

RESSALVA

Atendendo solicitação do autor, o texto completo desta **TESE** será disponibilizado somente a partir de 29/06/2023.

UNIVERSIDADE ESTADUAL PAULISTA “JULIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS FARMACÊUTICAS DE ARARAQUARA

VICENTE ESTEVAM MACHADO

ESTUDOS MOLECULARES E COMPORTAMENTAIS DO PROCESSO DE
OLFAÇÃO EM *Lutzomyia longipalpis* (DIPTERA:PSYCHODIDAE):

Araraquara

2021

UNIVERSIDADE ESTADUAL PAULISTA “JULIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS FARMACÊUTICAS DE ARARAQUARA

VICENTE ESTEVAM MACHADO

Estudos moleculares e comportamentais do processo de olfação em *Lutzomyia longipalpis* (Diptera:Psychodidae):

Tese de doutorado sob orientação da Profa. Dra. Mara Cristina Pinto, apresentada ao Programa de Pós-Graduação em Biociências e Biotecnologia aplicadas à Farmácia da Faculdade de Ciências Farmacêuticas - Universidade Estadual Paulista “Júlio de Mesquita Filho”, como requisito parcial para obtenção do grau de Doutor em Ciências.

Orientadora: Profa. Dra. Mara Cristina Pinto
Coorientador: Prof. Dr. Otavio Henrique Thiemann
Supervisor no exterior: Prof. Dr. Rickard Ignell

Araraquara

2021

M149e Machado, Vicente Estevam.
Estudos moleculares e comportamentais do processo de olfação em
Lutzomyia longipalpis (Diptera: Psychodidae) / Vicente Estevam
Machado. – Araraquara: [S.n.], 2021.
112 f. : il.

Tese (Doutorado) – Universidade Estadual Paulista. “Júlio de
Mesquita Filho”. Faculdade de Ciências Farmacêuticas. Programa de Pós
Graduação em Biociências e Biotecnologia Aplicadas à Farmácia.

Orientadora: Mara Cristina Pinto.
Coorientador: Otavio Henrique Thiemann.
Supervisor (exterior): Rickard Ignell.

1. Flebotomíneos. 2. Proteína ligadora de odorante. 3. Leishmaniose.
4. Atrativo. I. Pinto, Mara Cristina, orient. II. Thiemann, Otavio Henrique,
coorient. III. Ignell, Richard, Supervisor (exterior). IV. Título.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: Estudos moleculares e comportamentais de olfação em *Lutzomyia longipalpis* (Diptera:Psychodidae)

AUTOR: VICENTE ESTEVAM MACHADO

ORIENTADORA: MARA CRISTINA PINTO

Aprovado como parte das exigências para obtenção do Título de Doutor em CIÊNCIAS,
área: Análises Clínicas pela Comissão Examinadora:

Profa. Dra. MARA CRISTINA PINTO (Participação Virtual)
Departamento de Ciências Biológicas / Faculdade de Ciências Farmacêuticas do Câmpus
de Araraquara da UNESP

Profa. Dra. CAROLINA NASCIMENTO SPIEGEL (Participação Virtual)
Departamento de Biologia Celular e Molecular do Instituto de Biologia / Universidade
Federal Fluminense

Prof. Dr. ANDREY JOSÉ DE ANDRADE (Participação Virtual)
Departamento de Patologia Básica / Universidade Federal do Paraná

Prof. Dr. JAYME AUGUSTO DE SOUZA NETO (Participação Virtual)
Departamento de Bioprocessos e Biotecnologia / Faculdade de Ciências Agrônômicas de
Botucatu - UNESP

Araraquara, 29 de junho de 2021

*Dedico este trabalho aos meus Pais
Joaquim e Neuza, que nunca mediram
esforços para que esta conquista fosse
possível; à minha irmã Sarah e as minhas
sobrinhas Alice e Maria Júlia por todo o
apoio e incentivo desde o início.*

AGRADECIMENTOS

Aos meus pais, Joaquim e Neuza, por sempre estarem presentes e prontos para me ajudar e apoiar, além de me conduzirem para o melhor caminho. Obrigado por acreditar nas minhas escolhas, se sacrificar pelos meus sonhos e sempre me lembrar de que tudo vai dar certo. Essa vitória também é de vocês! Vocês são o maior e melhor exemplo que eu poderia ter!

A minha irmã Sarah, minhas sobrinhas Alice e Maria Júlia e meu cunhado Carlos, por todo o carinho e amizade compartilhados.

À professora e orientadora Mara, fundamental em todo meu crescimento acadêmico, profissional e pessoal. Obrigado pela oportunidade de poder trabalhar com você, e por todos os ensinamentos, dedicação e confiança. Agradeço também por ser a primeira incentivadora desta caminhada, por ter tanta paciência, e por fazer parte da minha vida. Obrigado por sempre acreditar no meu potencial.

Ao professor Otavio, pela orientação e apoio no desenvolvimento do projeto, e o suporte durante os experimentos de expressão das OBPs

Aos professores Rickard e Sharon por me aceitarem no período sanduíche, de suma importância para o meu crescimento pessoal e profissional.

Ao professor Jairo Magalhães Torres-Junior, pelo apoio na realização das coletas de campo e todo apoio na fase final do doutorado.

À Tamires, por todo apoio nos momentos finais da escrita da tese, e por pelos vários momentos de “vai dar tudo certo, fica tranquilo”!

Aos meus amigos do laboratório dos flebs e companheiras de projeto Flavia, Cintia e Thais por todo conhecimento, companheirismo, respeito, e amizades compartilhadas.

Aos meus amigos “suecos” Luis, Bruno e Fay, por fazer meu tempo na Suécia mais tranquilo, e me ajudarem a sentir menos saudades do Brasil, vocês são demais!

Aos meus colegas do laboratório Dennys, Wanderson, Jader, Juliana e Lucas por todo conhecimento, companheirismo, respeito e boa convivência.

Aos meus colegas do laboratório de cristalografia do Instituto de física da USP de São carlos.

Aos meus colegas do *Unit of chemical ecology* da *SLU*, por me acolherem em um país diferente e todos os ensinamentos durante a minha estada na Suécia.

Aos alunos do professor Jairo, Thauane e Thailan pelo auxílio na realização das coletas de campo.

Aos amigos de Passos que fazem tanta falta no meu dia-a-dia: Átila, Chagas, Kennia, Rock, Marolo, Maiko, Paula, Fabão e Rosy.

Aos amigos e padrinhos Levy e Tereza, obrigado pela confiança!

A todos os professores e funcionários do departamento, em especial o Seu Paulo por ser sempre tão simpático nas idas ao departamento.

O presente trabalho foi realizado com apoio do Conselho Nacional de desenvolvimento Científico e Tecnológico (CNPq), Código de Financiamento: 140298/2018-5

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) - Códigos de Financiamento 001 e CAPES/PrInt 88887.373456/2019-

A todas as pessoas que de alguma forma contribuíram para a realização deste trabalho.

Muito obrigado!

*“O cientista não é o homem que fornece
as verdadeiras respostas;
é quem faz as verdadeiras perguntas”.
(Claude Lévi-Strauss)*

LISTA DE FIGURAS

Figura 1. Representação de uma sensila em antena de insetos e os processos envolvidos na recepção de compostos voláteis. A etapa 1 mostra os odorantes entrando na linfa sensilar, a etapa 2 a ligação do odorante com a OBP, a etapa 3 o complexo odorante-OBP é transportado pela linfa sensilar, etapa 4, o complexo OBP odorante se liga ao OR, ou libera o odorante para se ligar diretamente ao OR, a etapa 5, o odorante é liberado e na etapa 6 é capturado pelas ODEs, sendo degradado na etapa 7. Fonte: LEAL, 2013		Erro! Indicador não definido.
Figura 2. <i>Vial</i> contendo a mistura para a liberação		28
Figura 3. Túnel de vento utilizado nos testes de atratividade dos sistemas de liberação.		29
Figura 4. Adaptação para a utilização da garrafa lavadora de gás.		30
Figura 5. Vista aérea dos locais das armadilhas		31
Figura 6. <i>Vial</i> contendo a mistura de compostos colocado em posição na armadilha luminosa CDC.		32
Figura 7. Delineamento do quadrado latino (4x4)		32
Figura 8. Confirmação de que a OBP 99a pertence à família das OBPs		34
Figura 9. Alinhamento e similaridade da sequência de aminoácidos da OBP 99a.		35
Figura 10. Alinhamento e similaridade da sequência de aminoácidos da OBP 3-1.		35
Figura 11. Alinhamento e similaridade da sequência de aminoácidos da OBP 10.		35
Figura 12. Alinhamento e similaridade da sequência de aminoácidos da OBP 14.		36
Figura 13. Predição da estrutura secundária e acessibilidade ao solvente da OBP 99a.		38
Figura 14. Predição da estrutura terciária da OBP 99a (C-score: -0,58). As setas vermelhas indicam as ligações dissulfeto.		38
Figura 15. Predição da estrutura secundária e acessibilidade ao solvente da OBP 3-1.		39

Figura 16. Predição da estrutura terciária da OBP 3-1 (C-score: -0,88). As setas vermelhas indicam as ligações dissulfeto.	40
Figura 17. Predição da estrutura secundária e acessibilidade ao solvente da OBP 10.	41
Figura 18. Predição da estrutura terciária da OBP 10 (C-score: -0,22). As setas vermelhas indicam as ligações dissulfeto.	41
Figura 19. Predição da estrutura secundária e acessibilidade ao solvente da OBP 14.	42
Figura 20. Predição da estrutura terciária da OBP 14 (C-score: -0,86). As setas vermelhas indicam as ligações dissulfeto.	43
Figura 21. Em destaque as bandas de expressão em 4 e 24h das OBPs 10 e 14, respectivamente	47
Figura 22. Gel SDS – Page do teste de solubilidade da OBP 10, as bandas de expressão mostradas na fração insolúvel.	48
Figura 23 Comportamento de ativação e atração das fêmeas de <i>Lu. longipalpis</i> em túnel de vento para as diluições da mistura de compostos ($p < 0.0001$).	50
Figura 24 Comportamento de ativação e atração dos machos de <i>Lu. longipalpis</i> em túnel de vento para as diluições da mistura de compostos ($p < 0.0001$).	51
Figura 25. Parte do relatório de mapeamento dos <i>reads</i> no software CLC.	62

LISTA DE TABELAS

Tabela 1. Compostos utilizados para a produção da mistura de compostos	27
Tabela 2. Dados de predição de parâmetros químicos e físicos das OBPs	36
Tabela 3 - Resultado das diferentes construções, linhagens, temperatura e de meios de cultura de concentração de IPTG.	45
Tabela 4. Flebotomíneos coletados separados por tratamento e localização	53
Tabela 5. Gênero e espécie dos flebotomíneos coletados separados por tratamento	53
Tabela 6. Regiões com transmissão de leishmaniose visceral, sem a presença de <i>Lu. longipalpis</i>	55
Tabela 7. Detalhamento das 48 amostras de RNA total enviadas à Eurofins em 2020.	63

Resumo

Flebotomíneos são insetos transmissores de vírus, bactérias e protozoários, sendo a *Leishmania* spp. o de maior importância. Esses insetos utilizam sinais químicos na busca de alimentos, cópula e em outras fases da vida, e a ecologia química estuda esses sinais. Atualmente, têm se ampliado o número de estudos com ecologia química de vetores, ou seja, estudos que visam encontrar compostos voláteis que sejam atrativos ou repelentes para esses insetos, porém de forma geral a busca por esses compostos é realizada de maneira aleatória. A percepção dos compostos voláteis ocorre nas antenas e palpos dos insetos, onde proteínas ligam-se a compostos odorantes e os direcionam para os neurônios sensoriais olfativos, desencadeando respostas comportamentais. Por essa razão, o entendimento do processo de olfação mostra-se interessante para os estudos de ecologia química reversa, pois conhecendo os alvos desses atrativos/repelentes, é possível prever quais classes de compostos são favorecidos para essa ligação. Outra abordagem, mais clássica, é a identificação dos caimônios, compostos liberados por hospedeiros que desencadeiam respostas comportamentais em insetos. Para o grupo dos flebotomíneos, os estudos de ecologia química clássica são ainda incipientes, e a abordagem utilizando ecologia química reversa é praticamente nula. Com essa perspectiva, o presente estudo teve como objetivos: i) expressar proteínas de odor da espécie *Lutzomyia longipalpis*; ii) avaliar a atratividade de uma mistura de compostos desenvolvida através do odor humano em laboratório, por meio de túnel de vento para *Lu. longipalpis*, e em armadilhas colocadas em área com a presença dessa espécie. Com relação às proteínas de odor, foi possível a expressão de duas proteínas ligadoras de odorantes (10 e 14), que futuramente serão caracterizadas. A mistura de compostos apresentou atratividade tanto para fêmeas como para os machos de *Lu. longipalpis* em túnel de vento, apresentando uma resposta dose dependente. Os dados de campo indicaram uma tendência de efetividade da mistura na atração de *Lu. longipalpis*, principalmente a dosagem de 1:1000, e 1:10. Os resultados dos testes de atratividade em campo foram de uma região com baixa densidade de insetos, sendo necessário repetições futuras, e também em áreas com uma maior densidade de insetos.

Palavras chaves: Flebotomíneos, proteína ligadora de odorante, leishmaniose, atrativo.

Abstract

Sand flies are insects involved in the transmission of virus, bacteria and protozoa, which the *Leishmania* spp. is the most important, these insects use chemical cues to look for host, mating and other aspects of their life cycle, the chemical ecology research these chemical cues. Nowadays, the number of vector chemical ecology studies has been increasing, studies that the aim is to find new volatile compounds that can be attractive or repellent for these insects, mainly this has been done in a screening way. However, these insects perceive volatile compounds in proteins in the maxillary palps and antennae, where the compounds bind with different proteins and elicit the odorant sensory neurons, and cause behavior responses. Thus, the understanding of the olfaction process is relevant to chemical ecology studies, because when we know the targets of these compounds it is possible to predict the chemical classes that will bind in these proteins. Another approach is the evaluation of the volatile compounds released by hosts, using gas chromatography and single sensillum record to find volatile compounds that elicit responses in the odorant sensory neurons and with that develop blends that can be attractive for hematophagous insects. For sand flies, chemical ecology studies are incipient, thereby, the aims of this study are: i) express odorant proteins of *Lutzomyia longipalpis*; ii) evaluate the attractiveness of a blend that was prepared from human odour in laboratory and field. We were able to express the odorant binding protein 10 and 14, which will be characterized in the future. The blend was attractive to males and females of *Lu. longipalpis* in laboratory, and the field data showed that the blend has a tendency of effectiveness in the collection of *Lu. longipalpis*. These results were in a low density region of insects, and it is necessary future replicates and also trials in a high density area of insects.

Key words: Sand flies, odorant binding protein, leishmaniasis, attractant.

Sumário

1.	Introdução	11
1.1.	Flebotomíneos.....	11
1.2.	Processo olfativo em vetores.....	12
1.2.1.	Proteínas ligadoras de odorantes (OBPs).....	15
1.2.1.1.	Aplicação das OBPs	16
1.2.2.	Receptores de odorantes	16
1.2.3.	Outras proteínas relacionadas à olfação.....	18
1.3.	Aplicações de cairomônios em insetos hematófagos	19
1.4	Ecologia química e flebotomíneos.....	20
2	Objetivos	24
2.1	Objetivos específicos	24
3	Material e Métodos.....	25
3.1	Expressão de OBPs	25
3.1.1	Sequências das OBPs de <i>Lutzomyia longipalpis</i>	25
3.1.2	Expressão das proteínas de odor em <i>Escherichia coli</i>	25
3.2	Testes de atratividade em laboratório	26
3.2.1	Origem dos insetos utilizados nos testes de atratividade	26
3.2.2	Produção da mistura para os testes em laboratório e campo	27
3.2.3	Avaliação da mistura de compostos no túnel de vento	28
3.3	Testes em campo.....	30
3.3.1	Local de Coleta.....	30
3.3.2	Avaliação da atratividade em campo.....	31
3.3.3	Identificação das espécies de flebotomíneos.....	33
4	Resultados e discussão.....	34
4.1	Expressão das OBPs	34
4.1.1	Análise das sequências das OBPs de <i>Lu. longipalpis</i>	34
4.1.2	Expressão de proteínas ligadoras de odorante recombinantes de <i>Lu. longipalpis</i>	43
4.2	Testes de atratividade em laboratório	49
4.2.1	Manutenção da colônia de <i>Lu. longipalpis</i>	49
4.2.2	Avaliação da mistura de compostos no túnel de vento	49
4.3	Testes de atratividade em campo.....	53

4.3.1 Avaliação da atratividade em campo e identificação das espécies de flebotomíneos coletadas	53
5 Conclusões e perspectivas.....	58
Anexos	59
Projeto A:.....	59
1.Etapas desenvolvidas na SLU:.....	59
1.1.Manutenção da colônia <i>Anopheles gambiae</i> e coleta dos tecidos	59
1.2 Extração do RNA e sequenciamento	60
1.3 Mapeamento dos genes	60
3 Etapas em desenvolvimento e futuras.....	66
3.1 Análise do RNA-Seq.....	66
Projeto B.....	66
1. Etapa realizada.....	66
1.1. Verificação da anotação gênica dos receptores de odorantes em <i>Aedes aegypti</i>	66
1.2 Resultados das análises de bioinformática	68
Referências	98

1. Introdução

1.1. Flebotomíneos

Flebotomíneos são insetos pequenos, com cerca de 3 mm, que apresentam o corpo recoberto por cerdas e asas em formato de “V”, quando em repouso (BRAZIL; RODRIGUES; ANDRADE FILHO, 2015; KILLICK-KENDRICK et al., 1986; MAROLI et al., 2013). São insetos hematófagos, holometabólicos, tendo no seu desenvolvimento os estágios de ovo, quatro estádios larvais, pupa e o inseto adulto (SOARES; TURCO, 2003). Os adultos de ambos os sexos necessitam de uma alimentação rica em carboidratos, a qual é obtida por meio de tecidos de plantas ou no “honeydew”, substância açucarada produzida por pulgões (KILLICK-KENDRICK, 1999). Somente as fêmeas necessitam da alimentação sanguínea para a maturação dos ovos (SOARES; TURCO, 2003).

Estes insetos estão envolvidos na transmissão de vários agentes etiológicos, desde vírus, bactérias até protozoários, sendo o mais importante os protozoários do gênero *Leishmania* spp. (Kinetoplastida, Trypanosomatidae) agente etiológico das leishmanioses (WHO, 2017). Até o momento, são mais de 1000 espécies de flebotomíneos descritas no mundo, porém, não mais do que 70 são transmissoras das leishmanioses (BRAZIL; RODRIGUES; ANDRADE FILHO, 2015).

As leishmanioses são divididas em cutânea (LC) e visceral (LV). Estima-se que existam de 200 a 400 mil novos casos/ano da forma visceral e 1,2 milhão da forma tegumentar no mundo (WHO, 2017). Essas enfermidades afetam principalmente a África, Ásia e América Latina, com casos em 98 países. O Brasil é o país que apresenta o maior número de casos de LV na América Latina, são cerca de 3000 casos/ano, na qual apresentou uma taxa de letalidade de 7,5%, no ano de 2019 (BRASIL, 2019).

A LV é a forma mais grave da doença, pois quando não tratada adequadamente é letal (DESJEUX, 2004; DUJARDIN, 2006). É causada pelo protozoário *Leishmania infantum* Nicolle, 1908, que é transmitido para humanos principalmente pelas fêmeas da espécie *Lutzomyia longipalpis* nas Américas (LAINSON; RANGEL, 2005).

A relação entre hospedeiro e vetor é um ponto importante na epidemiologia da LV, pois o flebotomíneo necessita encontrar seu hospedeiro

para que possa realizar os repastos sanguíneos resultando na transmissão do agente etiológico das leishmanioses.

Nessa busca do hospedeiro para a alimentação sanguínea, flebotomíneos e outros insetos hematófagos, utilizam sinais olfativos, que, além disso, são também fundamentais em outras várias etapas do ciclo de vida, (LOGAN; BIRKETT, 2007),

Esses sinais olfativos são divididos em, cairomônios, alomônios e sinomônio. Os feromônios são compostos de comunicação intraespecífica, enquanto os cairomônios e alomônios de comunicação interespecífica ((WHITTAKER; FEENY, 1971; ZARBIN; RODRIGUES; LIMA, 2009)

Os cairomônios são compostos liberados por um organismo e beneficiam outro de uma espécie diferente; os alomônios são compostos liberados que beneficiam o seu emissor; e os sinomônios beneficiam ambos, tanto o emissor, quanto o receptor (WHITTAKER; FEENY, 1971; ZARBIN; RODRIGUES; LIMA, 2009).

O entendimento do processo de olfação é relevante tanto para o conhecimento básico da área, como para possíveis aplicações práticas na melhoria de técnicas de monitoramento e controle dos insetos.

1.2. Processo olfativo em vetores

O processo de olfação é um ponto crítico na sobrevivência dos insetos, pois por meio de sinais químicos, detectados pelos órgãos olfativos, os insetos encontram alimento, cópula e locais para oviposição. A ecologia química é a área de conhecimento que estuda a olfação e os sinais químicos envolvidos nesse processo (BROWN; EISNER; WHITTAKER, 1970).

Os primeiros insetos a apresentarem receptores para odorantes foram os apterigotos, cerca de 400 milhões de anos atrás (MISSBACH et al., 2014). A evolução da olfação em insetos se deu com a especialização de regiões das antenas e dos palpos maxilares, chamadas de sensilas, que são estruturas capazes detectar diferentes tipos de estímulos, como olfativos, gustativos, proprioceptivos, mecânicos, higró- e termossensoriais (CARVALHO et al., 2017; HORE et al., 2017; SCHNEIDER, 1964).

Nas antenas dos flebotomíneos são encontradas seis tipos de sensilas: tricoides, celocônicas, campaniformes, caéticas, escamiformes e basicônicas. Há dimorfismo sexual nas antenas de machos e fêmeas, tanto no número médio de sensilas como também em sua distribuição, sendo que o número de sensilas é significativamente maior nos machos. As sensilas envolvidas na olfação dos flebotomíneos são: caéticas, as celocônicas de parede dupla, e as alguns tipos de tricoides (FERNANDES et al., 2008).

No interior dessas sensilas olfativas é encontrada a linfa-sensilar, onde estão as terminações dendríticas dos Neurônios Sensoriais Olfativos (OSNs – *odorant sensory neurons*), cuja função é a transdução do estímulo ambiental em impulso elétrico neural. As sensilas especializadas na olfação (basicônicas, tricoides e celocônicas) apresentam poros nos quais os compostos voláteis penetram e conseguem acesso aos OSNs (XU et al., 2005).

A ligação entre os compostos voláteis que penetram através dos poros das sensilas e os neurônios, começou a ser melhor entendida após a descoberta de pequenas proteínas globulares solúveis, secretadas em grande quantidade na linfa sensilar (VOGT; RIDDIFORD, 1981), indicando assim uma possível participação dessas proteínas no transporte dos compostos voláteis.

Essas proteínas são chamadas de “proteínas ligadoras de odorantes”(OBPs - *Odorant Binding Proteins*) e são responsáveis pelo transporte dos compostos voláteis de baixa solubilidade até a membrana dos OSNs, onde são espessos as três principais famílias de receptores de odorantes. Tais receptores são classificados em: receptores de odorantes (ORs - *Odorant receptor*), receptores ionotrópicos (IRs - *Ionotropic receptor*) e as proteínas sensoriais do sistema neuronal (SNMPs - *sensory neuron membrane proteins*) (PELOSI, 1994; VOGT, 2003). Essa interação do composto volátil e dos ORs promove a ativação da cascata de sinalização, resultando na despolarização da membrana do OSN. Esse fenômeno resulta na mudança do potencial de ação do OSN, fazendo com que o sinal elétrico seja transmitido ao sistema nervoso central, podendo gerar respostas fisiológicas e/ou comportamentais. Após a emissão do sinal, os compostos voláteis são degradados pelas enzimas degradadoras de odorantes (ODE – *odorant-degrading enzymes*) (Fig 1) (LEAL, 2013; VOGT, 2003). Atualmente, há a descrição de outras proteínas provavelmente envolvidas no

processo de ofação, como os receptores gustativos (GR- *gustative receptors*) e as proteínas *Pickpocket*, entre outras (TALLON; HILL; IGNELL, 2019).

