UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de São José do Rio Preto

William Vilas Boas Nunes

História evolutiva e divergência funcional do gene de especiação *OdsH* e sua duplicata *unc-4* em Drosophilidae

São José do Rio Preto
2022

William Vilas Boas Nunes

História evolutiva e divergência funcional do gene de especiação *OdsH* e sua duplicata *unc-4* em Drosophilidae

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Biociências, junto ao Programa de Pós-graduação em Biociências, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto.

Financiadora: CAPES

Orientadora: Profª. Drª. Claudia Marcia Aparecida Carareto

São José do Rio Preto
2022

William Vilas Boas Nunes

História evolutiva e divergência funcional do gene de especiação *OdsH* e sua duplicata *unc-4* em Drosophilidae

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Biociências, junto ao Programa de Pós-graduação em Biociências, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto.

Financiadora: CAPES

Comissão Examinadora

Profª. Drª. Claudia Marcia Aparecida Carareto
UNESP – Câmpus de São José do Rio Preto
Orientadora

Profª. Drª. Maria Dulcetti Vibranovski
Universidade de São Paulo

Profª. Drª. Vera Lucia da Silva Valente Gaiesky
Universidade Federal do Rio Grande do Sul

São José do Rio Preto
08 de julho de 2022

Às trabalhadoras e aos trabalhadores

AGRADECIMENTOS

À minha mãe, Maira Vilas Boas, cujos incansáveis esforços e apoio incondicional me permitiram alcançar este momento.

À minha orientadora, Claudia Carareto, por todos os ensinamentos, dedicação, paciência, apoio, rigor, carinho e respeito, que possibilitou a execução desse projeto e enriqueceu imensuravelmente minha formação científica.

Às minhas companheiras e companheiros do Laboratório de Evolução Molecular, Maryanna Simão, Cecília Banho, Camila Vieira, Daniel de Oliveira, Luis Gustavo Galego, Bianca Manfré e Ana Carolina Pires, pelos ensinamentos, trocas de ideias, conversas, amizades e descontrações, que tornaram esta minha jornada mais rica e leve. Também à Assistente de Suporte Acadêmico Joice Biselli, por todos os ensinamentos e paciência, desde o início de minha jornada no laboratório, e pela contribuição às análises moleculares deste projeto.

À equipe do Laboratoire de Biométrie et Biologie Évolutive (LBBE), da Universidade de Lyon, em especial à professora Cristina Vieira, pela supervisão durante o estágio que lá desenvolvi e pelas contribuições a este trabalho, e à Nelly Burlet, pela supervisão e ensinamentos nessa etapa do meu projeto.

Aos meus colaboradores pela contribuição neste projeto: Daniel de Oliveira, pelo método de anotação de sequências; Guilherme Dias e Antônio Bernardo Carvalho, pela concessão de sequências das espécies basais de Drosophilidae; Joice Biselli, pela padronização da RT-qPCR; Ícaro Putinhon, pelas análises estruturais proteicas.

Aos professores, funcionários e alunos do IBILCE-UNESP e do Programa de Pós-Graduação em Biociências que vieram a contribuir à minha jornada.

A todas minhas professoras e professores que se dedicam verdadeiramente ao ensino e que cujos esforços à minha formação se fazem presentes, diretamente ou indiretamente, nesta dissertação.

Ao meu irmão Wender Vilas Boas, à minha prima Isabella Vilela e aos meus amigos, em especial Bojana Cetinja, Bianca Barbosa, Mariana Sant'Anna, Vinicius Moreira, Marília Ferreira, Ana Carolina Pires, Raquel Megid, Gabriela Calvet, Maryanna Simão e Indira Barbosa, pelo companheirismo durante a minha jornada no Mestrado, por sonharem comigo com um futuro melhor, por todo o apoio e torcida e por tornarem minha vida mais feliz.

Ao Programa de Educação Tutorial, em especial à minha tutora Claudia Bonini e meus companheiros do PET Biologia, pelo profundo enriquecimento da minha formação acadêmica.

Aos governantes cujas políticas contribuíram à democratização do acesso à Ciência e à Universidade pública aos filhos da classe trabalhadora.

Ao Centre National de la Recherche Scientifique, da França, pelo apoio ao estágio realizado durante este projeto junto ao LBBE.

O presente trabalho foi realizado com a concessão de bolsa de Mestrado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, à qual agradeço.

Agradeço à FAPESP pelo auxílio ao projeto de pesquisa, sob o processo nº 2020/060238-2, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e pelo apoio ao estágio realizado durante este projeto junto ao LBBE.

RESUMO

O gene *Odyssesus* (*OdsH*) foi o primeiro gene de especiação descrito em *Drosophila*, assim caracterizado em face de sua associação com a esterilidade híbrida entre *Drosophila simulans* e *Drosophila mauritiana*. Sua origem é atribuída à duplicação do gene *unc-4*, o qual é muito conservado em metazoários e relacionado a funções neurais. Tanto *OdsH* como *unc-4* codificam homeodomínios e atuam como fatores de transcrição. A ocorrência dessa duplicação foi circunscrita ao subgênero *Sophophora*, contudo, análises preliminares indicaram sua ocorrência em outros subgêneros de *Drosophila*. O presente estudo teve como objetivo ampliar a busca pelo gene *OdsH* para além do gênero *Drosophila*, de modo definir a extensão de sua ocorrência e datar o evento da duplicação, bem como, caracterizar a evolução molecular e funcional desta duplicata. Foram encontradas evidências da ocorrência de *OdsH* nos genomas de espécies das duas tribos da subfamília Drosophilinae (Drosophilini e Colocasiomyini) da família Drosophilidae, mas não em genomas da subfamília Steganinae. A Inferência Bayesiana utilizada para coestimar as relações filogenéticas e o tempo de divergência entre as sequências das duplicatas *unc-4* e *OdsH* indicaram que elas formam dois clados monofiléticos irmãos em Drosophilinae. Concordantemente com as estimativas de divergência das duas subfamílias, a datação da duplicação permitiu estimar sua origem há 55,58 milhões de anos atrás. As taxas de evolução desses genes indicaram assimetria, com aceleração da evolução de *OdsH*, sugerindo-se maior diversificação dessa duplicata anteriormente à divergência da subfamília Drosophilinae. Busca em bancos públicos de dados por transcriptomas de órgãos reprodutivos e não reprodutivos em espécies da subfamília Drosophilinae, confirmaram expressão inovativa de *OdsH* em testículos de *Drosophila*, o que não foi observado para *unc-4*. Adicionalmente, foi inferida a regulação de *OdsH* por fatores de transcrição relacionados à reprodução masculina. Análises dos domínios funcionais proteicos indicaram que a proteína OdsH apresenta maior divergência dos motivos funcionais, do homeodomínio e do octapeptídeo, do que Unc-4, e que a divergência observada no homeodomínio pode perturbar a energia de ligação ao DNA, tornando o sistema mais instável. Essas características permitem sugerir que as duas proteínas têm diferentes alvos de ligação na regulação de genes alvo. A proteína OdsH de *D. mauritiana* apresenta singularidades que podem estar associadas à sua incompatibilidade em relação ao genoma de *D. simulans*, e relacionada à esterilidade híbrida reportada entre as duas espécies. Os níveis de expressão de *OdsH* variam entre as espécies *D. mojavensis* e *D. arizonae*, e entre os híbridos férteis e estéreis, o que pode configurar um cenário de envolvimento de *OdsH* com a esterilidade híbrida. A expressão de *OdsH* nessas espécies ocorre nos espermatócitos,

indicando que possivelmente desempenha a regulação de fatores meióticos e pós meióticos. Neste estudo reportamos de forma inédita que o gene *OdsH* teve origem no ancestral Drosophilinae e que reúne características que permitem propor sua neofuncionalização em funções sexuais masculinas logo após sua origem.

Palavras–chave: Duplicação gênica. Homeodomínios. Neofuncionalização. Fator de transcrição. Drosophilidae.

ABSTRACT

The *Odysseus* (*OdsH*) gene was the first speciation gene described in *Drosophila*, associated with hybrid sterility between *Drosophila simulans* and *Drosophila mauritiana*. Its origin is attributed to the duplication of *unc-4*, a gene highly conserved in metazoans and related to neural functions. Both *OdsH* and *unc-4* are transcription factors that encode homeodomains. The occurrence of this duplication has been limited to the subgenus *Sophophora*, however preliminary analyses indicated its occurrence in other *Drosophila* subgenera. This study had the goal to expand the search for the *OdsH* beyond *Drosophila* genus, in order to define the extent of its occurrence and to date the duplication event, as well as to characterize the molecular and functional evolution of this duplicate. Evidence of the occurrence of *OdsH* was found in the genomes of species from two tribes of the Drosophilinae subfamily (Drosophilini and Colocasiomyini) of the Drosophilidae family, but not in genomes from the Steganinae subfamily. The Bayesian Inference used to co-estimate the phylogenetic relationships and the divergence time between sequences of the *unc-4* and *OdsH* duplicates indicated that they form two monophyletic sister clades in Drosophilinae. In agreement with the divergence estimations of the two subfamilies, the duplication dating allowed us to estimate its origin at 55.58 million years ago. The evolutionary rates of these genes indicated asymmetry, with acceleration of *OdsH* evolution, suggesting greater diversification of this gene prior to the divergence of the Drosophilinae subfamily. Through expression level comparisons of reproductive and non-reproductive organs in species of the Drosophilinae subfamily, as well as data obtained via quantitative expression analysis specifically for *D. mojavensis*, *D arizonae* and their fertile and sterile hybrids, we have confirmed the innovative expression of *OdsH* in testes of *Drosophila*, which was not observed for *unc-4*. Additionally, we have found evidence of *OdsH* regulation by transcription factors related to male reproduction. Functional analysis of the protein domains indicated that the OdsH protein shows greater divergence of the functional motifs, as well as the homeodomain and the octapeptide, in comparison to Unc-4. Furthermore, the divergence observed in the homeodomain can disturb the DNA binding energy, making the system more unstable. These characteristics allow us to suggest that the two proteins have different binding targets in the regulation of target genes. The protein OdsH of *D. mauritiana* have specific features that may be associated with their incompatibility with the genome of *D. simulans*, as well as to the reported hybrid sterility. *OdsH* expression levels vary between *D. mojavensis* and *D. arizonae* species, and between fertile and sterile hybrids, which can configure a scenario of *OdsH* relationship with hybrid sterility. The expression of *OdsH* in these species occurs in

spermatocytes, indicating that it possibly plays a role in the regulation of meiotic and post-meiotic factors. In this study, we reported an unprecedented way that the *OdsH* gene originated in the Drosophilinae ancestor. We also found that it has characteristics allowing us to propose its neofunctionalization in male sexual functions rapidly after its origin.

.

Keywords: Gene duplication. Homeodomain. Neofunctionalization. Transcription factor. Drosophilidae.

LISTA DE ILUSTRAÇÕES

Figura 1 – Esquema da espermatogênese ao longo do testículo de *D. melanogaster* 20
Figura 2 – Estrutura do gene *OdsH* em *D. melanogaster* 22
Figura 3 – Relações filogenéticas entre as subfamílias e tribos de Drosophilidae 25
Figura 4 – Distribuição geográfica das populações de *D. arizonae* e *D. mojavensis* 26
Figura 5 – Esquema dos testes de hipóteses desenvolvidos para análises de assinaturas de seleção entre *OdsH* e *unc-4* 31
Figura 6 – Cruzamentos entre *D. m. baja* e *D. arizonae* 35
Figura 7 – Localização relativa de *unc-4* e *OdsH* em genomas de Drosophilinae 38
Figura 8 – Inferência filogenética Bayesiana das sequências dos genes parálogos *unc-4* e *OdsH* e datação do evento de duplicação *unc-4*/*OdsH* 41
Figura 9 – Sítios de ligação de fatores de transcrição nas regiões reguladoras de *OdsH* e *unc-4* em *Drosophila* 45
Figura 10 – Motivos funcionais nas proteínas Unc-4 e OdsH em Drosophilinae 47
Figura 11 – Expressão relativa de *OdsH* em *D. arizonae* e *D. mojavensis* e seus híbridos recíprocos 48
Figura 12 – Hibridização *in situ* de *OdsH* em testículos de *D. arizonae*, *D. m. baja* e seus híbridos 50

Quadro 1 – Expressão dos genes *unc-4* e *OdsH* em tecidos de *Drosophila* e *T. dalmanni*. 44

LISTA DE TABELAS

Tabela 1 – Sequências de oligonucleotídeos utilizados nos ensaios de q-PCR. 36

Tabela 2 – Valores de ω obtidos por meio de teste de seleção por ramos na filogenia de *OdsH* por grupos de Drosophilinae. 43

LISTA DE ABREVIATURAS E SIGLAS

BLAST	Basic Local Alignment Search Tool
BIC	Bayesian Information Criterion
Ct	Cycle Threshold
e.g.	exemple grata
ENC	Effective Number of Codons
ESS	Effective Sample Size
FISH	Fluorescent *in situ* hybridization
GO	Gene Onthology
GTR	General Time Reversible
indel	inserção-deleção
LBA	Long Branch Attraction
MCMC	Markov Chain Monte Carlo
NCBI	National Center for Biotechnology Information
TFBS	Transcription Factor Binding Site
PDB	Protein Data Bank
PBS	Phosphate-Buffered Saline
RSCU	Relative Synonymous Codon Usage
RT-PCR	Reverse transcription polymerase chain reaction
qPCR	Real-time polymerase chain reaction

LISTA DE SÍMBOLOS

M.a	Milhões de anos
pb	pares de bases
kb	Quilobases
kJ	Quilojoule
nm	Nanômetro
Gb	Energia livre de Gibbs
h	Hora
ng	Nanograma
µL	Microlitro

SUMÁRIO

1 INTRODUÇÃO **16**

1.1 Evolução de duplicatas gênicas **17**

1.2 Espermatogênese em *Drosophila* **19**

1.3 Evolução de *OdsH* e *unc-4* **21**

1.4 A família Drosophilidae

1.4.1 Espécies do grupo *repleta*, *Drosophila arizonae* e *D. mojavensis*, modelo em estudos de especiação 25

2 OBJETIVOS **27**

2.1 Objetivo geral **27**

2.2 Objetivos específicos **27**

3 MATERIAL E MÉTODOS **28**

3.1 Anotação de *unc-4* e *OdsH* em genomas de Drosophilidae **28**

3.2 Inferência filogenética e datação da duplicação **28**

3.3 Viés de uso de códons **30**

3.4 Taxa relativa de substituição nucleotídica **30**

3.5 Estimativas de seleção natural e investigação de seleção positiva **30**

3.6 Expressão de *unc-4* e *OdsH* **32**

3.7 Sítios de ligação de fatores de transcrição nas regiões reguladoras de *unc-4* e *OdsH* **32**

3.8 Motivos funcionais proteicos **33**

3.9 Análises experimentais da expressão quantitativa e *in situ* de *OdsH* e *unc-4* em *D. m. baja* e *D. arizonae,* e seus híbridos interespecíficos **34**

3.9.1 Extração de RNA e RT-qPCR 35

3.9.2 Detecção de transcritos *OdsH* em testículos por hibridização fluorescente *in situ* (FISH) 36

4 RESULTADOS **37**

4.1 Ocorrência de *unc-4* e *OdsH* em genomas de Drosophilidae e relações filogenéticas **37**

4.2 Dinâmica evolutiva **42**

4.3 Expressão em transcriptomas de *Drosophila* **43**

4.4 Fatores de transcrição candidatos à regulação da expressão de *OdsH* e *unc-4* **44**

4.5 Motivos funcionais proteicos **45**

4.6 Expressão quantitativa de *OdsH* e *unc-4* em *D. m. baja* e *D. arizonae* e seus híbridos **48**

4.7 Expressão *in situ* de *OdsH* em testículos de *D. m. baja* e *D. arizonae* e seus híbridos **48**

5 DISCUSSÃO **50**

5.1 *OdsH* – a emergência de uma nova duplicata na subfamília Drosophilinae **51**

5.2 *OdsH* e *unc-4* – mesma origem, mas histórias evolutivas divergentes **54**

5.3 *OdsH* – um caso de neofuncionalização de genes codificantes de fatores de transcrição **56**

6 CONCLUSÕES **59**

REFERÊNCIAS **60**

APÊNDICE A – Figuras Suplementares **72**

APÊNDICE B – Tabelas Suplementares **79**

1 INTRODUÇÃO

A origem de novas espécies e os mecanismos moleculares envolvidos nesse processo são questões centrais na Biologia, uma vez que são elementos chave na produção de diversidade biológica. Sob o conceito biológico de espécie, a especiação é decorrente da evolução de mecanismos de isolamento reprodutivo que impedem o fluxo gênico entre as populações (DOBZHANSKY, 1937; 1942; MAYR, 1942, 1963). Todavia, em espécies incipientes, o isolamento reprodutivo pode não ser completo, e o intercruzamentos entre elas pode resultar na fusão de duas espécies em uma única linhagem (COMEAULT et al., 2015; CENZER, 2016) ou na formação de híbridos com valor adaptativo reduzido e a persistência das espécies emergentes (ROSENBLUM et al., 2012).

Os inúmeros fatores que podem impedir a reprodução heteroespecífica classificam-se, de acordo com o momento do ciclo reprodutivo em que a barreira se estabelece, em: (i) pré-zigótico, que impossibilitam o cruzamento entre espécies; (ii) pré-zigótico pós-cópula, que impedem a formação do zigoto híbrido após o acasalamento interespecífico; e (iii) pós-zigótico, com a formação de híbridos com valor adaptativo reduzido em relação aos parentais, como, por exemplo, híbridos inviáveis ou estéreis (DOBZHANSKY, 1937; COYNE; ORR, 2004). Embora tenha sido observado que as barreiras pré-zigóticas e pré-zigóticas pós-cópula evoluem mais inicialmente durante a divergência das espécies (COYNE; ORR, 1989; DIXON et al., 2003; TERUSSINI et al., 2018), estudos indicam que as barreiras pós-zigóticas podem reduzir efetivamente o fluxo gênico entre pares de espécies e promover o estabelecimento subsequente de isolamento reprodutivo (revisado em COUGHLAN; MATUTE, 2020). O momento em que essas barreiras se estabelecem pode diferir entre os sexos. Essa diferença tem sido bem documentada em insetos, pássaros e mamíferos (COYNE; ORR, 1989) e fundamentou a proposição da conhecida Regra de Haldane (HALDANE, 1922); segundo a qual quando o valor adaptativo difere entre os sexos da prole híbrida, o sexo heterogamético tem o menor valor. Essa regra é corroborada pela observação de que a esterilidade de machos híbridos evolui 10-100 vezes mais rapidamente do que a esterilidade da fêmea híbrida em espécies com determinação sexual XY, o mesmo se aplicando para a inviabilidade híbrida (WU, 1991; WU; DAVIS, 1993; WU et al., 1996).

A regra de Haldane pode ser explicada pelo modelo proposto por Dobzhansky-Muller (DOBZHANSKY, 1937; MULLER, 1942) para a evolução das barreiras reprodutivas características do isolamento pós-zigótico a partir da incompatibilidade entre locos genéticos que surgem como um subproduto da divergência entre as populações isoladas. Assim, a teoria

da dominância da regra de Haldane propõe que a redução do valor adaptativo de híbridos heterogaméticos se dá em função da incompatibilidade de alelos recessivos localizados no cromossomo X de uma espécie parental com fatores autossômicos da outra espécie (HALDANE, 1922). Nesse sentido, aceita-se atualmente que o isolamento reprodutivo pós-zigótico seja um evento com a participação de diversos genes e das complexas redes regulatórias desses genes, além de outros elementos genéticos, como os elementos de transposição (e. g. KELLEHER et al., 2012; SATYAKI et al., 2014; GOMES; CIVETTA, 2014; 2015; ROMERO-SORIANO et al., 2017; LOPEZ-MAESTRE et al., 2017; BANHO et al., 2021).

Estudos com o objetivo de identificar genes específicos associados à redução do valor adaptativo de híbridos – os "genes de especiação" – levaram ao avanço da compreensão sobre os mecanismos genéticos envolvidos no isolamento reprodutivo. Embora, frequentemente esses genes apresentem funções não essenciais nas espécies puras, suas taxas de evolução se mostram aceleradas, provavelmente dirigidas por seleção Darwiniana positiva (ORR et al., 2004; WU; TING, 2004). Essas características se ajustam às de *OdsH*, o primeiro gene de especiação descrito em *Drosophila*, que tem um papel modesto na espermatogênese de *Drosophila melanogaster* (SUN et al., 2004), e evolução rápida, com sinais de seleção positiva nas espécies *D. simulans* e *D. mauritiana* (TING, 1998). A origem desse gene remete a uma duplicação do gene *unc-4* (TING et al., 2004).

1.1 Evolução de duplicatas gênicas

A duplicação gênica é um fator importante para o aumento do número de genes e da diversidade das funções gênicas e, assim, da complexidade dos genomas e dos organismos (OHNO, 1970). A duplicação de genes pode se dar em decorrência de poliploidização, retrotransposição ou mecanismos baseados em DNA (RAMOS; FERRIER, 2012). Estes pocessos são responsáveis pela origem da maior parte de novos genes em *Drosophila* (ZHOU et al., 2008) e podem acarretar duplicações segmentares *in tandem* ou dispersas (RAMOS; FERRIER, 2012). Os mecanismos genômicos envolvidos nas duplicações segmentares podem ser dependentes de sequências homólogas ou repetitivas - por meio da recombinação homóloga não alélica (STANKIEWICZ; LUPSKI, 2002; BAILEY et al., 2003; FISTON-LAVIER et al., 2007) ou transposição mediada por elementos de transposição (LAL et al., 2003; JIANG et al., 2004; YANG et al., 2008) - ou independentes de homologia de sequências, com o reparo ao DNA ocorrendo entre extremidades rompidas localizadas proximamente (ROTH et al., 1985;

SCLACK et al., 2006). A junção de terminais próximos não homólogos tem sido o mecanismo reportado como o mais provável para as duplicações adjacentes em *Drosophila* (ZHOU et al., 2008; MEISEL, 2009; RANZ et al, 2010). Após o evento de duplicação, caso uma das duplicatas não passe pelo processo de pseudogenização, com o acúmulo de mutações degenerativas, diferentes vias funcionais podem direcionar a sua retenção a longo prazo: conservação, subfuncionalização e neofuncionalização.

A conservação se dá quando tanto a cópia parental como a duplicata mantêm a função de seu gene ancestral de cópia única, casos em que a dosagem duplicada de expressão do gene ancestral é vantajosa, e, dessa forma, tanto o gene ancestral como uma ou várias duplicatas são conservadas por seleção natural purificadora (OHNO, 1970). Por outro lado, a subfuncionalização ocorre quando mutações degenerativas levam à perda ou redução da expressão de diferentes subfunções nas duas cópias de um gene, de tal maneira que as duas cópias seriam necessárias para desempenhar o que originalmente era o papel de um único loco ancestral (HUGHES, 1994; FORCE et al., 1999; SOLTZFUS, 1999). Já, a neofuncionalização ocorre com a aquisição de uma nova função ao longo da divergência evolutiva de uma das duplicatas, com efeito potencialmente inovador no fenótipo (OHNO, 1970). Por outro lado, He e Zhang (2005) propuseram uma via alternativa de evolução funcional de duplicatas, denominada subneofuncionalização, segundo a qual após a duplicação ocorre uma fase rápida de subfuncionalização, seguida de uma lenta neofuncionalização.

Assis e Bachtrog (2013) demonstraram que em duplicatas descendentes de um gene ancestral com múltiplas funções, em comparação àquelas descendentes de genes com funções específicas, há maior probabilidade de evoluírem novas funções. Também foi demonstrado que a abrangência do evento de duplicação é um fator determinante para a evolução funcional, de modo que quando ocorre em larga escala, como em duplicações genômicas, ocorre tipicamente a subfuncionalização e em pequena escala, como em duplicações localizadas, ocorre a neofuncionalização (CASNEUF et al., 2006; FARES et al., 2013). Outro fator determinante é o tamanho efetivo da população, sendo que a probabilidade de subfuncionalização diminui conforme o tamanho da população aumenta (LYNCH et al., 2001). Isso se deve à maior ação da seleção natural na evolução molecular proteica em populações grandes, como *Drosophila*, e menor ação nas menores, como mamíferos (BRITTEN, 1986; MORIYAMA, 1987).

Todavia, é aceito consensualmente que a seleção diversificadora (ou seleção positiva) é uma força muito sutil para ser identificada (HUGHES et al., 2000; FAY et al. 2002), não sendo possível sua identificação após 30-50 milhões de anos (M.a) de divergência das duplicatas (HUGHES, 1999; HUGHES et al., 2000), em função do acúmulo de substituições sinônimas

nas sequências após o evento de seleção positiva. Adicionalmente, Van de Peer et al. (2001) observaram que após a divergência das duplicatas antigas, mediada por força de seleção positiva, e o estabelecimento de divergência funcional vantajosa, elas passam a ser alvo de intensa seleção purificadora, sendo possível apenas a identificação de sinais de seleção negativa relaxada. Alguns autores apontam ainda pouca ou nenhuma correlação entre assimetrias de evolução de sequências e neofuncionalidade, sendo necessária a observação de dados funcionais, como expressão, regulação da expressão e características funcionais proteicas, para constatação de divergência funcional (CONANT; WAGNER, 2003; ASSIS; BACHTROG, 2013).

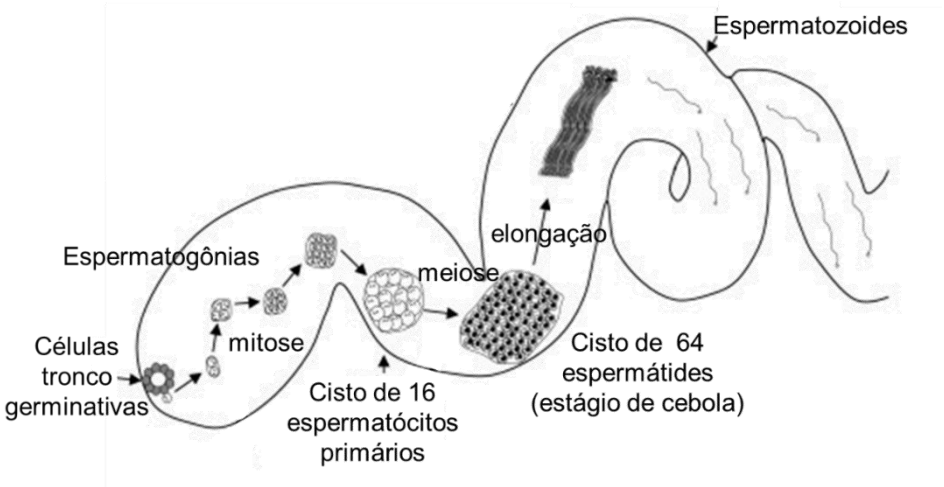
Em *Drosophila*, observou-se que novos genes emergentes via duplicação passam por neofuncionalização em testículos (ASSIS; BACHTROG, 2013; ASSIS, 2016; JIANG; ASSIS, 2017), concordante com a hipótese *out of testes*, que propõe que novos genes tendem a ser expressos em testículos, tecido no qual adquirem novas funções, podendo evoluir outras funções em outros tecidos posteriormente (KAESSMAN, 2010). Os programas de regulação epigenética particulares de testículos poderiam ser uma explicação para esse padrão, pois permitem maior flexibilidade na exposição da cromatina, relacionada ao reparo de danos do DNA acoplado a transcrição, permitindo que a expressão gênica seja mais promíscua em testículos (GAN et al., 2010; XIA et al., 2020; ZHANG; ZHOU, 2019).

1.2 Espermatogênese em *Drosophila*

O desenvolvimento de espermatozoides em *Drosophila* ocorre ao longo do testículo, que possui uma estrutura anatômica similar a um tubo retorcido, em cujo ápice se encontram os estágios iniciais da espermatogênese e, na região mais basal, os estágios mais tardios (Figura 1). As células tronco germinativas se encontram em um micronicho na extremidade apical dos testículos, e apresentam divisão assimétrica, produzindo uma célula tronco e um gonioblasto, que se diferenciará em gameta. A divisão mitótica dos gonioblastos dão origem aos cistos de espermatogônias, circunscritos por células do cisto (somáticas). Em *D. melanogaster*, os gonioblastos passam por quatro rodadas de divisão mitótica, originando 16 espermatogônias, em *D. melanogaster*, e oito em *D. hydei* (grupo *repleta*, subgênero *Drosophila*) (HACKSTEIN, 1987) que completam a fase S pré-meiótica anterior ao crescimento (cerca de 20-25 vezes seu volume inicial), na fase G2, que as diferenciam em espermatócitos primários (FULLER, 1993), marcando o início da prófase meiótica, de acordo com Kimmins e Sassoni-Corsi (2005) e Vibranovski (2014). Após duas divisões meióticas, forma-se o cisto de 64 espermátides, uma

estrutura característica denominada “estágio de cebola”, que pode ser observada como uma estrutura densa com pequenos núcleos (CENCI et al., 1994; FABIAN; BRILL, 2012). As espermátides passam, então, por intensa diferenciação em espermatozoides maduros (espermiogênese), com a individualização, formação do axonema flagelar por meio da síntese de fibras de actina e tubulina, remodelação da membrana plasmática, regulação da morfologia mitocondrial e empacotamento nuclear, coordenados por eficientes vias de regulação da expressão gênica (FABIAN et al., 2010; WHITE-COOPER; BAUSEK, 2010; FABIAN; BRILL, 2012).

