0 < \omega < 1$), positiva ($\omega > 1$) e neutra ($\omega=1$)... 44
- Figura 6:** Pares de ovários de moscas de idades distintas de *Anastrepha fraterculus*. Diferenças morfológicas entre os ovários de 1, 5, 7, 10, 14, 15, 20, 25 e 29 dias (a). Características observadas durante a dissecação (b). Em B5 são apresentadas as estruturas: ovários (1), espermatecas (2), glândulas acessórias (3), oviduto lateral (4) e oviduto mediano (5). 53
- Figura 7:** Gráficos boxplot dos valores obtidos para o Índice ovariano (a) e a Carga de ovos (b) das fêmeas de *Anastrepha fraterculus* dissecadas para as diferentes idades. Os grupos I (roxo), II (rosa), III (azul) e IV (laranja) foram formados de acordo com presença ou ausência de diferenças estatísticas significativas entre classes. 56
- Figura 8:** Domínios encontrados nos alinhamentos das isoformas FC106, FC125 e FC177 de *A. fraterculus* e *A. obliqua*. Estão destacados o domínio *Pfam* ID: DEC-1_N (Description: PF04625, DEC-1 protein, N-terminal region) em amarelo, o *Pfam* ID: DEC-1_C (Description: PF04626, DEC-1 protein, C-terminal region) em verde, e o *Pfam* ID: Rpr2 (Description: PF04032, RNase PRpr2/Rpp21/SNM1 subunit domain) em roxo. 73

LISTA DE TABELAS

Tabela 1: Amostras de <i>Anastrepha obliqua</i> e <i>Anastrepha fraterculus</i> coletadas em diferentes localidades e utilizadas na análise populacional.	26
Tabela 2: Arranjo dos éxons e CDS (<i>Coding DNA Sequence</i>), e respectivas proteínas (aminoácidos) das três isoformas em ordem crescente de tamanho (X3, X2 e X1) detectados no <i>dec-1</i> das moscas-das-frutas <i>Anastrepha obliqua</i> e <i>Anastrepha fraterculus</i> . As letras a, b e c representam os pontos de <i>splicing</i>	37
Tabela 3: Percentuais de identidade dos domínios putativos do <i>dec-1</i> (isoforma FC106 ou X3) entre espécies de <i>Anastrepha</i> e outros ortólogos em Cyclorrhapha da ordem Diptera.....	38
Tabela 4: Resultados do teste de Xia para a saturação das taxas de substituições..	39
Tabela 5: p-valor das diferenças entre as variâncias e médias dos ovários esquerdo e direito para verificação de simetria bilateral.	54
Tabela 6: Coeficientes de correlação de Pearson e p-valor para cada variável analisada.	59
Tabela 7: Lista de ortólogos da proteína do <i>defective chorion-1 (dec-1)</i> (isoforma FC106) de <i>A. fraterculus</i> e <i>A. obliqua</i> em Diptera, que estão publicamente disponíveis no NCBI GenBank.....	72
Tabela 7: Substituições de aminoácidos encontradas pela análise BEB em sequências dos grupos de análise “ <i>Anastrepha</i> ”, “ <i>Tephritidae</i> ” e “ <i>Cyclorrhapha</i> ”.....	79
Tabela 9: Testes de normalidade de Ryan-Joiner para o Índice ovariano (I_o), para a Carga de ovos (C_o), e para os comprimentos da asa (hA) e do tórax (hT), considerando a presença de simetria bilateral entre ovários esquerdo e direito.	80
Tabela 10: p-valores obtidos a partir dos testes de comparação entre a diferença das variâncias (a) e a igualdade das médias populacionais (b). Os valores de variâncias e médias maiores que 0,05 não rejeitam H_0 e são estatisticamente iguais (cor verde), enquanto os menores que 0,05 rejeitam H_0 e são estatisticamente diferentes (cor vermelha).	81
Tabela 11: p-valores obtidos a partir do teste t de igualdade de médias populacionais para a hipótese alternativa (H_1). Os valores menores que 0,05 não rejeitam H_1 e são estatisticamente iguais, enquanto os maiores que 0,05 rejeitam H_1 e são estatisticamente diferentes.	82

SUMÁRIO

1. INTRODUÇÃO	12
1.1 O gênero <i>Anastrepha</i>	12
1.2 Genes coriônicos e evolução molecular.....	13
1.3 Estruturas macromoleculares do ovo	15
1.4 Oogênese e gene <i>dec-1</i>	17
1.5 Função e importância do <i>dec-1</i> e suas pró-proteínas	19
2. OBJETIVOS	23
3. MATERIAIS E MÉTODOS.....	24
3.1 Identificação do gene <i>defective chorion 1 (dec-1)</i> em <i>Anastrepha</i>	24
3.2 Caracterização e estrutura molecular do <i>dec-1</i>	25
3.3 Procedimentos moleculares e testes de seleção	26
3.3.1 Populações Experimentais	26
3.3.2 Extração de DNA e amplificação por PCR.....	27
3.3.3 Clonagem e Sequenciamento	27
3.3.4 Alinhamentos	28
3.3.5 Testes filogenéticos e de seleção	29
3.3.6 Medição da magnitude físico-química das substituições de aminoácidos	31
3.4 Caracterização ovariana	31
3.4.1 Amostras utilizadas	31
3.4.2 Estudos morfológicos	32
3.4.3 Análise estatística.....	32
4. RESULTADOS E DISCUSSÃO.....	34
4.1 Caracterização do gene <i>defective chorion 1</i>	34
4.2 Testes filogenéticos e de seleção	39
4.2.1 Testes filogenéticos	39
4.2.2 Testes de seleção: análises globais do <i>dec-1</i>	42
4.2.3 Testes de seleção: análises de seleção de ramos	46
4.3 Medição da magnitude físico-química das substituições de aminoácidos	49
4.4 Caracterização ovariana	52
4.4.1 Caracterização morfológica	52
4.4.2 Análises estatísticas.....	54
5 CONCLUSÕES.....	62
6 REFERÊNCIAS BIBLIOGRÁFICAS	63
APÊNDICE A	72
APÊNDICE B.....	79
APÊNDICE C.....	80

1. INTRODUÇÃO

1.1 O GÊNERO *ANASTREPHA*

A família de moscas-das-frutas Tephritidae abriga cerca de 4.300 espécies descritas divididas em 500 gêneros de ocorrência mundial, dentre as quais algumas que constituem algumas pragas de maior interesse econômico (White, 2006; Raza et al., 2020), pertencentes aos gêneros *Anastrepha*, *Bactrocera*, *Ceratitis* e *Rhagoletis* (Zucchi et al., 2004). O gênero *Anastrepha* é o maior da tribo Toxotrypanini (Silva and Barr, 2006), com mais de 20 grupos taxonômicos e mais de 300 espécies descritas nos trópicos e subtropicais do continente americano (Norrbom et al., 2018; 2021). Dentre elas, há o grupo *fraterculus* que consiste em 34 espécies formalmente descritas até o momento (Norrbom et al., 2012), incluindo a mosca-das-frutas sul-americana *Anastrepha fraterculus* (Wiedmann, 1830) e a mosca-das-frutas das Índias Ocidentais *A. obliqua* (Macquart, 1835).

As espécies *A. fraterculus* e *A. obliqua* são duas das principais pragas agrícolas das Américas e infestam diversos frutos hospedeiros, possuindo ampla distribuição geográfica (Zucchi, 2000; Solferini and Morgante 1987), o que inviabiliza a exportação de frutos e prejudica as atividades comerciais do setor da fruticultura (Zucchi et al., 2004). A espécie *A. fraterculus* é composta por um complexo de espécies (o complexo *fraterculus*) potencialmente formado por pelo menos sete morfotipos, sendo que três deles estão distribuídos no Brasil (Hernández-Ortiz et al., 2012), afetando frutos de maior relevância econômica (Hendrichs et al., 2015). Estes morfotipos foram definidos e caracterizados por meio de análises morfométricas lineares de indivíduos adultos de populações do México, América Central e América do Sul (Hernández-Ortiz et al., 2012), e sua existência ainda não é bem suportada devido às dificuldades em se estabelecer o *status* taxonômico de cada um deles geradas pelas consideráveis divergências interpopulacionais mostradas por análises de sequências de DNA e compatibilidade sexual, e pelas incertezas acerca da distribuição geral destes morfotipos, principalmente os brasileiros e dos Andes Norte e Central (Hernández-Ortiz et al., 2015).

A possível existência de morfotipos e a similaridade entre as espécies do grupo *fraterculus* implica na importância de se considerar também os aspectos genéticos e a incompatibilidade sexual para a identificação dessas espécies (Hernández-Ortiz et al., 2012), que é um procedimento complexo e contribui com a aquisição de informações que auxiliam o desenvolvimento de técnicas de combate a essas pragas agrícolas (Dyck et al., 2005). Desse

modo, a identificação de morfotipos demanda diferentes abordagens integradas (Schutze et al., 2017), com o uso de análises moleculares.

Portanto, a compreensão de eventos de diferenciação em moscas-das-frutas tem se tornado cada vez mais importante para o planejamento e estudos de estratégias de controle populacional (Ruiz-Montoya et al., 2020), como o isolamento reprodutivo que pode levar grupos de espécies à especiação em decorrência do impedimento do fluxo gênico entre elas (Butlin e Ritchie, 2001; Wu e Ting, 2004), e para os estudos sobre supostos "genes de especiação". Estes genes constituem formas alélicas distintamente adaptadas e não integráveis aos genomas de diferentes populações (Wu, 2001) que estão associados, ou que caracterizam, diferentes linhagens evolutivas. A família de genes coriônicos é um exemplo entre potenciais genes candidatos a participar do processo de diferenciação de algumas espécies, como evidenciado na família *Drosophilidae* (Eickbush e Izzo, 2009) e até mesmo em espécies de *Anastrepha* (Gonçalves et al., 2013).

1.2 GENES CORIÔNICOS E EVOLUÇÃO MOLECULAR

A família multigênica dos genes coriônicos codifica proteínas estruturais denominadas proteínas coriônicas, que são expressas de forma sequencial durante os estágios da coriogênese (Nadel e Kafatos, 1980; Papantonis et al., 2015), formando uma única estrutura macromolecular denominada córion em um processo combinado (Eickbush e Izzo, 2009). É uma família gênica bem caracterizada em evolução e desenvolvimento, pois possui propriedades compartilhadas e únicas, como por exemplo, evolução facilitada por sequências diferentes entre si (Jones e Kafatos, 1980) e evolução rápida em alguns de seus componentes, pelos quais fragmentos dessas sequências podem ser comuns a indivíduos de uma espécie, mas não a indivíduos de espécies distintas (Eickbush e Izzo, 2009).

As proteínas coriônicas são abundantes e codificadas por genes estruturais, os genes coriônicos, que são expressos em níveis elevados em um programa temporalmente regulado (Spradling e Mahowald, 1980), e participam da formação do córion diretamente com a localização dessas proteínas na parte externa do ovo pela expressão regulada e temporalmente limitada (Parks e Spradling, 1987). Em um subgrupo de *Drosophila melanogaster*, alguns destes genes estão envolvidos em processos adaptativos ao ambiente, nos quais as proteínas coriônicas da membrana externa do ovo (córion) evoluem mais rapidamente em comparação com a genes que formam a membrana interna (vitelínica), por estarem sob seleção positiva (Jagadeeshan e Singh, 2007). A membrana vitelínica é responsável pela proteção do ovo

contra fatores externos (Gonçalves et al., 2013) e é formada pelas proteínas vitelínicas que, por sua vez, são codificadas pela família dos genes vitelínicos (Jagadeeshan e Singh, 2007). São proteínas que participam da formação de proteínas precursoras de gema como a Vitelogenina (Byrne et al., 1989), e desempenham funções imunológicas e relacionadas à longevidade (Amdam et al., 2003).

Um estudo que investigou genes expressos em tecidos reprodutivos de machos e fêmeas de *A. fraterculus* e *A. obliqua* revelou que o gene *defective chorion 1* apresenta alta taxa evolutiva entre elas (Rezende et al., 2016), o que foi também encontrado para um estudo mais amplo com diversas espécies de *Anastrepha*, mesmo entre espécies de grupos filogeneticamente próximos, com razão entre taxas de substituição dN/dS acima de 0,6 (Congrains et al., 2018), indicando considerável relevância desse gene no contexto de espécies crípticas no grupo *fraterculus*.

A relação dN/dS (taxa de substituições não-sinônimas/taxa de substituições sinônimas) é um dos métodos utilizados para investigar se sequências genéticas codificantes evoluem sob seleção positiva ($dN/dS < 1$), neutra ($dN/dS = 1$) ou purificadora ($dN/dS > 1$) (Kimura, 1977; Nielsen, 2005). A pressão de seleção pode impulsionar diversificações na sequência do gene e seus produtos proteicos (Findlay e Swanson, 2010). Portanto, o estudo de genes e proteínas reprodutivos contribuem com a maior compreensão sobre esse processo, principalmente por explicitar as forças evolutivas envolvidas na diferenciação de espécies (Swanson e Vacquier, 2002), como as espécies crípticas e de considerável proximidade evolutiva do grupo *fraterculus*, especialmente por meio da genômica comparativa (Findlay e Swanson, 2010), cujas análises acerca dos componentes genéticos das diferenças fenotípicas têm o potencial de auxiliar na identificação de genes envolvidos no processo de especiação (Zeng et. al., 2000).

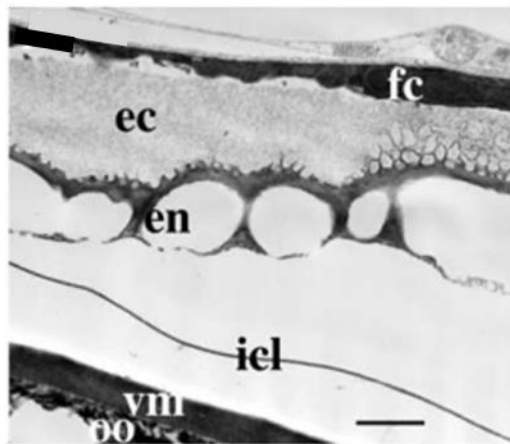
Ademais, os estudos sobre os genes e proteínas reprodutivos também contribuem com o esclarecimento da influência de suas divergência e diversidade no processo de especiação (Swanson e Vacquier, 2002), já que estes evoluem tanto por eventos de substituição quanto por mutações de inversão, duplicação ou deleção (Jones e Kafatos, 1980), permitindo investigar quais forças seletivas influenciam o seu processo evolutivo, o que pode contribuir com a compreensão da sua influência na diferenciação em moscas-das-frutas do gênero *Anastrepha*.

1.3 ESTRUTURAS MACROMOLECULARES DO OVO

A oviposição em substratos distintos expõe os ovos de insetos a condições potencialmente estressantes, por isso são estruturas fechadas internamente pela casca, que por sua vez deve permitir que as trocas gasosas e de umidade ocorram normalmente entre o embrião e o meio externo (Figueiredo et al., 2016). A Figura 1 mostra estruturas das cascas de ovos em *D. melanogaster* (a) e em *Anastrepha sororcula* (b). Em *Drosophila* (Figura 1.a), a casca do ovo é uma matriz extracelular especializada que se forma durante os últimos estágios da oogênese entre o oócito e as células somáticas sobrepostas (Mauzy-Melitz e Waring, 2003). É uma estrutura composta pelas camadas proteicas denominadas membrana vitelínica e córion (Noguerón e Waring, 1995), que conferem proteção ao embrião contra fatores ambientais adversos (Figueiredo et al., 2016).

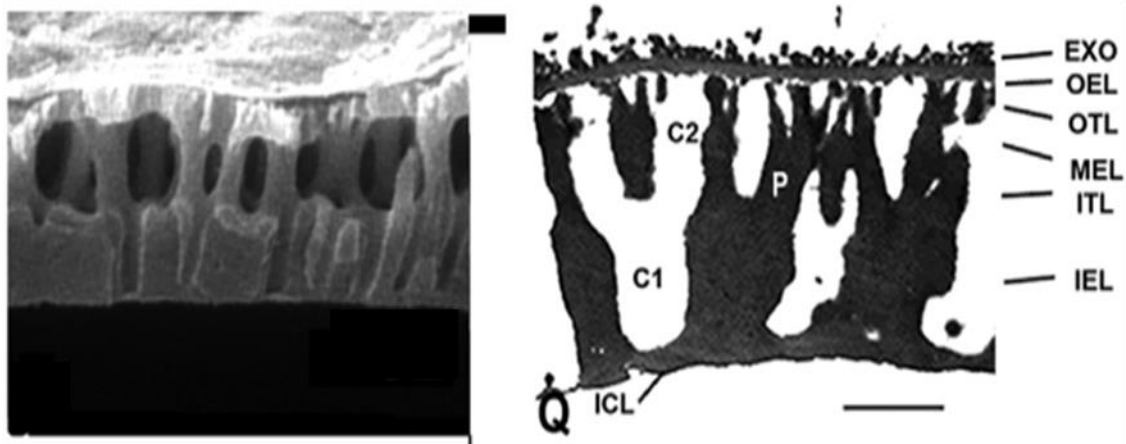
A membrana vitelínica envolve o embrião internamente e consiste em uma membrana mais fina e menos diversa (Figueiredo et al., 2016), que propicia proteção geral e elementos nutricionais (Papantonis et al., 2015). Por representar uma barreira física impenetrável, pode fornecer não só um mecanismo de comunicação entre as células foliculares e o oócito, mas também uma superfície de deposição de partículas do córion e de armazenamento de produtos que atuam na formação dos padrões embrionários (Noguerón et al., 2000).

(a) *Drosophila melanogaster*



fc: células foliculares
 ec: exocórion
 en: endocórion tripartido
 icl: membrana coriônica interna
 vm: membrana vitelínica
 oo: oócito

(b) *Anastrepha sororcula*



ICL: membrana endocoriônica mais interna
 ITL: membrana trabecular interna
 IEL: membrana endocoriônica interna
 MEL: membrana endocoriônica média
 OTL: membrana endocoriônica externa
 OEL: membrana trabecular externa

Figura 1: Comparação entre as estruturas do córion encontradas para *Drosophila melanogaster* (a) e para *Anastrepha sororcula* (b). Modificado de Figueiredo e colaboradores (2016).

Por outro lado, o córion envolve a membrana vitelínica e o embrião externamente e é subdividido em várias camadas justapostas que além de facilitar as trocas gasosas com o meio externo (Figueiredo et al., 2016), conferem proteção mecânica e contra a dessecação (Chapman, 2013). São elas: camada coriônica interna (*Inner Chorion Layer* ou ICL), endocóron tripartido (Noguerón e Waring, 1995), e exocóron amorfo (Margaritis, 1985; Figueiredo et al., 2016). O endocóron tripartido de *Drosophila* é organizado em superfície, pilares e telhado reticulado (Noguerón e Waring, 1995), enquanto o exocóron é uma estrutura fibrosa quimicamente mal definida e secretada sobre o endocóron no estágio 14 da oogênese (Mauzy-Melitz e Waring, 2003).

Em *Anastrepha* em geral, e no grupo *fraterculus* em particular (Figura 1.b), o córion é mais complexo, inclusive quando comparado ao de outros tefritídeos (Figueiredo et al., 2016). Estudos de microscopia eletrônica com ovos de diversos grupos do gênero *Anastrepha* mostraram um endocóron de estrutura básica formada por 6 camadas: a membrana endocoriônica mais interna (ICL), a membrana trabecular interna (ITL), a membrana endocoriônica interna (IEL), membrana endocoriônica média (MEL), a membrana endocoriônica externa (OTL) e a membrana trabecular externa (OEL) (Figueiredo et al., 2016).

A ICL é uma membrana fina e contínua que fica justaposta à IEL, que representa uma camada mais grossa e que conecta a ICL às cavidades da ITL. Os pilares da ITL se expandem e se fundem na parte superior para formar a MEL, que não é uma camada contínua, mas sim dividida por canais. A MEL, por sua vez, possui pilares protetores que formam o exterior da OTL, e que se fundem em filamentos finos para compor a rede contínua do OEL. A OEL consiste na camada mais superficial do endocóron e é recoberta pelo exocóron (a cobertura externa do ovo) (Figueiredo et al., 2016).

A membrana vitelínica e o córion são tecidos reprodutivos importantes na proteção do embrião (Wei et al., 2019), e são sintetizados nos últimos estágios da oogênese em um processo complexo de montagem macromolecular regulado pelo desenvolvimento (Hawley e Waring, 1988) e pela expressão de genes e proteínas vitelínicas e coriônicos (Jagadeeshan e Singh, 2007).

1.4 OOGÊNESE E GENE *DEC-1*

Em *Drosophila*, a síntese da casca do ovo ocorre dos estágios 8 ao 14 do processo de oogênese, no qual há ativação transcricional programada de genes vitelínicos e coriônicos

(Tootle et al., 2011) e a provável deposição sequencial de camadas para formar a membrana vitelínica e o córion (Giorgi, 1977) (Figura 2). Durante os estágios 9 e 10 (Mauzy-Melitz and Waring, 2003), os folículos ovarianos recebem o sinal interno de um hormônio esteróide e iniciam a absorção da gema e a síntese da membrana vitelínica, o que marca também o início da vitelogênese (Tootle et al., 2011). Nesta etapa, os genes vitelínicos são expressos e codificam as proteínas vitelínicas (Papantonis et al., 2015) que são consideradas estruturais e participam da formação da membrana vitelínica (Gonçalves et al., 2013), que no estágio 11 torna-se morfologicamente completa (Mauzy-Melitz e Waring, 2003).

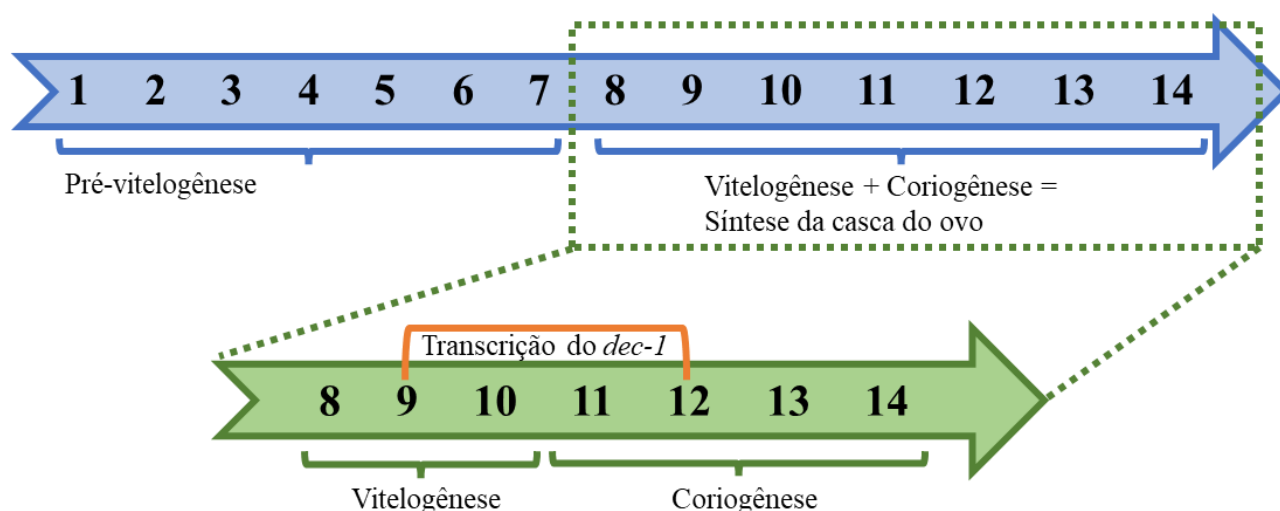


Figura 2: Esquema das etapas do processo de oogênese. Em verde está representada a etapa desse processo na qual ocorre a formação da casca do ovo, e em laranja, a etapa na qual o gene *dec-1* é transcrito, de acordo com o proposto para *Drosophila melanogaster*.

No final da vitelogênese, os folículos ovarianos aumentam em volume e têm sua taxa metabólica aumentada, caracterizando o início da coriogênese (Papantonis et al., 2015). Com a amplificação e expressão dos genes coriônicos estruturais no final do estágio 10 (Tootle et al., 2011) e início do estágio 11 (Margaritis, 1985), as proteínas coriônicas são codificadas e expressas de maneira sequencial e combinada, e o córion começa a ser produzido (Nadel e Kafatos, 1980; Eickbush e Izzo, 2009). Durante os estágios de 11 a 14, a camada coriônica interna, o endocóron tripartido e o exocóron são formados (Margaritis, 1985). No estágio 14, a formação de *crosslinks* de di e tri-tirosina faz com que as camadas externas do córion tornem-se insolúveis (Mindrinos et al., 1980), o que contribui para que a casca do ovo consiga

desempenhar sua função como elemento estrutural (Waring et al., 1990).

O gene *defective chorion 1* (*dec-1*) é transcrito durante os estágios de 9 a 12 do processo de oogênese (Hawley and Waring, 1988) e codifica ao menos 3 pró-proteínas foliculares em *Drosophila melanogaster*: FC106, FC125 e FC177 (Andersson e Lambertsson, 1991). Estas pró-proteínas são resultado de *splicing* alternativo do mRNA deste gene, que faz com que compartilhem as regiões central e N-terminal, mas sejam diferentes na região C-terminal (Waring et al., 1990). Dentre as regiões compartilhadas entre os três produtos, estão domínios repetitivos de 26 aminoácidos de comprimento na região central do gene, nos quais mutações podem afetar as sequências dos 3 produtos proteicos (Waring et al., 1990).

As pró-proteínas FC106, FC125 e FC177 são secretadas constitutivamente pelas células foliculares (Noguerón and Waring, 1995) em diferentes estágios da coriogênese de *D. melanogaster* (Waring et al., 1990), o que sugere funções distintas e essenciais para cada uma delas (Spangenberg and Waring, 2007). Após secretadas, localizam-se na membrana vitelínica proximal ao oócito, onde sofrem clivagem proteolítica pós-traducional extracelular regulada pelo desenvolvimento (Noguerón e Waring, 1995) e múltiplos derivados proteicos são gerados (Noguerón et al., 2000).

Os derivados proteicos de *dec-1* possuem papel chave na montagem e na estabilização da casca do ovo (Waring et al., 1990) e são liberados da membrana vitelínica em momentos e quantidades distintos (Noguerón et al., 2000). Esta membrana desempenha papel importante na migração dos derivados do *dec-1*, pois atua como local de armazenamento transitório destas moléculas (Waring et al., 1990) e controla sua concentração antes da montagem, o que previne a formação e interação prematuras e/ou inadequadas de proteínas (Noguerón et al., 2000). Após liberados, os derivados são realocados em regiões distintas da casca do ovo (Noguerón et al., 2000). Em conjunto, FC106, FC125 e FC177 são clivadas em pelo menos 5 fragmentos proteicos e alguns deles podem ser levados pelo oócito para o córion como um todo, ou para as cavidades endocoriônicas (Waring et al., 1990).

1.5 FUNÇÃO E IMPORTÂNCIA DO *DEC-1* E SUAS PRÓ-PROTEÍNAS

A pró-proteína FC106 é a mais abundante das 3 pró-proteínas previamente mencionadas (Mauzy-Melitz and Waring, 2003) e é sintetizada e secretada pelas células foliculares ovarianas nos estágios 9 e 10 da oogênese em *D. melanogaster* (Bauer and Waring, 1987). Durante o estágio 10, ela é clivada nos derivados N-terminal s25 e C-terminal

s80 (Bauer and Waring, 1987). Após gerado, o s25 é encontrado na membrana vitelínica, no endocóron e na camada coriônica interna (ICL), sendo o único derivado do *dec-1* a localizar-se dentro desta camada (Noguerón et al., 2000).

A distribuição do derivado protéico N-terminal s25 por todas as camadas do ovo de *D. melanogaster* é associada às interações proteína-proteína intercamadas (Noguerón et al., 2000) devido ao seu segmento central com alanina e rico em prolina, que também seria capaz de fornecer proteção contra adversidades ambientais (Badciong et al., 2001) e de formar uma superfície hidrofóbica sobre as outras proteínas da casca do ovo (Williamson, 1994). As prolinas dos envelopes de células cornificadas, por exemplo, promovem ligações proteína-proteína cruzadas, o que confere a mesma proteção físico-química (Badciong et al., 2001). Além disso, as características destes aminoácidos teriam o potencial de promover a reticulação de s25 a si mesmo ou com outras proteínas da casca do ovo, o que é reforçado pelo fato deste derivado encontrar-se reticulado no final da oogênese (Badciong et al., 2001).

Nos estágios 13 e 14 da oôgenese em *D. melanogaster*, o derivado protéico C-terminal s80 é clivado nos derivados N-terminal s20 e C-terminal s60 (Noguerón et al., 2000). O s20 sai rapidamente da membrana vitelínica, é absorvido pelo oócito e mantém-se ligado a vesículas de membrana (Noguerón et al., 2000), nas quais pode auxiliar e ordenar as interações protéicas que antecedem a formação do oócito, e a própria montagem do ovo (Badciong et al., 2001). Além disso, s20 possui estrutura básica e aminoácidos carregados, como K e R (lisina e arginina: carga positiva), e E e D (ácido Glutâmico e ácido Aspártico: carga negativa) (Spangenberg and Waring, 2007). O derivado s60 consiste majoritariamente de um *motif* de repetição de 26 aminoácidos com alto teor de glutamina (Spangenberg and Waring, 2007). Devido à sua clivagem ocorrer quando o endocóron está se organizando em uma estrutura tripartida e à sua migração para esta mesma camada assim que é gerado, sugere-se que seja uma molécula necessária para tal organização e para eventos de montagem posteriores (Noguerón et al., 2000).

Ademais, a produção do derivado protéico C-terminal s60 com alto teor de glutamina pode estar associada à ação da involucrina dos queratinócitos de eucariotos em geral (Waring et al., 1990), que quando fragmentada tem seus resíduos de glutamina expostos para atuar como receptores de aminas primárias (Eckert and Green, 1986). Estas ligações são catalisadas pela enzima transglutaminase em um processo de reticulação que garante a estabilidade dos envelopes destes queratinócitos (Eckert and Green, 1986). Desta maneira, a clivagem de s80 em s60 deixaria os resíduos de glutamina expostos, o que possibilitaria um processo de

reticulação capaz de conferir estabilidade ao endocórion (Waring et al., 1990).