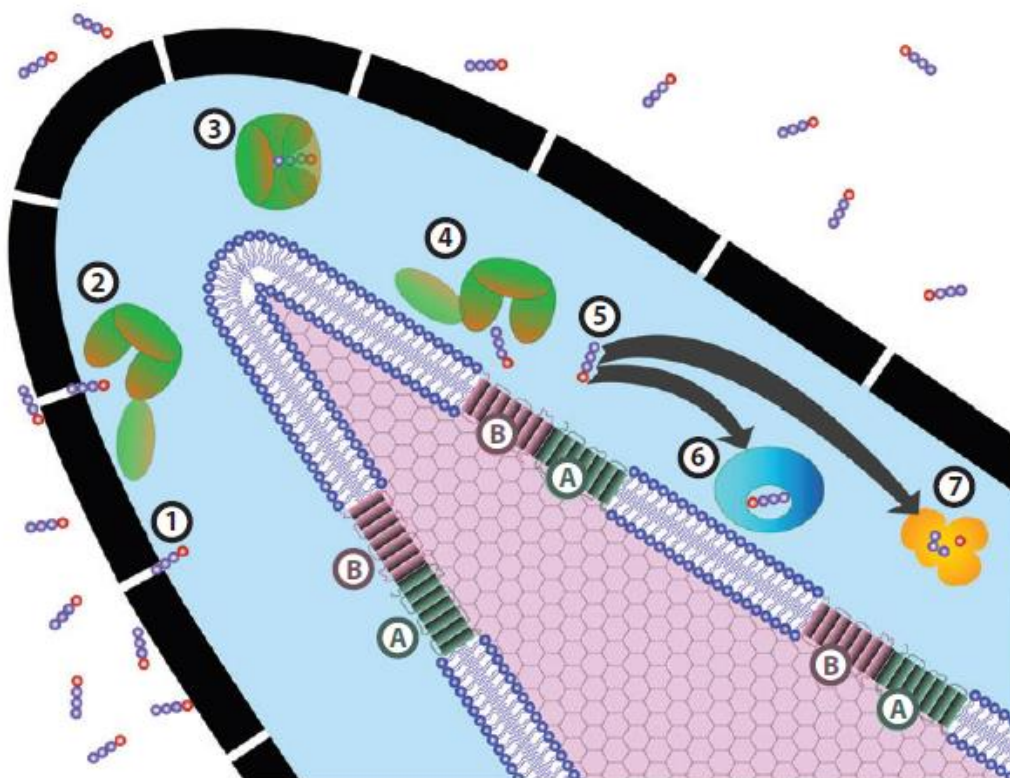


Figura 1. Representação de uma sensila em antena de insetos e os processos envolvidos na recepção de compostos voláteis. A etapa 1 mostra os odorantes entrando na linfa sensilar, a etapa 2 a ligação do odorante com a OBP, a etapa 3 o complexo odorante-OBP é transportado pela linfa sensilar, etapa 4, o complexo OBPs odorante se liga ao OR, ou libera o odorante para se ligar diretamente ao OR, a etapa 5, o odorante é liberado e na etapa 6 é capturado pelas ODEs, sendo degradado na etapa 7. Fonte: (LEAL, 2013)

A produção das proteínas relacionadas à olfação é controlada pela tradução de mRNAs, os quais podem ser expressos em diferentes padrões, de acordo com o período do ciclo de vida no qual o inseto se encontra (VOGT, 2002). Em insetos hematófagos, esse controle já foi demonstrado sendo que, há diferenças na expressão gênica de algumas proteínas ligadas à olfação, como OBPs, ORs, e outras, em *Anopheles gambiae* (RINKER et al., 2013), *Culex*

quinquefasciatus (TAPARIA; IGNELL; ROSE HILL, 2017) e *Aedes aegypti* (MATTHEWS et al., 2016)

1.2.1. Proteínas ligadoras de odorantes (OBPs)

As primeiras OBPs descritas foram em lepidópteros em meados da década de 80. Com o avanço dos estudos, há várias espécies de insetos em diferentes ordens que já possuem as OBPs descritas (VOGT; RIDDIFORD, 1981), como em culicídeos para as espécies *Anopheles gambiae* (HOLT et al., 2002), *Aedes aegypti* (COSTA-DA-SILVA et al., 2013) e *Aedes albopictus* (LI et al., 2012). Em moscas, como *Drosophila melanogaster*, foram identificados no genoma cerca de 50 genes que codificam para OBPs (GALINDO; SMITH, 2001; HEKMAT-SCAFE et al., 2002); há também a descrição das OBPs em moscas de fruta do gênero *Anastrepha* (CAMPANINI; DE BRITO, 2016).

Embora estudos indiquem a importância das OBPs na solubilização e transporte dos compostos voláteis (XU et al., 2005), estas proteínas também podem ser encontradas em outros tecidos não olfativos e são denominadas de “encapsulinas” de acordo com seu modo de ação (LEAL, 2013).

As OBPs são produzidas e secretadas na linfa-sensilar por células acessórias neuronais. Por serem proteínas secretadas, apresentam na porção N-terminal uma região de peptídeo sinal, que está presente para seu endereçamento celular, fazendo com que ela seja então secretada no seu local de ação, a linfa sensilar (VOGT et al., 1999).

Após a ligação do ligante com o alvo, pode haver mudanças na conformação da OBP, que geralmente é composta por seis alfa-hélices adotando uma conformação estável (TEGONI; CAMPANACCI; CAMBILLAU, 2004), mantida por ligações dissulfeto entre os resíduos de cisteína, que é a principal característica do grupo (LEAL, 2013). Outro fator que pode alterar sua conformação é o pH da região, que determina a ligação ou a liberação do composto volátil (LEAL; CHEN; ERICKSON, 2005).

Além disso, há uma baixa similaridade nas sequências intraespecíficas das OBPs, como mostrado em um estudo de populações simpátricas de *Lutzomyia longipalpis* (DIAS et al., 2013), o que dificulta a identificação de novas proteínas desse grupo (XU et al., 2009). Sendo então, utilizado para a identificação e classificação das OBPs o seu bem conservado espaço hidrofóbico,

que consiste em ligações dissulfeto entre suas cisteínas (FAN et al., 2011; LEAL, 2013).

1.2.1.1. Aplicação das OBPs

Além da utilização para possíveis alvos de busca de compostos atrativos/repelentes para insetos, as OBPs têm mostrado grande potencial biotecnológico, com aplicações em diversas áreas, principalmente por terem como característica se ligar a compostos em baixas concentrações, além de geralmente serem facilmente solubilizadas e termorresistentes (PELOSI et al., 2014, 2018).

Um exemplo de aplicação dessa abordagem foi a construção de um biossensor utilizando a OBP 14 de *Apis mellifera* que pode detectar um grande número de compostos atrativos para abelhas. Assim, em uma possível combinação de várias OBPs pode-se desenvolver sensores extremamente sensíveis e específicos (LARISKA et al., 2015).

Além disso, a OBP 1 de *Culex quinquefasciatus* e a PBP 1 de *Bombix mori* foram utilizadas para o desenvolvimento de uma coluna cromatográfica, pois essas proteínas são facilmente expressas de forma recombinante, são estáveis, além de apresentarem alterações na sua conformação de acordo com a alteração do pH, o que pode ser usado para a eluição dos ligantes (MARGARYAN et al., 2006; PELOSI et al., 2006).

1.2.2. Receptores de odorantes

Os ORs têm grande importância no processo de olfação e são expressos na membrana dendrítica dos OSNs, são proteínas transmembranas com a região do N-terminal posicionadas em direção ao citoplasma e possuem cerca de 400 aminoácidos (LUNDIN et al., 2007).

São majoritariamente expressos nos principais órgãos olfativos, as antenas e os palpos maxilares, sendo que a detecção dos compostos voláteis tem sido demonstrada como sendo mediada pela associação desses receptores com co-receptores (ORco) altamente conservados (HUSSAIN et al., 2016; VOSSHALL; WONG; AXEL, 2000).

Em *Lu. longipalpis* foram identificados 140 ORs presentes em seu genoma na linhagem de Jacobina-BA, sendo a família gênica mais prevalente nos tecidos olfativos, entre genes relacionados a olfação (HICKNER et al., 2020), o que mostra seu papel de extrema importância nos comportamentos mediados por odorantes.

Há dois tipos de ORs descritos atualmente, o metabotrópico e o ionotrópico. O metabotrópico depende da ativação da proteína G para a produção de AMP cíclico, que desencadeia a abertura dos canais e a emissão do sinal elétrico para a despolarização do neurônio (GETAHUN et al., 2013; WICHER et al., 2008). O ionotrópico possui ação direta, ou seja, não depende da ativação de outra proteína, porém é um complexo heteromérico, que possui um correceptor acoplado (ORco) (KRIEGER et al., 2003; SATO et al., 2008).

Os ORs são proteínas de sete domínios pós-membrana. Esses domínios formam receptores heteroméricos, constituídos por um ORco altamente conservado e uma variável que apresenta evolução rápida, o que determina a seletividade e a sensibilidade do complexo do receptor e por consequência do OSN (CAREY et al., 2010; HALLEM; CARLSON, 2006).

Os IRs são receptores parecidos com os ORs do tipo ionotrópico, porém, não necessitam dos correceptores e são homólogos a outros receptores, como os de glutamato (KAUPP, 2010)

Além disso, os IRs se diferenciam dos ORs por ser uma família de canais iônicos dependentes de ligantes altamente conservados, presentes em animais, plantas e bactérias. Além disso, em *Drosophila melanogaster* os IRs estão presente em neurônios quimiossensoriais que não expressam os ORs (BENTON et al., 2009).

Os IRs são filogeneticamente não relacionados aos outros receptores, sendo majoritariamente expressos em OSNs que não expressam ORs (BENTON et al., 2009), sendo o IR76b e o OR35a encontrados no mesmo OSN em *Drosophila* (COUTO; ALENIUS; DICKSON, 2005).

Foram identificados 23 IRs em *Lu. longipalpis*, e esses receptores se ligam em geral a aminas e outros voláteis que não são detectados pelos ORs (HUSSAIN et al., 2016), porém a caracterização funcional desses receptores é necessária para elucidar o seu papel na detecção de compostos voláteis relevantes biologicamente.

E por fim, há as SNMPs que são pouco estudadas, porém sabe-se que estão associadas à percepção de feromônios; são proteínas da família CD36, consiste em um *loop* extracelular contendo 9 cisteínas ancoradas por dois domínios transmembrana (JIN; HA; SMITH, 2008). Na maioria das espécies de culicídeos são encontradas duas proteínas dessa família (TAPARIA; IGNELL; ROSE HILL, 2017).

1.2.3. Outras proteínas relacionadas à olfação

Além dos receptores citados anteriormente, há algumas classes de proteínas, nas quais seus papéis ainda não estão claros.

Os receptores gustativos são os responsáveis pela percepção do gás carbônico e estão presentes nos palpos maxilares dos mosquitos (TALLON; HILL; IGNELL, 2019). Além disso, são importantes para a percepção de uma gama de ligantes, doces, amargos entre outros, porém o mecanismo para a recepção e a discriminação do sabor ainda não está elucidado (ISONO; MORITA, 2010).

Os receptores transitórios de potencial (TRP – *transient receptor potential*) que foram descritos em *Drosophila melanogaster*, são receptores da superfície das células e participam da detecção de alterações no estímulo ambiental, assim influenciando no comportamento (FOWLER; MONTELL, 2013). Pertencem a uma classe de proteínas que não é exclusiva do sistema olfativo (TALLON; HILL; IGNELL, 2019).

As proteínas *Pickpockets* (PPK) são canais de denegrina/sódio epitelial (DEG/ENaC - *Degenerin/epithelial sodium channels*) e representam uma grande família de proteínas de membrana, porém suas funções fisiológicas não são bem estabelecidas (ZELLE et al., 2013), e também não são exclusivos do sistema olfativo (TALLON; HILL; IGNELL, 2019).

Observa-se que ainda há muito a se estudar com relação ao processo de olfação em insetos. Além do conhecimento básico sobre as proteínas envolvidas na olfação, a relevância de maiores estudos nessa área poderá refletir na busca por compostos que se liguem às mesmas, para serem utilizados como iscas atrativas ou possíveis repelentes, para monitoramento e controle de insetos vetores.

Com isso, as proteínas relacionadas ao processo de olfação apresentam-se como um alvo estratégico para a seleção de compostos, e vêm sendo utilizadas como ferramentas na chamada Ecologia Química Reversa (FRANCO et al., 2018; KAMALA JAYANTHI et al., 2014; LEAL et al., 2008). De acordo com essa abordagem, parte-se da investigação molecular na tentativa de identificar possíveis compostos ligantes que desencadeiem comportamentos específicos nos insetos, como, por exemplo, atração.

1.3. Aplicações de cairomônios em insetos dípteros hematófagos

O cairomônio mais estudado é o dióxido de carbono (CO₂), sendo que praticamente todos os insetos hematófagos respondem a este importante componente liberado na respiração dos hospedeiros vertebrados (DEKKER; CARDÉ, 2011; GILLIES, 1980; HEALY; COPLAND, 1995).

O CO₂ também age como um composto sinergista, pois, alguns voláteis ocasionam uma alteração do comportamento somente quando estão em sinergia com o CO₂, como é o caso do ácido láctico (EIRAS; JEPSON, 1991).

Porém, o CO₂ é responsável somente por uma parte da resposta atrativa dos insetos, o que sugere o envolvimento de outros voláteis na busca pelo hospedeiro desses insetos (MBOERA et al., 1997; PINTO et al., 2001).

Vários estudos com compostos voláteis foram realizados, de diferentes classes químicas, como ácidos carboxílicos, álcoois, ésteres, aldeídos, entre outros para diversos insetos hematófagos (BERNIER et al., 2000; CORK; PARK, 1996; MACHADO et al., 2015; MAGALHÃES-JUNIOR et al., 2014; MUKABANA et al., 2012).

Sendo assim, com o avanço no conhecimento dos aspectos moleculares de ecologia química em insetos, iscas atrativas mais efetivas puderam ser desenvolvidas. Como por exemplo, foi desenvolvida uma mistura de compostos voláteis identificados em humanos. Para isso, utilizou-se o conhecimento em 49 receptores de odorantes de *Anopheles coluzzi*, através de análises de gravação de uma única sensila – *single sensillum record* (SSR) (OMONDI et al., 2019).

Para a identificação dos voláteis presentes no corpo humano, 20 voluntários foram colocados em sacos plásticos selados por calor, somente com a cabeça para o lado externo. Um fluxo de ar sintético foi extraído por bombas

contendo o polímero adsorvente PorapakQ por 2.5h, para adsorção dos voláteis presentes. A dessorção dos voláteis foi realizada utilizando pentano, e os extratos foram colocados juntos, e concentrados através de jato de nitrogênio (OMONDI et al., 2019). A avaliação dos compostos voláteis foi realizada através da combinação das técnicas de cromatografia gasosa acoplada a espectrômetro de massas e de cromatografia gasosa acoplada ao SSR. Foram identificados 15 compostos que geraram uma resposta consistente dos receptores de odorantes. Os compostos foram: 1-hexanol, 1-octen-3-ol, 3-octanol, 2-nonanol, octanal, nonanal, decanal, acetato de butila, sulcatona, fenol, benzaldeído, acetofenona, limoneno, linalool, α -terpineol, e a taxa de liberação desses compostos também foi avaliada, para a formulação da mistura (OMONDI et al., 2019).

Um ponto importante é a taxa de liberação desses voláteis pelo hospedeiro, pois o mesmo composto em diferentes taxas de liberação pode desencadear comportamentos diferentes em insetos hematófagos (GILLIES, 1980; SMITH et al., 1970), e já foi mostrado que alguns compostos quando apresentados aos insetos fora do contexto da mistura de compostos não causam nenhuma resposta do inseto (BRAKS; MEIJERINK; TAKKEN, 2001; MAJEED et al., 2016). Como na mistura de compostos citada acima, além dos compostos que geraram resposta dos receptores de odorantes, também foi levada em consideração a taxa de liberação dos voláteis, isso foi possível, pois todos os voluntários ficaram o mesmo tempo dentro do saco plástico, com o mesmo tipo de polímero, e os voláteis foram eluídos na mesma quantidade de solvente. Após a eluição um controle interno foi adicionado, sendo assim feito o cálculo da taxa de liberação no hospedeiro, e o desenvolvimento da formulação atrativa para mosquitos (OMONDI et al., 2019).

Sendo assim, o comportamento de busca pelo hospedeiro não depende somente dos compostos envolvidos, mas sim de uma série de fatores, como a concentração dos compostos na mistura, e de sua taxa de liberação. Todos esses fatores agem de forma com que se tenham compostos ativadores, que tem, por exemplo, a capacidade de gerar uma resposta de início do voo no inseto, e outros que aumentam a resposta atrativa dos insetos aos voláteis (OMONDI et al., 2019; QIU et al., 2011).

1.4 Ecologia química e flebotomíneos

As armadilhas luminosas ainda são as mais comumente utilizadas para captura e monitoramento de espécies de flebotomíneos. Porém, apresentam algumas desvantagens como: dificuldades logísticas pela necessidade de uma fonte de energia no campo; são pouco seletivas e uma grande quantidade de outros insetos acaba por danificar os flebotomíneos (ALEXANDER, 2000).

Os avanços na área de ecologia química, na busca de iscas atrativas, para flebotomíneos são ainda tímidos em relação aos culicídeos.

Os estudos da olfação em nível molecular são restritos. Há um estudo com a confecção de uma biblioteca de cDNA oriunda de antenas de machos e fêmeas de *Lu. longipalpis* que identificou 14 sequências de genes relacionadas a possíveis OBPs (DIAS, 2008). A partir desse estudo foi realizada a modelagem de algumas destas sequências (SANTANA, 2016).

O maior empenho tem sido dado à investigação de feromônios sexuais na espécie *Lutzomyia longipalpis* (BRAY et al., 2014; SPIEGEL et al., 2011, 2016). Sendo que o *Lu. longipalpis* é um complexo de espécies, com diferentes feromônios sexuais e diferentes sons de cópula, possuindo pelo menos 5 feromônios sexuais diferentes, 5-metilgermacreno, 3-himachaleno, sobraleno, cembreno-1 e 9-metilgermacreno + (HICKNER et al., 2020; PALFRAMAN et al., 2018).

Para os caïromônios, o dióxido de carbono (CO₂) é o mais estudado. Alguns estudos mostram, por exemplo, que a taxa de atração de *Lu. longipalpis* tende a aumentar de acordo com o tamanho do hospedeiro (QUINNELL; DYE; SHAW, 1992). Para testes em campo com as espécies *Nyssomyia intermedia* e *Ny. whitmani* utilizando o CO₂, foi quantificada a porcentagem de captura para um cilindro de CO₂ em comparação à uma pessoa, sendo que cerca de 45% da atratividade de um ser humano foi atribuída ao CO₂; assim há outros caïromônios de importância (PINTO et al., 2001).

Sabendo que 55% da atratividade dos humanos é ligada aos caïromônios, foram identificados mais de 400 compostos na respiração e no suor (DORMONT; BESSIÈRE; COHUET, 2013). A partir de pêlos das pernas de 33 voluntários foram identificados 42 compostos, que foram posteriormente testados em túnel de vento para a espécie *Ny. intermedia*, que foi mais atraída para os compostos 2-fenilacetaldéido e icosano (TAVARES et al., 2018).

Outro cairomônio utilizado como atrativo para mosquitos é o ácido láctico que produz um efeito sinérgico com outros compostos como a acetona, aumentando o número de mosquitos coletados em armadilhas (BERNIER et al., 2015). Para flebotomíneos o ácido láctico não se mostrou atrativo em testes laboratoriais para a espécie *Ny. neivai* (PINTO et al., 2012) ou em testes de campo para *Lu. longipalpis* e *Ny. intermedia* (ANDRADE et al., 2008).

Em outro estudo, um extrato orgânico feito a partir de glândulas da cauda e da região perianal de raposa foi atrativo flebotomíneos da espécie *Lu. longipalpis*; em análise de CG-SSR foram identificados 16 compostos que causaram resposta eletrofisiológica, sendo que os compostos benzaldeído, 4-metil-2-pentanona e 4-hidroxi-4-metil-2-pentanona foram capazes de atrair os insetos em testes comportamentais, porém não com a mesma capacidade do extrato total (DOUGHERTY et al., 1999).

Um cairomônio muito estudado é o 1-octen-3-ol, tendo estudos, que mostram sua atratividade para moscas tsé-tsé (TORR, 1990; TORR et al., 1997), e também para diferentes espécies de mosquito como *An. gambiae* (CORK; PARK, 1996; LU et al., 2007) e do gênero *Aedes* (COOK et al., 2011; KLINE; LEMIRE, 1995).

Para flebotomíneos, o octenol gerou respostas eletrofisiológicas em *Lu. longipalpis*, (SANT'ANA; EIRAS; CAVALCANTE, 2002), e foi atrativo em bioensaios de laboratório (MAGALHÃES-JUNIOR et al., 2014) e em campo (ANDRADE et al., 2008). Além de este composto ser atrativo também para a espécie *Ny. neivai* em campo (PINTO et al., 2011) e laboratório (PINTO et al., 2012). Outros possíveis compostos foram testados em túnel de vento para realização de testes de atratividade. Os álcoois saturados de diferentes tamanhos de cadeias de carbono; desde 1-propanol ao 1-decanol. Os álcoois 1-hexanol e 1-octanol, além do octenol foram mais atrativos para a espécie *Ny. neivai*. em laboratório (MACHADO et al., 2015).

Sendo que para *Lu. longipalpis* os álcoois octenol e nonanol foram atrativos para fêmeas em túnel de vento além dos dois compostos citados, o heptanol também foi atrativo para os machos (MAGALHÃES-JUNIOR et al., 2014)

Com o exposto, observa-se que estudos sobre os aspectos básicos, moleculares e de comportamento, com a utilização de iscas atrativas para a aumento da captura de flebotomíneos são necessários.

No presente trabalho foram utilizadas duas linhas de abordagem da ecologia química em *Lu.longipalpis*. Em uma parte, de caráter molecular, foram estudadas e expressas proteínas ligadoras de odorantes; em outra parte o foco foi para estudos comportamentais, em laboratório e campo, visando avaliar a atratividade de uma mistura de compostos obtida a partir do odor humano.

5 Conclusões e perspectivas

As OBPs 10 e 14 da espécie *Lu. longiplapis* foram expressas com a construção de BL21 (DE3) e pET 28a.

Os testes de atratividade em laboratório mostraram que os machos e fêmeas de *Lu. longipalpis* foram ativados e atraídos para todas as diluições da mistura de compostos de forma dose dependente.

Para testes futuros dos testes de atratividade em laboratório serão realizadas repetições com a mistura com uma maior diluição.

Os testes de atratividade em campo, mostraram uma tendência de coleta para a diluição 1:1000 e 1:10 da mistura de compostos.

Para testes futuros de atratividade em campo serão realizadas mais repetições na mesma região e em outra com maior densidade de insetos.

Anexos

Dois projetos em paralelo aos tratados no corpo dessa tese foram desenvolvidos durante a realização do período de doutorado sanduíche, sob supervisão do Prof. Dr. Rickard Ignell e Profa. Dra. Sharon Hill no *Unit of Chemical Ecology, Swedish University of Agricultural Sciences (SLU)*.

Projeto A: Transcriptoma de antenas de *Anopheles gambiae*

O objetivo desse projeto seria extração de RNA de antenas de *Anopheles gambiae* em diferentes fases da vida e condições fisiológicas do inseto para análise da expressão gênica.

Em virtude de problemas técnicos como: degradação do RNA na fase de transporte e da pandemia de COVID-19, com atrasos da empresa que faria o sequenciamento das amostras, não foi possível a finalização do projeto durante o período do doutorado sanduíche. Entretanto, a colaboração com o grupo da Suécia continua, e as fases finais estão em desenvolvimento.

Projeto B: Análise bioinformática de receptores de odorantes de *Aedes aegypti*

O objetivo geral desse projeto foi verificação da anotação gênica dos receptores de odorantes em *Aedes aegypti*, para clonagem das sequências de aminoácidos em *Drosophila*, para testes de CG-SSR.

Abaixo serão relatadas as etapas que foram desenvolvidas e as perspectivas futuras para cada projeto.

Projeto A:

1. Etapas desenvolvidas na SLU:

1.1. Manutenção da colônia *Anopheles gambiae* e coleta dos tecidos

A espécie *An. gambiae* s.s foi criado em insetário climatizado com temperatura de $27 \pm 2^{\circ}\text{C}$, $70 \pm 2\%$ de umidade relativa e fotoperíodo de 12 horas. Os adultos foram mantidos em gaiolas plásticas cobertas com *voil* e alimentados com sucrose 10% em papel filtro. As fêmeas foram alimentadas com sangue utilizando de um sistema de alimentação artificial (Hemotek Discovery Workshops,

Accrington, UK) com sangue desfibrinado de carneiro. As larvas foram alimentadas com dieta de peixe, mantidas em câmaras com água destilada.

Para minimizar os efeitos do ciclo cicardiano associado com a expressão gênica, as antenas foram coletadas, no início da escotofase, em múltiplos dias. As antenas coletadas foram mantidas a 8°C durante a noite e depois transferidas para o -80°C até a extração do RNA.

Foram extraídas antenas dos insetos com 1 dia após emergência (dpe), 3dpe, 5dpe, 6dpe, 7dpe, 8dpe, 9dpe e 10dpe com o intuito de avaliar a diferença da expressão a cada 24 horas, durante seu ciclo de vida. Além disso, as fêmeas foram separadas em dois grupos, sendo que um foi analisado após alimentação sanguínea que foi realizada no quinto dia após a emergência.

1.2 Extração do RNA e sequenciamento

Os tecidos foram homogeneizados utilizando um sonicador (VCX-130, Sonics and Materials, Newtown, CT) por 10 ciclos a 70% de amplitude, com 1s de pulso (ligado/desligado) repetidos por 3 vezes entre incubações de 30s em gelo. O RNA total foi extraído utilizando o RNeasy Mini Kit (Qiagen, Sweden) seguindo o protocolo do produtor, incluindo o tratamento em coluna *RNase-free DNase*.

O RNA extraído foi quantificado fluorometricamente (Qubit, Life Technologies, Sweden/TapeStation, Agilent Technologies) mantido em -80°C. Uma alíquota foi enviada em gelo seco para Eurofins Genomics (Munique, Alemanha) para a construção da biblioteca de cDNA e sequenciamento Illumina de *single end* (3' RNA-Seq) (Illumina HiSeq 2500). Foram utilizados Oligo-dT baseados no RNA total para extrair mRNA e gerar as bibliotecas de cDNA.

1.3 Mapeamento dos genes

Anteriormente ao mapeamento dos genes, foram removidos as sequências dos adaptadores do sequenciamento Illumina, os *reads* e as bases de baixa qualidade (*Phred score* ≤ 20), com análises do início ao fim de cada *read* para serem recortados com uma abordagem de janela deslizante (tamanho da janela de 4bp, com qualidade mínima de 20) utilizando o Trimmomatic 0.20

(<http://www.usadellab.org>). *Reads* menores que o limite de 40bp foram também removidos. O alinhamento dos *reads* filtrados foi realizado com o genoma que possui as anotações dos genes, os quais serão baseados para a análise do RNA-Seq.

Foi utilizado o CLC Genomics workbench 20.0 (<http://www.clcbio.com>; Qiagen, DK) para mapear as sequências dos *reads* para o genoma de referência no Vectorbase (VECTORBASE, 2019).

2 Resultados das etapas desenvolvidas

Foi realizado no período de Janeiro/2020 a Abril/2020, a dissecação de todas as antenas para a extração do RNA total, após a extração do RNA, a concentração de todas amostras foi analisada no Qubit, apresentando qualidade suficiente para o envio para a empresa BGI que faria a construção das bibliotecas de cDNA.

Entretanto, um problema com a transportadora, degradou o RNA. Após essa indicação pela BGI, tentou-se utilizar o Trio RNA-Seq Library Preparation Kit – NuGEN, em três das 48 amostras, para aumentar a detecção de RNA em amostra de baixa abundância.