Figura 1. Esquema da espermatogênese ao longo do testículo de *D. melanogaster.*



Fonte: Adaptado de Hirst e Carmichael (2011).

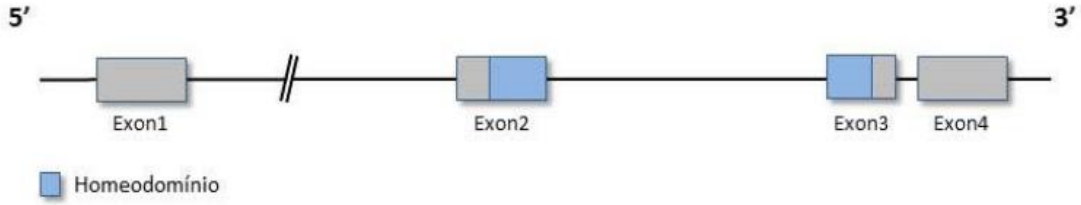
A expressão de genes novos na espermatogênese de *Drosophila* varia temporalmente. Foi observado que genes cuja origem é recente apresentam maior taxa de substituição de nucleotídeos e sinais de seleção positiva e tendem a ser expressos em fases mais tardias da espermatogênese, meiótica e pós-meiótica (JIANG; ASSIS, 2017; RAICES et al., 2019; SU et al., 2021). Enquanto isso, ocorre um enfraquecimento na diversificação de sequências sobre genes com origem mais antiga, que tendem a ser expressos em fases mais iniciais da espermatogênese, mitótica (RAICES et al., 2019; SU et al., 2021).

1.3 Evolução de *OdsH* e *unc-4*

A origem do gene *OdsH* é atribuída a um evento de duplicação do gene *unc-4*, um gene conservado em metazoários e cujas sequências dos homeodomínios são homólogas (TING et al., 2004). De acordo com Ting et al. (2004), a duplicação desse gene teria ocorrido provavelmente no ancestral do subgênero *Sophophora*. Contudo, a sequências homólogas a *OdsH* foram identificadas nas espécies *D. arizonae* e *D. mojavensis* (subgênero *Drosophila*, grupo *repleta*) por nosso grupo de pesquisa (SANTOS, 2014) e buscas, por nós desenvolvidas, identificaram a presença da duplicata também em outros grupos de espécies do subgênero *Drosophila* e no subgênero *Dorsilopha* (dados não publicados). Ademais, a busca no banco de dados GenTree (SHAO et al., 2019; http://gentree.ioz.ac.cn/download/dm6_ver78_age.xlsx) indicou a presença de *OdsH* no nó ancestral do gênero *Drosophila.* Observa-se assim a necessidade de uma datação mais acurada do evento de duplicação que originou *OdsH*, com a utilização de robusta inferência filogenética das relações entre suas sequências, com a calibração do relógio molecular, por meio de taxas de substituição ou com o uso de tempos conhecidos de divergência de táxons analisados (DRUMMOND et al., 2006).

As duplicatas *unc-4* e *OdsH* foram caracterizadas em *D. melanogaster* e outras espécies do subgrupo *melanogaste*r (grupo *melanogaster*, subgênero *Sophophora*, gênero *Drosophila*). Os dois genes encontram-se localizados no cromossomo X, *in tandem*, e separados por uma região intergênica de 12,3 kb. Eles são compostos por quatro éxons, preservando a estrutura exons-introns, sendo que os éxons 2 e 3 apresentam o homeodomínio homólogo em suas sequências (TING et al., 1998) (Figura 2). Embora as duas sequências sejam muito divergentes em ambos os terminais N (incluindo todo o exon 1) e o terminal C (a maior parte do exon 4), a extremidade do C terminal da região codificante contém um pequeno domínio com 8 aminoácidos (octapeptídeo) compartilhado entre os dois genes (COPLEY, 2005). Os homeodomínios codificam aminoácidos conservados nos chamados genes homeóticos, que formam motivos de ligação ao DNA. As proteínas que contêm esses motivos geralmente atuam como fatores de transcrição (McGINNIS et al., 1984), sendo responsáveis por inúmeras vias de determinação e diferenciação celular (revisado em GEHRING, 1987), em especial no processo de neurogênese (LAI et al., 1991).

Figura 2. Estrutura do gene *OdsH* em *D. melanogaster*.



O gene é constituído por quatro éxons, contendo o homeodomínio (azul) compartilhado entre os éxons 2 e 3.
Fonte: SANTOS, 2014.

O homeodomínio Unc-4 é classificado filogeneticamente como pertencente à classe Paired-Like (WINNIER et al., 1999). Alguns resíduos de aminoácidos são elementos chave para a atuação dessa classe de homeodomínios, tal como o resíduo 50 que atua na interação com outros homeodomínios, sendo essa cooperatividade fundamental à especificidade de sua atuação (WILSON et al., 1995). Adicionalmente, são encontrados em algumas famílias dessa classe motivos denominados PAI, RED, octapeptídeo e OAR, sendo os dois primeiros relacionados à interação com o DNA e os dois últimos com outras proteínas, reprimindo e ativando a expressão gênica, respectivamente (revisado em BÜRGLIN; AFFOLTER, 2016). A evolução dessa classe se dá por meio de rearranjos e perdas desses motivos, sendo observada entre as suas famílias uma diversidade de estruturas proteicas quanto a sua presença/ausência (JACOB, 1977). Na proteína Unc-4 de *Caenorhabditis elegans* foi descrito o motivo octapeptídeo na extremidade C-terminal do domínio PAIRED, ao contrário do que se observa nos demais homeodomínios, sugerindo duplicação ou origem *de novo* (WINNIER et al., 1999). O octapeptídeo também foi identificado na extremidade C-terminal de *unc-4* e *OdsH* em *D. melanogaster* (COPLEY, 2005).

O gene cópia única *unc-4* descrito em *C. elegans* é expresso especificamente em tecidos neurais e é responsável pelo desenvolvimento e determinação das vias sinápticas em neurônios motores e de células sensoriais ópticas (FOX et al., 2005; MARQUES et al., 2019; MILLER et al., 1992). Já em *Drosophila*, o gene *unc-4* apresenta funções relacionadas ao desenvolvimento do sistema nervoso central e epidérmico no estágio embrionário (TABUCHI et al., 1998; LACIN; TRUMAN, 2016). Em indivíduos adultos, o produto de expressão do gene *unc-4* é encontrado em neurônios colinérgicos e está relacionado a diferentes vias de destino de linhagem dos neurônios (LACIN et al., 2019, 2020). Também é associada à proteína Unc-4, a

participação no processo de desenvolvimento morfológico de neurônios motores e proprioceptores em adultos (LACIN et al., 2020).

A função do gene *OdsH* foi inicialmente relacionada ao fenótipo estéril de híbridos machos de *D. simulans* e *D. mauritiana*, em decorrência da introgressão de uma sequência *D. mauritiana* contendo *OdsH* e fatores ligados no *background* genômico de *D. simulans*, a partir retrocruzamentos com esta última espécie (PEREZ et al., 1993; PEREZ; WU, 1995; TING et al., 1998). Análises de hibridização *in situ* de transcritos *OdsH* em testículos demonstraram que a expressão do gene *OdsH* é variável, em quantidade e em localização, em espécies do subgrupo *melanogaster* (TING et al. 1998). Por outro lado, em híbridos estéreis, descendentes do cruzamento entre *D. simulans* e *D. mauritiana*, observou-se intensa produção de transcritos *OdsH* na região apical do testículo, onde se encontram as células germinativas pré-meióticas (SUN et al., 2004). Dentre os efeitos resultantes da ausência da expressão de *OdsH*, demonstrados em machos mutantes nulos para esse gene em *D. melanogaster*, foi identificada apenas uma modesta redução na fertilidade durante os três primeiros dias após a eclosão (SUN et al., 2004) e do número de células da linhagem germinativa, incluindo espermatogônias e espermatócitos (CHENG et al., 2012).

O transcrito *OdsH* é expresso em testículos de espécies do subgrupo *melanogaste*r de *Drosophila*, durante a fase G2 do ciclo celular em espermatócitos primários pós-mitóticos em *D. simulans* (BAYES; MALIK et al., 2009). Bayes e Malik (2009) também identificaram que em *D. simulans* e *D. mauritiana*, normalmente, o produto da expressão do gene *OdsH* atua como um fator de transcrição que se liga a regiões heterocromáticas ricas em sequências repetitivas (a região pericêntrica do cromossomo X e do cromossomo 4), descondensando-as e, exclusivamente em *D. mauritiana*, a proteína OdsH se liga à região pericêntrica do cromossomo Y. Os autores também evidenciaram que a descondensação dessa região rica em DNA repetitivo, no *background* genômico de *D. simulans*, pode estar relacionada à esterilidade. Adicionalmente, também foi proposto que *OdsH* atua regulando genes autossômicos expressos mais tardiamente na espermatogênese, sendo a desregulação desses genes nos híbridos, decorrente da desregulação de *OdsH*, provavelmente responsável pelos fenótipos de testículos menores e defeitos nos feixes de esperma observados nos híbridos estéreis (LU et al., 2010).

O homeodomínio *OdsH* apresenta taxa elevada de substituições de aminoácidos nas espécies do subgrupo *melanogaster*, sendo maior que a divergência entre o seu parálogo *unc-4* e *unc-4* cópia única encontrada em espécie evolutivamente distantes (*C. elegans*) (TING et al, 1998). Inferiu-se que a rápida evolução desse gene, ao contrário do que ocorre em geral com os homeodomínios, estaria relacionada com a aquisição de novas funções nos testículos e

participação no processo de especiação (TING et al, 1998; 2004). Em análises de seleção par a par anteriores (dados não publicados), além dos sinais de seleção positiva encontrados por Ting et al. (1998) entre as sequências *OdsH* de *D. mauritiana* e *D. sechellia* (ω = 1,660) e entre as de *D. mauritiana* e *D. simulans* (ω = 1,551), observamos também assinaturas de seleção negativa relaxada nas espécies do grupo *repleta* de *Drosophila*, entre *Drosophila navojoa* e *Drosophila mojavensis* (ω = 0,318).

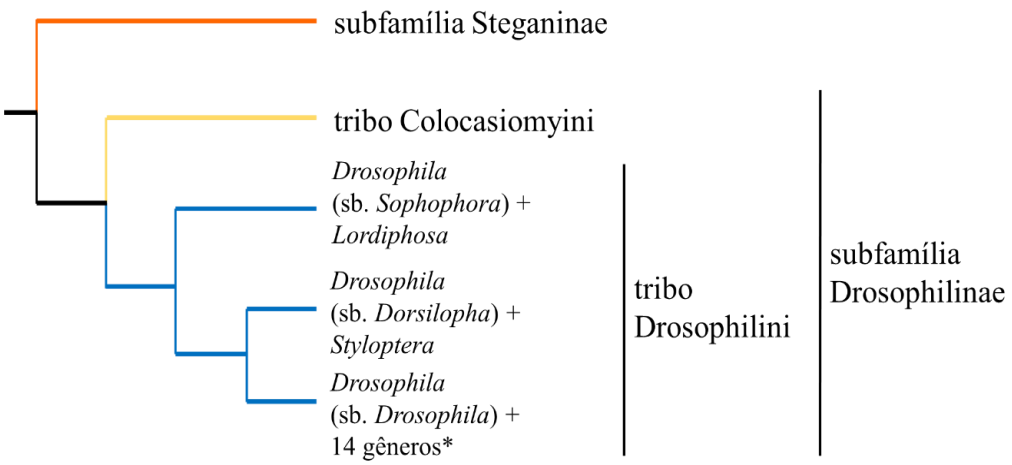
1.4 A família Drosophilidae

A monofilia família Drosophilidae foi proposta a partir de traços simplesiomórficos (e.g. presença das veias A_1 + CuA_2 nas asas) compartilhados também por outras famílias da superfamília Ephydroidea (YASSIN, 2013), pois nenhuma sinapormofia morfológica foi relatada para todos os seus gêneros. Contudo, análises filogenômicas suportam a monofilia dessa família (WIEGMANN, et al., 2011), cuja divergência é estimada em 70,5 M.a (RUSSO et al., 2013).

Tradicionalmente, Drosophilidae é dividida em duas subfamílias monofiléticas, Steganinae e Drosophilinae (Figura 2) (THROCKMORTON, 1975; OKADA, 1989; GRIMALDI, 1990), com 28 e 48 gêneros, respectivamente (BÄCHLI, 2016). Steganinae é dividida em duas tribos, Steganini e Gitonini (OKADA, 1989; GRIMALDI, 1990; YASSIN, 2013). Grimaldi (1990) propôs a divergência basal da subfamília Drosophilinae em duas subtribos, Drosophilini e Colocasiomyini, caracterizadas por Okada (1989), que foram elevadas a tribos e apontadas como monofiléticas por Yassin (2013) (Figura 2). Todavia, as relações filogenéticas e monofilia dos grupos basais em relação à Drosophilini permanecem duvidosas, dada a escassez de estudos publicados com ampla representatividade de espécies desses grupos.

A tribo Drosophilini inclui o gênero *Drosophila* e gêneros relacionados (31, e.g., *Lordiphosa*, *Zaprionus*, *Scaptomyza*) (OKADA, 1989; GRIMALDI, 1990; YASSIN, 2013). O gênero *Drosophila*, é dividido, tradicionalmente, em subgêneros *Sophophora* (que inclui os grupos de espécies *melanogaster*, *obscura*, *willistoni* e *saltans*), *Drosophila* (que inclui os grupos *virilis*, *repleta*, *bromeliae*, *nannoptera*, *melanica*, *robusta*, *quinaria*, *funebris* e *immigrans*) e *Dorsilopha* (com uma única espécie representante *D. busckii*) (Figura 3) (STURTEVANT, 1939; 1942). Contudo, a parafilia do gênero *Drosophila*, com outros Drosophilini, é amplamente reportada e consensualmente aceita (e. g. RUSSO et al., 1995; KWIATOWSKI; AYALA, 1999; VAN DER LINDE et al., 2010; RUSSO et al., 2013; YASSIN, 2013).

Figura 3. Relações filogenéticas entre as subfamílias e tribos de Drosophilidae.



*: *Dichaetophora*, *Dettopsomyia*, *Jeannelopsis*, *Hirtodrosphilia*, *Zygothrica*, *Mycodrosophila*, *Zaprionus*, *Liodrosophila*, *Hypselothyrea*, *Samoaia*, *Mulgravea*, *Microdrosophila*, *Scaptomyza*, *Idiomyia*. Fonte: adaptado de Yassin (2013).

1.4.1 Espécies do grupo *repleta*, *Drosophila arizonae* e *D. mojavensis*, modelo em estudos de especiação

O par de espécies crípticas *Drosophila mojavensis* e *D. arizonae* (grupo *repleta*, subgênero *Drosophila*) tem sido o foco dos estudos de especiação desde 1947 (BAKER, 1947, apud PETIT, 1972). Estas espécies são relativamente jovens, tendo divergido a 1,5 M.a (SANCHEZ-FLORES et al., 2016). As duas espécies apresentam forte isolamento reprodutivo pré-zigótico, porém incompleto (WARSSEMAN; KOEPFER, 1977), uma vez que podem ocorrer corte e cópulas interespecíficas em laboratório (COUNTERMAN; NOOR, 2006; MACHADO et al., 2007). O isolamento reprodutivo pós-zigótico também é incompleto, e mais fraco do que o pré-zigótico, uma vez que pode ocorrer ou não produção de descendência híbrida, em cruzamentos experimentais realizados em laboratório, e assimétrico, uma vez que apenas as fêmeas híbridas são férteis nos dois sentidos de cruzamento e os machos são frequentemente estéreis (BONO; MARKOW, 2009; REED; MARKOW, 2004). O grau de isolamento tanto pré como pós-zigótico depende da origem geográfica das populações de *D. mojavensis* (RUIZ et al., 1990; REED; MARKOW, 2004).

As populações de *D. mojavensis* apresentam diferenças genéticas, comportamentais, ecológicas e morfológicas que permitiram sua classificação em quatro subespécies: *D. m. mojavensis*, encontrada no Deserto de Mojave (sudoeste dos Estados Unidos); *D. m. wrigleyi*,

encontrada exclusivamente na ilha de Santa Catalina, na costa da Califórnia (Estados Unidos); *D. m. sonorensis*, encontrada no Deserto de Sonora no México e no sul do Arizona e *D. m. baja*, encontrada na península de Baja Califórnia (México) (REED; MARKOW, 2004; REED et al., 2007; ROSS; MARKOW, 2006). *Drosophila arizonae*, por sua vez, é encontrada no sudoeste do Novo México, sul do Arizona até o sul do Deserto de Sonora, por todo o oeste e sudeste do México até a Guatemala e possivelmente na Baja Califórnia. Assim, as populações das duas espécies encontram-se parcialmente sobrepostas, originando regiões em que são alopátricas e simpátricas (RUIZ; HEED, 1988) (Figura 4).

Figura 4. Distribuição geográfica das populações de *D. arizonae* e *D. mojavensis.*



*: Cape Region, local de origem da amostra de *D. m. baja*. Fonte: Adaptado de Markow e Hocutt (1998).

Estudos realizados por nosso grupo de pesquisa (BANHO et al., 2021), caracterizaram a assimetria em cruzamentos interespecíficos entre as quatro subespécies de *D. mojavensis* e *D. arizonae*. Em particular, identificou-se que híbridos dos dois sexos descendentes do cruzamento entre fêmeas *D. m. baja*, procedentes de Cape Region (Baja California Sur, México) e machos *D. arizonae*, procedentes de Metztitlan (Hidalgo, México) (H♀moj^{baja}♂ari) são férteis e produzem descendentes quando cruzados com fêmeas híbridas, enquanto que, no sentido contrário de cruzamento (H♀ari♂moj^{baja}), os machos são estéreis, sendo as fêmeas férteis. A comparação entre híbridos férteis e estéreis, particular à descendência interespecífica da linhagem *D. m. baja*, permitiria a distinção de genes efetivamente envolvidos com a esterilidade híbrida e aqueles desregulados também em híbridos férteis provavelmente em função da evolução mais rápida de machos (GOMES; CIVETTA, 2014; 2015). Pouco ainda se sabe sobre os genes envolvidos na esterilidade híbrida dessas espécies, todavia, em soma ao sinal de seleção negativa relaxada em *OdsH* observada em *D. mojavensis* (dados não

publicados), já foi identificada a incompatibilidade entre os cromossomos X e Y dessas duas espécies (VIGNEAULT; ZOUROUS, 1986; ZOUROS et al., 1988). Dessa forma, a assimetria em relação à fertilidade do macho híbrido no par de espécies *D. mojavensis*-*D. arizonae* é uma condição estratégica para ampliar nosso conhecimento sobre a atividade do *OdsH* como gene de especiação. Por ser um gene localizado no cromossomo X, associado à esterilidade híbrida no complexo *melanogaster*, e até então não descrito em espécies de outros grupos de *Drosophila*, o gene *OdsH* torna-se novamente um interessante material de estudo para se ampliar o conhecimento da evolução molecular e funcional dos genes de especiação.

2 OBJETIVOS

2.1 Objetivo geral

Caracterizar a história evolutiva e a divergência funcional do gene *OdsH* e sua duplicata *unc-4* em Drosophilidae.

2.2 Objetivos específicos

1. Identificar duplicatas do gene *unc-4* em Drosophilidae em bancos de dados genômicos e caracterizá-las quanto à ortologia em relação ao gene *OdsH* descrito em *D. melanogaster*;
2. Datar a origem da duplicação *unc-4*/*OdsH*;
3. Estimar as taxas de evolução atuantes sobre as sequências nucleotídicas dos genes *unc-4* e de sua duplicata;
4. Caracterizar a expressão de *OdsH* e *unc-4*, com base nos bancos de dados de expressão disponíveis de *Drosophila* e *Teleopsis dalmanni* (Diptera, Diopsoidea), como grupo externo*;*
5. Identificar possíveis sítios de ligação de fatores de transcrição (TFBS) nas regiões reguladoras dos genes *unc-4* e *OdsH* em *Drosophila*;
6. Modelar a estrutura tridimensional proteica do homeodomínio das proteínas OdsH e Unc-4, em *Drosophila*, e da proteína Unc-4 em *Teleopsis dalmanni*, como grupo externo;
7. Caracterizar a expressão quantitativa e *in situ* dos genes *unc-4* e *OdsH* em tecidos de *D. arizonae* e *D. mojavensis baja*, e de seus híbridos recíprocos.

3 MATERIAL E MÉTODOS

3.1 Anotação de *unc-4* e *OdsH* em genomas de Drosophilidae

As sequências de *unc-4* e sua duplicata foram recuperadas dos genomas públicos e anotados de Drosophilidae no NCBI, com BLAST (NCBI), usando como *query* as sequências RNAm *unc-4* (NM_001298461.1) e *OdsH* (NM_078665.3) de *D. melanogaster*. As sequências com maior *score* e *e-value* < 1e-05 para cada duplicata foram coletadas (Tabela Suplementar 1 – APÊNDICE B) e alinhadas com MAFFT (KATOH et al., 2002). Os alinhamentos foram inspecionados manualmente com BioEdit Sequence Alignment Editor v. 7.0.9 (HALL, 1999) e as regiões não alinháveis e não homólogas com *indels* foram removidas. A região conservada alinhável entre as duplicatas contém o homeodomínio (162 pb), com 15 pb a montante da extremidade N-terminal do homeodomínio e 228 pb a jusante da extremidade C-terminal.

Para a busca de regiões homólogas nos genomas que não possuem suas respectivas anotações (Tabela Suplementar 2 – APÊNDICE B) foram utilizadas como *query* a região conservada entre as duplicatas das espécies mais próximas filogeneticamente de cada genoma inspecionado. Tais buscas foram baseadas no maior *bistcore* e *e-value* < 1e-0.5, por meio de um *script* em linguagem BASH escrito por nosso grupo de pesquisa.

As sequências de *Chymomyza amoena*, *Colocasiomyia xenalocasiae*, *Cacoxenus indagator* e *Rhinoleucophenga bivisualis* foram anotadas a partir de genomas montados com o software SPADES v. 3.9.0 ou v. 3.11.1 (BANKEVICH et al., 2012). Esses genomas foram sequenciados conforme descrito por Dias et al. (2020). Para anotar as sequências gênicas, as sequências de aminoácidos das proteínas OdsH e Unc-4 de *D. melanogaster* foram utilizadas como *query*, em buscas TBLASTN nos genomas montados para encontrar os *scaffolds* com as sequências dos dois genes. Por fim, nesses *scaffolds*, o programa Genewise (BIRNEY et al., 2004) foi usado para identificar as sequências codificantes.

3.2 Inferência filogenética e datação da duplicação

O alinhamento da região conservada entre as sequências dos dois genes foi usado na Inferência Bayesiana das relações filogenéticas, realizada a partir de cadeias convergentes de MCMC (Métodos de Monte Carlo via Cadeias de Markov) com comprimento de 10.000 amostragens a cada 1.000 cadeias (DRUMMOND et al., 2012), no programa BEAST v1.10.4

(SUCHARD et al., 2018). Para tanto, foi utilizado o modelo de substituição GTR (*General Time Reversible*) (NEI; KUMAR, 2000), com 5 categorias e distribuição gama e sítios invariáveis, estimado como o melhor modelo de substituição pelo critério bayesiano BIC (POSADA, 2008), no programa MEGA X (KUMAR et al., 2018). Estimativas de árvores com alta correlação de amostras foram visualizadas nas primeiras 1.000 árvores, no Tracer v.1.7.2 (RAMBAUT et al., 2018), as quais foram utilizadas como *burn-in* para reconstruir a topologia da árvore, no TreeAnnotator (DRUMMOND et al., 2006), que estimou uma árvore consenso com melhor probabilidade posterior computada para cada nó. A árvore foi editada com o *software* Figtree 1.4 (RAMBAUT, 2009). Para a coestimativa do tempo de divergência com a topologia da árvore consenso, foi utilizado o método Bayesiano de relógio molecular relaxado lognormal, que permite a variação de taxas de evolução ao longo dos ramos (DRUMMOND et al., 2006), com processo de Yule como modelo especiação, que permite variação de taxas de especiação e extinção ao longo dos ramos (GENHARD et al., 2008), já que se tratam de sequências provenientes de linhagens evolutivas distantes da família Drosophilidae.

Uma vez que não foram encontradas evidências que suportem a presença de duplicatas *in tandem* em táxons externos à Drosophilinae, indicou-se a monofilia entre as sequências *unc-4* e *OdsH* desse táxon. Esse método foi usado de modo a evitar o artefato filogenético *long branch atraction* (LBA) (FELSENSTEIN, 1978; HENDY; PENNY, 1989) que pode causar viés na inferência de origem e artefatos na inferência da história evolutiva dos genes (FARES et al., 2006; HAHN, 2007), sob a hipótese de que *unc-4* e *OdsH* evoluem a taxas diferentes em relação ao grupo externo *unc-4* cópia única de Steganinae. Esse artefato já foi demonstrado também em metodologias heurísticas de inferência filogenética para genes parálogos que apresentam evolução assimétrica em *Drosophila* (BAO; FRIEDRICH, 2009). Esse contexto poderia acarretar a atração do ramo do gene mais conservado ao grupo externo, e repulsão do clado da duplicata mais divergente (LBA), como foi identificado em análises anteriores (dados não publicados).

Para datar o evento de duplicação que originou a duplicata *OdsH* foram utilizadas as estimativas do tempo de divergência *Sophophora/Drosophila* (39,2 ± 3,35 M.a) e *D. melanogaster/D. simulans* (2,3 ± 0,65 M.a), de Russo et al. (1995), como ponto de calibração para ancorar o relógio molecular. A calibração foi realizada com a informação *a priori* do tempo de divergência dos mesmos clados para as duas duplicatas. Utilizou-se dessa abordagem na calibragem de modo a reduzir os efeitos resultantes da assimetria nas taxas de substituição observada entre duplicatas (ZHAXYBAYEVAL, 2013).

3.3 Viés de uso de códons

Levando-se em consideração que o viés no uso de codons pode resultar em artefatos filogenéticos em árvores de genes (INAGAKI et al., 2004, INAGAKI; ROGER, 2006; LIU et al, 2014), foi estimado o *relative synonymous codon usage* (RSCU), por grupo e por gene, no MEGA X (KUMAR et al., 2018) devido a diferenças no uso de códons nas radiações *saltans* e *willistoni* em comparação a outros grupos *Drosophila*, com maior presença de códons terminados em AT do que as outras linhagens de *Drosophila* (RODRÍGUEZ-TRELLES; AYALA, 2000; POWELL et al., 2003; VICARIO et al., 2004; SINGH et al., 2006), como também pelo posicionamento filogenético de espécies dos grupos *willistoni* e *saltans* ser comumente artefatual em árvores de genes, decorrente do menor conteúdo GC observado em suas sequências (PÉLANDAKIS; SOLIGNAC, 1993; TARRÍO et al., 1999; GAILEY et al., 2000), A ferramenta CAIcal (PUIGBÒ et al., 2008) foi usada para calcular o número efetivo de códons (ENC) e a proporção GC na terceira posição (% GC3). As diferenças entre esses parâmetros foram testadas pelo teste ANOVA *one way,* com cálculo do valor p pelo teste de Tuckey.

3.4 Taxa relativa de substituição nucleotídica

Para identificar se os genes *unc-4* e a sua duplicata evoluem a taxas diferentes, foi usado o Teste de Taxa Relativa, aplicado no *software* PHYLTEST 2.0 (KUMAR, 1996), utilizando-se como grupos externos as sequências do gene *unc-4* coletadas nas demais espécies de Steganinae. Foi utilizado o melhor modelo de substituição disponível no programa, o Kimura-2-Parâmetros (KIMURA, 1980).

