UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

Genes responsivos ao hormônio ecdisona em
***Bradysia hygida* (Diptera, Sciaridae)**

Henrique Oliveira Frank

Ribeirão Preto

2019

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

**Genes responsivos ao hormônio ecdisona em
Bradysia hygida (Diptera, Sciaridae)**

Henrique Oliveira Frank

Orientadora: Profa Dr^a. Nadia Monesi

Tese apresentada à Faculdade de
Medicina de Ribeirão Preto da
Universidade de São Paulo para
obtenção do título de *Doutor em*
Ciências, área de concentração:
Biologia Celular e Molecular.

Ribeirão Preto

2019

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte, mediante concluído prazo de 2 anos a partir desta data

Catálogo na publicação
Serviço de Biblioteca e Documentação
Faculdade de Odontologia de Ribeirão Preto da Universidade de São Paulo

Frank, Henrique Oliveira

Genes responsivos ao hormônio ecdisona em *Bradysia hygida* (Diptera, Sciaridae)

54 p.

Tese de doutorado, apresentada à Faculdade de Medicina de Ribeirão Preto/USP. Área de concentração: Biologia Celular e Molecular.

Orientadora: Monesi, Nadia.

Palavras-chave:

1. Ecdisona. 2. Sciarídeos. 3. *Bradysia hygida* 4. Genes de resposta primária

Folha de aprovação

Nome: FRANK, Henrique Oliveira

Título: Genes responsivos ao hormônio ecdisona em *Bradysia hygida* (Diptera, Sciaridae)

Tese apresentada à Faculdade de
Medicina da Universidade de São
Paulo para obtenção do título de
Doutor em Ciências.

Aprovado em:

Banca Examinadora

Prof. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Resumo

FRANK, H. O. **Genes responsivos ao hormônio ecdisona em *Bradysia hygida* (Diptera, Sciaridae)**. 2019. Tese (Doutorado) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, São Paulo.

Este trabalho visa a identificação de genes responsivos a via de transdução de sinal do hormônio ecdisona no sciarídeo *Bradysia hygida*. Os genes *EcR*, *USP*, *E75*, *E74* e *Broad-complex* foram os selecionados para dar início a caracterização desta via em *B. hygida*. Inicialmente, através de abordagens *in silico* foi possível elaborar a estrutura gênica e suas respectivas isoformas. Para tal, foram utilizados como ponto de partida sequências brutas de 2 transcriptomas depositados do também sciarídeo *B. odoriphaga*, além de sequências de outros insetos dípteros como *Drosophila melanogaster* e *Aedes aegypti*, porém os dados finais foram obtidos utilizando somente por buscas no genoma de *B. hygida*, com o transcriptoma de embriões de *B. hygida*. Através de buscas no genoma montado de *B. hygida* foi possível identificar em diferentes segmentos genômicos os diferentes genes alvo. Foram identificadas 3 isoformas, que codificam 2 diferentes polipeptídeos, para o gene *USP*. Também foram identificadas quatro diferentes isoformas para o gene *EcR*, nomeadas A1, A2, A3 e B, visto que as 3 pertencentes a isoforma EcRA diferiam somente em suas regiões 5' UTR. Para a identificação dos genes *E74*, *E75* e *Broad-Complex*, também foram utilizadas sequências obtidas no transcriptoma de embriões de *B. hygida*. Obteve-se um total de 2 isoformas de *E74* (A e B), 5 possíveis isoformas para *E75* (A, B, C, D, E) e, também 5 possíveis (A, B, C, D, E) para o gene *Broad-Complex* em *B. hygida*. Com isso, pretende-se estender a caracterização da rede reguladora de expressão gênica induzida pelo hormônio ecdisona e futuramente caracterizar a expressão destes genes em diferentes tecidos. No momento os dados *in silico* permitem afirmar que a via descrita em outros insetos parece estar conservada em *B. hygida*, o que nos permitirá produzir dados inéditos a respeito deste díptero basal e ampliar a caracterização de genes iniciais responsivos à ecdisona, durante a metamorfose.

Palavras-chave: 1.Ecdisona; 2.Sciaridae; 3. *Bradysia hygida*; 4.genes de resposta primária

Abstract

FRANK, H. O. **The ecdysone hormone responsive genes in *Bradysia hygida* (Diptera, Sciaridae)**. 2019. Tese (Doutorado) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, São Paulo.

This work aims to identify responsive genes from the transduction signal pathway of the ecdysone hormone in the sciarid *Bradysia hygida*. The genes EcR, USP, E75, E74 e Broad-complex were selected to begin the characterization of this pathway in *B. hygida*. Initially, through in silico approaches it was possible to elaborate the gene structure and also there respective isoforms. Doing so, it was used as a starting point 2 groups of raw sequences from transcriptomes from the also sciarid *B. odoriphaga*, and also sequences from other dipterian insects as *Drosophila melanogaster* and *Aedes aegypti*, however the final data were obtained only by searching the genome of *B. hygida*, with the transcriptome of embryos of *B. hygida*. Searching the assembled genome from *B. hygida* it was possible to identify on diferente genomic segments the target genes. Three isoforms, which encode 2 different polypeptides, were identified for the USP gene. We also identified four different isoforms for the EcR gene, named A1, A2, A3 and B, since the 3 belonging to the EcRA isoform differed only in their 5' UTR regions. For the identification of E74, E75 and Broad-Complex genes, sequences obtained in the transcriptome of *B. hygida* embryos were also used. A total of 2 isoforms of E74 (A and B), 5 possible isoforms for E75 (A, B, C, D, E) and also 5 (A, B, C, D, E) isoforms were obtained for Broad-Complex gene in *B. hygida*. With this, we intend to extend the characterization of the regulating network of gene expression induced by the hormone ecdysone and to characterize the expression of these genes in different tissues in the future. At the moment the in silico data allow us to affirm that the pathway described in other insects seems to be conserved in *B. hygida*, which will allow us to produce unprecedented data about this basal dipteran and to extend the characterization of initial genes responsive to ecdysone, during metamorphosis.

Keywords: 1.Ecdysone; 2. Sciaridae; 3.*Bradysia hygida*; 4.initial genes

INTRODUÇÃO

1. Insetos e a metamorfose

Insetos estão provavelmente entre os primeiros animais a colonizar ambientes aquáticos de água doce e terrestres. Aspectos ainda pouco entendidos para o entendimento da evolução desse grupo permanecem um desafio para ciência contemporânea. Recentemente, o estudo comparativo de genomas (mitocondriais ou não) de diferentes organismos, aliados a estudos morfológicos culminaram na era pós-genômica, em que é possível integrar diferentes tipos de dados, no intuito de aprimorar as hipóteses que determinam as relações evolutivas entre os insetos (Trautwein et al., 2012). Registros fósseis do primeiro inseto datam do período Devoniano, o que corresponde a aproximadamente a 412 milhões de anos atrás, culminando na hipótese da origem dos primeiros insetos durante o período Siluriano tardio, ou de acordo com evidências moleculares, no período Cambriano, entretanto sem evidências de registros fósseis (Engel; Grimaldi, 2012; Rota-Stabelli et al., 2013).

Modelos de insetos são interessantes para o estudo de mecanismos moleculares que regulam as diferentes fases do desenvolvimento e suas transições em organismos multicelulares, muito por conta da clara diferenciação de estágios durante a muda e metamorfose. Estes eventos de transição estão sob o controle endócrino e a sua compreensão vem sendo desvendada desde os estudos iniciais de células neurosecretoras realizados no inseto *Lymantria dispar*, por Stefan Kopec (Kopec, 1922). Alguns hormônios e neuropeptídeos são responsáveis pelo controle dessas transições. A ecdisona e o hormônio juvenil atuam como “cronômetros endócrinos” para que haja o controle preciso no início das transições nas diferentes fases no desenvolvimento (muda, pupação, eclosão), porém a ecdisona é o regulador mestre dentre todos eles (Di Cara; King-Jones, 2013; Yamanaka et al., 2013).

Na maioria dos insetos, durante o estágio larval, a glândula protorácica fica encarregada de produzir ecdisona. Quando o hormônio é liberado na hemolinfa, ele é convertido na sua forma ativa, 20-hidroxiecdisona, que se liga ao receptor de ecdisona, o que inicia a cascata de expressão gênica. Estas cascatas levam a mudanças fisiológicas, morfológicas e de comportamento essenciais na muda e metamorfose. Estudos acerca de vários aspectos inerentes ao hormônio não só produzem dados para o entendimento de como essas transições durante o desenvolvimento ocorrem, mas também podem fornecer modelos interessantes no entendimento de doenças em humanos, como em processos fisiológicos inerentes ao tráfego de esteróides (Gilbert, 2002), e também defeitos na proliferação e diferenciação em diferentes tecidos para o estudo do câncer (Brumby; Richardson, 2005). Dessa forma, é possível afirmar que o estudo da ecdisona não se restringe apenas ao entendimento dos processos típicos de insetos como: controle de pragas, controle de vetores de doenças e apicultura (Yamanaka et al., 2013).

Estudos clássicos em insetos (Ashburner et al., 1974; Beermann, 1964) em que foram observados os padrões de pufes em cromossomos politênicos induzidos por ecdisona levaram a elucidação da ação do hormônio ecdisona. O modelo de Ashburner, explica como a cascata de expressão gênica é ativada por meio da ligação hormônio-receptor nuclear e o modelo foi ao longo dos anos confirmado por metodologias moleculares e genéticas e, mais recentemente a procura por novos genes induzidos pelo hormônio tem sido conduzida em escala genômica (Beckstead et al., 2005; Gauhar

et al., 2009; Li; White, 2003). Embora a diversidade de resposta ao hormônio ser explicada sobre o espaço e tempo pelas diferentes concentrações do hormônio em que cada gene é induzido, ou pela diferente expressão dos receptores e fatores de transcrição primários induzidos pelo hormônio, ainda não há uma completa compreensão dos mecanismos que conduzem os efeitos provocados pelo hormônio. Entretanto, estudos relacionados ao fator de competência betaFTZ-F1, em *D. melanogaster*, que constitui um receptor nuclear que atua como um ativador estágio e tecido específico (Broadus et al., 1999; Woodard et al., 1994; Yamada et al., 2000), além da atuação de co-ativadores e co-repressores de receptores nucleares (Rosenfeld et al., 2006), podem representar uma significativa expansão ao modelo inicialmente proposto por Ashburner e colegas.

2. Ordem Diptera e Família Sciaridae

Os dípteros são uma das ordens mais diversas entre os insetos, tanto no que se refere ao número de espécies descritas (158000 espécies descritas em mais de 180 famílias) como pela diversidade de hábitos de vida, comportamentos, adaptações morfológicas e impactos em humanos (vetores de doenças, impactos na agricultura, ciência forense) (Pape et al., 2009; Wiegmann; Yeates, 2017). Dado o grande impacto e importância da ordem Diptera, a mesma possui um grande número de espécies de insetos com genomas completos ou parcialmente completos sequenciados (110 genomas). Atualmente existem 5 grandes infraordens dentro do grupo dos dípteros: Tipulomorpha, Brachycera, Culicomorpha, Psychodomorpha e Bibionomorpha, sendo que nas 5 existem atualmente genomas disponíveis, embora a maioria esteja mais bem representada nas famílias Culicidae e Drosophilidae (27 e 33 genomas, respectivamente) do total de 110 genomas disponíveis (Wiegmann et al., 2018). Historicamente, após o advento da era genômica *D. melanogaster*, que já possuía um grande pioneirismo no âmbito da Genética, passou a ser ainda mais notável por ser o primeiro organismo modelo do grupo dos insetos a ter seu genoma sequenciado e anotado e, com esses esforços foi possível determinar que o mesmo possui 13601 genes codificadores de proteína (Adams et al., 2000). A partir daí, questionou-se o grau de similaridade com outras espécies até então os únicos eucariotos sequenciados (*Caenorhabditis elegans* e *Saccharomyces cerevisiae*) e chegou-se a um ortologia de 20% das proteínas de *D. melanogaster* com proteínas de *C. elegans* e *Saccharomyces cerevisiae* (Rubin et al., 2000). Mais adiante, em Holt et al. (2002) e Zdobnov et al. (2002), foi possível realizar esse tipo de comparação entre dois insetos, com a geração do genoma de *Anopheles gambiae*, onde foi observado que existe uma ortologia de 55% dos genes de *A. gambiae* com os genes de *Drosophila*, o que contribuiu para ratificar a ideia de que há mais genes em comum entre organismos próximos filogeneticamente (Waterhouse et al., 2015). Entre os membros dos anofelinos, os esforços investidos nesta direção também puderam garantir uma melhor compreensão de mecanismos de relevância biológica como estudos em desenvolvimento (Jory et al., 2012; Geib et al., 2014), neurobiologia (Grönke et al., 2010; Neely et al., 2010), determinação sexual (Mahajan et al., 2017), regulação gênica e reparo (Richards et al., 2005; Gratz et al., 2014), entre outros.

Um outro foco no estudo de genômica em insetos está entre populações da mesma espécie. Em linhagens de *Drosophila*, por meio destes estudos de populações foi possível observar, quando há redução de recombinação entre a linhagem que sofreu a inversão com a linhagem de cariótipo padrão, abundância e tolerância em inversões intra-cromossômicas, e que estas são fontes de diversidade pela criação de ilhas de

diversidade genômica em relação ao resto do genoma (Mackay; Huang, 2018). Já em anofelinos, em estudo em que 765 espécimes de *An. gambiae* e *An. coluzzii* foram sequenciados, foi possível notar o alto polimorfismo na população, por volta de 3%, o que justificaria o alto potencial adaptativo de insetos, especialmente mosquitos, além de negativamente implicar na dificuldade de implementação de sistemas de modificação gênica como o CRISPR-Cas9. No mesmo estudo, foi também possível observar a presença de fluxo gênico e introgressão na região do gene *Vgsc*, um importante gene associado a resistência a inseticidas piretroides, como o DDT (The Anopheles gambiae 1000 genomes consortium, 2017).

Dentro das características de genomas de dípteros existe uma grande variedade no quesito tamanho intra-ordem, podendo atingir uma diferença de aproximadamente 10^3 pares de bases e, intra-família (Culicidae), podendo atingir uma diferença de 174 Mb a 1,376 Mb (*An. darlingi* e *Culex quinquefasciatus*) (Chen et al., 2015). Uma outra característica interessante a respeito do tamanho genômico entre mosquitos e sua variação intra e interespecífica, é que essa variação parece ser causada principalmente por mudanças e na organização de sequências de DNA repetitivo, por exemplo *Aedes albopictus* possui 68% do seu genoma constituído de sequências repetitivas (Neafsey et al., 2015), e que o aumento do número de classes de DNA repetitivo parece estar linearmente correlacionado com o tamanho do genoma (Black; Rai, 1988). O número de genes também é bastante variável dentro do dípteros, podendo atingir uma diferença de aproximadamente 10000 genes entre espécies do mesmo grupo, como entre *Musca domestica* e *Belgica antarctica* (Waterhouse, 2015; Kelley et al., 2014).

A família Sciaridae compreende um grupo de mais de 2000 espécies distribuídas em todo o mundo, sendo que a maioria delas possuem estágio larval de hábito decompositor (Shin et al., 2013; Mohrig; Menzel, 2009). Em alguns poucos gêneros a larva possui o hábito de consumir partes vivas de plantas e fungos, causando prejuízos na agricultura, ao passo que no gênero *Bradysia* são conhecidas espécies pragas de cogumelos e outros importantes cultivos (Yi et al., 2008; Shin et al., 2012; Shin et al., 2013). De acordo com reconstruções filogenéticas mais recentes utilizando marcadores moleculares, aliado a marcadores morfológicos tradicionais e predominância de habitat durante o estágio larval, a família Sciaridae é composta de 3 subfamílias (Sciarinae, Cratyninae e Megalosphyinae) e 2 grupos (Chetosciara e Pseudolycoriella) (Shin et al., 2013).

No Brasil, o interesse inicial com a família Sciaridae se deu pela presença de estruturas denominadas pufes de DNA nos cromossomos de células da glândula salivar no organismo *Rhynchosciara americana* (Breuer; Pavan, 1955). Estas estruturas chamaram a atenção por serem regiões no genoma que durante o quarto estágio larval acumulavam grande quantidade de DNA além de apresentarem uma estrutura aberta sugerindo ser regiões ativas transcricionalmente, tal qual os pufes de RNA inicialmente descritos em *Chironomus* e *Drosophila* (Sass, 1980). Os estudos iniciais sugeriram que a intensa produção de polipeptídeos poderia estar associada ao fenômeno de amplificação de genes contidos nestes pufes de DNA e que seria um mecanismo que permite a síntese rápida de proteínas na glândula salivar, de modo que a larva entre em estágio de pupa (Almeida et al., 1997), o que mais tarde se consolidou em estudos com marcação de anticorpos destas proteínas, que são secretadas na saliva durante o mesmo período descrito (Fontes et al., 1999; Monesi et al., 2004).

3. Ecdisona na metamorfose

3.1. Papel da ecdisona na regulação gênica

Sob a perspectiva histórica, a identificação do hormônio ecdisona na década de 1950 como um regulador chave durante a muda marcou o início da endocrinologia de insetos (Butenandt; Karlson, 1954). O primeiro indício do papel do hormônio esteroide como um regulador da expressão gênica foi conseguido através de cromossomos politênicos da glândula salivar de *Chironomus*, em que se observou que pufes (iniciais e tardios) tornavam-se ativados após a adição do hormônio em culturas de glândula salivar (Clever; Karlson, 1960). Em seguida, determinou-se que os pufes iniciais eram induzidos mesmo na presença de inibidores de síntese proteica, enquanto que os pufes tardios não. Somado a isso, Ashburner (utilizando *Drosophila* como modelo) determinou que os pufes iniciais se autorregulavam, pois os mesmos não regrediam quando se inibia a síntese proteica, além de corretamente prever que os pufes iniciais eram regiões alvo dos receptores de ecdisona e que os produtos dos genes de pufes iniciais são essenciais para a indução dos pufes tardios (Ashburner, 1974).

Durante diferentes fases do desenvolvimento de insetos em geral, a atuação do hormônio ecdisona, secretado inicialmente pela glândula protorácica e conduzido por todo o corpo pela hemolinfa, acontece de modo a atingir seus tecidos alvo de forma praticamente sistêmica pela sua difusão facilitada em membranas influenciando diretamente na expressão de diversos genes (Gilbert et al. 2002; Shlyueva et al., 2014). O mecanismo de ação se inicia com o receptor nuclear heterodimérico EcR/USP, que se liga diretamente em seu elemento responsivo (*Ecdysone response element - EcRE*) (Cherbas et al., 1991). A atuação de EcR/USP, acoplado ao hormônio, regula na transcrição de um grupo inicial de genes - grande parte deles fatores de transcrição (Ashburner, 1990). O grupo inicial de genes, juntamente com EcR, ativa um grupo de genes tardios, que codificam proteínas que vão atuar na resposta fisiológica ao hormônio (por exemplo, proteínas que atuam na fixação da pupa a um substrato durante a metamorfose). Entretanto, ainda existem questões abertas sobre como um único hormônio se relaciona com seus possíveis módulos *cis*-reguladores alvos e como diferentes respostas fisiológicas em diferentes tecidos são produzidas pelo hormônio. A resposta a esse estímulo hormonal é diversificada, podendo levar desde a morte celular programada em tecidos larvais, até a vias de diferenciação e morfogênese para tecidos que vão formar o organismo adulto. Como exemplo, em células do intestino de dípteros flebotomíneos como o *Lutzomyia longipalpis* e *Phlebotomus papatasi*, onde o promotor de *atg1* contém o elemento canônico responsivo a ecdisona, elemento este que no lepidóptero *Bombyx mori* (Liu et al., 2015) está também presente e atua juntamente com o fator de transcrição E93 no controle do início da metamorfose nesses insetos dando pistas cruciais no mecanismo molecular no desenvolvimento (Malta et al., 2017).

A cultura de células, aliada a técnicas inovadoras como o RNAi, vem continuamente gerando grande impacto para análise funcional em diversos organismos (Moher et al., 2011). Recentemente, diversos estudos em diferentes linhagens celulares de *Drosophila* (Cherbas et al., 2011; Graveley et al., 2011; Shlyueva et al., 2014; Stoiber et al., 2016) abordaram interessantes tópicos da expressão gênica como: diversidade de mRNAs, caracterização de RNAs não-codificadores; caracterização de regiões reguladoras (módulos *cis*-reguladores) induzidas ou reprimidas por ecdisona, e seus diferentes fatores de transcrição parceiros. Estudos funcionais em cultura de células (S2 em sua maioria) de *D. melanogaster* inicialmente identificaram na região promotora de genes responsivos a ecdisona, que a resposta ao hormônio, é difundida entre centenas

de transcritos que sofrem regulação, (Laval et al., 1993). Entretanto, neste cenário ainda não era possível responder como o hormônio ecdisona era responsável pela repressão de certos genes (pela ausência de dados sobre regiões reguladoras de natureza repressora), além de não ser possível definir como o hormônio é capaz de ativar e reprimir ao mesmo tempo a expressão gênica, uma importante questão na elucidação de vias de sinalização em geral (Affolter et al., 2008). A resposta ao hormônio parece atingir regiões reguladoras de genes presentes em grande parte em cromatina fechada, e que se torna mais acessível à maquinaria de transcrição após tratamento com ecdisona (Arnold et al., 2013). Além disso, foi demonstrado por ensaios de ChIP-Seq que a repressão de *enhancer* de genes induzidos por ecdisona é independente da ligação com EcR e dependente da interação com fatores de transcrição parceiros e exclusivos para um tipo celular, como Eip74 (Shlyueva et al., 2014).

