



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



Caracterização estrutural e de expressão do gene *exuperantia* (*exu*) em *Anastrepha fraterculus* (Diptera, Tephritidae) e evidências de seleção positiva em *Cyclorrhapha*

Janaína Lima de Oliveira

Orientador: Prof. Dr. Reinaldo Alves de Brito

São Carlos

2014

JANAÍNA LIMA DE OLIVEIRA

Caracterização estrutural e de expressão do gene *exuperantia (exu)* em *Anastrepha fraterculus* (Diptera, Tephritidae) e evidências de seleção positiva em *Cyclorrhapha*

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

Orientador: Prof. Dr. Reinaldo Alves de Brito

São Carlos

2014

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

O48c Oliveira, Janaina Lima de
Caracterização estrutural e de expressão do gene
exuperantia (exu) em *Anastrepha fraterculus*
(Diptera, Tephritidae) e evidências de seleção
positiva em *Cyclorrhapha* / Janaina Lima de Oliveira.
-- São Carlos : UFSCar, 2016.
80 p.

Dissertação (Mestrado) -- Universidade Federal de
São Carlos, 2014.

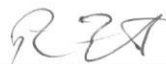
1. *Anastrepha*. 2. *Exuperantia*. 3. Evolução
molecular. 4. Co-optação. 5. *Bicoid*. I. Título.

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

Dissertação de Mestrado de
JANAÍNA LIMA DE OLIVEIRA

Banca Examinadora

Prof. Dr. Reinaldo Alves de Brito



Prof. Dr. Francis de Morais Franco Nunes



Prof. Dr. Klaus Hartmann Hartfelder



À minha família

AGRADECIMENTOS

Ao meu orientador, Prof. Dr. Reinaldo Alves de Brito, pela oportunidade de realizar esse projeto, pela confiança e por todo o aprendizado a mim concedido.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos, e à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo financiamento do projeto.

Ao Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular da Universidade Federal de São Carlos, pelo oferecimento do curso.

Ao pessoal do Laboratório de Genética de Populações e Evolução, pelo auxílio durante a realização desse projeto, em especial a Samira Chahad e Iderval Sobrinho pelos conselhos, amizade e contribuições nos trabalhos.

A toda a minha família, que esteve sempre presente, mas em especial aos meus pais e meu irmão, por toda força e apoio.

A todos aqueles que não foram aqui mencionados, mas fizeram parte, seja pelo incentivo, apoio técnico, ou simplesmente pelos momentos de descontração.

*“O começo de todas as ciências é o espanto
de as coisas serem como são”*

Aristóteles

SUMÁRIO

CAPÍTULO I – CARACTERIZAÇÃO ESTRUTURAL E DE EXPRESSÃO DO GENE <i>exuperantia</i> (<i>exu</i>) EM <i>Anastrepha fraterculus</i> (DIPTERA, TEPHRITIDAE).....	13
RESUMO.....	13
INTRODUÇÃO.....	14
O gênero <i>Anastrepha</i> e o grupo <i>fraterculus</i>	14
Genes envolvidos com a reprodução em <i>Anastrepha</i> : importância em estudos evolutivos e práticos.....	16
O gene <i>exuperantia</i> (<i>exu</i>): função e estrutura.....	17
MATERIAIS E MÉTODOS.....	22
Transcriptomas de cabeça e tecido reprodutivo de <i>Anastrepha fraterculus</i> por sequenciamento de próxima geração (<i>Next generation sequencing</i>).....	22
Caracterização das diferentes isoformas de <i>exu</i> geradas pelo sequenciamento de transcriptomas e desenho de <i>primers</i> para seu estudo em <i>A. fraterculus</i>	24
Identificação das regiões não traduzidas (UTRs) e região codificadora (CDS) de <i>exu</i> via RACE PCR e sequenciamento Sanger.....	25
Amplificação e sequenciamento do gene <i>exu</i> em <i>A. fraterculus</i>	27
RESULTADOS E DISCUSSÃO.....	28
Sequência e expressão diferencial de isoformas de <i>exu</i> a partir de dados de sequenciamento de transcriptomas de <i>A. fraterculus</i>	29
Amplificação e sequenciamento de fragmentos de regiões não traduzidas (UTRs) e região codificadora (CDS).....	31
Estrutura de <i>exuperantia</i> em <i>A. fraterculus</i>	31
CONCLUSÕES.....	33
CAPÍTULO II – EVIDÊNCIAS DE SELEÇÃO POSITIVA NO GENE <i>exuperantia</i> EM CYCLORRHAPHA (DIPTERA).....	35
RESUMO.....	35
INTRODUÇÃO.....	36
A subordem Brachycera (Diptera) e o grupo Cyclorrhapha.....	36

Mudanças moleculares adaptativas no grupo Cyclorrhapha: componentes do sistema nervoso e da determinação do eixo anteroposterior	37
MATERIAIS E MÉTODOS.....	40
RESULTADOS E DISCUSSÃO.....	42
Alinhamento das sequências, análises iniciais e árvore filogenética de <i>exuperantia</i>	42
Estaria <i>exu</i> evoluindo sob seleção positiva no grupo Cyclorrhapha?.....	45
CONCLUSÕES.....	47
CONSIDERAÇÕES FINAIS.....	48
REFERÊNCIAS BIBLIOGRÁFICAS.....	50
ANEXOS.....	55

LISTA DE TABELAS

CAPÍTULO I

Tabela 1.1 – Lista de *primers* construídos para estudo de *exu* em *A. fraterculus*.....25

CAPÍTULO II

Tabela 2.1 – Lista de espécies estudadas nesse trabalho e respectivas entradas de acesso no GenBank.....41

LISTA DE FIGURAS

CAPÍTULO I

Figura 1.1 – Morfologia do gênero <i>Anastrepha</i>	15
Figura 1.2 – Representação esquemática de ovários de <i>Drosophila</i> e composição das câmaras ovocitárias.....	17
Figura 1.3 – O gradiente do morfógeno BICOID.....	18
Figura 1.4 – Modelo revisado para a localização de mRNA <i>bcd</i> dependente de microtúbulos.....	20
Figura 1.5 – Estrutura do gene <i>exuperantia</i> em <i>D. melanogaster</i>	21
Figura 1.6 – Esquema representando os perfis das bibliotecas de cDNA.....	23
Figura 1.7 – Isoformas e regiões alternativas de <i>exu</i> em <i>A. fraterculus</i>	30
Figura 1.8 – Estrutura de <i>exuperantia</i> em <i>A. fraterculus</i> e em <i>D. melanogaster</i>	32

CAPÍTULO II

Figura 2.1 – Relações filogenéticas entre espécies de Diptera e o surgimento de <i>bicoid</i>	38
Figura 2.2 – Gráfico de saturação de <i>exu</i> : taxas de mutação x distâncias genéticas.....	43
Figura 2.3 – Estimativa dos níveis de divergência de <i>exu</i> no grupo Cyclorrhapha e entre este e outros Diptera.....	44
Figura 2.4 – Árvore filogenética não-enraizada de máxima verossimilhança do gene <i>exuperantia</i>	44
Figura 2.5 – Sítios de <i>exuperantia</i> sob seleção positiva em Cyclorrhapha.....	46

RESUMO

O gênero *Anastrepha* (Tephritidae) é de grande importância econômica devido ao seu impacto na fruticultura nacional. Estudamos aqui a história evolutiva de *Anastrepha fraterculus*, a espécie mais relevante do grupo *fraterculus*, o qual sofreu divergência recente e contém a maioria das espécies de importância econômica do gênero. Para entender melhor o processo de especiação nesse grupo, e auxiliar no processo de identificação taxonômica, precisamos estabelecer marcadores morfológicos e moleculares envolvidos com as diferenças entre as espécies. Genes envolvidos com a reprodução têm se mostrado bastante informativos para estes propósitos, devido às rápidas taxas evolutivas que apresentam, embora poucos tenham sido estudados e caracterizados nessas espécies até o momento. Nesse sentido, esse trabalho tem como foco o gene *exuperantia* (*exu*), que participa tanto da ovogênese (localização dos mRNAs *bicoid* e *oskar*) quanto da espermatogênese, uma vez que embriões formados a partir de fêmeas mutantes *exu* são inviáveis e machos mutantes *exu* são estéreis. Assim, o primeiro passo foi caracterizar estruturalmente e definir padrões de expressão desse gene entre os sexos nos tecidos cefálicos e reprodutivos em *A. fraterculus*. Utilizando sequenciamento de próxima geração (RNA-seq), identificamos um padrão estrutural e de expressão semelhante àquele de *Drosophila melanogaster* em tecidos reprodutivos: os transcritos diferem nas regiões não traduzidas (UTRs) 5' e 3' entre os sexos. Pela primeira vez identificamos a expressão de *exu* em tecidos cefálicos, envolvendo uma nova isoforma que é igual nos dois sexos. Em todos os casos, as regiões codificadoras são iguais, resultando na sequência proteica. Utilizamos essa região codificadora de *exu* junto com sequências de outras espécies de Diptera para investigar a evolução molecular de *exu* em Cyclorrhapha, que sofreu uma grande radiação adaptativa associada com alterações morfológicas e moleculares. Comparando a relação dN/dS entre Cyclorrhapha e outros Diptera, encontramos dois sítios sob seleção positiva em Cyclorrhapha, um deles presente no domínio semelhante à RNA exonuclease de *exu*. O segundo está presente em uma região não caracterizada para domínios proteicos, mas que é bem conservada e está sob seleção purificadora em Diptera, sugerindo ser uma região importante da proteína. As mudanças adaptativas encontradas em *exu* podem refletir um ganho evolutivo que permitiu sua co-optação para uma nova função em Cyclorrhapha: a de localização de *bicoid* na região anterior do ovócito, visto que *bcd* é um gene exclusivo de Cyclorrhapha.

Palavras-chave: *Anastrepha*; *exuperantia*; *Splicing* alternativo; Seleção positiva; Co-optação; *bicoid*.

ABSTRACT

The genus *Anastrepha* (Tephritidae) has several species of great economic importance due to their impact in friculture. We are particularly interested in the evolutionary history of *Anastrepha fraterculus*, the most relevant species in the closely related group of species *fraterculus*, which has undergone recent divergence and holds the majority of species of economic importance in the genus. To better understand the differentiation process and identification in the group, we need to find molecular and morphological markers that are involved with the differences between species. Reproductive genes have, in general, been very informative in this regard due to their rapid evolutionary rates, although few of them have been studied and characterized in this species so far. Therefore, the present study focuses on the *exuperantia* gene (*exu*), which participates in both oogenesis (localization of *bicoid* and *oskar* mRNAs) and spermatogenesis, since *exu* embryos from mutant mothers are unviable and males are sterile. Thus, the first step was to use Next generation sequencing strategies (RNA-seq) to structurally characterize and define expression patterns of this gene between sexes in the cephalic and reproductive tissues of *A. fraterculus*. *A. fraterculus* has similar structural and expression patterns to *Drosophila melanogaster* in reproductive tissues, in which the transcripts differ between sexes for the 5' and 3' untranslated regions (UTRs). We describe for the first time the expression of *exu* in cephalic tissues, involving a new isoform that is common to both sexes. All these alternative spliced transcripts share a common coding region, resulting in the same protein. We used the *exu* coding region, along with sequences from several Diptera to investigate the molecular evolution of *exu* in Cyclorrhapha. This group has experienced an enormous adaptive radiation, associated with molecular and morphological changes, and by comparing the dN/dS ratio between Cyclorrhapha and other Diptera we found that *exu* was subjected to positive selection in Cyclorrhapha for at least two sites in the coding region. One of the sites is present in the RNA exonuclease-like domain of *exu*. The second site is located in a not characterized region that is conserved and under purifying selection in Diptera, suggesting that it may be an important region of the protein. The adaptive changes found in *exu* may reflect an evolutionary gain that allowed its co-option for a new function the gene performs in Cyclorrhapha: the localization of *bicoid* in the anterior region of the oocyte, since *bcd* is a gene exclusive of Cyclorrhapha.

Keywords: *Anastrepha*; *exuperantia*; Alternative splicing; Positive selection; Co-option; *bicoid*.

PREFÁCIO

As moscas-das-frutas do gênero *Anastrepha* constituem um grupo cujas espécies são conhecidas principalmente por sua importância econômica na fruticultura, pois as fêmeas conseguem ovipor em uma ampla quantidade de frutos, tornando-os inviáveis para comercialização. As espécies do gênero são subdivididas em grupos, sendo o grupo *fraterculus* o mais estudado entre eles, principalmente por sua ampla distribuição e importância econômica. As espécies desse grupo sofreram divergência recente e apresentam uma grande semelhança morfológica, o que dificulta sua identificação taxonômica. Nesse sentido, o estabelecimento de marcadores moleculares entre linhagens é uma importante ferramenta, mas trabalhos prévios do nosso laboratório têm demonstrado que isso não é um processo trivial, principalmente porque essas espécies ainda compartilham muito polimorfismo ancestral.

Genes que estão envolvidos com a reprodução constituem bons candidatos a serem estudados em espécies próximas, pois em geral apresentam taxas evolutivas mais altas em relação ao restante do genoma, levando a uma fixação mais rápida entre as linhagens. Além disso, esses genes são alvos importantes para estudos de controle ambiental (em especial, para o desenvolvimento da técnica do inseto estéril). Diante disso, *exuperantia*, que é um gene que participa da gametogênese em machos e fêmeas, constitui um candidato interessante para estudo. Esse gene tem sido estudado sob o foco da biologia do desenvolvimento no inseto modelo *Drosophila melanogaster*, devido à sua importância na ovogênese e espermatogênese. Trabalhos posteriores adicionaram informações no que diz respeito à estrutura, padrões de expressão e evolução molecular desse gene, mas sempre restritos a esse gênero. Por isso, faz-se necessário o desenvolvimento de estudos mais abrangentes, seja no aspecto estrutural/funcional ou no contexto evolutivo. Assim, uma das propostas desse trabalho foi prover uma caracterização estrutural de *exu* em *Anastrepha fraterculus*, assim como de seus transcritos e sítios de expressão. Essas informações estão contidas no capítulo 1, intitulado “Caracterização estrutural e de expressão do gene *exuperantia* (*exu*) em *Anastrepha fraterculus* (Diptera, Tephritidae)”.

O grupo Cyclorrhapha (Diptera), que inclui, entre outras moscas, os gêneros *Drosophila* e *Anastrepha*, é o mais derivado dentro dessa ordem e sofreu uma grande radiação adaptativa. Poucos trabalhos até o momento têm revelado mudanças moleculares que possam estar envolvidas com o sucesso adaptativo desse grupo, mas crescentes evidências apontam a importância do estabelecimento do eixo anteroposterior e seus componentes nesse processo.

Exclusivo de Cyclorrhapha, *bicoid* (*bcd*) é um fator determinante no estabelecimento desse eixo, cujo mRNA deve estar localizado na região anterior do ovócito. Essa localização é um passo muito importante, visto que ao ser traduzido no embrião, a proteína BCD vai agir como um morfógeno que regula a expressão de vários genes. O surgimento de *bicoid* nessa linhagem parece ter promovido um desenvolvimento embrionário mais rápido, e provavelmente genes que participam dessa via também sofreram algumas mudanças, permitindo a integração de *bcd* nessa rede de interações. De fato, já foi demonstrado que regiões regulatórias de genes sobre os quais BCD age como um fator de transcrição sofreram algumas mudanças nessa linhagem, permitindo a regulação destes por parte de BCD. Entretanto, é possível que alterações adaptativas também tenham ocorrido em genes que agem regulando a transcrição de *bcd* ou participando do processo de localização de seu mRNA na região anterior do ovócito. Diante disso, *exuperantia*, que participa da localização do mRNA *bcd*, pode ter sofrido esse tipo de alteração por meio de seleção positiva, o que permitiu sua participação nessa nova função no grupo Cyclorrhapha e pode ajudar a entender tanto a evolução de *bcd*, quanto desse grupo. Assim, testamos a hipótese de que *exu* evoluiu sob seleção positiva nesse grupo, no capítulo 2, intitulado “Evidências de seleção positiva no gene *exuperantia* em Cyclorrhapha (Diptera)”.

Cada capítulo foi escrito de forma relativamente independente no que diz respeito aos seus objetivos, metodologias e conclusões. Entretanto, essas informações foram discutidas em conjunto no tópico “Considerações Finais”, trazendo algumas hipóteses que relacionam os aspectos estruturais/funcionais com a importância desse gene ao longo da evolução dos insetos.

Por fim, são listadas as referências bibliográficas e anexos referentes aos dois capítulos.

CAPÍTULO I: CARACTERIZAÇÃO ESTRUTURAL E DE EXPRESSÃO DO GENE *exuperantia (exu)* EM *Anastrepha fraterculus* (DIPTERA, TEPHRITIDAE)

Resumo: Genes envolvidos com processos reprodutivos apresentam em geral taxas de evolução mais rápidas que a maioria dos genes. Por esse motivo, são candidatos interessantes para o estabelecimento de marcadores de linhagens, principalmente em grupos que sofreram divergência recente, tal qual o grupo de moscas-das-frutas *fraterculus*, do gênero *Anastrepha* (Diptera, Tephritidae). Além disso, tais genes são alvos para estudos práticos de controle de espécies-praga. O gene *exuperantia (exu)* participa tanto da ovogênese, localizando os mRNAs *bicoid (bcd)* e *oskar (osk)* na região anterior e posterior do ovo, respectivamente, quanto da espermatogênese, embora ainda não se saiba sua função molecular neste último processo. Fêmeas mutantes *exu* são férteis, mas os embriões gerados são inviáveis porque não desenvolvem estruturas da região anterior. Por outro lado, os machos mutantes para esse gene são estéreis, visto que os espermatozoides não desenvolvem a cauda. Neste trabalho, a caracterização molecular e de expressão de *exu* em *Anastrepha fraterculus*, principal espécie do grupo, foi realizada a partir da análise de transcriptomas (RNA-seq) e confirmada utilizando ampliações por PCR e sequenciamento em Sanger. Nossos resultados indicam um padrão estrutural e de expressão semelhante àquele de *Drosophila melanogaster* nos tecidos reprodutivos: as isoformas diferem nas regiões 5' e 3' UTR entre os sexos, embora a região codificadora seja a mesma. Além disso, a análise de transcriptomas (RNA-seq) revelou pela primeira vez a expressão de *exu* em tecidos cefálicos, nos quais o transcrito é formado a partir de um sítio alternativo de início de transcrição (seja pela utilização de um novo promotor, ou resultado de regulação epigenética) diferente daqueles utilizados nos tecidos reprodutivos, mas é igual em machos e fêmeas, o que não havia sido relatado para *D. melanogaster*. É possível que trate-se de uma nova via de regulação da transcrição de *exu*, associada com a expressão desse gene em tecidos somáticos, dando origem a uma nova isoforma. Esse conjunto de resultados de caracterização de *exu* em *A. fraterculus* traz informações relevantes que podem ser utilizadas tanto para estudos evolutivos quanto práticos nesse gênero.

Palavras-chave: *Anastrepha*; *exuperantia*; *Splicing* alternativo.

INTRODUÇÃO

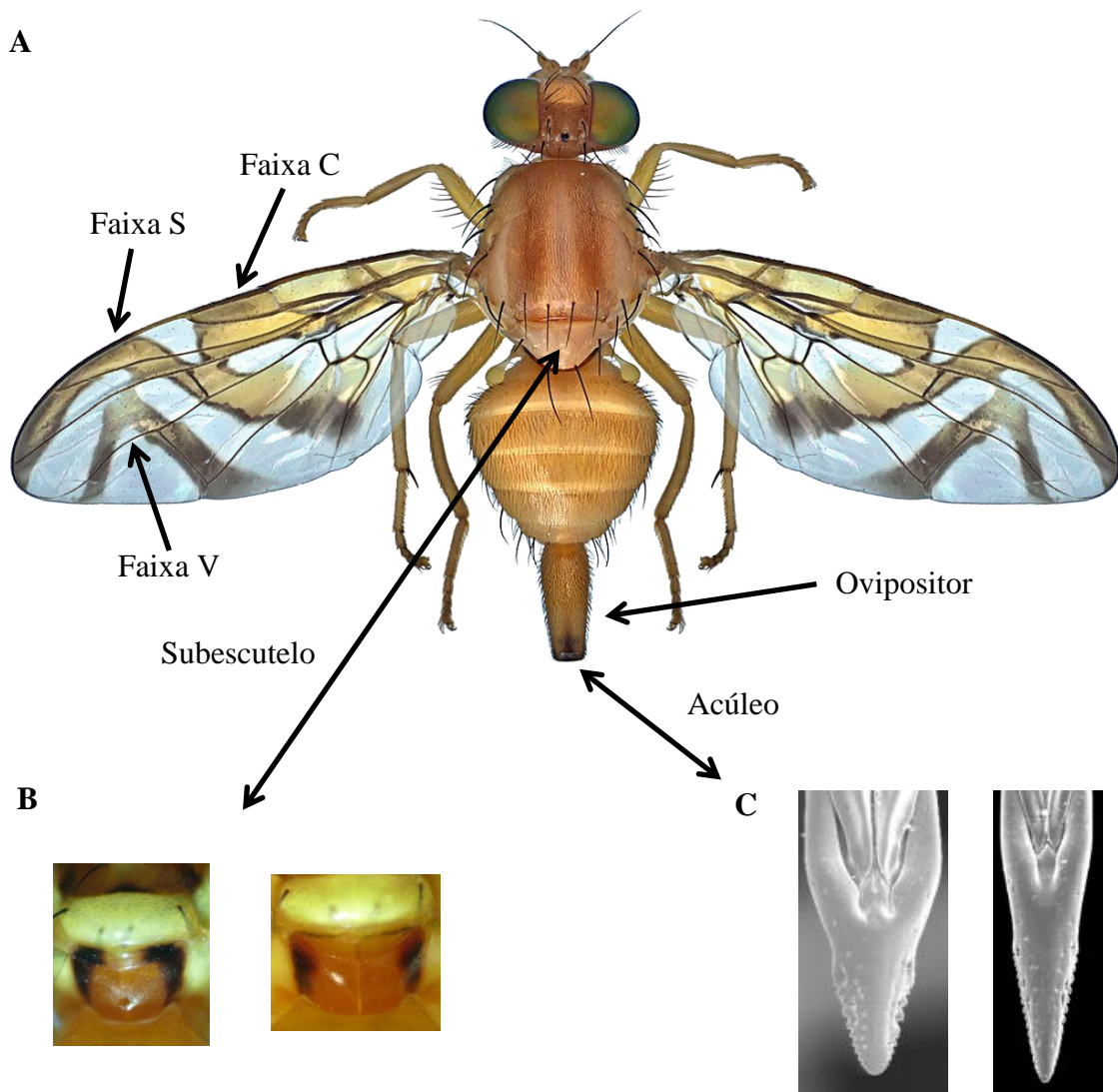
O gênero *Anastrepha* e o grupo *fraterculus*

A família Tephritidae possui um total de 471 gêneros e 4.257 espécies distribuídas por toda região tropical e temperada no mundo (THOMPSON, 1998), sendo mais conhecida pelas diversas espécies de grande importância econômica para a fruticultura, em sua maioria pertencentes aos gêneros *Bactrocera*, *Ceratitis*, *Rhagoletis* e *Anastrepha*. Dentre eles, *Anastrepha* é o grupo mais bem representado no Brasil, com aproximadamente 95 espécies, seguido de *Rhagoletis* com quatro, e *Bactrocera* e *Ceratitis* com apenas uma (ZUCCHI, 2000), e já foi observado em todos os estados brasileiros, embora nem todos tenham registros publicados (MALAVASI; ZUCCHI; SUGAYAMA, 2000).

As 197 espécies do gênero *Anastrepha* são distribuídas em 17 grupos, embora algumas não se enquadrem em nenhum deles (NORRBOM; ZUCCHI; HERNÁNDEZ-ORTIZ, 1999). Dentre estes, o grupo *fraterculus* engloba as espécies mais amplamente distribuídas (ZUCCHI, 1988) e economicamente importantes do gênero no Brasil (ZUCCHI, 2000). Essas moscas trazem grande prejuízo às lavouras devido ao fato de ovipositarem em uma ampla quantidade de frutos carnosos, tendo como principais hospedeiros espécies da família Myrtaceae (como goiaba e pitanga) e Anacardiaceae (como manga e seriguela) (MALAVASI; ZUCCHI; SUGAYAMA, 2000). As larvas se desenvolvem dentro dos frutos e alimentam-se de sua polpa, tornando esses frutos inviáveis para comercialização. Segundo o Instituto Brasileiro de Frutas (IBRAF), apenas 7,8% de toda a produção anual de frutas do país foi destinada à exportação (IBRAF, 2009), em parte devido ao controle contra a introdução de espécies parasitas exóticas nos países importadores (DUARTE; MALAVASI, 2000).

A taxonomia das espécies do grupo *fraterculus* leva em consideração caracteres morfológicos como padrões alares (coloração, nervuras e faixas C, S e V), manchas presentes no sub-escutelo e/ou no médio-tergito, e as características do acúleo (aspecto geral do ápice, presença ou não de constrição e dentes), sendo esta última estrutura fundamental para diferenciar entre as espécies (Figura 1.1) (ZUCCHI, 2000). Embora o conjunto desses caracteres seja frequentemente utilizado para identificar espécies do grupo, esse processo pode ser dificultado por sua plasticidade fenotípica e grande semelhança interespecífica (ARAÚJO; ZUCCHI, 2006).

Figura 1.1 – Morfologia do gênero *Anastrepha*



Legenda: (A) Aspecto geral da fêmea, destacando caracteres utilizados na taxonomia; (B) Detalhe mostrando subescutelos com e sem manchas, respectivamente; (C) Detalhe mostrando acúleos com e sem constrições, respectivamente. Fonte: ZUCCHI, 2000 (Adaptado).

Dada a complexidade do gênero *Anastrepha*, particularmente no grupo *fraterculus*, o estabelecimento de marcadores moleculares tem se mostrado uma estratégia promissora, seja para aplicação na taxonomia, na construção de mapas de ligação, ou trazendo informações a respeito dos processos evolutivos que ocorreram nesse grupo.

Genes envolvidos com a reprodução em *Anastrepha*: importância em estudos evolutivos e práticos

O estabelecimento de marcadores moleculares eficientes entre espécies é uma ferramenta fundamental para o entendimento da história evolutiva de um grupo, particularmente para o estudo de espécies que divergiram recentemente e acumularam poucas diferenças morfológicas entre si, o que dificulta sua taxonomia. Nesse aspecto, genes envolvidos com processos reprodutivos têm se mostrado bastante eficientes por apresentarem uma rápida taxa evolutiva sob seleção positiva (CIVETTA et al., 2006; SWANSON; VACQUIER, 2002; SWANSON et al., 2004).

Dado o interesse em estudar os processos evolutivos envolvidos na especiação do grupo *fraterculus*, nosso laboratório tem investigado diversos genes direta ou indiretamente relacionados com reprodução e reconhecimento entre machos e fêmeas no intuito de estabelecer marcadores de linhagens. Como resultado, foram encontrados sinais de seleção positiva em genes expressos em bibliotecas de cDNA de tecidos reprodutivos de fêmeas de *A. obliqua*, como os genes coriônicos *Cp15*, *Cp16*, *Cp19* e *Cp38*, e vitelínicos *Vm26Ab* e *Vm26Aa'*, embora apenas *Cp38* e os dois genes vitelínicos tenham apresentado mudanças radicais de aminoácidos, provavelmente em função da grande diversidade destas proteínas (GONÇALVES et al., 2013). Genes envolvidos na cascata de determinação sexual também mostraram estar evoluindo sob seleção positiva no grupo *fraterculus*: 16 sítios no gene *fruitless* e 3 em *doublesex* (SOBRINHO; DE BRITO, 2010; SOBRINHO; DE BRITO, 2012).

Além de bons candidatos a marcadores moleculares, genes envolvidos com a reprodução também podem ser alvos interessantes para estudos práticos de controle de insetos-praga. A técnica no inseto estéril, (*Sterile Insect Technique* – SIT) consiste na liberação em massa de insetos estéreis no ambiente, de forma que esses organismos modificados venham a competir com os indivíduos normais pelos parceiros na cópula, mas não deixem descendentes (KNIPLING, 1955). Para que essa metodologia seja eficiente, é necessário que os insetos estéreis desenvolvam-se normalmente até a fase adulta e sejam capazes de copular, embora não produzam uma prole viável.

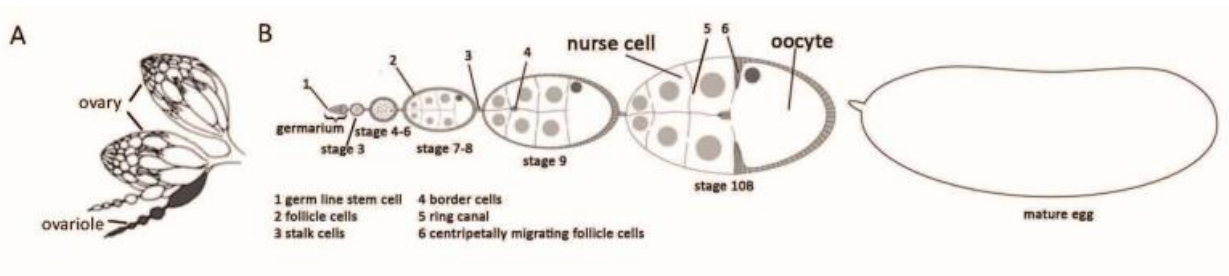
Dessa forma, estudos com genes envolvidos com a reprodução trazem informações para o entendimento da evolução do grupo estudado e podem ser eficientes marcadores moleculares

de linhagens (auxiliando a taxonomia), além de poderem ser alvos para estudos mais aplicados de controle biológico.