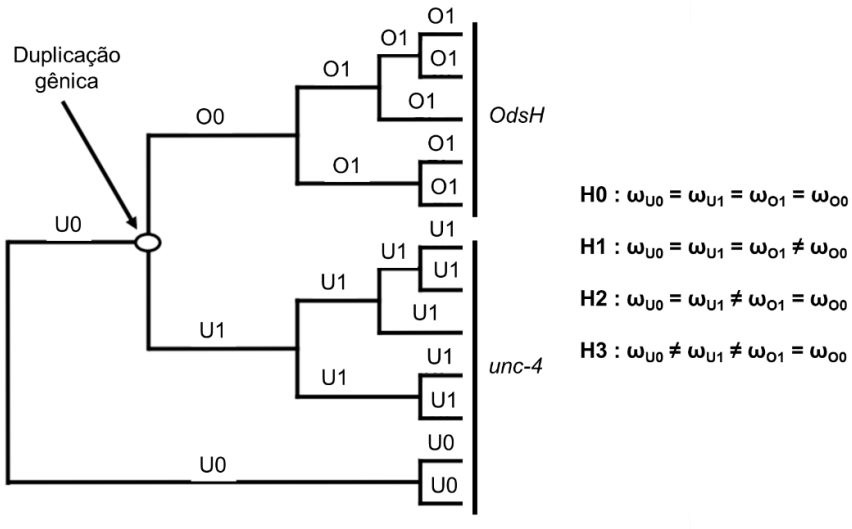
3.5 Estimativas de seleção natural e investigação de seleção positiva

Para caracterizar o tipo de seleção atuante sobre as sequências *unc-4* e *OdsH*, foram utilizados métodos de verossimilhança baseados em códons usando o programa CODEML, do pacote PAML ver 4.9 (YANG, 2000). As estimativas de pressão seletiva foram calculadas pela razão ômega ($\omega = Ka/Ks$) entre as taxas de substituições nucleotídicas não sinônimas (Ka) e sinônimas (Ks). Para essas análises foram usadas duas ávores em formato *newick*, uma delas (*Árvore 1*) a descrita acima (utilizada para as estimativas da datação) obtida com o alinhamento de *unc-4* e *OdsH*. Uma vez que apenas as sequências *OdsH* apresentaram substituições não

sinônimas em *Drosophila*, testes de seleção foram também conduzidos considerando-se apenas esse gene, sendo sua árvore reconstruída, doravante denominada *Árvore 2*, também por Inferência Bayesiana, com os mesmos *priors* utilizados para *Árvore 1*.

Para as análises, tanto da *Árvore 1*, como da *Árvore 2*, foi utilizado o modelo de seleção para ramos, que permite variar ω na filogenia, e é apropriado para a detecção de seleção positiva agindo sobre linhagens particulares (YANG, 1998; YANG; NIELSEN, 1998). Para isso, utilizou-se o modelo de ramos tipo 2, que permite, a partir da marcação de ramos no arquivo da árvore, e testar se esse ramo apresenta razão ω diferente em relação ao *background*. Na *Árvore 1* foram realizadas três marcações: (1) no nó ancestral *OdsH*; (2) nos nós da divergência *OdsH* e (3) nos nós ancestrais *OdsH* e *unc-4* (Figura 5). Na *Árvore 2* foram realizadas marcações nos ramos dos grupos de *Drosophila* que possuem três ou mais representantes e do gênero *Scaptomyza*, de modo a identificar se o gene *OdsH* possui dinâmica evolutiva particular nos diferentes agrupamentos filogenéticos. Dentre essas marcações, encontram-se grupos de espécies que apresentaram sinais de seleção positiva em análises par a par (*D. simulans, D. sechellia* e *D. mauritiana*) desenvolvidas por Ting et al. (1998) e seleção negativa relaxada (ω>0,3) em análises anteriormente por nós desenvolvidas (*D. navojoa* e *D. mojavensis*) (dados não publicados).

Figura 5. – Esquema dos testes de hipóteses desenvolvidos para análises de assinaturas de seleção entre *OdsH* e *unc-4*.



U: *unc-4*; O: *OdsH*; H: hipótese.

Também foi utilizado o programa CODEML, com o uso dos diferentes modelos de seleção para sítios, para identificar possíveis sítios sob seleção positiva, utilizando-se as

seguintes hipóteses: M1 – modelo neutro, ausência de códons sob seleção positiva; M2 - presença de códons sob pressão seletiva diversificadora; M7 – códons distribuídos em categorias em distribuição β de valor ω, sem a presença de sítios sob seleção positiva; M8 - códons distribuídos em categorias em distribuição β de valor ω, com a presença de sítios sob seleção positiva. As análises foram realizadas com as sequências de *OdsH* de Drosophilinae, e, também, somente do complexo *melanogaster*, uma vez que este apresentou assinatura de relaxamento de seleção negativa na análise anterior. Todas as hipóteses assumidas para identificar a razão ω foram testadas com o teste de χ^2, pela comparação dos valores lnL de cada hipótese com a hipótese nula (2ΔlnL), com graus de liberdade pela diferença dos números de parâmetros entre as hipóteses.

3.6 Expressão *in silico* de *unc-4* e *OdsH*

A expressão de *OdsH* e *unc-4* foi inspecionada manualmente *in silico* por meio da ferramenta *Tracks*, da plataforma *Gene*, do NCBI. As espécies da subfamília Drosophilinae que possuem transcriptomas de órgãos reprodutivos e não reprodutivos foram analisadas separadamente por sexo. O mesmo procedimento foi realizado para o gene *unc-4* cópia única na mosca *Teleopsis dalmanni* (Diptera, Diopsoidea), espécie mais próxima filogeneticamente à Drosophilidae, com transcriptomas disponíveis, externa à duplicação. Os genes foram então categorizados, por espécie e por tecido, como expressos, caso apresentassem mais de 10 *counts* no histograma de visualização dos éxons *OdsH* e *unc-4*, ou não expressos.

3.7 Sítios de ligação de fatores de transcrição nas regiões reguladoras de *unc-4* e *OdsH*

Para investigar a presença diferencial de TFBSs nas regiões reguladoras dos genes *unc-4* e *OdsH* foram extraídas as sequências 500pb a montante e a jusante desses genes nos genomas das espécies de *Drosophila* que tiveram seus transcriptomas investigados quanto à expressão de *unc-4* e *OdsH*. Foram incluídas também as sequências das espécies *D. sechellia*, *D. simulans* e *D. mauritiana* que tem o híbrido estéril afetado pelo *OdsH* (TING et al., 1998). As sequências reguladoras de *OdsH* foram submetidas à análise de enriquecimento do CiiiDER (GEARING et al., 2019), utilizado-se as sequências de *unc-4* como *background*, com o *déficit threshold default* (0,15) e limite de *p-value* de Fisher de 0,05, para identificação de TFBSs diferencialmente enriquecidos entre *OdsH* e *unc-4*. Os fatores de transcrição que apresentaram enriquecimento diferencial de sítios de ligação entre *OdsH* e *unc-4*, foram submetidos à análise

de enriquecimento de termos GO (ASHBURNER et al., 2000; MI et al., 2019), nas categorias de processo biológico, e as redundâncias foram removidas em REVIGO (SUPEK et al., 2011).

3.8 Motivos funcionais proteicos

Os homeodomínios e o octapeptídeo de Drosophilinae foram identificados em Unc-4 e OdsH separadamente, usando MEME (BAILEY; ELKAN, 1994) na plataforma MEME Suite (BAILEY et al., 2015). A fim de se observar um perfil amplo da diversidade nos homeodomínios OdsH e Unc-4 em Drosophilinae, eles foram identificados a partir da tradução de todas as sequências nucleotídicas alinháveis recuperadas em Drosophilinae. Já o octapeptídeo foi identificado a partir do alinhamento dos últimos 11 aminoácidos da extremidade C-terminal das proteínas OdsH e Unc-4 anotados disponíveis no NCBI (Tabela Suplementar 1 – APÊNDICE B).

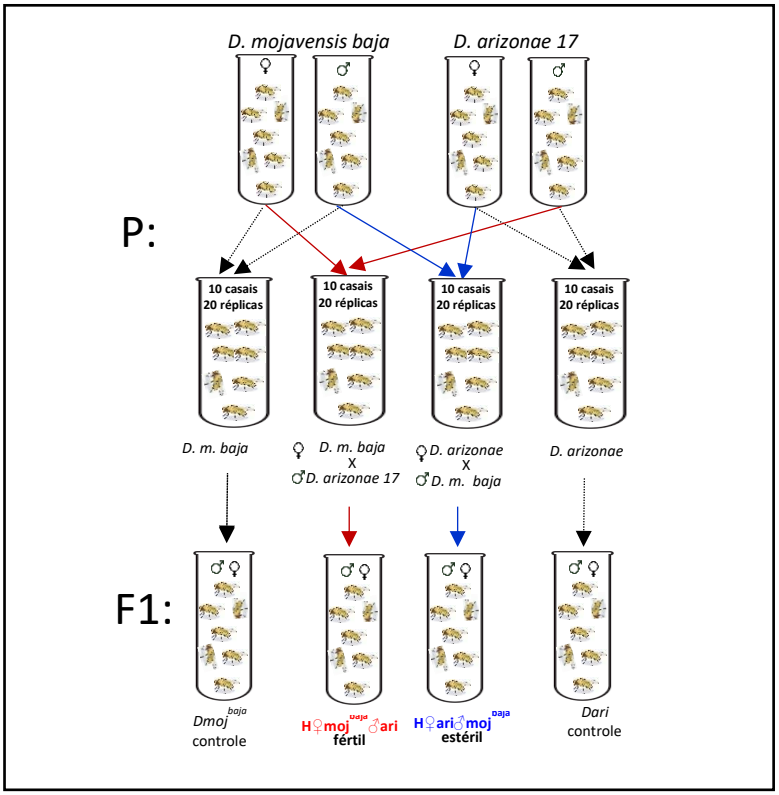
A estabilidade dos modelos tridimensionais dos homeodomínios Unc-4 e OdsH associados à dupla fita de DNA foi avaliada de modo a inferir se a divergência de suas sequências acarretaria divergência funcional em termos biofísicos. Foi então desenvolvida a modelagem proteica dos homeodomínios Unc-4 e OdsH, de *D. melanogaster* (NP_573242.2 e NP_523389.3) e da Unc-4 da mosca *T. dalmanni* (XP_037943702.1), como grupo externo à duplicação, utilizando SwissModel (WATERHOUSE et al., 2018) e a estrutura do PDB 3LNQ (MIYAZONO et al., 2010) como *template.*

Assim, após o cálculo de modelagem, o complexo formado pelo modelo estrutural Unc-4 de *T. dalmanni* e a estrutura de DNA foi minimizado via cálculos de dinâmicas moleculares usando o programa GROMACS (ABRAHAM et al., 2015) com o pacote AMBER14-OL15, incluindo o campo de força ff14sb (MAIER et al., 2015) e ff99bsc0χOL3+OL15 para proteína e ácido nucleico (ZGARBOVÁ et al., 2015), respectivamente, e modelo de água TIP3Pl (JORGENSEN et al., 1983). Para isso, o sistema molecular foi inserido em uma caixa cúbica solvatada por uma solução de NaCl 100 mM em água. A minimização de energia foi realizada com o *steepest descent integrator* e o algoritmo de gradiente conjugado, usando 500 kJ·mol–1·nm–1 como critério de força máxima. Foram calculados os valores da perturbação da variação da energia livre de ligação (ΔΔGb) acarretadas pelas substituições observadas em OdsH de *Drosophila* (espécies utilizadas nas análises funcionais anteriores) na estabilidade do complexo homeodomínio e DNA foram calculados usando o servidor mCSM (PIRES et al., 2014), em relação à estrutura do complexo homeodomínio Unc-4/DNA dupla fita.

3.9 Análises experimentais da expressão quantitativa e *in situ* de *OdsH* e *unc-4* em *D. m. baja* e *D. arizonae,* e seus híbridos recíprocos

Foram realizados cruzamentos intraespecíficos e interespecíficos entre *D. arizonae* (linhagem 17), de Metztitlan, Hidalgo, México (Stock Center n°: 15081-1271.17) e *D. m. baja* (linhagem 20), oriunda de Cape Region, Santiago, Baja California Sur, México, Stock Center n°: 15081-1352.20). A linhagem *D. m. baja* 20 possui cor de olhos marrons, uma característica autossômica recessiva que é usada para se comprovar o estado híbrido nos cruzamentos F_1 com *D. arizonae*, que possui olhos vermelhos, caráter selvagem, com herança autossômica dominante. Fêmeas e machos virgens foram coletados em 48 h após a eclosão e mantidos isolados em tubos com meio de cultura até atingirem a maturidade sexual (3 dias), sendo então utilizados em intra e intercruzamentos recíprocos (Figura 6). Para isso, 10 casais de cada tipo de cruzamento foram colocados em tubos de 2,3 × 9,5 cm, em 35 réplicas, contendo meio de cultura *Drosophila* à base de cacto suplementado com levedura sendo neles mantidos até 9 a 12 dias de idade, quando foram descartados. Com a emergência, moscas virgens (híbridos F_1 e controle) foram coletadas em 48 h após a eclosão e mantidas isoladas em tubos com meio de cultura até 9-12 dias de idade. Durante a coleta foi verificada a cor dos olhos de todos os descendentes, sendo o estado híbrido de machos e fêmeas comprovado pelos olhos vermelhos nos dois sentidos de cruzamentos. Esses indivíduos machos tiveram seu trato reprodutivo e carcaça dissecados, separados por sexo, em tampão PBS 1X, e esses tecidos foram utilizados em análises de expressão dos genes *OdsH* e *unc-4*.

Figura 6. Cruzamentos entre *D. m. baja* e *D. arizonae.*



3.9.1 Extração de RNA e RT-qPCRC

Foi realizada extração do RNA total dos tratos reprodutivos masculino e feminino e carcaças de *D. m. baja* e *D. arizonae* e seus híbridos recíprocos envelhecidos até 9-12 dias de idade, em três réplicas biológicas de 30 indivíduos de cada cruzamento (controles e interespecíficos). As extrações de RNA foram realizadas com o kit RNeasy (Qiagen), em seguida as amostras foram tratadas com DNase (DNA-free kit; Ambion), quantificadas em espectofotômetro (Nanodrop) e armazenadas a -80^{o}C. A conversão de 1000ng de RNA em cDNA foi realizada usando o *kit High Capacity cDNA Reverse Transcription* (Thermo Fisher). O nível relativo de mRNA foi quantificado usando ensaios *TaqMan* em um instrumento *LightCycler* 480 (Roche Diagnostics).

Para quantificação da expressão dos genes *unc-4* e *OdsH* foram desenhados oligonucleotídeos iniciadores e sondas fluorescentes contendo sequências específicas do homeodomínio de cada uma delas (Tabela 1) de modo a garantir a presença nos *amplicons* apenas de cDNA dos genes de interesse (Taqman Thermo Fischer Scientific). Os experimentos

de RT-qPCR foram realizados em três réplicas técnicas, usando 100ng de cDNA. Os níveis de expressão gênica foram medidos em relação à expressão do gene ribossomal endógeno 49 (rp49), também conhecido como asnrpL.

Tabela 1. Sequências de oligonucleotídeos utilizados nos ensaios de q-PCR.

Gene	***unc-4***	***OdsH***	***rp49***
Primer forward	AGCGAGGGCTCCAATGAG	AGCCGCAGAGCTGCA	CCCAACATTGGTTACGGTTCCA
Primer reverse	TGCCACGAGTTGAAGTTGGT	GCTCGATCGCCTTGGCTAT	GCACATTGTGTACGAGGAATTTCTT
Sonda	ATGGCGGCTTCCCC	CTGCAGGAGCTGCGAGCCA	CACCCGCCACATGCT

Os valores de Ct resultantes da RT-qPCR foram submetidos ao cálculo de média entre as réplicas técnicas e normalizados com os valores obtidos para o gene constitutivo de referência (*rp49*), com a razão RQ (RQ = $E^{Ct\ rp49}/E^{Ct\ alvo}$; E = eficiência da reação). Inicialmente, os valores de RQ dos genes *unc-4* e *OdsH* foram submetidos ao teste de Shapiro-Wilk, para a avaliação da normalidade das réplicas, e ao teste de Bartlett, para observação da homogeneidade de variâncias das réplicas entre as amostras. Uma vez identificadas a distribuição normal dos valores e a homogeneidade de variâncias, foram então calculadas as variâncias por meio da Análise de Variância um fator (ANOVA *one way*), com o cálculo do valor P pelo Teste de Tukey, para identificação de quais grupos (híbridos e controles) diferiam significativamente. Os testes estatísticos foram realizados no programa BioEstat 5 ®.

3.9.2 Detecção de transcritos *OdsH* em testículos por hibridização fluorescente *in situ* (FISH)

Testículos dissecados em PBS1X foram fixados em PBS1X contendo 3% Tween 10 e 25% formaldeído 16. A hibridização (*overnight*) se deu a 37ºC. Para a produção das sondas, 36 oligonucleotídeos (20 nt cada) foram desenhados em Stellaris® Sonda Designer version 4.2 (https://www.biosearchtech.com/stellaris-designer) e sintetizados e marcados com fluoróforo com emissão de comprimentos de onda de 550nm (Tabela Suplementar 3). Para a marcação do DNA, os testículos foram incubados em PBS1X contendo DAPI (ThermoFischer Scientific). As lâminas foram visualizadas em Microscópio Confocal Upright Zeiss LSM780-NLO.

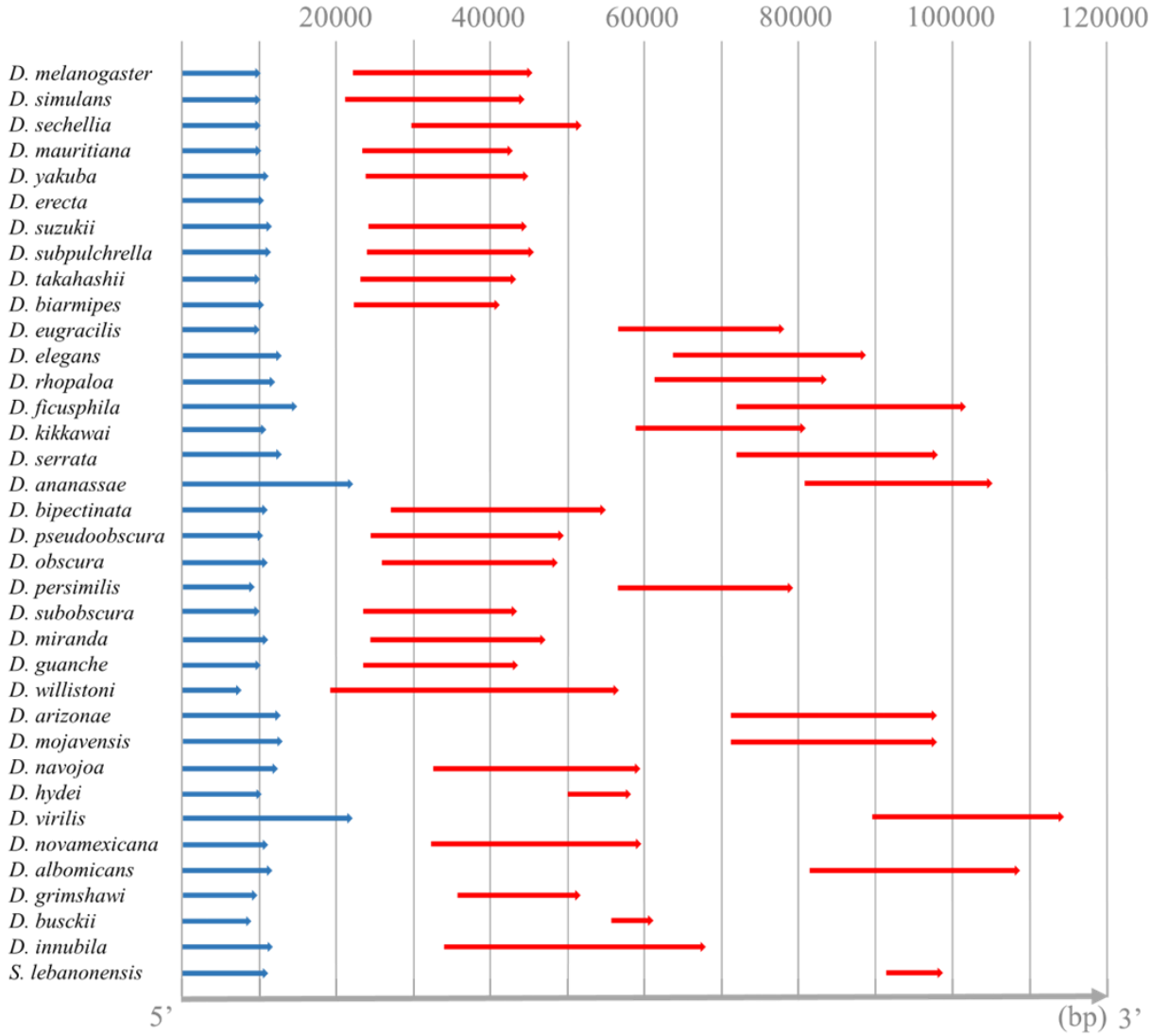
4 RESULTADOS

4.1 Ocorrência de *unc-4* e *OdsH* em genomas de Drosophilidae e relações filogenéticas

A busca por sequências do gene *unc-4* e sua duplicata em 35 genomas anotados resultou na obtenção de sequências em todas as espécies do gênero *Drosophila* (subgêneros *Sophophora* [23], *Drosophila* [10] e *Dorsilopha* [1]), *D. grimshawi* e *Scaptodrosophila lebanonensis* (Tabela Suplementar 1 – APÊNDICE B), à exceção de *D. erecta*. Apenas *D. melanogaster* e *D. sechellia* apresentaram sua duplicata predita como *OdsH*, sendo nas demais espécies anotada como *unc-4* ou *unc-4*-like, à exceção da duplicata do gene *unc-4* de *D. willistoni*, que foi predita como homólogo do gene homeobox pituitário *Ptx1* (LOC6649333), e de *D. albomicans* como *segmentation polarity homeobox protein engrailed* (LOC117569146). Sequências homólogas a *unc-4* e a *OdsH* também foram obtidas nos grupos de Drosophilinae subrrepresentados nos genomas anotados (gêneros *Drosophila* [24] (subgêneros *Sophophora* [12] e *Drosophila* [12]), *Drosophila* "havaianas" [2], *Zaprionus* [3], *Lordiphosa* [5], *Scaptomyza* [4] e *Chymomyza* [2]). *Drosophila sucinea, Lordiphosa stackelbergi, Colocasiomyia xenalocasiae* (subfamília Drosophilinae) e as espécies investigadas da subfamília Steganinae (4) retornaram apenas a sequência de *unc-4*, sendo as últimas classificadas como grupo externo (Tabela Suplementar 2 – APÊNDICE B).

As duplicatas se encontram localizadas vizinhas em relação ao gene *unc-4* nos genomas de todas as espécies (Figura 7). A distância entre as duplicatas variou entre 10.982 pb (*D. simulans*) e 80.454 pb (*Scaptodrosophila lebanonensis*). O tamanho de *OdsH* variou entre 5.195pb (*D. busckii*) e 37.364 pb (*D. willistoni*). Já o tamanho de *unc-4* variou entre 7.801pb (*D. willistoni*) e 21.691 pb (*D. virilis*).

Figura 7. Localização relativa de *unc-4* e *OdsH* em genomas de Drosophilinae.



unc-4: azul; *OdsH*: vermelho; comprimento da seta: comprimento relativo das sequências gênicas em pares de base (pb); direção da seta: sentido de transcrição. e posição relativa dos genes *unc-4* (azul) e *OdsH* (vermelho) em Drosophilinae, em pares de bases (pb).

Todas as sequências *unc-4* e *OdsH* usadas para inferir as relações filogenéticas entre os dois genes se segregaram em dois clados monofiléticos irmãos, corroborando as relações de ortologia entre as sequências *OdsH* obtidas e *OdsH* predito de *D. melanogaster* e *D. sechellia*, e de paralogia destas em relação à duplicata *unc-4*. As sequências da radiação *willistoni-saltans-Lordiphosa* formaram agrupamentos monofiléticos robustos coalescentes aos nós ancestrais tanto no clado de *unc-4*, como no de *OdsH*, assumindo, contudo, um posicionamento incongruente à história evolutiva de Drosophilinae, na base da divergência Drosophilini para *unc-4* e do subgênero *Drosophila*, para *OdsH*, mas sem suporte na inferência filogenética para ambos clados (Figura Suplementar 1 – APÊNDICE A). Essa incongruência pode ser devida ao

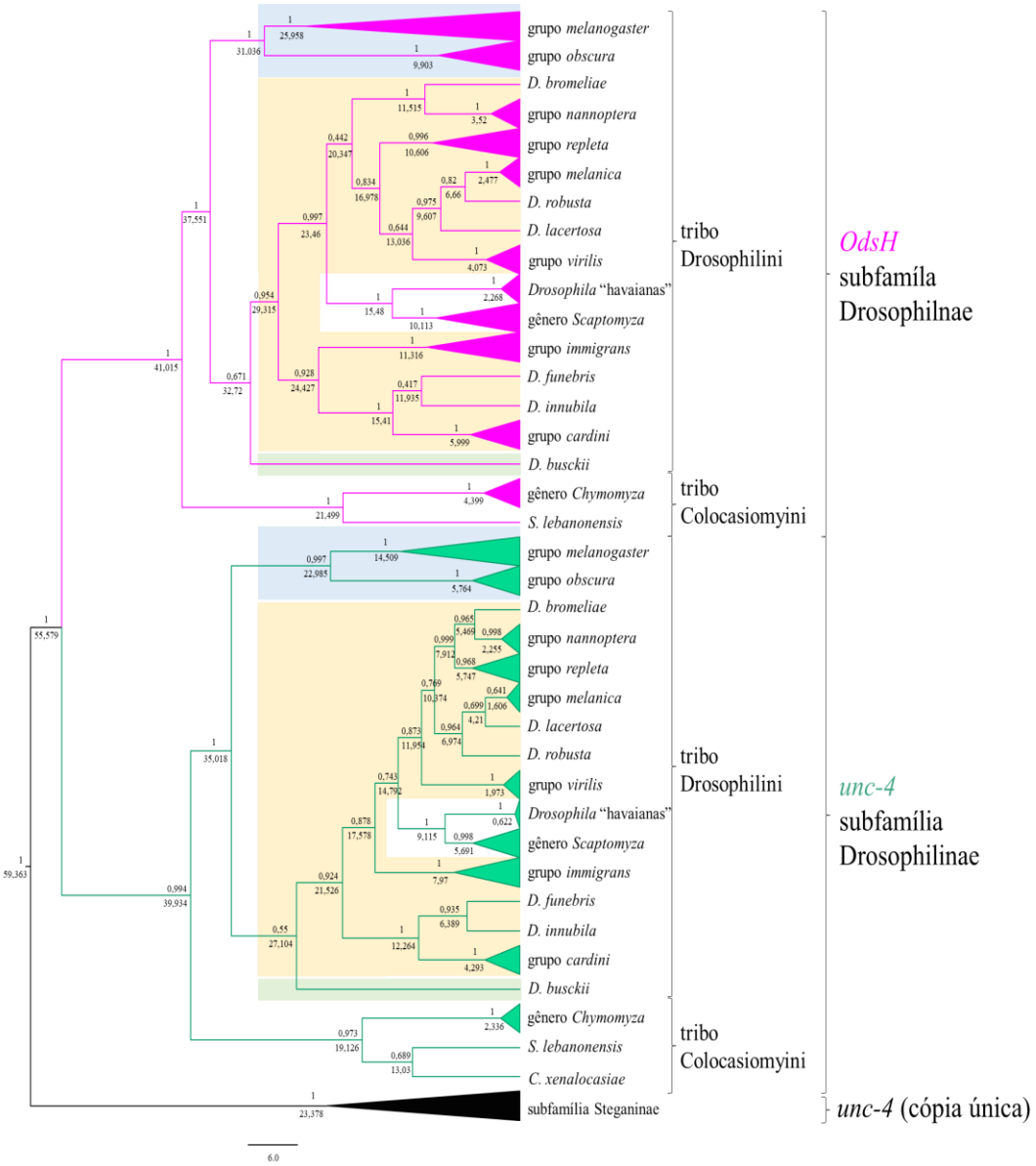
uso diferencial de códons dessa linhagem em relação às demais *Drosophila*, uso esse já reportado na literatura para espécies dos grupos *willistoni* e *saltans* (RODRÍGUEZ-TRELLES; AYALA, 2000; POWELL et al., 2003; VICARIO et al., 2004; SINGH et al., 2006).

