Marina Santos Deszo

Genômica funcional da preferência alimentar em espécies da superfamília Oestroidea

Functional genomics of feeding preferences in Oestroidea

São Paulo

2018

Marina Santos Deszo

Genômica funcional da preferência alimentar em espécies da superfamília Oestroidea

Functional genomics of feeding preferences in Oestroidea

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

Orientador(a): Tatiana Teixeira Torres

São Paulo

2018

FICHA CATALOGRÁFICA

Deszo, Marina Santos

Genômica funcional da preferência alimentar em espécies da superfamília Oestroidea / Marina Santos Deszo; orientadora Tatiana Teixeira Torres. -- São Paulo, 2018.

48 f.

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

1. Expressão Gênica. 2. Evolução. 3. Ectoparasitas. 4. Calliphoridae. 5. Oestridae. I. Torres, Tatiana Teixeira, orient. II. Título.

COMISSÃO JULGADORA

_____

Prof (a) Dr(a)

_____

Prof (a) Dr(a)

_____

Prof (a) Dr(a)

_____

Prof (a) Dr(a) Tatiana Teixeira Torres

Orientadora

À minha família, por todo apoio e amor.

[...] caminante, no hay camino,

se hace camino al andar. [...]

Antonio Machado, “Proverbios y cantares” em Campos de Castilla, 1912.

AGRADECIMENTOS

Primeiro gostaria de agradecer meus irmãos, Luciano e Carolina, e especialmente minha mãe, Célia. Vocês têm sido meu chão, meu apoio. Agradeço pelo amor e paciência. Também quero agradecer por aqueles que já se foram, mas tanto me ensinaram sobre amor e cuidado. Meu pai Gilberto, meu avô Deszo e minha avó Tereza, seus ensinamento estarão sempre comigo. Graças a todos que consegui chegar onde cheguei.

Também agradeço por todos meus amigos e familiares, que sempre me dão suporte e me fazem lembrar das incríveis jornadas que já passei. Aqui preciso agradecer minha prima/irmã Monica. Passamos por tantos momentos! Ah, e claro, Kamilla, Ana e Vanessa. São mais que amigas, são irmãs!

Agradeço por todos aqueles que me ajudaram nessa incrível mudança que foi a vinda para a Universidade. Professores, colegas e funcionários fizeram eu acreditar e conseguir vencer enormes desafios. E mais, viver momentos inesquecíveis e inimagináveis. Muito obrigada a minha família da moradia estudantil, especialmente a Kate, que tantas vezes me ouviu e me ajudou em momentos que pensei em desistir.

Também quero agradecer a todos do Laboratório de Genômica e Evolução de Artrópodes, conversas que estimularam o meu pensamento crítico e que aumentaram minha vontade pela ciência. Principalmente Gisele, Raquel, Nancy, Valeria, Sophie e nosso agregado, Tiago. Cada um me ensinou muito. Mas um agradecimento especial para Tatiana Torres, minha orientadora. Sou mais do que grata a todo suporte, por acreditar e pela paciência. Muita paciência. Esse suporte não me fez desistir e me fez e faz querer seguir em frente. Obrigada!

SUMÁRIO

1. INTRODUÇÃO 1
1.1. Miíases e sua importância 1
1.2. Família Calliphoridae 2
1.3. Família Oestridae 3
1.4. Genômica Funcional 5
2. PERGUNTAS, HIPÓTESES E DESENHO EXPERIMENTAL 7
3. MATERIAIS E MÉTODOS 11
3.1. Coletas e manutenção das larvas 11
3.2. Extração do RNA 11
3.3. Sequenciamento 12
3.4. Aquisição de dados públicos 12
3.5. Pré-processamento das “reads” 13
3.6. Montagem 13
3.7. Identificação de sequências ortólogas 14
3.8. Anotação das sequências ortólogas 16
3.9. Análise de expressão gênica 17
3.10. Análise das sequências codificantes 19
4. RESULTADOS 21
4.1. Coleta, extração e sequenciamento do RNA das larvas 21
4.2. Pré-processamento 21
4.3. Montagem 22
4.4. Identificação dos transcritos ortólogos 23
4.5. Análise quantitativa da expressão gênica 24
4.6. Evidência de seleção nas sequências codificantes 28
5. DISCUSSÃO 30
6. CONCLUSÕES 36
7. RESUMO 38
8. ABSTRACT 38
9. REFERÊNCIAS 42

1. INTRODUÇÃO

1.1. Miíases e sua importância

Um tipo comum de infestações em vertebrados é causado por larvas de dípteros que se alimentam de seus tecidos, as chamadas miíases. Essas infestações podem causar em seus hospedeiros desde simples desconfortos até mesmo sua morte (Zumpt, 1965). Os agentes causadores de miíases podem ser definidos como obrigatórios ou facultativos. Aqueles definidos como obrigatórios (parasitas obrigatórios) se alimentam de um hospedeiro vivo para completar o estágio de vida larval. Já os ditos facultativos (parasitas facultativos) podem completar seu estágio de vida larval se alimentando tanto de um hospedeiro vivo, quanto de matéria orgânica em decomposição (Zumpt, 1965). Um segundo sistema de classificação, separa as espécies causadoras de miíases em primárias, quando uma espécie tem capacidade de iniciar a infestação; ou secundárias, que invadem o hospedeiro após uma espécie primária ou um parasita obrigatório já ter iniciado a infestação (Zumpt 1965).

As infestações por larvas de dípteros representam grande transtorno a pecuária em todo o mundo. Há uma perda na produtividade e gastos no manejo, que juntos somam perdas importantes para o setor (Hall & Wall, 1995). No último levantamento feito por Grisi e colaboradores (2014), as perdas causadas pela baixa produtividade de leite, carne e couro, decorrente de miíases foi de cerca de 720 milhões de dólares por ano no Brasil.

Além dos prejuízos à pecuária, as miíases têm uma grande importância médica, principalmente em regiões desfavorecidas socioenomicamente. Por exemplo, *Cochliomyia hominivorax* e *Chrysomya bezziana*, espécies da família Calliphoridae do novo e velho mundo, respectivamente, causam infestações que, quando não tratadas, podem levar a infecções dérmicas, urogenitais ou gástricas (Hall & Wall, 1995). Essas duas espécies são bastante conhecidas por parasitarem humanos, animais domésticos e vertebrados selvagens. Outra espécie

causadora de miíases é a *Wohlfahrtia magnifica* com distribuição no sudoeste paleártico. Essa espécie pertencente a família Sarcophagidae é encontrada parasitando humanos e outros animais de sangue quente (Robbins & Khachemoune, 2010).

Esses dípteros são integrantes da superfamília Oestroidea, que pertence ao grupo Calyptratae (Diptera: Brachycera). Oestroidea é bastante diversa com relação às estratégias de vida dos adultos e/ou larvas, que podem ser parasitas obrigatórios, predadores de vertebrados e invertebrados, saprófagas, fitófagas e hematófagas de amniotas (Stevens *et al.*, 2006; Kutty *et al.*, 2010). Oestroidea atualmente compreende sete famílias: Calliphoridae, Mystacinobiidae, Rhinophoridae, Oestridae, Sarcophagidae, Tachinidae e a antiga subfamília de Calliphoridae, Rhiniinae (Kutty *et al.*, 2010). Porém, duas famílias pertencentes a Oestroidea recebem destaque por conter importantes espécies causadores de miíases: Calliphoridae (varejeiras) e Oestridae (berne e bicho da cabeça, por exemplo) (Stevens *et al.*, 2006a).

1.2. Família Calliphoridae

Calliphoridae compreende aproximadamente mil espécies presentes em todas regiões biogeográficas. Os substratos para a alimentação são diversos e incluem carcaça de vertebrados, tecidos necrosados e feridas de vertebrados vivos, minhocas, caramujos (vivos ou mortos), sangue de filhotes de pássaros, javali e porco-formigueiro. Entre suas subfamílias estão Chrysomyinae, Calliphorinae e Luciliinae, conhecidas por conterem espécies necro-saprófagas e causadoras de miíases, apresentando importância forense, veterinária, médica e econômica (Marinho *et al.*, 2012; Stevens *et al.*, 2006).

Dois exemplos de espécies causadoras de miíases são *Co. hominivorax* e *Lucilia cuprina*. *Co. hominivorax*, é uma espécie que pertence a subfamília Chrysomyinae. Também é popularmente conhecida como “mosca da bicheira”, sendo uma espécie que parasita obrigatoriamente

vertebrados de sangue quente (Hall & Wall, 1995). A fêmea pode depositar até mais de 500 ovos em bordas de feridas ou orifícios corporais (Hall & Wall, 1995). Apresenta ocorrência no novo mundo, sendo atualmente encontrada em parte do Caribe e em toda América do Sul, excluindo o Chile (Torres et al. 2004). As infestações pelas larvas da mosca da bicheira em animais como o gado bovino leva a prejuízo no setor pecuário decorrente de perdas na produção e gastos com manejo. Além de outros vertebrados, a mosca da bicheira pode infestar humanos, apresentando, também, uma importância sanitária (Hall & Wall, 1995; Stevens *et al.*, 2006).

L. cuprina, também conhecida como "mosca da latrina" é parasita facultativa de diversos vertebrados, como as ovelhas (Anstead *et al.*, 2015). Suas infestações são observadas principalmente na Europa e Austrália, causando prejuízo de centenas de milhões de dólares anualmente devido a perdas na produção de lã, perda na produtividade de leite e carne e manejo das infestações. Assim como em *Co. hominivorax,* a fêmea de *L. cuprina* é atraída por odores resultantes de infecções em feridas abertas e deposita aproximadamente 200 ovos. O desenvolvimento dos estágios larvais ocorre na superfície do hospedeiro, que se adere a pele do hospedeiro por meio de ganchos. As larvas se alimentam de secreções da pele, tecidos dérmicos e sangue (Anstead *et al.*, 2015; Hall & Wall, 1995).

1.3. Família Oestridae

Oestridae é uma família com cerca de 150 espécies distribuídas em 18 gêneros (Stevens & Wallman, 2006). Nesse grupo, todas as espécies exibem parasitismo obrigatório durante o período larval, sendo que todos os hospedeiros são mamíferos (Colwell *et al.*, 2006). Os adultos das espécies de Oestridae não se alimentam e possuem o aparelho bucal reduzido. Além disso, é observado nesse clado uma alta especificidade entre parasita/hospedeiro e uma grande gama de espécies de hospedeiros, de camundongos a elefantes (Pape, 2006).

