


Elias Alberto Gutierrez Carnelossi

Investigação de elementos de transposição em populações de
Drosophila mojavensis e *D. arizonae* e seus híbridos



Tese apresentada para
obtenção do Título de Doutor
em Genética.

São José do Rio Preto - SP
2014

Investigação de elementos de transposição em populações de
Drosophila mojavensis e *D. arizonae* e seus híbridos

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Genética, junto ao Programa de Pós-Graduação em Genética, Área de Genética Molecular e Evolução, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto.

Orientadora: Prof^a. Dr^a. Claudia Marcia Aparecida Carareto

Orientadora: Prof^a. Dr^a.Cristina Vieira

Carnelossi, Elias Alberto Gutierrez.

Investigação de elementos de transposição em populações de *Drosophila mojavensis* e *D. arizonae* e seus híbridos / Elias Alberto Gutierrez Carnelossi. -- São José do Rio Preto, 2014.
161 f. : il. ; 30 cm.

Orientador: Claudia Marcia Aparecida Carareto

Co-orientador: Cristina Vieira

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências, Letras e Ciências Exatas

1. Genética Molecular. 2. Hibridação 3. *Drosophila* – Genética. 4. *Drosophila mojavensis*. 5. *Drosophila arizonae*. 6. Elementos genéticos móveis. I. Carareto, Claudia Marcia Aparecida. II. Vieira, Cristina. III. Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências, Letras e Ciências Exatas. IV. Título.

CDU – 575.12

Ficha catalográfica elaborada pela Biblioteca do IBILCE
Campus de São José do Rio Preto - UNESP

Elias Alberto Gutierrez Carnelossi

Investigação de elementos de transposição em populações de
Drosophila mojavensis e *D. arizonae* e seus híbridos

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Genética, junto ao Programa de Pós-Graduação Genética, Área de Genética Molecular e Evolução, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto.

Comissão Examinadora

Prof^a. Dr^a. Claudia Marcia Aparecida CARARETO

UNESP – São José do Rio Preto

Orientadora

Prof^a. Dr^a. Cristina VIEIRA (orientadora - UCBL)

Prof. Dr. Élgion Lúcio da Silva LORETO

Prof^a. Dr^a. Marie France SAGOT

Prof^a. Dr^a. Lilian Madi RAVAZZI

São José do Rio Preto - SP
07 de Março de 2014

Este trabalho foi realizado sob convenção de co-tutela entre Universidade Estadual Paulista (UNESP) - Brasil e l'Université Claude Bernard (Lyon1 - UCBL) – França, no laboratório de Evolução Molecular, do Instituto de Biociências, Letras e Ciências Exatas (IBILCE/UNESP) – São José do Rio Preto/SP, e no *Laboratoire de Biométrie et Biologie Évolutive* (LBBE/Lyon1).

Universidade Estadual Paulista (UNESP – IBILCE)

Laboratório de Evolução Molecular
Departamento de Biologia
Rua Cristóvão Colombo, 2265, Jardim Nazareth
15054-000 São José do Rio Preto-SP
Brasil

Université Claude Bernard – Lyon 1

Laboratoire de Biométrie et Biologie Evolutive
CNRS UMR 5558
43 Boulevard du 11 novembre 1918
69622 Cedex Villeurbanne

Elias Alberto GUTIERREZ CARNELOSSI
elias-alberto.gutierrez-carnelossi@etu.univ-lyon1.fr
eliasgutzoo@hotmail.com

Palavras chave - Mots Clés – Key words

Elementos de Transposição - Eléments transposables – Transposable elements

Híbridos - Híbrides – Hybrids

Drosophila

Drosophila mojavensis

Drosophila arizonae

Ao meu pai, o mais forte guerreiro.

Agradeço aqui a todos que passaram ao meu lado antes e durante o doutorado, no trabalho ou fora dele, a todas as palavras de apoio, carinho e críticas, aos rostos e gestos; todo apoio e esforço pessoal físico e espiritual foram essenciais, meu muito obrigado. Mas, não posso deixar de mencionar e agradecer algumas pessoas em especial:

- Às minhas orientadoras: Claudia Marcia Carareto e Cristina Vieira, o meu mais profundo reconhecimento pela enorme confiança depositada, sempre acreditando e valorizando o trabalho com tamanha seriedade, competência, ética e determinação. Pelos seus valiosos ensinamentos, sem estas pessoas jamais poderia alcançar tal êxito. Pela enorme paciência durante todos esses anos, pelo profundo sentimento de segurança que passaram, por me acolherem e compreenderem minhas limitações, pela capacidade particular de enfrentarem e lidar com os problemas. Pude assim adquirir ensinamentos pessoais e profissionais que carregarei ao longo dos anos, recebam esses agradecimentos como « *un témoignage de profonde et affectueuse reconnaissance* » (testemunho de profundo reconhecimento).

- Aos meus pais e irmãos, pelo carinho, por compreenderem e me apoiarem na alegria ou nos momentos mais tristes, incentivarem e investirem nesse aprendizado, pelo amor incondicional que move nossa família. Aos tios e tias, primos e primas que estiveram juntos incentivando esse trabalho.

- A todos os funcionários, técnicos, alunos que contribuíram de alguma forma na obtenção dos resultados deste trabalho, aos servidores do Programa de Pós-Graduação em Genética, Departamento de Biologia do Instituto de Biociências (IBILCE), Letras e Ciências Exatas, da Universidade Estadual Paulista e à todos do “Laboratoire de Laboratoire de Biométrie et Biologie Evolutive (LBB) e Université Claude Bernard Lyon1.

- Aos amigos que estiveram nestes anos no laboratório de Evolução Molecular: Adriana Granzotto (obrigado especial por todo ensinamento e ajuda na bancada), grande Wellington, Leliane, Lilian, Joice, Elaine, Eric, Camila e a todos aqueles que passaram pelo laboratório durante minha estada.

- A todos da equipe TREEP do LBBE, por me acolherem de braços abertos, Emmanuelle, Marie, Matthieu, Virginie, H  l  ne, Laurent, Annabelle, Cl  ment, Clemence, Th  r  se, agradecimento especial ao Abdou Akkouche e Nelly Buret, pela enorme ajuda na bancada, profundo ensinamento e compreens  o.

-A todos da Equipe Baobab, especialmente    Marie-France, Paulo, Gustavo, Susan, Cec  lia, Ricardo, grandes companheiros, obrigado pelo incentivo e pela ajuda.

- Aos amigos e companheiros brasileiros em Lyon e estrangeiros que me acolheram, Rochele, Camila, Ana Paula, Matthaus, Bruno, Renata, Myriam, Luise, Giannis e Claudia.

- Aos amigos que pude conhecer e confiar ainda mais nesse companheirismo que nos une: C. Eduardo e primo Breno, Nei. Le fou agradece Caion, Gustavo, pelas noites em Paris, Elizeu, Zez  o, Andr   e Wellington.

-    Ana Luisa, que com toda compaix  o e delicadeza de uma flor, com todo seu amor e carinho, me deu apoio e confiou para que pudesse enfrentar grandes momentos de alegria e dificuldades, meu porto seguro, obrigado.

- Aos financiadores durante a Tese no Brasil e na Fran  a: FAPESP, CAPES, CNRS e ANR.

RESUMO

Investigação de elementos de transposição em populações de *Drosophila mojavensis* e *D. arizonae* e seus híbridos

Essa tese teve como principal objetivo investigar como os elementos de transposição (TEs), sequências capazes de se movimentarem no genoma hospedeiro, e entre diferentes espécies, se comportam em híbridos interespecíficos. Buscou-se compreender a atividade do TEs em cruzamentos entre *Drosophila mojavensis* e *D. arizonae* devido a algumas características particulares desse par de espécies. São espécies irmãs, com divergência muito recente (< 1 milhão de anos), apresentam isolamento reprodutivo pré- e pós-zigótico incompleto em laboratório e uma delas, *D. mojavensis*, teve seu genoma sequenciado, o que permitiu a realização de análises *in silico* de TEs dessa espécie. A tese foi dividida em quatro capítulos. O primeiro apresenta uma análise detalhada do retrotransposon NLTR denominado *I*, conhecido por causar disgenesia híbrida em *D. melanogaster*. Foram identificadas e caracterizadas sequências putativamente ativas de elementos semelhantes ao *I* no genoma de *D. mojavensis* e realizadas análises filogenéticas que mostraram que as sequências *I* de *D. mojavensis* e aquelas encontradas em outras espécies de *Drosophila* pertencem a diferentes famílias *I*. Análises de expressão por RTq-PCR mostraram que esse elemento é transcricionalmente ativo nos ovários e testículos de ambas as espécies, e em seus híbridos, e que têm elevada expressão em testículos, mas não em ovários dos híbridos, o que poderia ser associado ao fenótipo de esterilidade do macho híbrido. No segundo capítulo são apresentadas análises por RNA-Seq de TEs expressos em ovários de duas linhagens parentais e seus

híbridos F1. Os resultados mostram expressão espécie-específica de diferentes elementos nas espécies parentais e híbridos; e, de forma inédita, que apesar de alguns TEs estarem superexpressos nos híbridos, encontram-se de forma geral regulados em relação aos parentais. No terceiro capítulo são apresentados resultados de expressão quantificada por RTq-PCR de quatro retrotransposon (*Helena*, *I*, *Copia* e *Osvaldo*) e por fim, no último capítulo, abordamos os resultados das estimativas do tamanho do genoma (C-value) nas espécies parentais e em híbridos recíprocos. tese revela um cenário da expressão de TE específico em *D. mojavensis* e *D. arizonae*, com raras exceções, pode auxiliar a compreensão da complexidade da dinâmica e ação destes elementos móveis no processo de especiação de diferentes espécies.

Palavras-chave: Elementos de transposição; híbridos; *Drosophila*.

ABSTRACT

Activity of transposable elements in populations of *Drosophila mojavensis* and *D. arizonae*
and in their hybrids

This thesis aimed to investigate how transposable elements (TEs), sequences capable of moving into the host genome and between different species, behave in interspecific hybrids. We seek to understand the activity of TEs in crosses between *Drosophila mojavensis* and *D. arizonae* due to some particular characteristics of this species pair. They are sister species, with very recent divergence (< 1 million years), present incomplete pre- and post-zygotic reproductive isolation in laboratory and one of them, *D. mojavensis*, had its genome sequenced, which allowed for *in silico* analyses of TEs. The thesis is divided into four chapters. The first presents a detailed analysis of a non-LTR retrotransposon called *I*, known to cause hybrid dysgenesis in *D. melanogaster*. Putatively active sequences similar to the *I* element were identified and characterized in the genome of *D. mojavensis*. The performed phylogenetic analyzes showed that the *I* sequences in *D. mojavensis* and those harbored by other *Drosophila* species belong to different *I* families. Expression analyses by RTq-PCR showed that this element is transcriptionally active in ovaries and testes of both species and their hybrids, and have high expression in the testes, but not in the hybrids ovaries, which could be associated with the male hybrid sterility phenotype. In the second chapter are presented analyses of expressed TEs in the ovaries of two parental strains and their hybrids by RNA-Seq. The results show species-specific expression of TEs in the parents and hybrids; and, in an unprecedented manner, that TEs are generally regulated in hybrids regarding with

their parents, although some of them are overexpressed. In the third chapter are presented results of expression of four retrotransposons (*Helena*, *I*, *Copia* and *Oswaldo*) quantified by RTq-PCR; and finally, in the last chapter, we presented estimates of the genome size (C-value), in both parental species and reciprocal hybrids. Overall, this thesis reveals a scenario of expression of specific TEs in *D. mojavensis* and *D. arizonae*, and its regulation in hybrids with rare exceptions, which can help us to understand the complexity of the dynamics and action of these mobile elements in the speciation process of different species.

Keywords: Transposable elements; Hybrids; *Drosophila*.

LONG RESUME

Activité d'éléments transposables dans les populations de *Drosophila mojavensis* et *D. arizonae* et chez leurs hybrides

Les ET ont un rôle important dans l'évolution, puisque ce sont des séquences d'ADN qui ont la capacité de se déplacer dans le génome hôte, et peuvent ainsi s'insérer à l'intérieur ou près de gènes et modifier leur fonctionnement. Ces séquences sont en nombre variable dans le génome des différentes espèces et peuvent atteindre un niveau important dans le génome de certaines, comme par exemple 45% du génome humain, ou même plus 85% du génome du maïs (BRUGGMANN, BHARTI et al., 2006; LI et al., 2009; ZHANG, YU et al., 2009) . Les ET sont aussi responsables, de par la nature répétée de leurs séquences, de réarrangements chromosomiques qui sont source de mutation. Chez les hybrides inter-spécifiques, quelques travaux suggèrent que la mobilisation d'ET peut induire des dysfonctionnements importants qui peuvent potentiellement agir comme des inducteurs efficaces de spéciation (VOLFF et al., 2001; DOBIGNY et al., 2004; MICHALAK, 2009; REBOLLO et al., 2010). Cette thèse vise à étudier la façon dont les éléments transposables (ETs), les séquences capables de se déplacer dans le génome de l'hôte entre des espèces différentes, se comportent dans les hybrides interspécifiques. Nous cherchons à comprendre l'activité dans les croisements entre la *Drosophila mojavensis* et *D. arizonae* en raison de certaines caractéristiques particulières de cette paire d'espèces. Ils sont des espèces soeurs, avec une divergence très récente (< 1 million d'années), présente incomplète isolement pré- et post-zygotique reproduction en laboratoire et l'un d'eux (REED et al., 2007; WASSERMAN, KOEPPER, 1987). *D. mojavensis*, ont

séquencé le génome, ce qui a permis les analyses *in silico* des ETs. La thèse est divisée en quatre chapitres. Le premier présente une analyse détaillée d'un rétrotransposon non-LTR appelé I, connu pour causer dysgénésie hybride chez *D. melanogaster*. Putative séquences actives semblables à l'élément de I ont été identifiées et caractérisées dans le génome de *D. mojavensis*. Les analyses phylogénétiques ont montré que les séquences I chez *D. mojavensis* et nourri par ceux d'autres espèces de *Drosophila* appartiennent à des familles différentes d'ET. Les analyses d'expression par RTq-PCR ont montré que cet élément est une activité de transcription dans les ovaires et les testicules des deux espèces et leurs hybrides, et ont une grande expression dans les testicules, mais pas dans les ovaires des hybrides qui pourraient être associés avec le phénotype de stérilité mâle hybride. Dans le deuxième chapitre sont présentées les analyses des ETs exprimés dans les ovaires des deux souches parentales et leurs hybrides par RNA-Seq. Les résultats montrent l'expression d'ETs spécifiques chez les parents et les hybrides, et, d'une manière générale, que les ETs sont régulés de manière similaire chez les hybrides avec leurs parents, bien que certains d'entre eux soient surexprimés. Dans le troisième chapitre sont présentés les résultats de l'expression de quatre rétrotransposons (*Helena*, *I*, *Copia* et *Osvaldo*) quantifiés par RT-qPCR et, enfin, dans le dernier chapitre, nous sommes présentés des estimations de la taille du génome (C - valeur), à la fois pour les espèces parentales et les hybrides réciproques. Dans l'ensemble, la thèse révèle un scénario d'expression de TE spécifique chez *D. mojavensis* et *D. arizonae*, et sa régulation dans les hybrides de rares exceptions près, qui peut nous aider à comprendre la complexité de la dynamique et de l'action de ces éléments mobiles dans le processus de spéciation de différentes espèces.

References

- BRUGGMANN, R. et al. Uneven chromosome contraction and expansion in the maize genome. **Genome Research**, v.16, n.10, p.1241-1251, 2006.
- DOBIGNY, G.; OZOUF-COSTAZ, C.; WATERS, P.D.; BONILLO, C.; COUTANCEAU, J.P.; VOLOBOUEV, V. LINE-1 amplification accompanies explosive genome repatterning in rodents. **Chromosome Research**, v.12, p.787-793, 2004.
- LI, W-H; GU, Z; WANG, H; NEKRUTENKO, A. Evolutionary analyses of the human genome. **Nature**, v.409, p.847-849, 2001.
- KOEPFER, H.R. Selection for sexual isolation between geographic forms of *Drosophila mojavensis*. II. Effects of selection of mating preference and propensity. **Evolution**, 41:1409-141, 1987.
- MICHALAK, P. Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. **Heredity**, v.102, p.45-50, 2009.
- REBOLLO, R.; HORARD, B.; HUBERT, B.; VIEIRA, C. Jumping genes and epigenetics: Towards new species. **Gene**, v.454, p.1-7, 2010.
- REED, L.K.; NYBOER, M.; MARKOW T.A. Evolutionary relationships of *Drosophila mojavensis* geographic host races and their sister species *Drosophila arizonae*. **Molecular Ecology**, v.16, p.1007-1022, 2007.
- VOLFF, J.N.; KORTING, C.; MEYER, A.; SCHARTL, M. Evolution and discontinuous distribution of Rex3 retrotransposon in fish. **Molecular Biology Evolution**, v.18, p.427-431, 2001.
- WASSERMAN, M.; KOEPFER, H.R. Character displacement for sexual isolation between *Drosophila mojavensis* and *Drosophila arizonensis*. **Evolution**, v.31, p.812-823, 1977.
- ZHANG, J. et al. Alternative Ac/Ds transposition induces major chromosomal rearrangements in maize. **Genes & development**, v.23, n.6, p.755-65, 2009

Sumário

INTRODUÇÃO GERAL.....	17
OBJETIVOS.....	31
CAPÍTULO 1- Manuscrito: Specific activation of an I-like element in <i>Drosophila</i> interspecific hybrids.....	33
CAPÍTULO 2: Mobiloma expresso em <i>Drosophila mojavensis</i> e <i>D. arizonae</i> e seus híbridos: análises preliminares.....	70
CAPÍTULO 3: Expressão de elementos de transposição em <i>Drosophila mojavensis</i> , <i>D. arizonae</i> e híbridos quantificados por RT-qPCR.....	103
CAPÍTULO 4: Diversidade no tamanho do genoma em híbridos de <i>Drosophila</i>	127
DISCUSSÃO GERAL.....	147
CONCLUSÕES.....	153
REFERÊNCIAS.....	155

INTRODUÇÃO GERAL

INTRODUÇÃO GERAL

Os elementos de transposição, classificação e estrutura

A principal questão deste trabalho foi verificar se os elementos de transposição (TEs) são desregulados no genoma de híbridos interespecíficos em comparação com os parentais. Os TEs são sequências móveis que afetam o genoma hospedeiro por sua habilidade de se movimentar e replicar. Em decorrência dessas duas habilidades causam mutações, alteram o genoma estruturalmente, causando duplicações, deleções e inversões, e funcionalmente, afetando a expressão de genes específicos ou de redes gênicas (BIÉMONT; VIEIRA, 2006; FESCHOTTE, 2007). Essas mudanças muitas vezes resultam em efeitos fenotípicos complexos.

A primeira classificação dos TEs, proposta em 1989, distinguia duas classes pela forma intermediária da sequência durante o processo de transposição: com uma molécula de RNA, Classe I, ou apenas na forma de DNA, Classe II (FINNEGAN, 1989). O mecanismo de transposição da Classe I é comumente *copy-and-paste* (copia e cola) e, o da Classe II, *cut-and-paste* (corta e cola). A descoberta de que alguns TEs em bactérias (DUVAL-VALENTIN et al., 2004) e em eucariotos (LAI et al., 2005; MORGANTE et al., 2005) tinham transposição do tipo *copy and paste*, mas sem um RNA intermediário, desafiou o sistema de duas classes e uma terceira classe foi proposta, utilizando na classificação um critério enzimológico (CURCIO et al., 2003). A proposta de classificação mais recente, e aparentemente a mais adotada atualmente é a de Wicker et al. (2007), reconcilia os dois tipos de classificação mantendo as duas classes, mas aplica também critérios enzimáticos. É um sistema de classificação hierárquica que inclui os níveis classe, subclasse, ordem, superfamília, família e subfamília. O nível classe divide os TEs em Classe I e Classe II, como previamente, e o nível

subclasse é usado para distinguir apenas elementos da Classe II (na Subclasse I incluem-se os clássicos transposons de DNA do tipo *cut-and-paste* e na Subclasse II, os elementos com processo de transposição distinto dos anteriores). O nível ordem classifica os TEs pelo mecanismo de inserção, organização geral e enzimologia. As superfamílias dentro de uma ordem compartilham estratégias de replicação, mas são distinguidas pela estrutura dos domínios proteicos ou não codificantes; as famílias são definidas pela conservação de suas sequências e as subfamílias por agrupamento filogenético das sequências de uma família (WICKER et al., 2007)

Os elementos de *Classe I* (Retrotransposons) se replicam usando um transcrito de RNA como intermediário para transposição, utilizando a enzima transcriptase reversa. Nesta classe incluem-se cinco ordens (LTR, DIRS, PLE, LINE e SINE), dentre elas destacam-se os *retrotransposons com LTRs* (Repetições Terminais Longas) e *sem LTRs* (NLTRs). Os elementos com LTRs possuem sequências diretamente repetidas em cada extremidade, essenciais para o processo de transcrição reversa e integração; tipicamente contêm ORFs, quadros aberto de leitura ou “*open reading frame*”, para GAG, uma proteína estrutural de partícula viral e para POL. *Pol* codifica uma proteinase aspártica (AP), uma transcriptase reversa (RT) e uma RNase H. Os retrotransposons NLTRs, ou retrotransposons poli(A), não possuem repetições terminais porém apresentam uma sequência poli(A) na extremidade 3'. Os elementos NLTRs são de dois tipos: os LINEs (*Long Interspersed Nuclear Elements*) que consistem de uma região 5' não traduzida (UTR) com promotor ativo, duas ORFs e uma 3'UTR com uma cauda poly-A, uma repetição em *tandem* ou apenas um região rica em Adenina; e os SINEs (*Short Interspersed Nuclear Elements*), que são elementos não autônomos (WICKER et al., 2007). Em adição aos clássicos retrotransposons acima descritos, as ordens DIRs (GOODWIN; POULTER, 2004) e PLE (EVGEN'EV et al., 1997) apresentam terminações e mecanismos de transposição não usuais e não serão aqui descritos.

Os elementos da *Classe II* – denominados de transposons de DNA - não requerem o passo da transcrição reversa para se integrarem no genoma porque codificam para uma transposase que excisa o TE de seu sítio hospedeiro e o integra em um novo local no genoma. A Subclasse I contém duas ordens, a ordem TIR, caracterizada pelas suas repetições terminais invertidas (do Inglês *Terminal Inverted Repeats*) e ordem Crypton; e a subclasse II inclui transposons das ordens Helitron e Maverick, cujos processos de transposição envolvem replicação sem clivagem da dupla fita, muito diferente do processo dos elementos da subclasse I (WICKER et al., 2007).

Os TEs podem ainda ser classificados em autônomos e não-autônomos. Aqueles que podem sintetizar as proteínas necessárias para sua mobilização são chamados de elementos autônomos, e aqueles elementos que não tem suas sequências codificadoras de proteínas necessárias para a transposição íntegras são chamados de *não-autônomos*, como os *SINEs* e *MITEs*. Embora sem os domínios codificadores de proteínas funcionais, esses elementos podem se mobilizar no genoma se as proteínas necessárias para a sua transposição forem fornecidas por outros elementos do genoma a eles relacionados (FESCHOTTE, et al., 2002).

Ciclo evolutivo dos TEs

Os elementos de transposição são encontrados no genoma de praticamente quase todos os organismos, e podem ser mantidos nos genomas hospedeiros por longos períodos evolutivos (KIDWELL; LISCH, 1997). A única exceção em que não foram indentificados TEs em um genoma refere-se a *Plasmodium falciparum*, e provavelmente em algumas espécies próximas (SARKAR et al. 2003). A ampla distribuição de todas as grandes classes de TEs, por exemplo, ao longo da evolução dos eucariotos, indica que estas sequências residem no genoma há muito tempo (WICKER et al., 2007). Entretanto, as taxas de amplificação de TEs não são uniformes ao longo do tempo. Resultados sugerem que a vida

longa de certos elementos, como o elemento Alu em humanos, se deve a baixa atividade ao longo do tempo, o que ocasionalmente produz novas cópias, algumas das quais podem se tornar altamente ativas. Essa atividade pode ou não ser deletéria ou fazer com que os elementos persistam ao longo do tempo, sendo portanto, a atenuação da mobilização a estratégia evolutiva normal que ocorre para vários elementos de transposição (HAN; BOEKE, 2004;CORDAUX; BATZER, 2009).

No entanto, como qualquer outra sequência no genoma, os TEs são submetidos a processos evolutivos, tais como substituições de nucleotídeos, deleções e inserções, bem como outros tipos de rearranjos que podem conduzir a degradação de suas sequências e posterior extinção de cópias ativas do genoma hospedeiro. Ainda assim, podem posteriormente ser (re) introduzidos em uma espécie após terem sido degradados. Tal (re) introdução pode ser decorrente da origem de sequências ativas pela recombinação de sequências de TEs erodidas, pré-existentes (origem *de novo*), ou por transferência horizontal entre espécies, mediada por vetores específicos ou por hibridação e introgressão. Posteriormente, os novos integrantes do genoma podem se espalhar por transmissão vertical a todas as populações da espécie a partir das linhagens ascendentes. Vários casos de transferência horizontal de transposons de DNA e de retrotransposons LTR têm sido relatados, mas para os NLTRs o modo de transmissão vertical, com apenas algumas exceções, parece ser a regra (para revisões recentes ver SCHAACK et al., 2010; DE ALMEIDA; CARARETO, 2004; CARARETO 2011; WALLAU et al., 2012).

Impacto dos elementos de transposição nos genomas

Os TEs são onipresentes em quase todas as espécies já investigadas. Foram descobertos primeiramente no genoma do milho por Barbara McClintock (MCCLINTOCK 1956; 1984) que os classificou como “elementos controladores”, ou como alelos mutantes

“instáveis”, por que alteravam o padrão de expressão do gene da antocianina, responsável pela pigmentação do grão. Um exemplo semelhante, e reportado quando os TEs já eram amplamente reconhecidos como agentes produtores de variabilidade genética, é o de mutações no gene *pallida* na espécie *Antirrhinum majus* (boca-de-leão), em que a presença e excisão do elemento Tam3 alteravam a pigmentação das flores (MARTIN et al., 1985).

O impacto dos TEs na evolução dos genomas pode ser tanto positivo como deletério. De modo geral esse impacto é deletério, em curto prazo. Novas inserções de TEs, ou produto de recombinações entre cópias do mesmo TE, em cromossomos homólogos ou não homólogos, podem ser mutagênicas, e podem alterar a função dos genomas com impacto fenotípico imediato (como exemplo, as inserções dos elementos L1, Alu e SVA em humanos), causando doenças hereditárias como hemofilia, fibrose cística, síndrome de Apert, neurofibromatoses, câncer de cólon, etc. (CORDAUX; BATZER, 2009).

Contudo, a longo prazo, o acúmulo dessas sequências pode levar à divergência entre espécies. Por causa de sua contínua atividade e acumulação, os retrotransposons L1, Alu e SVA, por exemplo, tiveram um impacto importante na evolução do genoma de primatas em termos de estrutura e função. Embora as taxas de acúmulo desses TEs tenham sido uniformes ao longo do tempo, há evidências de variação desses três elementos entre esses genomas, e esta variação nas taxas de amplificação ocorreu em um curto período de tempo de divergência das duas espécies (~6 Myr atrás), sugerindo forte influência do nível populacional hospedeiro nestas taxas de amplificação (revisão em CORDAUX; BATZER, 2009). Também podemos citar o impacto dos TEs na evolução dos genomas dentro de uma espécie, pelas mudanças no número de cópias durante processos de migração, como descrito em linhagens de *D. simulans* (VIEIRA et al., 1999). Em pequenas populações a eficiência da seleção contra efeitos deletérios de inserções de TEs pode ser bastante reduzida e processos aleatórios podem levar ao acúmulo de TEs (BIÈMONT; VIERA, 2006).

Dentre os efeitos da mobilização dos TEs, que em longo prazo se tornam importantes para o genoma hospedeiro, destaca-se sua domesticação. Os TEs podem influenciar a expressão de genes vizinhos, em ambos os níveis, transcricional e pós-transcricional. Tem sido amplamente demonstrado que promotores dos genomas hospedeiros contêm sequências derivadas de TEs. Análises em escala genômica mostram que muitos promotores e sinais de poliadenilação em genes de humanos e camundongos são derivados de TEs primata- e rodentia-específicos (VAN DE LAGEMAAT et al., 2003; JORDAN et al., 2003; MARINO et al., 2005; THORNBURG et al., 2006; CARARETO et al., 2014).

Retrotransposon podem ainda participar na reconstituição do DNA, particularmente dos telômeros. As espécies do gênero *Drosophila* não contêm a telomerase, enzima que em outros organismos participa da recomposição dos telômeros em decorrência da perda de suas sequências terminais característica da duplicação cromossômica. Estudos em *D. melanogaster* mostraram que a transcrição reversa dos NLTRs HeT-A, TART e THARE nas extremidades dos telômeros replicados repõe as sequências perdidas durante a replicação, podendo criar telômeros até mesmo em extremidades de DNA deletado. Sem esse mecanismo, as extremidades cromossômicas ficariam mais curtas a cada replicação, podendo até ocorrer perda de genes (PARDUE; DEBARYSHE, 2003). Os retrotransposons também afetam a expansão centromérica, como explosões de transposição e diferenciação na metilação reportada em gibão-hoolok-oriental (*Hoolock leuconedys*), da família Hylobatidae (CARBONE et al., 2012). Em milho, elementos membros da família hAT de transposons de DNA são capazes de gerar rearranjos cromossômicos, incluindo deleções e inversões e, além disso, podem juntar sequências de dois genes parálogos codificantes e regulatórios para gerar uma série de genes quiméricos, alguns dos quais são funcionais (ZHANG, et al., 006).

Mecanismos de controle da transposição

Os organismos desenvolveram estratégias que permitem controlar e minimizar os efeitos da transposição. Os mecanismos de regulação são diversos, e muitos deles ainda não são totalmente compreendidos. TEs podem se auto-regular ou serem regulados por fatores do hospedeiro. Como mecanismo de controle da transposição pode ser citado a autoregulação tecido-específica do elemento *P* em *D. melanogaster*. A cópia completa do elemento *P* canônico tem 2.907pb que contém quatro exons. Em células do tecido germinativo todos os quatro exons são necessários para codificar uma transposase. Em células somáticas, um intron é retido produzindo uma transposase truncada, que age como um repressor da transposição (MISRA; RIO, 1990; REISS, et al., 2003). Este mecanismo de autorregulação permite ao elemento estar ativo em células germinais e ocupar novas localizações nos genomas descendentes, mas não sofre transposição em células somáticas nas quais os danos genéticos da transposição poderia causar redução no valor adaptativo do hospedeiro.

Metilação do DNA, modificações na cromatina e RNA de interferência (RNAi) tem sido implicados na repressão da atividade de TEs (O'DONNELL; BOEKE, 2007). Em *Drosophila*, os RNAi são mediadores essenciais de silenciamento em células germinativas, produzidos em ambos os sexos, mas somente carregados da mãe para o embrião (CHAMBEYRON et al, 2008).

A partir destas descobertas passou-se a compreender que os TEs são parte essencial da constituição dos genomas e que também são mediados por modificações epigenéticas. Em condições estáveis, ambientais, de combinações específicas de TEs, e da estrutura da cromatina, o DNA acumularia mutações neutras. Por outro lado, se o ambiente é altamente alterado, TEs e marcas epigenéticas podem ser modificadas dando lugar a novas situações, se inserindo próximos de genes e assim modificando sua expressão (FABLET; VIEIRA, 2011).

Todos os mecanismos acima citados podem resultar em variações no tamanho dos genomas e diferenças relativas na proporção de TEs no organismo hospedeiro, expressas pela variação do número de famílias e de cópias de TEs específicos. Esta variação pode ocorrer via explosões de transposição que podem induzir a rápida especiação (REBOLLO et al., 2010). Segundo Rebollo et al. (2010), para que explosões de transposição promovam rápida especiação, os TEs devem levar à reestruturação do genoma ou provocar mutações em genes de especiação. Significante atividade de TEs é observada nas espécies de uma família de morcegos (Vespertilionidae), na qual se demonstrou terem ocorrido múltiplas ondas de amplificação de diversos TEs nos últimos 40 milhões de anos (RAY et al., 2008). No genoma de *Myotis lucifugus*, pertencente a um dos gêneros de morcegos mais ricos em espécie do mundo, foram identificadas sete novas famílias de TE-exclusivas. Os autores concluíram que as ondas de atividades de TEs foi o mecanismo gerador da grande variabilidade genética e diversidade de espécies do gênero.

Híbridos e elementos de transposição

A hibridação interespecífica pode ser considerada como uma condição de estresse com múltiplas consequências para o genoma do híbrido, podendo ocasionar rearranjos diferenciais, inversões, deleções, alteração da expressão gênica, variações na metilação no DNA, disgenesia, entre outros efeitos (FONTDEVILA, 2005; ARKHIPOVA; RODRIGUEZ, 2013). Adicionalmente, pode ocorrer uma ativação global dos TEs que induza profundas modificações no DNA hospedeiro. Tais mudanças podem gerar novos fenótipos ou a formação de populações reprodutivamente isoladas, uma vez que o acúmulo de modificações genômicas estruturais e funcionais atuaria como uma pressão levando à especiação (HEDGES; DEININGER, 2007; OLIVER; GREENE, 2009; REBOLLO et al., 2010).

Considera-se que a rápida reorganização do genoma em híbridos levou à origem de três novas espécies diploides de girassol (gênero *Helianthus*). Observou-se nas espécies derivadas de hibridação um aumento no número de cópias de alguns elementos em relação aos parentais e novos rearranjos cariotípicos (UNGERER et al., 2006; UNGERER et al., 2009). Híbridos interespecíficos de cangurus da família Macropodidae também apresentaram variação na amplificação de uma repetição satélite e do elemento KERV-1, além de mudanças na estrutura da cromatina e rearranjos de braços cromossômicos inteiros (METCALFE et al., 2007). Isto demonstra durante a especiação por hibridação que ocorre aumento da transposição, e que a mobilização dos TEs pode induzir importantes mudanças no cariótipo (HEDGES; DEININGER, 2007; WEIL, 2009).

Diversos estudos sobre atividade de TEs em *Drosophila* são derivados de estudos em cruzamentos intraespecíficos. Dois sistemas diferentes em *Drosophila* são conhecidos por alterarem características reprodutivas: o sistema P-M, o H-E e o sistema I-R, todos responsáveis pelo fenômeno de disgenesia híbrida em *Drosophila melanogaster*. O sistema P-M está relacionado com a presença ou não de um transposon de DNA, o elemento *P*, em linhagens dessa espécie. Quando linhagens maternas não portadoras de elementos *P* (denominadas M) eram cruzadas com linhagens paternas portadoras desses elementos (denominadas P), os descendentes F1 exibiam características disgênicas como redução da fertilidade e viabilidade, atrofia das gônadas, aberrações cromossômicas e alto nível de mutações espontâneas, enquanto que descendentes do cruzamento recíproco não apresentavam essas anomalias (KIDWELL; NOVY, 1979; BINGHAM et al., 1982). Em *D. melanogaster*, a maioria dos pontos de quebra de inversões em híbridos disgênicos ocorrem próximos a sítios de inserção do elemento *P* (LYTTLE; HAYMER, 1992; LADEVEZE et al., 1998). O elemento *hobo*, apresentam muitas semelhanças ao elemento *P*. As sequências *hobo*, além de promoverem altas taxas de instabilidade cromossômica, também causam disgenesia

híbrida ligada a mobilização deste elemento na linhagem germinal de indivíduos resultantes do cruzamento de machos que possuem os elementos ativos com fêmeas que não possuem o elemento ativo (BLACKMAN, et al., 1987).

O sistema I-R, tratado em detalhes no Capítulo 1, está relacionado com a presença ou não do retrotransposon NLTR I em linhagens dessa mesma espécie. Cruzamentos entre machos I (Inducer), portadores de elementos I ativos e fêmeas R (Reactive), em cujos genomas esses elementos estão ausentes ou degradados, geram fêmeas F1 com desenvolvimento anormal dos ovários e ausência de oviposição (ORSI et al., 2010).

Disgenesia híbrida também ocorre em *D. virilis* decorrente de morte de células germinativas na embriogênese relacionada à iniciação de transcrição do retrotransposon *Penelope* (SOKOLOVA et al., 2010).

Portanto, a evolução do híbrido é mediada por grandes alterações genômicas seguidas por rápida estabilização e fixação de genótipos altamente adaptados. São relatados casos de episódios de transposição seguindo a hibridação e de seu impacto na reorganização do genoma híbrido que pode levar à especiação (REBOLLO et al., 2010). Poucos são os estudos descrevendo a atividade dos TEs em cruzamentos interespecíficos, já que a dificuldade para obtenção de descendentes é um fator limitante neste tipo de estudo. No entanto, alterações na atividade de TEs também estão presentes e ocasionam importantes mudanças no genoma híbrido. Alta frequência de ocupação de vários sites cromossômicos pelo LTR *Oswaldo* foi revelada em híbridos entre *Drosophila buzzatti* e *D. koepferae*, duas espécies irmãs, do grupo *repleta*, do subgênero *Drosophila*, que coexistem em zonas áridas da Bolívia e nordeste da Argentina (NAVEIRA; FONTDEVILA, 1985; LABRADOR et al., 1999; GUERREIRO; FONTDEVILA, 2007). Maiores detalhes sobre a atividade desse elemento como resultado da hibridação podem ser encontrados no Capítulo 3.

Estudos do padrão global de atividade de TEs em genomas híbridos ainda são escassos. O único trabalho sobre atividade global de TEs em híbridos interespecíficos é o de Kelleher et al. (2012), no qual investigaram o padrão funcional de TEs e a regulação, pela produção de pequenos RNA (small RNA - responsáveis pelo silenciamento de TEs), em cruzamentos entre *D. melanogaster* e *D. simulans*, espécies que se divergiram entre 3-5 milhões de anos (TAMURA et al., 2004). Mostrou-se nesse estudo de que híbridos interespecíficos são caracterizados por uma desrepressão de famílias de TEs. Contudo, não há relatos na literatura sobre alterações globais da ativação de TEs no genoma de híbridos de cruzamentos interespecíficos de espécies de *Drosophila* evolutivamente próximas.