A pró-proteína FC125 acumula-se na membrana vitelínica no início do estágio 10 da oôgenese de *D. melanogaster* e é clivada no derivado C-terminal s95 no final deste mesmo estágio, nas câmaras do ovo (Waring et al., 1990). Seu derivado N-terminal não é estável e não desempenha nenhuma função até então conhecida no processo de coriogênese (Noguerón et al., 2000), no entanto, estudos com mutantes nulos para s95 mostraram fêmeas estéreis, o que indica sua importância, potencialmente na formação da casca do ovo (Mauzy-Melitz e Waring, 2003).

Enquanto FC106 e FC125 (*id.*) são sintetizadas e clivadas durante a formação da membrana vitelínica, a FC177 é sintetizada nos estágios 11 e 12, e clivada no derivado C-terminal s85 somente nos estágios 12 e 13 da coriogênese, além de ser a maior das 3 pró-proteínas produzidas (Noguerón et al., 2000). A proteína derivada s85, assim que produzida, rapidamente fica localizada nas cavidades endocoriônicas, sendo o único derivado de *dec-1* a migrar para esta estrutura quimicamente desconhecida (Noguerón et al., 2000). É um produto proteico de natureza hidrofílica, com região central rica em glutamina e metionina, e com domínio C-terminal rico em cisteína, o que sugere que esta molécula esteja envolvida na estabilização da membrana vitelínica (Waring et al., 1990), através da formação de ligações dissulfeto intermoleculares, de interações homólogas ou com componentes subjacentes da matriz extracelular (Badciong et al., 2001).

Na ausência de FC177 e s85 (*ibid*), o endocórion apresentava morfologia semelhante ao normal, porém, a sua tripartição não é formada, o que sugere que FC177 e s85 são necessárias para a formação das cavidades endocoriônicas e prevenção da agregação anormal de proteínas dentro desta estrutura (Mauzy-Melitz e Waring, 2003). Diante disso, sua função também tem sido associada a proteínas de andaime virais (Mauzy-Melitz e Waring, 2003), ou seja, enquanto estas proteínas de andaime virais promovem a montagem de macromoléculas por interações proteína-proteína corretas (Dokland, 1999), o derivado s85 promoveria a agregação de proteínas do córion para formar os pilares, piso e teto reticulado do endocórion ao se vincular ao espaço extracelular (Mauzy-Melitz e Waring, 2003).

As pró-proteínas do *dec-1* são sintetizadas durante a formação da membrana vitelínica, porém, Bauer e Waring (1987) observaram que os problemas morfológicos em fêmeas mutantes nulas para o derivado s85 só ficam aparentes no final da coriogênese, o que sugere que *dec-1* pode desempenhar um papel importante na formação da casca do ovo. Ademais, estudos baseados em imunolocalização e hibridização de RNA associaram as alterações

morfológicas do córion com a perda de proteínas vitelínicas (Pascucci et al., 1996) e de proteínas coriônicas estruturais (Digan et al., 1979).

Estudos com mutantes nulos observaram que a camada coriônica interna (ICL) torna-se defeituosa na ausência dos derivados proteicos do *dec-1* (Margaritis et al., 1991) e que o endocórion não organiza-se em uma estrutura tripartida e seu material colapsa dentro da membrana vitelínica (*ibid*) (Hawley e Waring, 1988). Enquanto a pró-proteína FC106 por si só não é suficiente para a formação do endocórion tripartido e a FC125 não é necessária para tal processo, a FC177 é responsável pela produção do piso, dos pilares e do teto reticulado (Mauzy-Melitz and Waring, 2003). Entretanto, FC106 e FC125 são suficientes para evitar que o endocórion colapse, e o derivado s25 (FC106) é essencial para mantê-lo estável (Mauzy-Melitz and Waring, 2003).

De modo geral, mais de uma das pró-proteínas e derivados proteicos de pelo menos duas delas parecem ser necessários para formação de um ovo funcional e para a fertilidade da fêmea de *D. melanogaster* (Noguerón et al., 2000; Mauzy-Melitz and Waring, 2003), o que sugere que estes produtos proteicos interagem com outras proteínas envolvidas na coriogênese e governam os eventos de montagem da casca do ovo (Waring et al., 1990).

2. OBJETIVOS

O objetivo geral deste trabalho foi investigar o papel do gene *defective chorion 1* no processo de diferenciação das espécies do grupo *fraterculus*, em particular *A. fraterculus* e *A. obliqua*. Portanto, os objetivos específicos foram:

- Investigar a estrutura molecular do gene coriônico *dec-1* para espécies de *Anastrepha* a partir de transcriptomas de *Anastrepha fraterculus* e *A. obliqua* gerados em nosso laboratório e da sequência gênica derivada do genoma de *A. ludens*;
- Caracterizar as isoformas FC106, FC125 e FC177 usando como modelo sequências do *dec-1* de *D. melanogaster* disponível no FlyBase e de ortólogos disponíveis no NCBI;
- Realizar análises de seleção no intuito de identificar forças evolutivas que estão influenciando a diversidade deste gene coriônico nas espécies *A. fraterculus* e *A. obliqua*;
- Caracterizar o desenvolvimento ovariano em indivíduos da espécie *A. fraterculus*, a fim de compreender morfolologicamente o momento em que o *dec-1* se expressa em *Anastrepha* através de comparações com o que se conhece para o organismo modelo *D. melanogaster*.

3. MATERIAIS E MÉTODOS

3.1 IDENTIFICAÇÃO DO GENE *DEFECTIVE CHORION 1 (DEC-1)* EM *ANASTREPHA*

Para identificar os transcritos do *dec-1*, sequências de seu ortólogo em *D. melanogaster* – *Dmdec-1* (número de acesso no Flybase: FBgn0000427) foi usada para buscas por *contigs* em transcriptomas de *A. fraterculus*, *A. obliqua* e *A. grandis* disponíveis em nosso laboratório (Gonçalves et al., 2013; Rezende et al., 2016; Congrains et al., 2018; Congrains et al., 2021). Os potenciais variantes de *splicing* alternativo do *dec-1* nestas moscas foram identificados por predição da sequência deste gene no genoma de *Anastrepha ludens*, generosamente compartilhado conosco pelo laboratório do Dr. Scott Geib, apesar de agora estar já disponível no GenBank. Tal versão do genoma tinha cerca de 683 Mb, e teve repetições e regiões de baixa complexidade mascaradas (Dupuis et al., 2019).

Para realizar a identificação de sequências similares à *Dmdec-1* e seu isolamento, o *software* Funannotate v. 1.8.9 (Palmer e Stajich, 2019) foi utilizado para a previsão e anotação do genoma de *A. ludens*: a previsão *ab initio* da sequência nucleotídica no *pipeline* foi realizada pelo uso dos programas preditores de genes Augustus v. 3.3.3 (Keller et al., 2011), snap v. 2006-07-28 (Korf, 2004) e glimmerHMM v. 3.0.4 (Majoros et al., 2004), com a consulta de vários bancos de dados com curadoria para função de proteína e identificação de domínio.

O programa Augustus v. 3.3.3 prevê genes em sequências genômicas eucarióticas por meio de informações de homologia entre proteínas (Keller et al., 2011). As sequências proteicas de *Rhagoletis zephyria* também foram fornecidas como evidência para o treinamento do Augustus, juntamente com modelos BUSCO (Manni et al., 2021) para Diptera (Odb10), facilitando a identificação de regiões codificadoras de proteínas. O método BUSCO estima a integridade e redundância de dados genômicos e transcriptômicos por meio de métricas biologicamente significativas baseadas no conteúdo genético esperado, mesmo que a origem taxonômica das espécies seja desconhecida (Manni et al., 2021).

O programa snap v. 2006-07-28 identifica SNPs (Polimorfismos de Nucleotídeo Único) próximos em desequilíbrio de ligação e facilita a interpretação e comparação entre regiões genômicas e suas variantes (Korf, 2004). O *software* glimmerHMM v. 3.0.4, por sua vez, é um localizador de genes baseado no modelo generalizado de cadeias de Markov ocultas (GHMM), que inclui a busca de regiões inter e intra-gênicas em sequências de tamanho fixo (Majoros et al., 2004). No presente caso, foram utilizados os bancos de dados: Uniprot

(Uniprot Consortium, 2019) com um valor de corte de $1e10^{-3}$, e o Gene Ontology (Ashburner et al., 2000). Em seguida, os modelos de genes encontrados foram filtrados por comprimento de sequência, intervalos de abrangência e elementos transponíveis, e os produtos gênicos foram identificados (Palmer e Stajich, 2019). A partir destes produtos gênicos encontrados, a sequência genômica de *dec-1* foi identificada em *Anastrepha ludens*.

3.2 CARACTERIZAÇÃO E ESTRUTURA MOLECULAR DO *DEC-1*

Uma vez tendo identificadas as sequências genômicas do *dec-1* de *A. ludens*, tais sequências foram utilizadas tanto para buscar no GenBank, por meio da ferramenta BLAST, potenciais ortólogos de outras espécies de Tephritidae e de Cyclorrhapha, quanto para a garimpagem nos transcriptomas disponíveis em nosso laboratório para a identificação em outras espécies de *Anastrepha*. As sequências de nucleotídeos codificantes garimpadas de *A. ludens*, *A. fraterculus* e *A. obliqua* foram convertidas em suas supostas sequências correspondentes de aminoácidos usando a ferramenta Expasy Translate (<https://web.expasy.org/translate/>).

Para a discriminação das isoformas, a base de dados do NCBI (<http://www.ncbi.nlm.nih.gov/sites/entrez>) foi utilizada para identificar o ortólogo de maior semelhança em outros Tephritidae (ou seja, *R. zephyria*, NCBI gene ID: 108362233), a fim de utilizá-lo como modelo para localização dos limites éxon-íntron de *dec-1* em *A. fraterculus* e *A. obliqua*. Todas as sequências de nucleotídeos e de aminoácidos foram alinhadas usando Clustal-W (Larkin et al., 2007) implementado no programa Bioedit, e as arquiteturas íntron-éxon foram definidas por comparações. Além disso, buscas por domínios em comum com ortólogos foram realizadas no site Motif Search (<https://www.genome.jp/tools/motif/>), e as porcentagens de semelhanças entre eles foram analisadas com a ferramenta SMS (https://www.bioinformatics.org/sms2/ident_sim.html), a fim de verificar a existência de regiões conservadas e associá-las com possíveis funções biológicas do *dec-1* entre os Cyclorrhapha (uma subordem de Diptera pertencente ao clado dos Muscomorpha), bem como as consequências da ação da seleção sobre elas.

3.3 PROCEDIMENTOS MOLECULARES E TESTES DE SELEÇÃO

3.3.1 Populações Experimentais

Coletamos frutos de oito pontos de amostragem distribuídos nas regiões Sul, Sudeste e Nordeste do Brasil, que foram mantidos em gaiolas de acrílico (21 x 13 x 13 cm), recobertos com vermiculita, até a pupação, em laboratório sob temperatura controlada (26 ± 5 °C), 65% de umidade relativa e condições de iluminação natural. As pupas foram separadas em uma nova gaiola e os adultos que emergiam dispunham de água e uma fonte de alimento à base de levedura, glicose e proteína hidrolisada *ad libitum*. Nestas condições, esses adultos foram mantidos por 10 a 15 dias, ou seja, até atingirem a maturidade sexual. Os adultos maduros foram identificados de acordo com Zucchi (2000), 30 casais foram separados para estabelecer colônias de moscas que foram mantidas em mangas e usadas como repositórios para oviposição. Para os procedimentos moleculares e testes de seleção, 27 fêmeas (Tabela 1) foram sacrificadas por congelamento em gelo e armazenadas a -80 °C para posterior extração de DNA.

Tabela 1: Amostras de *Anastrepha obliqua* e *Anastrepha fraterculus* coletadas em diferentes localidades e utilizadas na análise populacional.

<i>A. obliqua</i>		<i>A. fraterculus</i>	
Localidade/Estado	Amostra	Localidade/Estado	Amostra
Capanema/PR	AOP1	São Thomé das Letras/MG	AFT1a
	AOP4		AFT2a
Marialva/PR	AOP4a		AFT3b
	AOM1b		AFT4a
	AOM2b		AFT5b
Caetité/BA	AOM3	Pinheiros/ES	AFPI1a
	AOCT1a		AFPI3a
	AOCT2a		AFPI4a
	AOCT3b		AFPI5a
	AOCT4c		
Passos/MG	AOCT5a	Espírito Santo do Turvo/SP	AFST1a
	AOPS1a		AFST2
	AOPS2c	Criciúma/SC	AFCR3a
	AOPS4a		
	AOPS5c		

AO=*Anastrepha obliqua* (P= Capanema; M= Marialva; CT= Caetité; PS= Passos). AF= *Anastrepha fraterculus* (T= São Thomé das Letras; PI= Pinheiros; ST= Espírito Santo do Turvo; CR= Criciúma). Os números representam a identificação de cada fêmea coletada. As letras a, b, e c indicam os clones utilizados para cada indivíduo.

3.3.2 Extração de DNA e amplificação por PCR

O material previamente congelado foi macerado em tampão de lise e digerido em proteinase K para a extração do DNA total de cada indivíduo adulto inteiro, segundo o protocolo de tiocianato de guanidina (Nelson e Krawetz, 1992) e de precipitação do DNA com álcool isobutílico. Após a precipitação, as amostras de DNA obtidas foram ressuspensas e submetidas à corrida eletroforética em gel de agarose 1% e visualizadas com brometo de etídeo em transiluminador com luz ultravioleta, a fim de confirmar a qualidade do DNA extraído.

Como o gene *dec-1* possui uma organização complexa, com *splicing* alternativo de diferentes éxons, uma região específica do Éxon-3 que é comum às diferentes isoformas, alinhada para diferentes espécies de *Anastrepha*, foi selecionada. Para isso, *primers* foram desenhados para amplificar essa região nas 27 moscas. Utilizamos o software Primer3Plus (<https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) para auxiliar na confecção dos primers, considerando parâmetros padrão do programa, e o site da IDT (<https://www.idtdna.com/pages/tools/oligoanalyzer>) para análise das condições (%GC, T_m, etc), que levaram à produção dos primers *dc1_2371F*: 5' - ATG TCT ATG GAT AGG CAA ATG - 3' e *dc1_3236R*: 5' - GGA GYG GCC AAG CCT AGC ATA TC - 3'. As amplificações foram feitas de acordo com o padrão da literatura (Saiki et al., 1988) em um Termociclador BioRad T100TM programado para 35 ciclos de 94°C por 1', hibridação por 90" a 54,2°C e 72°C por 1', com 3' no início a 94°C para desnaturação e 3' de extensão final. Os produtos de amplificação foram verificados em gel de agarose a 1% contendo brometo de etídeo sob iluminação UV, e purificados utilizando o protocolo de precipitação diferencial com PEG (Polietileno Glicol) 8000 e NaCl (Lis and Schleif, 1975).

3.3.3 Clonagem e Sequenciamento

A fim de evitar problemas de sequenciamento causados pela presença de possíveis *indels* e pelos próprios polimorfismos, os produtos de PCR obtidos foram clonados com o kit InsTAclone (Fermentas): as amostras purificadas foram incorporadas na concentração sugerida pelo kit na reação de ligação com o vetor pTZ57R/T, que foram mantidas em temperatura de 4 °C até o dia seguinte. O produto dessa ligação foi utilizado para transformar bactérias competentes *E. coli* DH5α por choque térmico a 42 °C por dois minutos e plaqueadas em meio de cultura LB sólido com ampicilina para crescimento *overnight* em estufa a 37 °C. As colônias transformadas obtidas foram isoladas com o auxílio de palitos

autoclavados, diluídas em água ultrapura e fervidas a 98 °C para lise celular.

Em seguida, clones positivos foram amplificados com *primers* M13 (M13F: 5'- GTA AAA CGA CGG CCA GT - 3'/M13R: 5'- GGA AAC AGC TAT GAC CAT G-3') por 35 ciclos, como segue: (a) 94°C para 1 minuto na etapa de desnaturação; 34 ciclos com (b) adição da etapa de 94°C por 30", hibridação por 30" a 56°C e 72°C por 3', com 3' no início a 94°C para desnaturação e 3' de extensão final. Os produtos foram novamente purificados com PEG 8000 e as amostras obtidas foram preparadas em placas de sequenciamento com o *primer* M13F a 5 µM e concentração de 50 nmol µL⁻¹ para posterior sequenciamento por Sanger (Sanger et al., 1977), na MacroGen Inc (Seoul, Coréia).

3.3.4 Alinhamentos

As sequências e os eletroferogramas da região amplificada obtidas do sequenciamento feito em uma *facility* foram visualmente inspecionados em conjunto com as sequências do transcriptoma de *A. fraterculus* e *A. obliqua*. As que apresentaram boa qualidade no sequenciamento foram alinhadas utilizando o software MAFFT que possui algoritmos mais precisos para o alinhamento de sequências codificantes, no qual o método L-INS-i foi implementado, pois permite avaliar a consistência entre um alinhamento múltiplo e alinhamentos emparelhados (Katoh et al., 2019).

Para testes de evolução molecular, foram consideradas apenas as sequências da isoforma mais curta do *dec-1* (FC106), por ser a mais predominante entre os ortólogos identificados no GenBank. As sequências foram divididas em três grupos de espécies que representam diferentes padrões de seleção em tempos evolutivos distintos: “Anastrepha”, “Tephritidae” e “Cyclorrhapha”, com 31, 10 e 13 espécimes, respectivamente, que foram independentemente alinhadas usando o software MAFFT (Katoh et al., 2019). As sequências utilizadas estão disponíveis no Apêndice A (Tabela 7) juntamente com seus códigos de acesso do NCBI. O primeiro conjunto de dados compreende os fragmentos amplificados de *A. obliqua* e *A. fraterculus*, a região correspondente para seus transcriptomas, a sequência genômica de *A. ludens* e, como grupo externo, a sequência do transcriptoma de *A. grandis*, que foi identificada após filtragem do transcriptoma de *A. grandis* disponível (Gonçalves et al., 2013; Rezende et al., 2016; Congrains et al., 2018; Congrains et al., 2021), e pela consulta da sequência genômica de *A. ludens*.

O segundo conjunto de dados, “Tephritidae”, incluiu sequências inteiras da isoforma FC106 do *dec-1* dos transcriptomas de *A. obliqua*, *A. fraterculus* e *A. grandis*, do genoma de

A. ludens e dos genomas de *Ceratitis capitata*, *Zeugodacus cucurbitae*, *Bactrocera tryoni*, *Rhagoletis zephyria*, *R. pomonella* e *Drosophila pseudoobscura* (esta utilizada como grupo externo). O terceiro banco de dados (“Cyclorrhapha”) contém todas as sequências do segundo grupo, com a adição de *D. pseudoobscura*, *D. melanogaster*, de *D. albomicans* e de *Teleopsis dalmanni* (família Diopsidae), que foi utilizada como grupo externo.

3.3.5 Testes filogenéticos e de seleção

Com os dados do alinhamento de supostas cópias ortólogas de *dec-1* nos diferentes conjuntos de espécies estudados, os níveis de saturação das sequências estudadas e as taxas de transição e transversão foram estimados no programa DAMBE 7.3.11 (Xia, 2018), considerando o modelo evolutivo GTR. Tais análises foram feitas individualmente, para cada um dos três conjuntos de amostras mencionados no tópico anterior, que representam diferentes níveis de divergência taxonômica, para os quais diferentes modelos de evolução de códons foram testados para detectar variação em ω (ômega), ou seja, pela razão entre as taxas de substituições não-sinônimas (dN) e sinônimas (dS).

A análise filogenética, novamente considerando individualmente cada um dos três conjuntos de dados de *dec-1*, utilizou o modelo evolutivo de substituição de nucleotídeos mais provável para cada conjunto estimado pelo software ModelTest-ng (Darriba et al., 2020), seguindo o Critério de Informação Bayesiana (BIC) (Schwarz, 1978) para inferir relações de parentesco segundo o princípio da Máxima Verossimilhança (Guindon et al., 2010) no software PhyML v.3.0. As árvores filogenéticas geradas foram visualizadas e editadas com o FigTree v.1.4.0 (<http://tree.bio.ed.ac.uk/software/figtree/>)

Em seguida, testes de seleção foram feitos tanto pelo método CODEML implementado no software PAML (Yang, 1997; Chen, 2009) quanto pelo pacote HyPhy implementado no DataMonkey (Kumar et al., 2018). Para ambos os modelos, duas categorias de análise foram empregadas: uma global, ou seja, que considera taxas médias de evolução no gene como um todo em todos os táxons dentro da filogenia, e uma que considera apenas diferenças em taxas em ramos específicos. O método CODEML detecta seleção positiva entre linhagens previamente selecionadas considerando a história evolutiva por meio de um teste de Máxima Verossimilhança (LRT) que compara o modelo nulo ($M0$ nulo) com modelos que permitem valores de $\omega > 1$. Para o teste de seleção de ramos, verifica-se se os dados estão sob influência de seleção positiva pela comparação entre as probabilidades calculadas para o modelo nulo e para o modelo alternativo (Zhang et al., 2005; Yang e Dos Reis, 2011). O

método *Bayes Empirical Bayes* (BEB) também foi implementado no PAML v.4 e utilizado para estimar a probabilidade (posterior) de que um determinado códon estaria evoluindo sob seleção positiva ($P > 0,95$), e para inferir se os dados se encaixam em um modelo neutro de evolução ou em modelos de seleção alternativos, utilizando-se de testes de significância LRT.

O pacote HyPhy, por sua vez, identifica assinaturas de seleção positiva ou de seleção purificadora também pela razão dN/dS , mas permitindo que mais de um ramo apresente valores de ω diferenciados. Para o teste de taxa global, assume-se uma única pressão de seleção para todos os ramos pela identificação de evidências genéticas de seleção positiva para ramos específicos (Murrell et al., 2015). Para testar se os sítios estão evoluindo neutramente, ou estão selecionados positivamente ($dN/dS > 1$) ou negativamente ($dN/dS < 1$), os métodos FEL (Kosakovsky e Froisy, 2005), MEME (Murrell et al., 2012), BUSTED (Murrell et al., 2015), RELAX (Wertheim et al., 2015) e FUBAR (Murrell et al., 2013), foram utilizados. Estes métodos estão implementados no software Datamonkey (Kumar et al., 2018).

O método FEL (*Fixed Effects Likelihood*) utiliza máxima verossimilhança (ML) para inferir taxas de substituição não sinônimas (dN) e sinônimas (dS) para uma região gênica como um todo, e assume que a pressão de seleção para cada sítio é constante ao longo de toda a filogenia (Kosakovsky e Froisy, 2005). O método MEME (*Mixed Effects Model of Evolution*) usa análise de máxima verossimilhança para testar se ramos individuais sofreram ação de seleção episódica positiva ou diversificada em determinadas ramificações (Murrell et al., 2012). Já o BUSTED (*Branch-site Unrestricted Statistical Test for Episodic Diversification*) testa se um gene sofreu seleção positiva em pelo menos um *locus* (Murrell et al., 2015). O método RELAX, por sua vez, consiste em um teste de hipóteses que infere se a força da seleção natural foi relaxada ou intensificada em ramos específicos, e permite identificar tendências ou mudanças na pressão da seleção natural em um determinado gene (Wertheim et al., 2015). E por fim, o método FUBAR (*Fast Unconstrained Bayesian AppRoximation*) infere taxas de substituição não-sinônimas (dN) e sinônimas (dS) por *locus* em um alinhamento de regiões codificantes e em sua filogenia correspondente por abordagem Bayesiana, que também considera a pressão de seleção constante ao longo da filogenia. Foram utilizadas árvores não enraizadas para análises globais, mas as mesmas foram enraizadas para a análise de seleção de ramos, considerando os grupos externos.

3.3.6 Medição da magnitude físico-química das substituições de aminoácidos

Para inferir o impacto de eventos de substituição em supostas mudanças de aminoácidos no gene *dec-1*, as árvores geradas pelo *software* PAML foram utilizadas no *software* TreeSAAP (*Selection on Amino Acid Properties using phylogenetic trees*) (Woolley et al., 2003). O padrão de substituição de aminoácidos inferido é analisado pelos métodos de Xia e Li (1998) e McClellan e McCracken (2001), que estimam as distribuições de potenciais mudanças nas propriedades físico-químicas dos aminoácidos e assumem que todas elas são igualmente prováveis em condições neutras.

As mudanças médias esperadas e observadas nas propriedades dos aminoácidos no gene estudado são resumidas pela diferença percentual média (Xia e Li, 1998), e analisadas pela qualidade do ajuste e pela comparação estatística entre as proporções de substituições de aminoácidos observadas e as vias evolutivas esperadas para cada categoria de substituição (McClellan e McCracken, 2001). Foram considerados positivos e significativos os z-scores nas categorias de seis a oito, que representam as alterações estruturais ou funcionais mais radicais, ou seja, as que apresentam maior diferença em suas propriedades físico-químicas.

3.4 CARACTERIZAÇÃO OVARIANA

3.4.1 Amostras utilizadas

Para a caracterização morfológica dos ovários foram utilizadas fêmeas da espécie *A. fraterculus* de uma colônia de reprodução estabelecidas em laboratório a partir de uma coleta de goiaba (*Psidium guajava*) realizada em Criciúma/SC. Adultos maduros tiveram sua identificação confirmada de acordo com Malavasi e Zucchi (2000), dos quais 30 casais foram separados para estabelecer estoques de moscas em mangas que foram usadas como hospedeiras para a oviposição e desenvolvimento de larvas. Após eclosão de larvas e seu empupamento, gaiolas com pupas foram mantidas em BOD sob condição controlada: iluminação em ciclo claro-escuro de 12 horas, temperatura controlada ($26^{\circ}\text{C} \pm 1^{\circ}\text{C}$), 65% de umidade relativa, ração à base de levedura, glicose e proteína hidrolisada, e água *ad libitum*.

À medida que os adultos emergiram, as fêmeas foram separadas, sacrificadas por congelamento em idades distintas (1, 5, 7, 10, 14, 15, 20, 25 e 29 dias) e armazenadas a -80°C para os estudos morfológicos posteriores. As idades para dissecação foram escolhidas segundo o padrão estudado em *Bactrocera cacuminata* (Raghu et al., 2003), e em *Anastrepha suspensa* (Kendra et al., 2016).

3.4.2 Estudos morfológicos

Para identificar a idade fisiológica ovariana e consequentemente, presumir o momento em que o gene *dec-1* se expressa em espécies do gênero *Anastrepha*, as moscas armazenadas foram descongeladas em gelo e dissecadas em lupa (Leica, modelo MSV266). No total, 135 fêmeas (15 indivíduos de cada idade) foram dissecadas com o auxílio de pinças de ponta fina e submetidas a medições dos ovários e do corpo em geral.

A fim de avaliar a maturidade sexual, a carga de ovos maduros (*CO*) em cada ovário foi contada, e o índice ovariano (*IO*) foi obtido por meio da multiplicação das medidas de comprimento pelas medidas da largura dos ovários (Landolt e Davis- Hernandez 1993; Kendra et al. 2005). O comprimento dorsal do tórax (borda anterior do mesonoto até a extremidade posterior do mesoscutelo) (Sivinski, 1993) e o comprimento da asa anterior (base da veia costal ao ápice da asa) foram mensurados como indicadores independentes do tamanho geral feminino, a fim de verificar se há ou não alguma relação entre a idade e os tamanhos do corpo e dos ovários, ou seja, se os ovários sofrem influência de caracteres que não estão envolvidos na reprodução.

3.4.3 Análise estatística

Para verificar a existência de simetria bilateral entre os ovários esquerdo e direito, foi avaliado se havia diferenças nas médias populacionais do Índice Ovariano (*IO*) e Carga de Ovos (*CO*) entre os dois ovários. Para isso, foi utilizado o *software* Minitab v.19 (Minitab LLC, 2021) primeiramente para realizar os testes de normalidade de Ryan-Joiner (Ryan e Joiner, 1976) para o *IO* e *CO* respectivo a cada idade avaliada e para cada posição do ovário. Como todos os dados apresentaram distribuição normal, foi utilizado o teste F para verificar, em um teste de hipóteses, quais variâncias dos grupos eram iguais ou diferentes com base no p-valor para 95% de confiança. A mesma análise foi realizada para a *CO* e para as demais idades. A informação de quais variâncias eram iguais ou diferentes foi aplicada no teste *t* de Student de igualdade de médias (Student, 1908), a partir do qual foi possível verificar com base no p-valor para 95% de confiança se as médias populacionais das medidas coletadas para os ovários esquerdo e direito de cada idade diferem estatisticamente. Após a identificação da presença de simetria bilateral, os dados foram reorganizados e as medidas dos diferentes lados passaram a ser analisadas em conjunto.

Para inferir os possíveis momentos fisiológicos dos ovários, foram realizados testes de

normalidade, teste F das variâncias e teste *t* de Student de igualdade de médias para o Índice Ovariano e a Carga de Ovos entre as diferentes idades, a fim de verificar se a variância e a média populacional do *IO* de moscas com 1 dia se diferenciavam da média populacional dos outros dias. As médias apontadas como sendo estatisticamente diferentes foram submetidas a um teste *t* de Student adicional de comparação de médias que identifica quais delas são estatisticamente distintas das outras. Para uma melhor visualização, foram elaborados gráficos boxplot para as variáveis estudadas, com a identificação do momento fisiológico ovariano.

As correlações entre Idade, Índice ovariano, Carga de ovos e comprimentos de asa e tórax foram determinadas com comparações de pares a partir do índice de correlação de Pearson, para indicar a força e a direção da relação entre as variáveis (Freedman et al., 2007).

4. RESULTADOS E DISCUSSÃO

4.1 CARACTERIZAÇÃO DO GENE *DEFECTIVE CHORION 1*

A identificação da sequência genômica do *dec-1* de *A. ludens* facilitou a busca de potenciais ortólogos de outras espécies de Tephritidae e de Diptera e a garimpagem nos transcriptomas disponíveis utilizando a ferramenta BLAST. A partir dessas análises, três isoformas de *dec-1* foram identificadas nos transcriptomas de *A. fraterculus* e *A. obliqua* dentre as seis isoformas conhecidas e presentes em *D. melanogaster*. Elas correspondem às isoformas A, B e C de *D. melanogaster* (Flybase: FBgn0000427), aqui denominadas X1, X2 e X3 respectivamente, segundo as designações do NCBI para este gene em outros Tephritidae: *R. zephyria* (Gene ID: 108362233), *R. pomonella* (Gene ID: 118735351), e *C. capitata* (Gene ID: 101457983).