Uma das exceções quanto a especificidade ao hospedeiro é a espécie *Dermatobia hominis*, também conhecida popularmente como mosca do berne (Guimarães & Papavero, 1999). Essa espécie de mosca pode parasitar uma variedade de mamíferos, principalmente cães, gatos, gado bovino e humanos, possuindo uma importância médico-veterinária e econômica (Neiva *et al.*, 1917; Pape, 2006).

Além da generalidade quanto ao hospedeiro, essa espécie apresenta uma biologia bastante peculiar. A postura dos ovos não é realizada pela fêmea da mosca do berne diretamente nos hospedeiros. A fêmea captura um outro inseto (chamado de forético) durante o voo e deposita os ovos em seu abdome. Quando o forético contendo os ovos de *De. hominis* se aproxima de um animal de sangue quente, há um aumento na temperatura dos ovos, que faz com que os opérculos se abram e então haja a liberação das larvas. Rapidamente, a larva penetra através do folículo piloso e se instala nas camadas superficiais da derme do hospedeiro. Uma vez instalada, a larva utiliza um espiráculo que se estende até a superfície para as trocas gasosas. O crescimento da larva pode durar de 48 a mais de 70 dias dependendo do hospedeiro. No período, as larvas passam por três ínstares (L1, L2 e L3) (Neiva *et al.*, 1917; Pape, 2006).

Um outro exemplar da família que recebe destaque é o *Oestrus ovis*, ou "bicho da cabeça". Esse parasita, diferentemente da mosca do berne, é uma espécie cosmopolita e tem, em geral, hospedeiros específicos (os caprinos *Ovis aries* e *Capra hirus*). A fêmea de *O. ovis* é vivípara – ela deposita as larvas de primeiro instar diretamente na entrada da narina do hospedeiro (Silva *et al.*, 2012). Após a deposição, a larva se locomove à região interna da narina e se adere à mucosa onde se desenvolve até o segundo instar. Nessa etapa, ela se dirige ao sinus paranasal e se alimenta do muco produzido como resultado da irritação causada pela infestação e pela resposta a bactérias oportunistas (Dorchies, 1998). Assim que a larva atinge o terceiro instar, ela se dirige ao ambiente externo, cai no solo e a pupa se forma (Capelle, 1996). O intervalo de tempo do estágio de pupa

varia dependendo da temperatura e estação do ano (Tabouret *et al.*, 2001). Assim que ele emerge da pupa, ocorre o encontro do macho com a fêmea para a cópula (Capelle, 1996).

1.4. Genômica funcional

A expressão gênica é o primeiro passo para que a informação contida no DNA seja convertida em fenótipos morfológicos e fisiológicos. É um processo complexamente regulado, de diferentes formas e em diversos níveis: entre órgãos, tecidos, células e até mesmo entre componentes celulares. Além das diferenças espaciais há também diferenças temporais, nas quais, em cada fase da vida do indivíduo, um certo conjunto de genes pode ser expresso, ou expresso em níveis diferentes. Assim, variações na expressão podem resultar em fenótipos diferentes (Karp 2007).

Sendo a expressão gênica um fator importante na determinação do fenótipo, seu estudo se torna essencial para a compreensão de diferenças fenotípicas. Essa investigação pode ser realizada por meio da análise da sequência e estrutura dos transcritos, bem como de seus padrões de expressão temporais e espaciais por meio da *genômica funcional*. Parte dessa estratégia envolve o isolamento dos transcritos, cujo conjunto é denominado "transcriptoma", e a caracterização da função e regulação dos genes nos organismos (Passos & Jordan, 2000).

Uma das formas de se atingir esses objetivos é por meio do sequenciamento dos transcritos, seguido de sua anotação e verificação do nível de expressão. Neste trabalho foram utilizados dados gerados com a técnica de RNA-seq para a obtenção de transcriptomas. O RNA-seq alia as tecnologias de construção de bibliotecas de cDNA e sequenciadores de segunda geração para a obtenção dos dados. Quando comparada a outras técnicas baseadas em hibridação, uma das principais vantagens do seu uso está na independência do conhecimento prévio da sequência genômica (Wang, Gerstein & Snyder, 2009b). Os dados obtidos podem ser utilizados para determinar a tanto a sequência dos transcritos, como o seu nível de expressão gênica (Torres *et*

al., 2008, Tandonnet & Torres 2017). Quando comparamos com outras técnicas que também utilizam sequências, as vantagens são rapidez e menor custo (por transcrito), além de uma maior abrangência de transcritos sequenciados em um único experimento (Wang *et al.*, 2009).

Dados de RNA-seq também fornecem informações sobre a sequência dos transcritos. Diferentemente de mutações em sequências regulatórias, que podem alterar o padrão de expressão das proteínas, mutações em regiões codificantes de proteínas alteram a sequência de aminoácidos das proteínas e podem alterar sua função. Assim, mudanças nas sequências dos transcritos também podem ter um papel na diversificação dos hábitos alimentares em Calliphoridae. A análise comparativa de regiões codificantes permite a identificação de diferenças entre espécies que podem ser classificadas em mutações sinônimas e não sinônimas. As mutações sinônimas (cuja taxa pode ser representada por dS ou K_s) são aquelas que resultam de uma mudança no nucleotídeo que não altera a mudança do aminoácido codificado pelo novo códon. As mutações não-sinônimas (cuja taxa pode ser presentada por dN ou K_a), por outro lado, produzem um códon que codifica um aminoácido diferente. Presupõe-se que as mutações sinônimas sejam neutras, pois apesar de alterarem a sequência nucleotídica, não produzem uma proteína diferente. Já as mutações não sinônimas, alteram a proteína e podem resultar em um fenótipo diferente, tornando-se alvos da seleção natural. Em um cenário sem atuação de processos seletivos, dN seria igual a dS, resultando em uma razão de dN/dS = 1. Se há seleção natural atuando na eliminação de substituições não sinônimas, geralmente deletérias, o valor de dN observado é menor que o esperado por deriva genética, e a razão dN/dS (também chamada ω) passa a ter valores baixos. Por outro lado, quando a seleção positiva atua favorecendo a manutenção de mutações não-sinônimas, a razão dN/dS assume valores mais altos, maiores que um. Assim, os valores das razões são avaliados, considerando que (i) $\omega = 1$ indica sequências evoluindo neutramente, (ii) $\omega < 1$ para sequências com ação de seleção purificadora, e (iii) $\omega > 1$ como esperado sob seleção positiva.

2. PERGUNTAS, HIPÓTESES E DESENHO EXPERIMENTAL

Esta dissertação apresenta como questão principal as diferenças de sequência e expressão subjacentes às preferências alimentares de espécies da família Oestridae. Partimos do pressuposto de que para cada hábito alimentar, esteja envolvido um conjunto de genes específicos, mesmo em espécies de diferentes famílias. A partir desse pressuposto, além dos dados gerados para a família Oestridae, foram utilizados dados de outras espécies causadoras de miíase que pertencem a família Calliphoridae. Com isso, além de informações sobre a família Oestridae, também pudemos investigar uma questão mais geral, tentando identificar os genes envolvidos na preferência alimentar em moscas causadores de miíases.

As análises comparativas foram realizadas entre espécies da família Oestridae (*De. hominis* e *O. ovis*) e espécies da família Calliphoridae (*Co. hominivorax*, *Chysomya megacephala* e *L. cuprina*). Apesar dessas espécies causarem infestações em hospedeiros vertebrados (com exceção de *Ch. megacephala*), elas exibem diferenças quanto ao local da infestação (por exemplo, dérmico/furuncular ou nasofaringeal) e grau de parasitismo (por exemplo, parasita obrigatório), segundo a classificação proposta por Hall e Wall (1995). Além dessas classificações, propomos uma classificação de acordo com a preferência alimentar (tabela 1).

Tabela 1: Classificação das espécies estudadas, segundo o proposto por Hall e Wall (1995) e pela preferência alimentar.

Espécie	Hábito	Decrição clínica	Preferência alimentar
De. hominis	Parasita obrigatório	Dérmico/furuncular	Fluído dérmico
O. ovis	Parasita obrigatório	Nasofaringeal	Muco
Ch. megacephala	Parasita facultativa	Dérmico/ferida	Tecido necrosado
Co. hominivorax	Parasita obrigatório	Dérmico/ferida	Músculo esquelético
L. cuprina	Parasita facultativa	Dérmico/ferida	Músculo esquelético

Uma primeira abordagem para esta investigação foi a identificação dos genes diferencialmente expressos, explorando possíveis mudanças entre grupos de espécies que apresentam preferências alimentares contrastantes, assim como suas características relacionadas com o estilo de vida.

A análise de expressão foi realizada a partir da comparação de perfis de expressão global, obtidos por meio da técnica de RNA-seq, em conjunto com ferramentas de bioinformática. O primeiro passo foi a identificação de genes ortólogos compartilhados entre as espécies. Em seguida, foi realizada a identificação dos ortólogos diferencialmente expressos, entre diferentes hábitos e entre as diferentes famílias.

Para a identificação de genes diferencialmente expressos foram realizados quatro testes:

1. Espécies parasitas (*Co. hominivorax*, *De. hominis*, *L. cuprina* e *O. ovis*) *vs* necrófaga (*Ch. megacephala*), na tentativa de identificar diferenças na expressão de genes entre a espécie que se alimenta de tecido/matéria em decomposição e as espécies que se alimentam de tecidos vivos.

2. Parasitas da família Calliphoridae (*Co. hominivorax* e *L. cuprina)* vs parasitas da família Oestridae (*De. hominis* e *O. ovis*), na tentativa de identificar diferenças na expressão de genes entre as famílias.

3. Miiase dérmica (*Co. hominivorax*, *De. hominis* e *L. cuprina*) *vs* nasofaríngea (*O. ovis*), identificando possíveis mudanças na expressão de genes relacionados com o local em que ocorre a infestação no hospedeiro.

4. Miíase dérmica/furuncular (*De. hominis*) *vs* dérmica/ferida (*Ch. megacephala*, *Co. hominivorax* e *L. cuprina*), visando a investigação de possíveis mudanças relacionadas

a diferentes modos de infestação em espécies que parasitam a derme do hospedeiro. Testamos aqui, as diferenças entre as espécies que sobrevivem e se alimentam de tecido muscular esquelético, mas com formas de infestação diferentes.

Os testes foram realizados para inferir quais genes estariam potencialmente envolvidos nas diferenças de preferência alimentar e de estratégia de sobrevivência dessas espécies em seu estágio larval nos seus hospedeiros.

