Raquel Dietsche Monfardini

Evolução das vias de controle do desenvolvimento tecidual em holometábolos

Evolution of the control pathways of tissue development in holometabolous insects

São Paulo
2018

Raquel Dietsche Monfardini

Evolução das vias de controle do desenvolvimento tecidual em holometábolos

Evolution of the control pathways of tissue development in holometabolous insects

Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Mestre em Ciências Biológicas, na Área de Biologia/Genética.

Orientadora: Tatiana Teixeira Torres

Coorientador: Alisson Gontijo

São Paulo
2018

Ficha Catalográfica

Monfardini, Raquel Dietsche

Evolução das vias de controle do desenvolvimento tecidual em holometábolos / Raquel Dietsche Monfardini; oritentadora Tatiana Teixeira Torres – São Paulo, 2018

87 páginas

Dissertação (Mestrado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.

1. ilp8 2. Metabolismo 3. Expressão gênica

Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva

Comissão Julgadora

Prof(a). Dr(a).	Prof(a). Dr(a).
Prof(a). Dr.(a).	Prof(a). Dr.(a). Tatiana Teixeira Torres Orientadora

Em memória da minha avó Magda e do meu tio Félix,
de quem sinto muita falta.

“Wait for you? Not likely!

I've always had to run ahead of you and show you the way.”

***Fool’s Errand*, Robin Hobb**

Agradecimentos

Agradeço primeiramente meus orientadores, Tatiana e Alisson, por proporcionarem a grande oportunidade que foi trabalhar com este projeto. Obrigada pela paciência, o incentivo, a dedicação e tempo que depositaram em mim. Agradeço, principalmente, por me ajudar a aprender e fazer boa ciência. Muito obrigada pelas oportunidades e experiências únicas que vocês me proporcionaram.

Agradeço a todos do Laboratório de Genômica e Evolução de Artrópodes. Gisele, Marina, Nancy, Valérial, Nancy Lo, Mauimy, Juliana, Carla, Gustavo, Pedro e Vanessa, muito obrigada por estarem comigo nos meus sucessos e ataques de revolta. Adiciono aqui o Tiago, que faz parte do nosso grupo apesar de pertencer a outro laboratório. Nossas discussões sobre séries e filmes foram momentos de descontração essenciais. Agradeço especialmente pelo apoio e ajuda que vocês me deram durante este projeto e por engordarem junto comigo!

Aproveito também para agradecer a todos do Laboratório de Biomedicina Integrativa, em Portugal, por terem me acolhido tão carinhosamente. Joana, Andréia, André, Ángela, Ivana, Ana e Felipe, muito obrigada por me receber de braços abertos e por tudo que vocês me ensinaram. Obrigada por ficarem comigo até tarde da noite ou nos fins de semana no laboratório! Parte desta dissertação não teria sido realizada sem a participação e ajuda de todos. Um agradecimento especial aos colaboradores Drª. Fabiana Heredia, Dr. Andres Garelli, Dr. Takashi Koyama e Catarina Nunes por terem realizados os ensaios de dano tecidual.

A todos do Laboratório de Entomologia Integrativa, em Campinas, pela colaboração na realização de alguns dos ensaios experimentais. Um agradecimento mais que especial para a professora Patrícia, o Vinícius e o Rafael por me ajudarem ativamente na manutenção das larvas e pupas enquanto estava realizando os ensaios de dano.

Agradeço a agência de fomento CNPq e FAPESP (processo 2015/20183-8) pelo financiamento, sem o qual essa dissertação não se realizaria. O presente trabalho também foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Por fim, mas igualmente importante, agradeço à minha família. Seu amor e apoio incondicionais, mesmo longe de casa, foram essenciais para me manter no caminho da pesquisa. Muito obrigada por ouvir minhas reclamações e dúvidas, que inevitavelmente apareceram durante todo meu crescimento pessoal. Além disso obrigada por aguentarem (várias vezes) minhas gatas pentelhas quando tive que viajar!

Índice

Abreviaturas x
Lista de Figuras e Tabelas xii
Introdução Geral 2
Holometábolos: discos imaginais e sua importância 2
Regulação hormonal durante a fase larval 4
Expressão gênica 7
Perguntas e hipóteses 9
Organização da dissertação 10
Capítulo 1: Evolução do gene *ilp8* 12
Metodologia 14
1.1. Obtenção das sequências 14
1.2. Homologia ao *ilp8* 15
Resultados 17
1.3. Processamento dos dados do 1KITE 17
1.4. Busca por homólogos 17
Discussão 22
Conclusão do capítulo 26
Capítulo 2: Regulação do atraso da metamorfose 28
Metodologia 30
2.1. Obtenção das amostras e sequenciamento 31
2.2. Processamento das sequências e montagem *de novo* dos transcriptomas 32
2.3. Análise de expressão diferencial 33
2.4. Anotação dos transcriptomas e análises de enriquecimento 35
Resultados 37
2.5. Tempo de desenvolvimento e avaliação dos transcriptomas 37
2.6. Análises de expressão diferencial global 40
2.7. Análises de expressão diferencial do *ilp8* 42
2.8. Anotação e enriquecimento 44
2.8.1. Transcritos com expressão mais diferente entre tratamentos em cada espécie 46
Discussão 52

Conclusão do capítulo 57
Conclusões Gerais 58
Evolução das vias de controle do desenvolvimento tecidual em holometábolos 58
Resumo 59
Abstract 60
Referências Bibliográficas 61
Anexos 66
Tabelas (Anexos T1 – 11) 66
Figuras (Anexos F1 – F9) 66

Abreviaturas

1KITE: 1K insect transcriptome evolution

20E: Hormônio 20-hidroxiecdisona

BEPE: Bolsa estágio de pesquisa no exterior

Blast: Basic local alignment search tool

BUSCO: Benchmarking universal single-copy orthologs

cDNA: Ácido desoxi-ribonucleico complementar

CDS: Sequência codificante

CeC: Cecropin

Dpt: Diptericin

EcR–USP: Ecdysone receptor - ultraspiracle

Eip71CD: Ecdysone-induced protein 28/29kD

EMS: Etil metano sulfonato

ESALQ: Escola superior de agricultura Luiz de Queiroz

FC: Fold change

GO: Gene Ontology

Gr: Gustatory receptor

Hsc70: Heat shock protein cognate70

IIC: Intervalo até a interrupção do crescimento

ILP: Insulin-like peptide

JAK/STAT: Janus kinase-signal transducer and activator of transcription

JH: Hormônio juvenil

JNK: Jun N-terminal kinase

L3: Terceiro instar larval

LabGEA: Laboratório de genômica e evolução de artrópodes

LGR3: Leucine-rich repeat-containing G protein-coupled receptor 3

LIRP: Locusta insulin-related peptide

MCL: Markov clustering

mM: Milimolar

mRNA: Ácido ribonucleico mensageiro

NCBI: National center for biotechnology information

ORF: Quadro de leitura

pb: Pares de base

PBS: Tampão fosfato-salino

PTTH: Hormônio prototorácico

RNA-seq: Sequenciamento massivo em paralelo do transcriptoma

RNA: Ácido ribonucleico

RpL18: Ribosomal protein L18

TCDE: Transcritos ou contigs diferencialmente expressos

Tot: Turandot

Lista de Figuras e Tabelas

Introdução Geral

Figura 1: Desenvolvimento dos discos imaginais....4

Figura 2: Regulação hormonal do desenvolvimento....5

Figura 3: Processamento de insulina/relaxina....7

Capítulo 1: Evolução do gene *ilp8*

Figura 1.1: Exemplo da análise realizada pelo SignalP....16

Figura 1.2: Gêneros com sequência recuperada do banco do 1KITE....20

Figura 1.3: Hipótese evolutiva do gene ilp8....24

Tabela 1.1: Quantidade de sequências recuperadas utilizando o motivo....19

Tabela 1.2: Identidade dos alinhamentos contra a proteína Ilp8 de *Dr. melanogaster*....21

Capítulo 2: Regulação do atraso da metamorfose

Figura 2.1: Análise de RNA-seq....30

Figura 2.2: Resposta ao dano nos ensaios com EMS....37

Figura 2.3: Análise de completude....39

Figura 2.4: Análise exploratória das diferenças de expressão....41

Figura 2.5: Valores de expressão de cada grupo de ortólogos....42

Figura 2.6: Hipótese evolutiva da função do gene ilp8....56

Tabela 2.1: Resumo dos dados de sequenciamento e processamento das *reads*....38

Tabela 2.2: Resumo da montagem de cada transcriptoma....39

Tabela 2.3: Total de TCDE de cada espécie e experimento....40

Tabela 2.4: Análise da expressão dos transcritos do gene *ilp8*....43

Tabela 2.5: Predições e anotação dos transcriptomas montados....44

Tabela 2.6: Transcritos com maior diferença de expressão em *Dr. melanogaster*....47

Tabela 2.7: *Contigs* anotados com maior diferença de expressão em *Ch. megacephala*....48

Tabela 2.8: *Contigs* anotados com maior diferença de expressão em *H. illucens*....49

Tabela 2.9: *Contigs* anotados com maior diferença de expressão em *Ma. sexta*....51

INTRODUÇÃO GERAL

Introdução Geral

A maioria dos organismos são capazes de coordenar seu desenvolvimento frente a perturbações intrínsecas e/ou ambientais. Essa capacidade, observada principalmente durante a fase imatura, permite a formação de indivíduos adultos proporcionais e funcionais. Mesmo sendo observados em inúmeras espécies, os insetos são considerados um ótimo modelo para estudos na área, uma vez que a maior parte das espécies possui ciclo de vida relativamente curto. O melhor exemplo disso é a espécie *Drosophila melanogaster*, que foi utilizada como modelo em grande parte dos estudos iniciais relacionados ao desenvolvimento. Desse modo, em um curto período de tempo é possível observar as consequências de uma perturbação durante o desenvolvimento e tentar compreender os mecanismos regulatórios envolvidos no processo. Um mecanismo observado em *Dr. melanogaster* e em outros insetos é o atraso na metamorfose frente a um dano tecidual e este trabalho buscou compreender como ele é regulado em diferentes espécies de insetos e como ele evoluiu.

Holometábolos: discos imaginais e sua importância

Em insetos, são observadas três diferentes formas de desenvolvimento (Gilbert 1949). Algumas poucas espécies possuem desenvolvimento direto (ametábolos), em que o jovem é estritamente uma cópia do adulto, e apenas aumenta seu tamanho corpóreo entre cada muda. Entretanto, a maior parte dos insetos possuem desenvolvimento indireto e são dependentes de metamorfoses para atingir a maturidade. Tais insetos podem ser classificados como hemimetábolos ou holometábolos.

Os hemimetábolos são caracterizados por possuir metamorfose incompleta. Neste caso, os indivíduos jovens se parecem fisicamente com os adultos, porém não possuem algumas características essenciais completamente desenvolvidas, por exemplo a genitália ou asas. Por meio das mudas que ocorrem durante a fase imatura, as características do adulto se desenvolvem gradualmente. Os holometábolos, por sua vez, possuem metamorfose completa, ou seja, ocorre total separação morfológica entre as

fases imatura (ou larval) e adulta. A fase larval caracteriza-se por diversas mudas intercaladas por crescimento corpóreo (delimitando os instar larvais), o que promove um acúmulo de energia que será utilizado durante a fase de pupa. A última muda larval (também chamada de muda metamórfica) forma a pupa, fase em que ocorre a metamorfose de larva para adulto (ou imago) por meio da morte celular programada dos tecidos larvais e proliferação e diferenciação dos tecidos adultos a partir de precursores imaginais.

Existem dois tipos de precursores imaginais que originarão os tecidos do adulto: os histoblastos e os discos imaginais. Os histoblastos são agrupamentos de células embrionárias que formarão o abdômen e os órgãos internos do adulto (Roseland & Schneiderman 1979; Curtiss & Heilig 1995). Já os discos imaginais são responsáveis por formar as estruturas cuticulares do adulto, como por exemplo as asas. Cada disco imaginal é um conjunto celular indiferenciado, porém já programado para formar uma estrutura específica (Figura 1). Essa especificidade ocorre durante fase embrionária e depende da expressão espacial e temporal dos genes homeóticos (Scott & Carroll 1987), sendo cada disco imaginal programado por uma combinação específica da expressão desses genes.

Considerando a importância dos discos imaginais na formação do adulto, uma de suas características mais marcantes é sua capacidade de regeneração durante o crescimento larval (Worley et al. 2012). Uma resposta às adversidades ao longo do desenvolvimento, principalmente ao dano tecidual em discos imaginais, é a extensão do tempo larval acarretando no atraso da metamorfose. Essa propriedade garante tempo para a regeneração dos discos danificados e consequente formação íntegra dos órgãos adultos.

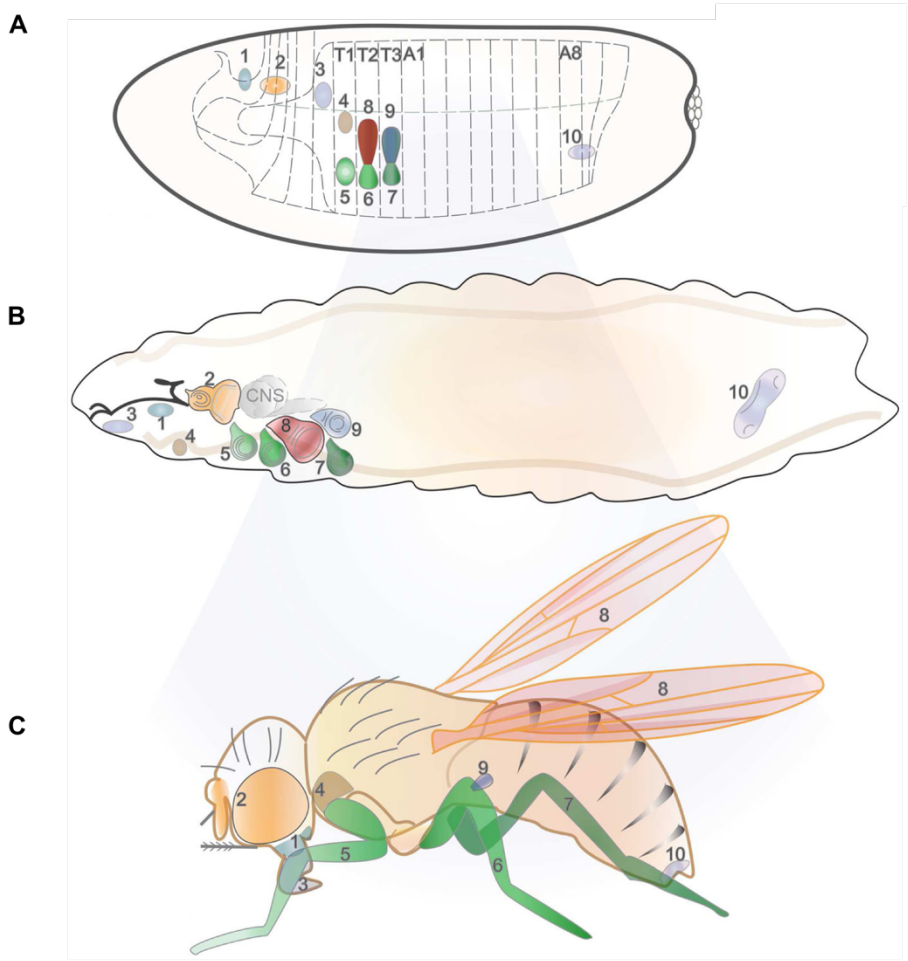


Figura 1: Desenvolvimento dos discos imaginais: Estão representadas as células embrionárias primordiais numeradas de 1 a 10 (A), os respectivos discos imaginais (B) e os produtos que formarão no adulto (C) na espécie *Dr. melanogaster*. T1 a T3 representam os seguimentos torácicos e A1 a A8 os abdominais. Imagem adaptada de Beira & Paro (2016).

Regulação hormonal durante a fase larval

Em insetos holometábolos, tanto os instares larvais quanto a formação da pupa são finamente regulados por um sistema composto principalmente pelos hormônios juvenil (JH), prototorácico (PTTH) e ecdisona. A ecdisona, considerada o principal hormônio da metamorfose, é produzida e secretada pela glândula protorácica (região presente na glândula anelar) e é convertida para sua forma ativa 20-hidroxiecdisona (20E) nos tecidos periféricos, onde promove as alterações na expressão gênica para que ocorra a muda

(revisado em Hill et al. 2013). A produção de ecdisona é controlada positivamente por um pulso de PTTH, produzido e secretado por dois pares de neurônios que enervam a glândula protorácica (Figura 2-A; McBrayer et al. 2007).

Além de desencadear a formação da pupa e a metamorfose, a ecdisona também é responsável por controlar as mudas larvais. O que difere as duas situações é uma relação entre a concentração do JH e o peso (Figura 2-B; revisado em Edgar 2006; Shingleton 2010). Durante o desenvolvimento larval o JH é mantido em altas concentrações, permitindo o crescimento do indivíduo e acúmulo de energia. Nessa situação, ainda que haja picos de ecdisona, a elevada concentração do JH inibe a muda metamórfica caracterizando, consequentemente, uma muda larval. Porém, quando a larva atinge um peso crítico, a produção do JH é inibida. Desse modo, o próximo pico de ecdisona desencadeia a formação da pupa e a metamorfose do indivíduo.

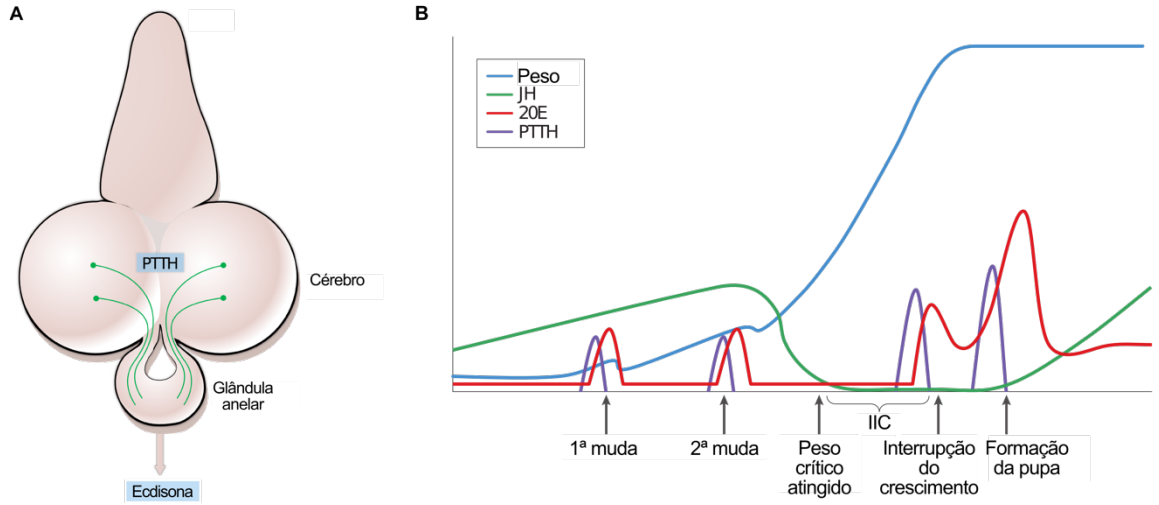


Figura 2: Regulação hormonal do desenvolvimento: (A) Produção e secreção de ecdisona controlada pelo PTTH; (B) flutuação nos níveis de concentração dos hormônios JH, PTTH, 20E e do peso larval ao longo do desenvolvimento larval (IIC significa intervalo até a interrupção do crescimento). Imagens adaptadas de Hariharan (2012) e Edgar (2006), respectivamente.

Entretanto, qualquer perturbação durante o desenvolvimento larval pode acarretar em mudanças nesse equilíbrio hormonal. Pesquisas utilizando *Dr. melanogaster* como modelo demonstraram que quando um disco imaginal é danificado ou tem um

crescimento descontrolado (caracterizando tumores), há extensão do tempo larval e atraso da metamorfose (Simpson et al. 1980; Menut et al. 2007). Com isso, foi levantada a hipótese de que esses discos danificados secretariam um fator que sinalizaria o atraso através da redução da produção de E20 (Halme et al. 2010; Hackney et al. 2012).

Apenas em 2012, dois grupos independentes identificaram a proteína Insulin-Like Peptide 8 (Ilp8) como a responsável por essa sinalização em *Dr. melanogaster* (Colombani et al. 2012; Garelli et al. 2012). Como descrito por Garelli e colaboradores (2012), o peptídeo Ilp8 pertence a uma família de proteínas similares à insulina/relaxina (composta por oito proteínas – Ilp1 a 8), devido a um espaçamento típico de cisteínas na sequência de aminoácidos. Durante o processamento da proteína, as cisteínas favorecem a formação de pontes dissulfeto resultando em uma estrutura similar à insulina, com a produção de um peptídeo C (Figura 3). Quando um disco imaginal sofre dano tecidual, ele passa a secretar Ilp8. Essa proteína é reconhecida no sistema nervoso central por meio da ativação dos receptores Lgr3 (Colombani et al. 2015; Garelli et al. 2015; Vallejo et al. 2015) e atua negativamente na regulação da secreção de ecdisona, provavelmente por meio da inibição da sinalização do PTTH. A consequência dessa regulação é a extensão do tempo larval, o que garante tempo necessário para que o disco imaginal danificado se recupere por meio de proliferação celular. Além disso, a ausência de Ilp8 durante o desenvolvimento normal acarreta em aumento na assimetria das estruturas formadas pelos discos, sugerindo que essa proteína também é responsável por regular o crescimento sincronizado dos discos imaginais (Garelli et al. 2012).

Ainda que o mecanismo de inibição da síntese de ecdisona por meio da sinalização por parte do Ilp8 não esteja totalmente esclarecido em *Dr. melanogaster*, é inegável a importância desse gene na regulação do atraso frente ao dano tecidual. Além disso, uma vez que esse mecanismo regulatório é amplamente observado em insetos, é possível que o gene *ilp8* tenha um papel fundamental nessas espécies.

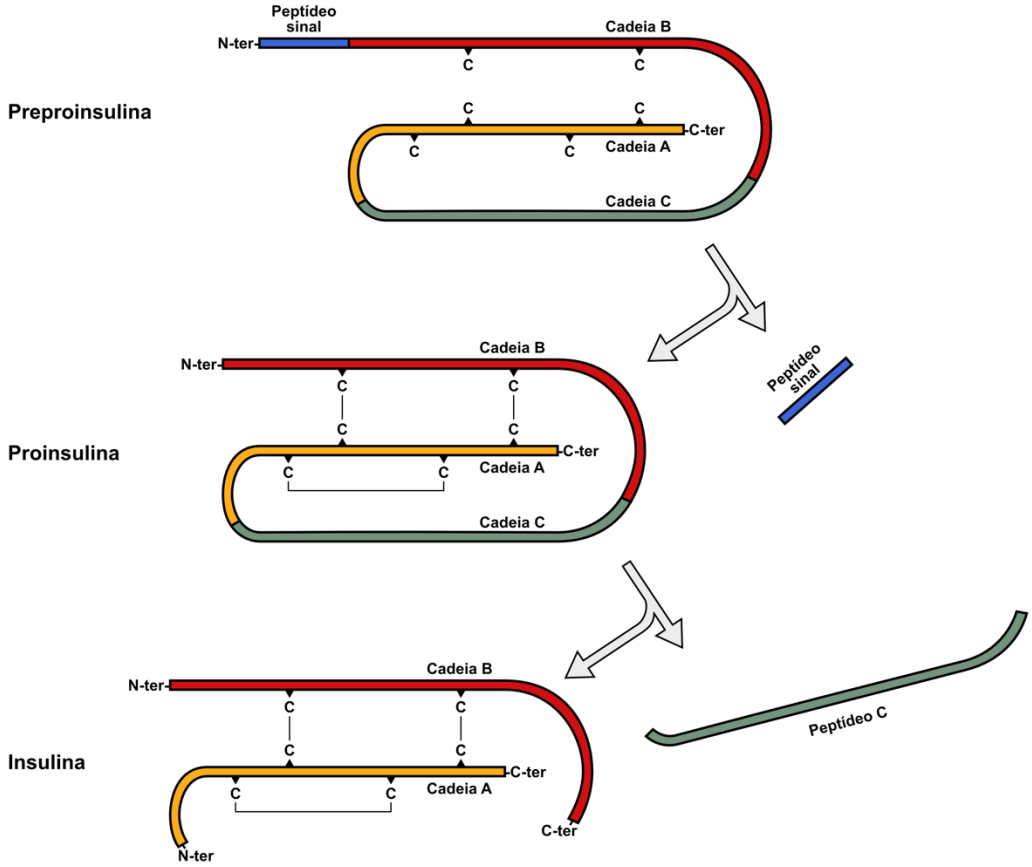


Figura 3: Processamento de insulina/relaxina: Insulina/relaxina são codificadas como preprohormônios que necessitam de etapas de processamento para retirada do peptídeo sinal e da cadeia C, secretada junto com a proteína madura. O "C" representa o aminoácido cisteína (editado de Støy et al. 2007).

Expressão gênica

Perturbações estocásticas ocorrem naturalmente durante o desenvolvimento, ocasionando uma manutenção basal de concentrações da proteína Ilp8, que regula as consequências dessas flutuações. Entretanto, danos nos discos imaginais acarretam em um aumento na concentração de Ilp8. Considerando a estrutura de um gene eucarioto, o controle da expressão gênica e, consequentemente, da síntese de proteína, pode ser realizado por meio de inúmeros processos que incluem: (1) regulação da taxa de transcrição do gene, (2) regulação do processamento do RNA mensageiro (mRNA) primário, (3) regulação da taxa de tradução do mRNA e (4) processamento e ativação ou inativação da proteína final (Alberts et al. 1999). Dessas quatro formas de controle, a

regulação da taxa de transcrição tem, provavelmente, um papel proeminente na quantidade de produto final (proteínas ou RNAs funcionais).

De forma simplificada, a transcrição de um gene se inicia por meio do reconhecimento da sua região promotora pela RNA polimerase, enzima responsável por sintetizar o mRNA primário a partir do sítio de iniciação. Entretanto, é necessária também a presença de sequências regulatórias responsáveis por ativar ou inibir o gene. Essas sequências são reconhecidas por proteínas específicas capazes de ligarem-se a elas e com isso direcionar a RNA polimerase para o promotor do gene correspondente (Alberts et al. 1999). Nesse contexto, a ativação das sequências regulatórias pode promover o aumento da expressão gênica em resposta a determinada situação/estímulo.