Após o envio das três bibliotecas de cDNA, foi realizada a etapa de mapeamento dos genes, na qual boa parte dos *reads* não foi mapeado, ou então estavam em regiões de íntrons, o que não deveria ocorrer, pois o material é de RNA (Fig. 25)

4 Mapping statistics

4.1 Single reads

	Number of sequences	%
Reads mapped	7.001.919	29,22
Reads not mapped	16.957.183	70,78
Total	23.959.102	100,00

5 Fragment statistics

5.1 Fragment counting (total)

	Fragments	% of total
Counted fragments	7.001.919	29,22
- unique fragments	4.278.138	17,86
- non-specifically	2.723.781	11,37
Uncounted fragments	16.957.183	70,78
Total fragments	23.959.102	100,00

Figura 25. Parte do relatório de mapeamento dos *reads* no software CLC.

O processo de envio das amostras até o recebimento das bibliotecas levou cerca de 3 meses. Antes da finalização dos mapeamentos, foi decidido por recomeçar a dissecção de todas as 48 amostras.

Após a segunda dissecção, o RNA total foi novamente extraído e dessa vez analisado, utilizando o TapeStation (agilent Technologies). Os resultados detalhados estão na tabela 6.

Essas amostras foram enviadas para a Eurofins Genomics, Alemanha, onde se encontram no momento.

Assim que as bibliotecas de cDNA estejam construídas serão realizados o mapeamento dos *reads* como descrito acima e a análise do RNA-seq.

Tabela 7. Detalhamento das 48 amostras de RNA total enviadas à Eurofins em 2020.

Data gaiolas	Data dissecção	Data alimentação sanguínea	Idade	Coorte	Hora	Tubos	RNA conc (pg/μL)
26/ago	26/ago		1dpe	NBF	20h	V1	3170
23/ago	27/ago	27/ago	5dpe	BF	20h	V2	3370
23/ago	27/ago	27/ago	5dpe	BF	20h	V3	1810
23/ago	28/ago	27/ago	6dpe	BF	20h	V4	995
20/ago	28/ago	24/ago	9dpe	BF	20h	V5	1970
21/ago	29/ago	25/ago	9dpe	BF	20h	V6	991
21/ago	29/ago	25/ago	9dpe	BF	20h	V7	1850
24/ago	02/set		10dpe	NBF	20h	V8	809
24/ago	02/set		10dpe	NBF	20h	V9	2250
25/ago	03/set	29/ago	10dpe	BF	17h	V10	723
26/ago	03/set		9dpe	NBF	20h	V11	1850
26/ago	03/set		9dpe	NBF	20h	V12	1340
27/ago	04/set		9dpe	NBF/ovi	17h	V13	1590
27/ago	04/set		9dpe	NBF/ovi	17h	V14	957
05/set	09/set		5dpe	NBF	20h	V15	1230
07/set	09/set		3dpe	NBF	20h	V16	628
08/set	10/set		3dpe	NBF	20h	V17	12400
06/set	10/set	10/set	5dpe	BF	20h	V18	2020

05/set	11/set	09/set	7dpe	BF	17h	V19	1420
05/set	13/set		9dpe	NBF/ovi	17h	V20	1160
05/set	13/set		9dpe	NBF	17h	V21	990
07/set	13/set		7dpe	NBF	20h	V22	695
11/set	13/set		3dpe	NBF	20h	V23	633
05/set	14/set	09/set	10dpe	BF	20h	V24	1730
05/set	14/set	09/set	10dpe	BF	20h	V25	1140
08/set	15/set		8dpe	NBF	17h	V26	2290
08/set	15/set		8dpe	NBF	17h	V27	946
06/set	15/set		10dpe	NBF/ovi	20h	V28	1690
06/set	15/set		10dpe	NBF	20h	V29	1800
07/set	16/set		10dpe	NBF/ovi	17h	V30	1770
07/set	16/set		10dpe	NBF/ovi	17h	V31	2440
09/set	16/set	13/set	8dpe	BF	20h	V32	2170
12/set	17/set		6dpe	NBF	17h	V33	1330
22/set	22/set		1dpe	NBF	20h	V34	1370
22/set	22/set		1dpe	NBF	20h	V35	1980
17/set	23/set		7dpe	NBF	20h	V37	2130
19/set	23/set		5dpe	NBF	20h	V36	1620
19/set	24/set		6dpe	NBF	20h	V38	1270
17/set	24/set		8dpe	NBF	20h	V39	1150

18/set	25/set	23/set	8dpe	BF	17h	V40	1310
18/set	25/set	23/set	8dpe	BF	17h	V41	1300
13/out	19/out		7dpe	NBF	17h	V42	759
15/out	20/out	19/out	6dpe	BF	17h	V43	586
15/out	20/out	19/out	6dpe	BF	17h	V44	835
16/out	20/out		5dpe	NBF	20h	V45	729
14/out	20/out	18/out	7dpe	BF	20h	V46	672
14/out	20/out	18/out	7dpe	BF	20h	V47	504
16/out	21/out		6dpe	NBF	17h	V48	429

3 Etapas em desenvolvimento e futuras

3.1 Análise do RNA-Seq

Após o mapeamento *dos reads*, as análises do RNA-Seq serão realizadas utilizando CLC Genomics Workbench 9.0 (<https://www.qiagenbioinformatics.com/>). Serão aplicados limites de contagens médias para diminuição do ruído, similar ao filtro de RPKM <1 em cada uma das bibliotecas geradas (RINKER et al., 2013).

Para o tratamento dos dados serão utilizados testes estatísticos para a normalização dos dados, como por exemplo, o método quantile, que aumenta a sensibilidade da detecção do diferencial de expressão em transcritos pouco abundantes (BOLSTAD et al., 2003).

Após a normalização dos dados, será aplicado o teste Z descrito por (KAL et al., 1999) para aproximação da distribuição binomial, e por fim será aplicada a correção de Benjamini-Hochberg para o controle de falsos positivos (YOAV BENJAMINI AND YOSEF HOCHBERG, 1995).

A partir dos dados tratados será avaliada a expressão diferenciada seguindo os valores de *fold change* (FC), sendo $FC > 2$ e $P < 0,05$ demonstram uma diferença de expressão significativa e $1,5 \geq FC \geq 2$ e $P < 0,05$ não é considerada significativa, porém pode haver um interesse potencial.

Projeto B

1. Etapa realizada

1.1. Verificação da anotação gênica dos receptores de odorantes em *Aedes aegypti*

Para realização de estudos com SSR, o primeiro passo é a expressão dos receptores de odorante na sensilas de linhagens apropriadas de *Drosophila*. Para alcançar esse fim, utiliza-se o banco de dados *Vectorbase* para encontrar as sequências de aminoácidos desses receptores, porém em alguns casos, essas sequências se encontram muito curtas, ou muito longas, no qual é necessária a confirmação se há a ausência de algum éxon, por exemplo.

Para isso, foi feita a extração de sequências consensuais dos receptores de odorantes de *Aedes aegypti* utilizando o software CLC Genomics workbench 20.0, a partir de bibliotecas de cDNA de estudos prévios do *Unit of chemical ecology*. Essas sequências foram utilizadas como meio para a confirmação das alterações na anotação gênica dos receptores de odorantes.

Todos os receptores de odorantes analisados tiveram suas sequências comparadas às consensuais de nossas bibliotecas, mesmo as que apresentavam tamanho dentro do esperado para receptores de odorantes, e que eram similares aos seus ortólogos.

Quando uma sequência com tamanho inesperado era encontrada, o primeiro passo era a visualização de seus parálogos e ortólogos no *VectorBase*, após a identificação do seu parálogo mais próximo, foi feito o alinhamento da sequência para identificar qual éxon estava faltando, se era na região do 5', 3' ou mesmo um intermediário.

Após esses passos, foi realizada a busca pela região cromossômica do gene e eram incluídas sequências extras na região que estava faltando para tentar encontrar o éxon perdido.

Com a sequência aumentada foi utilizado o servidor Sixpack no EMBL (<https://www.ebi.ac.uk/Tools/st/>) para traduzir a sequência em 6 possíveis *frames* de leitura da sequência de aminoácidos. Foram destacados em amarelo os aminoácidos mostrados na sequência do *Vectorbase*, avaliando as posições de início de íntrons, os quais finalizam sempre em “AG” e de éxons, os quais finalizam sempre em “GT”.

Após isso, utilizando o alinhamento com o parálogo, por exemplo, caso o alinhamento mostre a falta de um éxon no 5', ignore-se a metionina inicial que está incorreta, e estende a busca na direção 5' até encontrar um stop códon, e então busca na direção contrária a sinalização de final de íntron “AG”.

Assim que foi encontrado o fim do íntron, busca-se novamente na direção 5' a sequência que está faltando, ela deve ser similar ao parálogo, quando essa sequência é encontrada, continua-se na direção 5' até que se encontre um stop códon, que é a indicação que há a presença de outro íntron, repetindo as ações descritas, ou encontra a metionina inicial com a presença da sequência de Kozak, que é um sinal de inicialização de transcrição. Para o *Ae. aegypti* as sequências mais comuns são “AAAATGG” e “ACGATGG” (MATTHEWS et al., 2016).

Com isso é montada a sequência putativa de aminoácidos do receptor de odorante, que foi então comparada com o parálogo e com a sequência consensual obtida pelas bibliotecas de cDNA.

1.2 Resultados das análises de bioinformática

No total foram avaliadas 86 sequências de receptores de odorantes de *Aedes aegypti*, sendo que todas foram comparadas com as sequências consensuais extraídas das bibliotecas de cDNA de mosquitos de diferentes fases do ciclo de vida.

Algumas das que foram necessárias modificações estão exemplificadas abaixo.

OR114: A sequência estava curta no início da tradução em comparação com o parálogo, então foi buscado a sequência de Kozak, que indica a região da metionina inicial, essa parte foi acrescentada para a OR114 proposta.

OR14: A sequência estava curta no final da tradução em comparação com o parálogo, buscando a sinalização AG e GT, foi buscado o éxon que faltava e acrescentado na OR14 proposta.

OR76: A sequência estava curta no final da tradução em comparação com o parálogo, buscando a sinalização AG e GT, foi buscado o éxon, porém em comparação com o parálogo havia uma região que se encontra um stop códon, foi possível encontrar outro íntron, que não está presente no parálogo, assim essa parte foi retirada e outro éxon acrescentado na OR76 proposta.

As partes acrescidas estão destacadas em azul e a que já estava presente no *Vectorbase* em amarelo. A sequência de Kozak e o início e fim dos íntrons e éxons está destacado em vermelho.

AaGPROr114 – OR114

>AAEL017219

MFRVVVEVVKFTAFVVAGCKLLLLKLHRKSITSIRSYINDARMITIDDSYDLLELSKF
 KHRSM TMIRGIYGLIVIDTVLLFIPNDTTEMAFDRRSDTLPFMSLERNIYRLLFLTLL
 PVGFLPKFFSSMATIGSLLLGM EANFKILANRFRSILSQPFVINGTDWEKMNSELR
 DTLKQHLEFWCHFKALKKLVGETFFLVHYFSIMSIGALCYICQAIGVNFLSFV VMA
 TLAMFLMEYYMFCHFVDSFQDVANCIGEHIFQIAILMPHNRKNHSHYIGFRTALMII
 WMNTRREVSMDCMGLFNISTA AFLHVLNIA YTVLTF LIQMS

Sequência curta no *Vectorbase*

```

V K K G L * * I Y T L I C P T N N Q N K      F1
L K K D C N E F I R S F V P L I T K I S      F2
* K R I V M N L Y A H L S H * * P K * V      F3
1 GTTAAAAAAGGATTGTAATGAATTTATACGCTCATTTGTCCCACTAATAACCAAAATAAG 60
----:----|----:----|----:----|----:----|----:----|----:----|
1 CAATTTTTCTCCTAACATTACTTAAATATGCAGTAAACAGGGTGATTATTGGTTTATTC 60
T L F P N Y H I * V S M Q G V L L W F L      F6
X * F L I T I F K Y A * K D W * Y G F Y      F5
N F F S Q L S N I R E N T G S I V L I L      F4

* P K T I D K Q S I D Y L V * L Q I F K      F1
D P K L L I N N Q L I I W F N Y K F S K      F2
T Q N Y * * T I N * L F G L T T N F Q S      F3
61 TGACCCAAACTATTGATAACAATCAATTGATTATTTGGTTTAACTACAAATTTTCAA 120
----:----|----:----|----:----|----:----|----:----|----:----|
61 ACTGGGTTTTGATAACTATTTGTTAGTTAACTAATAAACCAAATTGATGTTTAAAGTTT 120
H G L V I S L C D I S * K T * S C I K L      F6
T V W F * Q Y V I L Q N N P K V V F K *      F5
S G F S N I F L * N I I Q N L * L N E F      F4

V I L N N S D Y * K * P S L K H S F Y K      F1
L F * I I L I I R N N L H * N T L S I N      F2
Y F E * F * L L E I T F I K T L F L * I      F3
121 GTTATTTTGAATAATTCTGATTATTAGAAATAACCTTCATTAACACTCTTTCTATAAA 180
----:----|----:----|----:----|----:----|----:----|----:----|
121 CAATAAACTTATTAAGACTAATAATCTTTATTGGAAGTAATTTGTGAGAAAGATATTT 180
T I K F L E S * * F Y G E N F C E K * L      F6
L * K S Y N Q N N S I V K M L V S K R Y      F5
N N Q I I R I I L F L R * * F V R E I F      F4

Y R F D S Q I V D Q S C G K E R L V F N      F1
T V L T V K L S I K A V G R K G L F L I      F2
P F * Q S N C R S K L W E G K A C F * L      F3
181 TACCGTTTTGACAGTCAAATTGTCGATCAAAGCTGTGGGAAGGAAAGGCTTGTTTTAAT 240
----:----|----:----|----:----|----:----|----:----|----:----|
181 ATGGCAAACTGTCAAGTTTAAACAGCTAGTTTCGACACCCTTCCTTTCCGAACAAAATTA 240
Y R K S L * I T S * L Q P F S L S T K L      F6
I G N Q C D F Q R D F S H S P F A Q K *      F5
V T K V T L N D I L A T P L F P K N K I      F4

* F C L L L F L D Y I H M V F * N S * S      F1
N F A F C Y F * T I S T W S S K I H K A      F2
I L P F V I F R L Y P H G L L K F I K L      F3
241 TAATTTTGCTTTTGTATTTTACTATATCCACATGGTCTTCTAAAATTCATAAAGC 300
----:----|----:----|----:----|----:----|----:----|----:----|
241 ATTAACCGGAAAACAATAAAAATCTGATATAGGTGTACCAGAAGATTTTAAGTATTTTCG 300
* N Q R K N N K S * I W M T K * F E Y L      F6

```

```

N I K G K T I K L S Y G C P R R F N M F      F5
L K A K Q * K * V I D V H D E L I * L A      F4

F L R T L A L Q R Q * I K Q * I L L I K      F1
S C E R * P F S A N K * S N K Y Y * L N      F2
P A N V S P S A P I N K A I N I I D * T      F3
301 TTCCTGCGAACGTTAGCCCTTCAGCGCCAATAAAATAAGCAATAAATATTATTGATTAAA 360
----:----|----:----|----:----|----:----|----:----|----:----|
301 AAGGACGCTTGCAATCGGGAAGTCGCGGTTATTTATTTTCGTTATTTATAATAACTAATTT 360
K R R V N A R * R W Y I F C Y I N N I L      F6
S G A F T L G E A G I F L A I F I I S *      F5
E Q S R * G K L A L L Y L L L Y * Q N F      F4

L T H Q R T L T L A I S R Y T P Y T Q G      F1
L L I K E H * H * Q Y R D T R L I L K V      F2
Y S S K N I D I S N I E I H A L Y S R S      F3
361 CTTACTCATCAAAGAACATTGACATTAGCAATATCGAGATACACGCCTTATACTCAAGGT 420
----:----|----:----|----:----|----:----|----:----|----:----|
361 GAATGAGTAGTTTCTTGTAAGTGTAAATCGTTATAGCTCTATGTGCGGAATATGAGTTCCA 420
S V * * L V N V N A I D L Y V G * V * P      F6
V * E D F F M S M L L I S I C A K Y E L      F5
K S M L S C Q C * C Y R S V R R I S L T      F4

L V A I S Q Y N S L Y D C L Q I F S L C      F1
L L Q * V S I I H S T I V F K F S V F A      F2
C C N K S V * F T L R L S S N F Q S L L      F3
421 CTGTGTGCAATAAGTCAGTATAATTCAGTCTACGATTGTCTCAAATTTTCAGTCTTTGC 480
----:----|----:----|----:----|----:----|----:----|----:----|
421 GAACAACGTTATTCAGTCATATTAAGTGAGATGCTAACAGAAAGTTAAAAGTCAGAAACG 480
R T A I L * Y L E S * S Q R * I K L R Q      F6
D Q Q L L D T Y N V R R N D E F K * D K      F5
K N C Y T L I I * E V I T K L N E T K A      F4

L S * E R K C Y S V F R V A * F V F V N      F1
* A E N E N V I P F S E * H D L C S * M      F2
K L R T K M L F R F Q S S M I C V R K W      F3
481 TTAAGCTGAGAACGAAATGTTATTCCGTTTTCAGAGTAGCATGATTTGTGTTCGTAAAT 540
----:----|----:----|----:----|----:----|----:----|----:----|
481 AATTCGACTCTTGCTTTTACAATAAGGCAAAAGTCTCATGCTACTAAACACAAGCATTTA 540
K L Q S R F H * E T K L T A H N T N T F      F6
S L S L V F I N N R K * L L M I Q T R L      F5
* A S F S F T I G N E S Y C S K H E Y I      F4

G G H L * W P K R G I F G P T T D L G V      F1
V V I C S G Q K E V F L V R Q R T W E *      F2
W S S V V A K K R Y F W S D N G P G S N      F3
541 GGTGGTCATCTGTAGTGGCCAAAAAGAGGTATTTTGGTCCGACAACGGACCTGGGAGTA 600
----:----|----:----|----:----|----:----|----:----|----:----|
541 CCACCAGTAGACATCACCGGTTTTTCTCCATAAAACCAGGCTGTTGCCTGGACCTCAT 600
P P * R Y H G F L P I K P G V V S R P T      F6
H H D D T T A L F L Y K Q D S L P G P L      F5
T T M Q L P W F S T N K T R C R V Q S Y      F4

I A S S G K M C Y C S L E V G * V S K V      F1
L L L L A R C V T A H W R * V K Y L K *      F2
C F F W Q D V L L T G G R L S I * S S      F3
601 ATTGCTTCTTCTGGCAAGATGTGTTACTGCTCACTGGAGGTAGTTAAGTATCTAAAGTA 660
----:----|----:----|----:----|----:----|----:----|----:----|
601 TAACGAAGAAGACCGTTCTACACAATGACGAGTGACCTCCATCCAATTCATAGATTTTCAT 660
I A E E P L I H * Q E S S T P * T D L T      F6
L Q K K Q C S T N S S V P P L N L I * L      F5
N S R R A L H T V A * Q L Y T L Y R F Y      F4

```

```

V C * V T I Y I T * I E C L L A * I V P      F1
F V K * P F T * L K L S V F * H E S F Q      F2
L L S N H L H N L N * V S S S M N R S K      F3
661 GTTTGTAAAGTAACCATTACATAACTTAAATTGAGTGTCTTCTAGCATGAATCGTTCCA 720
----:----|----:----|----:----|----:----|----:----|----:----|
661 CAAACAATTCATTGGTAAATGTATTGAATTAACTCACAGAAGATCGTACTTAGCAAGGT 720
T Q * T V M * M V * I S H R R A H I T G      F6
L K N L L W K C L K F Q T D E L M F R E      F5
N T L Y G N V Y S L N L T K * C S D N W      F4

```

```

N H F H Y W T Q L S I F K Q E Y K V N S      F1
I I F I I G L N Y L S L N R N T K * T H      F2
S F S L L D S I I Y L * T G I Q S E L T      F3
721 AATCATTTTCATTATTGGACTCAATTATCTATCTTTAAACAGGAATACAAAGTGAACCTCA 780
----:----|----:----|----:----|----:----|----:----|----:----|
721 TTAGTAAAAGTAATAACCTGAGTTAATAGATAGAAATTTGTCCTTATGTTTCACTTGAGT 780
F * K * * Q V * N D I K L C S Y L T F E      F6
L D N E N N S E I I * R * V P I C L S S      F5
I M K M I P S L * R D K F L F V F H V *      F4

```

```

H * S G A V N V T F A S * W H V Y S P F      F1
I E V E L * T L H S L L S G M S I L P S      F2
L K W S C E R Y I R F L V A C L F S L Q      F3
781 CATTGAAGTGGAGCTGTGAACGTTACATTCGCTTCTTAGTGGCATGTCTATTCTCCCTTC 840
----:----|----:----|----:----|----:----|----:----|----:----|
781 GTAATTCACCTCGACACTTGCAATGTAAGCGAAGAATCACCGTACAGATAAGAGGGAAG 840
C Q L P A T F T V N A E * H C T * E G K      F6
V N F H L Q S R * M R K K T A H R N E R      F5
M S T S S H V N C E S R L P M D I R G E      F4

```

```

K H S S F L * Y F I V L C R T I E P K C      F1
N I L H F C S T L * Y F A E Q * S R N V      F2
T F F I F V V L Y S T L Q N N R A E M F      F3
841 AAACATTCTTCATTTTTGTAGTACTTTTATAGTACTTTGCAGAACAAATAGAGCCGAAATGT 900
----:----|----:----|----:----|----:----|----:----|----:----|
841 TTTGTAAGAGTAAAAACATCATGAAATATCATGAAACGTCTTGTTATCTCGGCTTTACA 900
L C E E N K Y Y K I T S Q L V I S G F H      F6
* V N K M K T T S * L V K C F L L A S I      F5
F M R * K Q L V K Y Y K A S C Y L R F T      F4

```

```

L E L L S K L * N S R P L L L L A A N C      F1
* S C C R S C E I H G L C C C W L Q I A      F2
R V V V E V V K F T A F V V A G C K L L      F3
901 TTAGAGTTTGTGTCGAAGTTGTGAAATTCACGGCCTTTGTTGTTGCTGGCTGCAAATTGC 960
----:----|----:----|----:----|----:----|----:----|----:----|
901 AATCTCAACAACAGCTTCAACACTTTAAGTGCCGAAACAACAACGACCGACGTTTAACG 960
K S N N D F N H F E R G K N N S A A F Q      F6
N L T T T S T T F N V A K T T A P Q L N      F5
* L Q Q R L Q S I * P R Q Q Q Q S C I A      F4

```

```

C C * N S I G N R L P V S G V T L M M R      F1
V V K T P S E I D Y Q Y Q E L H * * C A      F2
L L K L H R K S I T S I R S Y I N D A R      F3
961 TGTTGTTAAAACTCCATCGGAAATCGATTACCAGTATCAGGAGTTACATTAATGATGCGC 1020
----:----|----:----|----:----|----:----|----:----|----:----|
961 ACAACAATTTGAGGTAGCCTTTAGCTAATGGTCATAGTCCTCAATGTAATTACGCG 1020
Q Q * F E M P F R N G T D P T V N I I R      F6
S N N F S W R F D I V L I L L * M L S A      F5
T T L V G D S I S * W Y * S N C * H H A      F4

```

```

V * S P L M T A M I C L N * A S S S I D F1
Y D H H * * Q L * F A * T E Q V Q A S I F2
M I T I D D S Y D L L E L S K F K H R S F3
1021 GTATGATCACCATTGATGACAGCTATGATTTGCTTGAAGTCAAGTTCATCGAT 1080
----:----|----:----|----:----|----:----|----:----|----:----|
1021 CATACTAGTGGTAACTACTGTCGATACTAAACGAACTTGACTCGTTCAAGTTCGTAGCTA 1080
T H D G N I V A I I Q K F Q A L E L M S F6
R I I V M S S L * S K S S S L L N L C R F5
Y S * W Q H C S H N A Q V S C T * A D I F4

```

```

L * Q * Y V A F M A * S * S I R F Y F L F1
Y D N D T W H L W L N R D R Y G S T F Y F2
M T M I R G I Y G L I V I D T V L L F I F3
1081 CTATGACAATGATACGTGGCATTATGGCTTAATCGTGATCGATACGGTTCTACTTTTTA 1140
----:----|----:----|----:----|----:----|----:----|----:----|
1081 GATACTGTTACTATGCACCGTAAATACCGAATTAGCACTAGCTATGCCAAGATGAAAAAT 1140
R H C H Y T A N I A * D H D I R N * K K F6
D I V I I R P M * P K I T I S V T R S K F5
* S L S V H C K H S L R S R Y P E V K * F4

```

```

Y Q T T P Q K W L L I G D L I H C L S C F1
T K R H H R N G F * * A I * Y T A F H V F2
P N D T T E M A F D R R S D T L P F M S F3
1141 TACCAAACGACACCACAGAAATGGCTTTTGTAGGCGATCTGATACACTGCCTTTCATGT 1200
----:----|----:----|----:----|----:----|----:----|----:----|
1141 ATGGTTTGTGTGGTGTCTTTACCGAAAATATCCGCTAGACTATGTGACGGAAAGTACA 1200
Y W V V G C F H S K I P S R I C Q R E H F6
I G F S V V S I A K S L R D S V S G K M F5
V L R C W L F P K Q Y A I Q Y V A K * T F4

```

```

H L S V I F T G Y S S * H F C Q W D S C F1
T * A * Y L P V T L L N T S A S G I L A F2
L E R N I Y R L L F L T L L P V G F L L P F3
1201 CACTTGAGCGTAATATTTACCGTTACTCTTCTTAACACTTCTGCCAGTGGGATTCTTGC 1260
----:----|----:----|----:----|----:----|----:----|----:----|
1201 GTGAACCTCGCATTATAAATGGCCAATGAGAAGAATTGTGAAGACGGTCACCTAAGAACG 1260
* K L T I N V P * E E * C K Q W H S E Q F6
D S S R L I * R N S K K V S R G T P N K F5
V Q A Y Y K G T V R R L V E A L P I R A F4

```

```

R N F S A A W R P L D L F C * E W K P T F1
E I F Q Q H G D H W I S F V R N G S Q L F2
K F F S S M A T I G S L L L G M E A N F F3
1261 CGAAATTTTTCAGCAGCATGGCGACCATTTGGATCTCTTTTGTAGGAATGGAAGCCAACT 1320
----:----|----:----|----:----|----:----|----:----|----:----|
1261 GCTTTAAAAAGTCGTCGTACCGCTGGTAACCTAGAGAAAACAATCCTTACCTTCGGTTGA 1320
R F K E A A H R G N S R K Q * S H F G V F6
G F N K L L M A V M P D R K N P I S A L F5
S I K * C C P S W Q I E K T L F P L W S F4