Uma outra abordagem utilizada foi o teste de seleção. Foram realizadas as comparações das sequências codificantes por meio da análise da razão de taxas de substituição não sinônimas por taxas de substituição sinônimas (dN/dS = ω) na tentativa de buscar evidências de seleção nos genes expressos. Nesta análise, diferentes ramos da filogenia com surgimento de hábitos e/ou preferências (ramos de primeiro plano, "foregroung") foram testados em comparação ao restante da árvore, os ramos de segundo plano ("backgroung"). Os ramos de primeiro plano são aqueles que esperamos encontrar genes sob seleção positiva que possam estar relacionados com adaptações específicas dos grupos. Em cada um dos testes foi comparado o ω calculado para o ramo de primeiro plano contra o ω médio do restante a árvore. A hipótese nula dos testes é um modelo neutro com um valor fixo de ω = 1 para toda árvore. No modelo 1, foi considerado um valor fixo de ω para toda a árvore (mas diferente de 1). Cada um dos demais modelos reflete a escolha de um ramo de primeiro plano diferente:

- Modelo 2: ramo de *Co. hominivorax*;
- Modelo 3: ramo das parasitas dérmicas (*Co. hominivorax, De. hominis* e *L. cuprina*);
- Modelo 4: ramo da espécie saprófaga, *Ch. megacephala*;
- Modelo 5, ramo de *De. hominis*;
- Modelo 6, ramo de *O. ovis*.

Os modelos de 1 até 6 foram comparados com o modelo neutro avaliando os valores de verossimilhança. Diferenças significativas observadas nos dados foram interpretadas como evidência de seleção atuando nos genes analisados.

3. MATERIAIS E MÉTODOS

3.1. Coletas e manutenção das larvas

As coletas das larvas de *De. hominis* foram realizadas na zona rural do município de Guarulhos, no Estado de São Paulo. Doze larvas de terceiro instar foram extraídas manualmente por meio da compressão do nódulo da infestação nos hospedeiros bovinos e acondicionadas em recipiente contendo serragem. Logo após o transporte ao Instituto de Biociências da Universidade de São Paulo (IB-USP), as larvas foram congeladas e armazenadas a uma temperatura de -70° C.

As larvas da espécie *O. ovis* foram gentilmente cedidas pelo professor Alessandro Francisco Talamini do Amarante do Instituto de Biologia da UNESP campus Botucatu, no Estado de São Paulo. Vinte e seis larvas foram obtidas de um mesmo hospedeiro ovino sacrificado por ocasião de outro projeto. As larvas foram armazenadas sob uma temperatura de -70º C até a extração do RNA.

Adultos da espécie *L. cuprina* foram coletados no município de Campinas, também no Estado de São Paulo e mantidos em laboratório. Larvas provenientes da primeira geração foram utilizadas para a extração de RNA para a produção dos dados da segunda réplica para a espécie.

3.2. Extração do RNA

Dois indivíduos de cada uma das espécies em terceiro instar foram utilizados para as análises. As extrações foram realizadas com TRIzol (Invitrogen), seguindo o protocolo sugerido pelo fabricante. Para as quantificações, foi utilizado o Qubit RNA BR kit (Invitrogen) e o fluorômetro Qubit 2.0, também da Invitrogen, conforme instruções do fabricante.

3.3. Sequenciamento

O preparo e o sequenciamento das bibliotecas de cDNA foram realizados no "Laboratório Multiusuários Centralizado de Genômica Funcional Aplicada à Agropecuária e Agroenergia", na ESALQ (Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo) na cidade de Piracicaba. Para a construção das bibliotecas foi utilizado o protocolo TruSeq DNA Low Sample (Illumina). Elas foram sequenciadas no equipamento HiSeq 2500 utilizando o kit HiSeq SBS v4 High Output Kit, gerando "reads" pareadas ("paired-end") de 125 pares de base (pb).

3.4. Aquisição de dados públicos

Dados de sequências expressas em larvas das espécies *Co. hominivorax*, *Ch. megacephala* e *L. cuprina* foram obtidos de repositórios públicos. Para *Co. hominivorax* foram obtidos os dados brutos do sequenciamento do transcriptoma larval em terceiro instar de duas réplicas, gerados para o trabalho realizado por Tandonnet e Torres (2017) e presente no repositório SRA sob o número de acesso SRP044071. Para *Ch. megacephala* foram obtidos dados brutos do sequenciamento do transcriptoma larval de duas réplicas, depositados no repositório SRA. Esses dados foram gerados para os trabalhos realizados por Wang e colaboradores (2011), utilizando larvas de diferentes estágios; e Zhang e colaboradores (2013) utilizando larvas em terceiro instar. Os dados foram depositados com os números de acessos SRR1663113 e SRR620248, respectivamente. Para a espécie *L. cuprina* os dados brutos do sequenciamento do transcriptoma larval de uma réplica foram obtidos do repositório SRA sob o número de acesso SRR1853100. Esses dados foram gerados para o trabalho de Anstead e colaboradores (2015). Do mesmo trabalho, também foram obtidas sequências das regiões codificantes ("coding sequences", CDS) para essa espécie.

3.5. Pré-processamento das “reads”

Antes da montagem dos transcriptomas, foi realizado um pré-processamento das sequências curtas (“reads”) brutas no programa Trimommatic (Bolger *et al.*, 2014) utilizando o parâmetro para a remoção de adaptadores (ILLUMINACLIP). Em relação a qualidade, foram utilizados os parâmetros para a remoção de bases que apresentavam uma qualidade inferior a três, tanto no início (LEADING:3) quanto no final (TRAILING:3) da “read”. Também foi realizada a varredura da “read” em janelas deslizantes com o tamanho de quatro bases. No processo de varredura, a primeira janela que apresentava um valor médio de qualidade menor que 15 (SLIDINGWINDOW:4:15), indicava o ponto inicial para o corte da extremidade final da sequência. Também foram removidas as sequências que apresentaram um tamanho menor que 20 pb (MINLEN:20).

Para a avaliação das “reads”, foi utilizado o programa FastQC (Andrews, 2010), antes e após o pré-processamento. Os parâmetros avaliados por meio do programa foram as qualidades médias das bases, a qualidade média das sequências, quantidade média de bases CG nas sequências e a presença “kmers” repetidos.

3.6. Montagem

Após o pré-processamento, “reads” idênticas e repetidas foram retiradas das análises como descrito previamente (Tandonnet & Torres, 2017). Todas as “reads” foram comparadas par-a-par e quando duplicações eram detectadas, apenas uma era mantida, aquela com melhor qualidade.

As montagens *de novo* (sem a utilização de um genoma de referência) dos transcriptomas foram realizadas por meio do programa Trinity (Grabherr *et al.*, 2001), versão 2.4.0. Trinity é um programa que realiza a reconstrução *de novo* de transcriptomas a partir de dados de RNA-seq de forma eficiente e robusta. Para a execução do programa foi utilizado como entrada os dados

combinados das réplicas de cada espécie e com apenas uma cópia de cada “read” idêntica, com o valor padrão dos parâmetros.

As montagens foram avaliadas por meio do programa BUSCO (Simão *et al.*, 2015), versão 3. BUSCO é um programa que gera uma análise do genoma ou transcriptoma montado por meio da identificação do conjunto de genes esperado de acordo com a classificação taxonômica do grupo estudado. O programa foi executado utilizando o modo transcriptoma (-m tran), um valor de 10^{-4} para corte do “e-value” (-e 0.0001), o uso da otimização Augustus (--long) e contra o banco OrthoDB v9 (Zdobnov *et al.*, 2017) com 2.799 sequências pertencentes a ordem Diptera.

3.7. Identificação de sequências ortólogas

A identificação das sequências ortólogas entre as espécies foi baseada no protocolo desenvolvido por Yang e Smith (2014) para a inferência de sequências ortólogas a partir de dados de RNA-seq. A análise é realizada a partir de sequências codificantes. Por isso, é necessário que o primeiro passo da inferência dos ortólogos seja a obtenção das CDS das espécies analisadas. Para a espécie *L. cuprina* as CDS foram diretamente obtidas de um repositório:

ftp://ftp.hgsc.bcm.edu/I5K-pilot/Sheep_blowfly/maker_annotation/version_0.5.3/

Para o restante das espécies, as sequências codificantes foram geradas a partir dos transcriptomas montados para cada espécie. Foi realizada a seleção da maior isoforma de um transcrito por meio do “script” “get_longest_isoform_seq_per_trinity_gene.pl” disponível no pacote do programa Trinity. Em seguida, foi realizado o agrupamento dos “contigs” utilizando o programa CD-HIT-est, do pacote CD-HIT (Fu *et al.*, 2012). Para a execução do programa foram utilizados os seguintes parâmetros: identidade da sequência de 99% (-c 0.99) e tamanho da semente de 10 pb (-n 10). Em seguida, as CDS foram preditas a partir dos transcritos por meio do

programa TransDecoder (https://transdecoder.github.io/), versão r2014. Para a execução do programa foi utilizado o parâmetro para o descarte de sequências de tamanho menor que 66 aminoácidos (-m 66).

Para a inferência dos ortólogos, as CDS foram alinhadas reciprocamente utilizando o algoritmo BLASTN do programa BLAST+. O banco de dados foi gerado com o programa makeblastdb, do pacote BLAST+, e foi utilizada a opção "-parse_seqids". Para execução do programa foi utilizado um valor máximo de "e-value" de 10^{-4} (-evalue 0.0001) e um valor máximo de 1000 sequências alinhadas (-max_target_seqs 1000). Ao final do alinhamento foi realizado o corte das extremidades não alinhadas das sequências com o "script" "cut_seq_ends.py", do protocolo proposto por Yang e Smith (Yang & Smith, 2014).

Em seguida, foi realizado o agrupamento das sequências entre as espécies. A primeira etapa foi passar os dados por um filtro de proporção de alinhamento ("hit fraction") utilizando o "script" "blast_to_mcl.py". Assim foram identificados pares de sequências que obtiveram menos de 40% de bases alinhadas e que foram desconsideradas nas análises posteriores. A segunda etapa foi a realização do agrupamento utilizando o algoritmo de Markov (MCL algorithm) com o programa MCL (Enright *et al.*, 2002). O programa foi executado utilizando um valor de inflação de 1,4 (-I 1.4). O algoritmo de Markov calcula o agrupamento entre nós utilizando uma matriz de similaridade, no nosso caso uma matriz de similaridade entre sequências de aminoácidos.

