

Bianca Manfré

**Potencial adaptativo de transcritos quiméricos gene-elemento de
transposição em *Drosophila melanogaster* e *D. simulans***

São José do Rio Preto
2023

Bianca Manfré

**Potencial adaptativo de transcritos quiméricos gene-elemento de
transposição em *Drosophila melanogaster* e *D. simulans***

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

Financiadora: CAPES
CNPq: Proc. 308020/2021-9

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

São José do Rio Preto
2023

M276p Manfré, Bianca.
Potencial Adaptativo de transcritos quiméricos gene-elemento de
transposição em *Drosophila melanogaster* e *D. simulans* / Bianca
Manfré. -- São José do Rio Preto, 2023
120 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp),
Instituto de Biociências, Letras e Ciências Exatas, São José do Rio
Preto
Orientadora: Claudia Márcia Aparecida Carareto

1. ChimeraTE. 2. atividade de TEs induzida por estresse. 3.
plasticidade genômica. 4. plasticidade transcriptômica. 5. potencial
adaptativo de TEs. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de
Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Bianca Manfré

**Potencial adaptativo de transcritos quiméricos gene-elemento de
transposição em *Drosophila melanogaster* e *D. simulans***

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

Financiadora: CAPES

Comissão Examinadora

Prof^a. Dr^a. Claudia Marcia Aparecida Carareto
UNESP – Câmpus de São José do Rio Preto
Orientadora

Prof^a. Dr^a. Maura Helena Manfrin
Universidade de São Paulo

Prof^a. Dr^a. Nathalia de Setta Costa
Universidade Federal do ABC

São José do Rio Preto
5 de dezembro de 2023

A quem me tornarei.

AGRADECIMENTOS

À minha orientadora, Professora Claudia, os meus mais sinceros agradecimentos por tudo que fez e tem feito por mim e todos seus alunos. Agradeço pelo voto de confiança, pela oportunidade, pelos ensinamentos, pela paciência e pela amizade. Minha entrada em 2019 para esse laboratório, que é uma família, foi só o começo de uma jornada de descobrimentos. Foi também o começo da minha paixão pela pesquisa, sentimento esse que é um espelho de tua devoção para com a ciência, o ensino, a ética e mais importante para com a verdade. Minha admiração só cresce.

Aos meus pais Vera e Adilson, pela crença em meu potencial e pelo apoio incondicional em todas as fases de minha vida. Estiveram ao meu lado me incentivando ao longo de toda essa trajetória: desde a concepção do projeto, no processo seletivo, aos finais de semanas e madrugadas trabalhando, até a finalização de toda a Dissertação. Obrigada por me proporcionarem a estabilidade necessária para minha total dedicação nessa etapa. Tenham certeza de que seus esforços não serão em vão. Meu amor por vocês é um sentimento que não cabem em palavras, somente o sentir permite a compreensão.

Aos meus outros familiares que também estiveram me apoiando ao longo dessa jornada, à minha irmã Karoll, à minha Tia Vanda e ao meu Tio Júnior. A presença de vocês em minha vida é um dos maiores presentes que o acaso poderia ter me dado.

Agradeço também aos meus amigos pessoais que sempre se fizeram presentes em minha vida e que impactam de forma grandiosa meu crescimento como ser humano: à Gabriella, Kamila, Luiz Henrique, Bárbara e Letícia. Espero sempre poder contar com vocês e ter o imenso prazer de tê-los em minha vida.

A minha colega de laboratório e amiga Ana, meus agradecimentos pela amizade e parceria. Sua presença mudou meu rumo profissional e pessoal. Bom saber que há pessoas que compartilham da mesma vontade de crescer e ajudar.

Ao meu parceiro de laboratório Daniel, cuja presença foi de extrema importância em todo esse período. Desde o meu primeiro contato com Bioinformática, à concepção e execução de projetos, e incontáveis ajudas que só deixaram mais claro o quão paciente e querido és, além dessa vocação inata em liderar e gerir.

Ao Ibilce por ceder sua instituição e me aceitar mais dois anos, me proporcionando a experiência de conviver nesse campus onde só construí e construo boas memórias.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 88887.659699/2021-00 à qual agradeço,

“Palavras mudas transmitiam-se de alma a alma como um som no eco, dando-me alegrias passageiras que deixavam uma impressão profunda.”

Honoré de Balzac (1831, p. 131)

RESUMO

Os elementos de transposição (TEs), fontes importantes na geração de variabilidade genética devido à capacidade de se movimentarem no genoma, são predominantemente deletérios e neutros. Todavia, quando domesticados pelo genoma, desempenham papel positivo na evolução dos organismos, aumentando o número e diversidade gênica e enriquecendo as vias de regulação gênica. Identificar transcritos quiméricos gene + TE constitui uma das maneiras de investigar potenciais eventos de domesticação, os quais podem ser classificados de acordo com o sítio de inserção do TE em relação ao gene: iniciado por TE, exonizado por TE e terminado por TE. Para identificar transcritos quiméricos sugestivos de eventos de domesticação, analisamos transcriptomas de ovários de *Drosophila melanogaster* e *D. simulans*, procedentes de Gotheron (França, clima temperado) e de São José do Rio Preto (Brasil, clima tropical), duas linhagens de cada local, expostas aos estresses *Cold-Shock* e *Starvation*. A análise foi realizada com ChimeraTE, um *pipeline* capaz de detectar transcritos quiméricos em transcriptomas obtidos por RNA-seq *paired-end*, e identificados por meio de alinhamento genômico. A análise de expressão diferencial permitiu identificar a modulação da expressão gênica induzida por TEs associada aos dois tipos de estresse. Foram identificados, em média, 396 transcritos quiméricos nas linhagens de *D. melanogaster* e 285 nas linhagens de *D. simulans*, em 90% dos quais o TE estava exonizado. Aproximadamente 1% dos genes produziram transcritos quiméricos, dentre os quais, seis foram considerados como potenciais casos adaptativos por apresentarem expressões aparentemente moduladas pelas inserções do TE. Três genes estavam diferencialmente expressos (DE) entre controle e tratamentos: *Paris* e *Sulf1* em *D. melanogaster* e o gene *GD24271* em *D. simulans*. O gene *Paris* apresentou uma inserção *upstream*, polimórfica, de *Burdock* (LTR, superfamília *Gypsy*), apenas nas linhagens de Gotheron de *D. melanogaster*, no qual estava DE em apenas uma das linhagens em *Cold-Shock*. Também o gene *Sulf1*, com uma inserção intrônica, polimórfica, de *Micropia* (LTR, superfamília *Gypsy*), estava DE na mesma linhagem que *Paris*, tanto sob o estresse *Cold-Shock*, como sob *Starvation*. Já, o gene *GD24271*, que apresentou uma inserção *upstream* do elemento *G5A* (LINE, superfamília *Jockey*) em todas as linhagens de *D. simulans*, produziu transcrito quimérico e estava DE sob *Cold-Shock* apenas em uma linhagem da região de clima temperado. Outros três genes estavam DE quando comparada a expressão entre populações da região temperada e tropical: *CG5096* e *CG42598* em *D. melanogaster* e *GD24387* em *D. simulans*. O gene *CG5096*, com uma inserção *upstream* do TE 412 (LTR, superfamília *Gypsy*), exclusiva em uma linhagem de Gotheron, exibiu expressão maior nessa população. Já, os genes *CG42598* em *D. melanogaster*, com uma inserção intrônica do TE *gypsy4* (LTR, superfamília *Gypsy*), e o gene *GD24387*, em *D. simulans*, com uma inserção *upstream* do elemento *INE-1* (DNA, superfamília *Helitron*), ambos com inserções do TE comuns a todas as linhagens, exibiram maior expressão nas linhagens de clima tropical. Em todos os casos, a mudança de expressão gênica só ocorreu em linhagens que geraram transcritos quiméricos. Essas observações sugerem uma possível ação reguladora dos TEs, que por sua vez modulam a expressão dos genes adjacentes. Nas situações em que a mudança de expressão dos genes *Sulf1*, *Paris* e *GD24271* parece ter sido ativada pelo estresse, a ocorrência de transcritos quiméricos ocorreu apenas em linhagens nativas da região temperada. Essas mudanças sugerem resposta adaptativa ao frio, uma vez que a sobrevivência no inverno, com baixas temperaturas e restrição

de alimento é um desafio a ser superado pelas populações. Já, as mudanças de expressão dos genes *CG5096*, *CG42598* e *GD24387* entre linhagens das duas regiões climáticas, independente da inserção do TE, sugerem adaptações a colonização de diferentes habitats. Em síntese, esses resultados ilustram como o estresse pode induzir a domesticação do TE, de modo a refletir a adaptação e história evolutiva dos organismos.

Palavras-chave: ChimeraTE; atividade de TEs induzida por estresse; plasticidade genômica; plasticidade transcriptômica, potencial adaptativo de TEs.

ABSTRACT

Transposable elements (TEs), an important source of genetic variability due to their ability to move within the genome, are predominantly deleterious and neutral. However, when domesticated by the genome, TEs play a positive role in the evolution of organisms, leading to an increase in gene number and diversity and enriching gene regulation pathways. Identifying chimeric gene + TE transcripts is one way of investigating their domestication, which can be classified according to the TE insertion site in relation to the gene: transcripts TE-initiated -, TE-exonized and TE-terminated. To identify chimeric transcripts suggestive of domestication events, we analyzed transcriptomes from the ovaries of *Drosophila melanogaster* and *D. simulans*, from Gotheron (France, temperate climate) and São José do Rio Preto (Brazil, tropical climate), two strains from each location, exposed to Cold-Shock and Starvation stresses. The analysis was carried out using ChimeraTE, a pipeline capable of detecting chimeric transcripts in transcriptomes obtained by paired-end RNA-seq, and identified through genomic alignment. Differential expression analysis made it possible to identify the modulation of gene expression induced by TEs associated with the two types of stress. An average of 396 chimeric transcripts were identified in the *D. melanogaster* strains and 285 in the *D. simulans* strains, in 90% of which the TE was exonized. Approximately 1% of the genes produced chimeric transcripts, of which six were considered as potential adaptive cases because their expressions are apparently modulated by the TE insertions. Three genes were differentially expressed (DE) between the control and treatments: *Paris*, *Sulf1* in *D. melanogaster* and *GD24271* in *D. simulans*. The gene *Paris*, with an *upstream*, polymorphic *Burdock* insertion (LTR, *Gypsy* superfamily), only in the Gotheron strains, is DE in only one of the strains under *Cold-Shock*. *Sulf1*, with an intronic, polymorphic insertion of *Micropia* (LTR, *Gypsy* superfamily), is a DE gene in the same strain as *Paris*, both under the *Cold-Shock* and *Starvation* stresses. The third gene, *GD24271*, which presented in all the *D. simulans* strains an *upstream*, polymorphic insertion of the *G5A* element (LINE, *Jockey* superfamily), but produced chimeric transcript, and was DE under *Cold-Shock* only in a European strain. Three other genes were DE when comparing the expression between temperate and tropical populations: *CG5096* and *CG42598* in *D. melanogaster* and *GD24387* in *D. simulans*. The gene *CG5096*, with an *upstream* insertion of *412* (LTR, *Gypsy* superfamily), unique to a Gotheron lineage, exhibited higher expression in this population. On the other hand, the genes *CG42598* in *D. melanogaster*, with an intronic insertion of *gypsy4* (LTR, *Gypsy* superfamily), and *GD24387*, in *D. simulans*, with an *upstream* insertion of the *INE-1* element (DNA, *Helitron* superfamily), both with TE insertions common to all strains, showed greater expression in the Brazilian strains. In all these cases, the change in gene expression only occurred in strains that generated chimeric transcripts. These observations suggest a possible regulatory action of TEs, which in turn modulate the expression of adjacent genes. In situations where the change in expression of the *Sulf1*, *Paris* and *GD24271* genes appears to have been activated by stress, the chimeric transcripts occurred only in strains native to the temperate region. These changes suggest an adaptive response to the cold, since survival in winter, with low temperatures and food restriction, is a challenge to be overcome by insect populations. Meanwhile, the changes in expression of the *CG5096*, *CG42598* and *GD24387*

genes between strains from the two climatic regions, independent of the TE insertion, suggest adaptations to colonizing different habitats. In summary, these results illustrate how stress can induce the domestication of TE, in a way that reflects the adaptation and evolutionary history of the organisms.

Keywords: ChimeraTE; stress-induced TE activity; genomic plasticity, transcriptomic plasticity; adaptive potential of TEs.

LISTA DE ILUSTRAÇÕES

Figura 1 – Fluxograma dos experimentos de seleção artificial realizados por Menezes (2018), aplicando os estresses 1) Cold-Shock e 2) Starvation, para a obtenção das linhagens e dados de RNA-seq utilizados no trabalho.	29
Figura 2 – Fluxograma de funcionamento do pipeline ChimeraTE, disponível em Oliveira et al. (2023).	33
Figura 3 – Número de transcritos quiméricos produzidos em cada linhagem de a) <i>D. melanogaster</i> e b) <i>D. simulans</i> .	37
Figura 4. Diagrama Upset referente ao número de Transcritos Quiméricos compartilhados e exclusivos das linhagens de A) <i>D. melanogaster</i> e B) <i>D. simulans</i> .	38
Figura 5 – Proporção de genes que produzem transcritos quiméricos nas quatro linhagens de <i>D. melanogaster</i> e <i>D. simulans</i> .	41
Figura 6 – Frequência (%) de superfamílias de TEs no genoma e transcriptoma das linhagens de <i>D. melanogaster</i> .	43
Figura 7 – Frequência (%) de superfamílias de TEs no genoma e transcriptoma das linhagens de <i>D. simulans</i> .	45
Figura 8 – Frequência (%) de famílias de TEs no genoma e transcriptoma das linhagens de <i>D. melanogaster</i> .	47
Figura 9 – Frequência (%) de famílias de TEs no genoma e transcriptoma das linhagens de <i>D. simulans</i> .	49
Figura 10 – Frequência de famílias de TEs que constituem quimeras, por classe de transcrito, em <i>D. melanogaster</i> .	51
Figura 11 – Distribuição das famílias em transcritos quiméricos mais frequentes entre condições experimentais e as linhagens de <i>D. melanogaster</i> após a exclusão das famílias <i>INE-1</i> e <i>roo</i> .	52
Figura 12 – Frequência de famílias de TEs que constituem quimeras, por classe de transcrito, em <i>D. simulans</i> .	54

Figura 13 – Distribuição das famílias em transcritos quiméricos mais frequentes entre condições experimentais e as linhagens de <i>D. simulans</i> após a exclusão das famílias <i>INE-1</i> e <i>roo</i> .	55
Figura 14 – Diagrama de Venn exibindo as famílias compartilhadas na produção de quimeras entre as linhagens de <i>D. melanogaster</i> .	56
Figura 15 – Diagrama de Venn exibindo as famílias compartilhadas na produção de quimeras entre as linhagens de <i>D. simulans</i> .	56
Figura 16 – Correlação de Pearson entre o número de cópias das famílias de TEs no genoma e a frequência das famílias que mais formaram quimeras em <i>D. melanogaster</i> .	58
Figura 17 – Correlação de Pearson entre o número de cópias das famílias de TEs no genoma e a frequência das famílias que mais formam quimeras em <i>D. simulans</i> .	59
Figura 18 – Expressão de genes que geram transcritos quiméricos entre controle e tratados nas linhagens <i>D. melanogaster</i> .	63
Figura 19 – Número de genes diferencialmente expressos entre controle e tratados que formaram quimeras nas linhagens de <i>D. melanogaster</i> .	65
Figura 20 – Ilustração da expressão do gene <i>CG15043</i> e a presença do TE entre Controle e <i>Cold-Shock</i> no genoma das linhagens de <i>D. melanogaster</i> .	69
Figura 21 – Ilustração da expressão do gene e a presença do TE entre Controle e <i>Cold-Shock</i> no genoma das linhagens de <i>D. melanogaster</i> .	70
Figura 22 – Ilustração da expressão do gene <i>RhoGEF64C</i> e a presença do TE entre Controle e <i>Cold-Shock</i> no genoma das quatro linhagens de <i>D. melanogaster</i> .	71
Figura 23 – Ilustração da expressão do gene <i>Sulf1</i> e a presença do TE entre Controle e os Tratamentos no genoma das quatro linhagens de <i>D. melanogaster</i> .	72
Figura 24 – Expressão de genes que geraram transcritos quiméricos entre controle e tratados nas linhagens <i>D. simulans</i> .	74
Figura 25 – Número de genes diferencialmente expressos entre controle e tratados que formaram quimeras nas linhagens de <i>D. simulans</i> .	76
Figura 26 – Ilustração da expressão do gene <i>GD18191</i> e a presença do TE entre Controle e <i>Cold-Shock</i> no genoma das quatro linhagens de <i>D. simulans</i> .	79

Figura 27 – Ilustração da expressão do gene <i>GD24271</i> e a presença do TE entre Controle e <i>Cold-Shock</i> no genoma das quatro linhagens de <i>D. simulans</i> .	80
Figura 28 – Expressão de genes que geraram transcritos quiméricos entre linhagens brasileiras e francesas de <i>D. melanogaster</i> .	81
Figura 29 – Número de genes diferencialmente expressos que formaram quimeras entre linhagens brasileiras e francesas de <i>D. melanogaster</i> .	82
Figura 30 – Ilustração da expressão do gene <i>CG40498</i> e a presença do TE entre Controle e os Tratamentos no genoma das quatro linhagens de <i>D. melanogaster</i> .	84
Figura 31 – Ilustração da expressão do gene <i>CG5096</i> e a presença do TE entre Controle e os Tratamentos no genoma das quatro linhagens de <i>D. melanogaster</i> .	85
Figura 32 – Ilustração da expressão do gene e a presença do TE entre Controle e os Tratamentos no genoma das quatro linhagens de <i>D. melanogaster</i> .	86
Figura 33 – Expressão de genes que geraram transcritos quiméricos entre linhagens brasileiras e francesas de <i>D. simulans</i> .	88
Figura 34 – Número de genes diferencialmente expressos que formaram quimeras entre linhagens brasileiras e francesas de <i>D. simulans</i> .	89
Figura 35 – Ilustração da expressão do gene <i>GD10984</i> e a presença do TE entre Controle e os Tratamentos no genoma das quatro linhagens de <i>D. simulans</i> .	91
Figura 36 – Ilustração da expressão do gene <i>GD14844</i> e a presença do TE entre Controle e os Tratamentos no genoma das quatro linhagens de <i>D. simulans</i> .	92
Figura 37 – Ilustração da expressão do gene e a presença do TE entre Controle e os Tratamentos no genoma das quatro linhagens de <i>D. simulans</i> .	93
Figura 38 – Ilustração da expressão do gene <i>GD24386</i> e a presença do TE entre Controle e os Tratamentos no genoma das quatro linhagens de <i>D. simulans</i> .	94
Figura 39 – Ilustração da expressão do gene e a presença do TE entre Controle e os Tratamentos no genoma das quatro linhagens de <i>D. simulans</i> .	95

LISTA DE TABELAS

Tabela 1 – Distribuição de genes quiméricos, total de transcritos quiméricos e os exclusivos entre as linhagens de <i>D. melanogaster</i> e <i>D. simulans</i> , nos quais puderam ser identificadas inserções exclusivas potencialmente adaptativas.	40
Tabela 2 – Genes quiméricos diferencialmente expressos entre controle e tratados das linhagens de <i>D. melanogaster</i> .	67
Tabela 3 – Genes quiméricos diferencialmente expressos entre controle e tratados das linhagens <i>D. simulans</i> .	77
Tabela 4 – Genes quiméricos diferencialmente expressos entre linhagens brasileiras e francesas de <i>D. melanogaster</i> .	83
Tabela 5 – Genes quiméricos diferencialmente expressos entre linhagens brasileiras e francesas de <i>D. simulans</i> .	90

LISTA DE ABREVIATURAS E SIGLAS

CAGE	A análise Cap da expressão gênica
CDS	Região de Codificação
CS	Choque Frio
DA	Conteúdo de Dopamina
DE	Diferencialmente Expresso
DEG	Gene Diferencialmente Expresso
DHHC	Aspartato-Histidina-Histidina-Cisteína
DIRS	Sequência de Repetição Intermediária de <i>Dictyostelium</i>
DP	Doença de Parkinson
EST	Marcadores de Sequências Expressas
ERV	Elemento Viral Endógeno
HS	Sulfato de Heparano
LTRs	Repetição Terminal Longa
LINES	Elementos Nucleares Longos Intercalados
miRNA	microRNA
piRNA	Proteínas PIWI que interagem com RNA
SINES	Elementos Nucleares Curtos Intercalados
SNPs	Polimorfismo de Nucleotídeo Único
SSRs	Repetições de Sequência Simples
ST	Fome

TE	Elemento de Transposição
TIPs	Polimorfismos de Inserção de Transposon
TFBSs	Sítios de Ligação Fatores de Transcrição
ZFPs	Proteínas <i>Zinc-Finger</i>

SUMÁRIO

1	INTRODUÇÃO	20
2	OBJETIVOS	27
3	METODOLOGIA	27
3.1.	Exposição a estresses e seleção de genótipos adaptativos de <i>D. melanogaster</i> e <i>D. simulans</i>	28
3.2.	Anotação Gênica e de TEs	30
3.3.	Identificação de transcritos quiméricos	31
3.4.	Identificação de Polimorfismos	33
3.5.	Análise de Expressão Diferencial	34
3.6.	Estatísticas Adicionais e Ilustrações	34
4	CAPÍTULO 1- Caracterização de Transcritos Quiméricos	35
4.1.	A exposição ao estresse não parece resultar em maior produção de transcritos quiméricos	36
4.2.	TEs inseridos dentro de genes correspondem a cerca de 92% dos transcritos quiméricos em <i>D. melanogaster</i> e 80% em <i>D. simulans</i>	36
4.3.	~1,5% dos genes de <i>D. melanogaster</i> produzem transcritos quiméricos	38
4.4.	Elementos da família <i>INE-1</i> são os maiores produtores de transcritos quiméricos em <i>D. melanogaster</i> e <i>roo</i> em <i>D. simulans</i>	41
4.5.	Elementos de Classe I são os mais frequentes em transcritos quiméricos	46
4.6.	As famílias de TEs mais abundantes no genoma são as maiores produtoras de transcritos quiméricos no transcriptoma	57

4.7.	As linhagens de <i>D. melanogaster</i> apresentam mais inserções polimórficas que <i>D. simulans</i>	59
5	CAPÍTULO 2 – Efeitos dos TEs na Expressão de Genes Adjacentes	61
5.1.	Expressão diferencial entre controle e tratamentos e inserções de TEs adaptativas em <i>D. melanogaster</i>	62
5.2.	Expressão diferencial entre controle e tratamentos e inserções de TEs adaptativas em <i>D. simulans</i>	73
5.3.	Expressão diferencial interpopulacional e inserções de TEs potencialmente adaptativas em <i>D. melanogaster</i>	80
5.4.	Expressão diferencial interpopulacional e inserções de TEs potencialmente adaptativas em <i>D. simulans</i>	87
6	DISCUSSÃO	96
7	CONCLUSÕES E CONSIDERAÇÕES FINAIS	107
	REFERÊNCIAS	108
	ANEXOS A – Figuras Suplementares	
	ANEXOS B – Tabelas Suplementares	

1 INTRODUÇÃO

As espécies irmãs *Drosophila melanogaster* e *D. simulans* são excelentes modelos para identificar eventos potencialmente adaptativos de elementos de transposição (TEs), já que são espécies proximamente aparentadas, originadas na mesma região geográfica, afro-tropical, bastante estudadas e, principalmente *D. melanogaster*, conta com uma história evolutiva bem fundamentada. Esses conhecimentos propiciam comparar o conteúdo genômico e transcriptômico dos TEs nas duas espécies e formular hipóteses de como essas sequências genéticas móveis podem atuar na diferenciação e evolução, agregando conhecimento sobre os processos adaptativos associados às dispersões pelas quais passaram as espécies. Das duas espécies, *D. melanogaster* é um dos principais e mais importantes organismos modelo na Biologia, um excelente material para estudos genômicos, evolutivos e ecológicos. Oriunda do sudoeste da África-Subsaariana (González et al., 2010), colonizou o restante do mundo, primeiramente migrando para a Europa entre 19.000 e 4.100 anos atrás e, nos últimos séculos, para a América do Norte e Austrália em associação, provavelmente, ao comércio de escravos (Lachaise et al., 1988; Sprengelmeyer et al., 2020; Suhre; McCarthy; Schwenk, 2021). *Drosophila simulans*, espécie irmã e amplamente simpátrica a *D. melanogaster*, tem uma história demográfica aproximadamente semelhante; teve sua origem na África Oriental ou Madagascar, e mais recentemente, se espalhou pela Europa, Américas e Austrália (Lachaise et al., 1988; Dean; Ballard, 2004).

Apesar da ancestralidade comum recente compartilhada, das semelhanças quanto à biologia, ecologia e história demográfica, estudos comparativos têm evidenciado muitas diferenças entre *D. melanogaster* e *D. simulans*. Em ambas espécies ocorrem flutuações sazonais em SNPs (do Inglês: *Single Nucleotide Polymorphism*), que poderiam indicar resposta regular às pressões seletivas, as quais variam sazonalmente (Kolaczowski et al., 2011; Bergland et al., 2014); contudo, elas diferem em suas abundâncias sazonais e tolerâncias térmicas (Parsons, 1977). Sedghifar et al. (2016) reportaram diferenciação de SNPs em escala genômica entre as duas espécies e sugeriram que as populações australianas e norte-americanas de *D. simulans* estão se adaptando aos seus ambientes locais tanto por mecanismos semelhantes como diferentes aos caracterizados para *D. melanogaster*, incluindo dentre eles, a regulação da expressão gênica. Em comparações intrapopulacionais, a variabilidade entre as duas espécies é semelhante (Capy; Gilbert, 2004; Kang et al., 2019), já em âmbito interpopulacional *D. melanogaster* é mais diversificada (Hyytiä et al., 1985; Singh, Choudhary; David, 1987). Quando se analisa a variabilidade genética nucleotídica intrapopulacional, observa-se maior

variabilidade em *D. simulans*, mas em nível enzimático e morfológico essa variabilidade é maior para *D. melanogaster* (Aquadro; Lado; Noon, 1988).

Os desafios impostos pelo ambiente são uma das forças propulsoras da evolução, que se traduz na seleção de adaptações. A heterogeneidade de ambientes gera pressões seletivas diferentes que conferem adaptações distintas entre populações. A seleção pode atuar aumentando em frequência as variantes genéticas que podem conferir uma vantagem na sobrevivência (Ayala; Fitch, 1997). Dentre as respostas adaptativas, a variação da expressão gênica é relativamente comum em populações naturais. No entanto, processos genéticos populacionais responsáveis pela origem e manutenção dessa variação continuam a ser elucidados. Por exemplo, muitos genes são expressos diferencialmente entre populações de regiões com alta e baixa latitude, como entre *D. melanogaster* e *D. simulans* na América do Norte, e em populações australianas, nas quais a maioria deles variam na mesma direção, indicando um perfil comum em espécies diferentes (Reinhardt et al., 2014; Zhao et al., 2015). Como a regulação da expressão gênica é crítica para a função do organismo, e é uma fonte importante de diversidade fenotípica entre as espécies, a compreensão dos mecanismos responsáveis pela divergência regulatória pode fornecer informações sobre a adaptação das populações a seus ambientes e a natureza das mudanças evolutivas (Ayala; Fitch, 1997). Dentre esses mecanismos, os elementos de transposição (TEs) podem desempenhar um papel importante na adaptação.

Os elementos de transposição (TEs) são sequências de DNA medianamente repetidas, que, por apresentarem a capacidade de se deslocar no genoma, podem desempenhar papel significativo na produção de variabilidade genética e na adaptação de organismos. Primeiramente identificados por Barbara McClintock, na década de 40, em milho, hoje já foram registrados em praticamente todos os genomas de eucariotos, em quantidades variáveis, compondo 10% do genoma de *Arabidopsis thaliana*, (Gonzalez et al., 2008), 15% em *Drosophila melanogaster* (Quesneville et al., 2005), 45% em humanos (Lander et al., 2001), e podendo compreender até 85% do genoma de algumas espécies de milho (Schnable et al., 2009). No gênero *Drosophila*, o trabalho clássico de anotação dos TEs nos genomas de 12 representantes (*Drosophila 12 Genomes Consortium*, Clark et al., 2007), indicou uma variação de ~2,7% (*D. simulans* e *D. grimshawi*) a ~25% (*D. ananassae*) e compondo ~5,4 % do genoma de *D. melanogaster*. Sequenciamentos e anotações mais atuais (*long reads* Oxford Nanopore MinION, (Awad; Gan, 2020); dnaPipeTE, (Goubert et al., 2015; e TEcount; Lerat et al., 2017)) revelaram que os conteúdos genômicos de TEs são mais altos, variando de ~5% a ~12 % em

D. simulans a ~7% a 15% em *D. melanogaster*. Essas diferenças de conteúdo genômico de TEs observadas entre *D. melanogaster* e *D. simulans* podem ter relação com as diferentes dinâmicas populacionais e históricas das espécies (Vieira; Biémont, 2004).

Levando em consideração sua estrutura, aspectos evolutivos, auto regulação e mecanismo de replicação, os TEs são classificados em diferentes níveis, sendo os principais classe, ordem, superfamílias e famílias. A mais alta categoria, classe, divide os TEs de acordo com a produção ou não de um RNA intermediário no processo de transposição, e o mecanismo de transposição os agrupam em ordens. Os retrotransposons, elementos pertencentes à Classe I, se mobilizam via um RNA intermediário mediante a ação da enzima transcriptase reversa, por um mecanismo conhecido como “copia e cola”. Os retrotransposons são classificados em ordens de acordo com a presença ou ausência de longas repetições terminais, como os retrotransposons com LTRs (*Long Terminal Repeat* - LTR), que são autônomos e codificam proteínas do complexo Gag-Pol; os elementos nucleares intercalados longos (*Long Interspersed Nuclear Elements* - LINES); e os elementos nucleares intercalados curtos (*Short Interspersed Nuclear Elements* - SINES), sendo estes não-autônomos, pois não codificam as enzimas necessárias para sua transposição, dependendo da maquinaria enzimática dos LINES (Wicker et al., 2007; Sinzelle et al., 2009). Ainda, na Classe I, encontram-se os elementos da ordem DIRS-like (*Dictyostelium Intermediate Repeat Sequence*), que apresentam mecanismos de transposição distintos e suas extremidades também não apresentam LTRs, se assemelhando mais a repetições diretas parciais (SDR); e os elementos da ordem *Penelope-like*, contando com membros que apresentam em suas extremidades resquícios de LTRs, mas com um mecanismo de transposição pouco distinto dos retrotransposons com LTRs. Essas ordens incluem um número total de 17 superfamílias (LTR: *Copia*, *Gypsy*, *Bel-Pao*, *Retrovírus*, *ERV*; DIRS: DIRS, *Ngaro*, *VIPER*; PLE: *Penelope*; LINE: *R2*, *RTE*, *Jockey*, *L1*, *I* e SINE: *tRNA*, *7SL* e *5S*) (Wicker et al., 2007).

Os elementos da Classe II (Transposons de DNA) não requerem o passo da transcrição reversa para se integrarem ao genoma. Eles são divididos em duas subclasses: Subclasse I que realiza a transposição via mecanismo “corta-e-cola”, e subclasse II que realiza a transposição pelo mecanismo “copia-e-cola”. A subclasse I é composta pelas ordens TIR e Crypton e a subclasse II é constituída pelas ordens Helitron e Maverick / Polintron. Dentro dessas subclasses, há um total de 12 superfamílias (ordem TIR: *Tc1-Mariner*, *hAT*, *Mutator*, *Merlin*, *Transib*, *P*, *PiggyBac*, *PIF/Harbinger*, *CACTA*; ordem Crypton: *Crypton*, ordem Helitron:

Helitron e ordem Maverick: *Maverick*) (Wicker et al., 2007). As superfamílias são discriminadas de acordo com as proteínas e estruturas de domínios não codificantes, presença e comprimento das repetições terminais diretas; já a classificação em família leva em consideração ~80% de identidade entre as sequências ou de seu domínio interno, ou da repetição terminal (Wicker et al., 2007).

Os TEs são importantes promotores de variação genética e da expressão gênica porque podem carregar suas próprias sequências *cis* regulatórias capazes de explorar *in trans* o genoma do hospedeiro (Sundaram; Wysocka, 2020). Em nível transcricional, por exemplo, um TE pode inserir sequências promotoras em genes vizinhos, como também pode interromper os elementos reguladores existentes, ou introduzir um novo elemento *cis* regulatório, como um local de ligação de fator de transcrição (TBFSs - do Inglês *Transcriptional Factor Binding Site*) (Feschotte, 2008). Em nível pós-transcricional, um TE inserido na região 3' UTR de um gene pode adicionar um local de poliadenilação, um local de ligação para um microRNA (miRNA) ou para uma proteína de ligação ao RNA. Por fim, um TE inserido dentro de um íntron pode interferir no padrão de *splicing* de um pré-mRNA, culminando em novas formas de *splicing* alternativo (Feschotte, 2008).