O hormônio ecdisona, além de possuir um papel fundamental nos mecanismos que regulam temporalmente a expressão gênica e definir o programa de morfogênese pela indução da expressão de uma certa combinação de fatores de transcrição, contribui na indução dos fatores de transcrição *master*, que, de acordo com estudos bioquímicos, podem possuir o papel de fatores “pioneiros” nos nucleossomos garantindo a ligação de outros fatores de transcrição (Zaret; Mango, 2016). Este fenômeno é observado, por exemplo, na formação da asa de *Drosophila*, quando o fator de transcrição E93, após um pulso do hormônio ecdisona, controla a atividade de módulos *cis*-reguladores promovendo ou não permitindo a acessibilidade de ativadores de acordo com o tempo de desenvolvimento (Uyehara et al., 2017).

3.2. Ecdisona e amplificação

O fenômeno de amplificação em dípteros (como em *Sciara* e *Drosophila*) requer inicialmente a atuação dos ciclos de endoreplicação (ou endociclos) de praticamente quase todo seu genoma, que entra em fase S, seguida de uma fase preparativa para uma subsequente repetida fase S, sem que haja uma fase em que os cromossomos sejam divididos por mitose, formando cromossomos politênicos gigantes (Edgar; Orr-Weaver, 2001; Claycomb; Orr-Weaver, 2005; Shu et al., 2018). Em *D. melanogaster*, o endociclo usa reguladores similares aos do ciclo celular para a sua transição e regulação (CDKs e E2F, por exemplo), além de garantir que exista um controle prevenindo a entrada na mitose pela diminuição da expressão de CDKs específicas da mitose (Lee et al., 2009; Fox; Duronio, 2013). Em células de folículos ovarianos de *Drosophila*, a entrada na mitose é bloqueada mediante a degradação de CDKs mitóticas pelo complexo promotor da anáfase/ciclosomo (APC/C) (Song; Rape, 2011). Já em células endocíclicas da glândula salivar de *D. melanogaster*, foi demonstrado que a progressão da endoreplicação depende da oscilação da atividade de Geminin, que atua no bloqueio do relicenciamento (formação do complexo de pré-replicação durante o final da mitose ou em G1) da fase S (Zielke et al., 2008). Durante a progressão do endociclo, fases S e G se alternam graças a variação do conjunto ciclina E-Cdk 2, em que a ciclina E é degradada durante a fase G para que a montagem do complexo pré-replicativo possa acontecer. Para a progressão da fase S é necessário que ciclina E-Cdk2 estejam ativas e para isso, o inibidor deste conjunto, Dacapo se torna alvo de ubiquitinação por CRL4^{cdt2} (Swanson et al., 2015). Um outro importante alvo de CRL4^{cdt2} é o fator de transcrição E2F1, que tem o pico de atuação durante a fase S, levando a expressão oscilatória da ciclina E (Shibutani et al., 2008; Zielke et al., 2011). Existem evidências (em *Drosophila*) que apontam a existência de mecanismos exclusivos, para diferentes tecidos, para a

entrada no endociclo, em que foi observado que os níveis de ciclina E são altamente expressos no disco imaginal da asa pela via reguladora de crescimento Yki-Scalloped (Shu; Deng, 2017; Shu et al., 2018). Quase todos os tecidos larvais e muitos tecidos adultos de *Drosophila* possuem ploidia aumentada via endociclo. O tecido mais bem estudado dentre os poliploides é a glândula salivar, que durante o desenvolvimento larval necessita passar por 10 endociclos para atingir uma ploidia final de 1024 C (Hammond; Laird, 1985; Hua et al., 2017). O tamanho celular pode ser influenciado pelo conteúdo nuclear, dessa forma a endoreplicação se torna uma maneira eficiente para o crescimento celular onde é visto frequentemente como uma estratégia de células diferenciadas que necessitam aumentar em volume ou se tornar metabolicamente mais ativas, como também durante a regeneração de tecidos (Losick et al., 2013).

O fenômeno de amplificação nos pufes de DNA de sciarídeos, que se inicia no quarto estágio larval e termina no fim do estágio larval, foi descrito por estudos em que foram mensuradas as frequências de amplificação em diversos organismos deste grupo (Coelho et al., 1993; Monesi et al., 1995). Estudos recentes em *B. hygida* levaram a identificação de uma região de baixa amplificação gênica (4,8X) de *BhC4-1* na glândula protorácica (Candido-Silva et al., 2015), ao mesmo tempo em que ocorre a amplificação do mesmo gene na glândula salivar, porém numa frequência superior (21 X), o que pode ser indicativo de um processo de amplificação dependente de fatores exclusivos de certos tecidos. A glândula salivar de sciarídeos e as células foliculares de *Drosophila* são exemplos clássicos de tecidos que sofrem amplificação extra de seu conteúdo de DNA porém, pela histólise desses mesmos tecidos o DNA amplificado não se torna parte permanente de seus cromossomos (Claycomb; Orr-Weaver, 2005). Os endociclos podem produzir células poliploides ou politênicas (ambas com múltiplas cópias do seu conteúdo original de DNA), sendo que as poliploides se tornam células multinucleadas (endomitose) e as politênicas com um núcleo único aumentado contendo todo o DNA replicado (endoreplicação ou endociclo) (Fox; Duranio, 2013; Shu et al., 2018). O gradiente de amplificação pode variar no centro e na periferia destas regiões amplificadas: centro com o maior número de cópias e maior frequência de transcrição e periferia com baixa frequência de transcrição e menor número de cópias (Claycomb; Orr-Weaver, 2005). Além disso, a diferença dos níveis de amplificação pode variar entre tecidos, como em células foliculares (4 vezes) e células do córion (60 a 80 vezes) de *Drosophila*. O aumento do número de cópias pela amplificação gênica também pode influenciar no aumento da transcrição pela interação de proteínas de replicação e transcrição ou pela nova conformação aberta da cromatina nos promotores das regiões amplificadas (Claycomb; Orr-Weaver, 2005).

3.3. Receptores nucleares do hormônio ecdisona

Através do modelo de *Drosophila*, foi possível estabelecer a relação entre os pulsos do hormônio ecdisona, com as diferentes mudas e o início da metamorfose (Riddiford et al., 1993). Os pulsos do hormônio são garantidos pela produção a partir da glândula protorácica, uma parte da glândula anelar, e estes pulsos são característicos quanto a sua amplitude e duração. Apesar da principal fonte de hormônios ecdisteroides ser da glândula protorácica, durante o estágio larval e entre os estágios de pupa e adulto, outros tecidos glandulares como as células do folículo ovariano, e não glandulares, parecem contribuir para síntese (Gilbert et al., 2002). Nos seus tecidos-alvos, a forma ativa do hormônio ecdisona, 20-hidroxeecdisona (20-E), se liga a membros da superfamília de receptores nucleares, os quais constituem fatores de transcrição que possuem uma região altamente conservada de ligação ao DNA (DNA-

Binding Domain, DBD), assim como uma região menos conservada de ligação (**Ligand-Binding Domain, LBD**) (King-Jones; Thummel, 2005). O gene *ecr*, assim como outros genes de resposta primária à ecdisona levaram ao início da era molecular da biologia que envolve o hormônio ecdisona no início da década de 1990 (Burtis et al., 1990; DiBello et al., 1991). A classe de receptores nucleares como EcR, se encontra altamente conservada em metazoários e a sua estrutura comum consiste de: uma região não conservada N-terminal, um domínio ativador transcricional AF-1, uma região altamente conservada de 66-68 resíduos que atua como domínio de ligação ao DNA (DBD), uma pequena região atuando como uma dobradiça, um domínio ligante (LBD), e um domínio F não conservado de tamanho variável (King-Jones; Thummel, 2005; Stoiber et al., 2016). Existem 3 isoformas descritas em *D. melanogaster* (Talbot et al., 1993), sendo que a heterodimerização com o produto do gene *usp* é necessária para que o receptor nuclear se torne capaz de ter alta afinidade com a 20-E (Yao et al., 1993). As 3 diferentes isoformas de EcR (A, B-1 e B-2) possuem diferentes funções durante o desenvolvimento. A isoforma A é expressa nos discos imaginais e na glândula anelar, e mutações no gene impedem a progressão durante estágios tardios de desenvolvimento na pupa (Talbot et al., 1993). As isoformas B possuem também alta afinidade com a 20-E, e são expressas em tecidos larvais que não vão contribuir na forma adulta, além disso, mutações em B1 levam a não progressão da metamorfose (Bender et al., 1997). O parceiro do heterodímero receptor de 20-E, *usp*, é necessário durante a embriogênese e metamorfose, além de também se dimerizar com outros receptores nucleares, DHR38 e Seven-up (Baker et al., 2003), com isto e outras evidências existe possibilidade de receptores nucleares formarem com outros parceiros novas cadeias de regulação gênica.

Em geral, os receptores nucleares em insetos são responsáveis pela percepção e resposta ao hormônio ecdisona. O domínio dedo de zinco do tipo C₄ constitui o domínio de ligação ao DNA (DBD) e possui uma sequência consenso de 8 cisteínas com íons de zinco, formando 2 dedos de zinco com uma estrutura tetraédrica. O primeiro dedo de zinco (P-box) é o responsável específico pela ligação com o DNA e o segundo possui uma capacidade fraca de dimerização, o que permite a dimerização fraca do domínio DBD, mas somente em presença da molécula alvo de DNA (Renaud; Moras, 2000; King-Jones; Thummel, 2005). Esse tipo de domínio funciona como um parceiro de ligação ao DNA, formando heterodímero com outros receptores nucleares de dípteros, como por exemplo HR38, HR78, Seven-up, ou MET (Sutherland et al., 1995; Hirai et al., 2002; Baker et al., 2003). Todos heterodímeros do tipo RXR preferencialmente se ligam a elementos de resposta compostos de repetições de dois sítios AGGTCA espaçados por 1 a 5 pares de base. Estes domínios podem funcionar somente como componente estrutural de um complexo heterodimérico (ou até homodimérico), ou como um componente funcional duplo, o qual funciona como um componente estrutural e receptor ao mesmo tempo (Oro et al., 1990) Uma outra característica comum em receptores nucleares é também a presença na região C-terminal de um domínio principal de ligação para dimerização (Guo et al., 2018).

3.4. Genes iniciais

A ativação de uma série de reguladores-chave na cascata de regulação e diretamente induzidos pelo hormônio 20-E são os fatores de transcrição iniciais. Em *D. melanogaster*, na transição larva-pré-pupa, o hormônio ecdisona induz a expressão de um pequeno grupo de genes, que inclui broad-complex (BR-C), E74 e E75. Estes genes codificam fatores de transcrição que por sua vez, dirigem a expressão de outros genes

tardios, ao passo que atenuam a expressão dos iniciais (Thummel, 1996). A queda subsequente nos níveis do hormônio durante a metade do período pré-pupal acontece um aumento na expressão de *βFTZ-F1*, também conhecido como um fator de competência de estágio e tecido específicos (Woodard et al., 1994; Yamada et al., 2000). Assim que o hormônio volta a aumentar novamente, durante a pré-pupa tardia, a expressão de *βFTZ-F1* diminui e reinduz a expressão dos genes iniciais *E75*, *E74* e *BR-C*. Também em pré-pupas tardias ocorre a ativação de um grupo de genes de resposta inicial, denominados *E93* (Woodard et al., 1994). O mais complexo dentre os genes iniciais e igualmente essencial na entrada da metamorfose (Kiss et al., 1988), está o gene *broad* (**Broad-Complex**), que possui 14 isoformas descritas e 4 possíveis classes de proteínas (Z1 a Z4) (Flybase.org), dependendo do tipo de domínio dedo de zinco presente na isoforma, que por sua vez acredita-se dar especificidade ao alvo da isoforma (Dibello et al., 1991). A atuação de *broad* também é importante na glândula prótorácica, pois regula positivamente a expressão de um gene envolvido no transporte de colesterol, influenciando diretamente na síntese de hormônios esteroides em insetos (Xiang et al., 2010).

Outro regulador que desempenha um papel chave na regulação durante a metamorfose é o *E74*. Em *D. melanogaster* (CG32180) se encontra localizado no cromossomo 3L, no pufe de resposta primária 74EF. O gene codifica quatro diferentes transcritos e quatro diferentes polipeptídeos descritos (A, B, C e E) (flybase.org), ambas com a região C-terminal com domínio ETS DBD (Burtis et al., 1990). A atuação destas isoformas está diretamente associada à metamorfose e aos níveis do hormônio 20-E na hemolinfa, assim como na expressão dos genes tardios, ou seja, enquanto que a isoforma A se torna abundante em altas concentrações de 20-E, a isoforma B se torna abundante em baixas e intermediárias concentrações de 20-E (Urness; Thummel, 1995; Fletcher et al., 1997). As isoformas proteicas de *E74* (A e B) apresentam regiões carboxi-terminais idênticas, mas diferem quanto as suas regiões amino-terminais (Burtis et al., 1990). Os domínios amino-terminais apresentam um caráter predominantemente ácido e os domínios carboxi-terminais são de caráter básico. Estes domínios são separados por uma longa região homopolimérica e, a extremidade carboxi-terminal de ambas isoformas possui um domínio ETS ("E twenty-six"), que constitui um domínio ligante de DNA, característica esta conservada na família de proteínas com domínio ETS, que assumem o papel de fatores de transcrição (Sharrocks, 2001). Estudos de perda de função indicaram que o fator de transcrição *E74*, quando apresentava mutação no seu gene codificador, resultava em letalidade durante a metamorfose, além de perturbar o padrão de expressão gênica de diferentes tecidos larvais, incluindo a glândula salivar (Fletcher; Thummel; 1995; Fletcher et al., 1997). A isoforma B é responsável pela indução dos genes *glue* em glândulas salivares de larvas de terceiro estágio, ao passo que reprime a transcrição de genes considerados como tardios. Já a isoforma A é responsável pela regulação da expressão de genes tardios em pré-pupas (Fletcher; Thummel, 1995a). A identificação por anticorpos específicos para o fator de transcrição *E74* determinaram a existência de cerca de 70 sítios formadores de pufes de RNA em cromossomos politênicos da glândula salivar, o que pode indicar potenciais sítios em que o fator de transcrição *E74* atua como regulador. Dentre estes loci identificados se encontra o pufe tardio 71E, região do gene *L71-5* e *L71-6*, que possui 4 sítios de ligação para o fator *E74A*, dado que estes sítios apresentam uma sequência central GGAA (Urness; Thummel, 1990).

Outro membro da superfamília dos receptores nucleares está o *E75* com seus 7 diferentes transcritos e 5 polipeptídeos únicos descritos (flybase.org). A expressão do gene codifica fatores de transcrição que estão associados a regulação da expressão de

genes iniciais e tardios, cada isoforma de E75 contém uma sequência N-terminal única que é codificada por distintos éxons situados a 5' do gene (Segraves; Hogness, 1990). Nestes estudos iniciais do gene E75 foi possível identificar que a isoforma B não possui um dos domínios dedo de zinco (incapacitando a isoforma de se ligar ao DNA), que estão presentes nas outras duas isoformas inicialmente descritas. Entretanto, a isoforma B consegue se dimerizar com o DHR3, abrindo uma possibilidade na sua função de regulação gênica (White et al., 1997). Além disso, ensaios bioquímicos e de expressão ectópica indicam a função de E75B como repressor de β FTZ-F1, durante a metamorfose (White et al., 1997). As isoformas A e C quando mutadas estão associadas à (A) letalidade durante o estágio larval; (B) defeitos na formação da pupa; (C) desenvolvimento da pupa e viabilidade do adulto (Bialecki et al., 2002). A isoforma C também está associada a via de TOR, durante a metamorfose, e pela sua associação heterodimérica com o receptor nuclear UNF, que regulam o remodelamento de axônios larvais (Rabinovich et al., 2016). Em *A. aegypti* a transcrição do gene *E75* apresenta 3 diferentes isoformas resultantes de splicing alternativo. A expressão das 3 isoformas é induzida no início da vitelogênese e altamente expressa no ovário e corpo gorduroso.

3.5. O gene *BhC4-1* em *B. hygida*

Estudos desenvolvidos no sciarídeo *B. hygida* foram importantes para definir a presença do fenômeno de amplificação gênica nos pufes de DNA dos cromossomos politênicos, neste grupo de insetos (Paçó-Larson et al., 1992; Fontes et al., 1992). Posteriormente, foram definidos os mecanismos de amplificação pelo disparo reiterado de origens de replicação (Coelho et al., 1993; Monesi et al., 1995). Além de ampliarem o conhecimento acerca do hormônio ecdisona durante a formação destes pufes de DNA e nos processos de transcrição e amplificação dos genes presentes em pufes de DNA (Basso et al., 2002; Monesi et al., 2009). O gene *BhC4-1* em *B. hygida*, foi inicialmente detectado como polipeptídeo produzido e liberado na glândula salivar, também foi descrito como gene transcrito amplificado na glândula protorácica, denotando um papel mais generalizado da amplificação gênica além de possuírem outras funções além daquelas relacionadas a formação do pupário (Candido-Silva et al., 2005). A caracterização funcional do gene levou a conclusão que os mecanismos moleculares controladores da amplificação gênica provavelmente foram perdidos pois não estavam mais conservados em *D. melanogaster* (Monesi et al., 1998). Além disso foram identificados possíveis fatores *trans*-ativadores deste gene, como o *Eip74EF* de *D. melanogaster*, que consegue induzir a expressão de *BhC4-1*, na glândula salivar (Basso et al., 2002; Frank et al., 2017). Também foi possível identificar um novo módulo *cis*-regulador conservado em dípteros, que induz a expressão de genes relacionados com a musculatura indireta do voo em *D. melanogaster* (Wester et al., 2018). Em suma, os dados obtidos na região reguladora do gene *BhC4-1* de *B. hygida* são suficientes para indicar uma conservação dos mecanismos de expressão gênica induzidos pelo hormônio ecdisona, que permitem a entrada da metamorfose, provavelmente estão conservados tanto em dípteros superiores como em *D. melanogaster*, tanto nos basais como *B. hygida*.

4. *B. hygida* como modelo de estudo

A partir da cultura do sciarídeo *B. hygida*, mantida desde a década de 60 em condições de laboratório, foi possível observar que o díptero possui um ciclo de vida de

aproximadamente 36 dias, a 22° C e, durante o ciclo, passa pelos estágios embrionário (o qual se estende por 9 dias) (Uliana et al., 2018), larval, pupal (18 dias) e adulto (5 dias) (Laicine et al., 1984). O quarto estágio larval possui mais dados acerca de suas características, como sua duração (10 dias), pois é possível acompanhá-lo e caracterizá-lo pela presença de manchas oculares, que se formam a partir do sexto dia após a terceira muda (idade E1). Outra característica do quarto estágio é a sincronia com a abertura de pufes de cromossomos politênicos de células da glândula salivar, sendo que ao todo são formados 8 pufes de DNA até o final do quarto estágio larval (Laicine et al., 1984). A expansão dos pufes se inicia através de primórdios na idade E3 (48 horas após idade E1). Na idade E7 (4 horas após E3), a expansão do primeiro grupo de pufes torna-se visível. Os pufes do segundo grupo são expandidos 12 horas após E7, até que o último pufe (B3) permanece aberto na muda pré-pupal (E7+24h).

A expressão de genes oriundos de pufes de DNA é uma resposta secundária a presença do hormônio ecdisona, como observado em Basso et al. (2002), a expressão do gene *BhC4-1* em *B. hygida* é induzida tardiamente na glândula salivar, em resposta ao aumento de ecdisona. A cascata de ecdisona se inicia com o reconhecimento do produto dos genes *usp* e *ecr*. A presença destes dois genes é conservada dentro de várias ordens de insetos. O mesmo acontecendo com os genes iniciais da cascata induzida por ecdisona, o que pode demonstrar um papel regulador dos genes tardios (Ali et al., 2013). O estudo amplo de genomas, aliado ao estudo comparativo entre organismos próximos evolutivamente, podem ampliar a função e diversidade de genes iniciais e tardios e, que o produto destes genes pode ou não executar as mesmas funções regulatórias (Gonsalves et al., 2011). Dessa forma, a utilização de abordagens amplas como de geração de transcriptomas poderiam elucidar a existência ou não da conservação desta rede reguladora de genes. Durante o estudo funcional do promotor de *BhC4-1* foi demonstrado que o fator de transcrição E74 de *D. melanogaster* se liga diretamente (Frank et al., 2017), demonstrando uma conservação dos mecanismos que regulam a expressão gênica destes dois representantes da ordem Diptera. Nesse sentido, este trabalho procura responder questões em aberto de como a rede reguladora induzida pelo hormônio ecdisona está presente no díptero basal *B. hygida*, permitindo inferir a existência de ortólogos de genes da via induzida pelo hormônio.