O próximo passo foi a inferência das árvores a partir das sequências agrupadas. Primeiramente foi utilizado o "script" "write_fasta_files_from_mcl.py" que recupera as sequências dos transcritos que foram agrupados após a execução do MCL. Em seguida, foi utilizado o "script" "fasta_to_tree.py" que alinha cada grupo de sequências, corta os alinhamentos e infere as árvores. Os alinhamentos foram realizados com o programa MAFFT (Katoh & Standley, 2013),

utilizando os parâmetros “--genafpair”, que contabiliza todos os pares de sequência com um algoritmo local; e “--maxiterate 1000”, que corresponde ao número de ciclos de iteração realizados. Os alinhamentos foram cortados com o programa Phyutility (Smith & Dunn, 2008) utilizando o parâmetro “-clean <0,1>”. Por fim, as inferências das árvores foram realizadas com o programa RAxML que utiliza o método de máxima verossimilhança (Stamatakis, 2014). Ao final dessa etapa temos como resultado a inferência dos homólogos. Ambos os “scripts” fazem parte do pacote de Yang e Smith.

Para a inferência dos ortólogos foi utilizado ao método “one to one”, que considera para as análises sequências homólogas que estejam presentes em cópia única em cada táxon amostrado. Por fim foi utilizado um “script” que separa as sequências ortólogas por espécie, selecionando os agrupamentos que apresentaram apenas uma sequência por espécie, ou seja, excluindo possíveis parálogos.

3.8. Anotação das sequências ortólogas

A anotação das sequências transformadas em CDS foi realizada com a ferramenta *GOanna* (http://www.agbase.msstate.edu/cgi-bin/tools/GOanna.cgi), do repositório *AgBase* (Mccarthy *et al.*, 2006). AgBase é uma fonte de recursos, “online” e curada, para análise funcional de produtos gênicos de animais e plantas agricultáveis. Os alinhamentos foram realizados com o programa blastx, que traduz as CDS do banco de dados de interesse nos seis quadros de leitura e alinha cada quadro contra um banco de dados de proteínas retornando as sequências alinhadas. O banco de dados utilizado apresenta somente sequências com termos GO experimentalmente validados. Essas sequências são provenientes banco de dados universal de proteínas – UniProt, e do banco de dados curado pela comunidade AgBase, que conta com genomas de espécies de importância agronômica e patógenos, incluindo espécies de animais, plantas, microrganismos e parasitas de

importância econômica (“Database: AgBase-UniProt”). Para o restante dos parâmetros foram utilizados com os valores recomendados na plataforma.

3.9. Análise de expressão gênica

Para a identificação de transcritos diferencialmente expressos (DE) foram inicialmente realizados os mapeamentos das “reads” contra os transcritos ortólogos. Os mapeamentos foram realizados com o software Bowtie2, versão 2.2.6 (Langmead *et al.*, 2009). Bowtie2 é uma ferramenta para alinhamento de “reads” contra uma referência de sequências longas de uma forma eficiente e ultra-rápida. Diferentemente do Bowtie1, Bowtie2 apresenta algumas características que acompanham mudanças dos sequenciamentos de nova geração (maior volume e tamanho de sequências), Assim, Bowtie2 e apresenta um melhor desempenho no alinhamento de “reads” com mais de 50 caracteres.

Um primeiro mapeamento foi realizado permitindo “gaps” nos alinhamentos. Foram utilizados os parâmetros para alinhamento local (“--local”), no modo sensível (“--very-sensitive-local”), permitindo alinhamentos com fragmentos de, no máximo, 1000 pb (“--maxins 1000”), não permitindo o alinhamento de sequências não-pareadas (“--no-mixed”) e discordantes para o par de sequências (“--no-discordant”).

Um segundo mapeamento foi realizado, sendo utilizados parâmetros para não permitir “gaps” entre o começo ou o final da “read” (“--dpad 0” e “--gbar 99999999”), permitir apenas um “mismatch” (“--mp 1,1”) e um caractere ambíguo (“--np 1”), um “score” mínimo de 0,1 para validar o alinhamento (“--score-min L,0,-0.1”), e também não permitindo o alinhamento de sequências não-pareadas e discordantes para o par de sequências.

Após os mapeamentos, foi realizada a contagem do número de “reads” mapeadas a cada “contig” por meio de dois programas: o eXpress (Roberts *et al.*, 2011) e o RSEM (Li & Dewey,

2011). Para a execução do programa eXpress foram utilizados os dados do mapeamento em que foi permitida a presença de “gaps” e com os valores padrão para os parâmetros. A contagem foi realizada para cada réplica separadamente. Ao final, foi montada uma única matriz contendo todas as contagens de cada réplica.

Para a execução do programa RSEM, primeiramente foi realizada a construção das referências (index) a partir das sequências ortólogas para cada espécie por meio do “script” “rsem-prepare-reference”. Como entrada, foram utilizados os dados do mapeamento em que não foi permitida a presença de “gaps”, sendo cada arquivo validado por meio do “script” “rsem-sam-validator”. Em seguida foi realizada a contagem com o uso do “script” “rsem-calculate-expression”. Todos os “scripts” utilizados nesta etapa fazem parte do pacote RSEM.

Ao final de todas as contagens, foram realizadas as análises estatísticas com os pacotes edgeR (Robinson *et al.*, 2010) e DESeq2 (Love *et al.*, 2014) desenvolvidos para o ambiente R. DESeq2 utiliza, por padrão, uma distribuição binomial negativa ajustada a um modelo linear generalizado (GLM) e estatísticas de Wald para o teste de significância. Já o edgeR utiliza, no método clássico, uma distribuição binomial negativa com a máxima verossimilhança ajustada ao quantil (qCML), que é um método para estimar a dispersão dos dados. O teste exato implementado pelo edgeR é um análogo ao teste exato de Fisher.

Foram utilizadas quatro combinações para a avaliação do nível de expressão:

1. Contagem com o software eXpress, combinado a análise estatística com o edgeR;
2. Contagem com software eXpress, combinado a análise estatística com o DESeq2;
3. Contagem com o programa RSEM, combinado a análise estatística com o edgeR;
4. Contagem com o programa RSEM, combinado a análise estatística com o DESeq2.

Para identificarmos os genes que foram diferencialmente expressos comumente entre os grupos foram construídos diagramas de Venn (Heberle *et al.*, 2015) com os resultados de cada método em cada teste.

3.10. Análise das sequências codificantes

Além das mudanças na expressão de genes, foram estudadas as mudanças nas regiões codificantes das sequências ortólogas, na tentativa de identificar possíveis mudanças associadas às diferentes preferências alimentares. Para isso, estimamos a razão entre as taxas de mutações não-sinônimas e sinônimas (ω). Para essa etapa, foram utilizadas as CDS e sequências de aminoácidos preditas pelo TransDecoder, como descrito anteriormente. O primeiro passo foi o alinhamento das sequências nucleotídicas com o programa muscle (Edgar 2004). Em seguida, convertemos o alinhamento de aminoácidos em um alinhamento de nucleotídeos ("back translation") com o programa translatorX (Abascal *et al.*, 2010). Este programa também utiliza o algoritmo do Gblocks (Talavera *et al.*, 2007) para eliminar regiões mal alinhadas e muito divergentes do alinhamento. Os alinhamentos corrigidos foram utilizados para estimar os valores de dN, dS e ω entre as sequências dos ortólogos com o programa codeml do pacote PAML v.4 (Yang 2007). O codeml é um programa que compara e realiza testes em sequências de nucleotídeos ou aminoácidos baseados no método de máxima verossimilhança ("maximum likelihood").

Com o programa codeml foram realizados os testes de seis modelos que foram comparados à hipótese nula de um modelo neutro com um valor fixo de $\omega = 1$ para toda árvore:

- Modelo 1: considerado um valor fixo de ω para toda a árvore (mas diferente de 1).
- Modelo 2: ramo de *Co. hominivorax* como primeiro plano;
- Modelo 3: ramo das parasitas dérmicas (*Co. hominivorax, D. hominis* e *L. cuprina*);
- Modelo 4: ramo da espécie saprófaga, *Ch. megacephala*;

- Modelo 5, ramo de *D. hominis*;
- Modelo 6, ramo de *O. ovis*.

As estimativas de verossimilhança do modelo nulo foram comparadas às estimativas observadas em cada modelo testado (modelo 1 até modelo 6) por meio do teste de χ^2, com significância (α) de 0,05. O "False Discovery Rate - FDR" foi utilizado para correção de testes múltiplos. Para os testes significativos pressupomos que: (i) $\omega = 1$ indica sequências sob evolução neutra, (ii) $\omega < 1$ para sequências evoluindo sob seleção purificadora e (iii) $\omega > 1$ como esperado sob seleção positiva.

4. RESULTADOS

4.1 Coleta, extração e sequenciamento do RNA das larvas

Das larvas coletadas, para as espécies *De. hominis* e *O. ovis*, cada réplica foi realizada utilizando um indivíduo para a realização da extração do RNA total. Para *L. cuprina*, foram utilizados dois indivíduos para a extração do RNA e construção da segunda réplica.

O sequenciamento dos transcriptomas gerou mais de 76 milhões "reads" para a espécie *De. hominis*, mais de 64 milhões para espécie *O. ovis*, quando somadas as duas réplicas (Tabela 2) e mais de 47 milhões para *L. cuprina*. Dos dados recuperados de bancos públicos, foram obtidas para a espécie *Co. hominivorax* quase 40 milhões de "reads", mais de 54 milhões para *Ch. megacephala* e quase 48 milhões para *L. cuprina* (Tabela 2).

4.2 Pré-processamento

O pré-processamento das "reads" foi realizado para as duas réplicas de todas espécies utilizadas neste trabalho (incluindo os dados públicos) com o programa Trimmomatic. Em média, foram descartadas 3% das "reads" após o pré-processamento, sendo a primeira réplica de *Co. hominivorax* o conjunto que mais perdeu dados (15% das sequências). Já a primeira réplica de *Ch. megacephala,* foi a amostra com o menor descarte de "reads", sendo menos de 1% (Tabela 2).

Tabela 2. Sequenciamento e pré-processamento. São apresentadas as quantidades de "reads" obtidas por meio da técnica de RNA-Seq, incluindo dados brutos e pré-processados.