O gene *exuperantia* (*exu*): função e estrutura

Um trabalho pioneiro no estudo do gene *exuperantia* investigou nove genes de efeito materno que influenciam a determinação do eixo anteroposterior em *D. melanogaster* (SCHÜPBACH; WIESCHAUS, 1986). Esses genes são transcritos pela mãe durante a formação do ovo (ovogênese), nas câmaras ovocitárias (“*egg chambers*”), dentro dos ovariolos, que ficam nos ovários (Figura 1.2). Cada câmara ovocitária é formada por 16 células: 15 trofócitos (“*nurse cells*”) e um ovócito, todas derivadas de uma única célula (cistócito) após 4 divisões celulares, e interconectadas por canais em anel (“*ring canals*”) (SPRADLING, 1993).

Figura 1.2 – Representação esquemática de ovários de *Drosophila* e composição das câmaras ovocitárias



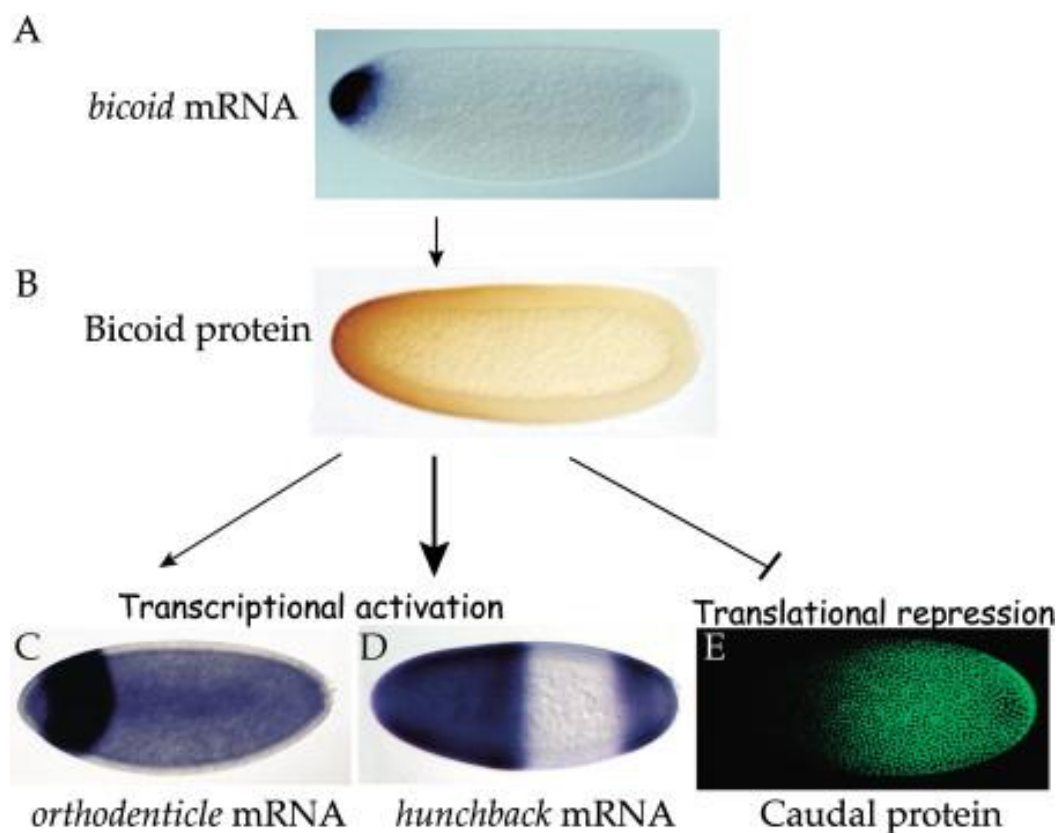
Legenda: (A) Ovário de *Drosophila*, consistindo de 15-20 ovariolos (B) Representação de um ovariolo, mostrando câmaras ovocitárias em diferentes estágios de diferenciação e o ovo maduro. Fonte: Wolff, 2012 (Adaptado).

Os resultados desse trabalho inicial mostraram que embriões produzidos por mães mutantes deficientes em *exu* apresentam deleções na região anterior (labrum), enquanto que apenas raramente apresentam defeitos na região posterior (espiráculos e télson). Além disso, através da construção de diferentes mosaicos, também mostrou-se que mesmo que as células do folículo (somáticas) tenham células com cópias normais do gene, se os trofócitos (linhagem germinativa) forem deficientes para *exu*, o fenótipo do embrião ainda será mutante (SCHÜPBACH; WIESCHAUS, 1986). Isso demonstra que o suprimento dos fatores maternos para o ovócito em desenvolvimento é dado a partir dos trofócitos, e não pelas células do folículo.

Outro gene de efeito materno que altera a formação de estruturas da região anterior em *D. melanogaster* é o *bicoid* (*bcd*). O transcrito desse gene é produzido nos trofócitos e transportado até a região anterior do ovócito (BERLETH et al., 1988). No embrião, a proteína

BICOID (BCD) sintetizada a partir desse mRNA difunde-se para a região posterior, formando um gradiente que ocupa 2/3 desse embrião (DRIEVER; NÜSSLEIN-VOLHARD, 1988). O estabelecimento desse gradiente é de fundamental importância, visto que BCD é um morfógeno (ou seja, promove o desenvolvimento de estruturas dependendo de sua concentração) que atua como um fator de transcrição de vários genes – tais como *orthodenticle* e *hunchback* – e regulador da tradução de outros genes, entre eles o *caudal* (EPHRUSSI; ST JOHNSTON, 2004) (Figura 1.3).

Figura 1.3 – O gradiente do morfógeno BICOID



Legenda: (A) Localização do mRNA *bcd* na região anterior do ovócito; (B) Gradiente da proteína BICOID da região anterior para a posterior; (C) Regulação transcricional e traducional promovida por BICOID. Fonte: Ephrussi; St. Johnson, 2004.

O mRNA *bcd* apresenta uma distribuição mais difusa no ovócito em mutantes deficientes para *exu* (BERLETH et al., 1988). Isso sugere que a proteína EXU pode ser um fator importante nessa função de localização, possivelmente capturando esse mRNA assim que ele entra no ovócito, ancorando-o na região anterior. Diversas linhas de evidência indicam que esse modelo é muito simplista, e que na verdade a ação de EXU nessa função inicia-se nos

trofócitos, localizando o mRNA *bcd* na região apical dessas células (ST JOHNSTON et al., 1989). De fato, há uma co-localização de EXU com o mRNA *bcd* nessas células e transitoriamente na região anterior do ovócito. O fato de que esse mRNA se mantém localizado no ovócito mesmo na ausência de EXU demonstra que essa proteína é importante apenas no início do processo de localização, não sendo necessária para manter esse padrão (MACDONALD; LUK; KILPATRICK, 1991; MARCEY; WATKINS; HAZELRIGG, 1991).

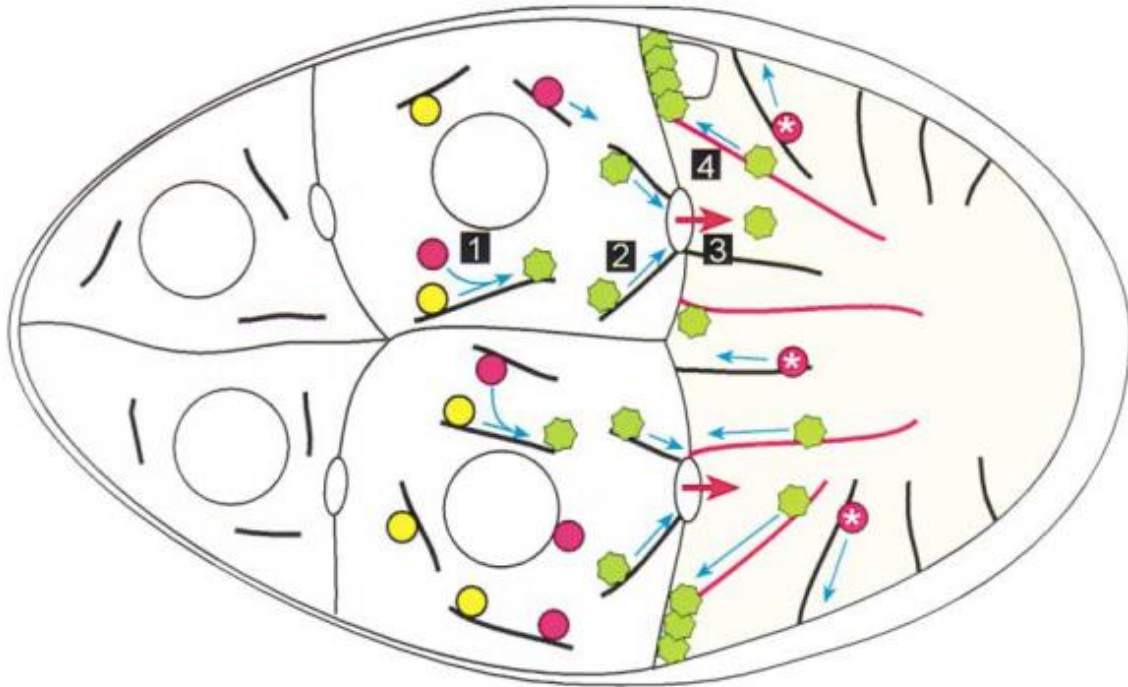
Diversos outros trabalhos acrescentaram informações sobre *exu* e *bcd*, como, por exemplo, demonstrando a importância do citoesqueleto nesse processo de localização (POKRYWKA; STEPHENSON, 1991). A partir dos conhecimentos obtidos até então, juntamente com uma série de experimentos, Cha et al. (2001) propuseram um modelo bastante elaborado que explica a localização do mRNA *bcd* na região anterior do ovócito e como EXU participa do processo. Nos trofócitos, EXU associa-se com microtúbulos e esse complexo participa da “maturação” do mRNA *bcd*, por um processo ainda não conhecido. Tanto esses mRNAs *bcd* maduros quanto os imaturos chegam ao ovócito passando através dos canais emanel, a diferença é que aqueles ligam-se a filamentos de microtúbulos originados no córtex anterior, enquanto esses últimos associam-se aleatoriamente aos filamentos, resultando em uma distribuição não-polarizada (Figura 1.4).

Outra função atribuída ao gene *exu* na ovogênese é a de localização do mRNA *oskar* (*osk*), através da associação entre EXU e YPSILON SCHACHTEL (YPS) (WILHELM et al., 2000). Assim como o gene *bcd*, *osk* é um gene de efeito materno que é transcrito nos trofócitos e seu mRNA é transportado até o ovócito, onde então deve ser localizado em uma região específica. Entretanto, ao contrário de *bcd*, o mRNA de *osk* deve ser localizado na região posterior do ovócito, onde vai atuar na montagem do plasma germinativo e controlar o número de células germinativas precursoras. Se o mRNA de *osk* estiver presente no polo anterior, leva à indução de células germinativas nessa região, enfatizando a importância de sua localização na região posterior para que o desenvolvimento do embrião ocorra normalmente (EPHRUSSI; LEHMANN, 1992).

Além de participar da localização dos mRNAs *bcd* e *osk* nos polos anterior e posterior do ovócito, respectivamente, alguns estudos demonstram a participação de *exu* na espermatogênese. Embora não se conheça o mecanismo molecular pelo qual esse gene participa do processo, sabe-se que machos mutantes para *exu* são estéreis, pois os

espermatozoides não desenvolvem a cauda (HAZELRIGG et al., 1990). É possível que EXU desempenhe uma função na espermatogênese similar àquela na ovogênese, participando da sublocalização de RNAs, embora não haja evidências disso até o momento.

Figura 1.4 – Modelo revisado para a localização de mRNA *bcd* dependente de microtúbulos



Legenda: No citoplasma dos trofócitos, o mRNA *bcd* forma partículas (vermelho) por um processo independente de EXU e microtúbulos. Microtúbulos então mediam a associação de EXU (amarelo) e pelo menos um fator adicional para formar complexos de transporte maduros (etapa 1, verde). Essas partículas são transportadas aos canais em anel por meio de microtúbulos que se originam próximo às junções celulares (etapa 2), e passam através destas por um processo independente de microtúbulos (etapa 3). Partículas maduras de RNA entrando no ovócito encontram uma rede densa de microtúbulos que são organizadas pelo córtex do ovócito (etapa 3), mas são preferencialmente transportadas através de um subgrupo de microtúbulos (vermelhos) que se originam no córtex anterior do ovócito (etapa 4). Por outro lado, as partículas de mRNA *bcd* que não passaram pelo citoplasma dos trofócitos (vermelho, asterisco) são transportadas por todos os microtúbulos, resultando em uma distribuição cortical não-polar. Fonte: Cha et al, 2001 (Adaptado).

As funções desempenhadas por *exu* em cada sexo, embora não completamente compreendidas, envolvem a expressão de transcritos específicos para cada sexo. Essas diferenças são nas regiões não traduzidas (*Untranslated Regions* – UTRs) 5' e 3', refletindo alguns aspectos: (I) a utilização de promotores diferentes por cada sexo, sendo que o da fêmea está presente em uma região mais interna, considerada um íntron em machos; (II) o reconhecimento de diferentes sítios de poliadenilação, de forma que em fêmeas esse sítio é reconhecido em uma região considerada um íntron em machos; (III) *splicing* diferencial sexo-

específico; (IV) como as diferenças são fora da região codificadora (CDS), a proteína resultante dos dois transcritos é a mesma (HAZELRIGG; TU, 1994). Em fêmeas, o mRNA de *exu* possui 2,1 kb, enquanto em machos ele é um pouco maior, com 2,9 kb (HAZELRIGG et al., 1990) (Figura 1.5).

Figura 1.5 – Estrutura do gene *exuperantia* em *D. melanogaster*



Legenda: Os transcritos apresentam éxons específicos para cada sexo nas regiões não traduzidas (UTRs), embora compartilhem a região codificadora (CDS). Fonte: Crowley; Hazelrigg, 1995 (Adaptado).

Além da estrutura, esses transcritos também diferem em relação ao padrão de expressão. Em *D. melanogaster*, o mRNA *exu* é detectado em machos a partir de larvas de estágio 3, mantendo-se por todos os estádios de pupa, e em adultos nos tecidos reprodutivos. Em fêmeas, o transcrito está presente nos tecidos reprodutivos das moscas adultas e em estágios iniciais do desenvolvimento do embrião (0-3 horas) (MACDONALD; LUK; KILPATRICK, 1991). Esse padrão de expressão diz respeito aos transcritos específicos para cada sexo nos tecidos reprodutivos. Além destes, um transcrito intermediário (2,5 kb) já foi encontrado em tecidos somáticos dos dois sexos (HAZELRIGG et al., 1990), embora não se tenha informações estruturais e funcionais mais detalhadas.

Devido à sua participação nos processos de ovogênese e espermatogênese, sendo importante para a viabilidade e fertilidade dos indivíduos, *exu* parece ser um alvo interessante para estudo no gênero *Anastrepha*, e embora tenha uma função importante no desenvolvimento de moscas, pouco se sabe sobre esse gene fora do gênero *Drosophila*. Dessa forma, esse trabalho teve como objetivos:

- Investigar sítios de expressão do gene (tecidos cefálicos e reprodutivos) em machos e fêmeas de *Anastrepha fraterculus*.
- Caracterizar os diversos transcritos resultantes de *splicing* alternativo desse gene em diferentes perfis.

- Descrever a organização estrutural de *exuperantia* nessa espécie, incluindo a região codificadora (CDS), regiões não traduzidas (UTRs) e íntrons.

Pelo sequenciamento de transcriptomas de cabeça e tecido reprodutivo de *A. fraterculus* por sequenciamento de próxima geração, caracterizamos a estrutura dos transcritos e sítios de expressão de *exu* nesses dois tecidos. Além disso, por meio de amplificação e sequenciamento Sanger, corroboramos algumas regiões dos transcritos, assim como descrevemos a organização estrutural do gene nessa espécie, revelando aspectos essenciais de sua estrutura.

MATERIAIS E MÉTODOS

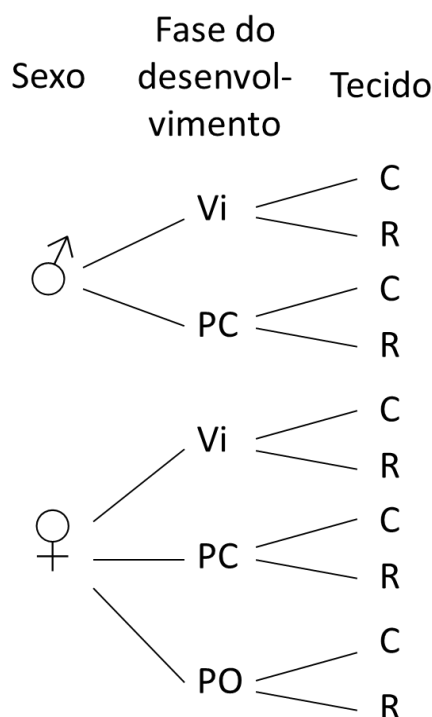
Transcriptomas de cabeça e tecido reprodutivo de *Anastrepha fraterculus* por sequenciamento de próxima geração (*Next generation sequencing*)

As moscas utilizadas nesse trabalho são provenientes de São Carlos (SP), e essa população tem sido mantida no Laboratório de Genética de Populações e Evolução na Universidade Federal de São Carlos há aproximadamente 15 gerações, em ambiente controlado (temperatura de $26^{\circ}\text{C} \pm 5^{\circ}\text{C}$; umidade relativa de 60-90%; fotoperíodo natural). As moscas são mantidas em gaiolas acrílicas contendo alimento (formado por 10 ml de mel, 10 ml de hidrolisado de proteína de milho, 0,5 ml de Levenil, 6g de levedo de cerveja, 0,5 g de Sustagen, 25 g de açúcar mascavo e 45 g de açúcar branco), água e uma manga (substrato para oviposição). A manga, contendo as larvas, é transferida para outra gaiola com vermiculita, até que estas empupem, quando então essas pupas são transferidas para uma nova gaiola contendo alimento e água. Ao emergirem, as novas moscas são mantidas para o estabelecimento da próxima geração.

Foram construídas bibliotecas de cDNA de diferentes perfis dos tecidos cefálicos e reprodutivos de *A. fraterculus*, considerando o sexo e fases do ciclo de vida (virgem e pós-cópula, além de pós-oviposição no caso das fêmeas). Para cada perfil, foi feito um *pool* com amostras de 10 indivíduos e uma réplica biológica com outros 10 indivíduos (Figura 1.6). Para isso, as moscas virgens foram separadas por sexo até que atingissem a maturidade sexual, quando então suas amostras de cabeça e tecido reprodutivo foram coletadas. No caso dos perfis pós-cópula, foram feitas gaiolas de cruzamento com machos e fêmeas sexualmente maduros, dos quais foram obtidas amostras de cabeça e tecido reprodutivo entre 15 e 20 horas após a

cópula. As últimas bibliotecas foram construídas a partir de fêmeas logo após as mesmas terem iniciado o processo de oviposição.

Figura 1.6 – Esquema representando os perfis das bibliotecas de cDNA



Legenda: Vi: Virgem; PC: Pós-cópula; PO: Pós-oviposição; C: Cabeça; R: Reprodutivo. Para cada um dos 10 perfis foi feito uma réplica biológica, gerando 20 bibliotecas, cada uma com amostras de 10 indivíduos diferentes. Fonte: Elaborada pela autora.

O RNA total foi extraído utilizando protocolo TRIzol/clorofórmio (CHOMCZYNSKI; MACKEY, 1995). A qualidade e integridade do material extraído foi observada em gel de agarose 1% corado com brometo de etídio visualizado em luz ultravioleta, e medida em NanoDrop 1000 (Thermo Scientific), enquanto a quantificação foi feita no Qubit (Invitrogen).

Um total de 4µg de cada *pool* de RNA total foi enviado para o Laboratório de Genômica Funcional Aplicada a Agropecuária e Agroenergia na Escola Superior de Agricultura Luiz de Queiroz – USP/Piracicaba, SP, sob coordenação do Prof. Dr. Luiz Lehmann Coutinho, para a construção das bibliotecas de cDNA e posterior sequenciamento. Para isso, foram utilizados protocolos padrão do kit TruSeq Stranded Total RNA Sample Prep (Illumina), e em seguida as bibliotecas produzidas foram sequenciadas no equipamento Illumina HiSeq2000, agrupando em *pools* de 12 bibliotecas por corrida, produzindo sequências

pareadas (2 x 100 pb) – *paired-end reads* – que foram separadas por biblioteca e enviadas para nosso laboratório.

As sequências produzidas por sequenciamento de próxima geração (*reads*) foram filtradas por qualidade e tamanho utilizando o software SeqyClean (disponível em <https://bitbucket.org/izhbannikov/seqyclean>). Em seguida, foi feita a montagem (*assembly*) de quatro perfis levando em consideração os tecidos e os sexos: cabeça de fêmea, cabeça de macho, reprodutivo de fêmea e reprodutivo de macho, agrupando as diferentes réplicas para um mesmo perfil e fases de vida diferentes nestas classes mais abrangentes, utilizando o pacote Trinity (GRABHERR et al., 2011). A anotação foi feita usando o pacote Standalone BLAST Setup for Unix (ALTSCHUL et al., 1997) utilizando como referência o Gene Ontology (GO).

Caracterização das diferentes isoformas de *exu* geradas pelo sequenciamento de transcriptomas e desenho de *primers* para seu estudo em *A. fraterculus*

Em todos os perfis foram encontrados *contigs* identificados como *exu* após a anotação, cujas ORFs (*Open Reading Frames*) e sequências proteicas foram inferidas *in silico* (BIKANDI et al., 2004). Eles foram alinhados entre si e também com as isoformas e sequência genômica já conhecidas de *D. melanogaster* (disponíveis em <http://flybase.org/>) para uma análise da estrutura desses transcritos.

Em seguida, foram desenhados *primers* nas regiões não traduzidas (UTRs) que seriam possíveis éxons de isoformas alternativas, assim como na CDS, a fim de elucidar a estrutura do gene (incluindo íntrons) e confirmar as sequências dos *contigs* via PCR e sequenciamento Sanger. Estes *primers* foram analisados no programa OligoAnalyzer 3.1 (disponível em <http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>) quanto aos valores de ΔG e as possibilidades de formar *hairpin*, *self-dimer* e *hetero-dimer*, visando diminuir a chance de formação de artefatos durante as reações. Todos os *primers* gerados e utilizados nesse trabalho estão listados na tabela a seguir (Tabela 1.1).

Tabela 1.1 – Lista de *primers* construídos para estudo de *exu* em *A. fraterculus*

Nome	Sequência (5'-3')	Tm (°C)	Região
exu5cf F	TTTGGTTATTGCATCGTGG	59.5	5' UTR de cabeça
exu5cf R	TTGCACTACTAGGTTAAGGC	59.4	5' UTR de cabeça
exuco5 R	TTTCCAATAAAAATAATCCAAGCG	59.4	Região comum 5' UTR
exuCDS1 F	GTAGTCAATTCTCCAGTAGTAGG	59.4	1/4 CDS
exuCDS2 F	CATACATGAACTTGAATCCTGC	59.7	2/4 CDS
exuCDS2 R	CGATAGAATCCAATAGAGATCACAC	60.8	2/4 CDS
exuCDS3 F	AGATTTTCTGGATTGGCTGG	60.4	3/4 CDS
exuCDS3 R	ATGATAGACTAGAATGACACCG	59.3	3/4 CDS
exuCDS4 F	CAATAACTCCTTAAAAACACAGC	58.8	4/4 CDS
exuCDS4 R	GTTTTTAAGGAGTTATTGCGTGAG	60.8	4/4 CDS
exu5rf F	GATCCGTGGAAGTCAAACCTG	60.6	5' UTR de rep. de fêmea
exu5rf R	CAGTTTGACTTCCACGGATC	60.6	5' UTR de rep. de fêmea
exu3rf1 F	ACGTTAGCAGTTACAATACCC	60.1	1/2 de 3' UTR final de rep. de fêmea e outros
exu3rf1 R	TGGGTATTGTAAGTCTAAGC	60.1	1/2 de 3' UTR final de rep. de fêmea e outros
exu3rf2 F	TTTTGAATTTTAGAAGGTTGCG	59.5	2/2 de 3' UTR final de rep. de fêmea e outros
exu3rf2 R	AGCGCAACCTTCTAAAAATTC	59.9	2/2 de 3' UTR final de rep. de fêmea e outros
exu5rm F	AAATTGTTTTGCATTTCTATTCAAG	58.9	5' UTR de reprodutivo de macho
exu5rm R	AATATAGAAAGAACTTGAATTTAGC	57.7	5' UTR de reprodutivo de macho
exu3rma F	AACGTAGCCACACTTTAGAG	59.6	3' UTR 'a' de rep. de macho e outros
exu3rma R	TCTGGCGTCTCTCTAAAGTG	61	3' UTR 'a' de rep. de macho e outros
exu3rmb F	TTTAATTCGCTCAAACACTAGG	59.2	3' UTR de macho
exu3rmb R	GTTATTTTCATCGGATCATGCC	59.4	3' UTR de macho
exu3rmc1 F	TACCTGGATGCTTTCAATGC	60.7	1/2 região de 3' UTR de macho [1]
exu3rmc1 R	CTAATTGTTGTTGTGCTCCG	59.5	1/2 região de 3' UTR de macho [1]
exu3rmc2 R	CAAGAAGTCACAACAGAATATTAGC	60.5	2/2 região de 3' UTR de macho [1]

Fonte: Elaborada pela autora.

Identificação das regiões não traduzidas (UTRs) e região codificadora (CDS) de *exu* via RACE PCR e sequenciamento Sanger

Com o intuito de confirmar se as regiões UTRs encontradas nos *contigs* eram realmente resultados de *splicing* alternativo do gene, foi aplicada a técnica de RACE PCR (*Rapid Amplification of cDNA Ends*), utilizando o kit FirstChoice® RLM-RACE (Ambion). Por meio da ligação de adaptadores no fosfato 5' dos mRNAs, ao sintetizar o cDNA essa fita complementar possui uma região conhecida em sua extremidade 5' (no caso do 5' RACE), que será importante na PCR, conforme explicado mais adiante. Para 3' RACE o procedimento é

mais simples, pois o *primer* que será estendido durante a reação de transcrição reversa hibridiza na cauda poli-A do mRNA e já contém o adaptador que servirá de molde na PCR. Para esse trabalho, tanto quanto com relação ao RACE 5' quanto ao RACE 3', foram utilizados *pools* de RNA total formados por alíquotas do RNA extraído para o sequenciamento do transcriptoma, conforme explicado no tópico anterior. Dessa forma, a amostra aqui utilizada foi constituída de RNA total de cabeça e tecido reprodutivo de machos e fêmeas em diferentes fases do ciclo de vida.

A partir desse passo, os procedimentos para 5' RACE e 3' RACE são semelhantes, e por isso serão explicados de uma forma generalizada. Para a RACE PCR, o kit provê *primers* que hibridam na região mais externa do adaptador (*primers* externos) e na região mais interna (*primers* internos), mas também são necessários *primers* externos e internos específicos para o gene *exu*. As amplificações das regiões 5' e 3' foram realizadas separadamente a partir de PCRs aninhadas (*nested* PCRs). Na primeira reação de amplificação, chamada de “*outer*”, foram utilizados *primers* mais externos (do adaptador e da CDS de *exu*) ao fragmento de interesse. Na segunda reação, chamada “*inner*”, foram utilizados *primers* mais internos em relação à anterior e a amostra utilizada foi o produto da amplificação da reação *outer*.

As amostras de cDNA geradas para o RACE PCR também foram utilizadas para amplificação da região codificadora, utilizando *primers* específicos para essa região.

Foram preparadas reações de amplificação com reagentes nas seguintes concentrações finais: Tampão 1X, MgCl₂ 2,5 mM, 0,2 mM de cada dNTP, 0,4 μM de *primer forward*, 0,4 μM de *primer reverse*, 1,25 U de SuperTaq Plus™ Polymerase (Ambion), 1 μl - 2 μl de amostra (cDNA, no caso das reações *outer*, ou o produto da reação *outer*, no caso da reação *inner*), além de água mili-Q, perfazendo um volume final de 50 μl de reação. As condições de ciclos das PCRs foram: desnaturação inicial de 94°C por 3 min; 35 ciclos de 94°C por 30 seg, 52°C-55°C* por 30 seg, 72°C por 30 seg-3 min*; e uma extensão final a 72°C por 10 min (*a temperatura de hibridação foi ajustada para cada par de *primer*, assim como o tempo de extensão em relação ao tamanho do fragmento – 1 min/kb). As amostras obtidas foram aplicadas em gel de agarose 1% corado com brometo de etídio e visualizado em luz ultravioleta para verificar o tamanho dos fragmentos obtidos. Em seguida, foram purificados com polietilenoglicol (PEG 8000) (LIS; SCHLEIF, 1975) e novamente visualizados em gel de agarose conforme já descrito.

Os fragmentos purificados foram clonados utilizando o kit InsTAclone (Fermentas), utilizando plasmídeos PTZ57R/T, e células competentes DH5 α para transformação por meio de choque térmico. Essas bactérias foram plaqueadas em placas de Petri contendo meio LB sólido com ampicilina (0,1 μ g/ μ l), e as colônias formadas, individualmente isoladas, fervidas a 95°C e submetidas à amplificação por PCR utilizando os *primers* universais M13 (DALLAS-YANG; JIANG; SLADEK, 1998). As amostras purificadas foram enviadas para sequenciamento na empresa Macrogen Inc., Coréia do Sul, onde o procedimento foi realizado em sequenciadores ABI 3730.

Os eletroferogramas das sequências obtidas foram analisados no programa FinchTV 1.4.0 (disponível para download em <http://www.geospiza.com/Products/finchtv.shtml>). As sequências com alto índice de qualidade foram selecionadas, eliminando-se regiões de baixa qualidade e aquelas correspondentes ao vetor. O alinhamento dos fragmentos entre si e com os *contigs* gerados pelo sequenciamento dos transcriptomas foi conduzido no programa BioEdit 7.1.3.0 (HALL, 1999).