As sequências de *dec-1* de *D. melanogaster*, *R. zephyria* e *A. ludens*, foram utilizadas como modelo de limite éxon-íntron para inferir a possível posição equivalente desses éxons nas sequências do transcriptoma de *A. obliqua* e *A. fraterculus* (Congrains et al., 2018), a partir de seu alinhamento. A comparação dessas sequências e os potenciais pontos de corte de *indel* inferidos no alinhamento permitiram observar a provável presença de 6 éxons: na região 5'-*upstream*, os quatro primeiros éxons compartilham o mesmo arranjo para as três isoformas encontradas em Tephritidae, enquanto na região 3'-*downstream* os dois últimos éxons estão dispostos de forma diferente. A sequência ATG que codifica o códon de iniciação canônico está localizada na mesma posição (éxon 1) para as três isoformas. A Figura 3 mostra como os éxons estariam posicionados em *A. fraterculus*, padrão equivalente ao encontrado em *A. obliqua* (não mostrado). A Tabela 2 contém o tamanho dos éxons, dos CDSs e das supostas proteínas codificadas em cada uma das isoformas. Observa-se que o éxon 5 apresenta três formas diferentes derivadas de *splicing* alternativo: uma mais curta que é reflexo de um *splicing* diferencial que faz o éxon 5 começar algumas bases antes e ficar menor, e outras duas em que o *splicing* começa da mesma posição, mas é encurtada em uma isoforma a qual se conecta com outro éxon, ou é estendida. Tal padrão é similar às isoformas encontradas em *D. melanogaster* e em outros Tephritidae. Os *splicing* alternativos são um tipo de processamento do RNA a partir do qual diferentes regiões de um mesmo pré-RNA podem ser utilizadas para produzir moléculas distintas de RNA maduro que codificam proteínas diferentes, e assim, aumentam a diversidade do proteoma das espécies (Lareau et al., 2004; Sharp, 2009).

Em *A. fraterculus*, as duas isoformas mais longas atingem comprimentos de proteína

de 3576 (isoforma X2) e 4416 (isoforma X1) aminoácidos, e em *A. obliqua*, 3543 (X2) e 4578 (X1) aminoácidos (Tabela 2). Nas sequências dos ortólogos utilizados nas análises, os comprimentos de proteína chegam a 1667 (X1) e 1307 (X2) para *R. zephyria* e *R. pomonella*, 1345 (X1) e 1179 (X2) para *Z. cucurbitae*, 1368 (X1) e 1198 (X2) para *C. capitata*, 1367 (X1) e 1125 (X2) para *B. tryoni*, 1968 (X1) e 1347 (X2) para *T. dalmanni*, 1590 (X1) e 1208 (X2) para *D. melanogaster*, 1207 (X1) e 949 (X2) para *D. albomicans*, e 1207 (X1) e 1077 (X2) para *D. pseudoobscura*. As diferenças encontradas no tamanho do gene *dec-1* em diferentes espécies são resultado da variabilidade decorrente de eventos de mutação aleatórios, ou mesmo pelo viés de inserção encontrado na maioria das espécies de eucariotos, no qual a eficiência da seleção contra o excesso de material genético pode ser relaxada (Lynch 2007), já que os pontos de *splicing* não parecem ter sido alterados entre as espécies.

A garimpagem bem-sucedida de diferentes transcriptomas, os bons resultados provenientes do sequenciamento Sanger e as facilidades atuais na disponibilidade dos bancos de dados para a coleta de sequências de supostos ortólogos enriqueceram as comparações abordadas nas análises e contribuíram com a descoberta das três isoformas estudadas, bem como para a caracterização do *dec-1* em *Anastrepha*. Entretanto, os trabalhos pioneiros que estudaram esse gene foram essenciais neste processo, principalmente na investigação dos produtos proteicos (Waring et al., 1990), dos processos regulatórios (Noguerón e Waring, 1995; Noguerón et al., 2000), e dos mecanismos de expressão do *dec-1* (Nilson e Schupbach, 1998; Camporeale et al., 2007; Kim et al., 2011; Spangenberg e Waring, 2007).

A busca por domínios no banco de dados Pfam detectou dois domínios básicos em *A. fraterculus* e *A. obliqua* que também estão presentes nos ortólogos e são anteriormente conhecidos em *D. melanogaster*: o da região N-terminal (Pfam ID: DEC-1_N; Descrição: PF04625) e o da região C-terminal (Pfam ID: DEC-1_C; Descrição: PF04626). Outros domínios que incluem 9 regiões de repetição (Pfam ID: DEC-1, Descrição: PF04624) não foram detectados em *A. fraterculus* e *A. obliqua*. Por outro lado, um domínio adicional denominado domínio da subunidade RNase P Rpr2/Rpp21/SNM1 (Pfam ID: Rpr2, Descrição: PF04032) foi detectado e significativo em uma das isoformas mais longas (X1) de *A. obliqua*. Ele contém um domínio similar a uma subunidade da RNase MRP (Rpp21), cuja família é responsável tanto pela replicação do DNA mitocondrial, quanto pelo processamento de rRNA 5.8S nuclear (Jarrous et al., 2001), e outra subunidade que consiste em uma proteína de ligação ao RNA (SNM1) com especificidade para o RNA do MRP (Cai et al., 1999).

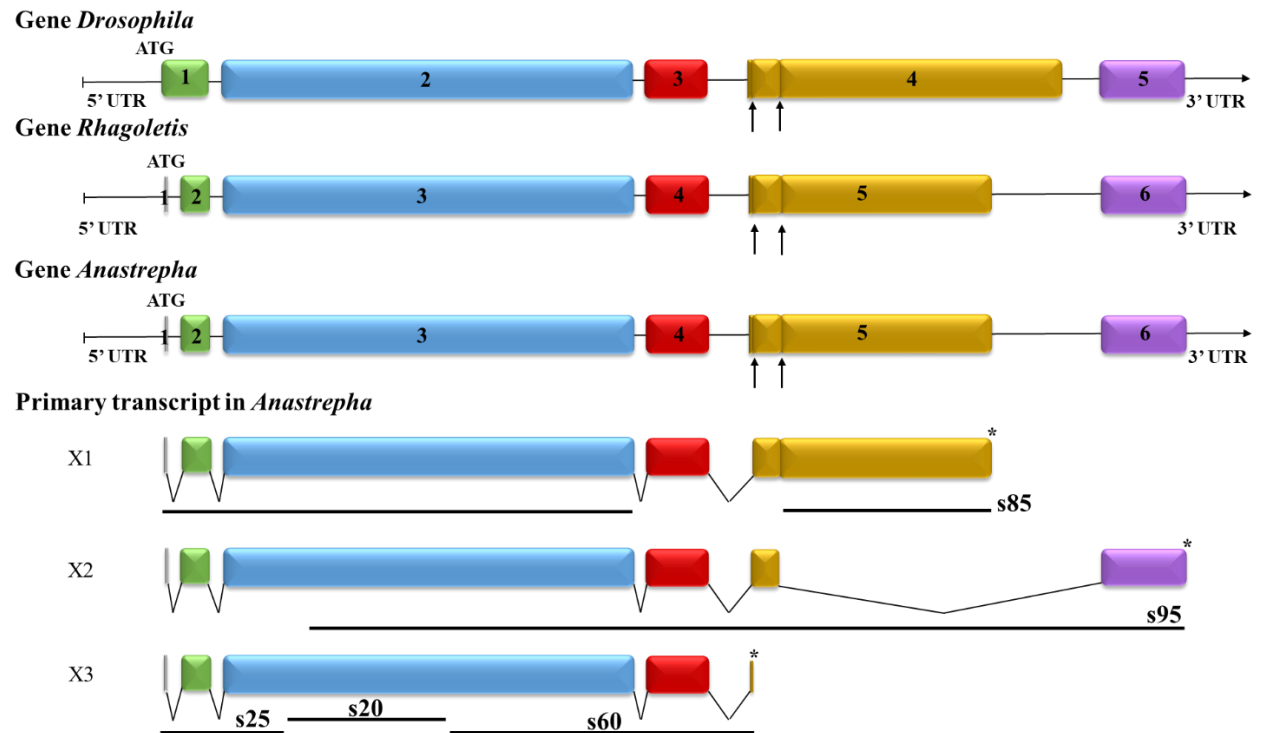


Figura 3: Estrutura do gene *dec-1* em *Drosophila melanogaster* (NCBI: NM_167133.2), *Rhagoletis zephyria* (NCBI: XM_017615135.1) e *Anastrepha* sp. (esse trabalho), bem como caracterização de três isoformas na mosca-das-frutas *Anastrepha fraterculus*. Os seis éxons detectados são representados proporcionalmente por seu tamanho em pb por cores diferentes. O éxon 5 (em amarelo) mostra os sítios de *splicing* internos (setas). As posições dos íntrons (linhas pretas) são previsões baseadas no genoma de *Anastrepha ludens*. A localização inferida para os derivados proteicos s85, s95, s25, s20 e s60 previstos para *D. melanogaster* estão representados embaixo de cada isoforma encontrada para *Anastrepha*. UTR, região não traduzida: 5' e 3'; códon de iniciação (ATG); códon de terminação (*).

Tabela 2: Arranjo dos éxons e CDS (*Coding DNA Sequence*), e respectivas proteínas (sequência de aminoácidos) das três isoformas em ordem crescente de tamanho (X3, X2 e X1) detectados no *dec-1* das moscas-das-frutas *Anastrepha obliqua* e *Anastrepha fraterculus*. As letras a, b e c no Éxon 5 representam os pontos de *splicing*.

		Isoformas						
		<i>A. obliqua</i>			<i>A. fraterculus</i>			
		X3	X2	X1		X3	X2	X1
Éxon 1	9	✓	✓	✓	9	✓	✓	✓
Éxon 2	189	✓	✓	✓	189	✓	✓	✓
Éxon 3	2286	✓	✓	✓	2340	✓	✓	✓
Éxon 4	348	✓	✓	✓	309	✓	✓	✓
Éxon 5(a)	18	✓			18	✓		
Éxon 5(b)	194		✓	✓	197		✓	✓
Éxon 5(c)	1552			✓	1372			✓
Éxon 6	517		✓		532		✓	
Tamanho do CDS	-	2850	3543	4578	-	2865	3576	4416
Tamanho da proteína	-	950	1181	1526	-	955	1192	1472

O domínio da região de repetição (Pfam ID: DEC-1, Descrição: PF04624), que seria restrito à *Drosophila*, tem suas particularidades: *D. pseudoobscura* é semelhante a *D. melanogaster*, porém, ao invés de 9 regiões de repetição, possui apenas duas, enquanto *D. albomicans* não as possui, bem como não as possuem outras espécies de moscas estudadas que não pertencem à família Drosophilidae. Pesquisas com as sequências de ortólogos do *dec-1* no GenBank mostraram que a isoforma mais curta (isoforma FC106 ou X3) é mais facilmente detectável do que as isoformas mais longas (FC125 ou X2 e FC177 ou X1), e mais similar por comparação no alinhamento. Portanto, a isoforma menor (FC106 ou X3) foi utilizada para comparar a presença de domínios entre ortólogos. As posições dos domínios encontrados estão destacadas nos alinhamentos de isoformas de *A. fraterculus* e *A. obliqua* (Figura 8, Apêndice A), e a comparação destes domínios entre *Anastrepha* e ortólogos está apresentada na Tabela 3.

Tabela 3: Percentuais de identidade dos domínios putativos do *dec-1* (isoforma FC106 ou X3) entre espécies de *Anastrepha* e outros ortólogos em Cyclorrhapha da ordem Diptera.

Espécie	Família	Tamanho da proteína	Domínio N-terminal (%)	Domínio C-terminal (%)
<i>Anastrepha obliqua</i>	Tephritidae	950	88.69	75.00
<i>Anastrepha ludens</i>	Tephritidae	984	97.86	71.26
<i>Anastrepha grandis</i>	Tephritidae	972	91.42	64.20
<i>Rhagoletis zephyria</i>	Tephritidae	1071	64.82	48.68
<i>Rhagoletis pomonella</i>	Tephritidae	1071	64.82	48.68
<i>Zeugodacus cucurbitae</i>	Tephritidae	1106	64.74	46.24
<i>Ceratitis capitata</i>	Tephritidae	1120	63.86	46.11
<i>Bactrocera tryoni</i>	Tephritidae	1125	64.88	39.43
<i>Drosophila melanogaster</i>	Drosophilidae	950	27.07	29.01
<i>Drosophila albomicans</i>	Drosophilidae	921	30.57	27.10
<i>Drosophila pseudoobscura</i>	Drosophilidae	949	29.32	24.84
<i>Teleopsis dalmanni</i>	Diopsidae	1347	29.27	17.71

A comparação dos domínios de supostos ortólogos *dec-1* entre *Anastrepha* e outras espécies apresentada na Tabela 3 mostrou diferentes graus de conservação, com identidade variando de ~27% (*D. melanogaster*) a ~97% (*A. ludens*) para o domínio N-terminal, e de ~17% (*T. dalmanni*) a ~75% (*A. obliqua*) para o domínio C-terminal. Ambos os domínios são altamente conservados entre as espécies *A. fraterculus*, *A. obliqua* e *A. grandis*, com identidades próximas ou superiores a 90% no domínio N-terminal, porém, mostram identidades relativamente baixas quando comparados com outras espécies e outros Tephritidae. Essa identidade inferior é mais evidente na região C-terminal, possivelmente devido à ocorrência dos *splicing* internos identificados nas sequências de *A. fraterculus* e *A. obliqua* e apontados pelas setas pretas na Figura 3, que determinam as diferentes isoformas e seus respectivos derivados proteicos. A região C-terminal da isoforma FC106 analisada, por exemplo, codifica o seu derivado s60 e é equivalente ao início da região gênica que codifica o derivado s95 da isoforma FC125 (Figura 3), o que também pode contribuir para divergências entre as sequências de gêneros distintos e resultar em valores baixos de identidade.

Ambos os domínios encontrados (Pfam ID: DEC-1_N; Descrição: PF04625 e Pfam ID: DEC-1_C; Descrição: PF04626) incluem os produtos de clivagem s80 (80 kDa) que são

incorporados na casca do ovo, e os produtos da proteólise adicional de s80 em s60 (60 kDa). Em *D. melanogaster*, terminações carboxila distintas em isoformas diferentes são produzidas por *splicing* alternativo no segundo domínio (C-terminal) (Waring et al., 1990). Como mencionado anteriormente, há a possibilidade da clivagem de s80 em s60 deixar resíduos de glutamina expostos que teriam função de sítio receptor de aminas primárias, criando um processo de reticulação capaz de conferir estabilidade ao endocóxon (Waring et al., 1990), assim como ocorre no envelope dos queratinócitos pela ação da involucrina (Eckert e Green, 1986).

4.2 TESTES FILOGENÉTICOS E DE SELEÇÃO

4.2.1 Testes filogenéticos

Os resultados do teste de saturação de Xia (Xia and Li, 1998) para cada grupo de análise estão apresentados na Tabela 4. Os níveis de saturação e taxas de transição e transversão não indicam que as taxas de substituições sinônimas (*dS*) estejam subestimadas com o aumento da distância evolutiva, o que sugere que elas ainda não atingiram o nível de saturação, exceto a taxa pela Iss.c A (taxa de substituições saturadas em filogenias assimétricas) do grupo de análise “Tephritidae” para divergência inferior 0,773.

Quando as taxas de substituição atingem a saturação, o valor de *dS* pode ser subestimado e resultar em um valor de ω (razão *dN/dS*) inflado, o que pode gerar falsos positivos e induzir a análises equivocadas (Smith e Smith, 1996). Porém, os resultados mostraram pouca saturação com um valor de diferença de apenas 0,05 (Iss.c A < Iss) (Tabela 4), o que indica que as análises das sequências de “Tephritidae” ainda podem ser utilizadas para as inferências acerca do histórico evolutivo do gene *dec-1*, embora devam ser tratadas de forma cuidadosa.

Tabela 4: Resultados do teste de saturação de Xia das taxas de substituições sinônimas.

Teste de Xia					
Grupo	Iss	Iss.c S	P	Iss.c A	P
Anastrepha	0,04	0,7377	0	0,4421	0
Tephritidae	0,773	0,8301	0,0007	0,7226	0,0026
Cyclorrhapha	0,6267	0,8103	0	0,6921	0

Para investigar padrões de evolução molecular, as sequências foram agrupadas em três conjuntos que representam níveis hierárquicos crescentes de diferenciação: no grupo “*Anastrepha*” as análises foram efetuadas apenas entre espécimes de *Anastrepha*, em um fragmento do éxon 3 de 729 pb, enquanto que para os grupos “Tephritidae” e “Cyclorrhapha”, as análises foram feitas com os éxons de 1 a 5, contendo 2.886 pb. As respectivas árvores filogenéticas inferidas para cada conjunto são mostradas na Figura 4. Vale ressaltar que as árvores filogenéticas geradas, principalmente as que contrastam gêneros mais distantes (“Tephritidae tree” e “Cyclorrhapha tree”), não possuem o objetivo de recuperar a filogenia desses grupos, que é bem conhecido, mas o de parametrizar adequadamente os testes de seleção, que devem ser baseados na história evolutiva dos ramos.

A árvore filogenética do grupo “*Anastrepha*” (Figura 4, *Anastrepha* tree) foi gerada utilizando o modelo de substituição HKY+G4, e pode-se identificar *A. obliqua* como um ramo único (quadrado verde) e separado das sequências de *A. fraterculus*, por sua vez divididas em três ramos distintos (quadrado azul, e ramos restantes dos quadrados amarelo e pontilhado). O primeiro ramo de *A. fraterculus* nesta árvore dividiu esta espécie e *A. obliqua* com duas amostras que pertencem à mesma região geográfica (São Tomé da Letras-MG). Já o segundo agrupou sequências de indivíduos do sul e sudeste do Brasil (São Tomé das Letras- MG e Criciúma-SC), enquanto o terceiro agrupou as amostras de Pinheiros-ES em um outro ramo. Embora não haja subsídios para afirmar a existência de morfotipos nos dados analisados, é possível que tal divisão seja resultado da presença destas “linhagens evolutivas” de *A. fraterculus* que já foram visualizadas e atestadas para populações brasileiras em análises de morfometria anteriores (Hernández-Ortiz et al., 2015; Prezotto et al., 2019). A filogenia construída por Congrains e colaboradores (2021) com genes oriundos de transcriptomas também diferenciou *A. fraterculus* em três diferentes ramos e mostrou um bom suporte para um ramo de *A. obliqua*. Acredita-se que essas diferentes linhagens de *A. fraterculus* podem estar sob influência de pressões seletivas distintas (Manni et al., 2015; Smith-Caldas et al., 2001).

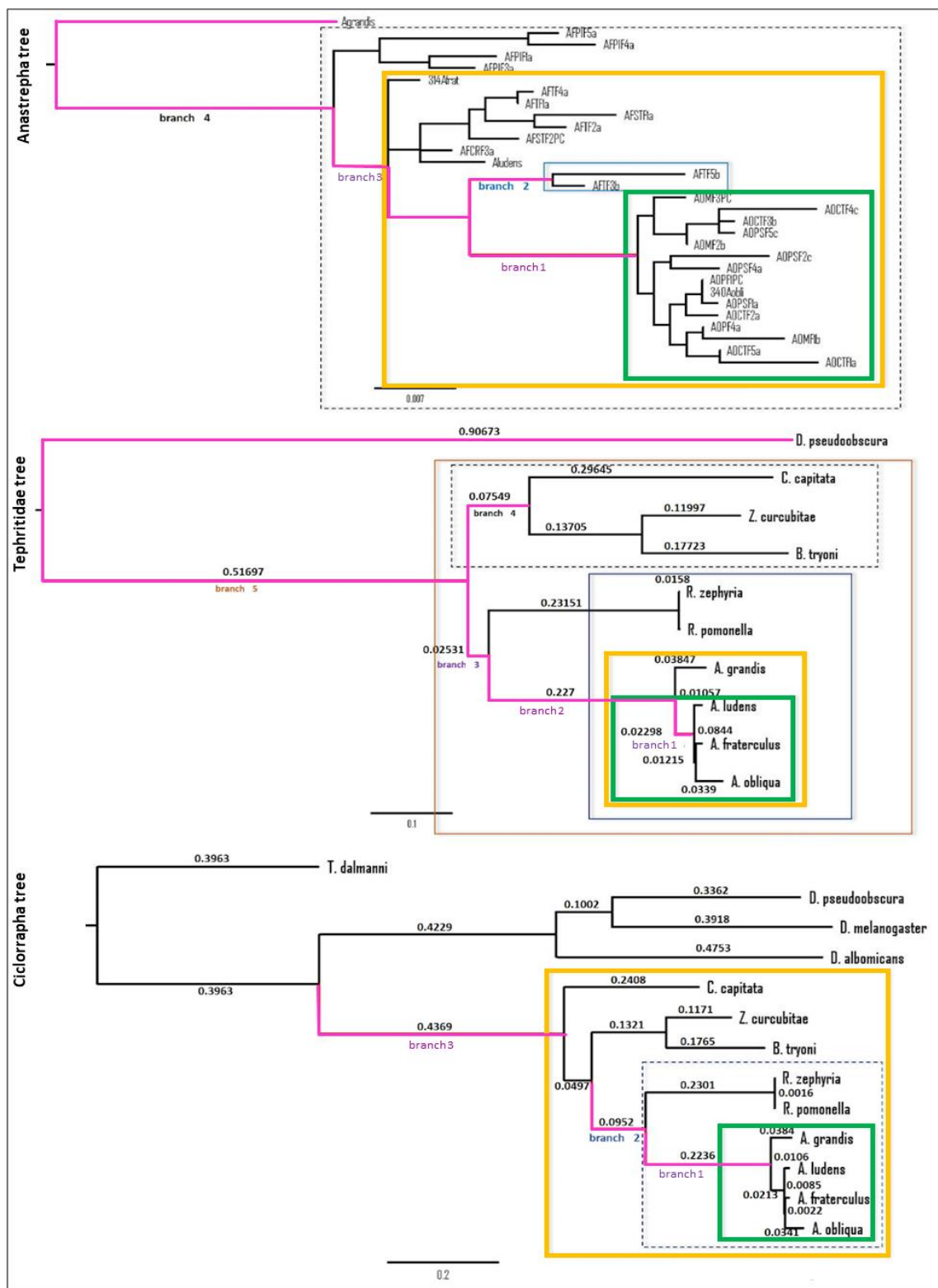


Figura 4: Árvores filogenéticas dos grupos “Anastrepha”, “Tephritidae” e “Cyclorhapha”. Os ramos em rosa foram submetidos a análise de seleção de ramos. Na árvore de “Anastrepha”, as sequências de *A. obliqua* estão apresentadas no quadrado verde, enquanto os quadrados azul, amarelo e pontilhado separam três linhagens evolutivas encontradas para *A. fraterculus*. Na árvore “Tephritidae”, as espécies do gênero *Anastrepha* estão apresentadas no quadrado amarelo, enquanto o quadrado verde separa *A. grandis* de *A. ludens*, *A. fraterculus* e *A. obliqua*. O quadrado azul apresenta as sequências do gênero *Rhagoletis* e o quadrado pontilhado as sequências dos gêneros *Ceratitis*, *Zeugodacus* e *Bactrocera*. Na árvore de “Cyclorhapha”, o quadrado amarelo inclui as sequências de tefritídeos, e os quadrados pontilhado e verde separam, respectivamente, as sequências do gênero *Rhagoletis* e do gênero *Anastrepha*.

A filogenia do segundo grupo, o “Tephritidae” (Figura 4, Tephritidae tree), foi gerada pelo modelo de substituição HKY+I+G4, e mostrou a monofilia de *Anastrepha*, em um ramo que inclui *A. grandis*, *A. ludens*, *A. fraterculus* e *A. obliqua* com bom suporte (quadrado amarelo), e a relação mais próxima desse gênero com *Rhagoletis*, separados dos gêneros *Ceratitis* e *Bactrocera*, o que também foi encontrado em estudos filogenéticos com caracteres moleculares de Toxotrypanini (Mengual et al., 2017), com o gene *vrrille* (Rego et al., 2021), e com a proteína quimiossensorial candidata CSP4 de tefritídeos (Segura-León et al., 2022).

A construção da árvore filogenética para “Cyclorrhapha” (Figura 4, Cyclorrhapha tree) foi gerada utilizando-se o modelo de substituição TPM3uf+I+G4 e manteve a topologia previamente inferida para tefritídeos, separando *Anastrepha* e *Rhagoletis* dos outros gêneros de Tephritidae, embora aqui *C. capitata* tenha sido inferida como a sequência basal do clado. Estudos anteriores realizados com espécies de *D. melanogaster*, *D. yakuba* e *D. virilis*, por exemplo, revelaram sequências de aminoácidos relativamente conservadas entre estas espécies, aparentemente retendo características funcionais (Badciong et al., 2001). De modo geral, os ramos foram curtos entre espécies mais próximas e longos e bem diferenciados entre os grupos mais distantes.

4.2.2 Testes de seleção: análises globais do *dec-1*

Para detectar o padrão global de evolução molecular do gene *dec-1*, os testes de seleção realizados pelo método CODEML implementado no software PAML utilizaram os modelos M_0 , M_{1a} , M_{2a} , M_3 , M_7 , M_8 e M_{8a} , que foram contrastados por Máxima Verossimilhança de modo a comparar modelos nulos que não assumem seleção positiva com modelos alternativos que permitem seleção positiva de forma distinta. Foram realizadas quatro combinações de testes: 1) M_0 - M_3 , 2) M_{1a} - M_{2a} , 3) M_7 - M_8 e 4) M_{8a} - M_8 .

Ao contrastar o modelo M_0 que assume taxa de evolução uniforme em todos os sítios da sequência (Goldman and Yang, 1994; Yang and Nielsen, 1998) com o modelo M_3 que considera uma distribuição discreta de taxas de substituição e assume seleção positiva (Yang et al., 2000), foram obtidos p-valores significativos (0,00) para os grupos “*Anastrepha*”, “*Tephritidae*” e “*Cyclorrhapha*”. Isso indica que é possível rejeitar o modelo M_0 em favor do M_3 , o que indica que o gene *dec-1* apresenta heterogeneidade nas taxas evolutivas ao longo do gene, o que não é de se surpreender, já que muito poucas proteínas apresentam padrões similares de restrição funcional ao longo de sua sequência.

Na segunda combinação de testes (M_{1a} - M_{2a}), o primeiro modelo assume ausência de

seleção positiva, enquanto o segundo modelo inclui uma classe de aminoácidos em que há presença de seleção positiva (Nielsen and Yang, 1998; Yang et al., 2005). Para nenhum dos três conjuntos de análise foram obtidos p-valores significativos (0,22 para “Anastrepha”, 1,00 para “Tephritidae” e 1,00 para “Cyclorrhapha”), indicando que não há evidência suficiente para se rejeitar $M1_a$ em favor do $M2_a$ e não foi detectado seleção positiva, mas esse teste é mais restritivo, por combinar sítios com indicação fraca de seleção positiva junto com a categoria de $dN/dS=1$ (Álvarez-Carretero et al., 2023).

O terceiro teste de hipóteses contrastou M_7 - M_8 , no qual M_7 não permite seleção positiva, porém, permite seleção purificadora, enquanto o M_8 permite seleção purificadora e positiva (Yang et al., 2000). Para o conjunto de dados “Anastrepha”, o p-valor não foi significativo (0,075) e não foi possível rejeitar o modelo M_7 em favor do M_8 , o que indica que essas sequências estão sujeitas principalmente a pressões de seleção purificadora. Em contrapartida, os p-valores para “Tephritidae” e “Cyclorrhapha” foram significativos (0,00 para ambos) e o modelo M_7 foi rejeitado, o que sugere que alguns sítios específicos das sequências analisadas podem estar sob seleção positiva.

Por fim, na quarta combinação de testes (M_{8a} - M_8), o modelo M_{8a} permite seleção purificadora e neutra, mas não seleção positiva, enquanto o M_8 permite seleção positiva (Yang et al., 2000). Ao contrastá-los para os grupos de análise “Anastrepha” e “Cyclorrhapha”, foram obtidos p-valores significativos (0,04 e 0,00, respectivamente), indicando influência de seleção positiva em suas sequências. No entanto, o p-valor obtido para o grupo “Tephritidae” não foi significativo (0,054) e não foi possível rejeitar o modelo M_{8a} , o que aponta que as sequências podem estar sob maior influência de seleção purificadora ou próximas à neutralidade.

Quando há a aceitação tanto de testes $M2_a$, ou M_8 , que sugerem existência de uma classe de sítios sob seleção positiva, podemos realizar testes específicos para tentar identificar que sítios seriam estes. A análise global do gene realizada no *Bayes Empirical Bayes* (BEB), fornece tais evidências adicionais sobre as possíveis regiões e códons do *dec-1* que estariam evoluindo em resposta à seleção (Nielsen and Yang, 1998). A Figura 5 apresenta o gráfico destes códons e a região selecionada em verde indica o fragmento analisado para as sequências do grupo de dados “Anastrepha” (éxon 3). Os valores no eixo y representam a probabilidade posterior de o valor de ω estar acima de 1 em cada sítio analisado. A representação gráfica foi realizada a partir dos dados de NEB (*Naïve Empirical Bayes*), porém, foram considerados somente os valores de probabilidade posterior de BEB. Os éxons

E1, E2, E3, E4 e E5, indicam as posições e o tamanho dos éxons no gene *dec-1*, e estão definidos como linhas coloridas abaixo do eixo x do gráfico.



Figura 5: Análise gráfica do Naïve Empirical Bayes (NEB) mostrando probabilidades posteriores de Bayes Empirical Bayes (BEB) de sítios sob seleção purificadora ($0 < \omega < 1$), positiva ($\omega > 1$) e neutra ($\omega=1$). No eixo x estão representados os códons das sequências analisadas e no eixo y, os valores de probabilidade posterior. Os círculos em cor laranja apontam os sítios sob seleção positiva com probabilidade posterior acima de 95%.