Espécie	Amostra	Brutos	Pré-processados
Co. hominivorax	1	13.028.851	11.055.926 (84,86%)
	2	26.807.116	26.277.392 (98,02%)
Ch. megacephala	1	19.549.331	19.386.665 (99,17%)
	2	35.112.471	34.108.071 (97,14%)
De. hominis	1	46.794.071	45.397.596 (97,02%)
	2	30.006.392	29.273.934 (97,56%)
L. cuprina	1	28.386.962	27.827.093 (98,03%)
	2	19.439.280	18.987.105 (97,67%)
O. ovis	1	33.799.262	32.827.955 (97,13%)
	2	30.643.772	29.882.997 (97,52%)

4.3 Montagem

As sequências provenientes das réplicas de cada espécie foram combinadas e processadas utilizando um "script" que reduziu para apenas uma o número de cópias idênticas das "reads". Em seguida, foi realizada a montagem dos transcriptomas por meio da execução do programa Trinity. A montagem de *De. hominis* foi a que apresentou a maior quantidade de "contigs" montados, com quase 167 mil, com uma média de 1041 pb. Já para a montagem da espécie *Co. hominivorax*, menos de 40 mil "contigs" foram montados, com uma média de tamanho de 965,2 pb (Tabela 3).

Tabela 3. Resultado das montagens *de novo* dos transcriptomas das espécies. São apresentados a quantidade e o tamanho médio dos "contigs" montados (em pares de bases).

Espécie	"contigs"	Tamanho médio (pb)
Co. hominivorax	38.749	965,22
Ch. megacephala	68.516	973,68
De. hominis	166.823	1041,98
O. ovis	103.386	1097,88

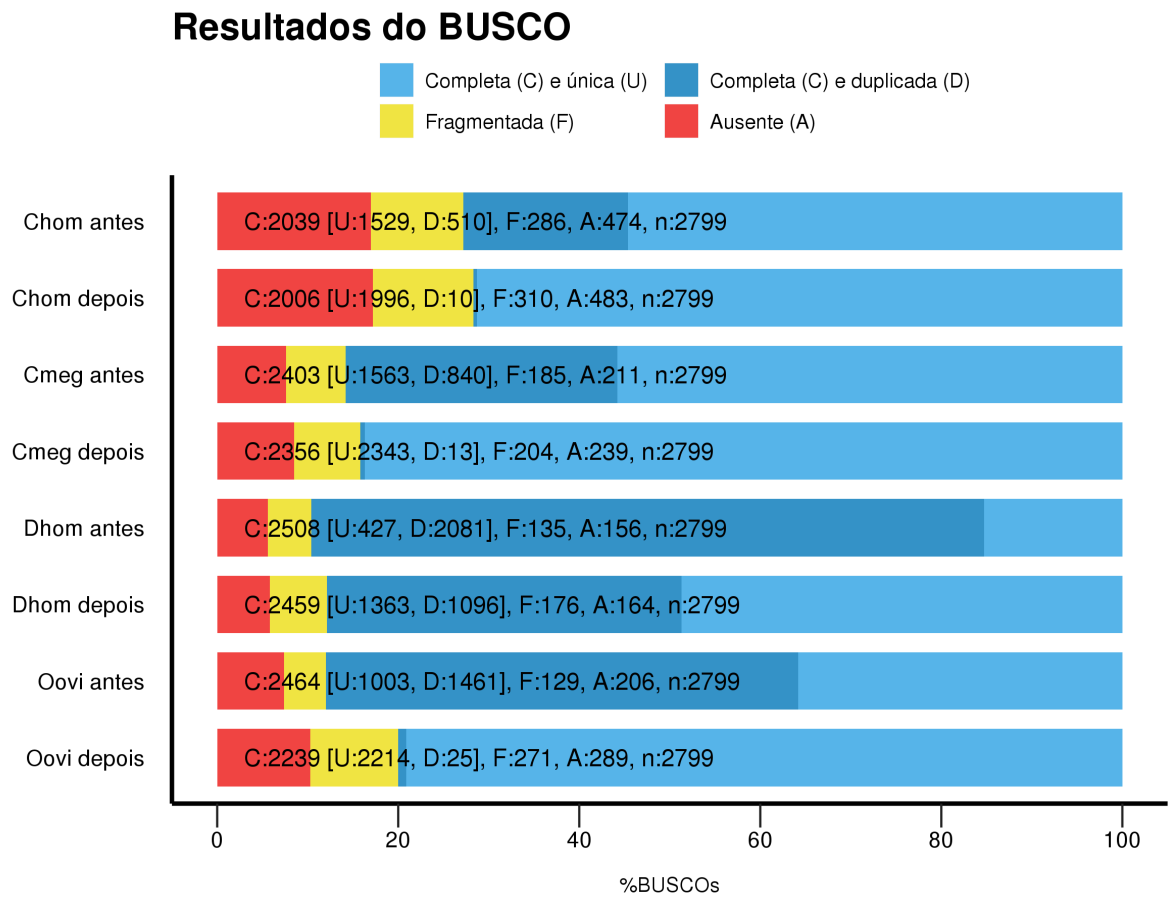

Figura 2: Avaliação das montagens com o programa BUSCO. O programa foi executado antes e depois da retirada das isoformas dos genes para cada espécie. Dentro das barras estão os valores e proporções de ortólogos alinhados em cada categoria. n: quantidade total de ortólogos presentes no banco utilizado pelo programa. Chom: *Co. hominivorax*; Cmeg: *Ch. megacephala*; Dhom: *De. hominis* e Oovi: *O. ovis*.

4.4 Identificação dos transcritos ortólogos

Após as montagens dos transcriptomas de *Co. hominivorax*, *Ch. megacephala*, *De. hominis* e *O. ovis*, para cada um dos possíveis genes com mais de uma isoforma identificada, foi mantida apenas uma sequência, correspondente a isoforma mais longa. Isso implica que, nas análises subsequentes, somente uma isoforma por gene foi analisada. Esse conjunto de sequências foi utilizado para encontrar regiões codificantes para cada gene (CDS). Mais uma vez, apenas uma CDS foi utilizada por gene. Para a espécie *L. cuprina*, que possui genoma completamente

sequenciado e anotado, esse processo não foi realizado, pois as CDS conhecidas puderam ser obtidas diretamente do banco de dados. Com as 15.247 CDS de *Co. hominivorax*, 18.466 de *Ch. megacephala*, 33.040 de *De. hominis*, 18.396 de *L. cuprina* e 15.906 de *O. ovis*, foram realizados os passos propostos no protocolo desenvolvido por Yang e Smith (2014), sendo identificadas 1.275 sequências ortólogas compartilhadas entre as cinco espécies.

4.5 Análise quantitativa da expressão gênica

Para obtenção dos perfis de expressão gênica, foi realizado o mapeamento das "reads" contra as sequências dos ortólogos, seguido da contagem das "reads" mapeadas. Dos mapeamentos realizados com o Bowtie2, os alinhamentos realizados com os parâmetros que permitiam "gaps" resultaram em uma proporção maior de "reads" alinhadas às sequências dos "contigs" ortólogos, sendo a maior proporção de alinhamento obtida para as réplicas de *Ch. megacephala*, com mais de 9,8% de "reads" alinhadas, em contraste com 2,5% de "reads" alinhadas sem permitir "gaps". A segunda réplica da espécie *De. hominis* foi a que obteve uma menor proporção de "reads" alinhadas, sendo 3,5% para os alinhamentos permitindo "gaps" e menos de 2% para os alinhamentos que não permitiam "gaps" (Tabela 4).

O número de "reads" alinhadas contra cada "contig" foi utilizado para as comparações interespecíficas. A contagem desse número, foi realizada com dois algoritmos diferentes. Assim, para cada amostra havia dois perfis de expressão distintos (duas matrizes; cada matriz gerada após a execução de um programa). Os diferentes modos de classificação (Tabela 1) foram testadas por meio do agrupamento de espécies com preferências similares e sua comparação entre dois grupos de preferências contrastantes.

Os testes foram realizados por meio de dois pacotes diferentes. Assim, cada teste de hipótese apresenta quatro resultados (combinação dos dois métodos de contagem e dos dois pacotes estatísticos):

I. eXpress + edgeR

II. eXpress + DESeq2

III. RSEM + edgeR

IV. RSEM + DESeq2

Tabela 4: Quantidade de "reads" alinhadas aos ortólogos. As "reads" obtidas para cada réplica de cada espécie foi alinhada ao seu respectivo conjunto de ortólogos. Foram utilizados dois modos diferentes de alinhamento, um modo permissivo ("Gapped") e no outro não ("Ungapped").

Espécie	Réplica	"Gapped"	"Ungapped"
Ch. megacephala	1	1.909.160 (9,85%)	1.599.688 (8,25%)
	2	3.371.587 (9,89%)	2.818.030 (8,26%)
Co. hominivorax	1	571.301 (5,17%)	516.382 (4,67%)
	2	1.613.652 (6,14%)	1.357.666 (5,17%)
De. hominis	1	2.841.784 (6,26%)	1.875.118 (4,13%)
	2	1.023.656 (3,50%)	572.371 (1,96%)
O. ovis	1	1.496.188 (4,56%)	1.268.021 (3,86%)
	2	1.801.913 (6,03%)	1.502.078 (5,03%)
L. cuprina	1	1.916.293 (6,89%)	1.598.115 (5,74%)
	2	1.375.278 (7,24%)	1.156.438 (6,09%)

A intersecção dos resultados das quatro combinações (figura 2) foi utilizada como o conjunto de genes diferencialmente expressos para interpretações posteriores. Esse conjunto de genes identificados em todas as abordagens é menos dependente de diferenças intrínsecas dos programas e nos parece um bom ponto de partida para a identificação de genes candidatos,

apesar de incorrermos no risco de um maior número de falsos negativos. Foram realizados os seguintes testes:

- Teste 1. Espécies parasitas (*Co. hominivorax*, *De. hominis*, *L. cuprina* e *O. ovis*) vs necrófaga (*Ch. megacephala*)
- Teste 2. Espécies da família Calliphoridae (*Co. hominivorax* e *L. cuprina*) vs espécies da família Oestridae (*De. hominis* e *O. ovis*)
- Teste 3. Miíase dérmica (*Co. hominivorax*, *De. hominis* e *L. cuprina*) vs nasofaríngea (*O. ovis*)
- Teste 4. Miíase furuncular (*De. hominis*) vs traumática (*Ch. megacephala*, *Co. hominivorax* e *L. cuprina*)

No teste para identificar diferenças entre a espécie necrófaga e as espécies parasitas (teste 1) e no teste realizado entre os grupos com espécies da família Calliphoridae e da família Oestridae (teste 2), não foram identificados genes diferencialmente expressos em nenhuma combinação de métodos utilizados.

No teste 3, realizado entre as espécies que parasitam a região dérmica contra a que parasita região nasofaríngea do hospedeiro, obtivemos onze genes diferencialmente expressos (figura 3A), sendo nove super-expressos nas espécies que parasitam a região dérmica e dois genes sub-expressos (Tabela 5). Quase metade dos genes DE encontrados são proteínas ribossomais, sendo que a proteína ribossomal L8 60S apresenta a maior diferença, com uma média de $\log_2 FC$ = 12.42. Analisando a função dos genes DE é possível identificar potenciais candidatos envolvidos com o parasitismo dérmico, como trypsin II-P29 e ferritina (Tabela 5).