Amplificação e sequenciamento do gene *exu* em *A. fraterculus*

A extração do DNA utilizado nessa etapa foi realizada utilizando o protocolo de tiocianato de guanidina (NELSON; KRAWETZ, 1992), preservando o esqueleto dos indivíduos para análises morfológicas. A qualidade do DNA foi então observada em gel de agarose 1% corado com brometo de etídio.

Como há *splicing* alternativo nas regiões 5' UTR e 3' UTR, foram realizadas várias reações de amplificação utilizando combinações diferentes de *primers forward* (*f*) e *reverse* (*r*) de éxons específicos de cada isoforma nessas regiões, a fim de entender a organização desses éxons na sequência genômica. Por exemplo, considerando a existência de 3 éxons 5' UTR diferentes (A, B e C), temos 6 combinações diferentes de como esses éxons podem estar ordenados no gene: A-B-C, A-C-B, B-A-C, B-C-A, C-A-B, C-B-A. Para responder a essa questão, foram realizadas reações com todas as combinações de *primers* possíveis (*forward* do éxon A, com *reverse* do éxon B e do éxon C, e assim sucessivamente), esperando-se que só amplificassem quando se tratasse das seguintes reações: (I) utilizando o *primer forward* do primeiro éxon com o *reverse* do segundo e do terceiro éxon, (II) utilizando o *primer forward* do segundo éxon com o *reverse* do terceiro éxon. Voltando ao exemplo anterior, se a organização correta dos éxons for A-B-C, espera-se que amplifiquem os fragmentos das

reações com *primer forward* A (Af) + reverse B (Br), *forward* A (Af) + *reverse* C (Cr), assim como *forward* B (Bf) + *reverse* C (Cr), e que o fragmento gerado por aquela primeira reação (Af + Cr) seja maior que o dessa última (Bf + Cr), já que nesse caso os éxons estão mais distantes entre si.

Todas as reações de amplificação foram preparadas com reagentes nas seguintes concentrações finais: Tampão 1X, MgCl₂ 2,5 mM, 0,2 mM de cada dNTP, 0,2 μM de *primer forward*, 0,2 μM de *primer reverse*, 1,25 U Taq DNA Polymerase (Fermentas), 2 μl de amostra de DNA extraído conforme explicado anteriormente, e água mili-Q, perfazendo um volume final de 50 μl de reação. As condições de ciclos das PCRs foram: desnaturação inicial de 94°C por 3 min; 35 ciclos de 94°C por 30 seg, 50°C-55°C* por 30 seg, 72°C por 45 seg-3:30 min*; e uma extensão final a 72°C por 10 min (*a temperatura de hibridização foi ajustada para cada par de *primer*, assim como o tempo de extensão em relação ao tamanho do fragmento – 1 min/kb). As amostras obtidas foram analisadas em gel de agarose corado com brometo de etídio, sob luz ultravioleta. Em seguida, os produtos foram purificados com polietilenoglicol, novamente analisados em gel de agarose, e esses fragmentos purificados foram clonados, amplificados utilizando os *primers* universais M13, purificados com polietilenoglicol e enviados para sequenciamento na empresa Macrogen Inc., conforme já explicado anteriormente.

Para a análise e alinhamento das sequências, foi utilizada a mesma metodologia descrita no tópico anterior, sendo conduzidos nos programas FinchTV 1.4.0 e BioEdit 7.1.3.0 (HALL, 1999), respectivamente.

RESULTADOS E DISCUSSÃO

O sequenciamento de transcriptomas de cabeça e tecido reprodutivo de *A. fraterculus* consistiu em uma série de experimentos realizados em conjunto com membros do Laboratório de Genética de Populações e Evolução (UFSCar), e não é específico desse trabalho. O manuscrito com a descrição dessas bibliotecas está sendo elaborado para publicação e consiste na dissertação de mestrado de outro membro do grupo. Dessa forma, aqui serão apresentados apenas os resultados referentes à etapa pós-anotação e identificação dos transcritos *exu*.

Sequência e expressão diferencial de isoformas de *exu* a partir de dados de sequenciamento de transcriptomas de *A. fraterculus*

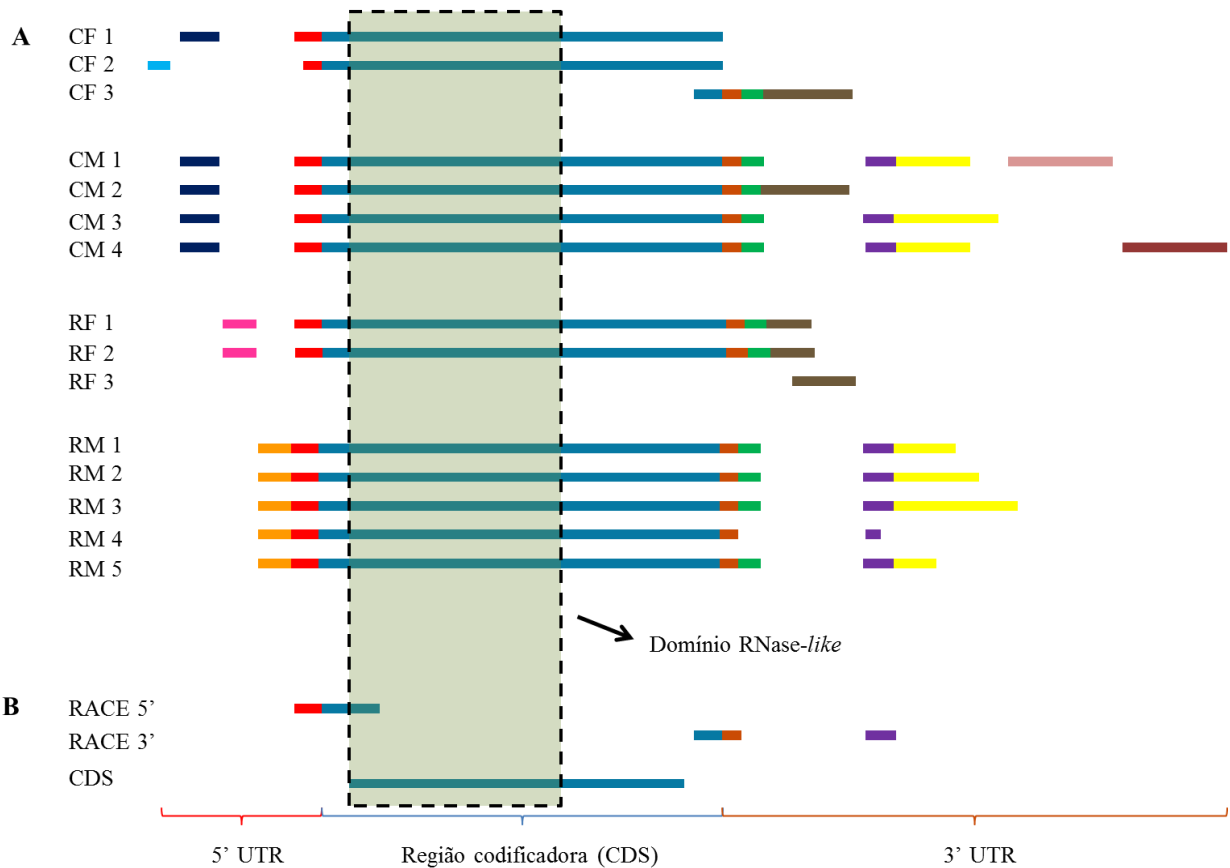
Após a anotação, foram encontrados *contigs* referentes a *exu* em todas as bibliotecas (Figura 1.7 A; Anexo A). A análise desses transcritos através de alinhamentos demonstrou que, quando completos, esses transcritos apresentam a mesma região codificadora (CDS), resultando em uma mesma proteína.

Um padrão interessante pôde ser observado em relação às regiões 5' UTR. As isoformas de reprodutivo de macho e reprodutivo de fêmea apresentam éxons específicos para cada sexo, da mesma forma que ocorre em *Drosophila*. Isso pode refletir a utilização de promotores diferentes seguida de *splicing* alternativo, ou início da transcrição a partir de um novo sítio por meio de regulação epigenética (ROBERTS; GAVERY, 2012). No caso de *Drosophila*, isso se dá pela utilização de promotores diferentes, sendo que aquele utilizado pela fêmea está localizado em uma região mais interna do gene em relação ao do macho. Este promotor da fêmea está presente no transcrito inicial do macho (pois este usa um outro promotor localizado mais a montante), mas é reconhecido como íntron nesse sexo e retirado após o processamento do RNA. Por outro lado, como a fêmea usa o promotor mais interno, a primeira base transcrita já faz parte de seu éxon 5' UTR e por isso não é retirado durante o *splicing* (HAZELRIGG; TU, 1994).

Observamos também a expressão de *exu* em tecidos cefálicos nos dois sexos, o que ainda não foi descrito na literatura para outras espécies. Entretanto, já foi demonstrada a expressão de uma isoforma de *exu* fora de tecidos reprodutivos em *Drosophila* (HAZELRIGG et al., 1990). Os *contigs* gerados nesses dois perfis de *A. fraterculus* mostram que os transcritos produzidos em cabeça nos dois sexos compartilham uma mesma região 5' UTR, que é diferente daquelas dos tecidos reprodutivos. Em análises preliminares de transcriptomas de larvas e pupas foi visto que as isoformas de *exu* presentes nessas fases de desenvolvimento apresentam uma região 5' UTR igual a essas de cabeça (dados não mostrados). Isso traz duas novidades interessantes: não só foi observada a expressão de *exu* em um novo sítio (cabeça), como também nossos resultados apontam a existência de uma isoforma diferente daquela de tecidos reprodutivos, provavelmente sob regulação diferencial (resultado da utilização de outro promotor ou de um novo sítio de início da transcrição regulado epigeneticamente) associada com a expressão desse gene em tecidos somáticos de adultos (cabeça), e em estágios iniciais do

desenvolvimento (larvas e pupas). É possível que em cabeça também ocorra a expressão de outra isoforma, mas específica de fêmeas.

Figura 1.7 – Isoformas e regiões alternativas de *exu* em *A. fraterculus*



Legenda: (A) *Contigs exu* gerados a partir do sequenciamento de bibliotecas de cDNA via *Next Generation Sequencing*. Cada cor nas UTRs (5' e 3') representa possíveis éxons diferentes. RM e RF: reprodutivo de macho e fêmea; CM e CF: cabeça de macho e fêmea. A região hachurada diz respeito ao domínio semelhante a RNase identificado por Mose et al. (1997). (B) Fragmentos sequenciados após amplificação via RACE PCR, incluindo UTRs, ou via PCR comum utilizando cDNA como amostra. Fonte: Elaborada pela autora.

Na região 3' UTR não houve uma correlação muito evidente entre éxons alternativos e os perfis analisados. Entretanto, é possível perceber que algumas regiões estão mais associadas a machos do que a fêmeas, independente do tecido. Em *Drosophila*, sabe-se que o éxon 3' UTR de *exu* específico de macho é importante para a manter a estabilidade do mRNA nos tecidos reprodutivos desse sexo. Não se sabe ao certo o mecanismo pelo qual isso acontece, mas é possível que essa região forme estruturas secundárias que impeçam a degradação do mRNA diretamente (impedindo a ligação de RNases) ou indiretamente (permitindo a ligação de proteínas que o protejam contra a degradação) (CROWLEY; HAZELRIGG, 1995).

Amplificação e sequenciamento de fragmentos de regiões não traduzidas (UTRs) e região codificadora (CDS)

Conforme já explicado, foi utilizada uma segunda metodologia para identificação de éxons UTR alternativos – a de RACE PCR, e os resultados desses experimentos estão ilustrados na Figura 1.7 B. Em relação ao RACE 5', o éxon 5' UTR corresponde a uma região comum a todas as isoformas identificadas no sequenciamento das bibliotecas. Na região 3' UTR, foi reconhecida uma região comum a todas as isoformas previamente identificadas, mas também uma região que parece estar mais associada ao tecidos masculinos (compare os resultados de 1.7 A e 1.7 B).

A amplificação e sequenciamento de cDNA referente a essa região codificadora (CDS) corroborou o resultado obtido na etapa anterior de apenas uma isoforma encontrada, demonstrando que em *A. fraterculus*, assim como em *Drosophila*, o *splicing* alternativo não afeta a região codificadora do gene *exu* (HAZELRIGG; TU, 1994). Isso significa que a proteína codificada é a mesma nos perfis analisados, embora possa estar sujeita a modificações pós-traducionais. De fato, em *Drosophila* foi visto que EXU é alvo da proteína quinase Par-1 (RIECHMANN; EPHRUSSI, 2004), embora as regiões nas quais ela se liga não sejam conservadas em *Anastrepha*.

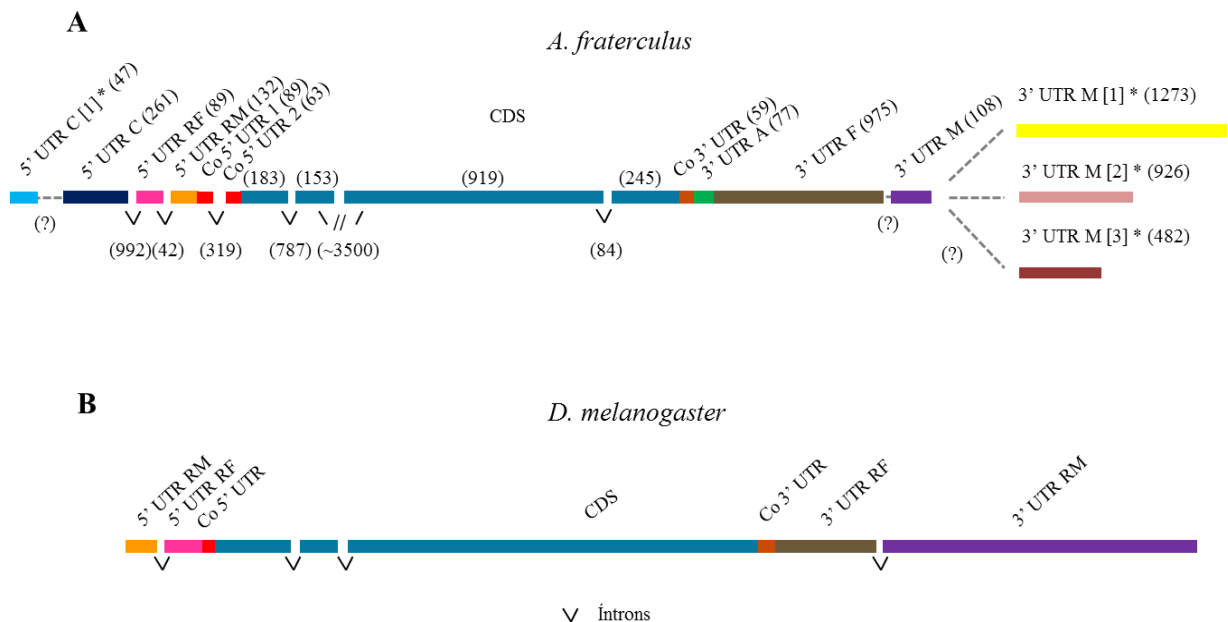
Pouco se sabe a respeito da proteína EXU, mas foi identificado um domínio semelhante ao de uma RNase ou DNase (RNase/DNase-like) amplamente distribuído entre os mais diferentes organismos e proteínas, de bactérias a humanos (MOSER et al., 1997). É possível que em EXU esse domínio tenha permanecido com a função catalítica original ou esteja apenas envolvido com ligação ao RNA, como foi observado com outra proteína de *Drosophila* (EGALITARIAN) que o contém (NAVARRO et al., 2004; DIENSTBIER et al., 2009).

Estrutura de *exuperantia* em *A. fraterculus*

Os fragmentos sequenciados foram alinhados entre si e com as isoformas já identificadas nos experimentos anteriores. Isso permitiu o reconhecimento dos éxons alternativos das UTRs e o entendimento de como estes se organizam na sequência genômica. Esses resultados, junto com uma comparação com a estrutura de *exu* em *D. melanogaster*, estão representados na Figura 1.8 (ver também Anexo B).

A ordem dos éxons 5' UTR de reprodutivo de fêmea (RF) e 5' UTR de reprodutivo de macho (RM) em *A. fraterculus* (Figura 1.8 A) está invertida em relação à encontrada em *D. melanogaster* (Figura 1.8 B). O éxon de cabeça (5' UTR C), identificado apenas em *A. fraterculus* até o momento, é anterior aos dois, separando-se destes por um íntron. Além disso, são observados dois íntrons adicionais, um na região comum 5' UTR (Co 5' UTR) e outro localizado na CDS, criando dois éxons para uma região que consiste em apenas um éxon em *D. melanogaster*. Ainda não foi possível estabelecer uma conexão entre os éxons 3' UTR "F", mais associado aos transcritos de fêmea, e 3' UTR "M", exclusivo de machos, e assim responder se há um íntron ou não entre eles. Entretanto, pelo menos a relação entre este último e o restante do gene *exu* já foi demonstrada no experimento de RACE PCR.

Figura 1.8: Estrutura de *exuperantia* em *A. fraterculus* e em *D. melanogaster*



Legenda: **(A)** *A. fraterculus*; **(B)** *D. melanogaster*. C: Cabeça; R: Reprodutivo; M: Macho; F: Fêmea; Co: Regiões UTRs comuns entre as isoformas. As conexões que ainda não puderam ser inferidas estão indicadas por linhas tracejadas e por um ponto de interrogação "?", enquanto os éxons ainda não validados, por um asterisco "*". "/" indica que a região ainda não foi sequenciada por inteiro. Os valores entre parênteses na primeira parte da figura **(A)** indicam o tamanho da região em pares de bases (pb). Fonte: Elaborada pela autora.

Algumas regiões encontradas nos *contigs* (Figura 1.7 A) não puderam ser associadas ao gene *exu* até o momento, seja pelo experimento de RACE PCR ou quando sua estrutura genômica foi estudada. Elas estão indicadas na Figura 1.8 A com um asterisco (*). É possível que elas sejam parte do gene, pois no experimento de RACE PCR foram encontrados outros fragmentos além daquele que foi sequenciado, indicando a existência de éxons UTR

alternativos. Por outro lado, eles podem ser apenas um artefato associado a *exu* durante o processo de montagem dos *contigs* pelas estratégias de sequenciamento de próxima geração. Como estas regiões não codificadoras apresentam uma grande diferenciação em suas sequências, não permitindo sua homologização, essa questão permanece em aberto até que sejam realizados outros experimentos, primeiramente identificando todas as regiões UTRs existentes do gene *exu* via RACE PCR e sequenciamento, seguido do desenho de novos *primers*, amplificação e sequenciamento dessas regiões no DNA, a fim de entender como elas estão organizadas na sequência genômica.

Os resultados aqui apresentados trazem informações muito relevantes em relação ao gene *exuperantia* em *A. fraterculus*. Pela primeira vez foi feita uma caracterização estrutural desse gene fora do gênero *Drosophila*, tanto das isoformas quanto da organização genômica. Também foi demonstrado pela primeira vez que esse gene é expresso em cabeça de machos e fêmeas, pois antes só havia sido descrito em tecidos reprodutivos. A expressão nesse novo sítio está associada à utilização de um novo promotor ou sítio de início da transcrição regulado epigeneticamente, visto que a isoforma encontrada em tecidos cefálicos apresenta uma região 5' UTR alternativa, diferente das dos tecidos reprodutivos. Esses dados podem ser aplicados para testar se *exu* é um bom marcador molecular para distinguir as espécies do grupo *fraterculus*, assim como alvo para silenciamento na técnica do inseto estéril.

CONCLUSÕES

- O gene *exu* é expresso em tecidos reprodutivos em *A. fraterculus*, existindo isoformas específicas de cada sexo que diferem entre si apenas nas regiões não traduzidas, assim como ocorre em *Drosophila*.
- A identificação de transcritos em tecidos cefálicos de machos e fêmeas expande os sítios de expressão desse gene em adultos, embora ainda não se saiba que função ele desempenhe nesses tecidos.
- A presença de um novo éxon 5' UTR associado a esse novo sítio de expressão indica a existência e utilização de um terceiro promotor ou novo sítio de início da transcrição regulado epigeneticamente, que em adultos é exclusivo de tecidos somáticos.

- A organização estrutural de *exu* em *A. fraterculus* é semelhante àquela de *D. melanogaster*, embora tenha adquirido alguns íntrons e sofrido uma modificação na organização dos éxons 5' UTR de reprodutivo de macho e de fêmea.

CAPÍTULO II: EVIDÊNCIAS DE SELEÇÃO POSITIVA NO GENE *exuperantia* EM CYCLORRHAPHA (DIPTERA)

Resumo: Cyclorrhapha constitui um grupo derivado dentro da ordem Diptera, e inclui diversos dípteros comuns, como as moscas-das-frutas do gênero *Anastrepha* e as moscas de frutas do gênero *Drosophila*. Por muito tempo, Cyclorrhapha foi considerado um grupo controverso em relação à filogenia, mas dados moleculares e morfológicos demonstram que trata-se de um grupo monofilético que sofreu uma grande radiação adaptativa, ocupando vários nichos. Embora tenham sido de grande importância, as mudanças moleculares que participaram desse processo permanecem bastante obscuras. Alguns estudos têm demonstrado que a forma como é estabelecido o eixo anteroposterior, assim como os componentes que participam desse processo, podem ajudar a entender como se deu a evolução desse grupo. O gene *bicoid* (*bcd*), que é exclusivo de Cyclorrhapha, está envolvido com a determinação da região anterior do ovo e seu surgimento nessa linhagem tem sido associado a um desenvolvimento embrionário mais acelerado. Assim, genes que estão envolvidos com o estabelecimento desse eixo, em especial os que estão mais diretamente associados a *bcd*, são bons candidatos para estudos evolutivos nesse grupo de moscas. O gene *exuperantia* (*exu*) é fundamental na localização do mRNA *bcd* na parte anterior do ovo, sem a qual o embrião não consegue desenvolver estruturas dessa região. Assim, mudanças adaptativas em *exu* podem ajudar a entender não só como se deu a evolução do gene *bcd*, mas também das principais mudanças no grupo Cyclorrhapha. Uma das formas pelas quais essas mudanças podem ocorrer é através da fixação de novas variantes por meio da seleção natural, processo conhecido como seleção positiva direcional. Dessa forma, esse trabalho se propôs a testar a hipótese de que *exu* passou por um processo de seleção positiva em Cyclorrhapha. Utilizando análises que levam em consideração a relação dN/dS (taxa de substituições não-sinônimas/taxa de substituições sinônimas), nossos resultados corroboram essa hipótese, indicando que pelo menos dois sítios estão sob seleção positiva nesse grupo. Essas mudanças adaptativas em *exu* podem ter ocorrido em conjunto com outras modificações em genes que participam da determinação do eixo anteroposterior em Cyclorrhapha, permitindo a integração de *bicoid* nesse processo, resultando em adaptações com papéis importantes para o sucesso evolutivo desse grupo de moscas.

Palavras-chave: Cyclorrhapha; Seleção positiva; *exuperantia*.

INTRODUÇÃO

A subordem Brachycera (Diptera) e o grupo Cyclorrhapha

A subordem Brachycera constitui um grupo monofilético dentro da ordem Diptera, cujas espécies compartilham sinapomorfias como (I) redução dos flagelômeros antenais para oito ou menos e (II) as veias CuA_2 e A_1 com ápices aproximados formando uma célula fechada. As larvas da maioria das espécies de Brachycera vivem em habitats terrestres úmidos, e os adultos são mais robustos e compactos do que os Diptera mais primitivos (YEATES; WIEGMANN, 1999).

Cyclorrhapha, por sua vez, constitui um grupo derivado em Diptera, inserido na subordem Brachycera, cujas espécies compartilham características como redução da cabeça no estágio larval, modificações das estruturas de alimentação da larva, rotação de 360° na terminália do macho, mas, principalmente, o desenvolvimento de um pupário, formado pelo endurecimento da cutícula da larva de último ínstar, dentro do qual a pupa se desenvolve (YEATES; WIEGMANN, 1999; YEATES et al., 2007). Esse grupo contém cerca de 105 famílias e 72.000 espécies (GRIMALDI; ENGEL, 2005), incluindo a grande maioria das espécies de Diptera, entre elas as moscas de frutas *Drosophila melanogaster*, as moscas-das-frutas *Anastrepha fraterculus*, e a mosca doméstica *Musca domestica*. O grupo Cyclorrhapha sofreu uma grande radiação que permitiu a ocupação de diversos nichos, especialmente pelas larvas, que podem apresentar hábitos que vão desde sarcofagia e fitofagia até parasitismo e predação (WIEGMANN; MITTER; THOMPSON, 1993; KUTTY et al., 2007).

Publicações mais antigas sobre taxonomia e filogenia de Diptera trazem informações conflitantes a respeito das relações do grupo Cyclorrhapha com os outros grupos dessa ordem. Algumas vezes ele é tratado como uma subordem, ora ao lado de Orthorrhapha (que continha os grupos Nematocera e Brachycera) (COMSTOCK, 1940), outras vezes ao lado de Nematocera e Brachycera (BORROR; DELONG, 1964). Entretanto, estudos mais recentes de filogenia que incluem dados morfológicos, moleculares e de expressão gênica, reconhecem esse grupo como monofilético e parte da subordem Brachycera (WIEGMANN et al., 2011).

Mudanças moleculares adaptativas no grupo Cyclorrhapha: componentes do sistema nervoso e da determinação do eixo anteroposterior

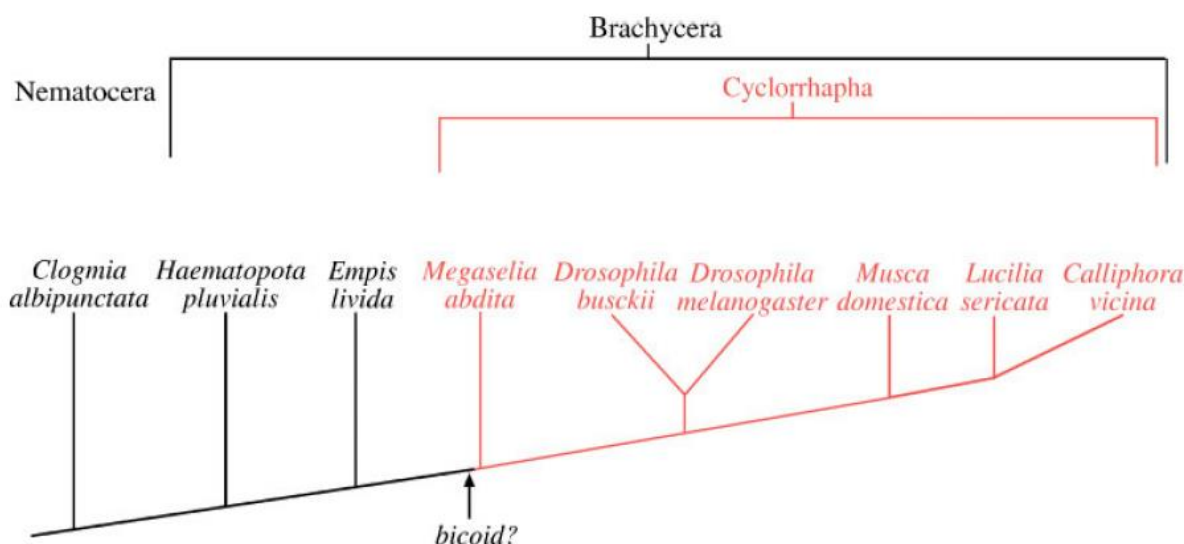
O grande sucesso das espécies do grupo Cyclorrhapha está ligado a algumas adaptações fundamentais, como por exemplo o aumento na velocidade do voo em adultos (MCALPINE; WOOD, 1989). Em relação a essa característica, o gene que codifica a acetilcolinesterase (AChE) sináptica, que interrompe a neurotransmissão das sinapses sensoriais, parece ter tido um papel importante. Esse gene está duplicado em várias espécies de Diptera, que apresentam *ace-1* e *ace-2*, sendo que a condição ancestral é de que a enzima funcional é codificada pela primeira cópia (*ace-1*). No entanto, em Cyclorrhapha existe apenas *ace-2*, que codifica a enzima ativa nesse processo. A explicação mais provável é que no ancestral desse grupo a cópia *ace-2* tenha sofrido alterações que trouxeram alguma vantagem em relação a *ace-1*, tomando parte nessa função. Como alterações nessa enzima levam a diferenças na velocidade da atividade sináptica, tem sido sugerido que variantes que aumentem essa velocidade possam resultar em um voo mais rápido, além de promover outras adaptações relacionadas ao sistema nervoso (HUCHARD et al., 2006).

Apesar de trabalhos que relacionam características estruturais/comportamentais do grupo Cyclorrhapha com os aspectos moleculares envolvidos serem bem escassos na literatura, estudos na área da biologia do desenvolvimento têm trazido informações relevantes que podem ajudar a entender a relação entre genes e seus padrões de expressão com o desenvolvimento de estruturas morfológicas. Nesse contexto, genes que estão envolvidos com a formação de padrões são alvos interessantes. Em particular, o estabelecimento do eixo anteroposterior em Cyclorrhapha apresenta características únicas, as quais podem ter sido fundamentais para o sucesso evolutivo desse grupo.