A análise do BEB indicou 7 possíveis códons com maiores probabilidades de estar sob seleção positiva no conjunto de dados “*Anastrepha*”: 6G ($\omega= 1,27$), 20F ($\omega= 1,32$), 43F ($\omega= 1,30$), 136K ($\omega= 1,17$), 146R ($\omega= 1,33$), 210P ($\omega= 1,30$) e 215E ($\omega= 1,15$), com probabilidades posteriores de 98,5%, 89,3%, 88%, 67,5%, 90%, 87,8% e 78,8% respectivamente, entretanto, somente o primeiro apresentou p-valor acima de 95%. Assim como o encontrado pela análise BEB, a análise global feita pelo teste FUBAR no HyPhy, que infere taxas de substituição dN/dS por sítio também apontou os códons 6G, 43F, 136K, 210P e 215E, com respectivas probabilidades posteriores de 97,6%, 91,3%, 93,9%, 92,5% e 94,2% de estar sob seleção positiva. Além disso, revelou que o códon 99K (93,5%) também sofre influência dessa força evolutiva.

Já o teste BUSTED, cujas análises são mais abrangentes e testam todos os ramos da

filogenia com toda a árvore definida em *foreground*, encontrou 12,7% das ramificações sob influência de seleção neutra e 87,3% sob seleção purificadora. Esse mesmo teste mostrou que o valor global das taxas de substituição sinônima ($dS= 1,22$) foi maior que as taxas de substituição não-sinônima ($dN= 0,71$), indicando que a maioria dos códons do éxon 3 analisados está sob seleção purificadora, como mostrado na Figura 5. Os resultados do FEL não foram significativos e não indicaram pressão de seleção positiva constante ao longo do gene, enquanto o teste MEME que infere se sítios individuais foram sujeitos a seleção episódica ou diversificada mostrou os códons 99K e 136K sob ação desse tipo de seleção. Os códons 99K e 136K também foram encontrados sob seleção positiva pelo FUBAR, em adição ao códon 194G. Isto indica que a seleção positiva pode ter afetado esses códons durante um tempo evolutivo específico e com ação episódica, mas não com influência que tenha se estendido à toda a região gênica analisada.

Para o conjunto de dados “Tephritidae”, apenas um códon (1147V) apresentou probabilidade de 97,7% de estar sob seleção positiva na análise do BEB. Esse mesmo códon mostrou probabilidade de 96,8% de estar sob este tipo de seleção nos resultados da análise do FUBAR. O BUSTED indicou que cerca de 26,3% dos ramos analisados estão sob seleção diversificadora, e 73,7% sob seleção purificadora. A divergência recente dentro do grupo dos tefritídeos pode ser consequência de pressões evolutivas que culminaram em adaptação ambiental. O BUSTED ainda mostrou que as taxas de substituições sinônimas ($dS= 2,81$) foram superiores às taxas de substituição não-sinônimas ($dN= 0,71$), indicando influência de seleção purificadora na maior parte das sequências analisadas para “Tephritidae”. O teste MEME apontou 32 sítios sob seleção positiva diversificadora, enquanto o teste FEL apontou 9 sítios sob essa mesma força evolutiva, entre eles, o códon 1147V que também foi encontrado pela análise do PAML. O gráfico para o método BEB (Figura 6) indicou apenas mais duas regiões como possíveis sítios sob seleção, porém, as probabilidades posteriores não apresentaram significância.

Assim como nos outros dois conjuntos de dados, “Cyclorrhapha” a análise global do PAML mostrou os códons 1309G e 1314V sob seleção positiva, com probabilidades posteriores de 96% e 98,2%, respectivamente. Esse mesmo códon foi apontado pelas análises do FUBAR com probabilidade de 96,7% e do FEL, que por sua vez relatou outros 6 sítios sob seleção diversificadora. Ademais, o MEME encontrou 46 sítios também sob seleção diversificadora episódica, os mesmos observados no gráfico do BEB (Figura 5), no qual estão representados como sítios sob seleção positiva, embora as probabilidades sejam pequenas.

Assim como nos outros dois grupos de análise, as taxas de substituições sinônimas ($dS=2,815$) foram maiores que as taxas não-sinônimas ($dN=0,714$) no teste global do BUSTED.

De modo geral, o gene *dec-1* parece estar sob influência de seleção purificadora em toda a sua extensão dos 5 códons analisados aqui, mas os dados indicam que ele provavelmente sofreu alguns episódios de seleção positiva em alguns códons ao longo de sua história, o que é refletido pelo sinal diferenciado de seleção encontrado nos três conjuntos de sequências estudados, que representam diferentes tempos evolutivos. Os dados encontrados por Congrains e colaboradores (2018) nos quais esse gene mostrou estar sob influência de seleção positiva sustentam os dados encontrados aqui. Entretanto, se as questões acerca do papel funcional do *dec-1*, de suas isoformas e dos seus derivados proteicos no processo de oogênese (Cummins e Nagle, 2005), e principalmente na formação do córion e da membrana vitelínica (Bauer e Waring, 1987) forem consideradas, a preponderância da seleção purificadora seria esperada. Uma mutação no éxon 3, por exemplo, seria capaz de alterar as três isoformas codificadas e desencadear, como já visto em outros estudos, esterilidade ou anomalias nos ovos (Waring et al, 1990; Kim et al., 2002; Spangenberg e Waring, 2007). Por sua vez, exatamente por sua importância em atributos do ovo que podem ter grande potencial adaptativo, certas mutações podem trazer vantagem adaptativa, o que está sendo refletido na combinação de padrões de seleção purificadora e positiva ao longo da história do gene.

4.2.3 Testes de seleção: análises de seleção de ramos

Nos testes de seleção de ramos, testa-se a hipótese da existência de episódios de seleção nas ramificações da árvore filogenética, uma vez que o valor da combinação de modelos (o ΔLRT) e a análise dos ramos *foreground* e *background* assume valores distintos para ω nos modelos nulo ($\omega_0 < 1$; $\omega_1=1$; $\omega_{FG}=1$) e alternativo ($\omega_0 < 1$; $\omega_1=1$; $\omega_{FG} > 1$). Também é possível testar a intensidade de seleção única sobre os códons de um ramo em particular: valores estabelecidos pela proporção de P1 e P2 são confrontados e especificam a intensidade de seleção constante, enquanto 2a e 2b especificam mudanças episódicas em intensidade de seleção, obtendo um valor de ω exclusivo para cada ramo. Os ramos selecionados para análise encontram-se na Figura 5.

Episódios de seleção foram verificados nos ramos de *A. obliqua* (ramo 1), de *A. fraterculus* (ramo 2) e entre *A. obliqua* e *A. fraterculus* (ramo 3) do grupo de dados “*Anastrepha*”. O ramo que separa *A. obliqua* e *A. fraterculus* mostrou resultado significativo para o cálculo do ΔLRT (4,27685; p: 0,03863), e demonstrou a presença de cerca de 3% dos

sítios sob seleção positiva, que possivelmente afetaram um subconjunto de códons locais durante um tempo evolutivo específico e que pode-se denominar “modelo episódico de evolução de proteínas”. A análise de BEB denotou o códon 194G com 94,6% de probabilidade de estar sob esse regime de seleção. O valor do $\omega_0=0,1280$ indicou que os outros ramos desse grupo estão sob influência da seleção purificadora com probabilidade de 97%.

Neste conjunto de dados, a análise da ramificação de *A. fraterculus* foi separada por se tratar de um complexo separado por morfotipos distintos, o que pode estar refletido nas topologias (Figura 4). O ramo mais próximo a *A. obliqua*, composto por duas amostras da mesma região geográfica que agruparam no mesmo clado (AFTF3b e AFTF5b), não indicou sinais de seleção diferenciada, o que pode ser reflexo de um regime de seleção similar, ou um pequeno comprimento de ramo que limita o poder estatísticos de análise. A comparação entre as ramificações de *A. fraterculus* e *A. obliqua* não foi significativa.

O teste RELAX detecta mudanças sutis na intensidade da seleção ao longo do histórico evolutivo a partir do valor de K: se menor que 1, indica possível seleção relaxada, e se maior que 1, indica potencial intensificação da seleção na comparação entre ramos distintos. Os resultados indicaram que o ramo 2 está sob seleção intensificadora ($K = 2,03$), enquanto o ramo 3 está sob seleção relaxada ($K = 0,49$). A seleção diversificadora foi observada em baixas porcentagens no ramo 2 ($\omega=3,30$; 8,1%) e no ramo 3 ($\omega=3,47$; 18,7%), porém, o p-valor para ambos não foi significativo (p: 0,6192 e p: 0,7026, respectivamente). O ramo 1 apresentou valor de $K=0$, o que indica que não foram detectadas evidências de um relaxamento ou intensificação da seleção.

Em “Tephritidae”, as ramificações que incluem *A. obliqua*, *A. fraterculus* e *A. ludens* foram investigadas para detectar episódios de seleção, mas o valor do cálculo para o ΔLRT não foi significativo ($\Delta LRT=1,49399$; p: 0,221). A análise realizada pela análise de ramos do BEB apontou 18 códons com probabilidades posteriores de estar sob influência de seleção positiva, mas apenas cinco com valores acima de 80%: 43G (83,5%), 70L (85,2%), 102Q (81,7%), 131G (84,7%) e 530I (80,7%). O ramo 2 que inclui *A. grandis*, *A. ludens*, *A. fraterculus* e *A. obliqua* foi significativo ($\Delta LRT= 7,629$, p: 0,005) e o BEB retratou 9 códons sob seleção positiva com altos valores de probabilidade, sendo que quatro deles mostraram valores acima de 80%: 436P (88,3%), 482M (89%), 521A (89,6%) e 527A (80%). Já o ramo 3, que separa *Anastrepha* de *Rhagoletis*, o ramo 4, que contém *C. capitata*, *Z. cucurbitae* e *B. tryoni*, e o ramo 5 mostraram ter passado por processo de seleção purificadora, com valor de

$\omega=0,126$. No teste RELAX, o ramo 1 apresentou $K=0$ ($p: 0,0141$), e não detectou evidências de pressão de seleção relaxada ou intensificadora. Os ramos 3 e 4 indicaram assinaturas de seleção intensificadora com valores de $K= 2,07$ ($p: 0,0036$) e $K= 45,28$ ($p: 0,0016$), e proporção de sítios sob seleção purificadora de 93,7% e 94%, respectivamente. O alto valor de K para o ramo 4 pode ser resultado de maior quantidade de mudanças no ramo, porém, com alta proporção de sítios sob seleção purificadora, possivelmente por se tratar de um clado com espécies bem estabelecidas evolutivamente.

Para a árvore de “Cyclorrhapha”, três ramos foram analisados: o que define o gênero *Anastrepha* (ramo 1), o que separa tal ramo do gênero *Rhagoletis* (ramo 2), e por último, o que separa Tephritidae e Cyclorrhapha (ramo 3). O ramo 1 obteve valor significativo para o cálculo do ΔLRT ($\Delta LRT= 9,3923$, $p:0,0022$) com cerca de 3,7% dos códons sob seleção positiva. Os valores de BEB indicaram códons com altos valores de probabilidade de estar sob essa força evolutiva, entre eles: 68A (87%), 318P (84,5%), 401M (91,6%), 429A (93,1%), 435A (86,2%) e 526N (96,8%). Já o ramo 2 não obteve resultados significativos para o ΔLRT , porém, o BEB sugeriu o códon 274H com 75,7% de probabilidade posterior de estar sob a influência de uma força seletiva.

O ramo 3 apresentou valores altos de significância para o cálculo do ΔLRT ($\Delta LRT= 36,539$, $p: 0,0000$). Os valores de BEB indicaram 90 códons (cerca de 20%) com possibilidade de seleção positiva, dentre eles: 52L (99,6%), 172N (99,5%), 182L (99,6%), 277S (99,4%), 310I (99,2%) e 448M (99,8%). Os resultados do teste RELAX não foram significativos para nenhum dos 3 ramos de “Cyclorrhapha” analisados, indicando que não há evidência estatística nem de relaxamento ou intensificação da seleção nesses ramos.

É possível observar que as razões dN/dS variaram substancialmente entre os diferentes ramos, e a seleção positiva pode estar operando em apenas algumas ramificações (ramos de primeiro plano), enquanto os outros (ramos de segundo plano), estão sob seleção purificadora. Como a razão dN/dS é uma medida de pressões seletivas que compara as taxas fixadas de substituição entre duas sequências homólogas (Kimura, 1977), e foi desenvolvida para analisar sequências distantes de linhagens independentes, pode ser influenciada por outros fatores que afetam o histórico evolutivo, e não indicar diretamente a intensidade da seleção em sequências de linhagens próximas ou de uma única população (Kryazhimskiy and Plotkin, 2008). A razão Ka/Ks , por sua vez, compara as taxas não fixadas de substituição dentro de uma sequência: uma razão Ka/Ks maior do que 1 indica seleção positiva, igual a 1 indica ação de seleção neutra, e menor do que 1 indica seleção purificadora (Nei and

Gojobori, 1986). Ademais, é uma medida aplicável a espécies próximas e de uma mesma população (Sharp, 1997), sendo uma boa alternativa para complementar as análises de evolução molecular. Entretanto, a análise combinada de ramos e gene abordada evidenciou a combinação entre seleções positiva e purificadora que fica ofuscada quando se considera uma análise global do gene.

Em genes sob forte seleção purificadora, as mutações não-sinônimas teriam maior chance de serem eliminadas do que mutações sinônimas, uma vez que alterações de aminoácido têm maior chance de levarem a alterações que afetem negativamente o valor adaptativo de uma proteína. Ainda assim, não se pode afirmar que essas mudanças impactam diretamente modificações morfológicas, ou mesmo fisiológicas, e que tais devem ser investigadas individualmente, pois o gene *dec-1* é pouco explorado em estudos de evolução molecular.

4.3 MEDIÇÃO DA MAGNITUDE FÍSICO-QUÍMICA DAS SUBSTITUIÇÕES DE AMINOÁCIDOS

Embora os testes de detecção de seleção nucleotídica tragam informações acerca das taxas de substituições sinônimas e não-sinônimas e das forças de seleção que estão potencialmente agindo nas sequências, ainda não possuem ferramentas que consigam investigar se as mudanças nos aminoácidos promovem alterações estruturais ou físico-químicas nas proteínas. O TreeSAAP é um teste que analisa as sequências códon a códon não só para inferir eventos de substituição de aminoácidos, mas também para estimar as potenciais mudanças em uma gama de suas propriedades físico-químicas e avaliar quais teriam maior chance de promover mudanças estruturais mais radicais nos domínios da proteína formada (Xia and Li, 1998; McClellane and McCracken, 2001).

As análises do TreeSAAP encontraram sete categorias de mudanças físico-químicas que indicam mudanças radicais, ou seja, com *z-score* de 6 a 8 e dentro do grau de significância $p < 0,01$: “*Total non-bonded energy*” (Energia total não-ligada), “*Isoelectric point*” (Ponto Isoelétrico), “*Beta-structure tendencies*” (Tendências da estrutura Beta), “*Coil tendencies*” (Tendências da estrutura em hélice), “*Mean r.m.s. fluctuation displacement*” (Deslocamento de aminoácidos), “*Solvent accessible reduction ratio*” (Taxa de redução da solubilidade por solventes) e “*Bulkiness*” (Volume). Todas as substituições encontradas estão na Tabela 8 (Apêndice B).

No grupo de dados “*Anastrepha*”, o códon 6G apresentou a substituição de uma Glicina por uma Serina, o que aponta alteração da propriedade “*Total non-bonded energy*” (6, p: 0,01), e pode ter implicações no desempenho funcional proteico devido a mudanças nos contatos de interface proteína-proteína. Ainda neste grupo de sequências, a substituição no códon 136K (Lisina por Glutamina), mostrou alteração na propriedade “*Isoelectric point*” (6, p:0,01), que se refere a mudanças de pH na membrana. Isto sugere que a proteína de membrana pode estar se adaptando a este tipo de desequilíbrio químico, uma vez que a molécula não consegue suportar a carga elétrica líquida.

As substituições do códon 1147G (Glicina por Valina, Serina ou Prolina) no grupo de dados de “*Tephritidae*” apresentaram o maior número de propriedades com possíveis alterações: “*Beta-structure tendencies*” (6, p:0,001), “*Coil tendencies*” (7, p:0,001), “*Mean r.m.s. fluctuation displacement*” (7, p:0,001) e “*Solvent accessible reduction ratio*” (7, p:0,01). A primeira refere-se à probabilidade de um resíduo estar em uma estrutura secundária beta, enquanto a segunda avalia a estrutura helicoidal transmembrana, e a terceira, a capacidade de um resíduo de aminoácido mudar de posição na estrutura tridimensional da proteína. A quarta propriedade, por sua vez, compete à acessibilidade superficial da proteína, que afeta sua interação com solutos e solventes. Já a substituição do códon 1314V (Prolina, Glicina, Isoleucina, Leucina ou Serina) do grupo de dados de “*Cyclorrhapha*” foi relacionada à propriedade *Bulkiness* (6, p:0,01) que pode não só afetar a hidrofobicidade e o enovelamento da proteína, mas também a sua ligação a substratos específicos. Todas as referências conceituais sobre as propriedades físico-químicas foram retiradas de Gromiha et al., (1997), Gromiha et al., (1999), Woolley et al., (2003), Novotny et al., (2007) e Beckstead et al., (2009).

Os códons com modificações físico-químicas apontados pelo TreeSAAP também foram encontrados sob seleção positiva pelos testes de seleção, como relatado no item 4.2, o que pode reforçar seu papel potencialmente seletivo. Em contrapartida, os códons encontrados sob seleção positiva que não foram apontados pelo TreeSAAP podem indicar possível seleção relaxada, na qual a propriedade geral do aminoácido se torna mais importante do que parâmetros específicos (Wertheim et al., 2015) ou mesmo mutações que são importantes para estrutura terciária do mRNA, que pode também afetar a expressão (Mauger et al., 2019). A ocorrência de mutações não-sinônimas no gene *dec-1* pode alterar os produtos proteicos das isoformas FC106, FC125 e FC177 aqui estudadas, e levar a alterações percebidas por seleção nos ovos, seja por meio de mudanças positivas, ou negativas (Waring et al, 1990; Kim et al.,

2002; Spangenberg e Waring, 2007). Desse modo, embora a estrutura terciária específica de cada uma das pró-proteínas do *dec-1* ainda não seja conhecida, é possível associar as modificações físico-químicas encontradas aos derivados proteicos desse gene e a algumas suas características funcionais. O códon 6G está localizado aproximadamente no meio do éxon 3, e a alteração físico-química radical encontrada pode afetar as 3 isoformas do gene *dec-1* nas sequências de “*Anastrepha*” analisadas: o derivado s20 da FC106, as regiões iniciais dos derivados s95 (FC125) e s85 (FC177). As alterações radicais encontradas nos códons 1147V (“*Tephritidae*”) e 1314V (“*Cyclorrhapha*”) localizam-se no meio e no final do éxon 4, respectivamente, e não possuem potencial de afetar a isoforma FC106, porém, ambas podem alterar a região C-terminal do derivado s85 (FC177).

Alterações no derivado s20 podem prejudicar a organização das interações proteicas precoces no processo de montagem do ovo (Badciong et al., 2001), assim como alterações no derivado s95 podem causar esterilidade nas fêmeas (Mauzy-Melitz and Waring, 2003). Ademais, alterações na função do derivado s85 podem levar à desestabilização da membrana vitelínica (Waring et al., 1990) e do endocórion tripartido, já que não haveria a agregação correta de proteínas do córion para a formação dos pilares, piso e teto reticulado (Mauzy-Melitz and Waring, 2003). De modo geral, as mutações apontadas ocorrem em regiões que poderiam comprometer a formação e a viabilidade dos ovos (Hawley and Waring, 1988).

Os códons apontados pelo TreeSAAP também foram apontados sob seleção positiva pelos testes realizados anteriormente, o que corrobora um padrão de seleção positiva episódica nas sequências analisadas, embora haja preponderância de seleção purificadora. As diferenças encontradas no gene *dec-1* entre sequências de *Tephritidae* e *Drosophilidae* são esperadas já que são de famílias taxonômicas distintas, porém, divergências são encontradas também entre os gêneros da família *Tephritidae* e entre espécies do gênero *Anastrepha*. Tais divergências podem ser resultado da variabilidade gerada por mutações aleatórias (Lynch, 2007), ou mesmo pela possível presença de linhagens diferentes diferenciadas em decorrência de pressões ecológicas, como fluxo gênico, introgressão de genes (Prezotto et al., 2019; Congrains et al., 2021) e seleção sexual (Swanson and Vacquier 1998; Jagadeeshan and Singh, 2007). Quaisquer que sejam as forças evolutivas envolvidas, a indicação de seleção positiva episódica associada a mudanças radicais de aminoácidos, mesmo que em reduzido número, sugere papel relevante da seleção na diferenciação do gene *dec-1*.

4.4 CARACTERIZAÇÃO OVARIANA

4.4.1 Caracterização morfológica

Durante o procedimento de dissecação, foi possível fazer algumas observações sobre as características dos ovários das fêmeas de *A. fraterculus* de diferentes idades, e estas são mostradas na Figura 6. Nota-se que os ovários do 1º dia (a1) são vestigiais e possuem aspecto filamentoso, sem nenhuma outra estrutura reprodutiva visível. Do 5º dia em diante (a5), os ovários permanecem pequenos e de aspecto vestigial, porém, os pares de espermatecas (b5, estrutura 2) e de glândulas acessórias (b5, estrutura 3) passaram a ser visíveis. As espermatecas são estruturas reprodutivas responsáveis por armazenar o esperma dos machos durante a reprodução, enquanto as glândulas acessórias secretam um muco durante a oviposição que recobre os ovos, auxiliando na fixação destes no substrato (Chapman, 2013)

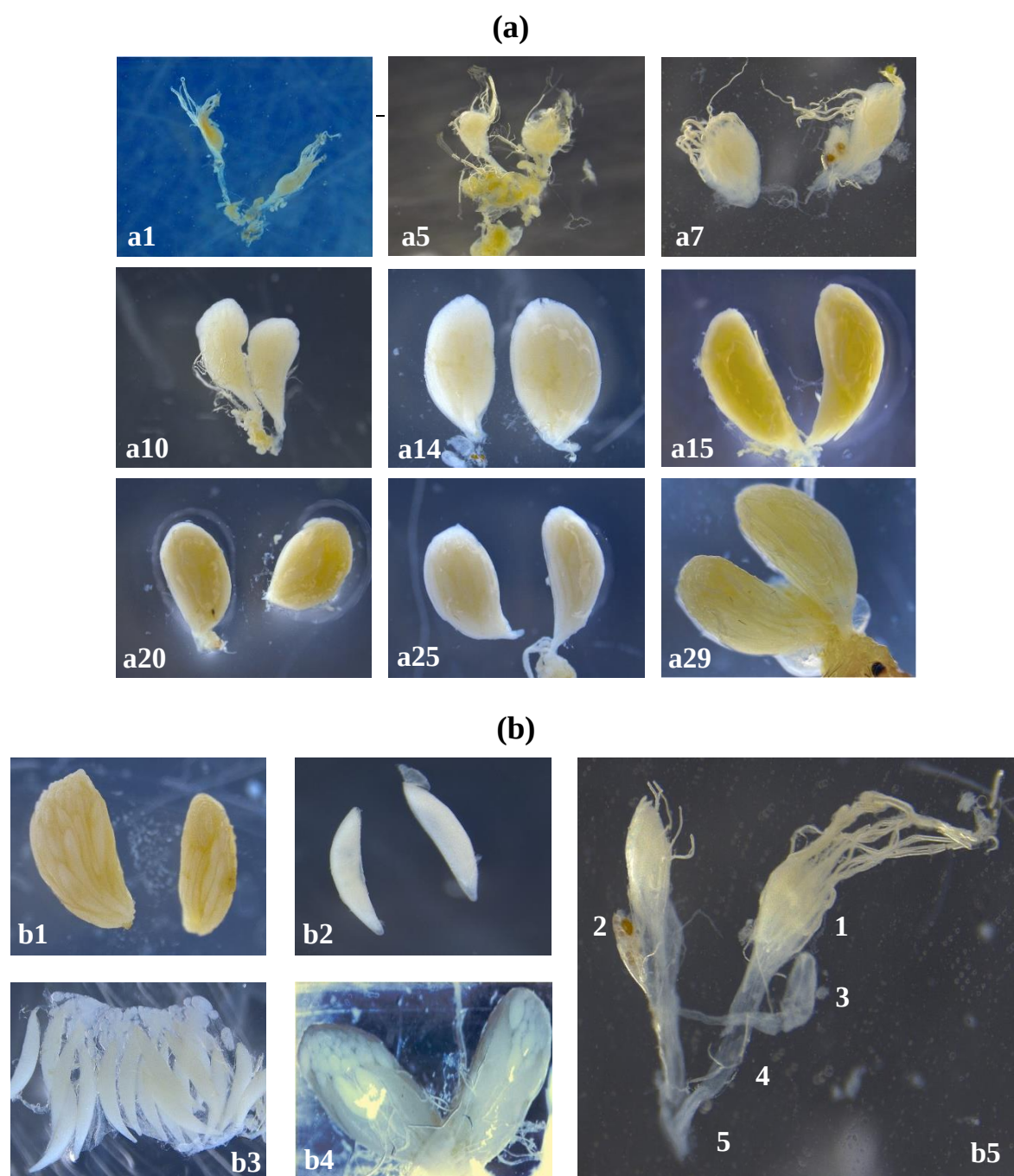


Figura 6: Pares de ovários de moscas de idades distintas de *Anastrepha fraterculus*. Diferenças morfológicas entre os ovários de 1, 5, 7, 10, 14, 15, 20, 25 e 29 dias (a). Características observadas durante a dissecação (b). Em B5 são apresentadas as estruturas: ovários (1), espermatecas (2), glândulas acessórias (3), oviduto lateral (4) e oviduto mediano (5).

A partir do 10º dia (a10), foram encontrados ovócitos em diferentes estágios de desenvolvimento e dispostos de maneira linear, no qual são observados os primeiros ovos completos, porém, em pequena quantidade. A presença de numerosos ovos corionados, completamente desenvolvidos e ocupando toda a área dos ovários foi detectada somente no 14º (a14) e 15º (a15) dias.

4.4.2 Análises estatísticas

Os p-valores para variância e média obtidos para medidas de caracteres reprodutivos em ambos os lados de cada mosca não indicam diferença estatisticamente significativa entre os lados (Tabela 5), sugerindo a presença de simetria bilateral nos ovários.

Tabela 5: p-valor das diferenças entre as variâncias e médias dos ovários esquerdo e direito para verificação de simetria bilateral.

Idade (dias)	p-valor			
	Índice ovariano		Carga de ovos	
	Variância	Média	Variância	Média
1	0,97	0,58	-	-
5	0,76	0,74	-	-
7	0,59	0,95	-	-
10	0,34	0,47	0,62	0,65
14	0,90	0,88	0,94	0,78
15	0,95	0,35	0,57	0,64
20	0,80	0,43	0,77	0,84
25	0,93	0,82	0,73	0,87
29	0,97	0,94	0,76	0,94

Os testes de normalidade de Ryan-Joiner (1976) para todos os caracteres separados por idade e considerando lados esquerdo e direito em conjunto (uma vez que não houve assimetria significativa) indicam que as distribuições das variâncias não rejeitam a hipótese nula de normalidade (Tabela 9 do Apêndice C). Sabendo que a distribuição dos dados segue um padrão normal, o método F foi aplicado para determinar a diferença entre as variâncias, cujos resultados obtidos (variâncias iguais ou diferentes) foram implementados no teste *t* de Student para determinar a igualdade entre as médias populacionais.

A Tabela 10 (Apêndice C) apresenta os resultados dos testes da diferença entre as variâncias e da igualdade entre as médias. As diferenças foram obtidas de acordo com a análise do p-valor para 95% de confiança, em que classes são consideradas como

estatisticamente distintas se $p < 0,05$ para a rejeição de H_0 (Student, 1908; Neter et al., 1985). As médias que mostraram ser estatisticamente diferentes entre si foram submetidas a um teste t de Student adicional a fim de identificar quais delas são estatisticamente maiores ou menores do que as outras. Esta análise também considera o p-valor para 95% de confiança, porém, para a hipótese alternativa (H_1): se $p < 0,05$, não rejeita H_1 e a média do indivíduo mais velho é maior que a média do indivíduo mais novo, e se $p > 0,05$, rejeita H_1 e a média do indivíduo mais velho é menor que a média do indivíduo mais novo (Neter et al., 1985). A Figura 7 apresenta os gráficos boxplot elaborados para o Índice Ovariano (a) e para a Carga de Ovos (b).