Drosophila mojavensis e D. arizonae como modelo para estudo da ativação de elementos de transposição em híbridos

As espécies *D. mojavensis* e *D. arizonae* apresentam-se como um par de espécies bastante apropriado para se estudar a dinâmica de TEs em híbridos por diversas características biológicas e práticas, quanto ao desenvolvimento de experimentos em laboratório. Essas são espécies jovens, incipientes, que divergiram de um ancestral comum em torno de 0,6 a 0,9 milhões de anos (REED et al., 2007); diferente das espécies *D. melanogaster* e *D. simulans*, também utilizadas para estudos de processos de hibridação, mas com tempo de divergência quase quatro vezes maior. A distância genética (D) entre esses dois pares de espécies reflete esse tempo; entre *D. arizonae* e *D. mojavensis* D é igual a 0,212 enquanto entre *D. melanogaster* e *D. simulans* é igual a 0,550 (REED; MARKOW, 2004). Como *D. mojavensis* e *D. arizonae* estão em processo recente de especiação, propiciam a investigação dos mecanismos genéticos que possam levar ao isolamento reprodutivo, e no caso desta Tese, o foco é a ação dos TEs no genoma híbrido. Além do tempo recente de divergência, uma segunda característica interessante desse par de espécies é que existe uma sobreposição de

habitats, com populações simpátricas e alopátricas, que apresentam isolamento pré- e pós-zigótico incompleto, o que facilita a obtenção de híbridos em laboratório. Todas as populações de *D. arizonae* exibem forte isolamento pré-zigótico em relação às de *D. mojavenensis*, mas populações de *D. mojavenensis* de áreas simpátricas são significativamente mais isoladas reprodutivamente de *D. arizonae* do que as alopátricas (WASSERMAN; KOEPPER, 1977). Além disso, o isolamento sexual varia entre populações alopátricas, por exemplo, as fêmeas de *D. mojavenensis* do deserto de Anza Borrego mostram o menor isolamento com machos de *D. arizonae*, enquanto as de *D. mojavenensis* de San Carlos – Sonora (simpátricas), mostram maior isolamento (MASSIE; MARKOW, 2005).

Quando o cruzamento entre fêmeas de *D. arizonae* com machos de *D. mojavenensis* ocorre, originam duas vezes mais descendentes que nos cruzamentos recíprocos; entretanto, nesta direção, é observado isolamento pós-zigótico incompleto, gerando machos F₁ estéreis. Já nos cruzamentos entre fêmeas *D. mojavenensis* e machos *D. arizonae*, os descendentes machos não apresentam esterilidade (RUIZ et al., 1990), com exceção de uma linhagem de *D. mojavenensis* procedente da ilha de Santa Catalina (Califórnia) que produz machos estéreis em cruzamentos com machos de *D. arizonae* (RUIZ et al., 1990). A Figura 1 ilustra a distribuição geográfica das duas espécies (*D. mojavenensis* e *D. arizonae*).

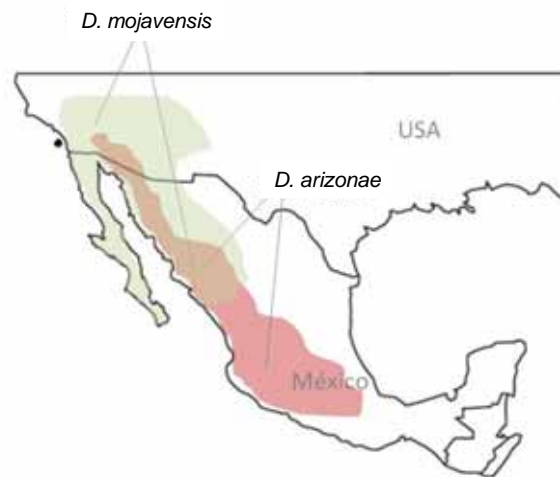


Figura 1. Representação da distribuição geográfica das subespécies de *D. mojavensis* e da espécie *D. arizonae*. O ponto preto na costa do continente dos EUA indica a ilha Santa Catalina de onde provém a linhagem de *D. mojavensis* cujo genoma foi sequenciado.

As características biológicas relacionadas ao isolamento reprodutivo entre *D. mojavensis* e *D. arizonae* podem estar associadas aos TEs, uma vez que essas sequências móveis podem sofrer explosões de transposição em híbridos, causando reorganização genômica e efeitos deletérios, como por exemplo, a esterilidade resultante da disgenesia, já descritos acima. Existem ainda duas vantagens práticas que nos levou a escolher este par de espécies para estudar a dinâmica dos TEs em híbridos: i) a possibilidade de se obter híbridos em cruzamentos em laboratório e, ii) o genoma de *D. mojavensis* está sequenciado e disponível, tornando possível a investigação prévia da presença de sequências específicas de TEs, ao menos em uma das espécies, para serem realizados estudos *in vivo*.

OBJETIVOS

OBJETIVOS

O objetivo geral desse trabalho foi investigar a expressão de elementos de transposição específicos e de todo mobiloma de parentais e híbridos de cruzamentos interespecíficos de *Drosophila mojavensis* e *D. arizonae*, além de mensurar o tamanho do genoma de linhagens dessas espécies e de híbridos.

Os objetivos específicos foram:

1. Avaliar a presença e variabilidade do número de cópias do retrotransposon I em *D. mojavensis* e *D. arizonae*, suas relações filogenéticas, seu perfil de expressão e localização em tecidos germinais nas espécies parentais e em seus híbridos;

2. Avaliar o perfil de expressão global do mobiloma (todos os elementos de transposição de um transcriptoma) nos parentais e híbridos recíprocos de *D. mojavensis* e *D. arizonae*;

3. Quantificar os níveis de expressão de TEs específicos nos parentais e híbridos recíprocos de *D. mojavensis* e *D. arizonae*;

4. Estimar o tamanho do genoma (C-value) das espécies parentais *D. mojavensis* e *D. arizonae* e de seus híbridos recíprocos.

CAPÍTULO 1



Specific activation of an I-like element in *Drosophila* interspecific hybrids

Journal:	<i>Genome Biology and Evolution</i>
Manuscript ID:	GBE-140213
Manuscript Type:	Research Article
Date Submitted by the Author:	09-Feb-2014
Complete List of Authors:	Carnelossi, Elias Lerat, Emmanuelle; University Lyon 1 - CNRS, Lab. Biométrie et Biologie Evolutive - UMR CNRS 5558 Henri, Helene; University Lyon 1 - CNRS, Lab. Biométrie et Biologie Evolutive - UMR CNRS 5558 Martinez, Sonia; University Lyon 1 - CNRS, Lab. Biométrie et Biologie Evolutive - UMR CNRS 5558 Carareto, Claudia; Sao Paulo State University, Biology Vieira, Cristina; University Lyon 1 - CNRS, Lab. Biométrie et Biologie Evolutive - UMR CNRS 5558
Keywords:	non-LTR retrotransposon, phylogeny, hybrids, repleta group

SCHOLARONE™
Manuscripts

1
2
3 1 **Specific activation of an *I*-like element in *Drosophila* interspecific hybrids**
4

5 2
6
7
8 3 **Carnelossi, E.A.G^{1,2}; Lerat, E²; Henri, H²; Martinez, S²; Carareto, C.M.A¹; Vieira, C^{2,3*}**
9

10 4 ¹UNESP – Universidade Estadual Paulista, Laboratório de Evolução Molecular,
11 5 Departamento de Biologia, 15054-000 São José do Rio Preto, São Paulo, Brazil
12

13 6 ²Université de Lyon, F-69000, Lyon; Université Lyon 1; CNRS, UMR5558, Laboratoire de
14 7 Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France
15

16 8 ³ Institut Universitaire de France, Paris, France
17
18
19

20 9
21
22
23
24 10 **Abstract:** The non-LTR retrotransposon *I*, which belongs to the *I* superfamily of non-LTR
25 11 retrotransposons, is well known in *Drosophila* because it transposes at a high frequency in the
26 12 female germline cells in I-R hybrid dysgenic crosses of *Drosophila melanogaster*. Here, we
27 13 report the occurrence and the up-regulation of an *I*-like element in the hybrids of two sister
28 14 species belonging to the *repleta* group of the genus *Drosophila*, *Drosophila mojavensis* and
29 15 *D. arizonae*. These two species display variable degrees of pre- and post-zygotic isolation,
30 16 depending on the geographic origin of the strains. We took advantage of these features to
31 17 explore the transposable element (TE) dynamics in interspecific crosses. We fully
32 18 characterized the copies of this TE family in the *D. mojavensis* genome and identified at least
33 19 one complete copy. We showed that this element is transcriptionally active in the ovaries and
34 20 testes of both species and in their hybrids. Moreover, we showed that this element is up-
35 21 regulated in hybrid males, which could be associated with the male-sterile phenotype.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52 22 **Keywords:** hybrids, non-LTR retrotransposon, *repleta* group, *phylogeny*
53
54
55
56
57
58 24 * Corresponding author: cristina.vieira@univ-lyon1.fr
59
60 25

1
2
3 26
4
56 **27 Introduction**

7
8 28 Virtually all species harbor transposable elements (TEs), sequences that are able to move
9
10 29 within genomes and that differ in structure, activity, and copy number between populations
11
12 30 and species. TEs are classified in two broad classes, Class I and Class II, based on their mode
13
14 31 of transposition. Class I is composed of two types of elements that transpose through RNA
15
16 32 intermediates: elements with Long Terminal Repeats (LTRs) and elements without these
17
18 33 direct repeats, the non-LTR elements, which can further be divided into two types, the LINES
19
20 34 (Long Interspersed Nuclear Elements) and the SINES (Short Interspersed Nuclear Elements).
21
22 35 Class II (DNA transposons, rolling circle transposons, and miniature inverted repeat elements
23
24 36 -MITES) is composed of TEs that transpose directly via a DNA molecule (Wicker et al. 2007;
25
26 37 Kapitonov and Jurka 2008).
27
28
29
30
31
32
33

34 39 The LINES are among the oldest genetic elements in eukaryotes and are the most abundant in
35
36 40 the human genome (Fawcett et al. 1986; Simonelig et al. 1988; Bucheton et al. 1992; Han
37
38 41 2010). They consist of a 5' untranslated region (UTR), which has promoter activity, two open
39
40 42 reading frames (ORFs) separated by a spacer, and a 3' UTR with a poly-A tail. Our study
41
42 43 focused on one *I* like element, a LINE element first described in *D. melanogaster* at the
43
44 44 beginning of the 1970s (Picard 1976). The *I* element (originally called the *I* factor) is
45
46 45 responsible for the hybrid dysgenesis syndrome in *D. melanogaster* females. This species
47
48 46 presents two types of strains based on the presence and activity of this element: the *I* strains,
49
50 47 harboring complete and functional *I* elements, and the R strains, with non-functional *I*
51
52 48 elements. Crosses between *I* males (Inducer) and R females (Reactive) result in high rates of
53
54 49 transposition in the germline of female offspring, known as *SF* females. These females are
55
56 50 sterile because most of the laid eggs fail to hatch due to the very high frequency of *I* element
57
58
59
60

1
2
3
4 51 transposition, which causes genetic abnormalities such as failure in meiotic divisions and
5
6 52 chromosomal rearrangements (Bucheton et al. 1976; Picard et al. 1978; Chaboissier et al.
7
8 53 1990; Chaboissier et al. 1995). The dysgenesis caused by the *I* element in *D. melanogaster* is
9
10 54 restricted to the female germline and has not been reported in males. Furthermore, the
11
12 55 transposition rate correlates with the level of sterility in the SF females (Chaboissier et al.
13
14 56 1990; Seleme et al. 1999; Seleme et al. 2006).

15
16
17 57

18
19
20 58 The number of *I* copies in *D. melanogaster* is approximately 30, but only five functional
21
22 59 sequences are known to be dispersed on the chromosomal arms. The complete and functional
23
24 60 copies are 5.4 kb long, possess two long ORFs with cysteine-rich motifs, apurinic-apyrimidic
25
26 61 endonucleases, reverse transcriptase and RNase H domains, and the 3' end has several TAA
27
28 62 repeats. These copies are expressed in the nurse cells of dysgenic ovaries, and the transcripts
29
30 63 are transported into the oocytes, where retrotransposition occurs (Pelisson and Picard 1979;
31
32 64 Fawcett et al. 1986; Vaury et al. 1990; Seleme et al. 1999; Seleme et al. 2005).

33
34
35
36 65

37
38 66 Because TEs are able to induce hybrid dysgenesis (Bucheton et al. 1984; Bucheton 1990;
39
40 67 Vieira et al. 1998; Khurana et al. 2011), mutations (Fedoroff et al. 1983; McClintock 1984),
41
42 68 and genomic rearrangements (Zhang et al. 2009) and can affect the expression of nearby
43
44 69 genes (Hollister and Gaut 2009), among other deleterious effects (Orsi et al. 2010), host cells
45
46 70 use different mechanisms to control their expression and transposition. In *Drosophila*, piwi-
47
48 71 interacting RNAs (piRNAs) are important posttranscriptional regulators that are responsible
49
50 72 for TE mRNA degradation (Khurana and Theurkauf 2010; Senti and Brennecke 2010;
51
52 73 Kelleher et al. 2012; Akkouche et al. 2013; Dufourt et al. 2013). Specifically, the piRNAs are
53
54 74 implicated in maternal *I* element control in *Drosophila* (Brennecke et al. 2008; Chambeyron
55
56 75 et al. 2008).

1
2
3 76
4
5
6 77 The implication of TEs in hybrid dysgenesis is one of the examples in which transposition
7
8 78 may drive population isolation and is often proposed as one of the first steps of speciation
9
10 79 (Kidwell and Novy 1979; Fontdevila 2005). However, the TE dynamics of expression and
11
12 80 transposition have not been well described during this process. Hence, several questions
13
14 81 remain unanswered regarding this subject. If TEs act in the first steps of speciation, what
15
16 82 happens when crossing very closely related species that still have incomplete reproductive
17
18 83 isolation? In particular, are *I*-like elements associated with the incompatibility observed, and
19
20 84 are they related to high rates of hybrid transposition as reported in other species (Ungerer et
21
22 85 al. 2009; Cavallini et al. 2010; Moschetti et al. 2010; Kelleher et al. 2012; Vela et al. 2014)?
23
24
25
26
27 86
28
29 87 *Drosophila mojavensis* and its sister species analyzed in this study – *D. arizonae* – are a
30
31 88 perfect pair of species with which to address these questions. They are found in the desert of
32
33 89 the southwest USA and Mexico and share sympatric areas in southern Arizona and the State
34
35 90 of Sonora (Mexico). These two species present three interesting features that facilitate the
36
37 91 investigation of the above questions: (1) their hybrids can be produced in the laboratory; (2)
38
39 92 the genome of *D. mojavensis* has been sequenced, which allow us to analyze the sequences of
40
41 93 its TEs; and (3) these species exhibit variable degrees of pre- and post-zygotic isolation
42
43 94 (Wasserman and Koepfer 1977; Reed and Markow 2004; Massie and Markow 2005).
44
45 95 Moreover, the male offspring of *D. arizonae* females and *D. mojavensis* males are sterile, but
46
47 96 in the reciprocal cross, the hybrids have motile sperm depending on the origin of the *D.*
48
49 97 *mojavensis* population that is used in the crosses (Reed and Markow 2004). Additionally, the
50
51 98 strong pre-zygotic isolation is higher between flies from sympatric areas than between the
52
53 99 allopatric flies, but the isolation level depends in part on the direction of the crosses and the
54
55 100 geographic origin of the populations (Wasserman and Koepfer 1977; Ruiz 1990; Reed and
56
57
58
59
60

1
2
3 101 Markow 2004; Massie and Markow 2005). Because of the incomplete pre-zygotic isolation in
4
5 102 allopatric strains, we chose to perform crosses of strains derived from *D. mojavensis* from the
6
7
8 103 Anza Borrego Desert population and from *D. arizonae* from Hidalgo-Mexico, for which we
9
10 104 could obtain enough F1 hybrids between *D. mojavensis* females and *D. arizonae* males, and
11
12 105 in the reciprocal cross, where the F1 male hybrids were sterile.

13
14
15 106

16
17 107 We found four potential full-length copies of *I*-like elements and reconstructed their
18
19 108 phylogenetic relationships with other *I* family members in *Drosophila* species, thus
20
21 109 establishing the vertical and ancestral inheritance of the *I* element of *D. mojavensis* from a
22
23 110 common ancestor shared with the *melanogaster* group. We also showed that the *D.*
24
25 111 *mojavensis* *I*-like sequences are transcriptionally and transpositionally regulated in the hybrid
26
27 112 female germline. However, a significant increase in transcription was detected in testes from
28
29 113 hybrids coming from *D. arizonae* mothers. This is the first report that shows the activity of *I*-
30
31 114 like elements in the male germline, and it suggests a link with the male-sterile phenotype
32
33 115 observed in the male hybrids of the studied species.

34
35
36
37
38 116

39 40 41 117 **Materials and Methods**

42 43 118 *Genomic analyses*

44
45 119 To investigate the occurrence of *I*-like elements in the *D. mojavensis* sequenced genome, we
46
47 120 used the canonical *I* sequence from *D. melanogaster* (*I_{DM}*), which is available in the
48
49 121 Repbase library Rel.16.06, v.4 (Fawcett et al. 1986). This sequence was blasted against the
50
51 122 scaffolds of the *D. mojavensis* genome on the Ensembl Metazoa platform using BLASTN *D.*
52
53 123 *mojavensis*, version-70,13/dmoj_caf1 (<http://metazoa.ensembl.org/index.html>).

54
55 124 The hits were manually curated, and only the largest and putative complete copies (more than
56
57 125 2,000 bp) were selected. We then selected the *D. mojavensis* copies that were over 80%

1
2
3 126 identical between them. The sequences with the above characteristics were selected and
4
5 127 analyzed using the ORF-FINDER web tool (<http://www.ncbi.nlm.nih.gov/projects/gorf/>) to
6
7
8 128 identify putative coding regions. The ORFs identified were compared and confirmed by
9
10 129 BLAST search for protein sequences at the NCBI website, and the conserved proteic domains
11
12
13 130 were checked on the Conserved Domain Database
14
15 131 (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>). The sequence with the most complete
16
17 132 structure was selected and hereafter referred to as “Reference Copy” or “RC” (Figure 1).

18
19
20 133
21
22 134 The RC was used to search for *I*-like element copies in the *D. mojavensis* genome with
23
24 135 identity (ID) higher than 80% and length greater than 200 bp. This criterion was chosen based
25
26 136 on the 80-80-80 rule for TE family determination, as proposed by Wicker et al. (2007) for the
27
28 137 identification of TEs belonging to the same family. The position of each copy (start and end)
29
30 138 on the genome scaffolds, the length of each fragment aligned by BLAST, and the number of
31
32 139 the unidentified nucleotides (N) were recorded (Table 1). We performed multiple alignments
33
34 140 of the *gag* and *reverse transcriptase* domains (nt) identified in the copies, and we determined
35
36 141 the identity of each domain, based on those of the RC, using Bioedit version 7.0.4.1 (HALL,
37
38 142 1999). A drawing of the copies is shown in Figure 2 and represents the integrity of the
39
40 143 putatively full-length copies compared to RC. To confirm that we recovered unique copies in
41
42 144 the genome, the 5' and 3' end flanking sequences of each copy were retrieved, and their
43
44 145 identities were determined compared to the RC flanking sequences. All copies corresponded
45
46 146 to different insertions.

47
48 147 To check the functional organization of all putatively full-length copies of *D. mojavensis*, the
49
50 148 5' end sequences were compared to the sequence of the *I* promoter of *D. melanogaster*
51
52 149 (Minchiotti et al. 1997) using BLAST2seq.
53
54
55
56
57
58
59
60

1
2
3 151 The chromosomal location of the most complete *I* copies was inferred from the scaffold
4
5 152 coordinates of each copy, and the correspondence of scaffolds with *D. mojavensis* polytene
6
7
8 153 chromosome maps was performed using information found in Flybase.org
9
10 154 (<http://flybase.org/maps/chromosomes/maps.html>) and in Schaeffer et al. (2008). For an intra-
11
12 155 chromosomal distribution of the copies, each *D. mojavensis* chromosome (X, 2, 3, 4, 5, and 6)
13
14
15 156 was divided into three regions (telomeric, central, and centromeric, corresponding to 10%,
16
17 157 80%, and 10% of the chromosome region, respectively).
18
19
20 158

21
22 159 The nomenclature used to identify each copy followed Wicker *et al.* (2007); for example,
23
24 160 RII_DMJA1 consists of RII (for the superfamily), DMJ (for *Drosophila mojavensis*), and A to
25
26
27 161 Z (for the clades formed by phylogenetic analyses, where the exception is X, which was used
28
29 162 for the undetermined groups), followed by the number of each copy within each clade.
30
31
32 163

33 34 164 *Evolutionary Analyses of Reverse Transcriptase Sequences*

35
36 165 We reconstructed the evolutionary relationships between the *D. mojavensis* *I*-like sequences
37
38 166 with all the *I* superfamily members of *Drosophila* species deposited in the RepBase library
39
40
41 167 (<http://www.girinst.org/replib/>). Each retrieved element was analyzed using ORF FINDER
42
43 168 (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>), and only the amino acid sequences of the
44
45
46 169 second ORF were used for further analysis. We identified the reverse transcriptase (RT)
47
48 170 domain in each protein sequence using the structural domain bank from the Pfam platform
49
50 171 (<http://pfam.sanger.ac.uk/>). The *Drosophila* Repbase RT domains were aligned with the *D.*
51
52 172 *mojavensis* RT domains using MEGA5 (Tamura et al. 2011). We performed a phylogenetic
53
54
55 173 reconstruction using the Maximum Likelihood method, the JTT model and 500 bootstrap
56
57 174 replicates, as implemented in MEGA5. Two sequences of other LINE superfamilies were
58
59
60 175 used as outgroups: the *D. melanogaster* *Fw* non-LTR retrotransposon (Genbank accession

1
2
3 176 number: M17214.1) and the *D. yakuba Helena* non-LTR retrotransposon (Genbank accession
4
5 177 number: AF012049.1).
6
7

8 178

9
10 179 *RNA extraction and RT-qPCR*
11

12 180 The fly samples were obtained from the US San Diego Drosophila Stock Center. We used
13
14 181 strains of *D. mojavensis* (Anza Borrego Desert, California – USA, stock number: 15081-
15
16 182 1352.01) and *D. arizonae* (Metztitlan, Hidalgo – Mexico, stock number: 15081-1271.17).
17
18 183 These populations were maintained at 25°C in mass cultures. One- to two-day-old virgins
19
20 184 (females and males) were isolated from the parental lines and split in vials. Subsequently, the
21
22 185 crosses were performed (*D. mojavensis* females x *D. arizonae* males and the reciprocal cross)
23
24 186 using 10 couples per cross. F1 hybrids were isolated, and five-day-old females and males
25
26 187 were used for DNA and RNA extraction.
27
28
29
30
31
32
33

34 189 The flies were dissected, and their tissues were placed in 1x PBS (Phosphate buffer saline).
35
36 190 The RNA was extracted from pools of 20 pairs of ovaries and from 30 testes from parental
37
38 191 and hybrids. Two independent sets of crosses were performed. For each, three replicates were
39
40 192 performed to extract total RNA using the RNeasy kit (Qiagen) and treated with DNase (the
41
42 193 DNA-free kit; Ambion). One microgram of the total RNA was then converted into cDNA
43
44 194 using the ThermoScript Invitrogen kit primed with a mix of oligo(dT)₂₀ and random primers.
45
46
47
48
49

50 196 The expression level of the RT of the *I*-like elements was measured by RT-qPCR using the
51
52 197 following primers: Forward 5'-ATC CAC TCT TCA ACG GCA TC-3' and Reverse 5'- TGG
53
54 198 ACG ATA TGG TGC AAA TG- 3'. We verified by sequencing (the TOPO TA cloning kit;
55
56 199 Invitrogen) that no sequence polymorphism was present in the primers that could affect the
57
58
59
60 200 PCR efficiency. Five clones of each strain were sequenced (Applied Biosystems 3730XL),

1
2
3 201 and the sequences obtained were aligned using BLAST alignment at the NCBI website. The
4
5 202 cDNA was diluted 50 times, and the relative mRNA level was quantified using SYBR green
6
7
8 203 quantitative PCR (qPCR) in a LightCycler 480 apparatus (Roche Diagnostics). The transcript
9
10 204 quantity was estimated relative to the rp49 reference gene of *D. mojavensis* using the primers
11
12 205 described by Granzotto et al. (2009). The RT-qPCR experiments were performed with
13
14 206 technical triplicates. Only RT-qPCR experiments with efficiencies greater than 1.9 were
15
16 207 retained.
17
18
19

20 208

21
22 209 *Fluorescent In situ hybridization* (FISH)

23
24 210 *I* transcripts were detected by *in situ* hybridization in 5-day-old ovaries of parents and hybrids
25
26 211 according to Akkouche et al. (2012). Ovaries were hybridized with a 952-bp riboprobe that
27
28 212 corresponded to the reverse transcriptase and RNA-H domains in the RC and included a T3
29
30 213 promoter site in the reverse primer (Forward: 5'-GCA ATA CAA CCG GCG CTA ACA,
31
32 214 Reverse: 5'-TGG CTG TGG ATT TGG CTG TGA ATT AAC CCT CAC TAA AGG GA -
33
34 215 3'). For the *in vitro* transcription, a digoxigenin (DIG)-RNA labeling mix (Roche) was used.
35
36 216 After hybridization and stringency washes, an anti-DIG POD antibody mix (ROCHE) was
37
38 217 used for chemiluminescent detection, with a fluorescence amplification (the Tyramide signal
39
40 218 amplification kit; PerkinElmer). The DNA was labeled by SYTOX Green Nucleic Acid Stain
41
42 219 (Life Technology), and the visualization was performed with a Zeiss LSM510 Meta confocal
43
44 220 microscope. Tissues without anti-DIG treatment were used as the control. Five-day-old testes
45
46 221 hybridization was performed with the same *I* element probe according to Morris et al. (2009).
47
48 222 The testes images were processed in a Zeiss Axioskop 2 mot plus microscope and analyzed
49
50 223 with Axiovision Release 4.8 software.
51
52
53
54
55
56
57
58
59
60

224

225

226 *Transposable element activity (Transposon Display)*

227 Transposon Display (TD) was performed with protocols adapted from several reports
228 (Munroe et al., 1994; Esnault et al. 2008; Akkouche et al. 2012). Total genomic DNA was
229 isolated from single individuals used in crosses between *D. mojavensis* females and *D.*
230 *arizonae* males, and the F1 single hybrids were backcrossed with *D. arizonae* males to obtain
231 the F2 generation (Supplementary Figure S1). We could only perform crosses in this direction
232 because individual reciprocal crosses using *D. arizonae* females did not yield any offspring.
233 From the six individual crosses performed with the parental lines, we were able to follow two
234 independent families. Individuals analyzed by TD are in Supplementary Table 1.

235

236 The genomic DNA (100 ng/ μ l) was digested with HindIII (10 U/ μ l, 3 hours at 37°C). The
237 MSBE adapter was linked with the HindIII enzyme adaptor site in one double-stranded
238 reaction (HindIII-MSEB) using 100 μ M of Hindlink, 100 μ M of MSEB, SSC 20x, and 1 M
239 Tris in total volume of 333 μ l. The product was denatured for 5 minutes in 92°C and cooled at
240 room temperature. The linker HindIII-MSEB was ligated to digested DNA with a T4 DNA
241 ligase (10 units), enzyme buffer (5x), and digested DNA in a 50- μ l reaction for 3 hours at
242 23°C. DNA was amplified with a primer adapter (LNP 5' - GAA TTC GTC AAC ATA GCA
243 TTT CT - 3') and a forward primer inside of the 3' end of the *I* element (5'- TAA CTG TCC
244 TGC AAC TTC CCA CCT - 3'). The amplification conditions were as follows: denaturation
245 at 94°C, 2 minutes; followed by 94°C, 30 seconds; 59°C, 1 minute; 72°C, 1 minute; 35
246 cycles; then, 72°C, 10 minutes. Subsequently, the product amplification was run on a gel, and
247 analyses were performed with QIAxcel ScreenGel 1.1.0 software.

248 The TD method was tested with *D. mojavensis* genomic DNA, extracted from a single flies ,
249 that was digested by HindIII and amplified by PCR with HindIII-MSEB and *I* primers. Then,
250 two fragments (842 bp and 450 bp), identified *in silico* and observed in electrophoresis, were

1
2
3 251 cloned (with five clones for each fragment), amplified by PCR with M13 universal primers,
4
5 252 and sequenced. The alignment of these sequences against the *D. mojavensis* RC confirmed the
6
7
8 253 veracity of the fragments obtained by enzyme digestion and amplification for the *I* element.
9

10 254

11 12 255 **Results**

13
14
15 256 Our study revealed the presence of an *I*-like element in *D. mojavensis* and *D. arizonae* that is
16
17 257 closely related to the *I* element found in the genomes of the *D. melanogaster* subgroup. *In*
18
19 258 *silico* and *in vivo* analyses showed the occurrence of complete and transcriptionally active
20
21 259 copies in both species and in their hybrids.
22
23

24 260

25 26 27 261 *The I element of D. mojavensis*

28
29 262 We detected 10 *I*-like copies in the *D. mojavensis* genome sequence (Table 1). One potentially
30
31 263 complete copy, called RC (RII_DMJA1), is 5,382 bp long and contains two ORFs (ORF 1
32
33 264 and ORF 2) with no internal stop codons and with all the expected protein domains for this
34
35 265 type of non-LTR retrotransposon (Malik et al., 1999; Figure 1). It has 62.42% nucleic identity
36
37 266 with the *D. melanogaster I* element and is only 7 bp longer. We identified three other copies
38
39 267 that are most likely complete: RII_DMJA2, RII_DMJA3, and RII_DMJX1 (Table 1, Figure
40
41 268 2), which show high identity and conserved protein domains compared to the RC, but their
42
43 269 sequences were interrupted with unidentified nucleotides (N). Similarly, the RII_DMJX5
44
45 270 copy, with 1,629 bp aligned against RC, may also be complete because it has a long N stretch
46
47 271 at its 5' end. The other copies were small and fragmented by insertions and deletions. The
48
49 272 identification of the relative chromosome position of these copies, as established by analyzing
50
51 273 the polytene chromosome maps found in Flybase (Schaeffer et al. 2008), revealed that three
52
53 274 of the four putative full-length copies, RII_DMJA1, RII_DMJA2, and RII_DMJA3, were
54
55
56
57
58
59
60

1
2
3 275 inserted in the central portion of different chromosome arms corresponding to chromosomes
4
5 276 2, 3, and 5, respectively.

6
7
8 277 Moreover, we looked for the *I* element promoter in the *D. mojavensis* copies by aligning the
9
10 278 flanking sequences of each copy with the promoter sequence of the *I* element of *D.*
11
12 279 *melanogaster* (Minchiotti et al. 1997; Han 2010). We identified a putative promoter in the RC
13
14 280 that has the same length (41 bp) and is well conserved (76% identity) compared to that of the
15
16 281 *D. melanogaster I* element. This promoter region is identical in the four putative full-length
17
18 282 copies found in the *D. mojavensis* genome (Figure 3). The four longest *I* sequences have a
19
20 283 poly-A tail in the 3'UTR, with an average of four TAAA repeats.
21
22
23
24
25

26 284

27 285 *Phylogenetic relationship between the D. mojavensis I-like element and the I elements of*
28
29 286 *other Drosophila species*

30
31 287 The phylogenetic relationship of the different *I*-like elements was reconstructed using the RT
32
33 288 protein sequences (see Material and Methods for details) (Figure 4). It allowed us to
34
35 289 distinguish four major clades of sequences (A to D). Different *I*-like sequences are present in
36
37 290 species of the *melanogaster* group of the subgenus *Sophophora* (see Supplementary Figure 2
38
39 291 for the species phylogeny), which are included in all four clades (Figure 4). Clade A can be
40
41 292 divided into two well-supported groups of sequences (clade A1, bootstrap: 84%, and clade
42
43 293 A2, bootstrap: 50%). Inside clade A1, the three sequences of *D. mojavensis* and the sequence
44
45 294 of *D. virilis*, where both species belong to the *virilis-repleta* radiation of the *Drosophila*
46
47 295 subgenus, cluster in a highly supported monophyletic group (A1b: 99%). This small clade is
48
49 296 closely related (bootstrap: 84%) to the group of sequences from the *melanogaster* group
50
51 297 (clade A1a), which includes the canonical *I* element from *D. melanogaster (I_DM)*. These
52
53 298 two groups of sequences (A1a and A1b) form a sister group of clade A2. This clade includes
54
55 299 sequences found in the same species that form clade A1a with the exception of *D. biarmipes*
56
57
58
59
60

1
2
3 300 and of *D. persimilis*, a species of the *obscura* group. The other three clades (clades B, C, and
4
5 301 D) include more divergent *I*-like sequences from species belonging only to the *Sophophora*
6
7 302 subgenus: clade B includes only species from the *melanogaster* group, whereas clades C and
8
9 303 D also include sequences from *D. willistoni* (the *willistoni* species group). The phylogeny
10
11 304 reveals that different *I*-like families are harbored by the same species, as is the case for *D.*
12
13 305 *kikkawai*, whose sequences branch into three different clades (A1, A2, and C).
14
15
16
17
18
19

306

307 *Activity of the complete I copies*

20 308 To determine if the full-length *I* copies found in *D. mojavensis* were transcriptionally active,
21
22 309 we performed expression analyses by RT-qPCR. The results showed transcription of the *I*
23
24 310 elements in *D. mojavensis* and in *D. arizonae*, both in ovaries and testes (Figure 5). The levels
25
26 311 of transcription between the parental lines and the reciprocal male and female hybrids were
27
28 312 not significantly different between ovaries from parents and hybrids (two-way ANOVA).
29
30 313 However, expression levels were significantly higher in hybrid testes produced by crosses
31
32 314 between *D. arizonae* females and *D. mojavensis* males than in the reciprocal cross ($F(3, 9) =$
33
34 315 84.78 , $P < 0.0001$). We obtained the same results with crosses and expression analyses
35
36 316 performed again six months later (Supplementary Figure 3).
37
38
39
40
41
42
43
44

317

45 318 We observed *I*-element mRNA in ovarian cells by immunofluorescence and detected the
46
47 319 presence of transcripts in both parents and hybrids. The *I* transcripts had the same localization
48
49 320 in both species, with an accumulation in the nurse cells and no labeling of the oocytes. We did
50
51 321 not detect any signal in the follicle cells (Figure 6). In testes, we observed labeling in parental
52
53 322 lines and both types of hybrids (Figure 7). Transcripts were specifically located in primary
54
55 323 and meiotic spermatocyte cells (Fuller 1998).
56
57
58
59
60

324

1
2
3 325 To detect transposition, we performed transposon display analyses. Unfortunately, these
4
5 326 experiments could not be performed for both reciprocal crosses because, when crossing
6
7 327 individual *D. arizonae* females with individual *D. mojavensis* males, we obtained no
8
9 328 offspring. This cross is only possible in one direction (Supplementary Table S1 and Figure
10
11 329 S1). In agreement with the RT-qPCR results, no transposition was detected in the F1 tissues,
12
13 330 which indicates that the *I* element does not transpose in the parental gonads. Additionally, no
14
15 331 transposition was detected in the F2 generation, indicating that neither the *I* sequences in F1
16
17 332 females nor in *D. arizonae* males induce transposition (data not shown).
18
19
20
21
22 333
23

24 334 **Discussion**

25 335 *Identification and phylogenetic analysis of I-like elements*

26
27 336 In this report, we identified and characterized *I*-like sequences in the sequenced genome of *D.*
28
29 337 *mojavensis*. In addition to several defective and divergent sequences, we identified four
30
31 338 putatively complete copies in this species, but only one showed all typical structural domains
32
33 339 of the *I* superfamily (Malik et al. 1999). This finding suggests that this copy may also have
34
35 340 transposition ability. In *D. melanogaster*, two intact ORFs are a requisite for protein encoding
36
37 341 and complete retrotransposition of the *I* factor (Seleme et al. 2005). The presence of
38
39 342 unidentified nucleotides in some of the detected copies in *D. mojavensis* suggests the
40
41 343 possibility that other sequences are potentially full-length copies. One argument favoring this
42
43 344 hypothesis is the conservation of the promoter region that we identified, which is typically not
44
45 345 conserved between species (Minchiotti et al. 1997).
46
47
48
49
50
51
52

53 346 Non-LTR retrotransposons are common elements in *Drosophila* genomes and are often
54
55 347 described as vertically inherited in this group of species (Malik et al. 1999, Granzotto et al.,
56
57 348 2011). *I*-like element sequences have been found in many species, and their distribution
58
59 349 correlates with the phylogenetic relationships between species (Bucheton et al., 1992), which
60

1
2
3 350 indicates that they are old components of *Drosophila* genomes. However, *I* elements with a
4
5 351 structure strikingly similar to that of *D. melanogaster* (I_DM) occur only in the species of the
6
7 352 *melanogaster* subgroup (*D. simulans*, *D. mauritiana*, and *D. sechellia*). The most widely
8
9 353 supported hypothesis is that ancestral *I* elements were lost from the genome of *D.*
10
11 354 *melanogaster* but were then reintroduced recently because older strains of *D. melanogaster*
12
13 355 contain only defective and immobile *I* sequences, and all the strains collected after 1930 have
14
15 356 both types of sequences (defective and active mobile *I* elements). This observation is in
16
17 357 agreement with the hypothesis that the complete and active *I* factor progressively invaded *D.*
18
19 358 *melanogaster* after this period (Kidwell, 1983a). The process by which *I* elements have
20
21 359 reinvaded the *D. melanogaster* genome is unknown. Horizontal transfers from another species
22
23 360 (Simonel et al., 1988) is one of the possibilities, but more complex hypotheses, such as the
24
25 361 reactivation of sequences sequestered in the heterochromatin, should also be considered.
26
27
28
29
30
31

32 362 The phylogeny placement of the *I*-like elements from *D. mojavensis* is generally in
33
34 363 agreement with the species phylogeny. The *I*-like sequences of *D. mojavensis* branch together
35
36 364 with a sequence of *D. virilis*, as was expected based on the species phylogeny because these
37
38 365 species belong to the *virilis-repleta* radiation of the subgenus *Drosophila*. However, when we
39
40 366 analyze the global phylogeny of the *I*-like element, we cannot exclude that horizontal transfer
41
42 367 events may have occurred between other drosophila species. For example, this could have
43
44 368 been the case between *D. ficusphila* and *D. bipectinata* (see clade A1a in Figure 4).
45
46
47
48
49

50 369 The *I*-like element superfamily is a very old component of the genome of *Drosophila*,
51
52 370 and its diversification most likely occurred very early in the evolution of *Drosophila*, most
53
54 371 likely before the separation of the Sophophora and *Drosophila* subgenera. More
55
56 372 diversification occurred in the Sophophora subgenus, which was accompanied by sequence
57
58 373 loss and horizontal transfer events. In the *Drosophila* subgenus, we can only assume that the
59
60

1
2
3 374 loss of I families lead to only one family, which is described in this manuscript and which
4
5 375 corresponds to the I factor from *D. melanogaster*.
6
7

8
9 376 *Activity of the I element*

10
11 377 We measured the transcriptional activity of the *I* elements by RT-qPCR and FISH
12
13 378 experiments in the parental lines and their hybrids. The level of *I* element mRNA was low in
14
15 379 ovaries and testes of the parental lines. In the parental ovaries, the *I* element transcripts
16
17 380 accumulate in the nurse cells. This type of labeling is also observed for the *I* element in
18
19 381 *Drosophila* ovaries (Seleme et al. 2005; Chambeyron et al. 2008). However, transcription of
20
21 382 the *I* element was never reported in testes in this species. Our RT-qPCR analysis shows that
22
23 383 the *I* element is transcribed in testes of both *D. mojavensis* and *D. arizonae*; this was
24
25 384 confirmed by FISH, which showed labeling in primary and meiotic spermatocytes. In the
26
27 385 reciprocal hybrids between the two species, the levels of expression were identical for the
28
29 386 female germline tissues, with no difference between parents and F1 hybrids. FISH
30
31 387 experiments also confirmed this result because no change in the localization of the transcripts
32
33 388 and no labeling in the oocytes were observed. For the male germline, the results were quite
34
35 389 different depending on the direction of the cross. F1 males from *D. arizonae* male parents
36
37 390 showed no difference in the expression of the *I* element in testes. However, when *D.*
38
39 391 *mojavensis* males were used as parents, a significant increase in the *I* element expression in
40
41 392 testes was observed. The fact that F1 males in this particular cross are sterile (Ruiz 1990)
42
43 393 suggest a role for TEs in the male-sterile phenotype. Furthermore, it is important to determine
44
45 394 if the high expression in F1 hybrids observed in RT-qPCR is derived from complete copies of
46
47 395 *D. mojavensis* males. Hybrid dysgenesis in *D. melanogaster* involves transposon-naive
48
49 396 females and active transposon-containing males in several different systems: the *I-R* system,
50
51 397 the P-M system and even the H-E system (Bucheton et al. 1976; Picard et al. 1978; Kidwell
52
53
54
55
56
57
58
59
60

1
2
3 398 and Novy 1979; Kidwell 1983a, b; Streck et al. 1986). However, at least in the I-R and P-M
4
5 399 systems, it is the maternal line that is affected.