Em *Drosophila*, *bicoid* (*bcd*) é um gene chave na determinação do eixo anteroposterior. A proteína BCD age como um morfógeno que participa da regulação transcricional e traducional de diversos genes no embrião. A distribuição dessa proteína ocorre como um gradiente de alta concentração na região anterior que é difundido até 2/3 do embrião. Para que esse processo ocorra de forma correta, é essencial que o mRNA *bcd* esteja localizado na região anterior do ovócito (BERLETH et al., 1988; DRIEVER; NÜSSLEIN-VOLHARD, 1988; EPHRUSSI; ST JOHNSTON, 2004). Diversos trabalhos têm demonstrado que *bicoid* é um gene exclusivo de Cyclorrhapha (Figura 2.1) (MCGREGOR, 2005), e que seu surgimento

ocorreu partir da duplicação de *Hox3*, cujos parálogos deram origem não só a *bcd* (de expressão materna), mas também ao gene zigótico *zerknüllt* (*zen*) (STAUBER; JACKLE; SCHMIDT-OTT, 1999). *Hox3*, presente em diversos insetos, com exceção de Cyclorrhapha, apresenta um padrão de expressão interessante: é expresso tanto zigoticamente em tecidos extraembrionários (assim como *zen*), quanto maternalmente nos trofócitos e no ovócito (assim como *bcd*), embora nesse último caso a localização não seja restrita à região anterior como ocorre com *bcd* (STAUBER; PRELL; SCHMIDT-OTT, 2002). Isso sugere que após a duplicação de *Hox3*, cada parálogo adquiriu uma função específica que está um pouco relacionada ao gene que os deu origem, resultando nos genes *bicoid* e *zerknüllt* (MCGREGOR, 2005).

Figura 2.1 – Relações filogenéticas entre espécies de Diptera e o surgimento de *bicoid*



Legenda: *bicoid* é compartilhado por espécies do grupo Cyclorrhapha, enquanto parece estar ausente em grupos de dípteros mais basais (Brachycera mais basais e Nematocera). Seu surgimento a partir da duplicação de *Hox3* parece ter ocorrido em alguma linhagem ancestral de Cyclorrhapha, antes da radiação desse grupo. Fonte: McGregor, 2005 (Adaptado).

Agindo como um fator de transcrição, BCD regula a expressão de cerca de 20 genes (OCHOA-ESPINOSA et al., 2005). Entre eles, a expressão zigótica do gene *hunchback* (*hb*), resultando em uma distribuição da proteína HUNCHBACK (HB) na metade anterior do zigoto (DRIEVER; NÜSSLEIN-VOLHARD, 1988). Por meio de uma série de construções com o gene *hb* de diversas espécies de Diptera (dentro e fora do grupo Cyclorrhapha) e o gene repórter *lacZ*, Lemke et al. (2008) demonstraram que o padrão de expressão desse gene induzido por BCD ocorre apenas nas espécies de Cyclorrhapha. Os resultados desse trabalho como um todo (incluindo alguns que não foram discutidos aqui) trazem evidências de que não

só o gene *bcd*, mas também a regulação transcricional do gene *hunchback* dependente de BCD são sinapomorfias do grupo Cyclorrhapha.

Essas mudanças observadas em *hb*, um gene que está sob regulação de BCD, ajudam a entender como se deu a evolução de *bicoid* e como em Cyclorrhapha esse novo gene foi sendo integrado a uma rede de interações já existente. Diante disso, outras mudanças em genes que estão sob controle transcricional/traducional dessa proteína, assim como aqueles que participam desde a regulação do próprio *bcd* até a localização de seu RNA mensageiro na região anterior do ovócito, também podem ser muito informativas. Nesse último caso, sabe-se que o gene *exuperantia* (*exu*) desempenha uma função primordial, visto que mutantes para esse gene apresentam uma distribuição difusa do mRNA *bcd* no ovócito em *Drosophila* (BERLETH et al., 1988). O gene *exu* é filogeneticamente mais antigo que *bcd*, pois está amplamente distribuído em espécies de Diptera, e até mesmo em outras ordens de insetos (observação em bancos de dados). Como essa função de localização do mRNA *bcd* é evolutivamente mais recente, visto que *bcd* só ocorre em Cyclorrhapha, não se sabe qual(is) outra(s) função(ões) *exu* realiza em espécies fora desse grupo, embora possa(m) estar relacionada(s) ao seu papel na espermatogênese (HAZELRIGG et al., 1990) ou na localização dos mRNAs de *oskar* durante a ovogênese (WILHELM et al., 2000). De qualquer forma, espera-se que o gene *exu* tenha passado por mudanças adaptativas que permitiram sua participação na localização do mRNA *bcd* no grupo Cyclorrhapha. Essas mudanças podem ter ocorrido na região promotora, afetando seu padrão de expressão, e/ou na sequência da região codificadora do gene, através da fixação de novas variantes por meio de seleção positiva.

Uma das estratégias mais utilizadas para se estudar a evolução molecular de genes baseia-se na relação entre a taxa de substituições não-sinônimas (dN) e de substituições sinônimas (dS), resultando em um valor ω (dN/dS). Essa relação pode assumir um valor menor que um ($\omega < 1$), indicando que o gene (ou regiões dele) está sob seleção purificadora; igual a 1 ($\omega = 1$) indicando uma evolução neutra; ou maior que 1 ($\omega > 1$), indicando que o gene está sob seleção positiva. Como se trata de relações de substituições de aminoácidos, essas análises tem como unidade fundamental o códon, avaliando se alterações neste resultam ou não em mudanças na sequência proteica.

Esse trabalho se propôs a testar a hipótese de que há seleção positiva atuando no gene *exu*, especificamente em sua região codificadora, e que isso pode ter um papel importante na

determinação dos padrões de expressão de *bicoid* e os processos básicos de formação de padrão no grupo Cyclorrhapha. Assim, os objetivos desse trabalho foram:

- Investigar se há seleção positiva atuando no gene *exu* em Cyclorrhapha.
- Identificar quais os sítios que estão sendo selecionados positivamente nessa linhagem, se houver evidência de evolução adaptativa nesse gene.

Utilizando uma análise na relação dN/dS, seguido pela aplicação de um método bayesiano empírico em um alinhamento de regiões codificadoras para o gene *exu* no grupo Cyclorrhapha, e em diversos grupos externos, neste trabalho investigamos a existência de seleção positiva que possa estar associada às alterações que permitiram a participação de *exu* na função de localização do mRNA *bicoid* na região anterior do ovócito nesse grupo.

MATERIAIS E MÉTODOS

A fim de testar a hipótese de que o gene *exu* estaria sujeito à seleção positiva no grupo Cyclorrhapha (Diptera), obtivemos a região codificadora desse gene em várias espécies de moscas disponíveis no banco de dados GenBank do NCBI (*National Center for Biotechnology Information*) utilizando a ferramenta BLAST (ALTSCHUL et al., 1997). Além disso, utilizamos a sequência do gene *exu* da espécie de moscas-das-frutas *A. fraterculus*, obtida na primeira etapa desse trabalho. A Tabela 2.1 mostra as espécies estudadas e a identificação de acesso no banco de dados. Nos casos em que a região codificadora do transcrito não estava indicada na própria página do GenBank, foi feita uma inferência *in silico* da ORF mais provável e de sua sequência proteica (BIKANDI et al., 2004).

As regiões codificadoras das sequências correspondentes ao gene *exuperantia* foram inicialmente alinhadas por aminoácidos utilizando a versão *online* do software MAFFT (KATO et al., 2002), retirando-se as regiões que apresentaram muita divergência para um alinhamento confiável, e depois manualmente revertidas para suas respectivas sequências de DNA. Esse alinhamento do DNA por códon foi utilizado em todas as análises descritas a seguir.

Tabela 2.1: Lista de espécies estudadas nesse trabalho e respectivas entradas de acesso no GenBank

Grupo/Espécie	ID de acesso no GenBank
Cyclorrhapha	
<i>Anastrepha fraterculus</i> (Afrat)*	---
<i>Ceratitis capitata</i> (Ccap)	XM_004526429.1 GI:498969170
<i>Musca domestica</i> (Mdom)	KA645318.1 GI:399939025
<i>Drosophila melanogaster</i> (Dmel)	NM_001169747.1 GI:281363853
<i>Drosophila mojavensis</i> (Dmoj)	XM_002004251.1 GI:195119535
<i>Drosophila virilis</i> (Dvir)	XM_002059408.1 GI:195401686
<i>Drosophila pseudoobscura</i> (Dps1)	XM_001360949.2 GI:198458310
Grupo externo (Nematocera)	
<i>Sitodiplosis mosellana</i> (Smos)	GAKJ01009913.1 GI:511202845
<i>Corethrella appendiculata</i> (Cape)	GANO01001921.1 GI:545918536
<i>Aedes aegypti</i> (Aaeg)	XM_001654190.1 GI:157125158
<i>Anopheles gambiae</i> (Agam)	XM_308467.4 GI:158285796
<i>Phlebotomus papatasi</i> (Ppap)	JP555152.1 GI:353485238
<i>Chironomus riparius</i> (Crip)	KA182541.1 GI:400999303
<i>Belgica antartica</i> (Bant)	GAAK01003495.1 GI:418272270

* Sequência recentemente obtida e ainda não depositada no GenBank.

Fonte: Elaborada pela autora

O alinhamento por códons da região de *exu* investigada foi analisado no jModelTest 2.1.1 (DARRIBA et al., 2012; POSADA; CRANDALL, 1998) para a inferência do modelo evolutivo mais provável pelo critério de informação do Akaike (AIC). Essas informações, juntamente com o alinhamento, foram utilizados nos passos seguintes. Nós estimamos os níveis de saturação das sequências em estudo, as taxas de transição e transversão no programa DAMBE 5.3.74 (XIA, 2013), que plota tais valores contra as distâncias genéticas considerando o modelo evolutivo GTR.

Uma estimativa do nível de divergência de *exu* entre as espécies estudadas foi conduzida no programa DnaSP (LIBRADO; ROZAS, 2009), na qual foram avaliadas as taxas de divergência em dois níveis: dentro do grupo Cyclorrhapha e entre esse grupo e outras espécies de Diptera, considerando uma *sliding-window* de 30 nucleotídeos, sem sobreposição e descartando-se os *gaps*.

A árvore filogenética não-enraizada de máxima verossimilhança de *exu* foi inferida no programa PhyML 3.0 (GUINDON et al., 2010), considerando o alinhamento obtido e as taxas

de substituições do modelo evolutivo GTR previamente inferidos. Estes dados foram utilizados no software CodeML implementado no programa PAML 4.7 (YANG, 2007) para a realização do *strict branch-site test* (ZHANG; NIELSEN; YANG, 2005). Nesse tipo de análise, são realizadas comparações de modelos que permitem detectar seleção positiva agindo em sítios distintos em linhagens previamente especificadas (YANG; NIELSEN, 2002), considerando-se a história evolutiva. Para isso, a árvore gerada no PhyML é dividida em um ramo *foreground* contendo as espécies de Cyclorrhapha, e um ramo *background* contendo as outras espécies de Diptera para a realização do teste de razão de máxima verossimilhança (*Likelihood Ratio Test* – LRT) que compara o modelo nulo que não permite valores de $\omega > 1$ em nenhum dos ramos (MA *null*), contra um modelo alternativo (MA) que o permite apenas no ramo *foreground* (Cyclorrhapha). Os valores de LRT obtidos foram comparados com a distribuição nula de qui-quadrado (χ^2_1), com valores críticos de 2,71 e 5,41 a níveis de significância de 0,05 e 0,01, respectivamente, conforme sugerido em trabalhos prévios (YANG, 2007).

Após verificar sinais de seleção positiva em *exu* no grupo Cyclorrhapha, o método Bayes Empirical Bayes (BEB) foi aplicado para identificar quais sítios estão evoluindo sob esse tipo de seleção, calculando as probabilidades posteriores de seleção purificadora, evolução neutra e seleção positiva para cada códon. O BEB utiliza as estimativas dos parâmetros calculados para o *strict branch-site test* por máxima verossimilhança (como por exemplo, as taxas de ω para cada sítio), para construir a distribuição *a priori*. Esse método constitui uma versão mais aprimorada do método Naïve Empirical Bayes (NEB), pois adiciona correções para erros de amostragem dos parâmetros gerados por estimativas de máxima verossimilhança (o que é importante quando o conjunto de dados analisados é pequeno) (YANG; WONG; NIELSEN, 2005). Para identificar os sítios (códon) sob seleção positiva no ramo *foreground*, foi utilizado um *threshold* de $P = 0,95$ para as probabilidades posteriores de $\omega > 1$.

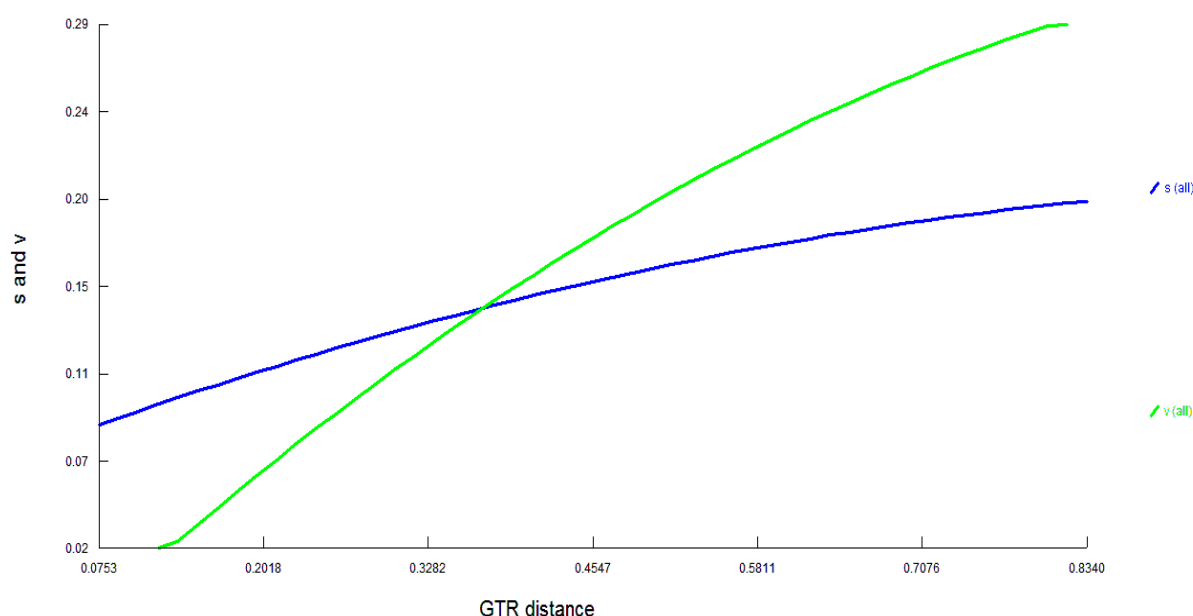
RESULTADOS E DISCUSSÃO

Alinhamento das sequências, análises iniciais e árvore filogenética de *exuperantia*

O alinhamento por códon da região selecionada de *exu* com sequências de 14 espécies, sendo sete do grupo Cyclorrhapha e sete de outros Diptera não-Cyclorrhapha, gerou um arquivo com um total de 571 sítios (códon). Esse alinhamento pode ser visto no Anexo C.

A análise das sequências no jModelTest 2.1.1 (DARRIBA et al., 2012) demonstrou que o modelo GTR (*Generalised Time-Reversible*) + I + G (-lnL = 15659,6433) é o que melhor representa a evolução das sequências em estudo. Com os dados do alinhamento e do modelo evolutivo, foram plotadas as taxas de transição e transversão contra as distâncias genéticas entre as sequências no programa DAMBE 5.3.74 (XIA, 2013). Os resultados demonstram que essas taxas de substituição ainda não atingiram o nível de saturação (Figura 2.2), descartando a possibilidade de que as taxas de substituições sinônimas (dS) estejam sendo subestimadas. Esse tipo de análise é importante porque quando essas taxas atingem a saturação (visualizado por um platô no gráfico) o valor de dS pode ser subestimado, que por sua vez pode resultar em um valor de ω inflado, gerando um falso positivo (SMITH; SMITH, 1996).

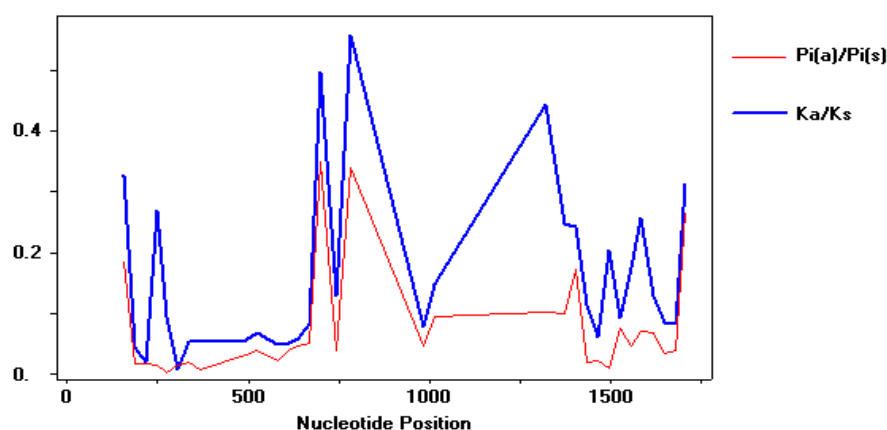
Figura 2.2 – Gráfico de saturação de *exu*: taxas de mutação x distâncias genéticas



Legenda: Taxas de transição (s; em azul) e transversão (v; em verde) contra as distâncias genéticas segundo o modelo GTR para a região selecionada de *exu*. Fonte: Elaborada pela autora.

A estimativa dos níveis de divergência entre as sequências demonstrou que a maior parte da variação encontrada ao longo do gene é melhor explicada pelas diferenças entre os grupos (Cyclorrhapha e o grupo externo, medida pela relação Ka/Ks) do que pelas diferenças dentro do grupo Cyclorrhapha (medida pela relação $Pi(a)/Pi(s)$) (Figura 2.3). Isso sugere que as espécies desse grupo sofreram um processo evolutivo diferente em relação às outras espécies de Diptera, que permitiu a fixação de variantes específicas para esse grupo, possivelmente por seleção positiva.

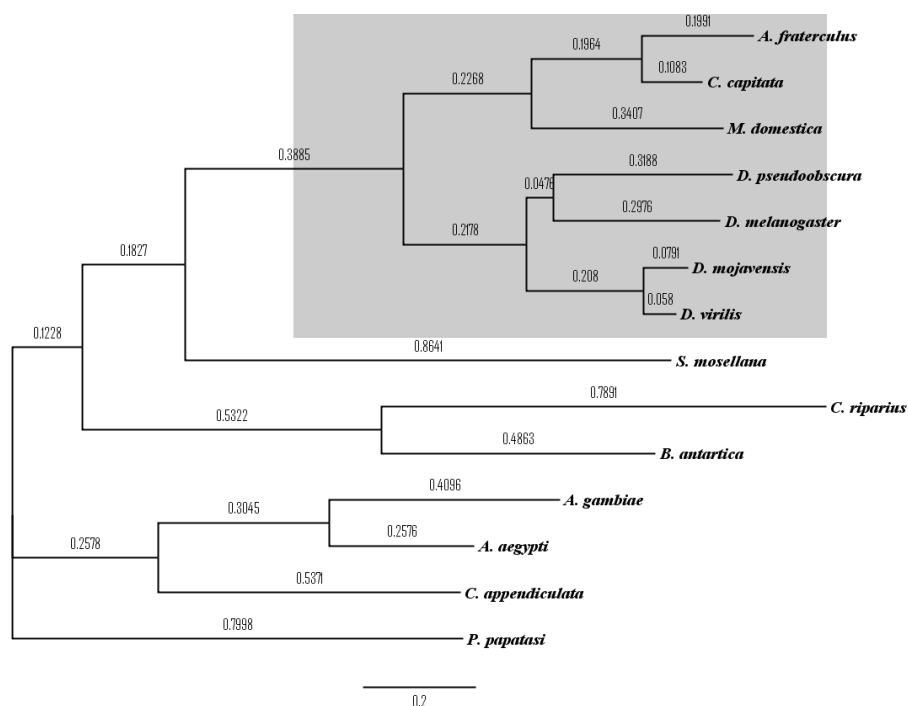
Figura 2.3 – Estimativa dos níveis de divergência de *exu* no grupo Cyclorrhapha e entre este e outros Diptera



Legenda: Níveis de divergência de *exu* dentro do grupo Cyclorrhapha (Pi(a)/Pi(s), vermelho) e entre esse grupo e outras espécies de Diptera (Ka/Ks, azul). Fonte: Elaborada pela autora

A árvore não-enraizada de máxima verossimilhança de *exu* (Figura 2.4), gerada no programa PhyML 3.0 (GUINDON et al., 2010), foi então utilizada nos testes de seleção, a fim de testar a hipótese de seleção positiva atuando sobre esse gene no grupo Cyclorrhapha. Segundo o manual do programa PAML, esse tipo de árvore (não-enraizada) deve ser utilizada nos casos em que não forem fornecidas informações sobre relógio molecular, pois permite que as taxas variem livremente de ramo para ramo.

Figura 2.4 – Árvore filogenética não-enraizada de máxima verossimilhança do gene *exuperantia*



Legenda: O ramo *foreground* testado está marcado em cinza. Os valores de cada ramo correspondem ao número de substituições de nucleotídeos por sítio de substituição. Fonte: Elaborada pela autora.

Estaria *exu* evoluindo sob seleção positiva no grupo Cyclorrhapha?

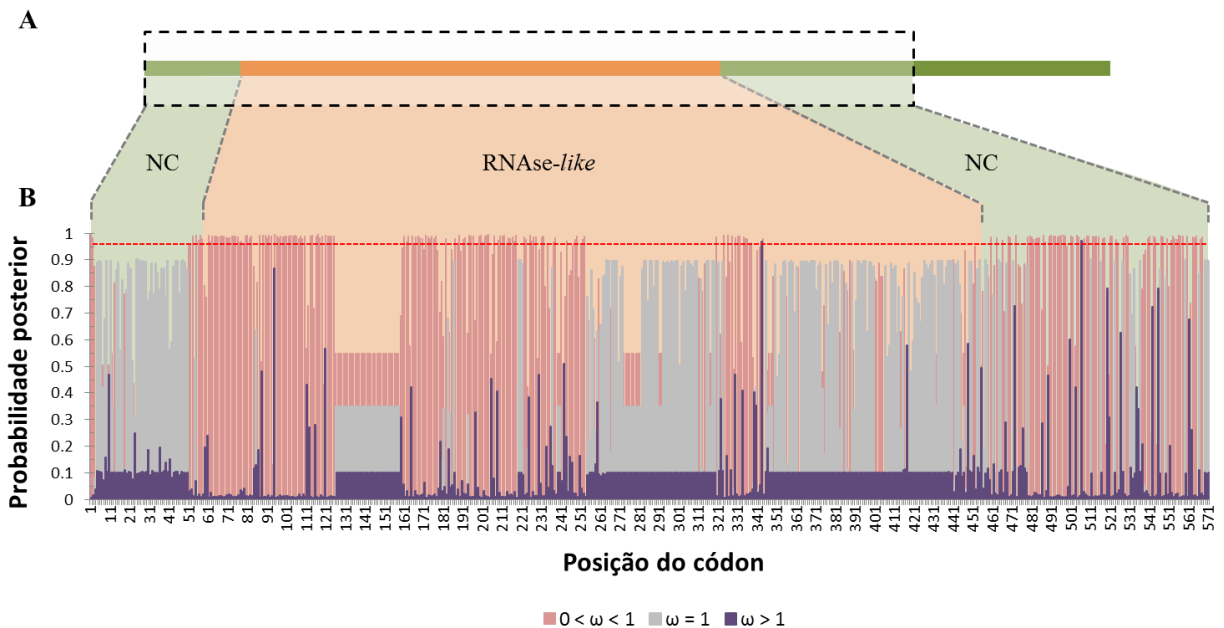
Os resultados do *strict branch-site test* demonstraram que o modelo alternativo (MA), no qual permite-se que os códons no ramo *foreground* atinjam valores de ω maiores que 1, explicam melhor os dados do que o modelo nulo, que não o permite (MA *null*), rejeitando-se o modelo nulo com um nível de significância de 0,000034. Isso significa que há seleção positiva atuando no gene *exuperantia* no ramo do grupo Cyclorrhapha.

O *Bayes Empirical Bayes* (BEB) identificou dois sítios (códon) que respondem por um sinal de seleção positiva (Figura 2.5). Um deles é o códon 343 ($P = 0,97$), localizado dentro do domínio semelhante a RNase (RNase-like), e o outro é o 506 ($P = 0,97$), localizado em uma região ainda não caracterizada em relação a domínios de proteína. Os resultados do BEB trazem ainda outro dado interessante: esses dois sítios estão dentro de regiões sob forte seleção purificadora em Cyclorrhapha (Figura 2.5), e no ramo *background* (grupo externo) esses próprios sítios estão sob seleção purificadora (dados não mostrados). Esse tipo de padrão de seleção é muito interessante, visto que regiões sob forte seleção purificadora em geral constituem regiões de grande importância funcional, e o fato de que dentro delas foram encontrados sítios sob seleção positiva sugere que esses novos variantes podem ter sido fixados por trazerem alguma vantagem evolutiva aos indivíduos que os portam (NIELSEN, 2005).

O domínio semelhante à RNase (RNase-like) em EXU foi identificado em *D. melanogaster* utilizando uma abordagem computacional (MOSER et al., 1997), a fim de buscar proteínas que contenham esse domínio em vários organismos. Em relação a *exu*, ainda não se sabe qual o papel desse domínio na função da proteína, mas o fato de que essa região está sob seleção purificadora indica que ela pode desempenhar uma função importante. Se esse domínio estiver permanecendo funcionalmente conservado em EXU, é possível que através de reações catalíticas essa proteína promova a maturação do mRNA *bcd*, o que explicaria resultados prévios de que só após um contato com EXU esse mRNA é “marcado” para uma localização específica na região anterior do ovócito (CHA; KOPPETSCH; THEURKAUF, 2001). Por outro lado, também é possível que esse domínio tenha perdido sua função catalítica nesse gene ao longo da evolução, permanecendo como uma região de ligação ao mRNA, conforme observado em outra proteína de *Drosophila* que o contém (DIENSTBIER et al., 2009). Independente de qual for o caso, uma mudança adaptativa nesse domínio no grupo Cyclorrhapha indica que ela trouxe uma vantagem seletiva, possivelmente relacionada com sua

nova função de localização do mRNA *bicoid*, seja promovendo a maturação desse mRNA ou aumentando sua especificidade de ligação. Infelizmente a proteína EXU ainda não foi cristalizada, o que dificulta a análise de alterações em sua estrutura terciária para tentar inferir se as mudanças aqui investigadas podem de fato ser as mudanças principais envolvidas na modificação de EXU em *Cyclorrhapha*.

Figura 2.5 – Sítios de *exuperantia* sob seleção positiva em *Cyclorrhapha*



Legenda: (A) Região codificadora de *exu* utilizando a sequência de *Anastrepha fraterculus* como referência. A área destacada com um retângulo pontilhado corresponde ao fragmento considerado para análise, o qual contém o domínio RNase-like (em laranja) e duas regiões sem domínios caracterizados (NC, em verde). (B) Probabilidades posteriores de códons sob seleção purificadora ($0 < \omega < 1$), evolução neutra ($\omega = 1$) e seleção positiva ($\omega > 1$). A linha vermelha pontilhada representa o *threshold* utilizado de $P = 0,95$. Fonte: Elaborada pela autora.

Em relação à outra região na qual foi observado um códon sob seleção positiva, ainda não se tem registros a respeito de motivos ou domínios proteicos, mas o fato de estar sob seleção purificadora sugere ser uma região importante da proteína EXU. Além disso, deve ser lembrado que a região final de *exu* não foi considerada na análise por ser muito divergente, e isso pode acontecer tanto porque esta região está evoluindo neutramente, quanto pode ser pela existência de outros sítios sob seleção positiva que evoluíram rapidamente no grupo *Cyclorrhapha*.

Tem sido sugerido que *bicoid* representa um ganho evolutivo por proporcionar um desenvolvimento mais rápido, pois a segmentação de todo o embrião ocorre de uma só vez,

enquanto que há um atraso em espécies de Diptera mais basais. Para que esse desenvolvimento ocorra de forma correta é necessário que o mRNA *bcd* seja localizado na região anterior do ovócito, que vai resultar em um gradiente proteico agindo na regulação transcricional e traducional de diversos genes (MCGREGOR, 2005). Stauber et al. (2002) demonstraram que o gene *Hox3*, a partir do qual *bcd* originou-se após um evento de duplicação, apesar de ser expresso maternalmente em linhagens basais de Diptera, apresenta uma distribuição difusa tanto nas câmaras ovocitárias (trofócitos e ovócito) quanto no embrião. Quando considerados em conjunto, esses dados sugerem que esse mecanismo de localização parece ter se desenvolvido concomitantemente ao estabelecimento de *bicoid* como fator chave na determinação do eixo anteroposterior em alguma linhagem ancestral de Cyclorrhapha. Dessa forma, embora mais estudos se façam necessários, as alterações adaptativas encontradas em *exu* no grupo Cyclorrhapha podem ter sido resultado de uma co-optação desse gene para função de localização de *bicoid* nesse grupo.

CONCLUSÕES

- A hipótese de que o gene *exuperantia* está evoluindo sob seleção no grupo Cyclorrhapha foi corroborada, e esse fato pode ser resultado de uma co-optação desse gene para a localização do mRNA *bicoid* nessa linhagem.
- O desenvolvimento dessa via de localização juntamente com o surgimento do gene *bicoid* podem ter trazido uma vantagem evolutiva no sentido de permitir um desenvolvimento mais rápido, visto que a segmentação ocorre de uma só vez ao longo do embrião.
- Os sítios sob seleção positiva estão localizados dentro de duas regiões sob forte seleção purificadora, uma tratando-se de um domínio já conhecido e a outra ainda não caracterizada em relação a domínios proteicos. Isso indica que essas regiões são importantes na função de EXU, e o fato de conterem sítios sob seleção positiva significa que essas substituições trazem uma vantagem evolutiva para os indivíduos que as portam.