A maneira abrupta pela qual as inserções de TEs podem interromper sequências gênicas, alterando sua estabilidade, composição, e regulação, irá resultar na degradação da grande maioria dessas inserções, que serão prontamente reconhecidas pela maquinaria de reparo celular e eliminadas. Assim, a maioria dos eventos de inserção são provavelmente deletérios para o genoma do hospedeiro e, como consequência, evoluíram os mecanismos de defesa contra a mobilidade dos TEs (revisado em Kavi et al., 2005; Slotkin; Martienssen, 2007). Em *Drosophila*, estes incluem silenciamento transcricional e pós-transcricional pelo sistema *Piwi interacting RNA* (piRNA), presente nas células germinativas (revisado em Iwasaki; Siomi, 2015) e que garantem a transmissão de genomas intactos de uma geração para outra. Quando a eficiência dos efetores dessas vias não é mais mantida, os TEs tornam-se ativos, podendo levar a uma redução significativa da aptidão, e até a letalidade (Vagin et al., 2004; Malone et al., 2009; Schnable et al., 2009). Caso não impactem a sobrevivência e divisão celular e, por conseguinte, a aptidão do organismo, algumas inserções serão mantidas, e, por acúmulo de mutações ao longo das gerações, poderão perder sequências regulatórias responsáveis pela sua movimentação e ficar inativas, desempenhando um efeito neutro para a célula. Em raros casos

em que inserções, de alguma maneira, são benéficas para o organismo, essas podem ser mantidas, por seleção positiva, e desempenhar um importante papel na adaptação.

Uma das contribuições dos TEs com a diversidade genética, a evolução do genoma e a diferenciação populacional, é a formação de Polimorfismos de Inserção de Transposon (*Transposon Insertion Polymorphisms* - TIPS) (Sanseverino et al., 2015). Em estudos populacionais, pode-se averiguar e diferenciar as inserções TEs ditas fixas, em que a transposição ocorreu antes da divergência das populações, em comparação às polimórficas, adquiridas ou perdidas depois da divergência e com possível papel na adaptação da população aos novos ambientes (Rishishwar et al., 2015).

A susceptibilidade dos TEs às mudanças do ambiente é demonstrada pela ativação e/ou repressão da sua atividade, geralmente influenciando a expressão de genes próximos à inserção (Horváth et al., 2017). Por conta disso, a relação entre a ativação de TEs pelo estresse e respostas a esses é bem documentada. O dinamismo associado a ativação dos TEs é principalmente relacionado com alterações epigenéticas, como marcas na cromatina e metilação. Como exemplo, verificou-se em populações de *D. melanogaster*, de fora da África, que a inserção do TE *invader4* na região 5' UTR do gene *sarah* e *upstream* ao gene *Bin1*, estava associada à diminuição do tempo de desenvolvimento e, ao aumento da sensibilidade ao estresse por frio (Ullastres et al., 2015; 2021). Também, que o aumento da expressão do gene *CG188446*, candidato à resposta a resistência ao frio em *D. melanogaster*, estava associado ao LTR *roo* atuando como promotor alternativo para esse gene (Merenciano et al., 2016).

A colonização de novos ambientes também expõe as espécies a tipos diferentes de estresses. Esses podem ser ambientais ou resultantes de cruzamento com espécies nativas, podendo propiciar a invasão do genoma de uma das espécies envolvidas por TEs, com consequente aumento de taxas de transposição (Garcia-Guerreiro, 2012). A alta frequência de ocupação de alguns TE em populações colonizadoras, em relação às populações nativas, foi observada em estudo com espécies invasoras com *D. buzzatii* (Labrador et al., 1998; Garcia Guerreiro; Fontdevila, 2001; 2011) e *D. subobscura* (Garcia Guerreiro et al., 2008). Como também se registrou o aumento no número de cópias dos TEs *Uhu* e *LOA* em diversos subgrupos de espécies de *Drosophila* havaianas, associado à colonização de novas ilhas comparada com ilhas antigas (Wisotzkey et al., 1997). Em cenários de estresse, os TEs oferecem possibilidades de atuação da seleção natural em respostas ao agente estressante

(Wessler, 1996), fornecendo variabilidade de transcriptomas, otimizando a resposta ao desafio ambiental (Fablet; McDonald; Biemont; Vieira, 2006).

Entre os mecanismos pelos quais os TEs podem contribuir com a evolução adaptativa, os eventos de domesticação e exaptação são de suma importância. Cunhado em 1982 por Gould e Vrba, exaptação, ou cooptação, caracteriza-se pela evolução de novas funções a partir de um caráter ancestral que contribuem com a aptidão do organismo, como a exaptação dos pulmões em peixes pulmonados a partir da bexiga natatória (Brosius, 2019; Phillips; Heath, 1992). Um TE dito exaptado, exercerá em seu novo gene hospedeiro uma função distinta da original, alterando a proteína final de seu hospedeiro (Schrader; Schmitz, 2019). Já a domesticação refere-se a inserções de TEs, que muito provavelmente perderam sua capacidade de mobilização, que afetam a expressão de genes adjacentes e são vantajosas (Capy, 2021). Na literatura esses termos são intercambiáveis, como um exemplo da oncoexaptação, cujo termo carrega a palavra exaptação mas se trata de um caso de domesticação, já que um elemento viral endógeno (ERV) é cooptado como promotor alternativo e age aumentando o nível transcricional do oncogene humano CSF1R no Linfoma de Hodgkin (Lamprecht et al., 2010).

Exemplos de domesticação de TEs incluem casos como o do retrotransposon *Hopscotch* no milho, que pode ter sido domesticado por atuar intensificando a expressão do gene *teosinte branched1*, acarretando no aumento da dominância apical quando comparado à forma selvagem sem a inserção (Studer et al., 2011). Já um retrotransposon *Copia-like*, em laranjas *Citrus sinensis*, atua como um intensificador do loco *rubi*, responsável pela modulação da cor da fruta em resposta ao frio (Butelli et al., 2012). O famoso caso do melanismo industrial das mariposas inglesas, também configura um caso de domesticação, onde a inserção de um transposon de DNA, denominado *carbonaria*, no gene *córtex*, induz o aumento da expressão gênica influenciando a melanização da mariposa *Biston betularia* adulta (Van't Hoff et al., 2016).

A domesticação da sequência completa do TE, ou parte dela, e a subsequente fusão dessa sequência a um gene funcional pré-existente, gera transcritos quiméricos (Nicolau et al., 2021). Eles podem ser de três tipos: Transcritos iniciados por TEs (*TE-initiated transcript*), Transcritos com TEs exonizados (*TE-exonized transcript*) e Transcritos terminados por TEs (*TE-terminated transcript*). Como os próprios termos indicam, os transcritos gênicos iniciados por TEs possuem o seu sítio de início da transcrição em um TE, ao invés de seu local de iniciação comum (McGinnis et al., 1983) e os terminados por TEs são distinguidos por terminarem na sequência de um TE (Farré et al., 2016). Ambos podem provocar mudanças na

expressão gênica. Os transcritos exonizados, por outro lado, são aqueles em que a sequência do TE é reconhecida como um éxon pela maquinaria de *splicing*, com prováveis consequências na proteína funcional final. Os transcritos iniciados por TEs são os melhor documentados, devido aos diversos *softwares* e técnicas disponíveis que são sensíveis a sua expressão. A análise de expressão utilizando a técnica CAGE (*Cap analysis gene expression*), identificaram esses transcritos variando de 3-14% em humanos e camundongos (Faulkner et al., 2009). A oncoexaptação é um exemplo do impacto da formação de um transcrito TE-quimérico no hospedeiro. Assim como a cooptação de um transposon *pogo* no gene *CG11699* em populações de *D. melanogaster*, que afeta o sinal de poliadenilação desse gene, aumentando sua expressão e conferindo resistência ao estresse xenobiótico (Mateo et al., 2014).

A capacidade de mobilização dos elementos de transposição (TE), suas implicações na geração de variabilidade genética, remodelação e plasticidade de genomas, atreladas a respostas adaptativas é de suma importância no contexto de exposição dos organismos ao estresse. O aumento e a repressão da atividade de TEs são possíveis respostas a eventos estressantes. Se, durante o processo de transposição, os TEs se integrarem às sequências gênicas codificadoras de proteínas, e gerarem transcritos quiméricos, a depender do sítio de inserção do elemento, eles podem alterar a expressão gênica e/ou o produto final, que é a proteína. Alterações como essas, comumente, apresentam grandes chances de serem desfavoráveis ou neutras; mas, menos frequentemente, podem vir a ser favoráveis ao valor adaptativo do organismo, sendo, portanto, mantidas e domesticadas. As espécies cosmopolitas *D. melanogaster* e *D. simulans*, filogeneticamente próximas, adaptadas a diferentes zonas climáticas e que exibem diferenças bem documentadas no conteúdo de TEs (Biemont et al., 1999; Vieira et al., 2002; Cappy; Gibert, 2004; Vieira; Biemont, 2004), as tornam de especial valia para compreender e avaliar a associação dos TEs a adaptação.

Considerando os fatores acima citados, as hipóteses testadas neste trabalho foram de que linhagens de *D. melanogaster* e *D. simulans*, submetidas a tratamentos estressantes (frio e restrição nutricional) ou adaptadas a regimes climáticos diferentes em suas populações de origem (no Velho e no Novo Mundo), produzem transcritos quiméricos gene-TEs potencialmente adaptativos a essas condições.

2 OBJETIVOS

O objetivo geral deste trabalho foi identificar genes com transcritos quiméricos derivados da cooptação de TEs, sugestivos de domesticações potencialmente adaptativas, em resposta ao estresse e à colonização de novos habitats e zonas climáticas.

Objetivos Específicos

Utilizando quatro linhagens de *D. melanogaster* e quatro de *D. simulans*, duas procedentes de regiões geográficas com clima temperado (França) e duas de clima tropical (Brasil), pretendeu-se alcançar os seguintes objetivos:

1. Anotar os TEs nos genomas de quatro linhagens das duas espécies;
2. Identificar transcritos quiméricos genes-TEs derivados de inserção de TEs nas regiões 5'UTR, éxon e 3'UTR; *up* e *downstream* ao gene;
3. Comparar a ocorrência das famílias de TEs nos genomas das linhagens com aquelas que constituem as quimeras;
4. Identificar possíveis casos de inserções de TEs adaptativas, associadas aos dois tipos de estresses, *Cold-Shock* e *Starvation*.
5. Identificar possíveis casos de inserções de TEs adaptativas, associadas à colonização de novos habitats e zonas climáticas em cada população.

3 METODOLOGIA

Foram analisados neste estudo os genomas e transcriptomas de quatro linhagens de *D. melanogaster* e quatro linhagens de *D. simulans*, previamente sequenciados e montados (Mohamed et al., 2020). Oito amostras de populações naturais foram coletadas em junho de 2014, utilizando iscas de frutas no Brasil (São José do Rio Preto 20° 41'04.3" S 49° 21'26.1" W – linhagens “sj”) e França (Gotheron, 44° 56'0" N 04° 53'30" E – linhagens “goth”), duas linhagens de *D. melanogaster* e duas de *D. simulans* brasileiras e outras duas linhagens europeias de cada espécie. Os genomas e transcriptomas foram obtidos no *Laboratoire de Biométrie et Biologie Evolutive* (LBBE) da Université Claude Bernard Lyon 1, sob a coordenação da Dra. Cristina Vieira (Mohamed et al., 2020). Trinta linhagens isofêmeas foram estabelecidas diretamente de fêmeas grávidas coletadas em campo. Descendentes F₁ de cada linhagem foram endo cruzadas por 30 gerações, a fim de reduzir a heterozigosidade. As moscas

foram mantidas em meio de cultura LM (ágar, farinha de trigo, levedura e solução de Nipagin), em câmara de temperatura constante a 25°C e fotoperíodo de 12h e expostas a condições estressantes (Menezes et al., 2018). As análises deste trabalho foram realizadas com os dados do sequenciamento genômico dessas linhagens e dos transcriptomas por RNA-seq das linhagens submetidas a seleção artificial (Menezes et al., 2018).

3.1 Exposição a estresses e seleção de genótipos adaptativos de *D. melanogaster* e *D. simulans*

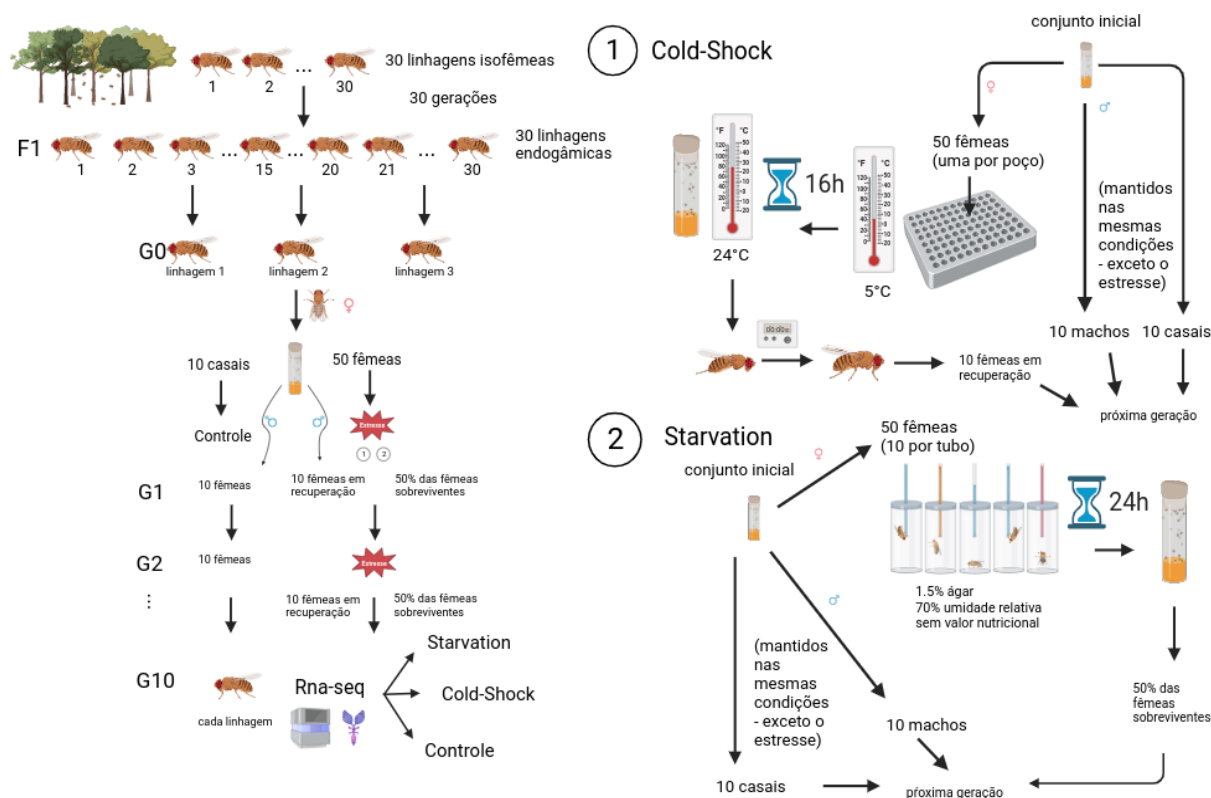
De cada linhagem endogâmica obtida das amostras de populações naturais de *D. melanogaster* e de *D. simulans*, após 30 gerações, foram escolhidas aleatoriamente três linhagens para realizar os experimentos. As fêmeas das linhagens de cada região foram submetidas a dois tipos de estresse: tempo de recuperação do coma frio (*Chill Coma Recovery Time* referido em toda a análise como *Cold-Shock* = CS) e de resistência à restrição nutricional (*Starvation* = ST). Os experimentos e análises descritas nesta seção foram realizados por Menezes e colaboradores (2018) (Figura 1).

Para o tratamento CS, foram utilizadas apenas moscas de 2 a 5 dias, as quais foram expostas a 5° C em câmara contendo gelo derretido, durante 16 horas, até entraram em coma por frio. O tempo de recuperação do movimento após a exposição ao frio foi medido à temperatura ambiente (24 °C) até que as moscas pudessem ficar de pé. As primeiras 10 fêmeas foram utilizadas como reprodutoras da próxima geração, juntamente com 10 outros pares aleatórios empregados como controle. A fim de minimizar efeitos de variações micro ambientais, foram estabelecidas para cada linhagem seus próprios controles mantidos nas mesmas condições (temperatura, densidade e meio de cultura). Foram excluídos da análise, indivíduos que demoraram mais de 120 min para se recuperarem do coma frio (Menezes et al., 2018).

Para o tratamento ST, as moscas foram colocadas em frascos de inanição, com meio de 1,5% de ágar, sem valor nutricional, apenas para disponibilizar umidade ingerível. Cinco tubos foram estabelecidos, contendo 10 fêmeas, de 2 a 5 dias de idades, e mantidas a 25°C e 70% de umidade relativa. A seleção foi aplicada por 10 gerações com uma amostra de 50 fêmeas; o número de moscas mortas foi registrado três vezes ao dia. Das cinco réplicas, ~50% das fêmeas sobreviventes foram utilizadas para as gerações seguintes, os machos foram escolhidos aleatoriamente. Após o tratamento, as moscas foram alimentadas em frascos contendo alimentos frescos e sua sobrevivência foi medida depois de 24 h. Os controles

constaram do mesmo número de moscas selecionadas por linhagem a cada geração, certificando que densidade não fosse uma variável considerável (Menezes et al., 2018).

Figura 1. Fluxograma dos experimentos de seleção artificial realizados por Menezes (2018), aplicando os estresses 1) *Cold-Shock* e 2) *Starvation*, para a obtenção das linhagens e dados de RNA-seq utilizados no trabalho.



Ao fim dos tratamentos, foram selecionados quatro genótipos de cada espécie, oriundos de populações diferentes: para *D. melanogaster* Dmgoth101 e Dmgoth63, provenientes de Gotheron, e Dmsj23 e Dmsj7, de São José do Rio Preto; para *D. simulans* Dsgoth31 e Dsgoth61, provenientes de Gotheron, e Dssj27 e Dssj9, de São José do Rio Preto.

Extração e Anotação Genômica e Transcriptômica

As extrações, sequenciamentos e análises descritas nesta seção foram realizados por Mohamed e colaboradores (2020). O DNA foi extraído das fêmeas de 3 a 5 dias das oito linhagens selecionadas, utilizando o kit Qiagen DNeasy Blood&Tissue (# 69506), de acordo com as instruções do fabricante. A fragmentação do DNA (1,5 g) foi realizada para um tamanho de inserção alvo de 300 pares de base (pb), que foi posteriormente sequenciado utilizando-se Illumina HiSeq *paired-end* com leituras de 125 pb, na instalação GeT-PlaGe, Génopole Toulouse (França). Já as leituras longas (*long-reads*) foram obtidas pelo sequenciamento da Oxford Nanopore Technology (ONT). Para isso, o DNA foi extraído de aproximadamente 100

machos selvagens com o cada kit Qiagen DNeasy Blood&Tissue, e reparado (3 mcg) usando o NEBNext FFPE DNA Repair Mix (NEB M6630). O sequenciamento MinION foi realizado, também na instalação GeT-PlaGe, Génopole Toulouse (França), e de acordo com as diretrizes do fabricante usando células de fluxo R9.4.1 (FLO-MIN106, ONT) e um sequenciador Nanopore MinIon Mk1b (ONT) controlado pelo software ONT MinKNOW (versão 18.3.1 para G0, versão 19.05.0.0 para linhagens isogênicas do tipo selvagem e versão 19.10.1 para as amostras G73 e G0-F100).

O RNA foi extraído de 30 carcaças e 30 pares de ovários de fêmeas com 3-5 dias de idade, de cada linhagem, usando o kit Qiagen RNeasy Plus. Duas réplicas biológicas foram feitas para cada tratamento (controle, CS e ST). As bibliotecas foram construídas com o kit Illumina TruSeq RNA Sample Prep Kit e o sequenciamento via RNA-seq foi feito com o sequenciador Illumina High HiSeq 3000 com método *paired-end* com leituras de 150 pares de base. As leituras de baixa qualidade foram removidas com o *software* Trimmomatic (Modolo; Lerat, 2015, Mohamed et al. 2020).

Ao fim dessa etapa, obteve-se a anotação genômica e três arquivos do sequenciamento de RNA-seq, antes (Controle) e depois da exposição aos estresses (*Cold-Shock* e *Starvation*), para cada linhagem (*D. melanogaster*: Dmgoth101, Dmgoth63, Dmsj23, Dmsj7; *D. simulans*: Dsgoth31, Dsgoth61, Dssj27, Dssj9), analisados no presente trabalho. Foram sequenciados e analisados somente transcriptomas de ovários tendo em vista que os tecidos reprodutores femininos estão mais intimamente relacionados com a transmissão transgeracional e com a regulação genômica de TEs.

3.2 Anotação Gênica e de TEs

A anotação gênica dos oito genomas foi realizada com o software Liftoff v1.6.3 (Shumate; Salzberg, 2021), com os parâmetros -s 0,5 -s 0,75 -exclude_partial e os genomas de referência de *D. melanogaster* r6.49 e *D. simulans* r2.02 disponíveis no *FlyBase* em (http://ftp.flybase.net/genomes/dmel/dmel_r6.49_FB2022_06/fasta/dmel-all-chromosome-r6.49.fasta.gz; http://ftp.flybase.net/genomes/Drosophila_simulans/dsim_r2.02_FB2020_03/fasta/dsim-all-chromosome-r2.02.fasta.gz).

A fim de identificar os TEs nos genomas, foi utilizado um *script* unindo os *softwares* RepeatMasker (Smit; Arian FA, 2004) e One Code to Find Them All (Bailly-Bechet; Haudry; Lerat, 2014) disponível no ambiente do *pipeline* ChimeraTE (Oliveira et al., 2023). O

RepeatMasker é conhecido por sua capacidade de identificar, classificar e mascarar elementos repetitivos, repetições intercaladas e sequências de baixa complexidade. A anotação dos TEs genômicos foi realizada com este *software* com os parâmetros: -nolow; -norna; -s; e -lib, alinhando cada genoma contra uma biblioteca de sequência TE consenso para *D. melanogaster*, disponível em bergmanlab v.10.2 (https://github.com/bergmanlab/drosophila-transposons/blob/master/current/D_mel_transposon_sequence_set.fa); a mesma biblioteca consenso foi utilizada para a anotação dos TEs em *D. simulans*. Em seguida, o *software* One Code to Find Them All foi empregado para garantir a qualidade das sequências anotadas e evitar a fragmentação de TEs, juntando sequências LTR (Repetição Terminal Longa, do Inglês *Long Terminal Repeat*) com seus respectivos TEs, e unindo fragmentos de TEs da mesma família, que se sobrepõem no genoma, para reduzir pseudo-duplicações e identificar inserções de TEs íntegras. Seguindo a regra 80-80-80 (Wicker et al., 2007), inserções de TE com comprimento < 80pb foram descartadas devido à falta de robustez para defini-las como TEs. Para identificar a presença de SSRs (Repetições de Sequência Simples, do Inglês *Simple Sequence Repeats*) na composição das inserções TE previamente identificadas, empregou-se o *software* TRF v.4.09 (Benson, 1999), com parâmetros 2, 5, 6, 75, 20, 50, 500, -m, -h. Inserções com mais de 50% de sua sequência compostas por SSRs foram também removidas. A avaliação da composição das sequências de TEs anotadas foi um passo importante, tendo em vista que foi identificada a presença de SSRs em muitos TEs, como os elementos representantes das famílias *roo*, *412*, *tirant* (Flutre et al., 2011), as quais, devido ao comprimento podem ser indistinguíveis das inserções TE identificadas pelo RepeatMasker. A abundância genômica de TEs foi quantificada mediante o número de nucleotídeos e/ou o número de cópias.

3.3 Identificação de transcritos quiméricos

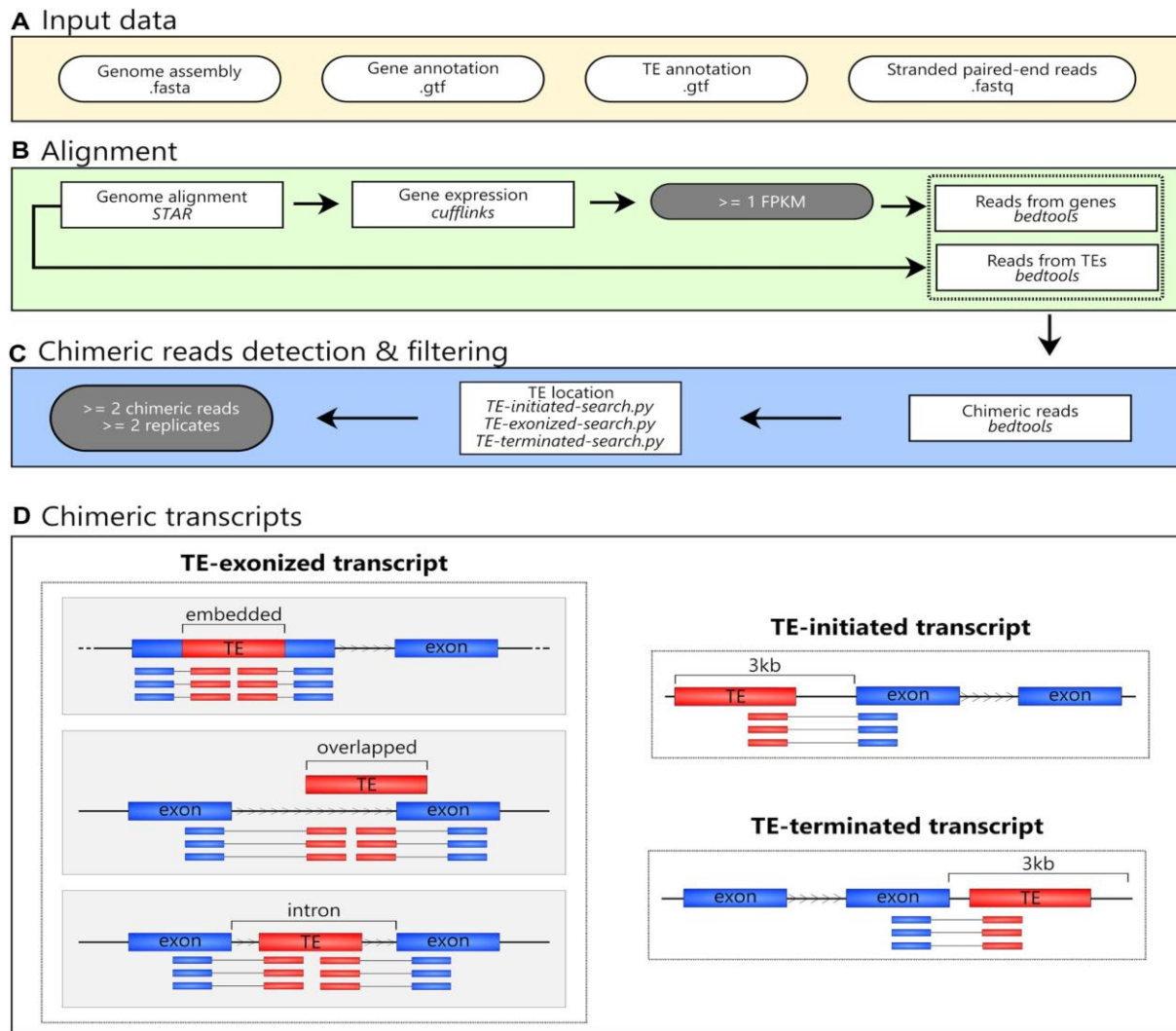
A identificação de transcritos quiméricos derivados de genes e TEs foi realizada neste trabalho com o *pipeline* ChimeraTE (Oliveira et al., 2023), com o método *genome-guided* (Mode1) que identifica fragmentos de mRNA quiméricos (derivados de éxons e TEs), com base nas sequências *paired-end* de *RNA-seq*, o genoma e suas respectivas anotações gênicas e de TEs. Brevemente, o *pipeline* ChimeraTE Mode1 realiza o alinhamento de leituras *paired-end* contra a sequência genômica com o *software* STAR v.2.7.6a (Dobin et al., 2013). O alinhamento é então submetido para análise com o *software* bedtools (Quilna; Hall 2010), função *intersect*, para identificação de quais leituras alinharam contra a sequência de um TE, e se as mesmas também alinharam contra as sequências dos éxons dos genes, formando então os pares de

leituras quiméricas. A formação destes pares no alinhamento fornece evidências suficientes para inferir que antes do sequenciamento haviam moléculas de RNA mensageiro na biblioteca em que parte era formada por um TE, e parte por um gene.

Dependendo da região gênica em que o TE está localizado, o transcrito quimérico é classificado (Oliveira *et al.*, 2023) como: (1) *TE-initiated transcript* (Transcritos iniciado por TE), transcritos em que o TE está localizado na região promotora do gene; (2) *TE-terminated transcript* (Transcritos terminados por TE), transcritos em que o TE está localizado *downstream* ao gene; (3) *TE-exonized transcripts* (Transcritos com TE exonizados), transcritos em que o TE está localizado dentro da região gênica. Com base na posição do TE em relação ao éxon, esses transcritos ainda podem ser subdivididos em *Embedded* (embutido), quando o TE está inserido totalmente dentro de um éxon, *Intronic* (intrônico), quando o TE está localizado inteiramente dentro de um íntron, e *Overlapped* (sobreposto), quando o TE é encontrado parcialmente inserido em um éxon e parcialmente inserido em um íntron (Figura 2).

Uma vez que o ChimeraTE fornece as coordenadas dos *scaffolds* em que os TEs e os genes que formam quimeras estão localizados, assim como a família de TE assinalada ao transcrito, cada quimera foi manualmente avaliada pelo software IGV (Thorvaldsdottir; Robinson; Mesirov, 2013) para a identificação de possíveis falsos positivos, onde não há de fato a produção de leituras quiméricas. Além disso, só foram consideradas as quimeras identificadas nas duas réplicas de RNA-seq, com cobertura média de pelo menos duas leituras quiméricas de suporte.

Figura 2. Fluxograma de funcionamento do *pipeline* ChimeraTE, Mode1, disponível em Oliveira et al. (2023).



3.4 Identificação de Polimorfismos

Com o intuito de encontrar transcritos quiméricos polimórficos entre as linhagens e entre populações, e verificar se as inserções associadas às respectivas quimeras estavam presentes no genoma de todos genótipos, foi desenvolvido um *script bash ad-hoc*, aliado à função *intersect* do *bedtools* (Quinlan; Hall, 2010). Foram criadas três listas para cada genoma para identificar genes com TEs localizados *upstream* (2kb), *downstream* (2kb) e na região gênica. Essas listas foram cruzadas a fim de encontrar inserções presentes em um genoma (lista) e ausente em outro genoma (lista).

3.5 Análise de Expressão Diferencial

Os dados de expressão diferencial dos genes entre as linhagens de *D. melanogaster* e *D. simulans* foram obtidos por (Merenciano et al., dados não publicados). Brevemente, os dados de RNA-seq foram alinhados nos respectivos genomas das linhagens com o *software* STAR (Dobin et al., 2013), e a quantificação da expressão gênica foi realizada com o *software* htseq (Anders et al., 2015), parâmetro -t exon, o qual realiza a soma das leituras alinhadas em todos os éxons de cada gene, e atribui o valor resultante à expressão do mesmo. Posteriormente, os genes foram analisados com o pacote DEseq2 (Love 2014), no qual a normalização se deu entre os genótipos e as subseqüentes comparações pareadas entre as condições de controle e tratados de cada linhagem, assim como entre as populações em cada tratamento. Foram considerados genes diferencialmente expressos aqueles com valor P ajustado (FDR) $< 0,05$, e com valores de *Log2 Fold Change* > 1 e < -1 .

3.6 Estatísticas Adicionais e Ilustrações

As análises de distribuição normal Shapiro-Wilk (1965) dos dados, as médias e erros padrão calculados, as análises de variância ANOVA OneWay foram realizadas utilizando os pacotes estatísticos disponíveis no BioEstat 5.3 (Ayres et al, 2003). As análises de correlação de Pearson e os gráficos construídos no presente trabalho foram produzidos por meio de pacotes disponibilizados no R Studio.