A análise realizada para testar o uso diferencial de códons permite observar um padrão diferente no uso de códons nas sequências de *unc-4* em espécies dos grupos *willistoni, saltans, Lordiphosa* e *tumiditarsus* da subfamília Drosophilinae, bem como em Steganinae (grupo externo – *unc-4*) (Figura Suplementar 3 – APÊNDICE A). O mesmo foi observado nas sequências de *OdsH* nos grupos em *willistoni, saltans* e Steganinae (Figura Suplementar 4 – APÊNDICE A). Todavia, análise estatística não indicou diferenças significativas (ANOVA, $F = 0$; $p > 1$ [*unc-4*]; $F = 0{,}001$; $p = 1$ [*OdsH*]). Por outro lado, foram observados maiores valores de ENC e menores de % GC3, tanto em *unc-4* (ANOVA, ENC: $F = 14{,}486$; $p < 0{,}0001$; %GC3: $F = 82{,}574$; $p < 0{,}0001$) (Figura Suplementar 5 – APÊNDICE A), quanto em *OdsH* (ANOVA, ENC: $F = 13{,}672$; $p < 0{,}0001$; %GC3: $F = 68{,}446$; $p < 0.01$) (Figura Suplementar 6 – APÊNDICE A), na radiação *willistoni* e *saltans* (Tukey, ENC: $Q = 6{,}184$; $p < 0{,}01$ [*unc-4*]; 5,627; $p < 0{,}01$ [*OdsH*]; % GC3: $Q = 13{,}471$; $p < 0{,}01$ [*unc-4*]; $F = 14{,}61$; $p < 0{,}01$ [*OdsH*]) e em Steganinae (Tukey, ENC: $Q = 5.106$; $p < 0{,}01$; % GC3: $Q = 13.397$; $p < 0{,}01$), em relação às demais Drosophilinae. Sabendo-se que diferenças no uso de códons podem ocasionar artefatows filogenéticos (INAGAKI et al., 2004, INAGAKI; ROGER, 2006; LIU et al, 2014) optou-se por remover essas sequências das análises filogenéticas realizadas para a datação do evento de duplicação e análise de seleção posteriores. Em adição também foram removidas as sequências de grupos de espécies que não se agruparam na filogenia, em relação aos subgêneros de *Drosophila*, de acordo com o descrito em estudos anteriores de modo também a se evitar viéses para as análises posteriores de datação e seleção. São eles: *Zaprionus,* cujas sequências *unc-4* se agruparam a *Sophophora* e *OdsH* a *Drosophila*, e *tumiditarsus* (*D. repletoides*), que no clado *unc-4* se agrupou a *Dorsilopha*, apresentando para esse gene padrão semelhante à RSCU das espécies dos grupos *willistoni*, *saltans*, *Lordiphosa* e Steganinae.

A Inferência Bayesiana para coestimar a topologia da árvore e o tempo de divergência entre as sequências das duplicatas *unc-4* e *OdsH* de Drosophilinae, também confirmou a sua monofilia, reconstruindo em clados irmãos, e de modo geral, as relações entre os gêneros, subgêneros e grupos de espécies de Drosophilinae reportados na literatura (Figura 8, Figura Suplementar 2). O nó compartilhado por esses dois clados, que representa o evento de duplicação, enraizado pelo clado *unc-4* cópia única de Steganinae, foi estimado em 55,58 M.a.

O clado *OdsH* apresenta ramos mais longos que o *unc-4*, com idades dos táxons mais antigas, um artefato devido à maior divergência de suas sequências, em relação a *unc-4*. Todavia, os dois clados se apresentam congruentes com a monofilia das tribos Drosophilini e Colocasiomyini, do subgênero *Sophophora*, posicionado basalmente na tribo Drosophilini, e do agrupamento dos subgêneros *Drosophila-Dorsilopha*. As sequências também reconstruíram os grupos de espécies do gênero *Drosophila* e as suas relações filogenéticas são concordantes entre os ramos, salvo aos grupos *bromeliae*, *nannoptera*, *repleta*, *virilis*, *robusta* e *melanica*, que se agruparam, porém com posição relativa diferente entre os dois ramos, contudo, as relações entre esses grupos também não permanecem preservadas nos diferentes estudos (revisado em O'GRADY; DESALLE, 2018). Uma vez que essas incongruências não se encontravam na base da divergência dos subgêneros *Drosophila*, e, portanto, não acarretariam vieses na calibração da árvore, que foi realizada nesse nó para a estimativa do tempo de divergência das duplicatas, elas foram mantidas nas análises subsequentes.

Figura 8. Inferência filogenética Bayesiana das sequências dos genes parálogos *unc-4* e *OdsH* e datação do evento de duplicação *unc-4*/*OdsH*.



A inferência foi realizada utilizando usando o modelo de substituição GTR+G+I. A análise foi realizada com 405 sítios nucleotídicos de 124 sequências nucleotídicas. Todas as posições contendo *gaps* e posições ambíguas foram retiradas da análise par a par de sequências. Na raiz de cada clado estão indicados os valores de probabilidade posterior (acima) e estimativa do tempo de divergência (abaixo). Os pontos de calibração utilizados a datação da duplicação foram as estimativas do tempo de divergência *Sophophora/Drosophila* (39.2 ± 3.35 Ma) e *D. melanogaster/D. simulans* (2,3 ± 0,65 Ma). A análise foi conduzida no programa BEAST v1.10.4. O clado *unc-4* (verde), subdividido em cópia única mais basal de Steganinae (preto) e Drosophilinae, é apresentado na base da filogenia, seguido por *OdsH* na parte superior (rosa). O subgênero *Sophophora* está circunscrito pela caixa azul, *Drosophila* amarelo e *Dorsilopha* verde. Os grupos taxonômicos monofiléticos foram comprimidos. Os clados descomprimidos podem ser visualizados na Fig. Sup. 2. As chaves indicam as tribos e subfamílias de Drosophilidae.

4.2 Dinâmica evolutiva

O teste de taxa de evolução relativa indicou que as duplicatas evoluem a diferentes taxas relativas ao grupo externo (L_{OdsH} = 0,159; $L_{unc\text{-}4}$ = 0,082; Z = 4,523, p<0,05), observando-se aceleração da taxa de substituições nucleotídicas na evolução de *OdsH*.

Observam-se evidências de seleção negativa na evolução dos dois genes (ω<1), entretanto os valores diferem significantemente (χ^2 = 10,266, p = 0,001), sendo que o valor de ω encontrado para *OdsH* (ω = 0,196) se mostra cerca de quatro vezes maior que o de *unc-4* (ω = 0,052), e ainda maior se consideramos o ω de *unc-4* somente de Drosophilinae, que é zero, já que não há substituições não-sinônimas entre suas sequências (o valor ω = 0,052 deve-se a uma única substituição não sinônima que ocorre entre essas sequências e a do grupo externo *R. bivisualis* -T118Q). As marcações no nó ancestral *OdsH*, e ao longo da divergência *OdsH*, apresentaram valores ω maiores que o do *background*, respectivamente ω = 0,204 (χ^2 = 10,092, p = 0,001) e ω = 0,049 (χ^2 = 113,817, p = 0), indicando maior divergência observada em relação *unc-4* no nó ancestral *OdsH*, do que ao longo da divergência *OdsH*.

Como não foram observadas substituições não sinônimas nas sequências da duplicata *unc-4* de Drosophilinae não foi realizada análise de seleção por ramos para esse gene. Já, quanto à seleção atuante sobre *OdsH*, observa-se de forma geral, que não há diferenças entre os grupos de espécies de Drosophilinae (Tabela 2), exceto para as sequências de *Drosophila* "havaianas" (ω = 0,173; χ^2 = 4,044, p = 0,044) e grupo *melanogaster* (ω = 0,023, χ^2 = 4,25, p = 0,039). Apesar do grupo *melanogaster* apresentar um valor ω inferior ao das sequências dos demais grupos, a intensidade de seleção negativa se mostra relaxada para o clado *D. simulans-D. sechellia-D. mauritiana* (ω = 0,611; χ^2 = 30,84, p = 0). Em adição a essas espécies, marcações foram feitas no ramo de *D. navojoa*, que apresentou ω = 0,248, e de *D. mojavensis*, ω = 0,458, cujas diferenças não foram significantes em relação ao *background* (respectivamente χ^2 = 1,031, p = 0,31; e χ^2 = 3,352, p = 0,067), embora *D. mojavensis* tenha apresentado ω indicativo de seleção relaxada. Nenhuma assinatura de seleção positiva foi encontrada em *OdsH* (Tabela Suplementar 4 – APÊNDICE B).

Tabela 2. Valores de ω obtidos por meio de teste de seleção por ramos na filogenia de *OdsH* por grupos de Drosophilinae.

Grupo	ω_0	ω_1	lnL	np	x^2	df	P
M0	0,051		-5315,46	118			
melanogaster	**0,053**	**0,023**	**-5313,336**	**119**	**4,250**	**1**	**0,039**
D. simulans* - *D. mauritiana* - *D. sechellia	**0,047**	**0,611**	**-5300,040**	**119**	**30,843**	**1**	**0**
obscura	0,051	0,052	-5315,461	119	0,001	1	1,000
repleta	0,051	0,053	-5315,461	119	0,001	1	1,000
D. navojoa	0,051	0,248	-5314,946	119	1,031	1	0,310
D. mojavensis	0,051	0,458	-5313,785	119	3,351	1	0,067
virilis	0,052	0,019	-5314,932	119	1,058	1	0,304
***Drosophila* "havaianas"**	**0,050**	**0,173**	**-5313,439**	**119**	**4,044**	**1**	**0,044**
Scaptomyza	0,052	0,018	-5314,818	119	1,285	1	0,257
immigrans	0,052	0,040	-5315,406	119	0,111	1	0,740
Chymomyza-*Scaptodrosophila*	0,052	0,023	-5314,809	119	1,305	1	0,254

M0: hipótese nula, sem marcações filogenéticas; ω_0: razão Ka/Ks do grupo externo à marcação; ω_1: razão Ka/Ks do grupo marcado; np: número de parâmetros; χ^2: qui-quadrado (valor crítico 3.84), df: graus de liberdade. Em negrito, destacados grupos que apresentam valor ω diferente.

4.3 Expressão em transcriptomas de *Drosophila*

Os níveis de expressão de *unc-4* e *OdsH* em todos os transcriptomas disponíveis na plataforma *Gene* são muito baixos. Mais que 50 *counts* foram observados apenas para *OdsH* em testículos de *D. melanogaster* e no trato reprodutivo masculino de *D. pseudoobscura.* Podem ser visualizados, na maior parte das espécies, traços de expressão (entre 10 e 50 *counts*) de *unc-4* em mais tecidos do que *OdsH*, com exceção de *D. ananassae* e *D. willistoni* (Quadro 1). Embora apresente baixos níveis de expressão, *OdsH* parece ter se especializado em funções reprodutivas, sendo expresso em testículos (*D. melanogaster*, *D. simulans*, *D. yakuba*, *D. willistoni* and *D. mojavensis*) e tratos reprodutivos masculinos (*D. yakuba*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. mojavensis* and *D. virilis*). Todavia, *OdsH* é expresso também em larvas de *D. melanogaster* e *D. mojavensis* e fêmeas de *D. pseudoobscura* (carcaça e cabeça). Observaram-se traços de expressão de *unc-4* em testículos, exclusivamente de *D. yakuba* e *D. ananassae*. Dados de *T. dalmanni* (grupo externo) apresentam expressão de *unc-4* cópia única em todos os tecidos, exceto em gônadas.

Quadro 1. Expressão dos genes *unc-4* e *OdsH* em tecidos de *Drosophila* e *T. dalmanni*.

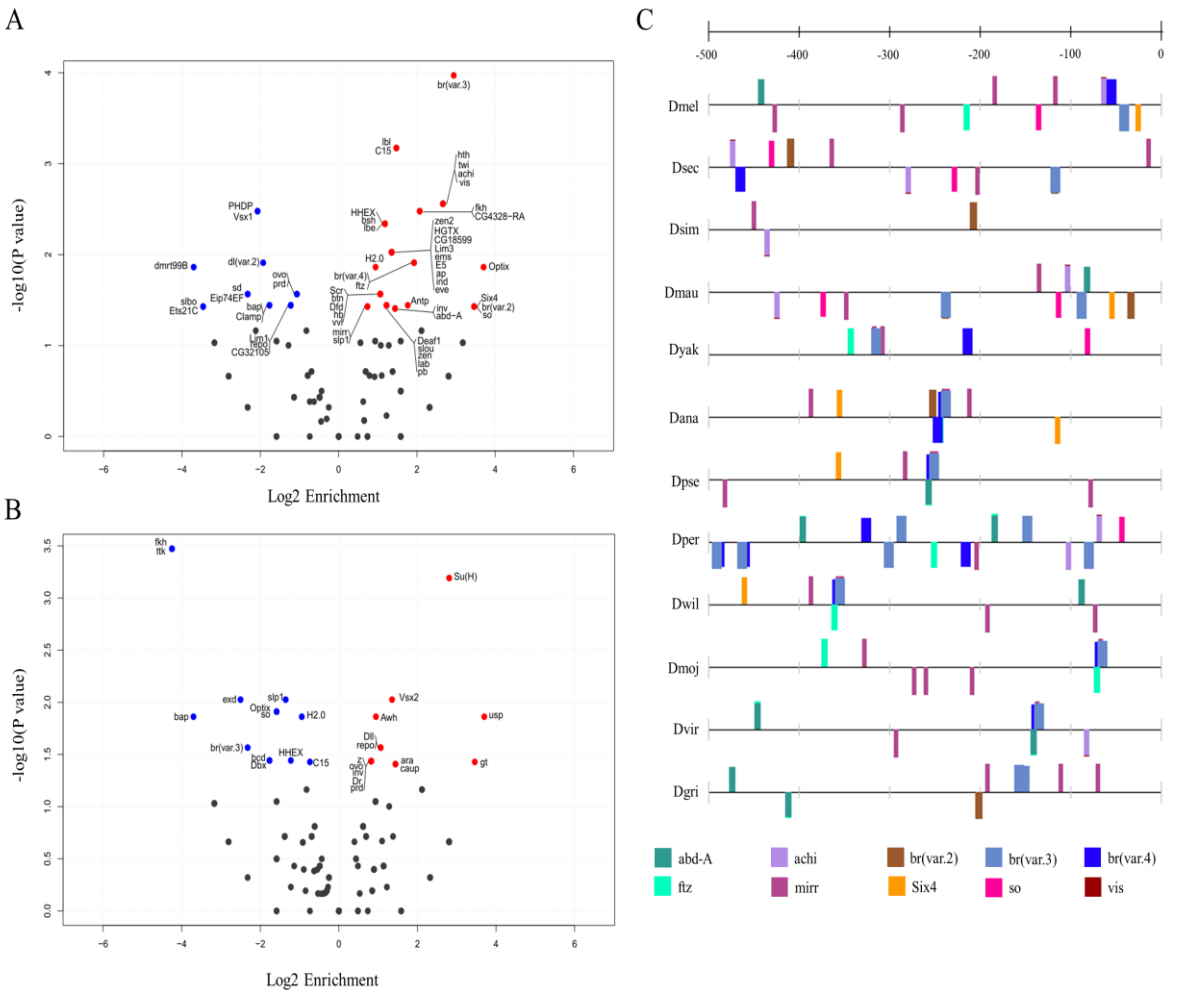
	unc-4									*OdsH*								
	Male				Female					Male				Female				
	C	H	G	R	C	H	G	R	L	C	H	G	R	C	H	G	R	L
D. melanogaster																		
D. yakuba									-									-
D. ananassae									-									-
D. persimilis									-									-
D. pseudoobscura								-										
D. willistoni				-				-	-				-				-	-
D. mojavensis																		
D. virilis									-									-
D. grimshawi				-				-	-			?	-				-	-
Teleopsis dalmanni				-				-										

Quadro preenchido: expressão; Quadro vazio: sem expressão; -: dados não disponíveis; ?: dados inconclusivos, sendo observados menos de 10 *counts.* C: carcaça; H: cabeça; G: gônada; R: trato reprodutivo sem gônadas; L: larva.

4.4 Fatores de transcrição candidatos à regulação da expressão de *OdsH* e *unc-4*

As regiões 500 pb a montante de *OdsH* apresentaram enriquecimento de sítios de ligação para um maior número de fatores de transcrição (43) do que a jusante (15), enquanto *unc-4* apresentou 15 e 13, respectivamente (Figura 9 A e B) (Tabelas Suplementares 5 e 6 – APÊNDICE B). Observa-se também que *OdsH* mostra enriquecimento para uma maior diversidade de candidatos a reguladores do que *unc-4*. Os fatores de transcrição que putativamente se ligam à região reguladora de *OdsH* mostram ampla diversidade de categorias GO, relacionados principalmente ao desenvolvimento e organogênese, enquanto os de *unc-4* também são relacionados ao desenvolvimento e morfogênese da perna (Figura Suplementar 7 – APÊNDICE A). Destaca-se, na região a montante a *OdsH*, o enriquecimento de TFBSs atribuídos à categoria de processo de desenvolvimento envolvido na reprodução (Figura 9 C) e, mais especificamente, de *achi*, *so* e *vis* em espécies do subgênero *Sophophora*, que são relacionados à categoria espermatogênese.

Figura 9. Sítios de ligação de fatores de transcrição nas regiões reguladoras de *OdsH* e *unc-4* em *Drosophila.*



A. Enriquecimento de TFBSs nas regiões reguladoras a montante de *OdsH* (vermelho) e *unc-4* (azul) em *Drosophila*. Os pontos cinza representam os fatores de transcrição cujos sítios de ligação não diferem entre ambos genes. B. Enriquecimento de TFBSs das regiões reguladoras a jusante de *OdsH* (vermelho) e *unc-4* (azul). C. Representação da distribuição de TFBSs enriquecidos na região reguladora a montante de *OdsH* relacionados à reprodução. *achi*, *Six4* e *vis* apresentam sítios sobrepostos.

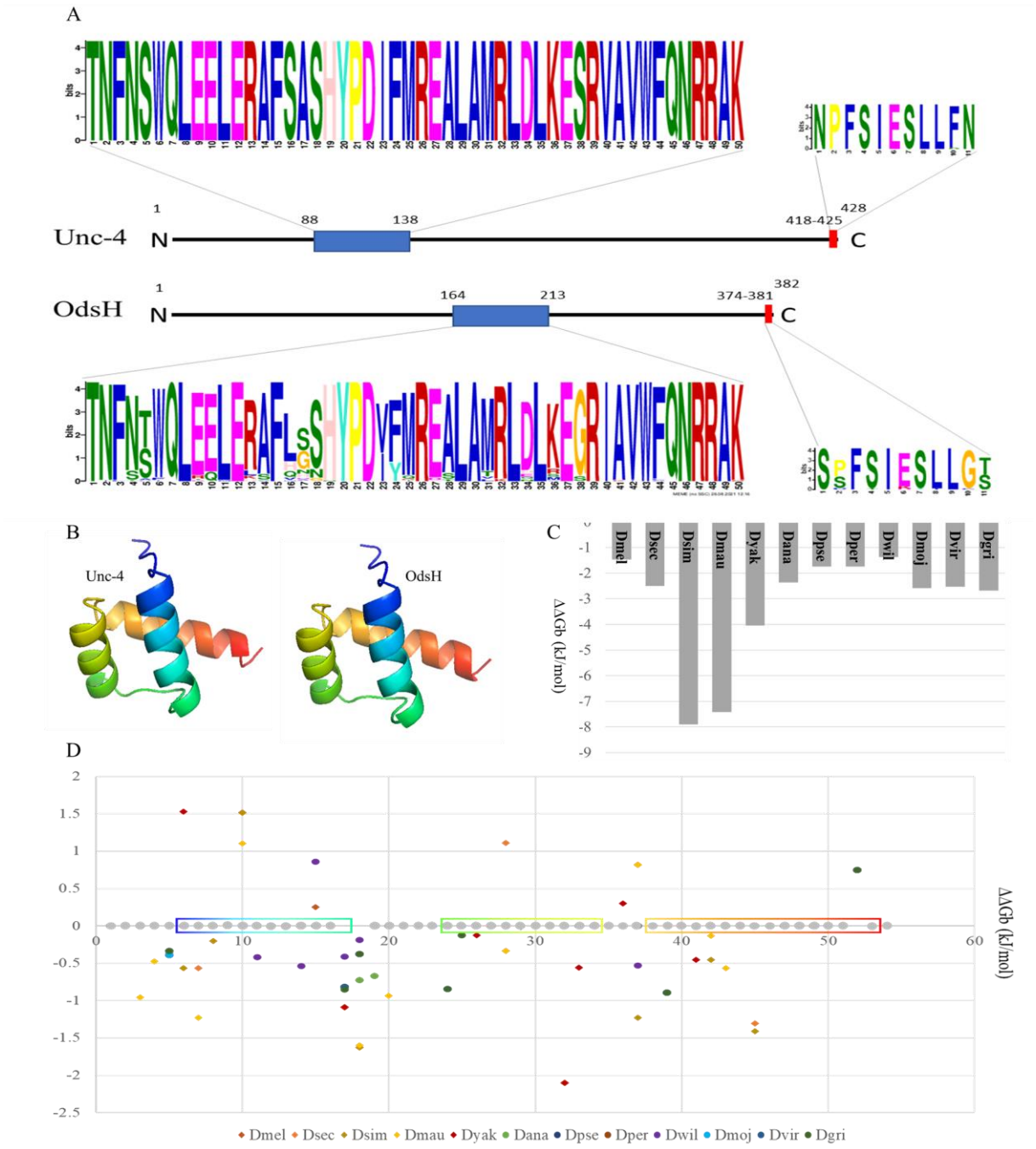
4.5 Motivos funcionais proteicos

O homeodomínio e o octapeptídeo C-terminal identificados com MEME se mostraram conservados nas sequências Unc-4 (*e-values* – homeodomínio: 1.0e-3418; octapeptídeo: 5.8e-238) e OdsH (*e-values* – homeodomínio: 2.3e-3505; octapeptídeo: 1.6e-189) anotadas em *Drosophila*, como observado no esquema das estruturas primárias em *D. melanogaster* (Figura 10). Em ambos os motivos se observa maior divergência de sequências da proteína OdsH, sendo

que os motivos da Unc-4 não apresentam substituições (Figura 10 A). Destaca-se a região entre os aminoácidos 16 a 18 de OdsH que aparentemente apresenta maior divergência. O octapeptídeo da OdsH apresenta o *core* de oito aminoácidos (3-10) conservado, e os aminoácidos das extremidades com certa divergência. Excepcionalmente, *D. mauritiana* tem sua sequência OdsH truncada na extremidade C-terminal em relação às demais *Drosophila* e não apresenta o octapeptídeo.

Os modelos tridimensionais dos homeodomínios desenvolvidos apresentam a estrutura secundária característica de três alfa hélices com uma cauda N-terminal, em um segmento de 54 aminoácidos (Figura 10 B). A estrutura se mostra conservada, com exceção do aminoácido 53, que se encontra na extremidade C-terminal da terceira hélice em OdsH, ao passo que em Unc-4 esse aminoácido não participa da estrutura. Uma vez que o homeodomínio Unc-4 não apresenta substituições de aminoácidos em *Drosophila*, e também não em relação à Unc-4 de *T. dalmanni,* não é observada variação de energia livre de ligação (ΔΔGb) entre eles. Já os homeodomínios da OdsH apresentam-se mais instáveis na associação ao DNA do que Unc-4 (ΔΔGb<0), sendo a instabilidade mais pronunciada em *D. simulans* (-7,896 kJ/mol) e *D. mauritiana* (-7,414 kJ/mol), (Figura 10 C). A maior parte das substituições dos homeodomínios OdsH desestabilizam o complexo formado com o DNA (ΔΔGb<0) (Figura 10 D). Observa-se em geral que as espécies apresentam diferentes substituições em OdsH que resultam em diferentes ΔΔGb por sítio, salvo a *D. persimilis* e *D. pseudoobscura* que possuem sequências idênticas, e *D. mojavensis, D. virilis* e *D. grimshawi* que apresentam números similares de substituições (6, 7 em *D. grimshawi,* sendo 4 delas compartilhadas entre as três espécies). As espécies do subgrupo *melanogaster* apresentam as substituições que resultam em maior ΔΔGb e um maior número de sítios substituídos é encontrado na primeira α-hélice. Na terceira α-hélice, que faz contato direto com o DNA, observam-se duas substituições compartilhadas por diferentes grupos (S40G, exceto por *D. simulans, D. mauritiana* e *D. sechellia* que apresentam o alelo ancestral, e V53W). As demais substituições nessa hélice são espécie-específicas exclusivamente no grupo *melanogaster.*

Figura 10. Motivos funcionais nas proteínas Unc-4 e OdsH em Drosophilinae.

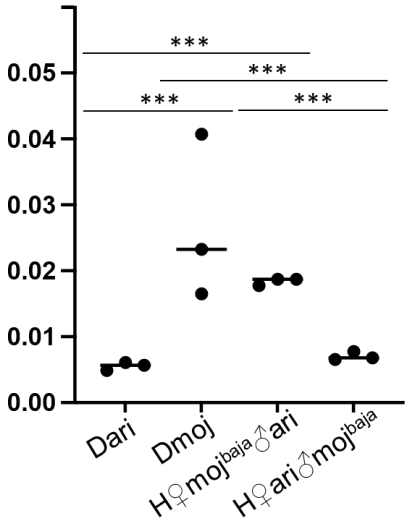


A. Esquema central da estrutura primária das proteínas Unc-4 e OdsH em *D. melanogaster* e motivos funcionais encontrados em Drosophilinae: homeodomíneo (azul) e octapeptídeo (vermelho). B. Modelo tridimensional dos homeodomínios Unc-4 e OdsH. C. Variação de energia total do complexo homeodomínio OdsH e DNA, em espécies do gênero *Drosophila*, em relação a Unc-4. D. Variação de energia do complexo homeodomínio OdsH e DNA por substituição em relação a Unc-4, em espécies do gênero *Drosophila*, ao longo da cadeia de aminoácidos (0-54). Os círculos cinza representam a variação de energia para cada substituição, em relação à Unc-4, com a cor das respectivas espécies. Os sítios sem círculo cinza apresentam substituição, em relação à Unc-4, para todas as espécies analisadas. As caixas em cores correspondentes a B representam a posição das 3 α-hélices. Pontos sobrepostos, que representam mutações compartilhadas: 5- *D. mojavensis, D. virilis, D. grimshawi;* 5- *D. persimilis, D. pseudoobscura;* 17- complexo *melanogaster*; 17- *D. persimilis, D. pseudoobscura, D. ananassae, D. mojavensis, D. virilis;* 18- *D. sechellia, D. simulans*; 18- *D. persimilis, D. pseudoobscura, D. mojavensis, D. virilis e D. grimshawi;* 19- *D. simulans, D. mauritiana, D. yakuba, D. ananassae;* 32- complexo *melanogaster*; 37- *D. melanogaster, D. sechellia, D. mauritiana;* 39- todas exceto *D. simulans, D. sechellia* e *D. mauritiana*; 52- todas.

4.6 Expressão quantitativa de *OdsH* e *unc-4* em *D. m. baja* e *D. arizonae* e seus híbridos

A expressão de *unc-4* se apresentou conservada entre os testículos (F = 2,389; p = 0,144) e carcaças (F = 3,024; p = 0,093), de *D. m. baja*, *D. arizonae* e seus híbridos. Já o gene *OdsH* apresentou maiores valores Q em testículos, concordante com a maior nível expressão de expressão nos testículos de *D. melanogaster* (TING et al., 2004) e comparáveis entre carcaças masculinas (F = 0,889; p = 0,511), porém variáveis em testículos (F = 27,907; p = 0,0003). Os valores de expressão relativa mostraram maior expressão de *OdsH* em *D. m. baja* em relação a *D. arizonae* (Q = 10,9; p < 0.01). A expressão de *OdsH* também foi variável entre híbridos, sendo maior no híbrido H♀ari♂moj^{baja} (Q = 6,965; p < 0,01). Foram observadas diferenças também entre a expressão nos híbridos e suas respectivas espécies paternas, *D. arizonae* (Q = 8,678; p < 0,01) e *D. m. baja* (Q = 9,187; p < 0,01), mas não em relação às espécies maternas (Figura 11).

Figura 11. Expressão relativa de *OdsH* em testículos *D. arizonae* e *D. mojavensis* e seus híbridos recíprocos.



Dari: *D. arizonae*; Dmoj: *D. m. baja*; H: híbrido. ***: p<0.01.