MATERIAL E MÉTODOS

1. Caracterização *in silico* de genes responsivos a ecdisona expressos durante a metamorfose em *B. hygida*.

Visando dar início à caracterização da rede reguladora de expressão gênica que atua na metamorfose em *B. hygida*, selecionamos um conjunto de cinco genes cujos ortólogos em *D. melanogaster* participam da resposta ao hormônio ecdisona. Especificamente, optamos por investir nossos esforços na caracterização do receptor de ecdisona (o heterodímero EcR/USP) e três genes diretamente induzidos pelo hormônio na entrada da metamorfose, os genes *E74*, *broad-complex* e *E75*. Em linhas gerais, foram utilizadas diferentes estratégias para a identificação destes genes no organismo *B. hygida*. Uma delas utilizou-se como ponto de partida sequências do recém sequenciado e montado genoma de *B. hygida*. Para tal, o DNA genômico foi extraído de embriões, pois o mesmo acredita-se não ter sofrido alterações somáticas como amplificação, recombinação e deleção. O sequenciamento foi realizado por tecnologias de nova geração *Illumina* e *PacBio* gerando, ao final da montagem 598 Mb com uma cobertura de 20 vezes. Ainda neste contexto, um aspecto a ser ressaltado é que provavelmente dificultou a montagem do genoma foi a quantidade elevada sequências repetitivas que alcançou 59,1%, um valor relativamente alto em comparação com outro organismo, como *D. melanogaster*, que possui 20% (Kaminker et al., 2002; Barrón et al., 2014). O segundo ponto de partida utilizado, no intuito de identificar sequências de genes induzidos pelo hormônio ecdisona, na entrada da metamorfose, foi um total de 4 transcriptomas do também sciarídeo *B. odoriphaga*. As sequências foram reunidas de RNA extraído de larvas de terceiro (Chen et al., 2015; Gao et al., 2016) e quarto estágio, além de pupas (Gao et al., 2016). Estas sequências foram reunidas e um transcriptoma *de novo* foi realizado em colaboração com o laboratório da professora Dra Tatiana Torres (Instituto de Biociências, USP). O terceiro ponto de partida foram sequências oriundas do transcriptoma de embriões de *B. hygida*. Os dados gerados a partir deste sequenciamento de transcritos são de embriões de até seis dias de desenvolvimento e foram obtidos em colaboração com o laboratório do Professor Dr. Luiz Paulo Andrioli (EACH-USP). Em conjunto, estes 3 pontos de partida foram essenciais para posteriormente, através de abordagens *in silico*, como o alinhamento por BLAST entre sequências do genoma e transcriptoma de *B. hygida* com as do gene EcR e USP de *B. coprophila*. Para determinação da estrutura destes genes utilizamos o programa EST2Genome, que é um dos componentes da plataforma EMBOSS (European Molecular Biology Open Software Suite) que realiza a predição das junções éxon-íntron, além de identificar em quais coordenadas dos segmentos genômicos se encontram éxons e íntrons (Mott, 1997). Com esses dados foi possível reagrupar estas sequências e formar a estrutura completa dos genes *EcR*, *USP*, *E75*, *E74* e *Broad-Complex* além de suas possíveis isoformas.

RESULTADOS E DISCUSSÃO

1.1. Estratégia utilizada para identificação dos segmentos genômicos que contêm os genes *EcR* e *USP* de *B. hygida*.

Para identificação dos segmentos genômicos que contêm os genes *EcR* e *USP* utilizamos como ponto de partida as sequências de mRNA destes genes do sciarídeo *Sciara coprophila* (Foulk et al., 2013), que foi posteriormente reclassificado como *Bradysia coprophila*. As sequências codificadoras de *EcRA* (GQ427082), *EcRB* (GQ427083) e *USP* (GQ427084) foram utilizados para realizar buscas no genoma de embriões de *B. hygida*, assim como do transcriptoma de embriões de *B. hygida*. Em seguida, os *scaffolds* (ou segmentos genômicos) que apresentaram alta identidade com as sequências de busca foram recuperados e investigados mais detalhadamente quanto à presença de sequências codificadoras similares aos genes *EcR* e *USP*. Para determinação da estrutura destes genes utilizamos o programa EST2Genome, que é um dos componentes da plataforma EMBOSS (European Molecular Biology Open Software Suite) que realiza a predição das junções éxon-íntron, além de identificar em quais coordenadas dos segmentos genômicos se encontram éxons e íntrons (Mott, 1997). Com esses dados foi possível reagrupar estas sequências e formar a estrutura completa do gene *EcR* e *USP* além de suas possíveis isoformas (somente para *EcR*).

1.1.1 Gene *USP*

Estudos anteriores revelaram a existência de uma única isoforma de *USP* em *S. coprophila* (Foulk et al., 2013). A sequência correspondente foi recuperada do banco de dados do NCBI e utilizada como sequência de busca no genoma de *B. hygida*. Nesta etapa foram identificados seis éxons distribuídos em dois *scaffolds* (scf95239 e scf95290) de *B. hygida*. Os *scaffolds* scf95239 e scf95290 apresentam respectivamente 101829 pb e 36024 pb e a união dos mesmos, neste caso do gene *USP* assim como nos outros genes analisados posteriormente, pode ser ou não de forma artificial, ou seja, não representa de forma real o que o genoma de *B. hygida* apresenta. Partindo-se dos dados obtidos do transcriptoma de embriões de *B. hygida*, foram identificadas três possíveis isoformas apresentando 6 éxons (USP-A), 6 éxons (USP-B) e 2 éxons (USP-C). Embora as isoformas A e B estejam aparentemente completas por possuírem em sua estrutura gênica as regiões 5' e 3' UTR, além de um códon de parada, os dados correspondentes a isoforma C indicam a ausência de um códon de parada, além da ausência da região 3' UTR, o que não implica na inexistência desta isoforma, mas sim pode indicar que o sequenciamento do transcrito não foi eficientemente realizado. Entretanto a sequência do transcrito parece não estar completa pela falta de um códon de parada no éxon 6, além da ausência de sequências da região 3' UTR. O transcrito completo possui 2277 pares de base e sua proteína predita possui 440 aminoácidos (ver tabela 2).

O primeiro éxon de *USP* está contido no *scaffold* scf95239, enquanto que os demais éxons preditos (E2 a E6) estão localizados no scf95290 (Figura 1). Com base nos dados apresentados na tabela 1 podemos afirmar que o gene *USP* está localizado em uma região genômica de no mínimo 19129 pb e apresenta 6 éxons. Uma vez que não é possível determinar a sobreposição entre os *scaffolds* scf95239 e scf95290 não é possível determinar o tamanho exato do primeiro íntron do gene *USP* de *B. hygida*. O tamanho predito para a sequência codificadora do gene *USP* é 1501 pb. Esta sequência codifica um polipeptídeo de 456 aminoácidos (Figura 2), que apresenta alta identidade (92%) com a proteína USP predita de *S. coprophila* (Figura 3).

Tabela 1. Mapeamento genômico de *USP-A* de *B. hygida*.

<i>USP-A</i> <i>Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	<i>DN7476_c0_g2_i1</i> (start)	<i>DN7476_c0_g2_i1</i> (end)	exon_cDNA
exon 1	426	81831	82256	426	1	426	426
intron 1		82257	115389	33133			
exon 2	289	115390	115678	289	427	715	289
intron 2		115679	119411	3733			
exon 3	358	119412	119769	358	716	1073	358
intron 3		119770	119837	68			
exon 4	171	119838	120008	171	1074	1244	171
intron 4		120009	120188	180			
exon 5	235	120189	120423	235	1245	1479	235
intron 5		120424	120488	65			
exon 6	889	120489	121377	889	1480	2368	889

A coluna *USP-A Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B. hygida*), à direita uma coluna dupla consistindo das coordenadas dos scaffolds em que o éxon e íntron se iniciam e terminam numa junção de scaffolds artificial, resultado da junção do *scaffold* scf95239 com o scf95290 (tamanho total de 137853 nucleotídeos) e que foi utilizado para a predição dos éxons e íntrons, já a coluna dupla com o transcrito (*DN7476_c0_g2_i1*) mostra as diferentes regiões do transcrito de *USP* identificado no transcriptoma de *B. hygida*.

Tabela 2. Mapeamento genômico de *USP-B* de *B. hygida*.

<i>USP-B</i> <i>Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	<i>DN7476_c0_g1_i1</i> (start)	<i>DN7476_c0_g1_i1</i> (end)	exon_cDNA
exon 1	426	81831	82256	426	1	426	426
intron 1		82257	115389	33133			
exon 2	289	115390	115678	289	427	715	289
intron 2		115679	119411	3733			
exon 3	358	119412	119769	358	716	1073	358
intron 3		119770	119837	68			
exon 4	171	119838	120008	171	1074	1244	171
intron 4		120009	120188	180			
exon 5	235	120189	120423	235	1245	1479	235
intron 5		120424	120488	65			
exon6	270	120489	120758	270	1480	1749	270
intron 6		120759	120849	91			
exon 7	528	120850	121377	528	1750	2277	528

A coluna *USP-B Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B. hygida*), à direita uma coluna dupla consistindo das coordenadas dos scaffolds em que o éxon e íntron se iniciam e terminam numa junção de scaffolds artificial, resultado da junção do *scaffold* scf95239 com o scf95290 (tamanho total de 137853 nucleotídeos) e que foi utilizado para a predição dos éxons e íntrons, já a coluna dupla com o transcrito (*DN7476_c0_g1_i1*) mostra as diferentes regiões do transcrito de *USP* identificado no transcriptoma de *B. hygida*.

Tabela 3. Mapeamento genômico de *USP-C* de *B. hygida*.

<i>USP-C</i> <i>Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	<i>DN10508_c0_g1_i1</i> (start)	<i>DN10508_c0_g1_i1</i> (end)	exon_cDNA
exon 1	55	89059	89113	55	1	55	55

Figura 1. Representação gráfica da estrutura gênica de *USPA*, *B* e *C* em *B. hygida*. Na parte superior da figura estão representados os *scaffolds* scf95239 e scf95290, com 137,8 kb. A região de junção entre os dois *scaffolds* está interrompida pois não dispomos de dados no momento acerca da extensão desta região. Mais abaixo, está representada uma figura preenchida na cor preta e de tamanho ≥ 39547 pb, onde está representada a região do gene *USP* e seus éxons e íntrons. Mais abaixo, estão representadas as estruturas gênicas das 3 isoformas (A, B e C) do gene *USP*, sendo que, A e B possuem 6 éxons cada e a isoforma C possui 2 éxons. Os retângulos marcados com a cor cinza representam as regiões 5' e 3' UTR, os retângulos marcados em preto representam os éxons das 3 isoformas. O retângulo azul representa um éxon único da isoforma C. As linhas pontilhadas em vermelho representam um possível *splicing* alternativo, no éxon 6, que gera a isoforma B. As linhas horizontais representam os íntrons e os números acima e abaixo das estruturas gênicas representam o número de nucleotídeos de éxons/regiões UTRs (abaixo) e íntrons (acima).

```
Bhyg_uspA -----MDTTDRGLDQTLMGQMGPISPPDLKPDLKPDTSLLNGNS 39
Bhyg_uspB -----MDTTDRGLDQTLMGQMGPISPPDLKPDLKPDTSLLNGNS 39
Bcop_USP  MLKKEKPMLSVAAI IQSQGLWDRTFAGLDQTLMGQMGPISPPDLKPDLKPDTSLLNGNS 60
          * *  *****

Bhyg_uspA MGGFSPGPGSPGSGYSQMGQNISTTSPGGTQSI-AYPPNHPLSGSKHLCSICGDRASGKH 98
Bhyg_uspB MGGFSPGPGSPGSGYSQMGQNISTTSPGGTQSI-AYPPNHPLSGSKHLCSICGDRASGKH 98
Bcop_usp  MGGFSPGPGSPGSGYSQIGMQNISTTSPGGTQNMSAYPPNHPLSGSKHLCSICGDRASGKH 120
          *****:*****.: *****

Bhyg_uspA YGVYSCEGCKGFFKRTVRKDLTYACREEKACIIDKRQRNRCQYCRYQKCLVCGMKREAVQ 158
```

```

Bhyg_uspB YGVYSCEGCKGFFKRTVRKDLTYACREEKACIIDKRQRNRCQYCRYQKCLVCGMKREAVQ 158
Bcop_usp  YGVYSCEGCKGFFKRTVRKDLTYACREEKACIIDKRQRNRCQYCRYQKCLVCGMKREAVQ 180
          *****

Bhyg_uspA EERQRGAKLQPKQQTEDVNPTSSVRDLTVDRIMEAEQISESRTGDNAIPYLRVGSNTLVP 218
Bhyg_uspB EERQRGAKLQPKQQTEDVNPTSSVRDLTVDRIMEAEQISESRTGDNAIPYLRVGSNTLVP 218
Bcop_usp  EERQRGAKLQPKQNDLNTSSVRDLTVDRIMEAEQMSERTGDNAIPYLRVGSNSMVP 240
          *****.:*:*****:*****:*****:*****:*****:*****

Bhyg_uspA PEYKGAVSHLCQMVNRLYQLIEYARRMPYFDKLQREDQVALLRAGWNELLIATVAWRSI 278
Bhyg_uspB PEYKGAVSHLCQMVNRLYQLIEYARRMPYFDKLQREDQVALLRAGWNELLIATVAWRSI 278
Bcop_usp  QEYKGAVSHLCQMVNKQLYQLIEYARRMPYFDRLQREDQVTLRAGWNELLIATVAWRSI 300
          *****.:*****:*****:*****:*****:*****:*****

Bhyg_uspA EYLDSERTNPDGSI DRRT-TRLPQLMCLGPNFTLHRNSAQQAGVVSIFDRILSEL SVKMK 337
Bhyg_uspB EYLDSERTNPDGSI DRRT-TRLPQLMCLGPNFTLHRNSAQQAGVVSIFDRILSEL SVKMK 337
Bcop_usp  EYLDSERTNPDGSI DRRAPVRFPQLMCLGPNFTLHRNSAQQAGVVAIFDRILSEL SVKMK 360
          *****:*****:.* *****:*****:*****:*****:*****

Bhyg_uspA RLNIDKSELTC LKAIILFN S DIRGLTVRGKEDVNMCREKVYSCLDEHCRTEHSGDDGRFA 397
Bhyg_uspB RLNIDKSELTC LKAIILFN S DIRGLTVRGKEDVNMCREKVYSCLDEHCRTEHSGDDGRFA 397
Bcop_usp  RLNIDKSELTC LKAIILFN S DIRGLTVRGKEDVNMCREKIY SCLDEHCRTEHSGDDGRFA 420
          *****.:*****:*****:*****:*****:*****

Bhyg_uspA QLLRLPALRSISLKCLDHLFFFR LIGDKHLETFVMEMLDTP L 440
Bhyg_uspB QLLRLPALRSISLKCLDHLFFFR LIGDKHLETFVMEMLDTP L 440
Bcop_usp  QLLRLPALRSISLKCLDHLFFFR TDW----- 447
          *****

```

Figura 2. Alinhamento múltiplo de seqüências polipeptídicas de USP para 2 seqüências identificadas em *B. hygida* (A e B), e uma única descrita em *B. coprophila*. As seqüências possuem 440, 440 e 447 aminoácidos, respectivamente. Os domínios conservados DBD e LBD estão marcados em amarelo e azul, respectivamente. Os asteriscos abaixo do alinhamento indicam que o resíduo de aminoácido é conservado completamente. Os dois pontos abaixo do alinhamento significam que os resíduos de aminoácidos fazem parte de um grupo de propriedade altamente similar. O ponto abaixo do alinhamento significa que os resíduos de aminoácidos fazem parte de um grupo com propriedades fracamente similares.

```

Bhyg_uspA -----MDTTDRGLDQTLMGQMGPI SP PDLKPDLKPDTSL L NGNS 39
Bhyg_uspB -----MDTTDRGLDQTLMGQMGPI SP PDLKPDLKPDTSL L NGNS 39
Bhyg_uspC MLKKEKPMLSVA AIIQSQGLWDRTFAGLDQTLMGQMGPI SP ----- 42
Bcop_usp  MLKKEKPMLSVA AIIQSQGLWDRTFAGLDQTLMGQMGPI SP PDLKPDLKPDTSL L NGNS 60
          * * *****

```

Figura 3. Alinhamento contendo a seqüência da proteína USP, isoforma C. O alinhamento compreende somente a região em que a isoforma C se alinha com as outras isoformas de USP de *B. hygida* e a isoforma única de *B. coprophila*. Os números à direita do alinhamento indicam o tamanho de cada seqüência.

As isoformas descritas em *B. hygida*, com exceção da isoforma C por estar incompleta, possuem dois domínios conservados em sua estrutura polipeptídica e numa forma semelhante ao que acontece nos insetos em geral. Mais próxima a região N-terminal dos polipeptídeos se encontra o domínio de ligação ao DNA (DBD), que é homólogo ao receptor de ácido retinoico X de mamíferos (RXR). Já na região mais próxima ao C-terminal se encontra o domínio de ligante (LBD). A região ou domínio A/B, que se localiza na região N-terminal do polipeptídeo de USP, como em todos os receptores esteroides é conhecido por ser pouco conservado entre insetos (Tsai; O'Malley, 1994). Regiões características de receptores nucleares de hormônios como os domínios A/B, C, D e E foram identificadas na estrutura deduzida polipeptídeo de USP (Evans, 1988; Forman; Samuels, 1990). O domínio E/F (domínio LBD) de USP apresenta alta identidade de resíduos de aminoácidos com outros insetos dípteros.

Também foram identificadas regiões características, uma P-box (EGCKG) e uma D-box (REERN), que estão presentes no domínio C, além da T-box (REAVQEER), presente no domínio D. No geral a sequência deduzida do polipeptídeo de USP apresenta uma alta identidade com USP de outros insetos dípteros. No caso de *D. melanogaster* são descritas duas isoformas para o gene *USP* (ambas de mesmo tamanho). Com relação ao domínio E/F (domínio LDB) de *B. hygida*, provavelmente pode ser considerado como um receptor órfão pela alta similaridade com o USP de outros insetos. Este mesmo domínio foi estudado no USP de *D. melanogaster* sugerindo uma estrutura em forma de bolso (ou *pocket*), uma estrutura de conformação antagonista, que previne a ligação de coativadores (Billas et al., 2001; Clayton et al., 2001). Os dados aqui apresentados em *B. hygida* apontam 3 diferentes isoformas para USP (sendo o polipeptídeo A igual ao polipeptídeo B) expressos somente em embriões. Pesquisas conduzidas em *D. melanogaster*, apontam a presença de uma única isoforma para USP e, a respeito da função deste gene apontam que, além de ser um parceiro de EcR na mediação em resposta ao hormônio ecdisona, também possui papel na morfogênese do olho, fertilização, desenvolvimento do tórax de adultos, além de ser expresso em diversas regiões do organismo e em todos os estágios do desenvolvimento, sendo que os níveis de expressão mais altos são encontrados durante o estágio embrionário (embriões de 0-2h), em larvas de terceiro estágio tardio, e em fêmeas adultas nos folículos ovarianos (Shea et al., 1990; Oro et al., 1990; Heinrich et al., 1994).

1.1.2 Gene *EcR*

Para o gene *EcR*, no sciarídeo *Bradysia coprophila* foram descritas 2 isoformas (A e B) (Foulk et al., 2012). Inicialmente, as sequências de ambas isoformas de *B. coprophila* foram utilizadas como sequências de busca no genoma de *B. hygida*. Estas análises resultaram na predição de duas isoformas de *EcR* em *B. hygida*. A mesma busca foi realizada no transcriptoma de embriões de *B. hygida*, o que levou a identificação de 4 diferentes transcritos no genoma de *B. hygida*, nomeados A1, A2, A3 e B. As isoformas A1, A2 e A3 apresentam 9 (A1 e A2) e 8 (A3) éxons, dos quais 4 (A1 e A2) e 3 (A3) éxons estão localizados no *scaffold* scf91162 e cinco no *scaffold* scf110350 (figura 3). No caso da isoforma B, os dois primeiros éxons estão localizados no *scaffold* scf91162 e os demais cinco éxons estão localizados no *scaffold* scf110350 (figura 3). O tamanho dos éxons e íntrons preditos, assim como as coordenadas destas regiões nos dois *scaffolds* de *B. hygida* são apresentados nas tabelas 5 a 8.