Ainda que não se saiba muito a respeito da regulação do gene *ilp8*, evidências sugerem que ele responda positivamente frente à ativação das vias de proliferação celular em *Dr. melanogaster* (Katsuyama et al. 2015). A recuperação de discos imaginais danificados por meio da proliferação celular é iniciada a partir da sinalização da via metabólica *Jun N-terminal kinase* (JNK) (Bosch et al. 2005; Mattila et al. 2005). Em seu trabalho, Katsuyama e colaboradores (2015) demonstraram que a ativação da via JNK estimula as vias JAK/STAT e *Wingless*, que, conjuntamente, proporcionam a regeneração e proliferação celular. Entretanto, além de controlar a resposta no disco imaginal, eles também documentam que a via JAK/STAT é responsável por controlar a comunicação entre o tecido danificado e o resto do organismo, por meio da regulação da expressão do *ilp8*. Nesse contexto, ainda que não completamente esclarecido, intermediários da via JAK/STAT devem atuar como elementos regulatórios para a expressão do gene *ilp8*.

Perguntas e hipóteses

Considerando a importância do atraso na metamorfose no controle do desenvolvimento, este trabalho busca compreender como esse mecanismo regulatório evoluiu ao longo da história dos insetos holometábolos. Partindo do conhecimento do gene *ilp8* de *Dr. melanogaster*, três perguntas devem ser consideradas:

- Quando o gene *ilp8* surgiu na filogenia dos holometábolos?

Como já discutido, esse gene foi primeiramente descoberto em *Dr. melanogaster*, mas sabe-se que ele não é exclusivo dessa espécie. Durante a descrição do *ilp8*, Garelli e colaboradores analisaram sequências homólogas a esse gene presente nas 12 espécies de *Drosophila* com genoma sequenciado bem como nos dípteros *Glossina morsitans* (mosca tsé-tsé) e *Rhagoletis pomonella* (conhecida como "larva da maçã") (Garelli et al, 2012; material suplementar). Assim, pode-se supor que outras espécies de holometábolos possuam homólogos ao gene *ilp8*, sendo sua origem anterior à evolução do grupo Drosophila. Entretanto, uma hipótese alternativa seria de que a origem do gene *ilp8* ocorreu durante a diferenciação de Schizophora. Nesse contexto, é importante avaliarmos quais espécies de insetos apresentam homólogos a esse gene.

- Este gene sempre esteve associado com a regulação do desenvolvimento?

É importante considerarmos que a função do *ilp8* só foi descrita em *Dr. melanogaster*, *Drosophila mauritiana* e *Drosophila sechellia* (Garelli et al, 2012; material suplementar) de modo que, mesmo partindo da hipótese de que outras espécies de insetos possuem homólogos a ele, não se sabe se ele sempre esteve relacionado com danos nos discos imaginais ou se possuía outra função primariamente. Uma das hipóteses é que o *ilp8* esteja associado à regulação do atraso na metamorfose desde sua origem. Por outro lado, a relação entre o gene *ilp8* e o atraso pode ser apenas uma inovação de *Drosophila*, e que este mecanismo seja mediado independentemente do *ilp8* em outras espécies, mesmo naquelas que possam ter homólogos a este gene.

- Como o atraso da metamorfose é regulado na ausência do gene *ilp8*?

Ainda que a hipótese inicial suponha que o *ilp8* esteja presente em diversas espécies de insetos, é de se esperar que algumas delas não apresentem homólogos a este gene. Entretanto, sabe-se que o atraso na metamorfose frente a um dano tecidual é um mecanismo amplamente observado em insetos holometábolos. Neste caso, é importante determinarmos como a regulação desse mecanismo ocorre na ausência do *ilp8*. É possível que nessas espécies o atraso seja sinalizado pela mesma via, porém dependente de uma outra proteína. Por outro lado, o atraso pode ser regulado por um sistema completamente diferente do observado em *Dr. melanogaster*. Em ambos os casos é essencial o estudo utilizando espécies que não possuam o gene *ilp8*.

Organização da dissertação

Para começarmos a responder essas questões foi realizada, primeiramente, uma busca pelo *ilp8* em dados públicos, para determinarmos quais espécies apresentam seu homólogo (capítulo 1). A partir desse resultado, foram selecionadas quatro espécies-chaves para realizarmos ensaios de dano tecidual associados com experimentos de RNA-seq, de modo a avaliarmos quais genes são diferencialmente expressos na condição com dano tecidual (capítulo 2).

Capítulo

1

EVOLUÇÃO DO GENE *ilp8*

Capítulo 1

Evolução do gene *ilp8*

Este capítulo descreve a busca por homólogos do gene *ilp8* em dados públicos de diferentes espécies de insetos. Considerando sua importância na regulação do desenvolvimento, e dado que é possível encontrar ortólogos desse gene nas 12 espécies de *Drosophila* que tiveram o genoma sequenciado, pode-se supor que a origem *ilp8* seja anterior ao surgimento de *Drosophila*.

Como discutido anteriormente, a proteína Ilp8 estrutura-se de forma similar à insulina (Garelli et al. 2012), sendo formada por duas cadeias proteicas independentes ligadas por pontes dissulfeto. Entretanto, a sequência desse gene também codifica a cadeia C que é eliminada da proteína final e pode acumular variações que não alterariam a função da proteína final, o que dificulta as buscas por meio de alinhamentos. Nesse contexto, uma forma de lidar com essa situação é por meio da utilização de motivos proteicos para selecionar sequências estruturadas da mesma maneira que o *ilp8*.

De forma geral, um motivo proteico refere-se a uma sequência de aminoácidos conservada em um dado grupo de proteínas (Bork & Koonin 1996). A maioria dos motivos são identificados a partir de alinhamentos de sequências da mesma família proteica e identificação de aminoácidos e/ou distâncias conservadas. Desse modo, a utilização de motivos na busca e caracterização de uma sequência permite predizer funções da proteína gerada ou alocá-la a um grupo funcional sem considerarmos a variação na sequência específica de cada espécie ou grupo taxonômico.

Entretanto, para serem realizadas buscas com motivos proteicos é necessário que haja um banco de dados com sequências genômicas ou transcriptomicas disponíveis, ou gerar essas sequências. Devido à redução dos custos de sequenciamento em larga escala ao longo dos anos e não ser mais necessário conhecimento prévio acerca do genoma das espécies, as novas tecnologias de sequenciamento permitiram que espécies não-modelo

se tornassem foco de estudo (Heather & Chain 2016). Uma vez que grande parte dessas sequências estão disponíveis publicamente, a busca por homólogos de qualquer gene é facilitada pois não é necessário gerar dados genômicos para o início das análises.

Este é o cenário ideal para começarmos a compreender a história evolutiva do gene *ilp8*. Para determinarmos quais espécies possuem homólogos a esse gene, foram utilizadas buscas direcionadas por motivos proteicos em genomas e transcriptomas disponíveis publicamente. Com isso, testamos a hipótese de que o *ilp8* havia surgido antes da diferenciação de Schizophora.

Metodologia

1.1. Obtenção das sequências

A fim de compreendermos a origem do gene *ilp8*, foram utilizados dados públicos para buscarmos homólogos a esse gene em diversas espécies de insetos. Para isso, foram utilizada as sequências depositadas na página do 1KITE (1K Insect Transcriptome Evolution – http://1kite.org/), que consiste em uma iniciativa internacional para realizar o sequenciamento do transcriptoma de mil espécies de insetos (Misof et al. 2014).

No momento da busca (fim de 2015 e início de 2016), o consórcio disponibilizava 103 transcriptomas de insetos de diversos grupos taxonômicos. Além disso, foram utilizados transcriptomas de espécies de dípteros disponíveis no laboratório por ocasião de outros projetos, como os califorídeos *Chrysomya albiceps*, *Chrysomya megacephala*, *Cochliomyia hominivorax* e *Cochliomyia macellaria* (transcriptomas larvais); *Hermetia illucens* (transcriptoma de fêmeas adultas virgens); a popularmente conhecida como mosca do berne, *Dermatobia hominis* (transcriptoma larval); e *Oestrus ovis* (transcriptoma larval), popularmente chamada de "bicho da cabeça".

Foi necessário prever as sequências de aminoácidos a partir dos transcriptomas de todas as espécies. Foi utilizado o programa TransDecoder versão 3.0.1 (Haas et al. 2013) para identificar possíveis regiões codificantes (CDS) nos transcriptomas disponíveis. Foram utilizados os parâmetros padrão do programa.

Por fim, também foram utilizadas sequências proteicas de 38 espécies de dípteros vetores disponíveis no repositório do VectorBase (http://www.vectorbase.org/downloads) além de 69 espécies de insetos holometábolos disponíveis no portal do NCBI (http://www.blast.ncbi.nlm.nih.gov/genome). A busca foi realizada com o termo *Holometabola* (que contém 264 entradas) e só foram utilizados os dados de espécies que apresentavam as sequências de aminoácidos descritas e que ainda não haviam sido amostradas nos outros bancos utilizados 1KITE, Flybase e VectorBase.

1.2. Homologia ao *ilp8*

É muito difícil a identificação de homólogos ao *ilp8* apenas por alinhamentos, pois a cadeia C da proteína Ilp8 acumula muita variação, uma vez que ela é eliminada da proteína final, fator que diminui a identidade dos alinhamentos. Devido a isso, a busca foi realizada por meio de motivos proteicos utilizando um script Perl desenvolvido para este propósito. Foi utilizado o seguinte motivo para selecionar possíveis sequências pertencentes à família das Ilps:

C-{C}(8,13)-C-{C}(3,50)-[RKH]-[RKH]-{C}(15,250)-[RKH]-[RKH]-

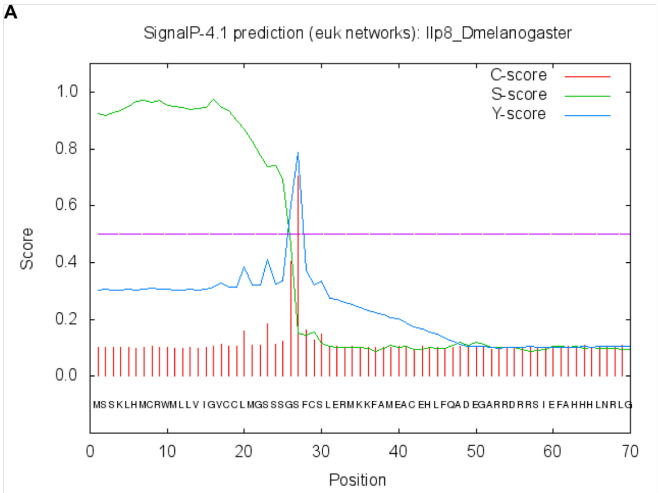
{C}(3,50)-C-C-{C}(2,13)-C-{C}(4,13)-C.

Na representação dos motivos, cada aminoácido está separado pelo hífen; as letras C, R, K e H representam, respectivamente, os aminoácidos cisteína, arginina, lisina e histidina; os números entre parênteses representam a quantidade mínima e máxima de repetições que o aminoácido anterior pode ter; entre as chaves estão os aminoácidos que não podem estar contidos no segmento; entre colchetes estão os aminoácidos que devem estar contidos no seguimento. As cisteínas essenciais para a formação da forma final da proteína estão destacadas em vermelho; já os aminoácidos necessários para o processamento da cadeia C estão marcados em verde.

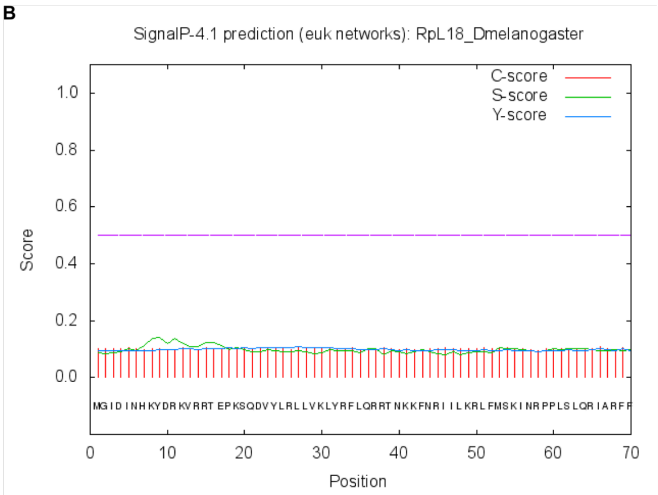
Após a recuperação de possíveis sequências homólogas à proteínas da família ILP utilizando o motivo acima, foi verificada a presença de um peptídeo sinal utilizando o programa SignalP versão 4.1 (Petersen et al. 2011). De forma simplificada, o programa SignalP gera, para cada aminoácido da sequência, uma pontuação referente ao reconhecimento de uma posição de clivagem (*C-score*) e a separação dos aminoácidos pertencentes ao peptídeo sinal da proteína madura (*S-score*). A combinação das duas pontuações (por média geométrica – *Y-score*) indica a presença do peptídeo sinal, caso atinja o limite estipulado (Figura 1.1).

Por fim, foram realizados alinhamentos no portal do NCBI (http://www.blast.ncbi.nlm.nih.gov/Blast.cgi; Camacho et al. 2009) utilizando o algoritmo blastp contra o banco de proteínas não redundantes, “nr”, para verificar se as sequências

recuperadas já não possuíam uma identidade ou função conhecida. Além disso, as sequências recuperadas foram alinhadas contra a proteína Ilp8 de *Dr. melanogaster* (obtido no repositório do FlyBase: http://flybase.org/), para confirmar ou não a identidade.




```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    27       0.705
  max. Y    27       0.787
  max. S    16       0.973
  mean S     1-26    0.887
       D     1-26    0.841   0.450   YES
Name=Ilp8_Dmelanogaster SP='YES' Cleavage site between pos. 26 and 27: SSG-SF D=0.841
D-cutoff=0.450 Networks=SignalP-noTM
```




```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    66       0.108
  max. Y    27       0.108
  max. S     9       0.142
  mean S     1-26    0.106
       D     1-26    0.107   0.450   NO
Name=RpL18_Dmelanogaster        SP='NO' D=0.107 D-cutoff=0.450 Networks=SignalP-noTM
```

Figura 1.1: Exemplo da análise realizada pelo SignalP: O perfil da proteína Ilp8 de *Dr. melanogaster* (A), foi utilizada como exemplo de proteína com peptídeo sinal (Garelli et al. 2012), e o resultado da proteína ribosomal RpL18 (B), como exemplo sem peptídeo sinal (Alonso and Santare 2006). Ambas as sequências foram obtidas no repositório do FlyBase.

Resultados

1.3. Processamento dos dados do 1KITE

De modo a determinarmos quais espécies apresentam homólogos ao gene *ilp8*, foram utilizados os transcriptomas disponíveis no repositório do 1KITE para o início das buscas. No momento da pesquisa, o repositório contava com 103 transcriptomas distribuídos em 29 ordens de insetos, 3 ordens basais de hexapoda e 3 ordens externas a hexapoda.

Uma etapa necessária para a busca utilizando motivo proteico foi prever quais transcritos codificam proteínas e obter essas sequências de aminoácidos. A quantidade de transcritos em cada transcriptoma variou de 4.236 a 45.581, com, em média, 62% codificando pelo menos uma proteína. É interessante ressaltar casos em que o número de transcritos codificantes foi maior que o número total de sequências: espécies *Folsomia candida*, *Aleochara curtula*, *Anurida marítima* e *Stylops melittae*. O valor máximo da razão codificantes/total foi de 133% (para a espécie *Stylops melittae*). Nestes casos, o TransDecoder conseguiu identificar mais de uma sequência codificadora para um mesmo transcrito.

1.4. Busca por homólogos

Utilizando o motivo proteico desenhado para proteínas da família das ILPs, foi possível recuperar ao menos uma sequência em 54 espécies das 103 disponíveis do 1KITE, totalizando 96 transcritos (Figura 1.2, Anexo T1). Essas sequências foram encontradas em quase todos os grupos de Hexapoda, exceto em Diplura, Dermaptera, Embioptera, Mantodea, Thysanoptera, Megaloptera, Neuroptera e Strepsiptera. A partir do alinhamento desses transcritos contra o banco de dados do NCBI, 45 sequências apresentaram similaridade com proteínas relacionadas à insulina: quatro sequências alinharam diretamente contra uma insulina de espécies variadas; seis sequências alinharam contra precursores da insulina; e seis sequências alinharam contra a proteína Bombyxin e 29 contra a proteína LIRP (*Locusta insulin-related peptide*).

Tanto Bombyxin quanto LIRP são proteínas da família da insulina caracterizados nas espécies de lepidoptera *Bombyx mori* e gafanhoto *Locusta migratoria* respectivamente

(Kromer-metzger & Lagueux 1994; Gu et al. 2015). Além disso, foi possível identificar um peptídeo sinal em 50% das sequências encontradas, indicando que essas proteínas são de fato secretadas da célula (Anexo T1). A ausência de um peptídeo sinal nas outras sequências pode indicar que a proteína não é secretada da célula ou que o transcrito está incompleto (observação comum em trancriptomas montados com sequências curtas).

Apesar dessas informações, não é possível afirmar que esses 96 transcritos são homólogos ao *ilp8* uma vez que nenhum deles alinhou contra as sequências de *Drosophila*. Porém, 18 sequências ou não alinharam contra o banco de dados, ou alinharam contra sequências não caracterizadas, representando, portanto, um potencial para estudo.

Ao analisarmos o resultado das buscas nas sequências do VectorBase, foi possível obter de 1 a 15 sequências para cada uma das 38 espécies analisadas, totalizando 166 sequências recuperadas com esse motivo (Anexo T2). Quando alinhadas contra o peptídeo Ilp8 de *Dr. melanogaster*, apenas quatro foram alinhadas: duas de *Stomoxys calcitrans*, uma de *Musca domestica*, e uma de *Glossina brevipalpis* (Anexo F1). Já das sequências provenientes do NCBI foi possível obter de 1 a 28 sequências para cada uma das 69 espécies, totalizando 652 sequências recuperadas (Anexo T3). Apesar disso, apenas sete sequências alinharam contra o Ilp8: *Rhagoletis zephyria*, *Bactrocera oleae*, *Zeugodacus cucurbitae*, *Bactrocera latifrons*, *Bactrocera dorsalis*, *Ceratitis capitata* e *Lucilia cuprina* (Anexo F2). Similarmente ao que aconteceu com as sequências do 1KITE, foi identificado um peptídeo sinal em 65% e 37% das sequências recuperadas do VectorBase e NCBI, respectivamente. É importante detacar que diversas sequências recuperadas tanto do VectorBase quanto do NCBI são reconhecidas como proteínas da família da insulina, como aconteceu com a anotação das sequências do 1KITE.

Olhando especificamente para os transcriptomas de dípteros disponíveis em nosso laboratório, a busca pelo motivo de proteínas da família das ILPs retornou em média 42 sequências por espécie (Tabela 1.1). Nenhuma das sequências de *Co. hominivorax*, *Co. macellaria* e *De. hominis* alinharam contra o *ilp8* de *Dr. melanogaster*. Já para as outras espécies, foi possível recuperar pelo menos uma sequência homóloga ao *ilp8*; no caso de

H. illucens foram recuperadas duas sequências (Anexo F3). Todas as cinco sequências anotadas como *ilp8* possuem um peptídeo sinal, previsto pelo SignalP, indicando que essas sequências são secretadas.

Tabela 1.1: Quantidade de sequências recuperadas utilizando o motivo.

Espécie	Estágio	Nº de sequências
Ch. albiceps	Larval	29
Ch. megacephala	Larval	32
Co. hominivorax	Larval	36
Co. macellaria	Larval	51
De. hominis	Larval	67
Oe. ovis	Larval	53
H. illucens	Adulto (fêmeas)	27

É importante notar que os alinhamentos das sequências recuperadas contra a proteína Ilp8 de *Dr. melanogaster* apresentaram uma grande variação na similaridade, ficando em torno de 40% a 74% de similaridade (que considera os aminoácidos iguais ou pertencentes a uma mesma classe; Tabela 1.2). Levando em consideração que a similaridade do *ilp8* entre as 12 drosófilas sequenciadas varia de 100% (quando considerando as espécies dentro do subgrupo *melanogaster*) a 73% (Tabela 1.2; Anexo F4), consideramos essas sequências como sendo homólogas ao *ilp8*.

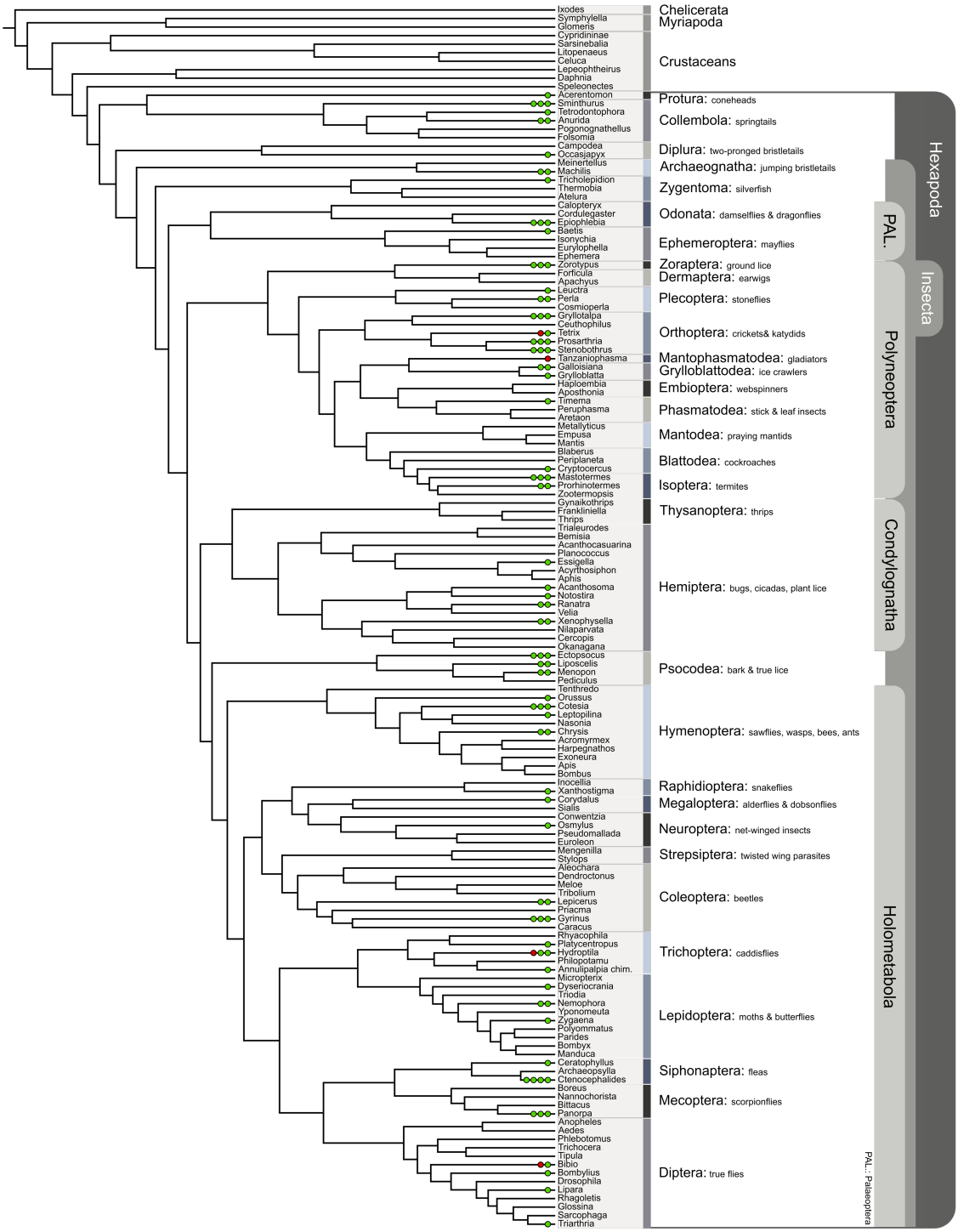


Figura 1.2: Gêneros com sequência recuperada do banco do 1KITE com a utilização do motivo proteico: Cada marcação representa uma sequência recuperada no táxon. Marcações verde representam sequências que foram anotadas; já marcações em vermelho representam sequências que não tiveram alinhamento. Filogenia adaptada de Misof e colaboradores (2014).

Tabela 1.2: Identidade dos alinhamentos contra a proteína Ilp8 de *Dr. melanogaster*.

Origem da sequência	Família	Espécie	Identidade	Similaridade	E-value
FlyBase	Drosophilidae	*Drosophila simulans*	99%	100%	2e-114
		Drosophila sechellia	99%	100%	3e-112
		Drosophila erecta	93%	99%	2e-108
		Drosophila yakuba	91%	95%	2e-103
		Drosophila ananassae	77%	91%	9e-94
		Drosophila permisilis	74%	85%	6e-88
		Drosophila pseudoobscura	74%	85%	6e-88
		Drosophila willistoni	66%	79%	7e-67
		Drosophila grimshawi	57%	73%	2e-62
		Drosophila virilis	54%	70%	1e-51
		Drosophia mojavensis	55%	73%	2e-51
Laboratório	Calliphoridae	*Ch. albiceps*	49%	65%	5e-38
		Ch. megacephala	47%	60%	8e-33
	Oestridae	*Oe. ovis*	45%	62%	1e-35
	Stratiomyidae	*H. illucens* (c14571_g1_i2)	29%	47%	2e-13
		H. illucens (c11423_g3_i1)	29%	47%	8e-13
VectorBase	Glossinidae	*G. brevipalpis*	43%	59%	1e-33
	Muscidae	*Mu. domestica*	47%	67%	1e-38
		S. calcitrans (SCAU012402-PA)	46%	64%	8e-39
		S. calcitrans (SCAU012402-PB)	46%	64%	8e-39
NCBI	Tephritidae	*Ba. dorsalis*	53%	73%	3e-52
		Ba. latifrons	53%	73%	2e-52
		Ba. oleae	54%	73%	8e-53
		Ce. capitata	54%	73%	8e-52
		R. zephyria	56%	74%	2e-53
		Z. cucurbitae	54%	72%	8e-53
	Calliphoridae	*Lu. cuprina*	45%	60%	2e-31

Discussão

Para compreendermos como o gene *ilp8* e a regulação do desenvolvimento evoluíram, é necessário primeiramente determinarmos quando este gene surgiu durante a evolução dos insetos. Considerando a importância do mecanismo de atraso como regulador do desenvolvimento, é de se esperar que diversas espécies de insetos possuam homólogos a esse gene. Por outro lado, como só foram identificados homólogos ao *ilp8* em *Drosophila*, *Glossina* e *Rhagoletis* é possível que este gene tenha surgido durante a diferenciação de Shizophora.

Para testar essas hipóteses, foram realizadas buscas ao *ilp8* guiadas por um motivo proteico em diversos transcriptomas de espécies de insetos disponíveis nos repositórios 1KITE (Misof et al. 2014), VectorBase (http://vectorbase.org/) e NCBI (http://ncbi.nlm.nih.gov/), bem como de sete dípteros disponíveis em nosso laboratório. As sequências recuperadas foram alinhadas tanto contra a sequência do *ilp8* de *Dr. melanogaster* quanto o banco de dados do NCBI, para avaliarmos sua identidade.