6
7
8 400

9
10 401 The presence of *I* transcripts only in the nurse cells in ovaries of *D. mojavensis*, *D. arizonae*
11
12 402 and their hybrids suggests that the *I* element in both species is most likely regulated by the
13
14 403 piRNA pathway. In *Drosophila* ovaries, Chambeyron et al. (2008) showed that there is an
15
16 404 association between the accumulation of piRNA and the regulation of the *I* element
17
18 405 transcripts in the nurse cells, where the transcripts are processed, before being transported to
19
20 406 the oocytes, where the retrotransposition occurs. Such findings suggest a posttranscriptional
21
22 407 regulation of the *I* element described here before its transportation to the oocytes, as reported
23
24 408 in *D. melanogaster* (Brennecke et al. 2008; Chambeyron et al. 2008).

25
26
27 409

28
29
30
31 410 In testes, the *I* transcripts in parents and hybrids are located in primary spermatocytes, where
32
33 411 high levels of transcription for genes encoding proteins for spermatogenesis and genes
34
35 412 encoding male germline-specific isoforms are detected (Fuller 1998). The transcripts of many
36
37 413 genes are stable and are present for days after meiosis but then decrease in the post-meiotic
38
39 414 spermatids stage, during which proteins can be detected (Kuhn, et al. 1988; Fuller, 1998). The
40
41 415 localization of the *I*-like transcripts observed in this study (Figure 7) corresponds to the
42
43 416 transcriptional dynamics described for their genes. Other elements have transcripts detected in
44
45 417 primary spermatocytes. Transcripts of *412* retrotransposon were not detectable in ovaries by
46
47 418 *in situ* hybridization in *D. melanogaster* and *D. simulans* populations but were detected in the
48
49 419 male germline in primary spermatocytes (Borie *et al.*, 2002). Furthermore, no association has
50
51 420 been found between expression in soma and in testes, which indicates independent regulation
52
53 421 (Borie et al., 2002). The same localization patterns were observed for the retrotransposons
54
55 422 *1731*, *GATE*, *mdg1* and *copia* (Haoudi et al., 1997; Kogan, et. al., 2003; Morozova, et al.,
56
57
58
59
60

1
2
3 423 2004; Kalmykova, et. al., 2005). The relatively high number of transcribed elements,
4
5 424 including the *D. mojavensis* *I* element, in primary spermatocytes reveals the importance of
6
7 425 these cells in the transcription of these TEs.
8
9

10 426
11
12 427 In this work, we have investigated the hypothesis that *I*-like elements (which are implicated in
13
14 428 the hybrid dysgenic incompatibility of *D. melanogaster*) could be activated in hybrids
15
16 429 between *D. mojavensis* and *D. arizonae*, two recently diverged species from the *Drosophila*
17
18 430 subgenus of the genus *Drosophila*. We showed that *I* elements are specifically activated in
19
20 431 testes from a specific direction of the intercross, and we suggest that this activation is
21
22 432 associated with the observed male sterility. Moreover, we showed that this activation is sex-
23
24 433 specific because the female germline does not seem to be affected by this phenomenon. The
25
26 434 impact of TEs in the male sterility observed in interspecific crosses is becoming clearer and
27
28 435 needs further investigations. Studies using closely related species will be helpful to
29
30 436 understand the first steps of the male sterility process and the role of TEs.
31
32
33
34
35

36 437

38 438 **Acknowledgments**

39
40
41 439 This study was supported by FAPESP (grant n° 2010/10731-4 to CMAC and fellowship
42
43 440 n°2010/10056-5 to EAGC), CAPES (fellowship n° 2427-11-7 to EAGC) and CNRS, ANR
44
45 441 Genemobile and IUF to CV. Additionally, we would like thank the DTAMB facility (from the FR41)
46
47 442 and A. Akkouche, A. Granzotto, M. P. Garcia Guerreiro, N. Burlet, and all laboratory staff of the
48
49 443 Biology Department in Brazil and the LBBE in France,.
50
51

52 444

54 445 **Literature Cited**

55
56
57 446 Akkouche A, et al. 2013. Maternally deposited germline piRNAs silence the tirant retrotransposon in
58
59 447 somatic cells. EMBO Rep. 14:458-464.
60

- 1
2
3 448 Akkouche A, et al. 2012. Tirant, a newly discovered active endogenous retrovirus in *Drosophila*
4
5 449 *simulans*. J Virol. 86:3675-3681.
6
7 450 Brennecke J, et al. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing.
8
9 451 Science . 322:1387-1392.
10
11 452 Borie, N. et al. 2002. Tissue-specificity of 412 retrotransposon expression in *Drosophila simulans* and
12
13 453 *D. melanogaster*. Heredity. 89: 247-252.
14
15 454 Bucheton A. 1990. I-transposable elements and I-R hybrid dysgenesis in *Drosophila*. Trends Genet.
16
17 455 6:16-21.
18
19 456 Bucheton A, Lavige JM, Picard G, L'Heritier P. 1976. Non-mendelian female sterility in *Drosophila*
20
21 457 *melanogaster*: quantitative variations in the efficiency of inducer and reactive strains. Heredity.
22
23 458 36:305-314.
24
25 459 Bucheton A, Paro R, Sang HM, Pelisson A, Finnegan DJ. 1984. The molecular basis of I-R hybrid
26
27 460 dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the I factor.
28
29 461 Cell. 38:153-163.
30
31 462 Bucheton A, et al. 1992. I-Elements and the *Drosophila* Genome. Genetica. 86:175-190.
32
33 463 Cavallini A, et al. 2010. Analysis of transposons and repeat composition of the sunflower (*Helianthus*
34
35 464 *annuus* L.) genome. Theor Appl Genet. 120:491-508.
36
37 465 Chaboissier MC, Busseau I, Prosser J, Finnegan DJ, Bucheton A. 1990. Identification of a potential
38
39 466 RNA intermediate for transposition of the LINE-like element I factor in *Drosophila*
40
41 467 *melanogaster*. EMBO J. 9:3557-3563.
42
43 468 Chaboissier MC, Lemeunier F, Bucheton A. 1995. IR hybrid dysgenesis increases the frequency of
44
45 469 recombination in *Drosophila melanogaster*. Genet Res. 65:167-174.
46
47 470 Chambeyron S, et al. 2008. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the
48
49 471 *Drosophila* female germline. P Natl Acad Sci Usa. 105:14964-14969.
50
51 472 Dufourt J, et al. 2013. Spatio-temporal requirements for transposable element piRNA-mediated
52
53 473 silencing during *Drosophila* oogenesis. Nucleic Acids Res. Nov27:1-13.
54
55 474 Esnault C, et al. 2008. High genetic differentiation between the M and S molecular forms of *Anopheles*
56
57 475 *gambiae* in Africa. PloS one. 3: e1968-e1968.

- 1
2
3 476 Fawcett DH, Lister CK, Kellett E, Finnegan DJ. 1986. Transposable elements controlling I-R hybrid
4
5 477 dysgenesis in *Drosophila melanogaster* are similar to mammalian lines. Cell. 47:1007-1015.
6
7 478 Fedoroff N, Wessler S, Shure M. 1983. Isolation of the transposable maize controlling elements Ac and
8
9 479 Ds. Cell 35:235-242.
10
11 480 Fontdevila A. 2005. Hybrid genome evolution by transposition. Cytogenet Genome Res. 110:49-55.
12
13 481 Fuller MT. 1998. Genetic control of cell proliferation and differentiation in *Drosophila*
14
15 482 spermatogenesis. Semin Cell Dev Biol 9:433-444.
16
17 483 Granzotto A, Lopes F, Lerat E, Vieira C, Carareto C. 2009. The evolutionary dynamics of the Helena
18
19 484 retrotransposon revealed by sequenced *Drosophila* genomes. BMC EVOL BIOL. 9:11
20
21 485 Granzotto A, Lopes FR, Vieira C, Carareto CM. 2011. Vertical inheritance and bursts of transposition
22
23 486 have shaped the evolution of the BS non-LTR retrotransposon in *Drosophila*. Mol Genet
24
25 487 Genomics. 286:57-66.
26
27 488 Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for
28
29 489 Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41:95-98.
30
31 490 Han JS. 2010. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent
32
33 491 developments, and unanswered questions. Mobile DNA. 1:15.
34
35 492 Haoudi A., et al. 1997. Developmental expression analysis of the *1731* retrotransposon reveals an
36
37 493 enhancement of Gag-Pol frameshifting in males of *Drosophila melanogaster*. Gene. 196:83-93.
38
39 494 Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: A trade-off between
40
41 495 reduced transposition and deleterious effects on neighboring gene expression. Genome Res.
42
43 496 19:1419-1428.
44
45 497 Kalmykova AI, Klenov MS, Gvozdev VA. 2005. Argonaute protein PIWI controls mobilization of
46
47 498 retrotransposons in the *Drosophila* male germline. Nucleic Acids Res. 33:2052-2059.
48
49 499 Kapitonov VV, Jurka J. 2008. A universal classification of eukaryotic transposable elements
50
51 500 implemented in Repbase. Nat Rev Genet . 9:411-412.
52
53 501 Kelleher ES, Edelman NB, Barbash DA. 2012. *Drosophila* interspecific hybrids phenocopy piRNA-
54
55 502 pathway mutants. PLoS Biol, 10: e1001428.
56
57
58
59
60

- 1
2
3 503 Kuhn R, Schafer U. Schafer M. 1988. Cis-acting regions sufficient for spermatocyte specific
4
5 504 transcriptional and spermatid-specific translational control of the *Drosophila melanogaster* gene
6
7 505 *mst(3)gl-9*. *Embo Journal*. 7:447-454.
8
9
10 506 Khurana JS, Theurkauf W. 2010. piRNAs, transposon silencing, and *Drosophila* germline development.
11
12 507 *J Cell Biol*.191:905-913.
13
14 508 Khurana JS, et al. 2011. Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell*.
15
16 509 147:1551-1563.
17
18 510 Kidwell MG. 1983a. Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. *Proc*
19
20 511 *Natl Acad Sci U S A*. 80:1655-1659.
21
22 512 Kidwell MG. 1983b. Hybrid Dysgenesis in *Drosophila melanogaster*: factors affecting chromosomal
23
24 513 contamination in the P-M system. *Genetics*. 104:317-341.
25
26 514 Kidwell MG, Novy JB. 1979. Sterility Resulting From Gonadal Dysgenesis In The P-M System.
27
28 515 *Genetics*. 92:1127-1140.
29
30 516 Kogan G.L, et al. 2003. The *GATE* retrotransposon in *Drosophila melanogaster*: mobility in
31
32 517 heterochromatin and aspects of its expression in germline tissues. *Mol. Genet. Genomics*.
33
34 518 269:234–242.
35
36 519 Malik HS, Burke WD, Eickbush TH. 1999. The age and evolution of non-LTR retrotransposable
37
38 520 elements. *Molecular Biology and Evolution* 16:793-805.
39
40 521 Massie KR, Markow TA. 2005. Sympatry, allopatry and sexual isolation between *Drosophila*
41
42 522 *mojavensis* and *D. arizonae*. *Hereditas* 142:51-55.
43
44 523 McClintock B. 1984. The significance of responses of the genome to challenge. *Science* 226: 792-801.
45
46 524 Minchiotti G, Contursi C, DiNocera PP. 1997. Multiple downstream promoter modules regulate the
47
48 525 transcription of the *Drosophila melanogaster* I, Doc and F elements. *J Mol Biol*. 267:37-46.
49
50 526 Morris CA, Benson E, White-Cooper H. 2009. Determination of gene expression patterns using in situ
51
52 527 hybridization to *Drosophila* testes. *Nat Protoc*. 4:1807-1819.
53
54 528 Moschetti R, Dimitri P, Caizzi R, Junakovic N. 2010. Genomic instability of I elements of *Drosophila*
55
56 529 *melanogaster* in absence of dysgenic crosses. *PLoS One*. 5:e13142.
57
58
59
60

- 1
2
3 530 Morozova, TV; et al. 2004 Impact of the regulatory regions of retrotransposon copia on the level of its
4
5 531 expression in testes of *Drosophila melanogaster*. Russ J Genet+. 40: 119-124.
6
7 532 Munroe DJ, et al. 1994. IRE-bubble PCR: a rapid method for efficient and representative amplification
8
9 533 of human genomic DNA sequences from complex sources. Genomics. 19:506-14.
10
11 534 Orsi GA, Joyce EF, Couble P, McKim KS, Loppin B. 2010. *Drosophila* I-R hybrid dysgenesis is
12
13 535 associated with catastrophic meiosis and abnormal zygote formation. J Cell Sci. 123:3515-3524.
14
15 536 Pelisson A, Picard G. 1979. Non mendelian female sterility in *Drosophila melanogaster* I-factor
16
17 537 mapping on inducer chromosomes. Genetica. 50:141-148.
18
19 538 Picard G. 1976. Non-mendelian female sterility in *Drosophila melanogaster*: hereditary transmission of
20
21 539 I factor. Genetics. 83:107-123.
22
23 540 Picard G, et al. 1978. Non-mendelian female sterility and hybrid dysgenesis in *Drosophila*
24
25 541 *melanogaster*. Genetical Res. 32:275-287.
26
27 542 Reed LK, Markow TA. 2004. Early events in speciation: polymorphism for hybrid male sterility in
28
29 543 *Drosophila*. P Natl Acad Sci USA. 101:9009-9012.
30
31 544 Ruiz A, Heed WB, Wasserman M. 1990. Evolution of the mojavensis cluster of cactophilic *Drosophila*
32
33 545 with descriptions of two new species. J Hered. 81:30-42.
34
35 546 Schaeffer SW, et al. 2008. Polytene chromosomal maps of 11 *Drosophila* species: the order of genomic
36
37 547 scaffolds inferred from genetic and physical maps. Genetics. 179:1601-1655.
38
39 548 Seleme MC, Busseau I, Malinsky S, Bucheton A, Teninges D. 1999. High-frequency retrotransposition
40
41 549 of a marked I factor in *Drosophila melanogaster* correlates with a dynamic expression pattern
42
43 550 of the ORF1 protein in the cytoplasm of oocytes. Genetics. 151:761-771.
44
45 551 Seleme MD, et al. 2005. In vivo RNA localization of I factor, a non-LTR retrotransposon, requires a
46
47 552 cis-acting signal in ORF2 and ORF1 protein. Nucleic Acids Res. 33:776-785.
48
49 553 Seleme MDC, et al. 2006. Extensive individual variation in L1 retrotransposition capability contributes
50
51 554 to human genetic diversity. . P Natl Acad Sci USA. 103:6611-6616.
52
53 555 Senti K-A, Brennecke J. 2010. The piRNA pathway: a fly's perspective on the guardian of the genome.
54
55
56
57
58
59 556 Trends Genet. 26:499-509.
60

- 1
2
3 557 Simonelig M, Bazin C, Pelisson A, Bucheton A. 1988. Transposable and nontransposable elements
4
5 558 similar to the I-factor involved in inducer-reactive (IR) hybrid dysgenesis in *Drosophila*
6
7 559 *melanogaster* coexist in various *Drosophila* species. P Natl Acad Sci USA. 85:1141-1145.
8
9
10 560 Streck RD, Macgaffey JE, Beckendorf SK. 1986. The structure of hobo transposable elements and their
11
12 561 insertion sites. EMBO J. 5:3615-3623.
13
14 562 Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood,
15
16 563 evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731-2739.
17
18 564 Ungerer MC, Strakosh SC, Stimpson KM. 2009. Proliferation of Ty3/gypsy-like retrotransposons in
19
20 565 hybrid sunflower taxa inferred from phylogenetic data. BMC Biol. 7:40-40.
21
22
23 566 Vaury C, Abad P, Pelisson A, Lenoir A, Bucheton A. 1990. Molecular characteristics of the
24
25 567 heterochromatic I elements from a reactive strain of *Drosophila melanogaster*. J Mol Evol.
26
27 568 31:424-431.
28
29 569 Vela D, Fontdevilla A, Vieira C, Garcia Guerreiro MP. 2014. A genome-wide survey of genetic
30
31 570 instability by transposition in *Drosophila* hybrids. Plos One (in press).
32
33 571 Vieira J, Vieira CP, Hartl DL, Lozovskaya ER. 1998. Factors contributing to the hybrid dysgenesis
34
35 572 syndrome in *Drosophila virilis*. Genet Res. 71:109-117.
36
37 573 Wasserman M, Koepfer HR. 1977. Character Displacement for Sexual Isolation Between *Drosophila*
38
39 574 *mojavensis* and *Drosophila arizonensis*. Evolution. 31:812-812.
40
41
42 575 Wicker T, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev
43
44 576 Genet. 8:973-982.
45
46 577 Zhang J, et al. 2009. Alternative Ac/Ds transposition induces major chromosomal rearrangements in
47
48 578 maize. Gene Dev. 23:755-765.
49
50
51
52
53
54
55
56
57
58
59
60

579

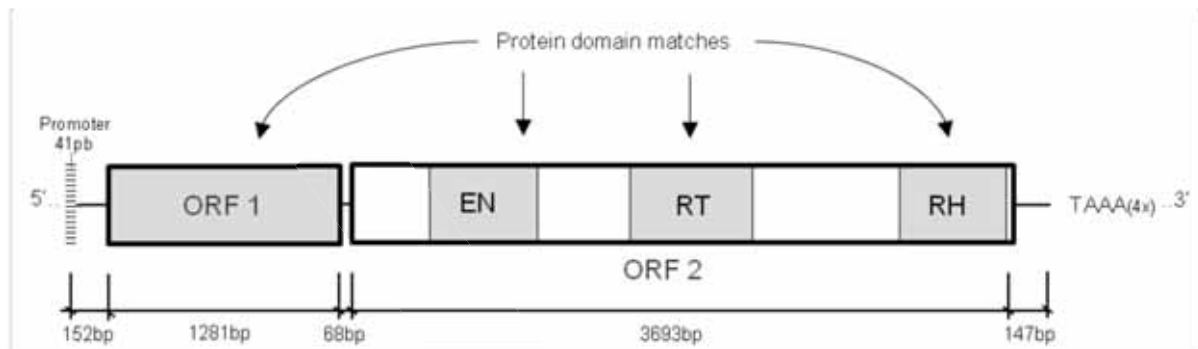
580 **Table 1.** Summary of the *I* element copies in the *D. mojavensis* genome.

Copy	Acc.	Scaff.	Start S.	End S.	Start Q.	End Q.	Or.	Chr.arms	Region	N	%(ID)	%(ID flank.)	Leng.	Integ.
RII_DMJA1(RC)*	CH933806.1	6540	12533998	12539232			-	2(E)	Central	0			5382	Comp
RII_DMJA2*	CH933807.1	6500	11558188	11563761	1	5382	+	3(B)	Central	321	0.9981	0.43	4948	Comp
RII_DMJA3*	CH933808.1	6496	21912581	21918262	1785	1052	-	5(C)	Central	1027	0.999	0.433	4505	Comp
RII_DMJX1	CH933815.1	6482	430575	438784	1	5382	+	N/A	N/A	5092	0.996	0.447	2808	Comp
RII_DMJX2	CH935525.1	12	175	1389	1	4456	-	N/A	N/A	0	0.99	B/#	1390	Def.
RII_DMJX3	CH937682.1	97	808	1490	4702	5382	-	N/A	N/A	0	0.99	B/#	681	Def.
RII_DMJX4	CH933809.1	6680	15471148	15472992	3547	5382	+	4(D)	Central	0	0.997	0.39	1848	Def.
RII_DMJX5	CH933813.1	6498	3343118	3344746	3755	5382	+	N/A	N/A	4049	0.996	0.391	1629	Def.
RII_DMJX6	CH933808.1	6496	9570337	9570944	4774	5382	+	5 (C)	central	0	0.998	0.417	609	Def.
RII_DMJX7	CH933814.1	6308	2487507	2487762	5128	5382	-	X(A)	telomeric	0	0.984	0.385	256	Def.

581 Accession number in Genbank (Acc); Scaffold (Scaff); Scaffold start hit (Start S.); Scaffold end hit (End S.); Query start hit (Start Q.); Query end hit (End
582 Q.); Orientation of the copy in genome (Or.); Chromosome arms (Chr.arms) and conventional numbering with Muller syntenic elements in parentheses;
583 Chromosome region (Region); Number of unidentified nucleotides (N); Nucleic percent identity to RC (%ID); Nucleic percent identity the 5' flanking
584 region to RC flanking region (%ID flank.); Total alignment length (Leng.); Copy integrity (Comp.: putatively complete; Def: defective copies) (Integ.); *
585 sequence used in phylogeny (Figure 4); N/A: not annotated; # not computed

586

1
2
3 587
4
5
6 588
7
8
9
10 589
11
12
13 590
14

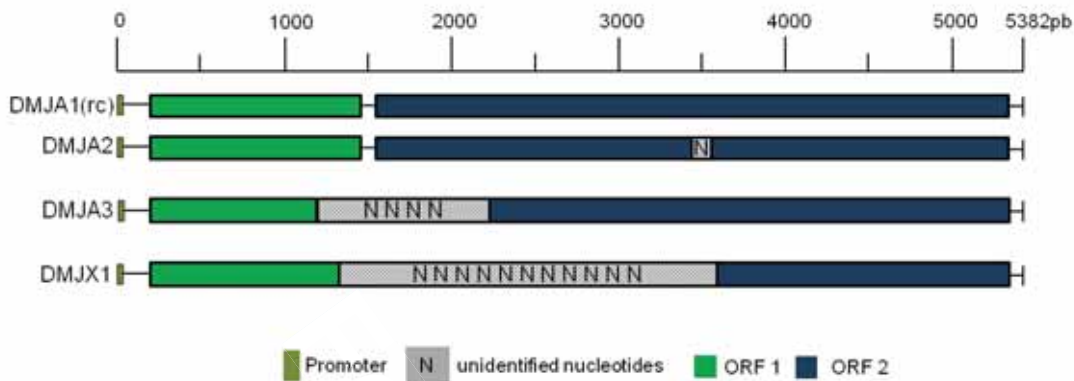


26 591
27
28
29 592
30
31 593
32
33 594
34
35 595
36
37
38 596
39
40
41
42 597
43
44
45 598
46
47
48 599
49
50
51 600
52
53
54 601
55
56
57
58
59
60

Fig. 1. The diagram represents the structure of the putative "Reference Copy" (5,382 bp) of the I-like non-LTR retrotransposon identified in the *D. mojavensis* genome. The boxes represent the two open reading frames (ORF1 and ORF2). EN: endonuclease domain; RT: reverse transcriptase domain; RH: RNase H domain; cross-hatch box: promoter region.

602

603



604

605

606 **Fig. 2.** Structure of the putatively complete copies of the I like non-LTR retrotransposon
 607 identified in the *D. mojavensis* genome.

608 DMJA1 to X1: copies shown in Table 1. Shaded gray boxes: unidentified nucleotides (N); red
 609 boxes: 5'-end promoter position; green boxes: ORF1; blue boxes ORF2; scale size range of the
 610 copies is presented in base pairs.

611

1
2
3 612
4
5
6
7 613
8
9
10 614
11
12
13 615
14
15

	Sites																																										
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41		
<i>D. melanogaster</i>	C	A	T	T	A	C	C	A	C	T	T	C	A	A	C	C	T	C	G	A	A	G	A	G	A	T	A	A	G	T	C	G	T	G	C	C	T	C	T	C			
DMJA1 (rc)	.	.	G	.	T	T	C	.	G	A	T	G	.	A	T
DMJA2	.	.	G	.	T	T	C	.	G	A	T	G	.	A	T	
DMJA3	.	.	G	.	T	T	C	.	G	A	T	G	.	A	T	
DMJX1	.	.	G	.	T	T	C	.	G	A	T	G	.	A	T	

616

617 **Fig. 3.** Sequence alignment of the *D. melanogaster* *I* promoter sequence with the putatively
618 complete *D. mojavensis* *I* copies.

619 Polymorphism sites are highlighted in gray, and identical nucleotides are indicated by a dot.

620

621

622

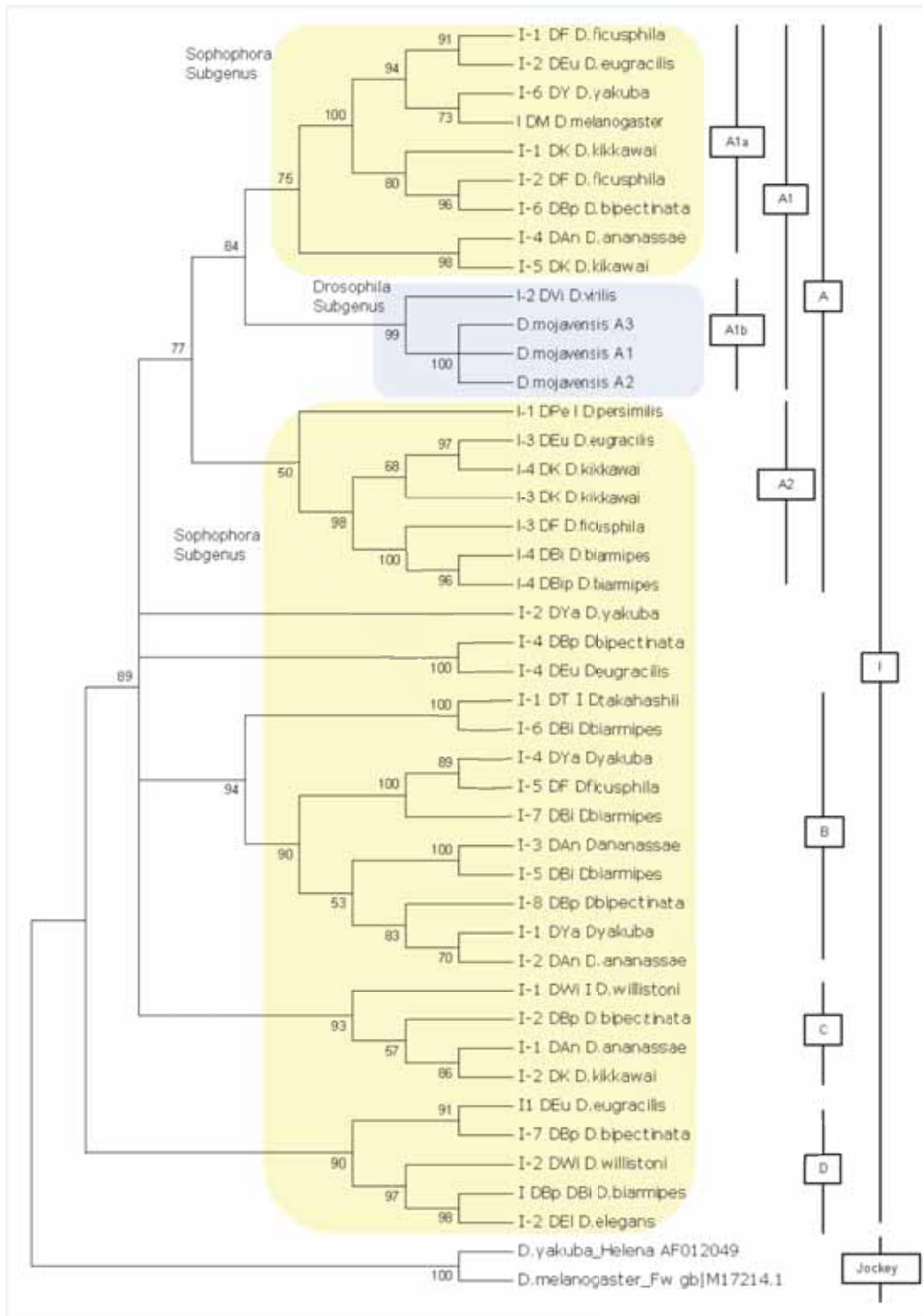
623

624

625

626

51
52
53
54
55
56
57
58
59
60



627

1
2
3 628 **Fig. 4.** Phylogenetic analyses of *I* elements based on the amino acid sequence of the reverse
4
5 629 transcriptase domains (219 sites), carried out with the Maximum Likelihood method, gamma
6
7 630 distributed with 2 discrete categories, 500 bootstrap, partial deletion analysis, and the JTT with Freqs.
8
9 631 (F+) model. Only bootstrap values greater than 50% are indicated.
10
11

12 632

13 633

14 634

15 635

16 636

17 637

18 638

19 639

20 640

21 641

22 642

23 643

24 644

25 645

26 646

27 647

28 648

29 649

30 650

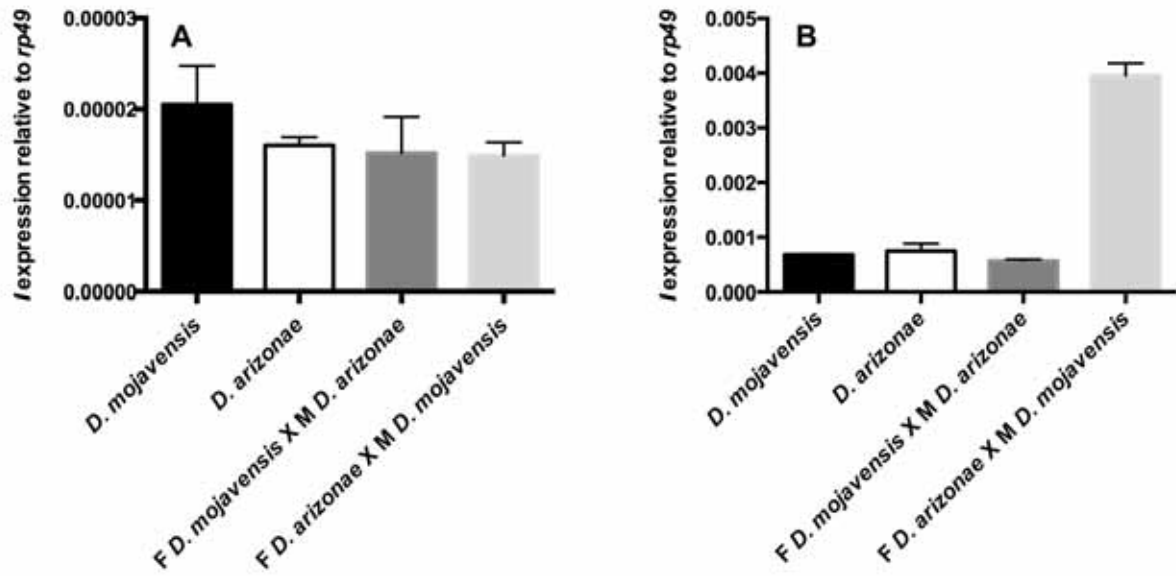
31 651

32 652

33

34

35



653

654 **Fig. 5.** Expression of *I*-like elements by RT-qPCR in ovaries (A) and testes (B) of *D. mojavensis*,
655 *D. arizonae*, and their hybrids (F for female and M for male).

656 The transcript levels of the *I* element were estimated relative to those of the *rp49* gene. No
657 significant differences were observed in ovaries (A). On the contrary, in testis, mRNA levels
658 were significantly different (two way ANOVA).

659

660

661

662

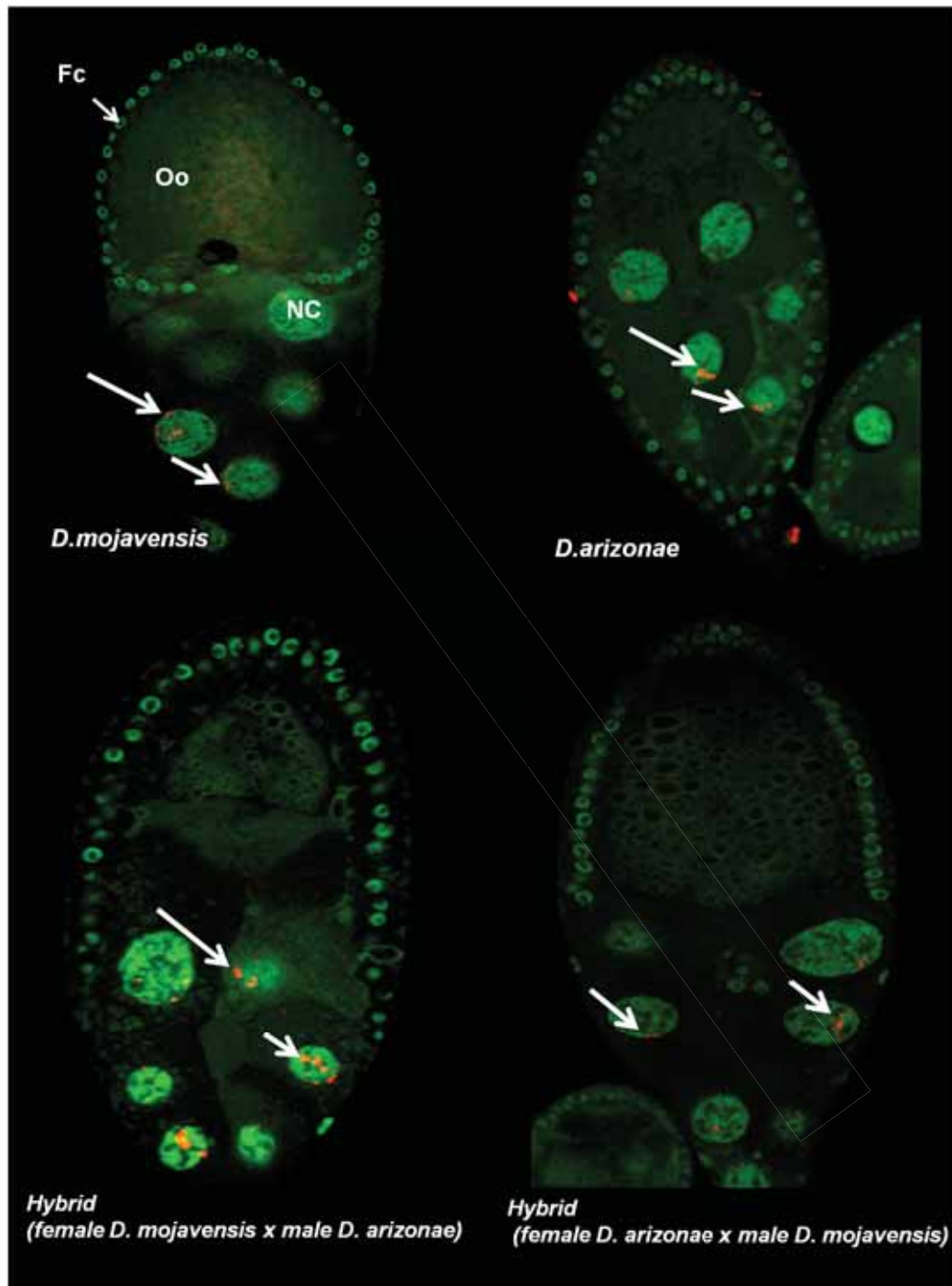
663

664

665

666

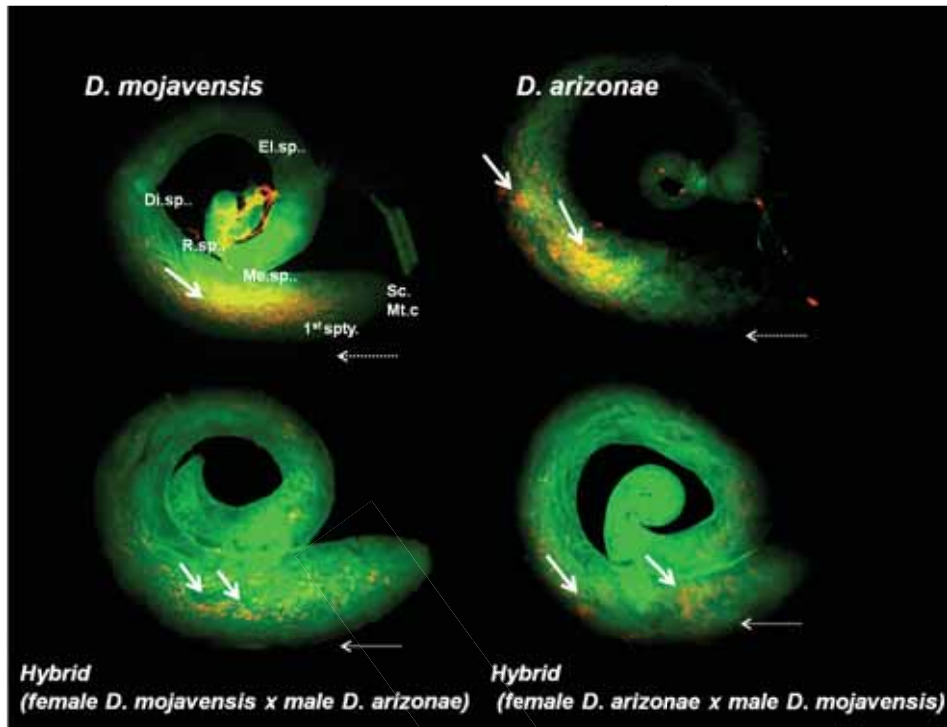
667



668

669 **Fig. 6.** Immunofluorescence for the *I*-element mRNA probe (RNA in red, DNA in green) in the egg
 670 chamber of the parental lines and hybrid crosses. Oo, oocyte; NC, nurse cells; Fc, follicle cells. The
 671 arrows indicate mRNA of the detected *I* transcripts.

672



673

674 **Fig. 7.** Immunofluorescence for the *I*-element mRNA probe (RNA in red, DNA in green) in testes of
 675 the parental lines and hybrid crosses. Yellow arrows indicate detected mRNA transcripts. Sc.Mt.c -
 676 Sperm and mitotic cells; 1st spty - Primary spermatocytes; Me.sp.- meiotic spermatocytes; R.sp- round
 677 spermatids; Di.sp- differentiating spermatids; El.sp- elongating spermatids; horizontal arrow indicates
 678 the direction of differentiation (stem cells are located at the apical tip, and germ cells move toward the
 679 basal end as they differentiate).