Com base nos dados apresentados nas tabelas 4 a 7 podemos afirmar que o gene *EcR* está localizado em um fragmento genômico de no mínimo 177994 pb e que os scaffolds unidos artificialmente produzem um fragmento genômico de 254666 pb, haja vista que não há sobreposição entre os *scaffolds* scf91162 e scf110350 não é possível determinar no momento o tamanho exato do íntron entre os éxons 2 e 3 das isoformas A1, A2 e A3 e tampouco do íntron entre os éxons 1 e 2 da isoforma B. Finalmente, o tamanho predito para a sequência codificadora das isoformas EcRA1, A2, A3 é de 5507, 4836 e 5343, enquanto que para a isoforma EcRB é de 4208 pb (tabelas 5 a 8). Estas sequências codificam polipeptídeos preditos de 610 (para as 3 isoformas de A) e 541 aminoácidos. As isoformas de EcRA de *B. hygida* apesar de codificarem 3 diferentes transcritos apresentando modificações expressivas na região 5' do transcrito, o produto do gene apresenta somente uma única sequência de 610 aminoácidos idêntica entre os 3 transcritos de EcRA. Estas 3 diferentes isoformas com modificações na região 3' podem sugerir uma especificidade espaço-temporal na função do gene *EcR* adicionando uma nova camada de regulação na expressão gênica do gene *EcR* em *B. hygida* (Watanabe et al., 2010). A respeito das diferentes isoformas de EcR e seus

diferentes mecanismos na regulação da transcrição, o domínio A/B de receptores nucleares geralmente possui a região independente de ligação de ativação ou AF-1 (Wärnmark et al., 2003). Em *Drosophila*, estudos na região AF-1 em isoformas de EcR revelaram uma alta atividade transativadora de isoformas B, enquanto que as isoformas de A possuem uma atividade fraca transativadora (Hu et al., 2003). A região AF-1 que se situa no N-terminal de EcRB-1 possui alta conservação entre insetos holometábolos derivados (Mouillet et al., 2001). Apesar da existência de microdomínios que foram conservados ou modificados em diferentes isoformas de EcR, além do fato da região AF-1 possuir uma natureza de envolvimento com o recrutamento de cofatores e regulação transcricional, a isoforma B1 de vários insetos holometábolos apresenta uma novidade evolutiva, que é uma região específica altamente conservada na região N-terminal, o motivo (K/R)RRW (Watanabe et al., 2010). Corroborando com estes dados, no organismo *B. hygida*, também foi identificado nos resíduos de aminoácidos iniciais da região N-terminal da isoforma predita de EcR-B (Figura 4). A isoforma EcR-A geralmente possui 4 motivos específicos estruturais: o motivo de SUMOilação, NLS, (D/E)(D/E)W e A-box. Dentre eles os dois últimos são os mais conservados, da região N-terminal, entre artrópodes (Gwozdz et al., 2006). O motivo de SUMOilação, que é resultado de uma modificação pós-traducional reversível, participa de alguns processos celulares como transporte núcleo-citoplasma e regulação transcricional, podendo conduzir a repressão na atividade de fatores de transcrição (Wilson; Rangasamy, 2001; Gill, 2003; Zhao, 2007). Nas três isoformas EcR-A identificadas em *B. hygida* existe a presença deste motivo representado por VKAE, o que sugere uma possível função na regulação transcricional destas isoformas, presente na região A/B. Na estrutura de EcRA também é possível identificar o motivo NLS (nuclear localization sequence) (KRPR), após o motivo de SUMOilação, o que sugere uma isoforma EcR-A do tipo 5, que é característica de dípteros (Watanabe et al., 2010).

Nas diferentes isoformas de *B. hygida* preditas para EcR foram identificados os domínios do tipo C2C2, em que o átomo de zinco se liga aos 4 resíduos de cisteína. Em *D. melanogaster* estão anotados 6 diferentes transcritos e 3 diferentes polipeptídeos de EcR. Com relação a expressão da proteína de EcR em *D. melanogaster*, a proteína de EcR de 105 kD já foi detectada em embriões, larvas de terceiro estágio e pupas, além de baixos níveis em larvas de primeiro estágio e adultos (Koelle et al., 1991), o que sinaliza uma concordância dos dados aqui apresentados para os diferentes transcritos de EcR identificados no transcriptoma de embriões de *B. hygida*.

Tabela 4. Mapeamento genômico da isoforma A1 do gene *EcR* de *B. hygida*.

EcRA1 <i>Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	DN26490_c1_g1_i1 (start)	DN26490_c1_g1_i1 (end)	exon_cDNA
exon 1	661	78929	79589	661	1	661	661
intron 1		79590	103133	23544			
exon 2	549	103134	103682	549	662	1210	549
intron 2		103683	149823	46141			
exon 3	530	149824	150353	530	1211	1740	530
intron 3		150354	156365	6012			
exon 4	173	156366	156538	173	1741	1913	173
intron 4		156539	237211	80673			
exon 5	174	237212	237385	174	1914	2087	174
intron 5		237386	237448	63			

exon 6	311	237449	237759	311	2088	2398	311
intron 6		237760	237816	57			
exon 7	150	237817	237966	150	2399	2548	150
intron 7		237967	238030	64			
exon 8	153	238031	238183	153	2549	2701	153
intron 8		238184	239163	980			
exon 9	2804	239164	241969	2806	2702	5507	2806

A coluna *EcRA1 Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B. hygida*), à direita uma coluna dupla consistindo das coordenadas dos scaffolds em que o éxon e íntron se iniciam e terminam numa junção de scaffolds artificial, resultado da junção dos scaffolds scf91162 e scf110350 (tamanho total de 254666 nucleotídeos) e que foi utilizado para a predição dos éxons e íntrons, já a coluna *TRINITY* mostra o tamanho das diferentes regiões do transcrito de *EcR* identificado no transcriptoma de *B. hygida*.

Tabela 5. Mapeamento genômico da isoforma A2 do gene *EcR* de *B. hygida*.

EcRA2 <i>Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	DN26490_c1_g1_i3 (start)	DN26490_c1_g1_i3 (end)	exon_cDNA
exon 1	497	64029	64525	497	1	497	497
intron 1		64526	103133	38608			
exon 2	549	103134	103682	549	498	1046	549
intron 2		103683	149823	46141			
exon 3	530	149824	150353	530	1047	1576	530
intron 3		150354	156365	6012			
exon 4	173	156366	156538	173	1577	1749	173
intron 4		156539	237211	80673			
exon 5	174	237212	237385	174	1750	1923	174
intron 5		237386	237448	63			
exon 6	311	237449	237759	311	1924	2234	311
intron 6		237760	237816	57			
exon 7	150	237817	237966	150	2235	2384	150
intron 7		237967	238030	64			
exon 8	153	238031	238183	153	2385	2537	153
intron 8		238184	239163	980			
exon 9	2804	239164	241969	2806	2538	5343	2806

A coluna *EcRA2 Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B. hygida*), à direita uma coluna dupla consistindo das coordenadas dos scaffolds em que o éxon e íntron se iniciam e terminam numa junção de scaffolds artificial, resultado da junção dos scaffolds scf91162 e scf110350 (tamanho total de 254666 nucleotídeos) e que foi utilizado para a predição dos éxons e íntrons, já a coluna *TRINITY* mostra o tamanho das diferentes regiões do transcrito de *EcR* identificado no transcriptoma de *B. hygida*.

Tabela 6. Mapeamento genômico da isoforma A3 do gene *EcR* de *B. hygida*.

EcRA3 <i>Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	DN26490_c1_g1_i2 (start)	DN26490_c1_g1_i2 (end)	exon_cDNA
exon 1	539	145391	145929	539	1	539	539
intron 1		145930	149823	3894			

exon 2	530	149824	150353	530	540	1069	530
intron 2		150354	156365	6012			
exon 3	173	156366	156538	173	1070	1242	173
intron 3		156539	237211	80673			
exon 4	174	237212	237385	174	1243	1416	174
intron 4		237386	237448	63			
exon 5	311	237449	237759	311	1417	1727	311
intron 5		237760	237816	57			
exon 6	150	237817	237966	150	1728	1877	150
intron 6		237967	238030	64			
exon 7	153	238031	238183	153	1878	2030	153
intron 7		238184	239163	980			
exon 8	2804	239164	241969	2806	2031	4836	2806

A coluna *EcRA3 Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B. hygida*), à direita uma coluna dupla consistindo das coordenadas dos scaffolds em que o éxon e íntron se iniciam e terminam numa junção de scaffolds artificial, resultado da junção dos *scaffolds* scf91162 e scf110350 (tamanho total de 254666 nucleotídeos) e que foi utilizado para a predição dos éxons e íntrons, já a coluna *TRINITY* mostra o tamanho das diferentes regiões do transcrito de *EcR* identificado no transcriptoma de *B. hygida*.

Tabela 7. Mapeamento genômico da isoforma B do gene *EcR* de *B. hygida*.

<i>EcRB Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	DN26490_c1_g3_i1 (start)	DN26490_c1_g3_i1 (end)	exon_cDNA
exon 1	151	202813	202963	151	1	151	151
intron 1	-20	202964	209015	6052			
exon 2	463	209016	209478	463	152	614	463
intron 2	-20	209479	237211	27733			
exon 3	174	237212	237385	174	615	788	174
intron 3	-20	237386	237448	63			
exon 4	311	237449	237759	311	789	1099	311
intron 4	-20	237760	237816	57			
exon 5	150	237817	237966	150	1100	1249	150
intron 5	-20	237967	238030	64			
exon 6	153	238031	238183	153	1250	1402	153
intron 6	-20	238184	239163	980			
exon 7	2804	239164	241969	2806	1403	4208	2806

A coluna *EcRB Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B. hygida*), à direita uma coluna dupla consistindo das coordenadas dos scaffolds em que o éxon e íntron se iniciam e terminam numa junção de scaffolds artificial, resultado da junção dos *scaffolds* scf91162 e scf110350 (tamanho total de 254666 nucleotídeos) e que foi utilizado para a predição dos éxons e íntrons, já a coluna *TRINITY* mostra o tamanho das diferentes regiões do transcrito de *EcR* identificado no transcriptoma de *B. hygida*.

Figura 3. Representação gráfica da estrutura gênica de *EcR* em *B. hygida*. Na parte superior da figura estão representados parte dos scaffolds *scf91162* e *scf110350*. A região de junção entre os dois scaffolds está interrompida pois não dispomos de dados no momento acerca da extensão desta região. Na parte inferior da figura estão representadas as estruturas das isoformas *EcRA* (A1, A2 e A3) e *EcRB* de *B. hygida* e seu posicionamento em relação aos scaffolds *scf91162* e *scf110350*. Nestas figuras, se encontram os éxons das diferentes isoformas, sendo que a cor vermelha representa o primeiro éxon das isoformas de *EcRA* e a cor azul o primeiro éxon de A2. Na parte inferior da figura está representada a estrutura do gene *EcR* propriamente dita. Os retângulos negros representam os éxons comuns entre as duas isoformas. Os retângulos em cinza representam as regiões 5' e 3' UTR e as linhas representam os íntrons.

```

Bhyg_EcRA1 MFRVNTTNPVAVTGVSHGHSQVITQQQTHHISQQSAPSHQQLSSLQQHSHQPMGSVLSQLD 60
Bhyg_EcRA2 MFRVNTTNPVAVTGVSHGHSQVITQQQTHHISQQSAPSHQQLSSLQQHSHQPMGSVLSQLD 60
Bhyg_EcRA3 MFRVNTTNPVAVTGVSHGHSQVITQQQTHHISQQSAPSHQQLSSLQQHSHQPMGSVLSQLD 60
Bcop_EcRA  MFRVNTTNPVAVTGVSHGHSQVITQQQTHHISQQSAPSHQQLSSLQQHSHQPLGSVLSQLD 60
Bhyg_EcRB  -----MKRRWSNNGGFALRMLE 17
Bcop_EcRB  -----MKRRWSNNGGFALRMLE 17
                                     ::: : * . * *:

Bhyg_EcRA1 SISVTGGNGPNDNSIIINETILTAVKAEPLHETASTSTNVTNVHVLYTSIKRPRLEGND 120
Bhyg_EcRA2 SISVTGGNGPNDNSIIINETILTAVKAEPLHETASTSTNVTNVHVLYTSIKRPRLEGND 120
Bhyg_EcRA3 SISVTGGNGPNDNSIIINETILTAVKAEPLHETASTSTNVTNVHVLYTSIKRPRLEGND 120
Bcop_EcRA  AISVSGSNGPNDNSILINETILTAVKAEPMHDTASTSTITNGHVIYTPIKRPRLEGND 120
Bhyg_EcRB  ESS-----SEVTSSSAGLVLPGLTMSPSPLDSPDYGELEL 54
Bcop_EcRB  ESS-----SEVTSSSAGLVLPGLTMSPSPLDSPDYGELEL 54
          *                : * . : * . : : : : * :

Bhyg_EcRA1 WLSSSPGSGVPS-----TAPPLSPSPGSGSHNYNTTMSNGYSSPMSNGSYDPYSP 170
Bhyg_EcRA2 WLSSSPGSGVPS-----TAPPLSPSPGSGSHNYNTTMSNGYSSPMSNGSYDPYSP 170
Bhyg_EcRA3 WLSSSPGSGVPS-----TAPPLSPSPGSGSHNYNTTMSNGYSSPMSNGSYDPYSP 170
Bcop_EcRA  WLSSSPGSGVPS-----SAPPLSPSPGSGSHNYNTTMSNGYSSPMSNGSYDPYSP 170
Bhyg_EcRB  WYEDTAYNQHSVIASAVQGCAAVQTTIIP LPS-----MNN----TL SRTPRSTSAN 101
Bcop_EcRB  WYDEAAYNPHSVIASAVQGCAAVQTTIIP LPS-----MNN----ALSHTPRSTSAN 101
          * . . : . : : * * . * . : * . : :

Bhyg_EcRA1 NGKIGREDLSPSSSLNGYSANDAESKAKKGPTP RQQEELCLVCGDRASGYHYNALTCEG 230
Bhyg_EcRA2 NGKIGREDLSPSSSLNGYSANDAESKAKKGPTP RQQEELCLVCGDRASGYHYNALTCEG 230
Bhyg_EcRA3 NGKIGREDLSPSSSLNGYSANDAESKAKKGPTP RQQEELCLVCGDRASGYHYNALTCEG 230
Bcop_EcRA  NGKIGREDLSPSSSLNGYSANDADSKAKKGPTP RQQEELCLVCGDRASGYHYNALTCEG 230

```

```

Bhyg_EcRB      SISSGREDLSPSSSLNGYSANDAESKKAKKGPTPRQQEELCLVCGDRASGYHYNALTCEG 161
Bcop_EcRB      SISSGREDLSPSSSLNGYSANDADSKAKKGPTPRQQEELCLVCGDRASGYHYNALTCEG 161
. . *****:*****

Bhyg_EcRA1     CKGFFRRSVTKNAVYCKFGHECEMDMYMRRKCQECRLKKCLAVGMRPECVVPETQCAIK 290
Bhyg_EcRA2     CKGFFRRSVTKNAVYCKFGHECEMDMYMRRKCQECRLKKCLAVGMRPECVVPETQCAIK 290
Bhyg_EcRA3     CKGFFRRSVTKNAVYCKFGHECEMDMYMRRKCQECRLKKCLAVGMRPECVVPETQCAIK 290
Bcop_EcRA      CKGFFRRSVTKNAVYCKFGHECEMDMYMRRKCQECRLKKCLAVGMRPECVVPETQCAIK 290
Bhyg_EcRB      CKGFFRRSVTKNAVYCKFGHECEMDMYMRRKCQECRLKKCLAVGMRPECVVPETQCAIK 221
Bcop_EcRB      CKGFFRRSVTKNAVYCKFGHECEMDMYMRRKCQECRLKKCLAVGMRPECVVPETQCAIK 221
*****

Bhyg_EcRA1     RKEKKAQKEKDKQIIPSSVVTITDTFKSEIPSLMKCDPPPHAECTQLLSQKLLEENRAK 350
Bhyg_EcRA2     RKEKKAQKEKDKQIIPSSVVTITDTFKSEIPSLMKCDPPPHAECTQLLSQKLLEENRAK 350
Bhyg_EcRA3     RKEKKAQKEKDKQIIPSSVVTITDTFKSEIPSLMKCDPPPHAECTQLLSQKLLEENRAK 350
Bcop_EcRA      RKEKKAQKEKDKQIIPSSVVTITDTYKSEIPSLMKCDPPPHAECTQLLSQKLLEENRAK 350
Bhyg_EcRB      RKEKKAQKEKDKQIIPSSVVTITDTFKSEIPSLMKCDPPPHAECTQLLSQKLLEENRAK 281
Bcop_EcRB      RKEKKAQKEKDKQIIPSSVVTITDTYKSEIPSLMKCDPPPHAECTQLLSQKLLEENRAK 281
*****:*****

Bhyg_EcRA1     NVPALTFNQLAVIYKLIWYQDGYEQPSEEDLKRIMINSAEESDQHDVNFHRHITEITILT 410
Bhyg_EcRA2     NVPALTFNQLAVIYKLIWYQDGYEQPSEEDLKRIMINSAEESDQHDVNFHRHITEITILT 410
Bhyg_EcRA3     NVPALTFNQLAVIYKLIWYQDGYEQPSEEDLKRIMINSAEESDQHDVNFHRHITEITILT 410
Bcop_EcRA      HVPALTNNQLAVIYKLIWYQDGYEQPSEEDLKRIMINSPEDSEDQHDVNFHRHITEITILT 410
Bhyg_EcRB      NVPALTFNQLAVIYKLIWYQDGYEQPSEEDLKRIMINSAEESDQHDVNFHRHITEITILT 341
Bcop_EcRB      HVPALTNNQLAVIYKLIWYQDGYEQPSEEDLKRIMINSPEDSEDQHDVNFHRHITEITILT 341
:***** *.:*****

Bhyg_EcRA1     VQLIVEFAKGLPAFTKIPQEDQITLLKACSEVMMLRMARRYDAESDSILFANNRTYTRD 470
Bhyg_EcRA2     VQLIVEFAKGLPAFTKIPQEDQITLLKACSEVMMLRMARRYDAESDSILFANNRTYTRD 470
Bhyg_EcRA3     VQLIVEFAKGLPAFTKIPQEDQITLLKACSEVMMLRMARRYDAESDSILFANNRTYTRD 470
Bcop_EcRA      VQLIVEFAKGLPAFTKIPQEDQITLLKACSEVMMLRMARRYDAESDSILFANNRTYTRD 470
Bhyg_EcRB      VQLIVEFAKGLPAFTKIPQEDQITLLKACSEVMMLRMARRYDAESDSILFANNRTYTRD 401
Bcop_EcRB      VQLIVEFAKGLPAFTKIPQEDQITLLKACSEVMMLRMARRYDAESDSILFANNRTYTRD 401
*****

Bhyg_EcRA1     SYIMAGMADTIEDLLHFCRQMSMTVDNVEYALLTAIVIFSDRPGLEQADLVQAIQSYII 530
Bhyg_EcRA2     SYIMAGMADTIEDLLHFCRQMSMTVDNVEYALLTAIVIFSDRPGLEQADLVQAIQSYII 530
Bhyg_EcRA3     SYIMAGMADTIEDLLHFCRQMSMTVDNVEYALLTAIVIFSDRPGLEQADLVQAIQSYII 530
Bcop_EcRA      SYIMAGMADTIEDLLHFCRQMSMTVDNVEYALLTAIVIFSDRPGLEQADLVQAIQSYII 530
Bhyg_EcRB      SYIMAGMADTIEDLLHFCRQMSMTVDNVEYALLTAIVIFSDRPGLEQADLVQAIQSYII 461
Bcop_EcRB      SYIMAGMADTIEDLLHFCRQMSMTVDNVEYALLTAIVIFSDRPGLEQADLVQAIQSYII 461
*****

Bhyg_EcRA1     DTLRIYILNRHGGTTKCMILFAKLLSILTELRTLGQNSEMCFSLKLKSKKLPRFLEEIIW 590
Bhyg_EcRA2     DTLRIYILNRHGGTTKCMILFAKLLSILTELRTLGQNSEMCFSLKLKSKKLPRFLEEIIW 590
Bhyg_EcRA3     DTLRIYILNRHGGTTKCMILFAKLLSILTELRTLGQNSEMCFSLKLKSKKLPRFLEEIIW 590
Bcop_EcRA      DTLRIYILNRHGGTTKCMILFAKLLSILTELRTLGQNSEMCFSLKLKSKKLPRFLEEIIW 590
Bhyg_EcRB      DTLRIYILNRHGGTTKCMILFAKLLSILTELRTLGQNSEMCFSLKLKSKKLPRFLEEIIW 521
Bcop_EcRB      DTLRIYILNRHGGTTKCMILFAKLLSILTELRTLGQNSEMCFSLKLKSKKLPRFLEEIIW 521
*****

Bhyg_EcRA1     DVQDIPQTKIANTNGGQSNE 610
Bhyg_EcRA2     DVQDIPQTKIANTNGGQSNE 610
Bhyg_EcRA3     DVQDIPQTKIANTNGGQSNE 610
Bcop_EcRA      DVQDIPQTKVVTNGSQSTE 610
Bhyg_EcRB      DVQDIPQTKIANTNGGQSNE 541
Bcop_EcRB      DVQDIPQTKVVTNGSQSTE 541
*****:..**.*

```

Figura 4. Alinhamento entre sequências polipeptídicas preditas de EcRA1, A2 e A3 (*B. hygida*), EcRA (*B. coprophila*) EcRX1 (*A. aegypti*), EcRB (*B. hygida*), EcRB (*B. coprophila*) e EcRX2 (*A. aegypti*). A busca pelas sequências no genoma de *B. hygida* utilizou como ponto de partida sequências de *B. coprophila* e sequências do transcriptoma de embriões de *B. hygida*. As sequências polipeptídicas possuem 610, 610, 610, 610, 776, 541, 541 e 675 aminoácidos, respectivamente. Os domínios conservados DBD e LDB estão marcados em amarelo e azul, respectivamente. Os asteriscos abaixo do alinhamento

indicam que o resíduo de aminoácido é conservado completamente. Os dois pontos abaixo do alinhamento significam que os resíduos de aminoácidos fazem parte de um grupo de propriedade altamente similar. O ponto abaixo do alinhamento significa que os resíduos de aminoácidos fazem parte de um grupo com propriedades fracamente similares.