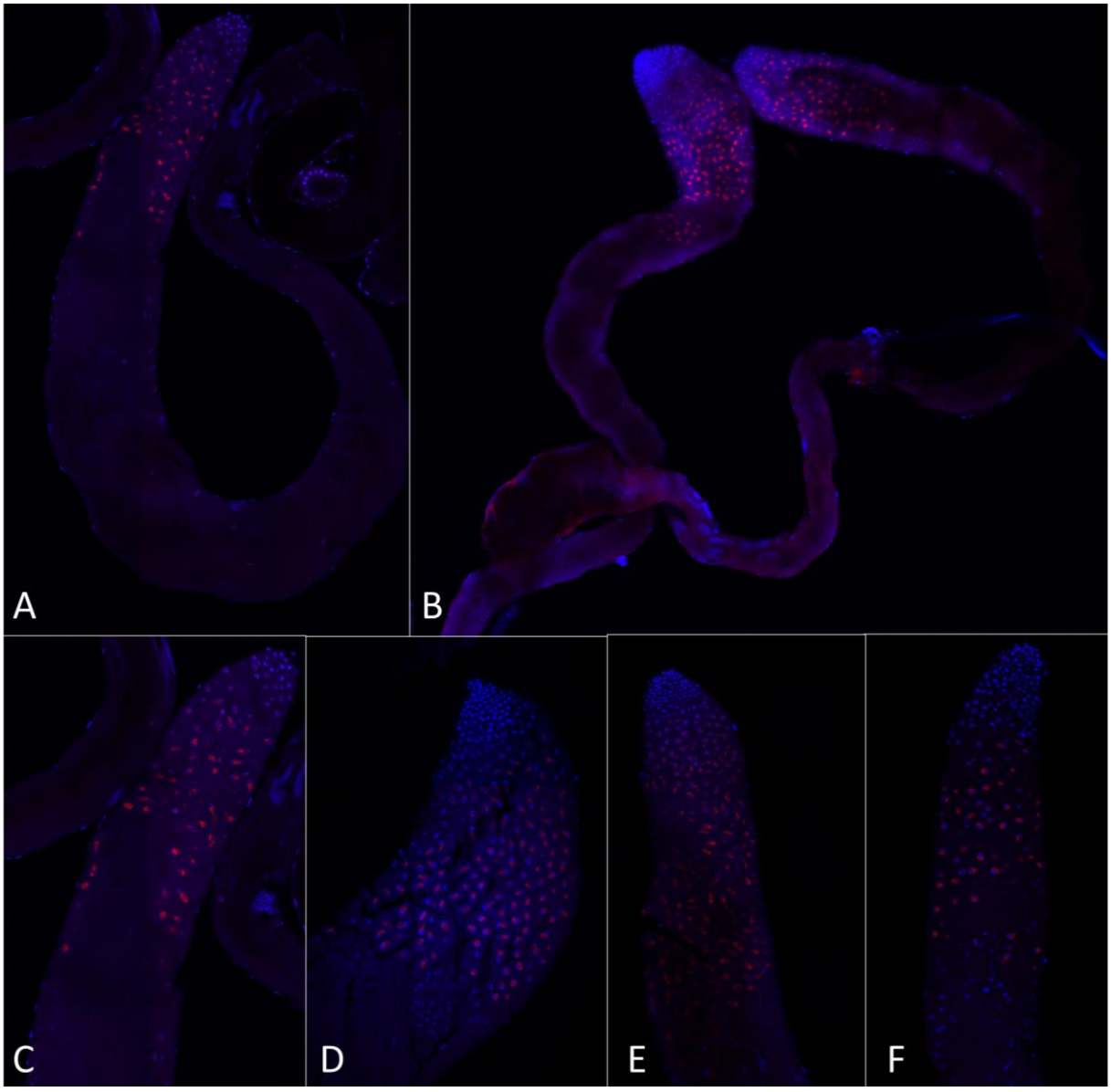
4.7 Expressão *in situ* de *OdsH* em testículos de *D. m. baja* e *D. arizonae* e seus híbridos

Foi observada a hibridização de transcritos *OdsH* exclusivamente em espermatócitos das duas espécies (Figura 12). A coloração vermelha menos intensa ao longo do testículo também pode ser indicativa de transcritos residuais em estágios mais tardios da espermatogênese (Figura 12 A e B). Nos testículos de *D. arizonae* (Figura 12 A e C) e *D. m.*

baja (Figura 12 B e D) observam-se transcritos de *OdsH* na região apical correspondente à diferenciação dos espermatócitos primários e secundários, com o aumento do volume celular e nuclear e descompactação da cromatina, sendo observados com maior intensidade nas células com maior tamanho e núcleo, cuja distribuição se prolonga mais basalmente (Figura 12 C e D). Nota-se ao centro dessa região ausência de células, provavelmente devido ao rompimento dos testículos. Observa-se que essa anatomia de distribuição dos espermatócitos é mais extensa em *D. arizonae* do que em *D. m. baja*. Aparentemente, *OdsH* não é expresso nas células do extremo ápice dos testículos, e também não em estágios pós-meióticos.

Em híbridos (Figura 12 F e G), não foi observada a expressão atípica de *OdsH*, sendo seus transcritos observados também espermatócitos. No híbrido H♀moj^{baja}♂ari (Figura 12 F) observa-se um padrão semelhante à espécie paterna *D. arizonae*, porém aparentemente com uma extensão ainda maior da marcação dos espermatócitos. Já em testículos do híbrido H♀ari♂moj^{baja} (Figura 12 G), o crescimento celular da linhagem germinativa parece ocorrer mais tardiamente, sendo observada uma quantidade maior de células menores no ápice do testículo e o início da fase dos espermatócitos, com a marcação de *OdsH*, mais basalmente em relação às demais amostras analisadas.

Figura 12. Hibridização *in situ* de *OdsH* em testículos de *D. arizonae*, *D. m. baja* e seus híbridos



Azul: DAPI; Vermelho: sondas *OdsH*. A. Panorama de testículo de *D. arizonae*. B. Panorama de testículos de *D. m. baja*. C. Região apical de testículo de *D. arizonae*. D. Região apical de testículo de *D. m. baja*; E. Região apical de testículo de híbrido H♀moj^{baja}♂ari. F. Região apical de testículo de híbrido H♀ari♂moj^{baja}.

5 DISCUSSÃO

Identificou-se que a ocorrência de *OdsH* é mais ampla do que previamente reportada, não estando restrita aos grupos *melanogaster* e *obscura*, do subgênero *Sophophora* (TING et al., 2004), mas se fazendo presente também em espécies de diferentes subgêneros de *Drosophila*, e também em outros gêneros da subfamília Drosophilinae. Suas sequências formam um clado monofilético em inferência filogenética bayesiana, apontando a uma origem

anterior à divergência das linhagens evolutivas de Drosophilinae. Foi observado que *OdsH* evolui a maiores taxas do que o seu parálogo *unc-4*, com características que permitem a inferência de neofuncionalização dessa duplicata em funções reprodutivas masculinas. Ademais, o estudo de sua sequência peptídica permitiu a identificação de particularidades em *D. mauritiana* que poderiam resultar em incompatibilidade dessa proteína no contexto genômico de outras espécies acarretando a esterilidade, como observada nos híbridos *D. mauritiana* x *D. simulans.*

5.1 *OdsH* – a emergência de uma nova duplicata na subfamília Drosophilinae

A hipótese de que a origem do gene *OdsH* se deu a partir de uma duplicação do gene *unc-4* no ancestral do subgênero *Sophophora*, após a divergência do subgênero *Drosophila*, foi elaborada por Ting et al. (2004), por terem identificado esse gene em espécies dos grupos *melanogaster* e *obscura* (subgênero *Sophophora*). Apesar de proporem sua origem a partir de *unc-4*, os autores não estimaram o período em que a duplicação teria ocorrido na história evolutiva de *Sophophora.* Pretendendo responder a essa questão, buscas por sequências homólogas a *unc-4* foram realizadas nos genomas disponíveis da família Drosophilidae. Foi possível identificar sequências do gene *OdsH* em genomas de seis gêneros da subfamília Drosophilinae (*Drosophila*, *Scaptodrosophila*, *Chymomyza*, *Scaptomyza*, *Lordiphosa* e *Zaprionus*), mas não em Steganinae e outras famílias de Diptera (dados não apresentados), o que sugere a ocorrência da duplicação no ancestral Drosophilinae, após a divergência das duas subfamílias de Drosophilidae. Evidências de duplicatas do gene *unc-4* não foram encontradas nos genomas analisados das espécies *D. erecta*, *D. sucinea*, *L. stackelbergi* e *C. xenalocasiae*, o que pode ser decorrente de artefatos na montagem desses genomas (DENTON et al., 2014), ou de perda da duplicata nessas espécies, que também seria plausível, uma vez que, em *D. melanogaster*, espécie em que *OdsH* foi estudado funcionalmente, o gene não influencia significativamente a viabilidade dos indivíduos (SUN et al., 2010; CHENG et al., 2012). Perda de uma das cópias devido a acumulação de mutações randômicas é uma via possível na evolução de genes duplicados (OHNO, 1970).

A ortologia de cada duplicata e a paralogia entre elas são corroboradas por sua posição conservada *in tandem* nos genomas anotados (Figura 7), bem como pelas relações filogenéticas estimadas entre as sequências (Figura 8), sendo essas evidências robustas de relações de homologia entre genes (ALTENHOFF; DESSIMOZ, 2012). A inferência Bayesiana do relacionamento filogenético das duplicatas com a sua segregação um clado monofilético (*OdsH*

e *unc-4*) indica a origem evolutiva de *OdsH,* a partir da duplicação do gene *unc-4* em Drosophilinae, já que esse gene ocorre em diferentes linhagens evolutivas dessa subfamília (tribos, subgêneros e grupos de espécies), agrupado com alta robustez com a sequência já caracterizada de *OdsH* no subgênero *Sophophora* (grupos *melanogaster e obscura*).

As reconstruções filogenéticas de *unc-4* e *OdsH* concordam, em geral, com as escassas árvores filogenéticas amplas de Drosophilinae já publicadas (VAN DER LINDE et al., 2010; RUSSO et al., 2013; YASSIN, 2013). A radiação *willistoni-saltans-Lordiphosa*, todavia, se apresentou em uma posição não esperada nos clados dos dois genes (Figura Suplementar 1 – APÊNDICE A), uma vez que a sua divergência basal a partir do ancestral *Sophophora* é um consenso nas reconstruções evolutivas do gênero *Drosophila* (*DROSOPHILA* 12 GENOMES CONSORTIUM, 2007; VAN DER LINDE et al., 2010; RUSSO et al., 2013; YASSIN, 2013; O'GRADY; DESALLE, 2018). Essas incongruências, todavia, já foram relatadas em árvores de genes que agrupam sequências de espécies dos grupos *willistoni* e *saltans* à base do subgênero *Sophopora*, ou externamente à divergência dos subgêneros *Sophophora/Drosophila*, atribuída ao menor conteúdo GC encontrado nessas espécies (PÉLANDAKIS; SOLIGNAC, 1993; TARRÍO et al., 1999; GAILEY et al., 2000). Incongruências filogenéticas podem ser atribuídas a diferentes taxas evolutivas, amostragem incompleta de linhagens (DEGNAN; ROSENBERG, 2009) e, também, a padrões diferentes no uso de códons entre linhagens evolutivas específicas (INAGAKI et al., 2004, INAGAKI; ROGER, 2006; LIU et al, 2014). Dentre esses fatores, o uso diferencial de códons já foi reportado para espécies dos grupos *willistoni* e *saltans*, com maior utilização de códons terminados em AT do que as demais espécies de *Drosophila* (RODRÍGUEZ-TRELLES; AYALA, 2000; POWELL et al., 2003; VICARIO et al., 2004; SINGH et al., 2006) e corroborado em nossas análises (Figuras Suplementares 2 a 5 – APÊNDICE A), mais pronunciado nas linhagens *willistoni* e *saltans*, mas também para as sequências *unc-4* de *Lordiphosa* e *D. repletoides*. Adicionalmente, o posicionamento filogenético de *Zaprionus* se mostrou ambíguo, uma vez que suas sequências *unc-4* se agruparam no subgênero *Sophophora*, em discordância à posição de *OdsH*, junto ao subgênero *Drosophila*, concordante com diversas filogenias reportadas (revisado em O'GRADY; DESALLE, 2018), mas que mesmo assim, apresenta diferentes posicionamentos filogenéticos em relação ao subgênero *Drosophila*, com a utilização de diferentes genes (THOMAS; HUNT, 1993; RUSSO et al., 1995; KWIATOWSKI; AYALA, 1999). Da mesma forma, a sequência de *unc-4* do grupo *tumiditarsus* se agrupou inesperadamente com o subgênero *Dorsilopha*, e já a de *OdsH* com o subgênero *Drosophila*, concordantemente com estudos anteriores. Logo, o agrupamento monofilético *willistoni-saltans-Lordiphosa*, bem

como *D. repletoides* e gênero *Zaprionus*, foram removidos da inferência filogenética para a datação da duplicação e análises do procesos seletivo para evitar artefatos, decorrentes de incongruências filogenéticas, nessas análises.

A datação da duplicação que originou *OdsH* foi estimada em 55,58 M.a. A presença dessa duplicata exclusivamente em Drosophilinae, e em suas duas tribos, indica a ocorrência da duplicação na linhagem ancestral dessa subfamília. Ao nosso conhecimento, somente Russo et al. (2013), utilizando seis genes nucleares (*Adh*, *Amyrel*, *Ddc*, *Gpdh*, *Sod* e *Xdh*), desenvolveram uma análise de datação ampla da família Drosophilidae, cuja divergência foi estimada em 70,5 M.a, anterior à proposta aqui apresentada à duplicação *unc-4/OdsH* após a divergência do ancestral Drosophilinae. A estimativa para a divergência da família Drosophilidae em nossa análise se mostrou mais recente (59,56 M.a – Figura 8) que a de Russo et al. (2013). Contudo, a datação da divergência de Drosophilinae de Russo et al. (2013) deve ser considerada com restrição, visto que a subfamília Drosophilinae se apresenta parafilética na proposta filogenética apresentada, com representantes da subfamília Steganinae sendo mais derivados do que *Scaptodrosophila* (que é um gênero basal de Drosophilinae), e esse agrupamento não tem suporte (probabilidade posterior < 0.5). Em concordância com nossos resultados, Van der Linde et al. (2010), Yassin (2013) e Dias et al. (2020) propuseram a monofilia da subfamília Drosophilinae, corroborando as relações filogenéticas dos clados *unc-4* e *OdsH* que indicam a duplicação em Drosophilinae, após a divergência Drosophilinae-Steganinae.

Embora tenham sido empregadas, neste estudo, as estimativas de divergência do gênero *Drosophila* de Russo et al. (1995), utilizando sequências do gene *Adh*, para a calibração da datação da inferência filogenética, outras abordagens foram testadas, sendo elas: (*i*) a taxa de mutação estimada por Obbard et al. (2012); (*ii*) os registros fósseis dos gêneros *Chymomyza* e *Scaptomyza* datados por Grimaldi (1987); (*iii*) e as estimativas de divergência por Obbard et al. (2012) dos mesmos táxons utilizados na calibração (entre os subgêneros *Sophophora*/*Drosophila* e espécies *D. melanogaster/D. simulans*). Todavia, utilizando dessas abordagens alternativas, foram obtidos valores de ESS (*Effective Sample Size*) menores que 200 para parâmetros estimados pela Inferência Bayesiana, indicando alta correlação entre as amostras de estados de árvores, e assim, baixa robustez na árvore consenso (DRUMMOND et al., 2006). Adicionalmente, as inferências de datação dos táxons analizados utilizando *i* e *iii* foram subestimadas e já a que utilizou *ii*, superestimada, em relação às datações previamente publicadas para o gênero *Drosophila* (RUSSO et al., 1995; OBBARD et al., 2012; RUSSO et al., 2013) e para o grupo *melanogaster* (CUTTER et al., 2008). Uma vez que *OdsH* e *unc-4*,

evoluem a diferentes taxas, a utilização da taxa de substituição para calibração também acarreta uma disparidade no tamanho dos ramos dos dois respectivos clados. Por outro lado, os parâmetros estimados com a calibração baseada em Russo et al. (1995) apresentaram ESS maiores que 200 e datas concordantes com os estudos anteriores, por isso essa calibração foi adotada.

5.2 *OdsH* e *unc-4* – mesma origem, mas histórias evolutivas divergentes

As sequências *OdsH* e *unc-4* evoluíram assimetricamente, com este evoluindo sob seleção negativa mais intensa. A duplicata *OdsH*, por outro lado, apresenta taxas mais altas de substituições de nucleotídeos e também de aminoácidos, e dessa forma, seleção negativa menos intensa do que *unc-4*, indicando menor restrição funcional sobre essa duplicata. Além disso, *OdsH* mostrou menor pressão de seleção negativa em seu nó ancestral, indicando maior taxa de evolução dessa duplicata no ancestral Drosophilinae, do que ao longo da divergência dessa subfamília, sofrendo maior conservação, em um padrão similar ao observado por Van de Peer et al. (2001) em duplicatas neofuncionais antigas. Ao longo da divergência Drosophilinae, a seleção atuante sobre *OdsH* se mostrou homogênea, salvo aos grupos *Drosophila* "havaianas" e *melanogaster*, com a observação de relaxamento de seleção negativa no clado das espécies *D. simulans-D. mauritiana-D. sechellia*, em concordância com os valores ω altos em análises de seleção par a par dessas espécies (TING et al., 1998).

A análise de dados de expressão dos genes *unc-4* e *OdsH* no banco do NCBI indicou baixos níveis de expressão dos dois genes, com exceção de *OdsH* em testículos de *D. melanogaster* e tratos reprodutivos masculinos de *D. pseudoobscura*. Deve-se levar em consideração que a expressão de *unc-4* e *OdsH* foi demonstrada em células específicas, com *unc-4* em neurônios colinérgicos (LACIN et al., 2020) e *OdsH* em espermatócitos primários na fase G1 pré meiótica (BAYES; MALIK, 2009). Assim, é esperado os níveis de expressão em análises transcriptômicas, como RNA-seq, sejam baixos. Não obstante, em transcriptomas de duas subspecies de *D. mojavensis* (*D. m. mojavensis* e *D. m. wrigleyi*) e *D. arizonae* (BioProject NCBI PRJNA691040), a expressão de *OdsH* varia de 169,5 a 340 *counts* e a de *unc-4*, para os mesmos tecidos, se apresenta inferior a 50 *counts*. Por outro lado, a expressão dos dois genes no sistema reprodutivo feminino é inferior a 10 *counts* (BANHO et al., 2021).

Sendo *unc-4* conservado em Metazoa e relacionado ao desenvolvimento e fisiologia de neurônios, funções essas que influenciam diretamente na viabilidade de metazoários (LACIN et al., 2020; MANSOURI et al., 1997; VON STELINA et al., 2007), inferimos que *unc-4* de

Drosophilinae permaneceu sob a restrição funcional do gene *unc-4* ancestral, já que ele apresenta maior conservação de sequência de nucleotídeos e ausência de substituição de aminoácidos nos Drosophilinae analisados, bem como conservação da expressão em cabeça e carcaças, de fêmeas e machos, e larvas, na maioria das espécies analisadas, concordantes com a expressão observada da cópia única do grupo externo *Teleopsis dalmanni*. Essa conservação funcional também pode ser inferida pela menor diversidade de TFBS putativos e ausência de substituições nos motivos funcionais proteicos octapeptídeo e homeodomínio em relação a Steganinae, com a conservação da estabilidade de ligação deste ao DNA, que pode indicar conservação também dos seus sítios alvo de regulação.

Já o gene *OdsH* é expresso exclusivamente em tratos reprodutivos masculinos, com raríssimas exceções (larvas de *D. melanogaster* e *D. mojavensis* e cabeça e carcaça de machos e fêmeas de *D. pseudoobscura*). Traços de expressão de *OdsH* em larvas, tais como *unc-4* de Drosophilinae e no grupo externo, poderiam indicar a conservaçao da função ancestral em larvas, ou ainda subfuncionalização em algum contexto celular ou tecidual específico, uma vez que apresenta regiões reguladoras e motivos estruturais funcionais da proteína divergentes em comparação ao *unc-4*. A expressão de *OdsH* em larvas também pode ser decorrente da expressão de genes testiculares em gônadas masculinas em desenvolvimento (MAHADEVARAJU et al., 2021). Esses resultados apontam para a necessidade de mais estudos com maior especificadade em larvas, de forma a identificar o papel funcional de *OdsH* nesse estágio de desenvolvimento.

Ao contrário de *unc-4*, *OdsH* mostrou enriquecimento de maior diversidade de TFBSs em suas regiões reguladoras, o que é observado para duplicatas filhas antigas que apresentam aumento da complexidade de sua regulação ao longo da divergência (ZHANG; ZHOU, 2019). Em particular, observou-se o enriquecimento de TFBSs relacionados ao desenvolvimento do sistema reprodutor masculino e às fases iniciais da espermatogênese (*achi*, *so* e *vis*) em espécies do subgênero *Sophophora*. Os genes *achi* e *vis* têm sua expressão conhecida em espermatócitos primários, atuando na especificação do programa de regulação gênica da espermatogênese (AYYAR et al., 2003; WANG; MANN, 2003). Além disso, foi demostrado que *so* é expresso nas células cisto da região apical do testículo de *Drosophila* e contribui com o desenvolvimento normal de espermatócitos primários (FABRIZIO et al., 2003).

Adicionalmente, a proteína OdsH apresenta maior divergência dos motivos funcionais, homeodomínio e octapeptídeo, do que Unc-4, e a divergência observada no homeodomínio pode perturbar a energia de ligação ao DNA, tornando o sistema mais instável. Essas particularidades da proteína OdsH podem tornar a ligação de seu homeodomínio aos sítios alvo

no DNA, menos específica do que Unc-4, permitindo sugerir que as duas proteínas, que são fatores de transcrição, tenham diferentes alvos de ligação na regulação de genes alvo. Todavia, OdsH e Unc-4 têm o aminoácido Q47 do homeodomínio conservado, que garante alta cooperativadade aos homeodomínios, sendo a cooperatividade o principal fator envolvido na especificidade de ligação de homeodomínios aos sítios alvo no DNA (WILSON, 1995). Os aminoácidos que interagem diretamente com as bases nitrogenadas de DNA são também conservados em OdsH e Unc-4 (V44 e N48 [WILSON, 1995]), com exceção à *D. mauritiana*, que apresenta uma isoleucina nesse sítio do homeodomínio da proteína OdsH.

5.3 *OdsH* – um caso de neofuncionalização de genes codificantes de fatores de transcrição

Em vista das mudanças evolutivas da duplicata *OdsH* discutidas acima, particularmente a expressão inovadora em tratos reprodutivos masculinos, o compartilhamento de TFBSs relacionados à reprodução masculina e à maior instabilidade dos homeodomínios OdsH ligados ao DNA, em relação a Unc-4, em soma à maior divergência experimentada pelo nó ancestral de *OdsH*, em relação ao seu parálogo, propomos ter ocorrido neofuncionalização de *OdsH* em testículos do ancestral Drosophilinae. Nossa proposta de neofuncionalização de *OdsH* encontra amparo na hipótese *out of testes* (KAESSMAN, 2010), que propôs que novos genes tendem a ser expressos inicialmente em testículos, o que também tem sido observado para a maior parte de novas duplicatas em *Drosophila* (ASSIS; BACHTROG, 2013; ASSIS, 2014). Posteriormente, *OdsH* parece ter evoluído e desempenhado diferentes funções em diferentes linhagens evolutivas de Drosophilinae, dada a ausência de sua expressão em testículos, mas presença em outros órgãos, em espécies dos grupos *obscura* (*D. pseudoobscura* e *D. persimulis*) e *virilis* (*D. virilis*), concordantemente às observações da aquisição de funções extra-testiculares ao longo do tempo (KAESSMAN, 2010), acompanhada de maior conservação das sequências (JIANG; BACHTROG, 2017), como também observado neste estudo. Em adição, Bao et al. (2018) demonstraram que duplicatas gênicas, em *Drosophila*, com origem estimada em aproximadamente 60 milhões de anos, passaram por uma maior taxa de neofuncionalização e evolução inovadora, configurando assim um contexto evolutivo propício à fixação de substituições e neofuncionalização em duplicatas à época em que ocorreu a duplicação do *unc-4*. Sugerimos também que o papel funcional de *OdsH*, nas espécies que apresentam expressão testicular do gene *OdsH*, deve ser desempenhado nos estágios iniciais da espermatogênese, devido à predição de TFBSs relacionados a esses estágios em *Sophophora*. Essa hipótese é suportada pelas análises de expressão *in situ* de *OdsH* em espermatócitos primários de *D.*

simulans (BAYES; MALIK, 2009) e pela identificação da participação de duplicatas antigas, sendo esse o caso do *OdsH*, nas fases mais iniciais da espermatogênese (RAICES et al., 2019; SU et al., 2021).

Em testículos de *D. arizonae* e *D. m. baja*, observa-se que a expressão de *OdsH* varia em quantidade, tal como observado entre as espécies do complexo *melanogaster* (TING et al., 2004). A expressão em híbridos parece ter um padrão de herança maternal, uma vez que a quantidade de transcritos *OdsH* não varia em relação às suas respectivas espécies maternais e que esse gene se localiza no cromossomo X, e, portanto, o gene nos machos, juntamente com sua região reguladora é herdado da mãe. Contudo, maior número de réplicas é necessário para poder-se estabelecer essa relação com mais segurança.

Observa-se que a expressão de *OdsH* nessas espécies ocorre durante a diferenciação dos espermatócitos, tal como observado em *D. simulans* (BAYES; MALIK, 2009), com maior intensidade nas maiores células da espermatogênese e maior descompactação da cromatina (espermatócitos primários diferenciados e secundários), estágio em que ocorre um intensivo crescimento celular e maior atividade sintética de RNA na espermatogênese de *Drosophila* (HACKSTEIN, 1987). *OdsH* parece ser expresso a partir da fase G2 e durante a meiose de *D. arizonae* e *D. m. baja*, como observado em uma quantidade reduzida de duplicatas gênicas antigas que são expressas nas fases pós-mitóticas da espermatogênese (RAICES et al., 2019; SU et al., 2021).

Estudos com *D. hydei* mostraram que o formato celular dos espermatócitos e a meiose masculina é controlada também por fatores localizados no cromossomo X (HACKSTEIN et al., 1990). Ademais, Lu *et al* (2010) propuseram que a desregulação da expressão de *OdsH* em híbridos estéreis descendentes de *D. simulans* e *D. mauritiana* acarreta, enquanto fator de transcrição, a desregulação de genes autossômicos em estágios mais tardios da espermatogênse, sendo esse mecanismo relacionado ao fenótipo de formação de feixes defeituosos de espermatozoides e resultando em imobilidade. Assim, propomos que o mecanismo de atuação de *OdsH* em *D. arizonae* e *D. m. baja*, também seja por meio da regulação de fatores envolvidos em eventos mais tardios da espermatogênese, tais como a meiose e a espermiogênese.

Dessa forma, embora os níveis de expressão entre híbridos férteis (H♀moj^{baja}♂ari) e estéreis (H♀ari♂moj^{baja}) difiram, como sugerido pela análise quantitativa, mas não em relação às respectivas espécies maternas, a expressão do *OdsH* de *D. arizonae* pode ser incompatível com fatores em outros cromossomos provenientes de *D. mojavensis*, já sendo relatada por Vigneault e Zouros (1986) a incompatibilidade entre os cromossomos X de *D. arizonae* e 3 e Y de *D. mojavensis*, em híbridos estéreis. Também se observou uma possível divergência de

expressão *in situ* em híbridos estéreis, com aparente diferenciação mais tardia de espermatócitos primários. Assim, se mostra necessária a caracterização mais profunda da espermatogênese nesses híbridos e, também, nas espécies parentais, sendo que nenhum estudo que caracterizasse a espermatogênese nessas espécies foi reportado.

Os resultados aqui apresentados também permitem sugerir possíveis causas para a incompatibilidade de *OdsH* no genoma de *D. simulans*, que foi associada à esterilidade híbrida em machos descendentes desse cruzamento (PEREZ et al., 1993; PEREZ; WU, 1995; TING et al., 1998): uma substituição do aminoácido conservado Valina no sítio 43, que interage diretamente com as bases nitrogenadas de seu sítio de ligação no DNA, por uma Isoleucina no *OdsH* de *D. mauritiana*, que pode mudar os sítios alvo de ligação no genoma. Foi identificado que a proteína OdsH de *D. mauritiana* se liga a região heterocromática do cromossomo Y, enquanto que a de *D. simulans* não se liga a essa região, o que pode ser causado pela diferença em um dos aminoácidos de *D. mauritiana* que se ligam a fita de DNA. Além disso, as proteínas OdsH nessas duas espécies apresentaram os maiores valores de instabilidade de ligação ao DNA (Figura 10 C), indicando provável maior divergência de TFBSs nessas espécies em relação a Unc-4, provavelmente dirigida pelo relaxamento da seleção negativa, cujos sinais são observados nessas sequências.