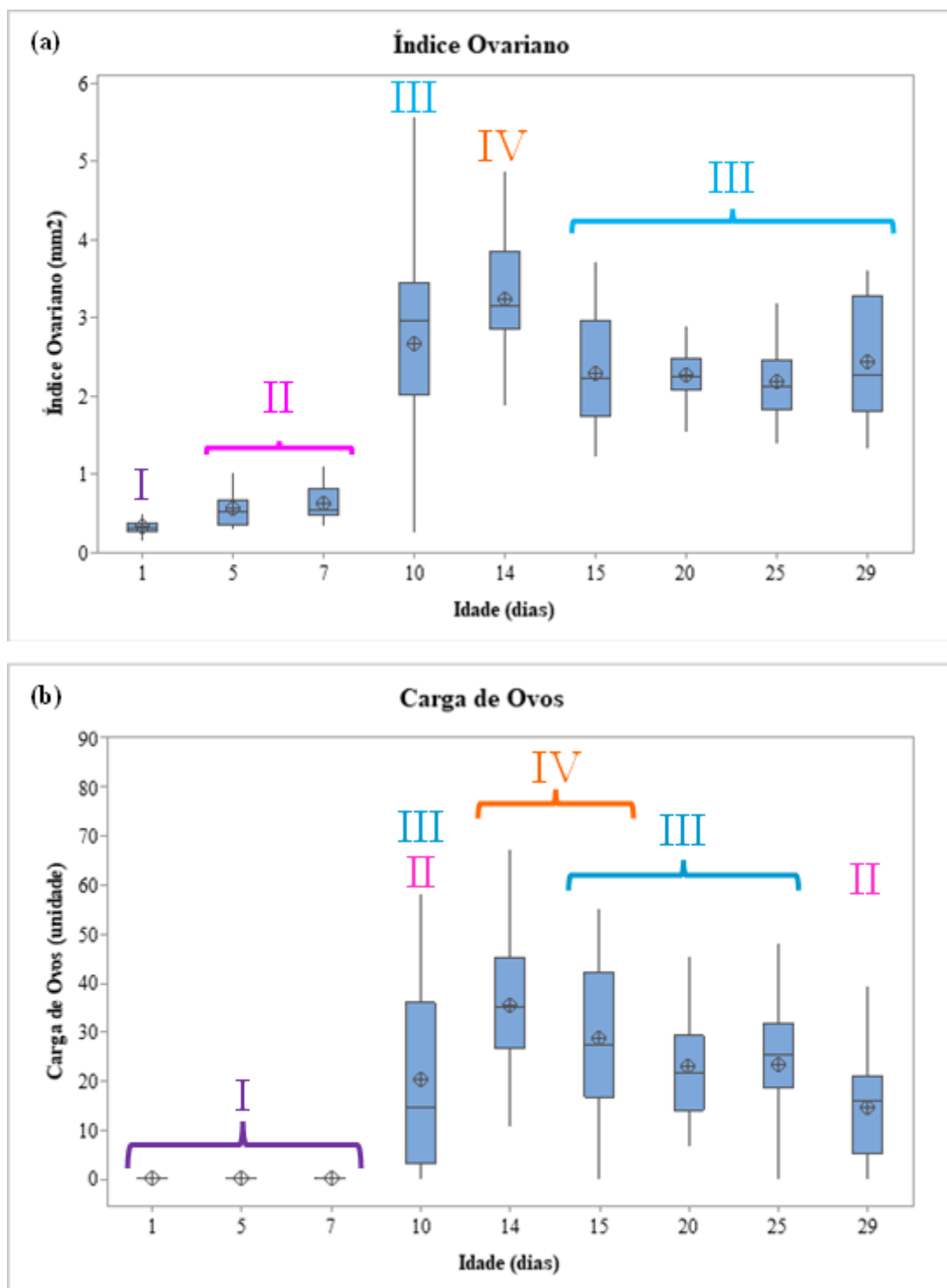


Figura 7: Gráficos boxplot dos valores obtidos para o Índice ovariano (a) e a Carga de ovos (b) das fêmeas de *Anastrepha fraterculus* dissecadas para as diferentes idades. Os grupos I (roxo), II (rosa), III (azul) e IV (laranja) foram formados de acordo com presença ou ausência de diferenças estatísticas significativas entre classes.

Para o Índice ovariano (Figura 7.a), a idade 1 é estatisticamente diferente de todas as outras idades, e assim constitui o Grupo I (cor roxa). As idades 5 e 7, por sua vez, são estatisticamente iguais entre si e diferentes do restante das idades analisadas, compondo o Grupo II (cor rosa). A média da idade 10 é estatisticamente igual às médias das idades 15, 20, 25 e 29, que são iguais entre si e diferentes da média da idade 14. Portanto, as idades 10, 15, 20, 25 e 29 compõem o Grupo III (cor azul), enquanto a idade 14 representa o Grupo IV (cor laranja). A comparação das médias diferentes mostrou que os valores de Índice ovariano do Grupo II são estatisticamente maiores que o do Grupo I, e ambos são menores que os valores do Grupo III. Por fim, o valor do Grupo IV se mostrou estatisticamente maior que todos os outros grupos.

O agrupamento das idades 10, 15, 20, 25 e 29 no Índice ovariano (Figura 7.a), pode indicar uma fase de transição entre o começo e o final da coriogênese, na qual alguns ovócitos estão iniciando seu processo de maturação enquanto outros estão finalizando, o que explicaria também a presença de alguns ovos corionados. Neste caso, após ovipor, as fêmeas retornariam para a fase de transição, na qual o tamanho dos ovários ainda sem ovos completos ou com baixa quantidade deles, seria equivalente ao tamanho dos ovários pós-oviposição. Entretanto, como os valores se mantêm da idade 15 a 29, este período pode representar um estado estacionário de reprodução no qual a fêmea se prepara para o próximo ciclo reprodutivo, ou começa a sofrer as consequências do envelhecimento e passa a depender da disponibilidade de componentes energéticos para a formação dos ovos (Bell e Bohm, 1975; Kendra et al., 2005).

Para a Carga de ovos (Figura 7.b), as idades 1, 5 e 7 não são estatisticamente diferentes entre si e constituem o Grupo I (cor roxa), já que não foram encontrados ovos nas fêmeas dessa etapa da vida. A média da idade 10 não é estatisticamente diferente das médias das idades 15, 20, 25 e 29, porém, a idade 29 é diferente das idades 15, 20 e 25. Por isso, dois grupos foram formados: o Grupo II com as idades 10 e 29 (cor rosa), e o Grupo III com as idades 10, 15, 20 e 25 (cor azul). A idade 14 formou o Grupo IV (cor laranja) com a idade 15, já que é estatisticamente diferente das idades 10, 20, 25 e 29. A comparação das médias diferentes mostrou que os valores da Carga de ovos das idades 14, 15, 20 e 25 são estatisticamente maiores que as das idades 10 e 29, que por sua vez, não diferem entre si.

A presença de flutuações na quantidade de ovos pode ser resultado da oviposição inicial em lotes e enquanto a oogênese ainda possui certo grau de sincronia, o que reduz o número de ovos, e conseqüentemente, o tamanho dos ovários (Kendra et al., 2016). Ademais,

tal variação pode até mesmo ser decorrente da presença de ovócitos em diferentes estágios de desenvolvimento no mesmo ovário (Kendra et al., 2016).

De modo geral, as diferenças estatísticas encontradas indicam que os processos de maturação sexual e de desenvolvimento dos ovários das fêmeas estudadas provavelmente se iniciam próximo ao 10º dia, no qual a presença de ovos corionados e de desenvolvimento completo foi observada pela primeira vez (Landolt e Davis-Hernandez 1993; Kendra et al. 2005). Embora a idade 14 apresente o maior valor de Índice Ovariano e de Carga de Ovos dentre as idades estudadas, ela não pode ser apontada como pico do desenvolvimento ovariano nos indivíduos analisados, pois não há análise para as idades 11, 12 e 13. No entanto, é possível inferir que este pico ocorra entre as idades 10 e 14.

Os dias de 1 a 4 podem ser interpretados como pré-vitelogênese, enquanto os dias de 5 a 7 poderiam representar a vitelogênese, já que os ovários são maiores e poderiam justificar o início do acúmulo de gema (vitelo) (Tootle et al., 2011; Papantonis et al., 2015). Assim, as idades de 8 a 10 podem representar uma fase de transição entre vitelogênese e coriogênese: quando a coriogênese é completada no folículo, a fêmea está apta a ovipor (idades de 10 a 14). Porém, deve-se ressaltar que isso só seria confirmado com uma análise de expressão de genes e proteínas específicos, já que as etapas da oogênese são caracterizadas justamente pela expressão diferencial entre eles (Nadel e Kafatos, 1980; Margaritis, 1985; Jagadeeshan e Singh, 2007; Eickbush e Izzo, 2009; Tootle et al., 2011; Papantonis et al., 2015).

Considerando o sistema de seis estágios da oogênese de Raghu e colaboradores (2003) proposto para *B. cacuminata*, e dos estudos realizados com *A. suspensa* (Kendra et al., 2016) e das fases indicadas pelas comparações entre as médias, pode-se sugerir que as fêmeas aqui estudadas provavelmente se encontram em estágio 5 do desenvolvimento (coriogênese) do 10º ao 14º dia de vida adulta, e se mantêm em oviposição do 15º dia de vida em diante. Os estágios de 1 a 4 (pré-vitelogênese e vitelogênese, respectivamente) ocorreriam do 1º até aproximadamente o 9º dia de vida, porém, não há dados específicos o suficiente para identifica-los. Segundo este mesmo sistema, o gene *dec-1* seria transcrito durante os estágios 3 e 4, e seus derivados proteicos se expressariam e desempenhariam funções durante o estágio 5.

Ao considerar o modelo de oogênese proposto para *D. melanogaster* (Noguerón et al., 2000), a pré-vitelogênese ocorreria durante os estágios de 1 a 7, a vitelogênese se iniciaria no estágio 8, e a transcrição do gene *dec-1* ocorreria dos estágios de 9 a 12 (ambos os casos do 1º até aproximadamente o 9º dia de vida). De acordo com este modelo, a coriogênese ocorreria

dos estágios de 11 a 14 (do 10º ao 14º dia de vida das fêmeas), nos quais o gene *dec-1* e seus derivados seriam expressos, auxiliando na formação da casca dos ovos (Waring et al., 1990).

Após tais considerações acerca do desenvolvimento ovariano das fêmeas de *A. fraterculus*, foi verificado se há relação entre as variáveis utilizadas nas análises, e se sim, qual a força e a direção desta relação. Para isso, foi aplicado o Coeficiente de Correlação de Pearson (r), cujos resultados são apresentados na Tabela 6. As análises de correlação entre Índice ovariano x Asa, Carga de ovos x Asa e Carga de ovos x Tórax mostraram valores de r não significativos em nenhuma desses contrastes. Embora a correlação Índice ovariano x Tórax tenha sido positiva, o valor de r (0,25) foi baixo e apontou que há uma relação fraca entre as variáveis, mas não é possível predizer valores de Tórax a partir do Índice ovariano, ou vice-versa.

Tabela 6: Coeficientes de correlação de Pearson e p-valor para cada variável analisada.

Variáveis correlacionadas	r	p-valor	Relação
Idade x Índice ovariano	0,51	0,00	Positiva
Idade x Carga de ovos	0,42	0,00	Positiva
Idade x Asa	0,17	0,01	Positiva
Idade x Tórax	0,36	0,00	Positiva
Índice ovariano x Carga de ovos	0,82	0,00	Positiva
Índice ovariano x Asa	0,11	0,09	Inconclusivo
Índice ovariano x Tórax	0,25	0,00	Positiva
Carga de ovos x Asa	-0,06	0,35	Inconclusivo
Carga de ovos x Tórax	0,09	0,14	Inconclusivo
Asa x Tórax	0,44	0,00	Positiva

As correlações Asa x Tórax e Idade x Tórax são positivas de valor de r médio, o que indica que há uma relação entre as duas variáveis e é possível predizer parcialmente o tamanho do tórax a partir do tamanho da asa, e a partir da Idade. A correlação Idade x Asa é positiva, mas de força fraca, apontando que uma pode ser predita a partir da outra, porém, com uma maior taxa de erro associado. Estes resultados apontaram que a Idade não impacta de forma considerável os caracteres morfológicos Asa ($r = 0,17$) e Tórax ($r = 0,36$), que apesar de apresentarem força de correlação regular entre si, obtiveram um valor mais alto ($r = 0,44$).

Tais resultados podem ser explicados pelo fato das espécies da ordem Diptera serem holometábolos, ou seja, passam por processo de metamorfose no qual o desenvolvimento ocorre em fases de larva, pupa e adulto (Stehr, 2009; Chapman, 2013). As etapas que compõem a metamorfose dos holometábolos dependem da regulação do sistema endócrino (Wesner et al., 2014; Debecker et al., 2017; Heneberg et al., 2019) e da ação do hormônio 20-hidroxiecdisona (20E) e do hormônio juvenil (JH), que sinalizam o início e o fim do processo de desenvolvimento (Zhou e Riddiford, 2002; Jindra et al., 2013; Jindra et al., 2015). Em *Drosophila*, por exemplo, a morfogênese das asas e a maturação das junções neuromusculares ocorrem durante a etapa larva-adulto (morfogênese pupal) (Caygill e Johnston, 2008). Desse modo, os indivíduos eclodem das pupas já na forma adulta completa e não dependem da idade para que os caracteres corporais se desenvolvam.

Ademais, os hormônios 20E e JH podem ativar a formação dos ovócitos e o desenvolvimento da pré-vitelogênese, da vitelogênese e da coriogênese nas fêmeas adultas, desempenhando papel importante na oogênese e na reprodução feminina (Wyatt e Davey, 1996; Belles, 2005; Roy et al., 2018). Isso pode explicar a correlação positiva regular obtida pelas análises Idade x Índice ovariano e Idade x Carga de ovos com valores de r maiores (0,51 e 0,42, respectivamente), indicando que é possível prever parcialmente o Índice ovariano e a Carga de ovos a partir da Idade, além de sustentar a fraca ou inconclusiva relação entre caracteres corporais e caracteres reprodutivos.

Por fim, a relação Índice ovariano x Carga de ovos foi positiva e forte, o que mostra a relação entre as duas variáveis a partir da qual é possível prever a Carga de ovos a partir do Índice ovariano, e vice-versa, para uma parte considerável dos dados coletados. Sabendo que o Índice ovariano é calculado a partir da multiplicação entre a largura e o comprimento dos ovários (Landolt e Davis- Hernandez 1993; Kendra et al., 2005), e que o volume dos ovários aumenta de acordo com a quantidade de ovos, é esperado que o Índice ovariano aumente quando as fêmeas possuem maior quantidade de ovos. Ademais, também é esperado que fêmeas com Índice ovariano maior possuam maior Carga de ovos. Na Figura 7, por exemplo, é possível notar que os menores valores para Índice ovariano e Carga de ovos foram obtidos das fêmeas de 1, 5 e 7 dias de idade, enquanto os maiores foram obtidos das fêmeas de 14 dias de idade.

Os dados obtidos na Caracterização ovariana podem ser utilizados para embasar estudos que utilizam análise de expressão gênica do *dec-1* utilizando *primers* específicos. Assim, poderia ser analisado o padrão de expressão em fêmeas de 1 dia de idade

(representaria um controle negativo de expressão), de 9 dias (momento no qual o *dec-1* seria transcrito), 14 dias (maiores valores de Índice ovariano e Carga de ovos encontrados), 15 dias (idade na qual os valores começam a diminuir) e 29 dias (final do ciclo reprodutivo). Vale ressaltar a importância de se combinar os estudos morfológicos com análises moleculares para que o contexto do desenvolvimento ovariano envolvendo genes específicos como o *dec-1* seja melhor compreendido.

5 CONCLUSÕES

As sequências de dados genômicos de *A. ludens* e de dados transcriptômicos de *A. fraterculus*, *A. obliqua* e *A. grandis* foram efetivas em obter informações sobre o gene *defective chorion 1* (*dec-1*), pois permitiram a coleta de dados necessária para uma investigação populacional entre estas espécies do grupo *fraterculus*, e do processo de diferenciação deste gene.

As análises evolutivas moleculares forneceram evidências sobre os padrões evolutivos do gene *dec-1* e revelaram que ele sofreu ação de seleção positiva episódica ao longo de sua história evolutiva, embora continue sob influência de seleção purificadora, principalmente em regiões comuns às suas três isoformas caracterizadas.

Além disso, a estrutura do *dec-1* até então nunca investigada para nenhuma espécie do gênero *Anastrepha* foi caracterizada para as três isoformas encontradas na literatura (FC106, FC125 e FC177), porém, pouco se sabe sobre os mecanismos genéticos subjacentes que envolvem o contexto das pró-proteínas, nem mesmo para *D. melanogaster*.

Embora a análise ovariana não tenha identificado especificamente cada momento fisiológico do desenvolvimento reprodutivo ovariano das fêmeas de *A. fraterculus* utilizadas, foi possível diferenciar alguns deles morfológica e estatisticamente, e os dados dos ovários coletados mostraram ser uma etapa inicial informativa para fundamentar e guiar estudos mais completos acerca da caracterização dos desenvolvimentos ovariano e reprodutivo de populações desta espécie, principalmente em estudos de expressão gênica.

Considerando o papel essencial de *dec-1* no processo reprodutivo e de formação de ovos em *Drosophila*, e provavelmente em outros Diptera, tais dados reforçam a importância de se estudar padrões de expressão gênicos e produção de isoformas proteicas distintas deste gene, a fim de compreender como este processo ocorre e como o contexto evolutivo no qual o gene está inserido pode afetá-lo nestas espécies.

6 REFERÊNCIAS BIBLIOGRÁFICAS

- ÁLVAREZ-CARRETERO, S.; KAPLI, P.; YANG, Z. (2023). Beginner's Guide on the Use of PAML to Detect Positive Selection. **Molecular Biology and Evolution**. 40 (4): msad041.
- AMDAM, Gro V.; SIMÕES, Zilá L. P.; GUIDUGLI, Karina R.; NORBERG, Kari; OMHOLT, Stig W. (2003). Disruption of vitellogenin gene function in adult honeybees by intra-abdominal injection of double-stranded RNA. **BMC Biotechnology**. 03, 01-08.
- ANDERSSON, S.; LAMBERTSSON, A. (1991). Evolution of the *dec-1* eggshell locus in *Drosophila*. I. Restriction site mapping and limited sequence comparison in the melanogaster species subgroup. **Journal of molecular Evolution**. (USA).
- ASHBURNER, M.; BALL, C. A.; BLAKE, J. A.; BOTSTEIN, D.; BUTLER, H.; CHERRY, J. M.; SHERLOCK, G. (2000). Gene ontology: tool for the unification of biology. **Nature genetics**, 25(1), 25-29.
- BAUER, B. J.; WARING, G. L. (1987). 7C female sterile mutants fail to accumulate early eggshell proteins necessary for later chorion morphogenesis in *Drosophila*. **Dev. Biol.** 121, 349–358.
- BADCIONG, J. C.; OTTO, J. M.; WARING, G. L. (2001). The functions of the multiproduct and rapidly evolving *dec-1* eggshell gene are conserved between evolutionarily distant species of *Drosophila*. **Genetics**, 159(3), 1089-1102.
- BECKSTEAD, W. A.; EBBERT, M. T.; ROWE, M. J.; MCCLELLAN, D. A. (2009). Evolutionary pressure on mitochondrial cytochrome b is consistent with a role of CytbI7T affecting longevity during caloric restriction. **PLoS One**, 4(6), e5836.
- BELL, W. J.; BOHM, M. K. (1975). Oosorption in insects. **Biological Reviews** 50, 373–396.
- BELLES, X. (2005). Vitellogenesis directed by juvenile hormone. **Reprod. Biol. Invertebr.** 12:157–98.
- BUTLIN, R.; RITCHIE, M. G. (2001). Searching for speciation genes. **Nature**. 412, 31-33.
- BYRNE, B. M.; GRUBER, M.; AB, G. 1989. The evolution of egg yolk proteins. **Prog. Biophys. Mol. Biol.** 53, 33-69.
- CAI, T.; REILLY, T. R.; CERIO, M.; SCHMITT, M. E. (1999). Mutagenesis of SNM1, which encodes a protein component of the yeast RNase MRP, reveals a role for this ribonucleoprotein endoribonuclease in plasmid segregation. **Molecular and Cellular Biology**. Nov, 19 (11):7857-7869.
- CAMPOREALE, G.; ZEMPLINI, J.; EISSENBERG, J. C. (2007). Susceptibility to heat stress and aberrant gene expression patterns in holocarboxylase synthetase-deficient *Drosophila melanogaster* are caused by decreased biotinylation of histones, not of carboxylases. **The Journal of nutrition**, 137(4), 885-889. <https://doi.org/10.1093/jn/137.4.885>
- CAYGILL, E. E.; JOHNSTON, L. A. (2008). Temporal Regulation of Metamorphic Processes in *Drosophila* by the *let-7* and *miR-125* Heterochronic MicroRNAs. **Current Biology**. 18 (13), 943-950.
- CHAPMAN, R. F. (2013). The insects – structure and function, 5th edn. **Cambridge University Press**, Cambridge.
- CONGRAINS, C.; CAMPANINI, E. B.; TORRES, F. R.; REZENDE, V. B.; NAKAMURA, A. M.; DE OLIVEIRA, J. L.; DE BRITO, R. A. (2018). Evidence of adaptive evolution and relaxed constraints in sex-biased genes of south American and West

- Indies fruit flies (Diptera: Tephritidae). **Genome Biology and Evolution**, 10(1), 380-395. <https://doi.org/10.1093/gbe/evy009>.
- CONGRAINS, C.; ZUCCHI, R. A.; DE BRITO, R. A. (2021). Phylogenomic approach reveals strong signatures of introgression in the rapid diversification of neotropical true fruit flies (*Anastrepha*: Tephritidae). **Molecular Phylogenetics and Evolution**, 162, 107200.
- CHEN, S. H., SU, S. Y., LO, C. Z., CHEN, K. H., HUANG, T. J., KUO, B. H., & LIN, C. Y. (2009). PALM: a paralleled and integrated framework for phylogenetic inference with automatic likelihood model selectors. **PLoS One**, 4(12), e8116.
- CUMMINS, S. F.; NAGLE, G. T. (2005). *Aplysia capsulin* is localized to egg capsules and egg cordon sheaths and shares sequence homology with *Drosophila dec-1* gene products. **Peptides**, 26(4), 589-596.
- DARRIBA, D., POSADA, D., KOZLOV, A. M., STAMATAKIS, A., MOREL, B., & FLOURI, T. (2020). ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. **Molecular biology and evolution**, 37(1), 291-294. <https://doi.org/10.1093/molbev/msz189>.
- DEBECKER, S.; DINH, K. V.; STOKS, R. (2017). Strong Delayed Interactive Effects of Metal Exposure and Warming: Latitude-Dependent Synergisms Persist Across Metamorphosis. **Environmental Science & Technology**. 51 (4), 2409-2417.
- DIGAN, M. E.; SPRADLING, A. C.; WARING, G. L.; MAHOWALD, A. P. (1979). The genetic analysis of chorion morphogenesis in *Drosophila melanogaster*. In "Eucaryotic Gene Regulation ICN- UCLA Symposium" (R. Axel *et. al.*, Eds.), pp. 171-181. **Academic Press**, New York.
- DOKLAND, T. (1999). Scaffolding proteins and their role in viral assembly. **Cell Mol. Life Sci.** 55, 580-603.
- DUPUIS, J. R.; RUIZ-ARCE, R.; BARR, N. B.; THOMAS, D. B.; GEIB, S. M. (2019). Range-wide population genomics of the Mexican fruit fly: Toward development of pathway analysis tools. **Evolutionary applications**, 12 (8), 1641-1660. <https://doi.org/10.1111/eva.12824>.
- DYCK, V. A.; HENDRICH, J.; ROBINSON, A. S. (2005). Sterile insect technique. Principles and practice in area-wide integrated pest management. **Springer**, Dordrecht, The Netherlands.
- ECKERT, R. L.; GREEN, H. (1986). Structure and evolution of the human involucrin gene. **Cell** 46,583-589.
- EICKBUSH, T. H.; IZZO, J. A. (2009). Chorion genes: molecular models of evolution. **Cambridge University Press. Books Online**. Disponível em: <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/CBO9780511529931.009>.
- FIGUEIREDO, J. V. A.; PERONDINI, A. L. P.; SELIVON, D. (2016). Patterns of inner chorion structure in *Anastrepha* (Diptera: Tephritidae) eggs. **Arthropod Structure & Development**. 1-10.
- FINDLAY, G. D., SWANSON, W. J. (2010). Proteomics enhances evolutionary and functional analysis of reproductive proteins. **BioEssays** 32, 26-36.
- FREEDMAN, D.; PISANI, R.; PURVES, R. (2007). Statistics (international student edition). Pisani, R. Purves, 4th Edn. **WW Norton & Company**, New York.
- GIORGI, F. (1977). An EM autoradiographic study on ovarian follicle cells of *Drosophila melanogaster* with special reference to the formation of egg coverings. **Histochemistry** 52, 105-117.
- GOLDMAN, N.; YANG, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. **Mol Biol Evol.** 11, 725-736.

- GONÇALVES, V. R., SOBRINHO JR, I. S., MALAGÓ JR, W., HENRIQUE-SILVA, F., & DE BRITO, R. A. (2013). Transcriptome analysis of female reproductive tissues of *Anastrepha obliqua* and molecular evolution of eggshell proteins in the *fraterculus* group. **Insect Molecular Biology**, 22(5), 551-561. <https://doi.org/10.1111/imb.12045>.
- GUINDON, S.; DUFAYARD, J. F.; LEFORT, V.; ANISOMOVA, M.; HORDIJK, W.; GASCUEL, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. **Syst. Biol.** 59, 307–21.
- GROMIHA, M. M.; SELVARAJ, S. (1997). Influence of medium and long range interactions in different structural classes of globular proteins. **Journal of Biological Physics** 23: 151–162.
- GROMIHA, M. M.; OOBATAKE, M.; KONO, H.; UEDAIRA, H.; SARAI, A. (1999). Relationship between amino acid properties and protein stability: buried mutations. **J Protein Chem** 18: 565–578.
- HAWLEY, R. J.; WARING, G. L. (1988). Cloning and analysis of the *dec-1* female-sterile locus, a gene required for proper assembly of the *Drosophila* eggshell. **Genes & development**, 2(3), 341-349.
- HENDRICHS, J.; VERA, M. T.; DE MEYER, M.; CLARKE, A. R. (2015). Resolving cryptic species complexes of major tephritid pests, In De Meyer M, Clarke AR, Vera MT, Hendrichs J [eds.] Resolution of Cryptic Species Complexes of Tephritid Pests to Enhance SIT Application and Facilitate International Trade. **ZooKeys** 540: 5–39.
- HENEBERG, P.; BOGUSCH, P.; ASTAPENKOVÁ, A. (2019). The effects of contact exposure to azole fungicides on insect metamorphosis. **Crop Protection**. 121, 66-72.
- HERNÁNDEZ-ORTIZ, V.; BARTOLUCCI, A.F.; MORALES-VALLES, P.; FRÍAS, D.; SELIVON, D. (2012). Cryptic species of the *Anastrepha fraterculus* complex (Diptera: Tephritidae): A multivariate approach for the recognition of South American morphotypes. **Ann. Entomol. Soc. Am.** 105, 305-318.
- HERNÁNDEZ-ORTIZ, V.; CANAL, N. A.; TIGRERO SALAS, J. O.; RUÍZ-HURTADO, F. M.; DZUL-CAUICH, J. F. (2015). Taxonomy and phenotypic relationships of the *Anastrepha fraterculus* complex in the Mesoamerican and Pacific Neotropical dominions (Diptera, Tephritidae). In: De Meyer M, Clarke AR, Vera MT, Hendrichs J (Eds) Resolution of Cryptic Species Complexes of Tephritid Pests to Enhance SIT Application and Facilitate International Trade. **ZooKeys**. 540, 95–124.
- JAGADEESHAN, S.; SINGH, R. S. (2007). Rapid evolution of outer egg membrane proteins in the *Drosophila melanogaster* subgroup: a case of ecologically driven evolution of female reproductive traits. **Molecular biology and evolution**, 24(4), 929-938. <https://doi.org/10.1093/molbev/msm009>.
- JARROUS, N.; REINER, R.; WESOŁOWSKI, D.; MANN, H.; GUERRIER-TAKADA, C.; ALTMAN, S. (2001). Function and subnuclear distribution of Rpp21, a protein subunit of the human ribonucleoprotein ribonuclease P. **RNA**, 7(8), 1153-1164. doi:10.1017/S1355838201010469
- JINDRA, M.; PALLI, S. R.; RIDDIFORD, L. M. (2013). The Juvenile Hormone Signaling Pathway in Insect Development. **Annual Review of Entomology**. 58(1), 181-204.
- JINDRA, M.; BELLÉS, X.; SHINODA, T. (2015). Molecular basis of juvenile hormone signaling. **Current Opinion in Insect Science**. 11, 39-46.
- JONES, Weldon C.; KAFATOS, Fotis C. (1980). Structure, Organization and Evolution of Developmentally Regulated Chorion Genes in a Silkworm. **Cell**. 22, 855-867.
- KATOH, K.; ROZEWICKI, J.; YAMADA, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization, **Briefings in Bioinformatics**. 20 (4), 1160–1166. <https://doi.org/10.1093/bib/bbx108>

- KELLER, O.; KOLLMAR, M.; STANKE, M.; WAACK, S. (2011). A novel hybrid gene prediction method employing protein multiple sequence alignments. **Bioinformatics**, 27(6), 757-763.
- KENDRA, P. E.; MONTGOMERY, W. S.; MATEO, D. M.; PUCHE, H.; EPSKY, N. D.; HEATH, R. R. (2005). Effect of age on EAG response and attraction of female *Anastrepha suspensa* (Diptera: Tephritidae) to ammonia and carbon dioxide. **Environ. Entomol** 34:584–590.
- KENDRA, P. E.; MONTGOMERY, W. S.; EPSKY, N. D.; HEATH, R. R. (2016). Assessment Of Female Reproductive Status In *Anastrepha suspensa* (Diptera: Tephritidae). **Florida Entomologist** 89(2), 144-151.
- KIM, C.; HAN, K.; KIM, J.; YI, J. S.; KIM, C.; YIM, J.; KIM-HA, J. (2002). Femcoat, a novel eggshell protein in *Drosophila*: functional analysis by double stranded RNA interference. **Mechanisms of development**, 110 (1-2), 61-70.
- KIM, J. C.; NORDMAN, J.; XIE, F.; KASHEVSKY, H.; ENG, T.; LI, S.; ORR-WEAVER, T. L. (2011). Integrative analysis of gene amplification in *Drosophila* follicle cells: parameters of origin activation and repression. **Genes & development**, 25(13), 1384-1398.
- KIMURA, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. **Nature**. 267, 275–276.
- KORF, I. (2004). Gene finding in novel genomes. **BMC bioinformatics**, 5 (1), 1-9.
- KOSAKOVSKY POND, S. L.; FROST, S. D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. **Molecular Biology and evolution**, 22 (5), 1208-1222.
- KRYAZHIMSKIY, S.; PLOTKIN, J. B. (2008). The Population Genetics of dN/dS. **PLoS Genet**. 4(12), e1000304.
- KUMAR, S.; STECHER, G.; Li M., KNYAZ C.; TAMURA, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. **Molecular Biology and Evolution**. 35:1547-1549.
- LANDOLT, P. J.; DAVIS-HERNANDEZ, K. M. (1993). Temporal patterns of feeding by Caribbean fruit flies (Diptera: Tephritidae) on sucrose and hydrolyzed yeast. **Ann. Entomol. Soc. Amer** 86:749–755.
- LAREAU, L. F.; GREEN, R. E.; BHATNAGAR, R. S.; BRENNER, S. E. (2004). The evolving roles of alternative splicing. **Current Opinion in Structural Biology**. 14 (3), 273–282.
- LARKIN, M. A.; BLACKSHIELDS, G.; BROWN, N. P.; CHENNA, R.; MCGETTIGAN, P. A.; MCWILLIAM, H.; VALENTIN, F.; WALLACE, I. M.; WILM, A.; LOPEZ, R.; THOMPSON, J. D.; GIBSON, T. J.; HIGGINS, D.G. (2007). Clustal W and Clustal X version 2.0. **Bioinformatics** 23, 2947-2948.
- LIS, J. T.; SCHLEIF, R. (1975). Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. **Nucleic Acids Research**. 02, 383–390.
- LYNCH, M. (2010). Evolution of the mutation rate. **Trends Genet**. 26 (8), 345-52.
- MAJOROS, W. H.; PERTEA, M.; SALZBERG, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. **Bioinformatics**, 20 (16), 2878-2879.
- MALAVASI, A.; ZUCCHI, R. A. (2000). Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado. Ribeirão Preto (SP): **Editora Holos**, 327p.
- MANNI, M., LIMA, K. M., GUGLIELMINO, C. R., LANZAVECCHIA, S. B., JURI, M., VERA, T.; ...; MALACRIDA, A. R. (2015). Relevant genetic differentiation among Brazilian populations of *Anastrepha fraterculus* (Diptera, Tephritidae). **ZooKeys**, (540), 157.