1.2.1. Gene *E75*

No intuito de identificar sequências que correspondem ao gene *E75* em *B. hygida* foram necessárias estratégias *in silico* que buscaram no genoma e no transcriptoma de *B. hygida* estas sequências. Inicialmente, sequências identificadas como sendo do gene *E75* vindas de um transcriptoma *de novo* do organismo *Bradysia odoriphaga* e das 7 isoformas de *D. melanogaster* disponíveis em banco de dados foram selecionadas e utilizadas para alinhar com o genoma de *B. hygida*. O resultado deste alinhamento localizou dois segmentos genômicos (ou *scaffolds*), denominados scf107168 e scf110038 (total de aproximadamente 220 kb). O alinhamento identificou no genoma de *B. hygida* 3 transcritos com similaridade com as sequências de *B. odoriphaga*. Com este segmento genômico foi feito um outro alinhamento com o transcriptoma de embriões de *B. hygida*. A partir destes dois alinhamentos foram identificados 5 transcritos advindos do transcriptoma de *B. hygida*. Em seguida, estas 5 sequências foram identificadas como sendo isoformas A, B, C, D e uma não identificada e sua estrutura gênica (figura 5) foi elaborada, baseada no alinhamento realizado com o pacote est2genome presente na plataforma EMBOSS (ver tabelas 10 a 15).

Os transcritos de *E75* identificados em *B. hygida* quando comparados com sequências do mesmo gene em *A. aegypti* são notadamente menores em tamanho. Somado a isso, não foi possível identificar nestes transcritos de *B. hygida* sequências do códon de parada (com exceção da isoforma sem identificação), o que sugere que os transcritos identificados no transcriptoma em *B. hygida* não estão com sua estrutura gênica completa. No entanto, ainda é possível identificar os domínios conservados de DBD e LBD nas 3 isoformas de *E75* em *B. hygida* (figura 6).

Em *D. melanogaster* são descritos 7 transcritos e 5 polipeptídeos únicos para o gene *E75*. A sua expressão atinge todos os estágios de vida do inseto, sendo que os picos de expressão são maiores em estágios larvais tardios e em pré-pupas (Flybase.org). Em particular, a isoforma *E75B* possui expressão que coincide com o pulso de ecdisona antes da formação da pupa e também durante um novo pico do hormônio 10 horas após a formação da pupa (Ihry; Bashirullah, 2014; Ritter; Beckstead, 2010).

Tabela 8. Mapeamento genômico da isoforma A do gene *E75* de *B. hygida*.

E75A <i>Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	DN28291_c0_g1_i2 (start)	DN28291_c0_g1_i2 (end)	exon_cDNA
exon 1	1259	6362	7620	1259	1	1259	1259
intron 1		7621	13780	6160			
exon 2	1617	13781	15397	1617	1260	2876	1617
intron 2		15398	131840	116443			
exon 3	89	131841	131929	89	2877	2965	89
intron 3		131930	166666	34737			
exon 4	145	166667	166811	145	2966	3110	145
intron 4		166812	181886	15075			

exon 5	236	181887	182122	236	3111	3346	236
intron 5		182123	190980	8858			
exon 6	1309	190981	192289	1309	3347	4655	1309

A coluna *E75A Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B. hygida*), à direita uma coluna dupla consistindo das coordenadas do scaffold em que o éxon e íntron se iniciam e terminam e que foi utilizado para a predição dos éxons e íntrons, já a coluna dupla com o transcrito (*DN28291_c0_g1_i2*) mostra as diferentes regiões do transcrito de *E75A* identificado no transcriptoma de *B. hygida*.

Tabela 9. Mapeamento genômico da isoforma B do gene *E75* de *B. hygida*.

<i>E75B Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	<i>DN28291_c0_g1_i1 (start)</i>	<i>DN28291_c0_g1_i1 (end)</i>	exon_cDNA
exon 1	1287	164737	166023	1287	1	1287	1287
intron 1		166024	166666	643			
exon 2	145	166667	166811	145	1288	1432	145
intron 2		166812	181886	15075			
exon 3	236	181887	182122	236	1433	1668	236
intron 3		182123	190980	8858			
exon 4	1309	190981	192289	1309	1669	2977	1309

A coluna *E75B Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B. hygida*), à direita uma coluna dupla consistindo das coordenadas do scaffold em que o éxon e íntron se iniciam e terminam e que foi utilizado para a predição dos éxons e íntrons, já a coluna dupla com o transcrito (*DN28291_c0_g1_i1*) mostra as diferentes regiões do transcrito de *E75B* identificado no transcriptoma de *B. hygida*.

Tabela 10. Mapeamento genômico da isoforma C do gene *E75* de *B. hygida*.

<i>E75C Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	<i>DN28291_c0_g1_i3 (start)</i>	<i>DN28291_c0_g1_i3 (end)</i>	exon_cDNA
exon 1	597	127047	127643	597	1	597	597
intron 1	-20	127644	131840	4197			
exon 2	89	131841	131929	89	598	686	89
intron 2	-20	131930	166666	34737			
exon 3	145	166667	166811	145	687	831	145
intron 3	-20	166812	181886	15075			
exon 4	236	181887	182122	236	832	1067	236
intron 4	-20	182123	190980	8858			
exon 5	1309	190981	192289	1309	1068	2376	1309

A coluna *E75C Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B. hygida*), à direita uma coluna dupla consistindo das coordenadas do scaffold em que o éxon e íntron se iniciam e terminam e que foi utilizado para a predição dos éxons e íntrons, já a coluna dupla com o transcrito (*DN28291_c0_g1_i3*) mostra as diferentes regiões do transcrito de *E75C* identificado no transcriptoma de *B. hygida*.

Tabela 11. Mapeamento genômico da isoforma D do gene *E75* de *B. hygida*.

<i>E75D Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	<i>DN28291_c0_g2_i1 (start)</i>	<i>DN28291_c0_g2_i1 (end)</i>	exon_cDNA
exon 1	736	177783	178518	736	1	736	736
intron 1		178519	181886	3368			
exon 2	236	181887	182122	236	737	972	236
intron 2		182123	190980	8858			
exon 3	1309	190981	192289	1309	973	2281	1309

A coluna *E75D Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B.hygida*), à direita uma coluna dupla consistindo das coordenadas do scaffold em que o éxon e íntron se iniciam e terminam e que foi utilizado para a predição dos éxons e íntrons, já a coluna dupla com o transcrito (*DN28291_c0_g2_i1*) mostra as diferentes regiões do transcrito de *E75D* identificado no transcriptoma de *B. hygida*.

Tabela 12. Mapeamento genômico de isoforma indefinida do gene *E75* de *B. hygida*.

<i>E75 Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	<i>DN28477_c8_g1_i1 (start)</i>	<i>DN28477_c8_g1_i1 (end)</i>	exon_cDNA
exon 1	1771	192163	193938	1776	1	1775	1775

A coluna *E75 Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B.hygida*), à direita uma coluna dupla consistindo das coordenadas do scaffold em que o éxon e íntron se iniciam e terminam e que foi utilizado para a predição dos éxons e íntrons, já a coluna dupla com o transcrito (*DN28477_c8_g1_i1*) mostra as diferentes regiões do transcrito de *E75* identificado no transcriptoma de *B. hygida*.

Figura 5. Representação gráfica da estrutura gênica de *E75* em *B. hygida*. Na parte superior da figura está representada a junção artificial dos scaffolds *scf110038* e *scf107168*. Na parte inferior da figura estão representadas as estruturas das isoformas E75A, E75B, E75C, E75D e uma isoforma sem identificação de *B. hygida* e seu posicionamento em relação aos segmentos genômicos. Os retângulos cinzas representam as regiões 5' e 3' UTR, os retângulos pretos representam os éxons comuns as isoformas representadas, e os retângulos coloridos representam éxons únicos a cada isoforma de E75.

```

Bhyg_E75A MQCIDGKMI-----KEEHMECDIRNEARTAVIS----EQK--VILTPR 37
Bhyg_E75B MGCAMQELQQQTNEP-----IRMDLSGCYSRVNGPSSPIDLKASQQKNSQLKESH 50
Bhyg_E75C -----
Bhyg_E75D -----
Cqui_E75B MGCAVQQEPAASDPKSTATLNAGKSLIVQSNQCSLKSSSES-----NSTDGSSNSLKKAH 55

Bhyg_E75A VEIHRVMKAPPVHESESLPPPQIIFTSRSSN-HQMPPQ-----IPQQRVQVIKDGRCY 90
Bhyg_E75B SALVKILESAPIQRL---D-KEMILTGGSDGSGNVAPITDNRHHHYRKRKNKHLVDMNCD 106
Bhyg_E75C -----MLMSPDSSSRGSFNQM-----PS-TVIRVQVQECS 29
Bhyg_E75D -----
Cqui_E75B SALVKILESAPLKCSNLSKIE--TENGGAGTSSASHSNNQVVGVSNEVTKVCGFKSAD 113

Bhyg_E75A EESPSRQ--YINHTNSDITIGVVIQPTHSSQVTEPKKFVLSRQVNVISNVDRDSNGTVFR 148
Bhyg_E75B NHDKSSE--DEGSSN---VE-QICPWKKTRIAREWRQKNKEMQEALERLEEKNKMEFD 159
Bhyg_E75C QSQQSS-----APSTYMTSSNVIFLNKTNSGGNRIKIE 62
Bhyg_E75D -----
Cqui_E75B SGRPSNQIESPPTVKMKSTPKIDFCPWKKTTIAKEWISACDQELQKHAL-ESSAQ---SK 169

Bhyg_E75A P---PVVSSS-----STP---MRPPPPPPPKNGIAPEEPSSSI---PDLEF----- 187

```



```

Bhyg_E75A LYDTNSLMDEAYKPHKKFRAFRHRESDAEAEVLLPQQ-----HS 749
Bhyg_E75B LYDTNSLMDEAYKPHKKFRAFRHRESDAEAEVLLPQQ-----HS 750
Bhyg_E75C LYDTNSLMDEAYKPHKKFRAFRHRESDAEAEVLLPQQ-----HS 777
Bhyg_E75D LYDTNSLMDEAYKPHKKFRAFRHRESDAEAEVLLPQQ-----HS 507
Cqui_E75B LYDTNSLMDEAYKPHKKFRAMRHRESEAEAPSSSSSSTITLSSHSPKPSSSVSVSSQH 816
*****:*****:** * .. :

Bhyg_E75A QPPPPPPQQQQQQQQSQLHMHLTRPQQQSS----- 778
Bhyg_E75B QPPPPPPQQQQQQQQSQLHMHLTRPQQQSS----- 779
Bhyg_E75C QPPPPPPQQQQQQQQSQLHMHLTRPQQQSS----- 706
Bhyg_E75D QPPPPPPQQQQQQQQSQLHMHLTRPQQQSS----- 536
Cqui_E75B SSAGPGQSSPQQPQSQLHMHLTRPLHQSPHQLHHHQHQAATPPNSHQSSSLSTHSVLA 876
. * *.. ** ***** :**

Bhyg_E75A -----
Bhyg_E75B -----
Bhyg_E75C -----
Bhyg_E75D -----
Cqui_E75B KSLMEQPTMTPEQMKRSDIIHNYIMRESQHHPDQGSYRSGLLVCSPRSSSVSPASSSSS 936

Bhyg_E75A -----
Bhyg_E75B -----
Bhyg_E75C -----
Bhyg_E75D -----
Cqui_E75B SSSSSSSSGNGCPSGSGNGSSISNTRWQGGPTSVITTTTRQQTSPSPSLNSTGSADTSSSS 996

Bhyg_E75A -----
Bhyg_E75B -----
Bhyg_E75C -----
Bhyg_E75D -----
Cqui_E75B SSSSSSNSIRYFQSPHSTSI SPPATSV AQASTTTGPIGSPQPVAVAAPTSPRLVELQVD 1056

Bhyg_E75A -----
Bhyg_E75B -----
Bhyg_E75C -----
Bhyg_E75D -----
Cqui_E75B IAGGSTTGATDAPLNL SKKSPSPSPARNCPQGFSSSTSTTSTAATTTIQKVLLEA 1111

```

Figura 6. Alinhamento entre as sequências polipeptídicas de *E75A*, *E75B* e *E75C* e *E75D* de *B. hygida* com as isoforma D de *Culex quinquefasciatus*. Utilizando sequências oriundas do transcriptoma de embriões de *B. hygida*, foi possível identificar sequências do gene *E75* no genoma de *B. hygida*. Os domínios conservados DBD e LDB estão marcados em amarelo e azul, respectivamente. Os asteriscos abaixo do alinhamento significam que o mesmo aminoácido é encontrado nas 5 sequências analisadas. Os asteriscos abaixo do alinhamento indicam que o resíduo de aminoácido é conservado completamente. Os dois pontos abaixo do alinhamento significam que os resíduos de aminoácidos fazem parte de um grupo de propriedade altamente similar. O ponto abaixo do alinhamento significa que os resíduos de aminoácidos fazem parte de um grupo com propriedades fracamente similares.

1.2.2. Gene *Broad-complex*

As 5 isoformas de *Broad complex* identificadas no genoma de *B. hygida* encontram-se contidas em 3 segmentos genômicos de aproximadamente 100 kb de tamanho (figura 7). Os transcritos que representam as isoformas de BrC apresentam uma região comum entre eles (excetuando-se a isoforma E), que são representados pelos éxons 4, 5 e 6 da isoforma A. Esta mesma isoforma provavelmente não está completa em sua região 3' pela ausência da região 3' UTR. No mesmo sentido, as isoformas B, C e D parecem estar incompletas por não possuírem a região 5' UTR. Entretanto algumas regiões estão conservadas nas suas estruturas gênicas como é

observado na figura 8, onde é possível notar no alinhamento a conservação do domínio BTB/POZ e os domínios dedo de zinco entre diferentes insetos como *A. aegypti*, *M. domestica* e *Rhagoletis zephyria*.

Tabela 13. Mapeamento genômico da isoforma A de *Broad-complex* em *B. hygida*.

BrCA <i>Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	DN26132_c0_g8_i1 (start)	DN26132_c0_g8_i1 (end)	exon_cDNA
exon 1	167	52537	52703	167	1	167	167
intron 1	-20	52704	55697	2994			
exon 2	216	55698	55913	216	168	383	216
intron 2	-20	55914	63611	7698			
exon 3	189	63612	63800	189	384	572	189
intron 3	-20	63801	63866	66			
exon 4	142	63867	64008	142	573	714	142
intron 4	-20	64009	65381	1373			
exon 5	264	65382	65645	264	715	978	264
intron 5	-20	65646	65858	213			
exon 6	235	65859	66093	235	979	1213	235
intron 6	-20	66094	66169	76			
exon 7	147	66170	66316	147	1214	1360	147

A coluna *BrCA Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *tamanho (pb)* mostra o tamanho em pares de bases dos éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B.hygida*), já a coluna dupla com o transcrito (*DN26132_c0_g8_i1*) mostra as diferentes regiões do transcrito de *Broad* identificado no transcriptoma de *B. hygida*.

Tabela 14. Mapeamento genômico da isoforma B de *Broad-complex* em *B. hygida*.

BrCB <i>Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	DN26132_c0_g5_i1 (start)	DN26132_c0_g5_i1 (end)	exon_cDNA
exon 1	256	65390	65645	256	1	256	256
intron 1	-20	65646	65858	213			
exon 2	235	65859	66093	235	257	491	235
intron 2	-20	66094	66169	76			
exon 3	146	66170	66315	146	492	637	146
intron 3	-20	66316	68726	2411			
exon 4	51	68727	68777	51	638	688	51
intron 4	-20	68778	76726	7949			
exon 5	807	76727	77533	807	689	1495	807

A coluna *BrCB Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *tamanho (pb)* mostra o tamanho em pares de bases dos éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B.hygida*), já a coluna dupla com o transcrito (*DN26132_c0_g5_i1*) mostra as diferentes regiões do transcrito de *Broad* identificado no transcriptoma de *B. hygida*.

Tabela 15. Mapeamento genômico da isoforma C de *Broad-complex* em *B. hygida*.

BrCC <i>Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	DN26132_c0_g6_i1 (start)	DN26132_c0_g6_i1 (end)	exon_cDNA
exon 1	256	65390	65645	256	1	256	256
intron 1	-20	65646	65858	213			
exon 2	235	65859	66093	235	257	491	235
intron 2	-20	66094	66169	76			
exon 3	146	66170	66315	146	492	637	146
intron 3	-20	66316	68726	2411			
exon 4	51	68727	68777	51	638	688	51
intron 5	-20	68778	96408	27631			
exon 5	446	96409	96854	446	689	1134	446

A coluna *BrCC Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *tamanho (pb)* mostra o tamanho em pares de bases dos éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B.hygida*), já a coluna dupla com o transcrito (*DN26132_c0_g6_i1*) mostra as diferentes regiões do transcrito de *Broad* identificado no transcriptoma de *B. hygida*.

Tabela 16. Mapeamento genômico da isoforma *C* de *Broad-complex* em *B. hygida*.

BrCD <i>Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	DN26132_c0_g1_i1 (start)	DN26132_c0_g1_i1 (end)	exon_cDNA
exon 1	256	65390	65645	256	1	256	256
intron 1	-20	65646	65858	213			
exon 2	235	65859	66093	235	257	491	235
intron 2	-20	66094	66169	76			
exon 3	146	66170	66315	146	492	637	146
intron 3	-20	66316	68726	2411			
exon 4	51	68727	68777	51	638	688	51

A coluna *BrCD Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *tamanho (pb)* mostra o tamanho em pares de bases dos éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequências alinhadas (genoma x transcrito de *B.hygida*), já a coluna dupla com o transcrito (*DN26132_c0_g1_i1*) mostra as diferentes regiões do transcrito de *Broad* identificado no transcriptoma de *B. hygida*.

Tabela 17. Mapeamento genômico da isoforma *C* de *Broad-complex* em *B. hygida*.

BrCE <i>Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	DN49472_c0_g1_i1 (start)	DN49472_c0_g1_i1 (end)	exon_cDNA
exon 1	350	82368	82717	350	1	350	350

A coluna *Z4C Bhygida*, mostra o número de éxons e íntrons preditos, a coluna *tamanho (pb)* mostra o tamanho em pares de bases dos éxons e íntrons preditos, a coluna *coordenada scaffolds* mostra o posicionamento dos éxons e íntrons dentro de um único scaffold, que foi utilizado para a predição dos éxons e íntrons, a coluna *Z4C Bodoriphaga* mostra os éxons da isoforma *Z4C* de *B. odoriphaga* e a coluna *tamanho (pb)* mostra o tamanho de cada um destes éxons.

Figura 7. Representação gráfica da estrutura gênica de *Broad* em *B. hygida*. Na parte superior da figura estão representados *scaffold* scf97767 e scf96698. O *scaffold* scf97767 apresentou similaridade no início de duas sequências de aproximadamente 100 pb de *B. odoriphaga*, o que pode indicar outra isoforma a montante do scf96698. A região de junção entre os dois *scaffolds* está interrompida pois não dispomos de dados no momento acerca da extensão desta região. Na parte inferior da figura estão representadas as estruturas das isoformas Z1B, Z1D e Z4C de *B. hygida* e seu posicionamento em relação ao *scaffold* scf96698. Os retângulos negros representam os éxons e as linhas representam os íntrons.

De acordo com os dados apresentados nas tabelas 16 a 20 podemos afirmar que o gene *Broad* encontra-se em um fragmento genômico é de no mínimo 44318 pb. As sequências das isoformas de *B. hygida* apresentaram alta identidade com as isoformas de *Aedes aegypti*, em especial com isoforma C.