CAPÍTULO 1

Caracterização de Transcritos Quiméricos

4 RESULTADOS

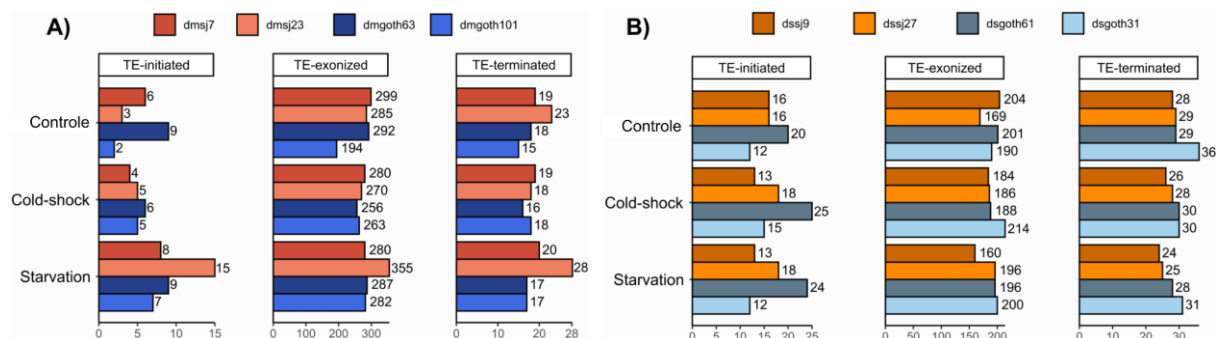
4.1 A exposição ao estresse não parece resultar em maior produção de transcritos quiméricos

Para determinar se a produção de transcritos quiméricos é influenciada pelos tratamentos, ou seja, se a resposta ao estresse é otimizada pelo aumento ou diminuição da cotranscrição de TEs com genes, foi calculada a média de transcritos quiméricos produzidos entre as condições. Nos transcriptomas das linhagens de *D. melanogaster* provenientes de Gotheron e São José do Rio Preto, a média de transcritos quiméricos identificados foi de 291 no Controle, 290 em *Cold-Shock* e 331 em *Starvation*, com diferenças não significativas (ANOVA, $p = 0.3138$, $F = 1.3232$) (Tabela Suplementar 1). As médias dos transcritos das linhagens de *D. simulans* foi praticamente as mesmas (ANOVA, $p = 0.8435$, $F = 0.1737$), 237 transcritos no Controle, 239 em *Cold-Shock* e 231 em *Starvation* (Tabela Suplementar 2) indicando que nas duas espécies, os dois tratamentos não afetaram significativamente a produção de transcritos quiméricos.

4.2 TEs inseridos dentro de genes correspondem a cerca de 92% dos transcritos quiméricos em *D. melanogaster* e 80% em *D. simulans*

A contagem do número de transcritos quiméricos por região gênica (*upstream*, *downstream* ou dentro do gene) permitiu analisar se o local de inserção do TE afeta a produção de transcritos quiméricos. Em *D. melanogaster*, observou-se que a maioria desses transcritos apresentam TEs inseridos dentro dos genes (*TE-exonized*) (Figura 3a), constituindo cerca de 92% do total de transcritos quiméricos produzidos. Ainda, nesta categoria, 69,7% foram do tipo *embedded*, 19,7% *intronic* e 10,7% *overlapped*. Transcritos da classe *TE-initiated* constituíram 2% do total, e *TE-terminated* 6%.

Figura 3. Número de transcritos quiméricos produzidos em cada linhagem de a) *D. melanogaster* e b) *D. simulans*, nas condições experimentais Controle, *Cold-Shock* e *Starvation*, classificados de acordo com a posição do TE no transcrito.

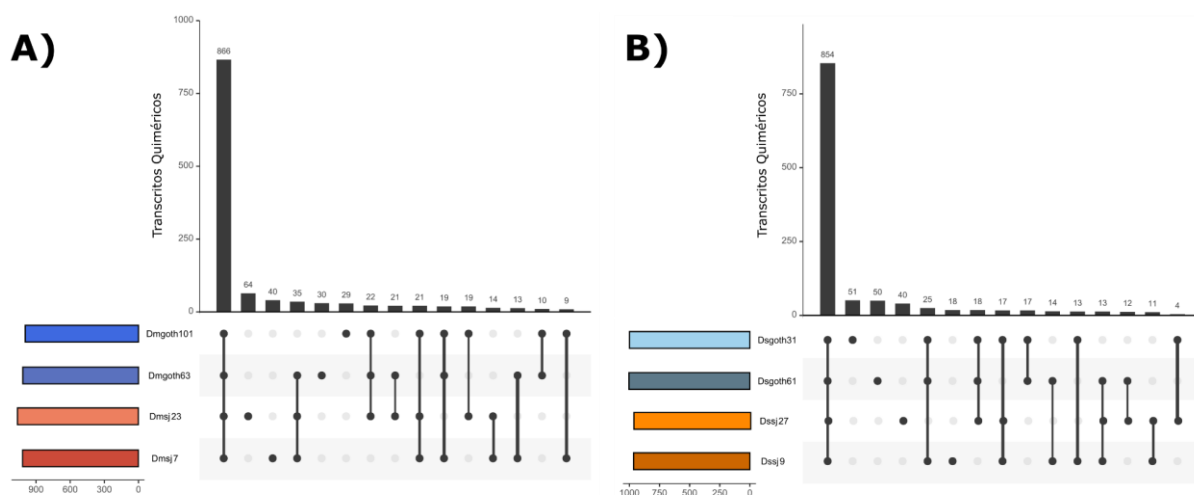


A análise de tipos de transcritos mais frequentes nas linhagens de *D. simulans* evidenciou algumas outras diferenças entre as espécies-irmãs: há mais transcritos do tipo *TE-initiated* e *TE-terminated* sendo produzidos (Figura 3b) do que *D. melanogaster*, subindo para 7% e 12%, respectivamente. Os transcritos do tipo *TE-exonized* continuam sendo os mais frequentes (81%), e dentre eles, 65,7% são *embedded*, 22,4% *intronic* e 11,9% *overlapped*.

Devido à natureza deletéria de inserções de TEs dentro de genes, que tendem a ser mais disruptivas, foi necessário identificar a exata localização das inserções que produzem o tipo mais frequente de transcrito quimérico *TE-exonized* (regiões não codificadoras de proteínas 5'UTR, 3'UTR e regiões codificantes CDSs). Dentre as inserções que produziram transcritos quiméricos do tipo *TE-exonized embedded*, totalmente dentro dos genes, em *D. melanogaster* cerca de 43% das inserções se encontram inseridas em CDSs e 45% estão inseridas em 3'UTRs. Curiosamente, em *D. simulans*, 79% estão em CDSs e 10 e 11% em 3'UTR e 5'UTR, respectivamente. Aparentemente, apesar de inserções que interrompem quadros de leituras serem eliminadas com mais frequência, a grande maioria das inserções de TEs que produzem transcritos quiméricos exonizados e inseridos inteiramente dentro de genes, são mantidas e estão inseridas em porções codificantes dos genes. Contudo, essas inserções potencialmente disruptivas, inseridas inteiramente dentro de porções de sequência codificadoras de proteínas (CDSs), ocorrem em apenas ~0,45% e ~0,6% do total de genes de *D. melanogaster* e *D. simulans*, respectivamente. Portanto, apesar da alta proporção de inserções quiméricas de TEs em regiões gênicas codificantes, essas são encontradas em menos de 1% do total de genes. Assim, a grande maioria de inserções de TEs que produzem transcritos quiméricos, se encontram em regiões não codificantes.

A fim de saber se a produção de transcritos quiméricos difere entre populações, os quais podem caracterizar a colonização de novos habitats e zonas climáticas em cada população, foram conferidos o número de transcritos comuns e exclusivos entre as linhagens das duas espécies. Do total de transcritos produzidos entre as condições e linhagens, cerca de 866 transcritos quiméricos são comuns nas linhagens de *D. melanogaster*, 14 comuns e exclusivos entre as linhagens brasileiras e 10 entre as linhagens francesas (Figura 4-a); Tabela Suplementar 3). As quatro linhagens de *D. simulans* compartilham um total de 854 transcritos quiméricos, 17 comuns e exclusivos das linhagens de Gotheron e 11 das linhagens de São José do Rio Preto (Figura 4-b); Tabela Suplementar 4). Enquanto que entre as linhagens de *D. melanogaster* a população de São José do Rio Preto apresentou mais transcritos quiméricos exclusivos, as linhagens de Gotheron de *D. simulans* exibiram uma maior produção de transcritos quiméricos polimórficos.

Figura 4. Diagrama Upset referente ao número de transcritos quiméricos compartilhados e exclusivos das linhagens de A) *D. melanogaster* e B) *D. simulans*.



4.3 ~1,5% dos genes de *D. melanogaster* produzem transcritos quiméricos

A identificação dos genes quiméricos permitiu quantificar a extensão da participação dos genes na produção de quimeras. Em *D. melanogaster*, existe uma média de 257 genes quiméricos por linhagem (Tabela 1), compondo cerca de 1,5 % do total de genes (Figura 5). Considerando as quatro linhagens de *D. melanogaster*, e excluindo-se as redundâncias, quando o mesmo gene produz transcritos em mais de uma condição e/ou linhagem, há um total de 401 genes da espécie formando transcritos quiméricos (Tabela Suplementar 5). Desses, 114 genes (28%) produziram dois ou mais transcritos quiméricos, a maioria constituintes de transcritos tipo *TE-exonized* (~90%). Ademais, alguns genes produziram mais de um transcrito com

múltiplas cópias de TE, dos quais 24% foram *Embebbbed*, 31% *Intronic* e 34% *Overlapped*. Apenas três genes *CR6900* em *Dmgoth101*, *Dmsj23* e *Dmsj7*, *CG31809* em *Dmsj23*, e *CG40470* em *Dmsj7*, produziram transcritos *TE-initiated* formados por mais que uma única cópia de TE. Entre os transcritos *TE-terminated*, somente os genes *NADH desidrogenase* (*CG40002*) em *Dmgoth63*, *Max* em *Dmsj23*, *six-banded* em *Dmsj7*, *CG14464* em *Dmsj7* produzem mais de um transcrito. Dentre estes, *CG14464* é o único que não apresenta inserções de TE da mesma família, próximas umas às outras, que podem configurar apenas uma única cópia e foram identificados, ao longo da etapa de anotação de elementos de transposição, como mais de um *hit* (Smit, 2004; Bailly-Bechet; Haudry; Lerat, 2014).

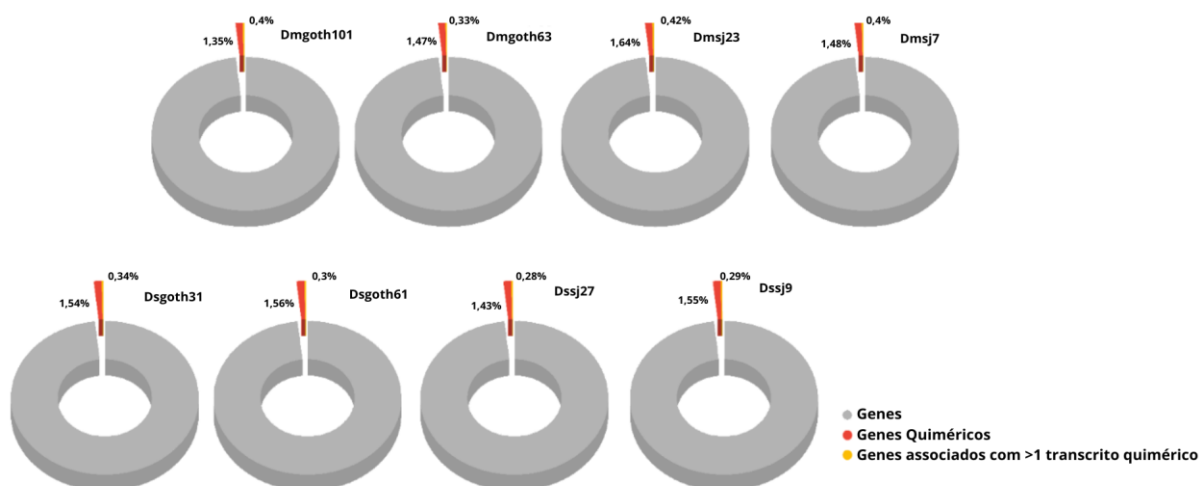
Tabela 1. Distribuição de genes quiméricos, total de transcritos quiméricos e os exclusivos entre as linhagens de *D. melanogaster* e *D. simulans*, dos quais puderam ser identificadas inserções exclusivas potencialmente adaptativas.

Linhagens	Genes quiméricos	Transcritos quiméricos	Transcritos quiméricos exclusivos	Localização das inserções quiméricas			Inserção genômica exclusiva
				Up	Gene	Down	
<i>D. melanogaster</i>	401						
Gotheron 101	233	368	29	2	26	1	4 (2 roo, 2 BS)
Gotheron 63	256	405	30	5	22	3	6 (4 roo, 2 412)
Gotheron *			10	1	8	1	16 (6 roo, 4 pogo, 1 FB, 1 17.6, 1 Bari1, 1 Burdock, 1 micropia, 1 Transpac)
SJRP 23	283	427	64	9	43	12	22 (8 roo, 4 P-element, 3 412, 1 micropia, 1 Stalker2, 1 jockey, 1 Doc3-element, 1 Juan, 1 3S18, 1 Rt1a)
SJRP 07	257	384	40	2	30	8	10 (2 roo, 2 POGON1, 1 pogo, 1 BS, 1 P-element, 1 Ivk, 1 mdg3, 1 G4)
SJRP *			14		14		10 (5 roo, 1 P-element, 1 Transpac, 1 1360, 1 Tabor, 1 Juan)
<i>D. simulans</i>	361						
Gotheron 31	233	298	51	5	40	6	22 (7 roo, 6 P-element, 2 Max-element, 2 transib2, 2 Juan, 1 flea, 1 hobo, 1 HB)
Gotheron 61	236	304	50	12	33	5	17 (6 roo, 4 P-element, 4 blood, 1 BS, 1 Tc1, 1 transib2)
Gotheron *			17	3	14		4 (3 roo, 1 TART-C)
SJRP 27	216	273	40	8	30	2	16 (7 roo, 2 transib2, 2 Doc, 1 micropia, 1 Tc1, 1 P-element, 1 INE-1, 1 G2, 1 hobo)
SJRP 09	235	267	18	1	16	1	6 (3 roo, 1 TART-A, 1 TART-B, 1 1731)
SJRP *			11	3	7	1	6 (3 roo, 1 Bar1, 1 TART-B, 1 TART-C)

Up: upstream ao gene, **Gene,** **Down:** downstream ao gene. * compartilhamento entre as linhagens das referidas populações (*D. melanogaster* Gothereon 66+101, SJRP 23 +07; *D. simulans* Gothereon 31+61, SJRP 27+09).

Do total de genes anotados nas linhagens de *D. simulans*, uma média de 230 genes por linhagem produziram transcritos quiméricos, isso corresponde em média a 1,5% do conteúdo gênico (Figura 5). Excluindo-se as redundâncias, 361 genes produziram todos os transcritos quiméricos identificados nas quatro linhagens (Tabela Suplementar 6), 40 genes a menos que o contabilizado em *D. melanogaster*. Dentre os genes quiméricos, 80 (22%) produziram dois ou mais transcritos quiméricos, especialmente da classe *TE-exonized* (~84%). Distribuindo-os entre as subclasses de transcritos *TE-exonized*, 23% dos transcritos do tipo *Embedded* são constituídos por mais de uma cópia de TE, cerca de 34% dos *Intronic* e 32% do tipo *Overlapped*. Como em *D. melanogaster*, tanto em transcritos *TE-initiated*, quanto em *TE-terminated*, foram encontrados alguns genes que produzem mais de um transcrito, formados por inserções de TE da mesma família, os quais podem configurar cópias únicas identificadas como mais de um *hit* durante a anotação genômica de TEs (Smit, 2004; Bailly-Bechet; Haudry; Lerat, 2014). Dentre os transcritos *TE-initiated* destacam-se *GD15406* (ortólogo de *Cyp6g1* em *D. melanogaster*) em *Dsgoth61*; já entre os *TE-terminated* *GD12257* (ortólogo de *CG8765*), *GD16079* (ortólogo de *Zpr1*) e *GD24119* (ortólogo de *CG5674*) formam mais de um transcrito.

Figura 5. Proporção de genes que produzem transcritos quiméricos nas quatro linhagens de *D. melanogaster* e *D. simulans*.



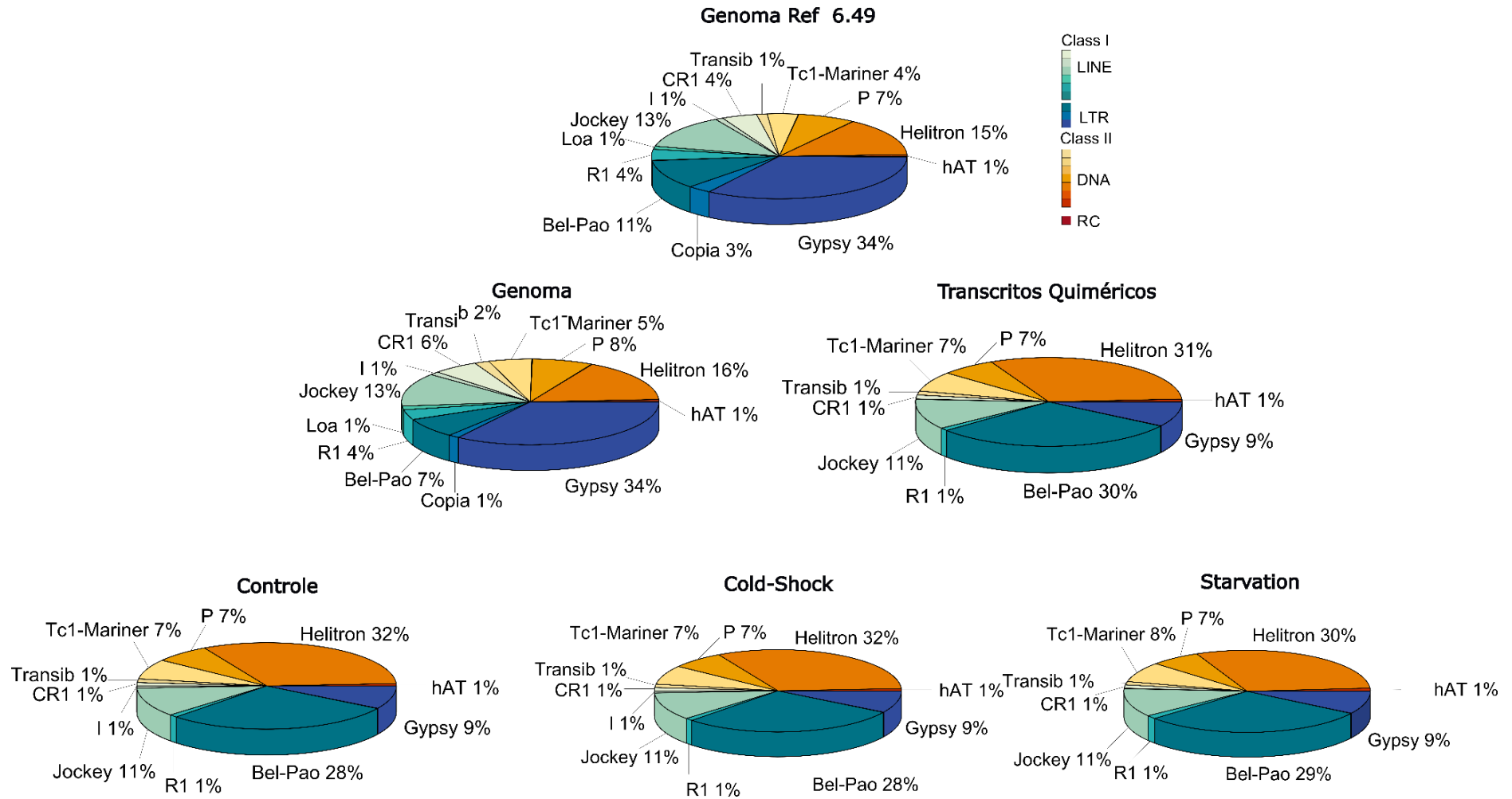
4.4 Elementos de Classe I são os mais frequentes em transcritos quiméricos

Para identificar se há prevalência de alguma família e superfamília na formação de transcritos quiméricos foi utilizada a lista de TEs consenso para *D. melanogaster* de Bergman (https://github.com/bergmanlab/drosophila-transposons/blob/master/current/D_mel_transposon_sequence_set.fa), classificados em superordem, superfamílias e famílias de

TEs. O número médio de cópias de TEs anotados nos genomas das quatro linhagens de *D. melanogaster* foi de 21.870 e entre as linhagens de *D. simulans* foi de 20.770. A ocorrência dos transcritos distribuídos entre as classificações de TEs e tipos de transcritos estão detalhados nas Tabela Suplementares 7 e 8.

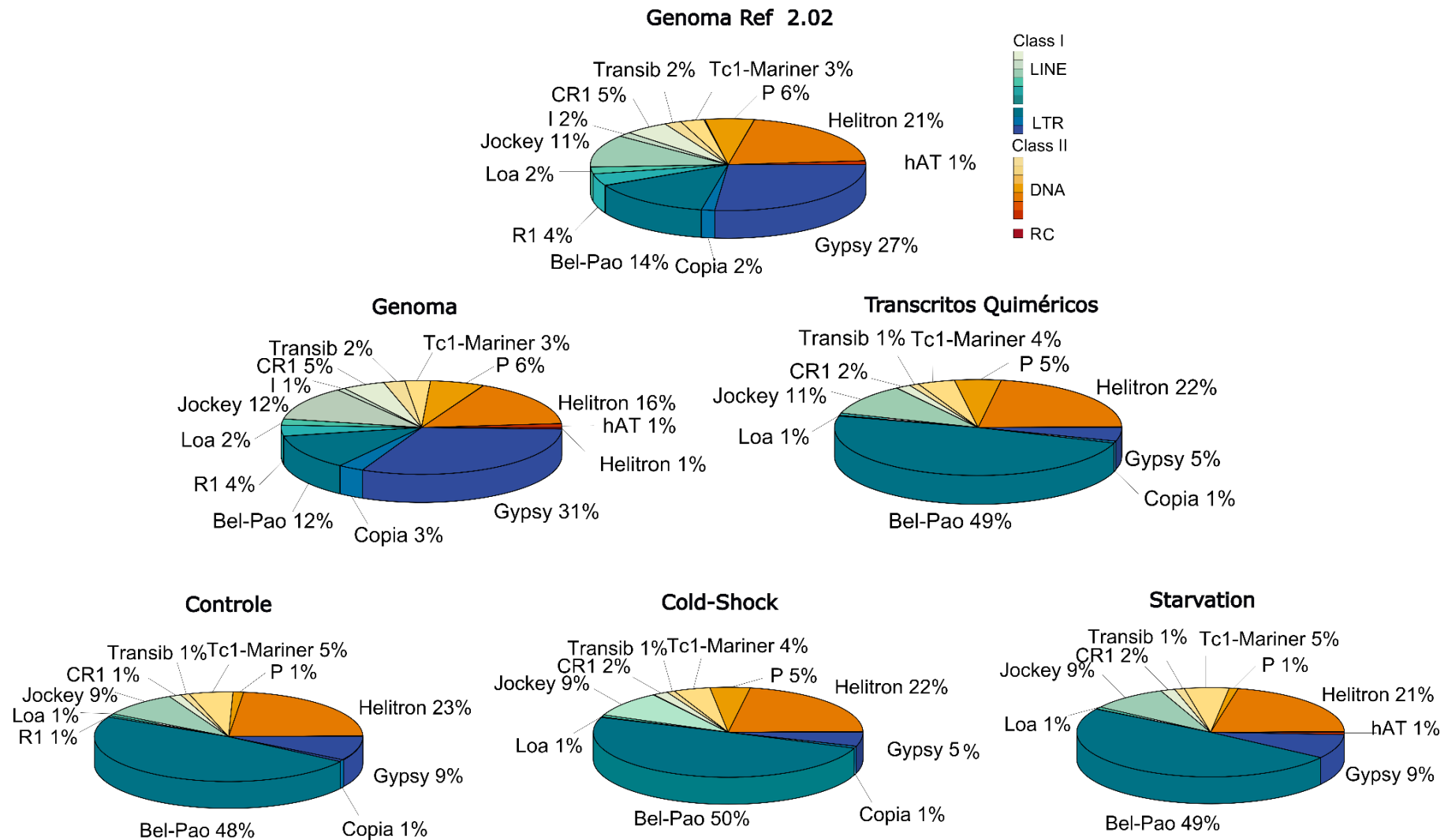
Todas as 17 superfamílias de TEs consenso foram encontradas no genoma de referência (r6.49) e nos genomas das linhagens analisadas de *D. melanogaster*; 12 delas presentes nos transcritos quiméricos (Figura 6). Os elementos de Classe I, que corresponderam a 68% do conteúdo genômico de TEs entre as linhagens, produziram o maior número de transcritos quiméricos (52,5%); dentre eles, os retrotransposons com LTR produziram 48,3% de toda a diversidade de famílias constituintes de transcritos quiméricos e 39% dos transcritos (43 famílias das superfamílias *Gypsy*, *Copia* e *Bel-Pao*). A superfamília *Bel-Pao* foi a segunda mais frequente em transcritos quiméricos (~30%). Apesar dos elementos *Gypsy* comporem 34% dos TEs genômicos, sua participação em transcritos quiméricos foi de apenas 9%. Ainda, na Classe I, os elementos LINE, produziram 29,2% das famílias quiméricas (26 famílias de superfamílias *CR1*, *I*, *Jockey* e *R1*). Elementos de Classe II, que compõe 32% do conteúdo genômico de TEs em *D. melanogaster*, constituíram 47,5% dos TEs que produziram transcritos quiméricos (20 famílias entre as superfamílias *hAT*, *Helitron*, *P*, *Tc1-Mariner* e *Transib*), sendo a superfamília *Helitron*, a mais frequente (~32%).

Figura 6. Frequência (%) de superfamílias de TEs, em número de cópias, identificadas no genoma de referência *D. melanogaster* (r6.49), nos genomas e na produção de transcritos quiméricos das quatro linhagens, distribuídos entre as condições Controle, *Cold-Shock* e *Starvation*.



Às 17 superfamílias de TEs consenso também foram encontradas no genoma referência de *D. simulans* (r2.02) e nos genomas das linhagens analisadas, mas somente representantes de 13 superfamílias produziram transcritos quiméricos (Figura 7). A contribuição genômica de elementos de Classe I continuou sendo a maior entre as linhagens de *D. simulans*, constituindo ~83% do número total de cópias. Os retrotransposons contribuíram com 67% na produção de transcritos quiméricos, sendo 54% do total formados pelos retrotransposons LTRs (24 famílias das superfamílias *Gypsy*, *Copia* e *Bel-Pao*). A superfamília *Bel-Pao* compõe ~50% das quimeras, quase 20% a mais que o observado em *D. melanogaster*. Novamente, quase 30% do conteúdo genômico de TEs em *D. simulans* é atribuído à família *Gypsy*, que formou entre 5-9% dos transcritos quiméricos. Os elementos LINE (21 famílias dentro das superfamílias *CR1*, *I*, *Jockey*, *Loa* e *R1*) produziram 13% de transcritos quiméricos. Os elementos de Classe II fazem parte de 17% do genoma, e ocorreram em 33% dos transcritos (13 famílias distribuídas entre as superfamílias *hAT*, *Helitron*, *P*, *Tc1-Mariner* e *Transib*). Dentre eles, *Helitron* constitui 22% das quimeras registradas, uma participação com cerca de 10% menor do que o relatado em *D. melanogaster*. Contrapondo-se a sua irmã *D. melanogaster*, *D. simulans* apresenta família representante de retrotransposons *Loa* na formação de seus transcritos.

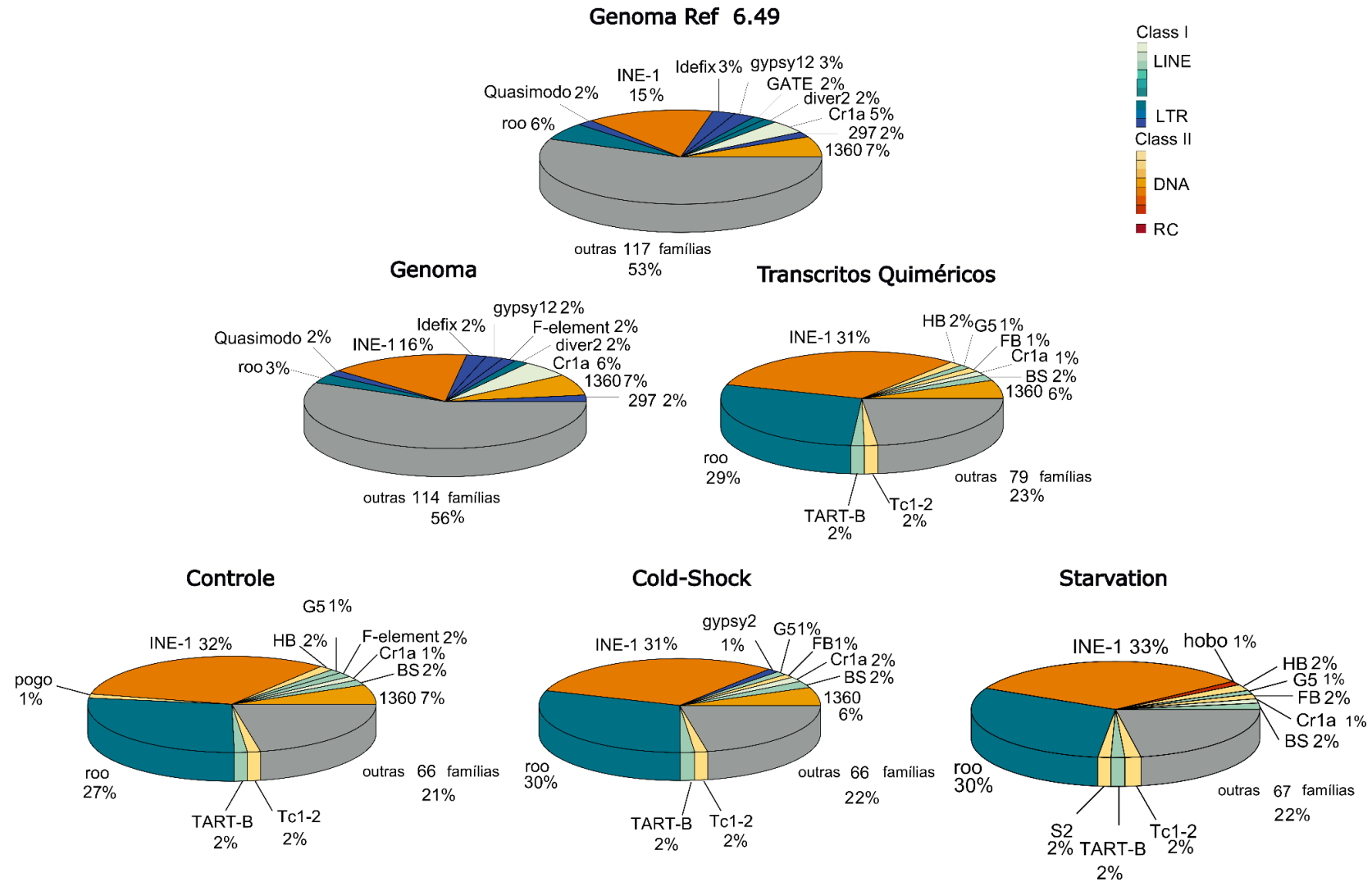
Figura 7. Frequência (%) de superfamílias, em número de cópias, identificadas no genoma de referência *D. simulans* (r2.02), nos genomas e transcriptomas das quatro linhagens, distribuídos entre as condições Controle, *Cold-Shock* e *Starvation*.