680

681

682

683

684

685

686

1 Supplementary Material

2 **Supplementary Table 1.** Individuals analyzed by Transposon display. From the six individual
 3 crosses, we follow 2 complete families. All the individuals that are presented in this table
 4 were analyzed by transposon display, according to Supplementary Figure 1 (n=86).

5

Parental crosses	F1 Analyzed offspring	F1 crosses	F2 Analyzed offspring
♀ 129F X ♂ P129 <i>D.mojavensis</i> X <i>D.arizonae</i>	F129♂1 F129♂2 F129♂4 F129♂5		
	F129♀A F129♀C F129♀J F129♀L F129♀O F129♀R	F129♀A X ♂ <i>D. arizonae</i>	5 ♀ and 3 ♂
		F129♀J X ♂ <i>D. arizonae</i>	4 ♀ and 5 ♂
		F129♀L X ♂ <i>D. arizonae</i>	5 ♀ and 5 ♂
		F129♀O X ♂ <i>D. arizonae</i>	5 ♀ and 5 ♂
		F129♀R X ♂ <i>D. arizonae</i>	4 ♀ and 3 ♂
♀ 149F X ♂ P149 <i>D.mojavensis</i> X <i>D.arizonae</i>	F149♂1 F149♂2 F149♂4		
	F149♀A F149♀B F149♀C F149♀D F149♀E		
		F149♀B X ♂ <i>D. arizonae</i>	3 ♀ and 4 ♂
		F149♀E X ♂ <i>D. arizonae</i>	4 ♀ and 2 ♂

6

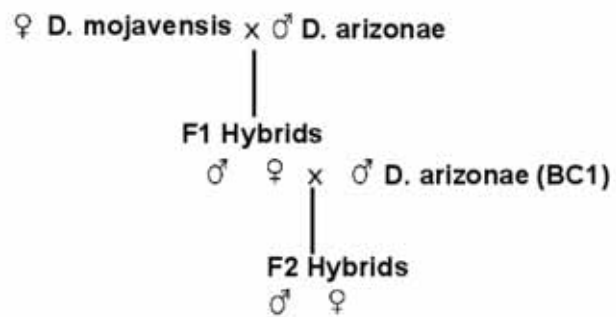
7

8

9

10

11

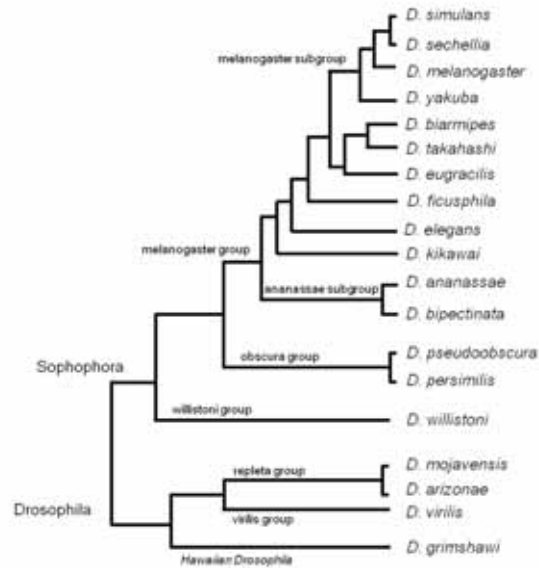


12

13 **Supplementary Fig. S1.** Transposon Display cross diagram

14

15

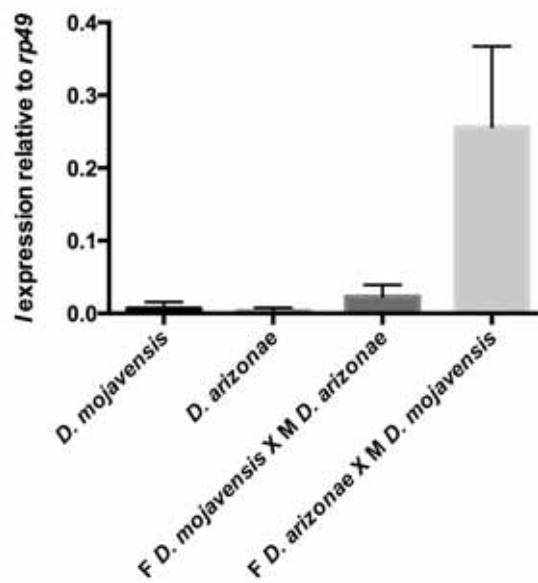


16

17 **Supplementary Fig. S2** Drosophila species phylogeny, according to Linde KVD, Houles D,
 18 Spicer GS, Steppan S. 2010. A supermatrix-based molecular phylogeny of the family
 19 Drosophila. Genetics Research, 92:25-28.

20

21



22

23 **Supplementary Fig. S3.** Expression of testes and ovaries in the second experiment.

CAPÍTULO 2

CAPÍTULO 2

Mobiloma expresso em *Drosophila mojavensis* e *D. arizonae* e seus híbridos: Análises preliminares

Introdução

Para entendermos como a composição e a complexidade dos genomas evoluem é essencial compreendermos a dinâmica dos elementos de transposição (TEs), dentro e entre populações, e em diferentes espécies, uma vez que diversos processos biológicos e históricos influenciam sua composição (VIEIRA et al., 2012). A evolução dos genomas consiste em uma série de episódios de reorganização nos quais a transposição pode ter um papel significativo, pois os TEs podem moldar consideravelmente a estrutura e função dos genomas e, assim, a evolução dos organismos hospedeiros (KIDWELL; LISCH, 2001). Não apenas a interferência dessas sequências móveis no conteúdo e atividade de genes específicos, como também em redes de regulação de muitos genes, pode ter consequências importantes no processo de especiação (BIÉMONT, 2010).

A hibridação ou introgressão de material genético em uma população pode desencadear a transposição de TEs e instabilidade genômica que, em tese, poderia proporcionar ao híbrido a capacidade de adaptação e especiação, ou mesmo levá-lo a extinção (FONTDEVILA, 2005). Apesar das muitas evidências de ativação de TEs em híbridos interespecíficos, ocasionando uma explosão de transposição, sua associação com a especiação não está completamente estabelecida (FONTDEVILA, 2005; REBOLLO et al., 2010). Um exemplo clássico dessa proliferação refere-se a resultados de cruzamentos entre espécies de girassóis (*Helianthus anomalus*, *H. paradoxus* e *H. deserticola*) em que o genoma dos

híbridos aumentou em até 50% devido à proliferação de sequências de retrotransposons Ty3/gypsy-like (UNGERER et al., 2009). Em *Drosophila*, podemos citar as explosões de transposição do retrotransposon *Oswaldo* relacionada à hibridação interespecífica em linhagens de *D. buzzatii* e *D. koepferae* (LABRADOR; FONTDEVILA, 1994; LABRADOR et al., 1999; PANTAZIDIS; LABRADOR, 1999).

Pouco são os trabalhos que exploram a anotação, classificação e expressão global dos TEs em espécies e híbridos. É particularmente difícil de determinar os limites reais das sequências de TEs com precisão, e apesar de cópias que pertencem a uma determinada família serem semelhantes em sequência, não são idênticas, pois mecanismos evolutivos geram mutações pontuais e/ou rearranjos e os chamados “*indels*”. Estes mecanismos resultam em cópias fragmentadas, divergentes e mosaicos que são difíceis de identificar por abordagens de similaridade, além disso, às vezes TEs estão inseridos em outras cópias de TEs para formar elementos aninhados. Estas características é uma dificuldade recorrente exigindo softwares, algoritmos e análises computacionais complexas (LERAT, 2010). O sequenciamento do transcriptoma completo (conjunto total de transcritos em uma célula em um estágio de desenvolvimento ou condição fisiológica específicos) por RNA-Seq, chamado de alto desempenho, recentemente aplicado ao estudo do mobiloma expresso (conjunto de transcritos de TEs de um transcriptoma), está auxiliando a desvendar a dinâmica de expressão de TEs em diferentes espécies (KELLEHER et al., 2012); porém; pouco se sabe sobre essa dinâmica em híbridos.

O “*RNA sequencing*” (RNA-Seq) refere-se ao uso de uma técnica de sequenciamento de alto rendimento para sequências de cDNA com objetivo de se obter informações sobre o transcriptoma. Esta técnica que permite o acesso ao conteúdo do transcriptoma proporciona uma estimativa da concentração relativa de transcritos e isoformas e fornece um indicativo da regulação e da expressão gênica. Com ela é possível quantificar mudanças no nível da

expressão de cada transcrito durante o desenvolvimento e sob condições distintas (WANG et al., 2009).

O sequenciamento do transcriptoma por RNA-Seq produz milhões de pequenos transcritos, chamados “*reads*”, com cerca de 30 a 400pb, dependendo da tecnologia utilizada (Illumina, Applied Biosystems SOLiD e Roche 454 Life Science). O resultado dos *reads* é então alinhado a um genoma referência ou a transcritos de referência, ou pode ser realizado por uma montagem *de novo* (sem a sequência genômica), para produzir um mapa de transcrição do genoma global, que consiste da estrutura transcricional e/ou do nível de expressão de cada gene (WANG et al., 2009). O transcriptoma de *D. melanogaster*, um dos organismos modelo mais estudados sob esse aspecto, revelou após o sequenciamento por RNA-Seq em 2010, mais 111,195 sequências, 14,862 genes e 17,745 novos transcritos (GRAVELEY et al., 2011). Em outro sequenciamento do transcriptoma de *D. melanogaster*, e de *D. sechellia*, o total de *reads* brutos foi de 26 e 31 milhões, respectivamente, e nos híbridos entre fêmeas *D. melanogaster* e machos *D. sechellia* obtiveram 78 milhões de sequências reads (MCMANUS; et al., 2010).

Apenas dois estudos utilizam RNA-Seq para análise de alto rendimento de transcriptoma de TEs em *Drosophila* (GRAVELEY et al., 2011; KELLEHER et al., 2012) e apenas Kelleher et al. (2012) focam o silenciamento de TEs e produção de piRNA em fêmeas F1 de cruzamentos interespecíficos. Os autores mostraram que ocorre uma desrepressão generalizada de TEs maternos e paternos em híbridos de cruzamentos entre *D. melanogaster* e *D. simulans* (espécies irmãs que divergiram entre 3-5 milhões de anos) (TAMURA et al., 2004). Nesse estudo, 32 das 265 famílias de TEs expressas exibiram um aumento de duas vezes mais transcritos nos ovários dos híbridos em relação a ambos os parentais.

Para testar a hipótese de que a hibridação conduz um aumento da taxa transcrição de transcritos de TEs, devido a um relaxamento do sistema de regulação, neste estudo comparou-

se a taxa de transcritos de TEs obtida por RNA-Seq em *D. mojavensis* e *D. arizonae* e seus híbridos. Estas espécies, que pertencem ao grupo *repleta* do subgênero *Drosophila*, divergiram entre 0,66 e 1,2 milhão de anos e, embora o grau de isolamento pré-zigótico entre as duas seja forte, é ainda incompleto e variável. Ocorre produção de híbridos em cruzamentos entre diferentes populações, dependendo da origem geográfica, em especial de população de *D. mojavensis*, sendo as populações simpátricas mais isoladas sexualmente que as alopátricas em cruzamentos interespecíficos (WASSERMAN; KOEPPER, 1977; REED et al., 2008; BONO; MARKOW, 2009).

Os resultados obtidos, mesmo que ainda preliminares, são inéditos, pois pouco se conhece sobre a dinâmica global de expressão dos TEs e seu comportamento em situação de hibridação interespecífica. Com a expressão global dos elementos quantificada, pode-se agora direcionar com eficiência diferentes pesquisas sobre TEs nesse par ou grupo de espécies. Identificamos expressão diferencial de TEs, representados por diferentes classes e famílias, tanto em *D. mojavensis* e *D. arizonae* quanto nos híbridos. Dos elementos com alta expressão, dois foram eleitos para validação por análises de amplificação em tempo real (RT-qPCR).

Material e Métodos

Obtenção das amostras e sequenciamento

Linhagens

Foram sequenciados mRNA de amostras obtidas da extração de RNA de ovários de fêmeas de *D. mojavensis* e *D. arizonae* e dos seus híbridos recíprocos com 10 dias de (fêmeas *D. mojavensis* x macho *D. arizonae*: híbridos A e fêmeas *D. arizonae* x machos *D. mojavensis*: híbridos B). As duas linhagens parentais foram obtidas da coleção do UC San Diego *Drosophila* Stock Center (San Diego, CA, USA): *D. mojavensis* (stock number: 15081-

1352.01, Anza Borrego Desert, CA, EUA) e *D. arizonae* (stock number: 15081-1271.17, Metztlán, Hidalgo, México).

Preparação das amostras

Os indivíduos parentais foram separados virgens, por sexo, um dia após a eclosão. No terceiro dia foram utilizados em cruzamentos para obtenção dos híbridos F1. Para isso, foram colocados 10 machos e 8 fêmeas, em tubos 2,3 x 9,5cm contendo meio de cultura. Fêmeas virgens parentais F1 e híbridos foram coletadas destes tubos após a eclosão, com dois dias de idade, e isoladas até atingirem 10 dias.

As fêmeas foram dissecadas em tampão TBE 1x, seus ovários foram separados em tubos de 1,5 mL contendo aproximadamente 20 ovários, em seguida as amostras em triplicata foram submetidas à extração de RNA total. As extrações foram realizadas com o kit RNeasy (Qiagen) e as amostras foram em seguida tratadas com DNase (DNA-free kit; Ambion) e armazenadas a -80°C. As amostras foram quantificadas por fluorescência em bioanalyser 2100 (Agilent), obedecendo a critérios pré-estabelecidos pela plataforma de sequenciamento (RIN>8). Um total de 5µg de RNA foi obtido a partir de um “pool” de amostras, das quais foram separadas e enviadas para sequenciamento em duplicata.

Sequenciamento

O sequenciamento foi realizado em Illumina HiSeq 2000, sob um banco unicamente de transcritos RNA-poly(A)+ pela plataforma de sequenciamento ProfileXpert (Plataforma ProfileXpert – Service de Génomique & Microgénomique, vinculada a Structure Fédérative de Recherche Santé Lyon-Est (SFR Santé Lyon-EST) e gerado por Ezus Lyon, filial da Universidade Claude Bernard Lyon 1, Lyon, França). As etapas realizadas para a obtenção dos *reads* foram as seguintes:

- a. Controle e qualidade dos RNAs (na recepção das amostras): RNA foi quantificado em fluorescência e RIN>8 em Bioanalyser 2100 (Agilent).
- b. Criação dos bancos: Os RNAs foram validados por dosagem Picogreen e em Bioanalyser por uma segunda vez pela plataforma de sequenciamento. Os bancos foram criados utilizando Truseq RNA indexado ao kit TruSeq RNA do Illumina e adicionados sequências controles (resultados disponíveis nos arquivos gerados).
- c. Sequenciamento HiSeq-2000: geração de clusters sob cBOT e sequenciamento no HiSeq-2000: corrida Paired End de 51pb com kit v3 (51x7x510. Um relatório de qualidade dos dados foi entregue com os dados.
- d. Análise primária: corresponde ao “base calling”, erros gerados pelos algoritmos que captam a intensidade de fluorescência, realizada com auxílio do software RTA.

Seleção e análise dos TES no transcriptoma sequenciado

A primeira fase de análise dos TEs no transcriptoma pode ser resumida em quatro etapas: 1) Anotação dos TEs do genoma de *D. mojavensis*; 2) seleção de sequências completas; 3) extração das sequências e confirmação; e 4) quantificação de transcritos no transcriptoma.

Primeiramente foi realizada a identificação dos TEs no genoma sequenciado de *D. mojavensis* (release 1.3 06/20090) obtido no banco Flybase com a ferramenta *RepeatMasker* (SMIT et al. 2004) que realiza uma busca no genoma tendo como referência a biblioteca de TEs do gênero *Drosophila* (*Update* 2007/12/04) do RepBase (JURKA et al. 2005; <http://www.girinst.org/RepBase/update/index.html>).

Com as sequências de cada TE de *D. mojavensis* e posição no genoma, selecionou-se aquelas que possuíam tamanho próximo ao elemento primeiramente descrito na literatura e mais de 60% de identidade aos depositados no banco de dados do Repbase

(<http://www.girinst.org/rebase/update/browse.php>). Em seguida, com a referência da posição das sequências de interesse de cada TE no genoma de *D. mojavensis*, fez-se a extração de suas sequências. Essas por sua vez, foram realinhadas por BLAST (NCBI website) com os respectivos elementos primeiramente descritos na literatura para confirmação de similaridade. Com os transcritos dos TEs selecionados, o arquivo tipo “fasta” foi submetido à busca de similaridade com os *reads* do RNA-Seq para todas as réplicas.

Uma quantificação da abundância de *reads* que se alinham ao longo da sequência dos TEs foi realizada utilizando o software package RSEM - RNA-Seq by Expectation Maximization (LI; DEWEY, 2011) e o resultado do alinhamento foi usado para estimar a abundância de transcritos e sua credibilidade ao longo do transcrito referência. Com os resultados dos alinhamentos, gráficos das abundâncias dos *reads* ao longo do elemento foram confeccionados no software Excel (Microsoft). Essas análises foram realizadas em colaboração com Vincent Lacroix (equipe Baobab) e H el ene Lopes (equipe Elementos de Transposi ao, Evolu ao e Popula oes), respectivamente pesquisador e aluna de doutorado do laborat rio LBBE (*Laboratoire de Biom trie et Biologie Evolutive*) da Universidade Claude Bernard Lyon1 – Lyon, Fran a.

Tratamento dos dados e an lises

Como o genoma sequenciado de *D. mojavensis* n o   de alta qualidade fez-se uma an lise de busca-e-alinhamento para reconstruir os transcritos sem utilizar um genoma de refer ncia. Para isso, os *reads* das duas esp cies (*D. mojavensis* e *D. arizonae*), e dos h bridos, foram misturados e tratados como um  nico conjunto e em seguida alinhados. Dessa maneira foram obtidos *contigs* que s o comuns  s duas esp cies e seus h bridos. Sendo assim, fez-se um pr -alinhamento dos dados das duas esp cies e h bridos usando Trinity (GRABHERR et al., 2011) para reconstru o do mobiloma expresso a partir dos dados do RNA-Seq.

Montagem

Como as duas linhagens parentais divergiram recentemente ($<1\text{mya}$), assumiu-se que estas possuem os transcritos compartilhados para as quatro amostras (espécie/híbridos). Os dados foram pré-montados, o que significa que foram usados os *reads* de todas as espécies ou híbridos para reconstruir um *contig* consenso. Este método apresenta duas grandes vantagens:

1. Se a espécie/híbrido compartilha uma sequência de expressão ortóloga, o software Trinity reconstruirá somente uma sequência consenso para as quatro espécies.
2. Os *reads* com baixa expressão são facilmente montados.

Para reconstrução dos *contigs* foi utilizado o programa Trinity, que faz uma montagem *de novo*, baseada num gráfico de Bruijn, com $k=25$.

Quantificação

Para quantificar a abundância de *contigs* em cada réplica para cada espécie, os *reads* de cada espécie foram alinhados separadamente com os *contigs* usando Bowtie (LANGMEAD et al., 2009), e então foram quantificados para cada sequência usando RSEM (LI; DEWEY, 2011). Ao número de *reads* alinhados em um *contig* chamou-se de *xcount*. A partir disso, o *xcount* foi usado para testar a expressão diferencial.

Reprodutibilidade das réplicas

Se as réplicas são similares, espera-se que o número de *reads* alinhados no *contig* (*xcount*) da réplica 2 em função do *xcount* da réplica 1 sejam correspondentes, ou seja, que os pontos estejam dispersos em torno de uma linha bissetriz quando representados em um gráfico de dispersão. Os resultados indicam que as réplicas são correspondentes, ou seja, há uma homogeneidade no valor de *xcount* entre as réplicas, visto que poucos pontos fogem estatisticamente da linha bissetriz no gráfico, demonstrando assim a reprodutibilidade dos

dados e da qualidade do sequenciamento. Este foi um critério determinante para que pudéssemos seguir com as análises (Figura S1).

Alinhamento com TEs

Primeiramente verificou-se a similaridade dos *contigs* contra os TEs referências de *Drosophila* depositados no banco Repbase, tanto de retrotransposons quanto de transposons de DNA. Para isso, os *contigs* foram misturados e a similaridade com os TEs foi obtida pela adição com o número de *reads* alinhados em cada *contig* (*xcount*).

Foram apenas considerados os TEs identificados segundo os seguintes critérios:

- 1) e-value máximo 10^{-10} ;
- 2) tamanho do alinhamento maior que 900 pb ou cobertura de 60% com a sequência de TE referência.
- 3) em caso de alinhamentos múltiplos (sequência *query* possuindo dois ou mais alinhamentos com dois ou mais TEs), escolheu-se o melhor *hit* se a razão do tamanho do alinhamento fosse maior que 1,3.

Foram anotados os TEs de acordo com o nome depositado no RepBase. Os valores de expressão e o número de *reads* alinhados em cada elemento foram então quantificados e normalizados segundo o critério de FPKM (Tabela Suplementar S1). FPKM é uma estimativa definida como Fragmentos por kilobase de transcritos por milhões de fragmentos mapeados, "Fragments Per Kilobase of transcript per Million fragments mapped" (TRAPNELL et al., 2010).

Análise de expressão diferencial

A expressão diferencial foi calculada segundo o modelo de distribuição binomial negativa DESeq (ANDERS; HUBER, 2010) com o software estatístico R (<http://www.r->

project.org/index.html). Os TEs diferencialmente expressos entre parentais e híbridos foram anotados com as seguintes informações (Tabela S2):

1. Nome do elemento (segundo identificação RepBase);
2. Valores de Log_2 de FoldChange (FC): onde FC é a razão entre a média *xcount* (para duas réplicas) da espécie 1 e a média *xcount* da espécie 2;

$$FC = \frac{\textit{xcount condição 1}}{\textit{xcount condição 2}}$$

3. P-value obtido por DESeq (apenas p-value < 0,0001);
4. *Xcount* para a primeira espécie;
5. *Xcount* para a segunda espécie.

Um diagrama de caixa, ou *boxplot*, foi criado para se ter uma visão global da variação de reads (em genes e TEs) para as condições parentais nos dois sentidos (fêmeas *D.mojavensis* vs machos *D. arizonae* e fêmeas *D. arizonae* vs machos *D.mojavensis*), e as condições híbridas (A vs B e B vs A), contra o transcriptoma total e contra TEs, respectivamente (Figura S2).

Os elementos diferencialmente expressos entre os parentais e híbridos recíprocos (*D. mojavensis* vs *D. arizonae*; *D. arizonae* vs *D. mojavensis* e híbrido A vs B e B vs A) foram identificados e calculados seus valores de *xcount*, FoldChange e p-value (Tabela S2).

Classificação dos padrões de herança da expressão

O modo de herança para TEs diferencialmente expressos ($p > 0,005$) foi determinado usando software R. O total de expressão foi normalizado dividindo o número total de *reads* do TE com total de *reads* no genoma e multiplicado por 100, fornecendo uma porcentagem de expressão. Em seguida, essa porcentagem, transformada em Log para cada parental foi subtraída daquelas dos híbridos para examinar a mudança de expressão. TEs cuja expressão

total nos híbridos desviou mais que 1,25 vezes de qualquer um dos seus parentais foram considerados ter herança não conservada, e foram classificados como tendo herança aditiva, dominante, sub ou superdominante, com base na magnitude da diferença entre a expressão total no híbrido e em cada espécie parental (MCMANUS et al., 2010). Para facilitar a compreensão, na Figura 1 são ilustrados em exemplos hipotéticos os diferentes padrões de herança. No padrão conservado tem-se o mesmo nível de expressão em parentais e híbridos; no aditivo, o nível corresponde à média da expressão dos parentais; no dominante um dos parentais (*D. mojavensis* ou *D. arizonae*) tem maior expressão que a do híbrido correspondente; no superdominante há maior expressão no híbrido que nos parentais; e no codominante, menor expressão no híbrido que nos parentais.

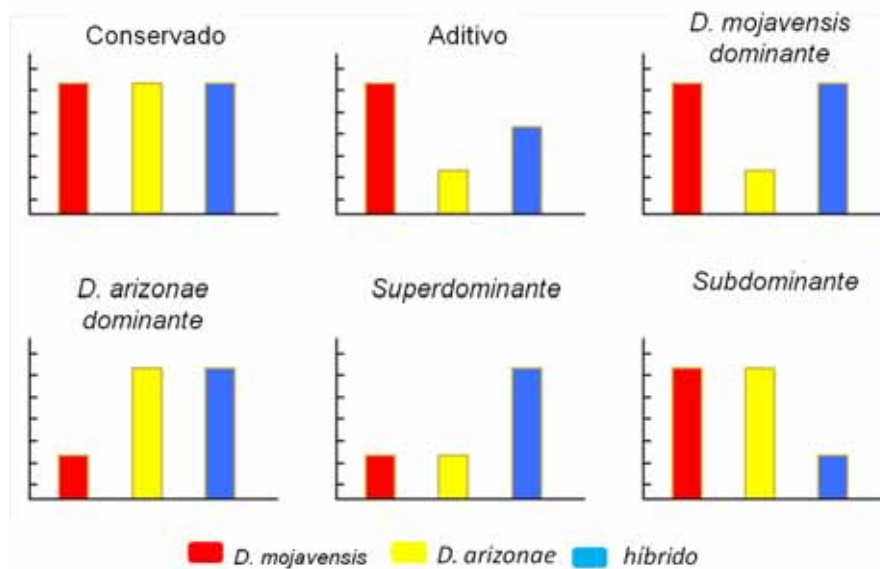


Figura 1. Ilustração hipotética de expressão de TEs em *D. mojavensis* (vermelho), *D. arizonae* (amarelo), F₁ híbridos (azul), ilustrando modelos de herança: conservado, aditivo, *D. mojavensis* dominante, *D. arizonae* dominante, superdominante, subdominante.

Validação por RT-qPCR

Dois elementos identificados pelo alinhamento dos RNA-Seq e o banco de TEs foram selecionados e quantificados por RT-qPCR. O objetivo foi validar os dados encontrados pelas análises *in silico* dos dados do RNA-Seq, por análise quantitativa *in vitro* da expressão dos elementos. Os elementos selecionados para isso foram os retrotransposon com LTR *Copia-1* e *Gtwin*, por terem apresentado o maior grau de superexpressão em híbridos.

Depois de identificada a expressão diferencial absoluta (FPKM) entre parentais e híbridos, utilizou-se a sequência consenso construída do alinhamento dos *reads* pelo Trinity, como base na sequência referência dos elementos *Copia-1* e *Gtwin* depositada no Repbase, para construir oligonucleotídeos iniciadores das regiões da transcriptase reversa (RT), com auxílio da plataforma IDT SciToolsPrimerQuest (<http://www.idtdna.com/Home/Home.aspx>). Esse oligonucleotídeos foram utilizados nas reações de RT-qPCR. Os níveis de expressão dos elementos foram mensurados a partir da expressão do gene endógeno constitutivo ribossomal rp49 de *D. mojavensis* (GRANZOTTO et al. 2009). As sequências dos oligonucleotídeos iniciadores foram as seguintes: *Gtwin* forward 5'- CGC TGA CGG CAA TAA TGA AAG C – 3' e *Gtwin* reverse 5' – ATC TTC CGA TGC CAA GAT A -3'; *Copia* forward 5' - GTG GAC CTA TAA GGC AAG TAT C – 3' e *Copia* reverse 5'- AGA CCT TTC TGA CGC TCT A - 3'.

O RNA convertido em cDNA o mesmo utilizado no RNA-Seq, foi tratado com DNase (DNA-free-kit; Ambion). Um micrograma do RNA total foi convertido em cDNA usando o kit Thermoscript Invitrogen. O cDNA foi então diluído 50 vezes e diluições seriadas (1:10) do produto de cDNA do elemento amplificado por PCR foram utilizadas para estabelecer a curva padrão de amplificação (diluição 1:10). O número de cópias foi quantificado usando SYBR Green quantitative PCR em sequenciador LightCycler 480 apparatus (Roche Diagnostics).

Resultados e Discussão

Averiguação da qualidade dos dados

A partir da busca e alinhamento de reconstrução de transcritos, realizada pelo Trinity, foram reconstruídos 26.661 *contigs*, com distribuição igualitária de *xcount* entre as réplicas 1 e 2 dos parentais e seus híbridos (Figura S1). Estes resultados indicam uma variação esperada entre réplicas. Podemos notar variabilidade maior na distribuição de *xcount*, apenas dentro das amostras de *D. arizonae*.

A distribuição global de “FoldChange”, tanto na comparação entre *D. mojavensis* e *D. arizonae*, como também nos híbridos A e B, tanto para todos os transcritos (incluindo genes), como para TEs é apresentada na Figura S2. Nota-se que a variabilidade de *xcounts* entre híbridos e parentais é homogênea, com medianas próximas e intervalos de confiança compartilhados entre todas as comparações, com maior amplitude quando considerados todos os transcritos em relação aos TEs. Se observamos apenas TEs, verificamos maior amplitude de transcritos entre *D. mojavensis* e *D. arizonae* em relação aos híbridos AvsB. Este resultado permite prever em qual conjunto de dados podemos encontrar maior variabilidade.

Expressão de TEs em D. mojavensis e D. arizonae

Os alinhamentos dos *reads* do RNA-Seq obtidos contra o banco de TEs do RepBase revelou que em *D. mojavensis* e *D. arizonae*, espécies evolutivamente próximas, há expressão diferencial dentro e entre classes e famílias de TEs. Foram identificados 47 TEs com expressão diferencial; destes, 54% são retrotransposons com LTRs (LTRs) e 13% são retrotransposons sem LTRs (NLTRs) e 33% são Transposons de DNA. Essas proporções são exatamente as mesmas que as descritas para TEs no genoma de *D. mojavensis* na publicação *Drosophila 12 Genomes Consortium* (CLARK et al., 2007): 54% LTRs, 12% LINES e 33% a

transposons de DNA. Em outras espécies de *Drosophila*, a maior proporção genômica de TEs também corresponde aos LTRs (60 a 70%), seguindo-se os NLTRs (20 a 30%) e Transposons de DNA (5 a 20%). Dentre os LTRs expressos nas duas espécies parentais aqui estudadas, 50% são representados por elementos *Gypsy-like*; entre os LINEs, 67% estão representados por elementos da superfamília *Jockey*; já entre os Transposons de DNA, 52% são de elementos da superfamília hAT (Figura 2).

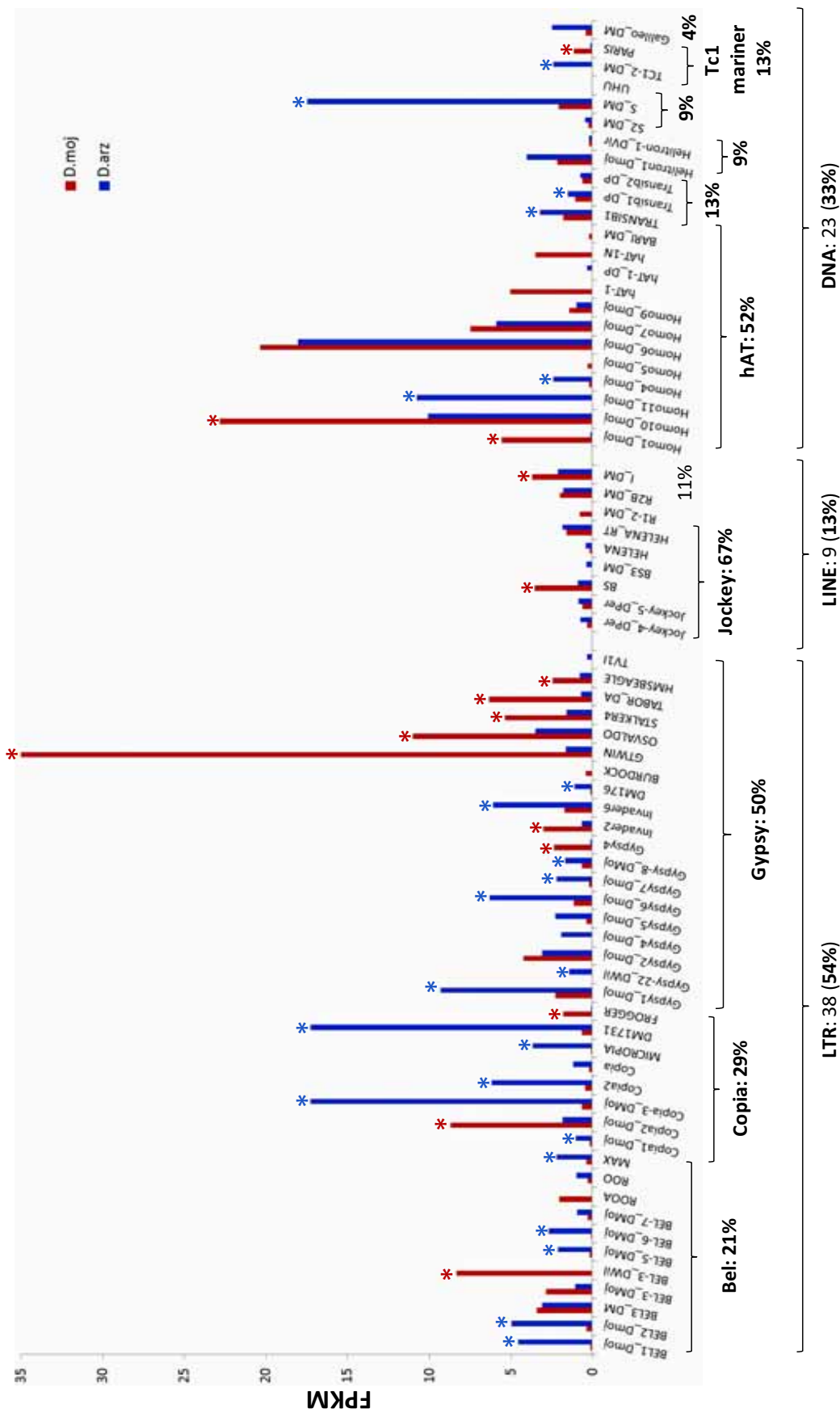


Figura 2. Elementos de transposição identificados com expressão diferencial do mobiloma de ovários de *D. mojavensis* e *D. arizonae* por análises de RNA-Seq ($\geq 60\%$ de identidade com TEs do Repbase). Os asteriscos indicam variação de expressão estatisticamente significativa.

Expressão diferencial nos híbridos

De maneira global não observamos uma explosão de transcrição de TEs nos híbridos F1 entre *D. mojavensis* e *D. arizonae*. No entanto, observamos alguns elementos particulares que estão diferencialmente expressos em relação aos parentais, dentro e entre famílias de TEs. Nota-se, por exemplo, que os NLTRs são encontrados com baixa expressão em relação aos LTRs ou transposons de DNA. Além disso, embora em baixo número, há uma proporção relativamente maior de transposons de DNA superexpressos nos híbridos que os LTRs (Figuras 3 e 4). Como acima observado, ocorre variabilidade no nível de expressão de um elemento da mesma família dentro e entre parentais e híbridos, além da diversidade de elementos de diferentes classes que são identificados com diferentes níveis de expressão (Classe I, ou retrotransposons e Classe II, ou transposons de DNA). Dentre os elementos diferencialmente expressos entre parentais e híbridos, oito têm maior expressão em *D. mojavensis* do que em *D. arizonae* e 12 maior expressão em *D. arizonae* que *D. mojavensis*, enquanto que quatro elementos apresentam maior expressão no híbrido A do que no B e cinco em B do que em A (Tabela S2).

As Figuras 3 e 4 apresentam os FPKM dos NLTRs e Transposons de DNA diferencialmente expressos. Dentre os Transposons de DNA destacam-se os elementos da família *Homo*, em particular *Homo 7*, com maior expressão no híbrido A em relação aos parentais e híbrido B. Dentre os LTRs destacam-se dois elementos: o retrotransposon *Copia 1*, com maior expressão no híbrido A em relação aos parentais e híbrido B e o retrotransposon *Gtwin*, como o elemento com expressão mais elevada dentre todos os transcritos (FPKM no híbrido B: 980,84). Na Figura 5 estão representados os FPKMs desses três elementos para melhor ilustração da expressão diferencial.

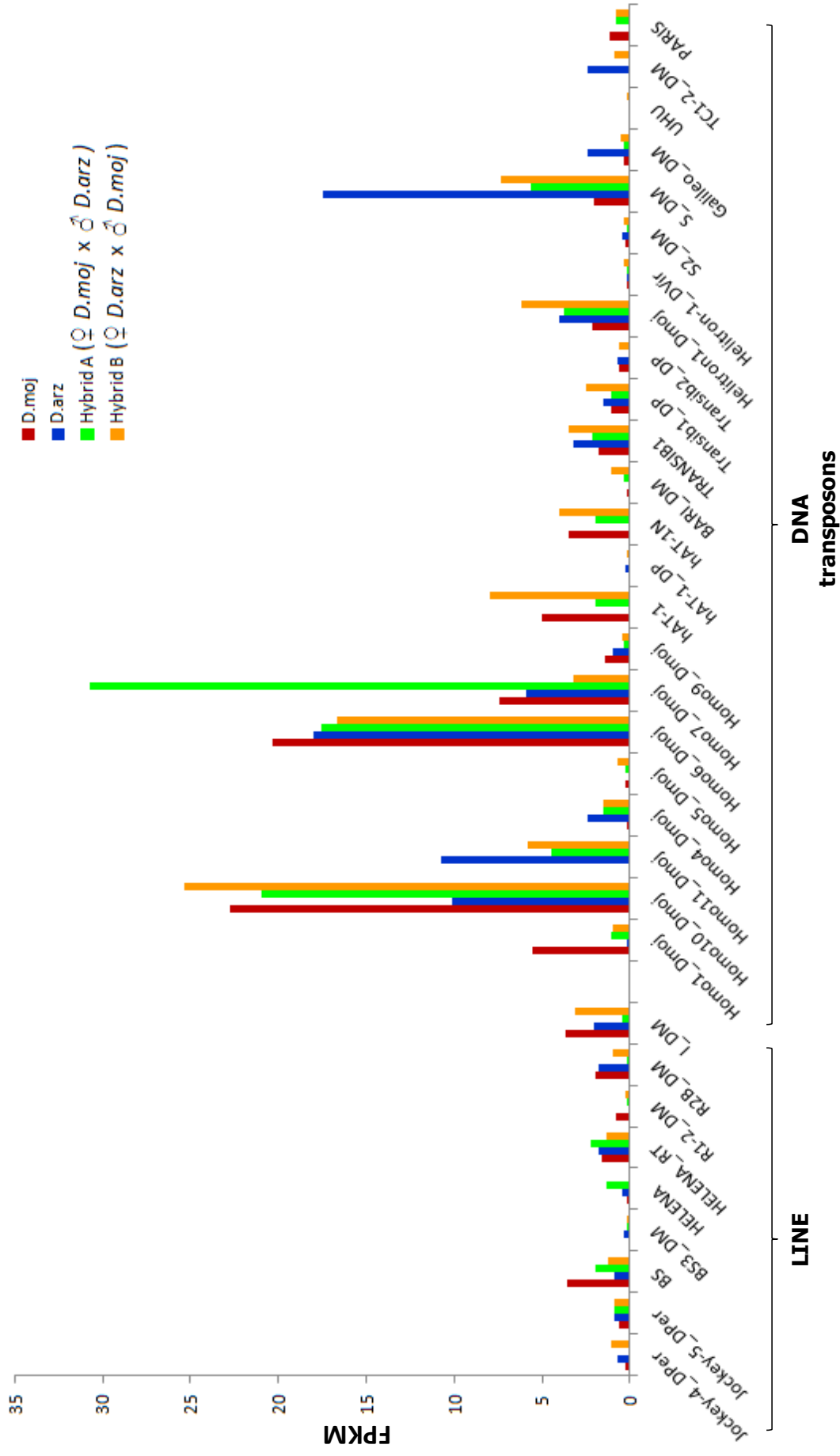


Figura 3. Retrotransposons NLTRs e transposons de DNA, identificados com expressão diferencial do mobiloma de ovários de *D. mojavensis* e *D. arizonae* e híbridos por análises de RNA-Seq ($\geq 60\%$ de identidade com TEs do RepBase).