Entretanto, a especificidade de ligação a sítios na fita de DNA depende principalmente dos cofatores de transcrição que atuam ligados aos homeodomínios (WILSON et al., 1995; BÜRGLIN; AFFOLTER, 2016). Nesse sentido, a proteína OdsH de *D. mauritiana* não apresenta o octapeptídeo, que é conservado na extremidade C-terminal de Unc-4 e OdsH de Drosophilinae. A evolução dos homeodomínios da classe filogenética Paired-Like, à qual pertencem Unc-4 (WINNIER et al., 1999) e OdsH, se dá por meio de rearranjos e perdas de motivos funcionais, dentre eles o octapeptídeo, sendo observada entre as suas famílias a diversidade de estruturas proteicas quanto a sua presença/ausência (JACOB, 1977). Esse motivo funcional também é conservado em Unc-4 de *C. elegans*, e se liga ao co-fator de transcrição Unc-37 (ortólogo à Groucho, em *Drosophila*) reprimindo a expressão (WINNIER et al., 1999). Portanto, um dos fatores que podem contribuir para a esterilidade híbrida pode ser também a perda da função repressora em OdsH de *D. mauritiana*, podendo acarretar a descondensação da região pericêntrica do cromossomo Y (BAYES; MALIK, 2009) e a desregulação dos genes autossômicos associados aos fenótipos de testículos menores e defeitos na formação dos feixes de esperma (LU et al., 2010). Assim, o mecanismo pelo qual *OdsH* acarreta a infertilidade desses híbridos se dá pela incompatibilidade de locos genéticos das duas espécies decorrente de sua divergência, como proposto pelo modelo de Dobzhansky-Muller

(DOBZHANSKY, 1937; MULLER, 1942) e também sugerido previamente (BAYES; MALIK, 2009; LU et al., 2010).

Em suma, apresentamos aqui uma proposta inovadora sobre a origem de *OdsH*, mais antiga do que reportada anteriormente, e sobre o processo evolutivo dessa duplicata, que evolui assimetricamente em relação ao seu parálogo e apresenta características que permitem a inferência de neofuncionalização. Reportamos também características associadas à divergência proteica, em particular em *D. mauritiana*, que podem estar associadas à incompatibilidade descrita de OdsH no genoma de *D. simulans*. Nosso estudo mostra que mesmo sendo o primeiro gene de especiação descrito em *Drosophila*, muito da história evolutiva que levou a duplicata *OdsH* a desempenhar essa função permanecia desconhecida. Nossa contribuição para o entendimento da evolução molecular e funcional desse gene não esgota a necessidade de explorar mais a fundo suas características funcionais, e reforça que esse gene permanece como um interessante modelo de estudos sobre mecanismos moleculares envolvidos na esterilidade híbrida e divergência de duplicatas gênica

6 CONCLUSÕES

1. A duplicação *unc-4*/*OdsH* é mais antiga do que proposta na literatura, com ocorrência estimada em 55,58 M.a no ancestral Drosophilinae, em consequência, sua distribução é mais ampla do que a reportada.
2. Substituições de aminoácidos e ausência do octapeptídeo na proteína OdsH de *D. mauritiana* podem ser responsáveis pela incompatibilidade nos híbridos estéreis com *D. simulans.*
3. Ocorrência de neofuncionalização de *OdsH* no sistema reprodutor masculino de Drosophilinae evidenciada por:

   a. aceleração das taxas de evolução no ancestral Drosophilinae, em relação ao seu parálogo mais conservado *unc-4;*

   b. expressão inovadora no sistema reprodutor masculino de Drosophilini;

   c. expressão na meiose de *D. arizonae*, *D. m. baja* e seus híbridos;

   d. maior diversidade de reguladores de expressão, incluindo relacionados à espermatogênese de Drosophilini;

   e. maior divergência do homeodomínio de Drosophilini, que pode acarretar menor especificidade de ligação ao DNA.

REFERÊNCIAS

ABRAHAM, M. J. et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. **SoftwareX**, v. 1–2, p. 19–25, 2015.

ALTENHOFF, A. M.; DESSIMOZ, C. Inferring orthology and paralogy. **Methods Mol. Biol.**, v. 855, p. 259-279, 2012.

ASHBURNER et al. Gene ontology: tool for the unification of biology. **Nat. Genet.**, v. 25, n. 1, p. 25-29, 2000.

ASSIS, R.; BACHTROG, D. Neofunctionalization of young duplicate genes in *Drosophila*. **PNAS**, v. 110, n. 43, p. 17409-17414, 2013.

ASSIS, R. *Drosophila* duplicate genes evolve new functions on the fly. **Fly (Austin)**, v. 8, n. 2, p. 91-94, 2014.

AYYAR, S. et al. *Drosophila* TGIF is essential for developmentally regulated transcription in spermatogenesis. **Development**, v. 130, n. 13, p. 2841-2852, 2003.

BÄCHLI, G. **TaxoDros: the database on taxonomy of Drosophilidae.** 2016. Disponível em: http://www.taxodros.uzh.ch/. Acesso em 29 nov. 2021.

BAILEY, J. A.; LIU, G.; EICHLER, E. E. An *Alu* transposition model for the origin and expansion of human segmental duplications. **Am. J. Hum. Genet.**, v. 73, n. 4, p. 823-834, 2003.

BAILEY, T. L. et al. The MEME Suite. **Nucleic Acids Res.**, v. 43, n. W1, p. W39-W49, 2015.

BAILEY, T. L.; LEKAN, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: International Conference on Intelligent Systems for Molecular Biology, 2nd., 1994, Stanford. **Proceedings**… Menlo Park, California: AAAI Press, 1994. p. 28-36.

BANHO, C. A. et al. Comparative transcriptomics between *Drosophila mojavensis* and *D. arizonae* reveals transgressive gene expression and underexpression of spermatogenesis-related genes in hybrid testes. **Sci. Rep.**, v. 10, n. 1, 9844, 2021.

BANKEVICH, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. **J. Comput. Biol.**, v. 19, n. 5, p. 455-477, 2012.

BAO, R.; FRIEDRICH M. Molecular evolution of the *Drosophila* retinome: exceptional gene gain in the higher Diptera. **Mol. Biol. Evol.**, v. 26, n. 6, p. 1273-1297, 2009.

BAO, R. et al. Comparative Evidence of an Exceptional Impact of Gene Duplication on the Developmental Evolution of *Drosophila* and the Higher Diptera. **Front. Evol. Ecol.**, v. 6, n. 63, 2018.

BAYES, J. J.; MALIK, H. S. Altered heterochromatin binding by a hybrid sterility protein in *Drosophila* sibling species. **Science**, v. 326, n. 5959, p. 1538-1541, 2009.

BIRNEY, E.; CLAMP, M.; DURBIN, R. GeneWise and Genomewise. **Gen. Res.**, v. 14, n. 5, p. 988–995, 2004.

BONO, J. M.; MARKOW, T. A. Post-zygotic isolation in cactophilic *Drosophila*: larval viability and adult life-history traits of *D. mojavensis*/*D. arizonae* hybrids. **Journal of evolutionary biology**, v. 22, n. 7, p. 1387–1395, 2009.

BRITTEN, R. J. Rates of DNA sequence evolution differ between taxonomic groups. **Science**, v. 231, n. 4744, p. 1393-1398, 1986.

BÜRGLIN, T. R.; AFFOLTER, M. Homeodomains: an update. **Chromosoma**, v. 125, p. 497-521, 2016.

CASNEUF, T. et al. Nonrandom divergence of gene expression following gene and genome duplication in the flowering plants *Arabidopsis thaliana*. **Genome Biol**., v. 7, n. 2, R13, 2006.

CENCI, G. et al. Chromatin and microtubule organization during premeiotic, meiotic and early postmeiotic stages of *Drosophila melanogaster* spermatogenesis. **J. Cell Sci.**, v. 107, pt.12, p. 3521-3534, 1994.

CENZER, M. L. Adaptation to an invasive host is driving the loss of a native ecotype. **Evolution**, v. 70, n. 10, p. 2296-2307, 2016.

CHENG, Y.-J. et al. Reduction of germ cells in the *Odysseus* null mutant causes male fertility defect in *Drosophila melanogaster*. **Genes & genetic systems**, v. 87, n. 4, p. 273-276, 2012.

COMEAULT, A. A. et al. Selection on a genetic polymorphism counteracts ecological speciation in a stick insect. **Curr. Biol.**, v. 25, n. 15, p. 1975-1981, 2015.

CONANT, G. C; WAGNER, A. Asymmetric sequence divergence of duplicate genes. **Genome Res.**, v. 13, n. 9, p. 2052–2058, 2003.

COPLEY, R. R. The EH1 motif in metazoan transcription factors. **BMC Genomics**, v. 6, n. 169, 2005.

COUGHLAN, J. M.; MATUTE, D. R. The importance of intrinsic postzygotic barriers throughout the speciation process. Philos. **Trans. R. Soc. Lond. B. Biol. Sci.**, v. 375, n. 1806, 20190533, 2020.

COUNTERMAN, B. A.; NOOR, M. A. F. Multilocus test for introgression between the cactophilic species *Drosophila mojavensis* and *Drosophila arizonae*. **Am. Nat.**, v. 168, n. 5, p. 682– 696, 2006.

COYNE, J. A., ORR, H. A. Patterns of speciation in *Drosophila*. **Evolution**, v. 43, n.2, p. 362–381, 1989.

COYNE, J. A.; ORR, H. A. **Speciation**. Sunderland, MA: Sinauer Associates, Inc., 2004.

CUTTER, A. D. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. **Mol. Biol. Evol.**, v. 25, n. 4, p. 778-786, 2008.

DEGNAN, J. H.; ROSENBERG, N. A. Gene tree discordance, phylogenetic inference and the multispecies coalescent. **Trends Ecol. Evol.**, v. 24, n. 6, p. 332-340, 2009.

DIXON, M. D.; COYNE, J. A.; NOOR, M. A. F. The evolution of conspecific sperm precedence in *Drosophila*. **Mol. Ecol.**, v. 12, n. 5, p. 1179-1184, 2003.

DENTON, J. F. et al. Extensive error in the number of genes inferred from draft genome assemblies. **PLoS Comput. Biol**., v. 10, n. 12, e1003998, 2014.

DIAS, G. R. et al. A phylogenomic study of Steganinae fruit flies (Diptera: Drosophilidae): strong gene tree heterogeneity and evidence for monophyly. **BMC Evol. Biol.**, v. 20, n. 1, 141, 2020.

DOBZHANSKY, T. H. **Genetics and the origin of species**. New York: Columbia Univ. Press, 1937.

DOBZHANSKY T. Speciation as a stage in evolutionary divergence. **Am. Nat.**, v. 74, p. 312–321, 1940.

DROSOPHILA 12 GENOMES CONSORTIUM. Evolution of genes and genomes on the *Drosophila* phylogeny. **Nature**, v. 450, n. 7167, p. 203-218, 2007.

DRUMMOND, A. J. et al. Relaxed phylogenetics and dating with confidence. **PLoS Biology**, v. 4, n. 5, p. e88, 2006.

FABIAN, L. et al. Phosphatidylinositol 4,5-bisphosphate directs spermatid cell polarity and exocyst localization in *Drosophila*. **Mol. Biol. Cell.**, v. 21, n. 9, p. 1546–1555, 2010.

FABIAN, L.; BRILL, J. A. *Drosophila* spermiogenesis: Big things come from little packages. **Spermatogenesis**, v. 2, n. 3, p. 197–212, 2012.

FABRIZIO, J. J.; BOYLE, M.; DINARDO, S. A somatic role for *eyes absent* (*eya*) and *sine oculis* (*so*) in *Drosophila* spermatocyte development. **Dev. Biol.**, v. 258, n. 1, p. 117-28, 2003.

FAY, J. C., WYCOFF, G. J., WU, C. -I. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. **Nature**, v. 415, n. 6875, p. 1024–1026, 2002.

FARES, M. A. et al. The roles of whole-genome and small-scale duplications in the functional specialization of *Saccharomyces cerevisiae* genes. **PLoS Genet.**, v. 9, n. 1, e1003176, 2013.

FELSENSTEIN, J. Cases in which parsimony or compatibility methods will be positively misleading. **Syst. Zool.**, v. 19, n. 4, p. 401–410, 1978.

JACOB, F. Evolution and tinkering. **Science**, v. 196, n. 4295, p. 1161–1166, 1977.

FISTON-LAVIER, A. -S.; ANXOLABEHERE, D.; QUESNEVILLE, H. A model of segmental duplication formation in *Drosophila melanogaster*. **Genome Res.**, v. 17, n. 10, p. 1458-1470, 2007.

FORCE, A. et al. Preservation of duplicate genes by complementary, degenerative mutations. **Genetics**, v. 151, n. 4, p. 1531–1545, 1999.

FOX, R. M. et al. A gene expression fingerprint of *C. elegans* embryonic motor neurons. **BMC Genomics**, v. 6, p. 42, 2005.

FULLER, M. T. Spermatogenesis. In: Bate, M. (Ed.) **The Development of *Drosophila***. Cold Spring Harbor: Cold Spring Harbor Press, p. 71-147, 1993.

GAILEY, D. A. et al. A phylogeny of the Drosophilidae using the sex-behaviour gene *fruitless*. **Hereditas**, v. 133, n. 1, p. 81-83, 2000.

GAN, Q. et al. Monovalent and unpoised status of most genes in undifferentiated cell-enriched *Drosophila* testis. **Genome Biol.**, v. 11, n. 4, p. R42, 2010.

GEARING, L. J. et al. CiiiDER: A tool for predicting and analysing transcription factor binding sites. **PLoS One**, v. 14, n. 9, e0215495, 2019.

GEHRING, W. J. Homeo boxes in the study of development. **Science**, v. 236, n. 4806, p. 1245-1252, 1987.

GENHARD, T.; HARTMANN, K.; STEEL, M. Stochastic properties of generalised Yule models, with biodiversity applications. **J. Math. Biol.**, v. 57, n. 5, p. 713-735.

GOMES, S.; CIVETTA, A. Hybrid male sterility and genome-wide misexpression of male reproductive proteases. **Sci. Rep.**, v. 5, p. 11976, 2015. doi: 10.1038/srep11976.

GOMES, S.; CIVETTA, A. Misregulation of spermatogenesis genes in *Drosophila* hybrids is lineage-specific and driven by the combined effects of sterility and fast male regulatory divergence. **J. Evol. Biol.**, v. 27, n. 9, p. 1775-1783, 2014.

GRIMALDI, D.A. A phylogenetic, revised classification of genera in the Drosophilidae (Diptera). **Bulletin of the American Museum of Natural History**, n. 197, 1–139, 1990.

GRIMALDI, D. A. Amber fossil Drosophilidae (Diptera), with particular reference to the hispaniolan taxa. **American Museum Novitates**, n. 2880, p. 1–23, 1987.

HAHN, M. W. Bias in phylogenetic tree reconciliation methods: implications for vertebrate genome evolution. **Genome Biol.**, v. 8, n. 7, p. R141, 2007.

HALDANE J. B. S. Sex ratio and the unisexual sterility in animal hybrids. **Journal of Genetics**, v. 12, p. 101–109, 1922.

HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic acids symposium series**, v. 41, n. 41, p. 95-98, 1999.

HE, X.; ZHANG, J. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. **Genetics**, v. 169, n. 2, p. 1157-1154, 2005.

HENDY, M. D.; PENNY, D. A framework for the quantitative study of phylogenetic trees. **Syst. Zool**., v. 38, n. 4, p. 297–309, 1989.

HIRST, J.; CARMICHAEL, J. A potential role for the clathrin adaptor GGA in *Drosophila* spermatogenesis. **BMC Cell Biol.**, v. 12, n. 22, 2011.

HUGHES, A. L. The evolution of functionally novel proteins after gene duplication. **Proc. Biol. Sci.**, v. 256, n. 1346, p. 119-124, 1994.

HUGHES, A. L. **Adaptive evolution of genes and genomes**. New York: Oxford University Press, 1999. 288p.

HUGHES, A. L., et al. Adaptive diversification within a large family of recently duplicated, placentally expressed genes. **Proc. Natl. Acad. Sci.**, v. 97, n. 7, p. 3319–3323, 2000.

INAGAKI, Y. et al. Phylogenetic artifacts can be caused by leucine, serine, and arginine codon usage heterogeneity: dinoflagellate plastid origins as a case study. **Syst. Biol.**, v. 53, n. 4, p. 582-93, 2004.

INAGAKI, Y.; ROGER, A. J. Phylogenetic estimation under codon models can be biased by codon usage heterogeneity. **Mol. Phylogenet. Evol.**, v. 40, n. 2, p. 428-434, 2006.

JIANG, N. *Pack-MULE* transposable elements mediate gene evolution in plants. **Nature**, v. 431, n. 7008, p. 569-573, 2004.

JIANG, X.; ASSIS, R. Natural selection drives rapid functional evolution of young *drosophila* duplicate genes. **Mol. Biol. Evol.**, v. 34, n. 12, p. 3089-3098, 2017.

JORGENSEN W. L. et al. Comparison of simple potential functions for simulating liquid water. **J. Chem. Phys.**, v. 79, n. 926, p. 926–935, 1983.

KAESSMANN, H. Origins, evolution, and phenotypic impact of new genes**. Genome Res.**, v. 20, n. 10, p. 1313-1768, 2010.

KATOH, K. et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic Acids Research**, v. 30, n. 14, p. 3059-3066, 2002.

KELLEHER, E. S.; EDELMAN, N. B.; BARBASH, D. A. *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. **PLoS biology**. v. 10, n. 11, 2012.

KIMURA, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. **J. Mol. Evol.**, v. 16, n. 2, p. 111-120, 1980.

KUMAR, S. **PHYLTEST: phylogenetic hypothesis testing software**. Version 2.0. Pennsylvania: Pennsylvania State University, 1996.

KUMAR, S. et al. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. **Mol. Biol. Evol.**, v. 35, n. 6, p. 1547-1549, 2018.

KIMMINS, S.; SASSONE-CORSI, P. Chromatin remodelling and epigenetic features of germ cells. **Nature**, v. 434, n. 733, p. 583-589, 2005.

KWIATOWSKI, J.; AYALA, F. J. Phylogeny of *Drosophila* and related genera: conflict between molecular and anatomical analyses. **Mol. Phylo. Evol.**, v. 13, n. 12, 319-328, 1999.

LACIN, H. et al. Neurotransmitter identity is acquired in a lineage-restricted manner in the *Drosophila* CNS. **eLife**, v. 8, p. e43701, 2019.

LACIN, H. et al. Unc-4 acts to promote neuronal identity and development of the take-off circuit in the *Drosophila* CNS. **eLife**, v. 9, p. e55007, 2020.

LACIN, H.; TRUMAN, J. W. Lineage mapping identifies molecular and architectural similarities between the larval and adult *Drosophila* central nervous system. **eLife**, v. 5, e13399, 2016.

LAI, Z. C.; FORTINI, M. E.; RUBIN, G. M. The embryonic expression patterns of *zfh-1* and *zfh-2*, two *Drosophila* genes encoding novel zinc-finger homeodomain proteins. **Mech. Dev.**, v. 34, n. 2, p. 123-134, 1991.

LAL, S. K. et al. The maize genome contains a *helitron* insertion. **Plant Cell.**, v. 15, n. 2, p. 381-391.

LIU, Y. et al. Mitochondrial phylogenomics of early land plants: mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias. Syst. Biol., v. 63, n. 6, p. 862-878, 2014.

LOPEZ-MAESTRE, H. et al. Identification of misexpressed genetic elements in hybrids between *Drosophila*-related species. **Sci. Rep.**. 7:40618, 2017.

LU, X. et al. Genome-wide misexpression of X-linked versus autosomal genes associated with hybrid male sterility. **Genome Res.**, v. 20, n. 8, p. 1097-1102, 2010.

LYNCH, M. et al. The probability of preservation of a newly arisen gene duplicate. **Genetics**, v. 159, n. 4, p. 1789-1804, 2001.

MACHADO, C. A. et al. Multilocus nuclear sequences reveal intra- and interspecific relationships among chromosomally polymorphic species of cactophilic *Drosophila*. **Mol. Ecol.**, v. 16, n. 14, p. 3009–3024, 2007.

MAHADEVARAJU, S. et al. Dynamic sex chromosome expression in *Drosophila* male germ cells. **Nat. Commun.**, v. 12, n. 1, p. 892, 2021.

MAIER, J. A. et al. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. **J. Chem. Theory Comput.**, v. 11, n. 8, p. 3696–3713, 2015.

MANSOURI, A. et al. Paired-related murine homeobox gene expressed in the developing sclerotome, kidney, and nervous system. **Dev. Dyn.**, v. 210, n. 1, p. 53-65, 1997.

MARKOW, T. A.; HOCUTT, G. D. Reproductive isolation in sonoran desert *Drosophila*: Testing the Limits of the Rules. In: HOWARD, D. J.; BERLOCHER, S. H. (Org.). **Endless Forms: Species and Speciation**. Oxford: Oxford Press, 1998. p. 234–244.

MARQUES, F. et al. Identification of avoidance genes through neural pathway-specific forward optogenetics. **PLoS Genet.**, v. 15, n. 12, e1008509, 2019.

MAYR, E. **Animal species and evolution**. Cambridge: Harvard University Pree, Cambridge, 1963, 797p.

MAYR, E. **Systematics and the Origin of Species**. Columbia Univ. Press, New York, 1942, 376p.

McGINNIS, W. et al. A conserved DNA sequence in homoeotic genes of the *Drosophila* Antennapedia and bithorax complexes. **Nature**, v. 308, n. 5958, p. 428-433, 1984.

MEISEL, R. P. Repeat mediated gene duplication in the *Drosophila pseudoobscura* genome. **Gene**, v. 438, n. 1-2, p. 1-7, 2009.

MI, H. et al. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. **Nucleic Acids Res.**, v. 47, n. D1, p. D419-D426, 2019.

MILLER, D. M. et al. *C. elegans unc-4* gene encodes a homeodomain protein that determines the pattern of synaptic input to specific motor neurons. **Nature**, v. 355, n. 6363, p. 842-845, 1992.

MIYAZONO, K. et al. Cooperative DNA-binding and sequence-recognition mechanism of aristaless and clawless. **EMBO. J.**, v. 29, n. 9, p. 1613–1623, 2010.

MORIYAMA, E. N. Higher rates of nucleotide substitution in *Drosophila* than in mammals. **Jpn J Genet.**, v. 62, p. 139-147, 1987.

MULLER, H.J. Isolating mechanisms. Evolution and temperature. **Biol.Symp**.*v.* 6, p.71–125. 1942.

NEI, M.; KUMAR, S. **Molecular evolution and phylogenetics**. Nova York: Oxford University Press, 2000.

O'GRADY, P. M.; DESALLE, R. Phylogeny of the Genus *Drosophila*. **Genetics**. V. 209, n. 1, p. 1-25, 2018.

OBBARD, D. J. et al. Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. **Mol. Biol. Evol.**, v. 29, n. 11, p. 3459-3473, 2012.

OHNO, S. **Evolution by gene duplication**. New York: Springer-Verlag, 1970, 160p.

OKADA, T. A proposal of establishing tribes for the family drosophilidae with key to tribes and genera (Diptera): taxonomy and systematics. **Zoolog. Sci.** v. 6, p. 391–399, 1989.

ORR, H. A.; MASLY, J. P.; PRESGRAVES, D. C. Speciation genes. **Curr. Opin. Genet. Dev.**, v. 14, n. 6, p. 675-679, 2004.

PÉLANDAKIS, M.; SOLIGNAC, M. Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. **J. Mol. Evol.**, v. 37, n. 5, p. 525-543, 1993.

PEREZ, D. E. et al. Genetics of reproductive isolation in the *Drosophila simulans* clade: DNA marker-assisted mapping and characterization of a hybrid-male sterility gene, *Odysseus* (*Ods*). **Genetics**, v. 134, n. 1, p. 261-275, 1993.

PEREZ, D. E.; WU, C. I. Further characterization of the *Odysseus* locus of hybrid sterility in *Drosophila*: one gene is not enough. **Genetics**, v. 140, n. 1, p. 201-206, 1995.

PETIT, C. Qualitative aspects of genetics and environment in determination of behavior. In: EHRMAN, L.; OMENN, G. S.; CASPARI, E. (Ed.). **Genetics, Environement and Behaviour: Implications for Educational Policy**. New York: Academic Press, Inc., 1972, p. 28-54.

PIRES, D. E. V. et al. mCSM: predicting the effects of mutations in proteins using graph-based signatures. **Bioinformatics**, v. 30, n. 3, p. 335–342, 2014.

POWELL J. R. et al. Analysis of a shift in codon usage in *Drosophila*. **J. Mol. Evol.**, v. 57, p. S214-225, 2003.

PUIGBÒ, P.; BRAVO, I. G.; GARCIA-VALLVE, S. CAIcal: a combined set of tools to assess codon usage adaptation. **Biol. Direct**, v. 3, n. 38, 2008. https://doi.org/10.1186/1745-6150-3-38

RAICES, J. B.; OTTO, P.A.; VIBRANOVSKI, M. D. Haploid selection drives new gene male germline expression. **Genome Res.**, v. 29, n. 7, p. 1115-1122, 2019.

RAMBAUT, A. **FigTree, a graphical viewer of phylogenetic trees**. Institute of Evolutionary Biology University of Edinburgh. < http://tree.bio.ed.ac.uk/>. 2009.

RAMBAUT, A. et al. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. **Syst. Biol.**, v. 67, n. 5, p. 901-904, 2018.

RAMOS, O. M.; FERRIER, D. E. K. Mechanisms of gene duplication and translocation and progress towards understanding their relative contributions to animal genome Evolution. **Int. J. Evol. Biol.**, n. 846421, 2012.

RANZ, J. M.; CASALS, F.; RUIZ, A. How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila.* **Genome Res.**, v. 11, n. 2, p. 230-239, 2001.

REED, L. K.; MARKOW, T. A. Early events in speciation: polymorphism for hybrid male sterility in *Drosophila*. **PNAS USA**, v. 101, n. 24, p. 9009-9012, 2004.

REED, L. K.; NYBOER, M.; MARKOW, T. A. Evolutionary relationships of *Drosophila mojavensis* geographic host races and their sister species *Drosophila arizonae*. **Mol. Ecol.**, v. 16, n. 5, p. 1007–1022, 2007.

RODRÍGUEZ-TRELLES, F.; TARRÍO, R.; AYALA, F. J. Fluctuating mutation bias and the evolution of base composition in *Drosophila.* **J. Mol. Evol.**, v. 50, n. 1, p. 1-10, 2000.

ROMERO-SORIANO, V. et al. Transposable element misregulation is linked to the divergence between parental piRNA pathways in *Drosophila* hybrids. **Genome Biol. Evol.**, v. 9, n.6, p.1450–1470, 2017.

ROSENBLUM, E. R. et al. Goldilocks meets Santa Rosalia: an ephemeral speciation model explains patterns of diversification across time scales. **Evol. Biol.**, v. 39, n.2, p. 255-261, 2012.

ROSS, C. L.; MARKOW T. A. Microsatellite variation among diverging populations of *Drosophila mojavensis*. **Journal Compilation a 2006 European Society for Evolutionary Biology**, p. 1691–1700, 2006.

ROTH, D. B.; PORTER, T. N.; WILSON, J. H. Mechanisms of nonhomologous recombination in mammalian cells. **Mol. Cell. Biol.**, v. 5, n. 10, p. 2599-2607, 1985.

RUIZ, A.; HEED, W. B. Host-plant specificity in the cactophilic *Drosophila mulleri* species complex. **J. Anim. Ecol.**, v. 57, n. 1, p. 237–249, 1988.

RUIZ, A.; HEED, W. B.; WASSERMAN, M. Evolution of the *mojavensis* cluster of cactophilic *Drosophila* with descriptions of two new species. **Journal of Heredity**, v. 81, n. 1, p.30-42, 1990.

RUSSO, C. A.; TAKEZAKI, N.; NEI, M. Molecular phylogeny and divergence times of drosophilid species. **Mol. Biol. Evol.**, v. 12, n. 3, p. 391-404.

RUSSO, C. A. et al. Phylogenetic analysis and a time tree for a large drosophilid data set (Diptera: Drosophilidae). **Zool. J. Lin. Soc.**, v. 169, n. 4, p. 765–775, 2013.

SANCHES-FLORES, A. et al. Genome evolution in three species of cactophilic *Drosophila*. **G3, Genes, Genomes, Genetics**, v. 6, n. 10, p. 3097-3115, 2016.

SANTOS, W. **Divergência interpopulacional e interespecífica e processo seletivo atuante nos genes ortólogos a *fruitless* e *Odysseus* em *Drosophila mojavensis* e *D. arizonae***. Dissertação (Mestrado em Biologia Animal) - Universidade Estadual Paulista "Júlio de Mesquita Filho", São José do Rio Preto, 2014.

SATYAKI, P. R. V. et al. The *Hmr* and *Lhr* hybrid incompatibility genes suppress a broad range of heterochromatic repeats. **PLOS Genetics**, v. 10, e1004240, 2014.