4.5 Elementos da família *INE-1* em *D. melanogaster* e *roo* em *D. simulans* são os maiores produtores de transcritos quiméricos

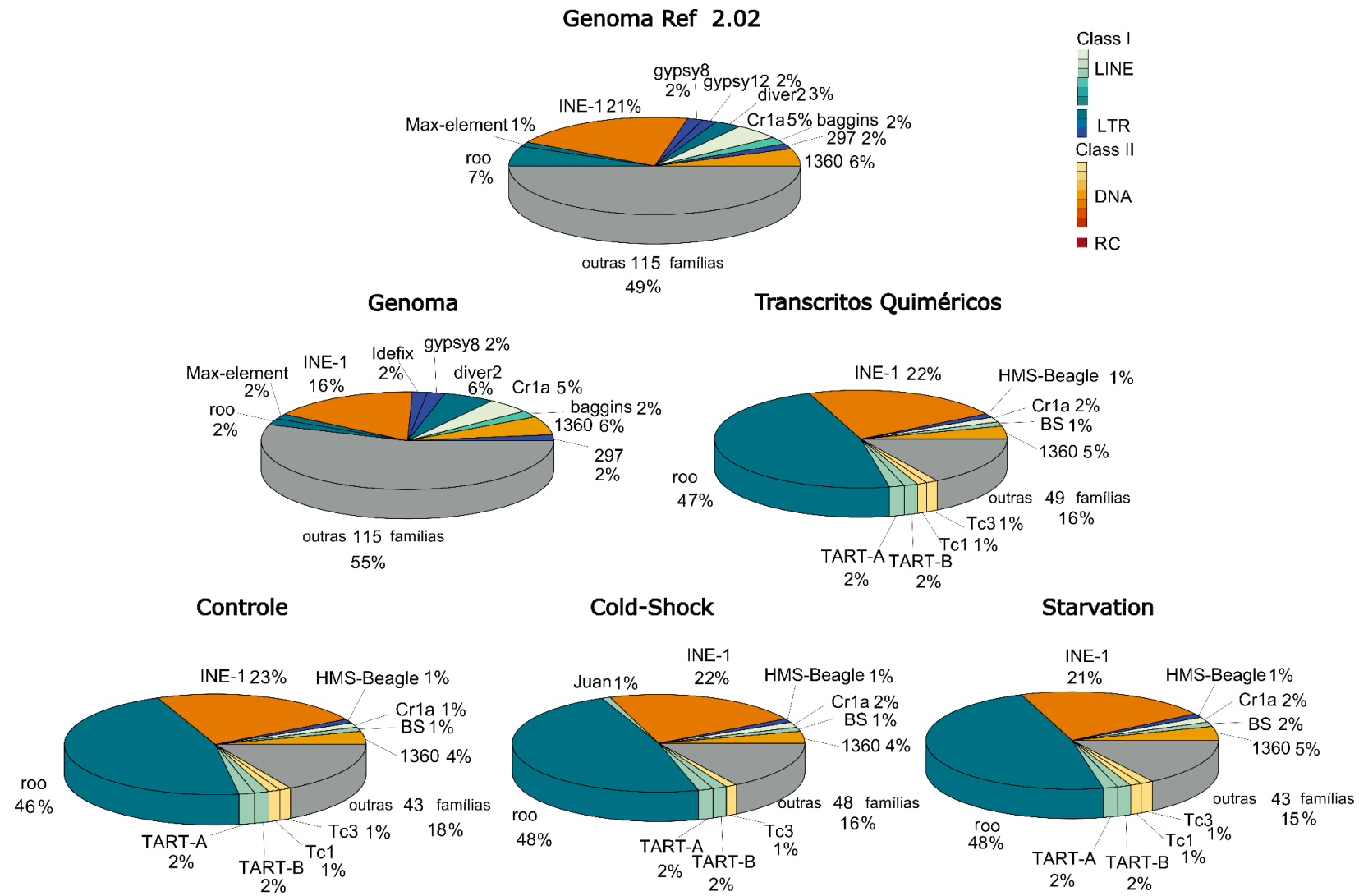
Das 128 famílias de TEs descritas em *D. melanogaster*, um total de 124 famílias foram anotadas nas quatro linhagens; desse total, 89 produziram transcritos quiméricos. As 10 famílias mais frequentes no genoma e no repertório de transcritos quiméricos das linhagens de *D. melanogaster* estão representadas na Figura 8. Os elementos das famílias *INE-1* (superfamília *Helitron*) são os mais abundantes no genoma e foram os mais frequentes em quimeras. Também as famílias *Cr1a* e *1360* aparecem com uma proporção significativa dentro do genoma e na produção de quimeras. A família *roo* (superfamília *Bel-Pao*) compõe apenas cerca de 3% do genoma, mas apresenta uma contribuição quase dez vezes maior na produção de transcritos quiméricos. As outras seis famílias mais frequentes na produção de transcritos entre as três condições se alternam entre 10, são elas: *HB*, *G5*, *BS*, *Gypsy2*, *hobo*, *pogo*, *F-element*, *TART-B*, *Tc1-2* e *S2*. Analisando essas 10 famílias, pode-se visualizar com mais clareza a contribuição das outras duas superfamílias mais frequentes comentadas, já que entre as 10, quatro são *Jockey* e cinco pertencem a *Tc1-Mariner*.

Figura 8. Frequência (%) de famílias, em número de cópias, identificadas no genoma de referência *D. melanogaster* (r6.49), nos genomas e transcriptomas das quatro linhagens, distribuídos entre as condições Controle, Cold-Shock e Starvation.



Foram anotadas 125 famílias de TEs no genoma de referência de *D. simulans* (r2.02), sendo o mesmo valor anotado, em média, nos genomas das quatro linhagens. Desse total, 59 famílias estavam presentes em quimeras. Como em *D. melanogaster*, a família *INE-1*, mais frequente do genoma, mantêm a mesma participação no repertório de transcritos quiméricos, mas perde a posição de maior participante na produção de transcritos. A família *roo* tem uma contribuição no repertório de transcritos quiméricos mais de vinte vezes maior que sua proporção genômica (Figura 9). As terceira e quarta famílias mais frequentes no genoma, *Cr1a* e *1360*, mantêm a mesma posição no transcriptoma; enquanto que as outras seis famílias mais frequentes na produção de transcritos oscilam entre *HS-Beagle*, *BS*, *TART-A*, *TART-B*, *Tc1* e *Tc3*, sendo apenas *BS* e *TART-B* presentes nessa categoria em *D. melanogaster*.

Figura 9. Frequência (%) de famílias, em número de cópias, identificadas no genoma de referência *D. simulans* (2.02), nos genomas e transcriptomas das quatro linhagens, distribuídos entre as condições Controle, *Cold-Shock* e *Starvation*.



As famílias de TEs foram distribuídas entre as classes de transcritos quiméricos para identificar se havia prevalência de famílias com inserção *upstream*, *downstream* e dentro do gene. Em *D. melanogaster* (Figura 10), com exceção de transcritos *TE-initiated*, foi observada a mesma contribuição das referidas 10 famílias mais abundantes no genoma, sendo *INE-1*, *roo* e *1360* as maiores produtoras de quimeras. As famílias *S2* e *hobo* tiveram as maiores frequências na produção de transcritos *TE-initiated*, e as outras nove famílias ocorreram entre 5 a 10% dos transcritos. Dentre os transcritos *TE-exonized*, *INE-1* e *roo* possuem maior participação (~30%), enquanto *1360* contribui com ~5%; todas as outras 17 famílias não chegaram a 1% de participação. Dentre os transcritos *TE-terminated*, *INE-1* é sempre a família mais frequente (15 a 20%), seguida de *1360* (~15%) e *BS* (~10%). Destaca-se a ausência do retrotransposons *roo* nessa categoria de transcrito. Retirando-se as duas famílias mais frequentes (*INE-1* e *roo*), é revelada a diversidade das famílias constituintes de quimeras entre cada condição de cada linhagem (Figura 11), destacando a abundância das outras famílias presentes nos transcriptomas das linhagens (Figuras Suplementares 1 e 2).

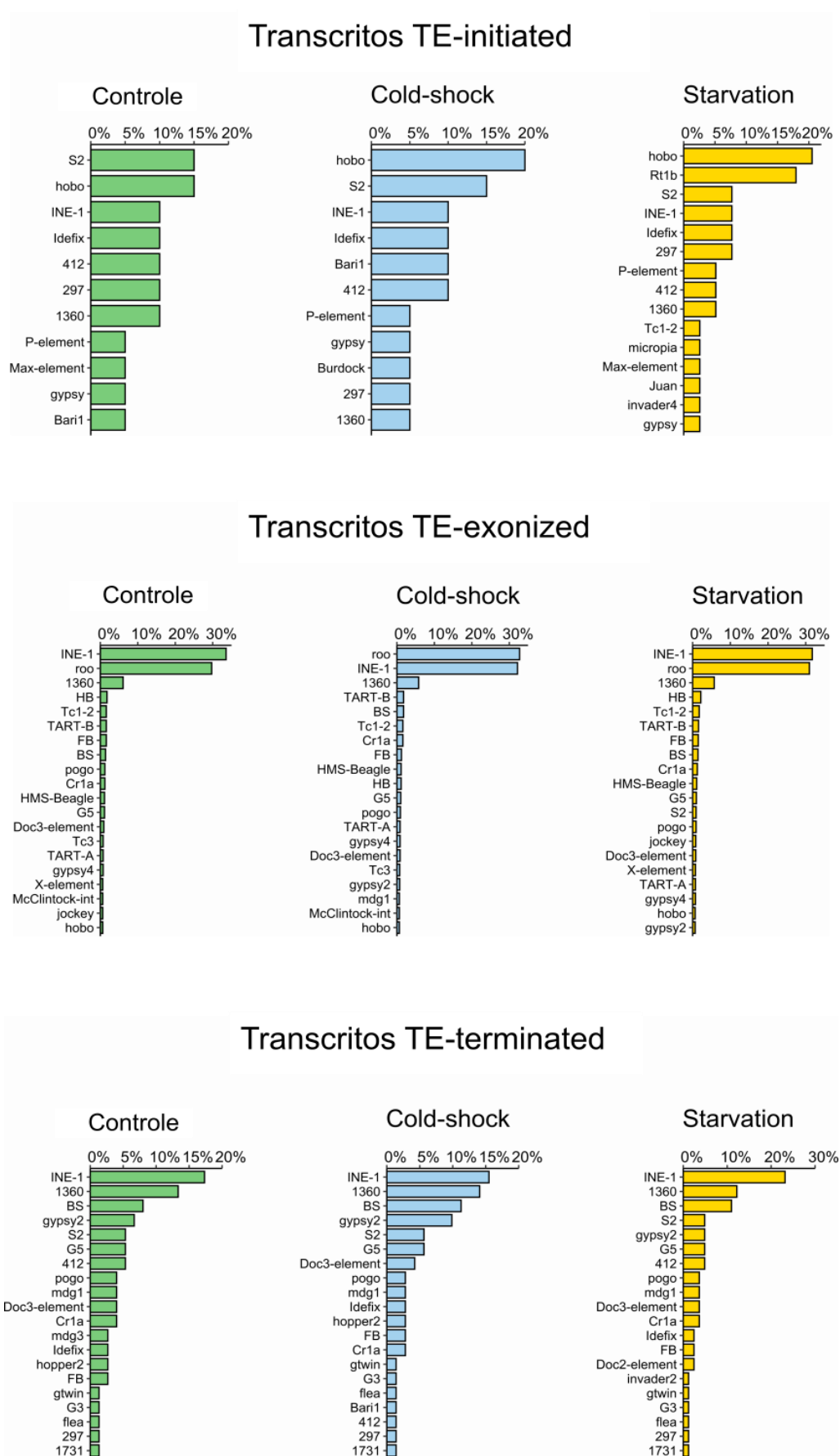
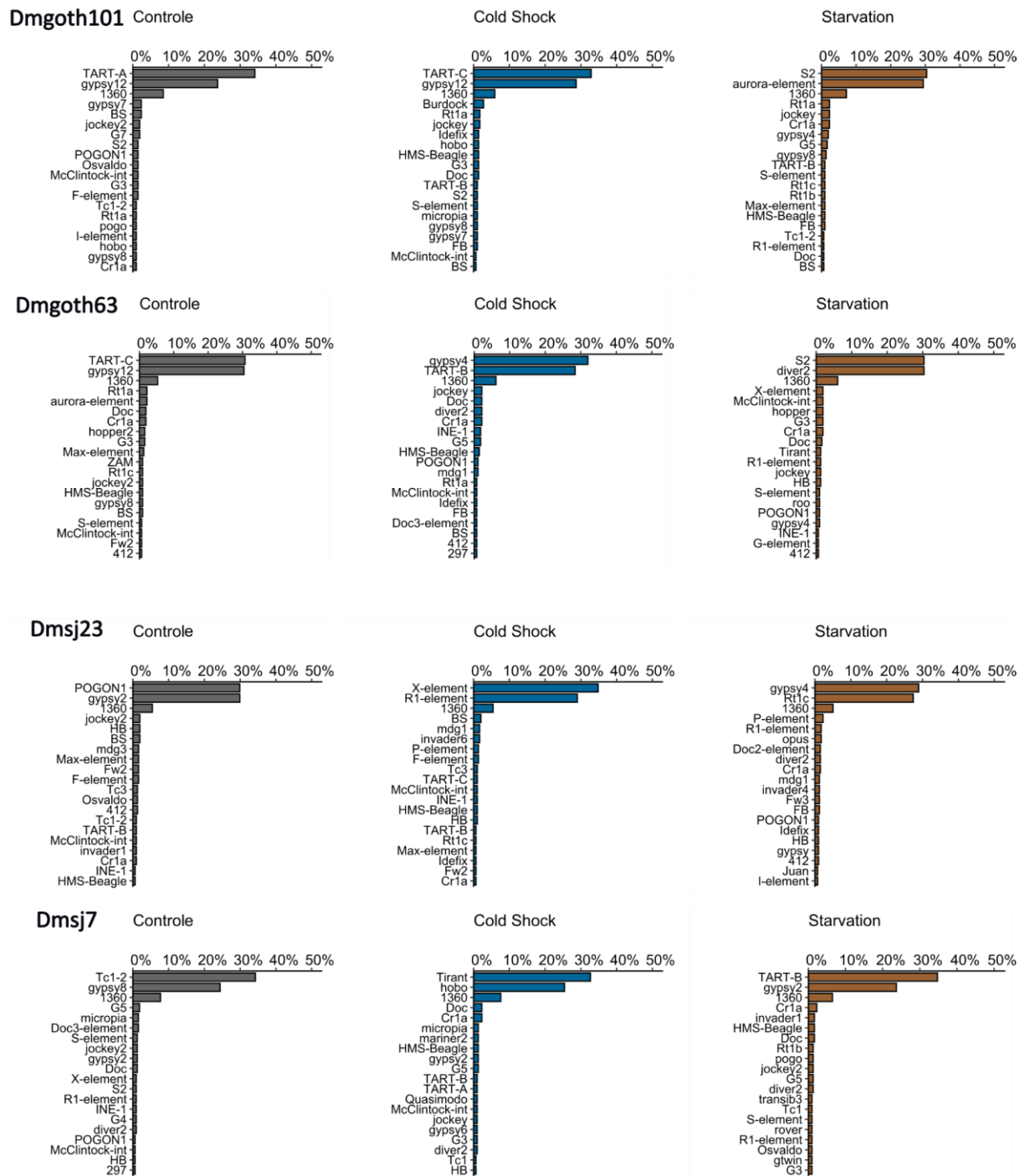
Figura 10. Frequência de famílias de TEs que constituem quimeras, por classe de transcrito, em *D. melanogaster*.

Figura 11. Distribuição das famílias mais frequentes entre condições experimentais e as linhagens analisadas de *D. melanogaster* após a exclusão das duas famílias mais frequentes (*INE-1* e *roo*) em transcritos quiméricos.



Também em *D. simulans*, as três famílias de TE mais abundantes no genoma (*roo*, *INE-1* e *1360*) foram as que mais contribuíram na formação de quimeras nos três tipos de transcritos (Figura 12). Em transcritos *TE-initiated* (~20%) e *TE-terminated* (~50%) a família *INE-1* apresenta a maior frequência, enquanto *roo* ocupa a primeira posição dentre os *TE-exonized* (~60%). *INE-1* e *1360* exibem uma parcela considerável nos três tipos de transcritos, *Tc3* entra com participação significativa entre transcritos *TE-terminated*, sendo que *roo*, assim como em *D. melanogaster*, está ausente apenas nessa classe de transcrito. Diferentemente de *D. melanogaster*, após a retirada das duas famílias mais frequentes (Figura 13), observa-se nos transcriptomas de *D. simulans* um predomínio de outras quatro famílias mais frequentes nas quimeras, que se alternam: *Gypsy11*, *diver2*, *invader1* e *gwtin* (Figuras Suplementares 3 e 4).

Dentre as famílias presentes em quimeras exclusivas de cada espécie, *D. melanogaster* contém 42 famílias exclusivas, ausentes nos transcriptomas de *D. simulans*; enquanto que essa espécie exibe 13 famílias originando quimeras, ausentes nos transcriptomas das linhagens de *D. melanogaster*. Com exceção das famílias *POGON1* (*DNA/Tc1-Mariner*) e *R1-element* (*LINE/R1*), produtoras de quimeras em *D. melanogaster*, mas ausentes nos genomas de *D. simulans*, e *R1-A1-element* (*LINE/R1*), produtora de quimera nas linhagens de *D. simulans* e ausente nos genomas dos genótipos de *D. melanogaster*, todas as outras famílias TE quiméricas foram identificadas dos genomas de ambas as espécies.

Investigando a tolerância de inserções em CDSs e as protegidas nas regiões regulatórias gênicas (5'UTR e 3'UTR), foram assinaladas aos transcritos do tipo *TE-exonized embedded* as famílias de TE mais frequentes em cada região. Em *D. melanogaster* a família *roo* compreende 81% das inserções em CDSs (média de tamanho das inserções = 112bp) e 86% em 5UTR'S ($\mu=115$ bp), enquanto que na região 3'UTR, as inserções *INE-1* (56% - $\mu=169$ bp) foram as mais frequentes. Em *D. simulans* a família *roo* foi a mais frequente em todas as regiões, compondo 83% das inserções em CDSs ($\mu=107$ bp), 79 % em 5'UTRs ($\mu=105$ bp) e 46% em 3'UTRs ($\mu=105$ bp). A análoga quantidade de inserções entre 3'UTR e CDSs em *D. melanogaster* não é verdadeira em sua composição, enquanto *roo* é mais tolerado interrompendo o quadro de leitura gênica, a família *INE-1* é protegida na região 3'UTR. Em *D. simulans* *roo* prevalece em todas as regiões, mas ainda é possível encontrar uma parcela significativa de *INE-1* na região 3'UTR (31% - $\mu=108$ bp). A família *roo* parece ser a mais conservada em genes, provavelmente por não configurar alterações desvantajosas na proteína final.

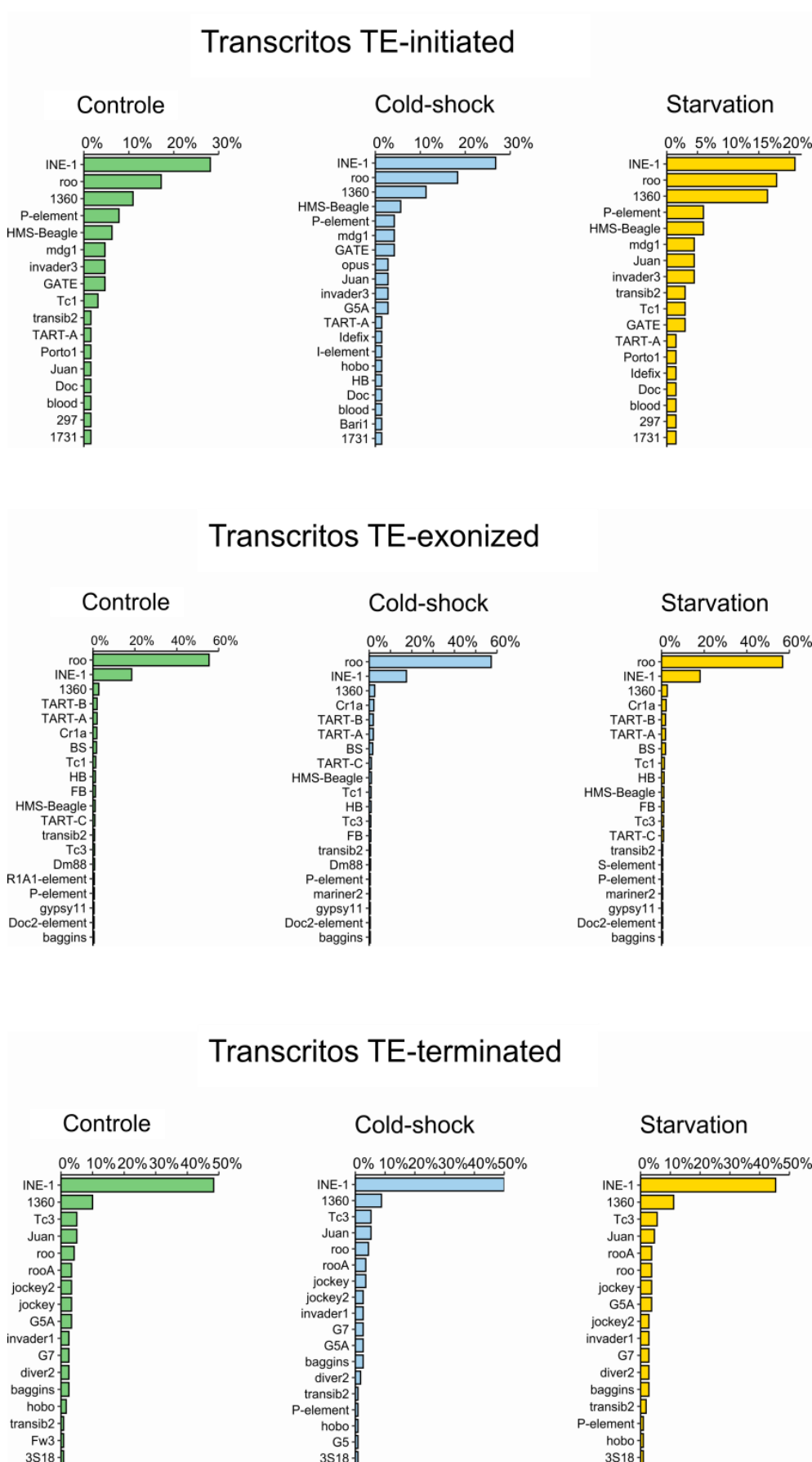
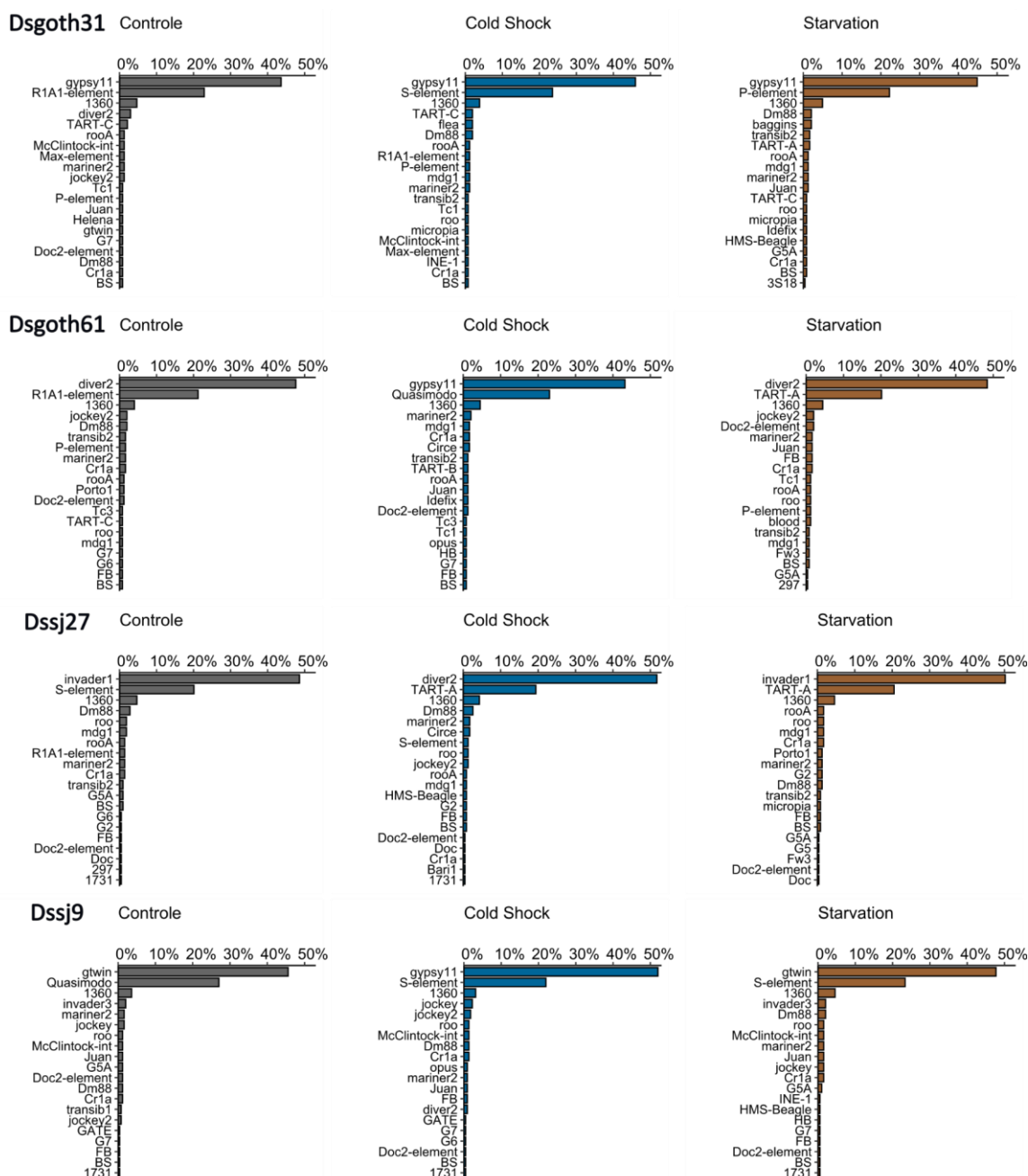
Figura 12. Frequência de famílias de TE que constituem quimeras, por classe de transcrito, em *D. simulans*.

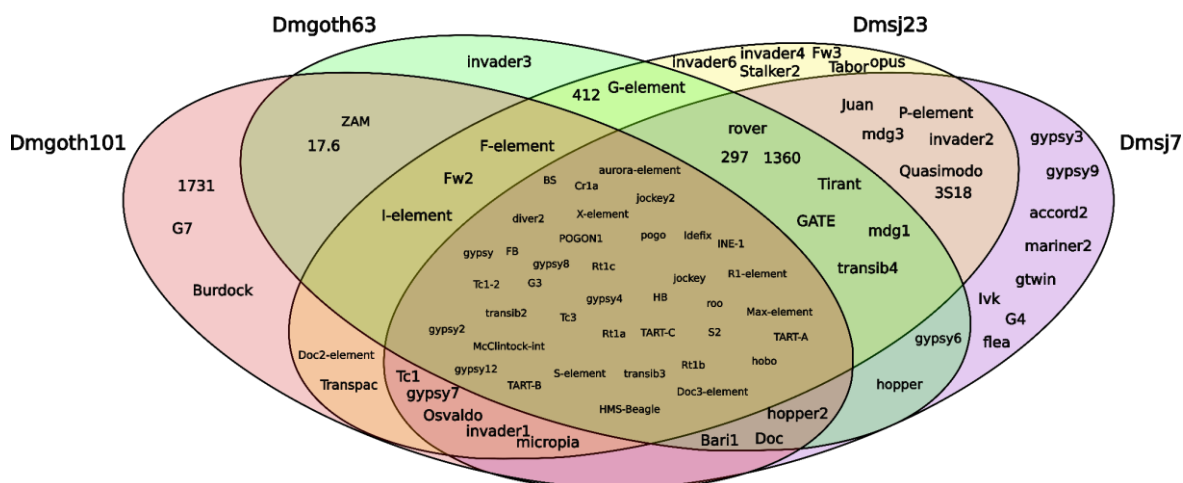
Figura 13. Distribuição das famílias mais frequentes entre condições experimentais e as linhagens analisadas de *D. simulans* após a retirada das duas famílias mais frequentes (*INE-1* e *roo*) em transcritos quiméricos.



Os diagramas de Venn (Figuras 14 e 15) revelam a diversidade de famílias de TE produtoras de transcritos quiméricos, exclusivas e compartilhadas entre as linhagens. Em *D. melanogaster*, das 89 famílias produtoras de quimeras, 38 (43%) produzem transcritos quiméricos nas quatro linhagens, 20 famílias são exclusivas das populações de São José do Rio Preto, e seis produziram transcritos quiméricos apenas em Gotheron. Das 20 famílias TE exclusivas dos transcritos das linhagens brasileiras, seis são compartilhadas pelos dois genótipos, seis são exclusivas de Dmsj23 e oito exclusivas de Dmsj7. Já, entre as seis

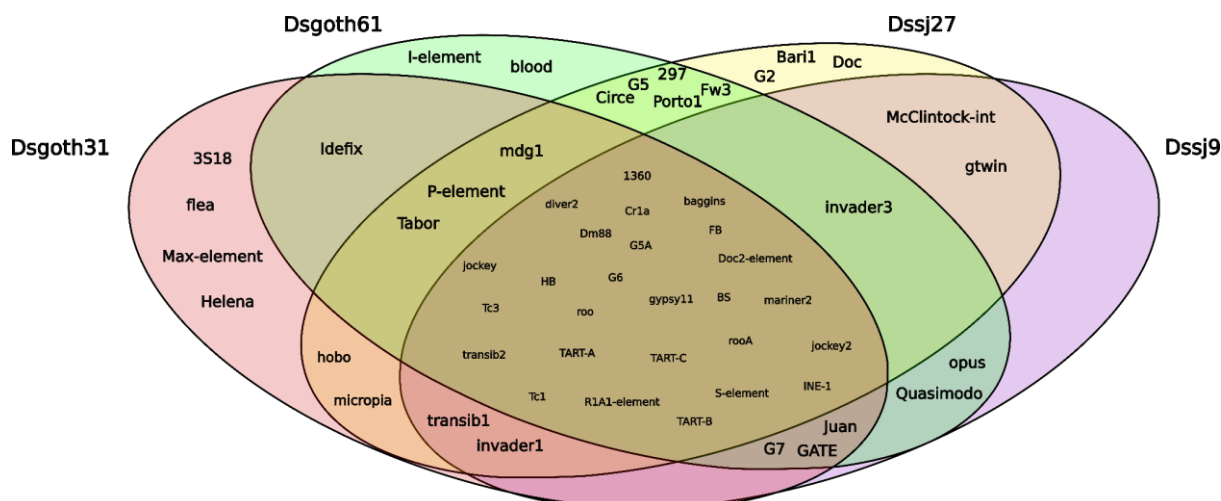
famílias exclusivas das linhagens europeias, duas são compartilhadas por ambas linhagens, três exclusivas de Dmgoth101 e uma exclusiva de Dmgoth63.

Figura 14. Diagrama de Venn exibindo as famílias compartilhadas na produção de quimeras entre as linhagens de *D. melanogaster*.



Como já descrito, há um número menor de famílias TE produzindo quimeras em *D. simulans*, contudo, desse total, a porcentagem de famílias compartilhadas pelas quatro linhagens é similar à *D. melanogaster*; assim, das 59 famílias identificadas, 26 (44%) produzem quimeras em todos os genótipos. As linhagens de origem brasileira exibem cinco famílias TE exclusivas, dentre as quais, duas são compartilhadas entre as linhagens e três são exclusivas de Dssj27. Os genótipos europeus exibem sete famílias exclusivas, com apenas uma compartilhada entre eles, quatro exclusivas de Dsgoth31 e duas exclusivas de Dsgoth61.

Figura 15. Gráfico de Venn exibindo as famílias compartilhadas na produção de quimeras entre as linhagens de *D. simulans*.



4.6 As famílias de TEs mais abundantes no genoma são as maiores produtoras de transcritos quiméricos no transcriptoma

A fim de testar se havia uma correlação entre a frequência das famílias de TE que formam quimeras e seu conteúdo genômico, se existiam famílias de TEs que produziam mais transcritos quiméricos apesar de sua abundância genômica, foi realizado o teste de correlação de Pearson para as quatro linhagens nas três condições de *D. melanogaster* e *D. simulans* utilizando-se o número de cópias (Figuras 16 e 17 e Figuras Suplementares 5 e 10, respectivamente para *D. melanogaster* e *D. simulans*) e o número de nucleotídeos (Figuras Suplementares 6 e 7 e 11 e 12, respectivamente para as duas espécies) para quantificar a abundância genômica dessas famílias (Figuras Suplementares 8 e 9 e 13 e 14, respectivamente para as duas espécies).

Na Figura 16 ilustra-se as correlações positivas entre o número de cópias nos genomas e a frequência dos respectivos TEs nos transcritos quiméricos. Os *outliers* observados correspondem às famílias de TE mais abundantes no genoma, *INE-1*, *roo* e *1360*, que consequentemente contribuíram mais com a formação de quimeras. Após exclusão desses *outliers*, observa-se uma distribuição mais ampla das demais famílias (Figuras Suplementares 5 e 7) e redução dos valores de R, resultando em valor de *p* não significativo. O mesmo ocorre quando analisados somente os *outliers*, ou seja, entre as famílias mais abundantes, não há uma família em particular que se destaca sendo a mais abundante e a maior produtora de transcritos quiméricos (Figuras Suplementares 8 e 9, 13 e 14, respectivamente para *D. melanogaster* e *D. simulans*). Esses resultados revelam que a correlação positiva só é verdadeira para as três famílias TEs mais comuns no genoma, sem nenhuma em particular que se sobressai, que de fato produzem mais quimeras, em relação ao restante das famílias, que não exibiram preferência na produção de transcritos quiméricos apesar de sua abundância genômica. Este padrão foi observado em todas as linhagens e condições analisadas de *D. melanogaster*. O mesmo foi encontrado em *D. simulans* (Figura 17).