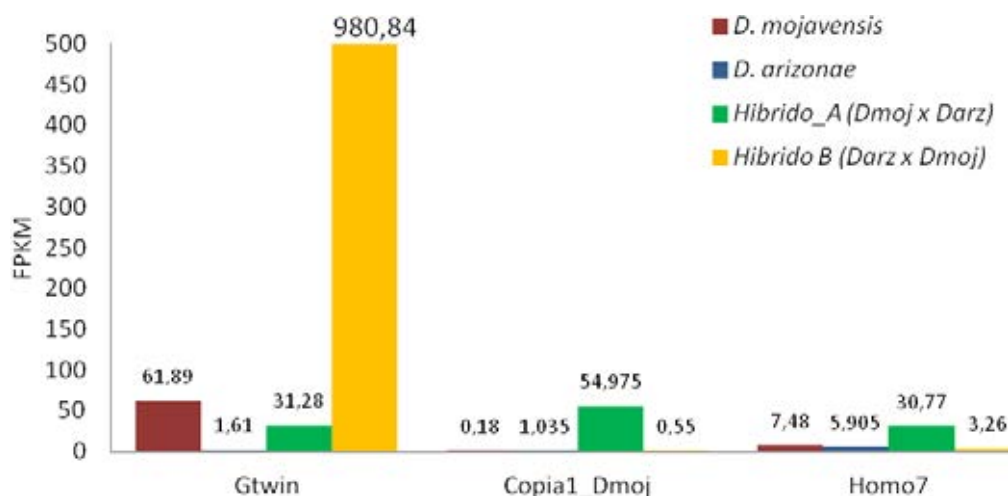


Figura 5. Representação dos elementos *Gtwin*, *Copia1* e *Homo7*, com superexpressão entre genomas de *D. mojavensis*, *D. arizonae* e híbridos, identificados por análises de dados do RNA-Seq. Números acima das colunas correspondem aos valores de FPKM da ordenada.

Padrão de herança da expressão dos TEs nos híbridos

Verificou-se as porcentagens dos diferentes padrões de herança dos TEs nos híbridos (conservativa, aditiva, sub ou superdominância, e dominância do padrão de expressão de TEs de *D. mojavensis* ou dominância de *D. arizonae* (ver Figura 1). O objetivo dessa análise foi conhecer qual mobiloma participa mais na herança de expressão dos TEs pelos híbridos, uma vez que os transcritos da linhagem maternal e paterna contribuem independentemente com o nível total de TEs nos híbridos. Pode-se notar nos híbridos das duas direções de cruzamentos que há uma predominância de TEs que têm expressão conservada como nos transcriptomas parentais (54 a 56%). Outra parcela relevante, e semelhante nos híbridos A e B, é constituída pelos elementos com herança aditiva (20% e 13%, respectivamente), e com dominância da expressão dos TEs de *D. arizonae* na expressão global (13% de *D. arizonae* no híbrido A, contra 6% de *D. mojavensis* e 21% no híbrido B contra 6% de *D. mojavensis*). A maior proporção de herança de expressão dos TEs de *D. arizonae* no híbrido B talvez

esteja relacionada ao fato de estes serem descendentes de fêmeas de *D. arizonae* (Figura 6).

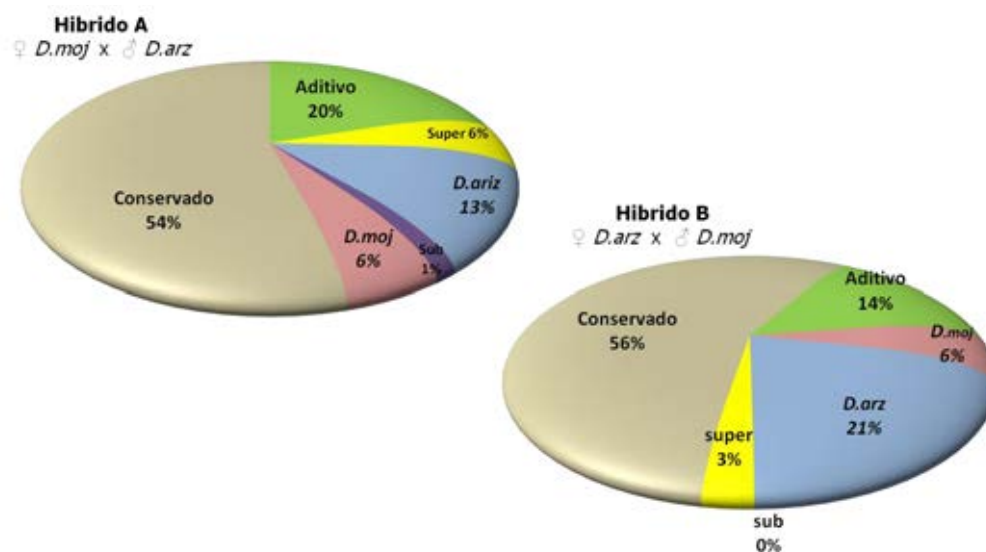


Figura 6. Proporções dos padrões de herança do nível de expressão de TEs em híbridos F₁ entre *D. mojavensis* e *D. arizonae*. O gráfico mostra a frequência de TEs em cada categoria de herança nos híbridos dos dois sentidos de cruzamentos (conservativa, aditiva, sub- super dominância, dominância de *D. mojavensis* e dominância *D. arizonae*. A: híbridos de fêmeas *D. mojavensis* x machos *D. arizonae*; B: fêmeas *D. arizonae* x machos *D. mojavensis*.

Na Figura 7 exemplificam-se os diferentes modos de herança utilizando os valores de FPKM para alguns elementos expressos diferencialmente em *D. mojavensis*, *D. arizonae* e híbridos.

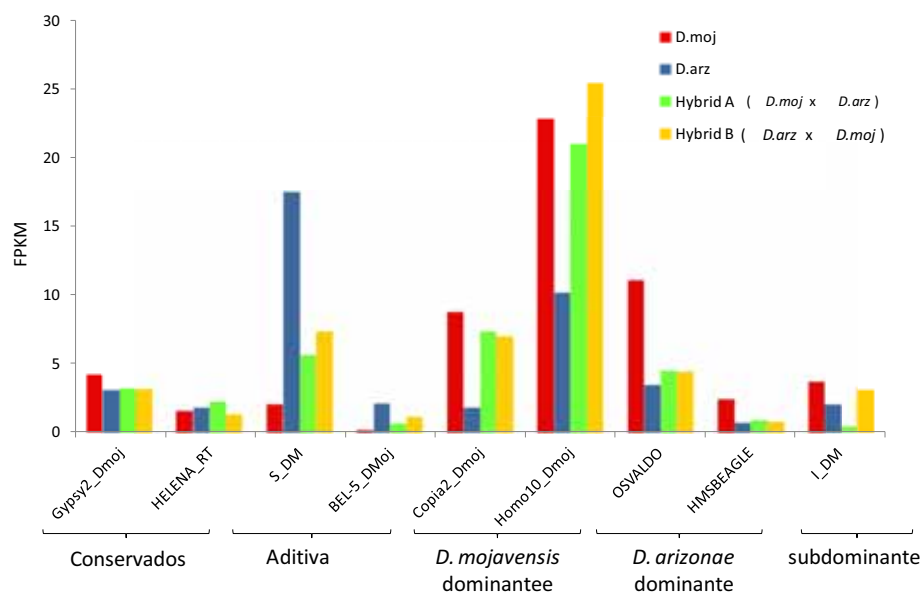


Figura 7. Representação dos elementos diferencialmente expressos e tipo de herança entre os transcriptomas de *D. mojavensis*, *D. arizonae* e híbridos.

Expressão por RT-qPCR

A superexpressão dos LTRs *Copia-1* e *Gtwin*, quantificada nas análises *in silico*, foi validada por RT-qPCR. O retroelemento *Copia-1* revelou-se praticamente sem expressão nos parentais, porém é expresso no híbrido A (fêmeas *D. arizonae* x machos *D. mojavenensis*), corroborando os dados por RNA-Seq (Figura 8).

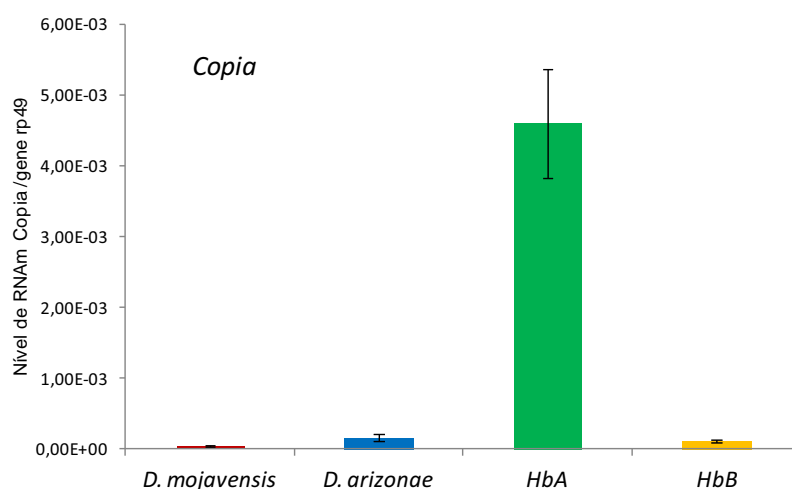


Figura 8. Expressão por RT-qPCR do retrotransposon *Copia1* das espécies *D. mojavenensis* e *D. arizonae* e híbridos. HbA: fêmeas *D. mojavenensis* x machos *D. arizonae*; HbB: cruzamento recíproco.

Ao contrário de *Copia1*, o retroelemento *Gtwin* é superexpresso no híbrido B (fêmeas *D. arizonae* x machos *D. mojavenensis*), e apresenta baixo nível de expressão em *D. mojavenensis* e híbrido A (Figura 9), também validando os resultados do RNA-Seq.

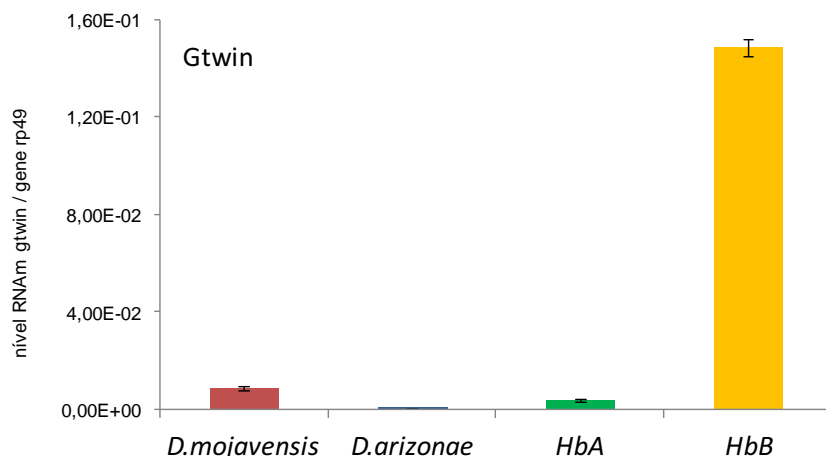


Figura 9. Expressão por RT-qPCR do retrotransposon *Gtwin* para espécies *D. mojavensis* e *D. arizonae* e híbridos. HbA: fêmeas *D. mojavensis* x machos *D. arizonae*; HbB: cruzamento recíproco.

O retrotransposon *Gtwin* apresentou grande quantidade de transcritos no híbrido do cruzamento B (fêmeas *D. arizonae* e machos *D. mojavensis*) e também, porém baixa, expressão na linhagem materna de *D. arizonae*. A sequência identificada desse elemento apresentou expressão 18 vezes maior que a do elemento *Copia* (segundo TE mais expresso), representando 1% de todos os transcritos do transcriptoma (incluindo genes e TEs). A estrutura de *Gtwin*, um retrotransposon de 6.421pb em *D. melanogaster*, é próxima a do retrovírus endógeno *Gypsy*, descrito inicialmente *in silico* (BOWEN; MCDONALD, 2001) e isolado de uma linhagem específica de *D. melanogaster* (G32), na qual acumula-se em cromossomos aberrantes e está associado a pontos de quebra de inversões. Suas cópias estão predominantemente localizadas em regiões de eucromatina, mas pelo menos três estão situadas na heterocromatina (KOTNOVA et al., 2005). Subfamílias mutantes também foram identificadas, com estruturas distintas, apresentando sequências circulares e contendo sítios polimórficos nas sequências alvos de tRNA (SALENKO et al., 2011). A sequência do elemento *Gtwin* identificada como superexpressa nos híbridos B pode

ter relevância evolutiva, uma vez que os híbridos dessa direção de cruzamento (fêmeas *D. arizonae* e machos *D. mojavensis*) exibem isolamento pós-zigótico, gerando machos F₁, híbridos estéreis. Uma análise mais pormenorizada da dinâmica de *Gtwin* em híbridos das duas espécies pode revelar se este retrotransposon está associado à esterilidade nessa direção de cruzamento.

Ao contrário de *Gtwin*, a expressão relativa de *Copia 1* indica uma possível reativação deste elemento nos híbridos do cruzamento A (fêmeas de *D. mojavensis* e machos *D. arizonae*), pois este retrotransposon tem expressão praticamente nula nos parentais. O elemento *Copia1* completo de *D. mojavensis* depositado no Repbase apresenta 4.030 pb e LTRs 100% idênticas, indicando uma origem recente dessa cópia. Elementos *Copia* apresentam alta heterogeneidade em eucariotos e já foram identificados e são ativos em muitas espécies de *Drosophila* (JORDAN; MCDONALD, 1998; JORDAN et al., 1999), estando associados a diversos eventos de transferência horizontal entre espécies de *Drosophila* e entre essas e espécies do gênero *Zaprionus* (DE ALMEIDA; CARARETO 2006; SETTA et al., 2011).

A partir desses resultados podemos concluir que a taxa de expressão de determinados TEs é alterada em diferentes magnitudes nos híbridos de *D. mojavensis* e *D. arizonae*, dentro e entre classes e famílias de TEs. Apesar de serem resultados preliminares, produzimos um cenário global do mobiloma diferencialmente expresso em cruzamentos interespecíficos. Ao contrário do relatado na literatura, particularmente não detectamos explosão global de expressão de TEs em híbridos, mas sim superexpressão de TEs particulares. Análises mais detalhadas desses dados serão realizadas com diferentes critérios de busca que poderão permitir a identificação de novos TEs superexpressos. Adicionalmente, está previsto o sequenciamento de

pequenos RNAs (small RNAs) para que possamos comparar as expressões diferenciais de TEs relatadas neste trabalho, com o controle destes elementos.

Em síntese, este é um trabalho inédito de análise global da atividade do mobiloma expresso em cruzamentos interespecíficos, que abre perspectivas para a compreensão dos efeitos da expressão dos TEs não apenas em híbridos de *D. mojavensis* e *D. arizonae* como em outras espécies de *Drosophila*.

Referências

- ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. **Genome Biology**, London, v. 11, n. 10, p. R106, 2010. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20979621> >. Acesso em: 13 fev. 2014.
- BIÉMONT, C. A brief history of the status of transposable elements: from junk DNA to major players in evolution. **Genetics**, Austin, v. 186, n. 4, p. 1085-93, 2010.
- 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**, Oxford, v. 22, n. 7, p. 1387-95, 2009.
- BOWEN, N. J.; MCDONALD, J. F. *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. **Genome Research**, Cold Spring Harbor, v. 11, n. 9, p. 1527-40, 2001.
- CLARK, A. G. et al. Evolution of genes and genomes on the *Drosophila* phylogeny. **Nature**, London, v. 450, n. 7167, p. 203-218, 2007.
- DE ALMEIDA, L. M.; CARARETO, C. M. Sequence heterogeneity and phylogenetic relationships between the copia retrotransposon in *Drosophila* species of the repleta and melanogaster groups. **Genetics Selection Evolution**, London, v. 38, n. 5, p. 535-50, 2006.
- DE SETTA, N. et al. Copia retrotransposon in the *Zaprionus* genus: another case of transposable element sharing with the *Drosophila melanogaster* subgroup. **Journal of Molecular Evolution**, v. 72, n. 3, p. 326-38, 2011. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21347850> >. Acesso em: 31 jan. 2014.
- FONTDEVILA, A. Hybrid genome evolution by transposition. **Cytogenetic and Genome Research**, Basel, v. 110, n. 1-4, p. 49-55, 2005.
- GRABHERR, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nat Biotechnol**, New York, v. 29, n. 7, p. 644-52, 2011.

GRANZOTTO, A. et al. The evolutionary dynamics of the Helena retrotransposon revealed by sequenced *Drosophila* genomes. **BMC EVOLUTIONARY BIOLOGY**, London, v. 9, 2009.

GRAVELEY, B. R. et al. The developmental transcriptome of *Drosophila melanogaster*. **Nature**, London, v. 471, n. 7339, p. 473-9, 2011.

JORDAN, I. K.; MCDONALD, J. F. Evolution of the copia retrotransposon in the *Drosophila melanogaster* species subgroup. **Molecular Biology and Evolution**, Chicago, v. 15, n. 9, p. 1160-71, 1998.

JORDAN, I. K.; MATYUNINA, L. V.; MCDONALD, J. F. Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 96, n. 15, p. 12621-12625, 1999.

JURKA, J. et al. Repbase update, a database of eukaryotic repetitive elements. **Cytogenetic and Genome Research**, Basel, v. 110, n. 1-4, p. 462-467, 2005.

KELLEHER, E. S.; EDELMAN, N. B.; BARBASH, D. A. *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. **PLoS Biol**, Cambridge, v. 10, n. 11, p. e1001428, 2012. Disponível em: <<http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1001428>>. Acesso em: 18 dez. 2013.

KIDWELL, M. G.; LISCH, D. R. Perspective: Transposable elements, parasitic DNA, and genome evolution. **Evolution**, Lancaster, v. 55, n. 1, p. 1-24, 2001.

KOTNOVA, A. P. et al. Retrotransposon gtwin: structural analysis and distribution in *Drosophila melanogaster* strains. **Genetika**, v. 41, n. 1, p. 23-9, 2005. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/15771247>>. Acesso em: 30 jan. 2014.

LABRADOR, M. et al. Interspecific hybridization increases transposition rates of Osvaldo. **Molecular Biology and Evolution**, Chicago, v. 16, n. 7, p. 931-937, 1999.

LABRADOR, M.; FONTDEVILA, A. High transposition rates of Osvaldo, a new *Drosophila buzzatii* retrotransposon. **Molecular & General Genetics**, Berlin, v. 245, n. 6, p. 661-674, 1994.

LANGMEAD, B. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome Biology**, London, v. 10, n. 3, p. R25, 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19261174>>. Acesso em: 13 Fev. 2014.

LERAT, E. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. **Heredity**, London, v. 104, n. 6, p. 520-533, 2010.

LI, B.; DEWEY, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. **Bmc Bioinformatics**, London, v. 12, 2011. Disponível em: <<Go to ISI>://WOS:000294361700001 >. Acesso em: 22 jul. 2012.

MCMANUS, C. J. et al. Regulatory divergence in *Drosophila* revealed by mRNA-seq. **Genome Research**, Cold Spring Harbor, v. 20, n. 6, p. 816-25, 2010.

PANTAZIDIS, A.; LABRADOR, M. The retrotransposon Osvaldo from *Drosophila buzzatii* displays all structural features of a functional retrovirus. **Molecular Biology and Evolution**, Chicago, v. 16, n. 7, p. 909-921, 1999.

REBOLLO, R. et al. Jumping genes and epigenetics: towards new species. **Gene**, Amsterdam, v. 454, n. 1/2, p. 1-7, 2010.

REED, L. K.; LAFLAMME, B. A.; MARKOW, T. A. Genetic architecture of hybrid male sterility in *Drosophila*: analysis of intraspecies variation for interspecies isolation. **PloS one**, San Francisco, v. 3, n. 8, p. e3076-e3076, 2008. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2517651&tool=pmcentre&rendertype=abstract>>. Acesso em: 22 jan. 2012.

SALENKO, V. B. et al. A rare family of gtwint retrotransposon carrying a mutation in the tRNA-primer binding site is amplified in G-32 *Drosophila melanogaster* strain. **Doklady Biochemistry And Biophysics**, New York, v. 436, p. 16-9, 2011.

TAMURA, K.; SUBRAMANIAN, S.; KUMAR, S. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. **Molecular Biology and Evolution**, Chicago, v. 21, n. 1, p. 36-44, 2004.

TRAPNELL, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nature Biotechnology**, London, v. 28, n. 5, p. 511-5, 2010.

UNGERER, M. C.; STRAKOSH, S. C.; STIMPSON, K. M. Proliferation of Ty3/gypsy-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. **BMC Biology**, London, v. 7, p. 40-40, 2009. Disponível em: <<http://www.biomedcentral.com/1741-7007/7/40>>. Acesso em: 25 jan. 2012.

VIEIRA, C. et al. A comparative analysis of the amounts and dynamics of transposable elements in natural populations of *Drosophila melanogaster* and *Drosophila simulans*. **Journal of Environmental Radioactivity**, Oxford, v. 113, p. 83-6, 2012.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, London, v. 10, n. 1, p. 57-63, 2009.

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

Material Suplementar

Tabela S1. Matriz do número de “reads” normalizados (FPKM) mapeados em diferentes TEs de *D.mojavensis* (Moj), *D. arizonae* (Arz) e híbridos A e B, média das duas réplicas sequenciadas (rep1 e rep2). Elementos recuperados do banco RepBase por alinhamento contra o genoma de *D.mojavensis*, com no mínimo 900pb ou cobertura de 60% com a sequência referência.

	TEs	<i>D. mojavensis</i>	<i>D. arizonae</i>	Híbrido A	Híbrido B
1	BEL1_Dmoj	0,13	4,55	1,02	2,37
2	BEL2_Dmoj	0,37	4,95	2,87	3,19
3	BEL3_DM	3,42	3,08	1,48	1,96
4	BEL-3_DMoj	2,84	1,07	0,70	0,95
5	BEL-3_DWil	8,34	0,04	2,06	1,43
6	BEL-5_DMoj	0,16	2,11	0,71	1,19
7	BEL-6_DMoj	0,08	2,67	2,31	1,38
8	BEL-7_DMoj	0,27	0,95	0,00	0,00
9	<i>Copia1</i> _Dmoj	0,18	1,04	54,98	0,55
10	<i>Copia2</i> _Dmoj	8,71	1,84	7,36	7,05
11	<i>Copia-3</i> _DMoj	0,64	17,25	2,67	1,35
12	Galileo_DM	0,40	2,46	0,34	0,51
13	Gypsy1_Dmoj	2,28	9,33	4,25	2,09
14	Gypsy-22_DWil	0,06	1,42	0,62	0,64
15	Gypsy2_Dmoj	4,24	3,11	3,24	3,15
16	Gypsy4_Dmoj	0,00	1,91	1,21	0,89
17	Gypsy5_Dmoj	0,37	2,28	0,18	0,43
18	Gypsy6_Dmoj	1,13	6,32	3,07	2,28
19	Gypsy7_Dmoj	0,20	2,19	5,29	1,87
20	Gypsy-8_DMoj	0,67	1,68	2,93	1,47
21	Helitron1_Dmoj	2,14	4,03	3,73	6,16
22	Homo1_Dmoj	5,57	0,14	1,06	1,01
23	Homo10_Dmoj	22,80	10,10	20,95	25,35
24	Homo11_Dmoj	0,06	10,77	4,45	5,81
25	Homo4_Dmoj	0,22	2,39	1,55	1,50
26	Homo5_Dmoj	0,27	0,00	0,25	0,76
27	Homo6_Dmoj	20,31	17,99	17,59	16,63
28	Homo7_Dmoj	7,48	5,91	30,77	3,26
29	Homo9_Dmoj	1,43	1,00	0,40	0,47
30	Jockey-4_DPer	0,31	0,73	0,03	1,06
31	Jockey-5_DPer	0,59	0,87	0,92	0,87
32	BARI_DM	0,20	0,00	0,37	1,06
33	BS	3,54	0,89	1,95	1,22
34	BS3_DM	0,00	0,37	0,20	0,20
35	BURDOCK_I	0,41	0,03	0,25	1,34
36	<i>Copia2</i> _I	0,43	6,18	1,34	3,39
37	<i>Copia</i> _I	0,19	1,17	0,66	1,28
38	DM1731_I	0,64	17,25	2,67	1,35

Continuação...

TEs	<i>D. mojavensis</i>	<i>D. arizonae</i>	Híbrido A	Híbrido B	
39	DM176_I	0,14	1,12	2,66	1,15
40	FROGGER_I	1,77	0,07	2,82	0,08
41	GTWIN_I	61,89	1,61	31,28	980,84
42	Gypsy4_I	2,37	0,12	0,74	1,26
43	hAT-1	5,03	0,00	2,00	8,00
44	hAT-1_DP	0,05	0,31	0,05	0,19
45	hAT-1N	3,50	0,00	2,00	4,00
46	HELENA	0,16	0,42	1,34	0,00
47	HELENA_RT	1,59	1,83	2,26	1,36
48	Helitron-1_DVir	0,19	0,21	0,20	0,37
49	HMSBEAGLE_I	2,44	0,76	0,90	0,82
50	I_DM	3,71	2,10	0,44	3,13
51	Invader2_I	3,01	0,67	1,83	0,78
52	Invader6_I	1,70	6,09	1,99	5,95
53	MAX_I	0,38	2,20	1,22	1,29
54	MICROPIA_I	0,10	3,64	0,23	0,82
55	OSVALDO_I	11,01	3,48	4,52	4,45
56	PARIS	1,12	0,13	0,81	0,81
57	R1-2_DM	0,79	0,00	0,19	0,28
58	R2B_DM	1,98	1,80	0,21	0,95
59	ROOA_I	2,04	0,00	0,20	0,59
60	ROO_I	0,29	0,98	0,38	1,03
61	S2_DM	0,23	0,46	0,22	0,40
62	S_DM	2,09	17,44	5,67	7,36
63	STALKER4_I	5,36	1,60	1,65	0,99
64	TABOR_DA-I	6,33	0,71	1,59	4,53
65	TC1-2_DM	0,00	2,40	0,00	0,90
66	TRANSIB1	1,80	3,23	2,18	3,46
67	Transib1_DP	1,04	1,49	1,08	2,51
68	Transib2_DP	0,60	0,75	0,00	0,60
69	TV1I	0,00	0,33	0,00	0,16
70	UHU	0,00	0,00	0,00	0,16

Tabela S2. Elementos de transposição diferencialmente expressos em *D. mojavensis* e *D. arizonae* e em híbrido na condição A e condição B (p-value < 0,05; count = número de reads mapeados). Blast >60% sequência referência.

Maior em expressão em <i>D. mojavensis</i> em relação <i>D. arizonae</i> (Dmoj > Arz)				
TE	<i>D. arz</i> x count	<i>D. moj</i> x count	log2FoldChange	pvalue
DM1731_I	86	2501	4,84	6,84E-128
Invader6_I	17	278	4,01	1,54E-35
S_DM	31	289	3,2	5,61E-28
MAX_I	37	218	2,55	4,32E-17
MICROPIA_I	1	42	4,79	1,07E-08
Transib1_DP	22	75	1,76	6,54E-05
DM176_I	2	21	3,42	0,000634
TC1-2_DM	0	12	Inf	0,00113
Maior em expressão em <i>D. arizonae</i> em relação <i>D. mojavensis</i> (Arz > Moj)				
GTWIN_I	11153	323	5,11	1,3E-175
FROGGER_I	170	7	4,44	2,22E-27
Invader2_I	99	10	3,27	1,81E-10
TABOR_DA-I	81	10	2,97	5,32E-10
HMSBEAGLE_I	209	69	1,58	2,25E-08
STALKER4_I	46	4	3,28	2,65E-06
BS	39	4	3,22	6,77E-06
OSVALDO_I	81	24	1,74	2,73E-05
R2B_DM	32	5	2,48	0,000585
Helitron-1_DVir	10	0	4,31	0,0124
Gypsy4_I	9	0	4,08	0,027
PARIS	14	2	2,3	0,0471
Maior expressão no híbrido A em relação híbrido B (A>B)				
FROGGER_I	332	8	5,25	3,2E-69
DM1731_I	382	188	1,02	3,29E-08
Invader2_I	48	20	1,24	0,0314
DM176_I	40	15	1,34	0,038
Maior expressão no híbrido B em relação híbrido A (B>A)				
GTWIN_I	6000	188028	4,97	0
Transib1_DP	26	172	2,69	1,68E-16
<i>Copia2_I</i>	148	535	1,85	8,71E-16
Invader6_I	55	171	1,63	2,3E-09
TABOR_DA-I	26	66	1,32	0,00469

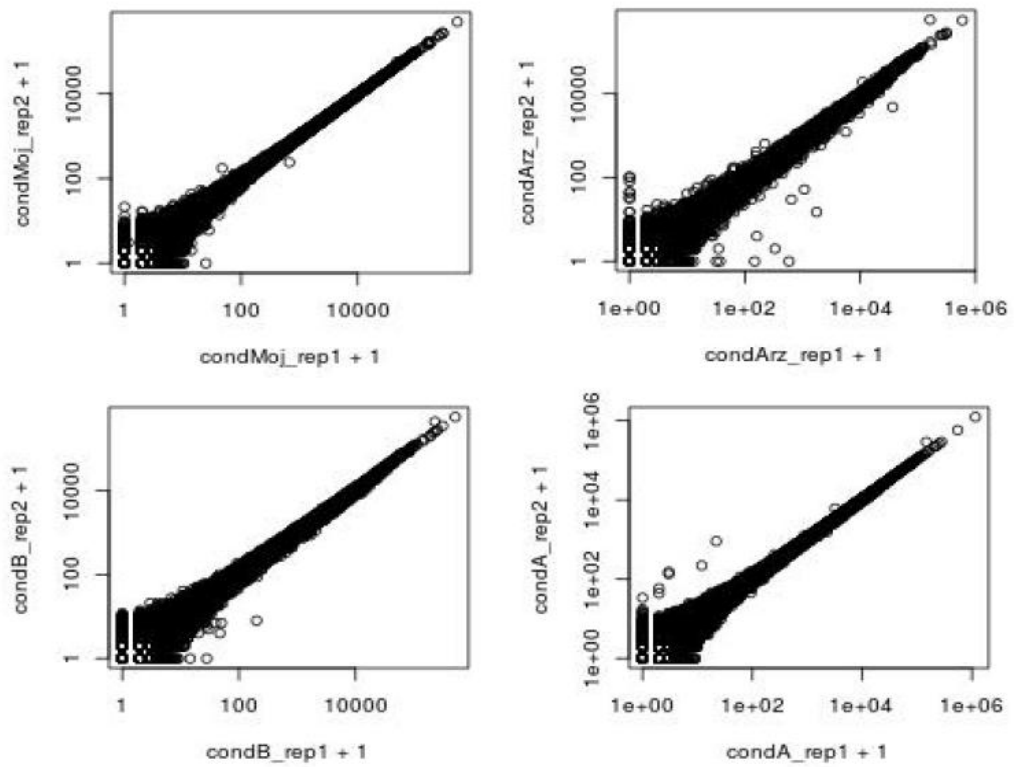


Figura S1. Relação do número de *reads* alinhados nos *contigs* (*xcount*) entre as réplicas das espécies de *D. mojavensis* e *D. arizonae* e híbridos (condição A e B).

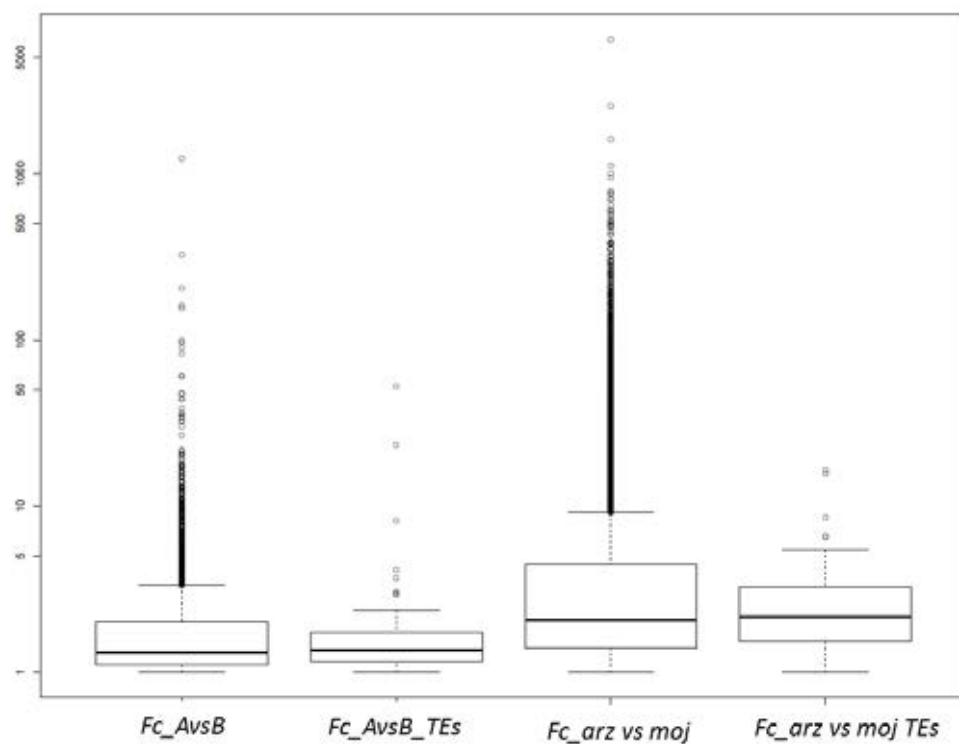


Figura S2. Diagrama de caixa da distribuição dos valores de FoldChange (FC) do transcriptoma total e dos elementos de transposição (TEs), entre as condições parentais (*D. arizonae* vs *D. mojavensis*) e híbridos (A e B), Linha em negrito no interior das caixas representa mediana e os pontos externos *outliers*.

CAPÍTULO 3

Expressão de elementos de transposição em *Drosophila mojavensis*, *D. arizonae* e seus híbridos, quantificada por RT-qPCR.

Introdução

Os elementos de transposição em *Drosophila* se transpõem em uma taxa variável e que depende do número de cópias no genoma. Famílias que transpõem frequentemente são normalmente jovens e não estão fixadas na espécie. Famílias com baixa frequência de transposição são geralmente antigas fixadas em diferentes populações (GONZÁLEZ; PETROV, 2012). Ambas as atividades, transcricional e transposicional dos TEs, são reguladas por uma variedade de mecanismos que alteram sua atividade em diferentes espécies e linhagens (FABLET; VIEIRA, 2011).

O perfil da variação da atividade dos TEs está bem documentado em plantas, como por exemplo, em cana-de-açúcar, cujos cultivares modernos são híbridos interspecíficos entre *Saccharum officinarum* e *S. spontaneum* retrocruzados com *S. officinarum*, altamente poliploides e com contribuição desigual do genoma ancestral. Domingues et al. (2012) identificaram nesses cultivares 60 retrotransposons LTRs, classificados em 35 famílias, das quais quatro eram linhagens *Copia* e três *Gypsy*. Análises por RT-PCR nesta espécie revelaram variação de expressão da mesma família e perfil distinto de transcritos e de pequenos RNA (sRNA) entre tecidos, sugerindo que estes são diferencialmente expressos e regulados.

Em populações de *Drosophila simulans* e *D. melanogaster*, análises de expressão e hibridização *in situ* do retrotransposon 412 não foi detectada em ovários de ambas as espécies; contudo transcritos foram detectados nos espermatócitos primários dos machos. No

entanto, em tecidos somáticos, a expressão foi altamente variável (BORIE et al., 2002). O nível de expressão do retrotransposon *tirant* também pode variar em ovários de linhagens derivadas de populações naturais de *D. simulans* (AKKOUICHE et al., 2012). Em uma das linhagens estudadas foram encontradas cópias de elementos LTR na eucromatina, mas não foram detectados transcritos em ovários e, em outra, considerada linhagem permissiva, foi detectada expressão, e algumas cópias de *tirant* foram capazes de completa retrotransposição. Foi identificada também uma terceira linhagem com atividade intermediária, das quais os TEs são transcritos, mas não traduzidos.

O NLTR *Helena* apresenta 62 cópias em *D. simulans*, mas com baixo nível transcricional em linhagens derivadas de diferentes populações, já que poucos transcritos foram detectados por RT-PCR. Já em *D. melanogaster* o elemento *Helena* apresenta-se em fase de extinção, com apenas cópias deletadas, e nenhuma evidência de expressão (REBOLLO et al., 2008). Por outro lado, níveis de expressão diferencial entre linhagens foram observados em *D. mojavensis*, espécie em que *Helena* está em diferentes estágios de evolução, com número de cópias variável, grande parte delas truncadas na extremidade 5' e com muitas deleções internas e inserções (GRANZOTTO et al., 2009).

O retrotransposon LTR *Oswaldo* tem sido estudado em duas espécies do *cluster buzzatii* do grupo *repleta*, *D. buzzatii* e *D. koepferae*. Em face de suas altas taxas de transposição pode-se inferir que apresentam altas taxas de expressão em linhagens que foram coletadas recentemente de populações naturais, as quais também foram elevadas em híbridos entre as duas espécies (LABRADOR; FONTDEVILA, 1994; LABRADOR et al., 1999; PANTAZIDIS; LABRADOR, 1999). As inserções cromossômicas de *Oswaldo* têm distribuição diferencial em linhagens de *D. buzzatii*, em alta frequência em populações específicas e em alguns sítios cromossômicos, sugerindo um importante papel deste TE no processo de colonização dessa espécie (GUERREIRO; FONTDEVILA, 2001).

Foi demonstrado em *D. melanogaster* que *Copia* interage com sequências reguladoras do genoma hospedeiro e com fatores celulares sexo-específicos (JORDAN, MCDONALD, 1998; MOROZOVA et al., 2004; VU, NUZHIDIN, 2011). Foram detectadas altas taxas de expressão de *Copia* em espermatozoides primários e também de transposição em linhagens germinativas de machos e fêmeas e de indivíduos inteiros de *D. melanogaster*, sendo que em uma sublinhagem (2b3) a taxa de transposição nos machos foi até 50 vezes mais elevada que em fêmeas (PASYUKOVA et al., 1997). O elemento *Copia* também tem sido encontrado em outras espécies de *Drosophila*, particularmente em espécies do grupo *repleta*. De Almeida e Carareto (2006) relataram a ocorrência desse elemento em 18 de 24 espécies estudadas, pertencentes a três subgrupos do grupo *repleta*, mas análises de expressão não foram realizadas.