SHAO, Y. et al. GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. **Genome Res.**, v. 29, n. 4, p. 682-696, 2019.

SINGH, N. D.; ARNDT, P. F.; PETROV, D. A. Minor shift in background substitutional patterns in the *Drosophila saltans* and *willistoni* lineages is insufficient to explain GC content of coding sequences. **BMC Biol.**, v. 4, n. 37, p. 1-14, 2006.

SLACK, A. On the mechanism of gene amplification induced under stress in *Escherichia coli*. **PLoS Genet.**, v. 2, n. 4, p. e48, 2006.

SOLTZFUS, A. On the possibility of constructive neutral evolution. **J. Mol. Evol.**, v. 49, n. 2, p 169–181, 1999.

STANKIEWICZ, P.; LUPSKI, J. R. Molecular-evolutionary mechanisms for genomic disorders. **Curr. Opin. Genet. Dev.**, v. 12, n. 3, p. 312-319, 2002.

STURTEVANT, A. H. On the subdivision of the genus *Drosophila*. **PNAS USA**, v. 25, n. 3, p. 137-141, 1939.

STURTEVANT, A. H. The classification of the genus *Drosophila*, with descriptions of nine new species. **The University of Texas Publication**, n. 4213, p. 5-51.

SU, Q.; HUANGYI, H; ZHOU, Q. On the origin and evolution of *Drosophila* new genes during spermatogenesis. **Genes (Basel)**. v. 12, n. 11, p. 1796, 2021.

SUCHARD, M. A. et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. **Virus Evol.**, v. 4, n. 1, vey016, 2018.

SUN, S.; TING, C.-T.; WU, C.-I. The normal function of a speciation gene, *Odysseus*, and its hybrid sterility effect. **Science**, v. 305, n. 5680, p. 81-83, 2004.

SUPEK, F. et al. REVIGO summarizes and visualizes long lists of gene ontology terms. **PLoS One**, v. 6, n. 7, p. e21800, 2011.

TABUCHI, K. et al. A novel *Drosophila* paired-like homeobox gene related to *Caenorhabditis elegans unc-4* is expressed in subsets of postmitotic neurons and epidermal cells. **Neurosci. Lett.**, v. 257, n. 1, p. 49-52, 1998.

TARRÍO, R.; RODRÍGUEZ-TRELLES, F.; AYALA, F. J. Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. **Mol. Biol. Evol.**, v. 18, n. 8, p. 1464-1473, 2001.

THOMAS, R. H.; HUNT, J. A. Phylogenetic relationships in *Drosophila*: a conflict between molecular and morphological data. **Mol. Biol. Evol.**, v. 10, n. 2, p. 362-374, 1993.

THROCKMORTON, L. H. The phylogeny, ecology, and geography of *Drosophila*. In: KING, R. (Ed.). **Handbook of genetics**. vol. 3. New York: Plenum Press, 1975, p. 421-469.

TING, C.-T. et al. A Rapidly evolving homeobox at the site of a hybrid sterility gene. **Science**, v. 282, n. 5393, p. 1501–1504, 1998.

TING, C.-T. et al. Gene duplication and speciation in *Drosophila*: evidence from the *Odysseus* locus. **PNAS USA**, v. 101, n. 33, p. 12232–12235, 2004.

TURISSINI, D. A. et al. The rate of evolution of postmating-prezygotic reproductive isolation in *Drosophila*. **Mol. Biol. Evol.**, v. 35, n. 2, p. 312-334, 2018.

VAN DE PEER, Y. et al. The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. **J. Mol. Evol.**, v. 53, n. 4-5, p. 436-446, 2001.

VAN DER LINDE, K. et al. A supermatrix-based molecular phylogeny of the family Drosophilidae. **Genet. Res. (Camb.)**, v. 92, n. 1, p. 25-38, 2010.

VIBRANOVSKI, M. D. Meiotic sex chromosome inactivation in *Drosophila*. **J. Genomics**, v. 2, p. 104-117, 2014.

VICARIO, S.; MORIYAMA, E. N.; POWELL, J. R. Codon usage in twelve species of *Drosophila*. **BMC Evol. Biology**, v. 7, p. 226, 2007.

VIGNEAULT, G.; ZOUROS, E. The genetics of asymmetrical male sterility in *Drosophila mojavensis* and *Drosophila arizonensis* hybrids: interactions between the Y-chromosome and autosomes. **Evolution**, v. 40, n. 6, p. 1160-1170, 1986.

VON STELINA, S. E. et al. UNC-4 represses CEH-12/HB9 to specify synaptic inputs to VA motor neurons in *C. elegans*. **Genes Dev.**, v. 21, n. 3, p. 332-346, 2007.

WANG, Z.; MANN, R. S. Requirement for two nearly identical TGIF-related homeobox genes in *Drosophila* spermatogenesis. **Development**, v. 130, n. 13, p. 2853-2865, 2003.

WARSSEMAN, M.; KOEPFER, R. Character displacement for sexual isolation between *Drosophila mojavensis* and *Drosophila arizonensis*. **Evolution**, v. 31, n. 4, p. 812-823, 1977.

WATERHOUSE, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. **Nucleic Acids Res.**, v. 46, n. W1, p. W296-W303, 2018.

WHITE-COOPER, H., BAUSEK, N. Evolution and spermatogenesis. **Philos. Trans. R. Soc. Lond. B. Biol. Sci.**, v. 365, n. 1546, p. 1465–1480, 2010.

WIEGMANN, B. M. et al. Episodic radiations in the fly tree of life. **PNAS USA**, v. 108, n. 14, p. 5690-5695, 2011.

WILSON, D. S. et al. High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. **Cell**, v. 82, n. 5, p. 709-719, 1995.

WINNIER, A. R. et al. UNC-4/UNC-37-dependent repression of motor neuron-specific genes controls synaptic choice in *Caenorhabditis elegans*. **Genes Dev.**, v. 13, n. 21, p.2774–2786, 1999.

WU, C. I.; DAVIS, A. W. Evolution of postmating reproductive isolation: the composite nature of Haldane's rule and its genetic bases. **Am. Nat.**, v. 142, n. 2, p. 187-212, 1993.

WU, C. I. Inferences of species phylogeny in relation to segregation of ancient polymorphisms. **Genetics**, v. 127, n. 2, p. 429-435, 1991.

WU, C. I.; JOHNSON, N. A.; PALOPOLI, M. F. Haldane's rule and its legacy: Why are there so many sterile males? **Trends Ecol. Evol.**, v. 11, n. 7, p. 281-284, 1996.

WU, C. I.; TING, C.-T. Genes and speciation. **Nature Reviews Genetics**, v. 5, n. 2, p. 114-122, 2004.

XIA, B. et al. Widespread transcriptional scanning in the testis modulates gene evolution rates. **Cell**, v. 180, n. 2, p. 248-262, 2020.

YANG, S. Repetitive element-mediated recombination as a mechanism for new gene origination in *Drosophila.* **PLoS Genet.**, v. 4, n. 1, p. e3, 2008.

YANG, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. **Mol. Biol. Evol.**, v. 15, n. 5, p. 568-73, 1998.

YANG, Z. **Phylogenetic analysis by maximum likelihood (PAML)**. 2000.

YANG, Z.; NIELSEN, R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. **J. Mol. Evol.**, v. 46, n. 4, p. 409-448, 1998.

YASSIN, A. Phylogenetic classification of the Drosophilidae Rondani (Diptera): the role of morphology in the post genomic era. **Systematic Entomology**, v. 38, n. 2, p. 349–364, 2013.

ZGARBOVÁ M. et al. Refinement of the sugar-phosphate backbone torsion Beta for AMBER force fields improves the description of Z- and B-DNA. **J. Chem. Theory Comput.**, v. 11, n. 12, p. 5723–5736, 2015.

ZHANG, J. -Y.; ZHOU, Q. On the regulatory evolution of new genes throughout their life history. **Mol. Biol. Evol.**, v. 36, n. 1, p. 15-27, 2019.

ZHAXYBAYEVAL, O. Anciently duplicated genes reduce uncertainty in molecular clock estimates. **PNAS USA**, v. 110, n. 30, p. 12168–12169, 2013. doi: 10.1073/pnas.1310930110.

ZHOU, Q. et al. On the origin of new genes in *Drosophila.* **Genome Res.**, v. 18, n. 9, p. 1446-1455, 2008.

ZOUROS, E.; LOFDAHL, K.; MARTIN, P. A. Male hybrid sterility in *Drosophila* - interactions between autosomes and sex-chromosomes in crosses of *Drosophila mojavensis* and *Drosophila arizonensis*. **Evolution**, v. 42, n. 6, p. 1321–1331, 1988.

APÊNDICE A – Figuras Suplementares

Figura Suplementar 1. Inferência filogenética Bayesiana das sequências dos genes parálogos *unc-4* e *OdsH*. (continua)

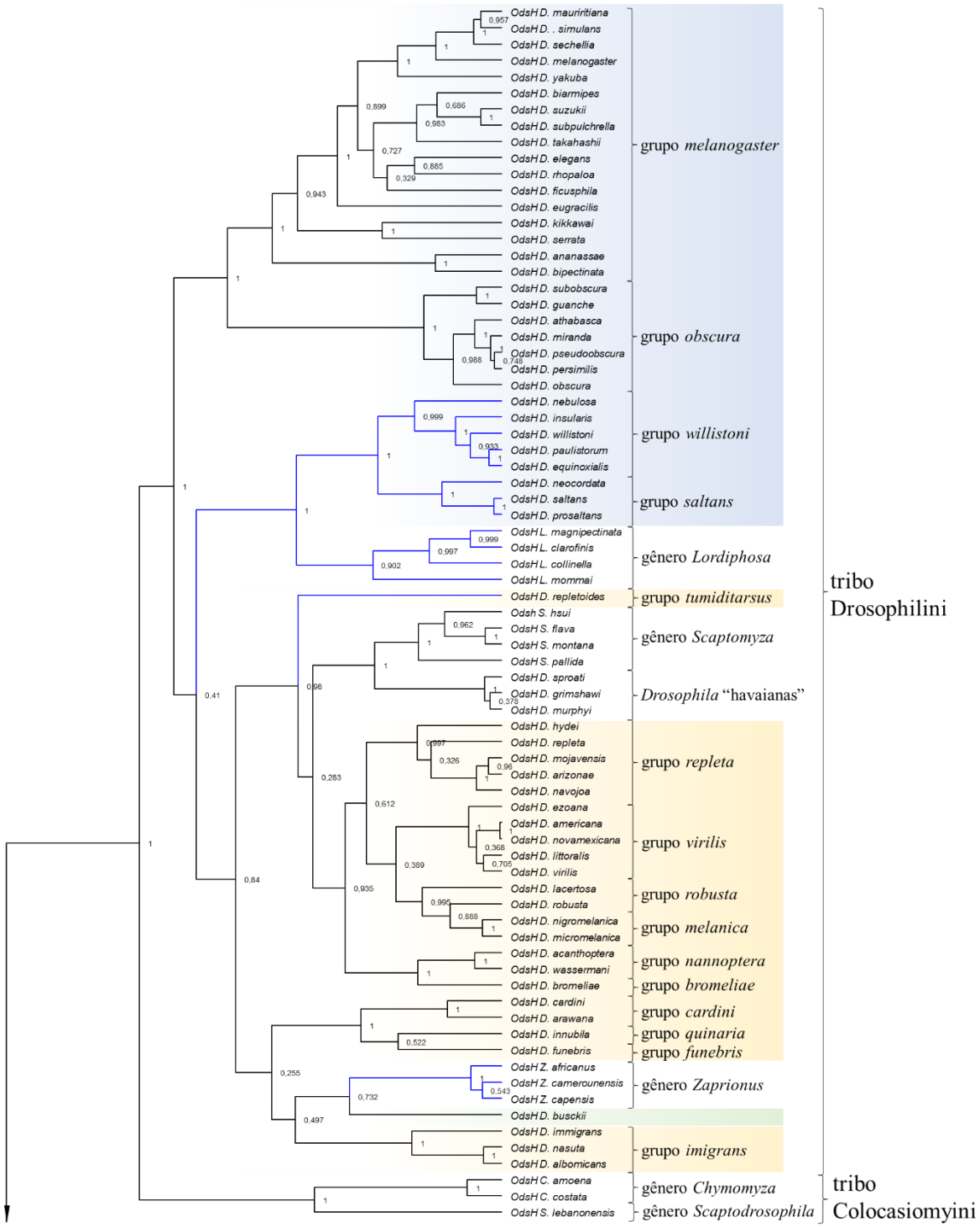
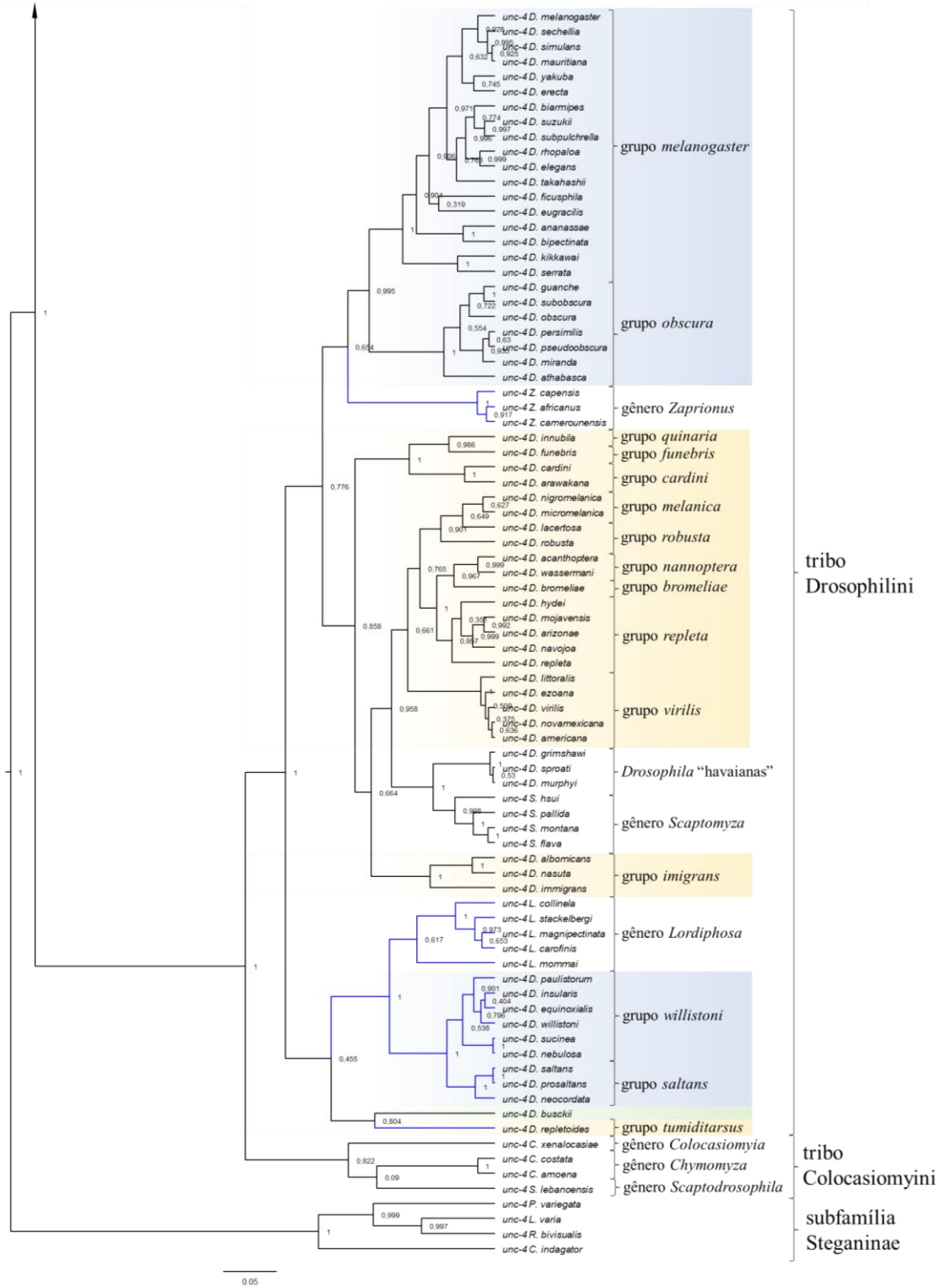


Figura Suplementar 1. Inferência filogenética Bayesiana das sequências dos genes parálogos *unc-4* e *OdsH.* (conclusão)



O clado *OdsH* é apresentado na primeira parte da figura, seguido da continuidade da árvore com clados *unc-4* Drosophilinae e *unc-4* cópia única de Steganinae (grupo externo), cujos genomas não apresentaram evidências de duplicata, na segunda parte. A análise foi realizada com 405 sítios nucleotídicos de 159 sequências nucleotídicas utilizando o modelo de substituição GTR+G+I. Todas as posições contendo *gaps* e posições ambíguas foram retiradas da análise par a par de sequências. Na raiz de cada clado estão indicados os valores de probabilidade posterior. A análise foi conduzida no BEAST v1.10.4. As setas indicam a junção entre as partes. Radiação *willistoni-saltans-Lordiphosa*, gênero *Zaprionus* e *D. repletoides* destacados em azul. Os subgêneros de *Drosophila* são delimitados pelos quadros azul (*Sophophora*), amarelo (*Drosophila*) e azul (*Dorsilopha*).

Figura Suplementar 2. Inferência filogenética Bayesiana calibrada dos genes parálogos *OdsH* e *unc-4*. (continua)

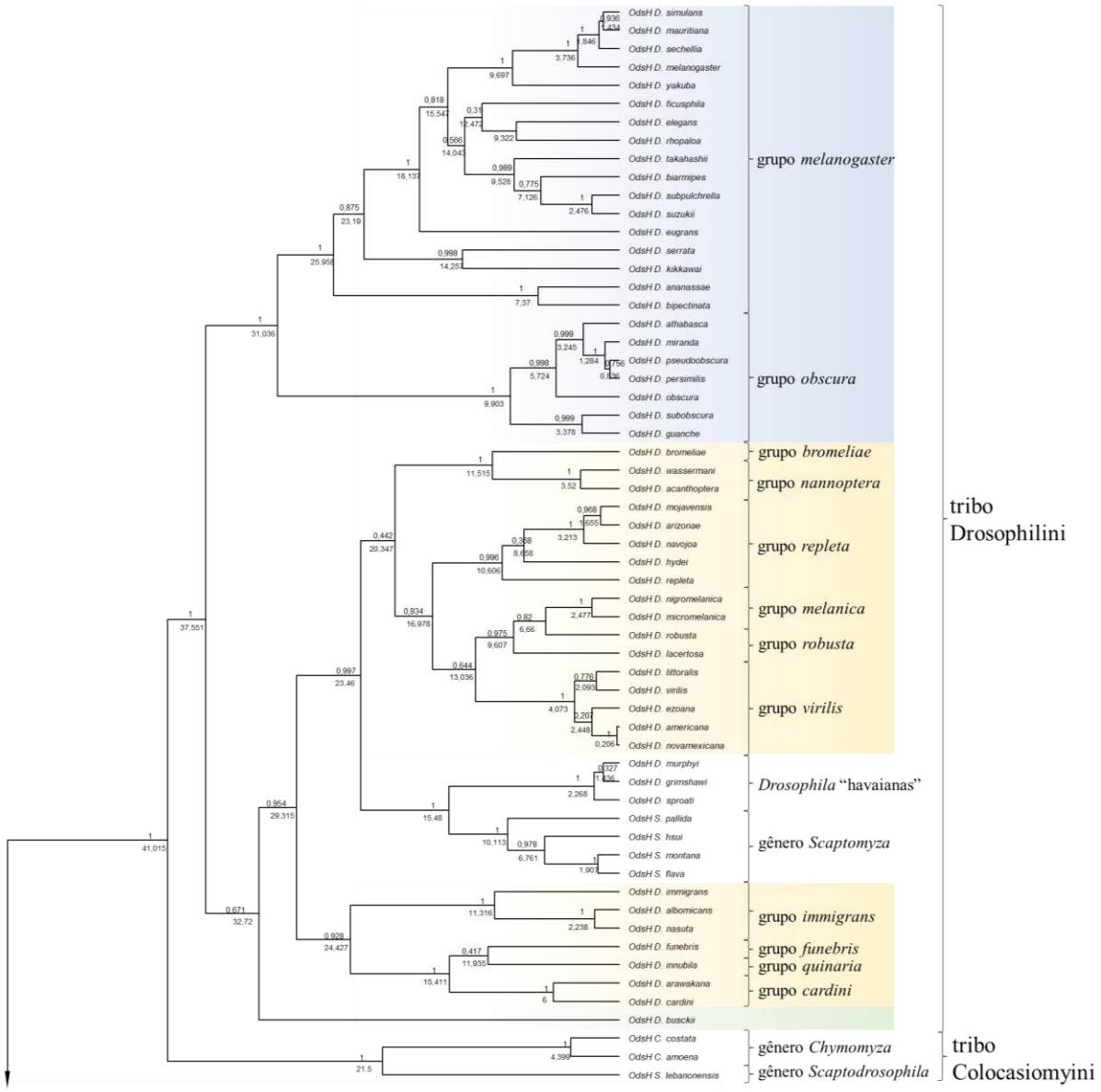
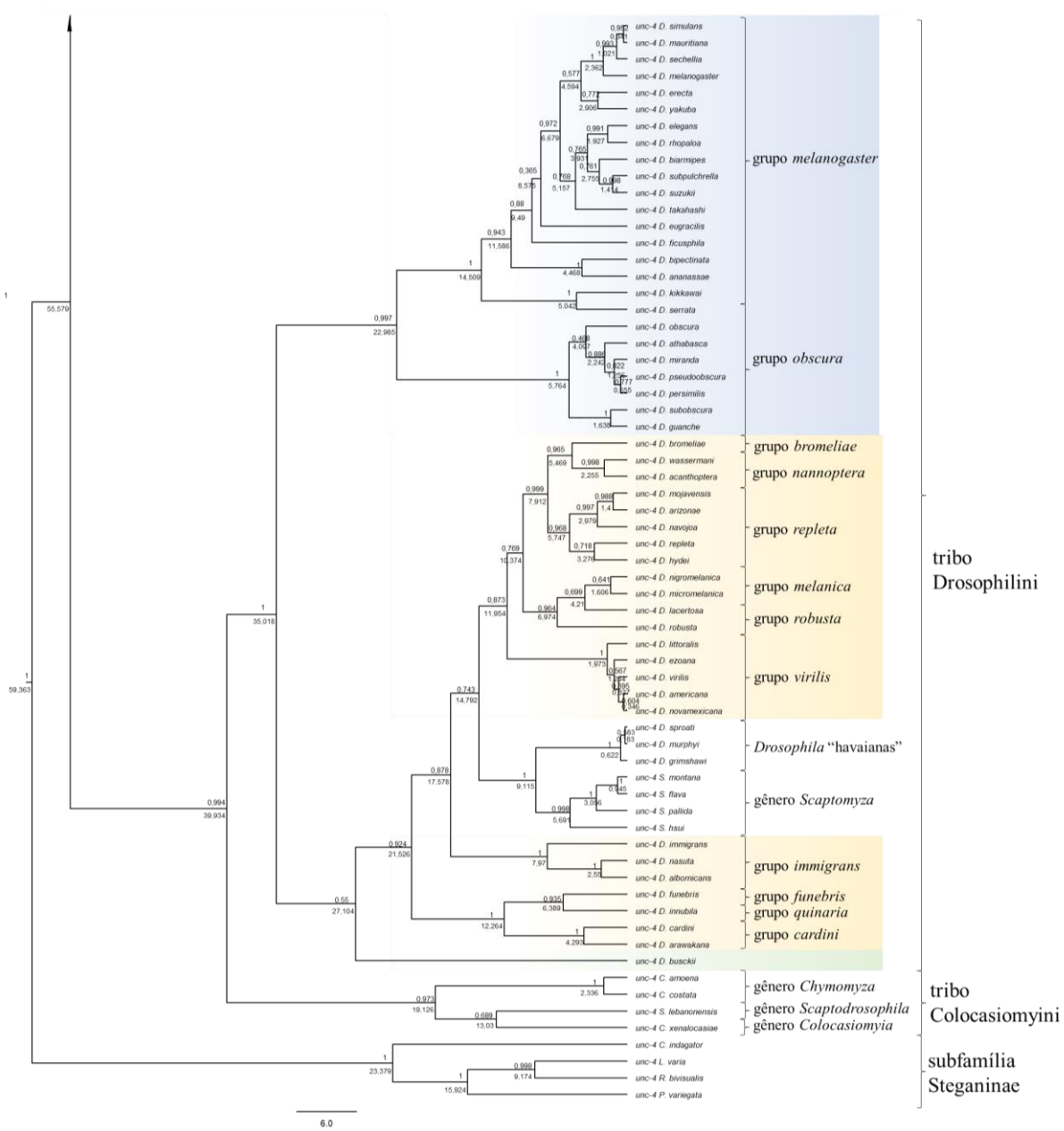


Figura Suplementar 2. Inferência filogenética Bayesiana calibrada dos genes parálogos *OdsH* e *unc-4.* (conclusão)



O clado *OdsH* é apresentado na primeira parte da figura, seguido da continuidade da árvore com clados *unc-4* Drosophilinae e *unc-4* cópia única de Steganinae (grupo externo), cujos genomas não apresentaram evidências de duplicata, na segunda parte. A análise foi realizada utilizando a usando o modelo de substituição GTR+G+I. A análise foi realizada com 405 sítios nucleotídicos de 124 sequências nucleotídicas. Todas as posições contendo *gaps* e posições ambíguas foram retiradas da análise par a par de sequências. Na raiz de cada clado estão indicados os valores de probabilidade posterior (acima) e os tempos de divergência (abaixo). Os pontos de calibração utilizados a datação da duplicação foram as estimativas do tempo de divergência *Sophophora/Drosophila* (39,2 ± 3,35 Ma) e *D. melanogaster/D. simulans* (2,3 ± 0,65 Ma). A análise foi conduzida no BEAST v1.10.4. As setas indicam a junção entre as partes. Os subgêneros de *Drosophila* são delimitados pelos quadros azul (*Sophophora*), amarelo (*Drosophila*) e azul (*Dorsilopha*).

Figura Suplementar 3. Gráfico *heatmap* agrupando os grupos Drosophilidae e códons a partir da similaridade no uso de códons estimado pelo cálculo RSCU das sequências *unc-4*.

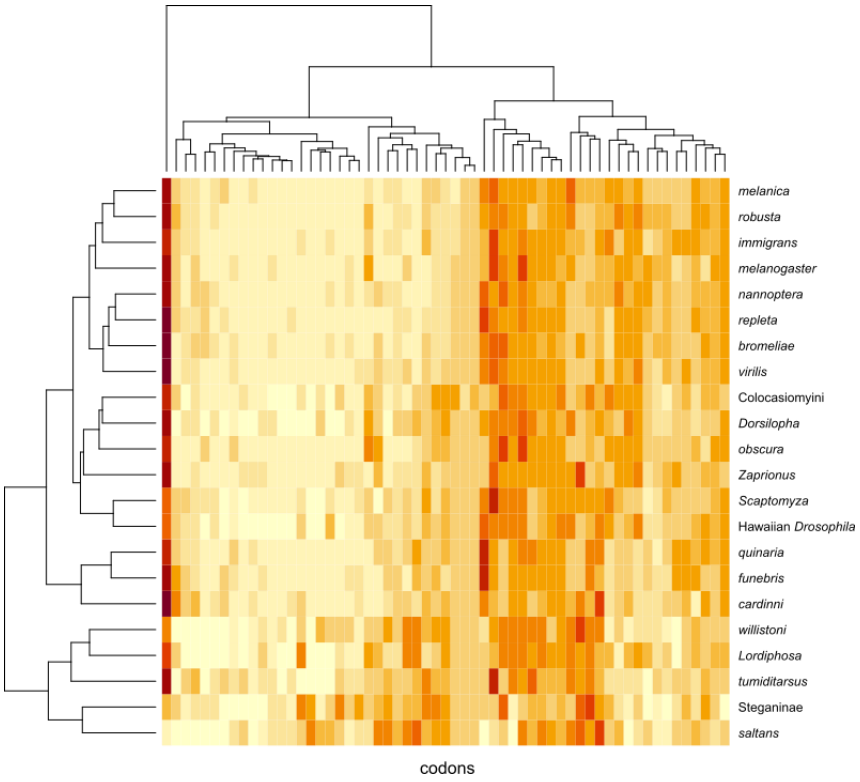


Figura Suplementar 4. Gráfico *heatmap* agrupando os grupos Drosophilidae e códons a partir da similaridade no uso de códons estimado pelo cálculo RSCU das sequências *OdsH*.