A partir dos dados do 1KITE foi possível recuperar sequências em 54 das 103 espécies com transcriptomas disponíveis. Muitas delas alinharam contra proteínas associadas à insulina, principalmente as proteínas *Locusta insulin-related peptide* (LIRP) e Bombyxin. A proteína LIRP foi primeiramente caracterizada na espécie de gafanhoto *Lo. migratoria* (Insecta, Orthoptera) e é considerada um neuro-hormônio pertencente à família das insulinas, atuando como fator de crescimento e controle do metamolismo (Lagueux et al. 1990; Goltzené et al. 1992). Por sua vez, a proteína Bombyxin foi caracterizada na espécie *Bo. mori* (Insecta, Lepdoptera), popularmente conhecida como “bicho da seda”. Inicialmente, a proteína foi erroneamente reconhecida como um homólogo ao hormônio PTTH. Entretanto, a ausência de resposta por parte das glândulas prototorácicas de *Bo. mori* quando estimuladas com essa proteína levou os pesquisadores a identificá-la como sendo estruturalmente semelhante à insulina dos vertebrados (revisado em Mizoguchi & Okamoto 2013).

Apesar disso, nenhuma das sequências apresentou semelhança com a proteína Ilp8 de *Dr. melanogaster*. Como a regulação do desenvolvimento depende da expressão do gene *ilp8* durante a fase larval, a ausência desse gene em amostras do 1KITE pode estar relacionado com o estágio de desenvolvimento e condição das amostras utilizadas para montar os transcriptomas. Apesar disso, a expressão do gene *ilp8* não é exclusiva do estágio larval: fêmeas adultas também apresentam níveis de Ilp8 nos ovários (Garelli et al. 2012). Nesse sentido, ainda que não seja mandatório, a utilização de transcriptomas larvais facilita a identificação desse gene.

Este fato é confirmado ao observarmos os dados dos transcriptomas disponíveis no nosso laboratório. Dos sete transcriptomas analisados, seis foram obtidos de espécies em seu estágio larval, o que deve ter auxiliado na identificação do *ilp8* em 3 espécies (*Ch. albiceps*, *Ch. megacephala* e *Oe. ovis*). No caso da sétima espécie, *H. illucens*, foi utilizado um transcriptoma de fêmeas adultas, mas ainda assim foi possível identificar homólogos ao *ilp8*, provavelmente devido à sua expressão nos ovários de fêmeas adultas. Além disso, com as análises dos dados disponíveis no VectorBase e NCBI, foi possível identificar esse gene em mais 11 espécies de dípteros, distribuídas nos grupos Tephritidae, Glossinidae, Muscidae e Calliphoridae (Figura 1.3). O mais interessante é que não foi possível encontrar homólogos ao *ilp8* em alguns clados de Diptera, como *Aedes* e *Anopheles*.

Ainda que esses dados apontem o *ilp8* como uma inovação de Brachycera, não podemos descartar a hipótese de que ele seja mais antigo. Em sua revisão, Gontijo e Garelli (2018) argumentam que uma vez que o par Ilp8-LGR3 estão na mesma via de sinalização (Vallejo et al. 2015), pode-se supor que espécies que possuem homólogos ao LGR3 devam tem uma proteína ligante similar à Ilp8. Nesse sentido, eles propõem que a origem do *ilp8* seria bem mais antiga do que a diferenciação dos dípteros, uma vez que é possível achar homólogos ao receptor LGR3 em diversos grupos de artrópodes, como Chelicerata, Crustacea, Polyneoptera e Holometabola, onde os dípteros são inseridos.

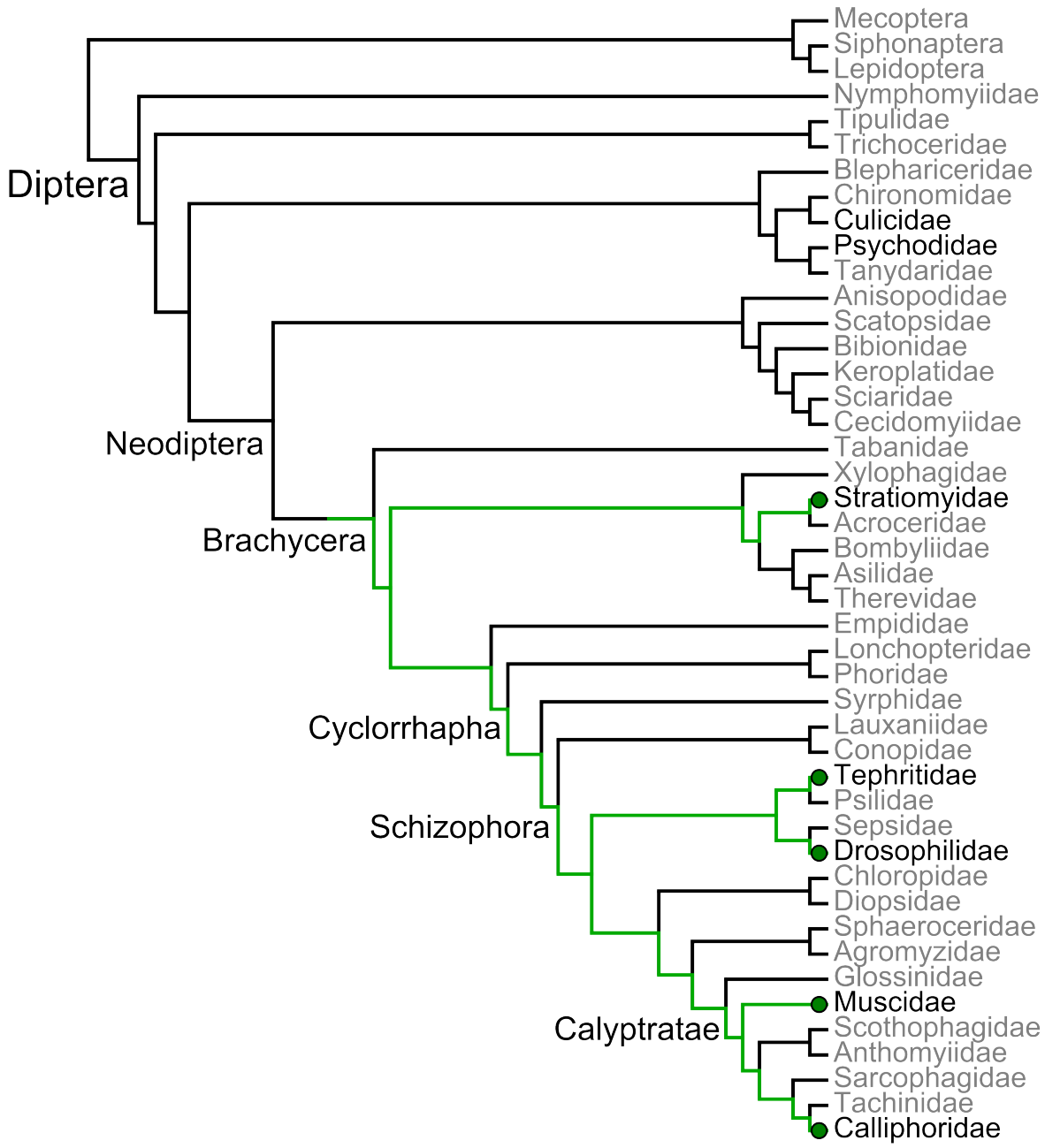


Figura 1.3: Hipótese evolutiva do gene *ilp8*: Espécies marcadas em preto representam dípteros que tiveram suas sequências analisadas; círculos verdes indicam os clados em que foi possível identificar homólogos ao gene *Ilp8*. Apesar do ramo Oestridae (*Oe. ovis*) não estar na filogenia, essa família faz parte do grupo Calyptratae (Cerretti et al. 2017). Filogenia adaptada de Wiegmann e colaboradores (2011) – material suplementar.

Algo que pode contribuir para a dificuldade de encontrar homólogos ao *ilp8* é a rápida taxa de evolução de genes da família *Ilp*. Grönke e colaboradores (2010) avaliaram a taxa

de evolução dos sete genes pertencentes à família *Ilp* em *Drosophila* e concluíram que, de forma geral, as proteínas dessa família evoluem rapidamente. Em seu trabalho, Grönke não analisou os dados do *ilp8* pois ainda não havia sido caracterizado na época, porém é de se supor que ele siga o mesmo padrão da família gênica. Assim, o acúmulo de diferenças ao longo de sua história evolutiva dificulta nas buscas por seus homólogos baseado apenas em alinhamentos.

Portanto, nossa hipótese inicial de que outros grupos de holometábolos além de Schizophora também possuam homólogos ao *Ilp8* não foi rejeitada. Ainda assim, não podemos dizer com exatidão quando que este gene surgiu na história evolutiva dos insetos. Com nossos dados, nossa hipótese mantida é de que o gene *ilp8* tenha originado-se durante a diferenciação da subordem Brachycera (Figura 1.3).

Conclusão do capítulo

Neste capítulo, nós buscamos avaliar a partir de que ponto na filogenia dos insetos ocorreu o surgimento do gene *ilp8*. Considerando a importância desse gene na regulação do desenvolvimento, por meio do atraso na metamorfose, foi levantada a hipótese de que ele teria surgido antes da diferenciação do grupo Schizophora. Como ele só foi descrito em *Drosophila, G. morsitans* e *R. pomonella*, uma hipótese alternativa aponta o *ilp8* como sendo uma inovação de Schizophora.

Por meio de buscas direcionadas por motivos proteicos e associadas com alinhamentos contra a proteína Ilp8 de *Dr. melanogaster*, foi possível identificar possíveis homólogos a esse gene em diversas espécies de dípteros distribuídas nos grupos Stratiomyidae, Tephritidae, Glossinidae, Muscidae, Calliphoridae e Oestridae. Ainda que nossas análises não reijeitem nossa hipótese inicial e apontem esse gene como sendo uma inovação de Brachycera, dados sugerem que sua origem pode ser bem mais antiga.

Devido a sua relação com o receptor LGR3, a origem do *ilp8* pode ter ocorrido durante a diferenciação dos artrópodes, uma vez que o LGR3 está presente em diversos grupos taxonômicos de artrópodes. Apesar do *ilp8* ainda não ter sido identificado nesses grupos, supõe-se que exista uma proteína similar à Ilp8 nesses grupos que atuaria como ligante para o receptor.

Capítulo

2

REGULAÇÃO DO ATRASO DA METAMORFOSE

Capítulo 2

Regulação do atraso da metamorfose

Como foi discutido no capítulo anterior, é possível traçar a presença do gene *ilp8* pelo menos até a diferenciação de Brachycera. Entretanto, nada pode ser afirmado sobre sua função, uma vez que estudos funcionais com esse gene só foram realizados utilizando *Dr. melanogaster* como modelo de estudo. Neste capítulo será discutido o papel do *ilp8* na regulação do atraso na metamorfose em outras espécies de insetos, por meio do estudo comparativo dos perfis de expressão gênica diferencial entre larvas que sofreram danos teciduais e controles sem dano.

Existem diversas tecnologias que fornecem dados de expressão, porém os estudos com sequenciamento massivo em paralelo do transcriptoma (RNA-seq) são ideais quando considerando espécies não-modelo, pois não exigem conhecimento prévio acerca do seu genoma. Além disso, são geradas grandes quantidades de informações tanto sobre a sequência dos genes quanto sua expressão, sendo ideal para estudos de evolução gênica. Com essa abordagem, é possível avaliar se há divergência tanto na sequência codificante quanto na regulação da expressão. Por fim, é possível também identificarmos outros genes associados ao controle do desenvolvimento, uma vez que obtemos dados globais de expressão (Wang et al. 2009).

Assim, foram escolhidas quatro espécies para a realização dos experimentos: *Dr. melanogaster* pois ainda que o mecanismo de regulação mediado pelo *ilp8* tenha sido descoberto nessa espécie, ele não está completamente elucidado; *Ch. megacephala* e *H. illucens*, pois como descrito no capítulo anterior, possuem homólogos ao *ilp8*, sem informação acerca de sua função; e *Manduca sexta* (Insecta, Lepdoptera), uma vez que Lepdopteras não apresentam homólogos ao *ilp8* (capítulo anterior), sendo portanto um ótimo modelo para compreendermos como o mecanismo de atraso é regulado na ausência do *ilp8*.

Uma vez que foram estimulados danos teciduais, que podem ou não acarretar em um atraso durante desenvolvimento larval, é importante compreendermos como é o ciclo de vida dessas espécies em condições normais. De forma geral, *Dr. melanogaster* leva em torno de 10 dias para desenvolver-se de ovo à adulto. O embrião leva 24h para se desenvolver, a fase larval (dividida em três ínstares) dura de quatro a cinco dias e, por fim, a fase de pupa se estende por quatro dias até a eclosão do adulto (Chyb & Gompel 2013). Este ciclo de vida se assemelha muito ao da *Ch. megacephala*, com o embrião também se desenvolvendo em 24h, a fase larval durando cerca de cinco dias e com quatro dias necessários para a metamorfose da pupa em adulto (Siddiki & Zambare 2017). Já o ciclo de vida da espécie *H. illucens* difere bastante, sendo bem mais longo que os anteriores. Nessa espécie, o desenvolvimento embrionário pode levar até quatro dias para desenvolver-se por completo, enquanto a fase larval (que contém seis instares) se estende por aproximadamente 14 dias e por fim a pupa dura 10 dias (Hall & Gerhardt 2002). Por fim a espécie *Ma. sexta* também é caracterizada por possuir um ciclo de vida longo. Seu desenvolvimento embrionário pode levar até cinco dias, a fase larval (dividida em cinco ínstares) pode durar até 18 dias enquanto a fase de pupa leva 18 dias ou mais, caso a pupa entre em diapausa (Reinecke et al. 1980). Ainda que o ciclo de vida de cada espécie seja diferente, a comparação dos perfis de expressão é intraespecífica, de modo que as peculiaridades de cada ciclo não prejudicam as análises.

A comparação dos resultados das quatro espécies permite testar as hipóteses aqui levantadas. Considerando a hipótese da origem da proteína Ilp8 como sinalizadora do atraso durante a diferenciação de Brachycera, é esperado que o *ilp8* de *Ch. megacephala* e *H. illucens* estejam mais expressos nas larvas com dano tecidual, similarmente ao que ocorre em *Dr. melanogaster*. Por outro lado, caso o *Ilp8* não tenha a função de sinalização do atraso em todas as espécies, espera-se que as respostas de *Ch. megacephala* e *H. illucens* sejam diferentes de *Dr. melanogaster*. Nesta situação, e relacionando com a resposta de *Ma. sexta*, será possível identificarmos outros genes envolvidos em mecanismos reguladores do atraso, independentes de *Ilp8*.

Metodologia

De forma a ilustrar e facilitar a compreensão das etapas descritas neste capítulo, é possível observar na figura 2.1 um fluxograma das fases necessárias para a análise de dados de RNA-seq.

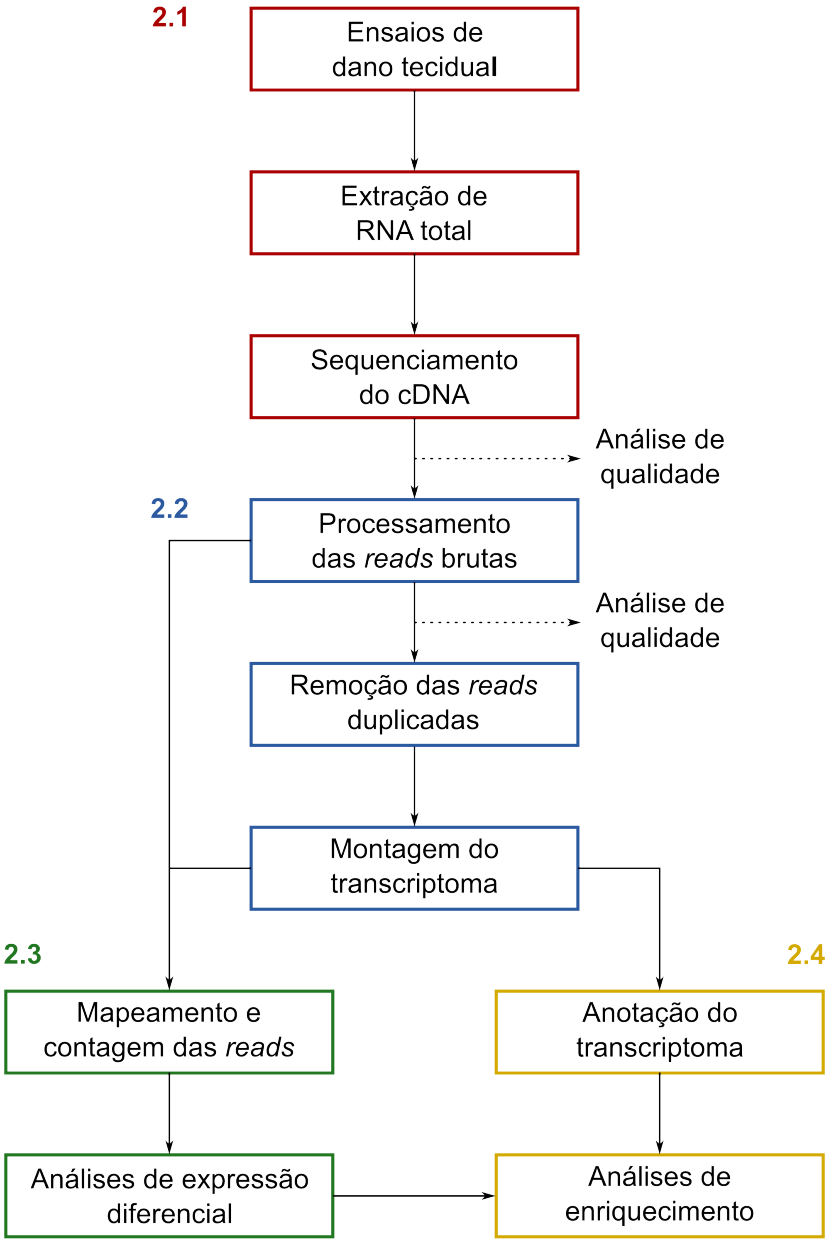


Figura 2.1: Análise de RNA-seq: Fluxograma com as etapas da análise de RNA-seq deste trabalho; os números representam os itens descritos neste capítulo.

2.1. Obtenção das amostras e sequenciamento

Os experimentos de dano tecidual foram baseados nos métodos descritos por Garelli e colaboradores (2012) e realizados colaborativamente em três laboratórios. Os ensaios com as espécies *Dr. melanogaster* e *H. illucens* foram realizados no Laboratório de Biomedicina Integrativa (Universidade Nova de Lisboa – Portugal), supervisionado pelo Dr. Alisson Gontijo. Já os ensaios com *Ch. megacephala* foram realizados no Laboratório de Entomologia Integrativa (Universidade de Campinas – Brasil), em colaboração com Profa. Dra. Patrícia Thyssen. Por fim, os ensaios com *Ma. sexta* foram realizados por integrantes do grupo *Development, Evolution and the Environment* (Dr. Takashi Koyama, Instituto Gulbenkian de Ciência – Portugal).

Para estimular danos teciduais foi utilizado o agente químico genotóxico etil metano sulfonato (EMS), que promove a alquilação das bases nitrogenadas acarretando em malpareamento (Griffiths et al. 2012). Células sofrendo divisão mitótica – como por exemplo as que compõe os discos imaginais no estágio larval (Worley et al. 2012) – são mais sucetíveis à ação do EMS (Sega 1984). De modo geral, essa adição do grupo alquila é reconhecida pelos sistemas de reparo das células, que por sua vez interrompem o ciclo celular para correção do dano estimulado. Entretanto, nos casos em que esse dano por alquilação é muito extenso – causado pela alta concentração do componete – a constante interrupção do ciclo celular acarreta na apoptose das células. Como controle negativo do experimento, foi utilizado o tampão fosfato-salino (PBS). Ambos os componentes foram adicionados na alimentação das larvas L3 de *Dr. melanogaster* e *Ch. megacephala*, estando o EMS a uma concentração de 10mM. Para os ensaios com *H. illucens* e *Ma. sexta* o EMS/PBS foi injetado nas larvas L3, uma vez que elas não se alimentaram do meio contendo o componente.

Após 24 horas do tratamento, foram coletadas três réplicas, cada uma contendo uma mistura de trêa a cinco larvas L3 sobreviventes para posterior extração de RNA – para *Dr. melanogaster* foram coletadas apenas duas réplicas. As larvas restantes foram utilizadas para monitorar o período larval, por meio da contagem de pupas formadas ao longo do tempo em cada condição. A identificação da extensão do período larval foi realizada por meio de análises de variância de um fator (*one-way* ANOVA) das distribuições das fases larvais em cada condição.

Devido ao tamanho das larvas de *Ma. sexta* (que podem chegar a 8 – 10 cm), foi possível obter dados de expressão específicos dos discos imaginais com mais facilidade. Nesse sentido, foi realizado um segundo experimento com larvas de *Ma. sexta*, em que estimulamos danos físicos ("beliscão" com pinça – *Pinch*) nos discos que formam as asas, por serem de mais fácil acesso. Como controle negativo, o "beliscão" foi realizado em uma das pernas da larva (*Sham*), pois as pernas não formam uma estrutura no adulto. Os discos imaginais que formam as asas foram dissecados de larvas de cada condição, três réplicas para cada.

As extrações de RNA total das amostras foram realizadas utilizando o reagente Trizol (Invitrogen) de acordo com as especificações do fabricante. Tanto a síntese da biblioteca de DNA complementar (cDNA) quanto o sequenciamento utilizando a técnica de extremidades pareadas (paired-end) foram realizados na Escola Superior de Agricultura Luiz de Queiroz (ESALQ – Universidade de São Paulo). Foi utilizado o protocolo TruSeq mRNA (Illumina) para construção das bibliotecas. Elas foram sequenciadas no equipamento HiSeq 2500 utilizando o kit HiSeq SBS v4, com leitura pareadas de 100 pb.

2.2. Processamento das sequências e montagem *de novo* dos transcriptomas

A primeira etapa após o sequenciamento das amostras foi a avaliação da qualidade das sequências brutas (*reads*) geradas. Essa análise de qualidade (média e por base) foi feita utilizando o programa FastQC versão 0.11.3 (Andrews, 2010). Seguiu-se então o processamento das *reads* para retirar bases com baixa qualidade ou resquícios de adaptadores, utilizando o pacote Trimmomatic versão 0.36 (Bolger et al. 2014). Os parâmetros utilizados para o processamento foram: valor mínimo de qualidade das extremidades igual a três, tamanho da janela de análise igual a quatro nucleotídeos, valor mínimo de qualidade média da janela igual a 15 e tamanho mínimo da sequência após os cortes de qualidade igual a 36 nucleotídeos. Após o processamento, a qualidade das *reads* foi reavaliada com o programa FastQC.

Como uma das características dos sequenciamentos de nova geração é a grande quantidade de sequências repetidas geradas, é importante reduzir essa redundância de modo a otimizar o tempo computacional para a montagem dos transcriptomas. Sendo assim, após passar pelo filtro de qualidade, as *reads* processadas foram colapsadas utilizando um *script* Perl escrito por ocasião de outro projeto. Esse *script* diminui a

redundância dos dados por meio da identificação de *reads* exatamente iguais e manutenção apenas da cópia que possui a maior qualidade média.

Com as sequências colapsadas seguiu-se então a montagem *de novo* dos transcriptomas de cada espécie. Esta etapa consistiu em gerar sequências contíguas (*contigs*) que podem representar transcritos inteiros ou parciais. Além disso, a grande vantagem da montagem *de novo* é tornar dispensável a utilização de um genoma de referência para a montagem, sendo, portanto, ideal para as espécies não-modelo. Para tanto, foi utilizado o montador Trinity versão 2.4.0 (Grabherr et al. 2011) que reconstrói os transcritos e suas possíveis isoformas por meio de grafos de Brujin. O tamanho mínimo dos *contigs* foi estabelecido em 200pb e foram mantidos os valores padrões para os outros parâmetros.

A completude de cada transcriptoma montado foi avaliada utilizando o programa BUSCO versão 2.0 (Simão et al. 2015). Ele utiliza bancos de dados contendo sequências ortólogas conservadas e avalia, por meio de alinhamentos, quantas sequências desse banco o transcriptoma montado consegue recuperar e em quais condições (cópia única ou duplicada, completas ou fragmentadas). Foram utilizados os parâmetros padrão do programa e o banco de sequência de Endopterygota (holometábolos), com um total de 2.442 sequências.

Todas as etapas foram realizadas para todas as espécies, exceto *Dr. melanogaster*. Por se tratar de uma espécie modelo e, portanto, possuir muita informação genômica disponível, para *Dr. melanogaster* só foram realizadas as etapas de processamento das *reads*. Os transcritos de cada gene foram obtidos no repositório do FlyBase (http://flybase.org/).

2.3. Análise de expressão diferencial

Para identificarmos os transcritos ou *contigs* diferencialmente expressos (TCDE) entre as condições de cada espécie foram necessárias três etapas: (1) mapeamento das *reads* de cada réplica no transcriptoma, (2) obtenção das contagens de *reads* de cada transcrito em cada réplica e construção das matrizes de expressão e (3) análise de expressão diferencial.

O mapeamento das *reads* processadas (mas não colapsadas) foi realizado utilizando o programa Bowtie2 versão 2.2.6 (Langmead & Salzberg 2012). Esse programa alinha grandes quantidades de dados contra um genoma/transcriptoma de referência de forma rápida e eficiente. Uma vez que a montagem do Trinity pode gerar mais de uma isoforma para um mesmo transcrito, o mapeamento foi realizado apenas contra a maior isoforma, para diminuir a quantidade de alinhamentos ambíguos das *reads*. Essa filtragem do transcriptoma foi realizada utilizado um *script* Perl do pacote Trinity. O mapeamento das *reads* foi realizado utilizando os parâmetros *gbar* (para evitar a formação de gaps), *no-mixed* (para evitar alinhamento de *reads* não pareadas) e *no-discordant* (para evitar alinhamentos discordantes entre as *reads* de um mesmo par). Com o mapeamento, foi possível obter as contagens de quantas *reads* alinharam contra cada transcrito ou *contig*. Para tanto, foi utilizado o programa RSEM versão 1.3.0 (Li & Dewey 2011), usando os parâmetros padrão.

Por fim, as análises de expressão das matrizes geradas foram realizadas utilizando os pacotes edgeR (Robinson et al. 2009) e DESeq2 (Love et al. 2014), do ambiente estatístico R. Especificamente para *H. illucens*, foi necessário realizar uma correção de viés de lote do sequenciamento (*batch effect*), uma vez que a terceira réplica foi sequenciada separadamente das outras. Essa análise foi baseada nos procedimentos descritos por Gilad e Mizrahi-Man (2015), que utiliza o pacote sva (Leek et al. 2012) – também para R – para identificação de *batchs* diferentes e normalização das contagens antes das análises de expressão.