No teste 4, espécie que causa miíase furuncular (*De. hominis*) foi comparada àquelas que formam feridas (*Co. hominivorax* e *L. cuprina*). Foram identificados dois genes diferencialmente expressos comuns aos quatro métodos (figura 3B). Um deles é regulado em hipóxia ou ausência de oxigênio ("protein anoxia like up-regulated"), o outro é uma proteína ribosomal (Tabela 5).

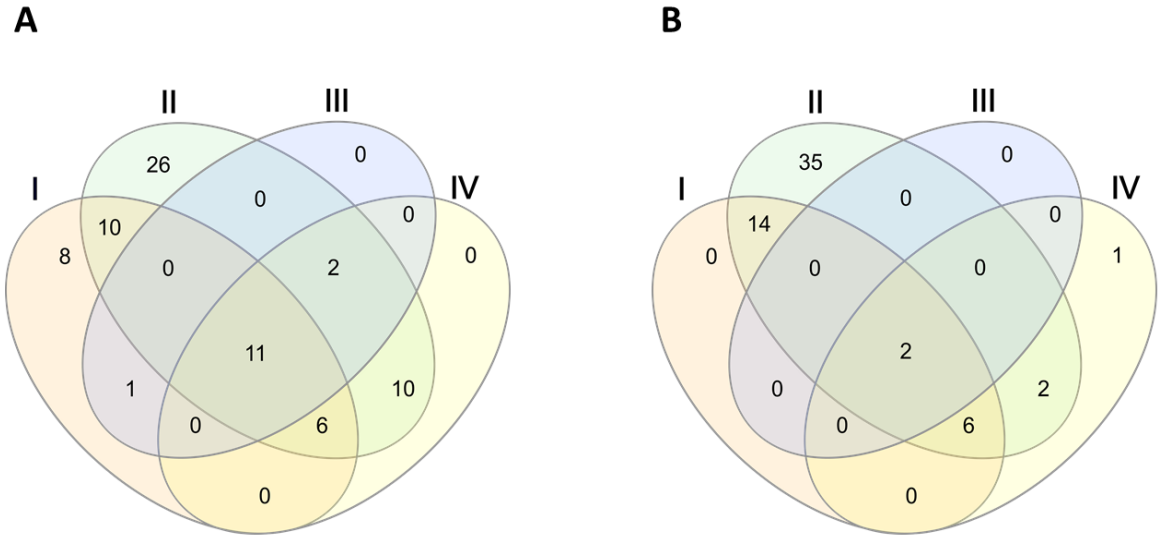

Figura 3. Diagramas de Venn ilustrando o compartilhamento de testes significativos pelos diferentes métodos. (A) Teste 3, Teste comparando os grupos de espécies que parasitavam a região dérmica contra a espécie que parasita a região nasofaríngea do hospedeiro (*Co. hominivorax*, *De. hominis* e *L. cuprina vs O. ovis*). (B) Teste 4, comparando espécies que parasitam a região dérmica de modos diferentes, espécie que causa miíase furuncular (*De. hominis*) contra o grupo de espécies que formam traumas ao hospedeiro (*Co. hominivorax* e *L. cuprina*). Os testes foram realizados combinando diferentes programas para a contagem (eXpress e RSEM) e para os testes estatísticos (edgeR e DESeq2), sendo observado genes diferencialmente expressos em comum a todos os métodos.

Tabela 5. Genes diferencialmente expressos. São apresentadas as identidades dos ortólogos (ID), os valores do log de "Fold Change" encontrados em cada método de análise de expressão diferencial e a anotação do respectivo ortólogo e a % de identidade com o banco ("% id) para cada teste.

	ID	FC I[1]	FC II[2]	FC III[3]	FC IV[4]	Anotação	%id
Teste 3	cluster4087	-5.24	-4.89	-4.93	-4.84	mimitin	77
	cluster3666	-4.45	-4.22	-4.15	-4.17	ribonucleoprotein RB97D	84
	cluster4049	10.82	11.89	11.17	12.01	60S ribosomal protein L27a	88
	cluster3925	11.01	11.49	11.40	11.65	trypsin II-P29	72
	cluster3845	11.42	13.22	11.79	13.25	60S ribosomal protein L8	96
	cluster4085	3.76	11.98	4.09	12.11	60S ribosomal protein L14	92
	cluster3919	5.12	13.34	5.43	13.52	40S ribosomal protein S7	95
	cluster3753	6.01	12.52	6.34	12.63	60S ribosomal protein L7a	90
	cluster4080	6.27	9.09	6.61	9.17	cytochrome c oxidase subunit 6A	70
	cluster3544	6.68	6.84	7.00	6.85	ADP, ATP carrier protein	93
	cluster3510	7.68	11.88	8.03	12.04	Ferritin	71
Teste 4	cluster3272	10.21	10.80	10.09	10.92	anoxia up-regulated-like	71
	cluster4034	5.73	11.53	5.76	11.68	40S ribosomal protein S23	99

[1]FC I: $\log_2$(FoldChange) edgeR+eXpress

[2]FC II: $\log_2$(Fold change) edgeR+RSEM

[3]FC III: $\log_2$(Folde change) DESeq2+eXpress

[4]FC IV: $\log_2$(Fold change) DESeq2+RSEM

4.6 Evidência de seleção nas sequências codificantes

Após a análise dos perfis de expressão dos genes, buscamos por evidências de seleção positiva utilizando o codeml testando seis diferentes modelos. Os modelos foram comparados contra o teste nulo de que todos os genes estariam evoluindo neutramente.

Nos dados foram encontrados ω com valores de 999 (valor infinito utilizado pelo programa). Esses valores de ω indicam que há uma taxa baixa de substituições não-sinônimas. Como descrito em Material e Métodos estes genes foram descartados.

Em geral, há uma grande restrição evolutiva nos genes em todos os modelos testados, estando ao menos 80,5% deles evoluindo sob um regime de seleção purificadora. Entretanto, alguns genes podem estar evoluindo sob seleção positiva. Ao ser analisado o ramo de *Co. hominivorax*, foram identificados nove genes sob seleção positiva (Tabela 6). Seguindo os resultados encontrados na análise de expressão, entre estes genes foram encontrados Spook (spo), envolvido em processos de oxido-redução e ligação de ferro e SH3PX1, envolvido na resposta a hipóxia. Para o ramo de *De. hominis*, foram identificados quatro genes com evidência de seleção positiva, um em comum com *Co. hominivorax*: o gene "starvin" (stv), com uma função de proteína "heat shock".

Tabela 6. Anotação dos genes com evidência de seleção positiva.

Modelo	ID	ω	dN	dS	p-value	Função (termo GO)
Chom[1]	cluster3179	2.36	0.01	0.03	1.62E-05	GPI inositol-deacylase
	cluster3423	1.59	0.02	0.03	1.40E-08	oxidation-reduction process/ iron ion binding
	cluster3499	1.49	0.03	0.04	2.53E-02	glycine-rich cell wall structural protein
	cluster3971	1.48	0.02	0.02	3.32E-05	28S ribosomal protein S11
	cluster2973	1.44	0.03	0.04	5.10E-06	heat shock protein binding
	cluster4045	1.35	0.03	0.04	5.09E-03	something about silencing protein 10
	cluster3505	1.26	0.04	0.05	1.80E-05	cellular response to hypoxia
	cluster3687	1.14	0.02	0.02	3.24E-02	regulation of transcription
	cluster4069	1.12	0.02	0.03	4.51E-02	zinc ion binding
Dhom[2]	cluster2973	2.19	0.03	0.01	7.47E-06	heat shock protein binding
	cluster3838	1.90	0.04	0.02	1.23E-07	negative regulation of Ras protein signal transduction
	cluster3685	1.60	0.03	0.02	8.55E-04	insulin-like growth factor binding
	cluster4083	1.35	0.06	0.05	8.12E-04	No GO associated

[1]Resultados das estimativas de ω para o ramo de *Co. hominivorax*

[2]Resultados das estimativas de ω para o ramo de *De. hominis*

5 DISCUSSÃO

O intuito do trabalho foi identificar genes que possam estar relacionados à preferência alimentar em espécies de dípteros pertencentes a família Oestridae e também em outras espécies causadoras de miíases de um modo geral. Outro objetivo foi gerar mais informações moleculares para essa família que apresentam uma biologia interessante e uma importância econômica. Foram estudadas cinco espécies cujas larvas infestam hospedeiros vertebrados, sendo três pertencentes a família Calliphoridae: *Ch. megacephala*, *Co. hominivorax* e *L. cuprina*; e duas pertencentes a família Oestridae: *De. hominis* e *O. ovis*. Para identificarmos tais genes candidatos foram utilizados dados do sequenciamento global do RNA mensageiro das larvas das espécies citadas.

Como não é claro quais fatores são determinantes para as diferentes preferências alimentares, os perfis de expressão gênica foram comparados entre grupos com preferências contrastantes e com diferentes interações com seus hospedeiros em quatro testes distintos.

Nos testes em que foram identificados genes DE, encontramos um grande número de genes ribossomais. No teste 3, entre parasitas dérmicas contra a nasofaríngea, seis genes ribossomais foram identificados com expressão diferencial entre os grupos. Proteínas ribossomais desempenham uma importante função em todos os processos biológicos, traduzindo mRNA em proteínas. Evidências recentes têm mostrado um possível papel regulatório de proteínas ribossomais, além de seu papel constitutivo na tradução. Qualquer defeito nos componentes ribossômicos ou qualquer perda de proteínas ribossômicas centrais pode alterar os padrões de expressão gênica (Thomas, 2000; Kondrashov et al, 2011), o que pode resultar em alterações específicas do tecido. Além disso, a superexpressão ou silenciamento de várias proteínas ribossomais leva a alterações na transcrição gênica, aumentando ou diminuindo a taxa de transcrição de genes em *Drosophila melanogaster* e *Schistosoma japonicum*, por exemplo (Ni et al,

2006; Sun et al, 2015). Assim, além da síntese protéica, padrões de expressão dinâmicos e heterogêneos de proteínas ribossomais provavelmente desempenham papéis importantes na regulação da tradução de proteínas. Cada uma das proteínas ribossomais identificadas pode ter uma ação regulatória sobre transcritos específicos e grupos de transcritos em espécies de Oestroidea, questão que pode ser melhor investigada no futuro.