Outro TE que pode afetar a adaptabilidade das espécies hospedeiras é o retrotransposon não-LTR *I*, conhecido por causar letalidade do embrião em cruzamentos entre determinadas linhagens de *D. melanogaster*, devido a catástrofes meióticas e rearranjos cromossômicos. Essas desordens na linhagem germinativa são decorrentes da presença de cópias ativas em machos de uma linhagem (Indutora) cruzados com fêmeas com elementos não funcionais (Reativas), resultando em alta taxa de transposição na linhagem germinal das fêmeas descendentes, chamadas fêmeas SF (SELEME et al., 2005; SELEME et al., 2006).

Podemos notar das descrições acima que a hibridação intra e interespecífica pode contribuir para a transposição de TEs e instabilidade genômica. Apesar dos mecanismos que desencadeiam a transposição em híbridos não serem completamente compreendidos, é consenso que quando dois genomas diferentes se combinam para formar o zigoto, esse deve responder a uma grande mudança nos mecanismos de regulação dos TEs devido a este “choque genômico” (MCCLINTOCK, 1984).

Neste capítulo descrevemos resultados de análises de expressão por RT-qPCR, dos quatro elementos de transposição acima descritos, em linhagens de *D. mojavensis*, *D. arizonae* e seus híbridos. O objetivo foi obter o perfil de expressão destes elementos em condições de hibridação e em diferentes tecidos destas espécies. Estes quatro elementos foram selecionados do genoma sequenciado de *D. mojavensis* a partir de dados de uma anotação prévia de TEs realizada em nosso laboratório (dados não publicados). Dois critérios foram utilizados para a seleção dos TEs: i) elementos que apresentassem cópias putativamente completas e ativas, comparadas com aquelas depositadas na biblioteca de TEs do gênero *Drosophila* do RepBase (<http://www.girinst.org/replib/reports/index.html>) e ii) elementos com baixo número de cópias, para facilitar o monitoramento e quantificação da expressão por RT-qPCR. Os elementos que atenderam a esses critérios foram os retrotransposons com LTRs *Copia* e *Osvaldo* e os NLTRs *Helena* e *I*.

Material e Métodos

Amostras e cruzamentos

Foram utilizadas nesses estudos linhagens obtidas da coleção da UC San Diego *Drosophila* Stock Center (San Diego, CA, USA), sendo uma amostra de *D. arizonae* e de *D. mojavensis* derivadas de populações que vivem em alopatria (Metztitlán, Hidalgo, México - stock number: 15081-1271.17 e Anza Borrego Desert, CA, EUA - 15081-1352.01), e outras de populações simpátricas (*D. arizonae* Las Bocas, Sonora-México 15081-1271.18 e *D. mojavensis* Agiabampo Bay, Sonora-México 15081-132.26). Essa distribuição está ilustrada na Figura 1.

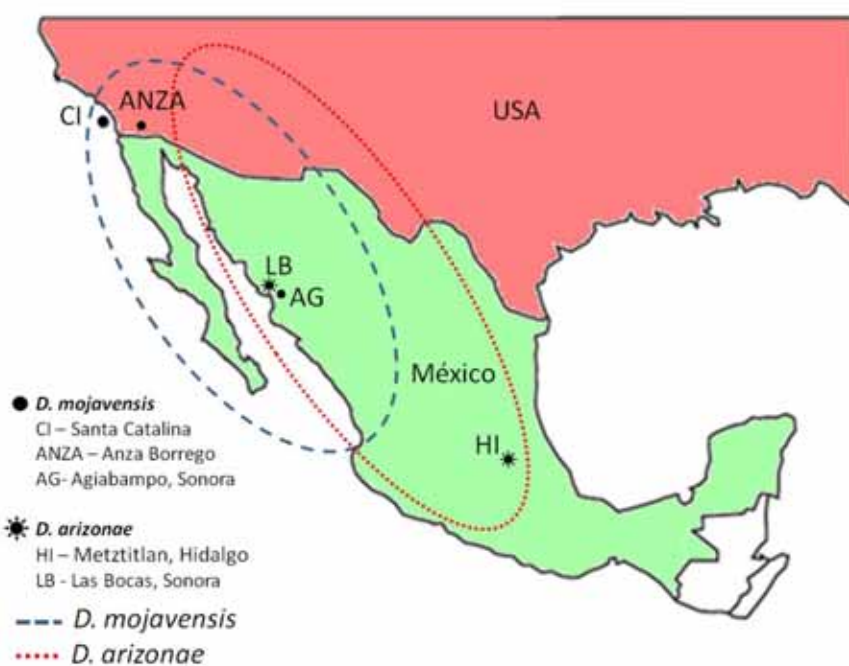


Figura 1. Esquema da distribuição das linhagens *D. mojavensis* e *D. arizonae* nas amostras utilizadas.

As amostras de RNA das linhagens parentais e dos híbridos foram obtidas a partir de cruzamentos em massa. Os sentidos dos cruzamentos interespecíficos entre cada linhagem estão descritos na Tabela 1. As moscas foram mantidas em frascos (9cmx13cm) e em tubos (2,5x10cm) de vidro, em câmara fechada com temperatura e iluminação controladas (25° C), luz artificial durante o dia (12h) e mantidas em meio de cultura contendo ágar, extrato de levedura, meio de cactos e fungicida (Material Suplementar).

Tabela 1. Representação dos sentidos dos cruzamentos interespecíficos entre *D. mojavensis* e *D. arizonae*.

Fêmeas	Linhagem	Machos	Linhagem	Cruzamento	Distribuição
<i>D. mojavensis</i>	ANZA	x <i>D. arizonae</i>	HI	A	Alopátrica
<i>D. arizonae</i>	HI	x <i>D. mojavensis</i>	ANZA	B	Alopátrica
<i>D. mojavensis</i>	ANZA	x <i>D. arizonae</i>	LB	C	Alopátrica
<i>D. arizonae</i>	LB	x <i>D. mojavensis</i>	ANZA	D	Alopátrica
<i>D. mojavensis</i>	AG	x <i>D. arizonae</i>	HI	E*	Alopátrica
<i>D. arizonae</i>	HI	x <i>D. mojavensis</i>	AG	F	Alopátrica
<i>D. mojavensis</i>	AG	x <i>D. arizonae</i>	LB	G	Alopátrica
<i>D. arizonae</i>	LB	x <i>D. mojavensis</i>	AG	H	Simpátrica

*ausência de emergência

O RNA foi extraído de indivíduos com 10 dias de idade, tanto de tecidos somáticos, quanto germinativos, nos parentais (carcaças, ovários e testículos), e nos híbridos de indivíduos inteiros. As moscas foram dissecadas e seus tecidos conservados em tubos contendo 1x PBS (Phosphate buffer saline) sob o gelo quando a extração era realizada logo em seguida de serem dissecados ou armazenados a -80°C se feitas em outro momento. O RNA foi extraído de um *pool* de 20 pares de ovários e aproximadamente 30 testículos dos parentais, sendo utilizadas as carcaças desses indivíduos. A extração de RNA dos híbridos se deu a partir de indivíduos inteiros de um *pool* de aproximadamente 10 indivíduos por amostras. Para cada tecido, três réplicas foram utilizadas para extração de RNA com auxílio do kit RNeasy (Qiagen) e tratamento com DNase (DNA-free kit; Ambion). O RNA total foi convertido em cDNA usando enzimas do kit “Thermoscript Invitrogen kit primed”.

A reação em cadeia de polimerase (*Real-time polymerase chain reaction*) combinada com transcrição reversa (RT-qPCR) mede os níveis de transcritos de mRNA para quantificação relativa da expressão gênica, que pode associar-se com condições biológicas específicas. Para se avaliar variações técnicas, genes de referências de expressão estáveis são medidos simultaneamente, juntamente com o gene alvo para normalização dos dados.

Para mensuração da expressão relativa dos retrotransposons *Copia* e *Oswaldo* e *Helena* e *I*, os elementos foram previamente identificados no genoma de *D. mojavensis* com o auxílio da ferramenta *RepeatMasker* que realiza uma busca no genoma, tendo como referência a biblioteca de TEs do gênero *Drosophila*, Flybase genomes (<http://flybase.org/> - Update 20071204).

Os oligonucleotídeos iniciadores utilizados foram desenhados a partir das sequências referência dos respectivos TEs extraídos do genoma, depositadas no banco de dados do RepBase, com auxílio da plataforma IDT SciToolsPrimerQuest (<http://www.idtdna.com/Home/Home.aspx>). As sequências dos oligonucleotídeos utilizados, tamanho dos fragmentos esperado e Tm para cada par estão descritos na Tabela 2.

Tabela 2. Sequência dos oligonucleotídeos iniciadores utilizados para amplificação do gene constitutivo rp49 e dos retrotransposons Helena, I, Copia e Oswaldo.

Elementos	Sequencia 5'-3'	Tm °C	Tamanho (pb)	Referência
rp49	F: GTCGTCGCTTCAAGGGCCAAT R: ATGGGCGATCTCACCGCAGTA	60,0	167	Granzotto et al. 2009
<i>Helena</i>	F: TTGGTCCGCTGCTGTTCTCCTT R: TGAGATTCCACCGCTTGCACCA	62,0	204	Granzotto et al. 2009
I	F : ATCCACTCTTCAACGGCATC R : CATTTCACCATATCGTCCA	60.0	300	
<i>Copia</i>	F: GATGTTTTGACGGCATACT R: AGCATGGGTGAGATGCTAGC	60.0	302	
<i>Oswaldo</i>	F: AGGATGGCATTTCACGGAT R: AATTGCTTGCCTCCGTTTGC	57.2	300	

Para uniformizar os níveis de expressão dos elementos estudados foi utilizada como controle a expressão do gene endógeno constitutivo ribossomal rp49. Além disso, para cada amostra, fez-se um controle negativo de expressão no momento da síntese do cDNA, utilizando-se todos os reagentes necessários para síntese, porém, sem a enzima transcriptase reversa.

Para estabelecer a curva padrão de expressão para quantificação absoluta de cada elemento, as sequências dos elementos foram amplificadas por PCR e visualizadas em gel de

agarose 1%, sendo o produto amplificado isolado do gel e purificado. Em seguida, o produto purificado foi submetido a uma diluição seriada com pelo menos oito pontos para determinação da curva de eficiência e normalização, foram considerados apenas as ampliações cuja curva apresentasse eficiência maior que 1,80 e “slope” próximo a -3,5. Todas as ampliações e análises de amplificação foram realizadas no sequenciador LightCycler 480 da Roche.

Resultados e Discussão

Expressão de TEs em linhagens de D. mojavensis e D. arizonae

As análises de expressão dos NLTRs *Helena* e *I*, e dos LTRs *Oswaldo* e *Copia* nas espécies parentais são apresentadas na Figura 2. Observa-se que a expressão varia entre os elementos, mas algumas homogeneidades são notadas. Por exemplo, o perfil de expressão em tecidos somáticos é relativamente homogêneo entre sexos (carcaças de machos e fêmeas) dentro de cada espécie. A expressão dos quatro elementos também é relativamente homogênea nos tecidos germinativos dos dois sexos nas duas linhagens de *D. arizonae* (HI e LB). Ao contrário, as duas linhagens de *D. mojavensis* (ANZA e LB) apresentam expressão heterogênea entre ovários e testículos para os elementos *Helena* e *Copia*, sendo que em testículos de machos ANZA a expressão desses dois TEs foi mais elevada que em ovários e carcaças.

Algumas populações expressam alguns elementos mais do que outras: enquanto o elemento *Helena* é mais expresso nas populações de *D. arizonae* (HI) em carcaças, em *D. arizonae* (LB) são mais expressos os elementos *I* e *Copia*; já o elemento *Oswaldo* apresenta expressão mais elevada em carcaças das duas populações de *D. mojavensis*.

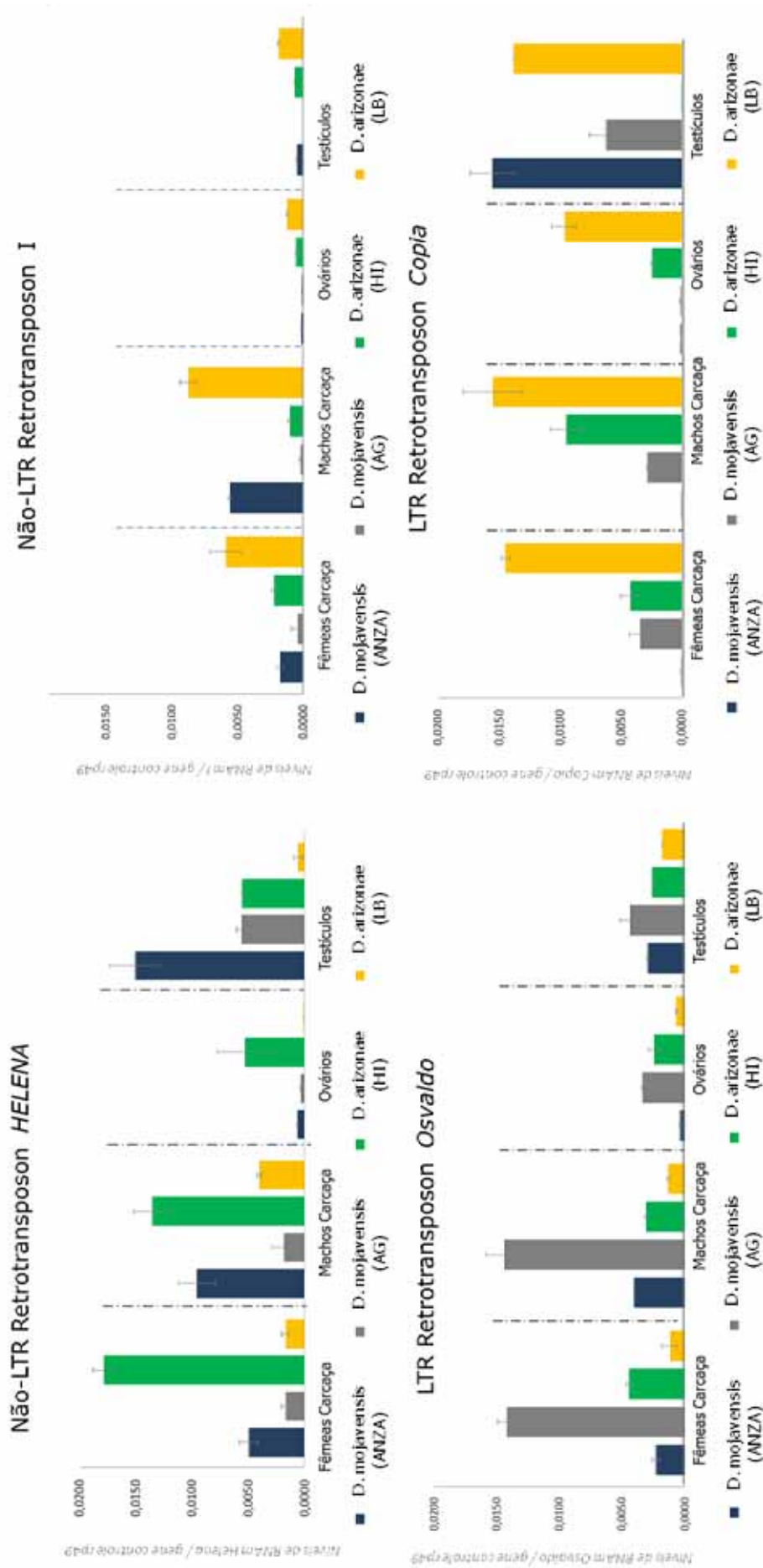


Figura 2. Expressão dos elementos *Helena*, *Oswaldo*, *Copia* e *I* em tecidos de carcaça (tecido germinativo ausente), ovários e testículos de linhagens de *D. mojavensis* e *D. arizonae*. Os níveis de relativos transcritos foram estimados em relação aos do gene constitutivo rp49.

As análises de expressão mostram que o elemento *Helena* ocorre em *D. arizonae*, espécie em que ainda não havia sido relatado. As análises indicam heterogeneidade de expressão tanto entre linhagens de *D. mojavensis*, quanto em *D. arizonae*. Pode-se notar baixa expressão nas linhagens de *D. mojavensis* (AG) e de *D. arizonae* (LB) com aumento em carcaças de *D. arizonae* (ANZA). Esta heterogeneidade é concordante com os resultados de GRANZOTTO (2009) que quantificou e mostrou variação no número de transcritos em cinco linhagens em ovários e carcaças de *D. mojavensis* e expressão em apenas uma de cinco linhagens em *D. simulans*. A distribuição altamente variável de *Helena* em diferentes espécies, resultado de variação do número de cópias e do estado de degradação (REBOLLO et al., 2008; GRANZOTTO et al., 2008), pode explicar a heterogeneidade de expressão observada em nosso estudo.

A expressão do elemento *I* parece estar controlada nos tecidos germinativos dos parentais, apresentado número relativamente maior de transcritos unicamente em carcaças dos dois sexos das linhagens de *D. arizonae* (LB) e de machos de *D. mojavensis* (ANZA) (Figura 2). A ausência virtual de transcritos em tecidos germinativos corresponde às análises anteriores por RT-qPCR (Capítulo 1) e do mobiloma expresso (Capítulo 2) para as linhagens (ANZA e HI); entretanto, naquelas análises a presença de transcritos em tecidos somáticos não foi detectada. Estes resultados mostram uma dinâmica diferente do elemento *I* em *D. mojavensis* e *D. arizonae*, em comparação a de *D. melanogaster*, espécie em que é chamado fator I, cuja expressão e transposição só ocorre, ou ao menos só foi reportada, em linhagens germinativas de fêmeas (SELEME et al., 2005; CHAMBEYRON et al., 2008). Essa diferença pode ser explicada pelos nossos estudos que permitiram classificar o elemento *I* dessas duas espécies como pertencentes a uma família *I* diferente daquela encontrada em *D. melanogaster* e outras espécies do seu subgrupo (Capítulo 1).

O elemento *Osvaldo*, que apresenta altas taxas de transposição em linhagens de *D. buzzatii* e *D. koepferae* (LABRADOR; FONTDEVILA, 1994; FONTDEVILA, 2005), parece estar completamente controlado na linhagem germinativa e somática das quatro linhagens aqui estudadas, exceto em carcaças de machos e fêmeas da linhagem de *D. mojavensis* AG. Esse resultado sugere que *Osvaldo* está regulado em células germinativas dessas linhagens, mas é instável em células somáticas, nas quais danos genéticos decorrentes de sua transposição podem reduzir o valor adaptativo hospedeiro (GONZÁLEZ; PETROV, 2012).

O elemento *Copia* também apresenta expressão heterogênea. É altamente expresso em tecidos germinativos e somáticos na linhagem LB de *D. arizonae* e tem expressão baixa na linhagem HI. Em *D. mojavensis*, esse elemento é mais expresso em testículos de ANZA, porém sua expressão é baixa em testículos da linhagem AG. Embora as espécies aqui estudadas e *D. melanogaster* sejam evolutivamente distantes, o mesmo perfil de expressão é observado - elevada taxa de transposição nos machos - o que correspondendo à taxa de transposição elevada verificada apenas em testículos e não nos ovários (PASYUKOVA et al., 1997). Neste mesmo trabalho, Pasyukova et al. (1997) chama a atenção para a acessibilidade do cromossomo Y a inserções de TEs, e destaca que estudos de hibridização *in situ* em cromossomos politênicos mostram que os elementos *I*, *P* e *hobo*, ligados à disgenesia híbrida, não estão presentes no cromossomo Y, mas *Copia*, *Doc*, *mdg1*, *mdg3* e *gypsy* estão todos presentes. Isto seria uma das razões para explicar porque ocorre transposição de muitos retrotransposons LTRs em linhagens germinativas dos machos.

2 Expressão em Parentais e Híbridos

A expressão dos quatro elementos também foi quantificada em híbridos entre diferentes populações das duas espécies. Contudo, enquanto a expressão dos TEs nos parentais foi mensurada em diferentes tecidos independentemente (carcaças, ovários e

testículos), para os híbridos foi quantificada a partir de RNA extraído de indivíduos inteiros. Um dos motivos de compararmos apenas os resultados de RT-qPCR de material obtido de indivíduos híbridos inteiros (de cinco a dez indivíduos), foi devido ao baixo número de descendentes obtidos em determinados intercruzamentos, o que dificultava a obtenção de RNA suficiente de tecido germinativo para síntese de cDNA e ampliações por RT-qPCR. Sendo assim, utilizamos apenas RNA de indivíduos híbridos inteiros e comparamos com tecidos somáticos e germinativos dos parentais. De maneira geral, a expressão nos híbridos, independente do sentido do cruzamento e do elemento, tem padrão semelhante a dos parentais (Figuras 3 a 5) e geralmente não é alterado entre dois sentidos de cruzamentos de mesma linhagem.

2.2.1 Expressão do Retrotransposon *I*

A sequência estudada do elemento *I* é mais expressa em híbridos machos dos cruzamentos A (fêmeas *D.moj.* ANZA x machos *D. arz.* HI), e tem expressão menor, mas presente, em híbridos do cruzamento recíproco (B), e ausente nos parentais. Como híbridos machos derivados de cruzamentos entre fêmeas de *D. arizonae* e machos de *D. mojavensis* apresentam esterilidade (RUIZ, 1990; REED; MARKOW, 2004), a maior expressão nas linhagens germinativas dos machos híbridos pode estar associada à esterilidade encontrada na direção do cruzamento B (REED; MARKOW, 2004). Esse fenômeno pode ser correspondente à digenesia observada em cruzamentos de *D. melanogaster*, decorrente da presença do elemento *I* em machos parentais (PICARD, 1976; BUCHETON et al., 1984). Entretanto, não podemos descartar que a expressão elevada no híbrido do cruzamento B seja apenas efeito das diferenças entre as espécies parentais e não um fator da esterilidade.

Porém, curiosamente, podemos notar uma tendência de maior expressão da sequência de *I* nos machos híbridos em quase todas as direções, apresentando as fêmeas híbridas,

expressão média a dos tecidos parentais (Figura 3). Contudo, no cruzamento F (fêmeas *D. arizonae* HI com machos *D. mojavenensis* AG), os híbridos de ambos os sexos, apresentam um forte aumento de expressão em relação aos parentais (Figura 5). Nesta direção de cruzamento, observamos um forte isolamento pré-zigótico, com poucos descendentes na geração F1. Esse isolamento é ainda mais forte no cruzamento recíproco, de fêmeas AG e machos HI, onde não foram produzidos descendentes suficientes para mensurar a expressão dos TEs (dados não mostrados). Este resultado sugere que a desrepressão dos TEs nos híbridos na direção E de cruzamento, possa estar relacionada ao forte isolamento reprodutivo observado.

2.2.2 Expressão do retrotransposon *Helena*

O elemento *Helena* parece apresentar cópias ativas e com expressão variável somente em *D. mojavenensis*. Da mesma forma, podemos notar que existe uma variação de expressão nos tecidos dentro e entre as linhagens de *D. mojavenensis* analisadas neste estudo, com expressão mais elevada nos machos híbridos dos cruzamentos A e C, em que foram utilizadas fêmeas de *D. mojavenensis* de ANZA (Figuras 3 e 4). Nos cruzamentos em que as linhagens são derivadas de populações simpátricas (G e H) observa-se que o elemento *Helena* se encontra regulado, tanto nos tecidos parentais quanto nos híbridos (Figura 6). Esse resultado é concordante com aqueles descritos por Granzotto et al. (2009), uma vez que a expressão de *Helena* é influenciada pela espécie e linhagem hospedeira, resultando em atividade diferencial.

2.2.3 Expressão do retrotransposon *Oswaldo*

Nos cruzamentos interespecíficos entre *D. buzzatii* e *D. koepferae*, a taxa de transposição de *Oswaldo*, como citada anteriormente, é aumentada nos híbridos (LABRADOR et al., 1999). As sequências amplificadas deste elemento no transcriptoma

híbrido de *D. mojavensis* e *D. arizonae*, em todas as direções de cruzamento, apresentam-se controladas. Embora ocorra alta expressão nas carcaças de machos e fêmeas parentais da população de *D. mojavensis* de ANZA e *D. arizonae* de LB, esta parece não se refletir na expressão nos híbridos (Figura 5 e 6). Isto sugere um controle celular deste TE dos híbridos nestes cruzamentos, talvez por piRNA herdados das linhagens paternas.

2.2.4 Expressão do retrotransposon *Copia*

O elemento *Copia* apresenta expressão nos machos *D. arizonae* LB ou HI e baixa expressão nos híbridos (Figuras 3, 4 e 6). Esse padrão é diferente do que ocorre em *D. melanogaster*, em que a diversidade e plasticidade do elemento *Copia* poderia se traduzir em desregulação no genoma dos híbridos de *D. mojavensis* e *D. arizonae* (PASYUKOVA et al., 1997; DE ALMEIDA; CARARETO, 2006).

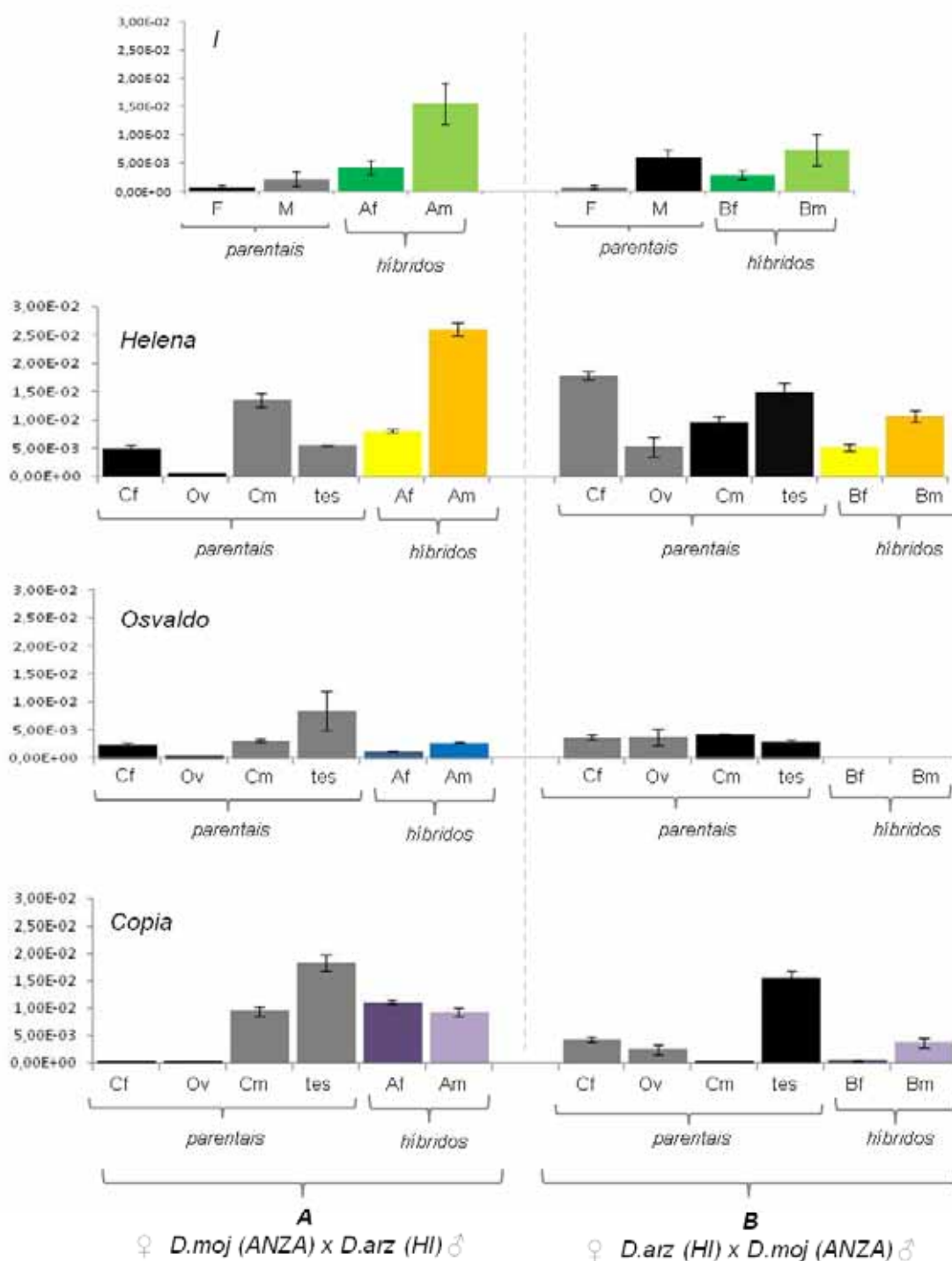


Figura 3. Expressão dos elementos *I*, *Helena*, *Osvaldo*, *Copia* em tecidos de carcaça fêmeas (Cf) e machos (Cm), ovários (ov) e testículos (te) de linhagens parentais *D. mojavensis* (colunas em preto) e *D. arizonae* (colunas em cinza) e híbridos dos cruzamentos A e B (direções dos cruzamentos, consultar Tabela 1); f (fêmeas), m (machos). Os níveis de transcritos relativos foram estimados em relação aos do gene rp49. Gráfico do elemento I – amostras da extração de parentais inteiros, fêmeas (f) e machos (m).

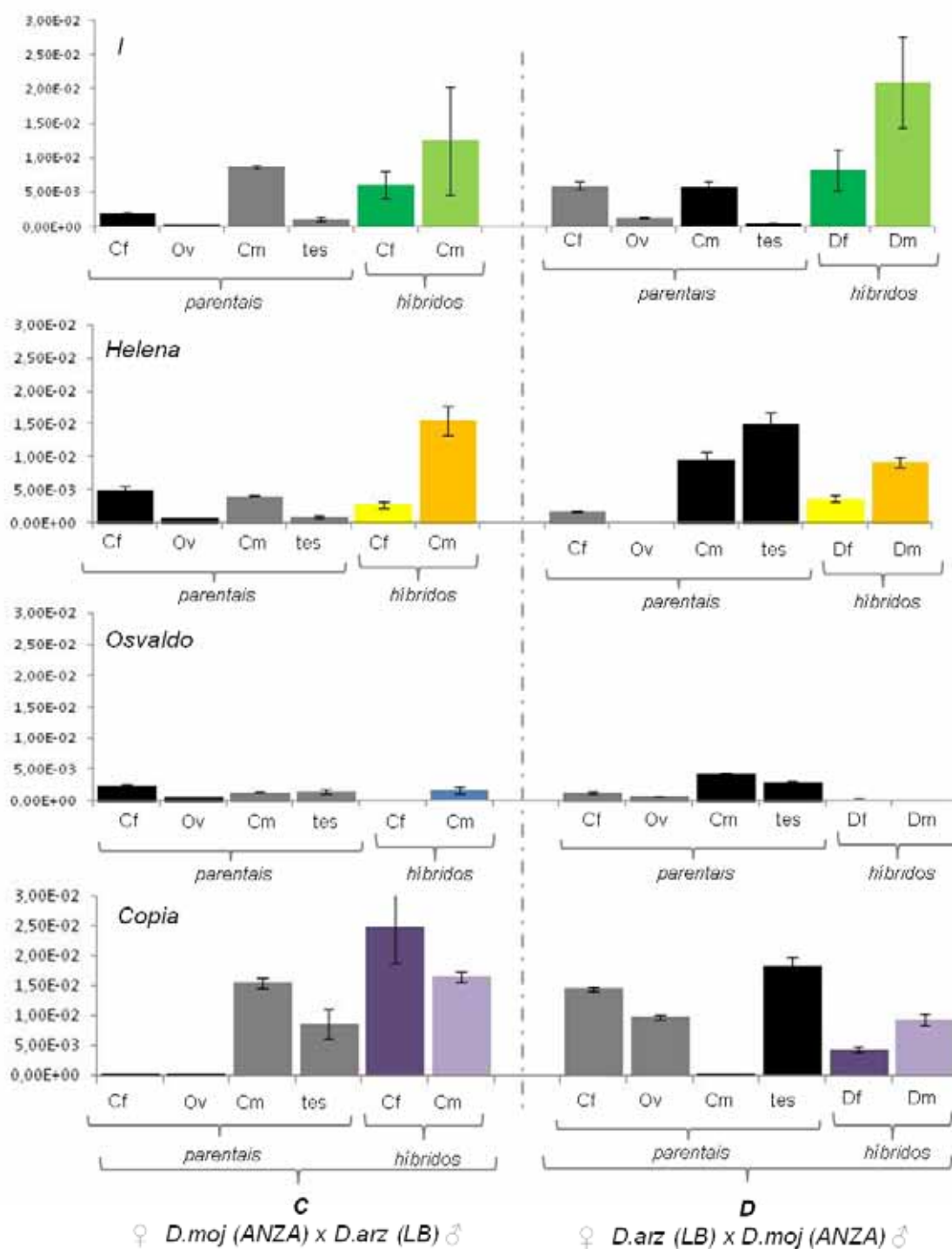


Figura 4. Expressão dos elementos *I*, *Helena*, *Osvaldo*, *Copia* em tecidos de carcaça fêmeas (Cf) e machos (Cm), ovários (ov) e testículos (te) de linhagens parentais *D. mojavensis* (colunas em preto) e *D. arizonae* (colunas em cinza) e híbridos dos cruzamentos C e D (direções dos cruzamentos, consultar Tabela 1); f (fêmeas), m (machos). Os níveis de transcritos relativos foram estimados em relação aos do gene rp49.

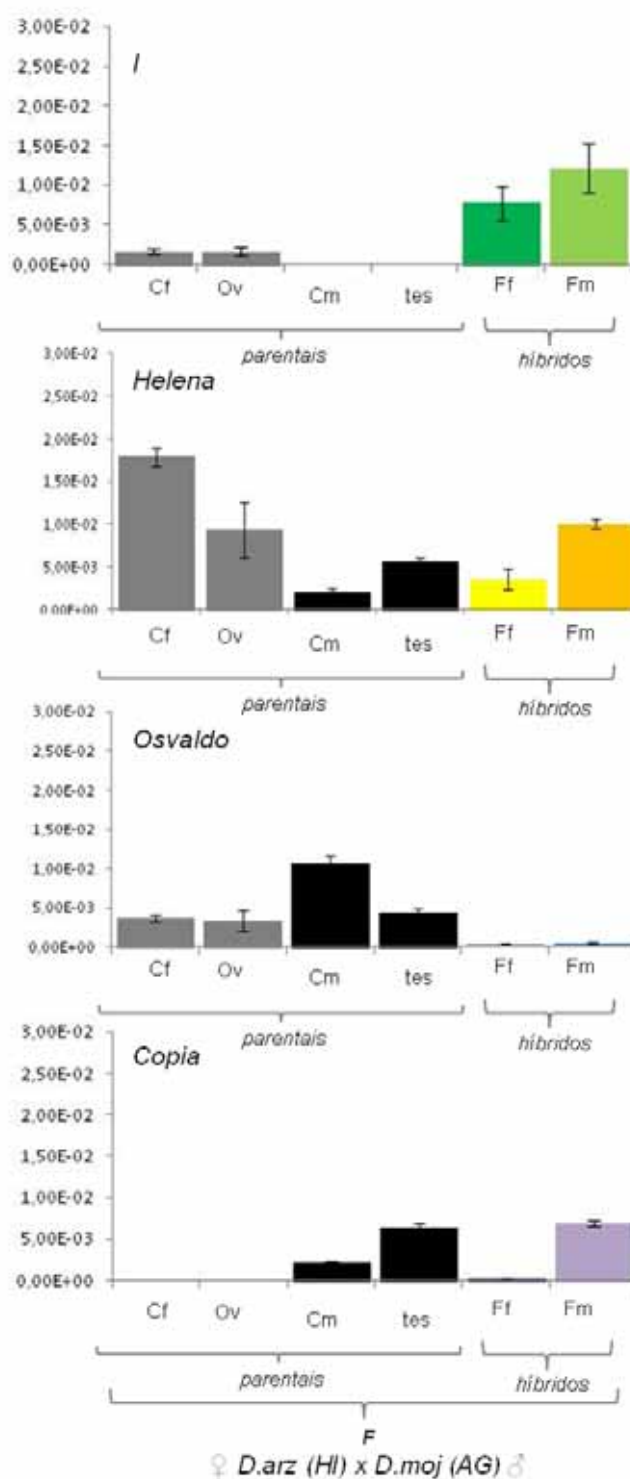


Figura 5. Expressão dos elementos *I*, *Helena*, *Osvaldo*, *Cópia* em tecidos de carcaça fêmeas (Cf) e machos (Cm), ovários (ov) e testículos (te) de linhagens parentais *D. mojavensis* (colunas em preto) e *D. arizonae* (colunas em cinza) e híbridos do cruzamento F (direção do cruzamento, consultar Tabela 1); f (fêmeas), m (machos). Os níveis de transcritos relativos foram estimados em relação aos do gene rp49.