Para a avaliação da expressão foram mantidas apenas as sequências que tiveram pelo menos uma *read* por milhão em pelo menos duas réplicas. A normalização do número de *reads* por sequência foi realizada pelo total de *reads* alinhadas por réplica e foi utilizado o teste exato de Fisher para a identificação dos TCDE, considerando o alfa de 0,05 para os valores de p corrigidos. A correção do valor de p foi feita utilizando o método de Benjamini-Hochberg, para controlar os falsos positivos decorrentes de testes múltiplos.

Assim sendo, para cada espécie foram realizadas duas análises de expressão diferencial, considerando as combinações da matriz de expressão com o pacote estatístico para identificação dos TCDE. Foi considerado como conjunto final de TCDE apenas aqueles compartilhados pelas duas combinações.

O conjunto de TCDE de cada espécie foram posteriormente utilizados em buscas de ortólogos diferencialmente expressos entre as espécies com base no roteiro de Yang e Smith (2014; https://bitbucket.org/yangya/phylogenomic_dataset_construction). De forma geral, o roteiro proposto pelos autores agrupa sequências putativamente homólogas por meio de alinhamentos recíprocos (*blast* e *Markov clustering* – MCL), infere árvores filogenéticas a partir desses alinhamentos e, por fim, remove parálogos. Para recuperar os ortólogos, foram consideradas apenas as árvores sem táxon duplicado, de modo que cada uma poderia ter, no máximo, quatro sequências. Além disso, foram selecionadas também árvores contendo pelo menos duas sequências (ortólogo compartilhado por no mínimo duas das quatro espécies).

2.4. Anotação dos transcriptomas e análises de enriquecimento

A anotação de um transcriptoma consiste em associar uma identidade ou função a cada um dos *contigs*, sendo, portanto, realizada apenas com as sequências das espécies não-modelo. Uma vez que o código genético é degenerado, é melhor anotar um *contig* com base em sua sequência proteica, uma vez que alterações na sequência de bases nem sempre acarretam em alterações na proteína final.

Nesse sentido, foi utilizado o programa TransDecoder versão 5.0.1 (Haas et al. 2013) para identificar possíveis CDS nos transcriptomas montados. Primeiramente foram obtidos os maiores quadros de leitura (*open reading frames* – ORFs) de cada *contig*. Com os ORFs foi possível realizar uma busca por similaridade proteica, realizada por meio de alinhamentos contra um banco de dados. Foi utilizado o banco de dados Pfam (Finn et al. 2016), que contêm diversas famílias proteicas já caracterizadas, e o programa Hmmer versão 3.1b2 (Eddy 1998) para realizar os alinhamentos implementando modelos probabilísticos baseados em Markov. Na última etapa, o TransDecoder foi utilizado para predizer as CDS dos transcriptomas a partir do resultado da busca por similaridade. Com as sequências proteicas de cada transcriptoma, foi possível fazer diversas predições com relação às suas estruturas, como a identificação de peptídeo sinal no início da sequência (SignalP versão 4.1 – Petersen et al. 2011) e domínios de transmembrana (Tmhmm versão 2.0 – Krogh et al. 2001). Todos os programas foram utilizados com os parâmetros padrão.

A anotação dos transcriptomas foi realizada por meio do recurso online AgBase (McCarthy et al. 2006 – agbase.msstate.edu), que fornece diversas ferramentas úteis para análises funcionais. Dentre elas, destaca-se a GOanna, que permite uma anotação por meio de alinhamentos similar ao Blast, mas que recupera também os termos do *Gene Ontology* (GO) associados ao gene. Os termos GO têm por princípio classificar genes em processo biológico, componente celular e função molecular, e com isso unificar bancos de dados e informações (geneontology.org). A anotação utilizando a GOanna foi feita com as sequências proteicas recuperadas pelo TransDecoder (algoritmo blastp) e com os transcriptomas completos (algoritmo blastx). As sequências foram anotadas usando cinco bancos de dados proteicos disponíveis (*AgBase-UniProt*, *AgBase_Community*, *SwissProt*, *UniProt* e *Invertebrates*), com uma cobertura e identidade mínima de 40% e E-value de 10e-05.

Além disso, os transcriptomas também foram anotados contra o banco de dados NCBI-nr, que contém as sequências proteicas não redundantes depositadas no NCBI, usando o programa GHOSTZ (Suzuki et al. 2015). Este programa realiza alinhamentos locais como o Blast, porém de forma muito mais rápida. Os alinhamentos também foram feitos usando os algoritmos blastp e blastx e a filtragem foi realizada após os alinhamentos utilizados os mesmos padrões do GOanna. Para transcritos que foram alinhados em mais de um banco de dados, foi mantida a anotação com a maior pontuação (*score*). Uma vez que o cálculo do *score* leva em consideração o total de bases alinhadas em relação ao número de bases malpareadas e espaçamento (*gaps*) presentes no alinhamento, ele representa o melhor fator para comparação de alinhamentos diferentes.

Com os transcritos anotados, foi possível realizar análises de enriquecimento considerando o conjunto de TCDE identificado para cada espécie. Para tanto, foi utilizado o programa Gorilla (Eden et al. 2009 – cbl-gorilla.cs.technion.ac.il) que utilizam os termos GO associados à anotação de cada transcrito para identificar processos, componentes ou funções enriquecidas no conjunto de TCDE.

Para identificar especificamente o gene *ilp8* no transcriptoma larval de *Ch. megacephala* e *H. illucens*, foram utilizadas as sequências descritas no capítulo anterior e realizados alinhamentos locais contra as sequências proteicas previstas pelo TransDecoder usando o programa Blast+ (algoritmo blastp; Camacho et al. 2009).

2.5. Tempo de desenvolvimento e avaliação dos transcriptomas

Com o intuito de avaliar o atraso na metamorfose, danos teciduais foram estimulados usando o componente genotóxico EMS. Larvas de *Dr. melanogaster*, *Ch. megacephala* e *H. illucens* atrasam a formação da pupa em resposta ao tratamento com o EMS (Figura 2.2). Especificamente para *H. illucens*, o tempo descrito no gráfico representa as horas após as injeções de cada componente, e não após a oviposição. No caso de *Ma. sexta*, após o tratamento, as larvas foram diretamente preparadas para extração de RNA total e sequenciamento, não sendo avaliado seu tempo de desenvolvimento. Entretanto, sabe-se que *Ma. sexta* também responde com atraso ao tratamento com EMS, similarmente às outras espécies (Dr. Takashi Koyama, comunicação pessoal).

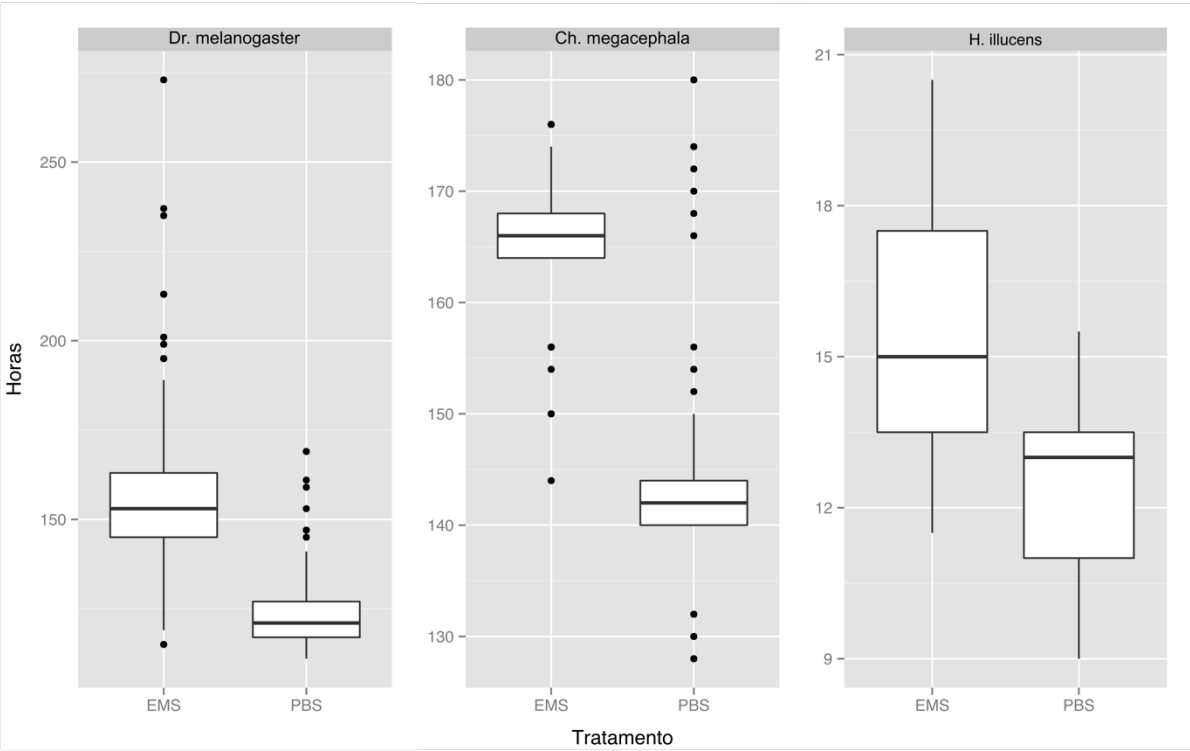


Figura 2.2: Resposta ao dano nos ensaios com EMS: Observa-se no gráfico o tempo para formação da pupa em cada tratamento (PBS ou EMS). Todas as espécies apresentaram um aumento no tempo de desenvolvimento após o tratamento com EMS. Todas as diferenças observadas foram significativas, com valores de $p < 0{,}001$.

Amostras de cada condição e espécie foram retiradas dos ensaios de dano para os experimentos de RNA-seq. O sequenciamento do cDNA das bibliotecas geradas a partir das réplicas geraram mais de 100 milhões de pares de sequências para cada espécie (Tabela 2.1). Após o processamento, foram mantidas mais de 95% de pares de sequências em todas as amostras. É interessante notar a taxa de *reads* duplicadas quando comparando o total de sequências brutas e após serem colapsadas: em torno de 50% para *Ch. megacephala* e *Ma. sexta*, e 30% para *H. illucens*.

Tabela 2.1: Resumo dos dados de sequenciamento e processamento das *reads*.

Espécie	Pares de sequências*				
	Brutas	Após corte de qualidade		Após remoção de duplicação	
Dr. melanogaster	131.758.329	126.710.242	96,17%	-----	-----
Ch. megacephala	118.658.716	113.837.799	95,94%	68.972.309	58,13%
H. illucens	176.849.148	172.283.744	97,42%	126.533.967	71,55%
Ma. sexta	176.560.005	174.255.957	98,69%	88.882.419	50,34%

* Somando todas as réplicas de cada espécie

As sequências colapsadas foram utilizadas para montar os transcriptomas de *Ch. megacephala*, *H. illucens* e *Ma. sexta* (Tabela 2.2). Ainda que foram montados mais *contigs* no transcriptoma de *H. illucens*, ele é estruturalmente parecido com o de *Ch. megacephala* considerando o número de isoformas e tamanho das sequências. É interessante notar que apenas 25% dos *contigs* de *Ma. sexta* representam isoformas alternativas, contra quase 50% de *Ch. megacephala* e *H. illucens*. Os transcriptomas também foram avaliados considerando sua completude, por meio de buscas por ortólogos no bando de dados de endopterygota do BUSCO (Figura 2.3). Nas três espécies foi observada uma grande quantidade de sequências completas e duplicadas, devido às isoformas alternativas identificadas na montagem. É interessante notar que o transcriptoma de *Ch. megacephala* e *H. illucens* possuem o mesmo padrão na recuperação dos ortólogos, enquanto *Ma. sexta* teve a menor quantidade de sequências recuperadas. Isso provavelmente acontece porque *Ch. megacephala* e *H. illucens* são espécies filogeneticamente mais próximas às espécies com maior número de genes sequenciados e presentes no banco do que de *Ma. sexta*.

Tabela 2.2: Resumo da montagem de cada transcriptoma.

Parâmetros	*Ch. megacephala*	*H. illucens*	*Ma. sexta*
Total de *contigs*	65.192	102.509	78.439
Maior isoforma*	34.566 (53,02%)	51.411 (50,15%)	58.132 (74,11%)
Variação de tamanho (pb)	200 – 27.826	200 – 24.948	200 – 10.287
Tamanho médio (pb)	976,55	1.062,37	671,4
Total de sequências < 1000pb	46.947 (72,01%)	69.320 (67,62%)	63,025 (80,35%)
Total de sequências ≥ 1000pb	18.245 (27,99%)	33.189 (32,38%)	15,414 (19,65%)

* Conjunto de dados nos quais foi mantida apenas a maior isoforma por gene.

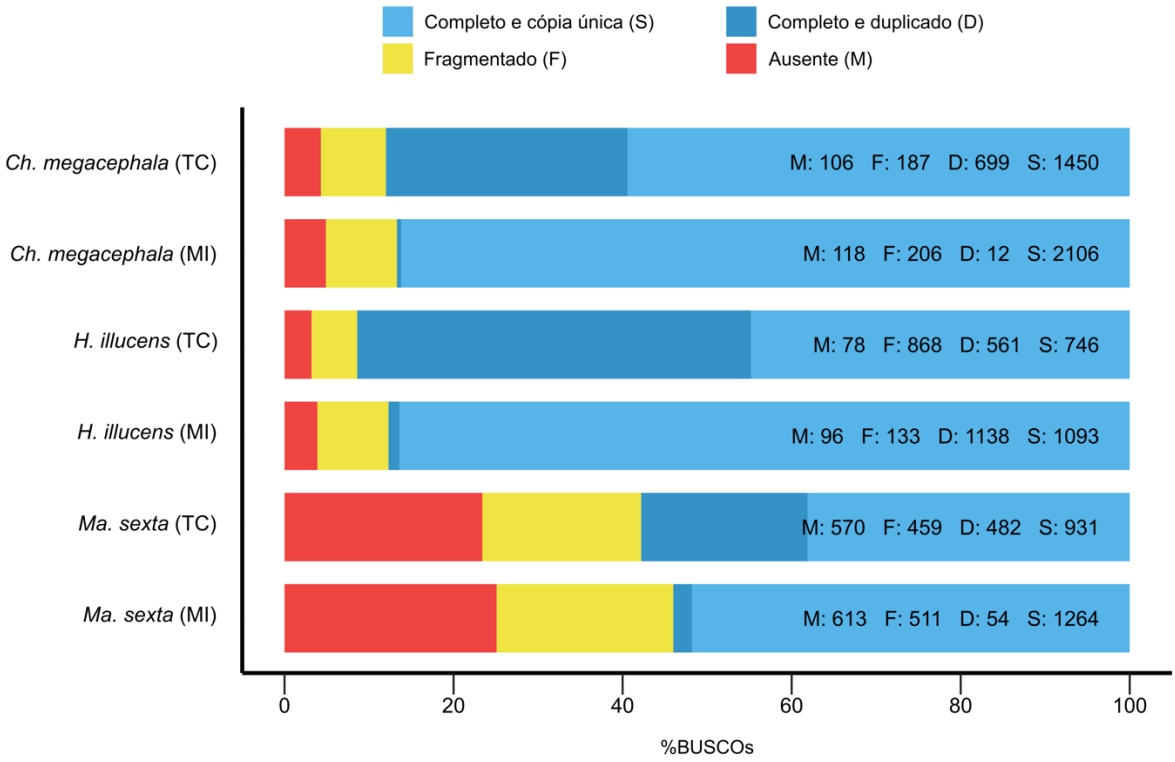


Figura 2.3: Análise de completude: As análises utilizando o BUSCO foram realizadas usando os transcriptomas completos (TC) ou apenas contendo a maior isoforma (MI) de cada transcrito. Os valores nas barras de cada análise correspondem ao total de transcritos que alinharam na categoria. O banco de sequências ortólogas de Endopterygota do BUSCO tinha no total 2.442 ortólogos.

2.6. Análises de expressão diferencial global

Com os transcriptomas de cada espécie, foi possível alinhar as *reads* de cada réplica e condição para obtermos os valores de expressão de cada transcrito ou *contig* e as matrizes de expressão. Com a matriz de cada espécie foi possível comparar a expressão de cada transcrito ou *contig* entre as condições controle (PBS) e dano tecidual (EMS). No caso de *Ma. sexta* foi realizada uma segunda comparação entre o disco imaginal controle (*Sham*) e danificado (*Pinch*).

Considerando os dois testes de expressão diferencial realizados, *Ch. megacephala* e *Dr. melanogaster* apresentaram as maiores quantidades de TCDE (Tabela 2.3). *H. illucens* foi a única espécie a apresentar uma quantidade similar de *contigs* super e sub-expressos. De forma geral, apenas *Ch. megacephala* apresentou um padrão diferente das outras espécies, tendo mais transcritos sub-expressos nas larvas que sofreram dano do que super-expressos (Tabela 2.3). Ao avaliarmos os valores relativos de diferença de expressão (chamados de *Fold Change* – FC) médios entre as duas análises de cada TCDE, é possível notar que *Ch. megacephala* e *H. illucens* apresentaram a maior variação no FC (Figura 2.4, Anexos T4-8). Por sua vez, as duas comparações de *Ma. sexta* apresentam as menores variações na diferença de expressão.

Tabela 2.3: Total de TCDE de cada espécie e experimento.

Análise	*Dr. melanogaster*	*Ch. megacephala*	*H. illucens*	*Ma. sexta*	
				PBS – EMS	*Sham – Pinch*
RSEM – edgeR	494	462	169	322	351
RSEM – DESeq2	446	666	491	344	280
TCDE compartilhados	407	384	104	245	250
Super-expressos*	261	114	57	198	232
Sub-expressos*	146	270	47	47	18

* Nas larvas/discos imaginais danificados.

Com relação às duas comparações de *Ma. sexta*, quatro *contigs* são compartilhados entre elas. O mais interessante é que apenas um deles (DN15923_c0_g1_i1) apresenta o mesmo padrão nas duas comparações, sendo menos expresso nas larvas/discos

imaginais que sofreram dano tecidual. Já os outros três (DN16827_c4_g4_i2, DN16886_c0_g2_i1 e DN17163_c3_g1_i9) são menos expressos nas larvas com dano tecidual (PBS – EMS), porém mais expressos nos discos imaginais danificados (*Sham – Pinch*).

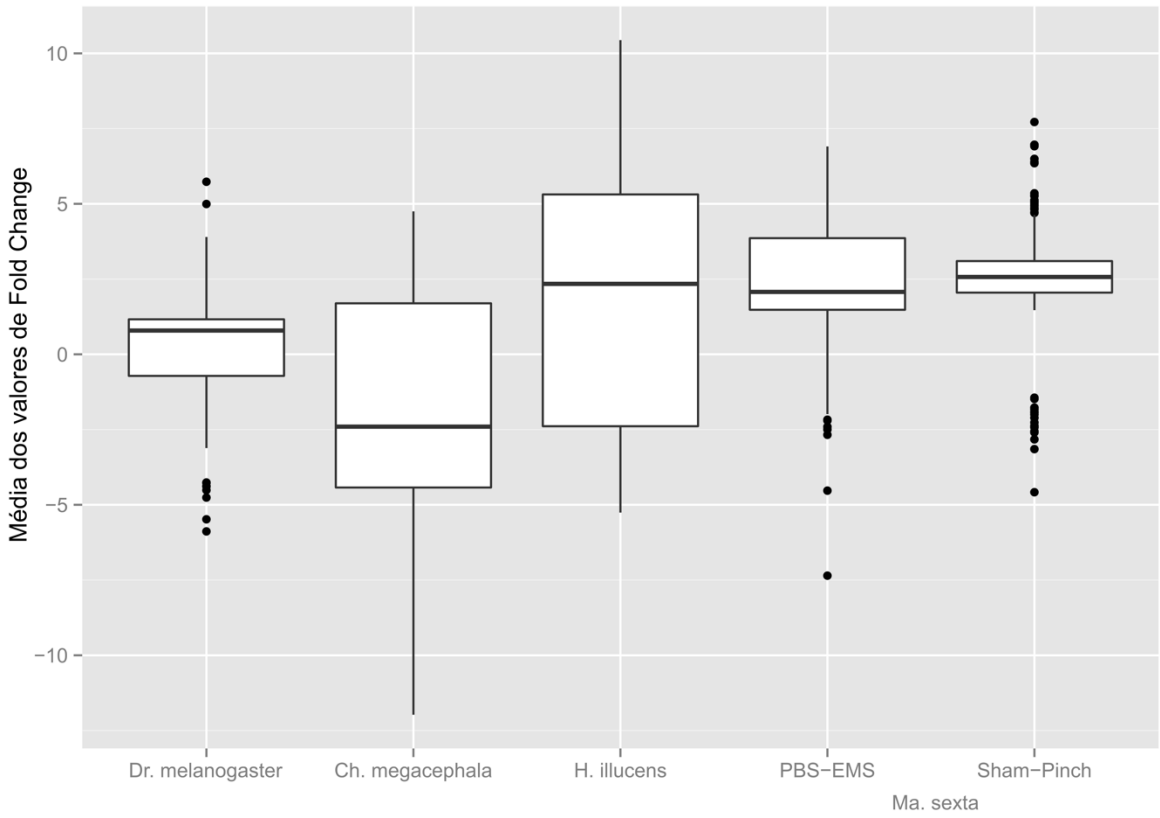


Figura 2.4: Análise exploratória das diferenças de expressão: Valores de log FC positivos indicam maior expressão de um transcrito nas amostras com dano tecidual e valores negativos indicam o oposto.

Entretando, ao olharmos as sequências compartilhadas (ortólogos) entre os TCDE das espécies, encontramos 3 ortólogos entre *Ch. megacephala* e *Dr. melanogaster* (*contig* DN16671_c5_g1_i1 e transcrito *trol*; *contig* DN11512_c0_g1_i1 e transcrito CG7720; *contig* DN9220_c0_g2_i1 e transcrito *MtnD*) e dois ortólogos entre *Ma. sexta* (corpo inteiro) e *Dr. melanogaster* (*contig* DN16786_c5_g2_i4 e transcrito *Fib*; *contig* DN16128_c0_g3_i1 e transcrito mt:ND2). É interessante observar que quase todos os grupos de ortólogos apresentam o mesmo padrão de expressão entre as duas espécies, apenas o grupo do transcrito Fib é diferente: é menos expresso nas larvas com dano de *Dr. melanogaster* e mais expresso nas larvas de *Ma. sexta* com dano tecidual (Figura 2.5).

Em todo caso, esses 10 transcritos/*contig*s possuem um ótimo potencial de estar envolvido no atraso.

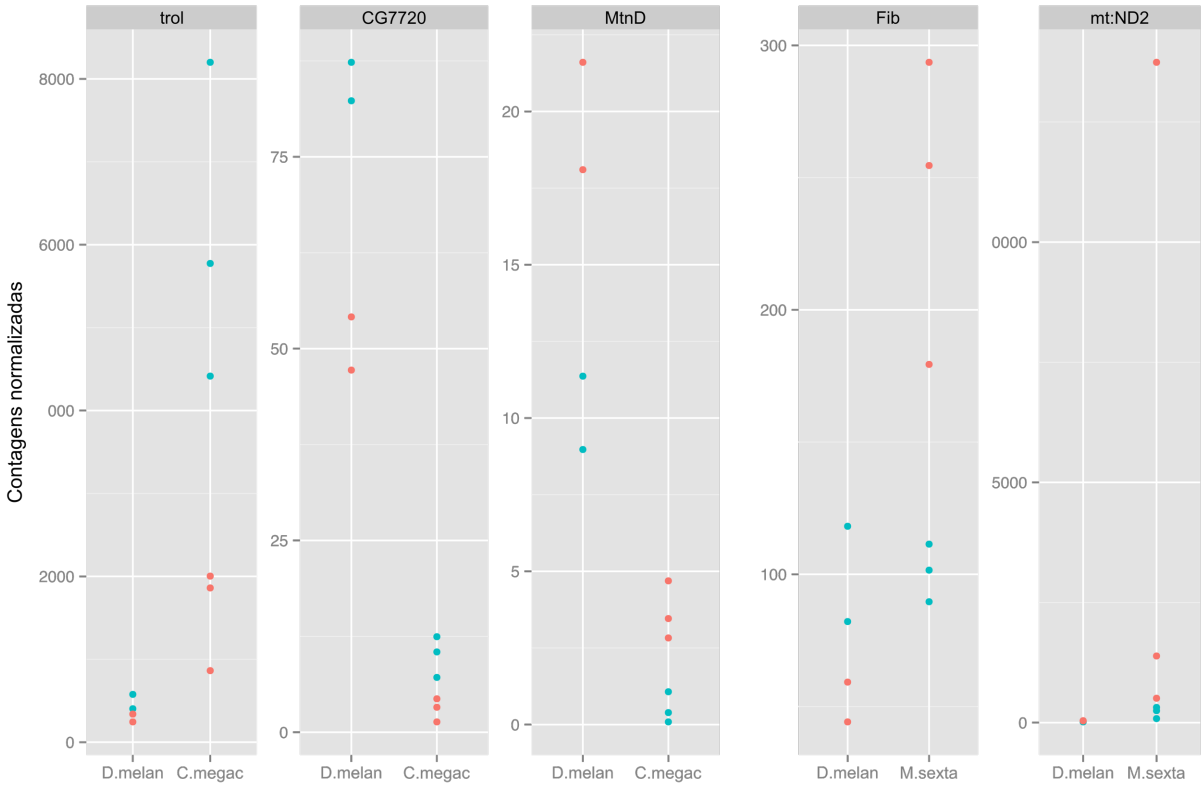


Figura 2.5: Valores de expressão de cada grupo de ortólogos: Para facilitar a compreensão, os grupos de ortólogos estão representados pelo nome do transcrito de *Dr. melanogaster*. Em vermelho estão representados as amostras tratadas com EMS, já as tratadas com PBS estão representadas em verde; "D.melan" e "C.megac" representam *Dr. melanogaster* e *Ch. megacephala*, respectivamente.

2.7. Análises de expressão diferencial do *ilp8*

Utilizando as sequências de *Ch. megacephala* e *H. illucens* descritas no capítulo anterior, foi possível identificar o gene *ilp8* em seus respectivos transcriptomas larvais (Anexo F5). Ainda que a sequência recuperada de *Ch. megacephala* seja exatamente igual à descrita anteriormente, ela não está completa, faltando uma parte final da sequência da proteína. Já as sequências recuperadas de *H. illucens* (isoforma A e B) apresentaram uma e duas alterações de aminoácidos, respectivamente, ainda que ambas alinhassem por completo às sequências previamente descritas.