Bhyg_BrCA	MGDTQHFCLRWNNYQSSITSAF	ENLRDDEDFVDVTLACDGRSLKAHRVLSACSPYFREL	60
Aaeg_BrCZ2	MVDTQHFCLRWNNYQSSITSAF	ENLRDDEDFVDVTLACDGRSLKAHRVLSACSPYFREL	60
Bhyg_BrCB	-----	-----	-----
Aaeg_BrCZ1	MVDTQHFCLRWNNYQSSITSAF	ENLRDDEDFVDVTLACDGRSLKAHRVLSACSPYFREL	60
Bhyg_BrCC	-----	-----	-----
Mdom_BrCZ2	MDETQHFCLRWNNYQSSITSAF	ENLRDDEDFVDVTLACEGRSLKAHRVLSACSPYFRDL	60
Bhyg_BrCD	-----	-----	-----
Rzep_BrCX2	MDETQHFCLRWNNYQSSITSAF	ENLRDDEDFVDVTLACEGRSLKAHRVLSACSPYFRDL	60
Bhyg_BrCA	LKSTPCKHPVIVLQDVAFTDLHALVEFIYHGEVNVHQRSLSSFLKTAEVLRVSGLTQQQA		120
Aaeg_BrCZ2	LKSTPCKHPVIVLQDVAFTDLHALVEFIYHGEVNVHQRSLSSFLKTAEVLRVSGLTQQQA		120
Bhyg_BrCB	-----	-----	-----
Aaeg_BrCZ1	LKSTPCKHPVIVLQDVAFTDLHALVEFIYHGEVNVHQRSLSSFLKTAEVLRVSGLTQQQA		120

```

Bhyg_BrCC -----
Mdom_BrCZ2 LKSTPCKHPVILLQDVNFNDLHSLVEFIYHGEVNVHQKSLQSFLKTAEVLRVSGLTQQQA 120
Bhyg_BrCD -----
Rzep_BrCX2 LKSTPCKHPVILLQDVNFNDLHSLVEFIYHGEVNVHQKSLQSFLKTAEVLRVSGLTQQQA 120

Bhyg_BrCA EDTHN-LAQIQGLANAG--ARTPINPHQSFTDKMEE-ALFPP--GASPP-HINN-LHTPN 172
Aaeg_BrCZ2 DETHN-LAQVQTSTGSG--SRTPVNHHPSPDKFVEDALFAQ--GVPPSSHLNNLHHTPA 175
Bhyg_BrCB -----
Aaeg_BrCZ1 DETHN-LAQVQTSTGSG--SRTPVNHHPSPDKFVEDALFAQ--GVPPSSHLNNLHHTPA 175
Bhyg_BrCC -----
Mdom_BrCZ2 EETHSQLAQIQNLANS--VRTPLNPLHTSMS--EDSPYQRNAVSPPPPLPPTQ----- 171
Bhyg_BrCD -----
Rzep_BrCX2 EETHSQLAHIQNLANAANATRTPHNAHHIAPH--DEPALYSRSSGSPPPPPPTQ--LPL 176

Bhyg_BrCA AHSGATVNQLLRRAAAAAAMRRERNSSSHSDELS--LKRHR--PDNGIPGNNNS-PDV-- 225
Aaeg_BrCZ2 AHGGPTVNQLLRRAAAAAALRRERNNSNQSDLS--LKRHRMSTDNNILGNNNN-LDV-- 230
Bhyg_BrCB -----MRRERNSSSHSDELS--LKRHR--PDNGIPGNNNS-PDV-- 34
Aaeg_BrCZ1 AHGGPTVNQLLRRAAAAAALRRERNNSNQSDLS--LKRHRMSTDNNILGNNNN-LDV-- 230
Bhyg_BrCC -----MRRERNSSSHSDELS--LKRHR--PDNGIPGNNNS-PDV-- 34
Mdom_BrCZ2 -----MPFFKR---IGMLRREHNNSTSSAEDPTNTLKRMRGPDNGLPPSNTNSPDIVH 221
Bhyg_BrCD -----MRRERNSSSHSDELS--LKRHR--PDNGIPGNNNS-PDV-- 34
Rzep_BrCX2 A-STHINSQLFKRS--MAMLRERHNSTPSAEDSNAMKRLRGPDNGLPSSSNNSPDIVH 233
               :***:***. * *           ::   **.:   ....   ::

Bhyg_BrCA -VNHMPQVTATDFSTIKHNNNTTP---IKDSERIENH----- 259
Aaeg_BrCZ2 -VSHMPQITATDFSTNMKTNNNSAPHSPHIKEQHRHESNKTPNGFTFIATGNNNNSININN 289
Bhyg_BrCB -VNHMPQVTATDFSTIKHNNNTTP---IKDSERIENH----- 68
Aaeg_BrCZ1 -VSHMPQITATDFSTNMKTNNNSAPHSPHIKEQHRHESNKTPNG-----NNNNSININN 282
Bhyg_BrCC -VNHMPQVTATDFSTIKHNNNTTP---IKDSERIENH----- 68
Mdom_BrCZ2 PRSTSPQLTPTDFSTMKHNNNNSPPK---DEKRNGSSTNATGN----- 262
Bhyg_BrCD -VNHMPQVTATDFSTIKHNNNTTP---IKDSERIENH----- 68
Rzep_BrCX2 PRSSSPQLTPADDFSTIKHNNNTPLK---DEKRNGPS-----G----- 269
               .   **:* :***** :.***:*           :..*

Bhyg_BrCA -----SANN-----NNNHGSNGNGLSSSLTDKESIT 286
Aaeg_BrCZ2 NTNNLNNNNNNNNNNNNLNTHSKDSSNVGDQPSSNLAS IATNGGNGGSLGSHTDKDSL 349
Bhyg_BrCB -----SANN-----NNNHGSNGNGLSSSLTDKESIT 95
Aaeg_BrCZ1 NTNNLNNNNNNNNNNNNLNTHSKDSSNVGDQPSSNLAS IATNGGNGGSLGSHTDKDSL 342
Bhyg_BrCC -----SANN-----NNNHGSNGNGLSSSLTDKESIT 95
Mdom_BrCZ2 -----G-----GSGSNSSANGNGI-TAGEKGESLT 286
Bhyg_BrCD -----SANN-----NNNHGSNGNGLSSSLTDKESIT 95
Rzep_BrCX2 -----NG-----NNGNSDTGNGNGNANS DKVGGSLT 295
               .           .   .**.*           :   **:*

Bhyg_BrCA PSPPNRLNDDVKSEPMELVCGNNNL-PLDEHSNDSSGENDLKYS G-MGDGKGSLSNNDD 344
Aaeg_BrCZ2 PSPSVRSNEDVKSEPMELVCSSNVE-N--DNSSDSIEDDHDFPH-PADIKGPISSPHEN 405
Bhyg_BrCB PSPPNRLNDDVKSEPMELVCGNNNL-PLDEHSNDSSGENDLKYS G-MGDGKGSLSNNDD 153
Aaeg_BrCZ1 PSPSVRSNEDVKSEPMELVCSSNVE-N--DNSSDSIEDDHHDIPH-PADIKGPISSPHED 398
Bhyg_BrCC PSPPNRLNDDVKSEPMELVCGNNNL-PLDEHSNDSSGENDLKYS G-MGDGKGSLSNNDD 153
Mdom_BrCZ2 PSPLTRRAEDVKSEPMELVCSNNND---EHSNDSNGENEQLRVGSVDCGKGSISSVNDE 342
Bhyg_BrCD PSPPNRLNDDVKSEPMELVCGNNNL-PLDEHSNDSSGENDLKYS G-MGDGKGSLSNNDD 153
Rzep_BrCX2 ASPLTRTADVKSEPMELVCSNNNPTAVDEHSNDSTGEHEANRSISGEGGKGSLSGND 355
               ** * :*****.*           :*.**   :..           ** :** :

Bhyg_BrCA DIDNSIHSHA-PPFLMTS-ENKLFSTPGSFNFSMAALADPAALA----- 387
Aaeg_BrCZ2 EMDGSIHSHA-PPFLISPGDQKLFAGPGSFNFSMAALADPATLAGFNPQALQAAELAGS 464
Bhyg_BrCB DIDNSIHSHA-PPFLMTS-ENKLFSTPGSFNFSMAALADPAALAGFNSQLQSAADLAGS 211
Aaeg_BrCZ1 EMDGSIHSHA-PPFLISPGDQKLFAGPGSFNFSMAALADPATLAGFNPQALQAAELAGS 457
Bhyg_BrCC DIDNSIHSHA-PPFLMTS-ENKLFSTPGSFNFSMAALADPAALAGFNSQLQSAADLAGS 211
Mdom_BrCZ2 EIENTMPPHPPSAPFLISPAESKLF-PSAPFTFPMGNL-DPAAFASLNSQMPSAADIAAS 400
Bhyg_BrCD DIDNSIHSHA-PPFLMTS-ENKLFSTPGSFNFSMAALADPAALAGFNSQLQSAADLAGS 211
Rzep_BrCX2 EIDNNMQQHPP-PPYLMSPAESKLF-PPGPFNFPM SGL-DPAALAGFNSQLQSAAE LSVS 412
               ::::.. *   *::: :.***   . *.* * . * **::*

Bhyg_BrCA -----
Aaeg_BrCZ2 PQGCSPK----- 471

```

Bhyg_BrCB	PQGHLLTATPP-----	222
Aaeg_BrCZ1	PQGE-----	461
Bhyg_BrCC	PQGPQSAVVSASS-----S-----	225
Mdom_BrCZ2	PQGSHNAMRSTASSPTSSTAPPSPPTAVTSPNSSLG-----M-----	439
Bhyg_BrCD	PQEFSIRPQVV-----	222
Rzep_BrCX2	PQGSSICGLNLTFPGNCNASNSGSGGANCSNNSAGGNTSSSGANTSSLLPAALMQANSQ	472
Bhyg_BrCA	-----	
Aaeg_BrCZ2	-----	
Bhyg_BrCB	-----HSHM-----	226
Aaeg_BrCZ1	-----	
Bhyg_BrCC	-----QLRM-----	229
Mdom_BrCZ2	----VGHV-----NNG-----GV-PFNMHASP-	456
Bhyg_BrCD	-----KNRF-----	226
Rzep_BrCX2	NQHQQHQQHQQHHHPHHQQQHGQNVSSGAGNGCNSNNANLSNANSAAVQQSQMFAAAA	532
Bhyg_BrCA	-----	
Aaeg_BrCZ2	-----KLFSCQLCGKVLCSKASL	489
Bhyg_BrCB	-----M-----PPGSGSSSVGIGEDFRCDPCNKNLSSLTRL	257
Aaeg_BrCZ1	-----DFRCDPCNKNLSSLTRL	478
Bhyg_BrCC	-----P-----PPT----SGGINEPQECPYCRRTFSCYYSL	256
Mdom_BrCZ2	-----PRPGSGGAAGTAPANNNNSTPQMPLRMPPPTSGGINEPQECPYCRRTFSCYYSL	510
Bhyg_BrCD	-----	
Rzep_BrCX2	AAAMANMAAASSGTTGVAAGNAH-NTSGGSLTTAGSGGAGGGDDFRCEPCNKNLSSLTRL	591
Bhyg_BrCA	-----	
Aaeg_BrCZ2	KRHIADKHAERQEEYRCIICERVYCSRNSLMTHTIYTHKSRPGEMGMKDI-----	539
Bhyg_BrCB	KRHIQNVHMRPTKEPVCNICKRVYSSLNSLRNHKSIYHRNMKQPKEEH-----QSP	308
Aaeg_BrCZ1	KRHIQNVHMRPTKEPICNICKRVYSSLNSLRNHKSIYHRNMKQPKEEHSAPSA--IPT	536
Bhyg_BrCC	KRHFQDKHEQSDTLVYCEFCRRRYRTKNSLTTHKSLQHRGSSGMLKRL-LKTS--IKNV	313
Mdom_BrCZ2	KRHFQDKHEQSDTLVYCEFCRRRYRTKNSLTTHKSLQHRGSSGMLKRLKTSKMNVLVPP	570
Bhyg_BrCD	-----	
Rzep_BrCX2	KRHIQNVHMRPTKEPVCNICKRVYSSLNSLRNHKSIYHRNLKQPKQEIGSGADVTQSPGM	651
Bhyg_BrCA	-----	
Aaeg_BrCZ2	-----KFF-----	542
Bhyg_BrCB	----LGP---SS---HP---YYPINHRMS-----	324
Aaeg_BrCZ1	----QPSMTSMSSAVDQKHQYFSLSHHTA-----	561
Bhyg_BrCC	----LGS---HHGSGHPRPHLFDFAAELGQPPVPIPYKKEHLA	349
Mdom_BrCZ2	PHHHHHHPHHHPAITHQRTPLFDSSSELGQPPPGIQ-----	607
Bhyg_BrCD	-----	
Rzep_BrCX2	----AGVAVMHN-----PYSSSVNHRSSS-----	671
Bhyg_BrCE	-----MVSNNNSNPSIGRLSLPLPLTACHRCDVCGKLLSTKLTCLKRHKEQQH	48
Aalb_BrC	MIAGMSTGRPSSPDNLHSGGGGANRLALPLPLTACHRCDVCGKLLSTKLTCLKRHKEQQH	60
	. *:... . **:*****	
Bhyg_BrCE	LQPLNNAVCNLCHKVFRTLSLNNHKSIIYHRRQKH-----	83
Aalb_BrC	LQPLNNAVCNLCHKVFRTLSLNNHKSIIYHRRQKLLHHHHQHHAQHPPASVAAVAAAGA	120

Bhyg_BrCE	-----	
Aalb_BrC	ASPLHHPSQHALLGDASGAGCDMKHNIKLEFL	153

Figura 8. Alinhamento entre sequências polipeptídicas preditas das isoformas de Broad-complex de *B. hygida*, utilizando sequencias oriundas do transcriptoma de *B. odoriphaga* e *B. hygida*, além da sequência de Broad-complex C de *Aedes albopictus*. Os domínios conservados

BTB/POZ e os dedos de zinco estão marcados em azul e amarelo, respectivamente (não foi possível prever domínios na isoforma D). Os asteriscos abaixo do alinhamento significam que o mesmo aminoácido é encontrado nas 5 sequências analisadas. Os asteriscos abaixo do alinhamento indicam que o resíduo de aminoácido é conservado completamente. Os dois pontos abaixo do alinhamento significam que os resíduos de aminoácidos fazem parte de um grupo de propriedade altamente similar. O ponto abaixo do alinhamento significa que os resíduos de aminoácidos fazem parte de um grupo com propriedades fracamente similares.

1.2.3. Gene *E74*

O gene *E74* foi identificado no genoma utilizando-se sequências oriundas de diferentes isoformas de insetos como *D. melanogaster* e *A. aegypti*. A partir daí foram determinados segmentos genômicos que serviriam de ponto de partida para buscar no transcriptoma de *B. hygida* sequências dentro do genoma compatíveis com a estrutura do gene *E74*. Foram identificadas duas isoformas possíveis (A e B) que estão representadas nas tabelas 18 e 19 com seus éxons e íntrons. A isoforma *E74A* apresenta 7 éxons enquanto que a isoforma *E74B* apresenta somente 4 éxons. A estrutura gênica (figura 8) nos permite visualizar que o gene foi identificado em 2 segmentos genômicos, que juntos apresentam um tamanho aproximado de 320 kb, enquanto que a região que está contida o gene é de aproximadamente 22 kb.

O alinhamento entre os polipeptídeos de *E74* com os polipeptídeos de *E74* de *Lucilia cuprina* é possível notar que todas as isoformas de *B. hygida* conservam a região do domínio ETS (figura 9). O gene de *E74* de *B. hygida* e suas sequências de cDNA correspondentes sugerem que promotores alternativos são os responsáveis pela formação das duas formas variantes A e B. Estes resultados se assemelham com os observados em outros insetos como em *D. melanogaster* (Burtis et al., 1990; Thummel, 2002), *A. aegypti* (Sun et al., 2004); *Manduca sexta* (Stilwell et al., 2003) e *Leptinotarsa decemlineata* (Xu et al., 2018).

Tabela 18. Mapeamento genômico da isoforma A de *E74* em *B. hygida*.

<i>E74A</i> <i>Bhygida</i>	hit number	scf (start)	scf (end)	exon/intron size	<i>DN26157_c0_g1_i1</i> (start)	<i>DN26157_c0_g1_i1</i> (end)	exon_cDNA
exon 1	48	39175	39280	106	570	680	111
intron 1		39281	64095	24815			
exon 2	509	64096	64604	509	681	1189	509
intron 2		64605	67168	2564			
exon 3	174	67169	67342	174	1190	1363	174
intron 3		67343	67401	59			
exon 4	283	67402	67684	283	1364	1646	283
intron 4		67685	219438	151754			
exon 5	196	219439	219640	202	1647	1849	203
intron 5		219641	234254	14614			
exon 6	204	234255	234473	219	1850	2063	214
intron 6		234474	258070	23597			
exon 7	2730	258071	260800	2730	2064	4793	2730

A coluna *E74A Bhygida*, mostra o número de éxons e íntrons preditos, a coluna dupla *scf* mostra o tamanho em pares de bases dos éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequencias alinhadas (genoma x transcrito de *B.hygida*), já a coluna dupla com o transcrito mostra as diferentes regiões do transcrito de *E74* identificado no transcriptoma de *B. hygida*.

Tabela 19. Mapeamento genômico da isoforma *B* de *E74* em *B. hygida*.

E74B	hit	scf	scf	exon/intron	DN26157_c0_g1_i2	DN26157_c0_g1_i2	exon_cDNA
<i>Bhygida</i>	number	(start)	(end)	size	(start)	(end)	
exon 1	888	169110	169997	888	1	888	888
intron 1	-20	169998	219438	49441			
exon 2	196	219439	219640	202	889	1091	203
intron 2	-20	219641	234254	14614			
exon 3	204	234255	234473	219	1092	1305	214
intron 3	-20	234474	258070	23597			
exon 4	2730	258071	260800	2730	1306	4035	2730

A coluna *E74B Bhygida*, mostra o número de éxons e íntrons preditos, a coluna dupla *scf* mostra o tamanho em pares de bases dos éxons e íntrons preditos, a coluna *hit number* mostra o número correspondente de nucleotídeos alinhados entre cada par de sequencias alinhadas (genoma x transcrito de *B.hygida*), já a coluna dupla com o transcrito mostra as diferentes regiões do transcrito de *E74* identificado no transcriptoma de *B. hygida*.