CONSIDERAÇÕES FINAIS

Esse trabalho trouxe várias informações relevantes em relação ao gene *exuperantia*, ampliando o conhecimento desse gene para outras moscas fora do gênero *Drosophila*. Na mosca-das-frutas sul-americana *Anastrepha fraterculus*, caracterizamos estruturalmente o gene e seus transcritos, assim como detectamos a expressão de *exu* em tecidos reprodutivos e cefálicos nos dois sexos. Até o momento ainda não havia sido descrito na literatura que esse gene pode ser expresso em cabeça. Além disso, demonstramos que a expressão nesse novo sítio envolve a utilização de um terceiro promotor ou sítio de início da transcrição regulado epigeneticamente, que é diferente daqueles dos tecidos reprodutivos.

A regulação diferencial de um gene em contextos celulares diferentes (em relação a tecidos, sexos ou fases do desenvolvimento) indica que os elementos envolvidos nessa regulação também são diferentes. No caso específico em que um novo sítio de expressão do gene está relacionado com uma linhagem evolutiva em particular, pode significar uma mudança adaptativa resultante da co-optação do gene para uma nova função. Um exemplo disso é *oskar*, que foi co-optado a partir do sistema nervoso para uma função de organização do polo germinativo em insetos holometábolos (LYNCH et al., 2011). Considerando uma escala evolutiva mais ampla, *exu*, que está presente em várias ordens de insetos e é regulado diferencialmente em pelo menos 3 contextos celulares diferentes em Diptera, pode ser um bom candidato a ser testado. Poderíamos considerar um cenário hipotético no qual em insetos mais basais *exu* fosse expresso apenas em tecidos reprodutivos masculinos, e que em uma certa linhagem (por exemplo, nos holometábolos) ele passasse a ser expresso também em tecidos reprodutivos femininos (ou vice-versa). Trata-se de um cenário meramente hipotético, mas que poderia ser testado a fim de entender como se estabeleceu o padrão de expressão atual de *exu* em insetos mais derivados, tais como *Drosophila* e *Anastrepha*.

Esse tipo de experimento torna-se bastante laborioso por requerer dados de expressão e de estrutura de transcritos em uma larga escala evolutiva, em especial para genes pouco estudados em organismos não-modelo, como *exu*. Em uma escala evolutiva menor, alguns trabalhos têm relacionado modificações em alguns genes com a integração do gene *bicoid* na via de estabelecimento do eixo anteroposterior em Cyclorrhapha, um grupo derivado de Diptera. Esses trabalhos focam principalmente em regiões regulatórias nas quais a proteína BCD se liga, agindo como um fator de transcrição. No entanto, consideramos que genes que

agem em um passo anterior dessa cascata, seja regulando a transcrição de *bcd* ou que participem da localização de seu RNA mensageiro na região anterior do ovócito, também possam ter sofrido mudanças adaptativas. Utilizando uma abordagem de evolução molecular, demonstramos que dois sítios estão evoluindo sob seleção positiva no grupo Cyclorrhapha, podendo ser resultado de uma co-optação de *exu* para essa função de localização de *bcd*.

Para finalizar, levantamos aqui algumas questões que permanecem em aberto a respeito de *exu* em relação aos dados levantados nesse trabalho: (I) qual a função que esse gene desempenha em tecidos cefálicos? (II) esse novo sítio de expressão, assim como o novo transcrito descrito, estão relacionados a quais linhagens evolutivas? (III) é um sítio de expressão evolutivo mais novo, em relação aos tecidos reprodutivos? (IV) a região sob forte seleção purificadora na qual foi encontrado um sítio sob seleção positiva, trata-se de um motivo/domínio funcional ainda não caracterizado? (V) outros genes que agem “*upstream*” ao gene *bicoid* na determinação do eixo anteroposterior em Cyclorrhapha, também apresentam mudanças adaptativas que podem estar associadas a sua relação com *bcd*? Assim, esse trabalho não só ajudou a entender um pouco mais sobre o gene *exu*, mas também abre as portas para novos trabalhos a serem realizados no futuro.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALTSCHUL, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research**, v. 25, n. 17, p. 3389–3402, set. 1997.
- ARAÚJO, E. L.; ZUCCHI, R. A. Measurement of the aculeus for the characterization of five *Anastrepha* species of the *fraterculus* group (Diptera: Tephritidae). **Neotropical Entomology**, v. 35, n. 3, p. 329–337, jun. 2006.
- BERLETH, T. et al. The role of localization of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo. **The EMBO Journal**, v. 7, n. 6, p. 1749–1756, jun. 1988.
- BIKANDI, J. et al. *In silico* analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction. **Bioinformatics**, v. 20, n. 5, p. 798–799, 22 mar. 2004.
- BORROR, D. J.; DELONG, D. M. **An introduction to the study of insects**. [s.l: s.n.].
- CHA, B.-J.; KOPPETSCH, B. S.; THEURKAUF, W. E. *In vivo* analysis of *Drosophila bicoid* mRNA localization reveals a novel microtubule-dependent axis specification pathway. **Cell**, v. 106, n. 1, p. 35–46, jul. 2001.
- CHOMCZYNSKI, P.; MACKEY, K. Short technical reports. Modification of the TRI reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. **BioTechniques**, v. 19, n. 6, p. 942–945, dez. 1995.
- CIVETTA, A. et al. Rapid evolution and gene-specific patterns of selection for three genes of spermatogenesis in *Drosophila*. **Molecular Biology and Evolution**, v. 23, n. 3, p. 655–662, jan. 2006.
- COMSTOCK, J. H. **Introduction to Entomology**. [s.l: s.n.].
- CROWLEY, T. E.; HAZELRIGG, T. A male-specific 3'-UTR regulates the steady-state level of the *exuperantia* mRNA during spermatogenesis in *Drosophila*. **Molecular and General Genetics MGG**, v. 248, n. 3, p. 370–374, ago. 1995.
- DALLAS-YANG, Q.; JIANG, G.; SLADEK, F. M. Avoiding false positives in colony PCR. **BioTechniques**, v. 24, n. 4, p. 580–582, abr. 1998.
- DARRIBA, D. et al. jModelTest 2: more models, new heuristics and parallel computing. **Nature Methods**, v. 9, n. 8, p. 772–772, ago. 2012.
- DIENSTBIER, M. et al. Egalitarian is a selective RNA-binding protein linking mRNA localization signals to the dynein motor. **Genes & development**, v. 23, n. 13, p. 1546–1558, jul. 2009.
- DRIEVER, W.; NÜSSLEIN-VOLHARD, C. A gradient of bicoid protein in *Drosophila* embryos. **Cell**, v. 54, n. 1, p. 83–93, 1 jul. 1988.
- DUARTE, A. L.; MALAVASI, A. Tratamentos quarentenários. In: MALAVASI, A.; ZUCCHI, R. A. (Eds.). Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado. [s.l.] Malavasi A. e Zucchi R. A., 2000. p. 93–98.

- EPHRUSSI, A.; LEHMANN, R. Induction of germ cell formation by *oskar*. **Nature**, v. 358, n. 6385, p. 387–392, jul. 1992.
- EPHRUSSI, A.; ST JOHNSTON, D. Seeing is believing: the bicoid morphogen gradient matures. **Cell**, v. 116, n. 2, p. 143–152, jan. 2004.
- GONÇALVES, V. R. et al. Transcriptome analysis of female reproductive tissues of *Anastrepha obliqua* and molecular evolution of eggshell proteins in the fraterculus group. **Insect Molecular Biology**, v. 22, n. 5, p. 551–561, 2013.
- GRABHERR, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nature Biotechnology**, v. 29, n. 7, p. 644–652, maio 2011.
- GRIMALDI, D. A.; ENGEL, M. S. **Evolution of the Insects**. [s.l.: s.n.].
- GUINDON, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. **Systematic Biology**, v. 59, n. 3, p. 307–321, maio 2010.
- HALL, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symposium Series**, v. 41, p. 95–98, 1999.
- HAZELRIGG, T. et al. The *exuperantia* gene is required for *Drosophila* spermatogenesis as well as anteroposterior polarity of the developing oocyte, and encodes overlapping sex-specific transcripts. **Genetics**, v. 126, n. 3, p. 607–617, 1990.
- HAZELRIGG, T.; TU, C. Sex-specific processing of the *Drosophila exuperantia* transcript is regulated in male germ cells by the *tra-2* gene. **Proceedings of the National Academy of Sciences**, v. 91, n. 22, p. 10752–10756, out. 1994.
- HUCHARD, E. et al. Acetylcholinesterase genes within the Diptera: takeover and loss in true flies. **Proceedings Biological sciences / The Royal Society**, v. 273, n. 1601, p. 2595–2604, out. 2006.
- KATOH, K. et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic Acids Research**, v. 30, n. 14, p. 3059–3066, jul. 2002.
- KUTTY, S. N. et al. Sensitivity analysis, molecular systematics and natural history evolution of Scathophagidae (Diptera: Cyclorrhapha: Calyptratae). **Cladistics**, v. 23, n. 1, p. 64–83, fev. 2007.
- LEMKE, S. et al. bicoid occurrence and Bicoid-dependent hunchback regulation in lower cyclorrhaphan flies. **Evolution & Development**, v. 10, n. 4, p. 413–420, 2008.
- LIBRADO, P.; ROZAS, J. DnaSP V5: A Software for Comprehensive Analysis of DNA Polymorphism Data. **Bioinformatics**, v. 25, n. 11, p. 1451–1452, jan. 2009.
- LIS, J. T.; SCHLEIF, R. Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. **Nucleic Acids Research**, v. 2, n. 3, p. 383–389, mar. 1975.

- LYNCH, J. A. et al. The phylogenetic origin of *oskar* coincided with the origin of maternally provisioned germ plasm and pole cells at the base of the Holometabola. **PLoS Genetics**, v. 7, n. 4, p. 1-16, abr. 2011.
- MACDONALD, P. M.; LUK, S. K.; KILPATRICK, M. Protein encoded by the *exuperantia* gene is concentrated at sites of *bicoid* mRNA accumulation in *Drosophila* nurse cells but not in oocytes or embryos. **Genes & Development**, v. 5, n. 12b, p. 2455–2466, dez. 1991.
- MALAVASI, A.; ZUCCHI, R. A.; SUGAYAMA, R. L. Biogeografia. In: MALAVASI, A.; ZUCCHI, R. A. (Eds.). Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado. [s.l.] Malavasi A. e Zucchi R. A., 2000. p. 93–98.
- MARCEY, D.; WATKINS, W. S.; HAZELRIGG, T. The temporal and spatial distribution pattern of maternal *exuperantia* protein: evidence for a role in establishment but not maintenance of *bicoid* mRNA localization. **The EMBO Journal**, v. 10, n. 13, p. 4259, 1991.
- MCALPINE, J. F.; WOOD, D. M. **Manual of Nearctic Diptera**. Ottawa, Canada: [s.n.].
- MCGREGOR, A. P. How to get ahead: the origin, evolution and function of *bicoid*. **BioEssays: news and reviews in molecular, cellular and developmental biology**, v. 27, n. 9, p. 904–913, set. 2005.
- MOSER, M. J. et al. The proofreading domain of *Escherichia coli* DNA polymerase I and other DNA and/or RNA exonuclease domains. **Nucleic Acids Research**, v. 25, n. 24, p. 5110–5118, jan. 1997.
- NAVARRO, C. et al. Egalitarian binds dynein light chain to establish oocyte polarity and maintain oocyte fate. **Nature cell biology**, v. 6, n. 5, p. 427–435, maio 2004.
- NELSON, J. E.; KRAWETZ, S. A. Purification of cloned and genomic DNA by guanidine thiocyanate/isobutyl alcohol fractionation. **Analytical Biochemistry**, v. 207, n. 1, p. 197–201, nov. 1992.
- NIELSEN, R. Molecular signatures of natural selection. **Annual Review of Genetics**, v. 39, p. 197–218, 2005.
- NORRBOM, A. L.; ZUCCHI, R. A.; HERNÁNDEZ-ORTIZ, V. Phylogeny of the genera *Anastrepha* and *Toxotrypana* (Trypetinae: Toxotripanini) based on morphology. In: Fruit flies (Tephritidae): phylogeny and evolution of behavior. [s.l.] Norrbom, A. L. e Aluja, M., 1999. p. 299–342.
- OCHOA-ESPINOSA, A. et al. The role of binding site cluster strength in Bicoid-dependent patterning in *Drosophila*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 14, p. 4960–4965, abr. 2005.
- POKRYWKA, N. J.; STEPHENSON, E. C. Microtubules mediate the localization of *bicoid* RNA during *Drosophila* oogenesis. **Development (Cambridge, England)**, v. 113, n. 1, p. 55–66, set. 1991.
- POSADA, D.; CRANDALL, K. A. MODELTEST: Testing the model of DNA substitution. **Bioinformatics**, v. 14, n. 9, p. 817–818, jan. 1998.

- RIECHMANN, V.; EPHRUSSI, A. Par-1 regulates *bicoid* mRNA localisation by phosphorylating Exuperantia. **Development (Cambridge, England)**, v. 131, n. 23, p. 5897–5907, dez. 2004.
- ROBERTS, S. B.; GAVERY, M. R. Is there a relationship between DNA methylation and phenotypic plasticity in invertebrates? **Frontiers in Physiology**, v. 2, jan. 2012.
- SCHÜPBACH, T.; WIESCHAUS, E. Maternal-effect mutations altering the anterior-posterior pattern of the *Drosophila* embryo. **Development Genes and Evolution**, v. 195, n. 5, p. 302–317, 1986.
- SMITH, J. M.; SMITH, N. H. Synonymous nucleotide divergence: what is “saturation”? **Genetics**, v. 142, n. 3, p. 1033–1036, jan. 1996.
- SOBRINHO, I. S.; DE BRITO, R. A. DE. Evidence for positive selection in the gene *fruitless* in *Anastrepha* fruit flies. **BMC Evolutionary Biology**, v. 10, n. 1, p. 293, set. 2010.
- SOBRINHO, I. S.; DE BRITO, R. A. Positive and purifying selection influence the evolution of *Doublesex* in the *Anastrepha fraterculus* species group. **PLoS ONE**, v. 7, n. 3, mar. 2012.
- SPRADLING, A. C. Germline cysts: communes that work. **Cell**, v. 72, n. 5, p. 649–651, mar. 1993.
- ST JOHNSTON, D. et al. Multiple steps in the localization of *bicoid* RNA to the anterior pole of the *Drosophila* oocyte. **Development (Cambridge, England)**, v. 107 Suppl, p. 13–19, 1989.
- STAUBER, M.; JACKLE, H.; SCHMIDT-OTT, U. The anterior determinant *bicoid* of *Drosophila* is a derived *Hox* class 3 gene. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 7, p. 3786–3789, mar. 1999.
- STAUBER, M.; PRELL, A.; SCHMIDT-OTT, U. A single *Hox3* gene with composite *bicoid* and *zerknüllt* expression characteristics in non-Cyclorrhaphan flies. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 1, p. 274–279, jan. 2002.
- SWANSON, W. J. et al. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. **Genetics**, v. 168, n. 3, p. 1457–1465, jan. 2004.
- SWANSON, W. J.; VACQUIER, V. D. The rapid evolution of reproductive proteins. **Nature Reviews Genetics**, v. 3, n. 2, p. 137–144, fev. 2002.
- THOMPSON, F. C. Introduction. In: Fruit fly expert identification system and systematic information database. [s.l.] Thompson F. C., 1998. v. 9, p. 1.
- WIEGMANN, B. M. et al. Episodic radiations in the fly tree of life. **Proceedings of the National Academy of Sciences**, p. 201012675, mar. 2011.

WIEGMANN, B. M.; MITTER, C.; THOMPSON, F. C. Evolutionary origin of the Cyclorrhapha (Diptera): test of alternative morphological hypotheses. **Cladistics**, v. 9, n. 1, p. 41–81, 1993.

WILHELM, J. E. et al. Isolation of a ribonucleoprotein complex involved in mRNA localization in *Drosophila* oocytes. **The Journal of Cell Biology**, v. 148, n. 3, p. 427–440, fev. 2000.

XIA, X. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. **Molecular Biology and Evolution**, v. 30, n. 7, p. 1720–1728, jul. 2013.

YANG, Z. PAML 4: Phylogenetic analysis by maximum likelihood. **Molecular Biology and Evolution**, v. 24, n. 8, p. 1586–1591, ago. 2007.

YANG, Z.; NIELSEN, R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. **Molecular Biology and Evolution**, v. 19, n. 6, p. 908–917, jun. 2002.

YANG, Z.; WONG, A.; NIELSEN, R. Bayes empirical bayes inference of amino acid sites under positive selection. **Molecular Biology and Evolution**, v. 22, n. 4, p. 1107–1118, jan. 2005.

YEATES, D. K. et al. Phylogeny and systematics of Diptera: two decades of progress and prospects. **Zootaxa**, n. 1668, p. 565–590, 2007.

YEATES, D. K.; WIEGMANN, B. M. Congruence and controversy: toward a higher-level phylogeny of Diptera. **Annual Review of Entomology**, v. 44, n. 1, p. 397–428, 1999.

ZHANG, J.; NIELSEN, R.; YANG, Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. **Molecular Biology and Evolution**, v. 22, n. 12, p. 2472–2479, ago. 2005.

ZUCCHI, R. A. Moscas-das-frutas (Diptera, Tephritidae) no Brasil: Taxonomia, distribuição geográfica e hospedeiros. In: Moscas-das-frutas no Brasil. [s.l.] H. M. L. de Souza, 1988. p. 114.

ZUCCHI, R. A. Taxonomia. In: MALAVASI, A.; ZUCCHI, R. A. (Eds.). Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado. [s.l.] Malavasi A. e Zucchi R. A., 2000. p. 13–24.

ANEXOS

ANEXO A: Contigs exu gerados a partir do sequenciamento de transcriptomas de *A. fraterculus*, e sequenciamento de fragmentos RACE e cDNA. Acima da região codificadora também está indicada a sequência proteica

```

|----- EXON 5'UTR DE CABEÇA -----|
>CF cont 1 -----CAGCCGTTTCTAGAATATGTAAAGTTTCTTGTAAACTTTTAAAGTTTTTGGTTATTGCATCGTGGCCAGAAAACAGAAGTGCATTCAACGAAATATTGC
>CF cont 2 GCTGCCATTTACGCTTACAAGTTTCAACTCGCCTCTTATCAGTTGG-----

>CM cont 1 -----AGAAAACAGAAGTGCATTCAACGAAATATTGC
>CM cont 2 -----AGAAAACAGAAGTGCATTCAACGAAATATTGC
>CM cont 3 -----AGAAAACAGAAGTGCATTCAACGAAATATTGC
>CM cont 4 -----AGAAAACAGAAGTGCATTCAACGAAATATTGC

----- EXON 5'UTR DE CABEÇA -----
>CF cont 1 ATTTAAAATTTA-TTTTTTGAATTCACACAACCTAGAACAAAAGACTTGCAATATTTAGCTATCCATTAGATCCAAGTGCATATGAAATAATTTGAATGTAAATGATCTTCACTTTGGCCTTAACCTAGTAGTGCAAAATCTTGGTGA
>CF cont 2 -----

>CM cont 1 ATTTAAAATTTA-TTTTTTGAATTCACACAACCTAGAACAAAAGACTTGCAATATTTAGCTATCCATTAGATCCAAGTGCATATGAAATAATTTGAATGTAAATGATCTTCACTTTGGCCTTAACCTAGTAGTGTAATAATCTTGGTGA
>CM cont 2 ATTTAAAATTTA-TTTTTTGAATTCACACAACCTAGAACAAAAGACTTGCAATATTTAGCTATCCATTAGATCCAAGTGCATATGAAATAATTTGAATGTAAATGATCTTCACTTTGGCCTTAACCTAGTAGTGTAATAATCTTGGTGA
>CM cont 3 ATTTAAAATTTA-TTTTTTGAATTCACACAACCTAGAACAAAAGACTTGCAATATTTAGCTATCCATTAGATCCAAGTGCATATGAAATAATTTGAATGTAAATGATCTTCACTTTGGCCTTAACCTAGTAGTGTAATAATCTTGGTGA
>CM cont 4 ATTTAAAATTTA-TTTTTTGAATTCACACAACCTAGAACAAAAGACTTGCAATATTTAGCTATCCATTAGATCCAAGTGCATATGAAATAATTTGAATGTAAATGATCTTCACTTTGGCCTTAACCTAGTAGTGTAATAATCTTGGTGA

-----|<>|----- EXON 5'UTR DE REP FÊMEA -----|
>CF cont 1 TTAAGACTTG-----
>CF cont 2 -----

>CM cont 1 TTAAGACTTG-----
>CM cont 2 TTAAGACTTG-----
>CM cont 3 TTAAGACTTG-----
>CM cont 4 TTAAGACTTG-----

>RF cont 1 -----GTAAAGCTACGACGTTGGCTTGAAACGTCTAGCGCTGACCGCGGCTAGTTTATTTGATCCGTTGAAGTCAAAGTGCCTGGTTTCGCGTATACACAAACAGGTACAAACTCTCAAAGTTGTACCAAAATTACAGTTCT
>RF cont 2 -----GTAAAGCTACGACGTTGGCTTGAAACGTCTAGCGCTGACCGCGGCTAGTTTATTTGATCCGTTGAAGTCAAAGTGCCTGGTTTCGCGTATACACAAACAGGTACAAACTCTCAAAGTTGTACCAAAATTACAGTTCT

|<>|----- EXON 5'UTR DE REP MACHO -----|
>CF cont 1 -----GGATTAGATTGTCTT
>CF cont 2 -----

>CM cont 1 -----GGATTAGATTGTCTT
>CM cont 2 -----GGATTAGATTGTCTT
>CM cont 3 -----GGATTAGATTGTCTT
>CM cont 4 -----GGATTAGATTGTCTT

>RF cont 1 -----GGATTAGATTGTCTT
>RF cont 2 -----GGATTAGATTGTCTT

```


----- EXON CDS3 -----

SEQ PROT I P Y M I L E S L N K Y N L L D R F L K T V K S F A N G L N L A K A N A N E S L Q H Y T L R K L S K
>CF cont 1 TATTCCATATATGATTTTGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCTTAAAACAGTCAAATCATTTGCCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAATCATTTGCAACATTACACGTTGCGCAAACCTCTCCAAA
>CF cont 2 TATTCCATATATGATTTTGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCTTAAAACAGTCAAATCATTTGCCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAATCATTTGCAACATTACACGTTGCGCAAACCTCTCCAAA
>CM cont 1 TATTCCATATATGATTTTGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCTTAAAACAGTCAAATCATTTGCCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAATCATTTGCAACATTACACGTTGCGCAAACCTCTCCAAA
>CM cont 2 TATTCCATATATGATTTTGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCTTAAAACAGTCAAATCATTTGCCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAATCATTTGCAACATTACACGTTGCGCAAACCTCTCCAAA
>CM cont 3 TATTCCATATATGATTTTGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCTTAAAACAGTCAAATCATTTGCCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAATCATTTGCAACATTACACGTTGCGCAAACCTCTCCAAA
>CM cont 4 TATTCCATATATGATTTTGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCTTAAAACAGTCAAATCATTTGCCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAATCATTTGCAACATTACACGTTGCGCAAACCTCTCCAAA
>RF cont 1 TATTCCATATATGATTTTGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCTTAAAACAGTCAAATCATTTGCCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAATCATTTGCAACATTACACGTTGCGCAAACCTCTCCAAA
>RF cont 2 TATTCCATATATGATTTTGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCTTAAAACAGTCAAATCATTTGCCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAATCATTTGCAACATTACACGTTGCGCAAACCTCTCCAAA
>RM cont 1 TATTCCATATATGATTTTGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCTTAAAACAGTCAAATCATTTGCCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAATCATTTGCAACATTACACGTTGCGCAAACCTCTCCAAA
>RM cont 2 TATTCCATATATGATTTTGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCTTAAAACAGTCAAATCATTTGCCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAATCATTTGCAACATTACACGTTGCGCAAACCTCTCCAAA
>RM cont 3 TATTCCATATATGATTTTGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCTTAAAACAGTCAAATCATTTGCCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAATCATTTGCAACATTACACGTTGCGCAAACCTCTCCAAA
>RM cont 4 TATTCCATATATGATTTTGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCTTAAAACAGTCAAATCATTTGCCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAATCATTTGCAACATTACACGTTGCGCAAACCTCTCCAAA
>RM cont 5 TATTCCATATATGATTTTGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCTTAAAACAGTCAAATCATTTGCCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAATCATTTGCAACATTACACGTTGCGCAAACCTCTCCAAA
CDS (cDNA) TATTCCATATATGATTTTGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCTTAAAACAGTCAAATCATTTGCCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAATCATTTGCAACATTACACGTTGCGCAAACCTCTCCAAA

----- EXON CDS3 -----

SEQ PROT V L S A K A H E V I N N Q N A S D G E I K N N V S S T T G I A A E C Q S P Q K S L Q K G V L A Q E R
>CF cont 1 GTGTTGAGCGCCAAAGCACATGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCC---GAGTGTGACAGGCCCCAAAAGTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTG
>CF cont 2 GTGTTGAGCGCCAAAGCACATGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCC---GAGTGTGACAGGCCCCAAAAGTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTG
>CM cont 1 GTGTTGAGCGCCAAAGCACATGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCCGCCGAGTGTGACAGGCCCCAAAAGTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTG
>CM cont 2 GTGTTGAGCGCCAAAGCACATGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCCGCCGAGTGTGACAGGCCCCAAAAGTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTG
>CM cont 3 GTGTTGAGCGCCAAAGCACATGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCCGCCGAGTGTGACAGGCCCCAAAAGTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTG
>CM cont 4 GTGTTGAGCGCCAAAGCACATGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCCGCCGAGTGTGACAGGCCCCAAAAGTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTG
>RF cont 1 GTGTTGAGCGCCAAAGCACATGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCC---GAGTGTGACAGGCCCCAAAAGTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTG
>RF cont 2 GTGTTGAGCGCCAAAGCACATGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCC---GAGTGTGACAGGCCCCAAAAGTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTG
>RM cont 1 GTGTTGAGCGCCAAAGCACATGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCCGCCGAGTGTGACAGGCCCCAAAAGTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTG
>RM cont 2 GTGTTGAGCGCCAAAGCACATGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCCGCCGAGTGTGACAGGCCCCAAAAGTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTG
>RM cont 3 GTGTTGAGCGCCAAAGCACATGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCCGCCGAGTGTGACAGGCCCCAAAAGTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTG
>RM cont 4 GTGTTGAGCGCCAAAGCACATGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCCGCCGAGTGTGACAGGCCCCAAAAGTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTG
>RM cont 5 GTGTTGAGCGCCAAAGCACATGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCCGCCGAGTGTGACAGGCCCCAAAAGTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTG
CDS (cDNA) GTGTTGAGCGCCAAAGCACACGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCCGCCGAGTGTGACAGGCCCCAAAAGTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTG

----- EXON CDS3 -----

SEQ PROT E L F D G N A S V R A K L A Y E V A Q Q L S N Q E H K V P I E S S A A L E N M I A A V Q T Y A Q P I E
>CF cont 1 AGCTGTTTCGATGGTAATGCTAGTGTGCGTGCCAAGTTGGCCTATGAGGTGGCACAACAGTTGAGCAATCAAGAACACAAGGTGCCTATCGAATCGTCGGCTGCATTAGAGAATATGATTGCCCGAGTGCAAACGTATGCACAGCCCATTTGA
>CF cont 2 AGCTGTTTCGATGGTAATGCTAGTGTGCGTGCCAAGTTGGCCTATGAGGTGGCACAACAGTTGAGCAATCAAGAACACAAGGTGCCTATCGAATCGTCGGCTGCATTAGAGAATATGATTGCCCGAGTGCAAACGTATGCACAGCCCATTTGA
>CM cont 1 AGCTGTTTCGATGGTAATGCTAGTGTGCGTGCCAAGTTGGCCTATGAGGTGGCACAACAGTTGAGCAATCAAGAACACAAGGTGCCTATCGAATCGTCGGCTGCATTAGAGAATATGATTGCCCGAGTGCAAACGTATGCACAGCCCATTTGA


```

CDS (cDNA) TAAGTCGCCATCTCCGCCAAACGGCGGCAAGCCGCGAGCTGGCAAGCGTAATCCGCG
RACE 3' -----CAATAACTCCTTAAAAACACAGCTGCCAAGCGCCGCGTGCACACGAGCAATACAACGGCAACAGCAACCAATTAAAAACTAAGAAGT

----- EXON 3' UTR COMUM -----|----- 3' UTR A -----|-----
>CF cont 3 TCATTCTAAGATTCAAGACAGGCGCTGTTGCAACATAAAAAAA-CAAGTCAATTTGTTTTATTTTTTTTTTGTGATAGTTTTCCACCAACGTAGCCACACTTTAGAGAGACGCCAGAAATATTTGCTGGTATGCAAGTTGAATAGAATTTGAA
>CM cont 1 TCATTCTAAGATTCAAGACAGGCGCTGTTGCAACATAAAAAAA-CAAGTCATTTGTTTTATTTTTTTTTT-GTGATAGTTTTCCACCAACGTAGCCACACTTTAGAGAGACGCCAGAAATATTTGCTG-----
>CM cont 2 TCATTCTAAGATTCAAGACAGGCGCTGTTGCAACATAAAAAAA-CAAGTCATTTGTTTTATTTTTTTTTT-GTGATAGTTTTCCACCAACGTAGCCACACTTTAGAGAGACGCCAGAAATATTTGCTG-----
>CM cont 3 TCATTCTAAGATTCAAGACAGGCGCTGTTGCAACATAAAAAAA-CAAGTCATTTGTTTTATTTTTTTTTT-GTGATAGTTTTCCACCAACGTAGCCACACTTTAGAGAGACGCCAGAAATATTTGCTG-----
>CM cont 4 TCATTCTAAGATTCAAGACAGGCGCTGTTGCAACATAAAAAAA-CAAGTCATTTGTTTTATTTTTTTTTT-GTGATAGTTTTCCACCAACGTAGCCACACTTTAGAGAGACGCCAGAAATATTTGCTG-----
>RF cont 1 TCATTCTAAGATTCAAGACAGGCGCTGTTGCAACATAAAAAAA-CAAGTCATTTGTTTTATTTTTTTTTT-GTGATAGTTTTCCACCAACGTAGCCACACTTTAGAGAGACGCCAGAAATATTTGCTGGTATGCAAGTTGAATAGAATTTGAA
>RF cont 2 TCATTCTAAGATTCAAGACAGGCGCTGTTGCAACATAAAAAAA-CAAGTCAATTTGTTTTATTTTTTTTTT-GTGA-----CGTAG--ACACTTTAGAGAGACGCCAGAAATATTTGCTGGTATGCAAGTTGAATAGAATTTGAA
>RM cont 1 TCATTCTAAGATTCAAGACAGGCGCTGTTGCAACATAAAAAAA-CAAGTCATTTGTTTTATTTTTTTTTT-GTGATAGTTTTCCACCAACGTAGCCACACTTTAGAGAGACGCCAGAAATATTTGCTG-----
>RM cont 2 TCATTCTAAGATTCAAGACAGGCGCTGTTGCAACATAAAAAAA-CAAGTCATTTGTTTTATTTTTTTTTT-GTGATAGTTTTCCACCAACGTAGCCACACTTTAGAGAGACGCCAGAAATATTTGCTG-----
>RM cont 3 TCATTCTAAGATTCAAGACAGGCGCTGTTGCAACATAAAAAAA-CAAGTCATTTGTTTTATTTTTTTTTT-GTGATAGTTTTCCACCAACGTAGCCACACTTTAGAGAGACGCCAGAAATATTTGCTG-----
>RM cont 4 TCATTCTAAGATTCAAGACAGGCGCTGTTGCAACATAAAAAAA-CAA-----
>RM cont 5 TCATTCTAAGATTCAAGACAGGCGCTGTTGCAACATAAAAAAA-CAAGTCATTTGTTTTATTTTTTTTTT-GTGATAGTTTTCCACCAACGTAGCCACACTTTAGAGAGACGCCAGAAATATTTGCTG-----
RACE 3' TCATTCTAAGATTCAAGACAGGCGCTGTTGCAACATAAAAAAA-CAA-----
----- EXON 3' UTR FÊMEA -----
>CF cont 3 TTTTAAATTTTATATATACCACATATTTTATTATAATTATTATTACATGTATTAATTAGCTAAACGTACATGCATATAATTAACAAAAGCAAAGAAATATGTATATATATATGTACACTTGTGTAACGTTAGCAGTTACAATACCCAAT
>CM cont 1 -----
>CM cont 2 TTTTAAATTTTATATATACCGCATATTTTATTATAATTATTATTACATGTATTAATTAGCTAAACGTACATGCATATAATTAACAAAAGCAAAGAAATATGTATATATATATGTACACTTGTGTAACGTTAGCAGTTACAATACCCAAT
>CM cont 3 -----
>CM cont 4 -----
>RF cont 1 TTTTAAATTTTATATATACCACATATTTTATTATAATTATTATTACATGTATTAATTAGCTAAACGTACATGCATATAATTAACAAAAGCAAAGAAATATGTATATATA-----CACTTGTACACTTGTTTAACGTTAGCAG
>RF cont 2 TTTTAAATTTTATATATACCACATATTTTATTATAATTATTATTACATGTATTAATTAGCTAAACGTACATGCATATAATTAACAAAAGCAAAGAAATATGTATATATA-----CACTTGTACACTTGTTTAACGTTAGCAG
>RF cont 3 -----ATATGTACACTTGTGTAACGTTAGCAGTTACAATACCCAAT
>RM cont 1 -----
>RM cont 2 -----
>RM cont 3 -----
>RM cont 4 -----
>RM cont 5 -----
RACE 3' -----
----- EXON 3' UTR FÊMEA -----
>CF cont 3 TGTAGATGACGTTAGTTTAGAAAATATCTGATTATTGAAGTCATTAATAATTATTATTTTATTTGCTTTGCTACGCTACAAAGTACAACAACATAAATTAGTTTGCCAGCCGAGGCAGAGGTGTAAGCTTTTCATAATTTATTTTCGTAGT

```

```

>CM cont 1 -----
>CM cont 2 TGTAGATGACGTTAGTTTAGAAAAATATCTGATTATTGAAGTCATTAATAATTATTATTTTATTTGCTTTGCTACGCTACAAAGTACAACAACACAAATTAGTTTGCCAGCCGCAGGCAGAGGTGTAAGCTTTTCATAATTTATTTTCGTAGT
>CM cont 3 -----
>CM cont 4 -----

>RF cont 3 TGTAGATGACGTTAGTTTAGAAAAATATCTGATTATTGAAGTCATTAATAATTATTATTTTATTTGCTTTGCTACGCTACAAAGTACAACAACATAAATTAGTTTGCCAGCCGCAGGCAGAGGTGTAAGCTTTTCATAATTTATTTTCGTAGT

>RM cont 1 -----
>RM cont 2 -----
>RM cont 3 -----
>RM cont 4 -----
>RM cont 5 -----

RACE 3' -----

----- EXON 3' UTR FÊMEA -----

>CF cont 3 TTTACTAGAGTTGAGTTAAATAAAATGTAACAAATATTAGCTAACCTCTGGTTTGACGCTTTAGAGGATTCTACAAGAATTTGACA-CCTTGCTTATTTGCGCGAATTGAAATTTTGAATTTTGTAGAAGGTTGCGCTTTAATATT-----T

>CM cont 1 -----
>CM cont 2 TTTACTAGAGTTGAGTTAAATAAAATGTAACAAATATTAGCTAACCTCTGGTTTGACGCTTTAGAGGATTCTACAAGAATTTGACA-CCTTGCTTATTTGCGCGAATTGAAATTTTGAATTTTGTAGAAGGTTGCGCTTTAATATT-----T
>CM cont 3 -----
>CM cont 4 -----

>RF cont 3 TTTACTAGAGTTGAGTTAAATAAAATGTAACAAATATTAGCTAACCTCTGGTTTGACGCTTTAGAGGATTCTACAAGAATTTGACA-CCTTGCTTATTTGCGCGAATTGAAATTTTGAATTTTGTAGAAGGTTGCGCTTTAATATT-----T

>RM cont 1 -----
>RM cont 2 -----
>RM cont 3 -----
>RM cont 4 -----
>RM cont 5 -----

RACE 3' -----

----- EXON 3' UTR FÊMEA -----

>CF cont 3 GTTTTTTATGATTCCTATTATGATATTTATGTATTTATGCTATTTTGAATTTTGTGTTAAAAAGAATACACAAGAAATATATAAACAAAT

>CM cont 1 -----
>CM cont 2 GTTTTTTATGATTCCTATTATGATATTTATGTATTTATGCTATTTTGAATTTTGTGTTAAAAAGAATACACAAGAAATATATAAACAAATATTTTTTCTAACTTAAGCA
>CM cont 3 -----
>CM cont 4 -----

>RF cont 3 GTTTTTTATGATTCCTATTATGATATTTATGTATTTATGCTATTTTGAATTTTGTGTTAAAAAGAATACACAAGAAATATATAAACAAATATTTTTTCTAACTTAAAAAAAAAAAAAAAAATAATTGGCGCGTACACTTCTGATAGGTGT

>RM cont 1 -----
>RM cont 2 -----
>RM cont 3 -----
>RM cont 4 -----
>RM cont 5 -----

```

RACE 3' -----
-----|<>|----- EXON 3' UTR MACHO -----|<>|-----

>CM cont 1 -----CTCAAATTTAAAGCGAAAAAACCC-ACAGTTTAGCTTCACAAGCAAACACAAAAATTGATTGGTGCCTTGCAGTTATACATA----GCAAAACGAAAAAAA-----CAAATTTAATTCGCTCAAACA
>CM cont 3 -----CTCAAATTTAAAGCGAAAAAACCC-ACAGTTTAGCTTCACAAGCAAACACAAAAATTGATTGGTGCCTTGCAGTTATACATA----GCAAAACGAAAAAAA-----CAAATTTAATTCGCTCAAACA
>CM cont 4 -----CTCAAATTTAAAGCGAAAAAACCC-ACAGTTTAGCTTCACAAGCAAACACAAAAATTGATTGGTGCCTTGCAGTTATACATA----GCAAAACGAAAAAAA-----CAAATTTAATTCGCTCAAACA

>RF cont 3 TGGCCGAGCTCCTCT

>RM cont 1 -----CTCAAATTTAAAGCGAAAAAACCC-ACAGTTTAGCTTCACAAGCAAACACAAAAATTGATTGGTGCCTTGCAGTTATACATA----GCAAAACGAAAAAAA-----CAAATTTAATTCGCTCAAACA
>RM cont 2 -----CTCAAATTTAAAGCGAAAAAACCC-ACAGTTTAGCTTCACAAGCAAACACAAAAATTGATTGGTGCCTTGCAGTTATACATA----GCAAAACGAAAAAAA-----CAAATTTAATTCGCTCAAACA
>RM cont 3 -----CTCAAATTTAAAGCGAAAAAACCC-ACAGTTTAGCTTCACAAGCAAACACAAAAATTGATTGGTGCCTTGCAGTTATACATA----GCAAAACGAAAAAAA-----CAAATTTAATTCGCTCAAACA
>RM cont 4 -----CTCAAATTTAAAGCGAAAAAACCC-ACAGTTTAGCTTCACAA
>RM cont 5 -----CTCAAATTTAAAGCGAAAAAACCC-ACAGTTTAGCTTCACAAGCAAACACAAAAATTGATTGGTGCCTTGCAGTTATACATA----GCAAAACGAAAAAAA-----CAAATTTAATTCGCTCAAACA

RACE 3' -----CTCAAATTTAAAGCGAAAAAACCCACAGTTTAGCTTCACAAGCAAACACAAAAATTGATTGGTGCCTTGCAGTTATACATACATAGCAAACGAAAAAAA-----

----- EXON 3' UTR MACHO [1] -----

>CM cont 1 CTAGGAGGCATGATCCGATGAAATAACCAAGAAAATGAAAAAGGAGAACGAAAAATTTGCGCGACCCATAAACTCCTTGCGCAAAATTTCTGTGAAAAAAAATTTACTTGTGTACTTTGAAATTTAACCAATTTTTCGGAATAAAT
>CM cont 3 CTAGGAGGCATGATCCGATGAAATAACCAAGAAAATGAAAAAGGAGAACGAAAAATTTGCGCGACCCATAAACTCCTTGCGCAAAATTTCTGTGAAAAAAAATTTACTTGTGTACTTTGAAATTTAACCAATTTTTCGGAATAAAT
>CM cont 4 CTAGGAGGCATGATCCGATGAAATAACCAAGAAAATGAAAAAGGAGAACGAAAAATTTGCGCGACCCATAAACTCCTTGCGCAAAATTTCTGTGAAAAAAAATTTACTTGTGTACTTTGAAATTTAACCAATTTTTCGGAATAAAT

>RM cont 1 CTAGGAGGCATGATCCGATGAAATAACCAAGAAAATGAAAAAGGAGAACGAAAAA-TTTCGCGACCCCGTAAACTCCTTGAGCAAAATTTCTGTGAAAAAAAATTTACTTGTGTACTTTGAAATTTAACCAATTTTTCGGAATAAAT
>RM cont 2 CTAGGAGGCATGATCCGATGAAATAACCAAGAAAATGAAAAAGGAGAACGAAAAA-TTTCGCGACCCCGTAAACTCCTTGAGCAAAATTTCTGTGAAAAAAAATTTACTTGTGTACTTTGAAATTTAACCAATTTTTCGGAATAAAT
>RM cont 3 CTAGGAGGCATGATCCGATGAAATAACCAAGAAAATGAAAAAGGAGAACGAAAAA-TTTCGCGACCCCGTAAACTCCTTGAGCAAAATTTCTGTGAAAAAAAATTTACTTGTGTACTTTGAAATTTAACCAATTTTTCGGAATAAAT
>RM cont 5 CTAGGAGGCATGATCCGATGAAATAACCAAGAAAATGAAAAAGGAGAACGAAAAA-TTTCGCGACCCCGTAAACTCCTTGCGCAAAATTTCTGTGAAAAAAAATTTACTTGTGTACTTTGAAATTTAACCAATTTTTCGGAATAAAT

----- EXON 3' UTR MACHO [1] -----

>CM cont 1 CAAGCCAAAGTAACCGACATTTGAAATGCAACAACAATCAACCATACCTGGATGCTTTCAATGCCAGTAACAAGAATTGATTTCCGCAATACAATGCAAGCAAAAATTTATTGAATTTAACCAATTACATATAACCATACATATCTGGGATC
>CM cont 3 CAAGCCAAAGTAACCGACATTTGAAATGCAACAACAATCAACCATACCTGGATGCTTTCAATGCCAGTAACAAGAATTGATTTCCGCAATACAATGCAAGCAAAAATTTATTGAATTTAACCAATTACATATAACCATACATATCTGGGATC
>CM cont 4 CAAGCCAAAGTAACCGACATTTGAAATGCAACAACAATCAACCATACCTGGATGCTTTCAATGCCAGTAACAAGAATTGATTTCCGCAATACAATGCAAGCAAAAATTTATTGAATTTAACCAATTACATATAACCATACATATCTGGGATC

>RM cont 1 CAAGCCAAAGTAACCGACATTTGAAATGCAACAACAATCAACCATACCTGGATGCTTTCAATGCCAGTAACAAGAATTGATTTCCGCAAGACAATGCAAGCAAAAATTTATTGAATTTAACCAATTACATATAACCATACATATCTGGGATC
>RM cont 2 CAAGCCAAAGTAACCGACATTTGAAATGCAACAACAATCAACCATACCTGGATGCTTTCAATGCCAGTAACAAGAATTGATTTCCGCAAGACAATGCAAGCAAAAATTTATTGAATTTAACCAATTACATATAACCATACATATCTGGGATC
>RM cont 3 CAAGCCAAAGTAACCGACATTTGAAATGCAACAACAATCAACCATACCTGGATGCTTTCAATGCCAGTAACAAGAATTGATTTCCGCAAGACAATGCAAGCAAAAATTTATTGAATTTAACCAATTACATATAACCATACATATCTGGGATC
>RM cont 5 CAAGCCAAAGTAACCGACATTTGAAATGCAACAACAAT

----- EXON 3' UTR MACHO [1] -----

>CM cont 1 GCCTCAACGAAGCAAGATCCTGTGCAAGTGAATCAATCAACAGCGGAGCACAACAACAATTAGAGACTGACTTATCTGTTGGTGATTTGCAGCAAACAACCTGTCTCTCATCAATGCCAGCTAGCTGGCCAAAAGGTGCGCGGGATAAAG
>CM cont 3 GCCTCAACGAAGCAAGATCCTGTGCAAGTGAATCAATCAACAGCGGAGCACAACAACAATTAGAGACTGACTTATCTGTTGGTGATTTGCAGCAAACAACCTGTCTCTCATCAATGCCAGCTAGCTGGCCAAAAGGTGCGCGGGATAAAG
>CM cont 4 GCCTCAACGAAGCAAGATCCTGTGCAAGTGAATCAATCAACAGCGGAGCACAACAACAATTAGAGACTGACTTATCTGTTGGTGATTTGCAGCAAACAACCTGTCTCTCATCAATGCCAGCTAGCTGGCCAAAAGGTGCGCGGGATAAAG

>RM cont 1 GCCTCAACGAAGCAAGATCCTGTGCAAGTGAATCAATCAACAGCGGAGCACAACAACAATTAGAGACTGACTTATCTGTTGGTGATTTGCAGCAAACAACCTGTCTCTCATCAATGCCAGCTAGCTGGCCACAAGGTGCGCGGGATAAAG
>RM cont 2 GCCTCAACGAAGCAAGATCCTGTGCAAGTGAATCAATCAACAGCGGAGCACAACAACAATTAGAGACTGACTTATCTGTTGGTGATTTGCAGCAAACAACCTGTCTCTCATCAATGCCAGCTAGCTGGCCACAAGGTGCGCGGGATAAAG
>RM cont 3 GCCTCAACGAAGCAAGATCCTGTGCAAGTGAATCAATCAACAGCGGAGCACAACAACAATTAGAGACTGACTTATCTGTTGGTGATTTGCAGCAAACAACCTGTCTCTCATCAATGCCAGCTAGCTGGCCACAAGGTGCGCGGGATAAAG

----- EXON 3' UTR MACHO [1] -----

>CM cont 1 CGGAATTTTCATTTTATTTAGTTTATTAACAAACCTCGCAAGTCCATGCCGATGGCGTGAGGGGTAGAAAATATTATGATTTATTGTTAAAAATATAAAAAATATTACATACAGAGAAGAATATCTGGTAGTGACGACATACATACATATCACACT
>CM cont 3 CGGAATTTTCATTTTATTTAGTTTATTAACAAACCTCGCAAGTCCATGCCGATGGCGTGAGGGGTAGAAAATATTATGATTTATTGTTAAAAATATAAAAAATATTACATACAGAGAAGAATATCTGGTAGTGACGACATACATATCACACT
>CM cont 4 CGGAATTTTCATTTTATTTAGTTTATTAACAAACCTCGCAAGTCCATGCCGATGGCGTGAGGGGTAGAAAATATTATGATTTATTGTTAAAAATATAAAAAATATTACATACAGAGAAGAATATCTGGTAGTGACGACATACATATCACACT

>RM cont 1 CGGAATTTTCATTTTATTTAGTTTATTAACAAACCTCGCAAGTCCATGCCGACGGCGTGAGGGGTAAAAATATTATGATTTATTGTTAAAAATATAAAAAATATTACATACAGAGAAGAATATCTGGTAGTGACGACA-----CACT
>RM cont 2 CGGAATTTTCATTTTATTTAGTTTATTAACAAACCTCGCAAGTCCATGCCGACGGCGTGAGGGGTAAAAATATTATGATTTATTGTTAAAAATATAAAAAATATTACATACAGAGAAGAATATCTGGTAGTGACGACATACATACATATCACACT
>RM cont 3 CGGAATTTTCATTTTATTTAGTTTATTAACAAACCTCGCAAGTCCATGCCGACGGCGTGAGGGGTAAAAATATTATGATTTATTGTTAAAAATATAAAAAATATTACATACAGAGAAGAATATCTGGTAGTGACGACATACATACATATCACACT

----- EXON 3' UTR MACHO [1] -----

>CM cont 1 ACATTTACACTTTTAAAAATTTTCTGAGATTAAGGAATCTTAAAGCTGATTTATAAATGCATATTTCTTTTAAACAGAATCCATATTGAACCGGTTCCGCTTAGTTCAGAAAACCA-TTTAATTTTAGTATGTATGAACTCTATAAACGAAA
>CM cont 3 ACATTTACACTTTTAAAAATTTTCTGAGATTAAGGAATCTTAAAGCTGATTTATAAATGCATATTTCTTTTAAACAGAATCCATATTGAACCGGTTCCGCTTAGTTCAGAAAACCA-TTTAATTTTAGTATGTATGAACTCTATAAACGAAA
>CM cont 4 ACATTTACACTTTTAAAAATTTTCTGAGATTAAGGAATCTTAAAGCTGATTTATAAATGCATATTTCTTTTAAACAGAATCCATATTGAACCGGTTCCGCTTAGTTCAGAAAACCA-TTTAATTTTAGTATGTATGAACTCTATAAACGAAA

>RM cont 1 ACATTTACACTTTTAAAA
>RM cont 2 ACATTTACACTTTTAAAAATTTTCTGAGATTAAGGAATCTTAAAGCTGATTTATAAATGCCTATTTCTTTTGAACAGAATCCATATTGAACCGATTCCGCTTAGTTCAGATACCCTTTAATTTTAGTATGTATGAACTCTATAAACGAAA
>RM cont 3 ACATTTACACTTTTAAAAATTTTCTGAGATTAAGGAATCTTAAAGCTGATTTATAAATGCCTATTTCTTTTGAACAGAATCCATATTGAACCGATTCCGCTTAGTTCAGATACCCTTTAATTTTAGTATGTATGAACTCTATAAACGAAA

----- EXON 3' UTR MACHO [1] -----

>CM cont 1 AATGAAAAAAA-CTATAGCTGAACTATTTACAACAACATTTAAAACAATAATAAGAAAAATTATAACATATCAAATAATCTAECTTAGATCAGTAGATTCAACAACAATACAACGCACATGAGTAATAAATAGGCATAGCAATCTAACATGCA
>CM cont 3 AATGAAAAAAA-CTATAGCTGAACTATTTACAACAACATTTAAAACAATAATAAGAAAAATTATAACATATCAAATAATCTAECTTAGATCAGTAGATTCAACAACAATACAACGCACATGAGTAATAAATAGGCATAGCAATCTAACATGCA
>CM cont 4 AATGAAAAAAA-CTATAGCTGAACTATTTACAACAACATTTAAAACAATAATAAGAAAAATTATAACATATCAAATAATCTAECTTAGATCAGTAGATTCAACAACAATACAACGCACATGAGTAATAAATAGGCATAGCAATCTAACATGCA

>RM cont 2 AATGAAAAAAAACCTATAGCTGAACTATTTACAACAACATTTAAAACAATAATAAGAAAAATTATAACATATCAAATAATCTAECTTAGATCAGTAGATTCAACAACAATACAACGCACCTTGAAGTAATAAATAGGCATAGCAGTCTA
>RM cont 3 AATGAAAAAAAACCTATAGCTGAACTATTTACAACAACATTTAAAACAATAATAAGAAAAATTATAACATATCAAATAATCTAECTTAGATCAGTAGATTCAACAACAATACAACGCACATGAGTAATAAATAGGCATAGCAGTCTAACATGCA

----- EXON 3' UTR MACHO [1] -----

>CM cont 1 TAGGTATGTGCACACACACACATACATACATACAT-----
>CM cont 3 TAGGTATGTGCACACACACACACATACATACATATTTCTTGTATGCCATAATTGTGCTCATATTTGCATTGCAATGCAGCAGAATCTTACATTTGAACTGCAGATATAAATATTTTACACACAATGAAATGTTATTGCATTTTCCCGC
>CM cont 4 TAGGTATGTGCACACACACACATACATACATACAT-----

>RM cont 3 TAGGTATGTGCACACACA----TACATACATACATTTACTTGTATGCCATAATTGTGCTCATATTTGCATTGCAATGCAGCAGAATCTTACATTTGAACTGCAGATATAAATATTTTACACACAATGAAATGTTATTGCATTTTCCCGC

----- EXON 3' UTR MACHO [1] -----

>CM cont 1 -----
>CM cont 3 CCAAGTTAATGGATCCTGTTTGAAGAGGTCCATTTTCTTATTTTCATGCGCCACAAAACGACCGAGTTCTAATTA AAAACTAAATATTCGTACATTTTTTTTTGTTATTGTTTCTGTTCAAAAAGAAAATTTGCTAATATTTCTGTTGTGACT
>CM cont 4 -----

>RM cont 3 CCAAGTTAATGGATCCAGTTCGAAAAGAGGTCCATTTTCTTATTTTCATGCGCCACAAAACGACCGAGTTCTAATTA AAAACTAAATATTCGTACATTTTTTTTTGTTATTGTTTCTGTTCAAAAAGAAAATTTGCTAATATTTCTGTTGTGACT

----- |<>| ----- EXON 3' UTR MACHO [2] -----

>CM cont 1 -----ACACACATGCAACAATCTTATATATATATATATACAAACAAAAGACTTTTCCGTACCAAGACAAAATTCGTTATAAAAATAATTGTAATTATAAAAATAAT
>CM cont 4 -----

```

>RM cont 3 TCTTGACTTTATTTGAAATAAAAAATCAAAAATCAATTTTAAAAAAAAA
----- EXON 3' UTR MACHO [2] -----
>CM cont 1 GTAGGTTGATAAGAGACGTTTAGTAGTTGTAGTTGTAGTAGTAAAACCATAAGCACTTAGACAATTCCAAATGAATTAGACTCGTCGCATAGTAATTTTTTAAATTAAAATAAAAGAAAATAAATATAAAATACCTGTAATACTAAATA
>CM cont 4 -----
----- EXON 3' UTR MACHO [2] -----
>CM cont 1 TATATAAGAACTACGAATAAAATGTAAATAATACCCAATTTAAATTTATGAAAAGATATTTTTGTAAGCTTTAAATGAGGAAGTGCGAAATAAAATGACTAACACGAATAATAATTACAATGAAAATCACACGAAACGCTTTTGTGTATAG
>CM cont 4 -----
----- EXON 3' UTR MACHO [2] -----
>CM cont 1 ATCAATTGCGGAAATAACTGGCGCGAGCGCTTGAGAATAAACTACATTTGTGTACAATAACACGAAGGAGAATTGATACGACAAAAACAATATACATATATATATGAATAAGTAAATAAATAACAATAATTTAACAGGGAATTC AAGA
>CM cont 4 -----
----- EXON 3' UTR MACHO [2] -----
>CM cont 1 TAACAAATTTATTAATTTGTTCTAATCAAAATATTAGCTTTAGTGCTCACAAAAGTATGCAAAGAAAACTTCTAAATCTCTCATGTGTGAATTTCTTGTTCATACAGAACATTGAACGTTTATCAGAAGAATTTAAATTTATGTATGTGA
>CM cont 4 -----
----- EXON 3' UTR MACHO [2] -----
>CM cont 1 ACCGAAATAAGTTTTACTAACAATTAAAAAATACTTATGGACCTTTAAATTAAGCAGCAAATGTATATTGAATTTATATTTGTTGAGCCAGACCAATGACTGGATTTTTTGTGGCTGATGTAGCTGAAGAATACAACGAAAAACAATAT
>CM cont 4 -----
----- EXON 3' UTR MACHO [2]-----|<>|----- EXON 3' UTR MACHO [3]-----
>CM cont 1 GATTATGGTATATAGCTACGAAGAAAGTAAATAAAAGGAGAAACATATTGTTTTTTGTTGGCTGATGTAGCTG
>CM cont 4 -----ACATGGATCATAAAAAGTATATAAATAGAAGGCATAAAGAAAAATTAGATGTATATGCTCAATGCCTGTGTGACC
----- EXON 3' UTR MACHO [3] -----
>CM cont 4 AATTGCTGTGAAAGCTTACAAATATACATACGTACACACATACATGCATACATACATTTGTTTTCATACATTCATTAACGCGGCATAAAACATTTCACTTCATTGCAGTACCTCAGCGAAAATAAAAAC TAATTAACAACAGTTTTT
----- EXON 3' UTR MACHO [3] -----
>CM cont 4 TTTAAGAAATCAAAATGCGTACTTCAGGAAGGTGCATTA AAAAGGAAGCAAAGTACAGGGCTTTGAGGACGATTCATAAAAATTCATTTTATATTTTCATAAGTGAAATATAACCAAACAATACGAGATTTATAGTGTAAACAAATACTTGC
----- EXON 3' UTR MACHO [3] -----|
>CM cont 4 AAAGTTGATATTAATAAGGTGCGCTCCCCACTTTAGATTAGATTTGTGGAGAATTGGGCAAGTCCAAGCACTCAGCAAAGAGACCATTATACTACGCCCGGATA

```

OBSERVAÇÕES: As regiões indicadas com “|<>|” indicam a possibilidade de presença de íntron