Figura 16. Correlação de Pearson entre o número de cópias das famílias de elementos de transposição no genoma e a frequência das famílias que mais formam quimeras em *D. melanogaster*.

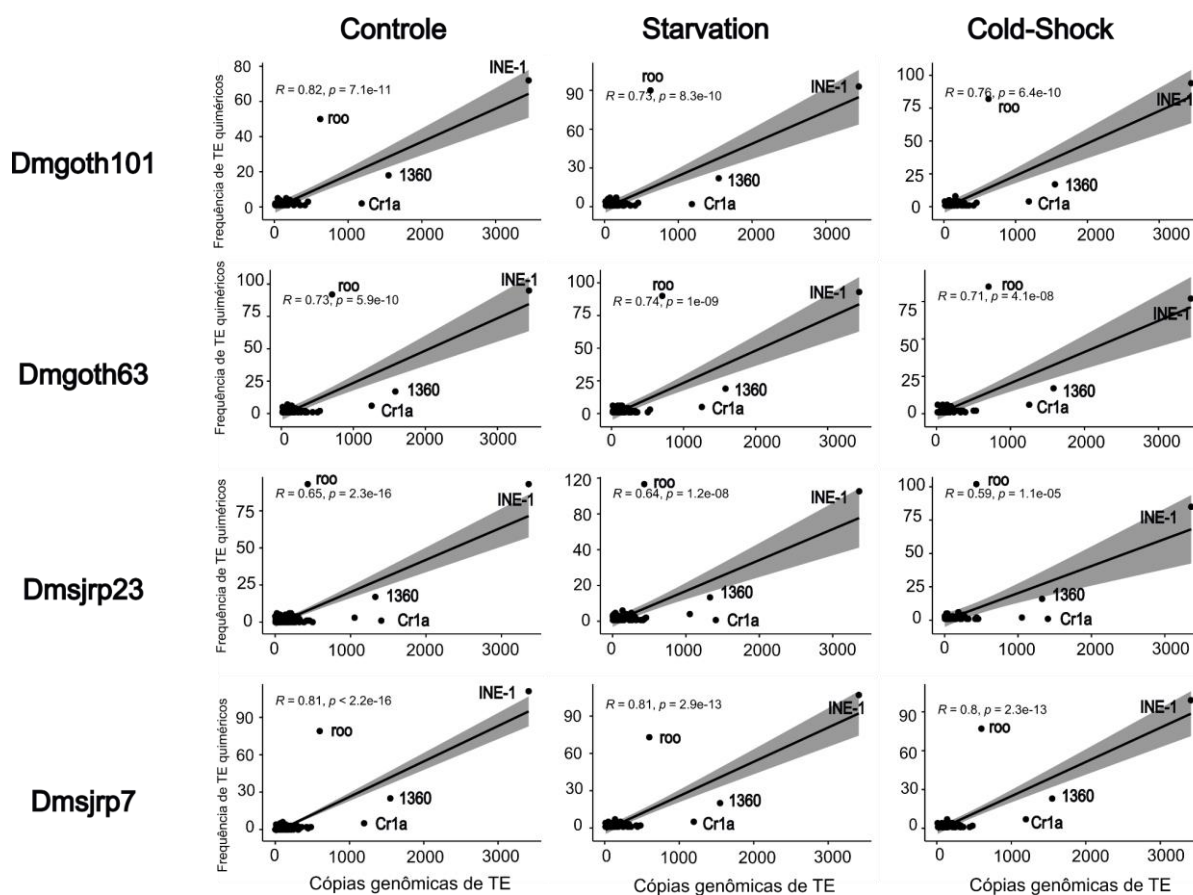
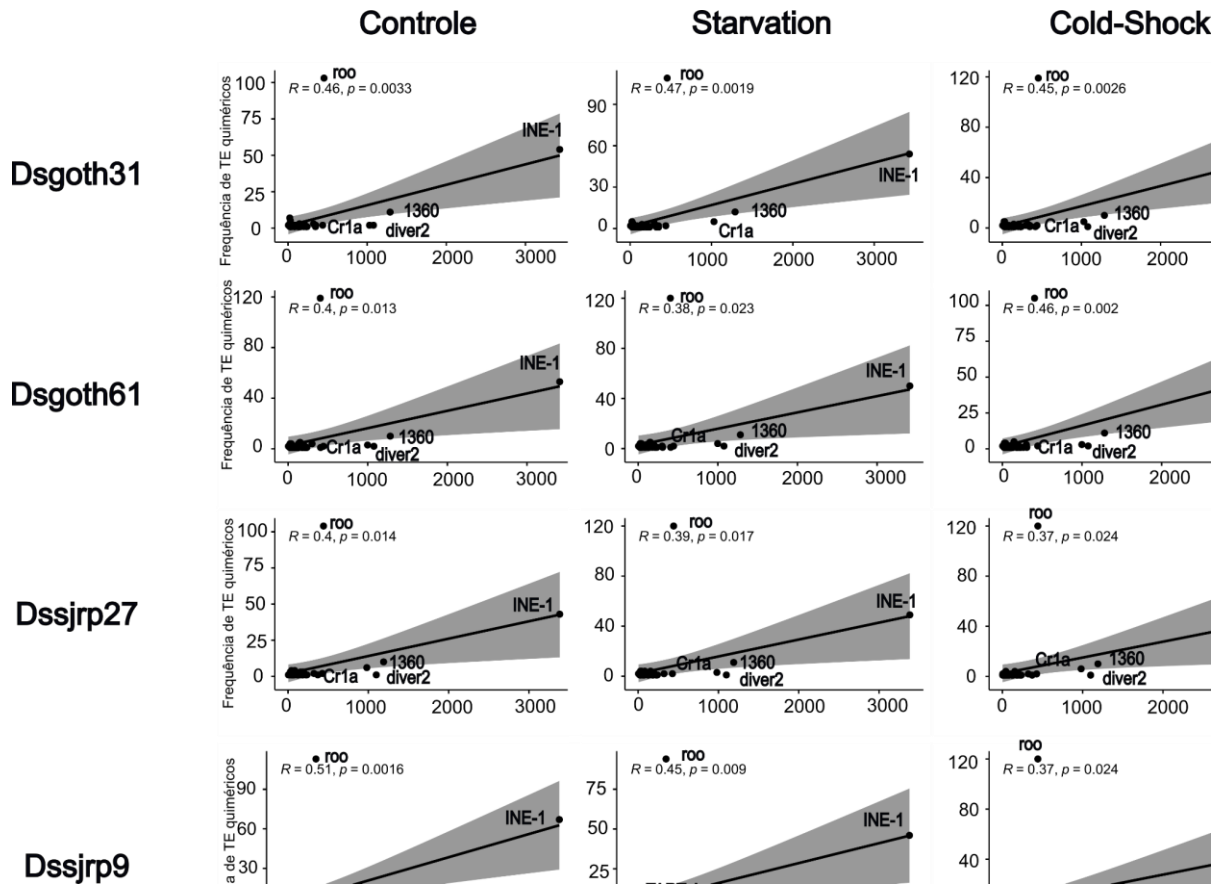


Figura 17. Correlação de Pearson entre o número de cópias das famílias de elementos de transposição no genoma e a frequência das famílias que mais formam quimeras em *D. simulans*.



4.7 As linhagens de *D. melanogaster* apresentam mais inserções polimórficas que *D. simulans*

Com o objetivo de caracterizar a divergência entre as populações, decorrente de eventos de mobilização e do recrutamento dos TEs nos transcritos, foram identificadas as inserções fixas e polimórficas e os transcritos quiméricos polimórficos. Assim como descrito anteriormente sobre a frequência de famílias exclusivas nas linhagens das espécies irmãs, o mesmo foi observado quando analisados transcritos quiméricos e inserções exclusivas. Ao todo foram identificados 180 transcritos polimórficos tanto nas linhagens de *D. melanogaster* quanto nas de *D. simulans*. Em *D. melanogaster* as linhagens de São José do Rio Preto exibiram mais transcritos quiméricos e inserções genômicas exclusivas, o mesmo é verdadeiro em *D. simulans* para as linhagens de Gothenon (Tabela 1, Tabelas Suplementares 9 e 10).

A família *roo* é a maior produtora de transcritos quiméricos exclusivos e inserções polimórficas nas linhagens das duas espécies, seguido de *INE-1*, as duas famílias mais abundantes no genoma. Em *D. melanogaster* a terceira família que mais produziu transcritos quiméricos exclusivos foi a 1360 (10 transcritos), a terceira família mais frequente no genoma e na produção de transcritos quiméricos, que em *D. simulans* divide a quarta posição com *transib2*, contribuindo com oito transcritos. Entre as linhagens de *D. simulans*, inserções de elementos *P* (12 transcritos) ocupam a terceira posição, inseridos preferencialmente *upstream* ao gene (sete transcritos, dos quais seis presentes nas linhagens de Gotheeron). Em *D. melanogaster* essa família produziu sete transcritos exclusivos. A grande maioria das inserções dos transcritos quiméricos exclusivos e inserções TE exclusivas se encontram dentro de genes. Juntas, as linhagens de *D. melanogaster* apresentam uma diversidade de 53 famílias de TEs compondo transcritos quiméricos exclusivos, das quais 22 não produziram transcritos quiméricos em *D. simulans*. Já, entre as linhagens de *D. simulans*, 35 famílias formaram transcritos quiméricos exclusivos, das quais sete estão ausentes na produção de transcritos em *D. melanogaster*.

CAPÍTULO 2

Efeitos dos TEs na Expressão de Genes Adjacentes

5 Resultados

5.1 Expressão diferencial entre controle e tratamentos e inserções de TEs potencialmente adaptativas em *D. melanogaster*

Para identificar possíveis inserções adaptativas de TEs relacionadas ao estresse em *D. melanogaster*, foram realizadas análises de expressão diferencial entre os tratamentos e o controle. Foram selecionados como candidatos adaptativos aqueles genes que produzem transcritos quiméricos e são diferencialmente expressos (DE). Os genes constituintes de quimeras que são ao mesmo tempo DEs são raros e, conseqüentemente, um maior número de quimeras DE é dependente do maior número de genes diferencialmente expressos (DEG). Esta abordagem permitiu investigar se a mudança de expressão observada está ou não relacionada à produção do transcrito quimérico, e conseqüentemente, à inserção do TE. Na Figura 18, que representa a expressão normalizada dos genes que produziram transcritos quiméricos, é possível identificar os genes sub e super expressos em uma das condições, distantes da diagonal. Dentre os genes produtores de transcritos quiméricos, em média 5% estavam DE, indicando que, no geral, os genes quiméricos não tiveram sua expressão modificada pelos tratamentos.

Figura 18. Expressão de genes que geraram transcritos quiméricos entre controle e tratados nas linhagens de Gotheon de *D. melanogaster*. No eixo y estão dispostos os valores de log10 normalizados da expressão dos genes no Controle, no eixo x os valores normalizados nos tratamentos: à esquerda *Cold-Shock* (A e C) e à direita em *Starvation* (B e D). Todos os genes produtores de transcritos quiméricos com suas expressões normalizadas estão inclusos no gráfico. Formas coloridas destacadas e identificadas representam genes quiméricos expressos diferencialmente entre controle e tratado entre todas as linhagens (p ajustado (FDR) $< 0,05$; $\log_2 \text{fold-change} > |1|$); formas transparentes representam genes quiméricos não expressos diferencialmente entre todas as linhagens (valor de p ajustado). A identificação “nenhum” corresponde a genes que estão DE e são quiméricos em outras linhagens e comparações, mas não na linhagem apresentada.

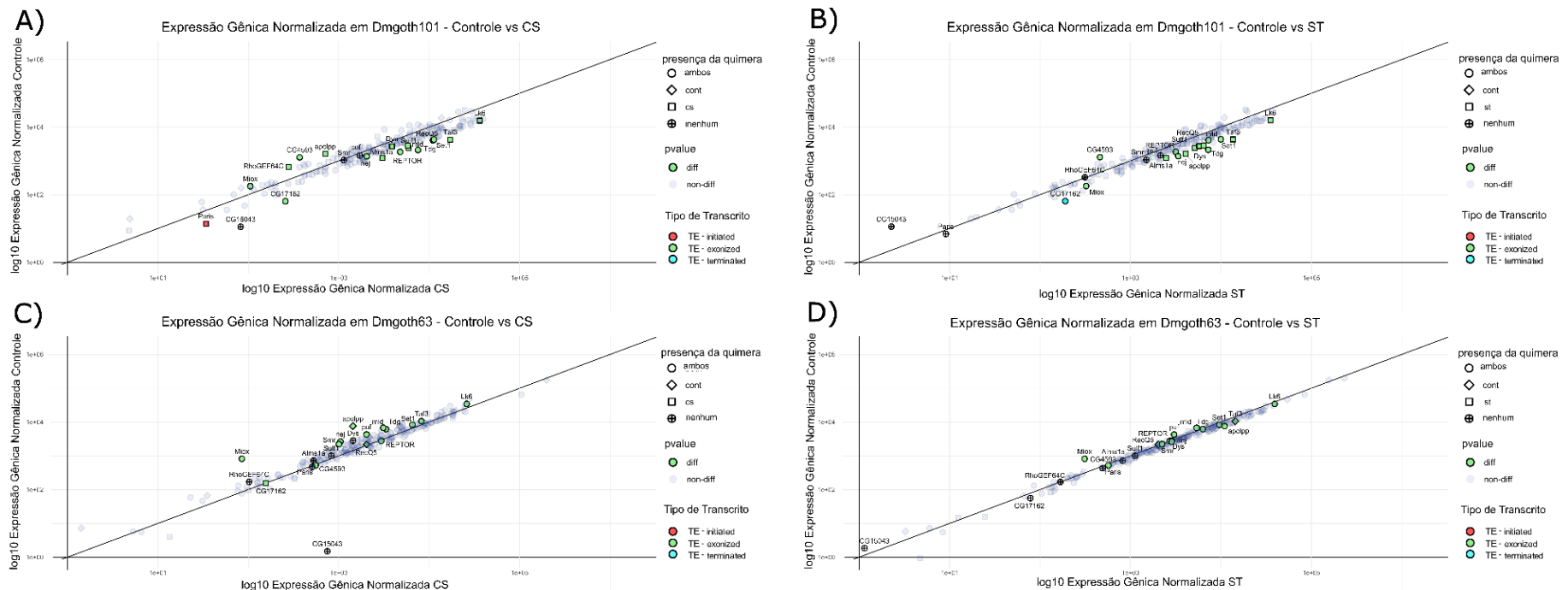
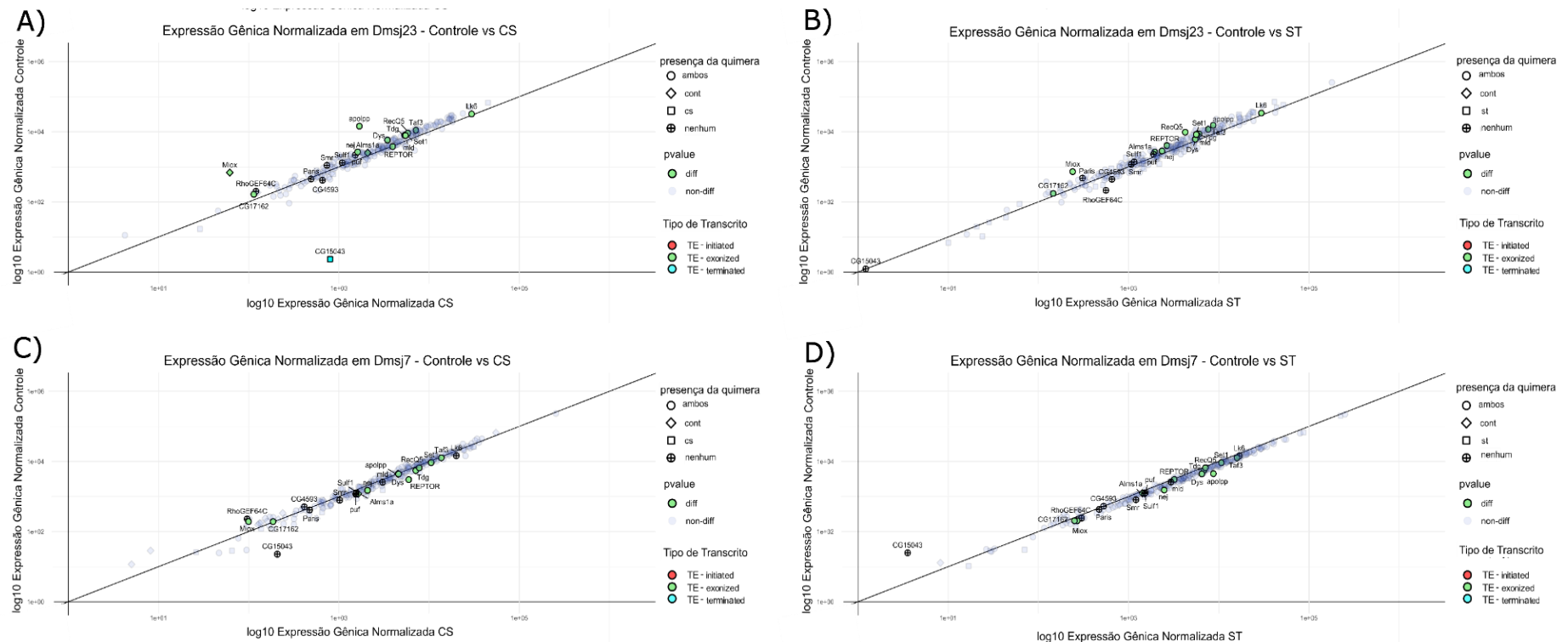
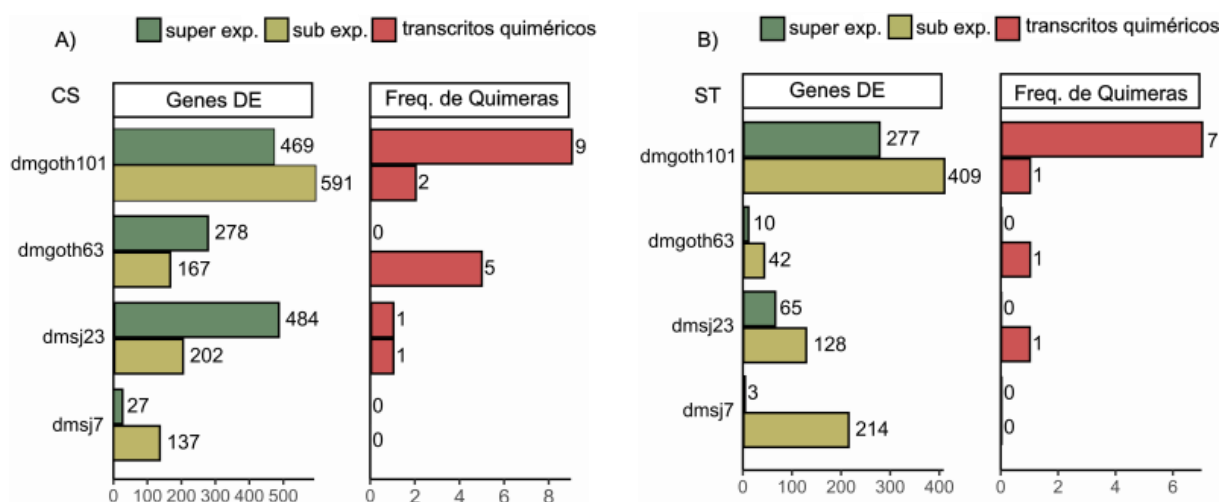


Figura 18, continuação. Expressão de genes que geraram transcritos quiméricos entre controle e tratados das linhagens de São José do Rio Preto de *D. melanogaster*. No eixo y estão dispostos os valores de log10 normalizados da expressão dos genes no Controle, no eixo x os valores normalizados nos tratamentos: à esquerda *Cold-Shock* (A e C) e à direita em *Starvation* (B e D). Todos os genes produtores de transcritos quiméricos com suas expressões normalizadas estão inclusos no gráfico. Formas coloridas destacadas e identificadas representam genes quiméricos expressos diferencialmente entre controle e tratado entre todas as linhagens (p ajustado (FDR) $< 0,05$; $\log_2 \text{fold-change} > |1|$); formas transparentes representam genes quiméricos não expressos diferencialmente entre todas as linhagens (valor de p ajustado). A identificação “nenhum” corresponde a genes que estão DE e são quiméricos em outras linhagens e comparações, mas não na linhagem apresentada.



A Figura 19 mostra, à esquerda, o número total de DEGs (sub e super expressos) entre o controle e as condições *Cold-Shock* (Figura 19a) e *Starvation* (Figura 19b) para as quatro linhagens de *D. melanogaster*, e à direita, o número de DEGs que produziram transcritos quiméricos. Destaca-se a linhagem Dmgoth101 com maior número de DEGs entre as duas condições.

Figura 19. Número de genes diferencialmente expressos (DE) que formaram quimeras nas linhagens de *D. melanogaster* entre Controle e tratado a) *Cold-shock* (CS) e b) *Starvation* (ST).



Na Tabela 2 são relacionados os genes diferencialmente expressos que produzem transcritos quiméricos em *D. melanogaster*. Do total de 28 DEGs quiméricos (Figura 18), excluindo as redundâncias, genes DE em mais de uma linhagem e tratamento correspondem a 20 genes únicos, dentre os quais, quatro correspondem a inserções de TE polimórficas entre as populações e dois são exclusivos de Dmgoth101. Como apresentado anteriormente, a grande maioria dos genes quiméricos estão DE em Dmgoth101, somente 10% (dois genes) estão DE exclusivamente em outras linhagens, e seis (30%) estão DE em mais de uma linhagem. A Tabela Suplementar 12 contém os dados de expressão diferencial (*Log2FoldChange*) dos referidos genes, acompanhado dos dados de expressão disponíveis no *FlyBase*.

A grande maioria dos transcritos quiméricos é constituída pelos TEs *roo* e *INE-1* (40% e 35%, respectivamente), as famílias mais frequentes no genoma e na constituição de quimeras, sendo o restante formado por famílias distintas das superfamílias *Gypsy*, *Jockey*, *Tc1-Mariner*, as quais são as mais frequentes no genoma e na produção de transcritos quiméricos. Todas as inserções dos genes presentes na Tabela 2 também foram registradas no

genoma *D. melanogaster* de referência (r6.49), indicando que essas inserções são também compartilhadas com a linhagem *iso-1* dos Estados Unidos, salvo alguns caracterizados a seguir.

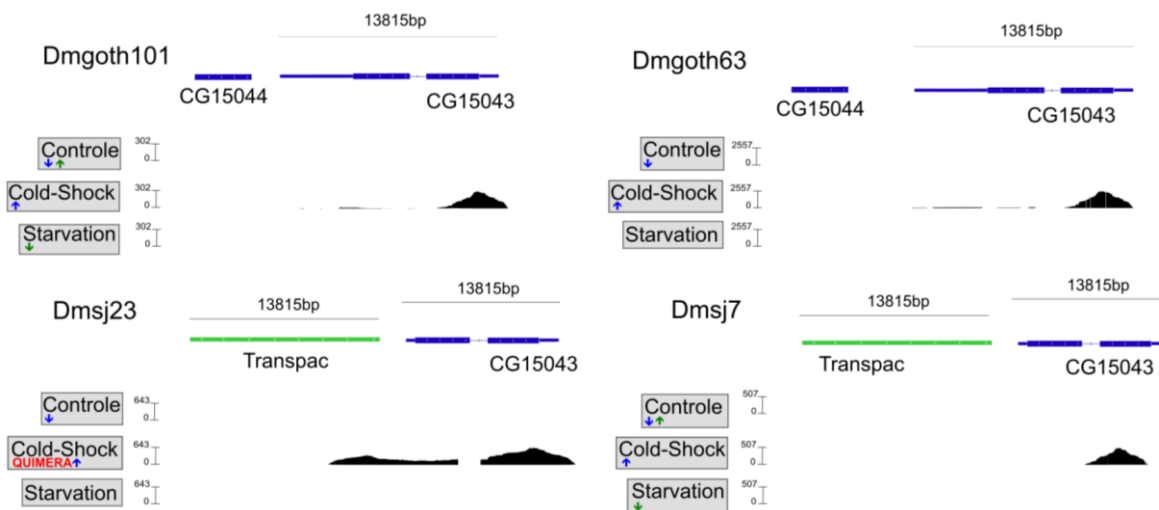
Tabela 2. Genes quiméricos diferencialmente expressos em CS e ST em *D. melanogaster*. O triângulo vermelho em uma linha representa a presença de um transcrito quimérico; o triângulo branco em uma linha representa a presença da inserção genômica do TE, mas sem um transcrito quimérico; a linha sem triângulos representa a ausência da inserção do TE.

Gene	Função	TE	Localização do TE	Linhagens			
				Dmgoth101	Dmgoth63	Dmsj23	Dmsj7
Alms1a	Divisão assimétrica de células-tronco germinativas masculinas; regulação positiva da replicação do centríolo; e regulação positiva da duplicação do centrossoma	INE-1	Exon	▲	△	▲	▲
Apolpp	Transporte lipídico envolvido no armazenamento lipídico; transporte de lipoproteínas; e regulação positiva da transdução de sinal	HB	Exon	▲	▲	▲	▲
CG15043	-	Transpac	Downstream	—	—	▲	△
CG17162	-	INE-1	Intron	▲	▲	▲	▲
CG4593	Regulação positiva da motilidade celular	INE-1	Exon	▲	▲	△	△
Dys	Morfogênese da asa derivada do disco imaginal; manutenção da estrutura da zona ativa pré-sináptica; e modulação da transmissão sináptica química	Roo	Exon	▲	—	▲	▲
Lk6	Regulação positiva da fosforilação da peptidilserina; fosforilação de proteínas; e regulação do crescimento	Roo	Exon/Intron	▲	▲	▲	—
Miox	Atividade da inositol oxigenase e processo catabólico	Rt1	Exon	▲	▲	▲	▲
Mld	Determinação da expectativa de vida adulta; processo biossintético de ecdisona; e memória de longo prazo	Roo	Exon	▲	▲	▲	—
Nej	Modificação de histonas; desenvolvimento do sistema nervoso; e regulação da via canônica de sinalização Wnt	Roo/TART-B/TART-C	Exon	▲	▲	▲	▲
Paris	Regulação negativa da transcrição, modelada por DNA	Burdock	Upstream	▲	△	—	—
Puf	Processo de modificação de proteínas celulares; regulação negativa da resposta imune; e regulação negativa da transdução de sinal	Roo	Exon	▲	—	—	—
RecQ5	Reparo de quebra de fita dupla via recozimento de fita simples; segregação mitótica de cromátides irmãs; e resposta ao raio X	INE-1	Exon	▲	▲	▲	▲
REPTOR	Sinalização TORC1; regulação positiva da transcrição, modelada por DNA; e resposta à fome	Roo	Exon	▲	▲	▲	▲
RhoGEF64C	Morfogênese da perna derivada do disco imaginal; regulação positiva da transdução de sinal da proteína Rho; e morfogênese do espiráculo, sistema traqueal aberto	Roo	Exon	▲	△	—	—
Smr	Regulação da transcrição pela RNA polimerase II; desenvolvimento de células foliculares ovarianas; e desenvolvimento do disco da asa	Roo	Exon	▲	—	—	—
Sulf1	Desenvolvimento de neurônios; regulação da transdução de sinal; e formação de padrão de disco de asa	Micropia	Intron	▲	△	—	—
Taf3	Regulação positiva da transcrição pela RNA polimerase II	INE-1	Intron	▲	▲	▲	▲
Tdg	Reparo de incompatibilidade	INE-1	Exon	▲	▲	▲	▲

Dentre os DEG quiméricos (Tabela 2, Figuras Suplementares 15-34), são destacados aqueles onde a diferença de expressão parece ter sido modulada pelo estresse e/ou pela inserção TE, são eles: *CG15043*, *Paris*, *RhoGEF64C* e *Sulf1*.

O gene *CG15043* apresenta *downstream* um grande fragmento do retrotransposon *Transpac* (3.000bp - *Gypsy/LTR*), exclusivo nas linhagens tropicais (Figura 20), e que produz transcritos quiméricos somente em Dmsj23, após a exposição ao estresse *Cold-Shock*. Essa inserção também está ausente no genoma *D. melanogaster* de referência (r6.49), indicando que a mobilização pode ter ocorrido após a divergência populacional. Apesar da produção de transcrito quimérico ser polimórfica, o gene está DE em todas as linhagens, inclusive nas linhagens europeias que não possuem inserção, indicando que *Transpac* não está modulando a expressão de *CG15043* em Dmsj23, e que o gene em questão possa ser responsivo ao frio. Os dados de expressão de RNA-seq publicados no *FlyBase* revelam a ausência de expressão de *CG15043* em ovários (RPKM = 0) e alta expressão em moscas adultas (4 dias – RPKM = 159) após o tratamento *Cold-Shock* (Tabela Suplementar 12). Essas observações, aliadas aos dados de expressão normalizados do presente trabalho (Figuras 18), sugerem uma possível responsividade de *CG15043* ao frio.

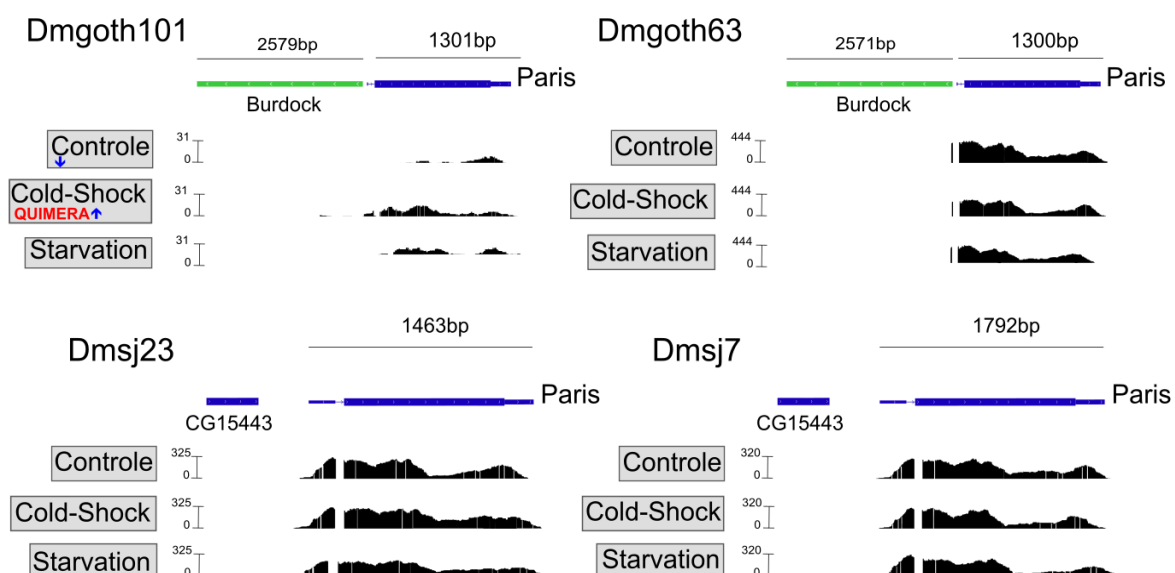
Figura 20. Ilustração da expressão do gene *CG15043* e a presença do TE entre Controle e os tratamentos no genoma das linhagens de *D. melanogaster* obtida por RNA-seq. A Figura representa no genoma os genes, em traços azuis e a presença dos TEs em verde (Retrotransposons com LTRs). Ambos estão identificados pela sua nomenclatura e pelo tamanho em bp de suas seqüências em cada linhagem. Acima dos genes e dos TEs, em preto, estão representadas suas respectivas expressões em cada situação, como também a cobertura de leituras encontrada em cada linhagem. As setas indicam os casos onde os genes estão super (↑) ou sub (↓) expressos, e suas cores indicam a comparação. Essa representação permite a identificação e visualização da influência do elemento na expressão do gene hospedeiro.