Olhando especificamente para a expressão do *ilp8*, (Tabela 2.4), em *Dr. melanogaster* foi observado um aumento na expressão do gene após o dano tecidual, ainda que a diferença

não tenha sido estatisticamente significativa devido ao baixo poder estatístico decorrente da baixa amostragem do transcrito (em número de *reads*). Quando avaliamos o homólogo do *ilp8* de *Ch. megacephala*, o número de *reads* alinhados ao gene foi tão baixo que não foi possível utilizar o dado para as análises de expressão. Por fim, as duas isoformas de *H. illucens* homólogas ao *ilp8* apresentaram uma variação não consistente com o tratamento nas réplicas e, por essa razão, não é possível concluir que a variação na expressão esteja associada ao dano tecidual. É importante ressaltar que as duas isoformas de *H. illucens* foram mantidas após a seleção da maior isoforma por gene. Isso aconteceu porque essa seleção é feita com base nas predições do programa Trinity, que pode ou não representar isoformas reais. Neste caso, o Trinity classificou as isoformas do *ilp8* de *H. illucens* como sendo dois genes diferentes. Para testarmos a expressão individual de cada *contig* (sem competição por *reads* alinhadas), cada um deles foi retirado manualmente do transcriptoma antes das análises de expressão. Entretanto, em nenhum dos casos houve uma alteração nas contagens que levassem a uma diferença de expressão.

Tabela 2.4: Análise da expressão dos transcritos do gene *ilp8* de cada espécie.

Espécie Transcrito/*Contig*	Contagens brutas/normalizadas*						edgeR		DESeq2	
	C1	C2	C3	E1	E2	E3	logFC	FDR	logFC	padj
Dr. melanogaster										
ilp8	6 5,4	6 6,3	---	32 29,5	18 20.6	---	---	---	2,11	0,24
Ch. megacephala										
DN25329_c0_g1_i1	1 ---	1 ---	1 ---	0	0	4 ---	---	---	---	---
H. illucens										
DN10846_c0_g2_i3	133 5.4	57 3.1	24 2.2	467 17.7	34 2.1	70 5.4	1,23	0,75	1,52	0,99
DN10846_c0_g1_i1	74 2.8	19 1.1	11 1.5	896 23.9	26 1.4	14 1.7	2,32	0,23	2,55	0,99

* "C" representa as amostras controle; "E" representa as amostras com dano tecidual.

2.8. Anotação e enriquecimento

A anotação consiste em atribuir uma identidade ou função aos *contigs* montados das espécies não-modelo. Com os transcriptomas montados, é possível fazer predições de CDS e utilizá-las para alinhar contra bancos de dados públicos contendo sequências já descritas.

Foi possível identificar entre 50% e 60% de *contigs* com CDS em cada transcriptoma (Tabela 2.5). Entretanto, poucas dessas sequências tiveram um peptídeo sinal identificado, provavelmente devido à presença de CDS incompletas. Além disso, entre 10% e 15% das CDS preditas de cada transcriptoma possui pelo menos um domínio de transmembrana, indicando a presença de possíveis proteínas de membrana, que atuariam como receptores ou formadoras de canais de transporte.

Tanto os *contigs* quanto as CDS preditas foram utilizadas nos alinhamentos contra bancos de dados conhecidos. Pela anotação usando tanto o programa GOanna (com seu próprio banco de dados) quanto o GHOSTZ (usando o banco de dados do NCBI) foi possível anotar em torno de 50% do transcriptoma de *Ch. megacephala* e *Ma. sexta* (Tabela 2.5). Ainda que a taxa de anotação de *H. illucens* tenha sido menor que a das outras espécies (35%), em números absolutos foi possível anotar aproximadamente a mesma quantidade de sequências (em torno de 30 mil *contigs*). A taxa de anotação só é menor em *H. illucens* devido a maior quantidade de *contigs* montado pelo Trinity (Anexos T9-10).

Tabela 2.5: Predições e anotação dos transcriptomas montados.

Predições	*Ch. megacephala*	*H. illucens*	*Ma. sexta*
Total de *contigs*	65.192	102.509	78.439
CDS	37.726 (57,87%)	60.619 (59,14%)	42.328 (53,96%)
Pepídeo sinal	1	19	3
Domínios de transmembrana	6.035 (16,00%)	9.566 (15,78%)	5.440 (12,85%)
Anotação	33.415 (51,26%)	36.764 (35,86%)	37.519 (47,83%)
TCDE	237 (61,72%)	26 (25,00%)	146 (59,59%)* 107 (42,80%)**

* TCDE considerando a comparação PBS – EMS.

** TCDE considerando a comparação Sham – Pinch.

Devido a essa alta taxa de anotação das três espécies não-modelo, uma quantidade razoável dos TCDE foi anotada. Apesar disso, foi possível utilizar somente a anotação realizada pelo GOanna para avaliarmos quais processos estavam enriquecidos em cada conjunto de *contigs*, pois somente ela recupera os termos GO necessários para a análise de enriquecimento.

A espécie *Ch. megacephala* foi a que apresentou a maior quantidade de TCDE anotados. É possível observar um aumento da resposta a estresse nas larvas que sofreram dano tecidual, principalmente associado a metabolismo de substâncias tóxicas e dobramento de proteínas (Anexo F6-A). Por outro lado, essas larvas apresentaram diminuição de processos associados na absorção de íons e desenvolvimento de cutícula (Anexo F6-B).

Não foi possível realizar essa mesma análise de enriquecimento para a espécie *H. illucens*, uma vez que apenas 26 TCDE foram anotados, não sendo o suficiente para determinarmos processos enriquecidos. Ainda assim, dois *contigs* super-expressos nas larvas com dano foram anotados como pertencentes à família das proteínas de choque-térmico (DN18626_c0_g1_i5 e DN16404_c0_g2_i1), relacionados com o estresse que as larvas sofreram.

Já para a *Ma. sexta*, considerando a resposta do corpo inteiro (PBS – EMS), foram observados um aumento nos processos biológicos associados a síntese de proteínas citoplasmáticas (Anexo F7-A) e diminuição dos processos metabólicos associados a oxirredução nas larvas que sofreram dano tecidual (Anexo F7-B). Já considerando apenas a resposta dos discos imaginais (*Sham – Pinch*), só foi possível avaliar os processos menos enriquecidos nas larvas com dano, que estão associados com a organização do sarcômero (Anexo F8).

Por fim, como *Dr. melanogaster* é uma espécie modelo, basta ter a identidade dos TCDE para avaliamos os processos enriquecidos, não sendo necessário a etapa de anotação. Nas larvas que sofreram dano tecidual houve aumento de processos biológicos envolvidos no metabolismo de proteínas, principalmente glutationa, e resposta à substâncias tóxicas (desintoxicação) (Anexo F9), como observado para *Ch. megacephala*. Já os processos menos expressos nas larvas com dano relacionam-se com resposta imune a bactérias,

resposta a hipóxia, enovelamento de proteínas (associado a cofatores de chaperonas) e processamento de RNA (Anexo F10).

Ainda que observar os processos biológicos de modo geral permite analisar um quadro geral do que está acontecendo em cada espécie em decorrência do tratamento, é importante avaliarmos a identidade dos *contigs* mais diferencialmente expressos, que podem nos fornecer informações mais interessantes sobre os processos envolvidos na resposta ao dano tecidual.

2.8.1. Transcritos com expressão mais diferente entre tratamentos em cada espécie

Dos cinco TCDE mais diferentes entre larvas tratadas com PBS e EMS (em termos de log da razão dos níveis de expressão, *fold change*) em *Dr. melanogaster*, dois deles estavam sendo mais expressos nas larvas com dano tecidual (Tabela 2.6).

O gene *turandot A* pertence a uma família de genes (*Turandot – Tot*) que codificam proteínas que são secretadas para a hemolinfa da larva, e a expressão desses genes está relacionada a altos níveis de estresse sinalizados pela via JAK/STAT (Grönholm et al. 2012). Ao contrário da resposta a choques térmicos, a defesa imunológica promovida pelos genes da família *Tot* é mais lenta, porém continua ativa por mais tempo depois da interrupção do estresse, ainda que ambas as repostas do sistema imune inato parecem ser acionadas pelos mesmos estímulos (Ekengren & Hultmark 2001). Ekengren e Hultmark (2001) também levantam a hipótese de que as proteínas dessa família atuariam na prevenção de danos teciduais, por meios de interação com proteínas desnaturadas e evitando assim agregrados proteicos. Nesse contexto, sua expressão 53 vezes maior nas larvas que sofreram dano tecidual mostra uma clara resposta ao tratamento com EMS. Já o segundo gene mais expresso nas larvas com dano tecidual não foi estudado previamente (ganhando apenas a sigla CG34166). Esse gene possui, portanto, um grande potencial para estudos mais aprofundados associando sua função com a regulação do desenvolvimento.

Já os genes que estão sendo menos expressos nas larvas com dano tecidual refletem bem os processos recuperados nas análises de enriquecimento. O gene *cecropin* pertence a uma família contendo três genes codificantes de proteínas (Samakovlis et al. 1990). Já os genes *diptericin A* e *diptericin B* fazem parte de uma família codificante para proteinas

ricas em glicina (Hedengren et al. 2000). Ambas as famílias fazem parte do sistema imune anti-bacteriano da larva e dependem da sinalização por parte da ecdisona ativada (20E) para serem expressas (Zhang and Palli 2009).

Tabela 2.6: Cinco transcritos com maior diferença de expressão em *Dr. melanogaster*.

Transcrito	Gene	EdgeR		DESeq2	
		logFC	FDR	logFC	padj
CecA1	*Cecropin A1*	-5,86	3,43E-07	-5,91	0,007805
TotA	*Turandot A*	5,74	9,87E-12	5,73	0,000836
DptA	*Diptericin A*	-5,46	4,88E-09	-5,51	0,015984
CG34166	---	5,01	2,38E-23	4,98	8,35E-08
DptB	*Diptericin B*	-4,75	6,50E-19	-4,77	4,37E-06

Os cinco *contigs* mais diferencialmente expressos em *Ch. megacephala* foram todos menos expressos nas larvas com dano tecidual (Tabela 2.7). É importante notar que esses *contigs* foram anotados contra proteínas não caracterizadas de *Lu. cuprina*. É importante ressaltar que a anotação recupera apenas o primeiro (ou seja, o melhor) alinhamento do *contig* contra o banco de dados. Ao avaliarmos os outros alinhamentos que cada *contig* obteve no processo de anotação, todas os genes com alta identidade aos *contigs* DN12689_c0_g2_i1, DN11484_c1_g1_i2 e DN13487_c1_g1_i1 estavam descritos como proteínas não caracterizadas de diversas espécies de dípteros.

Por outro lado, os outros dois *contigs* alinharam a proteínas já caracterizadas. O DN13502_c0_g2_i2 alinhou contra o gene *TweedleF* de *Dr. melanogaster*. Ele faz parte de uma família gênica que codifica proteínas cuticulares (Cornman 2009). O mais interessante é que essa família gênica aparenta estar relacionada com a resistência a toxinas presentes na alimentação das larvas, por meio de uma maior taxa na formação cuticular (Trienens et al. 2017). Além disso, o *contig* DN13171_c3_g1_i1 também alinhou a proteínas cuticulares de diversas espécies de *Drosophila*. É interessante notar que esses dois *contigs* estavam sendo menos expressos nas larvas tratadas com EMS e seria esperado o espessamento da cutícula com esse tratamento.

Tabela 2.7: Cinco *contigs* anotados com maior diferença de expressão em *Ch. megacephala*.

Contig	Anotação	EdgeR		DESeq2	
		logFC	FDR	logFC	padj
DN12689_c0_g2_i1	hypothetical protein FF38_04967 [*Lucilia cuprina*]	-10.31	4.50E-05	-10.40	4.71E-04
DN13487_c1_g1_i1	hypothetical protein FF38_10409 [*Lucilia cuprina*]	-9.75	2.16E-05	-9.79	2.50E-05
DN13502_c0_g2_i2	hypothetical protein FF38_04970 [*Lucilia cuprina*]	-9.66	8.71E-05	-9.72	3.11E-04
DN13171_c3_g1_i1	hypothetical protein FF38_00630 [*Lucilia cuprina*]	-8.54	1.71E-04	-8.59	3.90E-03
DN11484_c1_g1_i2	hypothetical protein FF38_01607 [*Lucilia cuprina*]	-8.48	1.07E-04	-8.48	4.79E-04

No caso da espécie *H. illucens*, três dos cinco *contigs* mais expressos foram anotados como sendo uma proteína tubulina de cadeia alfa ou beta (Tabela 2.8) e ambas são os principais constituintes de microtúbulos (Nogales 2000). São proteínas fundamentais do organismo, pois participam da formação do citoesqueleto celular e de cílios e flagelos. Nesse sentido, os microtúbulos são extretamente importantes para a organização e controle celular, pois grande parte dos processos celulares depende da estruturação fornecida pelo citoesqueleto, como por exemplo o transporte intracelular (que depende da ancoragem das vesículas no citoesqueleto) e a divisão celular (pois a separação dos cromossomos é mediada pelos microtubulos) (Akhmanova & Steinmetz 2015). Considerando que o tratamento com EMS promove morte celular, é de se esperar que genes codificantes para tubulinas estejam sendo mais expressos nas larvas com dano tecidual, uma vez que devem recuperar os tecidos danificados por meio de uma maior taxa de divisão celular.

Já os dois outros *contigs* foram anotados como genes cujos produtos são necessários para a espermatogênese. Em *Dr. melanogaster*, as gônadas masculinas já estão formadas no início da fase larval e já possuem o conjunto de células tronco que passarão pelo processo de espermatogênese diferenciando-se em espermatogônias (Whitworth et al. 2012). Não há muitos estudos acerca da proteína S-leucina aminopeptidase 1, apenas

que ela faz parte de uma família de oito proteínas enzimáticas que são expressas exclusivamente durante a espermatogênese nas fases tardias de meiose e pós-meiose (Dorus et al. 2011). O gene *Mst98Ca*, por sua vez, é expresso na fase na fase diploide da espermatogênese, ainda na fase larval, mas a proteína só é sintetizada durante o elongamento das espermátides, já haploides e na fase adulta do organismo (Schäfer et al. 1993; Blümer et al. 2002). Considerando que a espermatogênese na fase larval é constituida principalmente de divisões mitóticas para formar as espermatogônias, é de se esperar que o tratamento com o EMS afete essas células também. Combinado com o fato de que genes que codificam a tubulina também estavam mais expressos nas larvas tratadas, pode-se supor que também houve dano nas gônadas e, portanto, há uma maior taxa de multiplicação celular para recuperação do tecido.

Tabela 2.8: Cinco *contigs* anotados com maior diferença de expressão em *H. illucens*.

Contig	Anotação	EdgeR		DESeq2	
		logFC	FDR	logFC	padj
DN8891_c0_g1_i1	Sperm protein (Mst98Ca) [*Drosophila melanogaster*]	5.69	1.62E-09	7.30	6.39E-05
DN8220_c0_g1_i1	Tubulin alpha chain [*Salmo salar*]	5.44	9.32E-10	6.80	2.62E-07
DN16504_c0_g1_i3	Tubulin beta-4 chain [*Caenorhabditis elegans*]	5.12	2.01E-08	6.24	4.12E-05
DN14671_c0_g3_i1	Tubulin beta-2 chain [*Drosophila melanogaster*]	5.03	6.21E-09	5.99	3.02E-05
DN14882_c0_g2_i1	Sperm-Leucylaminopeptidase 1 [*Drosophila melanogaster*]	4.97	6.20E-08	6.14	3.70E-05

Por fim, similarmente à *Ch. megacephala*, alguns dos *contigs* mais diferencialmente expressos de ambos experimentos de *Ma. sexta* foram anotados como proteínas não caracterizadas (Tabela 2.9). O mais interessante é que três *contigs* do experimento do corpo inteiro foram alinhados a genes bacterianos e um *contig* a um gene de uma espécie de fungo (*Coccidioides posadasii*). Isso pode acontecer porque não foi realizado nenhum tratamento anti-bacteriano ou antifúngico nas larvas antes da extração do RNA e sequenciamento do cDNA. Assim, é provável que sequências de simbiontes tenham sido

processadas e sequenciadas juntamente com as de *Ma. sexta*. Ainda que o *contig* DN15208_c0_g2_i1 tenha sido alinhado a um gene de uma espécie de formiga (formigas birois – *Cerapachys biroi*), todos os alinhamentos dos cinco *contigs* do corpo inteiro foram contra proteínas não caracterizadas.

Quatro dos cinco *contigs* mais diferencialmente expressos do experimento, considerando apenas os discos imaginais de *Ma. sexta*, foram alinhados a genes de espécies de Lepdoptera (*Bo. mori*, *Spodoptera litura* e *Omphisa fuscidentalis*). Um dos *contigs* foi anotado como um precursor de esterase de tegumento (*Bo. mori*), estando mais expresso nos discos imaginais danificados. Em um estudo comparativo de larvas de *Bo. mori* infectadas ou não com o vírus da poliedrose nuclear (BmNPV – um dos principais patógenos dessa espécie), Wang e colaboradores (2016) avaliaram que essa esterase foi uma das proteínas que tiveram sua expressão elevada após a infecção com o vírus. Ainda que os autores não tenham proposto uma relação direta ou função dessa esterase com a contaminação, há indícios que sua função esteja diretamente relacionada com perturbações na homeostase da larva, considerando que essa proteína está sendo mais expressa nos discos imaginais

Outra proteína anotada é a trealase-1 (*Om. fuscidentalis*). Essa enzima catalisa a hidrólise de trealose em glicose, e mudanças na atividade dessa enzima está diretamente relacionada com a homeostase e desenvolvimento larval em situações ideais, uma vez que a forma ativa da ecdisona (20E) aumenta sua atividade (Tatun et al. 2008). Ainda que este resultado possa parecer contraintuitivo pois este gene está sendo mais expresso nos discos que sofreram dano, neste caso a trealase está atuando como um fator de recuperação em decorrência do estresse causado pelo EMS, disponibilizando energia para recuperação celular (revisado em Shukla et al. 2015).

Tabela 2.9: Cinco *contigs* anotados com maior diferença de expressão em cada experimento de *Ma. sexta*.

Contig	Anotação	EdgeR		DESeq2	
		logFC	FDR	logFC	padj
Corpo inteiro:					
DN15208_c0_g2_i1	hypothetical protein X777_02010 [*Cerapachys biroi*]	-8.01	4.66E-04	-6.7	1.77E-03
DN16952_c0_g1_i1	hypothetical protein PROSTU_00109 [*Providencia stuartii*]	7.09	3.51E-02	6.35	3.33E-02
DN16982_c0_g2_i2	senescence-associated protein [*Coccidioides posadasii*]	6.19	2.97E-02	6.19	3.51E-02
DN17617_c5_g2_i2	hypothetical protein [*Vibrio parahaemolyticus*]	5.8	4.06E-03	5.83	1.63E-02
DN28541_c0_g1_i1	hypothetical protein [*Bacillus coagulans*]	5.57	8.12E-03	5.72	8.03E-03
Disco imaginal:					
DN23177_c0_g1_i1	WAP four-disulfide core domain protein 18-like [*Cricetulus griseus*]	8.62	5.50E-05	6.82	4.00E-04
DN9388_c0_g1_i1	uncharacterized protein LOC105841443 [*Bombyx mori*]	7.86	4.66E-04	6.05	4.50E-03
DN17047_c0_g1_i2	integument esterase 2 precursor [*Bombyx mori*]	5.58	5.50E-05	4.94	1.88E-02
DN16923_c0_g4_i1	small heat shock protein 27.2 [*Spodoptera litura*]	5.42	2.34E-04	4.69	3.82E-02
DN46913_c0_g1_i1	trehalase-1 [*Omphisa fuscidentalis*]	4.99	3.69E-03	4.89	1.51E-02

Discussão

Como foi discutido no capítulo anterior, a presença do gene *ilp8* pode ser observada pelo menos desde a diferenciação de Diptera. Apesar disso, sua relação com o atraso na metamorfose só foi avaliada em *Dr. melanogaster*. Desse modo, para avaliarmos se o *ilp8* sempre esteve relacionado com o atraso, ou se ganhou essa função ao longo da sua história evolutiva, foram associados ensaios de dano tecidual com experimentos de RNA-seq para obtermos os padrões de expressão larval quando o dano é estimulado. Com isso, é possível identificar genes que podem estar envolvidos na regulação do atraso na metamorfose bem como variações na expressão do *ilp8*.

Uma das espécies chave escolhidas para esse projeto foi *Dr. melanogaster*. Ainda que ela tenha sido utilizada como modelo para descrever a função do *ilp8* (Colombani et al. 2012; Garelli et al. 2012) o mecanismo não está completamente elucidado, tornando-a essencial para a continuação dos estudos. Com a metodologia aplicada neste projeto, foram identificados transcritos diferencialmente expressos entre as condições dano e controle e avaliados os processos enriquecidos nesse conjunto de transcritos.

De forma geral, mesmo que os processos enriquecidos não estejam diretamente associados com o desenvolvimento, a resposta obtida está diretamente relacionada com o tratamento para estimular dano. O aumento da resposta a toxinas e ao estresse oxidativo nas larvas com dano está claramente associado à presença do componente tóxico EMS na alimentação. Já a diminuição da resposta do sistema imune pode estar associada com a morte celular dos hemócitos em decorrência do tratamento com EMS, uma vez que são células do sistema imune larval presentes na hemolinfa que se multiplicam por mitose (Landis et al. 2012; Worley et al. 2012). Além disso, a supressão do sistema imune, frente um dano tecidual, permite que os sistemas de defesa do organismo lidem com o estresse sem desencadear uma resposta autoimune (Dickinson 1988; Råberg et al. 1998).

Olhando para TCDE que estariam diretamente relacionados com o desenvolvimento destaca-se o *Eip71CD* (*Ecdysone-induced protein 28/29kD*). Curiosamente, ele estava sendo mais expresso nas larvas com dano tecidual (EdgeR: logFC: 2.98; FDR: 9.67E-30; DESeq2: log2FC: 2.94; padj: 7.05E-33). A expressão desse gene é induzida pela ecdisona

por meio da sinalização EcR–USP (Cherbas et al. 1991; Hu et al. 2003) e ele está relacionado com resposta a estresse oxidativo (Roesijadi et al. 2007). Mesmo não diretamente associado com o desenvolvimento, a expressão desse gene indica a produção e secreção de ecdisona nas larvas com dano tecidual.

Similarmente ao que acontece com *Dr. melanogaster*, nenhum dos processos enriquecidos parecem estar envolvidos com o desenvolvimento nos dados de *Ch. megacephala*. Processos mais enriquecidos nas larvas com dano retomam resposta à estresse e toxicidade do EMS observado em *Dr. melanogaster*. Por outro lado, os processos menos enriquecidos envolvidos com assimilação e transporte de íons podem ser relacionados com a perda dos potenciais de membrana, essenciais para diversas funções no organismo, como por exemplo transmissão de sinal pelo sistema nervoso, respiração celular, entre outros (Goller et al. 1990; Steinberg & Grinstein 2007). A diminuição desses processos é decorrente, provavelmente, da morte celular provocada pelo tratamento para estimular os danos teciduais.

Quanto aos dados de *H. illucens*, a anotação dos seus TCDE foi a mais limitada das três espécies não-modelo, o que impediu de avaliarmos processos enriquecidos. Mesmo assim, ao analisarmos os poucos TCDE anotados, destacam-se dois transcritos anotados como *Hsc70* (*heat shock 70*). Esse gene é responsivo a choques térmicos e outros tipos de estresse, como por exemplo toxicidade, e faz parte de uma família de proteínas conservada em animais (Craig et al. 1983; Takayama et al. 1999). Apesar de ser apenas um gene, uma vez que ele está sendo mais expresso nas larvas com dano tecidual, isso pode indicar um padrão de expressão similar ao de *Dr. melanogaster* e *Ch. megacephala*.

Ao analisarmos os dados de expressão da larva inteira de *Ma. sexta* (PBS – EMS), os processos mais enriquecidos como biossíntese de proteínas, estão envolvidos com os danos teciduais, uma vez que é uma resposta associada a proliferação celular em decorrência dos danos causados pelo tratamento (Katsuyama et al. 2015; Sanchez et al. 2015). Já a diminuição de processos associados com oxirredução evita a produção das formas reativas do oxigênio, tóxicas se em grandes quantidades nas células (Weston & Davis 2002; Pastor-Pareja et al. 2008). Por ser um agente tóxico, o EMS pode levar ao aumento dessas formas reativas, e a diminuição desses processos evita que ocorra

maiores danos em uma situação já estressante. Tal resposta é similar a observada nos dados de *Dr. melanogaster*.

Por outro lado, ao analisarmos os dados exclusivos dos discos imaginais de *Ma. sexta* (*Sham – Pinch*), vemos com clareza a limitação da anotação para as análises de enriquecimento. Como ela depende da anotação em comparação com *Dr. melanogaster*, espécie modelo de insetos, a taxa de anotação é muito baixa, uma vez que *Ma. sexta* é muito distante filogeneticamente de *Dr. melanogaster*. Apenas aproximadamente 18% dos TCDE dessa condição foram anotados no banco de dados do GOanna (contrapondo os 42% com a anotação do NCBI) e ainda que tenha sido possível realizar as análises de enriquecimento, a resposta obtida foi diminuição do desenvolvimento muscular. Considerando que esses dados se referem a discos imaginais indiferenciados, a resposta de enriquecimento obtida está enviesada, devido muito provavelmente à dificuldade em anotar os *contigs* de *Ma. sexta*.

Apesar disso, esses *contigs* não anotados possuem um grande potencial de estarem relacionados ao desenvolvimento, uma vez que estão diferencialmente expressos entre as condições. Assim, eles representam uma grande oportunidade para estudos funcionais de nocaute ou silenciamento para avaliarmos se estariam envolvidos no processo do mesmo modo que o *ilp8* regula o atraso em *Dr. melanogaster*.

Olhando especificamente para o gene *ilp8*, as contagens de *Dr. melanogaster* mostram um aumento de expressão consistente em todas réplicas com dano tecidual, em comparação com as larvas controle, fato não observado nas amostras de *Ch. megacephala* e *H. illucens*. Apesar disso, ele não foi considerado diferencialmente expresso nas análises estatísticas. Como a variação é pequena, durante as análises de expressão esse gene pode ter sido descartado em alguma das etapas, quando comparado com outros genes que tiveram uma variação maior. Ainda assim, é inegável a participação do *ilp8* na regulação do atraso em *Dr. melanogaster*.