ANEXO B: Estrutura do gene *exu* em *A. fraterculus*, incluindo a sequência proteica (acima da nucleotídica)

```
|----- EXON 5'UTR DE CABEÇA -----|
SEQ PROT
SEQ GEN TTTTGGTTATTGCATCGTGGCCAGAAAACAGAAGTGCATTCAACGAAATATTGCATTTAAAATTTA-TTTTTTGAATTCACACAACCTAGAACAAAAGACTTGCAATATTTAGCTATCCATTAGATCCAAGTGCATATGAAATAATTTGA

----- EXON 5'UTR DE CABEÇA -----|----- INTRON -----|
SEQ PROT
SEQ GEN ATGTAATGATCTTCACTTTGGCCTTAACCTAGTAGTGTAATCTTGGTGATTAAGACTTGGTAAGATCGCGTGACGATGTTGTTGTTGTAACAGCATAGACATTTCCCCTAAATGTTTGGGAAAAGCGGAGTGACAGAATGGAGTGACAT

----- INTRON -----|
SEQ PROT
SEQ GEN TTTTTAATCGAATATAAATCCGGGTAGTTCTGGTAGCGCAGAACCAGACTGTGGTGCGACGTCGCGAAAAGAGCTGTGTGATAAGGCAACTGAGGGATGCGTCTGTAGAAAGGGCAATTTACAATGGTTTTGCCTCACTACTCGCTAAAAACA

----- INTRON -----|
SEQ PROT
SEQ GEN ATCCAATGCGCGTAGCGTACAATTTTCATGTTAAAAAGTTGGCTTGCTTAGCAGATTATTAACGGAAAGTACTTGCAATTTAGGCACTCATAAATGATAATTGAAATTTGCATCGAATACCACCCAGCCAGCCAGCGGTCATTGATTGCTC

----- INTRON -----|
SEQ PROT
SEQ GEN TGATTTCCGCACAAATTTTCTTTACAGTTGTAACGAAATACATACATACACATACATGTACATACGTACATAAATACGGTAAAGATCGGCGATCATTGGTTTTTGTTCGTGCGAATTATTTCAATTCTTGGTGCGCTTTTGCCTTC

----- INTRON -----|
SEQ PROT
SEQ GEN AGTCCAGAGTGCTGCTTTGCTTTCCGGATGTCCACCTGTCAACGATCAATACACGGCTTGAAATGTTTGTATCTAACTGTGAGAGAAGGAGCACGAGCAAGAAAAGCGGAATTTGCACGAAGCAATTGGAACAATTGAAAGAGCGCAGCAT

----- INTRON -----|
SEQ PROT
SEQ GEN GAAGGATCAAGTTAACACCGGAAAAAATTACAATAATAGTGGATCGTAACTGGAAATTATACCGTACTGAGTTGAGTTGTTCTTACTCCACTTCGCCCCACGTCCTTACCACCGTTTGTGAAACGATTCTTGATGCACCTACTGCGCTTAC
```

```

----- INTRON -----|----- EXON 5'UTR DE REP FÊMEA -----
SEQ PROT
SEQ GEN TTCCGTTTGAGATTTCGACAATATATGTGCATATATACATACATACTTATGTATGTATATATCAGCAGCTATACAAAGCAAAGTGAGTGAGTAAAGTTACGACGTTGGCTTGAAATGTCTAGCGCTGACAGCGGCTAGTTTATTTGATCCGTTG

----- EXON 5'UTR DE REP FÊMEA -----|----- INTRON -----|----- EXON 5'UTR DE REP MACHO -----
SEQ PROT
SEQ GEN AAGTCAAACCTGCGTGGTTCGCGTATACACAAACAGTGACAAACTCTCAAAGTTGTACAAAATTACAGTCTGTAGTTGGGATTTTTTTTTTTATCGTTATTAAGCAAGCAATCAAACATGTGATAGACTACTCAAACCTTCGGAGTTTTC

----- EXON 5'UTR DE REP MACHO -----|----- EXON 1 5' UTR COMUM -----
SEQ PROT
SEQ GEN TTTTCAAATTGTTTTGCATTTCTATTCAAGTTTTATTTTTTAAATTTTAAGAAAAATCAGTTGCTAAATTCAGTTTCTTTCTATATTCACAGGATTAGATTGCTTTAAGAAAAATTCGATTCAGTACGTTTCAGAAAAATTCGTTTGT

-----|----- INTRON -----
SEQ PROT
SEQ GEN AGCAGAGAGAAAGGAGTTCAGCTGACTCAGGTAATGGCGGAGAGGGCACAAAAATGTCTGACCAATTGCAATGGGCTGGCGTTAATATTTTTTAAATTTAAACCAAATTTAAGAAAAGTAATTTTGATTCAAGAAAGGAACATAAAT

----- INTRON -----
SEQ PROT
SEQ GEN GCAATAATGATTTATGAGATTTTTTGTCTGTTCAATTAATTAATGGTGAATTTTTTTTTATTTTTGCTACTGCTTAAAGATTGCATTACAAATGGCATATGAAATAAAAAATTTATTTGCCTTTAAATTTATTTAGTCGACAATTTTTAAC

-----|----- EXON 2 5' UTR COMUM -----|----- EXON CDS1 -----
SEQ PROT
SEQ GEN ACATTTTTAACCCGAAAATTCCTTTTCCTTTGTCTATTTTCAGCGCATTAACTAGTCAAATTAGTCCGATTTTCGCTTGGATTATTTTTATTGAAATTTTAGAAAATGGTCGCTTCAACGAATGCAAGCCAAAACGGTGTAGTCAATTCCTCC
M V A S T N A S Q N G V V N S P

----- EXON CDS1 -----|-----
SEQ PROT
SEQ GEN AGTAGTAGGAACCTCTAATGGAGTAACAACCAATGGCACAAATGAATGTCGCTGGCGGACACAAGACTTTGGCCAAGGGCAGGTACACTTTAGTAGGCGTCGATATTGACACAACCTGGGCGTCGACTCATCGATGAAGTAAGTATGCGTGTGTG
V V G T S N G V T T N G T M N V A G G H K T L A K G R Y T L V G V D I D T T G R R L I D E

----- INTRON -----
SEQ PROT
SEQ GEN TGTGTTATAAATGTTTATTATACATATAATTTAAACAATCAGAACACAGTTTGTCCAATTCCTATCCTTATAATAACATTTGCATTTCTAGAACTTCTCACCTAAAAATTTGTTTAAAGCCGACTAGCGGCTCTTTTAACTTTTAATTTAATTT

----- INTRON -----
SEQ PROT
SEQ GEN TAATATAAATTTGTAATTATTTTCTGTATATACAAACGATTAATTTTAGTTTGTATAAGGGAAGCAAATATGTGAAAATCCCCGGCTTCAGTTGCAAGAATGTGCATAAAAACTGTTTCATATGCGCATGAAAAATATATTTTTTTAAAA

```

```

----- INTRON -----
SEQ PROT
SEQ GEN AAAAACTATATTAATGCATCCGCTTTCAGGTAATCCGCTTACTGTTACAAAGTCATGAACTTGTGCGTATGGGTGATGCATTACTTGAATATGTTGTTTGTGCATTGATAACTGCACCTTTTTTTCCTGATTCCGCTTTTGGTTTTCGTCT

----- INTRON -----
SEQ PROT
SEQ GEN AATTTTACTATTCCCCTACAAGTCATTGGCCTTTGAAATCGTTTACTTTCTTTGCGCAAATTATACTTTTTTGTGTTGTTTTATTTATTTATTTTAACCTTATGGCCGATGTACATATATACAAATATGTAATATATATATCCA

----- INTRON -----
SEQ PROT
SEQ GEN AAACATAAACTTCCTCAATTGCTTTGTCAAAGTTGTTTTCCGGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAACAAATTTTCTCTCAGTTATAAGAGGTTAGGAAAGCATAAAATTGTGATGCCTTATTTTGCTCTCTCTTTTCTCCTC

-----|----- EXON CDS2 -----
SEQ PROT      I V Q L A A Y S P K D H F E Q Y I M P Y M N L N P A A R Q R H Q V R V I S I G F Y R M L K S M Q T
SEQ GEN CACAGATCGTTTCAGTTGGCCGCGTATTCGCCAAAGATCATTTTGAACAATATATTATGCCATACATGAACTTGAATCCTGCTGCTCGCCAACGTACCAAGTGCCTGTGATCTCTATTGGATTCTATCGTATGTTGAAATCAATGCAAACCT

-----|----- INTRON -----
SEQ PROT Y K
SEQ GEN ATAAGGTGAGAGAAAAAAAATACATTAAGTACTGGACTTCATTTATGCACCTCCAATTACGTTTATTGGTAGTTGCCATATTTTTTTTTTTTTTGTGTAATTAACAAAAATATATATTTGTTGACGTTGTTGCATACAAAAGTAAATACA

----- INTRON -----
SEQ PROT
SEQ GEN ATTACTAAAAATATTTTTTATATAAAAGGTTTTCCAATAAAAGCTGTTATTTTGTATATTCCAAAAAAATTTTTTTTTTTTATTAATAAATGATCGGATGTTTATTTTATTATAAAGCGAAAGGTATGCCGGGACTAGGGAAAAATAACGTCGG

----- INTRON -----
SEQ PROT
SEQ GEN GCAAATGACCACCACAACCACGCGTTTCCACGGCAGTCGGTTCTACGTTACCAAAAACAACCCGATTTATATCCGGCCAAGGATTGCCACTCCAGCAGCTTTCCCGTATGTAAAAGGGGAATGTTTATGCTGATACACAACAACAACAA

----- INTRON -----
SEQ PROT
SEQ GEN CCACGCTTGCCACGCGTTTCCACGGCAGTCGGTTCTACGTTACTGAAACAATCCGATTTATATCCGGCCAAGGATTCCACTCCAGCAGCTTTCCCGTATGTAAAATGGAAATGTTTATGCTGATACACAACAACAACAAAACCAG

----- INTRON -----
SEQ PROT
SEQ GEN CTTCCAGGACAATATCCTTTTCATGAATTGTTCCATAACCAAAATGCAAAGTGGCTGCCCTATATCCTCGATAGCCTCACGAATCCATCTTTGAGGTCTTGAATCCTCCCTGGGCTGTTGGCGTAAACCTTCTCATTACGTTGGCCCCAAAA

```



```

----- INTRON -----//----- INTRON -----
SEQ PROT
SEQ GEN AAAAAAGTTACAAGGTGTA AATTACAAA AACTCGGGGCAATTGTGATCCCTCTC//TTGTTTCTTGAAACTTTAAACGCGTTTTTTTTTGGAACTGGGTTTTCAAAGTCGGTTGTCAAAGTTTCTGAAA AACTATTCAACCGATCTTGATGAAA

----- INTRON -----
SEQ PROT
SEQ GEN TTTTCACCGGTGTTAGAGATAACAATTTATTCGTGCTTGAACGAAGGGTTTTTTTTTTTTTTTCAATTACAACAATTTAAAGAAAAACAAAATGTCGAGCAAATTTGACCGAAATTTTCATTTTTTGCATAAATGTCTGCCAAAATTCCAA

----- INTRON -----
SEQ PROT
SEQ GEN TTTTTACTTTTTTCTTTCCTTCGTTCAAGCTCGAGTTTATCGTCTTAACTAAAGCACTAATTTTTGTTTTTTCATTTTTGATGAGCCTGTCAGGAGTTATGCTGACAACCGGGACGCATCTTTTTTTCGAGGGGTCACCGGCTTGACGTCA

----- INTRON -----
SEQ PROT
SEQ GEN CAATGGAGGAGCTTTAATTTTTTGAAATTATCTTTAAACATGTATCTATAAGACAGAAAAAGATTGAATTAATAAATAATTTTTTACATAGAAAAAATATATTGAAAATAGGCCTTTTTTACTCGACAAAACCCATGTAACCTCCTTA

----- INTRON -----
SEQ PROT
SEQ GEN ATGCCAACCGACTAACTAACCTTATCTGGAAGCTTGAGACGTTAGCTGTAATAGAAATGTGTTATATTACGTCCAAGTCGAAAAAGCCAAACCGGGTGCCCAACGGCGGATTACATTTTTTTTTTTCGTTTTTCTTATAACTTTGAATTT

----- INTRON -----
SEQ PROT
SEQ GEN CGCATTTATTTAATTTCTTTATTTTTTAAACAGTGTCTCTGATTTAAAATTTGTAATACTTAAAATTAATGAATATAACGAATTGTCTAGTCTAGTTTTAAGTATTACGCATAAAATGCCGTAATTATTAATAAATAAAAATAAAATGCATTT

----- INTRON ----- | ----- EXON CDS3 -----
SEQ PROT
SEQ GEN ATATTCCTCAGATTATTAAGTCGCAATTATGAATAAATAACATAAAAATAAAGTGCATTTCTATTCCCTCAGATTATCAAGTCTAAATCGGAGATCGCTGCCTTGAAAGATTTTCTGGATTGGCTGGAAGCGCTTAAGGCCGAAGATACATCATC
I I K S K S E I A A L K D F L D W L E A L K A E D T S S

----- EXON CDS3 -----
SEQ PROT
SEQ GEN N G V I L V Y H E E R K F I P Y M I L E S L N K Y N L L D R F L K T V K S F A N G L N L A K A N A N E
GAACGGTGTCAATTCTAGTCTATCATGAGGAACGCAATTTATCCATATATGATTTTGGAATCGCTTAACAAATACAACCTGTTGGACAGGTTCCCTTAAACAGTCAAATCATTGCCAATGGCCTCAATTTAGCCAAGCCAATGCCAACGA

----- EXON CDS3 -----
SEQ PROT
SEQ GEN S L Q H Y T L R K L S K V L S A K A H E V I N N Q N A S D G E I K N N V S S T T G I A A E C Q S P Q K
ATCATTGCAACATTACACGTTGCGCAA AACTCTCAAAGTGTGAGCGCCAAAGCACACGAAGTGATCAATAATCAAATGCAAGTGACGGTGAAATCAAAAACAATGTTAGCTCAACAACGGGCATAGCCGCCGAGTGTGAGAGCCCCCAAAA

```

----- EXON CDS3 -----
SEQ PROT S L Q K G V L A Q E R E L F D G N A S V R A K L A Y E V A Q Q L S N Q E H K V P I E S S A A L E N M I
SEQ GEN GTCATTGCAAAAGGGCGTTCTGGCACAGGAGCGTGAGCTGTTCGATGGTAATGCTAGTGTGCGTGTCTAAGTTAGCTACGAGGTGGCACAAACAGTTGAGCAATCAAGAACAAGGTGCCTATCGAATCGTCCGGCTGCATTAGAGAATATGAT

----- EXON CDS3 -----
SEQ PROT A A V Q T Y A Q P I E C D L N E L Q T Q N E N L E R Q N S F R P V F L N Y F K T T L Y H R V R A V K F
SEQ GEN TGCCCGAGTGCAAACGTATGCACAGCCCATCGAATGCGATTTGAATGAACTACAACTCAAACGAGAATTTGGAGCGTCAAACCTCTTCCGTCGGTTTTCTTAAATTACTTTAAAACCACTCTACCATCGAGTGCCTGCAGTCAAGTT

----- EXON CDS3 -----
SEQ PROT R I I L A E Q G Y D L Q T L N A I W S E K Q K D G L A E V L Q V I E E L K P E D R T E L A D L F D S Y
SEQ GEN TCGCATCACTTGGGAACAGGGTTATGACCTGCAGACGCTAAATGCAATTTGGAGTGAGAAACAAAAGGATGGATTGGCCGAAGTATTGCAAGTCATTGAGGAGCTAAAACCTGAAGATAGGACGGAGTTGGCTGATTTATTTGACAGTTA

----- EXON CDS3 ----- | ----- INTRON -----
SEQ PROT F D P V K T T I K P I V K N N N R R R N R R G
SEQ GEN CTTCGATCCTGTAAAGACGACTATTAAGCCGATTGTCAAAAACAACAATCGCCGCCGAAATAGACGCGGAGGTAGTGCAACATGCATGCATTAAGTTAAAGTGTATTGAAAAGCAATTAATCATATGATTTTCATTTTTTCTTTTTTACTT

- | ----- EXON CDS4 -----
SEQ PROT V R N M K N D G S N S A R T S S E Q G A G G D K Q V T G G Q L P D T T I K S P S P P N G G K P Q R G K
SEQ GEN AGTACGCAACATGAAGAACGACGGCAGCAATAGCGCTCGCACAAAGTAGTGAGCAAGGAGCTGGCGGTGACAAGCAAGTCACTGGCGGTGAGCTGCCCCGACACCACAATTAAGTCGCCATCTCCGCCAAACGGCGCAAGCCGCAGCGTGGCAA

----- EXON CDS4 ----- | ----- EXON 3' UTR COMUM -----
SEQ PROT R N P R S R N N S L K T Q L P S A A V D T S N T T A T A T N *
SEQ GEN GCGTAATCCGCGCTCAGCAATAACTCCTTAAAAACACAGCTGCCAAGCGCCGCGTGCACACGAGCAATACAACGGCAACAGCAACCAATTAAAAACTAAGAAGTTCATTCTAAGATTCAAGATAGGCGCTGTTGCGACATAAAAAAACAA

----- 3' UTR A ----- | ----- EXON 3' UTR FÊMEA -----
SEQ PROT
SEQ GEN GTAATTTGTTTTATTTTTTTTTT-GTAATAGTTTTCCACCAACGTAGACACACTTTAGAGAGACGCCAGAAAAAATT-CTGGTATGCAAGTTGAATAGAATTTGAATTTTTAATTTTATATATGCCACATATTTTTATTATAATTATTATTACAT

----- EXON 3' UTR FÊMEA -----
SEQ PROT
SEQ GEN GTATTGATTAGCTAAACGTACATGCATATAATTAACAAAAGCAAAGAAATATGTATATATATATGTACACTTATTGAACGTTAGCAGTTACAATACCCAATTGTAGATGACGTTAGTTTAGAAAATATCTGATTATTGAAGTCATTAATAAT

----- EXON 3' UTR FÊMEA -----
SEQ PROT
SEQ GEN TATTATTTTATCTGCTTTGCTACGCTGCAAAGTACAACAACATAAATTAGTTTCCAGCCGAGGAGGTGTAAGCTTTTCATAATTTATTTTCGTAGTTTTACTAGAGTTGAGTTAAATAAAATGTAACAAATATTAGCTAACCTCTGGTT

----- EXON 3' UTR FÊMEA -----

SEQ PROT
 SEQ GEN TGACGTCTTAGAGGATTCTACAAGAATTTGACAACCTTGCTTATTAACGCGTATTGAAATTTGAATTTTGTAGAGGTTGCGCTttaatattatatttgtttttatgattcctattatgatatttatgtatttatgctattttgaaatttt

----- EXON 3' UTR FÊMEA -----|<?>|----- EXON 3' UTR MACHO -----

SEQ PROT
 SEQ GEN gtttaaaagaatacacagaatatataaacaatatttttttctaacttaaaaaaaaa---CTCAAATTTAAAGCGAAAAACCCACAGTTAGCTTCACAAGCAAAACACAAAAATTGATTGGTGCAGTTGCAGTTATACATACATAG

-----|

SEQ PROT
 SEQ GEN CAAAACGAAAAAAAAAAAA

OBSERVAÇÕES: (I) O símbolo “//” indica uma descontinuidade da sequência, pois essa região não foi sequenciada por inteiro; (II) a região da sequência em letras minúsculas constitui um fragmento ainda não sequenciado via Sanger, e foi retirada dos *contigs* gerados pelo sequenciamento dos transcriptomas; (III) O símbolo “|<?>|” diz respeito a uma região que pode ou não existir, e ser constituída de íntron e/ou outros éxons.

exu_smos ACCATTCCGAAGGGCAAATACACGGTTCATCGGGCGTGATATCGATACAACGGTCGTCGCGCATTTTGGATGAGATCGTACATTTGTCGGCCTACAGCCCAGGAGCAATTTGATCAGTACGTTATCCCAGTATGAAATTTAAATCCAGCA
 T I P K G K Y T F I G V D I D T T G R R I L D E I V H L S A Y S P K E Q F D Q Y V I P V M N L N P A
 exu_cape CCATTGCCATATGGAATTAATAAATTGATTGGAATCGATATCGATACGACTGGTCGTAGATAAATCGATGAAATAGTACAAAATTTCCGCTTACACACCAAAATTCACAATATCCCAATACATTATGCCGTTGATGAAATCGAATCCGGCC
 P L P Y G N Y K L I G I D I D T T G R R L I D E I V Q I S A Y T P N S Q Y S Q Y I M P L M N L N P A
 exu_aaeg CCGCTGCCCTACGGGAAGTACACCTTGATTGGCATCGACATTTGATACGACGGGTCGTAGACTGATCGATGAAATTTCCAGATCTCCGCTACAGCCCGAATCAGTACTCGCAGTACATCATGCCGTTGATGAACTGAATCCGGCC
 P L P Y G K Y T L I G I D I D T T G R R L I D E I V Q I S A Y T P E H Q Y S Q Y I M P L M N L N P A
 exu_agam CCGCTGCCGCAAGTACACTCTGATCGGGCTCGATATAGATACGACGGGTCGTAGACTTATCGATGAGATCGTACAGATCTCGGCTTACGCGGAGAAAGAGTATGCCGAGTACATTATGCCGTTGATGAACTGAACCCGGCT
 P L P H G K Y T L I G L I D I D T T G R R L I D E I V Q I S A Y T P E K E Y A Q Y I M P L M N L N P A
 exu_ppap TCCCTCCGGAGGGCAATTACCAATCATCAGCGTGGACATTTGATACGACGGGAGACTCATTGATGAGATTTTCCCTGGCCGCCATCCAGTCCAGYACTCCGCAATACATTGACCATGAACTCAATCCGGCA
 S L P E G N Y T I I S V D I D T T G R R L I D E I V H L A A Y T P D S Q F S Q Y V M P Y M N L N P A
 exu_crip GCCAATTTGGAAGGAAGGTACAAGTATGTTGGATTGATGTTGACACGACTGGAAGGCGTTGATTGACGAGATTTGTCGAGATTTGATATACACCAGATGAGCAGTTTTACAGCACATTATGCCCTTAATGAAATTTAAATCCAGGA
 A N L E G R Y K Y V G F T D V D T T G R R L I D E I L Q I A A Y T P D E Q F S Q H I M P L M N L N P G
 exu_bant ACTGAAATGAAGGGACCTTCTCTACGTAGGGTTGGACATCGACACGAGCTGCGCTGATGACGAGATCGTGAACATCGCCGCTATCGTCCCGATGCCGAGTACTCGCAGCACATCATGCCGTTGATGAACTGAACCCGGGT
 T E M K G D F V Y V G L D I D T T G L R L I D E I V N I A A Y R P D A Q Y S Q H I M P L M N L N P G

310 320 330 340 350 360 370 380 390 400 410 420 430 440 450

exu_afra GTCGCAACGTCACCAAGTGCCTGATCTATTGGATTCTATCGTATGTTGAAATCAATGCAAACTATAAG
 A R Q R H Q V R V I S I G F Y R M L K S M Q T Y K
 exu_ccap GTCGTCACCGTCACCAAGTGCCTGATCTATTGGATTCTATCGTATGCTAAAATCAACGCAACCTATAAG
 A R Q R H Q V R V I S I G F Y R M L K S T Q T Y K
 exu_mdom GTCGTCACCGTCATCAAGTTCGCTGATTTCCATTGGCTTCTATCGCATGCTCAAGTCTATGCAAACTATAAG
 A R Q R H Q V R V I S I G F Y R M L K S M Q T Y K
 exu_dmel GTCGCGAGCGTCATCAAGTTCGCTGATTTGATTTGATTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
 A R Q R H Q V R V I S I G F Y R M L K S M Q T Y K
 exu_dmoj GCCCGTCAGCGTCATCAAGTTCGCGTATTTGATTTGATTTTTCGATGCTGAAGTCGATGCAGACCTACAAG
 A R Q R H Q V R V I S I G F F R M L K S M Q T Y K
 exu_dvir GCCCGTCAGCGTCATCAGGTTCCGCTATTTGATTTGATTTTTCGATGCTGAAGTCGATGCAGACCTACAAG
 A R Q R H Q V R V I S I G F F R M L K S M Q T Y K
 exu_dps1 GTCGTCAGCGTCATCAAAATCGTGTGATTTGATCGGCTTTTATCGCATGCTGAAGTCGATGCAGACCTACAAG
 A R Q R H Q I R V I S I G F Y R M L K S M Q T Y K
 exu_smos GCACGGCAACGACCAATTCGCGATAATTCAGGTTGGATTTTCCGATGCTAAAATGTACATTGACCAACAAA
 A R Q R H Q L R I I T V G F F R M L K C T L T N K
 exu_cape GCCCGTCACCGCCATCAAGTGCCTGATAACCGTCGGATTCTTTCGCATGTTGAAAAGTATGCAAACTATAAA
 A R Q R H Q V R V I T V G F F R M L K S M Q T Y K
 exu_aaeg GCCCGTCACCGCCACCAAGTTCGCGTCAATACCGTTGGATTCTTCCGGATGCTGAAAAGCATGCAGACTTACCGG
 A R Q R H Q V R V I T V G F F R M L K S M Q T Y R
 exu_agam GCCCGTCACCGCCACCAAGTTCGCGTCAATACCGTTGGATTTTTCGCATGCTGAAAAGCATGCAGACCTACCGT
 A R Q R H Q V R V I T V G F F R M L K S M Q T Y R
 exu_ppap GCCAGGCAACGACATCAGGTCGCTCATCACCATTTGGCTTCTCCGGATGCTAAAAGCATGCAGACGTACAAG
 A R Q R H Q V R V I T V G F F R M L K S M Q T Y K
 exu_crip GCAAGGCAACGCCATCAAGTGCCTGATAAATGTTGGAAATTTCCGATGCTTAAATGTATGAAAAGTTATAAG
 A R Q R H Q V R V I N V G N F R M L K C M K S Y K
 exu_bant GCCCGTCAGCGTCATCAAGTGCCTGATTTAAGTTCGCTGAGATTTGAAATCCATGCAAACTACAAGTGAGAACTTTCGAAAACCGTGCACATAACCAAGTCGGATTCTAATCGCAATCGCCTTTGCCCTTCCCTTCCCT
 A R Q R H Q V R V I N V G C F R M L K S M Q T Y K V R T C E T R A H N Q S R I L I A I A F A F P L P

460 470 480 490 500 510 520 530 540 550 560 570 580 590 600

exu_afra A T T A T C A A G T C T A A A T C G G A G A T C G C T G C C T T G A A A G A T T T T C T G G A T T G G T G G A A G C G C T T A A G G C C G A A G A T A C A T C A T C G A A C G G T G T C A T T C T A G T C T A T C A T G A G G A A C G C A A A T T T
 I I K S K S E I A A L K D F L D W L E A L K A E D T S S N G V I L V Y H E E R K F

```

exu_ccap -----ATTATCAAGTCCAAATCGGAGATCGCTGCCTTGAAGATTTTCTTGATTGGCTGGAAGCGCTTAAAGCCAAG--GATGCAGCATCACAAGGTATAATTCAGTCTATCACGAGGAAACGAAAATTC
I I K S K S E I A A L K D F L D W L E A L K A K D A A S Q G I I L V Y H E E R K F
exu_mdom -----ATCATTAAATCAAAATCTGAAATCGCTGCCTCAAAAGATTTTCTCGCTTGGCTCGAAGCTTTGAAAGCTAAA--GATGCCACATCGCAAGGTATTTATTTTATTGATCATCAAGAGCGTAAATTC
I I K S K S E I A A L K D F L A W L E A L K A K D A T S Q G I I L L Y H E E R K F
exu_dmel -----ATCATTAAATCCAAAGTCTGAGATCGCTGCCCTCAAGGACTTCTCAACTGGCTTGAGCAACTGAAGACCAAGGGGGTCCGACTCGGACGGGATTGTGCTTATCATCAAGAGCGCAAGTTTC
I I K S K S E I A A L K D F L N W L E Q L K T K A G P S S D G I V L I Y H E E R K F
exu_dmoj -----ATCATCAAAATCTAAATCAGAGTTCGCTGCCTAAAAGACTTCTTTCAGTGGCTGAGTTCAGTTAAAATCGAAA--AATCCAAATCGGGATGGCATTGTCTTGTCTTACCATGAGGAGCGCAAAATTT
I I K S K S E V A A L K D F L D W L D Q L K S K N P N A D G I V L L Y H E E R K F
exu_dvir -----ATCATCAAAATCTAAATCAGAGTTCGCTGCCTAAAAGACTTCTTTCAGTGGCTGAGTTCAGTTAAAATCGAAG--AATCCGAAATCGGATGGCATTGTCTTACTTTACCATGAGGAGCGCAAAATTT
I I K S K S E V A A L K D F L D W L D Q L K S K N P N T D G I V L L Y H E E R K F
exu_dps1 -----ATTATCAAAATCTAAATCGGAGTTCGCTGCCTCATGGACTTTCTCAACTGGCTTGAGATGCTGGTCGCCAAA--CAGCCCAGCACTGATGGCATCGTTATGCTCTATCACGAGGAGCGTAAATTC
I I K S K S E V A A L M D F L N W L E M L V A K Q P S T D G I V M L L Y H E E R K F
exu_smos -----GTGGTCAAAATCGAAAACGGAGTTCGCGTTGAAAACATTTTGGACTTGGAGGAGCCGAAAGCTGCT--GACGCTGCATCAGAAGGGGTGGTTTGGTTTATAATGAACCAATCAAAATTC
V V K C K T E V V A L K T F L D W L E Q R K A A D A A S E G V V L V Y N E P I K F
exu_cape -----GTTGTTAAAACAAAAACCGAGGTCATCGTGCCTCAATGAGTTTATCGACTGGTTGAAAAGGAAAAATCAAAAAATGGCGAAAATTCGATGGCGTTGATTGATTATCACGAACAGAGAAAATTC
V V K T K T E V I V L N E F I D W L E K E N Q G K N G E N S D G V V L I Y H E Q R K F
exu_aaeg -----GTGGTGAAAACCAAGACCGAAATCTCCGCCCTGAACGAATTTCAACTGGCTGGAAGAGCGGTAACGGKAG--GACGAAGGTTCCGAAGGCATTGTCTTGATCTACCATGAGCAGCGCAAGTTT
V V K T K T E I S A L N E F I N W L E E R Y R E D E G S E G I V L I Y H E Q R K F
exu_agam -----GTGATGAAAGCCAAAGTGAAGTTCGCCCTCGTGGACTTTCTGGGGTGGCTGGAAGAGCGGTAACGGKAG--GACGAGGGTCCGAGGGCATTGTACTGGTGTATCATGAGCAGCGCAAGTTT
V M K T K M E V A A L N E F L D W L E E R L R E D E G S E G I V L V L Y H E Q R K F
exu_ppap -----GTGATGAAAGCCAAAGTGAATATGCTGCCCTCGTGGACTTTCTGGGGTGGCTGGAAGAGCGGTAACGGKAG--CAACGAAGTCCGAAGGGCTCATCTCCTCTACCAGAACGAGGAAAATTC
V M K T K P E Y A A L V D F L G W L E E Q K A K Q Q D S K G L I L L Y H E Q R K F
exu_crip -----GTTATTTAAAACAAAAACCGAAATAGCAACATTGATGGACTTTATTTGCTGGCTTGAGAAAATTT-----AATGATGAGAAAAGATGGAATCATTTTGTTATATCATGAACAACTTAAGCTT
V I K T K T E I A T L M D L L C W L E K I N D E K D G I I L L Y H E Q S K L
exu_bant GTTCGTTCTTTTGTCTCCATTCAAGTCAATTAACCAAGACCGAGATCGCAGCACCTCCTTGAGTTCTCGATTGGTTGGACAAGATCAACACCAAG-----GACGGTGTATCTTGTGTTACCATGACAAGCTCAAGTTC
V R S F V S I Q V I K T K T E I A A L L E F L D W L D K I N T K D G V I F V Y H D K L K F