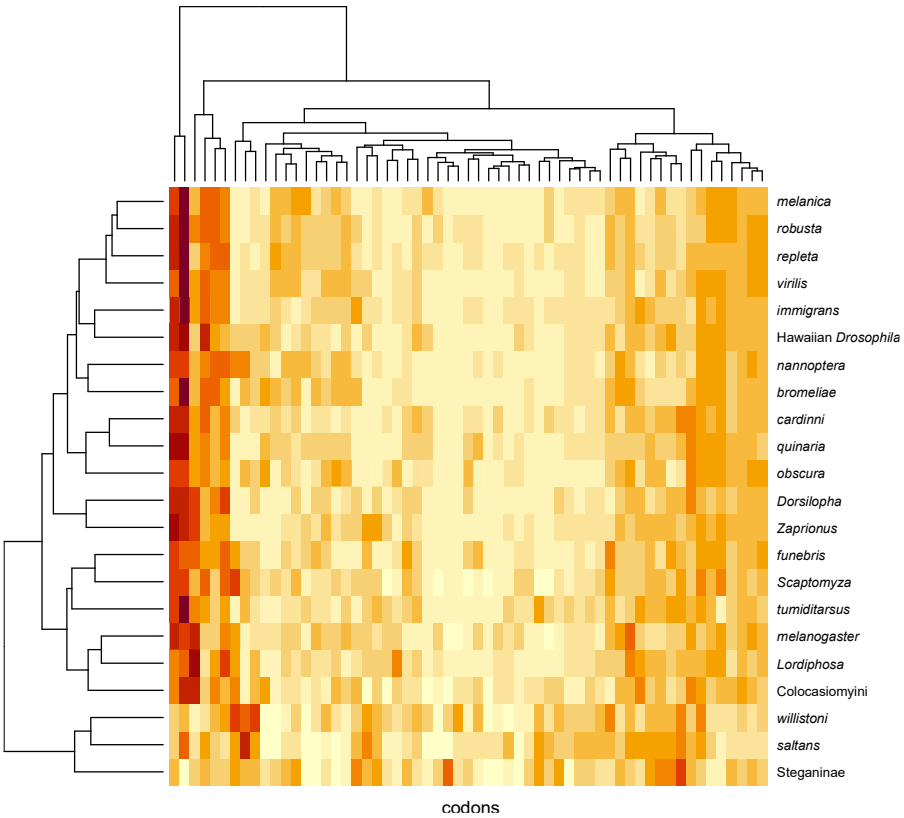


Figura Suplementar 5. % GC3 e ENC observados em sequências *unc-4* de Drosophilidae.

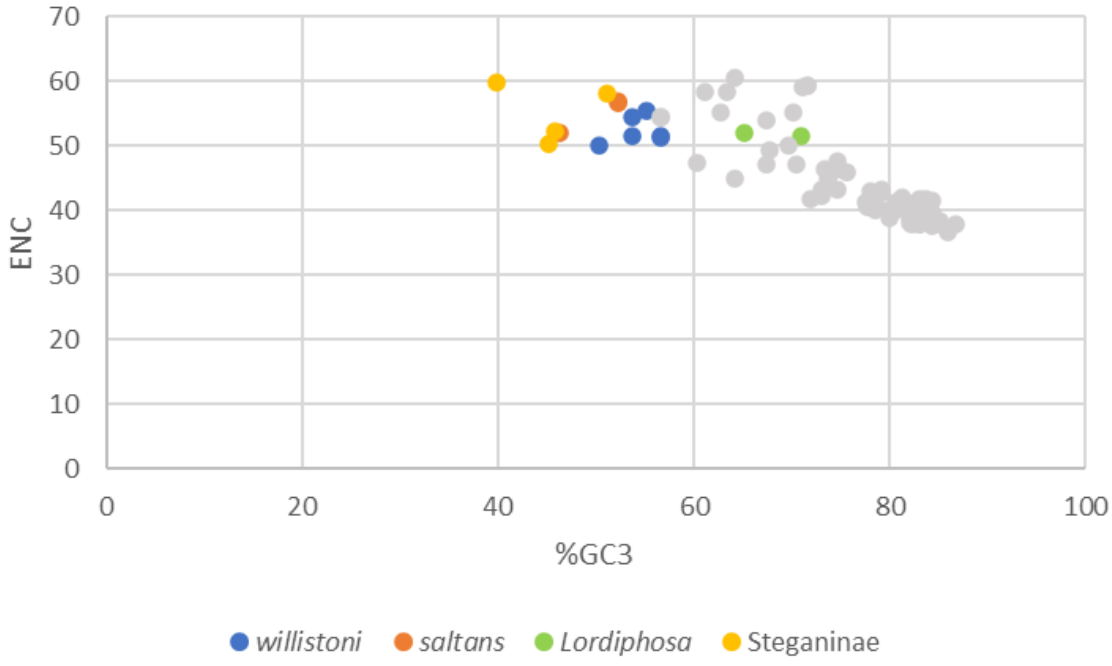


Figura Suplementar 6. %GC3 e ENC observados em sequências *OdsH* de Drosophilinae e ortólogo *unc-4* cópia única em Steganinae.

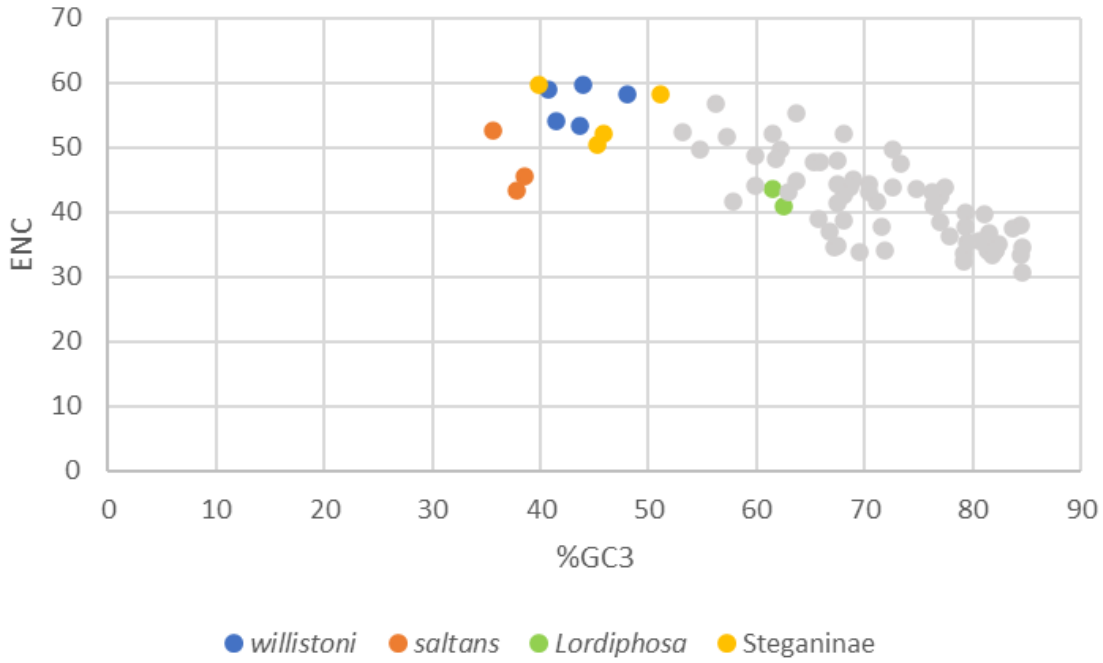


Figura Supplementar 7. Enriquecimento de termos GO de processo biológicos dos TBFSs enriquecidos nas regiões reguladoras de *OdsH* e *unc-4* em *Drosophila.*



APÊNDICE B – Tabelas Suplementares

Tabela Suplementar 1. Genomas de Drosophilinae anotados investigados e respectivos dos acessos das sequências de mRNA e aminoácidos utilizados. (continua)

Gênero **Subgênero** **Grupo** **Espécie**	**Genome ID**	***unc-4***	***OdsH***	**Unc-4**	**OdsH**
Drosophila					
Sophophora					
melanogaster					
D. melanogaster	47	NM_001298461.1	NM_078665.3	NP_001285390.1	NP_523389.3
D. simulans	200	XM_016174122.1	XM_016182480.1	XP_016039841.1	XP_016039839.1
D. sechellia	271	XM_002042375.2	XM_032726293.1	XP_002042411.1	XP_032582184.1
D. mauritiana	201	XM_033314566.1	XM_033314133.1	XP_033170457.1	XP_033170024.1
D. yakuba	148	XM_002101496.2	XM_039375143.1	XP_002101532.1	XP_039231077.1
D. erecta	250	XM_001977558.3	-	XP_001977594.3	-
D. suzukii	18317	XM_036820696.1	XM_036820876.1	XP_036676591.1	XP_036676771.1
D. subpulchrella	95496	XM_037868669.1	XM_037869533.1	XP_037724597.1	XP_037725461.1
D. ficusphila	3491	XM_017206456.1	XM_017206429.1	XP_017061945.2	XP_017061918.1
D. eugracilis	6863	XM_017229827.1	XM_017229847.1	XP_017085316.1	XP_017085336.1
D. takahashii	3493	XM_017140606.1	XM_017140608.1	XP_016996095.1	XP_016996097.2
D. biarmipes	3499	XM_017093384.1	XM_017092423.1	XP_016948873.1	XP_016947912.1
D. elegans	3490	XM_017256198.1	XM_017256104.1	XP_017111687.1	XP_017111593.1
D. rhopaloa	6853	XM_017126099.1	XM_017126098.1	XP_016981588.1	XP_016981587.1
D. kikkawai	3492	XM_017179476.1	XM_017179077.1	XP_017034965.1	XP_017034566.1
D. serrata	17008	XM_020949303.1	XM_020949435.1	XP_020804962.1	XP_020805094.1
D. ananassae	244	XM_001965984.3	XM_001965983.2	XP_001966020.3	XP_001966019.1
D. bipectinata	3489	XM_017239687.1	XM_017239680.1	XP_017095176.1	XP_017095169.1
obscura					
D. pseudoobscura	219	XM_001355153.4	XM_001355153.4	XP_001355189.3	XP_001355189.3
D. obscura	55893	XM_022371333.1	XM_022371338.1	XP_022227025.2	XP_022227030.2
D. persimilis	265	XM_026993102.1	XM_026993147.1	XP_026848903.1	XP_026848948.1
D. subobscura	64269	XM_034815904.1	XM_034815904.1	XP_034671795.1	XP_034671800.1
D. miranda	10915	XM_017301079.2	XM_017279317.1	XP_017156568.1	XP_017134806.1
D. guanche	74789	XM_034268271.1	XM_034267533.1	XP_034124162.1	XP_034123424.1
willistoni					
D. willistoni	251	XM_023179361.1	XM_002071068.2	XP_023035129.1	XP_002071104.2
Drosophila					
repleta					
D. arizonae	45090	XM_018017130.1	XM_018014500.1	XP_017872619.1	XP_017869989.1
D. mojavensis	259	XM_032732501.1	XM_002010399.3	XP_032588392.1	XP_002010435.1
D. navojoa	45091	XM_018110258.2	XM_018108884.2	XP_017965747.1	XP_017964373.1
D. hydei	10921	XM_023319443.1	XM_023319460.1	XP_023175211.1	XP_023175228.1

Tabela Suplementar 1. Genomas de Drosophilinae anotados investigados e respectivos dos acessos das sequências de mRNA e aminoácidos utilizados. (conclusão)

Gênero **Subgênero** **Grupo** **Espécie**	**Genome ID**	***unc-4***	***OdsH***	**Unc-4**	**OdsH**
Drosophila					
Drosophila					
virilis					
D. virilis	261	XM_002056953.3	XM_032440729.1	XP_002056989.3	XP_032296620.1
D. novamexicana	10918	XM_030710956.1	XM_030710953.1	XP_030566816.1	XP_030566813.1
immigrans					
D. albomicans	2712	XM_034263018.1	XM_034250207.1	XP_034118909.1	XP_034106098.1
quinaria					
D. innubila	76488	XM_034628969.1	XM_034627699.1	XP_034484860.1	XP_034483590.1
Dorsilopha					
D. busckii	38276	XM_017995536.1	XM_017995059.1	XP_017851025.1	XP_017850548.1
Drosophila **"havaianas"**					
D. grimshawi	247	XM_001992464.2	XM_032738902.1	XP_001992500.1	XP_032594793.1
Scaptodrosophila					
S. lebanonensis	38279	XM_030526514.1	XM_030526513.1	XP_030382374.1	XP_030382373.1

Tabela Suplementar 2. Genomas de Drosophilidae não anotados investigados. (continua)

Gênero **Subgênero** **Grupo** **Espécie**	**Genome ID**	***unc-4***	***OdsH***
Drosophila			
Sophophora			
obscura			
D. athabasca	71124	+	+
willistoni			
D. equinoxialis	10916	+	+
D. isularis	103454	+	+
D. nebulosa	https://web.stanford.edu/~bkim331/files/genomes/	+	+
D. paulistorum	100644	+	+
D. sucinea	100633	+	+
D. tropicalis	https://web.stanford.edu/~bkim331/files/genomes/	+	+
saltans			
D. neocordata	103442	+	+
D. prosaltans	100647	+	+
D. saltans	103445	+	+
Drosophila			
repleta			
D. repleta	10919	+	+
virilis			
D. americana	10917	+	+
D. ezoana	https://web.stanford.edu/~bkim331/files/genomes/	+	+
D. littoralis	100641	+	+
immigrans			
D. immigrans	100654	+	+
D. nasuta	55963	+	+
nannoptera			
D. achantoptera	https://web.stanford.edu/~bkim331/files/genomes/	+	+
D. wassermani	https://web.stanford.edu/~bkim331/files/genomes/	+	+
funebris			
D. funebris	103437	+	+
cardini			
D. arawakana	100640	+	+
D. cardini	103451	+	+
melanica			
D. micromelanica	75739	+	+
D. nigromelanica	75864	+	+

Tabela Suplementar 2. Genomas Drosophilidae não anotados investigados. (conclusão)

Gênero **Subgênero** **Grupo** **Espécie**	**Genome ID**	***unc-4***	***OdsH***
Drosophila			
Drosophila			
robusta			
D. robusta	75738	+	+
D. lacertosa	75740	+	+
tumiditarsus			
D. repletoides	100636	+	+
Drosophila **"havaianas"**			
D. murphyi	103459	+	+
D. sproati	103457	+	+
Lordiphosa			
L. clarofinis	103458	+	+
L. collinela	103460	+	+
L. magnipectinata	103464	+	+
L. mommai	103455	+	+
L. stackelbergi	103463	+	-
Scaptomyza			
S. flava	14175	+	+
S. hsui	100650	+	+
S. montana	103456	+	+
S. pallida	100649	+	+
Zaprionus			
Z. africanus	100646	+	+
Z. cameronensis	103452	+	+
Z. capensis	103444	+	+
Chymomyza			
C. amoena	Cedido por G.R. Dias e A.B. Carvalho	+	+
C. costata	100639	+	+
Colocasiomyia			
C. xenalocasiae	Cedido por G.R. Dias e A.B. Carvalho	+	-
STEGANINAE			
Cacoxenus indagator	Cedido por G.R. Dias e A.B. Carvalho	+	-
Leucophenga varia	103443	+	-
Rhinoleucophenga bivisualis	Cedido por G.R. Dias e A.B. Carvalho	+	-
Phortica variegata	38277	+	-

+: presença da sequência nos dados; -: ausência da sequência nos dados

Tabela Suplementar 3. Sequências das sondas utilizadas na reação de FISH

Sonda 1	taccttccgcgtaaatcagc
Sonda 2	cgaagctgtgagcgtaattt
Sonda 3	aatttgtttgtcgccgccgc
Sonda 4	cgagtctcacagcagccaaa
Sonda 5	ctcgacgcgagcgtccttat
Sonda 6	ttgcgcctgtcgtagtgacg
Sonda 7	tgagtgtgtgtgtgtgtgtg
Sonda 8	tgtgtatgtgatgcaacagt
Sonda 9	cacctgccccccacacaaat
Sonda 10	atcaactgctcaagcagctg
Sonda 11	cactcccagcctccgctgca
Sonda 12	cacttcgatggcgtgcatcc
Sonda 13	atccgcatccacagccacat
Sonda 14	atcatccaccacatccgcat
Sonda 15	ctccgctttgcagctgtatg
Sonda 16	atttctggccatgaatcagc
Sonda 17	ctagcgactcgagctatcca
Sonda 18	ctgccaggctgccacaagtg
Sonda 19	gtgcacatgctgcgccacat
Sonda 20	cgaagctgcagcagctccag
Sonda 21	gatgtggaggtgacagcaat
Sonda 22	caacctcacgcaacttgatg
Sonda 23	caatgaggtctcctccgcct
Sonda 24	ctccgcccaggatgatgaga
Sonda 25	cagctgccgccaatcgtcga
Sonda 26	cgcacgaatttcagcacctg
Sonda 27	gcctttctcagcagccacta
Sonda 28	cggacgtgtttatgcgcgag
Sonda 29	gccatgcggctggatctgaa
Sonda 30	gcatagcggtgtggttccag
Sonda 31	caagaaggagcacacgaaga
Sonda 32	cacaatgcccagccgcagag
Sonda 33	cgattccactgcaggagctg
Sonda 34	cgcatagccaaggcgatcga
Sonda 35	ctcgaggtcaacgtggagca
Sonda 36	cagctggagtccagcgacat

Tabela Suplementar 4. Testes de seleção por sítio em *OdsH.*

Modelo	Parâmetros	lnL	np	x2	P
Drosophilinae					
m1	p: 0.94621 0.05379 w: 0.06048 1.00000	-5208.84	97		
m2	p: 0.94621 0.05379 0.00000 w: 0.06048 1.00000 7.77649	-5208.84	99	0	1
m7	p = 0.48230 q = 5.13975	-5153.74	97		
m8	p0 = 0.99999 p = 0.48230 q = 5.13975 (p1 = 0.00001) w = 2.56876	-5153.74	99	0.0027	0.999
complexo melanogaster					
m1	p: 0.72272 0.27728 w: 0.19667 1.00000	-1090.02	11		
m2	p: 0.72272 0.18932 0.08796 w: 0.19667 1.00000 1.00000	-1090.02	13	0	1
m7	p = 0.58838 q = 0.85536	-1090.05	11		
m8	p0 = 0.74334 p = 3.52673 q = 12.75901 (p1 = 0.25666) w = 1.00000	-1090.01	13	0.074294	0.964

Tabela Suplementar 5. Sítios de ligação de fatores de transcrição diferencialmente enriquecidos na região reguladora a montante de *OdsH* em relação a *unc-4* em *Drosophila*.

ID do fator de transcrição	Fator de Transcrição	Deficit	No. TBFS encontrados	No. TBFS no *background*	P-Value	Log2 Proporção de ligação média	Log2 Enriquecimento	Score de significância
MA0012.1	br(var.3)	0.07	11	1	1.07E-04	-1.589593961	2.938599455	3.969633744
MA0170.1	C15	0.03	12	4	6.73E-04	-0.736965594	1.473931188	3.171960354
MA0232.1	lbl	0.09	12	4	6.73E-04	-0.736965594	1.473931188	3.171960354
MA0227.1	hth	0.08	9	1	0.002759456	-1.727411183	2.662965013	2.559176497
MA0252.1	vis	0.01	9	1	0.002759456	-1.727411183	2.662965013	2.559176497
MA0207.1	achi	0.06	9	1	0.002759456	-1.727411183	2.662965013	2.559176497
MA0249.2	twi	0.13	9	1	0.002759456	-1.727411183	2.662965013	2.559176497
MA0182.1	CG4328-RA	0.01	10	2	0.00332895	-1.286733431	2.070389328	2.477692733
MA0446.1	fkh	0.11	10	2	0.00332895	-1.286733431	2.070389328	2.477692733
MA0231.1	lbe	0.02	12	5	0.004576659	-0.592212286	1.184424571	2.339451441
MA0214.1	bsh	0.06	12	5	0.004576659	-0.592212286	1.184424571	2.339451441
MA0183.1	HHEX	0.12	12	5	0.004576659	-0.592212286	1.184424571	2.339451441
MA0177.1	CG18599	0.14	11	4	0.009422533	-0.797112711	1.353636955	2.025832318
MA0221.1	eve	0.09	11	4	0.009422533	-0.797112711	1.353636955	2.025832318
MA0228.1	ind	0.14	11	4	0.009422533	-0.797112711	1.353636955	2.025832318
MA0209.1	ap	0.11	11	4	0.009422533	-0.797112711	1.353636955	2.025832318
MA0189.1	E5	0.09	11	4	0.009422533	-0.797112711	1.353636955	2.025832318
MA0219.1	ems	0.07	11	4	0.009422533	-0.797112711	1.353636955	2.025832318
MA0195.1	Lim3	0.11	11	4	0.009422533	-0.797112711	1.353636955	2.025832318
MA0191.1	HGTX	0.07	11	4	0.009422533	-0.797112711	1.353636955	2.025832318
MA0257.1	zen2	0.11	11	4	0.009422533	-0.797112711	1.353636955	2.025832318
MA0225.1	ftz	0.07	9	2	0.012278138	-1.358928386	1.925999419	1.910867497
MA0013.1	br(var.4)	0.12	9	2	0.012278138	-1.358928386	1.925999419	1.910867497
MA0199.1	Optix	0.1	6	0	0.013729977	-2.793636331	3.700439718	1.862330187
MA0448.1	H2.0	0.09	12	6	0.013729977	-0.471708236	0.943416472	1.862330187
MA0254.1	vvl	0.07	11	5	0.027190739	-0.652359402	1.064130337	1.565578989
MA0049.1	hb	0.06	11	5	0.027190739	-0.652359402	1.064130337	1.565578989
MA0186.1	Dfd	0.13	11	5	0.027190739	-0.652359402	1.064130337	1.565578989
MA0215.1	btn	0.11	11	5	0.027190739	-0.652359402	1.064130337	1.565578989
MA0203.1	Scr	0.14	11	5	0.027190739	-0.652359402	1.064130337	1.565578989
MA0166.1	Antp	0.13	8	2	0.036074842	-1.439160722	1.765534746	1.442795564
MA0238.1	pb	0.07	10	4	0.036074842	-0.862734978	1.222392421	1.442795564
MA0230.1	lab	0.13	10	4	0.036074842	-0.862734978	1.222392421	1.442795564
MA0256.1	zen	0.1	10	4	0.036074842	-0.862734978	1.222392421	1.442795564
MA0245.1	slou	0.05	10	4	0.036074842	-0.862734978	1.222392421	1.442795564
MA0185.1	Deaf1	0.01	10	4	0.036074842	-0.862734978	1.222392421	1.442795564
MA0246.1	so	0.1	5	0	0.037267081	-2.91414038	3.459431619	1.428674626
MA0011.1	br(var.2)	0.04	5	0	0.037267081	-2.91414038	3.459431619	1.428674626
MA0204.1	Six4	0.01	5	0	0.037267081	-2.91414038	3.459431619	1.428674626
MA0458.1	slp1	0.1	12	7	0.037267081	-0.36848279	0.736965594	1.428674626
MA0233.1	mirr	0	12	7	0.037267081	-0.36848279	0.736965594	1.428674626
MA0206.1	abd-A	0.1	9	3	0.039125701	-1.116214972	1.440572591	1.407537865
MA0229.1	inv	0.1	9	3	0.039125701	-1.11621497	1.440572591	1.407537865
MA0916.1	Ets21C	0.06	0	5	0.037267081	-2.91414038	-3.459431619	-1.428674626

(continua)

Tabela Suplementar 5. Sítios de ligação de fatores de transcrição diferencialmente enriquecidos na região reguladora a montante de *OdsH* em relação a *unc-4* em *Drosophila.* (conclusão)

ID do fator de transcrição	Fator de transcrição	Déficit	No. TBFS encontrados	No. TBFS no *background*	P-Value	Log2 Proporção de Ligação Média	Log2 Enriquecimento	Score de significância
MA0244.1	Slbo	0.03	0	5	0.03727	-2.91414038	-3.459431619	-1.4286746
MA0178.1	CG32105	0.11	4	10	0.03607	-0.86273498	-1.222392421	-1.4427956
MA0240.1	Repo	0.09	4	10	0.03607	-0.86273498	-1.222392421	-1.4427956
MA0194.1	Lim1	0.13	4	10	0.03607	-0.86273498	-1.222392421	-1.4427956
MA1700.1	Clamp	0.07	2	8	0.03607	-1.43916072	-1.765534746	-1.4427956
MA0211.1	Bap	0.14	2	8	0.03607	-1.43916072	-1.765534746	-1.4427956
MA0239.1	Prd	0.06	5	11	0.02719	-0.6523594	-1.064130337	-1.565579
MA0126.1	Ovo	0.06	5	11	0.02719	-0.6523594	-1.064130337	-1.565579
MA0026.1	Eip74EF	0.05	1	7	0.02719	-1.89792964	-2.321928095	-1.565579
MA0243.1	Sd	0.04	1	7	0.02719	-1.89792964	-2.321928095	-1.565579
MA1455.1	dmrt99B	0.15	0	6	0.01373	-2.79363633	-3.700439718	-1.8623302
MA0023.1	dl(var.2)	0.14	2	9	0.01228	-1.35892839	-1.925999419	-1.9108675
MA0181.1	Vsx1	0.06	2	10	0.00333	-1.28673343	-2.070389328	-2.4776927
MA0457.1	PHDP	0.01	2	10	0.00333	-1.28673343	-2.070389328	-2.4776927

Tabela Suplementar 6. Sítios de ligação de fatores de transcrição enriquecidos na região reguladora a jusante de *OdsH* em relação a *unc-4* em *Drosophila.*

ID do fator de transcrição	Fator de transcrição	Deficit	No. De fatores de transcrição enriquecidos	No. Fatores de transcrição no *background*	P-value	Log 2 Proporção de ligação média	Log2 Enriquecimento	Score de Significância
MA0085.1	Su(H)	0.12	10	1	6.44E-04	-1.655216228	2.807354922	3.190983591
MA0180.1	Vsx2	0.13	11	4	0.009422533	-0.797112711	1.353636955	2.025832318
MA0016.1	usp	0.15	6	0	0.013729977	-2.793636331	3.700439718	1.862330187
MA0167.1	Awh	0.12	12	6	0.013729977	-0.471708236	0.943416472	1.862330187
MA0240.1	repo	0.09	11	5	0.027190739	-0.652359402	1.064130337	1.565578989
MA0187.1	Dll	0.02	11	5	0.027190739	-0.652359402	1.064130337	1.565578989
MA0239.1	prd	0.07	11	6	0.036613272	-0.531855353	0.823122238	1.436361454
MA0188.1	Dr	0.11	11	6	0.036613272	-0.531855353	0.823122238	1.436361454
MA0126.1	ovo	0.07	11	6	0.036613272	-0.531855353	0.823122238	1.436361454
MA0229.1	inv	0.09	11	6	0.036613272	-0.531855353	0.823122238	1.436361454
MA0255.1	z	0.09	11	6	0.036613272	-0.531855353	0.823122238	1.436361454
MA0447.1	gt	0.13	5	0	0.037267081	-2.91414038	3.459431619	1.428674626
MA0217.1	caup	0	9	3	0.039125701	-1.116214972	1.440572591	1.407537865
MA0210.1	ara	0	9	3	0.039125701	-1.116214972	1.440572591	1.407537865
MA0170.1	C15	0.04	7	12	0.037267081	-0.368482797	-0.736965594	-1.428674626
MA0183.1	HHEX	0.08	4	10	0.036074842	-0.862734978	-1.222392421	-1.442795564
MA0174.1	Dbx	0.03	2	8	0.036074842	-1.439160722	-1.765534746	-1.442795564
MA0212.1	bcd	0.15	2	8	0.036074842	-1.439160722	-1.765534746	-1.442795564
MA0012.1	br(var.3)	0.08	1	7	0.027190739	-1.897929642	-2.321928095	-1.565578989
MA0448.1	H2.0	0.07	6	12	0.013729977	-0.471708236	-0.943416472	-1.862330187
MA0211.1	bap	0	0	6	0.013729977	-2.793636331	-3.700439718	-1.862330187
MA0246.1	so	0.1	3	10	0.012278138	-1.044020017	-1.584962501	-1.910867497
MA0199.1	Optix	0	3	10	0.012278138	-1.044020017	-1.584962501	-1.910867497
MA0458.1	slp1	0.06	4	11	0.009422533	-0.797112711	-1.353636955	-2.025832318
MA0222.1	exd	0.05	1	8	0.009422533	-1.807643519	-2.502500341	-2.025832318
MA0460.1	ttk	0.11	0	9	3.37E-04	-2.519892433	-4.247927513	-3.47299035
MA0446.1	fkh	0.06	0	9	3.37E-04	-2.519892433	-4.247927513	-3.47299035