O gene *Paris* exibe uma inserção *upstream* do retrotransposon *Burdock* (2579 bp - *Gypsy/LTR*) exclusiva das linhagens europeias, e produz transcrito quimérico somente após exposição ao tratamento *Cold-Shock* em Dmgoth101, no qual foi detectado seu aumento de expressão (Figura 21). O fragmento de *Burdock* também está ausente no genoma *D. melanogaster* de referência (r6.49), revelando que o evento de mobilização pode ter ocorrido após divergência populacional. Somente no tratamento *Cold-Shock* de Dmgoth101 é possível identificar a presença de leituras quiméricas, alinhadas partes no gene e parte no TE, como também o aumento da expressão desse gene quando ocorre a formação de transcritos quiméricos, comparado com sua expressão no controle. Essas leituras sugerem que *Paris*, quando não em estresse, apresenta baixa expressão nessa linhagem e sua co-transcrição com *Burdock*, aparentemente ativada pelo estresse, modula a expressão de *Paris*. A expressão de *Paris* nas outras linhagens, com ou sem a inserção, é mais alta que em Dmgoth101 (Tabela Suplementar 11-a)), sem variação entre controle e tratado (onde de fato não está diferencialmente expresso – Figura 18,19, Tabela Suplementar 12). Os dados do Atlas de Expressão do *FlyBase* revelam uma expressão moderadamente baixa nos ovários de *D. melanogaster* (RPKM 10-22) e uma expressão muito baixa em moscas adultas que foram

expostas ao tratamento *Cold-Shock* (RPKM 3). Os dados de expressão publicados, associados aos nossos, indicam que os tratamentos, principalmente *Cold-Shock*, não alteram a expressão de *Paris*, e que somente em Dmgoth101 o TE parece atuar no aumento de sua expressão. A confirmação dessa modulação deve ser testada em experimentos adicionais.

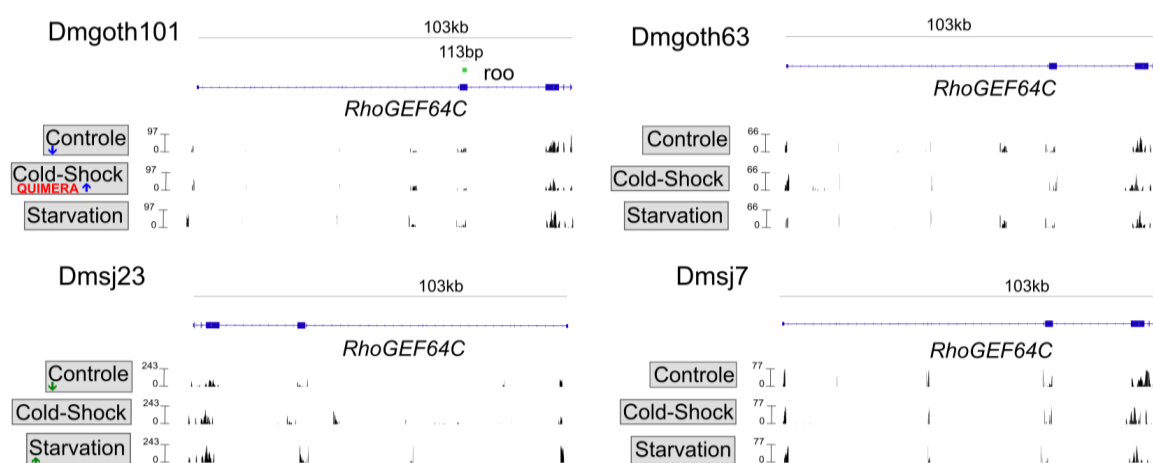
Figura 21. Ilustração da expressão do gene *Paris* e a presença do TE entre Controle e os tratamentos no genoma das linhagens de *D. melanogaster* obtida por RNA-seq. A Figura representa no genoma os genes, em traços azuis e a presença dos TEs em verde (Retrotransposons com LTRs). Ambos estão identificados pela sua nomenclatura e pelo tamanho em bp de suas sequências em cada linhagem. Acima dos genes e dos TEs, em preto, estão representadas suas respectivas expressões em cada situação, como também a cobertura de leituras encontrada em cada linhagem. As setas indicam os casos onde os genes estão super (↑) ou sub (↓) expressos, e suas cores indicam a comparação. Essa representação permite a identificação e visualização da influência do elemento na expressão do gene hospedeiro.



O gene *RhoGEF64C* tem um fragmento *roo* inserido dentro de um éxon (cds) (*embedded*), exclusivo da linhagem de Gotheron Dmgoth101 (Figura 22). A ausência dessa inserção no genoma das outras linhagens resultou da filtragem de tamanho adotada na metodologia, tendo em vista que a inserção é relatada no genoma público de *D. melanogaster* (r6.49). Esse gene produz transcritos quiméricos, sub expressos em Dmgoth101 após o tratamento *Cold-Shock* e, apesar da ausência da inserção, também está super expresso em *Starvation* de Dmsj23. Em Dmgoth101 *Cold-Shock*, na qual há a produção de transcritos quiméricos, a expressão é reduzida, comparada com as outras condições e linhagem onde não há leituras quiméricas. Tanto em ovários, quanto em moscas adultas que foram expostas a *Cold-Shock*, os dados de expressão do *FlyBase* exibem muito baixa e baixa expressão,

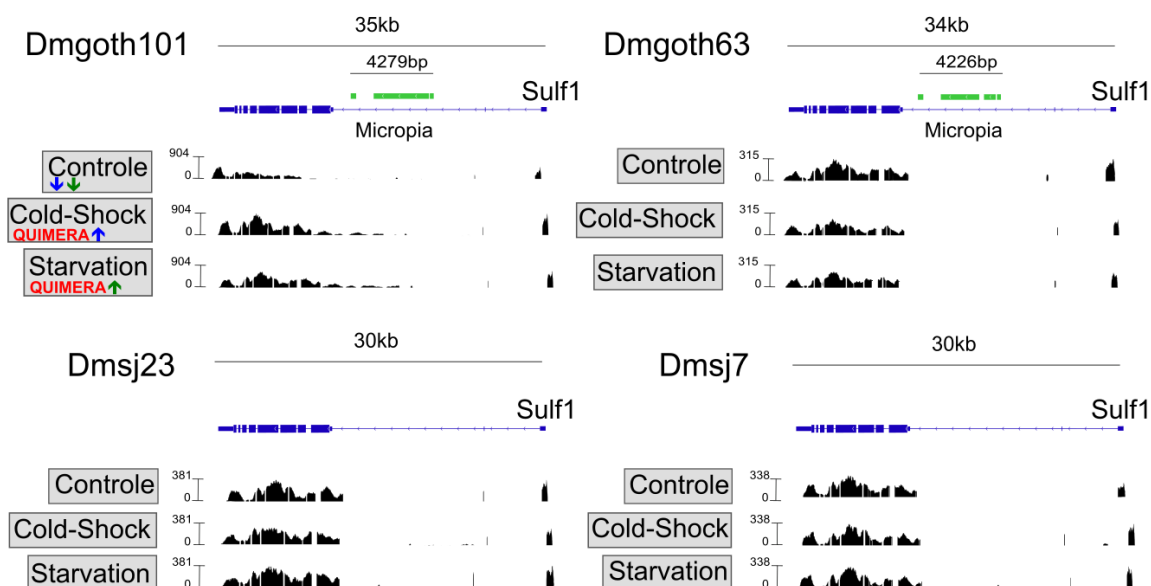
respectivamente, para esse gene (Tabela Suplementar 12). Conjuntamente, todos os dados de expressão relatados indicam que *RhoGEF64C* apresenta baixa expressão, independente da inserção TE, mas parece ter sua expressão ainda mais reduzida quando exposto ao frio (Tabela Suplementar 11-a)), e curiosamente aumentada após restrição alimentar em Dmsj23.

Figura 22. Ilustração da expressão do gene *RhoGEF64C* e a presença do TE entre Controle e os tratamentos no genoma das quatro linhagens de *D. melanogaster* obtida por RNA-seq. A Figura representa no genoma os genes, em traços azuis e a presença dos TEs em verde (Retrotransposons com LTRs). Ambos estão identificados pela sua nomenclatura e pelo tamanho em bp de suas sequências em cada linhagem. Acima dos genes e dos TEs, em preto, estão representadas suas respectivas expressões em cada situação, como também a cobertura de leituras encontrada em cada linhagem. As setas indicam os casos onde os genes estão super (↑) ou sub (↓) expressos, e suas cores indicam a comparação. Essa representação permite a identificação e visualização da influência do elemento na expressão do gene hospedeiro.



O gene *Sulf1* tem uma grande inserção da família *Micropia* (3.953 pb - *Gypsy/LTR*), localizada dentro de um íntron, exclusivo das linhagens europeias (Figura 23). A produção de transcritos quiméricos ocorre apenas entre o controle e as condições de estresse em Dmgoth101, onde está super expresso. Por também não ser encontrado no genoma *D. melanogaster* de referência (r6.49), essa inserção parece ser exclusiva das linhagens europeias, e sua mobilização ou perda pode ter ocorrido após a divergência das populações. As leituras quiméricas estão presentes nos tratamentos *Cold-Shock* e *Starvation* de Dmgoth101, onde observa-se o aumento de expressão, comparando-se com sua expressão no controle. Os dados de RNAseq disponíveis no *FlyBase* (Tabela Suplementar 12) indicam moderadamente alta e alta expressão de *Sulf1* nos ovários e moscas adultas de *D. melanogaster* expostas a *Cold-Shock*, respectivamente. Atrelados aos nossos dados de expressão, e comparando com as outras linhagens, onde não se observa produção de transcritos quiméricos nem diferença de expressão, sugerem que o TE *Micropia*, quando ativado pelo estresse, é recrutado na transcrição, contribuindo com o aumento da expressão de *Sulf1*.

Figura 23. Ilustração da expressão do gene *Sulf1* e a presença do TE entre Controle e os tratamentos no genoma das quatro linhagens de *D. melanogaster* obtida por RNA-seq. A Figura representa no genoma os genes, em traços azuis e a presença dos TEs em verde (Retrotransposons com LTRs). Ambos estão identificados pela sua nomenclatura e pelo tamanho em bp de suas seqüências em cada linhagem. Acima dos genes e dos TEs, em preto, estão representadas suas respectivas expressões em cada situação, como também a cobertura de leituras encontrada em cada linhagem. As setas indicam os casos onde os genes estão super (↑) ou sub (↓) expressos, e suas cores indicam a comparação. Essa representação permite a identificação e visualização da influência do elemento na expressão do gene hospedeiro.



5.2 Expressão diferencial entre controle e tratamentos e inserções de TEs potencialmente adaptativas em *D. simulans*

A expressão normalizada de todos os genes quiméricos em *D. simulans* é apresentada na Figura 24 com destaque para os genes quiméricos DE entre controle e tratado e que cujas inserções de TEs são candidatas a adaptativas. Dentre os genes produtores de transcritos quiméricos, em média 6% estavam DE, indicando, assim como observado em *D. melanogaster*, que geralmente é baixa a modulação dos genes quiméricos pelos tratamentos.

Figura 24. Expressão de genes que geraram transcritos quiméricos entre controle e tratados das linhagens de Gothen de *D. simulans*. No eixo y estão dispostos os valores de log10 normalizados da expressão dos genes no Controle, no eixo x os valores normalizados nos tratamentos: à esquerda *Cold-Shock* (A e C) e à direita em *Starvation* (B e D). Todos os genes produtores de transcritos quiméricos com suas expressões normalizadas estão inclusos no gráfico. Formas coloridas destacadas e identificadas representam genes quiméricos expressos diferencialmente entre controle e tratado entre todas as linhagens (p ajustado (FDR) $< 0,05$; $\log_2 \text{fold-change} > |1|$); formas transparentes representam genes quiméricos não expressos diferencialmente entre todas as linhagens (valor de p ajustado). A identificação “nenhum” corresponde a genes que estão DE e são quiméricos em outras linhagens e comparações, mas não na linhagem apresentada.

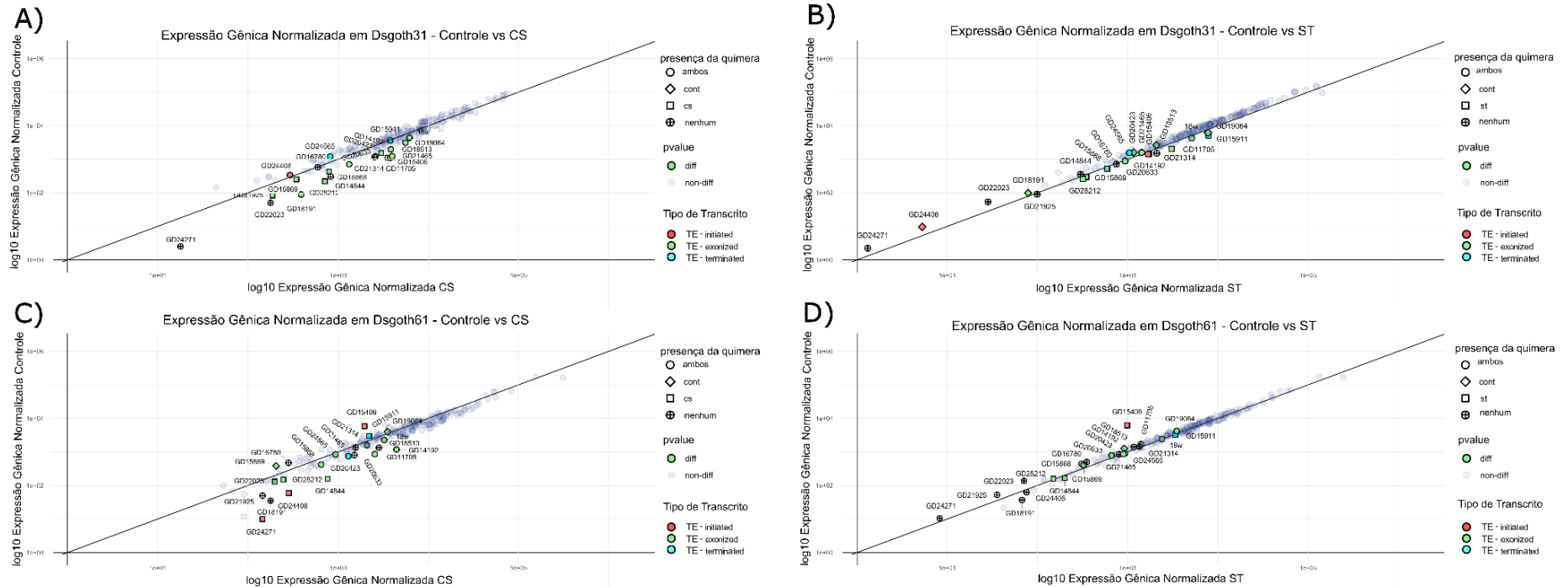
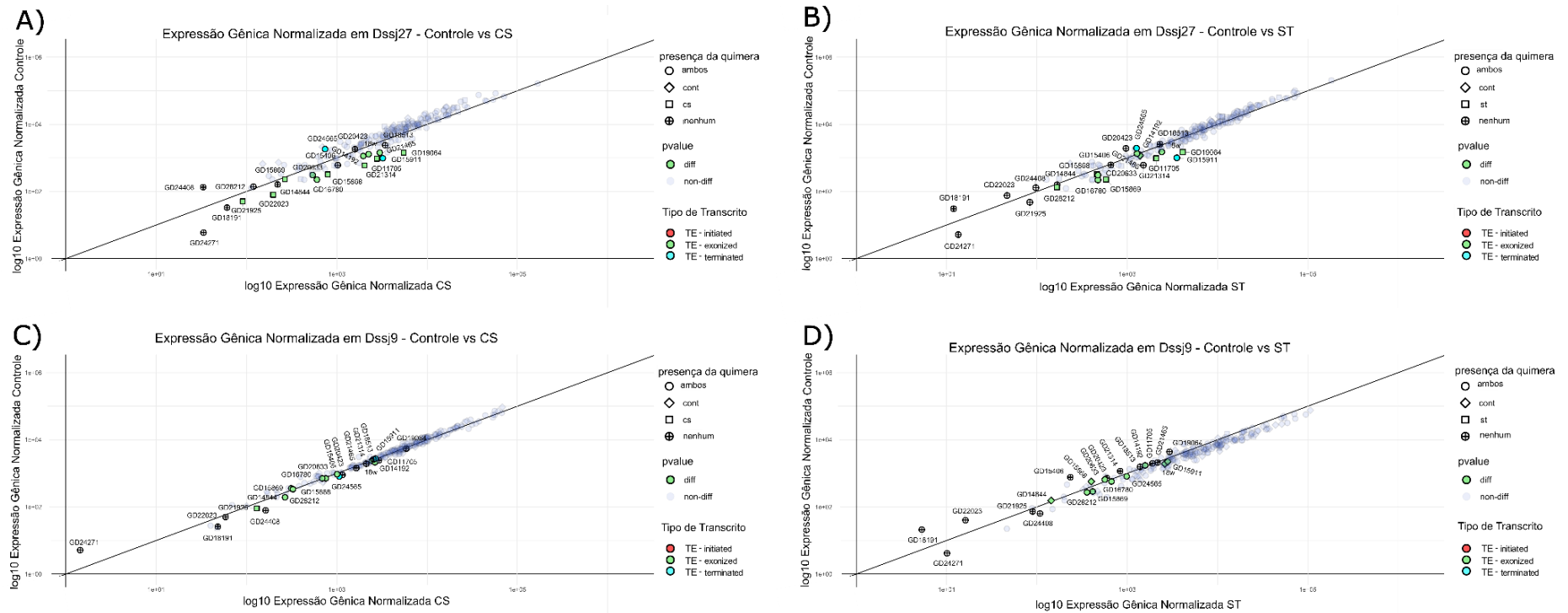
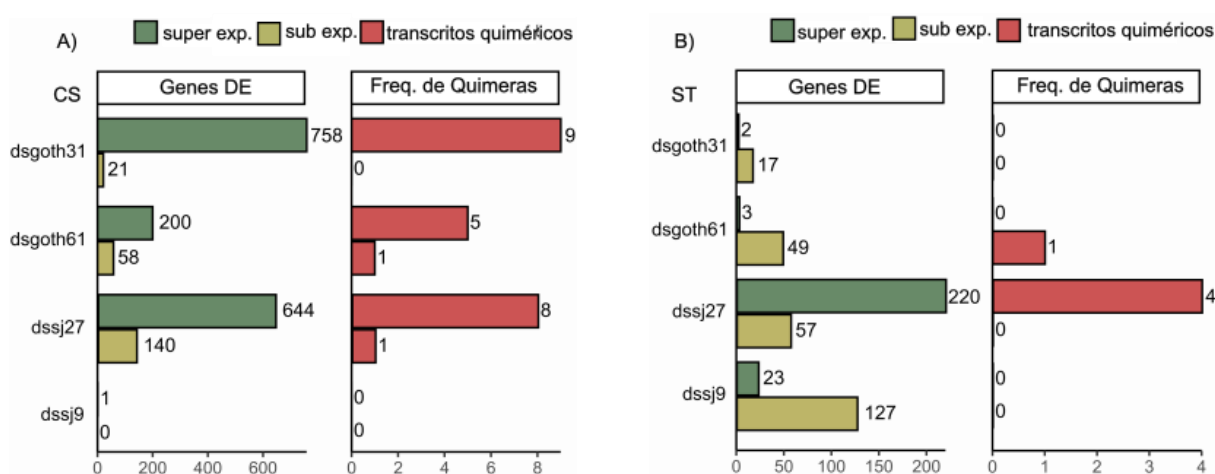


Figura 24, Continuação. Expressão de genes que geraram transcritos quiméricos entre controle e tratados das linhagens de São José do Rio Preto *D. simulans*. No eixo y estão dispostos os valores de log10 normalizados da expressão dos genes no Controle, no eixo x os valores normalizados nos tratamentos à esquerda *Cold-Shock* (A e C) e à direita em *Starvation* (B e D). Todos os genes produtores de transcritos quiméricos com suas expressões normalizadas estão inclusos no gráfico. Formas coloridas destacadas e identificadas representam genes quiméricos expressos diferencialmente entre controle e tratado entre todas as linhagens (valor de p ajustado (FDR) < 0,05; log2 fold-change > |1|); formas transparentes representam genes quiméricos não expressos diferencialmente entre todas as linhagens (valor de p ajustado). A identificação “nenhum” corresponde a genes que estão DE e são quiméricos em outras linhagens e comparações, mas não na linhagem apresentada.



Os números totais de DEGs sub e super expressos (Figura 25) entre controle e condições *Cold-Shock* (Figura 25a) e *Starvation* (Figura 25b) são apresentados à esquerda para as quatro linhagens de *D. simulans*, e à direita, o número de DEGs que produziram transcritos quiméricos. Como em *D. melanogaster*, as duas linhagens com os maiores números de DEGs (Dsgoth31 e Dssj27) também exibiram o maior número de genes produzindo transcritos quiméricos. As linhagens tropicais apresentaram mais DEGs entre Controle e *Starvation* do que as europeias, mas apenas Dssj27 apresentou genes quiméricos DEs.

Figura 25. Distribuição do número de DEGs que formaram quimeras em *D. simulans* entre Controle e tratados a) *Cold-Shock* (CS) e b) *Starvation* (ST).



Os genes quiméricos DE identificados em *D. simulans*, seus ortólogos em *D. melanogaster* e suas relações entre as linhagens são apresentados na Tabela 3. O total de 29 genes quiméricos DE contabilizados na Figura 26, correspondem a 22 genes exclusivos, dentre os quais dois correspondem a inserções polimórficas de TE entre as populações e dois são exclusivos de Dsgoth31. A maioria dos genes quiméricos DE em CS estão em Dsgoth31 e Dssj27, sendo quatro e sete genes, respectivamente, exclusivamente DE nessas linhagens. Somente um gene está DE exclusivamente em Dsgoth61, e ~45% (10 genes) são DE em mais de uma linhagem (Tabela 3); metade desses transcritos são formados pela superfamília *Bel-Pao/LTR* (*roo* e *rooA*), enquanto que ~23% são constituídos por *INE-1* (*Helitron/DNA*). Nota-se que *roo* produziu mais transcritos quiméricos DE, com um aumento de 10% em relação a *D. melanogaster*, enquanto a participação de *INE-1* atingiu apenas 9%. O restante desses transcritos apresenta inserções de *Jockey*, *Gypsy*, *P* e *hAT*. Os dados de expressão diferencial (*Log2FoldChange*) dos referidos genes encontram-se na Tabela Suplementar 13.

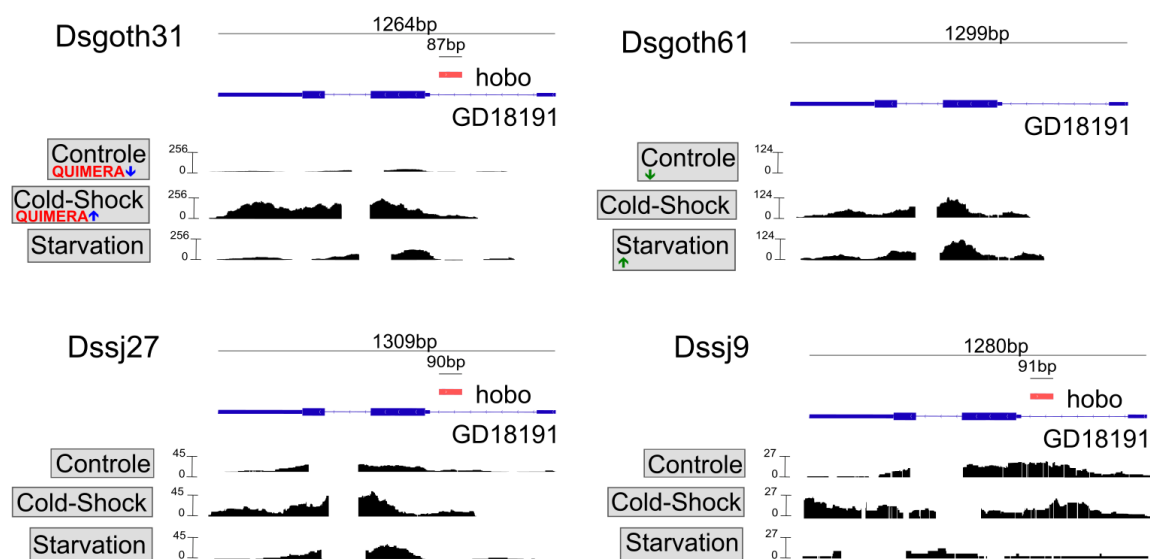
Tabela 3. Genes quiméricos diferencialmente expressos em CS e ST em *D. simulans*. O triângulo vermelho em uma linha representa a presença de um transcrito quimérico; o triângulo branco em uma linha representa a presença da inserção genômica do TE, mas sem um transcrito quimérico; a linha sem triângulos representa a ausência da inserção do TE.

Gene	Ortólogos em <i>D.mel</i>	Função	TE	Localização do TE	Linhagens			
					Dsgoth31	Dsgoth61	Dssj27	Dssj9
18w	18w	Resposta humoral antibacteriana; extensão convergente; e oogênese	Roo	Exon	▲	▲	▲	▲
GD11705	stl	Regulação da migração de neurônios e formação de pedúnculo de células foliculares ovarianas	INE-1	Exon	▲	▲	▲	△
GD14192	CG42673	Organização pós-sináptica do citoesqueleto de actina	Roo	Exon	▲	▲	▲	▲
GD14844	CG17732	-	Roo	Exon	—	▲	—	▲
GD15406	Cyp6g1	Processo catabólico inseticida; resposta ao DDT; e resposta ao íon mercúrio	Juan	Upstream	▲	▲	—	△
GD15868	mamo	Divisão nuclear meiótica feminina; regulação da organização da cromatina; e descondensação da cromatina do esperma	Roo	Exon	▲	▲	▲	▲
GD15869	Pde9	Processo catabólico cGMP e sinalização mediada por cGMP	Roo	Exon	△	▲	▲	▲
GD15911	CG12716	Formação de casca de ovo contendo córion	INE-1	Downstream	▲	▲	▲	▲
GD16780	CG5921	Transição G2/M do ciclo celular mitótico	Roo	Exon	—	—	▲	▲
GD18191	CG33494	-	Hobo	Exon/Intron	▲	—	△	△
GD18513	puc	Montagem de córion de casca de ovo; morfogênese de um epitélio; e regulação negativa da transdução de sinal	P-element	Exon	▲	—	—	—
GD19064	srp	Desenvolvimento de órgãos animais; desenvolvimento embrionário através do blastoderme sincicial; e desenvolvimento de hemócitos	Roo	Exon	▲	▲	▲	—
GD20423	cv-c	Organização do citoesqueleto; morfogênese embrionária; e desenvolvimento do sistema traqueal aberto	Roo	Exon	▲	—	—	—
GD20633	ClC-a	Transporte de cloreto e transporte de água	1360/Roo	Exon	▲	▲	▲	▲
GD21314	CG46339	Desdiferenciação celular; homeostase de nicho de células-tronco de linhagem germinativa; e manutenção da população de células-tronco da linhagem germinativa masculina	Transpac	Intron	—	—	▲	△
GD21465	kay	Desenvolvimento do disco imaginal; ritmo locomotor; e desenvolvimento do sistema nervoso	Roo	Exon	▲	▲	▲	—
GD22023	NimC1	Resposta de defesa à bactéria e fagocitose	Circe	Exon/Intron	△	▲	▲	△
GD24271	CG18810	S-palmitoilação de peptidil-L-cisteína e direcionamento de proteína para a membrana	G5A	Upstream	△	▲	△	△
GD24408	CG1909	Transmissão sináptica colinérgica	INE-1	Upstream	▲	▲	△	△
GD24565	CG12659	Remodelação da cromatina	RooA	Downstream	▲	▲	▲	▲
GD28212	Itgbn	Encapsulamento melanótico de alvo estranho; homeostase celular das células musculares; e regulação negativa da montagem sináptica na junção neuromuscular	INE-1	Exon	▲	▲	▲	▲

Dentre os genes quiméricos DE (Figuras Suplementares 35-57) apresentados na Tabela 3, destacam-se aqueles em que a diferença de expressão parece estar relacionada com o tratamento e com a inserção, são eles *GD18191* e *GD24271*.

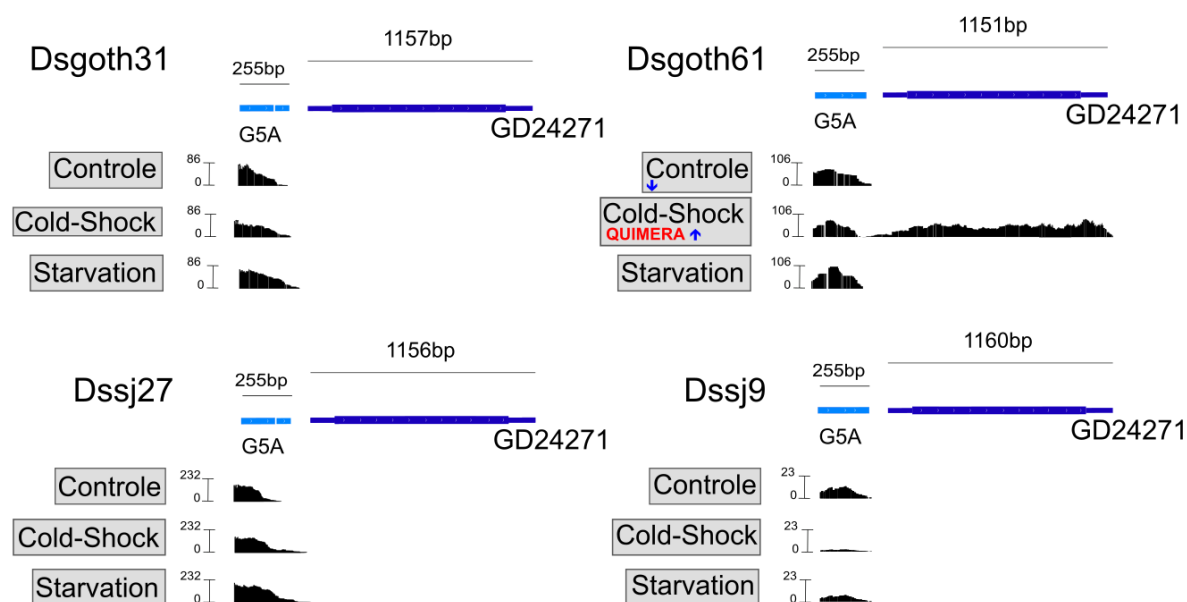
O gene *GD18191* têm um elemento *hobo* (87bp - *DNA/hAT*) inserido parcialmente em um éxon e parcialmente em um íntron, inserção essa encontrada no genoma publicado de *D. simulans* (r2.02) e anotada em todas as linhagens, exceto Dsgoth61, onde a inserção não passou pela etapa de filtragem de tamanho da inserção. Essa quimera produziu transcritos quiméricos somente no controle e *Cold-Shock* de Dsgoth31, estando *upregulated* nesse estresse. Observando sua expressão na Figura 26, alinhada aos dados de expressão normalizados, *GD18191* parece ser mais expresso nas linhagens europeias e após a exposição ao *Cold-Shock*, e tem sua expressão diminuída quando em *Starvation*; contudo, essa última observação não foi corroborada pelas análises de expressão diferencial realizadas (Figura 24, Tabela Suplementar 13). Portanto, os dados sugerem que *GD18191* é um gene responsivo ao *Cold-Shock*, sendo mais expresso nas linhagens adaptadas ao clima frio (Dsgoth31 e Dsgoth61), e a inserção do TE não parece ter qualquer efeito sobre o aumento da expressão.

Figura 26. Ilustração da expressão do gene *GD18191* e a presença do TE entre Controle e tratamentos no genoma das quatro linhagens de *D. simulans* obtida por RNA-seq. A Figura representa no genoma os genes, em traços azuis e a presença dos TEs em coral (Transposon de DNA). Ambos estão identificados pela sua nomenclatura e pelo tamanho em bp de suas sequências em cada linhagem. Acima dos genes e dos TEs, em preto, estão representadas suas respectivas expressões em cada situação, como também a cobertura de leituras encontrada em cada linhagem. As setas indicam os casos onde os genes estão super (↑) ou sub (↓) expressos, e suas cores indicam a comparação. Essa representação permite a identificação e visualização da influência do elemento na expressão do gene hospedeiro.



O último gene a ser destacado, *GD24271*, possui o retrotransposon sem LTR *G5A* (255bp - *Jockey/LINE*) inserido *upstream* a ele, em todas as linhagens, assim como no genoma público de *D. simulans* (r2.02); contudo, só produz transcritos quiméricos, e está super expresso, em Dsgoth61, após a exposição ao choque frio. A expressão de *GD24271* parece ter aumentado em todas as linhagens após a exposição ao *Cold-Shock*, sugerindo que esse gene também pode ser responsivo ao frio (Figura 27). Como relatado em Dsgoth61, a expressão em *Cold-Shock* aumenta significativamente (Tabela Suplementar 13) quando ocorre a produção do transcrito quimérico; portanto, nossos dados indicam que essa inserção *G5A* pode também estar modulando a expressão desse gene que parece ser responsivo ao frio.