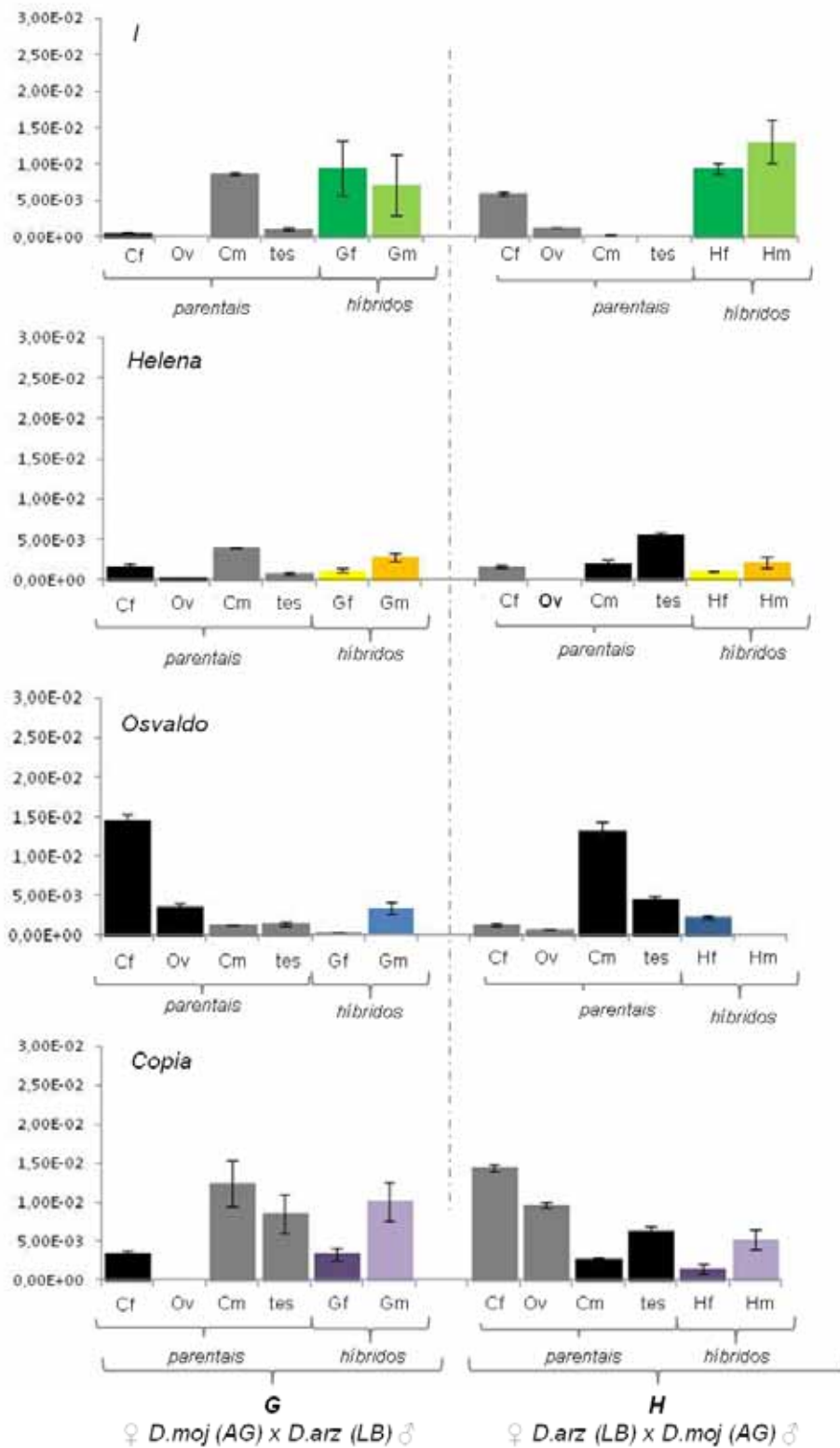


Figura 6. Expressão dos elementos *I*, *Helena*, *Osvaldo*, *Copia* em tecidos de carcaça fêmeas (Cf) e machos (Cm), ovários (ov) e testículos (te) de linhagens parentais *D. mojavensis* (colunas em preto) e *D. arizonae* (colunas em cinza) e híbridos dos cruzamentos G e H (direções dos cruzamentos, consultar Tabela 1); f (fêmeas), m (machos). Os níveis de transcritos relativos foram estimados em relação aos do gene rp49.

Nosso trabalho mostra que há uma tendência nos dois elementos não-LTR estudados de maior expressão nos machos híbridos nas direções de cruzamento de A-D, representados por fêmeas e machos *D. mojavensis* (ANZA) e *D. arizonae* das duas linhagens. Alguns relatos discutem a alta expressão relativa de genes em machos de *D. melanogaster*, em que a contribuição de tecidos individuais para expressão sexo-enviesado de genes (*sex-biased*), revelam que os níveis de expressão somente em testículos e glândulas acessórias são positivamente relacionados à expressão *sex-biased*, ilustrando a influência da expressão total sexo-específica do macho. Isso significa que nas fêmeas há maior restrição seletiva de genes de procedência masculina, provavelmente contribuindo para evolução mais lenta nestas fêmeas (PARISI et al., 2003; ELLEGREN; PARSCH, 2007; ASSIS; BACHTROG, 2012; PARSCH; ELLEGREN, 2013). Isto abre uma janela para novas pesquisas com o fim de investigar se os elementos móveis são influenciados por este mecanismo evolutivo ainda não completamente compreendido. Além disso, parece ocorrer uma diminuição na expressão nos híbridos e parentais de linhagens que habitam zonas de simpatria, onde a ocorrência de reforço do isolamento reprodutivo, embora bastante questionada, é proposta (BONO; MARKOW, 2009).

A grande maioria dos estudos que medem a atividade de TEs são provenientes de análises de hibridização *in situ* em tecidos ou cromossômica (LABRADOR; FONTDEVILA, 1994; . LOZOVSKAYA et al., 1996; BORIE et al., 2002; EVGEN'EV; ARKHIPOVA, 2005). Poucos são os estudos que revelam a taxa de expressão por RT-qPCR de TEs de diferentes famílias e tecidos, e em menor quantidade ainda são os realizados em condição de hibridação, como os representados neste trabalho. Nossos resultados mostram que TEs pertencentes a diferentes famílias apresentam expressão variável, e que essa expressão varia entre espécies, linhagens e tecidos. Dessa forma devemos ser cautelosos com resultados de atividade de TEs, pois podem estar sendo expressos ou não apenas nas condições em que foram analisadas

naquele momento. Desse modo, para serem feitas inferências é preciso se levar em consideração a espécie e linhagens estudadas, os tecidos particulares e as condições ambientais, além da técnica de análise empregada.

Referências

- AKKOUCHE, A. et al. Tirant, a newly discovered active endogenous retrovirus in *Drosophila simulans*. **Journal of Virology**, Washington Dc, v. 86, n. 7, p. 3675-3681, 2012.
- ASSIS, R.; ZHOU, Q.; BACHTROG, D. Sex-biased transcriptome evolution in *Drosophila*. **Genome Biology and Evolution**, Oxford, v. 4, n. 11, p. 1189-2006, 2012.
- BIÉMONT, C. A brief history of the status of transposable elements: from junk DNA to major players in evolution. **Genetics**, Austin, v. 186, n. 4, p. 1085-93, 2010.
- 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**, Oxford, v. 22, n. 7, p. 1387-95, 2009.
- BORIE, N. et al. Tissue-specificity of 412 retrotransposon expression in *Drosophila simulans* and *D. melanogaster*. **Heredity**, London, v. 89, n. 4, p. 247-252, 2002.
- BUCHETON, A. et al. The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the I factor. **Cell**, Cambridge, v.38, n.1, p.153-163, 1984.
- CHAMBEYRON, S. et al. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 105, n. 39, p. 14964-14969, 2008.
- DE ALMEIDA, L. M.; CARARETO, C. M. Sequence heterogeneity and phylogenetic relationships between the copia retrotransposon in *Drosophila* species of the repleta and melanogaster groups. **Genetics Selection Evolution**, London, v. 38, n. 5, p. 535-50, 2006.
- DOMINGUES, D. S. et al. Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. **BMC Genomics**, London, v. 13, p. 137, 2012. Disponível em: <<http://www.biomedcentral.com/1471-2164/13/137>>. Acesso em: 21 fev. 2014.
- ELLEGREN, H.; PARSCH, J. The evolution of sex-biased genes and sex-biased gene expression. **Nature Reviews Genetics**, London, v.8, n.9, p.689-98, 2007.
- EVGEN'EV, MB, ARKHIPOVA, IR. Penelope-like elements--a new class of retroelements: distribution, function and possible evolutionary significance. **Cytogenetic and Genome Research**, Basel, v.110, p.510-21, 2005.

FABLET, M.; VIEIRA, C. Evolvability, epigenetics and transposable elements. **Biomolecular Concepts**, Berlin, v. 2, p. 333-341, 2011.

FONTDEVILA, A. Hybrid genome evolution by transposition. **Cytogenetic and Genome Research**, Basel, v. 110, n. 1-4, p. 49-55, 2005.

GONZÁLEZ, J.; PETROV, D.A. Evolution of genome content: population dynamics of transposable elements in flies and humans. **Methods Molecular Biology**, Washington, v.855, p.361-83, 2012.

GRANZOTTO, A. et al. The evolutionary dynamics of the Helena retrotransposon revealed by sequenced *Drosophila* genomes. **BMC EVOLUTIONARY BIOLOGY**, London, v. 9, 2009.

GUERREIRO, M.P.; FONTDEVILA, A. Chromosomal distribution of the transposable elements Osvaldo and blanco in original and colonizer populations of *Drosophila buzzatii*. **Genetics Research**, v.77, n.3, p.227-38, 2001.

JORDAN, I.K.; MCDONALD, J.F. Evolution of the copia retrotransposon in the *Drosophila melanogaster* species subgroup. **Molecular Biology and Evolution**, Chicago, v.15, n.9, p.1160-71, 1998.

LABRADOR, M. et al. Interspecific hybridization increases transposition rates of Osvaldo. **Molecular Biology and Evolution**, Chicago, v. 16, n. 7, p. 931-937, 1999.

LABRADOR, M.; FONTDEVILA, A. High transposition rates of Osvaldo, a new *Drosophila buzzatii* retrotransposon. **Molecular & General Genetics**, Berlin, v. 245, n. 6, p. 661-674, 1994.

LOZOVSKAYA, E.R. et al. Germline Transformation of *Drosophila virilis* mediated by the transposable element hobo. **Genetics**, Austin, v.142, p.173-177, 1996.

MCCLINTOCK, B. The significance of responses of the genome to challenge. **Science**, Washington, v. 226, n. 4676, p. 792-801, 1984.

MOROZOVA, T.V. et al. Impact of the regulatory regions of retrotransposon copia on the level of its expression in testes of *Drosophila melanogaster*. **Russian Journal of Genetics**, Spring St - New York, v.40, n.2, p.119-124, 2004.

PANTAZIDIS, A.; LABRADOR, M. The retrotransposon Osvaldo from *Drosophila buzzatii* displays all structural features of a functional retrovirus. **Molecular Biology and Evolution**, Chicago, v. 16, n. 7, p. 909-921, 1999.

PARISI, M. et al. Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. **Science**, Washington, v.299, n.5607, p.697-700, 2003.

PARSCH, J.; ELLEGREN, H. The evolutionary causes and consequences of sex-biased gene expression. **Nature Reviews Genetics**, London, v.14, n.2, p.83-7, 2013.

PASYUKOVA, E. et al. Germ line transposition of the copia retrotransposon in *Drosophila melanogaster* is restricted to males by tissue-specific control of copia RNA levels. **Molecular & General Genetics**, Berlin, v.255, n.1, p.115-124, 1997.

PICARD, G. Non-mendelian female sterility in *Drosophila melanogaster*: hereditary transmission of I factor. **Genetics**, Austin, v.83, n.1, p.107-123, 1976.

REBOLLO, R. et al. Losing *Helena*: The extinction of a *Drosophila* line-like element. **BMC Genomics**, London, v.9, 2008.

REED, L. K.; MARKOW, T. A. Early events in speciation: polymorphism for hybrid male sterility in *Drosophila*. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 101, n. 24, p. 9009-9012, 2004.

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

SELEME, M.D. et al. In vivo RNA localization of I factor, a non-LTR retrotransposon, requires a cis-acting signal in ORF2 and ORF1 protein. **Nucleic Acids Research**, Oxford, v.33, n.2, p. 776-785, 2005.

SELEME, M.D.C. et al. Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v.103, n.17, p.6611-6616, 2006.

VU, W.; NUZHIDIN, S. Genetic variation of copia suppression in *Drosophila melanogaster*. **Heredity**, Edinburg, v.106, n.2, p.207-17, 2011.

MATERIAL SUPLEMENTAR

I. Meio cultura cactus

Para 1 Receita

Preparo do suco de cactus – picar em pedaços cactus jovens (*Opuntia ficus-indica*) e colocar na água, colocar no fogo e deixar ferver até ficar macio, tire a água e bata no liquidificador, em seguida autoclavar.

Meio de cultura

500ml de suco de cactus

500 ml de água

20 g de agar

30g de fermento biológico

25g de xarope de glicose/açúcar ou mel Karo

10ml de álcool e nipagim 7%

10ml de ácido propiônico (9,5%)

250g de banana nanica

Na água adicionar o agar e o xarope, deixar até ferver (cuidado que sobe rápido!), acrescentar a banana, deixar até dissolver e ferver bem, colocar o suco de cactus e deixe ferver. Em seguida, abaixe o fogo e coloque o fermento. Antes de colocar o álcool e o ácido propiônico, coloque os dois na mesma proveta e despeje vagarosamente.

Para tubos de 2 a 3ml de meio/tubo, 1000mL serve 500tubos.

Para garrafas 1 litro de meio serve 15 garrafas.

CAPÍTULO 4

Diversidade no tamanho do genoma em híbridos de *Drosophila*

Introdução

O tamanho do genoma (C-value) apresenta enorme variação entre organismos, a qual, de modo geral, não está relacionada com a complexidade biológica (GREGORY, 2001; PETROV, 2001). A partir do sequenciamento do genoma de um importante número de espécies, é consenso que esta falta de correlação está principalmente associada à presença de sequências repetidas. O debate ainda aberto refere-se ao significado evolutivo dessas variações.

Modelos de seleção consideram que o tamanho do genoma é relativo ao tamanho celular e que esta característica deve estar submetida à seleção (CAVALIER-SMITH; BEATON, 1999; GREGORY, 2001). De acordo com esta hipótese, o tamanho do genoma é mantido em condições ótimas por força de seleção estabilizadora, que elimina genomas pequenos ou extremamente maiores, o que pode tornar-se evidente que genomas de tamanhos extremos sempre apresentem uma desvantagem para o valor adaptativo do organismo (BEAULIEU, 2010). Modelos de seleção neutra postulam que a variação do tamanho do genoma é essencialmente resultado de diferenças do tamanho efetivo da população. No caso de populações pequenas, com fraca ação da seleção, pode ocorrer o acúmulo de sequências repetidas, dentre elas os elementos de transposição (TEs).

Diferentes processos podem explicar modificações no tamanho do genoma, e muitas dessas diferenças são devidas a variação no conteúdo de DNA satélite em regiões de heterocromatina (BOSCO et al., 2007), ou também pela quantidade de elementos

transponíveis (VIEIRA, et al., 2002, BOULESTEIX, et al. 2006). Variáveis ambientais como condição de temperatura e umidade, composição de meios de cultura e a idade dos organismos adultos, são também fatores que afetam a estimativa do tamanho do genoma (VIEIRA, et al., 2002; NARDON et al., 2005; CASACUBERTA; GONZALEZ; 2013).

Hibridação também é um fator que pode afetar o tamanho do genoma, pois após hibridação podem ocorrer grandes e rápidas mudanças genômicas, incluindo rearranjos cromossômicos, alterações no conteúdo de sequências satélites, e mobilização de elementos de transposição, conseqüentemente acompanhadas com aumento do tamanho do genoma (SANMIGUEL; BENNETZEN, 1998; PETROV, 2001; VIEIRA, et al., 2002; NARDON et al., 2005; BOULESTEIX, et al., 2006; BOSCO et al., 2007).

Em cruzamentos interespecíficos de *Narcissus* (Amaryllidaceae), por exemplo, efeito materno, rearranjos cromossômicos e recombinações influenciam o tamanho do genoma dos híbridos (MARQUES et al., 2012). Resultados semelhantes também foram observados em *Helianthus* e *Brassica* (SONG et al., 1995; BAACK, et al., 2005). Especificamente em híbridos de *Helianthus*, originados de cruzamentos de uma mesma espécie parental com outras, os híbridos apresentam 50% a mais de DNA nuclear que os parentais, principalmente devido à explosão de elementos de transposição (BAACK et al., 2005).

A mensuração do tamanho do genoma pode ser obtida por diferentes métodos. O sequenciamento completo nos dá uma boa aproximação do tamanho do genoma, mas em geral, regiões de heterocromatina que são ricas em sequências repetidas não são levadas em consideração. Outro caminho é por citometria de fluxo, técnica que mensura a intensidade de fluorescência que é incorporada no DNA e com o uso de um genoma referência obtemos uma estimativa do tamanho do genoma.

A quantidade de DNA em uma célula haploide (não replicado), ou núcleo gamético de um organismo, referida como “C-value” ou tamanho do genoma, é normalmente registrada

em termos de massa (picogramas, μg , com $1 \mu\text{g} = 10^{-12}\text{g}$) (GREILHUBER et al., 2005). O tamanho do genoma *D. mojavensis* está estimado em 0,17 μg para macho e fêmeas, um dos menores entre as espécies de *Drosophila*, acompanhado por *D. mercatorum* e *D. erecta* respectivamente (GREGORY; JOHNSTON, 2008). Contrariamente à *D. virilis*, que possui um dos maiores genomas, chegando a 0,34 μg (machos e fêmeas), que contém de 40 a 46% de conteúdo de DNA satélite em heterocromatina, em *D. mojavensis*, a porção de sequências satélites corresponde a apenas 1% (BOSCO et al., 2007).

Resultados de análises de 67 espécies de *Drosophila*, com poucas exceções, indicam que o genoma dos machos é significativamente menor que o das fêmeas (GREGORY; JOHNSTON, 2008), sendo a média entre as espécies de 0,21 μg . O tamanho do genoma de *D. arizonae*, a segunda espécie foco do presente trabalho, era ainda desconhecido.

O objetivo desse trabalho foi estimar o tamanho do genoma de duas espécies de *Drosophila* - *D. mojavensis* e *D. arizonae* - e de conhecer em que medida a estimativa do tamanho do genoma é afetada nos híbridos entre estas espécies irmãs do grupo *repleta* de *Drosophila*. As duas são espécies endêmicas de áreas áridas do sudoeste dos Estados Unidos e México. *D. arizonae* ocorre em uma ampla distribuição: desde a região do Cabo na Baja Califórnia, Sudoeste do Arizona, sudoeste do Novo México, sudoeste do deserto de Sonora, no estado de Hidalgo (HI), leste do México, Tamaulipas no sudeste do México a Guatemala. Por sua vez, *D. mojavensis*, ocorre nos desertos de Mojave e Sonora, no sudeste da Califórnia e Baja Califórnia (EUA), incluindo a população da ilha de Santa Catalina (CI), e ao longo da costa oeste de Sonora e Sinoloa (México), onde é simpátrica com *D. arizonae* (WASSERMAN; KOEPFER, 1977; RUIZ, 1990; PFEILER et al., 2009; RUIZ; HEED, 2009).

Híbridos entre as duas espécies exibem uma variação do isolamento pré- e pós-zigótico, com esterilidade dos machos híbridos, provavelmente controlados por fatores

presentes em diferentes frequências, como polimorfismos em múltiplos locos em *D. mojavensis* (REED; MARKOW, 2004; MACHADO et al., 2007). A esterilidade fenotípica é manifesta em machos no cruzamento entre fêmeas *D. arizonae* e machos *D. mojavensis*. Entretanto, no cruzamento recíproco, não foi observada esterilidade, sendo os machos híbridos de fêmeas *D. mojavensis* frequentemente férteis. Particularmente, cruzamentos entre fêmeas de *D. mojavensis* de linhagem da população da ilha de Santa Catalina, Califórnia-EUA e machos de *D. arizonae* de populações do Arizona-EUA (Peralta Canyon - PERA), produzem machos híbridos com menor motilidade espermática, e mesmo aqueles com motilidade espermática são incapazes de produzir descendentes (REED; MARKOW, 2004).

Mostramos neste estudo que a estimativa do tamanho do genoma é variável e depende das populações que estudamos. Além disso, identificamos também um efeito parental na estimativa do tamanho do genoma, o que sugere mudança na estrutura da cromatina que pode estar associada à esterilidade observada em determinados sentidos dos cruzamentos.

Material e Métodos

Amostras e cruzamentos

Foram utilizadas nesses estudos linhagens obtidas da coleção da UC San Diego *Drosophila* Stock Center (San Diego, CA, USA), sendo uma amostra de *D. arizonae* e de *D. mojavensis* derivadas de populações que vivem em alopatria (Metztlán, Hidalgo, México - stock number: 15081-1271.17 e Anza Borrego Desert, CA, EUA - 15081-1352.01), e outras de populações simpátrica (*D. arizonae* Las Bocas, Sonora-México 15081-1271.18 e *D. mojavensis* Agiabampo Bay, Sonora-México 15081-132.26), bem como a linhagem de *D. mojavensis* utilizada para o sequenciamento genômico (Catalina Island, California, 15081-1352.22).

Foi estimado o tamanho relativo dos genomas nos híbridos F1 (com 10 dias de idade) das linhagens derivadas de populações de *D. mojavensis* e *D. arizonae* que se encontram originalmente na natureza em alopatria (ALLO), como em simpatria (SYM), produzidos em todas as direções de cruzamentos, sendo, oito sentidos de cruzamentos interespecíficos e quatro controles (entre casais de linhagem da mesma espécie), em cinco réplicas cada, com 10 casais em cada réplica (Tabela 1). Além destes, com o mesmo número de réplicas, e mesmo número de casais, foram obtidos descendentes de cruzamentos entre casais da mesma linhagem. As linhagens e cruzamentos foram todos mantidos em mesmo meio de cultura e temperatura (25°C).

Tabela 1. Representação dos sentidos dos cruzamentos interespecíficos entre *D. mojavenis* e *D. arizonae*

Fêmeas	Linhagem	Machos	Linhagem	Cruzamento	Distribuição
<i>D. mojavenis</i>	ANZA	x <i>D. arizonae</i>	HI	A	Alopátrica
<i>D. arizonae</i>	HI	x <i>D. mojavenis</i>	ANZA	B	Alopátrica
<i>D. mojavenis</i>	ANZA	x <i>D. arizonae</i>	LB	C	Alopátrica
<i>D. arizonae</i>	LB	x <i>D. mojavenis</i>	ANZA	D	Alopátrica
<i>D. mojavenis</i>	AG	x <i>D. arizonae</i>	HI	E*	Alopátrica
<i>D. arizonae</i>	HI	x <i>D. mojavenis</i>	AG	F	Alopátrica
<i>D. mojavenis</i>	AG	x <i>D. arizonae</i>	LB	G	Alopátrica
<i>D. arizonae</i>	LB	x <i>D. mojavenis</i>	AG	H*	Simpátrica
<i>D. mojavenis</i>	ANZA	x <i>D. mojavenis</i>	AG	W	Interpopulacional
<i>D. mojavenis</i>	AG	x <i>D. mojavenis</i>	ANZA	X	Interpopulacional
<i>D. arizonae</i>	HI	x <i>D. arizonae</i>	LB	Y	Interpopulacional
<i>D. arizonae</i>	LB	x <i>D. arizonae</i>	HI	Z	Interpopulacional

*não nasceu

Citometria de fluxo

O tamanho do genoma foi estimado por citometria de fluxo de núcleos de células somáticas a partir de tecido de cabeças de machos e fêmeas, cinco cabeças em cada réplica e no mínimo cinco réplicas por cruzamento. A quantificação relativa em pg do tamanho do genoma foi calculada em relação ao genoma controle de amostras de *D. virilis* (0,34 pg), verificadas juntamente com as demais amostras e ao mesmo tempo.

As amostras foram maceradas a 4°C em um tubo de 1,5mL contendo 200µl do tampão de extração (citrato de sodium 30mM; Triton X-100 1:10000v; RNase A, 20mg/ml; MOPS, 20mM; Iodeto de propídio a 1mg/mL, pH 7,2). Em seguida, a mistura foi filtrada em malha de nylon a 140-micron e então a 30-micron, por fim, adicionado 520µl de tampão de extração e 80µl de iodeto de propídio (Sigma, 1mg/ml ou 100µg/ml final). Análises realizadas no “Centro de Génétique et de Physiologie Moléculaire et Cellulaire” na Universidade Lyon1-França. Teste de significância ANOVA, foram realizados no programa GraphPad Prism5.

Resultados e Discussão

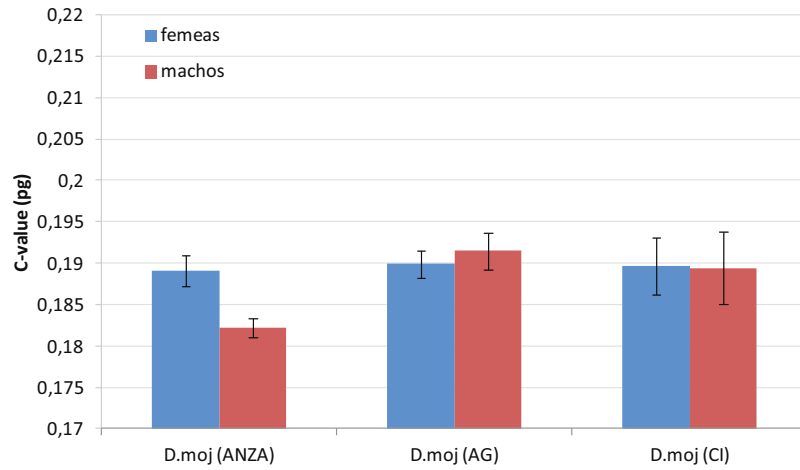
Tamanho do genoma em D. mojavensis e em D. arizonae

Nossas estimativas para as amostras das três linhagens *D. mojavensis* analisadas (ANZA, AG, CI) produziu um valor médio para fêmeas de 0,190 pg e para machos 0,188 pg. Não há diferença entre o tamanho do genoma estimado entre as três linhagens de *D. mojavensis* ($p < 0,05$), nem entre sexos (Tabela S1).

O tamanho do genoma de *D. arizonae* ainda não havia sido quantificado por citometria de fluxo antes deste estudo e foi estimado em média em 0,203 pg em fêmeas e 0,195 pg em machos (Figura 1B, Tabela S1). NA linhagem de *D. arizonae* HI, há diferença entre fêmeas e machos ($p < 0,05$, Tukey's multiple comparison). Mas, fêmeas e machos de *D. arizonae* LB não diferem ($p < 0,1385$). Entre as linhagens, fêmeas de HI diferem de machos e fêmeas de LB. Podemos citar algumas diferenças entre elas (ANOVA one way, Tukey's multiple comparison, $p < 0,05$) que i) *D. mojavensis* linhagem ANZA, difere entre sexos de ambas as linhagem de *D. arizonae*; ii) *D. mojavensis* AG não difere de *D. arizonae* LB; porém as fêmeas diferem das fêmeas de *D. arizonae* HI mas não os machos. ii) O mesmo ocorre em *D. mojavensis* de CI onde as fêmeas, não diferem de *D. arizonae* de LB, porém as fêmeas diferem das fêmeas da linhagem HI, mas não dos machos.

As diferenças entre o tamanho do genoma de *D. arizonae* e *D. mojavensis* foram significantes ($p < 0,005$) como ilustrado na Figura 1.

A



B

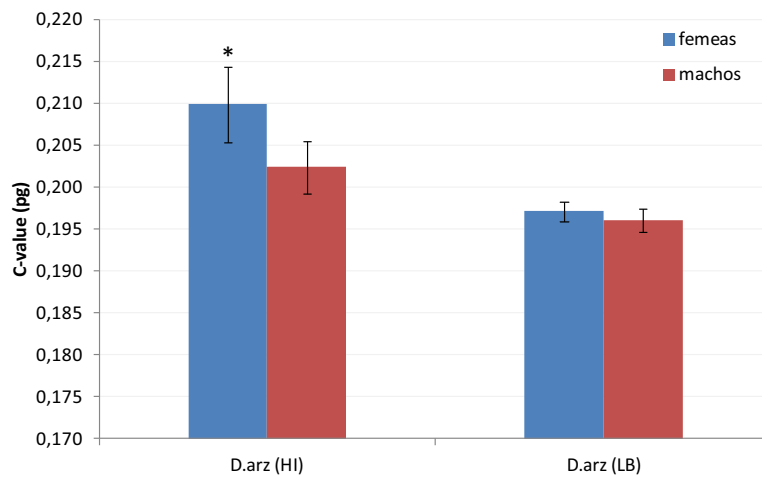


Fig. 1. Distribuição do tamanho do genoma (pg) das populações de *D. mojavensis* (A) e *D. arizonae* (B) (Teste Anova Two-way, comparações múltiplas de Tukey's, 5 réplicas, * $P < 0.05$, média \pm SE).

Cruzamentos controle

Foram realizados cruzamentos controles entre linhagens da mesma espécie para verificar se a herança do tamanho do genoma corresponde à média do tamanho do genoma parental, que indica herança aditiva, sem qualquer interferência de sexo ou linhagem. Os resultados corresponderam ao esperado. Nos cruzamentos de ambas as direções entre linhagens da mesma espécie, a estimativa do genoma dos descendentes correspondem à média do genoma parental (Figura 2). Neste mesmo tipo de cruzamento, utilizando as mesmas linhagens, foram quantificados os nascimentos até 42 dias após os parentais serem fertilizados e a quantidade de nascimentos foram os mesmos dos parentais correspondentes (dados não publicados).

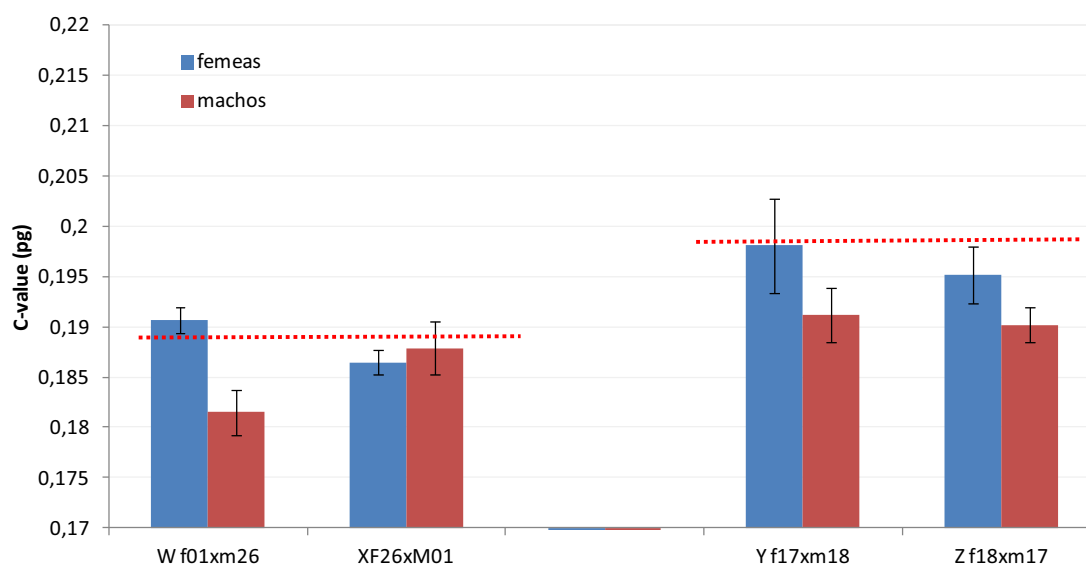


Fig.2. Distribuição do tamanho do genoma (pg) entre cruzamentos da mesma linhagem de *D. mojavensis* (W e X) e *D. arizonae* (Y e Z). Linha pontilhada vermelha corresponde a média do genoma parental correspondente a cada cruzamento.

Cruzamentos interespecíficos

Não foram obtidos descendentes nos cruzamentos entre fêmeas *D. mojavensis* AG (0,190pg) com machos de *D. arizonae*, tanto derivados da linhagem HI (0,191pg) como de LB, o que nos impediu de mensurar e comparar com os demais resultados.

A estimativa do tamanho do genoma do híbrido não difere da média do genoma parental. Porém quando se considera se o genoma do híbrido difere de um dos genomas parentais, os resultados entre *Drosophila mojavensis* (ANZA) e *D. arizonae* (HI) e o cruzamento recíproco foram os seguintes (Figura 3A):

- a) Genomas de machos e fêmeas do cruzamento A não diferem entre si; e genomas de machos e fêmeas híbridos não diferem do tamanho do genoma estimado paterno ou materno;
- b) Genomas de machos e fêmeas do cruzamento B não diferem entre si, porém os das fêmeas e machos híbridos diferem do tamanho do genoma paterno e não do materno;
- c) Genomas de fêmeas A não diferem das fêmeas de B, porém diferem dos machos B;
- d) Genomas de machos A não diferem das fêmeas de B, porém diferem dos machos B;
- e) Genomas parentais do cruzamento A são estatisticamente semelhantes, porém os parentais do cruzamento B são diferentes.

D. mojavensis (ANZA) e *D. arizonae* (LB) e recíproco (Figura 3B):

- a) Genomas de fêmeas e machos híbridos C não diferem entre si, e de machos e fêmeas C não difere do genoma materno, porém diferem do genoma paterno;
- b) Genomas de fêmeas e machos híbridos D, não diferem entre si, porém os de fêmeas D diferem do genoma materno e não do paterno. Os machos D diferem do materno e não do patern;
- c) Genomas de fêmeas e machos de C não diferem do genoma das fêmeas e machos D.

Fêmeas de *D. arizonae* (HI) e machos *D. mojavensis* (AG) - (Figura 3C)

- a) Não há diferença entre genomas de machos e fêmeas híbridos F, nem em relação aos parentais maternos ou paternos, embora o genoma materno seja diferente do genoma paterno.

Fêmeas *D. mojavensis* (AG) e machos *D. arizonae* (LB) - (Figura 3C)

- a) Genomas de fêmeas e machos híbridos não diferem entre si;
 b) Genomas de fêmeas híbridas não diferem do genoma materno e paterno, porém os machos diferem tanto do materno quanto do genoma paterno;
 c) Genomas de parentais não diferem entre si.

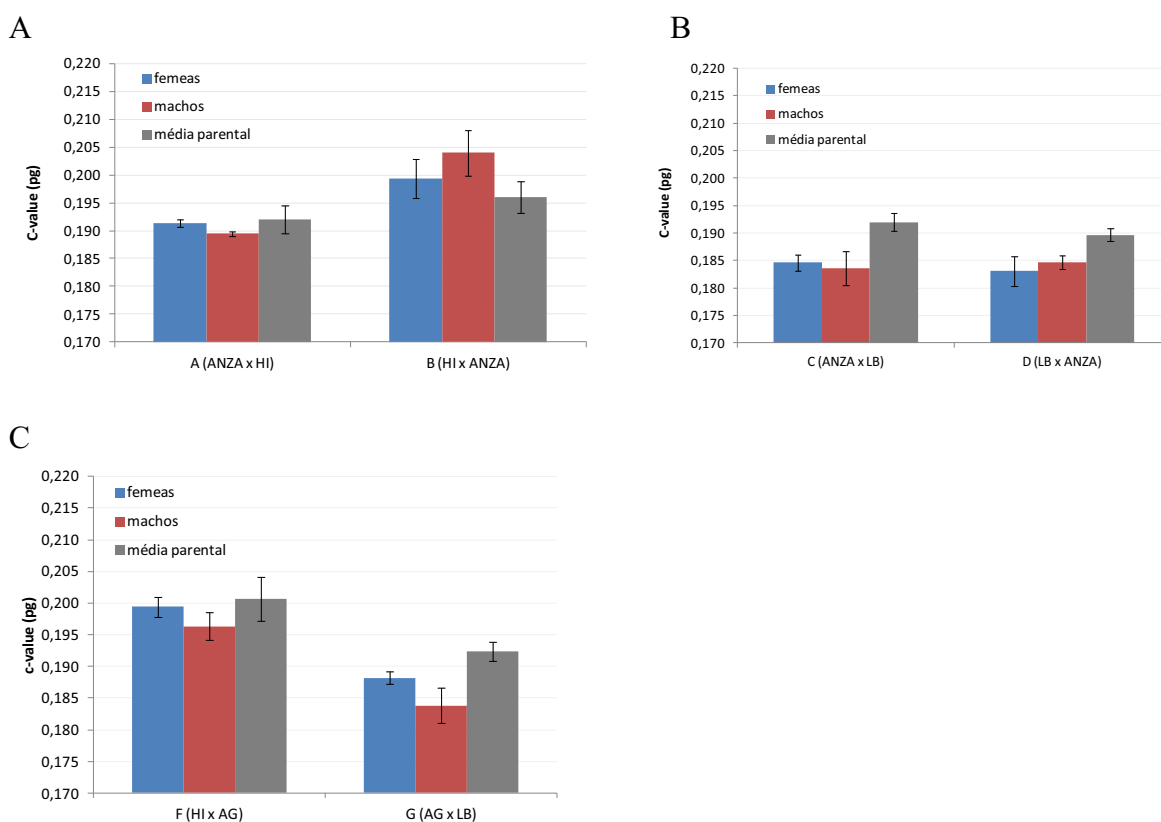


Fig. 3. Distribuição do tamanho do genoma (pg) dos cruzamentos interespecíficos de *D. mojavensis* e *D. arizonae*. (Teste Anova Two-way, bonferroni postest, 5 réplicas, letras iguas $p > 0,05$, média \pm SE), A-G: cruzamentos (fêmeas vs machos).

Pode-se notar uma relação, acima descrita, em relação ao genoma materno. Os híbridos descendentes de fêmeas de *D. arizonae* HI, apresentam estimativas de tamanho do genoma significativamente maior que os demais, tanto aqueles provenientes de cruzamentos entre fêmeas *D. arizonae* HI com machos *D. mojavensis* ANZA, quanto com machos de *D. arizonae* LB (Figura 3).

Análises independentes entre machos e fêmeas, revelam semelhança no tamanho do genoma entre fêmeas F1 dos cruzamentos A (*D.moj* ANZA x *D.arz* HI) e C (*D.moj* ANZA x *D.arz* LB) (ambos provenientes de fêmeas ANZA). Além destes, há semelhança entre os cruzamentos C (*D.moj* ANZA x *D.arz* LB), D (*D.arz* LB x *D.moj* ANZA) e G (*D.moj* AG x *D.arz* LB), porém, diferentes dos cruzamentos B (*D.arz* HI x *D.moj* ANZA) e F (*D.arz* HI x *D.moj* AG), esta diferença ocorre, pois fêmeas de *D. arizonae* (HI) estão sendo utilizadas nos cruzamentos. Entre os machos híbridos, há semelhança entre A e F e entre os cruzamentos C, D e G, novamente explicada pelo híbrido receber o genoma da linhagem HI que possui o maior genoma (Figura 3).

Dentre os fatores que podem estar relacionados à variação do tamanho do genoma, destacam-se a amplificação de elementos de transposição, poliploidização, hibridação, ou condições de estresse ambiental (ARNAULT; DUFOURNEL, 1994; VIEIRA, *et al.*, 2002; NARDON *et al.*, 2005; LEITCH *et al.*, 2008). Mas também, a diferença de estimativa do tamanho do genoma híbrido pode estar simplesmente e diretamente, relacionada com o nível de compactação da cromatina, e não com o aumento na quantidade de DNA, uma vez que o tamanho do genoma mensurado por citometria de fluxo é uma medida indireta da quantidade de DNA (VINOGRADOV, 2005).

Ou seja, o tamanho do núcleo pode ser alterado pelo nível de compactação da cromatina, o DNA pode estar mais espiralizado ou distendido, o que pode alterar a quantidade de moléculas que se intercalam no DNA, e que vão ser em seguida quantificadas. Portanto, a

diferença observada nas estimativas deste trabalho podem não representar aumento do material genético, mas sim resultante da variação do tamanho do núcleo em decorrência da configuração da cromatina.

Efeito do sexo na diferença entre tamanho dos genomas

O teste ANOVA mostra que as diferenças do tamanho do genoma entre as populações de *D. mojavensis* e *D. arizonae* estão associadas ao tamanho do genoma materno, isto é, o tamanho do genoma é moldado segundo os parentais ($p < 0,05$), particularmente pelo tamanho do genoma diferente entre os sexos das linhagens *D. arizonae* HI e *D. mojavensis* AG (Tabela 2; Tabela S2; Figura 1). Nos cruzamentos controles e interespecíficos não se verificou relação entre cruzamento e sexo do híbrido afetando a estimativa do tamanho do genoma do híbrido.