É importante destacar que estamos realizando estudos funcionais com o *Ilp8* de outras espécies além de *Dr. melanogaster* (Bolsa Estágio de Pesquisa no Exterior – BEPE – FAPESP 2016/10342-4) e um resultado preliminar interessante está relacionado com uma possível região regulatória desse gene. O intuito do estudo é inserir o *ilp8* de

diferentes espécies no genoma de *Dr. melanogaster* e avaliar o fenótipo resultante. Um dos controles do experimento é a inserção do *ilp8* da própria *Dr. melanogaster*, porém sem um íntron. O fenótipo resultante dessa transformação foi uma má formação das pupas. Esse resultado preliminar mostra que, sob controle de um elemento regulatório intrônico, o gene *ilp8* também atua na formação da pupa, e não apenas na regulação do desenvolvimento. Considerando que em nossos experimentos o gene *ilp8* não era mais expresso nas larvas com dano tecidual, é possível que a regulação da metamorfose seja uma função secundária desse gene.

Com isso, nós rejeitamos nossa hipótese inicial de que o *ilp8* esteve associado com a regulação do desenvolvimento desde sua origem. Nós propomos a hipótese de que este gene tem uma função primordial desde sua origem (podendo ser na formação da pupa) e que apenas em *Drosophila* ele adquiriu o papel na regulação do atraso na metamorfose (Figura 2.6).

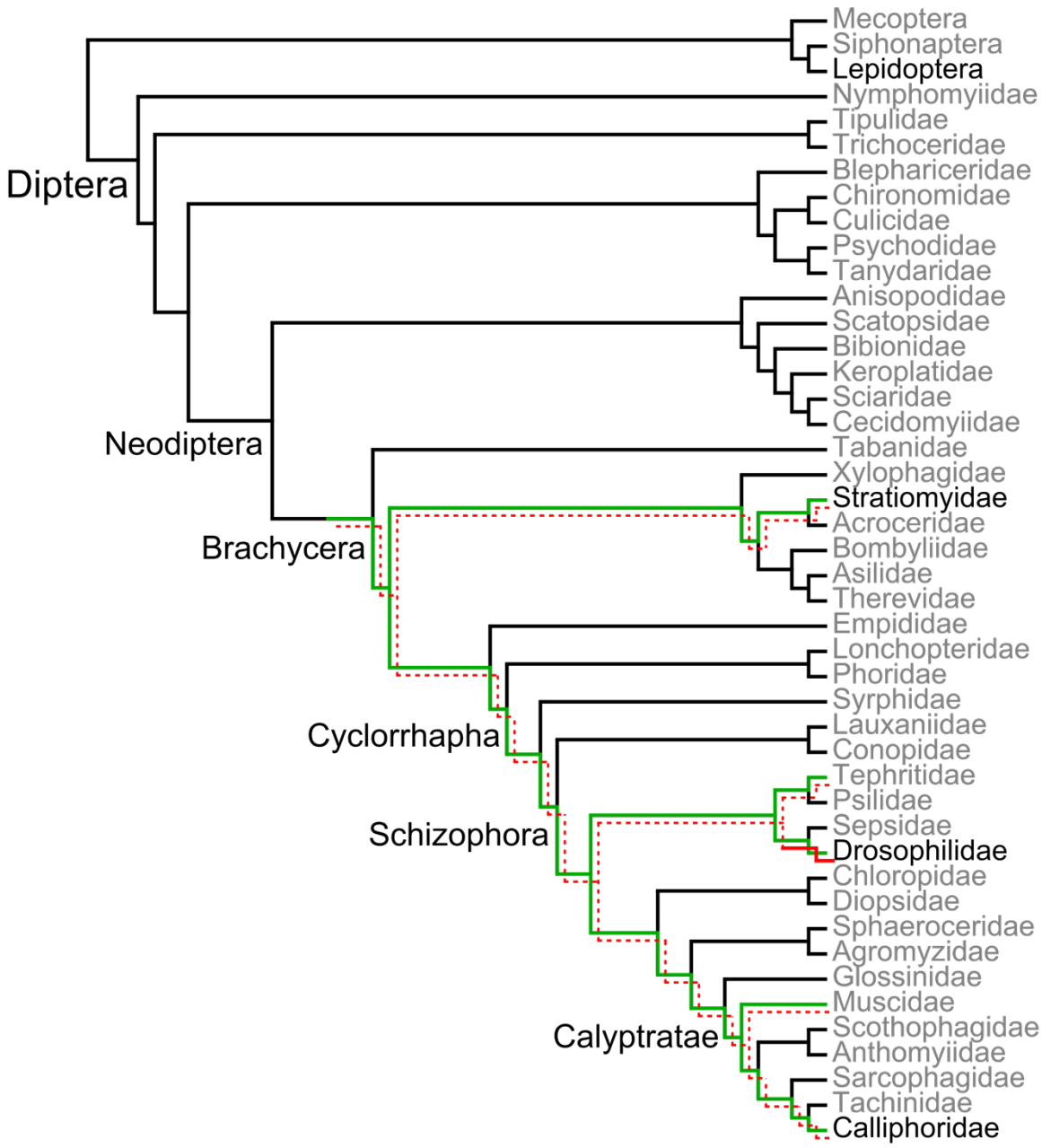


Figura 2.6: Hipótese evolutiva da função do gene *ilp8*: Espécies marcadas em preto representam os grupos de holometábolos estudados neste capítulo; em verde está destacado nossa hipótese evolutiva do gene *ilp8* discutida no capítulo anterior; em vermelho está destacada a hipótese evolutiva da função desse gene (tracejado: função primordial; cheio: função secundária de regulação do desenvolvimento). Filogenia adaptada de Wiegmann e colaboradores (2011) – material suplementar.

Conclusão do capítulo

Neste capítulo, nós buscamos avaliar a função do gene *ilp8* como regulador do atraso da metamorfose. Como vimos no capítulo anterior, a presença do *ilp8* pode ser traçada desde a diferenciação de Brachycera. Assim, como hipótese inicial, podemos dizer que o *ilp8* sempre esteve relacionado com o atraso na metamorfose desde sua origem. Por outro lado, é possível que ele tenha desenvolvido essa função ao longo de sua história evolutiva, sendo talvez uma inovação de *Drosophila*, e essa regulação é mediada por outras proteínas nos outros insetos.

Para testar isso, foram avaliadas as expressões globais de larvas de *Dr. melanogaster*, *Ch. megacephala*, *H. illucens* e *Ma. sexta* que sofreram danos teciduais. Com os dados obtidos foi possível rejeitar a hipótese inicial, uma vez que apenas *Dr. melanogaster* apresentou aumento na expressão do *ilp8*. Assim, nós propomos que este gene só passou a regular o desenvolvimento em *Drosophila*, possuindo outra função nas espécies que apresentam seu homólogo. Entretanto, tanto *Ch. megacephala* quanto *H. illucens* e *Ma. sexta* apresentaram TCDE não anotados, que podem codificar possíveis proteínas reguladoras do atraso e, portanto, ótimos candidatos para estudos funcionais.

Conclusões Gerais

Evolução das vias de controle do desenvolvimento tecidual em holometábolos

Este trabalho buscou compreender melhor como ocorre a regulação do desenvolvimento tecidual em larvas de insetos holometábolos. Com base em estudos que apontam a proteína Ilp8 como sendo fundamental na regulação do desenvolvimento da espécie *Dr. melanogaster*, foram utilizadas duas abordagens para entendermos melhor essa regulação: 1) foram realizadas buscas pelo gene *ilp8* em dados públicos de diversas espécies de insetos; 2) a partir dos dados da primeira fase, foram escolhidas espécies com e sem homólogos ao *ilp8* para ensaios de dano tecidual associados ao sequenciamento massivo em paralelo do RNA (RNA-seq).

As buscas pelo *ilp8* retornaram diversas sequências pertencentes ou similares a proteínas da família da insulina em praticamente todos os grandes grupos que fazem parte de Insecta. Entretanto, ao compararmos essas sequências com o *ilp8* de *Dr. melanogaster*, apenas grupos chave dentro da ordem Diptera apresentaram similaridade com o gene de *Dr. melanogaster*. Com isso foi levantada a hipótese de que este gene seja uma inovação de Brachycera (Diptera).

A partir dessa hipótese, foram realizados os ensaios de dano associados com RNA-seq em quatro espécies chave. Ainda que todas apresentassem atraso na metamorfose, apenas *Dr. melanogaster* apresentou um aumento de expressão consistente do *ilp8* nas larvas com dano tecidual. Assim, nós levantamos a hipótese de que o *ilp8* adquiriu a função de regulação apenas em *Drosophila* e que, nas outras espécies, o *ilp8* possui outra função. Com isso, nós propomos o *ilp8* como uma inovação de Brachycera, com uma função primordial, e só adquiriu a função de regulação do desenvolvimento em *Drosophila*.

Resumo

O atraso na metamorfose em decorrência de danos nos discos imaginais é um mecanismo de controle do desenvolvimento observado em diversas espécies de insetos. Em *Drosophila melanogaster*, esse mecanismo é regulado pela proteína Ilp8, porém não há estudos que buscam compreender como o gene *ilp8* evoluiu bem como esse controle é regulado em outros insetos. Este trabalho foi um dos primeiros passos para compreendermos como ocorre a regulação do desenvolvimento. O primeiro passo foi determinar quais espécies de insetos possuem homólogos ao *ilp8* de *Dr. melanogaster*. Por meio de buscas em bancos de dados disponíveis publicamente foi possível encontrar esse gene nos principais grupos de dípteros, sugerindo que sua origem tenha ocorrido durante a diferenciação de Brachycera. Nesse contexto, para avaliarmos se o *ilp8* já estava associado com a regulação do desenvolvimento, foram realizados ensaios de dano tecidual associados com RNA-seq em espécies com (*Dr. melanogaster*, *Chrysomya megacephala* e *Hermetia illucens*) ou sem (*Manduca sexta*) homólogos ao *ilp8*. Com essa abordagem, foi possível estudar a variação na expressão global em resposta ao dano causado. Ainda que as quatro espécies estudadas apresentavam atraso na metamorfose em decorrência do dano nos discos imaginais, apenas *Dr. melanogaster* apresentou aumento na expressão do *ilp8* nas larvas com dano. Esse resultado parece indicar que, apesar de ser relativamente antigo, a associação da proteína Ilp8 com a sinalização do atraso só aconteceu após a diferenciação de *Drosophila*.

Abstract

The metamorphosis delay due to tissue damage in imaginal discs is a control mechanism of larval development observed in several species of insects. This mechanism is regulated by the Ilp8 protein in *Drosophila melanogaster*, but there are no studies about how the *ilp8* gene evolved, and how this mechanism is regulated in other insects. The objective of this project is to understand how the metamorphosis delay evolved as a control mechanism of tissue development. The first step was to determine which insect species have homologs of the *ilp8* gene of *Dr. melanogaster*. By searching publicly available databases it was possible to find this gene in the main dipteran groups, suggesting that it is an inovation of the Brachycera group. To understand the evolution of the relationship between *ilp8* and the tissue-damage-dependent delay in metamorphosis, we performed tissue damage assays associated with RNA-seq in species with (*Dr. melanogaster*, *Chrysomya megacephala* and *Hermetia illucens*) or without (*Manduca sexta*) *ilp8* homologues. With this approach it was possible to study the global expression variation in response to the tissue damage. Although the four species studied delayed metamorphosis upon treatment with the DNA-damaging agent EMS, only *Dr. melanogaster* showed a detectable increase of the *ilp8* expression in the damaged larvae. These results indicate that the metamorphosis-delay function of the *ilp8* gene evolved relatively late during its evolution, when *Drosophila* differentiated from other Brachycera.

Referências Bibliográficas

Akhmanova A, Steinmetz MO. 2015. Control of microtubule organization and dynamics: two ends in the limelight. Nat. Rev. Mol. Cell Biol. 16:711–726.

Alberts B, Bray D, Johnson A, Lewis J, Raff M, Roberts K, Walter P. 1999. Fundamentos da biologia celular: uma introdução à biologia molecular. Capítulo 8, edição universitária, editora Artes Médicas Sul LTDA.

Alonso J, Santare JF. 2006. Characterization of The Drosophila melanogaster Ribosomal Proteome. J. Proteome Res. 5:2025–2032.

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Disponível online em: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Beira J V., Paro R. 2016. The legacy of Drosophila imaginal discs. Chromosoma 125:573–592.

Blümer N, Schreiter K, Hempel L, Santel A, Hollmann M, Schüfer MA, Renkawitz-Pohl R. 2002. A new translational repression element and unusual transcriptional control regulate expression of don juan during Drosophila spermatogenesis. Mech. Dev. 110:97–112.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Bork P, Koonin E V. 1996. Protein sequence motifs. Curr. Opin. Struct. Biol. 6:366–376.

Bosch M, Serras F, Martı E. 2005. JNK signaling pathway required for wound healing in regenerating Drosophila wing imaginal discs. 280:73–86.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST + : architecture and applications. BMC Bioinformatics 9:1–9.

Cerretti P, Stireman JO, Pape T, Hara JEO, Marinho MAT, Rognes K, Grimaldi DA. 2017. First fossil of an oestroid fly (Diptera : Calyptratae : Oestroidea) and the dating of oestroid divergences. PLoS One 12:e0182101.

Cherbas L, Lee K, Cherbas P. 1991. Identification of ecdysone response elements by analysis of the Drosophila Eip28/29 gene. Genes Dev. 5:120–131.

Chyb S, Gompel N. 2013. Atlas of Drosophila morphology: wild-type and classical mutants. Primeira edição, editora Academic Press.

Colombani J, Andersen DS, Boulan L, Boone E, Romero N, Virolle V, Texada M, L??opold P. 2015. Drosophila Lgr3 Couples Organ Growth with Maturation and Ensures Developmental Stability. Curr. Biol. 25:2723–2729.

Colombani J, Andersen DS, Leopold P. 2012. Secreted Peptide Dilp8 Coordinates Drosophila Tissue Growth with Developmental Timing. Science (80-.). 336:582–585.

Cornman RS. 2009. Molecular Evolution of Drosophila Cuticular Protein Genes. PLoS One 4:e8345.

Craig EA, Ingolia TD, Manseau LJ. 1983. Expression of Drosophila heat-shock cognate genes during heat shock and development. Dev. Biol. 99:418–426.

Curtiss J, Heilig JS. 1995. Establishment of Drosophila imaginal precursor cells is controlled by the Arrowhead gene. Development 121:3819–3828.

Dickinson WJ. 1988. On the Architecture of Regulatory Systems : Evolutionary Insights and Implications. BioEssays 8:204–208.

Dorus S, Wilkin EC, Karr TL. 2011. Expansion and functional diversification of a leucyl aminopeptidase family that encodes the major protein constituents of Drosophila sperm. BMC Genomics 12.

Eddy S. 1998. Profile hidden Markov models. Bioinformatics 14:755–763.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10:1–7.

Edgar BA. 2006. How flies get their size: Genetics meets physiology. Nat. Rev. Genet. 7:907–916.

Ekengren S, Hultmark D. 2001. A Family of Turandot -Related Genes in the Humoral Stress Response of Drosophila. Biochem. Biophys. Res. Commun. 284:998–1003.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. 2016. The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Res. 44:D279–D285.

Garelli A, Gontijo AM, Miguela V, Caparros E, Dominguez M. 2012. Imaginal Discs Secrete Insulin-Like Peptide 8 to Mediate Plasticity of Growth and Maturation. Science (80-.). 336:579–582.

Garelli A, Heredia F, Casimiro AP, Macedo A, Nunes C, Garcez M, Dias ARM, Volonte YA, Uhlmann T, Caparros E, et al. 2015. Dilp8 requires the neuronal relaxin receptor Lgr3 to couple growth to developmental timing. Nat. Commun. 6:8732.

Gilad Y, Mizrahi-Man O. 2015. A reanalysis of mouse ENCODE comparative gene expression data. F1000Research 121:1–37.

Gilbert SF. 1949. Developmental Biology. Capítulo 18, 8ª edição, editora Sinauer Associates Inc.

Goller F, Esch H, Goller F. & Esch H, Goller F, Esch H. 1990. Comparative study of chill-coma temperatures and muscle potentials in insect flight muscles. J. Exp. Biol. 150:221–231.

Goltzené F, Holder F, Charlet M, Meister M, Oka T. 1992. Immunocytochemical localization of Bombyx-PTTH-like molecules in neurosecretory cells of the brain of the migratory locust, Locusta migratoria. Cell Tissue Res. 269:133–140.

Gontijo AM, Garelli A. 2018. Mechanisms of Development The biology and evolution of the Dilp8-Lgr3 pathway : A relaxin-like pathway coupling tissue growth and developmental timing control. Mech. Dev.:1–7.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–652.

Griffiths AJF, Wessler SR, Carrol SB, Doebley J. 2012. Introduction to genetic analysis. Capítulo 16, décima edição, editora W. H. Freeman and Company.

Grönholm J, Kaustio M, Myllymäki H, Kallio J, Saarikettu J, Kronhamn J, Valanne S, Silvennoinen O, Ramet M. 2012. Not4 enhances JAK / STAT pathway-dependent gene expression in Drosophila and in human cells. FASEB J. 26:1239–1250.

Grönke S, Clarke D-F, Broughton S, Andrews TD, Partridge L. 2010. Molecular Evolution and Functional Characterization of Drosophila Insulin-Like Peptides. PLoS Genet. 6:e1000857.

Gu S, Chen C, Hsieh Y, Lin P, Young S. 2015. Modulatory effects of bombyxin on ecdysteroidogenesis in Bombyx mori prothoracic glands. J. Insect Physiol. 72:61–69.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494–1512.

Hackney JF, Zolali-Meybodi O, Cherbas P. 2012. Tissue Damage Disrupts Developmental Progression and Ecdysteroid Biosynthesis in Drosophila. PLoS One 7:1–12.

Hall DC, Gerhardt RR. Flies (Diptera). *In* Mullen G, Durden L. 2002. Medical and Veterinary Entomology. Capítulo 8, p. 127-145, editora Academic Press.

Halme A, Cheng M, Hariharan IK. 2010. Retinoids Regulate a Developmental Checkpoint for Tissue Regeneration in Drosophila. Curr. Biol. 20:458–463.

Hariharan IK. 2012. How growth abnormalities delay “puberty” in Drosophila. Sci. Signal. 5:pe27.

Heather JM, Chain B. 2016. Genomics The sequence of sequencers : The history of sequencing DNA. Genomics 107:1–8.

Hedengren M, Borge K, Hultmark D. 2000. Expression and Evolution of the Drosophila Attacin / Diptericin Gene Family. Biochem. Biophys. Res. Commun. 279:574–581.

Hill RJ, Billas IML, Bonneton F, Graham LD, Lawrence MC. 2013. Ecdysone Receptors: From the Ashburner Model to Structural Biology. Annu. Rev. Entomol. 58:251–271.

Hu X, Cherbas L, Cherbas P. 2003. Transcription Activation by the Ecdysone Receptor (EcR / USP): Identification of Activation Functions. Mol. Endocrinol. 17:716–731.

Katsuyama T, Comoglio F, Seimiya M, Cabuy E, Paro R. 2015. During *Drosophila* disc regeneration, JAK/STAT coordinates cell proliferation with Dilp8-mediated developmental delay. Proc. Natl. Acad. Sci. 112:E2327–E2336.

Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. 305:567–580.

Kromer-metzger E, Lagueux M. 1994. Expression of the gene encoding an insulin-related peptide Evidence for alternative promoter usage. Eur. J. Biochem. 434:427–434.

Lagueux M, Lwoff L, Meister M, Goltzenfi F, Hoffmann JA. 1990. cDNAs from neurosecretory cells of brains of Locusta migratoria (Insecta , Orthoptera) encoding a novel member of the superfamily of insulins. Eur. J. Biochem. 254:249–254.

Landis G, Shen J, Tower J. 2012. Gene expression changes in response to aging compared to heat stress, oxidative stress and ionizing radiation in <i>Drosophila melanogaster</i> Aging (Albany. NY). 4:768–789.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359.

Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. 2012. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28:882–883.

Li B, Dewey CN. 2011. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:1–21.

Mattila J, Omelyanchuk L, Kyttälä S, Turunen H, Nokkala S. 2005. Role of Jun N-terminal Kinase (JNK) signaling in the wound healing and regeneration of a Drosophila melanogaster wing imaginal disc. 399:391–399.

McBrayer Z, Ono H, Shimell MJ, Parvy JP, Beckstead RB, Warren JT, Thummel CS, Dauphin-Villemant C, Gilbert LI, O'Connor MB. 2007. Prothoracicotropic Hormone Regulates Developmental Timing and Body Size in Drosophila. Dev. Cell 13:857–871.

McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB, Barrell DG, Hill DP, Dolan ME, Williams WP, et al. 2006. AgBase: A functional genomics resource for agriculture. BMC Genomics 7:1–13.

Menut L, Vaccari T, Dionne H, Hill J, Wu G, Bilder D. 2007. A mosaic genetic screen for Drosophila neoplastic tumor suppressor genes based on defective pupation. Genetics 177:1667–1677.

Misof B, Liu S, Meusemann K, Peters RS, Flouri T, Beutel RG, Niehuis O, Petersen M. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science (80-.). 346.

Mizoguchi A, Okamoto N. 2013. Insulin-like and IGF-like peptides in the silkmoth Bombyx mori: discovery, structure, secretion, and function. Front. Physiol. 4:1–11.

Nogales E. 2000. Structural Insights into Microtubule Function. Annu. Rev. Biochem. 69:277–302.

Pastor-Pareja JC, Wu M, Xu T. 2008. An innate immune response of blood cells to tumors and tissue damage in Drosophila. Dis. Model. Mech. 1:144–154.

Petersen TN, Brunak S, Von Heijne G, Nielsen H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat. Methods 8:785–786.

Råberg L, Grahn M, Hasselquist D, Svensson E, B PRSL, Rôberg L. 1998. On the adaptive significance of stress-induced immunosuppression On the adaptive signi ® cance of stress-induced immunosuppression. Society:1637–1641.

Reinecke JP, Buckner JS, Grugel SR. 1980. Life cycle of laboratory-reared tobacco hornworms Manduca sexta, a study of development and behavior, using time-lapse cinematography. Biol. Bull. 158:129–140.

Robinson MD, McCarthy DJ, Smyth GK. 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.

Roesijadi G, Rezvankhah S, Binninger DM, Weissbach H. 2007. Ecdysone induction of MsrA protects against oxidative stress in Drosophila. Biochem. Biophys. Res. Commun. 354:511–516.

Roseland CR, Schneiderman HA. 1979. Regulation and metamorphosis of the abdominal histoblasts of Drosophila melanogaster. Wilhelm Roux's Arch. Dev. Biol. 186:235–265.

Samakovlis C, Kimbrell DA, Kylsten P, Engström A, Hultmark D. 1990. expression and biological activity. EMBO J. 9:2969–2976.

Sanchez CG, Teixeira FK, Czech B, Preall JB, Zamparini AL, Seifert JRK, Malone CD, Hannon GJ, Lehmann R. 2015. Regulation of ribosome biogenesis and protein synthesis controls germline stem cell differentiation. Cell Stem Cel 57:742–768.

Schäfer M, Borsch D, Hulster A, Schäfer U. 1993. Expression of a Gene Duplication Encoding Conserved Sperm Tail Proteins Is Translationally Regulated in Drosophila melanogaster. Mol. Cell. Biol. 13:1708–1718.

Scott MP, Carroll SB. 1987. The segmentation and homeotic gene network in early Drosophila development. Cell 51:689–698.

Sega GA. 1984. A review of the genetic effects of ethyl methanesulfonate. Mutat. Res. 134:113–142.

Shingleton AW. 2010. The regulation of organ size in Drosophila: Physiology, plasticity, patterning and physical force. Organogenesis 6:76–87.

Shukla E, Thorat LJ, Nath BB, Gaikwad SM. 2015. Insect trehalase: physiological significance and potential applications. Glycobiology 25:357–367.

Siddiki S, Zambare SP. 2017. Studies on Time Duration of Life Stages of Chrysomya megacephala and Chrysomya rufifacies (Diptera : Calliphoridae) during Different Seasons. J. Forensic Res. 8.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.

Simpson P, Berreur P, Berreur-Bonnenfant J. 1980. The initiation of pupariation in Drosophila: dependence on growth of the imaginal discs. J. Embryol. Exp. Morphol. 57:155–165.

Steinberg BE, Grinstein S. 2007. Unconventional roles of the NADPH oxidase: signaling, ion homeostasis, and cell death. Sci. STKE 2007:1–5.

Støy J, Edghill EL, Flanagan SE, Ye H, Paz VP, Pluzhnikov A, Below JE, Hayes MG, Cox NJ, Lipkind GM, et al. 2007. Insulin gene mutations as a cause of permanent neonatal diabetes. 104.

Suzuki S, Kakuta M, Ishida T, Akiyama Y. 2015. Faster sequence homology searches by clustering subsequences. Bioinformatics 31:1183–1190.

Takayama S, Xie Z, Reed JC. 1999. An Evolutionarily Conserved Family of Hsp70 / Hsc70 Molecular Chaperone Regulators *. Biochemistry 274:781–786.

Tatun N, Singtripop T, Tungjitwitayakul J, Sakurai S. 2008. Regulation of soluble and membrane-bound trehalase activity and expression of the enzyme in the larval midgut of the bamboo borer Omphisa fuscidentalis. Insect Biochem. Mol. Biol. 38:788–795.

Trienens M, Kraaijeveld K, Wertheim B. 2017. Defensive repertoire of Drosophila larvae in response to toxic fungi. Mol. Ecol. 26:5043–5057.

Vallejo DM, Juarez-Carreno S, Bolivar J, Morante J, Dominguez M. 2015. A brain circuit that synchronizes growth and maturation revealed through Dilp8 binding to Lgr3. Science (80-.). 350:aac6767-aac6767.

Wang X, Yu H, Geng L, Xu J, Yu D, Zhang S, Ma Y, Fei D-Q. 2016. Comparative Transcriptome Analysis of Bombyx mori (Lepidoptera) Larval Midgut Response to BmNPV in Susceptible and Near- Isogenic Resistant Strains. PLoS One 11:e0155341.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10:57–63.

Weston CR, Davis RJ. 2002. The JNK signal transduction pathway. Curr. Opin. Genet. Dev. 12:14–21.

Whitworth C, Jimenez E, Doren M Van. 2012. Development of sexual dimorphism in the Drosophila testis. Spermatogenesis 2:129–136.

Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim J, Blagoderov V, Caravas J, Narayanan S, Schmidt-ott U, Kampmeier GE, et al. 2011. Episodic radiations in the fly tree of life. Proc. Natl. Acad. Sci. 108:5690–5695.

Worley MI, Setiawan L, Hariharan IK. 2012. Regeneration and Transdetermination in *Drosophila* Imaginal Discs. Annu. Rev. Genet. 46:289–310.

Yang Y, Smith SA. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. Mol. Biol. Evol. 31:3081–3092.