```

```

610 620 630 640 650 660 670 680 690 700 710 720 730 740 750
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
exu_afra ATTCCATATATGATTTTGAATCGCTTAAACAAATACAACCTGTGGACAGGTTCCCTTAAAAAGTCAAAATCATTGCGCAATGGCCTCAATTTAGCCAAAGCCAATGCCAACGAA-----TCA---TTGCAACATTACAGTTGCGCAAA
I P Y M I L E S L N K Y N L L D R F L K T V K S F A N G L N L A K A N A N E S L Q H Y T L R K
exu_ccap ATTCCATATATGATTTTGAATCGCTTAAAAAATACAATCTGTTAGATAGATTTCTTAATTCAGTCAAAATCTTTCGCAATGGCCTCACCCCTGGCCAAAGCCAATGTCAACGAG-----TCC---TTACAGCATTATTCATTGCGTAAG
I P Y M I L E S L K K Y N L L D R F L N S V K S F A N G L T L A K A N V N E S L Q H Y S L R K
exu_mdom ATTCCATATATGATTTTGAATCGCTTAAAAAATATAAACCCTGTTGGACAGGTTTTTACTTACTGTTAAGTCAATTCGCTAACGGTCTCAATTTGGCCAAAGCCAATGTCAATGAG-----TCC---CTGCAACATTATACATTGCGAAAA
I P Y M I L E S L K K Y N L L D R F L L T V K S F A N G L N L A K A N V N E S L Q H Y T L R K
exu_dmel ATTCCCACATGATCCTGGAGTCGTTGAAGAAGTACGGCCCTGTTGGACGGCTTACCAGCGTTCGGTGAAGTCTTTGCAAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT-----GCCAACATAAAGAACTACAGTCTGCGAAAG
I P Y M I L E S L K K Y G L L E R F T A S V K S F A N S I N L L A K A S I G D A N I K N Y S L R K
exu_dmoj ATTCCATACATGATTTTGAATCGCTGGTCAAGTACGGCATGCTCGATAGATTCACCAAGACAGTCAAGTCTTTTGTAAACAGCTTCAATCTGGCCAAGTCCACAGTTGGCGAT-----ACC-----AAACACTATAGCCTACGCAAT
I P Y M I L Q S L V K Y G M L D R F T K T V K S F V N S F N L A K S T V G D T K H Y S L R N
exu_dvir ATACCCATATATGATTTTGAATCGCTCAAGTATGGCATGCTCGATAGATTCACCAAGACAGTCAAGTCTTTTGTAAATAGCTTCAATCTGGCCAAGTCCCTCAATGGCGAC-----ACC-----AAGCACTATAGCCTCGGAAT
I P Y M I L Q S L V K Y G M L D R F T K T V K S F V N S F N L A K S S I G D T K H Y S L R N
exu_dps1 ATACCCACATGGTATTGGAAGCGCTCAAGAAGTACGGTCTGATTGAAAGATTCAATCGGACGGTGAAGTCTTTTGTCAACACTTTCAATATGGCCAAGGCTCCCTTGGCGAT-----GCTAATCTCAAAAACGTGGTCTACGCAAG
I P Y M V L E A L K K Y G L I E R F N R T V K S F V N T F N M A K A S L G D A N L K N C G L R K
exu_smos ATTCCGTACATGCTGCTCGAAGCATTGAAAAAATACGATTTGGTTGATCGATTTTCTCAGTGGTTGTCGGTTGCTGGTATGGCCATGCTTTGGCTGAAGCCAAA--GGCGAT-----GCC---ATCAATATTGAGGCGTACGCGAA
I P Y M I L E A L K K Y D L V D R F S Q L T V G T F V D G H A L A E A K G D A I N A T I E A V R E
exu_cape GTGCCGTACATGATGATTGAGCGCTGAAAAAATAAATTTATTGGATCGTTTTACGAAAACCTGTAATAATCATTTCGAAATGGTTATAAATGGCCGAAGAGAAGTGCAGCAAAGAGGTCACA--ATCAAAATATTCTCGATTGCTCAA
V P Y M M I E A L K K Y N L L D R F T K T V K S F A N G Y K L A E E K C S K E V T I K Y F S I R Q
exu_aaeg GTCCCATACATGCTGATTGAGCCGTGAAAAAGTACAACCTCCTGGATCGCTTTAAATCCATTTGTAAGTCTTTTCTAACCGTTTCAAGTGGCCGAGGAGAAGTGCAGCAAA-----ACC---ATCAAGTACTTCTCGATCCGACAG
V P Y M L I E A V K K Y N L L D R F T K S I V K S F A N G F K L A E A K C S K T I K Y F S I R Q
exu_agam GTGCCGTACATGGTCAATTGAGCGCTGAAGAAGTACAAGCTCCTGGACCGCTTCTCCGAGTCCGGTGAAGTCAATTTGTGAATGGTTCAAATTTGGCCGAGGAAAAGTGTACCAAG-----ACC---ATCAAGTACTTGAAGATTGCTCAG
V P Y M V I E A L K K Y K L L D R F S E S V K S F V N G F K L A E E K C T K T I K Y L T I R Q

```

```

exu_ppap GTACCCACATGCTCCGAGGCCCTTAAAGAAATACGACATGCTTCCGCGGTTTGAAAAATCCGTTGCCGGGATTTGCCAATATGTACAGCCTGGCCGAGAGGAGTCCGGGGCC-----ACT---CTGAAGTACTTCACGCTCAGGGAA
V P Y M L L E A F K K Y D M L P R F E K S V A G F A N M Y S L A E K E C G A T L K Y F T L R E
exu_crip GCACCTTACATGCTTATTGAGTCAATGAAAAATTTAACTCTTTGATAGATTCAATAAGATAGTCAAAGGATTTGTCAATGGATATGAACATATGTTGGAGCAGATCAGAAAAGGA-----AAAGGATTAATAATCTGACATTGGCTCAA
A P Y M L I E S M K R F N L F D R F N K I V K G F V N G Y E L C G D D Q K G K G L K Y L T L A Q
exu_bant ACGCCTTATATGGTGATCGAGGCAATGAAGAAGTACCGGCTCATGGAGCGCTTCGAGAAGATCGTGAAGTCTTCCGTCATGGCTACGACTTTGGCAGGCGAAAAGCQGGCGAG-----ACCGGAATCAAGTACCTGACTTTGTCTCAG
T P Y M V I E A M K K Y R L M E R F E K I V K S F V N G Y D L A G E K P G E T G I K Y L T L S Q

          760      770      780      790      800      810      820      830      840      850      860      870      880      890      900
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
exu_afra CTCTCCAAAGTG-----TTGAGCGCCAAAGCACCGAAGTGATCAATAATCAAAATGCAAGT-----GACGGT---GAAATCAAAAAACAATGTT-----AGCTCAACCAACGGGCATAGCCGCC
L S K V L S A K A H E V I N N Q N A S D G E I K N N V S S T T G I A A
exu_ccap CTTTCTAAAGTC-----TTGAGTACTAAATCAGAAGGA-----GAAAACAAAAATACCGGC-----GATGGTGTGAGATAA'TTAATGTAATT-----TCAGCAAA'TACAAACAACGCCAGT
L S K V L S T K S E G E N K N T G D G A E I I N V I S A N T N N A S
exu_mdom CTATCCAAGTTT-----TTGAAATCAAACGCTGTGAACAGTGTGAA'TTAAACAAC-----AATAATGTGGAGGCATCGTCCACCAC'TGAAGGTGCTGCAGCAGTTGCCACTAAAGAG
L S K F L K S N A V N S V E L N N N N V E A S S T T E G A A A V A T K E
exu_dmel CTTTCCAAGATC-----TTGTCCACGACCAAGGAGGAAAGACGCGCTTGCTCTGCGTCCACATCTGGATCTGGATCTGGCTTGGGATCTGGATCTAGTATGTTGTCGGATTCCGGTA-----TCGATATCCCCCAGGACAGTACC
L S K I L S T T K E E D A A C S A S T S G S G S G L G S G S S M V S D S V S I S P R D S T
exu_dmoj CTGTCCAAGATA-----TTGAGCAAGGCCAAGGAGGAGCCGCAAGGGCAAT-----GGCAACGATAATGAAGCTGACAGTAAC
L S K I L S K A K E E T A K G N G N D N E A D S N
exu_dvir CTATCCAAAGATA-----CTGAGCAAGGCCAAGGAGGACAACGCCAGATCTGAT-----AATGATAATGAGGCAGATAGTAAC
L S K I L S K A K E D N A R V D N D N E A D S N
exu_dps1 CTTTCTGCTC-----TTGGCCAAGTCGAATGATGAAAGTGCCAACAAGGAG-----AATGAACCTGAGAATGTCAACAGAAAC
L S L L L A K S N D E S A N K E N E P E N V N R N
exu_smos ATGGTCAAAGGA-----CAGCTAAAAGCCGAAGATCAG
M V K G Q L K A E D Q
exu_cape TTGGCAAAATTG-----GTGATGGATGAAAATGATTCC
L A K L V M D E N D S
exu_aaeg CTGGCCAAGCTG-----GTGCTGGACGAAAGCGAGGCG
L A K L V L D E S E A
exu_agam CTGGCAGAGATA-----GTTTTGGACGAAAAGGCAGGA
L A E I V L D E K A G
exu_ppap CTGTCCAAGTG-----ATTTTGAACCGGAAGGGAT
L S K V I L N T E E D
exu_crip AATTACAAAGTTTCATGCTGATTGCTTAAAAATGGATGTAAAG-----
N Y K V H A D C L K M D V K
exu_bant AACTACAAGATTTCAGCCGAGAGCTGAACATGGACAACATT-----
N Y K I Q R Q K L N M D N I

          910      920      930      940      950      960      970      980      990      1000     1010     1020     1030     1040     1050
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
exu_afra GAGTGTACAGAGCCCCAAAAGTCATTGCAA-----AAGGGCGTTCG---GCACAGGAGCGTGAGCTGTTTCGATGGTAATGCTAGTGTGCGTGTAAAGTTAGCCTACGAGGTGGCACAACAGTTGAGCAATCAAGAA
E C Q S P Q K S L Q K G V L A Q E R E L F D G N A S V R A K L A Y E V A Q Q L S N Q E
exu_ccap -----CAAAGTCCACAAAATCTTTAAATATAATTAAGCGCTACA---GCACAGCAACGTTGAATTTGTTGATGGCAATGCTAGTGTGCGCGCAAGTTGGCCTTTGAGGTAGCACTGCAGTTGAGTAAACAGAG
Q S P Q K S L N N I K S A T A H E R E L F D G N A S V R A K L A F E V A L Q L S N Q E
exu_mdom -----CAACAGCTGCAACAAAAGACTAACATAAGTCTAGTCAAAAC---AATAATGAAGGTGAATTAATTCGATGGAAGTGCCAGTGTTCGTTGCCAAATAGCCTATGAGTGGCTTTTGCAATTAAGTAATCAAGAC
Q Q L Q Q K T N N K S S Q N N N E G E L F D G S A S V R A K L A Y E V A L H L S N Q D
exu_dmel GTGACCAATGGTGACGATAAGCAATCTAGC-----AAGAACGCCGTT---CAGGGTAAGCGGGAAC'TGTTGATGGCAATGCCAGTGTTCGTTGCCAAATTTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACTCCGAC
V T N G D C D K Q S S K N A V Q G K R E L F D G N A S V R A K L A F D V A L Q L S N S D
exu_dmoj -----AGTAGCAGCACTGATAAGCATGTA-----AAGAAATGGATCCAAACAAAAGGAGCAGATGAGTTCCGATGGCAATGCCAGCTCCGTTGCCAAGCTAAACCTTCAATGTGGCTTTGCAAGCTAAGCAATTTGGAC
S S S T D K H V K N G L Q Q K E R D E F D G S A S V R A K L T F N V A L Q L S N L D

```



```

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320 1330 1340 1350
exu_afra ..... CACAAGGTGCCTATCGAATCGTCG GCTGCATTAGAG
exu_ccap ..... AATAAGGCGCCATCGACTCGACC GATGCATATGAT
exu_mdom ..... TCGAAGGAAGATCTCGATTCTCAA AATGCTCTCGAT
exu_dmel ..... GGCAAGCCCGAACCCAAAAGCTCG GAGGCATTGGAG
exu_dmoj ..... GTTAGCAGC AACACCGCTGAGCCGGATTTCATCAGCGGAGGCCTTGAGC
V S S N T A E P D S S A E A L S
exu_dvir ..... AGCACCGGACCCAGAGTCCCTCG GAGGCCTTGAGC
S T E P E S S E A L S
exu_dps1 ..... CGCACATCCGAGCCAGAGTCTTCG GAGGAATTGGTT
R T S E P E S S E E L V
exu_smos ..... GCAAACGC AAAACTGAAGACATCATCAGTTCC
A N D K T E D I I S S
exu_cape ..... GAAAAAATCAGGCTACAATTACTGTTACAGAAGCTCCAAATGAC ACTTCCGCCGATGCCACAAAAACCGATAATGAAACACAAAACACGTTACACACAAGAATGAATTAACGGATGAA GAAATGTACGAA
E N N Q A T I T V T E A P N D T S A D A T K T D N E T T K H V T H K N E L T D E E E M Y E
exu_aaeg AACCGGAGAACAAATGGTGGTCCCGAGGGTCAGAACGACTCCGATGACGATCGGCATCAAACCGATGTGGACGCTGTCCCGGCCAAGTGACGACAAATCCGGAAACCGCAACCGATCCCGTCCAGCCAG GCGGAGCCGAG
N A E N N G G P Q G Q N D S D D D R H Q T D V D A V T G Q V T T I P E P Q P R S R P T T E A E R E
exu_agam AATGCTGAA CAATCGGATGACGATGCGCAGAAGCAGGACGAGAAGAAGGAAGCTCCTGCTGCTTCTTTGACAGC TCTGCATCTCGTCCCTTGACTGAA GAGGATCGTGAA
N A E Q S D D D A Q K Q D E K K E E A P A A S L N S S A S R P L T E E D R E
exu_ppap ..... CCACAGTCTGAGCAAAAAGCTGAGCCAAAAGCTGAGCCAAAAGCTGAGCCAAAAGCTGAG GATAAAGATTCT
P Q S E Q K A E P K A E P K A E P S D E D K N S
exu_crip ..... GAGATTAAGGAACAAGACGAGAAG GAAATTTATGAT
E I K E Q D E K E I Y D
exu_bant ..... GAGAAGAAAGAGTTCGAAGAGAA GCAATGGCCCGC
E K K E V E E E A M A A

```

```

1360 1370 1380 1390 1400 1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
exu_afra AATATGATTGCCGAGTGCAAAACGATGCACAGCCCATCGAATGCGATTTGAATGAACACAAACTCAAACGAGAAATTTGGAGCGTCAAACCTCTTTCCGTCCGGTTTTCTTAAATTACTTTAAACCACACTTACCATCGAGTGGCT
N M I A A V Q T Y A Q P I E C D L N E L Q T Q N E N L E R Q N S F R P V F L N Y F K T T L Y H R V R
exu_ccap AAGATGCTCAGCGATATACAAGCGTATGCACCACCCGTCGAATGTGATTTGAATGAACGACAGACAAAAATGAGAAATTTAGAACGTCAAAACCTCGTTCCGTCCAGTGTTCCTTGAATTTATTTCAAACCTACACTTATCATCGGGTGGCT
K M L S D I Q A Y A P P V E C D L N E L Q T Q N E N L E R Q N S F R P V F L N Y F K T T L Y H R V R
exu_mdom AAAATTTTAAACGGCCATAAATCCTTATGCCCAAACCATCAATGTGATTTGGATGAGTTGCAAAACCCAAAATGAAAATCTGGATCGCCAAAATTCATTCCTGCCAGTATTTTTTGAATTTCAAGTCAACACTTTACAATCGTGTGGCT
K F L T A I N P Y A Q T I Q C D L D E L Q T Q N E N L D R Q N S F R P V F L N Y F K S T L Y N R V R
exu_dmel AATATGTTTAAATGCCATACGTCCGTTTGCCAAAGTTGGTGGTGTCCGATGTCTTGGAGCTGGATATCCAGATCGAAAAATTTGGAGCGTCAAACCTCCTTCCGTCCGGTGTTCCTGAACACTTTCAAGACCCTCTGTACCATCGTGTGGCT
N M F N A I R P F A K L V V S D V L E L D I Q I E N L E R Q N S F R P V F L N Y F K T T L Y H R V R
exu_dmoj AATCTGTTTAAATGCTCTCAAGCCCTTCACTGAGGCATCAGCGGTGATGTCAAGGAGCTGGACACACAGAATGTTTCATTTAGAGGACAAAAATTCGTTTTCGTTCCGTCTTCTTGAACATAATTTCAAGACCACGCTCTATCATCGTGTACGT
N L F N A L K P F T E A I S G D V K E L D T Q N V H L E R Q N S F R P V F L N Y F K T T L Y H R V R
exu_dvir AATCTGTTTAAATGCACTCAAGCCCTTCAAGGAGCCATTAGCGGTGATGTCAAAGACTGGACACACAGAATGTGCACCTGGAACGTCAAAATTCGTTCCGTCCCGTATTCCTTGAACATAATTTCAAGACCACGCTCTATCATCGCGTGGCT
N L F N A L K P F T E A I S G D V K E L D T Q N V H L E R Q N S F R P V F L N Y F K T T L Y H R V R
exu_dps1 AACCTCTTCAATGCCGTGAAACCATTCGCCAAAATTGTCAGCTCTGACATCATGGAGCTGCAAAACCCAGAACGAGAACATGGGCCGCCAAAACCTCGTTCCGTCTCTTCTTGAATTTTCCGCAACACTCTGTACCATCGTGTTCGT
N L F N A V K P F A K L V S S D I M E L Q T Q N E N M G R Q N S F R P V F L N Y F R T T L Y H R V R
exu_smos GATTTGGTTCAAGGATTTGTTCAATTCGGTCAATGTGGACATAAATGAATTGACGAACCCAGAACAATGTATCGATTTGCAACACACATTCGCCGCCAAATTTTCGTCGAATTCCTCAAACCAAACTCTATCATCGCTATAAG
D L V Q G F V H S V N V D I N E L T N Q N K C I D L Q H T F R P I F V Q F F K T N L Y H R Y K
exu_cape AAAATGATTAATATTTTGTGATTTAGCTTCGCCAATTAACAAAGAAATTACTGAACGTGAATGAACAAGAACAAATCTTTTACGACAAAAATTCATTAAGATCTGTTTTTATTCAAATTTTAAACCAACACTTTATCATAGAGTTAAA
K M I N I L C D L A S P I T K E I T E L N E Q E Q I L L R Q N S L R S V F I Q Y F K T T L Y H R V K

```

```
exu_aaeg AAGATGTGCTCGGTGCTGTGCGACCTGGCCTCACCGATCACCGACGAAATTTCCGAGCTTGACGAGCAGGAGAAGATTCTGATTGCCAGAAATCTCTCCGGCCGGTGTTTTTGTCAGTACTTCAAGACGACTATTTATCATAGAGTTAAG
K M C S V L C D L A S P I T S E I S E L D E Q E K I L I R Q N S L R P V F L Q Y F K T T I Y H R V K
exu_agam AAGATGTGCGCAGTGTGTGCGAATATGTTCCACCGATTTCCACTGAAATCTCCGAGCTTGACGAAACAGGAGAAGATTCTGTGCCAGAAATCACTGAGACCGGTATTTTTGCTCTATTTTAAGACGCAATCTACCATCGGGTTAAG
K M C A V L C E Y A S P I S T E I S E L D E Q E K I L V R Q N S L R P V F L L Y F K T T I Y H R V K
exu_ppap GTTTTTCATCGTGTGATCTGGGCATGAGCAGCAGCTGGAGAAGAAATAGCGTGAACTGAGTGGACAGCAGACCATCCTGGATCGGCAGAAATCACTGAGGTCCATCTTCATCAACTACTTCAAGACGACGCTCTACCATCGCGTGAAG
V F H R V I L G M S S S L E K N K R E L S G Q Q T I L D R Q N S L R S I F I N Y F K T T L Y H R V K
exu_crip GCAATGAACAAGTTTTGATCGAATAAAGCTGCCCCAATTGATCGTGAATAGATGAATGGTTGAAATGGAAAGAGACATCACTCGTCAAAACAGATATGCGAGATATTTTTCTAAACATATTTTTCAACATCTCGTTTTCTCGTCTCGC
A M N K F F V S N C K A R P I D R E L D E L V E M E S I T R Q T D M R D I F L T Y F S R S R R R R
exu_bant ATTTGCAACCGATTTCAATCCGACCAATGGCCCTCCAAATCGACCGCAATGGATGAGATCCAGGAGCAGGAGAAATGCATCAAGCCAGCTCAACCTTGCCGCGACATCTTCCGAGGACACTTCCAGACTTCGCGTTACCATCGTCCCGC
I V N D F I R P M A R P I D S E L D E I T E Q E E C I K R Q S T L R D I F R G H F Q T S R Y H R R R
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600 1610 1620 1630 1640 1650
exu_afra GCAGTCAAGTTTCGCATCATACTTGGCGAACAGGGTATATGACCTGCAGACGCTAAATGCAATTTGGAGTGAGAAAACAAAGGATGGATTGGCCGAAGTATTGCAAGTCAATTGAGGAGCTAAAACCTGAAGATAGACGGAGTTGGCTGAT
A V K F R I I L A E Q G Y D L Q T L N A I W S E K Q K D G L A E V L Q V I E E L K P E D R T E L A D
exu_ccap GCCGTTAAATTCGCTATTATCTGGCAGACAGGGTTCGACTCTGCAACATTAATGAAATTTGGAGTGAGAAAACGGAAGACCGGTTTGGCCGAAATATTGAGGTTATTGAAAGAGCTCAAGACTGAAGATAAAATCGATTGGCAGAT
A V K F R I I L A E Q G F D L Q T L N E I W S E K R K D G L A E I L Q V I E E L K T E D K T E L A D
exu_mdom GCTGTTAAATTCGCTATTATCCTGGCAGAGCATGGTTATGATTTGCAAACTCTCAATGCCATCTGGACTGAAAAGAAACGTGATGGCCTTGTGAAATTTGTCAGGCCATAGAAGAGCTGAAGGATGAGGACAAAACAGAAATGGCTGAT
A V K F R I I L A E H G Y D L Q T L N A I W T E K K R D G L A E V L Q V I E E L K D E D K T E L A D
exu_dmel GCCGTTAAATTCGCTATTATCTGGCAGACAGGGTTCGACTCTGCAACATTAATGAAATTTGGAGTGAGAAAACGGAAGACCGGTTTGGCCGAAATATTGAGGTTATTGAAAGAGCTCAAGACTGAAGATAAAATCGATTGGCAGAT
A V K F R I V L A E N G F D L N T L S A I W A E K N I E G L D I A L Q S I G R L K S K D K A E L L E
exu_dmoj GCCGTTCAATTCGCTATTGCTGGCCGAGAAATGGCTTCGATCTAACCCACGCTTAATGATATTTGGACTGAGAAGCGTACCAGGGGGTAGATGCTGCCCTGCAAGGCATTGCCAGCCTCAAGACCAGGATAAAAACGAGCTTGTGAG
A V K F R I V L A E N G F D L T C T L N D I W T E K R T E G L D A A L Q G I A S L K T E D K N E L V E
exu_dvir GCCGTTCAATTCGCTATTGCTGGCCGAGAAATGGCTTCGACTTTAACTTGAATGATATTTGGACTGAGAAGCGTATCGAAGGACTGGATGCCGCCCTGCAGGCCATTGCCAGCCTGAAGGCCGAGGATAAAAATGAGCTGGTCGAG
A V K F R I V L A E N G F D L T T L N D I W T E K R I E G L D A A L Q G I A S L K A E D K N E L V E
exu_dps1 GCTGTGAAGTTCCGCTATGCTGGCTGAAAATGGCTTTACTTTGGATTCTCTAAAAGCCATTTGGACCGAGAAGCGCAAGGAGGGCTTGAACGGCTCTGACCAATATTGATACCCCTGAAGACCAGGAAAAGACCGAGCTGGTGGAG
A V K F R I V L A E N G F T L D S L K A I W T E K R K E G L E L A L T N I D T L K T E E K T E L V E
exu_smos GCCGTTACATTCGCTGTTTTCGCGAAAAGGCTTCGACTCTGAAGACTCTCGAAGCAATTTGGAATGGAAGAGAAGGATGGTTGACTGAGATAATTTGGAAGATGGATGAGATCAAGGAAGAGAAATCGCAGTGAATTTGGTTCGAT
A V T F R R V L A E K G F D L K S L E D I W N G K K D G L T E I I S K M D E I K E E N R S E L V D
exu_cape GCTGTAACTTATAGACGAGTTTTGGCTGAATGTGGATACGATTTTCGATTCTATAAAGCAAATTTGGACGAATAATAGGAGGAA---TTGCAACAGGCATTGGATAAATAGAAGATTGAAGAGAAGAACGCAGGAAATTTATTTGAA
A V T Y R R V L A E C G Y D F D S L K Q I W T N N K E E L Q Q A L D K L E E L K E E E R K E L F E
exu_aaeg GCTGTAAAGTACAGACGAGTGTGGCCGAAATACGGACACGATTTTCGCTCCAGCAGGTTTGGAAACGACAGCAAGCGAGGGTCTGGAGGCCGTTGTTAACAGATCACCGAATTTGAAGGAGGAAAGCGCACCGAGCTGGTCGAG
A V T Y R R V L A E Y G H D F A S L Q Q V W N D S K R E G L E A V V N K I T E L K E E E R T E L V E
exu_agam GCTGTTACTTATAGACGAGTACTGGCAGAACTGGCCACGATTACGAACTTTTCCGTCAGTCTGGCAAGAGAAGAGCGCAAGGCTTGGAAAGCGATCAACAGAGATTGCAGAGTTGAAGGAAGACGAGCGCACCGAGCTGAACGAA
A V T Y R R V L A E T G H D Y E S L R Q V W Q E K K R E G L E A I I N E I A E L K E D E R T E L N E
exu_ppap GGCGTGACGTACAGACCGCTCCTGGCTGAGCAGGGCTTCGACTGGTCAAGCTGCAGGATATCTGGATGGAGGGAAGCGCAGCGGCTCAAGGGGGCCGTCGATGGGCTGGAAGAGCTGAAGGACGAGGACAGGTTGGAGCTGACTGAG
G V T Y R R V L A E Q G F D M V K L Q D I W M E G K R D G L K G A V D G G L E E L K D E E L K V E L T E
exu_crip GCATTAGACTTTTCGACTGCAATTTGGCAGATCAAAAACACTGCAAGCAATTTGGGATAGTGAAAAACGTAAGGATTGAAAATTTAGTAAAAGGCTTGAATCAATCAAGGAAGAGATCATGAAGAAATGATTAAT
A L D F R R A L A D Q K H D S K T L Q A I W D S E K R E G I E K L V K G L E S I K E E D H E E L I N
exu_bant GCTGTTACTTTCCGCCCTGCTTGGCAGTCCAAAGCATGACATGGAATCAATTCGAGAAGATGTTGGAAAGAGGGCAAGCGTGTGCAATGACCGAGCAGATCAAGAAGCTCGACGAACCTGAAGGAAGAGAAGACAGATCGAGCTGGTCGAC
A V T F R R L L A D A K H D M E S L Q K M W E E G K R D A M T E Q I K K L D E L K E E E Q I E L V D
1660 1670 1680 1690 1700 1710
exu_afra TTATTTGACAGTTACTTCGATCCTGTAAAGACGACTATTAAGCCGATTTGTCAAAAACAACAAT
L F D S Y F D P V K T T I K P I V K N N N
exu_ccap TTATTTGACAGTTACTTCGATCCCGAAAAACGACTATTAAGCCGATTTGTAAGATTAGCAAT
L F D S Y F D P E K T T I K P I V K I S N
exu_mdom TTATTTGACAGCTACTATGATCCAGCCAAAGCAGACTATTAACCCGTTGTTAAACCAATCAAT
L F D S Y Y D P A K T T I K P V V K P I N
```

```

exu_dmel CTCTGGACAGCTATTTGATCCCAAGAAGACCACGGTGAAGCCGGTTGTCAAGGGTAACAGC
L L D S Y F D P K K T T V K P V V K G N S
exu_dmoj CTGCTCGATAGCTTCTTTGATCCCAATAAGGTGACCATCAAACCGATCATTAAAGCCAAGCACC
L L D S F F D P N K V T I K P I I K P S T
exu_dvir CTGCTAGACAGTTTCTTTGATCCAGCCAAGTTACCATCAAGCCGATCATCAAACCCAGCACA
L L D S F F D P A K V T I K P I I K P S T
exu_dps1 CTCTGGACAGCTTCTTTGACCCAAGCAAGCGACCATCAAGCCGAGTTTCAAGCCCAATGGC
L L D S F F D P S K A T I K P S F K P N G
exu_smos ATTTTGGACAGTTACTTCGATCCGAGAAAAGAGGCGATTGTTCCGCCCAAAAACAATCCT
I L D S Y F D P E K E A I V P P T K T N P
exu_cape ATTCTCGATTGCCATTTGATCCCGAAGACAGTCACTGAAACCCAGTTGTTAAGCGAACAAAA
I L D C H F D P E K Q S L K P V V K R T K
exu_aaeg TTGCTCGATTGCCACTACGATCCGGAGAAGCAACAGCTGAAGCCGGTGGTGAAGCGGGTGCGC
L L D C H Y D P E K Q Q L K P V V K R V R
exu_agam CTCTTTGATTGTCAATTACGATCCGGACAAGCAGGCATTTAAAGCCGATCGTCAAGCGTATGAAG
L L D C H Y D P D K Q A F K P I V K R M K
exu_ppap CTCTCGATTGTCACTTTGACCCGGACAACAGCCATCAAGCCGCCGTGAGGAAGAATAAC
L L D C H F D P D K Q P I K P A V R K N N
exu_crip ATTTTGAACATCATTATGATCCAGAGCAGAAACATTTAAACCTGTGTTGACGGGAAAAC
I L E H H Y D P E Q K P F K P V V R R E N
exu_bant ATCTTGACTCGTACTTCGATCCCGAGAAGACCAGTGAACCCGATGTCCAACGCCGTGAA
I L D S Y F D P E K K P V K P I V Q R R E

```