Figura 8. Representação gráfica da estrutura gênica de *E74* em *B. hygida*. Na parte superior da figura estão representados os *scaffolds* em que foram encontradas as duas isoformas de *E74*, dando um total de aproximadamente de 320 kb. A região de junção entre os dois *scaffolds* está interrompida pois não dispomos de dados no momento acerca da extensão desta região. Na parte inferior da figura estão representadas as estruturas das isoformas *A* e *B* de *E74* de *B. hygida*. Os retângulos negros representam os éxons comuns entre as duas isoformas, os retângulos azul e vermelho representam os éxons únicos a cada isoforma, os retângulos cinzas representam as regiões 5' e 3' UTRs.

```
Bhyg_E74A -----MRGGAAIGGAVMPFIDDALLWCPDND 26
Lcup_E74X1 MIMVQHLVAANAHTFVNPTAQSPSPSPANNLQSPPLASPTASISSHPTTASSPLSSPSDD 60
Bhyg_E74B -----MMMVQQAYPYFVGQSPVKCEESE 23
Lcup_E74X2 -----MPFIDDALLWCPDND 15
.: : . . .

Bhyg_E74A GRMV-----DMSA--CLPNAVNGSAPNHQE-----NSGN-----SCDLSSL----- 60
```



```

Lcup_E74X1 RDLNMQKIPCNMDATQCSSSNRSTCSPLQNKVETPEWSQTQSQMSEE-EDDLAQRRSDI 119
Bhyg_E74B AVRI-----HLPKRSLSPAIVYKQPVFT-----QPI----- 50
Lcup_E74X2 G-----RIVGGLDIATCMQDEVNESN--ENVSE-----STQLAESTTAELENALE---- 57
      :
      :

Bhyg_E74A -----DPLC---NDSDELLRQLQENSFELESFFSEFTSN-----VDIKEEN 98
Lcup_E74X1 INQTSAIHKGSLQDEEDQDEELEERISH-GHEYRPQQSTTSPTNPTPSSSPALSANSLN 178
Bhyg_E74B -----SPPS----PRNRLIKRSDS-----AEFYEN-----DVLRSDD 78
Lcup_E74X2 -----PSLCNNDSSD-DLFRQLSESNFEIESLLSD-----LATVEVKEEN 96
      * .: . : . . :

Bhyg_E74A -NNDVANH----- 105
Lcup_E74X1 QHIPSPQHSPTTTANSQDCRQQFTVLYNHLRQ-----HPPSSRLNVLNDH 223
Bhyg_E74B KHYESEPE----- 86
Lcup_E74X2 NNISDV----TSDVVGRSQHDTFVALVNGSGGVSGVVNGMVNGGSGDNVTSSNNLISGN 152
      :

Bhyg_E74A -----SNVQQN-----YIQNDSSII-----Q----QSL 124
Lcup_E74X1 KQPLPQPHPHQKDNNGNEMEQQNRRSVITSREEEDSLSQERSSSENLDKQMEEQPTSDH 283
Bhyg_E74B -----SGRQFE-----YQEN----- 96
Lcup_E74X2 AIQLPGE-----NSALGSV-GTNSLTLEPTAAS----STQEQAANSSS 189

Bhyg_E74A NSNRSAESNHNVLALSTNHHNIRNESR-FS-----AANPLLAEK 163
Lcup_E74X1 FNSDDDEDFDDEVPLDLSLPTKLARNRTRTYSGTESDDSGSGVNDHHRHHEASRLKP-- 341
Bhyg_E74B -PDEDATEDDNAPLDLSLPVGR--RRT-YS-----GAES--DD 128
Lcup_E74X2 FLNQNTNNSNS-----NLSNNN----NNNSNNDSTNIFDDSSSRYLIAANPLLAEK 237
      . . : : . . :

Bhyg_E74A LLSPALSDLDNLTISSGRLGRHRDIKVDTEIPLIKQSTSPPPHQQRIVLQ----- 213
Lcup_E74X1 ----E-----ERKAAYKKS-LMKRYYTEIPIIKQSTSPGPHLQHLPPHHQQQTPHPQQ 389
Bhyg_E74B SGGPN--EGKSVSDKAAYKKS LMKRYLDTEIPLIKQSTSPPPHQQRIVLQ----- 176
Lcup_E74X2 LMSKS-----LILNKNDKT-IADFKDTEIPIIKQSTSPGPHLQHLPPHHQQQTPHPQQ 289
      *****:***** ** *: :

Bhyg_E74A -----IKTEP----- 218
Lcup_E74X1 QQQQPHVTTQQQTTNQTSTFTGSTTVINIKSEQNVLTSPQHQQQQQQQSQLGAVNLQHGS 449
Bhyg_E74B -----IKTEP----- 181
Lcup_E74X2 QQQQPHVTTQQQTTNQTSTFTGSTTVINIKSEQNVLTSPQHQQQQQQQSQLGAVNLQHGS 349
      **.*

Bhyg_E74A -----SSSNGQHISHRPLLHNLLSGAPIHTTPYHRNYCTSSSTGSLPPSPADSGVSDVDSS 273
Lcup_E74X1 LTGPNTNLQSLNIPHRPLLHNLLSGGSLHTT-HHRSYGAGTTGSFPPSPADSGVSDVDSS 508
Bhyg_E74B -----SSSNGQHISHRPLLHNLLSGAPIHTTPYHRNYCTSSSTGSLPPSPADSGVSDVDSS 236
Lcup_E74X2 LTGPNTNLQSLNIPHRPLLHNLLSGGSLHTT-HHRSYGAGTTGSFPPSPADSGVSDVDSS 408
      .: .: .* ***** . : ** : ** . : .: *****:*****

Bhyg_E74A SSGGQPCSEELKARLGLTSHCPQONNQVPPGTFLNPNFYHNSPPIRNLWNNRNVQLPVPE 333
Lcup_E74X1 SSGGQPCSEELKARLGLPPHCNTHTNHLTSGTFLHPNLYQNS-QIRNIWNRG---VGMPE 564
Bhyg_E74B SSGGQPCSEELKARLGLTSHCPQONNQVPPGTFLNPNFYHNSPPIRNLWNNRNVQLPVPE 296
Lcup_E74X2 SSGGQPCSEELKARLGLPPHCNTHTNHLTSGTFLHPNLYQNS-QIRNIWNRG---VGMPE 464
      *****:***** ** .: *: : *****:*****:***** ***:** . : **

Bhyg_E74A NYYPHLMNAGY-----TSHFPTSPPRGTH---- 358
Lcup_E74X1 NYYLPLSTNGGNTSSNGPVGNNNVSCSSTAPNGLTSNGSYTGAPYFTTPSPPRTNVGLAS 624
Bhyg_E74B NYYPHLMNAGY-----TSHFPTSPPRGTH---- 321
Lcup_E74X2 NYYLPLSTNGGNTSSNGPVGNNNVSCSSTAPNGLTSNGSYTGAPYFTTPSPPRTNVGLAS 524
      *** * . * : : * ***** .

Bhyg_E74A -----QNFHQQSVIQAATSSVNEDLTY 380
Lcup_E74X1 HHSLHPHQHALHTHQQHSSVSNTVVGSTSNGLVLTNGVLTHQQSVIQAATSSVNYDLSY 684
Bhyg_E74B -----QNFHQQSVIQAATSSVNEDLTY 343
Lcup_E74X2 HHSLHPHQHALHTHQQHSSVSNTVVGSTSNGLVLTNGVLTHQQSVIQAATSSVNYDLSY 584
      ***** ***** **:*

Bhyg_E74A MLEY-GFPQRKLKKPRKPRMEMGVKRRSREGS TTYLWEFLKLLQDREYCPRFIKWTNRE 439
Lcup_E74X1 MLELGGFQQRKVKKPRKPKIEMGVKRRSREGS TTYLWEFLKLLQDREYCPRFIKWTNRE 744
Bhyg_E74B MLEY-GFPQRKLKKPRKPRMEMGVKRRSREGS TTYLWEFLKLLQDREYCPRFIKWTNRE 402

```

```

Lcup_E74X2 MLELGGFQQRKVKKPRKPKIEMGVKRRSREGS TTYLWEFLLKLLQDREYCPRFIKWTNRE 644
*** ** ***:*****.:*****

Bhyg_E74A KGIFKLVDISKAVSRLWGLHKNKPD MN YETMGRALRYYYQ R GILAKVDGQRLVYQFVDVPK 499
Lcup_E74X1 KGVFKLVDISKAVSRLWGMHKNKPD MN YETMGRALRYYYQ R GILAKVDGQRLVYQFVDVPK 804
Bhyg_E74B KGIFKLVDISKAVSRLWGLHKNKPD MN YETMGRALRYYYQ R GILAKVDGQRLVYQFVDVPK 462
Lcup_E74X2 KGVFKLVDISKAVSRLWGMHKNKPD MN YETMGRALRYYYQ R GILAKVDGQRLVYQFVDVPK 704
**.:*****.:*****

Bhyg_E74A DIIEIDCSGA 509
Lcup_E74X1 DIVEIDCNGV 814
Bhyg_E74B DIIEIDCSGA 472
Lcup_E74X2 DIVEIDCNGV 714
**.:****.*.

```

Figura 9. Alinhamento entre sequências polipeptídicas preditas das isoformas de E74 A e B de *B. hygida*, utilizando sequências oriundas do transcriptoma de *B. odoriphaga* e *B. hygida*, além e as isoformas X1 e X2 de E74 de *Lucilia cuprina*. O domínio conservado ETS está marcado em amarelo. Os asteriscos abaixo do alinhamento significam que o mesmo aminoácido é encontrado nas 4 sequências analisadas. Os asteriscos abaixo do alinhamento indicam que o resíduo de aminoácido é conservado completamente. Os dois pontos abaixo do alinhamento significam que os resíduos de aminoácidos fazem parte de um grupo de propriedade altamente similar. O ponto abaixo do alinhamento significa que os resíduos de aminoácidos fazem parte de um grupo com propriedades fracamente similares.

Proteína	Isoformas	Trinity de B. hygida	Blast hit	Espécie	Blastp max score	Identity	E-value
E74	A	DN26157_c0_g1_i3	XP_023301970.1	<i>Lucilia cuprina</i>	370	63.13%	4e-117
	B	DN26157_c0_g1_i2	XP_023301963.1	<i>Lucilia cuprina</i>	367	63.13%	4e-115
USP	A	DN7476_c0_g2_i1	AFT63499.1	<i>B. coprophila</i>	829	94.31%	0
	B	DN7476_c0_g1_i1	AFT63499.1	<i>B. coprophila</i>	829	94.31%	0
	C	DN10508_c0_g1_i1	AFT63499.1	<i>B. coprophila</i>	89,0	100%	2e-19
E75	A	DN28291_c0_g1_i2	XP_001849917.1	<i>Culex quinquefasciatus</i>	897	86.12%	0
	B	DN28291_c0_g1_i1	XP_001849917.1	<i>Culex quinquefasciatus</i>	944	68.18%	0
	C	DN28291_c0_g1_i3	XP_001849917.1	<i>Culex quinquefasciatus</i>	899	86.12%	0
	D	DN28291_c0_g2_i1	XP_001849917.1	<i>Culex quinquefasciatus</i>	792	84,70%	0
EcR	A1	DN26490_c1_g1_i1	ACY80738.2	<i>B. coprophila</i>	1142	94.26%	0
	A2	DN26490_c1_g1_i3	ACY80738.2	<i>B. coprophila</i>	1142	94.26%	0
	A3	DN26490_c1_g1_i2	ACY80738.2	<i>B. coprophila</i>	1142	94.26%	0
	B	DN26490_c1_g3_i1	ACY80739.2	<i>B. coprophila</i>	1098	96.67%	0
Broad complex	A	DN26132_c0_g8_i1	AAS80327.1	<i>A. aegypti</i>	368	78.23%	1e-120
	B	DN26132_c0_g5_i1	AAS80326.1	<i>A. aegypti</i>	295	52.46%	2e-92
	C	DN26132_c0_g6_i1	XP_011294084.1	<i>M. domestica</i>	313	49.77%	1e-98
	D	DN26132_c0_g1_i1	XP_017481712.1	<i>R. zephyria</i>	194	57.46%	3e-54
	E	DN49472_c0_g1_i1	XP_019534532.1	<i>A. albopictus</i>	141	88.46%	3e-41

Tabela 20. Resultado final do BLAST dos diferentes transcritos de *B. hygida*. A tabela apresenta diferentes isoformas preditas dos genes analisados na segunda coluna, os transcritos de embrião de *B. hygida* na terceira coluna, o hit de maior valor para cada transcrito analisado, a espécie que apresentou o maior hit com a sequência analisada, o valor correspondente ao BLAST, a porcentagem de identidade do alinhamento e na última coluna o valor E do alinhamento.

REFERÊNCIAS

- ADAMS, M. D. et al. The genome sequence of *Drosophila melanogaster*. **Science**, v. 287, n. 5461, p. 2185-95, Mar 2000. ISSN 0036-8075.
- ASHBURNER, M. et al. Temporal Control of Puffing Activity in Polytene Chromosomes. 1974-01-01 1974.
- AFFOLTER, M. et al. Signal-induced repression: the exception or the rule in developmental signaling? **Dev Cell**, v. 15, n. 1, p. 11-22, Jul 2008. ISSN 1878-1551.
- ALI, M. S.; IWANAGA, M.; KAWASAKI, H. Ecdysone-responsive transcriptional regulation determines the temporal expression of cuticular protein genes in wing discs of *Bombyx mori*. **Gene**, v. 512, n. 2, p. 337-47, Jan 2013. ISSN 1879-0038.
- ARNOLD, P. et al. Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. **Genome Res**, v. 23, n. 1, p. 60-73, Jan 2013. ISSN 1549-5469.
- ASHBURNER, M. et al. Temporal control of puffing activity in polytene chromosomes. **Cold Spring Harb Symp Quant Biol**, v. 38, p. 655-62, 1974. ISSN 0091-7451.
- ASHBURNER, M. Puffs, genes and hormones revisited. **Cell**, v. 61 (1), 1990.
- BAKER, K. D. et al. The *Drosophila* orphan nuclear receptor DHR38 mediates an atypical ecdysteroid signaling pathway. **Cell**, v. 113, n. 6, p. 731-42, Jun 2003. ISSN 0092-8674.
- BARRÓN, M. G. et al. Population genomics of transposable elements in *Drosophila*. **Annu Rev Genet**, v. 48, p. 561-81, 2014. ISSN 1545-2948.
- BASSO, L. R. et al. The induction of DNA puff BhC4-1 gene is a late response to the increase in 20-hydroxyecdysone titers in last instar dipteran larvae. **Mech Dev**, v. 110, n. 1-2, p. 15-26, Jan 2002.
- BECKSTEAD, R. B.; LAM, G.; THUMMEL, C. S. The genomic response to 20-hydroxyecdysone at the onset of *Drosophila* metamorphosis. **Genome Biol**, v. 6, n. 12, p. R99, 2005. ISSN 1474-760X.
- BEERMANN, W.; CLEVER, U. CHROMOSOME PUFFS. **Sci Am**, v. 210, p. 50-8, Apr 1964. ISSN 0036-8733.
- BENDER, M. et al. *Drosophila* ecdysone receptor mutations reveal functional differences among receptor isoforms. **Cell**, v. 91, n. 6, p. 777-88, Dec 1997. ISSN 0092-8674.
- BIALECKI, M. et al. Loss of the ecdysteroid-inducible E75A orphan nuclear receptor uncouples molting from metamorphosis in *Drosophila*. **Dev Cell**, v. 3, n. 2, p. 209-20, Aug 2002. ISSN 1534-5807.

BILLAS, I. M. et al. Crystal structure of the ligand-binding domain of the ultraspiracle protein USP, the ortholog of retinoid X receptors in insects. **J Biol Chem**, v. 276, n. 10, p. 7465-74, Mar 2001. ISSN 0021-9258.

BLACK, W. C.; RAI, K. S. Genome evolution in mosquitoes: intraspecific and interspecific variation in repetitive DNA amounts and organization. **Genet Res**, v. 51, n. 3, p. 185-96, Jun 1988.

BREUER, M. E.; PAVAN, C. Behavior of polytene chromosomes of *Rhynchosciara angela* at different stages of larval development. **Chromosoma**, v. 7, n. 4, p. 341-86, 1955.

BROADUS, J. et al. The *Drosophila* beta FTZ-F1 orphan nuclear receptor provides competence for stage-specific responses to the steroid hormone ecdysone. **Mol Cell**, v. 3, n. 2, p. 143-9, Feb 1999. ISSN 1097-2765.

BRUMBY, A. M.; RICHARDSON, H. E. Using *Drosophila melanogaster* to map human cancer pathways. **Nat Rev Cancer**, v. 5, n. 8, p. 626-39, Aug 2005. ISSN 1474-175X.

BURTIS, K. C. et al. The *Drosophila* 74EF early puff contains *E74*, a complex ecdysone-inducible gene that encodes two ets-related proteins. **Cell**, v. 61, n. 1, p. 85-99, Apr 1990.

BUTENANDT, A.; KARLSON, P. Über die Isolierung eines Metamorphose-Hormons der Insekten in kristallisierter Form **Z. Naturforsch**, 9b, pp. 389-391, 1954.

CANDIDO-SILVA, J. A.; M. C. MACHADO; K. H. HARTFELDER; J. C. de ALMEIDA; M. L. PAÇÓ-LARSON; N. MONESI (2015) Amplification and expression of a salivary gland DNA puff gene in the prothoracic gland of *Bradysia hygida* (Diptera: Sciaridae). **J Insect Physiol**, 74, 30-7.

CHEN, H. et al. De novo sequencing and characterization of the *Bradysia odoriphaga* (Diptera: Sciaridae) larval transcriptome. **Comp Biochem Physiol Part D Genomics Proteomics**, v. 16, p. 20-7, Dec 2015. ISSN 1878-0407.

CHEN, H. et al. Using Next-Generation Sequencing to Detect Differential Expression Genes in *Bradysia odoriphaga* after Exposure to Insecticides. **Int J Mol Sci**, v. 18, n. 11, Nov 2017.

CHEN, X. G. et al. Genome sequence of the Asian Tiger mosquito, *Aedes albopictus*, reveals insights into its biology, genetics, and evolution. **Proc Natl Acad Sci U S A**, v. 112, n. 44, p. E5907-15, Nov 2015. ISSN 1091-6490.

CHERBAS, L. et al. The transcriptional diversity of 25 *Drosophila* cell lines. **Genome Res**, v. 21, n. 2, p. 301-14, Feb 2011. ISSN 1549-5469.

CHERBAS, L.; LEE, K.; CHERBAS, P. Identification of ecdysone response elements by analysis of the *Drosophila* Eip28/29 gene. 5 (1); p 120-31, 1991.

CLAYCOMB, J. M.; ORR-WEAVER, T. L. Developmental gene amplification: insights into DNA replication and gene expression. **Trends Genet**, v. 21, n. 3, p. 149-62, Mar 2005.

CLAYTON, G. M. et al. The structure of the ultraspiracle ligand-binding domain reveals a nuclear receptor locked in an inactive conformation. **Proc Natl Acad Sci U S A**, v. 98, n. 4, p. 1549-54, Feb 2001. ISSN 0027-8424.

CLEVER, U.; KARLSON, P. [Induction of puff changes in the salivary gland chromosomes of *Chironomus tentans* by ecdysone]. **Exp Cell Res**, v. 20, p. 623-6, Sep 1960. ISSN 0014-4827.

COELHO, P. S. et al. DNA puff C4 of *Bradysia hygida* (Diptera: Sciaridae) contains genes unequally amplified and differentially expressed during development. **Chromosome Res**, v. 1, n. 2, p. 121-6, Jul 1993. ISSN 0967-3849.

CONSORTIUM, A. G. G. et al. Genetic diversity of the African malaria vector *Anopheles gambiae*. **Nature**, v. 552, n. 7683, p. 96-100, 12 2017. ISSN 1476-4687.

DE-ALMEIDA, J. C. A 28-fold increase in secretory protein synthesis is associated with DNA puff activity in the salivary gland of *Bradysia hygida* (Diptera, Sciaridae). **Braz J Med Biol Res**, v. 30, n. 5, p. 605-14, May 1997.

DI CARA, F.; KING-JONES, K. How clocks and hormones act in concert to control the timing of insect development. **Curr Top Dev Biol**, v. 105, p. 1-36, 2013. ISSN 1557-8933.

DIBELLO, P. R. et al. The *Drosophila* Broad-Complex encodes a family of related proteins containing zinc fingers. **Genetics**, v. 129, n. 2, p. 385-97, Oct 1991. ISSN 0016-6731.

EDGAR, B. A.; ORR-WEAVER, T. L. Endoreplication cell cycles: more for less. **Cell**, v. 105, n. 3, p. 297-306, May 2001. ISSN 0092-8674.

ENGEL, M. S.; GRIMALDI, D. Evolution of insects. **Cambridge University Press**, 2005.

EVANS, R. M. The steroid and thyroid hormone receptor superfamily. **Science**, v. 240, n. 4854, p. 889-95, May 1988. ISSN 0036-8075.

FLETCHER, J. C.; D'AVINO, P. P.; THUMMEL, C. S. A steroid-triggered switch in E74 transcription factor isoforms regulates the timing of secondary-response gene expression. **Proc Natl Acad Sci U S A**, v. 94, n. 9, p. 4582-6, Apr 1997.

FONTES, A. M. et al. The DNA puff B_hB10-1 gene encodes a glycine-rich protein secreted by the late stage larval salivary glands of *Bradysia hygida*. **Gene**, v. 231, n. 1-2, p. 67-75, Apr 1999.

FONTES, A., J. DE-ALMEIDA, J. EDSTROM & M. L. PAÇO-LARSON. Cloning of a B10 DNA puff sequence developmentally amplified and expressed in the salivary gland of *Bradysia hygida*. **Brazilian journal of medical and biological research** = Revista brasileira de pesquisas medicas e biologicas, 25, 777-780, 1992.

FORMAN, B. M.; SAMUELS, H. H. Interactions among a subfamily of nuclear hormone receptors: the regulatory zipper model. **Mol Endocrinol**, v. 4, n. 9, p. 1293-301, Sep 1990. ISSN 0888-8809.

FOULK, M. S. et al. Ecdysone induces transcription and amplification in *Sciara coprophila* DNA puff II/9A. **Dev Biol**, v. 299, n. 1, p. 151-63, Nov 2006.

FOULK, M. S. et al. Isolation and characterization of the ecdysone receptor and its heterodimeric partner ultraspiracle through development in *Sciara coprophila*. **Chromosoma**, v. 122, n. 1-2, p. 103-19, Mar 2013. ISSN 1432-0886.

FOX, D. T.; DURONIO, R. J. Endoreplication and polyploidy: insights into development and disease. **Development**, v. 140, n. 1, p. 3-12, Jan 2013. ISSN 1477-9129.