Zhang Z, Palli SR. 2009. Identification of a cis -regulatory element required for 20-hydroxyecdysone enhancement of antimicrobial peptide gene expression in Drosophila melanogaster. Insect Mol. Biol. 18:595–605.

Anexos

Tabelas (Anexos T1 – 11)

As tabelas anexadas estão disponíveis no endereço eletrônico:

http://www.ib.usp.br/biologia/torres/wordpress/monfardini

Figuras (Anexos F1 – F9)

```
Subject = Stomoxys calcitrans SCAU012402-PB

 Score = 122 bits (306),  Expect = 8e-39, Method: Compositional matrix adjust.
 Identities = 68/149 (46%), Positives = 96/149 (64%), Gaps = 11/149 (7%)

Query  9    RWMLLVIGVCCLMGS----SSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHH  64
            +  +  I V C +G+     + ++CSL+RMKK AMEACEHLF+  +   R+RRSI+   H
Sbjct  5    KSFVAAIFVVCYLGALCERGNATYCSLDRMKKLAMEACEHLFETFDV--RERRSIDH-EH  61

Query  65   HLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIK---PHHEKK  121
             +N + +      +HY+ RSSYP GGYLK+   H+ +LS+LDI PRYK  K      E+K
Sbjct  62   GVNSVETSARERFYHYVRRSSYPQGGYLKMGLNHYRKLSQLDISPRYKAKKLSQLQQEEK  121

Query  122  HRFKRDHSSRSYNNIPYCCLNQCEEEFFC  150
             RFKR+  S S N+I YCC ++C+E+FFC
Sbjct  122  MRFKRNEESYS-NSISYCCYHKCDEDFFC  149
```

```
Subject = Musca domestica MDOA009193-PA

 Score = 122 bits (305),  Expect = 1e-38, Method: Compositional matrix adjust.
 Identities = 68/144 (47%), Positives = 96/144 (67%), Gaps = 10/144 (7%)

Query  10   WMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHHHLNRL  69
            +++ ++GV C  G  S ++CSL+RMKK AMEACEHLF+A +   R+RRSI+  H H++
Sbjct  12   FLICLLGVMC--GRGSANYCSLDRMKKLAMEACEHLFEAYD--VRERRSID--HGHMDNN  65

Query  70   GSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKP---HHEKKHRFKR  126
               +  +   Y+ RSSYP GGYLKV  +H+ +LS LDI PRYK  K     H +K RFKR
Sbjct  66   AEYEPTSSFRYVRRSSYPQGGYLKVGIDHYRKLSTLDINPRYKAKKVWQMQHGEKLRFKR  125

Query  127  DHSSRSYNNIPYCCLNQCEEEFFC  150
            +  S   ++I YCC ++C+E+FFC
Sbjct  126  NEDSYG-SSISYCCFHKCDEDFFC  148
```

```
Subject = Stomoxys calcitrans SCAU012402-PA

 Score = 122 bits (306),  Expect = 8e-39, Method: Compositional matrix adjust.
 Identities = 68/149 (46%), Positives = 96/149 (64%), Gaps = 11/149 (7%)

Query  9    RWMLLVIGVCCLMGS----SSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHH  64
            +  +  I V C +G+     + ++CSL+RMKK AMEACEHLF+  +   R+RRSI+   H
Sbjct  4    KSFVAAIFVVCYLGALCERGNATYCSLDRMKKLAMEACEHLFETFDV--RERRSIDH-EH  60

Query  65   HLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIK---PHHEKK  121
             +N + +      +HY+ RSSYP GGYLK+   H+ +LS+LDI PRYK  K      E+K
Sbjct  61   GVNSVETSARERFYHYVRRSSYPQGGYLKMGLNHYRKLSQLDISPRYKAKKLSQLQQEEK  120

Query  122  HRFKRDHSSRSYNNIPYCCLNQCEEEFFC  150
             RFKR+  S S N+I YCC ++C+E+FFC
Sbjct  121  MRFKRNEESYS-NSISYCCYHKCDEDFFC  148
```

```
Subject = Glossina brevipalpis GBRI004816-PA

 Score = 110 bits (275),  Expect = 1e-33, Method: Compositional matrix adjust.
 Identities = 68/159 (43%), Positives = 94/159 (59%), Gaps = 19/159 (12%)

Query  11   MLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHHHLNRLG  70
            +LL I    LM  ++ ++CSL+RMK FAMEACEHL + ++   R+ RS+        R G
Sbjct  17   LLLSITQSLLMTFTTATYCSLDRMKMFAMEACEHLLEFEDLQARETRSVFHGLKKNERNG  76

Query  71   SGKTH----NKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIK-PHHEKKHRFK  125
            +  +     +  +YI RS YP GGYLKV+ EHF +LS LD+FP+YK  +  HHE K R K
Sbjct  77   NFVSARYRGDNRNYIRRSIYPQGGYLKVSFEHFQKLSRLDVFPKYKAKQNSHHEIKRRHK  136

Query  126  R--------------DHSSRSYNNIPYCCLNQCEEEFFC  150
            R              ++S+ + NNI YCC + C+EEFFC
Sbjct  137  RDDNNYNIGNNDNKNNNSNNNNNNITYCCYHTCDEEFFC  175
```

Anexo F1: Alinhamentos dos homólogos ao Ilp8 provenientes do VectorBase: Alinhamento das quatro sequências proteícas do Ilp8 (atuando como *subject*) contra a proteína de *Dr. melanogaster* (atuando como *query*)

```
Subject = Rhagoletis zephyria XP_017486945.1

 Score = 162 bits (411),  Expect = 2e-53, Method: Compositional matrix adjust.
 Identities = 79/142 (56%), Positives = 105/142 (74%), Gaps = 2/142 (1%)

Query  9    RWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHHHLNR  68
            R +L+ I  C  +G S+ ++CSLE+MKKFAMEACEHLF  DE   +   S+++ H+ ++
Sbjct  2    RLLLITIIACGFIGKSN-AYCSLEKMKKFAMEACEHLFMQDEAREKRSLSMQYDHNVIHH  60

Query  69   LGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEKKHRFKRDH  128
                K ++K HYISRS YP GGYLKV +EHF  LS++D+ PRYKP K HH+KK R+KR+H
Sbjct  61   NEYPKINDKSHYISRSLYPRGGYLKVGQEHFQILSQIDVVPRYKPKKLHHDKKERYKREH  120

Query  129  SSRSYNNIPYCCLNQCEEEFFC  150
             + SYNNIPYCC N+CEE+FFC
Sbjct  121  -ALSYNNIPYCCYNKCEEDFFC  141
```

```
Subject = Bactrocera oleae XP_014094322.1

 Score = 161 bits (407),  Expect = 8e-53, Method: Compositional matrix adjust.
 Identities = 79/146 (54%), Positives = 106/146 (73%), Gaps = 1/146 (1%)

Query  5    LHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHH  64
            +H+   +LL   + C   + S ++CSLE+MK FAMEACEHLFQ DEG  +    I + H+
Sbjct  1    MHITLRILLPTLLICGAVAKSHAYCSLEKMKMFAMEACEHLFQQDEGREKRSLPIPYEHN  60

Query  65   HLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEKKHRF  124
            +++     K ++K H+ISRS YP GGYLKV +EHF  LS++D+FPRYKP K HH+KK R+
Sbjct  61   YIHHNEYPKLNDKSHFISRSLYPHGGYLKVGQEHFKILSKVDVFPRYKPKKLHHDKKERY  120

Query  125  KRDHSSRSYNNIPYCCLNQCEEEFFC  150
            KR+H + SYNNIPYCC N+CEE+FFC
Sbjct  121  KREH-ALSYNNIPYCCYNKCEEDFFC  145
```

```
Subject =  Zeugodacus cucurbitae XP_011196195.1

 Score = 161 bits (407),  Expect = 8e-53, Method: Compositional matrix adjust.
 Identities = 79/146 (54%), Positives = 105/146 (72%), Gaps = 1/146 (1%)

Query  5    LHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHH  64
            +H+   +LL   + C   + S ++CSLE+MK FAMEACEHLFQ DEG  +    I + H+
Sbjct  1    MHITLRILLPTLLICGAVAKSHAYCSLEKMKMFAMEACEHLFQQDEGREKRSLPIPYEHN  60

Query  65   HLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEKKHRF  124
            ++      K ++K H+ISRS YP GGYLKV +EHF  LS++D+FPRYKP K HH+KK R+
Sbjct  61   YIRHNEYPKLNDKRHFISRSLYPHGGYLKVGQEHFKILSKVDVFPRYKPKKLHHDKKERY  120

Query  125  KRDHSSRSYNNIPYCCLNQCEEEFFC  150
            KR+H+  SYNNIPYCC N+CEE+FFC
Sbjct  121  KREHAL-SYNNIPYCCYNKCEEDFFC  145
```

```
Subject =  Bactrocera latifrons XP_018804119.1

 Score = 160 bits (404),  Expect = 2e-52, Method: Compositional matrix adjust.
 Identities = 78/146 (53%), Positives = 107/146 (73%), Gaps = 1/146 (1%)

Query  5    LHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHH  64
            +H+   +LL   + C + + S ++CSLE+MK FAMEACEHLFQ DEG  +    + + H+
Sbjct  1    MHITLRILLPTLLVCGVLAKSHAYCSLEKMKMFAMEACEHLFQQDEGREKRSLPLPYEHN  60

Query  65   HLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEKKHRF  124
            +++     K ++K H+ISRS YP GGYLKV +EHF  LS++D+FPRYKP K HH+KK R+
Sbjct  61   YIHHNEYPKLNDKRHFISRSLYPHGGYLKVGQEHFKILSKVDVFPRYKPKKLHHDKKGRY  120

Query  125  KRDHSSRSYNNIPYCCLNQCEEEFFC  150
            KR+H + SYNNIPYCC N+CEE+FFC
Sbjct  121  KREH-ALSYNNIPYCCYNKCEEDFFC  145
```

```
Subject = Bactrocera dorsalis XP_011201638.1

 Score = 159 bits (403),  Expect = 3e-52, Method: Compositional matrix adjust.
 Identities = 78/146 (53%), Positives = 106/146 (73%), Gaps = 1/146 (1%)

Query  5    LHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHH  64
            +H+   +LL   + C   + S ++CSLE+MK FAMEACEHLFQ DEG  +    + + H+
Sbjct  1    MHITLRILLPTLLICGALAKSHAYCSLEKMKMFAMEACEHLFQQDEGREKRSLPLPYEHN  60

Query  65   HLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEKKHRF  124
            +++     K ++K H+ISRS YP GGYLKV +EHF  LS++D+FPRYKP K HH+KK R+
Sbjct  61   YIHHNEYPKLNDKRHFISRSLYPHGGYLKVGQEHFKILSKVDVFPRYKPKKLHHDKKERY  120

Query  125  KRDHSSRSYNNIPYCCLNQCEEEFFC  150
            KR+H + SYNNIPYCC N+CEE+FFC
Sbjct  121  KREH-ALSYNNIPYCCYNKCEEDFFC  145
```

```
Subject = Ceratitis capitata XP_004525593.1

 Score = 158 bits (400),  Expect = 8e-52, Method: Compositional matrix adjust.
 Identities = 79/146 (54%), Positives = 106/146 (73%), Gaps = 2/146 (1%)

Query  5    LHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHH  64
            +H+    L +    C + + S ++CSLERMKKFAMEACEHLF  DEG  +    + + H+
Sbjct  1    MHIALRTLFLTFFFCGVVAKSNAYCSLERMKKFAMEACEHLFMQDEGREKRSLPMPYEHN  60

Query  65   HLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEKKHRF  124
            +L+     K ++K H+ISRSS+P GGYLKV +EHF  LS++D+FPRYKP K HH+KK R+
Sbjct  61   YLHHNEYPKHNDKSHFISRSSFPRGGYLKVGQEHFQILSKVDVFPRYKP-KKHHDKKERY  119

Query  125  KRDHSSRSYNNIPYCCLNQCEEEFFC  150
            KR+H + SYNNIPYCC N+CEE+FFC
Sbjct  120  KREH-ALSYNNIPYCCYNKCEEDFFC  144
```

```
Subject = Lucilia cuprina XP_023293322.1

 Score = 107 bits (266),  Expect = 2e-31, Method: Compositional matrix adjust.
 Identities = 69/152 (45%), Positives = 91/152 (60%), Gaps = 16/152 (11%)

Query  8    CRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFA--HHH  65
            C   LLVI    + G S  S CSL+RM+++A +ACEHLF  D   R +RRS++    + +
Sbjct  7    CLSGLLVICCIYIFGFSKASACSLDRMREYAKQACEHLFDFDTEVR-ERRSLDLKKNYDY  65

Query  66   LNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHH---EKKH  122
            L R       N    + RS YP GGYLKV  EH+++LS LD+ PRYK  K  H   EKK
Sbjct  66   LPRQPLKHFLN----VRRSIYPPGGYLKVAPEHYDKLSILDVSPRYKSKKISHIMNEKKM  121

Query  123  RFKRDHSSRSY----NNIPYCCLNQCEEEFFC  150
            R KR+   ++Y    NNIPYCC ++C+ EFFC
Sbjct  122  RNKRE--GKNYLNHNNNIPYCCFHRCDAEFFC  151
```

Anexo F2: Alinhamentos dos homólogos ao Ilp8 provenientes do NCBI: Alinhamento das sete sequências proteícas do Ilp8 (atuando como *subject*) contra a proteína de *Dr. melanogaster* (atuando como *query*).

```
Subject = Chrysomya albiceps c4426_g1_i1

Score          Expect Method                        Identities   Positives    Gaps
119 bits(299)  5e-38  Compositional matrix adjust.  71/146(49%)  96/146(65%)  11/146(7%)

Query  13   LVIGVCCLM----GSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHHHLNR  68
            ++  +CC+       +S ++CSL+ MK+ AM+ACEHL+   E   R+RRS++  +H  N
Sbjct  10   ILFAICCISINCCSQTSATYCSLDIMKEHAMDACEHLYSL-ESEVRERRSLD--NHPNNH  66

Query  69   LGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPH--HEKKHRFKR  126
                K+ N   +ISRSSYP GGYLKV REH++RLS LD+ PRYKP K +  HE   R KR
Sbjct  67   KHYKKSSNFLVHISRSSYPKGGYLKVNREHYDRLSRLDVRPRYKPTKFNLKHEMSLRKKR  126

Query  127  -DHSSRSY-NNIPYCCLNQCEEEFFC  150
             D    +Y NNI +CC+N+CEEEFFC
Sbjct  127  EDKEFYTYGNNIAFCCINKCEEEFFC  152
```

```
Subject = Chrysomya megacephala c6857_g1_i1

Score          Expect Method                        Identities   Positives    Gaps
106 bits(265)  8e-33  Compositional matrix adjust.  63/135(47%)  82/135(60%)  14/135(10%)

Query  24   SSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIE---FAHHHLNRLGSGKTHNKHHY  80
            +S ++CSL+ MK+ AM+ACEHL+   E   R+RRSI+     +H   +      H
Sbjct  25   TSATYCSLDVMKEHAMDACEHLYSL-ESEVRERRSIDDNDIYYHKHKKASKFLVH-----  78

Query  81   ISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPH---HEKKHRFKRDHSS--RSYNN  135
            I+RSSYP GGYLKV R H+  LS L++ PRYKP K +   HE   R KR+     R  NN
Sbjct  79   ITRSSYPRGGYLKVNRVHYTHLSHLNVTPRYKPTKLNHLKHEMSLRNKREDKDFYRYSNN  138

Query  136  IPYCCLNQCEEEFFC  150
            I +CC N+CEEEFFC
Sbjct  139  IAFCCFNKCEEEFFC  153
```

```
Subject =  Oestrus ovis c4455_g1_i1

Score          Expect Method                        Identities   Positives    Gaps
114 bits(285)  1e-35  Compositional matrix adjust.  69/153(45%)  96/153(62%)  24/153(15%)

Query  11   MLLVIGVCCL----MGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHHHL  66
            + +++ + C+    +  +S + CSL+RM+KFAM+ACEHLF+  E   R+RRSI
Sbjct  8    VTILLAISCISSFCISGTSATHCSLDRMRKFAMDACEHLFKF-ETEVRERRSIV------  60

Query  67   NRLGSGKTHNKHHYI--SRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHH---EKK  121
               GS  +H   ++I  SR  YP+GGYLKV +EH+N+LS LD+ PRYK  K  H   EKK
Sbjct  61   --SGSRNSHTTKYWIDISRGIYPVGGYLKVGQEHYNKLSLLDVSPRYKAKKFAHIMREKK  118

Query  122  HRFKRDHSSRSY----NNIPYCCLNQCEEEFFC  150
             R KR+   + Y    NNI YCC ++CEE+FFC
Sbjct  119  MRNKRE--DKDYHNFNNNISYCCYHKCEEDFFC  149
```

```
Subject = Hermetia illucens c14571_g1_i2

Score          Expect Method                        Identities   Positives    Gaps
55.5 bits(132) 2e-13  Compositional matrix adjust.  40/138(29%)  65/138(47%)  17/138(12%)

Query  22   GSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHHHLNRLGSGKTHNKHH--  79
            G+SS   C+LE M++ A  AC+H+F       R++R +  +  +      G  +N  H
Sbjct  25   GASSRKICTLEEMRELAANACDHMFM------REKRQVSNSLLNSWDTPKGVVNNLRHPS  78

Query  80   ---YISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEK----KHRFKRDHSSRS  132
                + R  +P GGYL +  ++   LS+    P+Y  I  H +K     +R KR +S+
Sbjct  79   KGVKVRRRDFPEGGYLIIDHKYAKILSQRIEKPKYHVIDWHSKKFVGMNYRSKRANSATI  138

Query  133  YNNIPYCCLNQCEEEFFC  150
               I YCC ++C     C
Sbjct  139  --GITYCCTHKCSPLDIC  154
```

```
Subject = Hermetia illucens c11423_g3_i1

Score          Expect Method                        Identities   Positives    Gaps
48.9 bits(115) 8e-13  Compositional matrix adjust.  43/150(29%)  71/150(47%)  23/150(15%)

Query  14   VIGVCCLMGSSSGS--FCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHHHLNRLG-  70
            VI +      S+GS   C+L+ M++ A  AC+H+F       R++R I  ++  LN
Sbjct  37   VITLSIQKSVSTGSRRVCTLDEMRELAANACDHMFM------REKRQI--SNSLLNSWDH  192

Query  71   --SGKTH----NKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEK----  120
              SG  H    +K   + R  +P GGYL +  ++   LS+    P+Y  +  H +K
Sbjct  193  PKSGTAHLHHPSKGVKVRRRDFPDGGYLIIDHKYAKILSQRIEKPKYHVVDWHSKKFANM  372

Query  121  KHRFKRDHSSRSYNNIPYCCLNQCEEEFFC  150
             +R KR +S+     + YCC ++C     C
Sbjct  373  DYRSKRANSAGI--GVTYCCTHKCSPLDIC  456
```

Anexo F3: Alinhamentos dos homólogos ao Ilp8 provenientes do nosso laboratório: Alinhamento das cinco sequências proteícas do Ilp8 (atuando como *subject*) contra a proteína de *Dr. melanogaster* (atuando como *query*).

```
Subject = Drosophila simulans

 Score = 310 bits (793),  Expect = 2e-114, Method: Compositional matrix adjust.
 Identities = 149/150 (99%), Positives = 150/150 (100%), Gaps = 0/150 (0%)

Query  1    MSSKLHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIE  60
            MSSKLHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIE
Sbjct  1    MSSKLHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIE  60

Query  61   FAHHHLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEK  120
            +AHHHLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEK
Sbjct  61   YAHHHLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEK  120

Query  121  KHRFKRDHSSRSYNNIPYCCLNQCEEEFFC  150
            KHRFKRDHSSRSYNNIPYCCLNQCEEEFFC
Sbjct  121  KHRFKRDHSSRSYNNIPYCCLNQCEEEFFC  150


Subject = Drosophila sechellia

 Score = 306 bits (785),  Expect = 3e-112, Method: Compositional matrix adjust.
 Identities = 145/146 (99%), Positives = 146/146 (100%), Gaps = 0/146 (0%)

Query  5    LHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHH  64
            LHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIE+AHH
Sbjct  65   LHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEYAHH  124

Query  65   HLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEKKHRF  124
            HLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEKKHRF
Sbjct  125  HLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEKKHRF  184

Query  125  KRDHSSRSYNNIPYCCLNQCEEEFFC  150
            KRDHSSRSYNNIPYCCLNQCEEEFFC
Sbjct  185  KRDHSSRSYNNIPYCCLNQCEEEFFC  210


Subject = Drosophila erecta

 Score = 295 bits (754),  Expect = 2e-108, Method: Compositional matrix adjust.
 Identities = 140/150 (93%), Positives = 149/150 (99%), Gaps = 0/150 (0%)

Query  1    MSSKLHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIE  60
            MSSKL+MCRWMLLV+GVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIE
Sbjct  1    MSSKLNMCRWMLLVLGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIE  60

Query  61   FAHHHLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEK  120
            +AHHHLNR+GSGK+ +KHHYISRSSYPMGGYLKVTR+HFNRLSELDIFPRYKPIKPHHEK
Sbjct  61   YAHHHLNRVGSGKSVHKHHYISRSSYPMGGYLKVTRDHFNRLSELDIFPRYKPIKPHHEK  120

Query  121  KHRFKRDHSSRSYNNIPYCCLNQCEEEFFC  150
            K+RFKRDHS+RSYNNIPYCCLNQCEEEFFC
Sbjct  121  KNRFKRDHSNRSYNNIPYCCLNQCEEEFFC  150


Subject = Drosophila yakuba

 Score = 282 bits (722),  Expect = 2e-103, Method: Compositional matrix adjust.
 Identities = 136/150 (91%), Positives = 142/150 (95%), Gaps = 0/150 (0%)

Query  1    MSSKLHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIE  60
            MSSKL+MCRWMLLV+GVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEG RRDRRSI
Sbjct  1    MSSKLNMCRWMLLVLGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGGRRDRRSIG  60

Query  61   FAHHHLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEK  120
            + HHHLN +GSGKT +KHHYISRSSYPMGGYLKVT  HF RLSELDIFPRYKPIKPHHEK
Sbjct  61   YPHHHLNGIGSGKTVHKHHYISRSSYPMGGYLKVTGHHFKRLSELDIFPRYKPIKPHHEK  120

Query  121  KHRFKRDHSSRSYNNIPYCCLNQCEEEFFC  150
            K+RFKRDHSSRSYNNIPYCCLNQCEEEFFC
Sbjct  121  KNRFKRDHSSRSYNNIPYCCLNQCEEEFFC  150


Subject = Drosophila ananassae

 Score = 258 bits (658),  Expect = 9e-94, Method: Compositional matrix adjust.
 Identities = 115/150 (77%), Positives = 137/150 (91%), Gaps = 0/150 (0%)

Query  1    MSSKLHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIE  60
            M+SK ++CRWML+ +GVCCL+GSS+GS+CSLERMKK+AMEACEHLF  DE ARRDRRSI+
Sbjct  1    MNSKFNVCRWMLIALGVCCLLGSSTGSYCSLERMKKYAMEACEHLFMQDEAARRDRRSID  60

Query  61   FAHHHLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEK  120
            + HHH+NRLG GK H+KHHYISRS+YP GGYLKV++EHF++LSELD+FPRYKP KPHH+K
Sbjct  61   YPHHHMNRLGYGKIHDKHHYISRSNYPPGGYLKVSQEHFHQLSELDVFPRYKPRKPHHDK  120

Query  121  KHRFKRDHSSRSYNNIPYCCLNQCEEEFFC  150
            K R+KRDH+S SYNNIPYCC NQCEEEFFC
Sbjct  121  KQRYKRDHASSSYNNIPYCCFNQCEEEFFC  150


Subject = Drosophila persimilis

 Score = 243 bits (620),  Expect = 6e-88, Method: Compositional matrix adjust.
 Identities = 111/150 (74%), Positives = 128/150 (85%), Gaps = 0/150 (0%)

Query  1    MSSKLHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIE  60
            M++K ++CRW LLVI V CL+  SSGSFCS+E+MK  AMEACEHLFQ DE ARRD+RSIE
Sbjct  1    MNAKFNLCRWTLLVISVFCLLKESSGSFCSMEKMKTVAMEACEHLFQQDEAARRDKRSIE  60

Query  61   FAHHHLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEK  120
            ++HHHLNRLG GK H+KHHYI+RS+YPMGGYLKV R+H+ RLSELD+ PRYKP K HHEK
Sbjct  61   YSHHHLNRLGYGKLHDKHHYITRSNYPMGGYLKVNRDHYRRLSELDVVPRYKPKKIHHEK  120

Query  121  KHRFKRDHSSRSYNNIPYCCLNQCEEEFFC  150
            K RFKRDH+S SYNNI YCCLN C EEFFC
Sbjct  121  KERFKRDHASSSYNNISYCCLNSCNEEFFC  150


Subject = Drosophila pseudoobscura pseudoobscura

 Score = 243 bits (620),  Expect = 6e-88, Method: Compositional matrix adjust.
 Identities = 111/150 (74%), Positives = 128/150 (85%), Gaps = 0/150 (0%)

Query  1    MSSKLHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIE  60
            M++K ++CRW LLVI V CL+  SSGSFCS+E+MK  AMEACEHLFQ DE ARRD+RSIE
Sbjct  1    MNAKFNLCRWTLLVISVFCLLKESSGSFCSMEKMKTVAMEACEHLFQQDEAARRDKRSIE  60

Query  61   FAHHHLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEK  120
            ++HHHLNRLG GK H+KHHYI+RS+YPMGGYLKV R+H+ RLSELD+ PRYKP K HHEK
Sbjct  61   YSHHHLNRLGYGKLHDKHHYITRSNYPMGGYLKVNRDHYRRLSELDVVPRYKPKKIHHEK  120

Query  121  KHRFKRDHSSRSYNNIPYCCLNQCEEEFFC  150
            K RFKRDH+S SYNNI YCCLN C EEFFC
Sbjct  121  KERFKRDHASSSYNNISYCCLNSCNEEFFC  150


Subject = Drosophila willistoni

 Score = 189 bits (481),  Expect = 7e-67, Method: Compositional matrix adjust.
 Identities = 100/152 (66%), Positives = 120/152 (79%), Gaps = 4/152 (3%)

Query  1    MSSKLHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIE  60
            M+ KL+MCR +LL +     +   + S+CSLE+MKKFAMEACEHLFQ DE A R +RS++
Sbjct  1    MNIKLNMCRSVLLALIAA-CLLQEAHSYCSLEKMKKFAMEACEHLFQQDE-ASRGKRSVD  58

Query  61   FAHHHLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEK  120
            +  HHLNRLG GK H+KHHYISR ++P GGYLKV REH++ LSELDI PRYKP K HHEK
Sbjct  59   YKPHHLNRLGFGKLHDKHHYISRGNFPNGGYLKVNREHYHILSELDIIPRYKPRKFHHEK  118

Query  121  K--HRFKRDHSSRSYNNIPYCCLNQCEEEFFC  150
            K   R+KRDHS+ SYNNIPYCC NQC+EEFFC
Sbjct  119  KERERYKRDHSASSYNNIPYCCFNQCDEEFFC  150


Subject = Drosophila grimshawi

 Score = 178 bits (452),  Expect = 2e-62, Method: Compositional matrix adjust.
 Identities = 86/150 (57%), Positives = 110/150 (73%), Gaps = 2/150 (1%)

Query  1    MSSKLHMCRWMLLVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIE  60
            M SK ++CR +LL + + C + + S ++CSL+ MK FAMEACE LF  DEG  R+RRSI+
Sbjct  1    MHSKFNVCRCLLLAVILGCCLVNESQAYCSLDEMKTFAMEACERLFLQDEG--RERRSID  58

Query  61   FAHHHLNRLGSGKTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEK  120
            ++ HH N+L   K H   HYI RS+YP GGYLKV+REH++RL  LD+ PRYKP K H +K
Sbjct  59   YSGHHENKLAHDKMHKDQHYIGRSNYPRGGYLKVSREHYHRLMRLDVMPRYKPRKQHQDK  118

Query  121  KHRFKRDHSSRSYNNIPYCCLNQCEEEFFC  150
            K RF+RD SS+SYNNI YCC + C EEFFC
Sbjct  119  KMRFRRDSSSKSYNNISYCCYHHCNEEFFC  148


Subject = Drosophila virilis

 Score = 150 bits (380),  Expect = 1e-51, Method: Compositional matrix adjust.
 Identities = 74/138 (54%), Positives = 97/138 (70%), Gaps = 4/138 (3%)

Query  13   LVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHHHLNRLGSG  72
            L++  C L G  +   CSL+ MK FAMEACE LFQ DEG  R+RRSI++  HH N++
Sbjct  10   LILASCLLSGIRA--HCSLDTMKAFAMEACERLFQRDEG--RERRSIDYGDHHANKVSHE  65

Query  73   KTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEKKHRFKRDHSSRS  132
            K H   HY+ RSSYP GGYLKV+ +HF+RL  +D+ PRYKP K H  KK R KR+ S+++
Sbjct  66   KMHKGQHYVGRSSYPHGGYLKVSEQHFHRLIGMDVLPRYKPRKQHQLKKLRPKRNSSAKN  125

Query  133  YNNIPYCCLNQCEEEFFC  150
            Y+NI YCC +QC ++FFC
Sbjct  126  YSNISYCCYHQCNDDFFC  143


Subject = Drosophila mojavensis

 Score = 150 bits (379),  Expect = 2e-51, Method: Compositional matrix adjust.
 Identities = 77/139 (55%), Positives = 101/139 (73%), Gaps = 7/139 (5%)

Query  13   LVIGVCCLMGSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHHHLNRLGSG  72
            L++G C L  S + ++CSL++MK FAMEACE LFQ DE   R+RRSI+ + HH N+L  G
Sbjct  10   LILGSCLL--SETRAYCSLDKMKAFAMEACERLFQKDEP--RERRSIDSSWHHKNKLSHG  65

Query  73   KTHNKH-HYISRSSYPMGGYLKVTREHFNRLSELDIFPRYKPIKPHHEKKHRFKRDHSSR  131
            K   K  H + RSS+P GGYLKV+ +H++RL  LD+ PRYKP+K + EK+H  KRD S +
Sbjct  66   KMDKKTPHGVKRSSFPTGGYLKVSSDHYHRLIHLDVLPRYKPLKHNQEKRH--KRDSSDK  123

Query  132  SYNNIPYCCLNQCEEEFFC  150
            SYNNI YCC N+C E+FFC
Sbjct  124  SYNNISYCCFNRCNEDFFC  142
```