```

```

L K F * P T A F V R F * V N R L L S M E F1
* N F S Q P L S F D F E S T V C Y Q W N F2
K I L A N R F R S I L S Q P F V I N G T F3
1321 TTAAAATTTTAGCCAACCGCTTTCGTTTCGATTTTGAGTCAACCGTTTGTATCAATGGAA 1380
----:----|----:----|----:----|----:----|----:----|----:----|
1321 AATTTTAAAAATCGGTTGGCGAAAGCAAGCTAAAACCTCAGTTGGCAAACAATAGTTACCTT 1380
K F N * G V A K T R N Q T L R K N D I S F6
K L I K A L R K R E I K L * G N T I L P F5
* F K L W G S E N S K S D V T Q * * H F F4

```

```

L T G K K * I V N * G I L * N S I W N F F1
* L G K N E * * I E G Y S E T A S G I L F2

```

D W E K M N S E L R D T L K Q H L E F W F3
 1381 CTGACTGGGAAAAAATGAATAGTGAATTGAGGGATACTCTGAAACAGCATCTGGAATTTT 1440
 ----:----|----:----|----:----|----:----|----:----|
 1381 GACTGACCCTTTTACTTACTTAACTCCCTATGAGACTTTGTCGTAGACCTTAAAA 1440
 S V P F F H I T F Q P I S Q F L M Q F K F6
 V S Q S F I F L S N L S V R F C C R S N F5
 Q S P F F S Y H I S P Y E S V A D P I K F4

 G G K F R Y I Y R * N L A L T F I Y L Y F1
 V V S L D I F I V E I * L S R S S I Y I F2
 W * V * I Y L S L K F S S H V H L S I Y F3
 1441 GGTGGTAAGTTTAGATATATTTATCGTTGAAATTTAGCTCTCACGTTTCATCTATCTATAT 1500
 ----:----|----:----|----:----|----:----|----:----|----:----|
 1441 CCACCATTCAAATCTATATAAATAGCAACTTTAAATCGAGAGTGCAAGTAGATAGATATA 1500
 P P L N L Y I * R Q F K A R V N M * R Y F6
 Q H Y T * I Y K D N F N L E * T * R D I F5
 T T L K S I N I T S I * S E R E D I * I F4

 I F S H F K A L K K L V G E T F F L V H F1
 F S A T L K P * R N W S E K H S F L F I F2
 F Q P L * S L E E I G R R N I L S C S L F3
 1501 ATTTTCAGCCACTTTAAAGCCTTGAAGAAATTGGTCGGAGAAACATTCTTTCTGTTCAT 1560
 ----:----|----:----|----:----|----:----|----:----|----:----|
 1501 TAAAGTCGGTGAAATTTCGGAACCTTCTTTAACCAGCCTCTTTGTAAGAAAGAACAAGTA 1560
 I K L W K L A K F F N T P S V N K R T * F6
 Y K * G S * L R S S I P R L F M R E Q E F5
 N E A V K F G Q L F Q D S F C E K K N M F4

 Y F S I M S I G A L C Y I C Q A I G V N F1
 I F Q S * A S G L C A I F V K P L E * T F2
 F F N H E H R G S V L Y L S S H W S E L F3
 1561 TATTTTCAATCATGAGCATCGGGGCTCTGTGCTATATTTGTCAAGCCATTGGAGTGAAC 1620
 ----:----|----:----|----:----|----:----|----:----|----:----|
 1561 ATAAAAAGTTAGTACTCGTAGCCCCGAGACATATATAAACAGTTCGGTAACCTCACTTG 1620
 * K E I M L M P A R H * I Q * A M P T F F6
 N N K L * S C R P E T S Y K D L W Q L S F5
 I K * D H A D P S Q A I N T L G N S H V F4

 F L S F V V M A T L A M F L M E Y Y M F F1
 F C H L S L W Q H W Q C F L W N T T C F F2
 F V I C R Y G N T G N V S Y G I L H V L F3
 1621 TTTTGTGTCATTTGTCGTTATGGCAACACTGGCAATGTTTCTTATGGAATACTACATGTTT 1680
 ----:----|----:----|----:----|----:----|----:----|----:----|
 1621 AAAACAGTAAACAGCAATACCGTTGTGACCGTTACAAAGAATACCTTATGATGTACAAA 1680
 K K D N T T I A V S A I N R I S Y * M N F6
 S K T M Q R * P L V P L T E * P I S C T F5
 K Q * K D N H C C Q C H K K H F V V H K F4

 C H F V D S F Q D V V S * K V F Y S S L F1
 A I L L I L F K M L * A R K C S I L L * F2
 P F C * F F S R C C K L E S V L F F S K F3
 1681 TGCCATTTTGTTGATTCTTTTCAAGATGTTGTAAGCTAGAAAGTGTTCTATTCTTCTCTA 1740
 ----:----|----:----|----:----|----:----|----:----|----:----|
 1681 ACGGTAACAACTAAGAAAAGTTCTACAACATTCGATCTTTCACAAGATAAGAAGAGAT 1740
 Q W K T S E K * S T T L * F T N * E E R F6
 K G N Q Q N K E L H Q L S S L T R N K E F5
 A M K N I R K L I N Y A L F H E I R R * F4

 K V L S K C I Y S * A N C I G E H I F Q F1
 K F F Q N V F I R R L T V L A S I F S R F2
 S S F K M Y L F V G * L Y W R A Y F P D F3
 1741 AAAGTTCTTTCAAATGTATTTATTCGTAGCTAACTGTATTGGCGAGCATATTTTCCAG 1800

```

-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1741 TTTCAAGAAAGTTTACATAAAATAAGCATCCGATTGACATAACCGCTCGTATAAAAGGTC 1800
    F T R E F H I * E Y A L Q I P S C I K W F6
    L L E K L I Y K N T P * S Y Q R A Y K G F5
    F N K * F T N I R L S V T N A L M N E L F4

    I A I L M P H N R K N H S H Y I G F R T F1
    L L F L C P T T G R T T H I T L V S E Q F2
    C Y S Y A P Q P E E P L T L H W F P N S F3
1801 ATTGCTATTCTTATGCCCCACAACCGGAAGAACCACTCACATTACATTGGTTTCCGAACA 1860
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1801 TAACGATAAGAATACGGGGTGTGGCCTTCTTGGTGAGTGTAATGTAACCAAAGGCTTGT 1860
    I A I R I G W L R F F W E C * M P K R V F6
    S Q * E * A G C G S S G S V N C Q N G F F5
    N S N K H G V V P L V V * M V N T E S C F4

    A L M I I W M N T R R E V S M D C M G L F1
    R * * * S G * T R V A K F P W T A W G Y F2
    V D D N L D E H A S R S F H G L H G V I F3
1861 GCGTTGATGATAATCTGGATGAACACGCGTCGCGAAGTTTCCATGGACTGCATGGGGTTA 1920
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1861 CGCAACTACTATTAGACCTACTTGTGCGCAGCGCTTCAAAGGTACCTGACGTACCCCAAT 1920
    A N I I I Q I F V R R S T E M S Q M P N F6
    L T S S L R S S C A D R L K W P S C P T F5
    R Q H Y D P H V R T A F N G H V A H P * F4

    F N I S T A A F L H V L N I A Y T V L T F1
    S T F Q L Q L F S T C L T L L T Q C * H F2
    Q H F N C S F S P R A * H C L H S V D I F3
1921 TTCAACATTTCAACTGCAGCTTTTCTCCACGTGCTTAACATTGCTTACACAGTGTTGACA 1980
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1921 AAGTTGTAAAGTTGACGTCGAAAAGAGGTGACGAATTGTAACGAATGTGTCAAACTGT 1980
    N L M E V A A K R W T S L M A * V T N V F6
    I * C K L Q L K E G R A * C Q K C L T S F5
    E V N * S C S K E V H K V N S V C H Q C F4

    F L I Q M S * F * T I T L L C R I R N Y F1
    F * S K * V N S K P * L Y Y A V * E I I F2
    S N P N E L I L N H N F I M P Y K K L F F3
1981 TTTCTAATCCAAATGAGTTAATTCTAAACCATAACTTTATTATGCCGTATAAGAAATTAT 2040
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1981 AAAGATTAGGTTTACTCAATTAAGATTTGGTATTGAAATAATACGGCATATTCTTTAATA 2040
    N R I W I L * N * V M V K N H R I L F * F6
    M E L G F S N I R F W L K I I G Y L F N F5
    K * D L H T L E L G Y S * * A T Y S I I F4

    L K V Q Q F K L S I K H * F R T R R G * F1
    * K F S N S S * V * N I D L E Q G E V R F2
    E S S A I Q A K Y K T L I * N K E R L E F3
2041 TTGAAAGTTCAGCAATTCAAGCTAAGTATAAAACATTGATTAGAACAAAGGAGAGGTTAG 2100
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
2041 AACTTTCAAGTCGTTAAGTTCGATTCATATTTGTAACATAATCTTGTTCCCTCTCCAATC 2100
    K F T * C N L S L I F C Q N L V L L P * F6
    N S L E A I * A L Y L V N I * F L S L N F5
    Q F N L L E L * T Y F M S K S C P S T L F4

    N S Q L Y Q I R N V L Y C S R E S H G F F1
    I H S Y I K Y V M F Y T A V E S L M V L F2
    F T V I S N T * C S I L Q * R V S W F Y F3
2101 AATTCACAGTTATATCAAATACGTAATGTTCTATACTGCAGTAGAGAGTCTCATGGTTTT 2160
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
2101 TTAAGTGTCAATATAGTTTATGCATTACAAGATATGACGTCATCTCTCAGAGTACCAAAA 2160

```

```

F E C N Y * I R L T R Y Q L L S D * P K F6
S N V T I D F V Y H E I S C Y L T E H N F5
I * L * I L Y T I N * V A T S L R M T K F4

I L Q C K Y I I Y V E E K K N N I I N R F1
Y Y S A N I * Y M * K K K K T T * * I G F2
I T V Q I Y N I C R R K K K Q H N K S V F3
2161 ATATTACAGTCAAATATATAATATATGTAGAGAAAAAACAACATAATAATCGG 2220
-----|-----|-----|-----|-----|-----|-----|
2161 TATAATGTCACGTTTATATATTATATACATCTTCTTTTTTTTTTGTGTATTATTTAGCC 2220
I N C H L Y I I Y T S S F F F L M I F R F6
* I V T C I Y L I H L L F F F C C L L D F5
Y * L A F I Y Y I Y F F F F V V Y Y I P F4

Y D F * Y I Q D R N T F L Q Q K S Y C I F1
M I F S I F K T E T H F C S K K A I V Y F2
* F L V Y S R Q K H I F A A K K L L Y I F3
2221 TATGATTTTTAGTATATTCAAGACAGAAACACATTTTTGCAGCAAAAAAGCTATTGTATA 2280
----:----|----:----|----:----|----:----|----:----|----:----|
2221 ATACTAAAAATCATATAAGTTCTGTCTTTGTGTAAAAACGTCGTTTTTCGATAACATAT 2280
Y S K * Y I * S L F V N K C C F L * Q I F6
T H N K T Y E L C F C M K A A F F S N Y F5
I I K L I N L V S V C K Q L L F A I T Y F4

F V W E P V S L Q D I I R K H V * L D E F1
L F G S L F L C K I L * G N M F D * M N F2
C L G A C F F A R Y Y K E T C L I R * I F3
2281 TTTGTTTGGGAGCCTGTTTCTTTGCAAGATATTATAAGGAAACATGTTTGATTAGATGAA 2340
----:----|----:----|----:----|----:----|----:----|----:----|
2281 AAACAAACCTCGGACAAAGAAACGTTCTATAATATTCCTTTGTACAAACTAATCTACTT 2340
N T Q S G T E K C S I I L F C T Q N S S F6
I Q K P A Q K K A L Y * L S V H K I L H F5
K N P L R N R Q L I N Y P F M N S * I F F4

F G I G * K V F L H Y * * N F I Q Y S L F1
L E L D K K C F Y I I D K I L F N I P * F2
W N W I K S V F T L L I K F Y S I F L R F3
2341 TTTGGAATTGGATAAAAAGTGTTTTACATTATTGATAAAATTTATTCAATATTCCTTA 2400
----:----|----:----|----:----|----:----|----:----|----:----|
2341 AAACCTTAACCTATTTTTTCAAAAAATGTAATAACTATTTTAAAAATAAGTTATAAGGAAT 2400
N P I P Y F T N K C * Q Y F K I * Y E K F6
I Q F Q I F L T K V N N I F N * E I N R F5
K S N S L F H K * M I S L I K N L I G * F4

E R T D * K S K * I T Y I E * Y L * K I F1
N V Q T E N Q N R * R T S N S T C R R F F2
T Y R L K I K I D N V H R I V P V E D L F3
2401 GAACGTACAGACTGAAAATCAAAATAGATAACGTACATCGAATAGTACCTGTAGAAGATT 2460
----:----|----:----|----:----|----:----|----:----|----:----|
2401 CTTGCATGTCTGACTTTTGTATCTATTGCATGTAGCTTATCATGGACATCTTCTAA 2460
S R V S Q F D F Y I V Y M S Y Y R Y F I F6
L V Y L S F I L I S L T C R I T G T S S F5
F T C V S F * F L Y R V D F L V Q L L N F4

Y S G S S R I R C N S V K C S L L T S H F1
T V V L R V F D A I Q * N V L S * L H I F2
Q W F F A Y S M Q F S K M F S L D F T L F3
2461 TACAGTGGTTCTTCGCGTATTCGATGCAATTCAGTAAAATGTTCTCTCTTGACTTCACAT 2520
----:----|----:----|----:----|----:----|----:----|----:----|
2461 ATGTCACCAAGAAGCGCATAAGCTACGTTAAGTCATTTTACAAGAGAGAACTGAAGTGTA 2520
* L P E E R I R H L E T F H E R K V E C F6
K C H N K A Y E I C N L L I N E R S K V F5

```

```

V T T R R T N S A I * Y F T R E Q S * M F4

Y T I V V F P T D A S Q L Y L L V F L R F1
T Q * * C F L Q T L L S Y I F L S S F D F2
H N S S V S Y R R F S V I S S C L P S T F3
2521 TACACAATAGTAGTGTTCCTACAGACGCTTCTCAGTTATATCTTCTTGTCTTCCTTCGA 2580
----:----|----:----|----:----|----:----|----:----|----:----|
2521 ATGTGTTATCATCACAAAGGATGTCTGCCAAGAGTCAATATAGAAGAACAGAAGGAAGCT 2580
* V I T T N G V S A E * N Y R R T K R R F6
N C L L L T E * L R K E T I D E Q R G E F5
V C Y Y H K R C V S R L * I K K D E K S F4

Q R L L K Q R E T F G N E T S S C Q I S F1
N D C * N N E K L L A T K P V P A K F P F2
T T A E T T R N F W Q R N Q F L P N F R F3
2581 CAACGACTGCTGAAACAACGAGAAACTTTTGGCAACGAAACCAGTTCTGCCAAATTTCC 2640
----:----|----:----|----:----|----:----|----:----|----:----|
2581 GTTGCTGACGACTTTGTTGCTCTTTGAAAACCGTTGCTTTGGTCAAGGACGGTTAAAGG 2640
C R S S F C R S V K P L S V L E Q W I E F6
V V V A S V V L F K Q C R F W N R G F K F5
L S Q Q F L S F S K A V F G T G A L N G F4

A L H Q S Y G G K T G T Y V I D S G I S F1
P S T N R M E E K P V P T S S T V E * A F2
P P P I V W R K N R Y L R H R Q W N K L F3
2641 GCCCTCCACCAATCGTATGGAGGAAAAACCGGTACCTACGTCATCGACAGTGAATAAGC 2700
----:----|----:----|----:----|----:----|----:----|----:----|
2641 CGGGAGGTGGTTAGCATACCTCCTTTTTGGCCATGGATGCAGTAGCTGTCACCTTATTCG 2700
A R W W D Y P P F V P V * T M S L P I L F6
R G G G I T H L F F R Y R R * R C H F L F5
G E V L R I S S F G T G V D D V T S Y A F4

L I Y T Y S F Q K * W A K K G V L * * R F1
* F T H T P F R N N G L R R E C Y D D G F2
N L H I L L S E I M G * E G S A M M T E F3
2701 TTAATTTACATACTCCTTTCAGAAATAATGGGCTAAGAAGGGAGTGCTATGATGACGG 2760
----:----|----:----|----:----|----:----|----:----|----:----|
2701 AATTAAATGTGTATGAGGAAAGTCTTTATTACCCGATTCTTCCCTCACGATACTACTGCC 2760
K I * V Y E K * F Y H A L F P T S H H R F6
S L K C M S R E S I I P * S P L A I I V F5
* N V C V G K L F L P S L L S H * S S P F4

R T F1
E R F2
N X F3
2761 AGAACG 2766
----:-
2761 TCTTGC 2766
L V F6
S F F5
S R F4

```

```

#####
Minimum size of ORFs : 1

```

```

Total ORFs in frame 1 : 53
Total ORFs in frame 2 : 64
Total ORFs in frame 3 : 27
Total ORFs in frame 4 : 55
Total ORFs in frame 5 : 44
Total ORFs in frame 6 : 68

```

```

Total ORFs : 311

```

#####

Sequência proposta

MLFRFQSSMICVRKWWSSVVAKKRYFWSNDNGPGSNCFWQDVLLLTGGRGIQ
 SELTLKWSCERYIRFLVACLFSLQTFIFVVLYSTLQNNRAEMFRVVVEVVKFTAF
 VVAGCKLLLLKLHRKSITSIRSYINDARMITIDDSYDLELSKFKHRSMTMIRGIYGLI
 VIDTVLLFIPNDTTEMAFDRRSDTLPFMSLERNIYRLLFLTLLPVGFLPKFFSSMATI
 GSLLLGMEANFKILANRFRSILSQPFVINGTDWEKMNSELRDTLKQHLEFWSHFK
 ALKKLVGETFFLVHYFSIMSIGALCYICQAIGVNFLSFVVMATLAMFLMEYYMFCH
 FVDSFQDVANCIGEHIFQIAILMPHNRKNHSHYIGFRTALMIIWMNTRREVSMDCM
 GLFNISTAFLHVLNIAYTVLTFLIQMS

Comparação com o parálogo

AAEL017219	MLFRFQSSMICVRKWWSSVVAKKRYFWSNDNGPGSNCFWQDVLLLTGGRGIQSELTLKWS 60
AALF000712	MIDLFRSIWSRVKWS---TVSYFWSHNDPAD-CFLMDTMLLLAG--VQSGLNQKRT 53
	*: *:* **** * . *****.*.*. **: *.:** .* :** *. * :
AAEL017219	CERYIRFLVACLFSLQTFIFVVLYSTLQNNRAEMFRVVVEVVKFTAFVVAGCKLLLLKL 120
AALF000712	STLIVRYLSNGLFAFISIILILQVWSDTQEGWSDVTLKINSLLKITIVMSCLKVLMINH 113
	. :*: * **:: : : : : : : * *.. : : : : : :*: **: :
AAEL017219	HRKSITSIRSYINDARMITIDDSYDLELSKFKHRSMTMIRGIYGLIVIDTVLLFIPNDT 180
AALF000712	FREPAQNLKNFILSQVTSGFKSTDDLDRKFNSIALRMLRAMFALLIADLILLSFPNQT 173
	.*: .:.* * : : . * * *: **: : : *:*:.*: * :** :*: *
AAEL017219	TEMAFDRRSDTLPFMS-LERNIYRLLFLTLLPVGFLPKFFSSMATIGSLLLGMEANFKIL 239
AALF000712	MDNAFRMPPLLQHMKGFSRLILCSFFVCALPLVFVPRYFTHLSCMGILIMGMRMKLQML 233
	: ** * .*. . . * :*: **: *:*:*: : : * *:*: . :*: *
AAEL017219	ANRFRSILSQPFVINGTDWEKMNSELRDTLKQHLEFWSHFKALKKLVGETFFLVHYFSIM 299
AALF000712	AHRFRRITHLPVNAKYNLEQINDELREILNQHMEYMGCFKMLANMVGTAFLLTHTYSSL 293
	*:*** * *.* : *:*.***: *:***: . ** * :*: :*:.**: :
AAEL017219	SIGALCYICQAIGVNFLSFVVMATLAMFLMEYYMFCHFVDSFQDVANCIGEHIFQIAILM 359
AALF000712	NIGALVYMWHTMGFNGFSMVYLLTFIALLFEYYIWCRLVDSFQEVADSIGQLVYEICVYM 353
	.**** *: :*:.* :*: * : :*:***:*.***:*.***: :*:.*: *
AAEL017219	PHNRKNHSHYIGFRTALMIIWMNTRREVSMDCMGLFNISTAFLHVLNIAYTVLTFLIQM 419
AALF000712	PHNADHHSEYHQLRTSLIIWINNAKAYTVSCYGLLKVSTLAFVDMVDITYTVLMFLINM 413
	*** .:*** :*:***:*. : : :.* **:*** **::*:*** **:*
AAEL017219	S-- 420
AALF000712	IDH 416

AaGPROr14 – OR14

>AAEL008442

MNYFELVEPEAVMPLALRLLETYGLRGEKRFLLFQVMIFWEMLMIVIPKIFLGYR
 SQDLVIRGLSELLFQLHIMIRICIFAWHRFKFEGLVAIIRRVYKKICSSEGDSTMKAD
 LLNFMVMINKQCKGYFLYIMGCVSLFTVAPLFQSLGKFITNRRGNATENVEYITMM
 EQEFYGLDIRGNFPHYLYVALGSIAFFTSASFFAVTGVIMNCGVRYTMPMFRLIV
 VRLNKLHELPEQHIREEELRETINLHVDALRCVKILEKIANVAMVIQIVDCVLIWISMIL
 YMRNNLSVDAISLMVLFVVLGTGETYAL

Sequência curta no Vectorbase

```

A L F Y D R Q T A G S * N A V F I S V *      F1
  H F F M I D K L L V A R M Q Y S L A F S      F2
    T F L * * T N C W * L E C S I H * R L A      F3
1  GCACCTTTTATGATAGACAACTGCTGGTAGCTAGAATGCAGTATTCATTAGCGTTTAG 60
   ----:----|----:----|----:----|----:----|----:----|----:----|
1  CGTGAAAAATACTATCTGTTGACGACCATCGATCTTACGTCATAAGTAATCGCAAATC 60
   A S K * S L C V A P L * F A T N M L T *      F6
   X V K K H Y V F Q Q Y S S H L I * * R K      F5
     C K K I I S L S S T A L I C Y E N A N L      F4

      H * Y * G F N S L P S V * Y S P N I G K      F1
      I N I K V S I H C H Q C N I P Q I L A S      F2
      L I L R F Q F I A I S V I F P K Y W Q V      F3
61  CATTAATATTAAGGTTTCAATTCATTGCCATCAGTGTAAATATTCCTCAAATATGGCAAG 120
   ----:----|----:----|----:----|----:----|----:----|----:----|
61  GTAATTATAATTCCAAAGTTAAGTAACGGTAGTCACATTATAAGGGGTTTATAACCGTTC 120
   C * Y * P K L E N G D T Y Y E G F I P L      F6
   A N I N L N * N M A M L T I N G L Y Q C      F5
     M L I L T E I * Q W * H L I G W I N A L      F4

      C A I N G W C A I I G N L P Y L L N V L      F1
      V Q * M G G V Q * L A I Y P I Y * M F Y      F2
      C N K W V V C N N W Q F T L F I E C F I      F3
121 TGTGCAATAAATGGGTGGTGTGCAATAATGGCAATTTACCTATTTATTGAATGTTTTA 180
   ----:----|----:----|----:----|----:----|----:----|----:----|
121 ACACGTTATTTACCCACCACGTTATTAACCGTTAAATGGGATAAATAACTTACAAAAT 180
   H A I F P H H A I I P L K G * K N F T K      F6
   T H L L H T T H L L I Q C N V R N I S H K      F5
     T C Y I P P T C Y N A I * G I * Q I N *      F4

      S Y I S L Y * V S * S D I * F R I F L Q      F1
      L I * A Y T E L V D I F D L E F S C N      F2
      L Y K L I L S * L I * Y L I * N F P A I      F3
181 TCCTATATAAGCTTATACTGAGTTAGTTGATCTGATATTTGATTTAGAATTTTCCTGCAA 240
   ----:----|----:----|----:----|----:----|----:----|----:----|
181 AGAATATATTCGAATATGACTCAATCAACTAGACTATAAACTAAATCTTAAAGGACGTT 240
   D * I L K Y Q T L Q D S I Q N L I K R C      F6
   I K Y L S I S L * N I Q Y K I * F K G A      F5
     R I Y A * V S N T S R I N S K S N E Q L      F4

      F K I T Y R I Y F S K C L I I E A K D G      F1
      L K S P I E F I L V N A * L L K L K M A      F2
      * N H L * N L F * * M L N Y * S * R W L      F3
241 TTTAAATCACCTATAGAATTTATTTTAGTAAATGCTTAATTATTGAAGCTAAAGATGGC 300
   ----:----|----:----|----:----|----:----|----:----|----:----|
241 AAATTTTAGTGATATCTTAAATAAAATCATTTACGAATTAATAACTTCGATTTCTACCG 300
   N L I V * L I * K L L H K I I S A L S P      F6