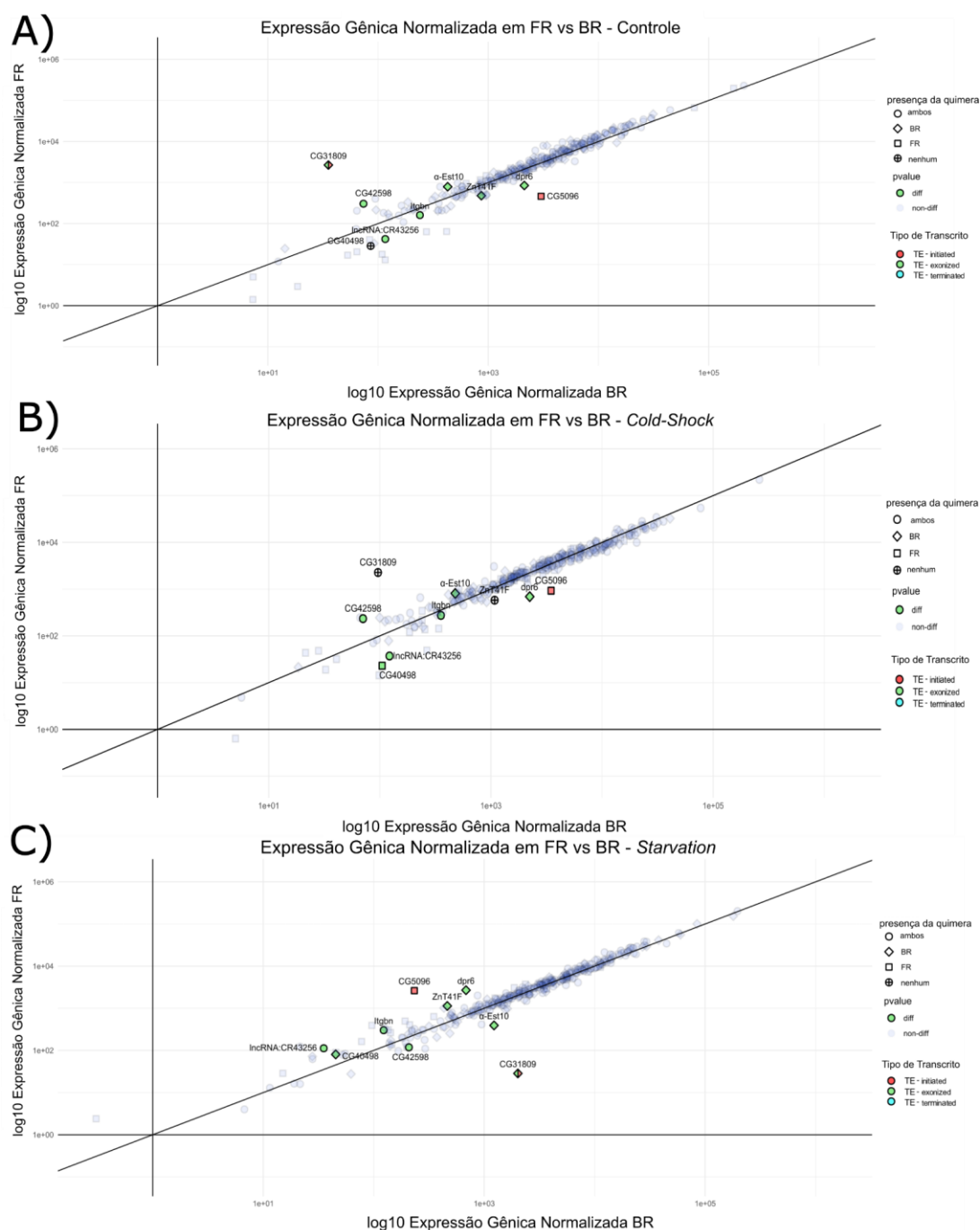
Figura 27. Ilustração da expressão do gene *GD24271* e a presença do TE entre Controle e tratamentos no genoma das quatro linhagens de *D. simulans*, obtida por RNA-seq. A Figura representa no genoma os genes, em traços azuis e a presença dos TEs em azul-claro (LINE). Ambos estão identificados pela sua nomenclatura e pelo tamanho em bp de suas sequências em cada linhagem. Acima dos genes e dos TEs, em preto, estão representadas suas respectivas expressões em cada situação, como também a cobertura de leituras encontrada em cada linhagem. As setas indicam os casos onde os genes estão super (↑) ou sub (↓) expressos, e suas cores indicam a comparação. Essa representação permite a identificação e visualização da influência do elemento na expressão do gene hospedeiro.



5.3 Expressão diferencial interpopulacional e inserções de TEs potencialmente adaptativas em *D. melanogaster*

A domesticação de inserções de TEs adaptativas associadas à colonização de novos habitats e zonas climáticas em cada população foi investigada por meio de análises de expressão diferencial entre as linhagens brasileiras e francesas. Essa abordagem possibilitou também a identificação do efeito do estresse na produção de transcritos quiméricos interpopulacionais. A expressão normalizada dos genes que produziram transcritos quiméricos permite a identificação dos genes super ou subexpressos entre as populações (linhagens francesas Dmgoth101 e Dmgoth63 e linhagens brasileiras Dmsj23 e Dmsj7) em cada uma das condições (Controle, *Cold-Shock* e *Starvation*) (Figura 28). Com relação à comparação interpopulacional, do total de genes produtores de transcritos quiméricos em *D. melanogaster*, aproximadamente 2,24% estavam DE, evidenciando mais uma vez que a mudança de expressão relacionada à produção de transcritos quiméricos é baixa.

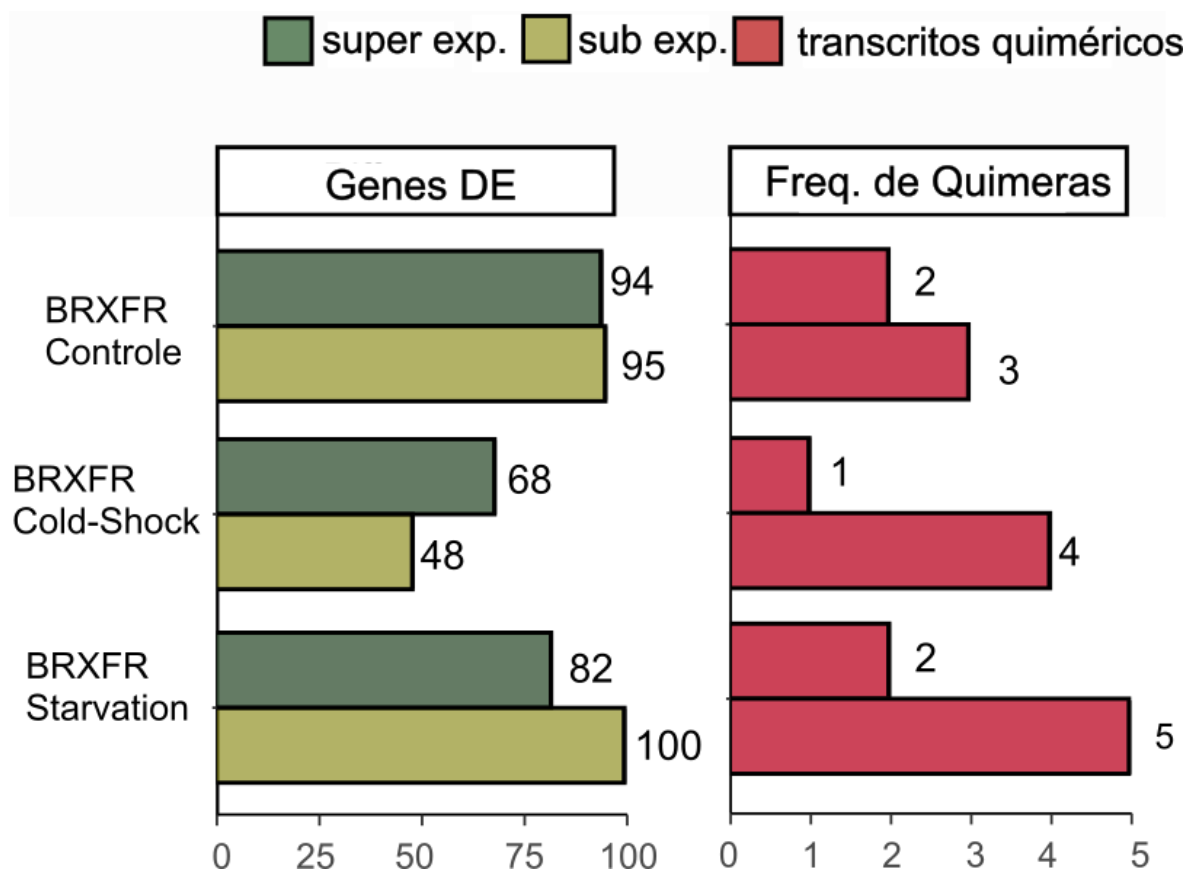
Figura 28. Expressão de genes que geram transcritos quiméricos entre as linhagens brasileiras e francesas, nas três condições (A) Controle, B) *Cold-Shock*, C) *Starvation*), em *D. melanogaster*. No eixo y estão dispostos os valores de log10 normalizados da expressão dos genes no Controle, no eixo x os valores normalizados nos tratamentos: à direita em *Starvation* e à esquerda *Cold-Shock*. Todos os genes produtores de transcritos quiméricos com suas expressões normalizadas estão inclusos no gráfico. Formas coloridas destacadas e identificadas representam genes quiméricos expressos diferencialmente entre controle e tratado entre todas as linhagens (valor de p ajustado (FDR) < 0,05; $\log_2 \text{fold-change} > |1|$); formas transparentes representam genes quiméricos não expressos diferencialmente entre todas as linhagens (valor de p ajustado). A identificação “nenhum” corresponde a genes que estão DE e são quiméricos em outras linhagens e comparações, mas não na linhagem apresentada.



O número total de DEGs (sub e super expressos) entre as linhagens das populações brasileiras e francesas de *D. melanogaster* nas condições de Controle e os tratamentos *Cold-Shock* e *Starvation* é exibido à esquerda da Figura 29, assim como, à direita, está evidenciado quais desses genes também produzem transcritos quiméricos. Os genes apresentados abaixo estão sub ou super expressos nas linhagens brasileiras em relação às linhagens francesas (Tabela Suplementar 15).

Foram encontrados menos genes DEs entre os tratamentos *Cold-Shock* das populações, já o número total de DEGs que também produzem transcritos quiméricos nesse tratamento e no controle são próximos; contudo, há mais genes quiméricos DEs em *Starvation*. Por outro lado, há mais genes sub expressos nas linhagens brasileiras em todas as condições, sugerindo que, independentemente da condição, genes que produzem transcritos quiméricos tendem a ser menos expressos nas linhagens adaptadas ao clima tropical do que as linhagens adaptadas ao clima temperado.

Figura 29. Distribuição do número de DEGs que formaram quimeras em *D. melanogaster* entre linhagens brasileiras e francesas no Controle e tratados *Cold-Shock* e *Starvation*



Os genes quiméricos DEs computados entre as populações analisadas de *D. melanogaster*, e suas relações entre as linhagens de funções são apresentados na Tabela 4. Ao todo, nove genes são destacados, dos quais três correspondem a inserções exclusivas de linhagens (*CG5096*, *dpr6* e *α -Est10*), e dois que produzem transcritos quiméricos exclusivos de linhagens (*CG31809* e *ZnT41F*). Não foram encontrados casos de inserção ou produção de transcritos quiméricos polimórficos entre as populações brasileiras e francesas. As análises dos casos descritos na Tabela 4 sugerem que, os dados obtidos de expressão diferencial entre as populações (p ajustado (FDR) $< 0,05$; $\log_2 \text{fold-change} > |1|$) podem ter sido enviesados durante a normalização das *counts* pela expressão (alta ou baixa) de uma única linhagem, dentre as representantes de determinada população. Das 10 inserções de TEs descritas na Tabela 4, quatro delas pertencem à família *INE-1*; com relação ao sítio de inserção do TE, seis estão localizados em regiões intrônicas, influenciando a produção de isoformas dos transcritos.

Os genes já destacados por apresentarem inserções polimórficas exclusivas de linhagens e a inserção de TE assinalada a *CG42598* estão ausentes no genoma *D. melanogaster* de referência (r6.49). As inserções de TEs dos genes remanescentes são encontradas no genoma de referência, linhagem essa *iso-1* proveniente dos Estados Unidos.

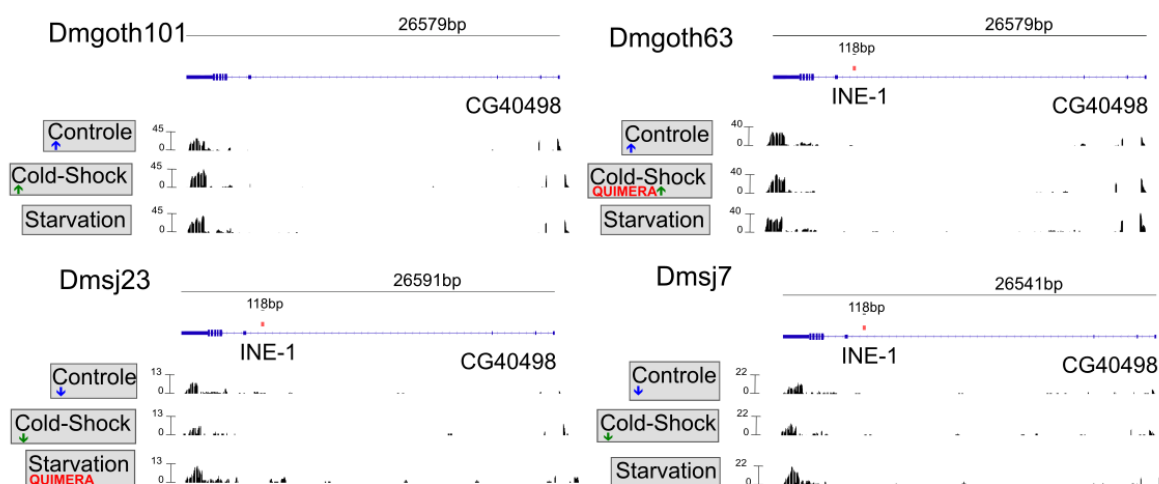
Tabela 4. Genes quiméricos diferencialmente expressos entre linhagens brasileiras e francesas, nas três condições, em *D. melanogaster*. O triângulo vermelho em uma linha representa a presença de um transcrito quimérico; o triângulo branco em uma linha representa a presença da inserção genômica do TE, mas sem um transcrito quimérico; a linha sem triângulos representa a ausência da inserção.

Gene	Função	TE	Localização do TE	Linhagens			
				Dmgoth101	Dmgoth63	Dmsj23	Dmsj7
CG5096	Adesão celular; desenvolvimento; transmissão neural	412	Upstream	—	▲	—	—
CG31809	Habilitar atividade da oxido redutase	Doc3-element/Hobo	Intron/Upstream	▲	▲	▲	▲
CG40498	-	INE-1	Exon/Intron	—	▲	▲	▲
CG42598	-	Gypsy4	Intron	▲	▲	▲	▲
dpr6	Envolvido na organização de sinapses	Roo	Exon	—	—	—	▲
Itgbn	Permite atividade de heterodimerização de proteínas	INE-1	Exon	▲	▲	▲	▲
lncRNA:CR43256	-	INE-1	Exon/Intron	▲	▲	▲	▲
ZnT41F	Ativar a atividade do transportador transmembrana de ions de zinco	INE-1	Intron	▲	▲	▲	▲
α -Est10	Permite a atividade da serina hidrolase	Invader6	Intron	—	—	▲	—

Dentre os genes da Tabela 4, três casos são de particular interesse. O primeiro deles é o gene *CG40498*, que possui um TE da família *INE-1* (116 bp - *DNA/Helitron*) inserido parcialmente em um éxon e parcialmente em um íntron. Esse TE está ausente apenas em Dmgoth101, e a produção de transcrito quimérico ocorre em Dmgoth63 e Dmsj7 em tratamentos diferentes (Figura 30). Esse gene está sub expresso ($\log_2\text{FoldChange}$

significativo <1) nas linhagens brasileiras na condição Controle e após o tratamento *Cold-Shock*. A presença do TE e a produção do transcrito quimérico não parecem ter relação com essa diferença de expressão, tendo em vista que na condição Controle não há produção de quimeras, mas há diferença de expressão significativa. Contudo, a alteração na expressão de *CG40498* parece configurar um caso de adaptação às zonas climáticas distintas, especificamente à zona temperada, por estar menos expresso em ambas linhagens brasileiras, independente da condição (Figura Suplementar 60, Tabelas Suplementares 14 e 15).

Figura 30. Ilustração da expressão do gene *CG40498* entre Controle e Tratamentos e a presença do TE no genoma das quatro linhagens de *D. melanogaster* obtida por RNA-seq. A Figura representa no genoma os genes, em traços azuis e a presença dos TEs em coral (Transposon de DNA). Ambos estão identificados pela sua nomenclatura e pelo tamanho em bp de suas seqüências em cada linhagem. Acima dos genes e dos TEs, em preto, estão representadas suas respectivas expressões em cada situação, como também a cobertura de leituras encontrada em cada linhagem. As setas indicam os casos onde os genes estão super ($\uparrow$) ou sub ($\downarrow$) expressos, e suas cores indicam a comparação. Essa representação permite a identificação e visualização da influência do elemento na expressão do gene hospedeiro.

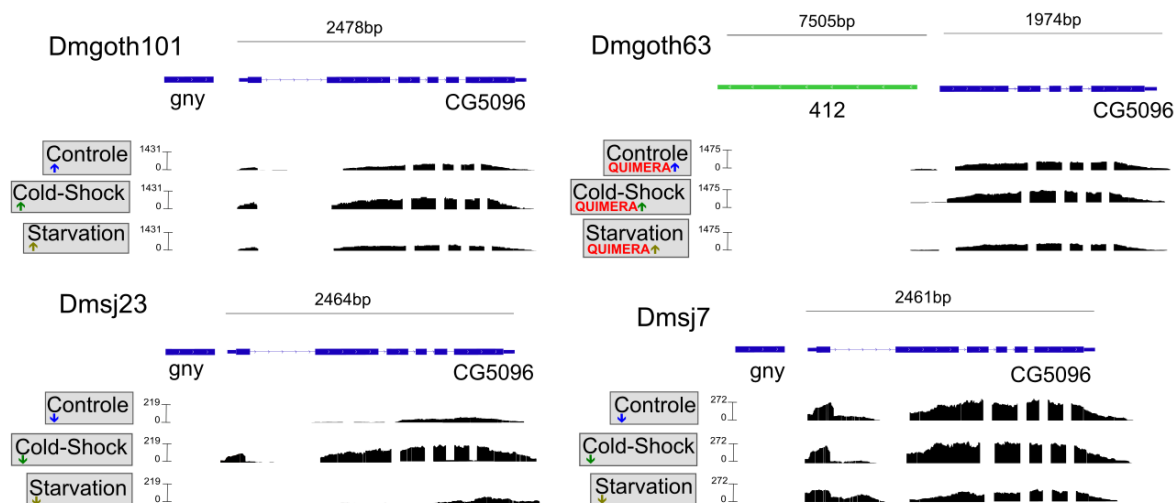


Apenas os genes *CG5096* e *CG42598*, dentre os nove genes quiméricos DE (Figuras Suplementares 58-66), destacam-se por terem sua diferença de expressão possivelmente afetada pela inserção do TE.

O gene *CG5096* apresenta uma inserção polimórfica *upstream* do TE 412 (7505 bp - *LTR/Gypsy*) encontrada apenas na linhagem francesa Dmgoth63, na qual ocorre a produção de transcritos quiméricos nas três condições (Figura 31). Apesar de Dmgoth101 não apresentar a inserção, a expressão do gene é maior nas duas linhagens francesas, do que nas linhagens brasileiras (Tabelas Suplementares 14 e 15). Os dados de expressão de *RNA-seq*

publicados no *FlyBase* indicam que a expressão de *CG5096* varia entre baixa e muito baixa nos ovários de moscas adultas (RPKM = 2), e baixa após o tratamento *Cold-Shock* (RPKM = 9 - Tabela Suplementar 15). Desconsiderando-se o possível papel do TE na modulação da expressão desse gene, aliado aos dados públicos de expressão, o aumento de expressão de *CG5096* parece ocorrer somente nas linhagens francesas, podendo a super expressão desse gene ser uma adaptação ao clima temperado. Quando ocorre a produção de transcrito quimérico em *Dmgoth63*, a expressão parece aumentar mais, quando comparado a outra linhagem francesa (Tabela Suplementar 14). Portanto, nossos dados sugerem possível ativação e recrutamento do TE 412 na expressão, o qual atua aumentando a expressão de *CG5096*, assim como um papel adaptativo desse gene na colonização e sobrevivência das linhagens francesas.

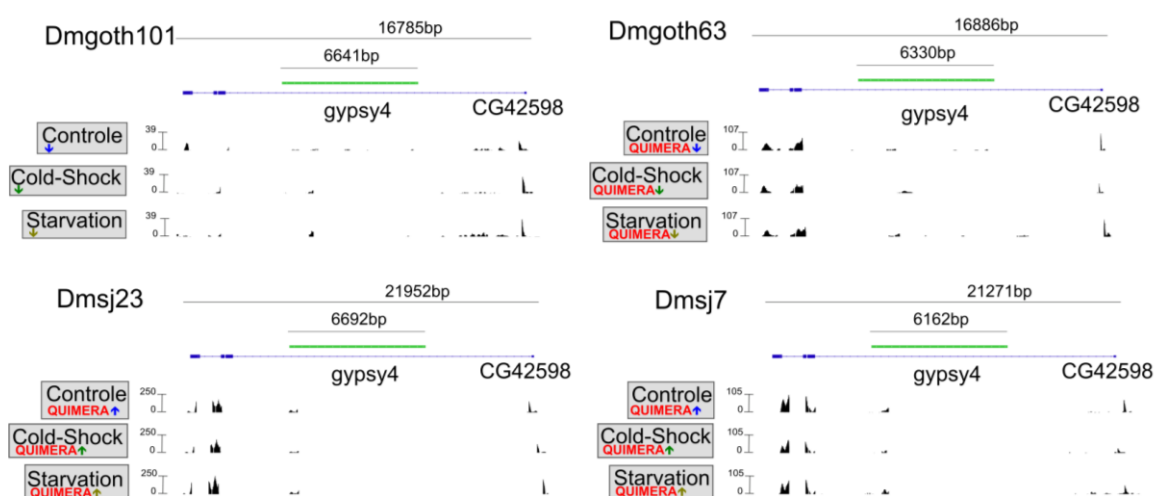
Figura 31. Ilustração da expressão do gene *CG5096* entre Controle e Tratamentos e a presença do TE no genoma das quatro linhagens de *D. melanogaster* obtida por RNA-seq. A Figura representa no genoma os genes, em traços azuis e a presença dos TEs em verde (Retrotransposon com LTRs). Ambos estão identificados pela sua nomenclatura e pelo tamanho em bp de suas seqüências em cada linhagem. Acima dos genes e dos TEs, em preto, estão representadas suas respectivas expressões em cada situação, como também a cobertura de leituras encontrada em cada linhagem. As setas indicam os casos onde os genes estão super (↑) ou sub (↓) expressos, e suas cores indicam a comparação. Essa representação permite a identificação e visualização da influência do elemento na expressão do gene hospedeiro.



O gene *CG42598* apresenta uma inserção intrônica do TE *gypsy4* (~6456bp - *LTR/Gypsy*) encontrada em todas as linhagens analisadas, porém ausente no genoma de referência (*D. melanogaster* 6.49). A produção de transcritos quiméricos ocorre em todas as linhagens e condições, exceto *Dmgoth101*, na qual essa isoforma não é produzida (Figura 32).

Esse gene é super expresso nas linhagens brasileiras em todas as condições (Tabela Suplementar 15). A expressão de *CG42598*, de acordo com os dados de expressão publicados no *FlyBase*, é nula (RPKM = 0) tanto em ovários de moscas adultas quanto após o tratamento *Cold-Shock*. Os dados publicados, juntamente com nossas análises, parecem indicar que esse gene tem sua expressão aumentada quando ocorre a produção de transcritos quiméricos, tendo em vista que tanto Dmgoth101 quanto os dados do *FlyBase*, referente a uma linhagem que não possui essa inserção de TE, não produzem transcritos quiméricos e apresentam baixa expressão. A super expressão (valor de *Log2FoldChange* - Tabela Suplementar 15) nas linhagens brasileiras parece ter sido enviesada pela baixa expressão registrada em Dmgoth101, onde coincidentemente não ocorre a produção de transcrito quimérico. Análise adicional de diferença de expressão intra populacional revelou que *CG42598* está DE entre as linhagens francesas nas três condições, sendo mais expresso na linhagem Dmgoth63, onde há a produção de transcrito quimérico (valor de *Log2FoldChange* entre Dmgoth63 e Dmgoth101 para Controle: 2,46301841310344; *Cold-Shock*: 2,24564622874066; *Starvation*: 1,51844799083363). Essa observação parece reforçar mais o possível papel do TE *gypsy4* no aumento da expressão do gene *CG42598*.

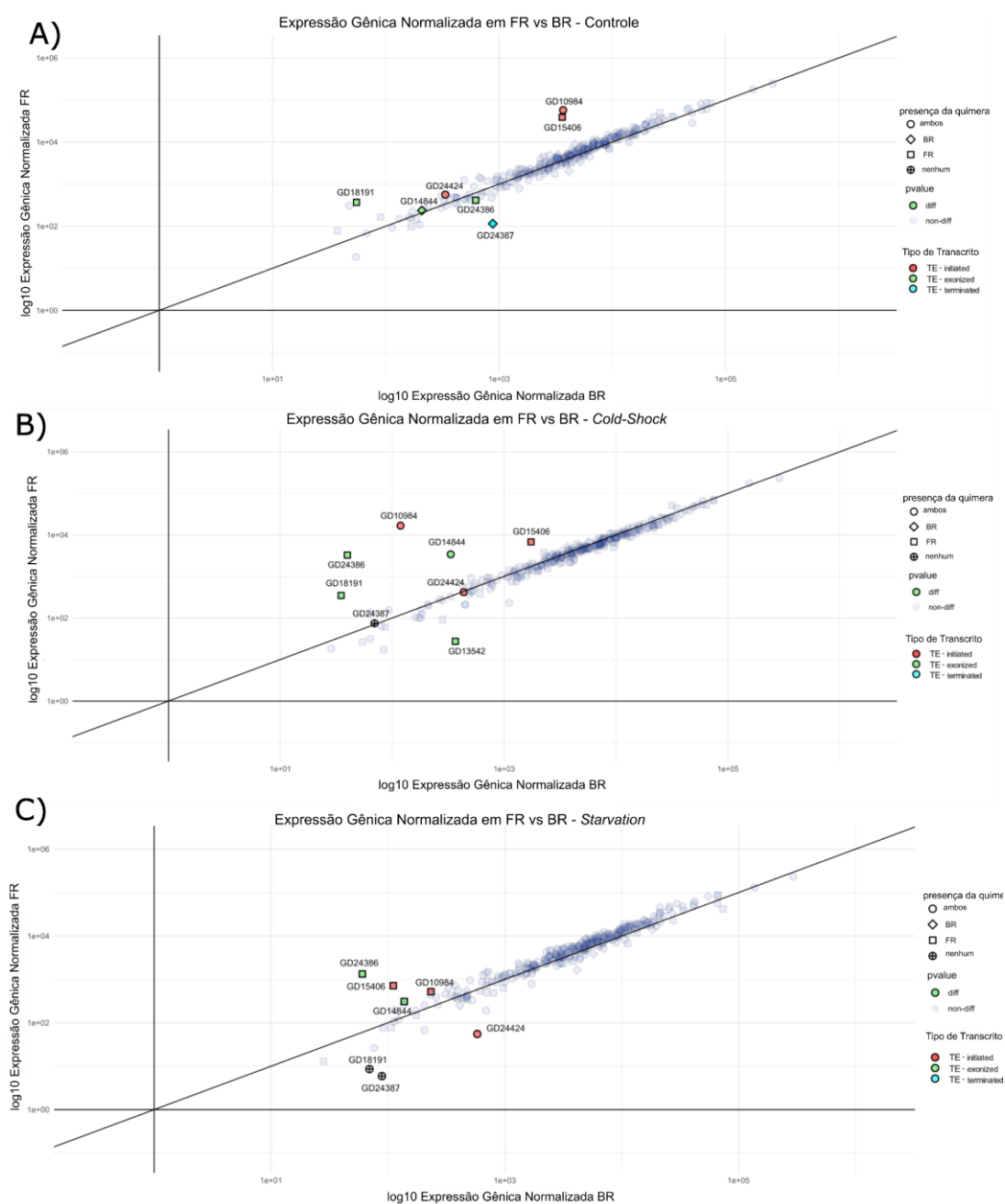
Figura 32. Ilustração da expressão do gene *CG42598* entre Controle e Tratamentos e a presença do TE no genoma das quatro linhagens de *D. melanogaster* obtida por RNA-seq. A Figura representa no genoma os genes, em traços azuis e a presença dos TEs em verde (Retrotransposon com LTRs). Ambos estão identificados pela sua nomenclatura e pelo tamanho em bp de suas sequências em cada linhagem. Acima dos genes e dos TEs, em preto, estão representadas suas respectivas expressões em cada situação, como também a cobertura de leituras encontrada em cada linhagem. As setas indicam os casos onde os genes estão super (↑) ou sub (↓) expressos, e suas cores indicam a comparação. Essa representação permite a identificação e visualização da influência do elemento na expressão do gene hospedeiro.



5.4 Expressão diferencial interpopulacional e inserções de TEs potencialmente adaptativas em *D. simulans*

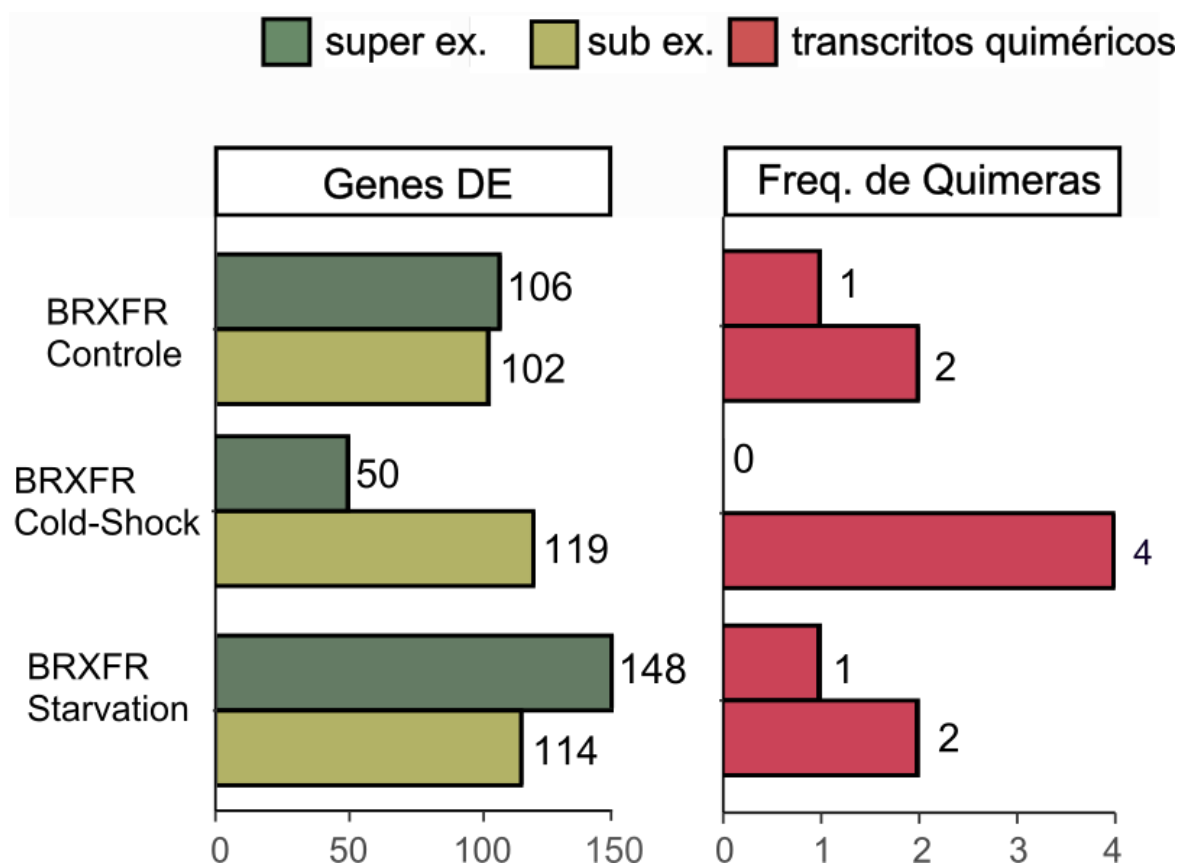
A expressão normalizada dos genes quiméricos super ou sub expressos nas linhagens das populações brasileiras e francesas de *D. simulans* (linhagens francesas Dsgoth31 e Dsgoth61 e linhagens brasileiras Dssj27 e Dssj9) em cada uma das condições (Controle, *Cold-Shock* e *Starvation*) é apresentada na Figura 33. Aproximadamente 2% do total de genes que produziram transcritos quiméricos estavam DE entre as populações amostradas de *D. simulans*.