- FRANK, H. O. et al. The *Drosophila melanogaster* Eip74EF-PA transcription factor directly binds the sciarid BhC4-1 promoter. **Genesis**, 55, e23075. 2017.
- GAO, H. et al. Transcriptome Analysis and Discovery of Genes Relevant to Development in *Bradysia odoriphaga* at Three Developmental Stages. **PLoS One**, v. 11, n. 2, p. e0146812, 2016. ISSN 1932-6203.
- GAUHAR, Z. et al. Genomic mapping of binding regions for the Ecdysone receptor protein complex. **Genome Res**, v. 19, n. 6, p. 1006-13, Jun 2009. ISSN 1088-9051.
- GEIB, S. M. et al. Characterizing the developmental transcriptome of the oriental fruit fly, *Bactrocera dorsalis* (Diptera: Tephritidae) through comparative genomic analysis with *Drosophila melanogaster* utilizing modENCODE datasets.
- GERBI, S. A. et al. DNA amplification in DNA puff II/9A of *Sciara coprophila*. **Cold Spring Harb Symp Quant Biol**, v. 58, p. 487-94, 1993.
- GILL, G. Post-translational modification by the small ubiquitin-related modifier SUMO has big effects on transcription factor activity. **Curr Opin Genet Dev**, v. 13, n. 2, p. 108-13, Apr 2003. ISSN 0959-437X.
- GILBERT, L. I.; RYBCZYNSKI, R.; WARREN, J. T. Control and biochemical nature of the ecdysteroidogenic pathway. **Annu Rev Entomol**, v. 47, p. 883-916, 2002.
- GONSALVES, S. E. et al. Genome-wide examination of the transcriptional response to ecdysteroids 20-hydroxyecdysone and ponasterone A in *Drosophila melanogaster*. **BMC Genomics**, v. 12, p. 475, Sep 2011. ISSN 1471-2164.
- GRATZ, S. J. et al. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. **Genetics**, v. 196, n. 4, p. 961-71, Apr 2014. ISSN 1943-2631.
- GRAVELEY, B. R. et al. The developmental transcriptome of *Drosophila melanogaster*. **Nature**, v. 471, n. 7339, p. 473-9, Mar 2011. ISSN 1476-4687.
- GRÖNKE, S. et al. Molecular evolution and functional characterization of *Drosophila* insulin-like peptides. **PLoS Genet**, v. 6, n. 2, p. e1000857, Feb 2010. ISSN 1553-7404.
- GUO, Z. et al. Insect Transcription Factors: A Landscape of Their Structures and Biological Functions in. **Int J Mol Sci**, v. 19, n. 11, Nov 2018. ISSN 1422-0067.
- GWÓŹDŹ, T. et al. EcR and Usp, components of the ecdysteroid nuclear receptor complex, exhibit differential distribution of molecular determinants directing subcellular trafficking. **Cell Signal**, v. 19, n. 3, p. 490-503, Mar 2007. ISSN 0898-6568.
- HAMMOND, M. P.; LAIRD, C. D. Control of DNA replication and spatial distribution of defined DNA sequences in salivary gland cells of *Drosophila melanogaster*. **Chromosoma**, v. 91, n. 3-4, p. 279-86, 1985. ISSN 0009-5915.
- HENRICH, V. C. et al. Expression and function of the ultraspiracle (usp) gene during development of *Drosophila melanogaster*. **Dev Biol**, v. 165, n. 1, p. 38-52, Sep 1994. ISSN 0012-1606.
- HIRAI, M. et al. Bombyx mori orphan receptor, BmHR78: cDNA cloning, testis abundant expression and putative dimerization partner for Bombyx ultraspiracle. **Mol Cell Endocrinol**, v. 189, n. 1-2, p. 201-11, Mar 2002. ISSN 0303-7207.

- HOLT, R. A. et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. **Science**, v. 298, n. 5591, p. 129-49, Oct 2002. ISSN 1095-9203.
- HU, X.; CHERBAS, L.; CHERBAS, P. Transcription activation by the ecdysone receptor (EcR/USP): identification of activation functions. **Mol Endocrinol**, v. 17, n. 4, p. 716-31, Apr 2003. ISSN 0888-8809.
- HUA, B. L.; ORR-WEAVER, T. L. DNA Replication Control During. **Genetics**, v. 207, n. 1, p. 29-47, 09 2017.
- IHRY, R. J.; BASHIRULLAH, A. Genetic control of specificity to steroid-triggered responses in *Drosophila*. **Genetics**, v. 196, n. 3, p. 767-80, Mar 2014. ISSN 1943-2631.
- JORY, A. et al. A survey of 6,300 genomic fragments for cis-regulatory activity in the imaginal discs of *Drosophila melanogaster*. **Cell Rep**, v. 2, n. 4, p. 1014-24, Oct 2012. ISSN 2211-1247.
- KAMINKER, J. S. et al. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. **Genome Biol**, v. 3, n. 12, p. RESEARCH0084, 2002. ISSN 1474-760X.
- KELLEY, J. L. et al. Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. **Nat Commun**, v. 5, p. 4611, Aug 2014. ISSN 2041-1723.
- KING-JONES, K.; THUMMEL, C. S. Nuclear receptors--a perspective from *Drosophila*. **Nat Rev Genet**, v. 6, n. 4, p. 311-23, Apr 2005a. ISSN 1471-0056.
- KISS, I. et al. Interactions and developmental effects of mutations in the Broad-Complex of *Drosophila melanogaster*. **Genetics**, v. 118, n. 2, p. 247-59, Feb 1988. ISSN 0016-6731.
- KOELLE, M. R. et al. The *Drosophila* EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. **Cell**, v. 67, n. 1, p. 59-77, Oct 1991. ISSN 0092-8674.
- KOPEC, S. Studies on the necessity of the brain for inception of insect metamorphosis, *Biology Bulletin*, Woods Hole, 42: 322-342, 1922.
- LAICINE, E. M. et al. Development of DNA puffs and patterns of polypeptide synthesis in the salivary glands of *Bradysia hygida*. **Chromossoma**, v. 84, n. 4, p. 280-284, 1984.
- LAVAL, M. et al. In vivo functional characterization of an ecdysone response enhancer in the proximal upstream region of the *Fbp1* gene of *D. melanogaster*. **Mech Dev**, v. 44, n. 2-3, p. 123-38, Dec 1993. ISSN 0925-4773.
- LEE, H. O.; DAVIDSON, J. M.; DURONIO, R. J. Endoreplication: polyploidy with purpose. **Genes Dev**, v. 23, n. 21, p. 2461-77, Nov 2009. ISSN 1549-5477.
- LI, T. R.; WHITE, K. P. Tissue-specific gene expression and ecdysone-regulated genomic networks in *Drosophila*. **Dev Cell**, v. 5, n. 1, p. 59-72, Jul 2003. ISSN 1534-5807.
- LIU, X. et al. 20-Hydroxyecdysone (20E) Primary Response Gene E93 Modulates 20E Signaling to Promote *Bombyx* Larval-Pupal Metamorphosis. **J Biol Chem**, v. 290, n. 45, p. 27370-83, Nov 2015. ISSN 1083-351X.

LOSICK, V. P.; FOX, D. T.; SPRADLING, A. C. Polyploidization and cell fusion contribute to wound healing in the adult *Drosophila* epithelium. **Curr Biol**, v. 23, n. 22, p. 2224-2232, Nov 2013. ISSN 1879-0445.

MACKAY, T. F. C.; HUANG, W. Charting the genotype-phenotype map: lessons from the *Drosophila melanogaster* Genetic Reference Panel. **Wiley Interdiscip Rev Dev Biol**, v. 7, n. 1, 01 2018. ISSN 1759-7692.

MAHAJAN, S.; BACHTROG, D. Convergent evolution of Y chromosome gene content in flies. **Nat Commun**, v. 8, n. 1, p. 785, 10 2017. ISSN 2041-1723.

MALTA, J. et al. Midgut morphological changes and autophagy during metamorphosis in sand flies. **Cell Tissue Res**, v. 368, n. 3, p. 513-529, 06 2017. ISSN 1432-0878.

MOHRIG, W., MENZEL, F. Sciaridae (Black Fungus Gnats). In: Brown, B.V., Borkent, A., Cumming, J.M., Wood, D.M., Woodley, N.E., Zumbado, M.A. (Eds.), *Manual of Central American Diptera*, vol. 1. **NRC Research Press**, Ottawa, pp. 279–292. 2009

MONESI, N. et al. Immunocharacterization of the DNA puff BhC4-1 protein of *Bradysia hygida* (Diptera: Sciaridae). **Insect Biochem Mol Biol**, v. 34, n. 6, p. 531-42, Jun 2004.

MONESI, N. et al. Molecular characterization of an 18 kb segment of DNA puff C4 of *Bradysia hygida* (Diptera, sciaridae). **Chromosoma**, v. 103, n. 10, p. 715-24, Jul 1995.

MONESI, N. et al. Regulation of Sciarid DNA Puffs by Ecdysone: Mechanisms and Perspectives. In: SMAGGHE, G. (Ed.). **Ecdysone: Structures and Functions**: Springer Netherlands, 2009. cap. II, p.165-183.

MONESI, N., M. JACOBS-LORENA & M. L. PAÇÓ-LARSON. The DNA puff gene BhC4-1 of *Bradysia hygida* is specifically transcribed in early prepupal salivary glands of *Drosophila melanogaster*. **Chromosoma**, 107, 559-69. 1998.

MOTT, R. EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. **Comput Appl Biosci**, v. 13, n. 4, p. 477-8, Aug 1997. ISSN 0266-7061.

MOUILLET, J. F. et al. Differential control of gene activity by isoforms A, B1 and B2 of the *Drosophila* ecdysone receptor. **Eur J Biochem**, v. 268, n. 6, p. 1811-9, Mar 2001. ISSN 0014-2956.

NEAFSEY, D. E. et al. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. **Science**, v. 347, n. 6217, p. 1258522, Jan 2015. ISSN 1095-9203.

NEELY, G. G. et al. A genome-wide *Drosophila* screen for heat nociception identifies $\alpha 2\delta 3$ as an evolutionarily conserved pain gene. **Cell**, v. 143, n. 4, p. 628-38, Nov 2010. ISSN 1097-4172.

ORO, A. E.; MCKEOWN, M.; EVANS, R. M. Relationship between the product of the *Drosophila* ultraspiracle locus and the vertebrate retinoid X receptor. **Nature**, v. 347, n. 6290, p. 298-301, Sep 1990. ISSN 0028-0836.

PAÇÓ-LARSON, M. L., J. C. DE ALMEIDA, J. E. EDSTRÖM & H. SAUAIA (1992) Cloning of a developmentally amplified gene sequence in the DNA puff C4 of *Bradysia hygida* (Diptera: Sciaridae) salivary glands. **Insect Biochemistry and Molecular Biology**, 22, 439-446.

PAPE, T.; BICKEL, D.; RUDOLF, M. *Diptera Diversity: Status, Challenges and Tools*, 1 ed. Leiden, Boston: **Brill Academic Publishers**, 459 p. 2009.

RABINOVICH, D. et al. Nitric Oxide as a Switching Mechanism between Axon Degeneration and Regrowth during Developmental Remodeling. **Cell**, v. 164, n. 1-2, p. 170-182, Jan 2016. ISSN 1097-4172.

RENAUD, J. P.; MORAS, D. Structural studies on nuclear receptors. **Cell Mol Life Sci**, v. 57, n. 12, p. 1748-69, Nov 2000. ISSN 1420-682X.

RICHARDS, S. M. et al. Androgen regulation of gene expression in the mouse lacrimal gland. **J Steroid Biochem Mol Biol**, v. 96, n. 5, p. 401-13, Sep 2005. ISSN 0960-0760.

RIDDIFORD, L. M. Hormone receptors and the regulation of insect metamorphosis. **Receptor**, v. 3, n. 3, p. 203-9, 1993. ISSN 1052-8040.

RITTER, A. R.; BECKSTEAD, R. B. Sox14 is required for transcriptional and developmental responses to 20-hydroxyecdysone at the onset of drosophila metamorphosis. **Dev Dyn**, v. 239, n. 10, p. 2685-94, Oct 2010. ISSN 1097-0177.

ROSENFELD, M. G.; LUNYAK, V. V.; GLASS, C. K. Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. **Genes Dev**, v. 20, n. 11, p. 1405-28, Jun 2006. ISSN 0890-9369.

ROTA-STABELLI, O.; DALEY, A. C.; PISANI, D. Molecular timetrees reveal a Cambrian colonization of land and a new scenario for ecdysozoan evolution. **Curr Biol**, v. 23, n. 5, p. 392-8, Mar 2013. ISSN 1879-0445.

RUBIN, G. M. et al. Comparative genomics of the eukaryotes. **Science**, v. 287, n. 5461, p. 2204-15, Mar 2000. ISSN 0036-8075.

SANTELLI, R. V. et al. Replication and transcription in the course of DNA amplification of the C3 and C8 DNA puffs of *Rhynchosciara americana*. **Mech Dev**, v. 36, n. 1-2, p. 59-66, Dec 1991.

SASS, H. Hierarchy of fibrillar organization levels in the polytene interphase chromosomes of *Chironomus*. **J Cell Sci**, v. 45, p. 269-93, Oct 1980. ISSN 0021-9533.

SEGRAVES, W. A.; HOGNESS, D. S. The E75 ecdysone-inducible gene responsible for the 75B early puff in *Drosophila* encodes two new members of the steroid receptor superfamily. **Genes Dev**, v. 4, n. 2, p. 204-19, Feb 1990. ISSN 0890-9369.

SHARROCKS, A. D. The ETS-domain transcription factor family. **Nat Rev Mol Cell Biol**, v. 2, n. 11, p. 827-37, Nov 2001. ISSN 1471-0072.

SHEA, M. J. et al. Proteins that bind to *Drosophila* chorion cis-regulatory elements: a new C2H2 zinc finger protein and a C2C2 steroid receptor-like component. **Genes Dev**, v. 4, n. 7, p. 1128-40, Jul 1990. ISSN 0890-9369.

SHIBUTANI, S. T. et al. Intrinsic negative cell cycle regulation provided by PIP box- and Cul4Cdt2-mediated destruction of E2f1 during S phase. **Dev Cell**, v. 15, n. 6, p. 890-900, Dec 2008. ISSN 1878-1551.

SHIN, S. et al. Molecular phylogeny of black fungus gnats (Diptera: Sciaroidea: Sciaridae) and the evolution of larval habitats. **Mol Phylogenet Evol**, v. 66, n. 3, p. 833-46, Mar 2013. ISSN 1095-9513.

- SHLYUEVA, D. et al. Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. **Mol Cell**, v. 54, n. 1, p. 180-192, Apr 2014. ISSN 1097-4164.
- SHU, Z.; DENG, W. M. Differential Regulation of Cyclin E by Yorkie-Scalloped Signaling in Organ Development. **G3 (Bethesda)**, v. 7, n. 3, p. 1049-1060, 03 2017. ISSN 2160-1836.
- SHU, Z.; ROW, S.; DENG, W. M. Endoreplication: The Good, the Bad, and the Ugly. **Trends Cell Biol**, v. 28, n. 6, p. 465-474, 06 2018. ISSN 1879-3088.
- SIMON, C. R.; SIVIERO, F.; MONESI, N. Beyond DNA puffs: What can we learn from studying sciarids? **genesis**, v. 54, n. 7, p. 17, June 1, 2016.
- SONG, L.; RAPE, M. Substrate-specific regulation of ubiquitination by the anaphase-promoting complex. **Cell Cycle**, v. 10, n. 1, p. 52-6, Jan 2011. ISSN 1551-4005.
- STILWELL, G. E. et al. E74 exhibits stage-specific hormonal regulation in the epidermis of the tobacco hornworm, *manduca sexta*. *Dev Biol*, v. 258, n. 1, p. 76-90, Jun 2003. ISSN 0012-1606.
- STOCKER, A. J. et al. DNA replication and amplification during the final cycle of polytene in sciarid salivary gland chromosomes and their control by ecdysone: **Ciência e Cultura**. 48: 306-312 p. 1996.
- STOIBER, M. et al. Diverse Hormone Response Networks in 41 Independent *Drosophila* Cell Lines. **G3 (Bethesda)**, v. 6, n. 3, p. 683-94, Jan 2016. ISSN 2160-1836.
- SUN, G.; ZHU, J.; RAIKHEL, A. S. The early gene E74B isoform is a transcriptional activator of the ecdysteroid regulatory hierarchy in mosquito vitellogenesis. **Mol Cell Endocrinol**, v. 218, n. 1-2, p. 95-105, Apr 2004. ISSN 0303-7207.
- SUTHERLAND, J. D. et al. *Drosophila* hormone receptor 38: a second partner for *Drosophila* USP suggests an unexpected role for nuclear receptors of the nerve growth factor-induced protein B type. **Proc Natl Acad Sci U S A**, v. 92, n. 17, p. 7966-70, Aug 1995. ISSN 0027-8424.
- SWANSON, C. I. et al. Expression of an S phase-stabilized version of the CDK inhibitor Dacapo can alter endoreplication. **Development**, v. 142, n. 24, p. 4288-98, Dec 2015. ISSN 1477-9129.
- TALBOT, W. S.; SWYRYD, E. A.; HOGNESS, D. S. *Drosophila* tissues with different metamorphic responses to ecdysone express different ecdysone receptor isoforms. **Cell**, v. 73, n. 7, p. 1323-37, Jul 1993. ISSN 0092-8674.
- THUMMEL, C. S. Flies on steroids--*Drosophila* metamorphosis and the mechanisms of steroid hormone action. **Trends Genet**, v. 12, n. 8, p. 306-10, Aug 1996.
- TRAUTWEIN, M. D. et al. Advances in insect phylogeny at the dawn of the postgenomic era. **Annu Rev Entomol**, v. 57, p. 449-68, 2012. ISSN 1545-4487.
- ULIANA, J. V. C., et al. Characterizing the embryonic development of *B. hygida* (Diptera: Sciaridae) following enzymatic treatment to permeabilize the serosal cuticle. **Mechanisms of development**, v. 154, 2018.

TSAI, M. J.; O'MALLEY, B. W. Molecular mechanisms of action of steroid/thyroid receptor superfamily members. **Annu Rev Biochem**, v. 63, p. 451-86, 1994. ISSN 0066-4154.

URNESS, L. D.; THUMMEL, C. S. Molecular analysis of a steroid-induced regulatory hierarchy: the *Drosophila* E74A protein directly regulates *L71-6* transcription. **EMBO J**, v. 14, n. 24, p. 6239-46, Dec 1995.

UYEHARA, C. M. et al. Hormone-dependent control of developmental timing through regulation of chromatin accessibility. **Genes Dev**, v. 31, n. 9, p. 862-875, 05 2017. ISSN 1549-5477.

WÄRNMARK, A. et al. Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation. **Mol Endocrinol**, v. 17, n. 10, p. 1901-9, Oct 2003. ISSN 0888-8809.

WATANABE, T.; TAKEUCHI, H.; KUBO, T. Structural diversity and evolution of the N-terminal isoform-specific region of ecdysone receptor-A and -B1 isoforms in insects. **BMC Evol Biol**, v. 10, p. 40, Feb 2010. ISSN 1471-2148.

WATERHOUSE, R. M. A maturing understanding of the composition of the insect gene repertoire. **ScienceDirect**, 7:15-23, 2015.

WESTER, J. V. W. C. Characterization of a novel *Drosophila melanogaster* cis-regulatory module that drives gene expression to the larval tracheal system and adult thoracic musculature. **Genesis**, 56, e23222. 2018.

WHITE, K. P. et al. Coordination of *Drosophila* metamorphosis by two ecdysone-induced nuclear receptors. **Science**, v. 276, n. 5309, p. 114-7, Apr 1997. ISSN 0036-8075.

WIEGMANN, B. M.; RICHARDS, S. Genomes of Diptera. **Curr Opin Insect Sci**, v. 25, p. 116-124, 02 2018. ISSN 2214-5753.

WIEGMANN, B.M. & YEATES, D.K., (2017) Phylogeny of Diptera. Manual of Afrotropical Diptera Volume 1. Suricata 4. **South African National Biodiversity Institute**. Pretoria.p. 253-265.

WILSON, V. G.; RANGASAMY, D. Intracellular targeting of proteins by sumoylation. **Exp Cell Res**, v. 271, n. 1, p. 57-65, Nov 2001. ISSN 0014-4827.

WOODARD, C. T.; BAEHRECKE, E. H.; THUMMEL, C. S. A molecular mechanism for the stage specificity of the *Drosophila* prepupal genetic response to ecdysone. **Cell**, v. 79, n. 4, p. 607-15, Nov 1994. ISSN 0092-8674.

XIANG, Y.; LIU, Z.; HUANG, X. br regulates the expression of the ecdysone biosynthesis gene *npc1*. **Dev Biol**, v. 344, n. 2, p. 800-8, Aug 2010. ISSN 1095-564X.

XU, Q. Y. et al. Requirement of *Leptinotarsa decemlineata* gene within the 74EF puff for larval-pupal metamorphosis and appendage growth. **Insect Mol Biol**, v. 27, n. 4, p. 439-453, 08 2018. ISSN 1365-2583.

YAMADA, M. et al. Temporally restricted expression of transcription factor betaFTZ-F1: significance for embryogenesis, molting and metamorphosis in *Drosophila melanogaster*. **Development**, v. 127, n. 23, p. 5083-92, Dec 2000. ISSN 0950-1991.

YAMANAKA, N.; REWITZ, K. F.; O'CONNOR, M. B. Ecdysone control of developmental transitions: lessons from *Drosophila* research. **Annu Rev Entomol**, v. 58, p. 497-516, 2013. ISSN 1545-4487.

ZARET, K. S.; MANGO, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. **Curr Opin Genet Dev**, v. 37, p. 76-81, 04 2016. ISSN 1879-0380.

ZDOBNOV, E. M. et al. Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. **Science**, v. 298, n. 5591, p. 149-59, Oct 2002. ISSN 1095-9203.

ZHAO, J. Sumoylation regulates diverse biological processes. **Cell Mol Life Sci**, v. 64, n. 23, p. 3017-33, Dec 2007. ISSN 1420-682X.

ZIELKE, N. et al. Control of *Drosophila* endocycles by E2F and CRL4(CDT2). **Nature**, v. 480, n. 7375, p. 123-7, Oct 2011. ISSN 1476-4687.

ZIELKE, N. et al. The anaphase-promoting complex/cyclosome (APC/C) is required for rereplication control in endoreplication cycles. **Genes Dev**, v. 22, n. 12, p. 1690-703, Jun 2008. ISSN 0890-9369.