```

```

I * F * R Y F K N * Y I S L * Q L * L H      F5
K F D G I S N I K T F A * N N F S F I A      F4

F V R S V Q M S S I I V P N N I V R N Q      F1
S L E A S K C H R L S C L I T L F E I N      F2
R * K R P N V I D Y R A * * H C S K S T      F3
301 TTCGTTAGAACGCTCCAAATGTCATCGATTATCGTGCCTAATAACATTGTTTCGAAATCAA 360
----:----|----:----|----:----|----:----|----:----|----:----|
301 AAGCAATCTTCGCAGGTTTACAGTAGCTAATAGCACGGATTATTGTAACAAGCTTTAGTT 360
K T L L T W I D D I I T G L L M T R F *      F6
S R * F R G F T M S * R A * Y C Q E F D      F5
E N S A D L H * R N D H R I V N N S I L      F4

R H N V L S N L L G K N K L F N S I H E      F1
V T T Y Y Q I C L E R I N F S I P Y T N      F2
S Q R I I K F A W K E * T F Q F H T R T      F3
361 CGTCACAACGTATTATCAAATTTGCTTGGAAAGAATAAACTTTTCAATTCACATACGAA 420
----:----|----:----|----:----|----:----|----:----|----:----|
361 GCAGTGTTCATAATAGTTTAAACGAACCTTCTTATTTGAAAAGTTAAGGTATGTGCTT 420
R * L T N D F K S P F F L S K L E M C S      F6
V D C R I I L N A Q F S Y V K * N W V R      F5
T V V Y * * I Q K S L I F K E I G Y V F      F4

Q S S K S M T F K I Q S S L S G F K F F      F1
N H Q K A * L S K F N Q A C Q D S N F S      F2
I I K K H D F Q N S I K L V R I Q I F P      F3
421 CAATCATCAAAAAGCATGACTTTCAAATTCATCAAGCTTGTGAGGATTCAAATTTTTC 480
----:----|----:----|----:----|----:----|----:----|----:----|
421 GTTAGTAGTTTTTCGTACTGAAAGTTTTAAGTTAGTTTGAACAGTCCTAAGTTTAAAAAG 480
C D D F L M V K L I * D L K D P N L N K      F6
V I M L F C S K * F E I L S T L I * I K      F5
L * * F A H S E F N L * A Q * S E F K E      F4

H S K P L H E W T F L L C C T S * F C E      F1
I Q S H Y M N G H F Y Y V A H H N F V K      F2
F K A I T * M D I F I M L H I I I L * N      F3
481 CATTCAAAGCCATTACATGAATGGACATTTTATTATGTTGCACATCATAATTTTGTGAA 540
----:----|----:----|----:----|----:----|----:----|----:----|
481 GTAAGTTTCGGTAATGTACTTACCTGTAAAAATAACAACGTGTAGTATTAAACACTT 540
W E F G N C S H V N K N H Q V D Y N Q S      F6
G N L A M V H I S M K I I N C M M I K H      F5
M * L W * M F P C K * * T A C * L K T F      F4

M A I A I L S R I K A H H R V I L A L S      F1
W Q S L S Y Q E * K P I T V * S L R * V      F2
G N R Y P I K N K S P S P C N P C A E C      F3
541 ATGGCAATCGCTATCCTATCAAGAATAAAAGCCCATCACCGTGAATCCTTGCGCTGAGT 600
----:----|----:----|----:----|----:----|----:----|----:----|
541 TACCGTTAGCGATAGGATAGTTCTTATTTTCGGGTAGTGGCACATTAGGAACGCGACTCA 600
I A I A I R D L I F A W * R T I R A S L      F6
F P L R * G I L F L L G D G H L G Q A S      F5
H C D S D * * S Y F G M V T Y D K R Q T      F4

V V L S T M N Y F E L V E P E A V M P L      F1
L S F Q L * T T L S * L N R K R * C R W      F2
C P F N Y E L L * A S * T G S G D A V G      F3
601 GTTGTCCTTTCAACTATGAAGTACTTTGAGCTAGTTGAACCGAAGCGGTGATGCCGTTG 660
----:----|----:----|----:----|----:----|----:----|----:----|
601 CAACAGGAAAGTTGATACTTGATGAAACTCGATCAACTTGGCCTTCGCCACTACGGCAAC 660
T T R E V I F * K S S T S G S A T I G N      F6
H Q G K L * S S S Q A L Q V P L P S A T      F5
N D K * S H V V K L * N F R F R H H R Q      F4

```

A L R L L E T Y G L R G E K R K F L L F F1
 L S V C W R P M D Y A A K S G N F C C F F2
 S P F A G D L W T T R R K A E I S A V S F3
 661 GCTCTCCGTTTCTGGAGACCTATGGACTACGCGCGAAAAAGCGGAAATTTCTGCTGTTT 720
 ----:----|----:----|----:----|----:----|----:----|----:----|
 661 CGAGAGGCAAACGACCTCTGGATACCTGATGCGCCGCTTTTCGCCTTTAAAGACGACAAA 720
 A R R K S S V * P S R P S F R F N R S N F6
 P E G N A P S R H V V R R F A S I E A T F5
 S E T Q Q L G I S * A A F L P F K Q Q K F4

Q V M I F W E M L M I V I P K I F L G Y F1
 K * * F F G K C * * * * F P R F S S V I F2
 S N D F L G N V D D S D S D S F P R L S F3
 721 CAAGTAATGATTTTTTGGGAAATGTTGATGATGATTCCCAAGATTTTCGCGTTAT 780
 ----:----|----:----|----:----|----:----|----:----|----:----|
 721 GTTCATTACTAAAAACCTTTACAACCTACTATCACTAAGGGTTCTAAAAGGAGCCAATA 780
 * T I I K Q S I N I I T I G L I K R P * F6
 E L L S K K P F T S S L S E W S K G R N F5
 L Y H N K P F H Q H Y H N G L N E E T I F4

R S Q D L V I R G L S E L L F Q L H I M F1
 A V R I W * F V A Y R S C C S S C T S * F2
 Q S G F G D S W P I G V A V P V A H H D F3
 781 CGCAGTCAGGATTTGGTGATTCGTGGCCTATCGGAGTTGCTGTTCCAGTTGCACATCATG 840
 ----:----|----:----|----:----|----:----|----:----|----:----|
 781 GCGTCAGTCCTAAACCACTAAGCACCGGATAGCCTCAACGACAAGGTCAACGTGTAGTAC 840
 R L * S K T I R P R D S N S N W N C M M F6
 D C D P N P S E H G I P T A T G T A C * F5
 A T L I Q H N T A * R L Q Q E L Q V D H F4

I R I C I F A W H R F K F E G L V A I I F1
 F G Y V Y L H G T V L S S R D W L P L Y F2
 S D M Y I C M A P F * V R G I G C H Y T F3
 841 ATTCGATATGTATATTTGCATGGCACCCTTTAAGTTCGAGGGATTGGTTGCCATTATA 900
 ----:----|----:----|----:----|----:----|----:----|----:----|
 841 TAAGCCTATACATATAAACGTACCGTGGCAAAATTCAAGCTCCCTAACCAACGGTAATAT 900
 I R I H I N A H C R K L N S P N T A M I F6
 S E S I Y I Q M A G N * T R P I P Q W * F5
 N P Y T Y K C P V T K L E L S Q N G N Y F4

R R V Y K K S D R H E V C F C E W Q F L F1
 G E Y T R K V I D M K C V F V N G S S Y F2
 E S I Q E K * * T * S V F L * M A V L I F3
 901 CGGAGATATACAAGAAAAAGTGATAGACATGAAGTGTGTTTTGTGAATGGCAGTTCTTA 960
 ----:----|----:----|----:----|----:----|----:----|----:----|
 901 GCCTCTCATATGTTCTTTTCACTATCTGTACTTCACACAAAAACACTTACCGTCAAGAAT 960
 R L T Y L F L S L C S T H K Q S H C N K F6
 V S L I C S F H Y V H L T N K H I A T R F5
 P S Y V L F T I S M F H T K T F P L E * F4

F G Y * L L V C S S E G D S T M K A D L F1
 L V I N F * F V H L K E I Q Q * R P I Y F2
 W L L T S S L F I * R R F N N E G R F T F3
 961 TTTGGTTATTAACCTTCTAGTTTGTTCATCTGAAGGAGATTCAACAATGAAGGCCGATTTA 1020
 ----:----|----:----|----:----|----:----|----:----|----:----|
 961 AAACCAATAATTGAAGATCAAACAAGTAGACTTCCTCTAAGTTGTTACTTCCGCTAAAT 1020
 N P * * S R T Q E D S P S E V I F A S K F6
 I Q N N V E L K N M Q L L N L L S P R N F5
 K T I L K * N T * R F S I * C H L G I * F4

```

L N F N V M I N K Q C K G Y F L Y I M G      F1
* T S T * * L I S N A K D T S C I L W V      F2
E L Q R N D * * A M Q R I L P V Y Y G L      F3
1021 CTGAACCTCAACGTAATGATTAATAAGCAATGCAAAGGATACTTCCTGTATATTATGGGT 1080
----:----|----:----|----:----|----:----|----:----|----:----|
1021 GACTTGAAGTTGCATTACTAATTATTCGTTACGTTTCCTATGAAGGACATATAATACCCA 1080
S F K L T I I L L C H L P Y K R Y I I P      F6
V S S * R L S * Y A I C L I S G T Y * P      F5
Q V E V Y H N I L L A F S V E Q I N H T      F4

```

```

C V S L F T V A P L F Q S L G K F I T N      F1
A S A C S R W L H Y F K V W V N S L Q I      F2
R Q L V H G G S I I S K F G * I H Y K S      F3
1081 TGCCTCAGCTTGTTTCACGGTGGCTCCATTATTTCAAAGTTTGGGTAAATTCATTACAAAT 1140
----:----|----:----|----:----|----:----|----:----|----:----|
1081 ACGCAGTCGAACAAGTGCCACCGAGGTAATAAGTTTCAAACCCATTTAAGTAATGTTTA 1140
Q T L K N V T A G N N * L K P L N M V F      F6
N R * S T * P P E M I E F N P Y I * * L      F5
A D A Q E R H S W * K L T Q T F E N C I      F4

```

```

R R G N A T E N V E Y I T M M E Q E * A      F1
E E E T Q Q R M * N I * Q * W N K S K L      F2
K R K R N R E C R I Y N N D G T R V S F      F3
1141 CGAAGAGGAAACGCAACAGAGAATGTAGAATATATAACAATGATGGAACAAGATAAGCT 1200
----:----|----:----|----:----|----:----|----:----|----:----|
1141 GCTTCTCCTTTGCGTTGTCTCTTACATCTTATATATTGTTACTACCTTGTTCCTCATTCGA 1200
R L P F A V S F T S Y I V I I S C S Y A      F6
D F L F R L L S H L I Y L L S P V L T L      F5
S S S V C C L I Y F I Y C H H F L L L S      F4

```

```

L N V L * K K I V F H N L I I L F K I L      F1
* M F Y R K K L F S I I * L F C L R F Y      F2
K C F I E K N C F P * S D Y F V * D S T      F3
1201 TTAATGTTTATAGAAAAAATGTTTTCCATAATCTGATTATTTTGTTTAAGATTCTA 1260
----:----|----:----|----:----|----:----|----:----|----:----|
1201 AATTTACAAAATATCTTTTTTAAACAAAAGGTATTAGACTAATAAAACAAATCTAAGAT 1260
K F T K Y F F I T K W L R I I K N L I R      F6
K L H K I S F F Q K G Y D S * K T * S E      F5
* I N * L F F N N E M I Q N N Q K L N *      F4

```

```

R F G Y P W K F Q S L F D L C R T R K H      F1
G L D I R G N F S H Y L I Y V A L G S I      F2
V W I S V E I S V T I * F M S H * E A L      F3
1261 CGGTTTGGATATCCGTGGAATTTCACTACTATTTGATTTATGTGCGACTAGGAAGCAT 1320
----:----|----:----|----:----|----:----|----:----|----:----|
1261 GCCAAACCTATAGGCACCTTTAAAGTCAGTGATAAACTAAATACAGCGTGATCCTTCGTA 1320
R N P Y G H F N * D S N S K H R V L F C      F6
V T Q I D T S I E T V I Q N I D C * S A      F5
P K S I R P F K L * * K I * T A S P L M      F4

```

```

C L F H F G I I L C C D R C D N E L W S      F1
A F F T S A S F F A V T G V I M N C G V      F2
P F S L R H H S L L * P V * * * I V E S      F3
1321 TGCCTTTTTCACTTCGGCATCATCTTTGCTGTGACCGGTGTGATAATGAATTGTGGAGT 1380
----:----|----:----|----:----|----:----|----:----|----:----|
1321 ACGGAAAAAGTGAAGCCGTAGTAAGAAACGACACTGGCCACACTATTACTTAACACCTCA 1380
Q R K * K P M M R Q Q S R H S L S N H L      F6
N G K E S R C * E K S H G T H Y H I T S      F5
A K K V E A D N K A T V P T I I F Q P T      F4

```

```

P I Y D A N V P I D S C * T E * A S * T      F1
R Y T M P M F R L I V V R L N K L H E L      F2

```

D I R C Q C S D * * L L D * I S F M N F F3
 1381 CCGATATACGATGCCAATGTTCCGATTGATAGTTGTTAGACTGAATAAGCTTCATGAAC 1440
 -----|-----|-----|-----|-----|-----|-----|
 1381 GGCTATATGCTACGGTTACAAGGCTAATCAACAATCTGACTTATTCGAAGTACTTGA 1440
 G I Y S A L T G I S L Q * V S Y A E H V F6
 D S I R H W H E S Q Y N N S Q I L K M F F5
 R Y V I G I N R N I T T L S F L S * S S F4

 S R A A Y S * R T S G N H * F A C G C I F1
 P E Q H I R E E L R E T I N L H V D A L F2
 Q S S I F V K N F G K P L I C M W M H * F3
 1441 TCCAGAGCAGCATATTCGTGAAGAACTTCGGGAAACCATTAAATTTGCATGTGGATGCATT 1500
 ----:----|----:----|----:----|----:----|----:----|----:----|
 1441 AGGTCTCGTCGTATAAGCACTTCTTGAAGCCCTTTGGTAATTAACGTACACCTACGTAA 1500
 E L A A Y E H L V E P F W * N A H P H M F6
 K W L L M N T F F K P F G N I Q M H I C F5
 G S C C I R S S R S V M L K C T S A N F4

 K V A I T V H T I I G * V R F F F T F R F1
 R * L * Q F I Q * L D R L D F F L L G F2
 G S Y N S S Y N N W I G * I F F Y L * V F3
 1501 AAGTAGCTATAACAGTTCATACAATAATTGGATAGGTTAGATTTTTTTTTTACTTTT 1560
 ----:----|----:----|----:----|----:----|----:----|----:----|
 1501 TTCCATCGATATTGTCAAGTATGTTATTAACCTATCCAATCTAAAAAAAATGAAAATCC 1560
 L T A I V T * V I I P Y T L N K K V K L F6
 * P L * L L E Y L L Q I P * I K K * K * F5
 L Y S Y C N M C Y N S L N S K K K S K P F4

 C V K I L E K I A N V A M V I Q I V D C F1
 A S R Y W K K S P M W P W S F K S W I A F2
 R Q D I G K N R Q C G H G H S N R G L R F3
 1561 TCGCTCAAGATATTGGAATAATCGCCAATGTGGCCATGGTCATTCAAATCGTGGATTGC 1620
 ----:----|----:----|----:----|----:----|----:----|----:----|
 1561 ACGCAGTTCTATAACCTTTTTTAGCGGTTACACCGGTACCAGTAAGTTTAGCACCTAAC 1620
 H T L I N S F I A L T A M T M * I T S Q F6
 T R * S I P F F R W H P W P * E F R P N F5
 A D L Y Q F F D G I H G H D N L D H I A F4

 V L I W I S M I L Y M R N V S F Q S V R F1
 Y S Y G F P * F C I * E M * V S R A L E F2
 T H M D F H D F V Y E K C K F P E R * N F3
 1621 GTACTCATATGGATTTCCATGATTTTGTATATGAGAAATGTAAGTTTCCAGAGCGTTAGA 1680
 ----:----|----:----|----:----|----:----|----:----|----:----|
 1621 CATGAGTATACCTAAAGGTACTAAAACATATACTCTTTACATTCAAAGGTCTCGCAATCT 1680
 T S M H I E M I K Y I L F T L K W L T L F6
 R V * I S K W S K T Y S F H L N G S R * F5
 Y E Y P N G H N Q I H S I Y T E L A N S F4

 I I I G I A N * F I P H Q N L S V D A I F1
 L L L V L Q I D L F L T R I L V L M P L F2
 Y Y W Y C K L I Y S S P E S * C * C H * F3
 1681 ATTATTATGGTATTGCAAATTGATTTATTCCTCACCAGAACTCTTAGTGTTCATGCCATT 1740
 ----:----|----:----|----:----|----:----|----:----|----:----|
 1681 TAATAATAACCATAACGTTTAACTAAATAAGGAGTGGTCTTAGAATCACAACCTACGGTAA 1740
 I I I P I A F Q N I G * W F R L T S A M F6
 F * * Q Y Q L N I * E E G S D * H Q H W F5
 N N N T N C I S K N R V L I K T N I G N F4

 S L M V L F V V L T G E T Y A L * D L L F1
 V * W F Y S L F * L G K L M L F E I Y * F2
 S N G S I R C S N W G N L C S L R F I D F3
 1741 AGTCTAATGGTCTATTTCGTTGTTCTAACTGGGGAACTTATGCTCTTTGAGATTTATTG 1800

```

-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1741 TCAGATTACCAAGATAAGCAACAAGATTGACCCCTTTGAATACGAGAAACTCTAAATAAC 1800
    L R I T R N T T R V P S V * A R Q S K N F6
    * D L P E I R Q E L Q P F K H E K L N I F5
    T * H N * E N N * S P F S I S K S I * Q F4

    T Q L T N E V N F S H H F Y Q M S K H T F1
    L S S L M R * I F P I T S I K * A N I Q F2
    S A H * * G K F F P S L L S N E Q T Y K F3
1801 ACTCAGCTCACTAATGAGGTAAATTTTTCCCATCACTTCTATCAAATGAGCAAACATACA 1860
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1801 TGAGTCGAGTGATTACTCCATTTAAAAAGGGTAGTGAAGATAGTTTACTCGTTTGTATGT 1860
    V * S V L S T F K E W * K * * I L L C V F6
    S E A * * H P L N K G D S R D F S C V Y F5
    S L E S I L Y I K G M V E I L H A F M C F4

    S I V L I F N R V * Q * H E P S L T A S F1
    V S F * Y S T E F S S D T S H H * L P V F2
    Y R F D I Q Q S L A V T R A I I D C Q W F3
1861 AGTATCGTTTTGATATTCAACAGAGTTTAGCAGTGACACGAGCCATCATTGACTGCCAGT 1920
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1861 TCATAGCAAACCTATAAGTTGTCTCAAATCGTCACTGTGCTCGGTAGTAACTGACGGTCA 1920
    L I T K I N L L T * C H C S G D N V A L F6
    L Y R K S I * C L K A T V R A M M S Q W F5
    T D N Q Y E V S N L L S V L W * Q S G T F4

    G T A C H W I Y R N H C R L S Y L E P N F1
    V Q L A I G Y T E I I V V C L I * S P T F2
    Y S L P L D I Q K S L S F V L F R A Q R F3
1921 GGTACAGCTTGCCATTGGATATACAGAAATCATTGTCGTTTGTCTTATTTAGAGCCCAAC 1980
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1921 CCATGTGCAACGGTAACCTATATGTCTTTAGTAACAGCAAACAGAATAAATCTCGGGTTG 1980
    P V A Q W Q I Y L F * Q R K D * K S G L F6
    H Y L K G N S I C F D N D N T K N L A W F5
    T C S A M P Y V S I M T T Q R I * L G V F4

    A K K V S Q R Q S F S S W M L N V L E G F1
    Q R R Y H S G K V F L H G C * T F W K G F2
    K E G I T A A K F F F M D V E R F G R V F3
1981 GCAAAGAAGGTATCACAGCGCAAAGTTTTCTTCATGGATGTTGAACGTTTGGGAAGGG 2040
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1981 CGTTTCTTCCATAGTGTGCGCGTTTCAAAAAGAAGTACCTACAACCTTGCAAAACCTTCCC 2040
    A F F T D C R C L K E E H I N F T K S P F6
    R L S P I V A A F N K K M S T S R K P L F5
    C L L Y * L P L T K R * P H Q V N Q F P F4

    W R R L H I R S T W C L R I T C K I S C F1
    G A D F I F D L R G A * G L L V K * V V F2
    A Q T S Y S I Y V V L K D Y L * N K L * F3
2041 TGGCGCAGACTTCATATTCGATCTACGTGGTGCTTAAGGATTACTTGTAATAAAGTTGT 2100
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
2041 ACCGCGTCTGAAGTATAAGCTAGATGCACCACGAATTCCTAATGAACATTTTATTCAACA 2100
    H R L S * I R D V H H K L I V Q L I L Q F6
    T A C V E Y E I * T T S L S * K Y F L N F5
    P A S K M N S R R P A * P N S T F Y T T F4

    D L V S F L D * Y C S F H M S A C T S K F1
    I * S V F L I N T A L F T * V L A H Q K F2
    S S Q F S * L I L L F S H E C L H I K K F3
2101 GATCTAGTCAGTTTTCTTGATTAATACTGCTCTTTTCACATGAGTGCTTGCACATCAAAA 2160
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
2101 CTAGATCAGTCAAAAGAACTAATTATGACGAGAAAAGTGTACTCACGAACGTGTAGTTTT 2160

```

```

S R T L K R S * Y Q E K * M L A Q V D F F6
H D L * N E Q N I S S K E C S H K C M L F5
I * D T K K I L V A R K V H T S A C * F F4

K I Q K N C L N F I F F G E K K I F S * F1
K Y K K I V * T S S F L V R K K Y L V K F2
N T K K L F K L H L F W * E K N I * L K F3
2161 AAAATACAAAAAATTGTTTAACTTCATCTTTTTTGGTGAGAAAAAATATTTAGTTAA 2220
-----|-----|-----|-----|-----|-----|-----|
2161 TTTTATGTTTAAAAACAAATTTGAAGTAGAAAAAACCACTCTTTTTTTATAAATCAATT 2220
F I C F F Q K F K M K K P S F F I N L * F6
F F V F F N N L S * R K Q H S F F I * N F5
F Y L F I T * V E D K K T L F F Y K T L F4

K K K Y P R R A F V K * Y T A I V F V L F1
K K S I R E E H S S N D I L P L Y L F * F2
K K V S A K S I R Q M I Y C H C I C F E F3
2221 AAAAAAAGTATCCGCGAAGAGCATTTCGTCAAATGATATACTGCCATTGTATTGTTTTG 2280
----:----|----:----|----:----|----:----|----:----|----:----|
2221 TTTTTTTTCATAGGCGCTTCTCGTAAGCAGTTTACTATATGACGGTAACATAAAACAAAC 2280
F F F Y G R L A N T L H Y V A M T N T K F6
F F F T D A F L M R * I I Y Q W Q I Q K F5
F F L I R S S C E D F S I S G N Y K N Q F4

R I V P Q T L F G I * V S K E * I D L T F1
E L F H K R C L V F K S P K S E L I * L F2
N C S T N V V W Y L S L Q R V N * F D W F3
2281 AGAATTGTTCACAAACGTTGTTTGGTATTTAAGTCTCCAAAGAGTGAATTGATTTGACT 2340
----:----|----:----|----:----|----:----|----:----|----:----|
2281 TCTTAACAAGGTGTTTGCAACAAACCATAAATTCAGAGGTTTCTCACTTAACATAAAGTGA 2340
L I T G C V N N P I * T E L S H I S K V F6
S F Q E V F T T Q Y K L R W L T F Q N S F5
S N N W L R Q K T N L D G F L S N I Q S F4

G * Q T E L G Q * A H F R Y K T I E * L F1
V D K Q N * A S K L I S V I K Q L S D W F2
L T N R T R P V S S F P L * N N * V I G F3
2341 GGTTGACAAACAGAACTAGGCCAGTAAGCTCATTTCCGTTATAAAACAATTGAGTGATTG 2400
----:----|----:----|----:----|----:----|----:----|----:----|
2341 CCAACTGTTTGTCTTGATCCGGTCATTCGAGTAAAGGCAATATTTTGTAACTCACTAAC 2400
P Q C V S S P W Y A * K R * L V I S H N F6
Q N V F L V L G T L E N G N Y F L Q T I F5
T S L C F * A L L S M E T I F C N L S Q F4

E V V L L L G V L E N T N L K H R R T T F1
R W F Y C L V Y L K I Q T * N T E E R H F2
G G S I A W C T * K Y K P K T Q K N D I F3
2401 GAGGTGGTTCTATTGCTTGGTGTACTTGAAAAATACAAACCTAAAACACAGAAGAACGACA 2460
----:----|----:----|----:----|----:----|----:----|----:----|
2401 CTCCACCAAGATAACGAACCACATGAACTTTATGTTTGGATTTTGTGTCTTCTGTGCTGT 2460
S T T R N S P T S S F V F R F C L L V V F6
P P P E I A Q H V Q F Y L G L V C F F S F5
L H N * Q K T Y K F I C V * F V S S R C F4

L N T F D P A F Y Y L P S S R Q K V H C F1
* I H L I Q L S T I Y P V V A K K Y I A F2
E Y I * S S F L L S T Q * S P K S T L H F3
2461 TTGAATACATTTGATCCAGCTTTCTACTATCTACCCAGTAGTCGCCAAAAAGTACATTGC 2520
----:----|----:----|----:----|----:----|----:----|----:----|
2461 AACTTATGTAACTAGGTCGAAAGATGATAGATGGGTCATCAGCGTTTTTTCATGTAACG 2520
N F V N S G A K * * R G L L R W F T C Q F6
M S Y M Q D L K R S D V W Y D G F L V N F5