- MANNI, M.; BERKELEY, M. R.; SEPPEY, M.; SIMÃO, F. A.; ZDOBNOV, E. M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. **Molecular Biology and Evolution**, 38 (10), 4647-4654.
- MARGARITIS, L. H. (1985). Structure and physiology of the eggshell. In “**Comprehensive Insect Physiology, Biochemistry, and Pharmacology**” (G. A. Kerkut and L. I. Gilbert, Eds.), Vol. 1, pp. 153-230. Pergamon, Elmsford, NY.
- MARGARITIS, L. H.; HAMODRAKAS, S. J.; PAPASSIDERI, I.; ARAD, T.; LEONARD, K. R. (1991). Three-dimensional reconstruction of innermost chorion layer of *Drosophila grimshawi* and *Drosophila melanogaster* eggshell mutant *fs(1)384*. **Int. J. Biol. Macromol.** 13, 247–253.
- MAUGER, D.; CABRAL, B. J.; PRESNYAK, V.; SU, S. V.; REID, D. W.; GOODMAN, B.; LINK, K.; KHATWANI, N.; REYNDERS, J.; MOORE, M.; MCFADYEN, I. (2019). mRNA structure regulates protein expression through changes in functional half-life. **Biophysics and Computational Biology**. 116 (48), 24075-24083.
- MAUZY-MELITZ, D; WARING, G. L. (2003). *fc177*, a minor *dec-1* proprotein, is necessary to prevent ectopic aggregation of the endochorion during eggshell assembly in *Drosophila*. **Developmental biology**, 255(2), 193-205.
- MCCLELLAN, D. A.; MCCRACKEN, K. G. (2001). Estimating the influence of selection on the variable amino acid sites of the cytochrome b protein functional domains. **Mol. Biol. Evol.**, 18, 917–925.
- MENGUAL, X.; KERR, P.; NORRBOM, A. L.; BARR, N. B.; LEWIS, M. L.; STAPELFELDT, A. M.; SCHEFFER, S. J.; WOODS, P.; ISLAM, M. S.; KORYTWOSKI, C. A.; URAMOTO, K.; RODRIGUEZ, E. J.; SUTTON, B. D.; NOLAZCO, N.; STECK, G. J.; GAIMARI, S. (2017). Phylogenetic relationships of the tribe Toxotrypanini (Diptera: Tephritidae) based on molecular characters. **Molecular Phylogenetics and Evolution**. 113, 84-112.
- MINDRINOS, M. N.; PETRI, W. H.; GALANOPOULOUS, V. K.; LOMBARD, M. F.; MARGARITIS, L. H. (1980). Crosslinking of the *Drosophila* chorion involves a peroxidase. Wilhelm Roux's Arch. 189,187-196.
- MINITAB, LLC. (2021). Minitab. Retrieved from: <<https://www.minitab.com>>.
- MURRELL, B.; WERTHEIM, J. O.; MOOLA, S.; WEIGHILL, T.; SCHEFFLER, K.; KOSAKOVSKY POND, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. **PLoS genetics**, 8 (7), e1002764.
- MURRELL, B.; MOOLA, S.; MABONA, A.; WEIGHILL, T.; SHEWARD, D.; SERGEI, L. K. P.; SCHEFFLER, K. (2013). FUBAR: A Fast, Unconstrained Bayesian AppRoximation for inferring selection. **Mol. Biol. Evol.** 30, 1196–1205.
- MURRELL, Ben; WEAVER, Steven; SMITH, Martin D.; WERTHEIM, Joel O.; MURRELL, Sasha; AYLWARD, Anthony; EREN, Kemal; POLLNER, Tristan; MARTIN, Darren P.; SMITH, Davey M.; SCHEFFLER, Konrad; Sergei L. KOSAKOVSKY POND, Sergei L. (2015). Gene-wide identification of episodic selection. **Mol Biol Evol.** 32, 1365- 1371.
- NADEL, M. R.; KAFATOS, F. C. (1980.) Specific protein synthesis in cellular differentiation. IV. The chorion proteins of *Bombyx mori* and their program of synthesis. **Developmental Biology**. 75, 26–40.
- NEI, M.; GOJOBORI, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. **Molecular biology and evolution**. 3(5), 418-426.

- NELSON, James E.; KRAWETZ, Stephen A. (1992). Purification of Cloned and Genomic DNA by Guanidine Thiocyanate/Isobutyl Alcohol Fractionation. **Analytical Biochemistry**. 207, 197-201.
- NETER, J.; WASSERMAN, W.; KUTNER, M. H. (1985). Applied Linear Statistical Models, Second Edition. **Irwin, Inc.**
- NILSON, L. A.; SCHÜPBACH, T. (1998). Localized requirements for windbeutel and pipe reveal a dorsoventral prepatter within the follicular epithelium of the *Drosophila* ovary. **Cell**, 93(2), 253-262.
- NIELSEN, R.; YANG, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. **Genetics**. 148, 929-936.
- NIELSEN, R. (2005). Molecular signatures of natural selection, **Annu Rev Genet**. 39, 197-218.
- NOVOTNY, M.; SEIBERT, M.; KLEYWEGT, G. J. (2007). On the precision of calculated solvent-accessible surface areas. **Acta Crystallogr D Biol Crystallogr** 63: 270–274.
- NOGUERÓN, M. I.; WARING, G. L. (1995). Regulated processing of *dec-1* eggshell proteins in *Drosophila*. **Dev. Biol.** 172, 272–279.
- NOGUERÓN, M. I.; MAUZY-MELITZ, D.; WARING, G. L. (2000). *Drosophila dec-1* eggshell proteins are differentially distributed via a multistep extracellular processing and localization pathway. **Developmental biology**, 225(2), 459-470.
- NORRBOM, A. L.; KORYTKOWSKII, C. A.; ZUCCHI, R. A.; URAMOTO, K.; VENABLE, G. L.; MCCORMICK, J.; DALLWITZ, M. J. (2012). *Anastrepha* and *Toxotrypana*: descriptions, illustrations, and interactive keys. **Version**: 29 May 2012. Available from <http://deltaintkey.com/anatox/intro.htm>.
- NORRBOM, A. L., BARR, N. B., KERR, P., MENGUAL, X., NOLAZCO, N., RODRIGUEZ, E. J.; ZUCCHI, R. A. (2018). Synonymy of *Toxotrypana* Gerstaecker with *Anastrepha* Schiner (Diptera: Tephritidae). **Proceedings of the Entomological Society of Washington**, 120(4), 834-841. DOI: 10.4289/0013-8797.120.4.834.
- NORRBOM, A. L., MULLER, A., GANGADIN, A., SUTTON, B. D., RODRIGUEZ, E. J., SAVARIS, M.; RUIZ-ARC, R. (2021). New species and host plants of *Anastrepha* (Diptera: Tephritidae) primarily from Suriname and Pará, Brazil. **Zootaxa**, 5044 (1), 001-074. DOI: [10.11646/ZOOTAXA.5044.1.1](https://doi.org/10.11646/ZOOTAXA.5044.1.1).
- PALMER J, STAJICH J. (2019). nextgenusfs/funcannotate: 854 funcannotate v1.6.0 Version 1.6.0. **Zenodo**. doi:105281/zenodo3354704.
- PAPANTONIS, A.; SWEVERS, L.; IATROU, K. (2015). Chorion genes: a landscape of their evolution, structure, and regulation. **Annual Review of Entomology**. 60, 177-194.
- PARKS, S.; SPRADLING, A. C. (1987). Spatially regulated expression of chorion genes during *Drosophila* oogenesis. **Genes & Dev**. 1987 (1), 497–509.
- PASCUCCI, T.; PERRINO, J.; MAHOWALD, A. P.; WARING, G. L. (1996). Eggshell assembly in *Drosophila*: processing and localization of vitelline membrane and chorion proteins. **Dev Biol** 177: 590–598.
- PREZOTTO, L. F., PERONDINI, A. L., HERNÁNDEZ-ORTIZ, V., FRÍAS, D., SELIVON, D. (2019). What can integrated analysis of morphological and genetic data still reveal about the *Anastrepha fraterculus* (Diptera: Tephritidae) cryptic species complex? **Insects**, 10 (11), 408.
- RAZA, M. F.; YAO, Z.; BAI, S., CAI, Z.; ZHANG, H. (2020). Tephritidae fruit fly gut microbiome diversity, function and potential for applications. **Bulletin of Entomological Research**, 110(4), 423-437.
- RAGHU, S.; P. HALCOOP, P.; DREW, R. A. I. (2003). Apodeme and ovarian development as predictors of physiological status in *Bactrocera cacuminata* (Hering) (Diptera: Tephritidae). **Australian J. Entomol** 42:281–286.

- REGO, N. F. C.; CHAHAD-EHLERS, S.; CAMPANINI, E. B.; TORRES, F. R.; DE BRITO, R. A. (2021). VRILLE shows high divergence among Higher Diptera flies but may retain role as transcriptional repressor of *clock*. **J Insect Physiol.** 133,104284.
- REZENDE, V. B.; CASTILLO, C. C.; LIMA, A. L. A.; CAMPANINI, E. B.; NAKAMURA, A. M.; OLIVEIRA, J. L.; CHAHAD-EHLERS, S.; SOBRINHO-JUNIOR, I.; DE BRITO, R. A. (2016). Candidate genes involved in differentiation between two non-model species of fruit flies (*Anastrepha*: Tephritidae) screened from head transcriptomes. **G3**. 06, 3283-3295.
- ROY, S.; SAHA, T. T.; ZOU, Z.; RAIKHEL, A. S. (2018). Regulatory Pathways Controlling Female Insect Reproduction. **Annual Review of Entomology**. 63 (1), 489-511.
- RUIZ-MONTOYA, L.; VALLEJO, R. V.; HAYMER, D.; LIEDO, P. (2020). Genetic and Ecological Relationships of *Anastrepha ludens* (Diptera: Tephritidae) Populations in Southern Mexico. **Insects**, 11(11), 815. doi:10.3390/insects11110815.
- RYAN, T. A.; JOINER, B. L. (1976). Normal Probability Plots and Tests for Normality," **Technical Report**, Statistics Department, The Pennsylvania State University. See further: <<https://blog.minitab.com/en/statistics-in-the-field/pencils-and-plots-assessing-the-normality-of-data>>.
- SAIKI, Randall K.; GELFAND, David H.; STOFFEL, Susanne; SCHARF, Stephen J.; HIGUCHI, Russell; HORN, Glenn T.; MULLIS, Kary B.; ERLICH, Henry A. (1988). Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. **Science**. 239, 487-491.
- SANGER, F.; NICKLEN, S.; COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. **Biochemistry**. 74, 5463-5467.
- SEGURA-LEÓN, O. L.; TORRES-HUERTA, B.; ESTRADA-PÉREZ, A. R.; CIBRIÁN-TOVAR, J.; HERNANDEZ-HERNANDEZ, F. D. L. C.; CRUZ-JARAMILLO, J. L.; MEZA-HERNÁNDEZ, J. S.; SÁNCHEZ-GALICIA, F. (2022). Identification of Candidate Chemosensory Gene Families by Head Transcriptomes Analysis in the Mexican Fruit Fly *Anastrepha ludens* Loew (Diptera:Tephritidae). **Int. J. Mol. Sci.**, 23, 10531.
- SCHWARZ, G. (1978). Estimating the dimensional of a model. **Ann. Stat.**, Hayward, v.6, n.2, p.461-464.
- SCHUTZE, M. K.; VIRGILIO, M.; NORRBOM, A.; CLARKE, A. R. (2017). Tephritid integrative taxonomy: where we are now, with a focus on the resolution of three tropical fruit fly species complexes. **Annu. Rev. Entomol.** 62, 147–164.
- SHARP, P. M. (1997) In search of molecular Darwinism. **Nature**. 385, 111–112.
- SHARP, P. A. (2009). The Centrality of RNA. **Cell**, 136(4), 577- 580.
- SIVINSKI, J. M. (1993). Longevity and fecundity in the Caribbean fruit fly (Diptera: Tephritidae): effects of mating, strain and body size. **Florida Entomol** 76:635–644.
- SILVA, Janisete G.; BARR, Norman B. (2006). Recent advances in molecular systematics of *Anastrepha* Schiner. In **Fruit Flies of Economic Importance: From Basic to Applied Knowledge**, pp. 13-28.
- SMITH, J. M.; SMITH, N. H. (1996). Synonymous nucleotide divergence: what is “saturation”? **Genetics**. 142 (3), 1033–1036.
- SMITH-CALDAS, M. R.; MCPHERON, B. A.; SILVA, J. G.; ZUCCHI, R. A. (2001). Phylogenetic relationships among species of the *fraterculus* group (*Anastrepha*: Diptera: Tephritidae) inferred from DNA sequences of mitochondrial cytochrome oxidase I. **Neotropical Entomology**, 30, 565-573.7
- SOLFERINI V. N.; MORGANTE J. S. (1987). Karyotype study of eight species of *Anastrepha* (Diptera: Tephritidae). **Caryologia** 40(3): 229–241.

- SPANGENBERG, D. K., & WARING, G. L. (2007). A Mutant *dec-1* Transgene Induces Dominant Female Sterility in *Drosophila melanogaster*. **Genetics**, 177, 1595-1608.
- SPRADLING, Allan C.; MAHOWALD, Anthony P. 1980. Amplification of genes for chorion proteins during oogenesis in *Drosophila melanogaster*. **Genetics**. 77, 1096-1100.
- STEHR, F. (2009). Metamorphosis. In Resh, V. H., & Carde, R. T. (Eds.), **Encyclopedia of insects** (2nd ed., pp. 627-628). Academic Press.
- STUDENT. 1908. The probable error of a mean. **Biometrika**. 6:1–25.
- SWANSON, Willie J.; VACQUIER, Victor D.; 2002. The rapid evolution of reproductive proteins. **Nat. Rev. Genetics**. 03, 137-144.
- TOOTLE, T. L.; WILLIAMS, D.; HUBB, A.; FREDERICK, R.; SPRADLING, A. (2011). *Drosophila* Eggshell Production: Identification of New Genes and Coordination by Pxt. **Plos One** 6(5): e19943.
- UNIPROT Consortium. (2019). UniProt: a worldwide hub of protein knowledge. **Nucleic acids research**, 47(D1), D506-D515.
- WARING, G. L.; HAWLEY, R. J.; SCHOENFELD, T. (1990). Multiple proteins are produced from the *dec-1* eggshell gene in *Drosophila* by alternative RNA splicing and proteolytic cleavage events. **Dev Biol**. 142, 1–12.
- WEI, D.; ZHANG, Y. X.; LIU, Y. W.; LI, W. J.; CHEN, Z. X.; SMAGGHE, G.; WANG, J. J. (2019). Gene expression profiling of ovary identified eggshell proteins regulated by 20-hydroxyecdysone in *Bactrocera dorsalis*. **Comparative Biochemistry and Physiology**. 30, 206-216.
- WILLIAMSON, M. P. (1994). The structure and function of proline-rich regions in proteins. **Biochemical journal**, 297(Pt 2), 249.
- WOOLLEY, S.; JOHNSON, J.; SMITH, M. J.; CRANDALL, K. A.; MCCLELLAN, D. A. (2003). TreeSAAP: Selection on Amino Acid Properties using phylogenetic trees. **Bioinformatics** 19: 671–672.
- WHITE, I. M. (2006). Taxonomy of the Dacina (Diptera: Tephritidae) of Africa and the Middle East. **African Entomology** 2, 1–156.
- WU, C. I. (2001). Genes and Speciation. **Journal of Evolutionary Biology**. 14, 889-891.
- WU, C. I.; TING, C. T. (2004). Genes and Speciation. **Nature Review Genetics**. 05, 114-122.
- WERTHEIM, J. O.; MURRELL, B.; SMITH, M. D.; KOSAKOVSKY POND, S. L.; SCHEFFLER, K. (2015). RELAX: detecting relaxed selection in a phylogenetic framework. **Molecular biology and evolution**, 32(3), 820-832.
- WESNER, J. S.; KRAUS, J. M.; SCHMIDT, T. S.; WALTERS, D. M.; CLEMENTS, W. H. (2014). Metamorphosis Enhances the Effects of Metal Exposure on the Mayfly, *Centroptilum triangulifer*. **Environmental Science & Technology**. 48(17), 10415-10422.
- WYATT, G. R.; DAVEY, K. G. (1996). Cellular and Molecular Actions of Juvenile Hormone. II. Roles of Juvenile Hormone in Adult Insects. Editor(s): P.D. Evans, **Advances in Insect Physiology**, Academic Press. 26, 1-155.
- XIA, X; LI, W. H. (1998). What amino acid properties affect protein evolution? **J. Mol. Evol.**, 47, 557–564.
- XIA X. 2018. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. **Molecular Biology and Evolution**. 35:1550–1552.
- YANG, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. **Comput. Appl. Biosci.** 13:555–556.
- YANG, Z.; NIELSEN, R. (1998). Synonymous and nonsynonymous rate variation in nuclear genes of mammals. **J Mol Evol**. 46, 409-418.

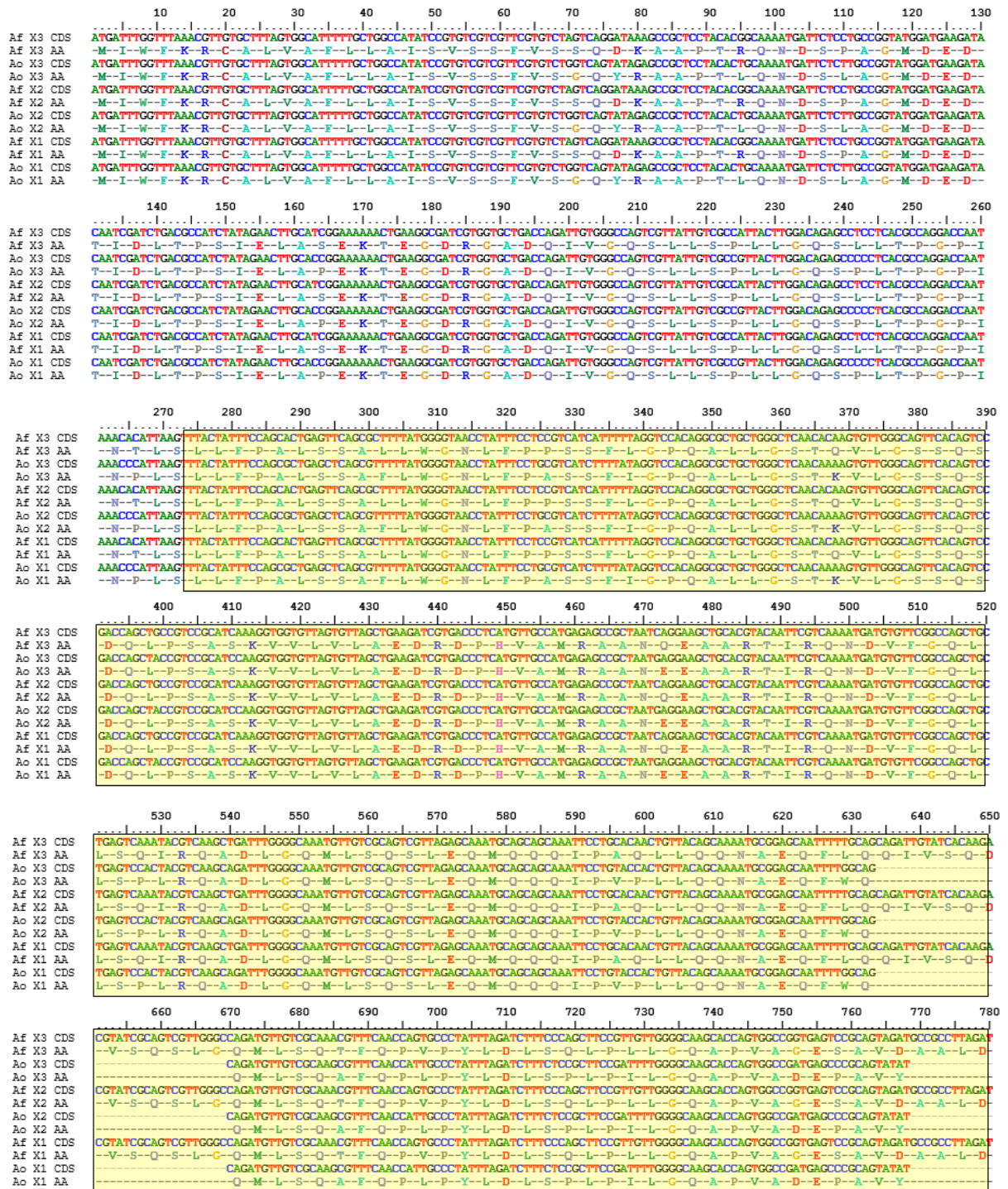
- YANG, Z.; WONG, W. S. W.; NIELSEN, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. **Mol Biol Evol.** 22, 1107-1118.
- YANG, Z.; NIELSEN, R.; GOLDMAN, N.; PEDERSEN, A-M. K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. **Genetics.** 155, 431-449.
- YANG, Z.; DOS REIS, M. (2011). Statistical properties of the branch-site test of positive selection. **Mol. Biol. Evol.** 28:1217-1228.
- ZENG, Z.; LIU, J.; STAM, L. F.; KAO, C.; MERCER, J. M.; LAURIE, C. C. (2000). Genetic Architecture of a Morphological Shape Difference Between Two *Drosophila* Species. **Genetics.** 154, 299-310.
- ZHANG, J.; NIELSEN, R.; YANG, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. **Mol. Biol. Evol.** 22:2472-2479. doi: 10.1093/molbev/msi237
- ZHOU, X.; RIDDIFORD, L. M. (2002). Broad specifies pupal development and mediates the 'status quo' action of juvenile hormone on the pupal-adult transformation in *Drosophila* and *Manduca*. **Development.** 129(9), 2259-69.
- ZUCCHI R. A. (2000). Espécies de *Anastrepha*, sinónimas, plantas hospedeiras e parasitóides, pp. 41–48 in **Moscas-das-frutas de Importância Econômica no Brasil: Conhecimento Básico e Aplicado**, edited by Malavasi A., Zucchi R. A. Holos, Ribeirão Preto.
- ZUCCHI, R. A.; A. MALAVASI; A. S. NASCIMENTO & J. M. M. WALDER. (2004). Prejuízos das moscas-das-frutas na exportação de citros. **Visão Agrícola.** 02, 73–77.

APÊNDICE A

Tabela 7: Lista de ortólogos da proteína do *defective chorion-1 protein (dec-1)* (isoforma FC106) de *A. fraterculus* e *A. obliqua* em Diptera, que estão publicamente disponíveis no NCBI GenBank.

Family\Species and gene ID	Transcript ID	Length (nt) RNA	Length (nt) CDS	Protein	Length (aa)	Isoform
Tephritidae						
<i>Rhagoletis zephyria</i> (RZ) Gene ID: 108362233	XM_017615135.1	3433	3216	XP_017470624.1	1071	X3
<i>Rhagoletis pomonella</i> (RP) Gene ID: 118735351	XM_036465058.1	3686	3216	XP_036320951.1	1071	X3
<i>Zeugodacus cucurbitae</i> (ZC) Gene ID: 105213646	XM_029041248.1	4075	3321	XP_028897081.1	1106	X4
<i>Ceratitis capitata</i> (CC) Gene ID: 101457983	XM_020859715.1	3639	3363	XP_020715374.1	1120	X3
<i>Bactrocera tryoni</i> (BT) Gene ID: 120777959	XM_040109580.1	3659	3378	XP_039965514.1	1125	X1
Diopsidae						
<i>Teleopsis dalmanni</i> (TD) Gene ID: 119681061	XM_038094130.1	4333	4044	XP_037950058.1	1347	X2
Drosophilidae						
<i>Drosophila melanogaster</i> (DM) Gene ID: 31691	NM_167133.2	3751	2853	NP_727203.1	950	C
<i>Drosophila albomicans</i> (DA) Gene ID: 117565635	XM_034244855.1	3032	2850	XP_034100746.1	949	X2
<i>Drosophila pseudoobscura</i> (DP) Gene ID: 6901818	XM_015186680.2	2946	2766	XP_015042166.2	921	X5

Figura 8: Domínios encontrados nos alinhamentos das isoformas FC106, FC125 e FC177 de *A. fraterculus* e *A. obliqua*. Estão destacados o domínio *Pfam* ID: DEC-1_N (Description: PF04625, DEC-1 protein, N-terminal region) em amarelo, o *Pfam* ID: DEC-1_C (Description: PF04626, DEC-1 protein, C-terminal region) em verde, e o *Pfam* ID: Rpr2 (Description: PF04032, RNase PRpr2/Rpp21/SNM1 subunit domain) em roxo.