Em relação aos demais genes diferencialmente expressos, identificamos genes no grupo das espécies que parasitam região dérmica contra a espécie que parasita a região nasofaríngea (teste 3) que receberam destaque. Entre eles estão o transportador ATP/ADP ou "Adenine nucleotide translocator" (ANT em *Dr. melanogaster*), a subunidade 6A da citocromo oxidase ou Levy (*Dr. melanogaster*) e Ferritina. Esses genes apresentavam-se superexpressos no grupo que parasita a região dérmica em relação a espécie que parasita a nasofaringe.

Os produtos gênicos do carreador de ATP/ADP e Levy estão associados à produção energética nas células. O carreador de ATP/ADP (ANT) participa da regulação do processo de biossíntese de ATP em *Dr. melanogaster* (Liu *et al.*, 2007; Vartiainen *et al.*, 2014). No trabalho de Vartiainen e colaboradores (2014) com mutantes de *Dr. melanogaster* para o gene ANT ("stress- sensitive B^1") exibiam uma redução na produção de ATP, acúmulo de lactato e mudanças no padrão de expressão de genes que são consistentes com a troca metabólica para glicólise.

Já o gene Levy codifica uma subunidade do complexo citocromo c oxidase. Mutações nesse gene em indivíduos de *Dr. melanogaster* causam uma diminuição na taxa de montagem desse complexo. Essas diminuições resultam em mudanças fisiológicas e fenotípicas, incluindo diminuição da expectativa de vida do adulto e aumento da taxa de neurodegeneração, devido a diminuição na produção dos níveis de ATP e aumento de radicais livres (Liu *et al.*, 2007).

Ambos os genes estão relacionados com a regulação do metabolismo celular, realizado na membrana mitocondrial. Sabe-se que diferentes dietas podem alterar a atividade mitocondrial (Holmbeck & Rand, 2015). Em um estudo com indivíduos de *Dr. melanogaster* criados em dietas suplementadas com diferentes ácidos graxos (saturados e insaturados), foi observado que os indivíduos apresentaram diferenças no metabolismo celular entre os diferentes tratamentos. Larvas que se desenvolveram com uma dieta suplementada com ácido graxo poli-insaturado apresentaram uma maior taxa de respiração celular do que aqueles que se desenvolveram com uma dieta suplementada com ácido graxo saturado. Essa maior atividade da cadeia respiratória leva a um aumento do estresse oxidativo, diminuindo a expressão dos carreadores de ADP/ATP. Comparando a expressão desses genes entre as espécies que apresentam diferentes preferências alimentares, a espécie que se aliment de muco, apresentam uma expressão menor destes genes, em contraste com as que se alimentam de tecido muscular, que pode estar relacionado com as diferentes dietas.

O produto do gene ferritina é amplamente encontrado entre bactérias, plantas e animais; e além de estar associado com a absorção de ferro também está envolvido com a homeostase celular de íons ferro. O silenciamento do gene em *Dr. melanogaster* resulta na diminuição da absorção e acúmulo do ferro pelas células intestinais, sendo também observada deposição de ferro em células oculares e neuronais. Essa deposição leva a alterações no desenvolvimento dos tecidos de olhos e cérebro, causando má formação dessas estruturas (Tang & Zhou 2013a), demonstrando os efeitos tóxicos do excesso desse elemento. Para a manutenção do equilíbrio de ferro celular, quando há uma alta concentração desse elemento no substrato, tem-se como resultado um aumento da expressão do gene e proteína ferritina (Dunkov *et al.*, 2002; Georgieva *et al.*, 1999; Tang & Zhou, 2013a). Observando indivíduos de *Dr. melanogaster* em meios com e sem suplementação de ferro, foi detectado um aumento na expressão de mRNA de ferritina em

larvas, pupas e imagos que se desenvolveram em um ambiente suplementado com ferro (Georgieva *et al.*, 1999). Isso leva a interpretação de que um aumento da expressão desse gene em espécies que se desenvolvem na derme em comparação com a espécie que se desenvolve na nasofaringe do hospedeiro, se deve a mais alta concentração do íon em tecido esquelético e ao balanço de ferro no organismo, evitando seus efeitos tóxicos.

Nos testes que envolvem apenas as espécies que parasitam a derme, foi observado que o gene "protein anoxia up-regulated" (ou fau) exibe uma expressão diferencial significativa, sendo mais expresso nas espécies que causam traumas. O gene fau está relacionado com a resposta ao estresse por falta de oxigênio (Kariithi *et al.*, 2016; Vontas *et al.*, 2005; Wu 2017). Num estudo realizado por Ma e colaboradores (1999), adultos de *Dr. melanogaster* foram expostas a uma concentração menor que 0,02% de O_2 (anoxia). A expressão gênica desses indivíduos foi avaliada, sendo observado um aumento na expressão deste gene. No nosso trabalho, observamos uma expressão maior deste gene nas espécies que causam miíase traumática quando comparadas à espécie que forma furúnculo. Nas espécies traumáticas há um grande número de larvas infectando o hospedeiro, formando uma massa larval. Foi reportado que espécies saprófagas agregados de larvas formam a massa larval durante a alimentação, nos quais os indivíduos passam por períodos de hipóxia (Lein 2013). Sendo assim, espécies que formam traumas estão mais suscetíveis à baixas concentrações de O_2.

Apesar de termos analisado um número limitado de genes, foi observado que muitos exibiam um perfil de expressão conservado (sem diferenças significativas) entre os cenários testados. Assumimos que esse padrão seja devido a restrição nas regiões regulatórias. Além disso, tivemos um resultado semelhante para as análises das sequências codificantes. A maioria das sequências testadas estão evoluindo sob seleção negativa. Detectamos evidências de seleção positiva em apenas nove e quatro genes no ramo de no ramo de *Ch. hominivorax* e *De. hominis*,

respectivamente. Dentre esses, foi encontrado um gene comum entre *Co. hominivorax* e *De. hominis*: o gene “starvin” (stv).

O gene “starvin” é essencial para o desenvolvimento larval de *Dr. melanogaster*. A proteína é expressa principalmente no desenvolvimento de músculos somáticos larvais, células do tendão epidérmico (que atuam na fixação do músculo) e, em menor intensidade, no esôfago (Coulson et al., 2005). Mutantes do gene não presentam alterações nos tecidos larvais e têm uma aparência indistinguível de uma larva normal. No entanto, têm a viabilidade severamente reduzida pois as larvas recém-nascidas perdem a capacidade de ingerir alimentos e crescer. Esta observação foi feita por Coulson e colaboradores (2005) que testaram a ingestão de alimentos por larvas mutantes que foram colocadas em meio contendo corante azul de bromofenol. O corante, não digerido por larvas, é observado no intestino de larvas selvagens, mas não de larvas mutantes. Assim, os efeitos da mutação, ainda desconhecidos, não são morfológicos, mas provavelmente fisiológicos, com papel na ingestão do alimento. É interessante que o gene apresente evidência de seleção positiva nas duas espécies parasitas que se alimentam de tecido muscular vivo do hospedeiro. Uma hipótese que pode ser levantada é que substituições não sinônimas podem alterar o funcionamento da proteína, mas ainda retendo seu papel essencial na ingestão de alimentos. Será interessante investigar a relação dessas substituições e a transição para um hábito parasita.

Outros genes observados no ramo de *Co. hominivorax* apresentam funções semelhantes às observadas nos genes DE. O gene Spook (spo) está envolvido em processos de oxirredução e ligação de ferro e o gene SH3PX1 está envolvido na resposta a hipóxia.

O gene “spook” faz parte da família Halloween de genes de desenvolvimento em *Dr. melanogaster*. É uma enzima da família das citocromo P450 que catalisa os passos de hidroxilação no processo de conversão do colesterol para o hormônio da muda, a 20-hidroxiecdisona (Ono et

al., 2006). É razoável supor, portanto, que esse gene apresente evidências de seleção positiva na linhagem de *Co. hominivorax* por seleção em fenótipo não necessariamente relacionado ao hábito de parasitismo. No entanto, uma de suas funções inferidas por anotação automática indica similaridade a proteínas ligantes de ferro (InterPro:IPR001128, InterPro:IPR002401, FlyBase Curators et al., 2004-), um processo também observado dentre os genes com expressão diferencial na análise de dados de RNA-seq. Essa nova função descrita, ainda não foi explorada, mesmo em *Dr. melanogaster*.

Outro transcrito com evidência de seleção positiva na linhagem de *Co. hominivorax* apresentou similaridade ao gene "SH3PX1" de *Dr. melanogaster*. Foi pouco investigado, mesmo na espécie modelo, mas um estudo indicou um possível envolvimento na resposta à hipóxia, função também observada dentre os genes diferencialmente expressos. Lee e colaboradores (2008) realizaram o silenciamento por RNAi em escala genômica utilizando uma biblioteca de 7216 dsRNAs. Por meio de um gene repórter controlando a proteína verde fluorescente (GFP), foi possível identificar genes de *Dr. melanogaster* envolvidos na supressão da transcrição em resposta à hipóxia. Dentre os genes identificados estava o "SH3PX1". Uma diferença na resposta a hipóxia também foi observada na análise da expressão gênica diferencial, mostrando, mais uma vez, uma coincidência entre os processos biológicos (mas não os genes) com diferença na expressão entre as espécies e genes com evidência de seleção positiva.

6 CONCLUSÕES

Após a realização dos testes para avaliação quantitativa da expressão gênica e análise de sequências, foi possível identificar mudanças, tanto entre espécies que causam diferentes tipos de miíase (furuncular ou ferida), quanto entre espécies que parasitam diferentes regiões do hospedeiro (dérmica ou nasofaringeal), de genes possivelmente relacionados com a preferência alimentar.

Foram identificados genes cujos produtos estão envolvidos com o controle do metabolismo, como os que codificam para o translocador/carregador ATP/ADP e Levy, associados com o metabolismo celular. Também foi observado um gene que não está diretamente envolvido no metabolismo, mas pode apresentar um importante papel na manutenção do hábito parasita, como "fau" que codifica para a proteína "anoxia up-regulated", sendo expressos em situações de estresse celular.

Por fim temos o gene Ferritina, relacionado com a absorção de ferro, que está mais expresso nas espécies que têm preferência alimentar por músculo esquelético ou que parasitam a região dérmica. Nessas espécies (*Co. hominivorax*, *De. hominis* e *L. cuprina*) esse gene pode estar relacionado com a maior necessidade em lidar com um ambiente rico em ferro, como a derme e o tecido muscular esquelético. Esse gene pode ser um futuro candidato para estudos funcionais ligados a preferência alimentar e ao hábito parasita.