```

```

Q I C K I W S E V I * G T T A L F Y M A F4

T S S H R P V I N E L E I * P Q P A H L F1
L P P I G Q * * M S S K F D R N Q L I W F2
F L P S A S D K * A R N L T A T S S F G F3
2521 ACTTCCTCCCATCGGCCAGTGATAAATGAGCTCGAAATTTGACCGCAACCAGCTCATTTG 2580
----:----|----:----|----:----|----:----|----:----|----:----|
2521 TGAAGGAGGGTAGCCGGTCACTATTTACTCGAGCTTTAAACTGGCGTTGGTCGAGTAAAC 2580
V E E W R G T I F S S S I Q G C G A * K F6
C K R G D A L S L H A R F K V A V L E N F5
S G G M P W H Y I L E F N S R L W S M Q F4

E R I T P D D R K K Q L P F Q Q K V H T F1
N A S R Q M T G K N N S H S S K K F I R F2
T H H A R * P E K T T P I P A K S S Y V F3
2581 GAACGCATCAGCCAGATGACCGGAAAAACAACCTCCATTCCAGCAAAAAGTTCATACG 2640
----:----|----:----|----:----|----:----|----:----|----:----|
2581 CTTGCGTAGTGGGTCTACTGGCCTTTTTTGTGAGGGTAAGGTCGTTTTTCAAGTATGC 2640
S R M V G S S R F F C S G N W C F T * V F6
P V C * A L H G S F V V G M G A F L E Y F5
F A D R W I V P F F L E W E L L F N M R F4

C K Y V S Q V S L S R R L P P S S K L V F1
V N T F P K F P C R V V Y H H P R S W S F2
* I R F P S F P V A S F T T I L E A G Q F3
2641 TGTAATAACGTTTCCCAAGTTTCCCTGTGCGTCGTTTACCACCATCCTCGAAGCTGGTC 2700
----:----|----:----|----:----|----:----|----:----|----:----|
2641 ACATTTATGCAAAGGGTTCAAAGGGACAGCGCAGCAAATGGTGGTAGGAGCTTCGACCAG 2700
H L Y T E W T E R D R R K G G D E F S T F6
T Y I R K G L K G T A D N V V M R S A P F5
T F V N G L N G Q R T T * W W G R L Q D F4

K R R I A R S P S E F1
N E E L H V P Q A X F2
T K N C T F P K R X F3
2701 AAACGAAGAATTGCACGTTCCCAAGCGAA 2730
----:----|----:----|----:----|
2701 TTTGCTTCTTAACGTGCAAGGGGTTGCGTT 2730
L R L I A R E G L S F6
* V F F Q V N G L R F5
F S S N C T G W A F F4

```

```

#####
Minimum size of ORFs : 1

```

```

Total ORFs in frame 1 : 32
Total ORFs in frame 2 : 52
Total ORFs in frame 3 : 56
Total ORFs in frame 4 : 51
Total ORFs in frame 5 : 54
Total ORFs in frame 6 : 45

```

```

Total ORFs : 290
#####

```

Sequência proposta

MNYFELVEPEAVMPLALRLLETYGLRGEKRFLLFQVMIFWEMLMIVIPKIFLGYR
 SQDLVIRGLSELLFQLHIMIRICIFAWHRFKFEGLVAIIRRVYKKVCSSEGDSTMKA
 DLLNFMINKQCKGYFLYIMGCVSLFTVAPLFQSLGKFITNRRGNATENVEYITM
 MEQEFYGLDIRGNFSHYLIYVALGSIAFFTSASFFAVTGVIMNCGVRYTMPMFRLI
 VVRLNKLHELPEQHIREELRETINLHVDALRCVKILEKIANVAMVIQIVDCVLIWISMI
 LYMRNLSVDAISLMVLFVV**GKFFPSLLSNEQTYKYRFDIQQSLAVTRAIDCQWY**
SLPLDIQKSLSFVLFRAQRKEGITAAKFFFMDVERFGRVAQTSYSIYVVLKDYL

Comparação com o parálogo

AAEL008442	MNYFELVEPEAVMPLALRLLETYGLRGEKRFLLFQVMIFWEMLMIVIPKIFLGYRSQDL 60
AGAP009408	MLFQELPNDRAVLPLLLYIQERLGFWDRTTRVYPVAFVAFWITVAIPKLATDYSDEL 60 * : ** : .**:* * : * * : **: . : : * :. : :.***: .* . : *
AAEL008442	VIRGLSELLFQLHIMIRICIFAWHRFK----FEGIVAIIRRVYKKVCSSEGDSTMKADL 115
AGAP009408	FICSMAEALAFVGNVFCGG-MALWSVYNSFRQFIEQVIRLT KYLYRD---HHPLQTVREQL 116 .* .:*** * ::: : * : : : * : : : :*: . . .*: : *
AAEL008442	LNFMINKQCKGYFLYIMGCVSLFTVAPLFQSLGKFITNRRGN-----A 160
AGAP009408	LQFNHRIHRYTNRYCFSMILIVFYLVAPVITSFGVYFQSVWQSHHDTNDTIRIWEISQV 176 *:** *: : : * : :* : : : ***: *:* : : . . .
AAEL008442	TENVEYITMMEQEFYGLDIRGNFSHYLIYVALGSIAFFTSASFFAVTGVIMNCGVRYTMP 220
AGAP009408	APHRQFSLHMEQDFYGLQHRTNLIHYTLIAVIVPMMFVTACTVHMKVLTIASSVRYAEM 236 : : : : ***.***: * * : * :*: : :*: . . . : : .***:
AAEL008442	MFRLIVVRLNKLHELPEQ-HIREELRETINLHVDALRCVKILEKIANVAMVIQIVDCVLI 279
AGAP009408	LLHVVMLKVDNLHRIPNQKSIREELHDIHVVHQRITLNCITLLVQALRPILMVQLVFCVFI 296 : : : : : : :*: .*: * *****: *:* :*.*:.* : . :*: * **:
AAEL008442	WISMILYMRN--NLSVDAISLMVLFVVVGKFFPSLLSNEQTYKYRFDIQQSLAVTRAIDC 337
AGAP009408	WCLMMLFFTIADKFSVAFFNLAILFVITIE----TFSACYFGTRLSTQAVELSKSVYGC 352 * *:*: : : ** :.* :*** .: : . * *: : : : : *
AAEL008442	QWYSLPLDIQKSLSFVLFRAQRKEGITAAKFFFMDVERFGRVAQTSYSIYVVLKDYL 394
AGAP009408	GWPAMDRDIQQGLRMVLHRTQSPVGIQAGKFCFVDVELFQNMVNKSYFFIVLKDAF 409 * : : ***.* :**.* * ** *.** *:*** * .: :.***: :*** :

AaGPROR76 – OR76

>AAEL013418

MSAARVDCLPNSRSFSILGHFRDPFAFQKLLDRCIGFIHWDTESAFTWVKILLLVF
 TGTYYTLSCCLCLTRIDPAEVPLDHFFGMWFLVGGGCSCFSQWYVLAIERRHAK
 VIGFLTNLQQNGIDHPTRVRQRSRIVLYSIHVWTTNVSQTAVWAVTLLFTSSIAHA
 TSNSYL

Sequência curta no *Vectorbase*

```

A L Y L I V I I F Y E K V N K L Y T K P      F1
      P S I S S S * Y F M K R * T N S T Q S L      F2
      P L S H R H N I L * K G E Q T L H K A F      F3
1  GCCCTCTATCTCATCGTCATAATATTTTATGAAAAGGTGAACAACTCTACACAAAGCCT 60
   ----:----|----:----|----:----|----:----|----:----|----:----|
1  CGGGAGATAGAGTAGCAGTATTATAAAATACTTTTCCACTTGTTTGAGATGTGTTTCGGA 60
   A R * R M T M I N * S F T F L S * V F G      F6
X G R D * R * L I K H F P S C V R C L A      F5
   G E I E D D Y Y K I F L H V F E V C L R      F4

      F S Y N F A E D H I S I R R Q D K L L F      F1
      S A T T L P K T T F R L D V R I N C Y S      F2
      Q L Q L C R R P H F D * T S G * I A I H      F3
61  TTCAGCTACAACTTTGCCGAAGACCACATTTCGATTAGACGTCAGGATAAATGCTATTC 120
   ----:----|----:----|----:----|----:----|----:----|----:----|
61  AAGTCGATGTTGAAACGGCTTCTGGTGTAAGCTAATCTGCAGTCCTATTTAACGATAAG 120
   K L * L K A S S W M E I L R * S L N S N      F6
K * S C S Q R L G C K S * V D P Y I A I      F5
   E A V V K G F V V N R N S T L I F Q * E      F4

      I I I K W V C I L H A R Q A T V V L L A      F1
      * S * S G F A F C M L G R Q Q W C S W R      F2
      N H K V G L H F A C S A G N S G A L G E      F3
121  ATAATCATAAAGTGGGTTTGCATTTTGCATGCTCGGCAGGCAACAGTGGTGCTCTTGGCG 180
   ----:----|----:----|----:----|----:----|----:----|----:----|
121  TATTAGTATTTACCCAAACGTAAAACGTACGAGCCGTCCTGTGTCACCACGAGAACCGC 180
   M I M F H T Q M K C A R C A V T T S K A      F6
* L * L T P K C K A H E A P L L P A R P      F5
   Y D Y L P N A N Q M S P L C C H H E Q R      F4

      S R I D Q S N P G Y Y V L Q Q K S S V A      F1
      V E * I K V I Q A T M F Y S K K A V L L      F2
      * N R S K * S R L L C S T A K K Q C C S      F3
181  AGTAGAATAGATCAAAGTAATCCAGGCTACTATGTTCTACAGCAAAAAGCAGTGTGCT 240
   ----:----|----:----|----:----|----:----|----:----|----:----|
181  TCATCTTATCTAGTTTCATTAGGTCCGATGATACAAGATGTCGTTTTTCGTCACAACGA 240
   L L I S * L L G P * * T R C C F L L T A      F6
S Y F L D F Y D L S S H E V A F F C H Q      F5
   T S Y I L T I W A V I N * L L F A T N S      F4

      L K V I D * A S Q H N L D F V I N N S K      F1
      * K * L I E H L S I I * T L L S I I A N      F2
      E S N * L S I S A * F R L C Y Q * * Q I      F3
241  CTGAAAGTAATTGATTGAGCATCTCAGCATAATTTAGACTTTGTTATCAATAATAGCAAA 300
   ----:----|----:----|----:----|----:----|----:----|----:----|
241  GACTTTTCATTAACATACTCGTAGAGTCGTATTAATCTGAAACAATAGTTATTATCGTTT 300
   R F T I S Q A D * C L K S K T I L L L L      F6
E S L L Q N L M E A Y N L S Q * * Y Y C      F5
   Q F Y N I S C R L M I * V K N D I I A F      F4

```

```

L S L R P I K M W S S A K M L L G R I R      F1
Y P Y V Q S K C G L R Q R C C W D E F A      F2
I L T S N Q N V V F G K D V A G T N S H      F3
301 TTATCCTTACGTCCAATCAAAATGTGGTCTTCGGCAAAGATGTTGCTGGGACGAATTCGC 360
----:----|----:----|----:----|----:----|----:----|----:----|
301 AATAGGAATGCAGGTTAGTTTACACCAGAAGCCGTTTCTACAACGACCCTGCTTAAGCG 360
N D K R G I L I H D E A F I N S P R I R      F6
I I R V D L * F T T K P L S T A P V F E      F5
* G * T W D F H P R R C L H Q Q S S N A      F4

```

```

M R R V Y S L F C R L F L G L N T K G F      F1
* E E Y I H F F V D Y F * G L T Q K V F      F2
E K S I F T F L * I I F R A * H K R F S      F3
361 ATGAGAAGAGTATATTCTTTTGTAGATTATTTTTAGGGCTTAACACAAAAGGTTT 420
----:----|----:----|----:----|----:----|----:----|----:----|
361 TACTCTTCTCATATAAGTGAAAAACATCTAATAAAAATCCCGAATTGTGTTTTCCAAA 420
M L L T Y E S K Q L N N K P S L V F P K      F6
C S F L I N V K K Y I I K L A * C L L N      F5
H S S Y I * K K T S * K * P K V C F T K      F4

```

```

P F P * L F P Y K L Q M A C A Q P N F F      F1
L F H S Y F R I N F K W R V H S Q I S S      F2
F S I V I S V * T S N G V C T A K F L P      F3
421 CCTTTTCCATAGTTATTTCCGTATAAACTTCAAATGGCGTGTGCACAGCCAAATTTCTTC 480
----:----|----:----|----:----|----:----|----:----|----:----|
421 GGAAAAGGTATCAATAAAGGCATATTTGAAGTTTACCGCACACGTGTCGGTTTAAAGAAG 480
G K G Y N N G Y L S * I A H A C G F K K      F6
E K E M T I E T Y V E F P T H V A L N R      F5
R K W L * K R I F K L H R T C L W I E E      F4

```

```

Q I A S N Y G R T I Y I I I Y I V * A *      F1
K S L Q I M G E R F I S * F T S F K H R      F2
N R F K L W E N D L Y H N L H R L S I G      F3
481 CAAATCGCTTCAAATTATGGGAGAACGATTTATATCATAATTTACATCGTTTAAGCATAG 540
----:----|----:----|----:----|----:----|----:----|----:----|
481 GTTTAGCGAAGTTTAAATACCTCTTGCTAAATATAGTATTAATGTAGCAAATTCGTATC 540
W I A E F * P L V I * I M I * M T * A Y      F6
G F R K L N H S F S K Y * L K C R K L M      F5
L D S * I I P S R N I D Y N V D N L C L      F4

```

```

G S R N F K I C I S L K W V Q E * S T S      F1
G A E T S K F V S V * S G Y R N D Q L L      F2
E Q K L Q N L Y Q S K V G T G M I N F S      F3
541 GGGAGCAGAACTTCAAAATTTGTATCAGTCTAAAGTGGGTACAGGAATGACAACTTCT 600
----:----|----:----|----:----|----:----|----:----|----:----|
541 CCCTCGTCTTTGAAGTTTTAAACATAGTCAGATTTACCCCATGTCCTTACTAGTTGAAGA 600
P L L F K L I Q I L R F H T C S H D V E      F6
P S C F S * F K Y * D L T P V P I I L K      F5
P A S V E F N T D T * L P Y L F S * S R      F4

```

```

P R * S T T S R I T V L K F F * I Q R L      F1
P A D Q L H P E L R Y * N S F R F K D *      F2
P L I N Y I P N Y G T E I L L D S K I D      F3
601 CCCCCTGATCAACTACATCCCGAATTACGGTACTGAAATTCCTTTAGATTCAAAGATTG 660
----:----|----:----|----:----|----:----|----:----|----:----|
601 GGGGCGACTAGTTGATGTAGGGCTTAATGCCATGACTTTAAGAAAATCTAAGTTTCTAAC 660
G R Q D V V D R I V T S F N K * I * L N      F6
E G S I L * M G F * P V S I R K S E F I      F5
G A S * S C G S N R Y Q F E K L N L S Q      F4

```

```

I * N R F L Y K T V Y F S L W Q R R A I      F1
Y R I D F Y T K R C I F H Y G N D G R Y      F2
I E * I F I Q N G V F F T M A T T G D M      F3
661 ATATAGAATAGATTTTTATACAAAACGGTGTATTTTCACTATGGCAACGACGGCGATA 720
----:----|----:----|----:----|----:----|----:----|----:----|
661 TATATCTTATCTAAAAATATGTTTTGCCACATAAAAAAGTGATACCGTTGCTGCCCGCTAT 720
I Y F L N K Y L V T Y K E S H C R R A I      F6
S I S Y I K I C F P T N K V I A V V P S      F5
Y L I S K * V F R H I K * * P L S P R Y      F4

```

```

W P K P V L L P Y D L I K T L * H H I Y      F1
G Q N R C F Y P T T * * K H C S T T Y I      F2
A K T G A S T L R L N K N I V A P H I S      F3
721 TGGCCAAAACCGGTGCTTCTACCTACGACTTAATAAAAACATTGTAGCACCACATATAT 780
----:----|----:----|----:----|----:----|----:----|----:----|
721 ACCGGTTTTGGCCACGAAGATGGGATGCTGAATTATTTTGTAAACATCGTGGTGTATATA 780
H G F G T S R G * S K I F V N Y C W M Y      F6
I A L V P A E V R R S L L F M T A G C I      F5
P W F R H K * G V V * Y F C Q L V V Y I      F4

```

```

Q F T S T V F T I L L S N * R F W S E I      F1
N L L P Q F S L F C * V I E G F G V K *      F2
I Y F H S F H Y F A K * L K V L E * N N      F3
781 CAATTTACTTCCACAGTTTTCCTATTTTGTAAAGTAATTGAAGGTTTTGGAGTGAAATA 840
----:----|----:----|----:----|----:----|----:----|----:----|
781 GTTAAATGAAGGTGTCAAAAGTGATAAAACGATTCATTAACCTTCCAAAACCTCACTTTAT 840
* N V E V T K V I K S L L Q L N Q L S I      F6
D I * K W L K * * K A L Y N F T K S H F      F5
L K S G C N E S N Q * T I S P K P T F Y      F4

```

```

I C K F Q L I P S * L G K L A T S P F F      F1
F A N F N * F Q V N W V N * L R P H I S      F2
L Q I S I N S K L I G * I S Y V P I S L R      F3
841 ATTGTCAAATTTCAATTAATTCCAAGTTAATTGGGTAAATTAGCTACGTCCCCATTCTTC 900
----:----|----:----|----:----|----:----|----:----|----:----|
841 TAAACGTTTTAAAGTTAATTAAGGTTCAATTAACCCATTTAATCGATGCAGGGTAAGAAG 900
I Q L N * N I G L * N P L N A V D G N K      F6
L K C I E I L E L N I P Y I L * T G M R      F5
N A F K L * N W T L Q T F * S R G W E E      F4

```

```

V W E F Y Q C G A Y V I H * * V T G N C      F1
S G N S I S V V H M L F I S K L Q V T V      F2
L G I L S V W C I C Y S L V S Y R * L C      F3
901 GTCTGGGAATTCTATCAGTGTGGTGCATATGTTATTCATTAGTAAGTTACAGGTAAGTGT 960
----:----|----:----|----:----|----:----|----:----|----:----|
901 CAGACCCTTAAGATAGTCACACCACGTATACAATAAGTAATCATTCAATGTCATTGACA 960
T Q S N * * H P A Y T I * * Y T V P L Q      F6
R R P I R D T H H M H * E N T L * L Y S      F5
D P F E I L T T C I N N M L L N C T V T      F4

```

```

A T K T Q F W L Y S N * P F E H E A V M      F1
R Q R R N S G Y I Q I S H S N T K Q S C      F2
D K D A I L A I F K L A I R T R S S H V      F3
961 GCGACAAAGACGCAATTCTGGCTATATTCAAATTAGCCATTGGAACACGAAGCAGTCATG 1020
----:----|----:----|----:----|----:----|----:----|----:----|
961 CGCTGTTTCTGCGTTAAGACCGATATAAGTTTAATCGGTAAGCTTGTGCTTCGTCAGTAC 1020
A V F V C N Q S Y E F * G N S C S A T M      F6
H S L S A I R A I N L N A M R V R L L *      F5
R C L R L E P * I * I L W E F V F C D H      F4

```

```

Y S W S C P H R P C P L L V L T V C Q T      F1
I V G A A L I D H V R C S C * L F A K L      F2

```

```

      * L E L P S S T M S A A R V D C L P N S      F3
1021 TATAGTTGGAGCTGCCCTCATCGACCATGTCGCTGCTGTTGACTGTTTGCCAAACT 1080
      ----:----|----:----|----:----|----:----|----:----|
1021 ATATCAACCTCGACGGGAGTAGCTGGTACAGGCGACGACACAACGACAAACGGTTTGA 1080
      Y L Q L Q G * R G H G S S T N V T Q W V      F6
      T Y N S S G E D V M D A A R T S Q K G F      F5
      I T P A A R M S W T R Q E H Q S N A L S      F4

      P G V S R S W D I S G T H S P S R S F W      F1
      P E F L D L G T F P G P I R L P E A S G      F2
      R S F S I L G H F R D P F A F Q K L L D      F3
1081 CCCGGAGTTTCTCGATCTTGGGACATTTCCGGGACCCATTGCGCTTCCAGAAGCTTCTGG 1140
      ----:----|----:----|----:----|----:----|----:----|----:----|
1081 GGGCCTCAAAGAGCTAGAACCCCTGTAAAGGCCCTGGGTAAGCGGAAGGTCTTCGAAGACC 1140
      G P T E R D Q S M E P V W E G E L L K Q      F6
      E R L K E I K P C K R S G N A K W F S R      F5
      G S N R S R P V N G P G M R R G S A E P      F4

      I A V S V S S I G T R S P H S P G S R Y      F1
      S L Y R F H P L G D H G V R I H L G Q D T      F2
      R C I G F I H W D T E S A F T W V K I L      F3
1141 ATCGCTGTATCGGTTTCATCCATTGGGACACGGAGTCCGCATTACCTGGGTCAAGATAC 1200
      ----:----|----:----|----:----|----:----|----:----|----:----|
1141 TAGCGACATAGCCAAAGTAGGTAACCCTGTGCCTCAGGCGTAAGTGGACCCAGTTCTATG 1200
      I A T D T E D M P V R L G C E G P D L Y      F6
      S R Q I P K M W Q S V S D A N V Q T L I      F5
      D S Y R N * G N P C P T R M * R P * S V      F4

      F C W S S P G R T T R C P V C V S P G S      F1
      S A G L H R D V L H A V L S V S H P D R      F2
      L L V F T G T Y Y T L S C L C L T R I D      F3
1201 TTCTGCTGGTCTTCACCGGACGTAACACGCTGTCTGTCTGTCTCACCCGGATCG 1260
      ----:----|----:----|----:----|----:----|----:----|----:----|
1201 AAGACGACCAGAAGTGGCCCTGCATGATGTGCGACAGGACAGACAGAGTGGGCCTAGC 1260
      K Q Q D E G P R V V R Q G T Q T E G P D      F6
      S R S T K V P V Y * V S D Q R H R V R I      F5
      E A P R * R S T S C A T R D T D * G S R      F4

      I R L K Y H W T T S S E C G S W L V V G      F1
      S G * S T T G P L L R N V V P G W W W V      F2
      P A E V P L D H F F G M W F L V G G G C      F3
1261 ATCCGGCTGAAGTACCACTGGACCACTTCTTCGGAATGTGGTTCCTGTTGGTGGTGGGT 1320
      ----:----|----:----|----:----|----:----|----:----|----:----|
1261 TAGGCCGACTTCATGGTGACCTGGTGAAGAAGCCTTACACCAAGGACCAACCACCA 1320
      I R S F Y W Q V V E E S H P E Q N T T P      F6
      S G A S T G S S W K K P I H N R T P P      F5
      D P Q L V V P G S R R F T T G P Q H H T      F4

      V L A S P S G T S W P S N G V T L P R S      F1
      F L L L P V V R P G H R T A S P C Q G H      F2
      S C F S Q W Y V L A I E R R H L A K V I      F3
1321 GTTCTTGCTTCTCCAGTGTACGTCTGGCCATCGAACGGCGTCACCTTGCCAAGGTCA 1380
      ----:----|----:----|----:----|----:----|----:----|----:----|
1321 CAAGAACGAAGAGGGTCACCATGCAGGACCGGTAGCTTGCCGACGTGGAACGGTTCCAGT 1380
      T R A E G L P V D Q G D F P T V K G L D      F6
      H E Q K E W H Y T R A M S R R * R A L T      F5
      N K S R G T T R G P W R V A D G Q W P *      F4

      S V F L R I Y S K T E S I I Q R E F V N      F1
      R F S Y E S T A K R N R S S N E S S S T      F2
      G F L T N L Q Q N G I D H P T R V R Q R      F3
1381 TCGGTTTTCTTACGAATCTACAGCAAAACGGAATCGATCATCCAACGAGAGTTTCGTCAAC 1440

```

```

-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1381 AGCCAAAAGAATGCTTAGATGTCGTTTTGCCTTAGCTAGTAGGTTGCTCTCAAGCAGTTG 1440
    D T K K R I * L L V S D I M W R S N T L F6
    M P K R V F R C F P I S * G V L T R * F5
    R N E * S D V A F R F R D D L S L E D V F4

    D H V L C F T P S F T G P R M S A K P Q F1
    I T Y C A L L H R S L D H E C Q P N R S F2
    S R I V L Y S I V H W T T N V S Q T A V F3
1441 GATCACGTATTGTGCTTTACTCCATCGTTCACTGGACCACGAATGTCAGCCAAACCGCAG 1500
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1441 CTAGTGCATAACACGAAATGAGGTAGCAAGTGACCTGGTGCTTACAGTCGGTTTGGCGTC 1500
    S * T N H K V G D N V P G R I D A L G C F6
    R D R I T S * E M T * Q V V F T L W V A F5
    I V Y Q A K S W R E S S W S H * G F R L F4

    Y G P L P C S S Q A L S L T P Q A T H T F1
    M G R Y L A L H K L Y R S R H K Q L I L F2
    W A V T L L F T S S I A H A T S N S Y L F3
1501 TATGGGCCGTTACCTTGCTCTTCACAAGCTCTATCGCTCACGCCACAAGCAACTCATACT 1560
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1501 ATACCCGGAATGGAACGAGAAGTGTTTCGAGATAGCGAGTGCGGTGTTTCGTTGAGTATGA 1560
    Y P G N G Q E E C A R D S V G C A V * V F6
    T H A T V K S K V L E I A * A V L L E Y F5
    I P R * R A R * L S * R E R W L C S M S F4

    C R C W L T P S F R W R S C S L E * P P F1
    V D V G L H L L S V G D H A H W N D R Q F2
    * M L A Y T F F P L E I M L I G M T A N F3
1561 TGTAGATGTTGGCTTACACCTTCTTTCCGTTGGAGATCATGCTCATTGGAATGACCGCCA 1620
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1561 ACATCTACAACCGAATGTGGAAGAAAGGCAACCTCTAGTACGAGTAACCTTACTGGCGGT 1620
    Q L H Q S V G E K R Q L D H E N S H G G F6
    K Y I N A * V K K G N S I M S M P I V A F5
    T S T P K C R R E T P S * A * Q F S R W F4

    T Y R R S I P S P H C W C L R W S S R Y F1
    R I A D Q Y L H H T V G V C G G V R D T F2
    V S Q I N T F T T L L V F A V E F E I L F3
1621 ACGTATCGCAGATCAATACCTTCACCACACTGTTGGTGTTTGCAGTGGAGTTTCGAGATAC 1680
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1621 TGCATAGCGTCTAGTTATGGAAGTGTTGTGACAACCACAAACGCCACCTCAAGCTCTATG 1680
    V Y R L D I G E G C Q Q H K R H L E L Y F6
    L T D C I L V K V V S N T N A T S N S I F5
    R I A S * Y R * W V T P T Q P P T R S V F4

    * G K T S V R R W T S G T W L T * R P A F1
    R G R L P S G V G Q V E R G * H E D L R F2
    G E D F R Q A L D K W N V V D M K T C V F3
1681 TAGGGGAAGACTTCCGTCAGGCGTTGGACAAGTGGAACGTGGTTGACATGAAGACCTGCG 1740
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1681 ATCCCCTTCTGAAGGCAGTCCGCAACCTGTTACCTTGACCAACTGTACTTCTGGACGC 1740
    * P F V E T L R Q V L P V H N V H L G A F6
    S P S S K R * A N S L H F T T S M F V Q F5
    L P L S G D P T P C T S R P Q C S S R R F4

    F D G T R S C W R W * C C F G I S S N C F1
    S T A P E V A G D G D A V S G * A Q T V F2
    R R H Q K L L E M V M L F R D K L K L Y F3
1741 TTCGACGGCACCAGAAGTTGCTGGAGATGGTGATGCTGTTTCGGGATAAGCTCAAAGTGT 1800
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1741 AAGCTGCCGTGGTCTTCAACGACCTCTACCACTACGACAAAGCCCTATTTCGAGTTTGACA 1800