		790	800	810	820	830	840	850	860	870	880	890	900	910
Af X3 CDS		AAGGCAGAGGCGGCGAGTGAAGGCGCGCTGCTGCAGCGGATGATGTCTAAACCCGGAATACGCCGCGCTTCCAAATTC												
Af X3 AA		K-A-E-G-G-S-E-A-A-A-A-A-D-D-V-L-T-P-N-T-P-A-F-Q-I-P-N-I-Q-S-A-I-D-A-S-T-Q-M-L-C-Q-V-F												
Ao X3 CDS		AAGGCAGAGGCGGCGAGTGAAGGCGCGCTGCTGCAGCGGATGATGTCTAAACCCGGAATACGCCGCGCTTCCAAATTC												
Ao X3 AA		K-A-E-G-G-S-E-A-P-A-A-A-D-D-D-V-L-R-P-N-T-P-A-F-Q-I-P-N-I-Q-S-A-I-D-A-S-T-Q-M-L-C-Q-V-F												
Af X2 CDS		AAGGCAGAGGCGGCGAGTGAAGGCGCGCTGCTGCAGCGGATGATGTCTAAACCCGGAATACGCCGCGCTTCCAAATTC												
Af X2 AA		K-A-E-G-G-S-E-A-A-A-A-A-D-D-V-L-T-P-N-T-P-A-F-Q-I-P-N-I-Q-S-A-I-D-A-S-T-Q-M-L-C-Q-V-F												
Ao X2 CDS		AAGGCAGAGGCGGCGAGTGAAGGCGCGCTGCTGCAGCGGATGATGTCTAAACCCGGAATACGCCGCGCTTCCAAATTC												
Ao X2 AA		K-A-E-G-G-S-E-A-P-A-A-A-D-D-D-V-L-R-P-N-T-P-A-F-Q-I-P-N-I-Q-S-A-I-D-A-S-T-Q-M-L-C-Q-V-F												
Af X1 CDS		AAGGCAGAGGCGGCGAGTGAAGGCGCGCTGCTGCAGCGGATGATGTCTAAACCCGGAATACGCCGCGCTTCCAAATTC												
Af X1 AA		K-A-E-G-G-S-E-A-A-A-A-A-D-D-V-L-T-P-N-T-P-A-F-Q-I-P-N-I-Q-S-A-I-D-A-S-T-Q-M-L-C-Q-V-F												
Ao X1 CDS		AAGGCAGAGGCGGCGAGTGAAGGCGCGCTGCTGCAGCGGATGATGTCTAAACCCGGAATACGCCGCGCTTCCAAATTC												
Ao X1 AA		K-A-E-G-G-S-E-A-P-A-A-A-D-D-D-V-L-R-P-N-T-P-A-F-Q-I-P-N-I-Q-S-A-I-D-A-S-T-Q-M-L-C-Q-V-F												
		920	930	940	950	960	970	980	990	1000	1010	1020	1030	1040
Af X3 CDS		AACAAACCAATGCTGATGCAAAATTCGGTTTCAGCTCCCAAAATTC												
Af X3 AA		Q-Q-P-N-A-D-A-N-S-G-F-Q-I-P-N-I-Q-S-A-L-D-A-G-T-Q-I-L-L-G-Q-A-L-Q-A-F-Q-Q-T-N-A-D-G-S-S-G-G-F-C												
Ao X3 CDS		AACAAACCAATGCTGATGCGGAATTCGGTTTCAGCTCCCAAAATTC												
Ao X3 AA		Q-Q-P-N-A-D-A-N-S-G-F-R-L-P-N-I-Q-S-A-L-D-A-G-T-Q-I-L-L-G-Q-A-L-Q-A-F-Q-Q-T-N-A-D-G-S-S-G-G-F-C												
Af X2 CDS		AACAAACCAATGCTGATGCAAAATTCGGTTTCAGCTCCCAAAATTC												
Af X2 AA		Q-Q-P-N-A-D-A-N-S-G-F-Q-I-P-N-I-Q-S-A-L-D-A-G-T-Q-I-L-L-G-Q-A-L-Q-A-F-Q-Q-T-N-A-D-G-S-S-G-G-F-C												
Ao X2 CDS		AACAAACCAATGCTGATGCGGAATTCGGTTTCAGCTCCCAAAATTC												
Ao X2 AA		Q-Q-P-N-A-D-A-N-S-G-F-R-L-P-N-I-Q-S-A-L-D-A-G-T-Q-I-L-L-G-Q-A-L-Q-A-F-Q-Q-T-N-A-D-G-S-S-G-G-F-C												
Af X1 CDS		AACAAACCAATGCTGATGCAAAATTCGGTTTCAGCTCCCAAAATTC												
Af X1 AA		Q-Q-P-N-A-D-A-N-S-G-F-Q-I-P-N-I-Q-S-A-L-D-A-G-T-Q-I-L-L-G-Q-A-L-Q-A-F-Q-Q-T-N-A-D-G-S-S-G-G-F-C												
Ao X1 CDS		AACAAACCAATGCTGATGCGGAATTCGGTTTCAGCTCCCAAAATTC												
Ao X1 AA		Q-Q-P-N-A-D-A-N-S-G-F-R-L-P-N-I-Q-S-A-L-D-A-G-T-Q-I-L-L-G-Q-A-L-Q-A-F-Q-Q-T-N-A-D-G-S-S-G-G-F-C												
		1050	1060	1070	1080	1090	1100	1110	1120	1130	1140	1150	1160	1170
Af X3 CDS		ATTTCCCAACATTCAGTGTGTCATAGATTCGGCTCCAAATGCTCAGCGCAAGTGGGAATTTTCAATTC												
Af X3 AA		F-P-P-N-I-Q-S-V-I-D-S-C-S-Q-M-L-S-Q-S-C-N-L-F-F-N-L-R-P-L-Q-S-L-D-P-M-S-D-E-P-A-E-P-A-A												
Ao X3 CDS		ATTTCCCAACATTCAGTGTGTCATAGATTCGGCTCCAAATGCTCAGCGCAAGTGGGAATTTTCAATTC												
Ao X3 AA		F-P-P-N-I-Q-S-V-I-D-S-C-S-Q-M-L-S-Q-S-C-N-L-F-F-N-L-R-P-L-Q-S-L-D-P-M-S-D-E-P-A-E-P-A-A												
Af X2 CDS		ATTTCCCAACATTCAGTGTGTCATAGATTCGGCTCCAAATGCTCAGCGCAAGTGGGAATTTTCAATTC												
Af X2 AA		F-P-P-N-I-Q-S-V-I-D-S-C-S-Q-M-L-S-Q-S-C-N-L-F-F-N-L-R-P-L-Q-S-L-D-P-M-S-D-E-P-A-E-P-A-A												
Ao X2 CDS		ATTTCCCAACATTCAGTGTGTCATAGATTCGGCTCCAAATGCTCAGCGCAAGTGGGAATTTTCAATTC												
Ao X2 AA		F-P-P-N-I-Q-S-V-I-D-S-C-S-Q-M-L-S-Q-S-C-N-L-F-F-N-L-R-P-L-Q-S-L-D-P-M-S-D-E-P-A-E-P-A-A												
Af X1 CDS		ATTTCCCAACATTCAGTGTGTCATAGATTCGGCTCCAAATGCTCAGCGCAAGTGGGAATTTTCAATTC												
Af X1 AA		F-P-P-N-I-Q-S-V-I-D-S-C-S-Q-M-L-S-Q-S-C-N-L-F-F-N-L-R-P-L-Q-S-L-D-P-M-S-D-E-P-A-E-P-A-A												
Ao X1 CDS		ATTTCCCAACATTCAGTGTGTCATAGATTCGGCTCCAAATGCTCAGCGCAAGTGGGAATTTTCAATTC												
Ao X1 AA		F-P-P-N-I-Q-S-V-I-D-S-C-S-Q-M-L-S-Q-S-C-N-L-F-F-N-L-R-P-L-Q-S-L-D-P-M-S-D-E-P-A-E-P-A-A												
		1180	1190	1200	1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
Af X3 CDS		TCCGACGATTCGAAGTCCCGCTGAAGCCGATGTAATCCAACTACTCAAAATGAACAAAAAGCAAGCAGCCAGGTATGCAATGAAT												
Af X3 AA		S-Q-I-S-E-V-R-V-K-P-D-V-N-P-T-T-Q-N-E-R-K-M-Q-T-A-A-C-M-D-E-L-E-K-M-L-S-A-M-Q-K-A-F-I-I-B												
Ao X3 CDS		TCCGACGATTCGAAGTCCCGCTGAAGCCGATGTAATCCAACTACTCAAAATGAACAAAAAGCAAGCAGCCAGGTATGCAATGAAT												
Ao X3 AA		S-Q-I-S-E-V-R-V-K-P-D-V-N-P-T-T-Q-N-E-R-K-M-Q-T-A-A-C-M-D-E-L-E-K-M-L-S-A-M-Q-K-A-F-I-I-B												
Af X2 CDS		TCCGACGATTCGAAGTCCCGCTGAAGCCGATGTAATCCAACTACTCAAAATGAACAAAAAGCAAGCAGCCAGGTATGCAATGAAT												
Af X2 AA		S-Q-I-S-E-V-R-V-K-P-D-V-N-P-T-T-Q-N-E-R-K-M-Q-T-A-A-C-M-D-E-L-E-K-M-L-S-A-M-Q-K-A-F-I-I-B												
Ao X2 CDS		TCCGACGATTCGAAGTCCCGCTGAAGCCGATGTAATCCAACTACTCAAAATGAACAAAAAGCAAGCAGCCAGGTATGCAATGAAT												
Ao X2 AA		S-Q-I-S-E-V-R-V-K-P-D-V-N-P-T-T-Q-N-E-R-K-M-Q-T-A-A-C-M-D-E-L-E-K-M-L-S-A-M-Q-K-A-F-I-I-B												
Af X1 CDS		TCCGACGATTCGAAGTCCCGCTGAAGCCGATGTAATCCAACTACTCAAAATGAACAAAAAGCAAGCAGCCAGGTATGCAATGAAT												
Af X1 AA		S-Q-I-S-E-V-R-V-K-P-D-V-N-P-T-T-Q-N-E-R-K-M-Q-T-A-A-C-M-D-E-L-E-K-M-L-S-A-M-Q-K-A-F-I-I-B												
Ao X1 CDS		TCCGACGATTCGAAGTCCCGCTGAAGCCGATGTAATCCAACTACTCAAAATGAACAAAAAGCAAGCAGCCAGGTATGCAATGAAT												
Ao X1 AA		S-Q-I-S-E-V-R-V-K-P-D-V-N-P-T-T-Q-N-E-R-K-M-Q-T-A-A-C-M-D-E-L-E-K-M-L-S-A-M-Q-K-A-F-I-I-B												
		1310	1320	1330	1340	1350	1360	1370	1380	1390	1400	1410	1420	1430
Af X3 CDS		AGAAATACCATATATGCTCCGATGCCATGCCACTCAGTGAAGGAGCAGTACTGAAATGAACACCATGCAAAATTC												
Af X3 AA		K-K-L-P-I-L-W-F-R-M-P-S-D-S-P-E-R-D-D-D-L-E-L-T-H-H-A-K-P-K-N-K-N-M-D-E-Q-L-E-S-Q-L-Q-Q-A												
Ao X3 CDS		AGAAATACCATATATGCTCCGATGCCATGCCACTCAGTGAAGGAGCAGTACTGAAATGAACACCATGCAAAATTC												
Ao X3 AA		K-K-L-P-I-L-W-F-R-M-P-S-D-S-P-E-R-D-D-D-L-E-L-T-H-H-A-K-P-K-N-K-N-M-D-E-Q-L-E-S-Q-L-Q-Q-A												
Af X2 CDS		AGAAATACCATATATGCTCCGATGCCATGCCACTCAGTGAAGGAGCAGTACTGAAATGAACACCATGCAAAATTC												
Af X2 AA		K-K-L-P-I-L-W-F-R-M-P-S-D-S-P-E-R-D-D-D-L-E-L-T-H-H-A-K-P-K-N-K-N-M-D-E-Q-L-E-S-Q-L-Q-Q-A												
Ao X2 CDS		AGAAATACCATATATGCTCCGATGCCATGCCACTCAGTGAAGGAGCAGTACTGAAATGAACACCATGCAAAATTC												
Ao X2 AA		K-K-L-P-I-L-W-F-R-M-P-S-D-S-P-E-R-D-D-D-L-E-L-T-H-H-A-K-P-K-N-K-N-M-D-E-Q-L-E-S-Q-L-Q-Q-A												
Af X1 CDS		AGAAATACCATATATGCTCCGATGCCATGCCACTCAGTGAAGGAGCAGTACTGAAATGAACACCATGCAAAATTC												
Af X1 AA		K-K-L-P-I-L-W-F-R-M-P-S-D-S-P-E-R-D-D-D-L-E-L-T-H-H-A-K-P-K-N-K-N-M-D-E-Q-L-E-S-Q-L-Q-Q-A												
Ao X1 CDS		AGAAATACCATATATGCTCCGATGCCATGCCACTCAGTGAAGGAGCAGTACTGAAATGAACACCATGCAAAATTC												
Ao X1 AA		K-K-L-P-I-L-W-F-R-M-P-S-D-S-P-E-R-D-D-D-L-E-L-T-H-H-A-K-P-K-N-K-N-M-D-E-Q-L-E-S-Q-L-Q-Q-A												
		1440	1450	1460	1470	1480	1490	1500	1510	1520	1530	1540	1550	1560
Af X3 CDS		TTTTCAGCGTCAGGTTATAGCTGAATTAAGTTCTCCAGGATATGAGCGTAAGCGTAAGGAATTCGGGACGCTCAAGGCGAGTCCACG												
Af X3 AA		F-E-R-Q-V-I-A-E-L-K-F-L-Q-D-I-E-R-K-A-K-E-M-R-A-A-S-K-A-C-A-T-D-K-T-S-K-D-A-T-N-H-H-A												
Ao X3 CDS		TTTTCAGCGTCAGGTTATAGCTGAATTAAGTTCTCCAGGATATGAGCGTAAGCGTAAGGAATTCGGGACGCTCAAGGCGAGTCCAC												
Ao X3 AA		F-E-R-Q-V-I-A-E-L-K-F-L-Q-D-I-E-R-K-A-K-E-M-R-A-A-S-K-A-C-A-T-D-K-T-S-K-D-A-T-N-H-H-A												
Af X2 CDS		TTTTCAGCGTCAGGTTATAGCTGAATTAAGTTCTCCAGGATATGAGCGTAAGCGTAAGGAATTCGGGACGCTCAAGGCGAGTCCAC												
Af X2 AA		F-E-R-Q-V-I-A-E-L-K-F-L-Q-D-I-E-R-K-A-K-E-M-R-A-A-S-K-A-C-A-T-D-K-T-S-K-D-A-T-N-H-H-A												
Ao X2 CDS		TTTTCAGCGTCAGGTTATAGCTGAATTAAGTTCTCCAGGATATGAGCGTAAGCGTAAGGAATTCGGGACGCTCAAGGCGAGTCCAC												
Ao X2 AA		F-E-R-Q-V-I-A-E-L-K-F-L-Q-D-I-E-R-K-A-K-E-M-R-A-A-S-K-A-C-A-T-D-K-T-S-K-D-A-T-N-H-H-A												
Af X1 CDS		TTTTCAGCGTCAGGTTATAGCTGAATTAAGTTCTCCAGGATATGAGCGTAAGCGTAAGGAATTCGGGACGCTCAAGGCGAGTCCAC												
Af X1 AA		F-E-R-Q-V-I-A-E-L-K-F-L-Q-D-I-E-R-K-A-K-E-M-R-A-A-S-K-A-C-A-T-D-K-T-S-K-D-A-T-N-H-H-A												
Ao X1 CDS		TTTTCAGCGTCAGGTTATAGCTGAATTAAGTTCTCCAGGATATGAGCGTAAGCGTAAGGAATTCGGGACGCTCAAGGCGAGTCCAC												
Ao X1 AA		F-E-R-Q-V-I-A-E-L-K-F-L-Q-D-I-E-R-K-A-K-E-M-R-A-A-S-K-A-C-A-T-D-K-T-S-K-D-A-T-N-H-H-A												
		1570	1580	1590	1600	1610	1620	1630	1640	1650	1660	1670	1680	1690
Af X3 CDS		AGCAATCTCGATGAAGTACACCGGCTCAATGAGTAAACACAGCTGATATCGAGAAGCTCTCGAAGCATGATATACGTAAGAAACCTGTTACATATGGCCGCAATCAAAATCGGAAGCAAGTGAAGAAGACCG												
Af X3 AA		S-I-L-S-K-I-P-V-H-K-I-T-R-A-D-I-E-K-A-L-N-L-D-Y-V-K-K-L-L-H-M-A-I-I-T-N-R-K-A-S-E-K-T												
Ao X3 CDS		AGCAATCTCGATGAAGTACACCGGCTCAATGAGTAAACACAGCTGATATCGAGAAGCTCTCGAAGCATGATATACGTAAGAAACCTGTTACATATGGCCGCAATCAAAATCGGAAGCAAGTGAAGAAGACCG												
Ao X3 AA		S-I-L-S-K-I-P-V-H-K-I-T-R-A-D-I-E-K-A-L-N-L-D-Y-V-K-K-L-L-H-M-A-I-I-T-N-R-K-A-S-E-K-T												
Af X2 CDS		AGCAATCTCGATGAAGTACACCGGCTCAATGAGTAAACACAGCTGATATCGAGAAGCTCTCGAAGCATGATATACGTAAGAAACCTGTTACATATGGCCGCAATCAAAATCGGAAGCAAGTGAAGAAGACCG												
Af X2 AA		S-I-L-S-K-I-P-V-H-K-I-T-R-A-D-I-E-K-A-L-N-L-D-Y-V-K-K-L-L-H-M-A-I-I-T-N-R-K-A-S-E-K-T												
Ao X2 CDS		AGCAATCTCGATGAAGTACACCGGCTCAATGAGTAAACACAGCTGATATCGAGAAGCTCTCGAAGCATGATATACGTAAGAAACCTGTTACATATGGCCGCAATCAAAATCGGAAGCAAGTGAAGAAGACCG												
Ao X2 AA		S-I-L-S-K-I-P-V-H-K-I-T-R-A-D-I-E-K-A-L-N-L-D-Y-V-K-K-L-L-H-M-A-I-I-T-N-R-K-A-S-E-K-T												
Af X1 CDS		AGCAATCTCGATGAAGTACACCGGCTCAATGAGTAAACACAGCTGATATCGAGAAGCTCTCGAAGCATGATATACGTAAGAAACCTGTTACATATGGCCGCAATCAAAATCGGAAGCAAGTGAAGAAGACCG												
Af X1 AA		S-I-L-S-K-I-P-V-H-K-I-T-R-A-D-I-E-K-A-L-N-L-D-Y-V-K-K-L-L-H-M-A-I-I-T-N-R-K-A-S-E-K-T												
Ao X1 CDS		AGCAATCTCGATGAAGTACACCGGCTCAATGAGTAAACACAGCTGATATCGAGAAGCTCTCGAAGCATGATATACGTAAGAAACCTGTTACATATGGCCGCAATCAAAATCGGAAGCAAGTGAAGAAGACCG												
Ao X1 AA		S-I-L-S-K-I-P-V-H-K-I-T-R-A-D-I-E-K-A-L-N-L-D-Y-V-K-K-L-L-H-M-A-I-I-T-N-R-K-A-S-E-K-T												

1700 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800 1810 1820

Af X3 CDS TCAATGGCGCTCCCTAGTAAGCGCCAAACCATGACTTCTAAATCCCTCAATACCAACGAGCAAAATGTCAGAGAGCAAAATATGCGGTTGCTAGCCTATGCATATCGAATGGCGAATGCTAAATGCCATACA

Af X3 AA L-N-C-A-P-S-K-R-Q-T-M-T-S-K-S-S-I-T-P-R-Q-M-S-R-D-E-I-M-R-L-L-A-Y-A-Y-R-M-A-N-A-N-G-I-I-Q

Ao X3 CDS TCAATGGCGCTCCCTAGTAAGCGCCAAACCATGACTTCTAAATCCCTCAATACCAACGAGCAAAATGTCAGAGAGCAAAATATGCGGTTGCTAGCCTATGCATATCGAATGGCGAATGCTAAATGCCATACA

Ao X3 AA L-N-C-A-P-S-K-R-Q-T-M-T-S-K-S-S-I-T-P-R-Q-M-S-R-D-E-I-M-R-L-L-A-Y-A-Y-R-M-A-N-A-N-G-I-I-Q

Af X2 CDS TCAATGGCGCTCCCTAGTAAGCGCCAAACCATGACTTCTAAATCCCTCAATACCAACGAGCAAAATGTCAGAGAGCAAAATATGCGGTTGCTAGCCTATGCATATCGAATGGCGAATGCTAAATGCCATACA

Af X2 AA L-N-C-A-P-S-K-R-Q-T-M-T-S-K-S-S-I-T-P-R-Q-M-S-R-D-E-I-M-R-L-L-A-Y-A-Y-R-M-A-N-A-N-G-I-I-Q

Ao X2 CDS TCAATGGCGCTCCCTAGTAAGCGCCAAACCATGACTTCTAAATCCCTCAATACCAACGAGCAAAATGTCAGAGAGCAAAATATGCGGTTGCTAGCCTATGCATATCGAATGGCGAATGCTAAATGCCATACA

Ao X2 AA L-N-C-A-P-S-K-R-Q-T-M-T-S-K-S-S-I-T-P-R-Q-M-S-R-D-E-I-M-R-L-L-A-Y-A-Y-R-M-A-N-A-N-G-I-I-Q

Af X1 CDS TCAATGGCGCTCCCTAGTAAGCGCCAAACCATGACTTCTAAATCCCTCAATACCAACGAGCAAAATGTCAGAGAGCAAAATATGCGGTTGCTAGCCTATGCATATCGAATGGCGAATGCTAAATGCCATACA

Af X1 AA L-N-C-A-P-S-K-R-Q-T-M-T-S-K-S-S-I-T-P-R-Q-M-S-R-D-E-I-M-R-L-L-A-Y-A-Y-R-M-A-N-A-N-G-I-I-Q

Ao X1 CDS TCAATGGCGCTCCCTAGTAAGCGCCAAACCATGACTTCTAAATCCCTCAATACCAACGAGCAAAATGTCAGAGAGCAAAATATGCGGTTGCTAGCCTATGCATATCGAATGGCGAATGCTAAATGCCATACA

Ao X1 AA L-N-C-A-P-S-K-R-Q-T-M-T-S-K-S-S-I-T-P-R-Q-M-S-R-D-E-I-M-R-L-L-A-Y-A-Y-R-M-A-N-A-N-G-I-I-Q

1830 1840 1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950

Af X3 CDS GATGATGGAAGTTTCCGATTCCTAAGAGCTAAACAGGCCACCTCAGTCGGATTCGAGGAGCTCTGCAAAACAGTTAAATGACGAGAAATGGTTAAGAAATTTCAAAGCGGAGCAATGGGAAGCGCAAGCG

Af X3 AA -M-M-E-V-S-D-S-K-K-L-N-K-P-P-Q-D-S-Q-G-S-A-K-Q-L-N-D-E-N-W-L-R-N-T-S-K-A-R-Q-W-E-G-E-R

Ao X3 CDS GATGATGGAAGTTTCCGATTCCTAAGAGCTAAACAGGCCACCTCAGTCGGATTCGAGGAGCTCTGCAAAACAGTTAAATGACGAGAAATGGTTAAGAAATTTCAAAGCGGAGCAATGGGAAGCGCAAGCG

Ao X3 AA -M-M-E-V-S-D-S-K-K-L-N-K-P-P-Q-D-S-Q-G-S-A-K-Q-L-N-D-E-N-W-L-R-N-T-S-K-A-R-Q-W-E-G-E-R

Af X2 CDS GATGATGGAAGTTTCCGATTCCTAAGAGCTAAACAGGCCACCTCAGTCGGATTCGAGGAGCTCTGCAAAACAGTTAAATGACGAGAAATGGTTAAGAAATTTCAAAGCGGAGCAATGGGAAGCGCAAGCG

Af X2 AA -M-M-E-V-S-D-S-K-K-L-N-K-P-P-Q-D-S-Q-G-S-A-K-Q-L-N-D-E-N-W-L-R-N-T-S-K-A-R-Q-W-E-G-E-R

Ao X2 CDS GATGATGGAAGTTTCCGATTCCTAAGAGCTAAACAGGCCACCTCAGTCGGATTCGAGGAGCTCTGCAAAACAGTTAAATGACGAGAAATGGTTAAGAAATTTCAAAGCGGAGCAATGGGAAGCGCAAGCG

Ao X2 AA -M-M-E-V-S-D-S-K-K-L-N-K-P-P-Q-D-S-Q-G-S-A-K-Q-L-N-D-E-N-W-L-R-N-T-S-K-A-R-Q-W-E-G-E-R

Af X1 CDS GATGATGGAAGTTTCCGATTCCTAAGAGCTAAACAGGCCACCTCAGTCGGATTCGAGGAGCTCTGCAAAACAGTTAAATGACGAGAAATGGTTAAGAAATTTCAAAGCGGAGCAATGGGAAGCGCAAGCG

Af X1 AA -M-M-E-V-S-D-S-K-K-L-N-K-P-P-Q-D-S-Q-G-S-A-K-Q-L-N-D-E-N-W-L-R-N-T-S-K-A-R-Q-W-E-G-E-R

Ao X1 CDS GATGATGGAAGTTTCCGATTCCTAAGAGCTAAACAGGCCACCTCAGTCGGATTCGAGGAGCTCTGCAAAACAGTTAAATGACGAGAAATGGTTAAGAAATTTCAAAGCGGAGCAATGGGAAGCGCAAGCG

Ao X1 AA -M-M-E-V-S-D-S-K-K-L-N-K-P-P-Q-D-S-Q-G-S-A-K-Q-L-N-D-E-N-W-L-R-N-T-S-K-A-R-Q-W-E-G-E-R

1960 1970 1980 1990 2000 2010 2020 2030 2040 2050 2060 2070 2080

Af X3 CDS ATGCCAGATGCGCTCGTATGCAAAATTTTGAACAGATTCACCAAGGATGCTATGATGATAGGCAAAATCCCGGAAAAAAGCAGACAAATGGGACGCTCAATCGTTGGATGAACAAACACCACCAATTAATGA

Af X3 AA -M-Q-M-P-R-M-Q-N-F-E-Q-N-P-T-R-M-S-M-D-R-Q-M-P-E-K-T-T-A-M-C-R-Q-S-L-D-E-Q-T-T-P-I-N-E

Ao X3 CDS ATGCCAGATGCGCTCGTATGCAAAATTTTGAACAGATTCACCAAGGATGCTATGATGATAGGCAAAATCCCGGAAAAAAGCAGACAAATGGGACGCTCAATCGTTGGATGAACAAACACCACCAATTAATGA

Ao X3 AA -M-Q-M-P-R-M-Q-N-F-E-Q-N-P-T-R-M-S-M-D-R-Q-M-P-E-K-T-T-A-M-C-R-Q-S-L-D-E-Q-T-T-P-I-N-E

Af X2 CDS ATGCCAGATGCGCTCGTATGCAAAATTTTGAACAGATTCACCAAGGATGCTATGATGATAGGCAAAATCCCGGAAAAAAGCAGACAAATGGGACGCTCAATCGTTGGATGAACAAACACCACCAATTAATGA

Af X2 AA -M-Q-M-P-R-M-Q-N-F-E-Q-N-P-T-R-M-S-M-D-R-Q-M-P-E-K-T-T-A-M-C-R-Q-S-L-D-E-Q-T-T-P-I-N-E

Ao X2 CDS ATGCCAGATGCGCTCGTATGCAAAATTTTGAACAGATTCACCAAGGATGCTATGATGATAGGCAAAATCCCGGAAAAAAGCAGACAAATGGGACGCTCAATCGTTGGATGAACAAACACCACCAATTAATGA

Ao X2 AA -M-Q-M-P-R-M-Q-N-F-E-Q-N-P-T-R-M-S-M-D-R-Q-M-P-E-K-T-T-A-M-C-R-Q-S-L-D-E-Q-T-T-P-I-N-E

Af X1 CDS ATGCCAGATGCGCTCGTATGCAAAATTTTGAACAGATTCACCAAGGATGCTATGATGATAGGCAAAATCCCGGAAAAAAGCAGACAAATGGGACGCTCAATCGTTGGATGAACAAACACCACCAATTAATGA

Af X1 AA -M-Q-M-P-R-M-Q-N-F-E-Q-N-P-T-R-M-S-M-D-R-Q-M-P-E-K-T-T-A-M-C-R-Q-S-L-D-E-Q-T-T-P-I-N-E

Ao X1 CDS ATGCCAGATGCGCTCGTATGCAAAATTTTGAACAGATTCACCAAGGATGCTATGATGATAGGCAAAATCCCGGAAAAAAGCAGACAAATGGGACGCTCAATCGTTGGATGAACAAACACCACCAATTAATGA

Ao X1 AA -M-Q-M-P-R-M-Q-N-F-E-Q-N-P-T-R-M-S-M-D-R-Q-M-P-E-K-T-T-A-M-C-R-Q-S-L-D-E-Q-T-T-P-I-N-E

2090 2100 2110 2120 2130 2140 2150 2160 2170 2180 2190 2200 2210

Af X3 CDS ATGCCAATGCCAAACAAATATGCTGATGGAACACAAACGCCAGCAATGCCAAATGCCAGAGATGCACTGCGCTTTTACCGCAGATCAACGGCAATGGATGACCGCAATATGGGATGCGCATGG

Af X3 AA N-P-M-Q-T-K-T-I-M-L-M-E-Q-Q-R-P-A-M-K-Q-M-Q-S-P-F-T-A-D-Q-R-Q-W-I-D-P-N-M-C-M-R-S-G

Ao X3 CDS ATGCCAATGCCAAACAAATATGCTGATGGAACACAAACGCCAGCAATGCCAAATGCCAGAGATGCACTGCGCTTTTACCGCAGATCAACGGCAATGGATGACCGCAATATGGGATGCGCATGG

Ao X3 AA N-P-M-Q-T-K-T-I-M-L-M-E-Q-Q-R-P-A-M-K-Q-M-Q-S-P-F-T-A-D-Q-R-Q-W-I-D-P-N-M-C-M-R-S-G

Af X2 CDS ATGCCAATGCCAAACAAATATGCTGATGGAACACAAACGCCAGCAATGCCAAATGCCAGAGATGCACTGCGCTTTTACCGCAGATCAACGGCAATGGATGACCGCAATATGGGATGCGCATGG

Af X2 AA N-P-M-Q-T-K-T-I-M-L-M-E-Q-Q-R-P-A-M-K-Q-M-Q-S-P-F-T-A-D-Q-R-Q-W-I-D-P-N-M-C-M-R-S-G

Ao X2 CDS ATGCCAATGCCAAACAAATATGCTGATGGAACACAAACGCCAGCAATGCCAAATGCCAGAGATGCACTGCGCTTTTACCGCAGATCAACGGCAATGGATGACCGCAATATGGGATGCGCATGG

Ao X2 AA N-P-M-Q-T-K-T-I-M-L-M-E-Q-Q-R-P-A-M-K-Q-M-Q-S-P-F-T-A-D-Q-R-Q-W-I-D-P-N-M-C-M-R-S-G

Af X1 CDS ATGCCAATGCCAAACAAATATGCTGATGGAACACAAACGCCAGCAATGCCAAATGCCAGAGATGCACTGCGCTTTTACCGCAGATCAACGGCAATGGATGACCGCAATATGGGATGCGCATGG

Af X1 AA N-P-M-Q-T-K-T-I-M-L-M-E-Q-Q-R-P-A-M-K-Q-M-Q-S-P-F-T-A-D-Q-R-Q-W-I-D-P-N-M-C-M-R-S-G

Ao X1 CDS ATGCCAATGCCAAACAAATATGCTGATGGAACACAAACGCCAGCAATGCCAAATGCCAGAGATGCACTGCGCTTTTACCGCAGATCAACGGCAATGGATGACCGCAATATGGGATGCGCATGG

Ao X1 AA N-P-M-Q-T-K-T-I-M-L-M-E-Q-Q-R-P-A-M-K-Q-M-Q-S-P-F-T-A-D-Q-R-Q-W-I-D-P-N-M-C-M-R-S-G

2220 2230 2240 2250 2260 2270 2280 2290 2300 2310 2320 2330 2340

Af X3 CDS GATCCAGAGAAATACCCAAATGCCCTGTGCAAAAGCCAGTGGATGACACACAGTCACAGAGATTCCTCAAGCGGAAATATCTCAACAGCCACCAGCTAATACAGACAAATCGCTACCTGCAATGCCGATGTTA

Af X3 AA -M-Q-E-N-T-Q-M-P-V-Q-R-Q-T-W-I-D-T-Q-S-Q-R-F-P-S-E-N-T-Q-Q-P-P-A-N-T-Q-Q-W-I-D-P-N-M-C-M-R-S-G

Ao X3 CDS GATCCAGAGAAATACCCAAATGCCCTGTGCAAAAGCCAGTGGATGACACACAGTCACAGAGATTCCTCAAGCGGAAATATCTCAACAGCCACCAGCTAATACAGACAAATCGCTACCTGCAATGCCGATGTTA

Ao X3 AA -M-Q-E-N-T-Q-M-P-V-Q-R-Q-T-W-I-D-T-Q-S-Q-R-F-P-S-E-N-T-Q-Q-P-P-A-N-T-Q-Q-W-I-D-P-N-M-C-M-R-S-G

Af X2 CDS GATCCAGAGAAATACCCAAATGCCCTGTGCAAAAGCCAGTGGATGACACACAGTCACAGAGATTCCTCAAGCGGAAATATCTCAACAGCCACCAGCTAATACAGACAAATCGCTACCTGCAATGCCGATGTTA

Af X2 AA -M-Q-E-N-T-Q-M-P-V-Q-R-Q-T-W-I-D-T-Q-S-Q-R-F-P-S-E-N-T-Q-Q-P-P-A-N-T-Q-Q-W-I-D-P-N-M-C-M-R-S-G

Ao X2 CDS GATCCAGAGAAATACCCAAATGCCCTGTGCAAAAGCCAGTGGATGACACACAGTCACAGAGATTCCTCAAGCGGAAATATCTCAACAGCCACCAGCTAATACAGACAAATCGCTACCTGCAATGCCGATGTTA