Na Tabela 2 pode-se observar os resultados das análises estatísticas das interações de p-value global, entre genoma do híbrido versus genoma parental (fêmeas e machos), sexo do híbrido F1, e entre os diferentes cruzamentos. São dois os resultados a serem destacados. Em primeiro lugar podemos notar, sobretudo, que o tamanho do genoma das fêmeas parentais (PF), influencia no tamanho do genoma híbrido F1, porém dos machos parental (PM) não exhibe interação com tamanho do genoma híbrido. Em outras palavras, há interação do tamanho do genoma da fêmea parental com o genoma do híbrido, mais precisamente, o tamanho do genoma da fêmea parental afeta o do genoma do híbrido correspondente. E em segundo, que não há diferença estatística quando comparado o tamanho do genoma híbrido em relação ao sexo do híbrido, entretanto, o tamanho do genoma é moldado com relação aos cruzamentos, sendo conseqüentemente influenciado pelo tamanho do genoma materno envolvida no cruzamento.

Tabela 2. Resultados de ANOVA do tamanho do genoma híbrido versus parentais (machos e fêmeas), sexo do híbrido F1 e cruzamentos (ANOVA mix models). Em negrito, $p > 0,05$.

	C-value Híbrido F1		
	df	F value	Pr (>F)
Parental Fêmea (PF).....	1	44,5333	>0,0001
Parental Macho (PM).....	1	0,4863	0,4863
Sexo Híbrido.....	1	1,22	0,2739
Cruzamento (Cruz.).....	9	10,328	>0,0001
PF x Cruz.....	9	0,6081	>0,0001
PF x Sexo Híbrido F1.....	1	2,8638	0,0961
Cruz. X Sexo Híbrido F1.....	9	2,0823	0,0546
PF x P M.....	1	3,3751	0,0715
PF x Cruz. X Sexo Híbrido	9	0,9697	0,4922

A diferença marcante no tamanho do genoma entre a linhagem de *D. arizonae* (HI) e as de *D. mojavensis*, pode ser uma das razões da ocorrência de isolamento pós-zigótico ser incompleto em determinados cruzamentos (MASSIE; MARKOW, 2005; REED, et al., 2008; BONO; MARKOW, 2009). A formação de um híbrido entre duas espécies diferentes implica no pareamento de dois cromossomos parentais que já divergiram. Se existem variações na compactação do DNA dos cromossomos homólogos, o pareamento pode não se fazer de forma correta e resultar em esterilidade híbrida. As variações na estimativa do tamanho do genoma observadas nos híbridos podem ser devidas a diferenças na compactação da cromatina nos núcleos das células somáticas, como também a mutações somáticas induzidas por rearranjos cromossômicos relacionados ou não com a presença de elementos de transposição. Assim, tanto efeitos germinativos como somáticos podem estar relacionados ao isolamento pós-zigótico influenciando a fertilidade ou viabilidade do híbrido.

Sem o acesso ao genoma completo de *D. arizonae* é difícil explorar fatores que possam justificar a diferença no tamanho do genoma em relação a *D. mojavensis*. Explorar o DNA satélite e elementos de transposição no genoma, o estado da cromatina nos telômeros, estrutura cromossômica das diferentes populações e composição nos cromossomos sexuais quanto à sua composição de DNA satélite, ou ainda explorar os fatores ligados à modificação

do tamanho do genoma e o impacto no valor adaptativo destes organismos, pode contribuir para explicar as diferenças encontradas do tamanho do genoma nas diferentes linhagens dessas duas espécies de *Drosophila*.

Referências

- ARNAULT, C.; DUFOURNEL, I. Genome and stresses: reactions against aggressions, behavior of transposable elements. **Genetica**, Dordrecht, v. 93, p. 149-60, 1994.
- BAACK, E. J.; WHITNEY, K. D.; RIESEBERG, L.H. Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. **New Phytologist**, Hoboken, v. 167, p. 623-30, 2005.
- BEAULIEU, J. The right stuff: evidence for an 'optimal' genome size in a wild grass population. **New Phytologist**, Hoboken, v.187, p.883-885, 2010.
- 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**, Oxford, v. 22, p. 1387-95, 2009
- BOSCO, G., et al. Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. **Genetics**, Austin, v. 177, p. 1277-90, 2007.
- BOULESTEIX, M.; WEISS, M.; BIÉMONT, C. Differences in genome size between closely related species: the *Drosophila melanogaster* species subgroup. **Molecular Biology and Evolution**, Chicago, v. 23, p. 162-7, 2006.
- CASACUBERTA, E; GONZALEZ, J. The impact of transposable elements in environmental adaptation. **Molecular Ecology**, Oxford, v. 22, n. 6, p.1503-1517, 2013.
- CAVALIER, S.; BEATON, M.J. The skeletal function of non-genic nuclear DNA: new evidence from ancient cell chimaeras. **Genetica**, Dordrecht, v. 106, p. 3-13, 1999.
- GREGORY, T. R. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. **Biological Reviews**, Hoboken, v. 76, n. 1, p. 65-101, 2001.
- GREGORY, T. R.; JOHNSTON, J. S. Genome size diversity in the family Drosophilidae. **Heredity**, London, v. 101, n. 3, p. 228-238, 2008.
- GREILHUBER, J., et al. The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents. **Annals of Botany**, Oxford, v. 95, n. 1, p. 255-60, 2005.
- LEITCH, I.J., et al. The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). **Annals of Botany**, Oxford, v. 101, n. 6, p. 805-14, 2008.
- MACHADO, C. A., et al. Multilocus nuclear sequences reveal intra- and interspecific relationships among chromosomally polymorphic species of cactophilic *Drosophila*. **Molecular Ecology**, Oxford, v. 16, n. 14, p. 3009-24, 2007.
- MARQUES, I., et al. Genome size and base composition variation in natural and experimental *Narcissus* (*Amaryllidaceae*) hybrids. **Annals of Botany**, Oxford, v.109, p.257-64, 2012.

MASSIE, K. R.; MARKOW, T. A. Sympatry, allopatry and sexual isolation between *Drosophila mojavensis* and *D. arizonae*. **Hereditas**, Lund, v. 142, p. 51-55, 2005.

NARDON, C., et al. Is genome size influenced by colonization of new environments in dipteran species? **Molecular Ecology**, Oxford, v. 14, p. 869-878, 2005.

PETROV, D. A. Evolution of genome size: new approaches to an old problem. **Trends in Genetics**, London, v. 17, n. 1, p. 23-8, 2001.

PFEILER, E., et al. Genetic, ecological and morphological differences among populations of the cactophilic *Drosophila mojavensis* from southwestern USA and northwestern Mexico, with descriptions of two new subspecies. **Journal of Natural History**, Abingdon, v. 43, p. 923-938, 2009.

REED, L. K.; LAFLAMME, B. A.; MARKOW, T. A. Genetic architecture of hybrid male sterility in *Drosophila*: analysis of intraspecies variation for interspecies isolation. **PloS one**, San Francisco, v. 3, n. 8, p. e3076-e3076, 2008.

REED, L. K.; MARKOW, T. A. Early events in speciation: polymorphism for hybrid male sterility in *Drosophila*. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 101, n. 24, p. 9009-12, 2004.

RUIZ, A.; HEED, W. B.; WASSERMAN, W. Evolution of the Mojavensis Cluster of Cactophilic *Drosophila* with Descriptions of Two New Species. **Journal of Heredity**, Washington, v. 81, p. 30-42, 1990.

SANMIGUEL, P.; BENNETZEN, J. L. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. **Annals of Botany**, Oxford, v. 82, p. 37-44, 1998.

SONG, K., et al. Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 92, p. 7719-23, 1995.

VIEIRA, C., et al. Evolution of genome size in drosophila. Is the invader's genome being invaded by transposable elements? **Molecular Biology and Evolution**, Chicago, v. 19, p. 1154-1161, 2002.

VINOGRADOV, A. E. Genome size and chromatin condensation in vertebrates. **Chromosoma**, Berlin, v. 113, p. 362-9, 2005.

WASSERMAN, M.; KOEPFER, H. R. Character Displacement for Sexual Isolation Between *Drosophila mojavensis* and *Drosophila arizonensis*. **Evolution**, Lancaster, v. 3, p. 812-812, 1977.

MATERIAL SUPLEMENTAR

Table S1. Medidas do tamanho do genoma (C-value) das cinco réplicas entre fêmeas e machos, das linhagens de *D. mojavensis* e *D. arizonae*, descendentes F1 dos cruzamentos entre linhagens da mesma espécie e híbridos interespecíficos, média e erro padrão das réplicas.

	Fêmeas				Machos				Fêmeas + Machos						
	réplicas		erro padrão		réplicas		erro padrão		média		erro padrão				
	média	erro padrão	média	erro padrão	média	erro padrão	média	erro padrão	média	erro padrão	média	erro padrão			
<i>D.moj</i> (ANZA)	0,194	0,188	0,183	0,188	0,192	0,189	0,184	0,178	0,184	0,183	0,182	0,182	1,12E-03	0,186	1,54E-03
<i>D.moj</i> (AG)	0,193	0,192	0,192	0,187	0,185	0,190	0,196	0,196	0,194	0,185	0,187	0,191	2,28E-03	0,191	1,36E-03
<i>D.moj</i> (CI)	0,188	0,187	0,183	0,203	0,188	0,190	0,205	0,187	0,181	0,193	0,181	0,189	4,42E-03	0,190	2,63E-03
<i>D.arz</i> (HI)	0,209	0,217	0,221	0,207	0,195	0,210	0,186	0,194	0,190	0,202	0,202	0,195	3,10E-03	0,202	3,60E-03
<i>D.arz</i> (LB)	0,199	0,195	0,201	0,195	0,195	0,197	0,195	0,192	0,198	0,192	0,198	0,195	1,36E-03	0,196	9,29E-04
W	0,189	0,193	0,187	0,191	0,194	0,191	0,176	0,179	0,180	0,190	0,182	0,181	2,23E-03	0,186	1,95E-03
X	0,188	0,184	0,183	0,190	0,188	0,186	0,190	0,183	0,181	0,192	0,194	0,188	2,62E-03	0,187	1,39E-03
Y	0,212	0,188	0,205	0,190	0,195	0,198	0,198	0,195	0,190	0,190	0,183	0,191	2,68E-03	0,195	2,79E-03
Z	0,197	0,187	0,200	0,201	0,191	0,195	0,184	0,194	0,194	0,190	0,189	0,190	1,74E-03	0,193	1,78E-03
A	0,192	0,192	0,193	0,190	0,190	0,191	0,190	0,190	0,188	0,189	0,189	0,189	4,09E-04	0,190	4,88E-04
B	0,206	0,207	0,189	0,193	0,202	0,199	0,200	0,218	0,208	0,200	0,194	0,204	4,02E-03	0,202	2,63E-03
C	0,191	0,184	0,183	0,183	0,183	0,185	0,178	0,180	0,180	0,194	0,187	0,184	3,07E-03	0,184	1,62E-03
D	0,176	0,178	0,186	0,187	0,189	0,183	0,182	0,188	0,187	0,182	0,184	0,185	1,19E-03	0,184	1,39E-03
F	0,202	0,199	0,194	0,202	0,199	0,199	0,193	0,195	0,205	0,193	0,195	0,196	2,17E-03	0,198	1,37E-03
G	0,188	0,186	0,192	0,187	0,188	0,188	0,181	0,178	0,179	0,191	0,191	0,184	2,84E-03	0,186	1,60E-03
<i>D.moj</i> (ANZA, AG, CI)						0,190						0,188	1,90E-03	0,189	1,15E-03
<i>D.arz</i> (HI, LB)						0,203						0,195	1,60E-03	0,199	1,59E-03

Table S2. ANOVA das médias do tamanho do genoma híbrido em relação ao genoma parental (fêmeas x machos). ANOVA 1 way, tukey's (95% IC) multicomparação; coluna 5: linhagem parental que diferente do genoma do híbrido correspondente $p < 0,05$.

Híbridos (F*M)	DF	F	P	P<0,05	BP*
A (ANZA x HI)	2	4,945	0,0187	D.arz (HI)	*
B (HI x ANZA)	2	18,61	<0,0001	D.moj (ANZA)	***
C (ANZA x LB)	2	8,262	0,0026	D.arz (LB)	ns
D (LB x ANZA)	2	35,12	<0,0001	D.arz (LB)	***
F (HI x AG)	2	11,36	>0,0001	D. arz (HI)	***
G (AG x LB)	2	7,884	0,0031	D.arz (LB)	ns

BP – * diferença entre parentais $p < 0,05$; ns – não significante.

DISCUSSÃO GERAL

DISCUSSÃO GERAL

A abundância de TEs nos genomas (ARENSBURGER, et al., 2011), sua capacidade mutacional (CALLINAN; BATZER, 2006), interferência na regulação de genes hospedeiros (JORDAN et al., 2003), e resposta a estresse ambiental (OHTSUBO et al. 2005) tem fundamentado hipóteses de que os TEs não somente podem gerar polimorfismos favoráveis à adaptação, mas também promoverem a especiação.

Diversos autores têm proposto que explosões de transposição decorrentes do estresse genômico, causado pela hibridação, podem levar a uma reestruturação do genoma, como resultado de rearranjos cromossômicos e recombinações ectópicas, que podem resultar em especiação (FONTDEVILA, 2005; HEDGES; DEININGER, 2007; MICHALAK, 2009; REBOLLO, et al., 2010). Alguns estudos sustentam essa hipótese, dentre eles, o de Dobigny et al. (2004) que relatam a ocorrência de uma significativa atividade de TEs, particularmente transposição maciça do retrotransposon LINE-1, durante o período de radiação do gênero *Taterillus* (gerbilo).

A relação entre expressão de TEs e provável contribuição para o processo de divergência que leva à especiação pode ser exemplificada no caso específico do elemento *I-like*, tratado no Capítulo 1. Esse elemento está, neste trabalho, pela primeira vez sendo amplamente caracterizado fora do grupo *melanogaster*. As análises filogenéticas mostram que as sequências *I-like* de *D. mojavensis* são intimamente aparentadas à sequência *I* ativa de *D. melanogaster* (fator I) e de outras espécies do mesmo subgrupo. Contudo, duas evidências levaram-nos a concluir que as sequências de *D. mojavensis* pertencem a uma família diferente daquela de *D. melanogaster*: 1) o grau de divergência média (62%) entre essas sequências é maior do que o esperado entre sequências de TEs da mesma família, e 2) as sequências de *D. mojavensis* se agrupam com as de *D. virilis* em um clado fortemente

suportado, concordante com a filogenia das espécies. Consequentemente concluímos que as sequências *I* encontradas nas espécies do grupo *repleta*, aqui estudadas, *D. mojavensis* e *D. virilis*, formam uma família *I*-like exclusiva do subgênero *Drosophila*.

As análises de expressão desse elemento nas espécies parentais, *D. mojavensis* e *D. arizonae*, e seus híbridos, sugerem que o elemento *I*-like possa ter algum papel no isolamento pós-zigótico diferencial entre linhagens das duas espécies. Sua expressão é significativamente maior em híbridos provenientes do cruzamento entre fêmeas *D. arizonae* e machos *D. mojavensis*, cujos machos híbridos são estéreis (RUIZ et al., 1990).

Os elementos *I* de *D. melanogaster* e de *D. mojavensis* são filogeneticamente muito próximos e apresentam expressão similar em tecido germinal das fêmeas. Hibridizações *in situ* em ovários de *D. melanogaster* de linhagens RSF (nos quais a transposição não foi detectada) revelam que a expressão do fator *I* ocorre nas células *nurse* e que vias de piRNA controlam sua atividade pós-transcricionalmente. Por outro lado, em linhagens SF dessa espécie, que apresentam ovários disgênicos, a transposição ocorre nos oócitos indicando perda desse controle nas células *nurse* (CHAMBEYRON et al, 2008).

Em nosso estudo, o elemento *I*-like foi detectado somente em células *nurse*, indicando que este possa estar sendo regulado pelos mesmos mecanismos que em *D. melanogaster*. Contudo, não há na literatura informações sobre os locais e nem mesmo sobre as vias de controle dos elementos *I* em testículos. Detectamos transcritos *I*-like por hibridização *in situ* na região dos testículos onde se formam os espermátocitos primários dos machos, sugerindo que ao menos o controle transcricional não está atuando. Este resultado, associado à expressão mais elevada em híbridos B (fêmeas *D. arizonae* HI x machos *D. mojavensis* ANZA), detectada por RT-qPCR (Capítulo 3), pode ser uma indicação de que este elemento esteja implicado na esterilidade do macho híbrido característica da direção de cruzamento fêmeas *D. arizonae* vs machos *D. mojavensis*.

As análises de expressão global dos TEs por meio de RNA-Seq (Capítulo 2) revelaram um perfil de variação espécie e TE específico. Elementos pertencentes a diferentes famílias apresentaram expressão variável entre as duas espécies, em diferentes linhagens e tecidos. Ocorreram aumentos significantes de expressão de TEs tanto nos híbridos A (*Copia1*, Homo7) quanto nos híbridos B (*Gtwin*), porém não identificamos uma explosão global de expressão dos TEs. Nossos resultados de expressão do mobiloma mostraram que nos híbridos há uma predominância de TEs cuja expressão está em grande parte conservada (54 a 56%) ou apresenta herança aditiva (20% e 13%), indicando regulação nos híbridos semelhante aos parentais.

As análises de expressão de alguns elementos via RT-qPCR (Capítulo 3) também mostraram variações entre parentais, híbridos e tecidos analisados. O elemento *Oswaldo*, por exemplo, que em híbridos de cruzamentos entre *D. buzzatii* e *D. koepferae* apresenta altas taxas de transposição (LABRADOR; FONTDEVILLA, 1994), tem baixa expressão em células somáticas e germinativas, com exceção da linhagem AG de *D. mojavensis*, que tem expressão cerca de três vezes maior em células somáticas. O elemento *Copia* também apresenta expressão heterogênea, sendo altamente expresso em tecidos germinativos e somáticos em uma linhagem de *D. arizonae* (LB) enquanto que na linhagem Anza Borrego de *D. mojavensis* é expresso apenas em testículos, em nível semelhante à expressão em linhagem somática de *D. arizonae* LB. Esta variação na expressão corrobora resultados obtidos em plantas e também em *Drosophila* com este elemento. Em linhagens híbridas recentes de cana-de-açúcar, o padrão de transcrição difere entre tecidos e famílias de retrotransposons LTR; por exemplo, os elementos *Ivana* e *Reina* apresentam distintos padrões e a expressão de *Ivana6* é mais intensa que *Ivana1*, este último com maior atividade transcricional na gema, enquanto *Ivana6* é mais ativo em folhas (DOMINGUES et al., 2012). Em *D. simulans* e *D. melanogaster* ocorre expressão tecido-específica do

retrotransposon 412, que se expressa em testículos, mas não em ovários (BORIE et al., 2002).

Reed e Markow (2004) discutem os motivos que poderiam causar a variação de infertilidade e esterilidade de híbridos entre fêmeas de *D. mojavensis* e machos *D. arizonae*. Estes autores consideram que a esterilidade não parece ter base simples. Essa variação pode ser devida à associação de vários locos relacionados ao armazenamento de esperma, recuperação do armazenamento ou capacidade de penetrar a micrópila, como também de apenas um gene de grande efeito, mas que exibe penetrância incompleta. Por outro lado, fatores associados ao cromossomo IV foram associados à esterilidade no cruzamento recíproco, entre fêmeas *D. arizonae* e machos *D. mojavensis* (PANTAZIDIS et al., 1993). Contudo, esses fatores não foram analisados no cruzamento recíproco e não está claro se afetam ou não habilidade reprodutiva dentro e entre populações de *D. mojavensis* (REED; MARKOW, 2004). Os TEs com expressão diferencial em testículos dos híbridos não poderiam ser um desses fatores?

No capítulo 4 constatamos variações nas estimativas do tamanho do genoma entre a linhagem HI de *D. arizonae* e as de *D. mojavensis*, e que o tamanho do genoma das fêmeas parentais influenciam o tamanho do genoma do híbrido F1. As variações observadas podem ser uma das razões do isolamento pós-zigótico incompleto entre essas espécies.

Variações no tamanho do genoma podem ser explicadas como resultado de vários fatores atuando separada ou conjuntamente, tais como, i) duplicações genômicas, polissomia, duplicações e deleções locais; ii) transposição, que determina inserções, excisões e proliferação de TEs; iii) expansão ou contração de microssatélites e iv) expansão ou contração da heterocromatina (revisão PETROV, 2001). Estes fatores podem produzir pressão mutacional levando ao aumento ou diminuição do tamanho do genoma, e em híbridos sua atuação pode ser intensificada pelo estresse genômico. Focando os TEs,

sua proliferação pode também provocar modificações da cromatina, como demonstrado em híbridos disgênicos de *Drosophila* (DIMITRI, 1997; DIMITRI et al., 1997; SENTMANAT; ELGIN, 2012). Sentmanat e Elgin (2012) mostraram que o transposon 1360 influencia a cromatina na região em que está inserido silenciando um gene repórter no domínio heterocromático, e assim influenciando a montagem da heterocromatina, e que esse fenômeno poderia ser efetivo para um conjunto mais amplo de TEs. Portanto, essas alterações da cromatina, mediadas ou não pelos TEs, podem também afetar as mensurações do tamanho do genoma (VIEIRA, et al., 2002; NARDON et al., 2005). O acesso transcricional aos TEs, e posterior mobilização e quantidade no genoma também pode estar relacionado com a disposição destes nos cromossomos, em regiões heterocromáticas ou eucromáticas (VAURY et al., 1989; MIKLOS; COTSELL, 1990; VIEIRA et al., 2002). Em síntese, as variações do tamanho genômico aqui relatadas podem ser decorrentes da atuação de múltiplos fatores, dentre eles a própria proliferação dos TEs, ou sua atuação na modificação da conformação da cromatina.

Muitas especulações são feitas sobre o impacto dos TEs no processo de especiação e encontrar um bom modelo de estudo nem sempre é fácil. Neste trabalho, não apenas reforçamos que *D. mojavensis/D. arizonae* constituem um par de espécies apropriado para estudos de especiação, mas também demonstramos sua adequação para se explorar a dinâmica de expressão dos elementos de transposição em genomas híbridos.

CONCLUSÕES

CONCLUSÕES

O presente estudo permitiu estabelecer as seguintes conclusões:

- 1- Há modificações na expressão de elementos de transposição específicos em híbridos entre *Drosophila mojavensis* e *D. arizonae*, mas não uma explosão global de expressão.
- 2- Um retrotransposon NLTR da superfamília I identificado no genoma sequenciado de *D. mojavensis*, intimamente aparentado ao fator *I* de *D. melanogaster*, mas não altamente similar a ele, foi considerado como pertencente a uma nova família *I* exclusiva do subgênero *Drosophila*.
- 3- A expressão do elemento *I-like*, detectada em ovários somente em células *nurse* de *D. mojavensis* e *D. arizonae*, como também de seus híbridos recíprocos, indica que esse elemento está controlado nesse órgão e que sua regulação segue o mesmo padrão do elemento *I* de *D. melanogaster*. Contudo, a expressão aumentada em testículos nos híbridos do cruzamento de fêmeas *D. arizonae* e machos *D. mojavensis*, sugere que esse elemento possa estar implicado na esterilidade do híbrido nessa direção de cruzamento.
- 4- Os retrotransposons *Copia* e *Gtwin* estão diferencialmente superexpressos em híbridos A e B, respectivamente, sendo necessário analisar se são transposicionalmente ativos e quais mecanismos estão envolvidos no seu silenciamento nos parentais.
- 5- O tamanho do genoma híbrido é moldado dependendo da linhagem de *D. mojavensis* ou *D. arizonae* utilizada como parental, sendo este influenciado pelo tamanho do genoma materno envolvida no cruzamento.

REFERÊNCIAS

REFERÊNCIAS

- ARENSBURGER, P. et al. The mosquito *Aedes aegypti* has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs. **BMC Genomics**, London, v. 12, 2011. Disponível em: <<http://www.biomedcentral.com/1471-2164/12/606>>. Acesso em: 10 fev. 2014.
- ARKHIPOVA, I. R.; RODRIGUEZ, F. Genetic and epigenetic changes involving (retro)transposons in animal hybrids and polyploids. **Cytogenetic and Genome Research**, Basel, v. 140, n. 2/4, p. 295-311, 2013.
- BIÉMONT, C.; VIEIRA, C. Genetics: junk DNA as an evolutionary force. **Nature**, London, v. 443, p. 521-524, 2006.
- BINGHAM, P. M.; KIDWELL, M. G.; RUBIN, G. M. The molecular-basis of P-M hybrid dysgenesis- the role of the P-element, a P-strain-specific transposon family. **Cell**, Cambridge, v. 29, n. 3, p. 995-1004, 1982.
- BLACKMAN, R. K. et al. Mobilization of hobo elements residing within the decapentaplegic gene complex: suggestion of a new hybrid dysgenesis system in *Drosophila melanogaster*. **Cell**, Cambridge, v. 49, n. 4, p. 497-505, 1987.
- BORIE, N. et al. Tissue-specificity of 412 retrotransposon expression in *Drosophila simulans* and *D. melanogaster*. **Heredity**, London, v. 89, n. 4, p. 247-252, 2002.
- CALLINAN, P. A.; BATZER, M. A. Retrotransposable elements and human disease. **Genome Dynamics**, Basel, v. 1, p. 104-115, 2006.
- CARARETO, C. M.; Tropical Africa as a cradle for horizontal transfers of transposable elements between species of the genera *Drosophila* and *Zaprionus*. **Mobile Genetic Elements**, London, v. 1, n. 3, p. 179-186, 2011.
- CARARETO, C. M.; HERNANDEZ, E. H.; VIEIRA, C. Genomic regions harboring insecticide resistance-associated Cyp genes are enriched by transposable element fragments carrying putative transcription factor binding sites in two sibling *Drosophila* species. **Gene**, Amsterdam, v. 537, p. 93-99, 2014.
- CARBONE, L. et al. Centromere remodeling in *Hoolock leuconedys* (*Hylobatidae*) by a new transposable element unique to the Gibbons. **Genome Biology and Evolution**, Oxford, v. 4, n. 7, p. 760-770, 2012.
- CHAMBEYRON, S. et al. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 105, n. 39, p. 14964-14969, 2008.
- CORDAUX, R.; BATZER, M. A. The impact of retrotransposons on human genome evolution. **Nature Reviews Genetics**, London, v. 10, n. 10, p. 691-703, 2009.

CURCIO, M.; DERBYSHIRE, K. The outs and ins of transposition: from mu to kangaroo. **Nature Reviews. Molecular Cell Biology**, London, v. 4, p. 865-877, 2003.

DE ALMEIDA, L. M.; CARARETO, C. M. A. Multiple events of horizontal transfer of the Minos transposable element between *Drosophila* species. **Molecular phylogenetics and evolution**, San Diego, v. 35, n. 3, p. 583-94, 2005.

DIMITRI, P. Constitutive heterochromatin and transposable elements in *Drosophila melanogaster*. **Genetica**, Dordrecht, v. 100, n. 1-3, p. 85-93, 1997.

_____. et al. High genetic instability of heterochromatin after transposition of the LINE-like I factor in *Drosophila melanogaster*. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 94, p. 8052-8057, 1997.

DOBIGNY, G. et al. LINE-1 amplification accompanies explosive genome repatterning in rodents. **Chromosome Research**, Oxford, v. 12, n. 8, p. 787-793, 2004.

DOMINGUES, D. S. et al. Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. **BMC Genomics**, London, v. 13, p. 137, 2012. Disponível em: <<http://www.biomedcentral.com/1471-2164/13/137>>. Acesso em: 21 fev. 2014.

DUVAL-VALENTIN, G.; MARTY-COINTIN, B.; CHANDLER, M. Requirement of *IS911* replication before integration defines a new bacterial transposition pathway. **EMBO Journal**, Oxford, v. 23, p. 3897-3906, 2004.

EVGEN'EV, M. et al. Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 94, p. 196-201, 1997.

FABLET, M.; VIEIRA, C. Evolvability, epigenetics and transposable elements. **Biomolecular Concepts**, Berlin, v. 2, p. 333-341, 2011.

FESCHOTTE, C.; JIANG, N.; WESSLER, S. R. Plant transposable elements: Where genetics meets genomics. **Nature Reviews Genetics**, London, v. 3, n. 5, p. 329-341, 2002.

_____. PRITHAM, E. J. DNA transposons and the evolution of eukaryotic genomes. **Annual Review of Genetics**, Palo Alto, v. 41, p. 331-368, 2007.

FINNEGAN, D. J. Eukaryotic transposable elements and genome evolution. **Trends in Genetics**, London, v. 5, p. 103-107, 1989.

FONTDEVILA, A. Hybrid genome evolution by transposition. **Cytogenetic and Genome Research**, Basel, v. 110, n. 1-4, p. 49-55, 2005.

GOODWIN, T.; POULTER, R. A new group of tyrosine recombinase-encoding retrotransposons. **Molecular Biology and Evolution**, Chicago, v. 21, p. 746-759, 2004.

- GUERREIRO, M. P. G.; FONTDEVILA, A. The evolutionary history of *Drosophila buzzatii*. XXXVI. Molecular structural analysis of Osvaldo retrotransposon insertions in colonizing populations unveils drift effects in founder events. **Genetics**, Austin, v. 175, n. 1, p. 301-310, 2007.
- HAN, J. S.; BOEKE, J. D. A highly active synthetic mammalian retrotransposon. **Nature**, London, v. 429, n. 6989, p. 314-8, 2004.
- HEDGES, D. J.; DEININGER, P. L. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. **Mutation Research**, Amsterdam, v. 616, n. 1/2, p. 46-59, 2007.
- JORDAN, I. K. et al. Origin of a substantial fraction of human regulatory sequences from transposable elements. **Trends in Genetics**, London, v. 19, n. 2, p. 68-72, 2003.
- KELLEHER, E. S.; EDELMAN, N. B.; BARBASH, D. A. *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. **PLoS Biol**, Cambridge, v. 10, n. 11, p. e1001428, 2012. Disponível em: <<http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1001428>>. Acesso em: 18 dez. 2013.
- KIDWELL, M. G.; LISCH, D. Transposable elements as sources of variation in animals and plants. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 94, n. 15, p. 7704-7711, 1997.
- _____. NOVY, J. B. Sterility resulting from gonadal dysgenesis in the P-M system. **Genetics**, Austin, v. 92, n. 4, p. 1127-1140, 1979.
- LABRADOR, M.; FONTDEVILA, A. High transposition rates of Osvaldo, a new *Drosophila buzzatii* retrotransposon. **Molecular & General Genetics**, Berlin, v. 245, n. 6, p. 661-674, 1994.
- _____. et al. Interspecific hybridization increases transposition rates of Osvaldo. **Molecular Biology and Evolution**, Chicago, v. 16, n. 7, p. 931-937, 1999.
- LADEVEZE, V. et al. Hobo transposons causing chromosomal breakpoints. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 265, n. 1402, p. 1157-1159, 1998.
- LAI, J. et al. Gene movement by Helitron transposons contributes to the haplotype variability of maize. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 102, p. 9068-9073, 2005.
- LYTTLE, T. W.; HAYMER, D. S. The role of the transposable element hobo in the origin of endemic inversions in wild populations of *Drosophila melanogaster*. **Genetica**, Dordrecht, v. 86, n. 1-3, p. 113-126, 1992.

MARIÑO-RAMÍREZ, L. et al. Transposable elements donate lineage-specific regulatory sequences to host genomes. **Cytogenetics and Genome Research**, Basel, v. 110, n. 1-4, p. 333-41, 2005.

MARTIN, C. et al. Molecular analysis of instability in flower pigmentation of *Antirrhinum majus*, following isolation of the pallida locus by transposon tagging. **EMBO Journal**, Oxford, v. 4, n. 7, p. 1625-1630, 1985.

MASSIE, K. R.; MARKOW, T. A. Sympatry, allopatry and sexual isolation between *Drosophila mojavensis* and *D. arizonae*. **Hereditas**, Lund, v. 142, p. 51-55, 2005.

MCCLINTOCK, B. Controlling elements and the gene. **Cold Spring Harbor Symposia on Quantitative Biology**, New York, v. 21, p. 197-216, 1956.

_____. The significance of responses of the genome to challenge. **Science**, Washington, v. 226, n. 4676, p. 792-801, 1984.

METCALFE, C. J. et al. Genomic instability within centromeres of interspecific marsupial hybrids. **Genetics**, Austin, v. 177, n. 4, p. 2507-2517, 2007.

MICHALAK, P. Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. **Heredity**, London, v. 102, n. 1, p. 45-50, 2009.

MIKLOS, G. L.; COTSELL, J. N. Chromosome structure at interfaces between major chromatin types: alpha- and beta-heterochromatin. **BioEssays**, Cambridge, v. 12, n. 1, p. 1-6, 1990.

MISRA, S.; RIO, D. C. Cytotype control of *Drosophila* P element transposition: the 66 kd protein is a repressor of transposase activity. **Cell**, Cambridge, v. 62, n. 2, p. 269-84, 1990.

MORGANTE, M. et al. Gene duplication and exon shuffling by *Helitron*-like transposons generate intraspecies diversity in maize. **Nature Genetics**, New York, v. 37, p. 997-1002, 2005.

NARDON, C. et al. Is genome size influenced by colonization of new environments in dipteran species? **Molecular Ecology**, Oxford, v. 14, n. 3, p. 869-878, 2005.

NAVEIRA, H.; FONTDEVILA, A. The evolutionary history of *Drosophila buzzatii*. 9. High-frequencies of new chromosome rearrangements induced by introgressive hybridization. **Chromosoma**, Berlin, v. 91, n. 2, p. 87-94, 1985.

O'DONNELL, K. A.; BOEKE, J. D. Mighty piwis defend the germline against genome intruders. **Cell**, Cambridge, v. 129, n. 1, p. 37-44, 2007.

OHTSUBO, Y. et al. High-temperature-induced transposition of insertion elements in *Burkholderia multivorans* ATCC 17616. **Applied and Environmental Microbiology**, Washington, v. 71, n. 4, p. 1822-1828, 2005.

OLIVER, K. R.; GREENE, W. K. Transposable elements: powerful facilitators of evolution. **BioEssays**, Cambridge, v. 31, n. 7, p. 703-714, 2009.

ORSI, G. A. et al. *Drosophila* I-R hybrid dysgenesis is associated with catastrophic meiosis and abnormal zygote formation. **Journal of Cell Science**, London, v. 123, n. 20, p. 3515-3524, 2010.

PANTAZIDIS, A. C.; GALANOPOULOS, V. K.; ZOUROS, E. An autosomal factor from *Drosophila arizonae* restores normal spermatogenesis in *Drosophila mojavensis* males carrying the *D. arizonae* Y chromosome. **Genetics**, Austin, v. 134, n. 1, p. 309-318, 1993.

PARDUE, M.-L.; DEBARYSHE, P. G. Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. **Annual Review of Genetics**, Palo Alto, v. 37, p. 485-511, 2003.

PETROV, D. A. Evolution of genome size: new approaches to an old problem. **Trends in Genetics**, London, v. 17, n. 1, p. 23-28, 2001.

RAY, D. A. et al. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. **Genome Research**, Cold Spring Harbor, v. 18, n. 5, p. 717-728, 2008.

REBOLLO, R. et al. Jumping genes and epigenetics: towards new species. **Gene**, Amsterdam, v. 454, n. 1/2, p. 1-7, 2010.

REED, L. K.; MARKOW, T. A. Early events in speciation: polymorphism for hybrid male sterility in *Drosophila*. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 101, n. 24, p. 9009-9012, 2004.

_____; NYBOER, M.; MARKOW, T. A. Evolutionary relationships of *Drosophila mojavensis* geographic host races and their sister species *Drosophila arizonae*. **Molecular Ecology**, Oxford, v. 16, n. 5, p. 1007-1022, 2007.

REISS, D. et al. Hoppel, a P-like element without introns: a P-element ancestral structure or a retrotranscription derivative? **Molecular Biology and Evolution**, Chicago, v. 20, n. 6, p. 869-79, 2003.

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

SARKAR, A. et al. Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences. **Molecular Genetics and Genomics**, Heidelberg, v. 270, n. 2, p. 173-80, 2003.

SCHAACK, S.; GILBERT, C.; FESCHOTTE, C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. **Trends in Ecology & Evolution**, Amsterdam, v. 25, n. 9, p. 537-546, 2010.

SENTMANAT, M. F.; ELGIN, S. C. Ectopic assembly of heterochromatin in *Drosophila melanogaster* triggered by transposable elements. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 109, n. 35, p. 14104-14109, 2012.

SOKOLOVA, M. I. et al. [Morphologic and molecular manifestations of hybrid dysgenesis in ontogenesis of *Drosophila virilis*]. **Ontogenez-Russian Journal of Developmental Biology**, Moskva, v. 41, n. 6, p. 451-4, 2010 2010.

TAMURA, K.; SUBRAMANIAN, S.; KUMAR, S. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. **Molecular Biology and Evolution**, Chicago, v. 21, n. 1, p. 36-44, 2004.

THORNBURG, B. G.; GOTEA, V.; MAKALOWSKI, W. Transposable elements as a significant source of transcription regulating signals. **Gene**, Amsterdam, v. 365, p. 104-110, 2006.

UNGERER, M. C.; STRAKOSH, S. C.; ZHEN, Y. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. **Current Biology**, London, v. 16, n. 20, p. 872-873, 2006.

_____; STRAKOSH, S. C.; STIMPSON, K. M. Proliferation of Ty3/gypsy-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. **BMC Biology**, London, v. 7, p. 40-40, 2009. Disponível em: <<http://www.biomedcentral.com/1741-7007/7/40>>. Acesso em: 25 jan. 2012.

VAN DE LAGEMAAT, L. N. et al. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. **Trends in Genetics**, London, v. 19, n. 10, p. 530-536, 2003.

VAURY, C.; BUCHETON, A.; PELISSON, A. The beta heterochromatic sequences flanking the I elements are themselves defective transposable elements. **Chromosoma**, Berlin, v. 98, n. 3, p. 215-224, 1989.

VIEIRA, C. et al. Evolution of genome size in drosophila. Is the invader's genome being invaded by transposable elements? **Molecular Biology and Evolution**, Chicago, v. 19, n. 7, p. 1154-1161, 2002.

_____. et al. Wake up of transposable elements following *Drosophila simulans* worldwide colonization. **Molecular Biology and Evolution**, Chicago, v. 16, n. 9, p. 1251-1255, 1999.

WALLAU, G. L.; ORTIZ, M. F.; LORETO, E. L. Horizontal transposon transfer in eukarya: detection, bias, and perspectives. **Genome Biology and Evolution**, Oxford, v. 4, n. 8, p. 689-699, 2012.

WASSERMAN, M.; KOEPFER, H. R. Character displacement for sexual isolation between *Drosophila mojavensis* and *Drosophila arizonensis*. **Evolution**, Lancaster, v. 31,

n. 4, p. 812-823, 1977.

WEIL, C. F. Too many ends: aberrant transposition. **Genes & Development**, Cold Spring Harbor, v. 23, n. 9, p. 1032-1036, 2009.

WICKER, T. et al. A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics**, London, v. 8, n. 12, p. 973-982, 2007.

ZHANG, J.; ZHANG, F.; PETERSON, T. Transposition of reversed Ac element ends generates novel chimeric genes in maize. **Plos Genetics**, San Francisco, v. 2, n. 10, p. 1535-1540, 2006. Disponível em:

<<http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.0020164>>.

Acesso em: 05 fev. 2014.