```

```

N S P V L L Q Q L H H H Q K P I L E F Q F6
T R R C W F N S S I T I S N R S L S L S F5
E V A G S T A P S P S A T E P Y A * V T F4

T F C C H C R S T S F R S R F A A * C W F1
P S A V T A D L L L F D H V L L R N A G F2
L L L S L Q I Y F F S I T F C C V M L V F3
1801 ACCTTCTGCTGTCAGTGCAGATCTACTTCTTTTCGATCACGTTTGTGCTGCGTAATGCTGG 1860
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1801 TGGAAGACGACAGTGACGTCTAGATGAAGAAAAGCTAGTGCAAAACGACGCATTACGACC 1860
V K Q Q * Q L D V E K R D R K A A Y H Q F6
Y R R S D S C I * K K E I V N Q Q T I S F5
G E A T V A S R S R K S * T K S R L A P F4

* S N * R P E T I R C S T R * S I L Q P F1
N P T E D R R Q S D V L H A D Q S Y N R F2
I Q L K T G D N Q M F Y T L I N L T T A F3
1861 TAATCCAACCTGAAGACCGGAGACAATCAGATGTTCTACACGCTGATCAATCTTACAACCG 1920
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1861 ATTAGTTGACTTCTGGCCTCTGTTAGTCTACAAGATGTGCGACTAGTTAGAATGTTGGC 1920
Y D L Q L G S V I L H E V R Q D I K C G F6
T I W S F V P S L * I N * V S I L R V V F5
L G V S S R L C D S T R C A S * D * L R F4

H C Y V C Y C S D C S A T I W I * R * V F1
T A M F A T V R T V L R L F G S E G E F F2
L L C L L L F G L F C D Y L D L K V S F F3
1921 CACTGCTATGTTTGCTACTGTTTCGGACTGTTCTGCGACTATTTGGATCTGAAGCTGAGTT 1980
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1921 GTGACGATACAAACGATGACAAGCCTGACAAGACGCTGATAAACCTAGACTTCCACTCAA 1980
C Q * T Q * Q E S Q E A V I Q I Q L H T F6
A S S H K S S N P S N Q S * K S R F T L F5
V A I N A V T R V T R R S N P D S P S N F4

F E G L L Y D V I A A G L T F Q V A E I F1
L K G Y C M T L L L Q D L H F R W P K * F2
* R A T V * R Y C C R T Y I S G G R N K F3
1981 TTTGAAGGGCTACTGTATGACGTTATTGCTGCAGGACTTACATTTCAAGTGGCCGAAATA 2040
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
1981 AAACCTTCCCAGATGACATACTGCAATAACGACGCTCCTGAATGTAAAGTCCACCGGCTTTAT 2040
K S P S S Y S T I A A P S V N * T A S I F6
K Q L A V T H R * Q Q L V * M E P P R F F5
K F P * Q I V N N S C S K C K L H G F Y F4

S D Q V F D S K W S E R I S R D R S M K F1
A T K F S T Q N G R S E S H G I D R * R F2
R P S F R L K M V G A N L T G S I D E E F3
2041 AGCGACCAAGTTTTGCTGACTCAAAATGGTCGGAGCGAATCTCACGGGATCGATCGATGAAG 2100
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
2041 TCGCTGGTTCAAAAGCTGAGTTTTACCAGCCTCGCTTAGAGTGCCCTAGCTAGCTACTTC 2100
L S W T K S E F H D S R I E R S R D I F F6
L R G L K R S L I T P A F R V P D I S F5
A V L N E V * F P R L S D * P I S R H L F4

R N L L M I L M R S Q K K I K F T C G D F1
E T C * * F * C G L R R R S S L P V A I F2
K P A D D S D A V S E E D Q V Y L W R Y F3
2101 AGAAACCTGCTGATGATTCTGATGCGGTCTCAGAAGAAGATCAAGTTTACCTGTGGCGAT 2160
-----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
2101 TCTTTGGACGACTACTAAGACTACGCCAGAGTCTTCTTCTAGTTCAAATGGACACCGCTA 2160
L F R S I I R I R D * F F I L N V Q P S F6
S F G A S S E S A T E S S S * T * R H R F5

```

S V Q Q H N Q H P R L L L D L K G T A I F4

I Y A M S I V T C M N V I N M C Y S A F F1
 F T P C R S L P A * T * S T C A T R H S F2
 L R H V D R Y L H E R D Q H V L L G I H F3
 2161 ATTTACGCCATGTCGATCGTTACCTGCATGAACGTGATCAACATGTGCTACTCGGCATTC 2220
 ----:----|----:----|----:----|----:----|----:----|----:----|
 2161 TAAATGCGGTACAGCTAGCAATGGACGTACTTGCCTAGTTGTACACGATGAGCCGTAAG 2220
 I * A M D I T V Q M F T I L M H * E A N F6
 Y K R W T S R * R C S R S * C T S S P M F5
 N V G H R D N G A H V H D V H A V R C E F4

T L L M N M V Q D * T I E C V K C V L V F1
 R C * * T W F R I K P S N V L N A C L L F2
 A A D E H G S G L N H R M C * M R A C * F3
 2221 ACGCTGCTGATGAACATGGTTCAGGATTAAACCATCGAATGTGTTAAATGCGTGCTTGTT 2280
 ----:----|----:----|----:----|----:----|----:----|----:----|
 2221 TGCGACGACTACTTGTACCAAGTCCTAATTTGGTAGCTTACACAATTTACGCACGAACAA 2280
 V S S I F M T * S * V M S H T L H T S T F6
 * A A S S C P E P N F W R I H * I R A Q F5
 R Q Q H V H N L I L G D F T N F A H K N F4

E N K N S H Q N F N R Q M V Y I V I L H F1
 R I K I A I K T L T A R W F T * L F Y I F2
 E * K * P S K L * P P D G L H S Y S T S F3
 2281 GAGAATAAAAAATAGCCATCAAACTTTAACCGCCAGATGGTTTACATAGTTATTCTACAT 2340
 ----:----|----:----|----:----|----:----|----:----|----:----|
 2281 CTCTTATTTTATCGGTAGTTTGTAAATTGGCGGTCTACCAATGTATCAATAAGATGTA 2340
 S F L F L W * F K L R W I T * M T I R C F6
 Q S Y F Y G D F S * G G S P K C L * E V F5
 L I F I A M L V K V A L H N V Y N N * M F4

P W V D L K K S L V E I L S N A L L K A F1
 P G * I * K N R * S K F * V M P F * R H F2
 L G R F K K I V S R N F E * C P F E G I F3
 2341 CCCTGGGTAGATTTAAAAAATCGTTAGTCGAAATTTTGAGTAATGCCCTTTTGAAGGCA 2400
 ----:----|----:----|----:----|----:----|----:----|----:----|
 2341 GGGACCCATCTAAATTTTTTTAGCAATCAGCTTTAAACTCATTACGGGAAAACCTCCGT 2400
 G Q T S K F F D N T S I K L L A R K F A F6
 D R P L N L F I T L R F K S Y H G K S P F5
 G P Y I * F F R * D F N Q T I G K Q L C F4

* S * T S L I F N E L T E Y C N T G S Y F1
 N R K H R * F L M N * R N I A T P E A I F2
 I V N I V D F * * I N G I L Q H R K L F F3
 2401 TAATCGTAAACATCGTTGATTTTAAATGAATTAACGGAATATTGCAACACCGGAAGCTAT 2460
 ----:----|----:----|----:----|----:----|----:----|----:----|
 2401 ATTAGCATTTGTAGCAACTAAAAATTACTTAATTGCCTTATAACGTTGTGGCCTTCGATA 2460
 Y D Y V D N I K L S N V S Y Q L V P L * F6
 M I T F M T S K * H I L P I N C C R F S F5
 L R L C R Q N K I F * R F I A V G S A I F4

F R M F P N T N D H S I L * R I I L I T F1
 L E C F Q I Q T I I Q Y Y E E L S * L P F2
 * N V S K Y K R S F N I M K N Y P N Y L F3
 2461 TTTAGAATGTTTCCAAATACAAACGATCATTCAATATTATGAAGAATTATCCTAATTACC 2520
 ----:----|----:----|----:----|----:----|----:----|----:----|
 2461 AAATCTTACAAAGGTTTATGTTTGCTAGTAAGTTATAATACTTCTTAATAGGATTAATGG 2520
 K L I N G F V F S * E I N H L I I R I V F6
 N * F T E L Y L R D N L I I F F * G L * F5
 K S H K W I C V I M * Y * S S N D * N G F4

C F F I W A Q G L * C I T D R N S F C I F1
 A F L Y G L K G Y D A L R I G I H F V F F2
 L F Y M G S R A M M H Y G S E F I L Y L F3
 2521 TGCTTTTTTATATGGGCTCAAGGGCTATGATGCATTACGGATCGGAATTCATTTGTATT 2580
 ----:----|----:----|----:----|----:----|----:----|----:----|
 2521 ACGAAAAATATACCCGAGTTCCCGATACTACGTAATGCCTAGCCTTAAGTAAACATAA 2580
 Q K K I H A * P S H H M V S R F E N Q I F6
 R S K * I P E L A I I C * P D S N M K Y F5
 A K K Y P S L P * S A N R I P I * K T N F4

Y W F Y D H N M * G P N C R I F V L V A F1
 T G F M I T T C E A Q I A V Y S C L S H F2
 L V L * S Q H V R P K L P Y I R A C R T F3
 2581 TACTGGTTTTATGATCACAAATGTGAGGCCCAAATTGCCGTATATTCGTGCTTGTGCGA 2640
 ----:----|----:----|----:----|----:----|----:----|----:----|
 2581 ATGACCAAAATACTAGTGTGTACACTCCGGGTTTAACGGCATATAAGCACGAACAGCGT 2640
 * Q N * S * L M H P G F Q R I N T S T A F6
 K S T K H D C C T L G L N G Y I R A Q R F5
 V P K I I V V H S A W I A T Y E H K D C F4

R H T N A K H V S W * K I I S V D N F Q F1
 D I R M Q N M S V G K K L S Q L I T F K F2
 T Y E C K T C Q L V K N Y L S * * L S K F3
 2641 CGACATACGAATGCAAAACATGTCAGTTGGTAAAAAATTATCTCAGTTGATAACTTTCAA 2700
 ----:----|----:----|----:----|----:----|----:----|----:----|
 2641 GCTGTATGCTTACGTTTTGTACAGTCAACCATTTTTTAATAGAGTCAACTATTGAAAGTT 2700
 R C V F A F C T L Q Y F I I E T S L K * F6
 V V Y S H L V H * N T F F * R L Q Y S E F5
 S M R I C F M D T P L F N D * N I V K L F4

S T H K N S N L R C R L C H S * D V A R F1
 V L I R T L I * D V G S V T V R M * R E F2
 Y S * E L * S E M * A L S Q L G C S A R F3
 2701 AGTACTCATAAGAACTCTAATCTGAGATGTAGGCTCTGTACAGTTAGGATGTAGCGCA 2760
 ----:----|----:----|----:----|----:----|----:----|----:----|
 2701 TCATGAGTATTCTTGAGATTAGACTCTACATCCGAGACAGTGTCAATCCTACATCGCGCT 2760
 L V * L F E L R L H L S Q * L * S T A R F6
 F Y E Y S S * D S I Y A R D C N P H L A F5
 T S M L V R I Q S T P E T V T L I Y R S F4

E R S R I R R S * I H F F Y L V N I Y T F1
 K E V E * G D L K F I S F I * L T S T Q F2
 K K * N K E I L N S F L L F S * H L H R F3
 2761 GAAAGAAGTAGAATAAGGAGATCTTAAATTCATTTCTTTTATTAGTTAATCTACACA 2820
 ----:----|----:----|----:----|----:----|----:----|----:----|
 2761 CTTTCTTCATCTTATTCCTCTAGAATTTAAGTAAAGAAAATAAATCAATTGTAGATGTGT 2820
 S L L L I L L D * I * K K * K T L M * V F6
 L F F Y F L S I K F E N R K N L * C R C F5
 F S T S Y P S R L N M E K I * N V D V C F4

D N T E S T I S R H N T R F V A A S L H F1
 I T Q N Q Q F H A T I L G S W P H L S I F2
 * H R I N N F T P Q Y S V R G R I S P S F3
 2821 GATAACACAGAATCAACAATTCACGCCACAATACTCGGTTCTGTCGCGCATCTCTCCAT 2880
 ----:----|----:----|----:----|----:----|----:----|----:----|
 2821 CTATTGTGTCTTAGTTGTTAAAGTGCGGTGTTATGAGCCAAGCACCGCGGTAGAGAGGTA 2880
 S L V S D V I E R W L V R N T A A D R W F6
 L Y C L I L L K V G C Y E T R P R M E G F5
 I V C F * C N * A V I S P E H G C R E M F4

P R F C P T L A K S I R T * S A H L A R F1

L G S A P R S P N R Y A L D L P T * L A F2
 S V L P H A R Q I D T H L I C P P S S L F3
 2881 CCTCGTTTCTGCCCCACGCTCGCCAAATCGATACGCACTTGATCTGCCCACCTAGCTCGC 2940
 ----:----|----:----|----:----|----:----|----:----|----:----|
 2881 GGAGCCAAGACGGGGTGCAGCGGTTTGTAGTATGCGTGAAGTAGACGGGTGGATCGAGCG 2940
 G R N Q G V S A L D I R V Q D A W R A R F6
 D E T R G W A R W I S V C K I Q G G L E F5
 R P E A G R E G F R Y A S S R G V * S A F4

C A P R L L V P T G S E A N T I F A G L F1
 A L H A F L Y Q L D P K R T P S L Q D C F2
 R S T P S C T N W I R S E H H L C R I A F3
 2941 TGCGCTCCACGCCTTCTTGTACCAACTGGATCCGAAGCGAACACCATCTTTGCAGGATTG 3000
 ----:----|----:----|----:----|----:----|----:----|----:----|
 2941 ACGCGAGGTGCGGAAGAACATGGTTGACCTAGGCTTCGCTTGTGGTAGAAACGTCCTAAC 3000
 Q A G R R R T G V P D S A F V M K A P N F6
 S R E V G E Q V L Q I R L S C W R Q L I F5
 A S W A K K Y W S S G F R V G D K C S Q F4

L S G I L A T C P A H R I L P A L A S F F1
 C P A F L Q H A L P I V S F Q L W P P S F2
 V R H S C N M P C P S Y P S S F G L L L F3
 3001 CTGTCCGGCATTCTTGCAACATGCCCTGCCCATCGTATCCTTCCAGCTTTGGCCTCCTTC 3060
 ----:----|----:----|----:----|----:----|----:----|----:----|
 3001 GACAGGCCGTAAGAACGTTGTACGGGACGGGTAGCATAGGAAGGTCGAAACCGGAGGAAG 3060
 S D P M R A V H G A W R I R G A K A E K F6
 A T R C E Q L M G Q G D Y G E L K P R R F5
 Q G A N K C C A R G M T D K W S Q G G E F4

* I L G S P * S W A S S W F I I R R Q T F1
 R Y W V R R R V G R A R G S S F A A K H F2
 D T G F A V E L G E L V V H H S P P N T F3
 3061 TAGATACTGGGTTTCGCCGTAGAGTTGGGCGAGCTCGTGGTTCATCATTCGCCGCCAAACA 3120
 ----:----|----:----|----:----|----:----|----:----|----:----|
 3061 ATCTATGACCCAAGCGGCATCTCAACCCGCTCGAGCACCAAGTAGTAAGCGGCGGTTTGT 3120
 * I S P E G Y L Q A L E H N M M R R W V F6
 R S V P N A T S N P S S T T * * E G G F F5
 L Y Q T R R L T P R A R P E D N A A L C F4

P F S C T P P E I V L S T R R S K T P S F1
 R S P A H R R R S S * A P D V R R L Q V F2
 V L L H T A G D R P K H P T F E D S K C F3
 3121 CCGTTCTCCTGCACACCGCCGAGATCGTCTAAGCACCCGACGTTTCAAGACTCCAAGT 3180
 ----:----|----:----|----:----|----:----|----:----|----:----|
 3121 GGCAAGAGGACGTGTGGCGCCTCTAGCAGGATTCGTGGGTGCAAGCTTCTGAGGTTCA 3180
 G N E Q V G G S I T R L V R R E F V G L F6
 V T R R C V A P S S R G L C G V N S S E L F5
 R E G A C R R L D D * A G S T R L S W T F4

A C R S S S S I V H V S C P * R T T G F F1
 L A G P P R A S S T F H A R R G L P V L F2
 L Q V L L E H R P R F M P V E D Y R F Y F3
 3181 GCTTGCAGGTCCTCCTCGAGCATCGTCCACGTTTTCATGCCCGTAGAGGACTACCGGTTTT 3240
 ----:----|----:----|----:----|----:----|----:----|----:----|
 3181 CGAACGTCCAGGAGGAGCTCGTAGCAGGTGCAAGTACGGGCATCTCCTGATGGCCAAAA 3240
 A Q L D E E L M T W T E H G Y L V V P K F6
 H K C T R R S C R G R K M G T S S * R N F5
 S A P G G R A D D V N * A R L P S G T K F4

M X F1
 * X F2
 E F3

```

3241 ATGAG 3245
----:
3241 TACTC 3245
      I  L
      *  S
      H
                                         F6
                                         F5
                                         F4

#####
Minimum size of ORFs : 1

Total ORFs in frame 1 :    37
Total ORFs in frame 2 :    47
Total ORFs in frame 3 :    37
Total ORFs in frame 4 :    61
Total ORFs in frame 5 :    57
Total ORFs in frame 6 :    54

Total ORFs :    293
#####

```

Sequência proposta

```

MSAARVDCLPNSRSFSILGHFRDPFAFQKLLDRCIGFIHWDTESAFTWVKILLLVF
TGTYITLSCCLTRIDPAEVPLDHFFGMWFLVGGGCSCFSQWYVLAIERRHAK
VIGFLTNLQQNGIDHPTRVRQRSRIVLYSIVHWTNNVSQTAVWAVTLLFTSSIAHA
TSNSYLIMLIGMTANVSQINTFTTLLVFAVEFEILGEDFRQALDKWNVVDMKTCVR
RHQKLLMVMFLFRDKLKLKLYLLLSLQIYFFSITFCCVMLVIQLKTGDNQMFYTLINLT
TALLCLLLFLGLFCDYLDLKVVAEISDQVFDSKWSEIRSRDRSMKRNLLMILMRSQK
KIKFTCGDIYAMSIVTCMNVINMCYSAFTLLNMNVQD

```

Comparação com o parálogo

```

AAEL013418      MSAARVDCLPNSRSFSILGHFRDPFAFQKLLDRCIGFIHWDTESAFTWVKILLLVFTGTY 60
AALF000085      -MSARTDFLQNPRTIPILRNLRDPFAFQKLLDRCIGFIHWDTKSAFTRAKLLLLAFTGTY 59
                  :*. * * * *: : ** : :*****:**** .:***.*****

AAEL013418      YTLSCCLTRIDPAEVPLDHFFGMWFLVGGGCSCFSQWYVLAIERRHAKVIGFLTNLQQ 120
AALF000085      YTLSCCLARINPAEVPLDHFFGMWFLVGGGCSCFSQWYVLAIERRHAKVIEFLTNLQR 119
                  *****:*. :*****

AAEL013418      NGIDHPTRVRQRSRIVLYSIVHWTNNVSQTAVWAVTLLFTSSIAHATSNSYL----- 172
AALF000085      RGVDPTRARYRPRIVLYSIVHWTNNVSQTVVWSATLIFTSSIAHAASSWFVKVLAYVFF 179
                  .*:*****.* * *****.**:.*:*****.*. : :

AAEL013418      ---IMLIGMTANVSQINTFTTLLVFAVEFEILGEDFRQALDKWNVVDMKTCVRRHQKLE 229
AALF000085      PVEIVLIGLTANVSQINTFTTLLVFAVEFEILGEDFRRAFDEWKLDLKS CVRRHQRLLE 239
                  *:***:*****:***:***:***:***:***:***:***:***

AAEL013418      MVMLFRDKLKLKLYLLLSLQIYFFSITFCCVMLVIQLKTGDNQMFYTLINLTALLCLLLFG 289
AALF000085      MVMLFRDKLKMYYLLLSLQIYFFSITFCCVMLVIQLKTGDNQMVYTMINLTALFCLLLFG 299
                  *****:*****.*:*****:*****

AAEL013418      LFCDYLDLKVVAEISDQVFDSKWSEIRSRDRSMKRNLLMILMRSQKIKFTCGDIYAMSI 349
AALF000085      LLCDFLELKV-TEISDQVFSSGWSRSICLDRSMKKNLLMVLMSQKRIKFTCGDIYAMSI 358
                  *:*.*:*** :*****.* *. *. *****:*****:*****:*****

AAEL013418      VTCMNVINMCYSAFTLLNMNVQD      372
AALF000085      VTCMNVINMCYSAFTLLNMNVQD      381
                  *****

```

Havia um *stop codon* na região destacada em vermelho, foi então buscado um íntron para que fosse possível a continuação da sequência de aminoácidos e por isso, o receptor ainda é um pouco menor quando comparado ao parálogo.

Referências

ALEXANDER, B. Sampling methods for phlebotomine sandflies. **Medical and Veterinary Entomology**, v. 14, n. 2, p. 109–122, jun. 2000.

ANDRADE, A. J. et al. Are light traps baited with kairomones effective in the capture of *Lutzomyia longipalpis* and *Lutzomyia intermedia*? An evaluation of synthetic human odor as an attractant for phlebotomine sand flies (Diptera: Psychodidae: Phlebotominae). **Memorias do Instituto Oswaldo Cruz**, v. 103, n. 4, p. 337–340, 2008.

BATES, P. A. et al. Recent advances in phlebotomine sand fly research related to leishmaniasis control. **Parasites & Vectors**, v. 8, n. 1, p. 131, 2015.

BENTON, R. et al. Variant Ionotropic Glutamate Receptors as Chemosensory Receptors in *Drosophila*. **Cell**, v. 136, n. 1, p. 149–162, 9 jan. 2009.

BERNIER, U. R. et al. Analysis of Human Skin Emanations by Gas Chromatography/Mass Spectrometry. 2. Identification of Volatile Compounds That Are Candidate Attractants for the Yellow Fever Mosquito (*Aedes aegypti*). **Analytical Chemistry**, v. 72, n. 4, p. 747–756, 2000.

BERNIER, U. R. et al. Laboratory Studies of *Aedes aegypti* Attraction to Ketones, Sulfides, and Primary Chloroalkanes Tested Alone and in Combination with L-Lactic Acid . **Journal of the American Mosquito Control Association**, v. 31, n. 1, p. 63–70, mar. 2015.

BIRKETT, M. A. et al. The role of volatile semiochemicals in mediating host location and selection by nuisance and disease-transmitting cattle flies. **Medical and Veterinary Entomology**, v. 18, n. 4, p. 313–322, 1 dez. 2004.

BOLSTAD, B. M. et al. A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Variance and Bias. **Bioinformatics**, v. 19, n. 2, p. 185–93, 2003.

BRAKS, M. A. H.; MEIJERINK, J.; TAKKEN, W. The response of the malaria mosquito, *Anopheles gambiae*, to two components of human sweat, ammonia and L-lactic acid, in an olfactometer. **Physiological Entomology**, v. 26, n. 2, p. 142–148, 1 jun. 2001.

BRASIL. **Casos de Leishmaniose visceral no Brasil**. Disponível em: <<https://www.saude.gov.br/images/pdf/2019/outubro/14/LV-Casos.pdf>>. Acesso em: 22 abr. 2020.

BRAY, D. P. et al. Synthetic sex pheromone in a long-lasting lure attracts the visceral leishmaniasis vector, *Lutzomyia longipalpis*, for up to 12 weeks in Brazil. **PLoS neglected tropical diseases**, v. 8, n. 3, p. e2723, mar. 2014.

BRAZIL, R.; RODRIGUES, A.; ANDRADE FILHO, J. D. Sand Fly Vectors of Leishmania in the Americas - A Mini Review. **Entomology, Ornithology and Herpetology: Current Research**, v. 4, n. 2, p. 4–7, 2015.

BROWN, W. L.; EISNER, T.; WHITTAKER, R. H. Allomones and Kairomones: Transspecific Chemical Messengers. **BioScience**, v. 20, n. 1, p. 21–22, 1 jan. 1970.

CAMPANINI, E. B.; DE BRITO, R. A. Molecular evolution of Odorant-binding proteins gene family in two closely related *Anastrepha* fruit flies. **BMC Evolutionary Biology**, v. 16, n. 1, p. 198, 7 dez. 2016.

CAREY, A. F. et al. Odorant reception in the malaria mosquito *Anopheles gambiae*. **Nature**, v. 464, n. 7285, p. 66–71, 3 mar. 2010.

CARVALHO, W. J. DE et al. Characterization of antennal sensilla, larvae morphology and olfactory genes of *Melipona scutellaris* stingless bee. **PLOS ONE**, v. 12, n. 4, p. e0174857, 19 abr. 2017.

CASANOVA, C. et al. Distribution of *Lutzomyia longipalpis* Chemotype Populations in São Paulo State, Brazil. **PLOS Neglected Tropical Diseases**, v. 9, n. 3, p. e0003620, 17 mar. 2015.

CERONI, A. et al. Disulfind: A disulfide bonding state and cysteine connectivity prediction server. **Nucleic Acids Research**, v. 34, n. Web. Serv. Iss., W177-81, Jul. 2006. <https://doi.org/10.1093/nar/gkl266>.

COOK, J. I. et al. Enantiomeric selectivity in behavioural and electrophysiological responses of *Aedes aegypti* and *Culex quinquefasciatus* mosquitoes. **Bulletin of entomological research**, v. 101, n. 5, p. 541–550, 2011.