Ao X2 AA -M-Q-E-N-T-Q-M-P-V-Q-R-Q-T-W-I-D-T-Q-S-Q-R-F-P-S-E-N-T-Q-Q-P-P-A-N-T-Q-Q-W-I-D-P-N-M-C-M-R-S-G

Af X1 CDS GATCCAGAGAAATACCCAAATGCCCTGTGCAAAAGCCAGTGGATGACACACAGTCACAGAGATTCCTCAAGCGGAAATATCTCAACAGCCACCAGCTAATACAGACAAATCGCTACCTGCAATGCCGATGTTA

Af X1 AA -M-Q-E-N-T-Q-M-P-V-Q-R-Q-T-W-I-D-T-Q-S-Q-R-F-P-S-E-N-T-Q-Q-P-P-A-N-T-Q-Q-W-I-D-P-N-M-C-M-R-S-G

Ao X1 CDS GATCCAGAGAAATACCCAAATGCCCTGTGCAAAAGCCAGTGGATGACACACAGTCACAGAGATTCCTCAAGCGGAAATATCTCAACAGCCACCAGCTAATACAGACAAATCGCTACCTGCAATGCCGATGTTA

Ao X1 AA -M-Q-E-N-T-Q-M-P-V-Q-R-Q-T-W-I-D-T-Q-S-Q-R-F-P-S-E-N-T-Q-Q-P-P-A-N-T-Q-Q-W-I-D-P-N-M-C-M-R-S-G

2350 2360 2370 2380 2390 2400 2410 2420 2430 2440 2450 2460 2470

Af X3 CDS ATGCCAATGCCAGAGCGGCTATGATGATGGAAGAGCCCTCGGATGCCAATGCAACTACCGCAGCGGAGGACAGCAAGGAGGTGACGGGATGATGTTACCAACAGATGCCAATGCGAGATGCCAATGCG

Af X3 AA -M-A-M-Q-Q-P-P-M-M-E-E-P-R-M-A-M-Q-L-P-Q-A-D-T-Q-C-G-D-C-M-M-L-Q-Q-M-H-P-E-M-Q-R-M-M

Ao X3 CDS ATGCCAATGCCAGAGCGGCTATGATGATGGAAGAGCCCTCGGATGCCAATGCAACTACCGCAGCGGAGGACAGCAAGGAGGTGACGGGTTGATGTTACCAACAGATGCCAATGCGAGATGCCAATGCG

Ao X3 AA -M-A-M-Q-Q-P-P-M-M-E-E-P-R-M-A-M-Q-L-P-Q-A-D-T-Q-C-G-D-C-M-M-L-Q-Q-M-H-P-E-M-Q-R-M-M

Af X2 CDS ATGCCAATGCCAGAGCGGCTATGATGATGGAAGAGCCCTCGGATGCCAATGCAACTACCGCAGCGGAGGACAGCAAGGAGGTGACGGGTTGATGTTACCAACAGATGCCAATGCGAGATGCCAATGCG

Af X2 AA -M-A-M-Q-Q-P-P-M-M-E-E-P-R-M-A-M-Q-L-P-Q-A-D-T-Q-C-G-D-C-M-M-L-Q-Q-M-H-P-E-M-Q-R-M-M

Ao X2 CDS ATGCCAATGCCAGAGCGGCTATGATGATGGAAGAGCCCTCGGATGCCAATGCAACTACCGCAGCGGAGGACAGCAAGGAGGTGACGGGTTGATGTTACCAACAGATGCCAATGCGAGATGCCAATGCG

Ao X2 AA -M-A-M-Q-Q-P-P-M-M-E-E-P-R-M-A-M-Q-L-P-Q-A-D-T-Q-C-G-D-C-M-M-L-Q-Q-M-H-P-E-M-Q-R-M-M

Af X1 CDS ATGCCAATGCCAGAGCGGCTATGATGATGGAAGAGCCCTCGGATGCCAATGCAACTACCGCAGCGGAGGACAGCAAGGAGGTGACGGGTTGATGTTACCAACAGATGCCAATGCGAGATGCCAATGCG

Af X1 AA -M-A-M-Q-Q-P-P-M-M-E-E-P-R-M-A-M-Q-L-P-Q-A-D-T-Q-C-G-D-C-M-M-L-Q-Q-M-H-P-E-M-Q-R-M-M

Ao X1 CDS ATGCCAATGCCAGAGCGGCTATGATGATGGAAGAGCCCTCGGATGCCAATGCAACTACCGCAGCGGAGGACAGCAAGGAGGTGACGGGTTGATGTTACCAACAGATGCCAATGCGAGATGCCAATGCG

Ao X1 AA -M-A-M-Q-Q-P-P-M-M-E-E-P-R-M-A-M-Q-L-P-Q-A-D-T-Q-C-G-D-C-M-M-L-Q-Q-M-H-P-E-M-Q-R-M-M

2480 2490 2500 2510 2520 2530 2540 2550 2560 2570 2580 2590 2600

Af X3 CDS AGGAGCTAGATATGCTAGGCTTGGCCACTCCACAAATGCCGATTAATGAGGGAAGAGCTGATTTAAAGGCCACATGTTTCGATGATTTTGGCGAGATCCATAAAGGTAAGAGCGAAAGCGGCA

Af X3 AA Q-E-L-D-M-L-C-L-A-T-P-Q-M-P-D-N-E-G-K-A-R-F-F-K-G-H-M-F-L-D-D-F-G-E-I-H-K-C-G-K-G-K-G-K-G

Ao X3 CDS AGGAGCTAGATATGCTAGGCTTGGCCACTCCACAAATGCCGATTAATGAGGGAAGAGCTGATTTAAAGGCCACATGTTTCGATGATTTTGGCGAGATCCATAAAGGTAAGAGCGAAAGCGGCA

Ao X3 AA Q-K-L-D-M-L-C-L-A-T-P-Q-M-P-D-N-E-G-K-A-R-F-F-K-G-H-M-F-L-D-D-F-G-E-I-H-K-C-G-K-G-K-G-K-G-V

Af X2 CDS AGGAGCTAGATATGCTAGGCTTGGCCACTCCACAAATGCCGATTAATGAGGGAAGAGCTGATTTAAAGGCCACATGTTTCGATGATTTTGGCGAGATCCATAAAGGTAAGAGCGAAAGCGGCA

Af X2 AA Q-E-L-D-M-L-C-L-A-T-P-Q-M-P-D-N-E-G-K-A-R-F-F-K-G-H-M-F-L-D-D-F-G-E-I-H-K-C-G-K-G-K-G-K-G-C

Ao X2 CDS AGGAGCTAGATATGCTAGGCTTGGCCACTCCACAAATGCCGATTAATGAGGGAAGAGCTGATTTAAAGGCCACATGTTTCGATGATTTTGGCGAGATCCATAAAGGTAAGAGCGAAAGCGGCA

Ao X2 AA Q-K-L-D-M-L-C-L-A-T-P-Q-M-P-D-N-E-G-K-A-R-F-F-K-G-H-M-F-L-D-D-F-G-E-I-H-K-C-G-K-G-K-G-K-G-V

Af X1 CDS AGGAGCTAGATATGCTAGGCTTGGCCACTCCACAAATGCCGATTAATGAGGGAAGAGCTGATTTAAAGGCCACATGTTTCGATGATTTTGGCGAGATCCATAAAGGTAAGAGCGAAAGCGGCA

Af X1 AA Q-E-L-D-M-L-C-L-A-T-P-Q-M-P-D-N-E-G-K-A-R-F-F-K-G-H-M-F-L-D-D-F-G-E-I-H-K-C-G-K-G-K-G-K-G-C

Ao X1 CDS AGGAGCTAGATATGCTAGGCTTGGCCACTCCACAAATGCCGATTAATGAGGGAAGAGCTGATTTAAAGGCCACATGTTTCGATGATTTTGGCGAGATCCATAAAGGTAAGAGCGAAAGCGGCA

Ao X1 AA Q-K-L-D-M-L-C-L-A-T-P-Q-M-P-D-N-E-G-K-A-R-F-F-K-G-H-M-F-L-D-D-F-G-E-I-H-K-C-G-K-G-K-G-K-G-V

76

3520 3530 3540 3550 3560 3570 3580 3590 3600 3610 3620 3630 3640

Af X1 CDS
Af X1 AA
Ao X1 CDS
Ao X1 AA
Af X2 CDS
Af X2 AA
Ao X2 CDS
Ao X2 AA
Af X3 CDS
Af X3 AA
Ao X3 CDS
Ao X3 AA

3650 3660 3670 3680 3690 3700 3710 3720 3730 3740 3750 3760 3770

Af X1 CDS
Af X1 AA
Ao X1 CDS
Ao X1 AA
Af X2 CDS
Af X2 AA
Ao X2 CDS
Ao X2 AA
Af X3 CDS
Af X3 AA
Ao X3 CDS
Ao X3 AA

3780 3790 3800 3810 3820 3830 3840 3850 3860 3870 3880 3890 3900

Af X1 CDS
Af X1 AA
Ao X1 CDS
Ao X1 AA
Af X2 CDS
Af X2 AA
Ao X2 CDS
Ao X2 AA
Af X3 CDS
Af X3 AA
Ao X3 CDS
Ao X3 AA

3910 3920 3930 3940 3950 3960 3970 3980 3990 4000 4010 4020 4030

Af X1 CDS
Af X1 AA
Ao X1 CDS
Ao X1 AA
Af X2 CDS
Af X2 AA
Ao X2 CDS
Ao X2 AA
Af X3 CDS
Af X3 AA
Ao X3 CDS
Ao X3 AA

4040 4050 4060 4070 4080 4090 4100 4110 4120 4130 4140 4150 4160

Af X1 CDS
Af X1 AA
Ao X1 CDS
Ao X1 AA
Af X2 CDS
Af X2 AA
Ao X2 CDS
Ao X2 AA
Af X3 CDS
Af X3 AA
Ao X3 CDS
Ao X3 AA

4170 4180 4190 4200 4210 4220 4230 4240 4250 4260 4270 4280 4290

Af X1 CDS
Af X1 AA
Ao X1 CDS
Ao X1 AA
Af X2 CDS
Af X2 AA
Ao X2 CDS
Ao X2 AA
Af X3 CDS
Af X3 AA
Ao X3 CDS
Ao X3 AA

4300 4310 4320 4330 4340 4350 4360 4370 4380 4390 4400 4410 4420

Af X1 CDS
Af X1 AA
Ao X1 CDS
Ao X1 AA
Af X2 CDS
Af X2 AA
Ao X2 CDS
Ao X2 AA
Af X3 CDS
Af X3 AA
Ao X3 CDS
Ao X3 AA

ACCGCACAAACAAACGAAATTCGCAAAATATTTCTTTAAATCAGAAATCAGCCACAGAGAGAAAGGGTACTACGGCAGCTATTAACGATCCTCAGCAACCGTCTTTGAATCTAAACACAGGTAATATG
-S--A-Q-Q-Q-R-N-S-A-N-I-S-F-N-Q-N-Q-P-Q-E-R-K-G-G-T-T-A-A-I-N-D-P-Q-Q-P-S-L-N-L-N-T-C-N-I-I-
AGTCACAAACAAACGAAATTCGCAAAATATTTCTTTAAATCAGAAATCAGCCAAATAGAGAGAAAGGGTACTACGGCAGCTGTTAAACGATCCTCAGCAACCGACTTTGAATCTAAAGACAGCCAAATATG
-S--A-Q-Q-Q-R-N-S-A-N-I-S-F-N-Q-N-Q-P-I-E-K-K-G-G-T-A-A-A-V-N-D-P-Q-Q-T-T-L-N-L-K-T-A-N-I-I-

CATTAAAGGAGGTTCTCAAGAGGGCTCTAAAGAGGAAAGCCAACTACTTTGGAGAAGCTGAGCTACATTTAAACAAAGCGGAAGATGCTGGCAAACTCAAGACGAGCAAACTCTAGATAAAGACGACGAG
-I-K-E-C-S-Q-E-C-S-K-E-E-S-Q-H-S-L-E-K-T-E-L-H-L-T-K-A-E-D-A-C-K-S-R-R-A-N-L-D-K-D-D-E-
CATTAAAGGAGGTTCTCAAGAGGGCTCTAAAGAGGAAAGCCAACTACTTTGGAGAAGCTGAGCTACATTTAAACAAAGCGGAAGATGCTGGCAAACTCTAGACGAGCAAACTCTAGATAAAGACGACGAG
-I-K-E-R-S-K-E-C-S-K-E-E-S-Q-H-S-S-E-K-K-T-E-L-H-L-T-K-A-E-D-A-C-K-S-R-R-A-N-L-D-K-D-D-E-

CAATTCGCAAGTGGTTACAGGAAATGGAAGCCAAAGGTTATGACTCCAAATACTTGACCAAGTCTCATCTTGGGAATACTTGAATAGCGGCAAAATGGCGCTAGCTTTGGAAAGCAATGAACAAACAGG
-Q-F-A-K-W-L-H-E-M-E-A-K-C-Y-D-S-K-Y-L-T-T-K-S-H-H-W-E-Y-L-N-S-G-K-L-A-L-A-L-E-S-N-E-Q-Q-
CAATTCGCAAGTGGTTACAGGAAATGGAAGCCAAAGGTTATGACTCCAAATACTTGACCAAGTCTCATCTTGGGAATACTTGAATAGCGGCAAAATGGCGCTAGCTTTGGAAAGCAATGAACAAACAGG
-Q-F-A-K-W-L-H-E-M-E-A-K-C-Y-D-S-K-Y-L-T-T-K-S-H-H-W-E-Y-L-N-S-G-K-L-A-L-A-L-E-S-N-E-Q-Q-
CTGATAATAAGATTCCAAAGAAACTATTAAGCACTGCTTGAAGCGGTATACGCAAGTGAATTCCTTAGTGAATGATGAAGAGCTTAGAAACACAAATACCATTAAGTCCCAATAGCACCAGGAAAGG
A-D-N-K-I-A-K-K-T-I-K-Q-S-L-K-R-Y-T-Q-V-N-S-L-V-N-D-E-E-L-R-N-T-N-T-I-T-S-N-S-T-K-C-K-A
CTGATAATAAGATTCCAAAGAAACTATTAAGCACTGCTTGAAGCGGTATACGCAAGTGAATTCCTTAGAGAAATGAAGAGCTTAGAAACACAAATGCCATTAGCTCCCAATAGCACCAGGAAAGG
A-D-N-K-I-A-K-K-T-I-K-Q-S-L-K-R-Y-T-Q-V-N-S-L-E-E-N-N-E-E-L-R-N-T-N-A-I-T-S-N-S-T-K-C-K-A

CTTTGCCAAAAAGAGCCGAGTGTAGCTACCCGAACAAAGAGCATATCCATGTAGTAAATGATGAGTTGCGCTCATTTAAATGATTACTCAGAAGAGTCACATCTAGTCCATATGTGATCGGTGGTAAA
-F-A-K-K-T-P-V-V-S-Y-P-N-K-E-H-I-H-V-V-N-D-E-E-L-R-H-L-N-D-Y-S-E-E-V-T-S-S-S-P-Y-V-M-R-C-K-
CTTTGCCAAAAAGAGCCGAGTGTAGCTACCCGAACAAAGAGCATATCCATGTGTTAAATGATGAGTTGCGCTCATTTAAATGATTACTCAGAAGAGTCACATCTAGTCCATATGTGATCGGTGGTAAA
-F-A-K-K-T-P-V-V-S-Y-P-N-K-E-H-I-H-V-V-N-D-E-E-L-R-H-L-N-D-Y-S-E-E-V-T-S-S-S-P-Y-V-M-R-C-K-

TTCAAGAATAGCAGAGTCAGAAAGCCATAGTGGTAAGGCCGCTGAACGAGCCGCAAGTAAAGTTTACTGCAATGTCGCCCACTGCGTGGCCACAACTATGCGGCGCAGACAAAGTCTAAGGTTCTCGAAC
-F-K-N-S-R-V-R-S-H-S-G-K-A-A-V-T-T-P-V-K-F-S-A-M-S-Q-Q-L-P-A-D-N-N-A-G-A-D-K-S-K-V-V-E-
TTCAAGAATAGCAGAGTCAGAAAGCCATAGGCGTAAGGCCGCTGAACGAGCCGCAAGTAAAGTTTACTGCAATGTCGCCCACTGCGTGGCCCACTATGCGGCGCAGACAAAGTCTAAGGTTCTCGAAC
-F-K-N-S-R-V-R-S-H-S-G-K-A-A-V-T-T-P-V-K-F-S-A-M-S-Q-Q-L-P-A-C-N-N-A-G-A-D-K-A-K-V-V-E-

4430 4440 4450 4460 4470 4480 4490 4500 4510 4520 4530 4540 4550

..f X1 CDS
..f X1 AA
..o X1 CDS
..o X1 AA
..f X2 CDS
..f X2 AA
..o X2 CDS
..o X2 AA
..f X3 CDS
TATTAAAGTTTACTTTGTGAGTCCCAAAACAGAGAGCGCCACAGTAA
..f X3 AA
L-L-S-L-L-L-F-E-S-Q-N-R-E-R-Q-Q
..f X3 CDS
TATTAAAGTTTACTTTACGATTTGAGTCCCAAAACAGAGAGCGCCACAGTAAACCCACACACTATCAGTAATGGTTATGCCGATTGTGGCGAGCTCGACGGAGGGTCCAGAGAAAGTGAACACGACGAG
..o X3 AA
L-L-S-L-L-L-Y-D-L-S-P-K-T-E-S-A-N-S-K-P-T-Q-L-S-V-M-V-M-P-I-V-A-T-S-T-E-G-P-E-K-V-N-T-T-S

4560 4570 4580 4590 4600 4610 4620 4630 4640 4650 4660 4670 4680

..f X1 CDS
..f X1 AA
..o X1 CDS
..o X1 AA
..f X2 CDS
TGCATGAAAGTCTAGAACTCCATTTCATTTCG
..f X2 AA
C-N-E-S-L-E-T-P-F-H-S
..o X2 CDS
TGCATGAAAGTCTAGAACTCCATTTCATTTCG
..f X3 CDS
C-N-E-S-L-E-T-P-F-H-S
..f X3 AA
TACGAGCCAGAACCAACGATTCGATCCAGACCAAGCCGAAATCAGACAAAGCCACAGCTCCACATCCGACGAGAAATAGTGTGGTCCGCGCTGA
..o X3 CDS
-R-S-R-T-K-R-L-R-S-R-R-K-P-K-S-D-K-P-T-T-A-H-A-D-E-N-S-V-G-A-R

4690 4700 4710 4720 4730 4740 4750 4760 4770 4780 4790 4800 4810

..f X1 CDS
..f X1 AA
..o X1 CDS
..o X1 AA
..f X2 CDS
AAGGATCCCGAAGAAGTAGGTTCCCTTTTCAGTTTGTGAGTCCGAAAAATTTTGAACCTTTTATGGGTTTACAAATCCGTACCCGCCCTACGAAAGCGATCCCTGGAATTATAAACCATATGACTTGCATCACC
..f X2 AA
-K-D-P-E-E-V-G-S-P-F-S-F-S-P-K-I-F-E-P-F-M-G-L-Q-S-V-P-A-Y-E-S-D-P-W-N-Y-K-P-Y-D-L-H-H
..o X2 CDS
AAGGATCCCGAAGAAGTAGGTTAGTCCGAAAAATTTTGAACCTTTTATGGGTTTACAAATCCGTACCCGCCCTACGAAAGCGATCCCTGGAATTATAAACCATATGACTTGCATCACC
..f X3 CDS
-K-D-P-E-E-V-G-S-P-K-I-F-E-P-F-M-G-L-Q-S-V-P-A-Y-E-S-D-P-W-N-Y-K-P-Y-D-L-H-H
..f X3 AA
TACGAGCCAGAACCAACGATTCGATCCAGACCAAGCCGAAATCAGACAAAGCCACAGCTCCACATCCGACGAGAAATAGTGTGGTCCGCGCTGA
..o X3 CDS
-R-S-R-T-K-R-L-R-S-R-R-K-P-K-S-D-K-P-T-T-A-H-A-D-E-N-S-V-G-A-R

4820 4830 4840 4850 4860 4870 4880 4890 4900 4910 4920 4930 4940

..f X1 CDS
..f X1 AA
..o X1 CDS
..o X1 AA
..f X2 CDS
CCCTCTATGAGGGTGGCGCAGTTATGAACCTTATATCAACAAACAAAGCCAGAAACGAAATACCCACACGCCACCAACCCACATCTACCGCCTAGTATTTCTGGAATATTTCTAAAGATACCGCGAA
..f X2 AA
P-L-Y-E-G-G-G-S-Y-E-P-Y-I-N-N-K-R-T-K-R-N-T-H-T-R-T-N-H-I-L-T-P-S-I-L-D-Y-F-L-K-I-R-A-N
..o X2 CDS
CCCTCTATGAGGGTGGCGCAGTTATGAACCTTATATCAACAAACAAAGCCAGAAACGAAATACCCACACGCCACCAACCCACATCTACCGCCTAGTATTTCTGGAATATTTCTAAAGATACCGCGAA
..f X3 CDS
P-L-Y-E-G-G-G-S-Y-E-P-Y-I-N-N-K-R-T-K-R-N-T-H-T-R-T-N-H-I-L-T-P-S-I-L-D-Y-F-L-K-I-R-A-N
..f X3 AA
TACGAGCCAGAACCAACGATTCGATCCAGACCAAGCCGAAATCAGACAAAGCCACAGCTCCACATCCGACGAGAAATAGTGTGGTCCGCGCTGA
..o X3 CDS
-R-S-R-T-K-R-L-R-S-R-R-K-P-K-S-D-K-P-T-T-A-H-A-D-E-N-S-V-G-A-R

4950 4960 4970 4980 4990 5000 5010 5020 5030 5040 5050 5060 5070

..f X1 CDS
..f X1 AA
..o X1 CDS
..o X1 AA
..f X2 CDS
TTTCGAAAGAGCTTTCCTCAGCTATACAAACCTTGCCTGATCAGCAACGCCCTCTAAATTTGACACCGCTATCCGTTAAACCGCCTGTGATACCTGAGGAACTCCGCTATCCCCCAACAGAACTGGGT
..f X2 AA
-F-E-K-S-F-P-Q-L-Y-K-T-L-R-D-Q-Q-R-A-L-N-L-T-R-V-S-V-K-P-P-P-V-I-P-E-E-V-A-Y-P-P-T-E-L-G
..o X2 CDS
TTTCGAAAGAGCTTTCCTCAGCTATACAAACCTTGCCTGATCAGCAACGCCCTCTAAATTTGACACCGCTATCCGTTAAACCGCCTGTGATACCTGAGGAACTCCGCTATCCCCCAACAGAACTGGGT
..f X3 CDS
-F-E-K-S-F-P-Q-L-Y-K-T-L-R-D-Q-Q-R-A-L-N-L-T-R-V-S-V-K-P-P-P-V-I-P-E-E-V-A-Y-P-P-T-E-L-G
..f X3 AA
TACGAGCCAGAACCAACGATTCGATCCAGACCAAGCCGAAATCAGACAAAGCCACAGCTCCACATCCGACGAGAAATAGTGTGGTCCGCGCTGA
..o X3 CDS
-R-S-R-T-K-R-L-R-S-R-R-K-P-K-S-D-K-P-T-T-A-H-A-D-E-N-S-V-G-A-R

5080 5090 5100 5110 5120 5130 5140 5150 5160 5170 5180

..f X1 CDS
..f X1 AA
..o X1 CDS
..o X1 AA
..f X2 CDS
GCTGCCGAAATGTTGCCACCGGATTTGAGCTCCACTCCCAAGCAATGAACCCACGCCCTAAAGTGATGATGCAATAGATTATTTCTATTTCACGACGATGATGATTTAG
..f X2 AA
-A-A-E-M-L-P-P-D-L-S-S-T-P-S-N-E-P-T-P-K-S-D-D-G-I-D-Y-F-L-F-D-D-D-D-D
..o X2 CDS
GCTGCCGAAATGTTGCCACCGGATTTGAGCTCCACTCCCAAGTAAATGAACCCACGCCCTAAAGTGATGATGCAATAGATTATTTCTATTTCACGACGATGATGATTTAG
..f X3 CDS
-A-A-E-M-L-P-P-D-L-S-S-T-P-S-N-E-P-T-P-K-S-D-D-G-I-D-Y-F-L-F-D-D-D-D-D
..f X3 AA
TACGAGCCAGAACCAACGATTCGATCCAGACCAAGCCGAAATCAGACAAAGCCACAGCTCCACATCCGACGAGAAATAGTGTGGTCCGCGCTGA
..o X3 CDS
-R-S-R-T-K-R-L-R-S-R-R-K-P-K-S-D-K-P-T-T-A-H-A-D-E-N-S-V-G-A-R

APÊNDICE B

Tabela 7: Substituições de aminoácidos encontradas pela análise BEB em sequências dos grupos de análise “Anastrepha”, “Tephritidae” e “Cyclorrhapha”.

Anastrepha		
Códon	Substituição	Probabilidade (%)
6G	G → S	89,30%
43F	F → L	66,80%
136K	K → Q	61,40%
210P	P → S	66,40%
215E	E → Q	51,40%
Tephritidae		
Códon	Substituição	Probabilidade (%)
1147G	G → V / P / S	64,40%
Cyclorrhapha		
Códon	Substituição	Probabilidade (%)
1314V	V → P / G / I / L / S	62,20%

APÊNDICE C

Tabela 9: Testes de normalidade de Ryan-Joiner para o Índice ovariano (I_o), para a Carga de ovos (C_o), e para os comprimentos da asa (hA) e do tórax (hT), considerando a presença de simetria bilateral entre ovários esquerdo e direito.

Normalidade de Ryan-Joiner (p-valor)				
Idade (dias)	I_o	C_o	Asa	Tórax
1	> 0,100	*	> 0,100	> 0,100
5	> 0,100	*	> 0,100	> 0,100
7	0,054	*	0,064	> 0,100
10	> 0,100	> 0,100	> 0,100	> 0,100
14	> 0,100	> 0,100	0,070	> 0,100
15	> 0,100	> 0,100	> 0,100	0,093
20	> 0,100	0,095	> 0,100	0,072
25	> 0,100	> 0,100	> 0,100	> 0,100
29	0,069	> 0,100	> 0,100	> 0,100

Tabela 10: p-valores obtidos a partir dos testes de comparação entre a diferença das variâncias (a) e a igualdade das médias populacionais (b). Os valores de variâncias e médias maiores que 0,05 não rejeitam H_0 e são estatisticamente iguais (cor verde), enquanto os menores que 0,05 rejeitam H_0 e são estatisticamente diferentes (cor vermelha).

(a)

DIFERENÇA DAS VARIÂNCIAS																			
p-valor para Índice ovariano										p-valor para Carga de ovos									
	1	5	7	10	14	15	20	25	29		1	5	7	10	14	15	20	25	29
1		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1		Idêntico	Idêntico	0,00	0,00	0,00	0,00	0,00	0,00
5			0,90	0,00	0,00	0,00	0,02	0,00	0,00	5			Idêntico	0,00	0,00	0,00	0,00	0,00	0,00
7				0,00	0,00	0,00	0,01	0,00	0,00	7				0,00	0,00	0,00	0,00	0,00	0,00
10					0,01	0,00	0,00	0,00	0,02	10					0,14	0,59	0,01	0,14	0,01
14						0,64	0,00	0,03	0,77	14						0,34	0,25	1,00	0,18
15							0,00	0,09	0,45	15							0,04	0,34	0,02
20								0,12	0,00	20								0,25	0,86
25									0,02	25									0,18
29										29									

(b)

DIFERENÇA DAS MÉDIAS																			
p-valor para Índice ovariano										p-valor para Carga de ovos									
	1	5	7	10	14	15	20	25	29		1	5	7	10	14	15	20	25	29
1		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1		Idêntico	Idêntico	0,00	0,00	0,00	0,00	0,00	0,00
5			0,24	0,00	0,00	0,00	0,00	0,00	0,00	5			Idêntico	0,00	0,00	0,00	0,00	0,00	0,00
7				0,00	0,00	0,00	0,00	0,00	0,00	7				0,00	0,00	0,00	0,00	0,00	0,00
10					0,03	0,13	0,08	0,05	0,38	10					0,00	0,05	0,46	0,45	0,45
14						0,00	0,00	0,00	0,00	14						0,10	0,00	0,00	0,00
15							0,84	0,47	0,39	15							0,11	0,16	0,00
20								0,47	0,23	20								0,93	0,00
25									0,11	25									0,01
29										29									

Tabela 11: p-valores obtidos a partir do teste t de igualdade de médias populacionais para a hipótese alternativa (H_1). Os valores menores que 0,05 não rejeitam H_1 e são estatisticamente iguais, enquanto os maiores que 0,05 rejeitam H_1 e são estatisticamente diferentes.

MÉDIAS MAIORES E MENORES																			
p-valor para Índice ovariano										p-valor para Carga de ovos									
	1	5	7	10	14	15	20	25	29		1	5	7	10	14	15	20	25	29
1		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1		Iguais	Iguais	0,00	0,00	0,00	0,00	0,00	0,00
5	Iguais		Iguais	0,00	0,00	0,00	0,00	0,00	0,00	5	Iguais		Iguais	0,00	0,00	0,00	0,00	0,00	0,00
7	1,00	0,88		0,00	0,00	0,00	0,00	0,00	0,00	7	Iguais	Iguais		0,00	0,00	0,00	0,00	0,00	0,00
10	1,00	1,00	1,00		0,01	Iguais	Iguais	Iguais	Iguais	10	1,00	1,00	1,00		0,00	Iguais	Iguais	Iguais	Iguais
14	1,00	1,00	1,00	0,99		1,00	1,00	1,00	1,00	14	1,00	1,00	1,00	1,00		Iguais	1,00	1,00	1,00
15	1,00	1,00	1,00	Iguais	0,00		Iguais	Iguais	Iguais	15	1,00	1,00	1,00	Iguais	Iguais		Iguais	Iguais	1,00
20	1,00	1,00	1,00	Iguais	0,00	Iguais		Iguais	Iguais	20	1,00	1,00	1,00	Iguais	0,00	Iguais		Iguais	1,00
25	1,00	1,00	1,00	Iguais	0,00	Iguais	Iguais		Iguais	25	1,00	1,00	1,00	Iguais	0,00	Iguais	Iguais		1,00
29	1,00	1,00	1,00	Iguais	0,00	Iguais	Iguais	Iguais		29	1,00	1,00	1,00	Iguais	0,00	0,00	0,00	0,00	