Nas análises das sequências codificantes, observamos que a maioria das sequências testadas estão evoluindo sob seleção negativa. Detectamos evidências de seleção positiva em nove e quatro genes no ramo de no ramo de *Ch. hominivorax* e *De. hominis*, respectivamente. Dentre esses, foi encontrado um gene comum entre *Co. hominivorax* e *De. hominis*: o gene "starvin" (stv), cujo produto é essencial na ingestão de alimentos por larvas de *Dr. melanogaster*. Outros genes

observados no ramo de *Co. hominivorax* apresentam funções semelhantes às observadas nos genes DE. O gene Spook está envolvido em processos de oxido-redução e ligação de ferro e o gene SH3PX1 está envolvido na resposta a hipóxia. Assim como o gene da Ferritina, eles podem ser futuros candidatos para estudos mais aprofundados para estudar relações de causalidade entre a variação de sequência codificante e ao hábito parasita.

7 RESUMO

A infestação por larvas de dípteros, conhecida como miíase, é um grande problema na pecuária em todo o mundo e pode causar perdas econômicas severas. A superfamília Oestroidea é um modelo interessante para estudar a evolução de moscas causadoras de miíase devido à diversidade de estratégias de parasitismo entre espécies intimamente relacionadas nessa família. Essas moscas são classificadas pelo seu hábito alimentar como parasitas saprófagos, obrigatórios ou parasitas facultativos. Os parasitas, em particular, podem ser subdivididos pela manifestação clínica/local de infestação como dérmica, nasofaríngea, traumática/ferida e furuncular. Com tal diversidade de estratégias, postula-se que espécies intimamente relacionadas tenham diferenças genéticas que desempenham um papel na formação desses hábitos. Aqui, utilizamos dados de expressão gênica e as sequências codificantes em escala genômica de cinco espécies (*Cochliomyia hominivorax*, *Chrysomya megacephala*, *Lucilia cuprina*, *Dermatobia hominis* e *Oestrus ovis*) para encontrar genes que possam estar envolvidos em diferentes estratégias e/ou preferências alimentares. Nós testamos se os 1.287 ortólogos identificados possuiam expressão diferente e restrições evolutivas em diferentes cenários. Ao comparar seus perfis de expressão gênica, encontramos dois genes regulados positivamente; um genes em espécies que causam miíase dérmica envolvido no transporte metabolização de ferro (Ferritina) e outro gene em espécies que causam miíase traumática que responde a níveis reduzidos de oxigênio (anoxia up-regulation-like). Nossa análise evolutiva mostrou um resultado semelhante. Em *Ch.* hominivorax, encontramos genes diferentes, mas envolvidos nas mesmas funções que podem estar evoluindo sob seleção positiva. Este é o primeiro passo para entender as origens e a evolução da diversidade de estratégias parasitárias em Oestroidea.

8 ABSTRACT

The infestation by dipterous larvae, known as myiasis, is a major problem in livestock worldwide and can cause severe economic losses. The Oestroidea superfamily is an interesting model to study the evolution of myiasis-causing flies because of the diversity of parasitism strategies among closely-related species in this family. These flies are classified by their feeding habit as saprophagous, obligate parasites or facultative parasites. The parasites in particular can be subdivided into dermal, nasopharyngeal, traumatic/wound and furuncular. With such a diversity of parasitic strategies, we expect that closely-related species have genetic differences that play a role in shaping these habits. Here, we used gene expression and coding sequence data from five species (*Cochliomyia hominivorax*, *Chrysomya megacephala*, *Lucilia cuprina*, *Dermatobia hominis* and *Oestrus ovis*) to find genes that may be involved in different parasitic strategies. We tested whether 1,287 orthologs have different expression and evolutionary constrains in different scenarios. By comparing their gene expression profiles, we found two up-regulated genes; one in species causing dermic myiasis that is involved in iron transportation/metabolization (Ferritin), and other in species causing traumatic myiasis that responds to reduced oxygen levels (anoxia up-regulated-like). Our evolutionary analysis showed a similar result. In the *Ch. hominivorax* branch, we found different genes, but involved in the same functions that may be evolving under positive selection. This is the first step towards understanding the origins and evolution of the diversity of parasitic strategies in Oestroidea.

9 REFERÊNCIAS

Abascal, F., Zardoya, R. & Telford, M.J., 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*, 38(suppl_2), pp.W7–W13.

Andrews S (2010) FASTQC. A quality control tool for high throughput sequence data. URL http://www. bioinformatics. babraham. ac. uk/projects/fastqc.

Anstead, C.A. et al., 2015. Lucilia cuprina genome unlocks parasitic fly biology to underpin future interventions. *Nature communications*, 6, p.7344.

Coulson, M., Robert, S., & Saint, R., 2005. Drosophila starvin encodes a tissue-specific BAG-domain protein required for larval food uptake. *Genetics*, 171(4), pp.1799-812.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), pp.1792–7.

Enright, A.J., Van Dongen, S. & Ouzounis, C.A., 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic acids research*, 30(7), pp.1575-84.

FlyBase Curators, Swiss-Prot Project Members, InterPro Project Members, 2004- . Gene Ontology annotation in FlyBase through association of InterPro records with GO terms.

Fu, L. et al., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics (Oxford, England)*, 28(23), pp.3150–2.

Grisi, L. et al., 2014. Reassessment of the potential economic impact of cattle parasites in Brazil. *Revista Brasileira de Parasitologia Veterinária*, 23(2), pp.150–156.

Hall, M. & Wall, R., 1995. Myiasis of Humans and Domestic Animals. In pp. 257–334.

Heberle, H. et al., 2015. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*, 16(1), p.169.

Kondrashov, N. et al., 2011. Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. Cell 145: 383–397.

KUTTY, S.N. et al., 2010. Molecular phylogeny of the Calyptratae (Diptera: Cyclorrhapha) with an emphasis on the superfamily Oestroidea and the position of Mystacinobiidae and McAlpine's fly. *Systematic Entomology*, 35(4), pp.614–635.

Langmead, B. et al., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), p.R25.

Lein, M.M., 2013. *Anoxia Tolerance of Forensically Important Calliphorids*,.

Li, B. & Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1), p.323.

Liu, W. et al., 2007. Mutations in cytochrome c oxidase subunit VIa cause neurodegeneration and motor dysfunction in Drosophila. *Genetics*, 176(2), pp.937–46.

Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15.

Ma, E., Xu, T. & Haddada, G.G., 1999. Gene regulation by O2 deprivation: an anoxia-regulated novel gene in Drosophila melanogaster. *Molecular Brain Research*, 63(2), pp.217–224.

Marinho, M.A.T. et al., 2012. Molecular phylogenetics of Oestroidea (Diptera: Calyptratae) with emphasis on Calliphoridae: Insights into the inter-familial relationships and additional evidence for paraphyly among blowflies. *Molecular Phylogenetics and Evolution*, 65, pp.840–854.

Mccarthy, F.M. et al., 2006. AgBase: a functional genomics resource for agriculture. *BMC Genomics*, 7. Robbins, K. & Khachemoune, A., 2010. Cutaneous myiasis: a review of the common types of myiasis. *International Journal of Dermatology*, 49(10), pp.1092–1098.

Ni, J.Q., Liu, L.P., Hess, D., Rietdorf, J. & Sun, F.L., 2006. Drosophila ribosomal proteins are associated with linker histone H1 and suppress gene transcription. Genes and Development, 20, pp. 1959–1973.

Ono, H. et al., 2006. Spook and Spookier code for stage-specific components of the ecdysone biosynthetic pathway in Diptera. Developmental Biology, 298(2), pp. 555-570.

Roberts, A. et al., Improving RNA-Seq expression estimates by correcting for fragment bias.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), pp.139–140.

Simão, F.A. et al., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), pp.3210–3212.

Stevens, J.R. et al., 2006. The evolution of myiasis in humans and other animals in the Old and New Worlds (part II): biological and life-history studies. *Trends in Parasitology*, 22(4), pp.181–188.

Stevens, J.R. & Wallman, J.F., 2006. The evolution of myiasis in humans and other animals in the Old and New Worlds (part I): phylogenetic analyses. *Trends in parasitology*, 22(3), pp.129–36.

Sun, J., Li, C. & Wang, S., 2015.The Up-Regulation of Ribosomal Proteins Further Regulates Protein Expression Profile in Female Schistosoma japonicum after Pairing. PLOS ONE 10(6): e0129626.

Talavera, G. et al., 2007. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments K. Kjer, R. Page, & J. Sullivan, eds.

Systematic Biology, 56(4), pp.564–577.

Tandonnet, S. & Torres, T.T., 2017. Traditional versus 3' RNA-seq in a non-model species. *Genomics Data*, 11, pp.9–16. Available at:

Tang, X. & Zhou, B., 2013. Ferritin is the key to dietary iron absorption and tissue iron detoxification in Drosophila melanogaster. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 27(1), pp.288–98.

Thomas, G. 2000. An encore for ribosome biogenesis in the control of cell proliferation. Nat Cell Biol 2: E71–72.

Torres, T.T. et al., 2004. Isolation and characterization of microsatellite markers in the new world screw-worm Cochliomyia hominivorax (Diptera: Calliphoridae). *Molecular Ecology Notes*, 4(2), pp.182–184.

Vartiainen, S. et al., 2014. Phenotypic rescue of a Drosophila model of mitochondrial ANT1 disease. *Disease models & mechanisms*, 7(6), pp.635–48.

Wang, X. et al., 2011. The developmental transcriptome of the synanthropic fly Chrysomya megacephala and insights into olfactory proteins..

Wang, Z., Gerstein, M. & Snyder, M., 2009a. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1), pp.57–63.

Wang, Z., Gerstein, M. & Snyder, M., 2009b. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), pp.57–63.

Yang, Y. & Smith, S. a, 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Molecular biology and evolution*, 31(11), pp.3081–92.

Yang, Z., 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, 24(8), pp.1586–1591.

Zdobnov, E.M. et al., 2017. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic acids research*, 45(D1), pp.D744–D749.

Zhang, M. et al., 2013. Analysis of the Transcriptome of Blowfly Chrysomya megacephala (Fabricius) Larvae in Responses to Different Edible Oils D. Zhou, ed. *PLoS ONE*, 8(5), p.e63168.

ZUMPT, F., 1965. Myiasis in Man and Animals in the Old World. A Textbook for Physicians, Veterinarians and Zoologists. *Myiasis in Man and Animals in the Old World. A Textbook for Physicians, Veterinarians and Zoologists.*