CORK, A.; PARK, K. C. Identification of electrophysiologically-active compounds for the malaria mosquito, *Anopheles gambiae*, in human sweat extracts. **Medical and Veterinary Entomology**, v. 10, n. 3, p. 269–276, 1 jul. 1996.

COSTA-DA-SILVA, A. et al. Expression and accumulation of the two-domain odorant-binding protein AegOBP45 in the ovaries of blood-fed *Aedes aegypti*. **Parasites & Vectors**, v. 6, n. 1, p. 364, 24 dez. 2013.

COUTO, A.; ALENIUS, M.; DICKSON, B. J. Molecular, anatomical, and functional organization of the *Drosophila* olfactory system. **Current Biology**, v. 15, n. 17, p. 1535–1547, 6 set. 2005.

DE MARCO, A. Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. **Microbial Cell Factories**, v. 8, n. 1, p. 1–18, 14 maio 2009.

DEKEL, A. et al. Evolutionarily conserved odorant receptor function questions ecological context of octenol role in mosquitoes. **Scientific Reports**, v. 6, n. 1, p. 1–7, 16 nov. 2016.

DEKKER, T.; CARDÉ, R. T. Moment-to-moment flight manoeuvres of the female yellow fever mosquito (*Aedes aegypti* L.) in response to plumes of carbon dioxide and human skin odour. **Journal of Experimental Biology**, v. 214, n. 20, p. 3480–3494, 15 out. 2011.

DESJEUX, P. Leishmaniasis: current situation and new perspectives. **Comparative Immunology, Microbiology and Infectious Diseases**, v. 27, n. 5, p. 305–318, 1 set. 2004.

DIAS, A. K. K. et al. Molecular analysis of an odorant-binding protein gene in two sympatric species of *Lutzomyia longipalpis* s.l. **Memórias do Instituto Oswaldo Cruz**, v. 108, n. suppl 1, p. 88–91, 2013.

DIAS, D. B. DOS S. **Análise de genes expressos nas antenas de *Lutzomyia longipalpis* (Lutz & Neiva, 1912) (Diptera: Psychodidae: Phlebotominae. [s.l.: s.n.].**

DORMONT, L.; BESSIÈRE, J. M.; COHUET, A. Human Skin Volatiles: A Review. **Journal of Chemical Ecology**, v. 39, n. 5, p. 569–578, 2013.

DOUGHERTY, M.; HAMILTON, G. Dodecanoic acid is the oviposition pheromone of *Lutzomyia longipalpis*. **Journal of Chemical Ecology**, v. 23, n. 12, p. 2657–2671, 1997.

DOUGHERTY, M. J. et al. Behavioural and electrophysiological responses of the phlebotomine sandfly *Lutzomyia longipalpis* (Diptera: Psychodidae) when exposed to canid host odour kairomones. **Physiological Entomology**, v. 24, n. 3, p. 251–262, 1 set. 1999.

DUJARDIN, J.-C. Risk factors in the spread of leishmaniasis: towards integrated monitoring? **Trends in Parasitology**, v. 22, n. 1, p. 4–6, 1 jan. 2006.

DYE, C.; DAVIES, C. R.; LAINSON, R. Communication among phlebotomine sandflies: a field study of domesticated *Lutzomyia longipalpis* populations in Amazonian Brazil. **Animal Behaviour**, v. 42, n. 2, p. 183–192, 1 ago. 1991.

EIRAS, A. E.; JEPSON, P. C. Host location by *Aedes aegypti* (Diptera: Culicidae): A wind tunnel study of chemical cues. **Bulletin of Entomological Research**, v. 81, n. 2, p. 151–160, 1991.

FAN, J. et al. Review An overview of odorant-binding protein functions in insect peripheral olfactory reception. **Genetics and Molecular Research**, v. 10, n. 4, p. 3056–3069, 8 dez. 2011.

FERNANDES, F. DE F. et al. Fine Structure and Distribution Pattern of Antennal Sensilla of *Lutzomyia longipalpis* (Diptera: Psychodidae) Sand Flies. **EVOLUTION J. Med. Entomol**, v. 45, n. 6, p. 982–990, 2008.

FOWLER, M. A.; MONTELL, C. *Drosophila* TRP channels and animal behavior. **Life Sciences**, v. 92, n. 8–9, p. 394–403, 19 mar. 2013.

FRANCO, T. A. et al. Reverse chemical ecology-based approach leading to the accidental discovery of repellents for *Rhodnius prolixus*, a vector of Chagas diseases refractory to DEET. **Insect Biochemistry and Molecular Biology**, v. 103, p. 46–52, dez. 2018.

FUGANTI, C.; GRASSELLI, P.; SERVI, S. Synthesis of the two enantiomeric forms of erythro-6-acetoxy-5-hexadecanolide, the major component of a mosquito oviposition attractant pheromone. **Journal of the Chemical Society - Series Chemical Communications**, n. 22, p. 1285–1286, 1982.

GALATI, E. A. B. Classificação de Phlebotominae. In: RANGEL, E. F.; LAINSON, R. (Eds.). **Flebotomíneos do Brasil**. 1º ed. **Rio de Janeiro**: Editora Oswaldo Cruz, 2003. p. 23–53.

GALINDO, K.; SMITH, D. P. A Large Family of Divergent *Drosophila* Odorant-Binding Proteins Expressed in Gustatory and Olfactory Sensilla. **Genetics**, v. 159, n. 3, p. 1059–72, 2001.

GANESAN, K. et al. Studies of *Aedes aegypti* (Diptera: Culicidae) ovipositional responses to newly identified semiochemicals from conspecific eggs. **Australian Journal of Entomology**, v. 45, n. 1, p. 75–80, fev. 2006.

GASTEIGER, E. et al. Protein Identification and Analysis Tools on the ExPASy Server. In: **The Proteomics Protocols Handbook**. Humana Press, 2005. p. 571–607.

GETAHUN, M. N. et al. Insect Odorant Response Sensitivity Is Tuned by Metabotropically Autoregulated Olfactory Receptors. **PLoS ONE**, v. 8, n. 3, p. e58889, 12 mar. 2013.

GILLIES, M. T. The role of carbon dioxide in host-finding by mosquitoes (Diptera: Culicidae): A review. **Bulletin of Entomological Research**, v. 70, n. 4, p. 525–532, 1980.

GOULART, T. M. et al. Techniques to improve the maintenance of a laboratory colony of *Nyssomyia neivai* (Diptera: Psychodidae). **Parasites and Vectors**, v. 8, n. 1, 2015.

GOULART, T. M. et al. Breeding protocol for the sand fly *Nyssomyia neivai* (Diptera: Psychodidae) in laboratory conditions. **Acta Tropica**, v. 174, 2017.

HALLEM, E. A.; CARLSON, J. R. Coding of Odors by a Receptor Repertoire. **Cell**, v. 125, n. 1, p. 143–160, 7 abr. 2006.

HAMILTON, J. G. C.; RAMSOONDAR, M. C. Attraction of *Lutzomyia longipalpis* to Human Skin Odours. **Medical and Veterinary Entomology**, v. 8, p. 375–380, 1994.

HARRACA, V. et al. Smelling your way to food: Can bed bugs use our odour? **Journal of Experimental Biology**, v. 215, n. 4, p. 623–629, 15 fev. 2012.

HEALY, T. P.; COPLAND, M. J. W. Activation of *Anopheles gambiae* mosquitoes by carbon dioxide and human breath. **Medical and Veterinary Entomology**, v. 9, n. 3, p. 331–336, 1 jul. 1995.

HEKMAT-SCAFE, D. S. et al. Genome-Wide Analysis of the Odorant-Binding Protein Gene Family in *Drosophila melanogaster*. **Genome Research**, v. 12, n. 9, p. 1357–59, 2002.

HICKNER, P. V. et al. Molecular signatures of sexual communication in the phlebotomine sand flies. **PLoS Neglected Tropical Diseases**, v. 14, n. 12, p. 1–18, 1 dez. 2020.

HOLT, R. A. et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. **Science**, v. 298, n. 5591, p. 129–149, 2002.

HORE, G. et al. Scanning electron microscopic studies on antenna of *Hemipyrellia ligurriens* (Wiedemann, 1830) (Diptera: Calliphoridae)—A blow fly species of forensic importance. **Acta Tropica**, v. 172, p. 20–28, 18 ago. 2017.

HUFF, R. M.; PITTS, R. J. An odorant receptor from *Anopheles gambiae* that demonstrates enantioselectivity to the plant volatile, linalool. **PLoS ONE**, v. 14, n. 11, p. e0225637, 1 nov. 2019.

HUSSAIN, A. et al. Ionotropic Chemosensory Receptors Mediate the Taste and Smell of Polyamines. **PLoS Biology**, v. 14, n. 5, p. 1002454, 4 maio 2016.

ISONO, K.; MORITA, H. Molecular and cellular designs of insect taste receptor system. **Frontiers in Cellular Neuroscience**, v. 4, p. 20, 18 jun. 2010.

JIN, X.; HA, T. S.; SMITH, D. P. SNMP is a signaling component required for pheromone sensitivity in *Drosophila*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 31, p. 10996–1001, 5 ago. 2008.

JONES, T. M.; QUINNELL, R. J. Testing predictions for the evolution of lekking in the sandfly, *Lutzomyia longipalpis*. **Animal Behaviour**, v. 63, n. 3, p. 605–612, 1 mar. 2002.

KAL, A. J. et al. Dynamics of Gene Expression Revealed by Comparison of Serial Analysis of Gene Expression Transcript Profiles from Yeast Grown on Two Different Carbon Sources □ D. **Molecular Biology of the Cell**, v. 10, p. 1859–1872, 1999.

KAMALA JAYANTHI, P. et al. Computational reverse chemical ecology: Virtual screening and predicting behaviorally active semiochemicals for *Bactrocera dorsalis*. **BMC Genomics**, v. 15, n. 1, p. 209, 2014.

KAUPP, U. B. Olfactory signalling in vertebrates and insects: differences and commonalities. **Nature Reviews Neuroscience**, v. 11, n. 3, p. 188–200, 10 mar. 2010.

KILLICK-KENDRICK, R. et al. Preliminary field observations on the flight speed of a phlebotomine sandfly. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 80, n. 1, p. 138–42, 1986.

KILLICK-KENDRICK, R. The biology and control of Phlebotomine sand flies. **Clinics in Dermatology**, v. 17, n. 3, p. 279–289, 1999.

KLINE, D. L.; LEMIRE, G. F. Field evaluation of heat as an added attractant to traps baited with carbon dioxide and octenol for *Aedes taeniorhynchus*. **Journal of the American Mosquito Control Association**, v. 11, n. 4, p. 454–456, 1 dez. 1995.

KRIEGER, J. et al. A candidate olfactory receptor subtype highly conserved across different insect orders. **Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology**, v. 189, n. 7, p. 519–526, 1 jul. 2003.

LAINSON, R.; RANGEL, E. F. *Lutzomyia longipalpis* and the eco-epidemiology of American visceral leishmaniasis, with particular reference to Brazil: a review. **Memórias do Instituto Oswaldo Cruz**, v. 100, n. 8, p. 811–827, dez. 2005.

LARISKA, M. et al. Electronic Olfactory Sensor Based on *A. mellifera* Odorant-Binding Protein 14 on a Reduced Graphene Oxide Field-Effect Transistor. **Angewandte Chemie International Edition**, v. 54, n. 45, p. 13245–13248, 2 nov. 2015.

LEAL, W. S. et al. Reverse and conventional chemical ecology approaches for the development of oviposition attractants for *Culex* mosquitoes. **PLoS ONE**, v. 3, n. 8, p. e3045, 2008.

LEAL, W. S. Odorant Reception in Insects: Roles of Receptors, Binding Proteins, and Degrading Enzymes. **Annual Review of Entomology**, v. 58, n. 1, p. 373–391, 7 jan. 2013.

LEAL, W. S.; CHEN, A. M.; ERICKSON, M. L. Selective and pH-dependent binding of a moth pheromone to a pheromone-binding protein. **Journal of Chemical Ecology**, v. 31, n. 10, p. 2493–2499, 28 out. 2005.

LI, C. et al. Identification and quantitative analysis of genes encoding odorant binding proteins in *Aedes albopictus* (Diptera: Culicidae). **Journal of medical entomology**, v. 49, n. 3, p. 573–80, maio 2012.

LOGAN, J. G.; BIRKETT, M. A. Semiochemicals for biting fly control: their identification and exploitation. **Pest Management Science**, v. 63, p. 647–657, 2007.

LU, T. et al. Odor Coding in the Maxillary Palp of the Malaria Vector Mosquito *Anopheles gambiae*. **Current Biology**, v. 17, n. 18, p. 1533–1544, 18 set. 2007.

LUNDIN, C. et al. Membrane topology of the *Drosophila* OR83b odorant receptor. **FEBS Letters**, v. 581, n. 29, p. 5601–04, 2007.

MACHADO, V. E. et al. Attraction of the sand fly *Nyssomyia neivai* (Diptera: Psychodidae) to chemical compounds in a wind tunnel. **Parasites & vectors**, v. 8, p. 147, 7 mar. 2015.

MAGALHÃES-JUNIOR, J. T. et al. A laboratory evaluation of alcohols as attractants for the sandfly *Lutzomyia longipalpis* (Diptera:Psychodidae). **Parasites & vectors**, v. 7, p. 60, 6 fev. 2014.

MAGALHÃES-JUNIOR, J. T. et al. Attraction of the sandfly *Lutzomyia longipalpis* to possible biomarker compounds from dogs infected with *Leishmania infantum*. **Medical and Veterinary Entomology**, v. 33, n. 2, p. 322–325, 16 jun. 2019.

MAJEED, S. et al. Detection and perception of generic host volatiles by mosquitoes modulate host preference: context dependence of (R)-1-octen-3-ol. **Royal society open Science**, v. 3, n. 11, 160467, 2016. <https://doi.org/10.1098/rsos.160467>.

MARGARYAN, A. et al. Synthesis of an immobilized *Bombyx mori* pheromone-binding protein liquid chromatography stationary phase. **Talanta**, v. 70, n. 4, p. 752–755, 15 nov. 2006.

MAROLI, M. et al. Phlebotomine sandflies and the spreading of leishmaniasis and other diseases of public health concern. **Medical and Veterinary Entomology**, v. 27, n. 2, p. 123–147, jun. 2013.

MATTHEWS, B. J. et al. The neurotranscriptome of the *Aedes aegypti* mosquito. **BMC Genomics**, v. 17, n. 1, p. 32, 6 dez. 2016.

MBOERA, L. E. G. et al. The response of *Anopheles gambiae* s.l. and *A. funestus* (Diptera: Culicidae) to tents baited with human odour or carbon dioxide in Tanzania. **Bulletin of Entomological Research**, v. 87, n. 2, p. 173–178, 1997.

MCBRIDE, C. S. et al. Evolution of mosquito preference for humans linked to an odorant receptor. **Nature**, v. 515, n. 7526, p. 222–227, 13 nov. 2014.

MISSBACH, C. et al. Evolution of insect olfactory receptors. **eLife**, v. 2014, n. 3, 26 mar. 2014.

MODI, G. B.; TESH, R. B. A simple technique for mass rearing *Lutzomyia longipalpis* and *Phlebotomus papatasi* (Diptera: Psychodidae) in the laboratory. **Journal of medical entomology**, v. 20, n. 5, p. 568–9, 5 out. 1983.

MUKABANA, W. R. et al. Evaluation of low density polyethylene and nylon for delivery of synthetic mosquito attractants. **Parasites and Vectors**, v. 5, n. 1, 2012.

NIELSEN, B. L. et al. Behavioural response of sexually naïve and experienced male rats to the smell of 6-methyl-5-hepten-2-one and female rat faeces. **Physiology and Behavior**, v. 120, p. 150–155, 5 ago. 2013.

OMONDI, A. B. et al. Age-dependent regulation of host seeking in *Anopheles coluzzii*. **Scientific Reports**, v. 9, n. 1, p. 1–9, 1 dez. 2019.

PALFRAMAN, M. J. et al. Sobralene, a new sex-aggregation pheromone and likely shunt metabolite of the taxadiene synthase cascade, produced by a member of the sand fly *Lutzomyia longipalpis* species complex. **Tetrahedron Letters**, v. 59, n. 20, p. 1921–1923, 16 maio 2018.

PELOSI, P. Odorant-Binding Proteins. **Critical Reviews in Biochemistry and Molecular Biology**, v. 29, n. 3, p. 199–228, 26 jan. 1994.

PELOSI, P. et al. Soluble proteins in insect chemical communication. **Cellular and Molecular Life Sciences**, v. 63, n. 14, p. 1658–1676, 19 jul. 2006.

PELOSI, P. et al. Structure and biotechnological applications of odorant-binding proteins. **Applied Microbiology and Biotechnology**, v. 98, n. 1, p. 61–70, 22 jan. 2014.

PELOSI, P. et al. Beyond chemoreception: diverse tasks of soluble olfactory proteins in insects. **Biological Reviews**, v. 93, n. 1, p. 184–200, 1 fev. 2018.

PINTO, M. C. et al. Phlebotomine sandfly responses to carbon dioxide and human odour in the field. **Medical and Veterinary Entomology**, v. 15, n. 2, p. 132–139, 2001.

PINTO, M. C. et al. Octenol as Attractant to *Nyssomyia neivai* (Diptera: Psychodidae: Phlebotominae) in the Field. **Journal of Medical Entomology**, v. 48, n. 1, p. 39–44, 2011.

PINTO, M. C. et al. Attraction of the cutaneous leishmaniasis vector *Nyssomyia neivai* (Diptera: Psychodidae) to host odour components in a wind tunnel. **Parasites & Vectors**, v. 5, n. 1, p. 210, 2012.

QIU, Y. T. et al. Behavioural responses of *Anopheles gambiae* sensu stricto to components of human breath, sweat and urine depend on mixture composition and concentration. **Medical and Veterinary Entomology**, v. 25, n. 3, p. 247–255, 1 set. 2011.

QUINNELL, R. J.; DYE, C. An experimental study of the peridomestic distribution of *Lutzomyia longipalpis* (Diptera: Psychodidae). **Bulletin of Entomological Research**, v. 84, n. 3, p. 379–382, 1994.

QUINNELL, R. J.; DYE, C.; SHAW, J. J. Host preference of the phlebotomine sandfly *Lutzomyia longipalpis* in Amazonian Brazil. **Med Vet Entomol**, v. 6, p. 195–200, 1992.

RINKER, D. C. et al. Blood meal-induced changes to antennal transcriptome profiles reveal shifts in odor sensitivities in *Anopheles gambiae*. **Proceedings of the National Academy of Sciences**, v. 110, n. 20, p. 8260–65, 2013.

SAMPAIO, R. P. **Identificação e caracterização de proteínas ligantes de odorantes de *Rhodnius prolixus***. 2018. Dissertação (Mestrado em Física Aplicada Biomolecular)-Instituto de Física de São Carlos, USP, São Carlos, 2018.

SANT'ANA, A. L.; EIRAS, A. E.; CAVALCANTE, R. R. Electroantennographic responses of the *Lutzomyia* (*Lutzomyia*) *longipalpis* (Lutz & Neiva) (Diptera: Psychodidae) to 1-octen-3-ol. **Neotropical Entomology**, v. 31, n. 1, p. 13–17, 2002.

SANTANA, I. B. **Modelagem comparativa e triagem virtual hierárquica para identificação de moduladores das OBPs de *Lutzomyia longipalpis***. 2016. Dissertação (Mestrado em Biotecnologia)- Universidade Estadual de Feira de Santana, Feira de Santana, BA, 2016.

SATO, K. et al. Insect olfactory receptors are heteromeric ligand-gated ion channels. **Nature**, v. 452, n. 7190, p. 1002–1006, 13 abr. 2008.

SCHNEIDER, D. Insect Antennae. **Annual Review of Entomology**, v. 9, n. 1, p. 103–122, jan. 1964.

SMITH, C. N. et al. L-lactic acid as a factor in the attraction of *Aedes aegypti* (Diptera: Culicidae) to human hosts. **Annals of the Entomological Society of America**, v. 63, n. 3, p. 760–770, 15 maio 1970.

SOARES, R. P. P.; TURCO, S. J. *Lutzomyia longipalpis* (Diptera: Psychodidae: Phlebotominae): a review. **Anais da Academia Brasileira de Ciências**, v. 75, n. 3, p. 301–330, set. 2003.

SPIEGEL, C. N. et al. Pheromone Gland Development and Pheromone Production in *Lutzomyia longipalpis* (Diptera: Psychodidae: Phlebotominae). **Journal of Medical Entomology**, v. 48, n. 3, p. 489–495, 2011.

SPIEGEL, C. N. et al. The *Lutzomyia longipalpis* complex: a brief natural history of aggregation-sex pheromone communication. **Parasites & Vectors**, v. 9, n. 1, p. 580, 14 dez. 2016.

TALLON, A. K.; HILL, S. R.; IGNELL, R. Sex and age modulate antennal chemosensory-related genes linked to the onset of host seeking in the yellow-fever mosquito, *Aedes aegypti*. **Scientific Reports**, v. 9, n. 1, p. 43, 10 dez. 2019.

TAPARIA, T.; IGNELL, R.; ROSE HILL, S. Blood meal induced regulation of the chemosensory gene repertoire in the southern house mosquito. **BMC Genomics**, v. 18, n. 393, p. 1–9, 2017.

TAVARES, D. DA S. et al. Attraction of phlebotomine sandflies to volatiles from skin odors of individuals residing in an endemic area of tegumentary leishmaniasis. **PLoS ONE**, v. 13, n. 9, p. 1–14, 2018.

TEGONI, M.; CAMPANACCI, V.; CAMBILLAU, C. Structural aspects of sexual attraction and chemical communication in insects. **Trends in Biochemical Sciences**, v. 29, n. 5, p. 257–264, maio 2004.

TORR, S. J. Dose responses of tsetse flies (*Glossina*) to carbon dioxide, acetone and octenol in the field. **Physiological Entomology**, v. 15, n. 1, p. 93–103, 1 mar. 1990.

TORR, S. J. et al. Methods for dispensing odour attractants for tsetse flies (Diptera: Glossinidae). **Bulletin of Entomological Research**, v. 87, n. 3, p. 299–311, jun. 1997.

VECTORBASE. **Anopheles gambiae**. **VectorBase**. Disponível em: <<https://www.vectorbase.org/organisms/anopheles-gambiae>>. Acesso em: 20 mar. 2019.

VOGT, R. G. et al. Odorant binding protein diversity and distribution among the insect orders, as indicated by LAP, an OBP-related protein of the true bug *Lygus lineolaris* (Hemiptera, heteroptera). **Chemical Senses**, v. 24, n. 5, p. 481–495, 1999.

VOGT, R. G. Odorant Binding Protein Homologues of the Malaria Mosquito *Anopheles gambiae*; Possibel Orthologues of the OS-E and OS-F OBPs of *Drosophila melanogaster*. **Journal of Chemical Ecology**, v. 28, n. 11, p. 29–36, 2002.

VOGT, R. G. Biochemical diversity of odor detection. In: BLOMQUIST, G.; VOGT, R. eds. **Insect Pheromone Biochemistry and Molecular Biology**. Academic Press; Elsevier, 2003. p. 391–445.

VOGT, R. G.; RIDDIFORD, L. M. Pheromone binding and inactivation by moth antennae. **Nature**, v. 293, n. 5828, p. 161–3, 1981.

VOSSHALL, L. B.; WONG, A. M.; AXEL, R. An olfactory sensory map in the fly brain. **Cell**, v. 102, n. 2, p. 147–159, 21 jul. 2000.

WARD, R. D. et al. Reproductive isolation between different forms of *Lutzomyia longipalpis* (Lutz & Neiva), (Diptera: Psychodidae), the vector of *Leishmania donovani chagasi* Cunha & Chagas and its significance to kala-azar distribution in South America. **Memórias do Instituto Oswaldo Cruz**, v. 78, n. 3, p. 269–280, set. 1983.

WARD, R. D. et al. The distribution of two morphological forms of *Lutzomyia longipalpis* (Lutz & Neiva) (Diptera: Psychodidae). **Memórias do Instituto Oswaldo Cruz**, v. 80, n. 2, p. 145–148, jun. 1985.

WHITTAKER, R.; FEENY, P. Allelochemicals: Chemical Interactions between Species. **Science**, v. 171, p. 757–770, 1971.

WHO. **Leishmaniasis. Fact sheet No. 375**. Geneva, Switzerland: World Health Organization; 2017.

WICHER, D. et al. *Drosophila* odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. **Nature**, v. 452, n. 7190, p. 1007–1011, 13 abr. 2008.

XU, P. et al. *Drosophila* OBP LUSH Is Required for Activity of Pheromone-Sensitive Neurons. **Neuron**, v. 45, n. 2, p. 193–200, 20 jan. 2005.

XU, P. et al. Odorant Inhibition in Mosquito Olfaction. **iScience**, v. 19, p. 25–38, set. 2019.

XU, Y.-L. et al. Large-scale identification of odorant-binding proteins and chemosensory proteins from expressed sequence tags in insects. **BMC Genomics**, v. 10, n. 1, p. 632, 25 dez. 2009.

YANG, J. et al. The I-TASSER suite: Protein structure and function prediction. **Nature Methods**, v. 12, n. 1, p. 7–8, 1 jan. 2014.

YOAV BENJAMINI AND YOSEF HOCHBERG. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. **Journal of the Royal Statistical Society**, v. 57, n. 1, p. 289–300, 1995.

ZARBIN, P. H. G.; RODRIGUES, M. A. C. M.; LIMA, E. R. Insect pheromones: Technology and challenges for a competitive agriculture in Brazil. **Quimica Nova**, v. 32, n. 3, p. 722–731, 2009.

ZELLE, K. M. et al. The Genetic Architecture of Degenerin/Epithelial Sodium Channels in *Drosophila*. **G3: Genes|Genomes|Genetics**, v. 3, n. 3, p. 441, mar. 2013.

ZHANG, B.; LI, L.; LÜ, Q. Protein solvent-accessibility prediction by a stacked deep bidirectional recurrent neural network. **Biomolecules**, v. 8, n. 2, 1 jun. 2018.

ZHANG, Y. I-TASSER server for protein 3D structure prediction. **BMC Bioinformatics**, v. 9, 23 jan. 2008.