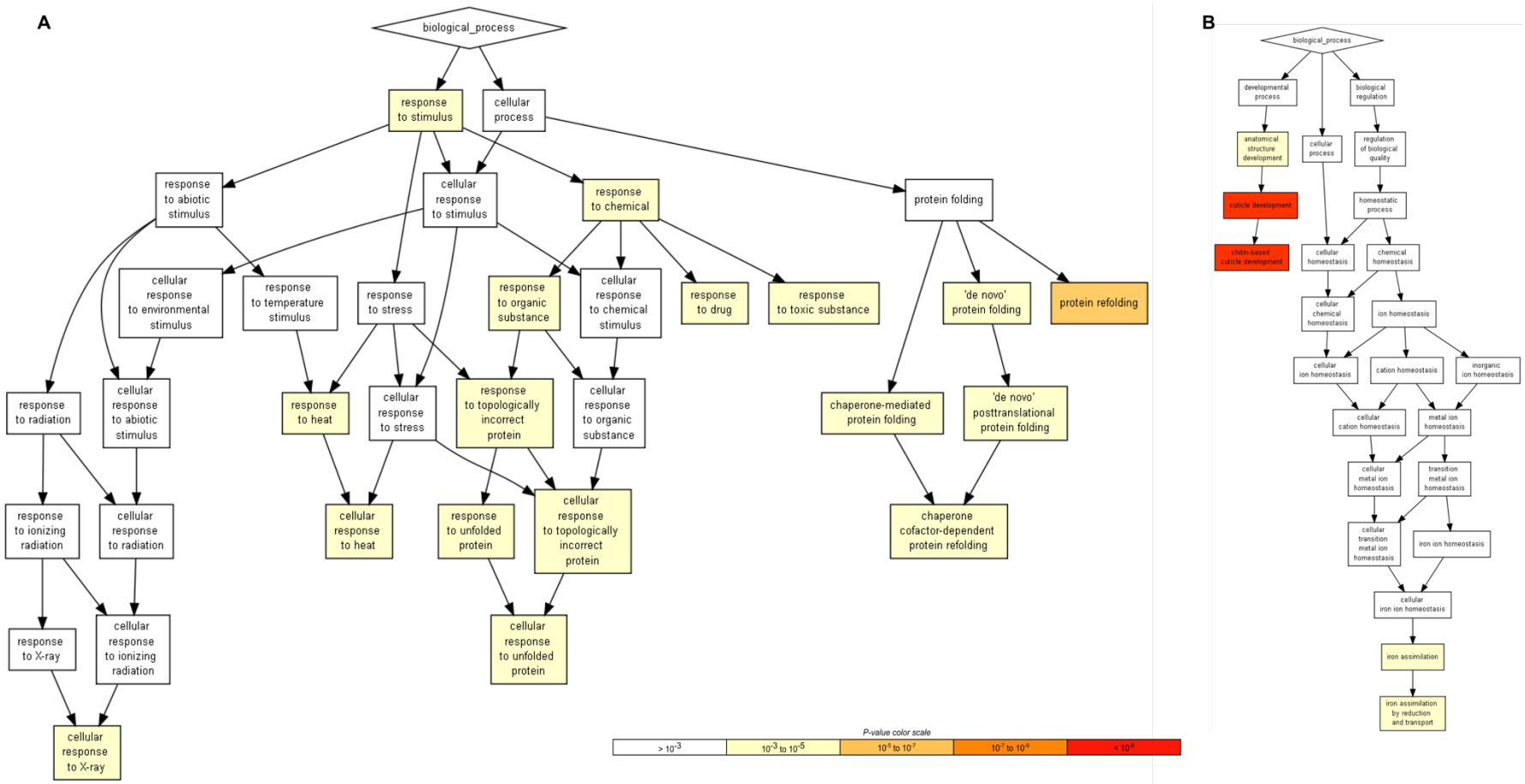
Anexo F4: Alinhamentos dos homólogos ao Ilp8 do grupo *Drosophila*: Alinhamento das 11 sequências proteícas do Ilp8 (atuando como *subject*) contra a proteína de *Dr. melanogaster* (atuando como *query*).

```
Query = Chrysomya megacephala c4426_g1_i1 (previamente descrito)
Subject = TRINITY_DN25329_c0_g1::TRINITY_DN25329_c0_g1_i1::g.74136::m.74136

Score  Expect Method    Identities Positives Gaps
231 bits(590) 2e-78 Compositional matrix adjust. 112/113(99%) 113/113(100%) 0/113(0%)

Query  1    MFLKKNILTSLFAICCISIICCSQTSATYCSLDVMKEHAMDACEHLYSLESEVRERRSID  60
            MFLKKNILTSLFAICCISIICCSQTSATYCSLDVMKEHAMDACEHLYSLESEVRERRSID
Sbjct  1    MFLKKNILTSLFAICCISIICCSQTSATYCSLDVMKEHAMDACEHLYSLESEVRERRSID  60

Query  61   DNDIYYHKHKKASKFLVHITRSSYPRGGYLKVNRVHYTHLSHLNVTPRYKPTK  113
            DNDIYYHKHKKASKFLVHITRSSYPRGGYLKVNRVHYTHLSHL+VTPRYKPTK
Sbjct  61   DNDIYYHKHKKASKFLVHITRSSYPRGGYLKVNRVHYTHLSHLDVTPRYKPTK  113
```

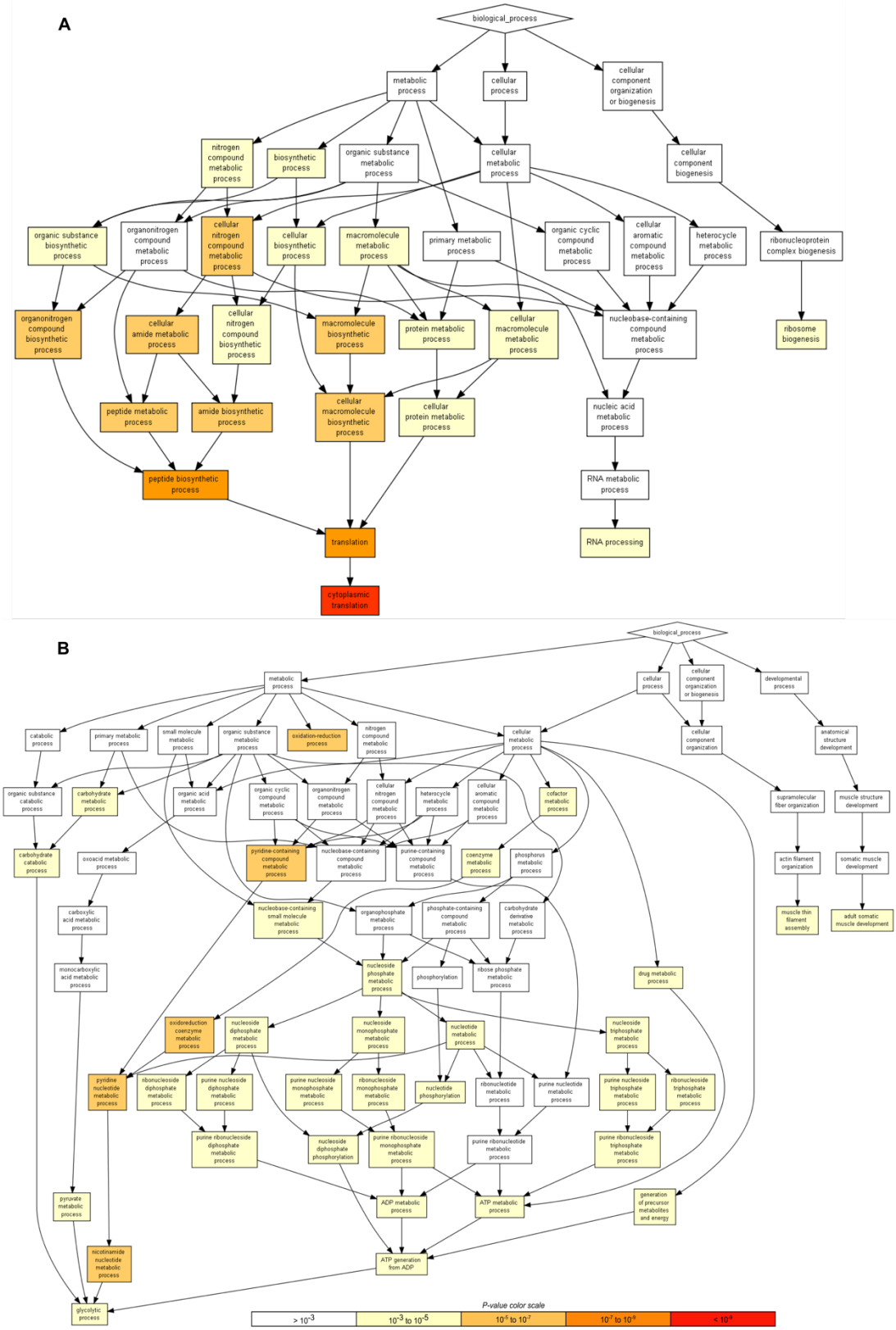
```
Query = Hermetia illucens c14571_g1_i2 (previamente descrito)
Subject = TRINITY_DN10846_c0_g2::TRINITY_DN10846_c0_g2_i3::g.16064::m.16064

Score  Expect Method    Identities Positives Gaps
320 bits(821) 1e-112 Compositional matrix adjust. 153/155(99%) 154/155(99%) 0/155(0%)

Query  3    IQMRVIVAVSVITLSLRKAEGASSRKICTLEEMRELAANACDHMFMREKRQVSNSLLNSW  62
            IQ+RVIVAVSVITLSLRKAE ASSRKICTLEEMRELAANACDHMFMREKRQVSNSLLNSW
Sbjct  5    IQVRVIVAVSVITLSLRKAESASSRKICTLEEMRELAANACDHMFMREKRQVSNSLLNSW  64

Query  63   DTPKGVVNNLRHPSKGVKVRRRDFPEGGYLIIDHKYAKILSQRIEKPKYHVIDWHSKKFV  122
            DTPKGVVNNLRHPSKGVKVRRRDFPEGGYLIIDHKYAKILSQRIEKPKYHVIDWHSKKFV
Sbjct  65   DTPKGVVNNLRHPSKGVKVRRRDFPEGGYLIIDHKYAKILSQRIEKPKYHVIDWHSKKFV  124

Query  123  GMNYRSKRANSATIGITYCCTHKCSPLDICLASRK  157
            GMNYRSKRANSATIGITYCCTHKCSPLDICLASRK
Sbjct  125  GMNYRSKRANSATIGITYCCTHKCSPLDICLASRK  159
```

```
Query = Hermetia illucens c11423_g3_i1 (previamente descrito)
Subject = TRINITY_DN10846_c0_g1::TRINITY_DN10846_c0_g1_i1::g.16061::m.16061

Score  Expect Method    Identities Positives Gaps
323 bits(829) 8e-114 Compositional matrix adjust. 154/157(98%) 156/157(99%) 0/157(0%)

Query  1    MKIQMRILVAMSVITLSIQKSVSTGSRRVCTLDEMRELAANACDHMFMREKRQISNSLLN  60
            MKIQMRILVAMSVITLSIQKSVSTGSRRVCTLDEMRELAANACDHMFMREKRQISNSLLN
Sbjct  1    MKIQMRILVAMSVITLSIQKSVSTGSRRVCTLDEMRELAANACDHMFMREKRQISNSLLN  60

Query  61   SWDHPKSGTAHLHHPSKGVKVRRRDFPDGGYLIIDHKYAKILSQRIEKPKYHVVDWHSKK  120
            SWDHPKSGTAHLHHPSKGVKVRRRDFPDGGYLII+HKYAKILSQR +KPKYHVVDWHSKK
Sbjct  61   SWDHPKSGTAHLHHPSKGVKVRRRDFPDGGYLIIEHKYAKILSQRSQKPKYHVVDWHSKK  120

Query  121  FANMDYRSKRANSAGIGVTYCCTHKCSPLDICLASRK  157
            FANMDYRSKRANSAGIGVTYCCTHKCSPLDICLASRK
Sbjct  121  FANMDYRSKRANSAGIGVTYCCTHKCSPLDICLASRK  157
```

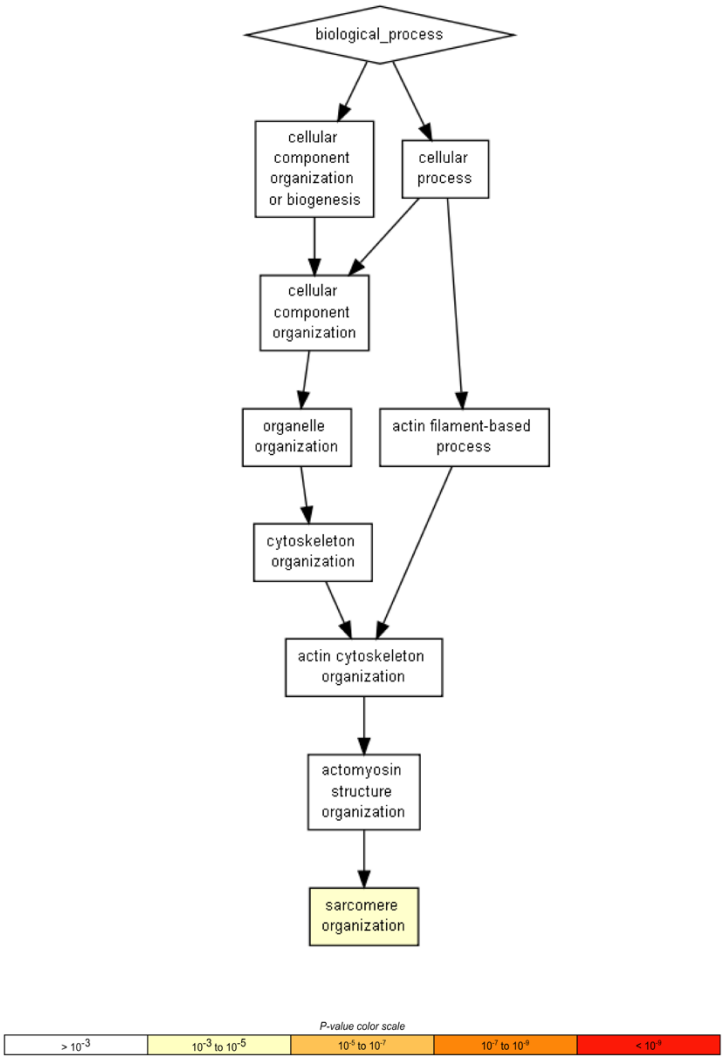
Anexo F5: Recuperação do Ilp8: Recuperação dos transcritos codificantes para a proteína Ilp8 nos transcriptomas larvais de *Ch. megacephala* e *H. illucens*.



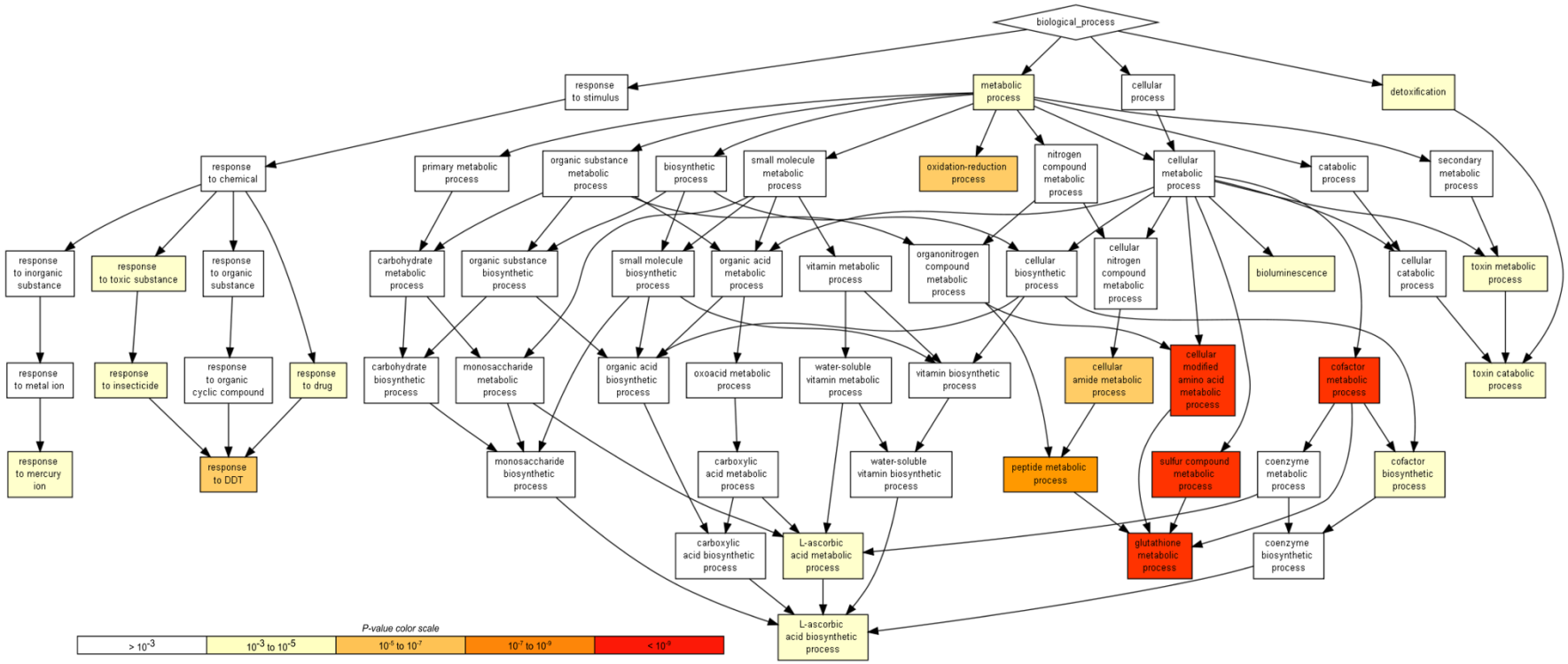
Anexo F6: Análise de enriquecimento de *Ch. megacephala*: Em (A) observa-se os processos biológicos enriquecidos no conjunto de TCDE mais expressos nas larvas com dano tecidual e em (B) observa-se os processos no conjunto de TCDE menos expressos nessa amostra.



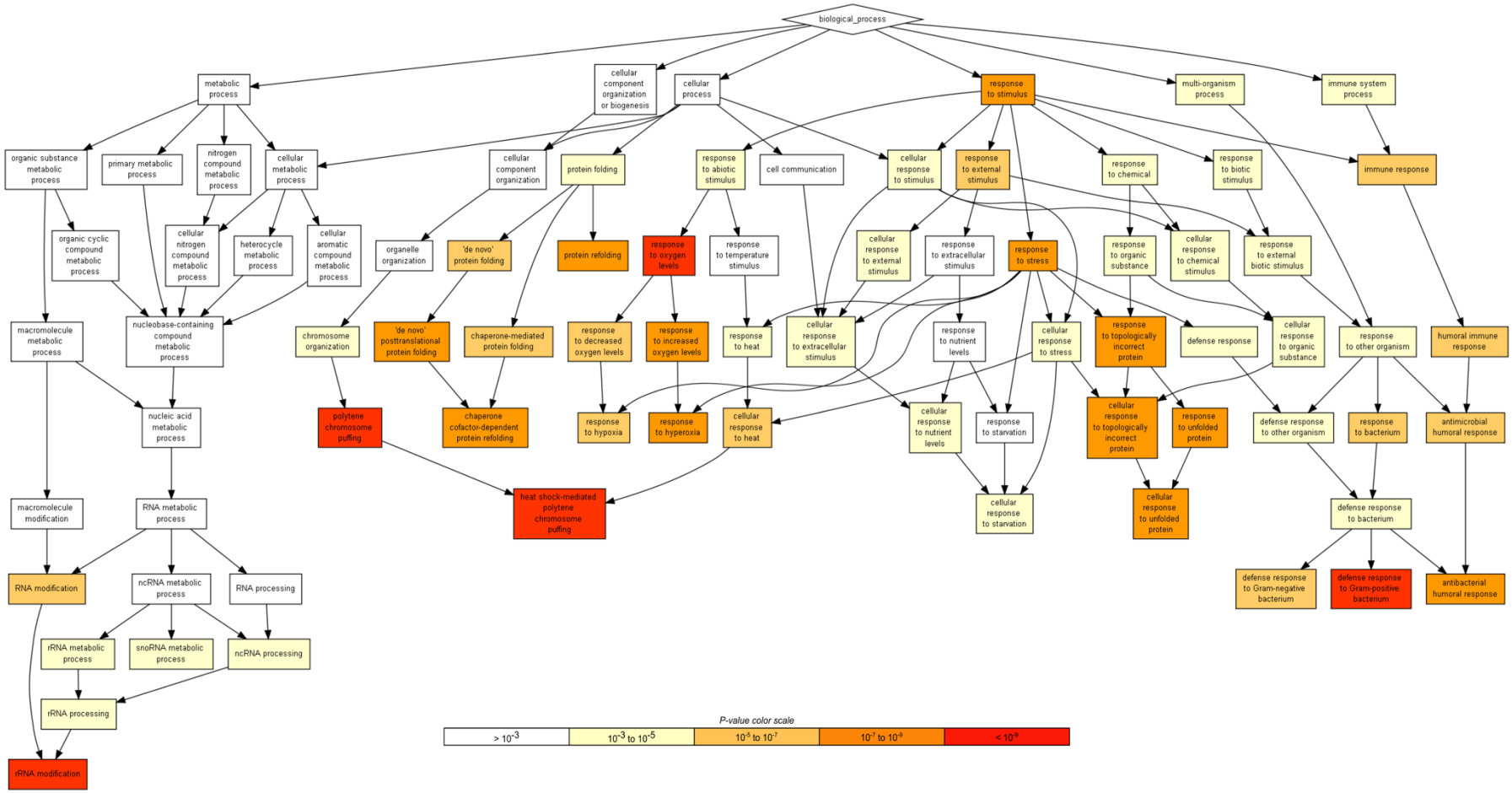
Anexo F7: Análise de enriquecimento de *Ma. sexta* (PBS - EMS): Em (A) observa-se os processos biológicos enriquecidos no conjunto de TCDE mais expressos nas larvas com dano tecidual e em (B) observa-se os processos no conjunto de TCDE menos expressos nessa amostra.



Anexo F8: Análise de enriquecimento de *Ma. Sexta* (*Sham – Pinch*): Observa-se os processos biológicos enriquecidos no conjunto de TCDE menos expressos nas larvas com dano tecidual.



Anexo F9: Análise de enriquecimento de *Dr. melanogaster* (I): Observa-se os processos biológicos enriquecidos no conjunto de TCDE mais expressos nas larvas com dano tecidual.



Anexo F10: Análise de enriquecimento de *Dr. melanogaster* (II): Observa-se os processos biológicos enriquecidos no conjunto de TCDE menos expressos nas larvas com dano tecidual.