Figura 33. Expressão de genes que geram transcritos quiméricos entre as linhagens brasileiras e francesas, nas três condições, em *D. simulans*. No eixo y estão dispostos os valores de log10 normalizados da expressão dos genes no Controle, no eixo x os valores normalizados nos tratamentos: à direita em *Starvation* e à esquerda *Cold-Shock*. Todos os genes produtores de transcritos quiméricos com suas expressões normalizadas estão incluídos no gráfico. Formas coloridas destacadas e identificadas representam genes quiméricos expressos diferencialmente entre controle e tratado entre todas as linhagens (valor de p ajustado (FDR) < 0,05; log2 fold-change > |1|); formas transparentes representam genes quiméricos não expressos diferencialmente entre todas as linhagens (valor de p ajustado). A identificação “nenhum” corresponde a genes que estão DE e são quiméricos em outras linhagens e comparações, mas não na linhagem apresentada.



A Figura 34 exibe à esquerda, os DEGs entre as populações brasileiras e francesas amostradas de *D. simulans* nas três condições (Controle e tratamentos *Cold-Shock* e *Starvation*), e à direita os genes quiméricos que estão diferencialmente expressos (Tabela Suplementar 16). Os genes estão sub e super expressos nas linhagens brasileiras em relação às linhagens francesas. Assim como observado para *D. melanogaster*, foram registrados menos DEGs entre as populações de *D. simulans* após o tratamento *Cold-Shock*, como também se verificou mais genes quiméricos sub expressos nas linhagens brasileiras nas três condições. Contudo, tanto no controle quanto em *Starvation* foram identificados mais genes super expressos na população brasileira. O maior número de genes quiméricos diferencialmente expressos foi registrado entre as populações de *Cold-Shock*.

Figura 34. Distribuição do número de DEGs que também formam quimeras em *D. simulans* entre linhagens brasileiras e francesas no Controle e tratados *Cold-Shock* e *Starvation*



A relação dos genes quiméricos DE identificados em *D. simulans* e seus ortólogos em *D. melanogaster* é apresentada na Tabela 5. Dos sete genes quiméricos DE dois produzem transcritos quiméricos exclusivos de linhagens (*GD18191* e *GD24387*), e um caso configura a produção de transcrito quimérico polimórfico populacional (*GD15406*). Mais uma vez, os resultados das análises de expressão diferencial populacional podem ter sido enviesados por

outliers (linhagens com expressão gênica muito maior ou menor que as outras numa mesma população). A família *INE-1* compreende três das sete inserções de TEs relatadas. A respeito dos sítios de inserção, três estão localizados dentro de éxons e três *upstream* aos genes. Todas as inserções dos genes presentes na Tabela 5 também foram registradas no genoma *D. simulans* de referência (r2.02), com exceção de *GD15406*.

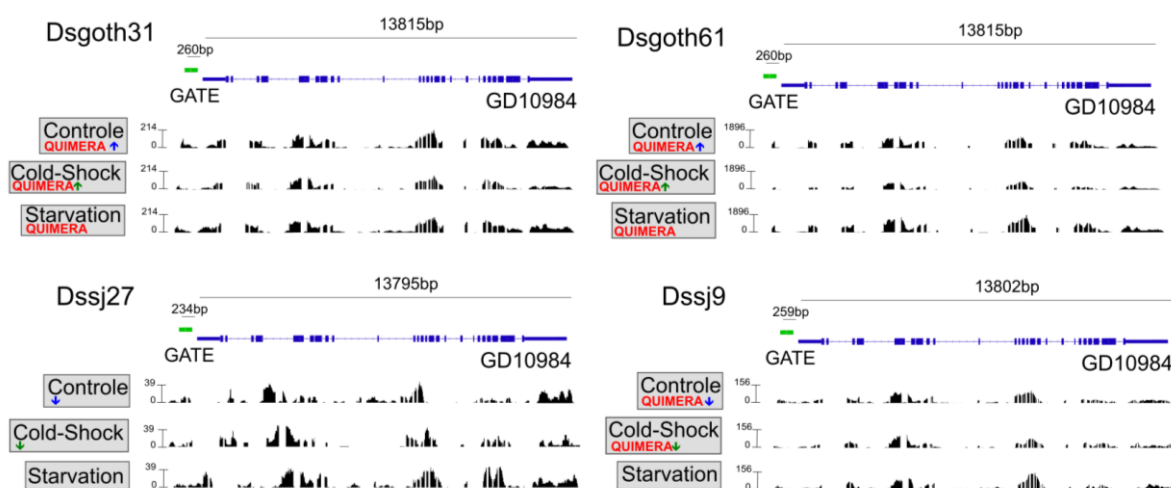
Tabela 5. Genes quiméricos diferencialmente expressos entre linhagens brasileiras e francesas, nas três condições, em *D. simulans*. O triângulo vermelho em uma linha representa a presença de um transcrito quimérico; o triângulo branco em uma linha representa a presença da inserção genômica do TE, mas sem um transcrito quimérico; a linha sem triângulos representa a ausência da inserção.

Gene	Ortólogos em <i>D.mel</i>	Função	TE	Localização do TE	Linhagens			
					Dsgoth31	Dsgoth61	Dssj27	Dssj9
GD10984	stj	Envolvido na endocitose da vesícula sináptica em massa; detecção de estímulo térmico envolvido na percepção sensorial da dor; e regulação negativa da montagem sináptica na junção neuromuscular	GATE	Upstream	—▲—	—▲—	—△—	—▲—
GD14844	CG17732	-	Roo	Exon	—	—▲—	—	—▲—
GD15406	Cyp6g1	Envolvido no processo catabólico do inseticida; resposta ao DDT; e resposta ao ion mercurio	Juan	Upstream	—▲—	—▲—	—	—△—
GD18191	CG33494	-	Hobo	Exon/Intron	—▲—	—	—△—	—△—
GD24386	Syt7	Envolvido na resposta celular ao ion cálcio; exocitose regulada; e regulação da secreção pela célula	INE-1	Exon	—▲—	—▲—	—△—	—△—
GD24387	Nfl	Envolvido na regulação da transcrição pela RNA polimerase II	INE-1	Downstream	—△—	—△—	—▲—	—△—
GD24424	CG33521	Envolvido na organização do citoesqueleto de actina	INE-1	Upstream	—▲—	—▲—	—▲—	—▲—

Dentre os genes descritos na Tabela 5, os genes *GD14844*, *GD15406* e *GD18191* também apareceram como possíveis candidatos nas análises de expressão diferencial entre controle e tratamento (Tabela 3). Os genes *GD10984*, *GD14844*, *GD18191* e *GD24386* são casos interessantes referentes a adaptação a zonas climáticas distintas. Apenas *GD24387* aparenta ter sua expressão modulada pela inserção do TE.

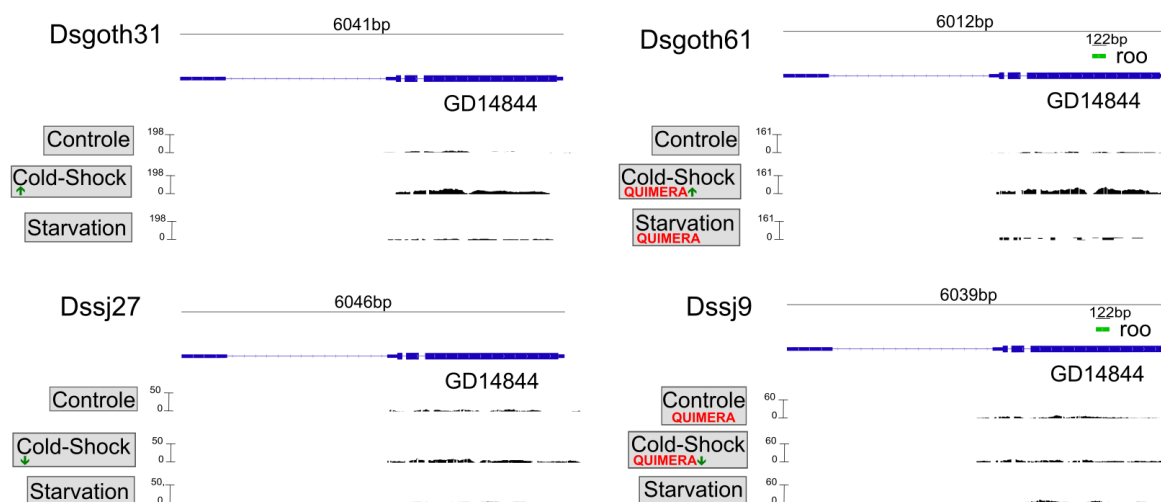
GD10984, super expresso em ambas linhagens francesas no Controle e no tratamento *Cold-Shock*, apresenta uma inserção *upstream* do TE *GATE* (257 bp - *LINE/Jockey*), comum a todas as linhagens, produzindo transcritos quiméricos, exceto em *Dssj27* (Figura 35, Figura Suplementar 67). Caso a transcrição do TE com o gene alterasse de alguma forma sua expressão (Tabela Suplementar 16), também haveria diferença significativa de expressão de *GD10984* entre as populações em *Starvation*, já que as linhagens de Gotheron são as únicas a produzirem transcritos quiméricos nessa condição. Isso parece indicar que a desregulação transcricional deste gene pode ter relação com a colonização e adaptação a condições climáticas diferentes, e não à produção de transcrito quimérico.

Figura 35. Ilustração da expressão do gene *GD10984* entre Controle e Tratamentos e a presença do TE no genoma das quatro linhagens de *D. simulans* obtida por RNA-seq. A Figura representa no genoma os genes, em traços azuis e a presença dos TEs em verde (Retrotransposon com LTRs). Ambos estão identificados pela sua nomenclatura e pelo tamanho em bp de suas seqüências em cada linhagem. Acima dos genes e dos TEs, em preto, estão representadas suas respectivas expressões em cada situação, como também a cobertura de leituras encontrada em cada linhagem. As setas indicam os casos onde os genes estão super (↑) ou sub (↓) expressos, e suas cores indicam a comparação. Essa representação permite a identificação e visualização da influência do elemento na expressão do gene hospedeiro.



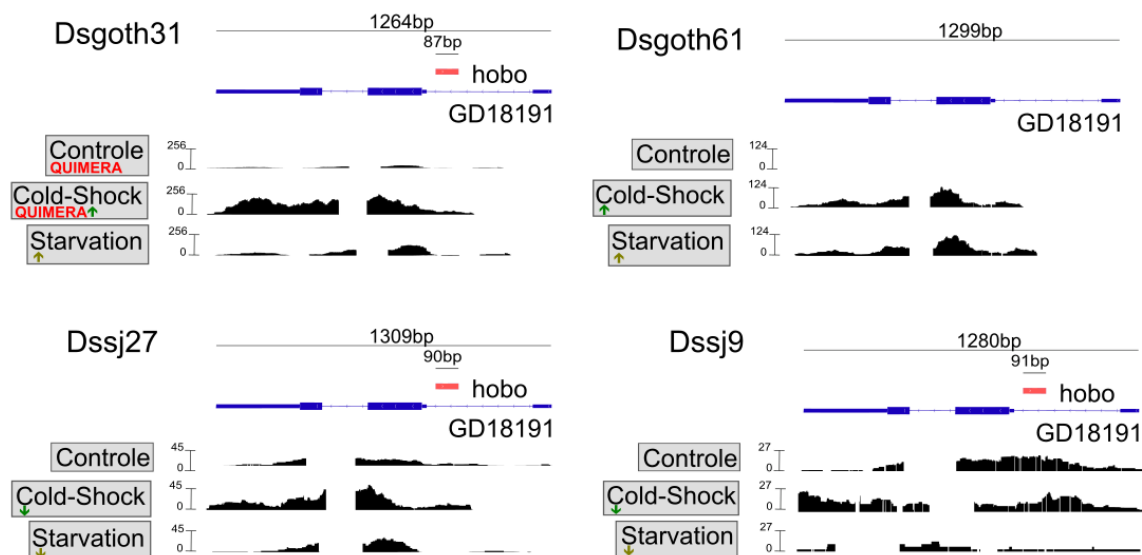
Como comentado anteriormente, *GD14844* é também um gene que produz quimera e que está diferencialmente expresso entre controle e tratamento. Esse gene apresenta uma inserção *roo* (120 bp - *LTR/Bel-Pao*) presente apenas em Dsgoth61 e Dssj9, nas quais há a produção de transcrito quimérico (Figura 36). Nas linhagens brasileiras o gene está sub expresso após o tratamento *Cold-Shock*, apesar de Dssj9 exibir a produção de transcrito quimérico. Também não há diferença de expressão significativa de *GD14844* nos demais tratamentos, mesmo com a produção de transcrito quimérico em outra linhagem. Portanto, a transcrição do TE com o gene não parece ser determinante da diferença de expressão encontrada. Esse caso parece exemplificar uma resposta das linhagens ao *Cold-Shock*, onde há diferença de expressão, especialmente nas linhagens adaptadas ao clima temperado (Figura Suplementar 68, Tabelas Suplementares 14 e 16).

Figura 36. Ilustração da expressão do gene *GD14844* entre Controle e Tratamentos e a presença do TE no genoma das quatro linhagens de *D. simulans* obtida por RNA-seq. A Figura representa no genoma os genes, em traços azuis e a presença dos TEs em verde (Retrotransposon com LTRs). Acima dos genes e dos TEs, em preto, estão representadas suas respectivas expressões em cada situação, como também a cobertura de leituras encontrada em cada linhagem. As setas indicam os casos onde os genes estão super (↑) ou sub (↓) expressos, e suas cores indicam a comparação. Essa representação permite a identificação e visualização da influência do elemento na expressão do gene hospedeiro.



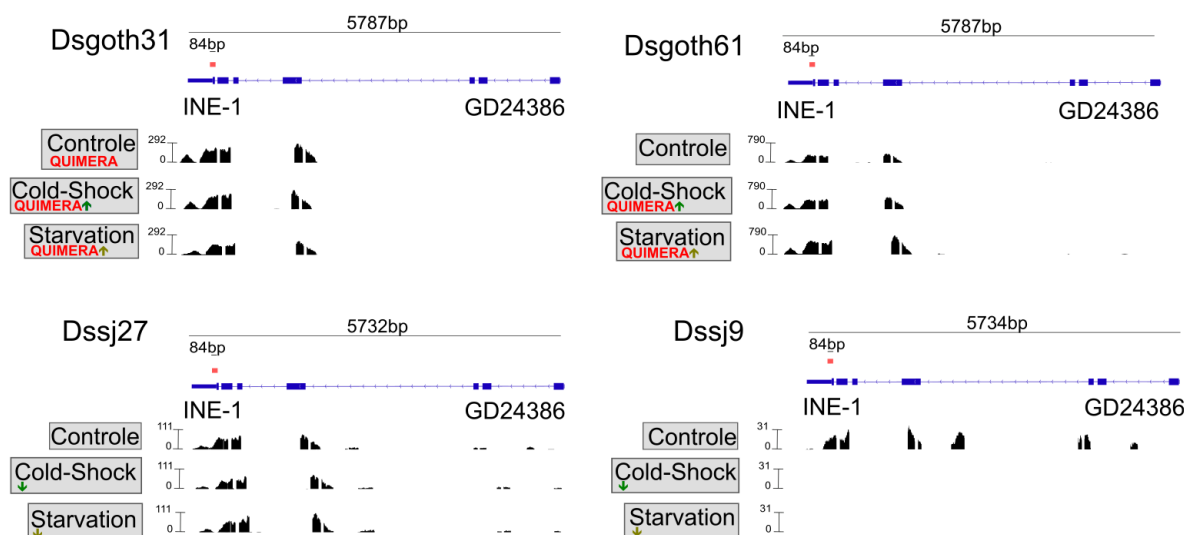
O gene *GD18191* já foi destacado por apresentar expressão diferencial entre controle e tratamentos (seção 5.2). É um gene que conta com um elemento *hobo* (85bp - *DNA/hAT*) inserido parcialmente em um éxon e parcialmente em um íntron, e que está anotado em todas as linhagens, exceto Dsgoth61 (Figura 37). *GD18191* está sub expresso nas linhagens brasileiras e só há a produção de transcritos quiméricos no Controle e *Cold-Shock* de Dsgoth31. Há uma clara diferença de expressão entre as populações, mas essa diferença não parece ser resultado da inserção do TE, tendo em vista que Dsgoth61 também exibe uma expressão maior que as linhagens brasileiras e não apresenta a inserção (Figura Suplementar 70, Tabela Suplementar 14 e 16). Portanto, esse gene parece tanto ser um caso de adaptação à zona temperada, quanto responsivo ao estresse *Cold-Shock*, dado o aparente aumento da expressão; contudo, nem a comparação interpopulacional, nem entre tratamentos (Tabela 3), sugerem que a diferença de expressão esteja relacionada com a inserção do TE e a produção de transcrito quimérico.

Figura 37. Ilustração da expressão do gene *GD18191* entre Controle e Tratamentos e a presença do TE no genoma das quatro linhagens de *D. simulans* obtida por RNA-seq. A Figura representa no genoma os genes, em traços azuis e a presença dos TEs em coral (Transposon de DNA). Acima dos genes e dos TEs, em preto, estão representadas suas respectivas expressões em cada situação, como também a cobertura de leituras encontrada em cada linhagem. As setas indicam os casos onde os genes estão super (↑) ou sub (↓) expressos, e suas cores indicam a comparação. Essa representação permite a identificação e visualização da influência do elemento na expressão do gene hospedeiro.



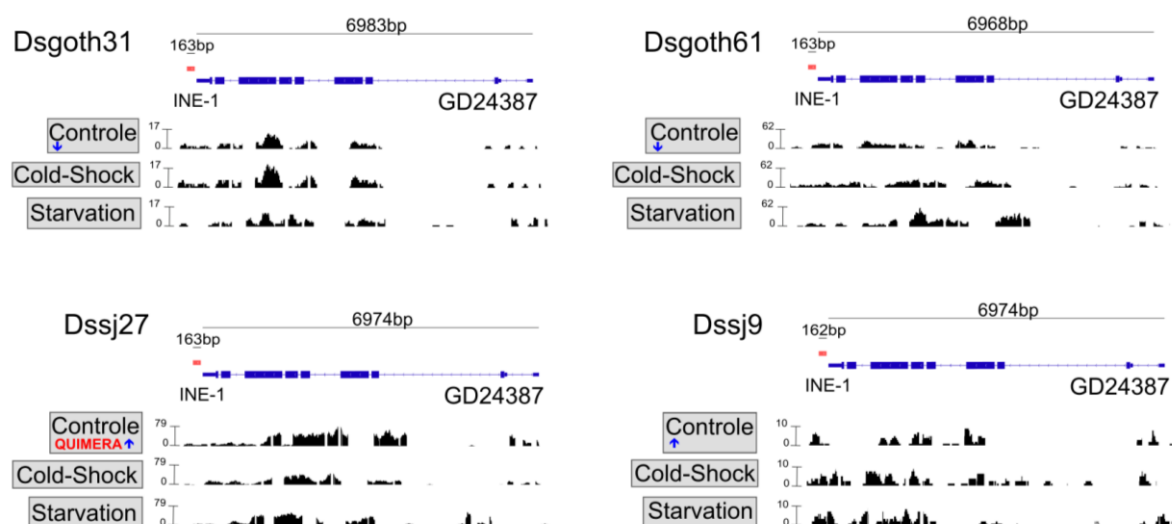
O gene *GD24386* exibe uma inserção da família *INE-1* (163 bp - DNA/Helitron) parcialmente em um éxon e parcialmente em um íntron, presente em todas as linhagens, no entanto, a produção de transcritos quiméricos ocorre apenas nas linhagens de Gotheron, exceto controle de Dsgoth61, e está sub expresso nas linhagens brasileiras, exceto controle (Figura 38). A presença do TE não parece ser determinante na diferença de expressão encontrada. Isso pode ser observado analisando a expressão na condição controle quando ocorre a produção de transcritos quiméricos em Dsgoth31, e em outras linhagens com expressão similar, mas que não produzem transcritos quiméricos, como Dsgoth61 e Dssj27 (Figura Suplementar 71, Tabelas Suplementares 14 e 16). *GD24386* é outro caso sugestivo de adaptação à zona temperada, ativada pelos estresses, tendo em vista que a diferença de expressão é significativa entre as populações após exposição aos dois tratamentos.

Figura 38. Ilustração da expressão do gene *GD124386* entre Controle e Tratamentos e a presença do TE no genoma das quatro linhagens de *D. simulans* obtida por RNA-seq. A Figura representa no genoma os genes, em traços azuis e a presença dos TEs em coral (Transposon de DNA). Acima dos genes e dos TEs, em preto, estão representadas suas respectivas expressões em cada situação, como também a cobertura de leituras encontrada em cada linhagem. As setas indicam os casos onde os genes estão super (↑) ou sub (↓) expressos, e suas cores indicam a comparação. Essa representação permite a identificação e visualização da influência do elemento na expressão do gene hospedeiro.



O gene *GD24387* apresenta uma inserção INE-1 (162 bp - DNA/Helitron) *upstream*, anotada no genoma de todas as linhagens (Figura 39). Apenas no controle de *Dssj27* há a produção de transcrito quimérico, condição em que foi relatada a superexpressão desse gene nas linhagens brasileiras em relação às linhagens francesas. Mais uma vez, a diferença de expressão calculada entre populações parece ter sido enviesada pela expressão de apenas uma linhagem. Analisando os dados normalizados de expressão e as ilustrações e painéis referentes ao gene (Figura 39 e Figura Suplementar 72), fica evidente que *GD24387* é mais expresso em *Dssj27* do que nas demais linhagens, independente da condição (Tabela Suplementar 14). Quando ocorre a produção de transcrito quimérico no Controle, a expressão é ainda maior, o que provavelmente pode ter influenciado na diferença de expressão calculada. Adicionalmente, análise de expressão diferencial intra populacional revelou que *GD24387* está super expresso no Controle de *Dssj27*, onde há a produção de transcrito quimérico, em relação ao Controle de *Dssj9* ($\log_2\text{FoldChange}$: 4,74355681863546). Esse detalhe corrobora a hipótese de que a inserção TE possa ser a responsável pelo aumento da expressão de *GD24387*, já que só é DE quando ocorre a produção de transcrito quimérico.

Figura 39. Ilustração da expressão do gene *GD24387* entre Controle e Tratamentos e a presença do TE no genoma das quatro linhagens de *D. simulans* obtida por RNA-seq. A Figura representa no genoma os genes, em traços azuis e a presença dos TEs em coral (Transposon de DNA). Acima dos genes e dos TEs, em preto, estão representadas suas respectivas expressões em cada situação, como também a cobertura de leituras encontrada em cada linhagem. As setas indicam os casos onde os genes estão super (↑) ou sub (↓) expressos, e suas cores indicam a comparação. Essa representação permite a identificação e visualização da influência do elemento na expressão do gene hospedeiro.



6 DISCUSSÃO

Dentre as diferentes maneiras de identificar e compreender a contribuição dos TEs na evolução genômica, seus efeitos na mudança de expressão de genes hospedeiros e/ou adjacentes, por meio de eventos de domesticação, têm sido bem relatados. Transcritos quiméricos gene-TEs são uma excelente ferramenta para verificar a variabilidade que sequências de TE podem gerar no transcriptoma. Neste contexto, é interessante avaliar a contribuição que os TEs tiveram na história evolutiva de diferentes organismos. Espécies cosmopolitas, como *Drosophila melanogaster* e *D. simulans*, com populações adaptadas a localidades distintas, expostas a estresses ambientais diferentes, se tornam excelentes modelos para investigar eventos potencialmente adaptativos de TEs, como a domesticação, que podem estar associados com a colonização e à adaptação dessas espécies.

A produção de transcritos quiméricos difere entre as espécies, mas não entre linhagens e tratamentos

O conteúdo genômico de TEs anotado nesse trabalho foi maior nos genótipos de *D. melanogaster* que nos de *D. simulans*, assim como na quantidade de transcritos quiméricos identificados. Mostramos neste trabalho que a disponibilidade de cópias dos TEs no genoma parece proporcionar maior produção de transcritos quiméricos, dependendo de sua localização no genoma, em relação a regiões eucromáticas e a genes ativamente transcritos (Fontanillas; Hartl; Reuter, 2007). Diversos estudos apontam para a existência de uma certa preferência dos TEs em se inserirem em regiões regulatórias de genes, na heterocromatina, bem como em regiões centroméricas e teloméricas (Pimpinelli et al, 1995; Pardue; DeBaryshe, 1999; Biemont; Vieira, 2006; Mazzucchelli; Martins, 2009; Auvinet et al. 2019). Em *D. melanogaster*, os TEs configuram de 50 a 60% das regiões heterocromáticas e somente 4 a 6% da eucromatina (Kapitonov; Jurka, 2003; Bergman et al., 2006).

A distribuição genômica dos TEs depende, principalmente, da tolerância do genoma às inserções (Fontanillas; Hartl; Reuter, 2007). Levando-se em consideração que a maioria das inserções são deletérias, a abundância, distribuição e atividade desses elementos é regulada, controlada e reprimida por diferentes mecanismos do genoma. Um desses mecanismos é a expansão da heterocromatina, que pode modular a expressão de genes adjacentes à inserção (Lee, 2015; Lee; Karpen, 2017). Já foi relatado que as histonas repressivas da heterocromatina (H3K9me2) exibem níveis mais altos de expressão em *D. simulans* (Lee; Karpen, 2017), podendo significar que mecanismos de regulação das

atividades de TEs pela heterocromatina são mais intensos nessa espécie (Chakraborty et al., 2019). Análises realizadas por Fablet e colaboradores (2022), combinando dados de inserção de TE polimórfica, RNA-seq, ChIP-seq de dois marcadores de histonas (H3K4me3 e H3K9me3), e RNA-seq de pequenos RNAs das mesmas linhagens de *D. melanogaster* e *D. simulans* analisadas neste trabalho, revelaram que a regulação epigenética é mais forte nas inserções de TEs em *D. simulans*. Isso sugere que não apenas o acúmulo de cópias de TEs observado em *D. melanogaster* em relação a *D. simulans* seja um fator determinante na produção de quimeras, mas também, que há mais TEs disponíveis na eucromatina das linhagens de *D. melanogaster* para serem recrutados na transcrição do que em *D. simulans*, podendo justificar a diferença observada neste trabalho na produção de transcritos quiméricos entre as espécies irmãs.

As diferenças conhecidas nos conteúdos genômicos de TEs também refletem a história de invasão genômica por família de TEs em ambas espécies. Evidências indicam que *D. melanogaster* exibe marcas de terem passado por explosões de transposição de determinadas famílias TEs mais recentes de que *D. simulans* (Bowen; McDonald, 2001; Bergman; Bensasson, 2007; Kofler; Hill et al., 2015; Kofler; Nolte et al., 2015), por conter um grande número de cópias completas, ao passo que *D. simulans* contém muitas cópias degradadas, inativas e deletadas (Lerat et al., 2011). A via de RNAs que interage com as proteínas *PIWI* (piRNA), específicas de linhagens germinativas, e muito bem descrita em *Drosophila* (Brennecke et al., 2008), constitui um dos mecanismos de regulação transcricional e pós-transcricional de TEs mais importantes, mantendo a estabilidade das células germinativas e, posteriormente, protegendo o zigoto da ação dos TEs (majoritariamente herdados da fêmea). Apesar de Saint-Leandre e colaboradores (2020) terem observado uma forte relação entre características de TEs recentemente ativos e os níveis de piRNAs, que indicavam uma preferência por famílias TEs que recentemente experimentaram explosões de transposição, a mesma correlação existia para *D. simulans*, que, como comentado, abriga cópias mais degradadas e antigas de TEs. Sabe-se que a produção de piRNAs depende da presença de famílias, que, ao atingirem um alto número de cópias, se acumulam em *clusters* nas regiões heterocromáticas, estabelecendo-se e estabilizando a produção das moléculas precursoras desse mecanismo de silenciamento (Kelleher; Barbash, 2013). Assim, os autores concluíram que a resposta de silenciamento por piRNAs persiste em espécies como *D. simulans* por atuarem como uma memória de longo prazo, garantindo a proteção genômica de possíveis futuras invasões (Saint-Leandre et al., 2020). As diferenças entre mecanismos de

silenciamento pré-transcricionais, frutos das diferentes histórias de invasão genômicas de TEs, juntamente com as diferenças inerentes específicas à regulação gênica podem explicar as diferenças no número de transcritos quiméricos observadas entre *D. melanogaster* e *D. simulans*.

A produção de transcritos quiméricos não pareceu ser afetada por variações intra e interpopulacionais de ambas espécies, já que as linhagens não exibiram diferenças no número total de transcritos. O *background* genético referente a história evolutiva das linhagens e demografia das populações, que remontam invasões genômicas de TEs e, por conseguinte, os mecanismos específicos regulatórios de suas atividades, embasam as diferenças encontradas na quantidade e diversidade de famílias ativas nesses transcritos. Por outro lado, a estabilidade dos mecanismos regulatórios das atividades de TEs intra espécies, também explica a similaridade na produção de transcritos quiméricos encontradas em nossas análises.

Nossos resultados também indicam não haver diferenças no número de transcritos quiméricos entre os controles e a exposição aos estresses *Cold-Shock* e *Starvation*. Há exemplos na literatura relatando a ativação de TEs por estresse, que, por conterem sequências *cis* regulatórias, podem também modular em *trans* a expressão de genes hospedeiros ou adjacentes. Contudo, a ativação de TEs por estresse por si só é um evento raro (Oliveira et al., 2021), correlacionado positivamente com o quanto o estresse também causa uma resposta na expressão gênica. A escassez, no geral, de transcritos quiméricos indica que as inserções de TEs que os geram são provavelmente fortemente deletérias (Lipatov et al., 2005); assim a ativação de TEs quiméricos por estresse e, por conseguinte, a modulação de genes associados, é ainda mais reduzida. O efeito do estresse parece ser mais notado na expressão do que na quantidade de TEs recrutados e transcritos quiméricos produzidos. Portanto, nossos resultados sugerem que a regulação da transcrição de TEs é prioridade para o genoma, sendo que o aumento de sua transcrição em genes quiméricos, não é utilizada como resposta imediata ao estresse.

Dois interessantes cenários puderam ser distinguidos entre os transcritos quiméricos gerados neste trabalho: transcritos polimórficos e inserções de TEs polimórficas. Esses transcritos quiméricos foram reconhecidos por serem produzidos exclusivamente em uma linhagem derivada da mesma população, ou exclusiva de linhagens derivadas de uma população e não da outra. Analisando os TEs constituintes desses transcritos exclusivos, puderam ser identificadas 67 inserções polimórficas entre as linhagens de *D. melanogaster* e

70 nas de *D. simulans*. Importante ressaltar que a presença de um TE no genoma de uma linhagem não significa necessariamente produção de transcrito quimérico; contudo, a ausência da inserção TE no genoma resulta em ausência de transcrito quimérico. As diferenças no recrutamento de inserções fixas no transcriptoma, aquelas encontradas no genoma de todas as comparações, podem ser explicadas pela atuação dos mecanismos epigenéticos de regulação transcricional, descritos anteriormente. As isoformas geradas por TEs quiméricos fixos também podem ser consequência do sítio de inserção, caso forneçam novos sítios de *splicing* alternativo, como TEs intrônicos (Treiber; Waddell, 2020). Uma inserção polimórfica de TE sugere tanto eventos de mobilização recentes, após divergência populacional, por exemplo, como também a perda das inserções. Portanto, a regulação transcricional por mecanismos epigenéticos específicos justifica somente os transcritos quiméricos polimórficos gerados por inserções fixas, e o carácter independente dos eventos de mobilização pode explicar polimorfismos de inserção encontrados.

Nossa estimativa de genes que produziram quimeras no genoma está de acordo com o que já foi relatado na literatura. Encontramos cerca de 1,5% dos genes, tanto entre as linhagens de *D. melanogaster* quanto entre as de *D. simulans*, produzindo transcritos quiméricos. Lipatov e colaboradores (2005) encontraram ~1 % em *D. melanogaster* utilizando marcadores de sequências expressas (do Inglês: *expressed sequence tags* (ESTs)) e Batur e Ginjeiras (2013) encontraram até 1% dos genes produzindo transcritos quiméricos iniciados por TE ao analisar 36 estágios do ciclo de vida de *D. melanogaster*, que correspondiam a 1,6% de todos os transcritos. Também, Treiber e Waddell (2020) analisando o mesencéfalo de *D. melanogaster*, contabilizaram 264 genes produzindo transcritos quiméricos, o que correspondia a ~1,5% do conteúdo gênico. Oliveira e colaboradores (2023) utilizaram o genoma anotado (Modolo; Lerat, 2015; Mohamed et al. 2020), das linhagens de *D. melanogaster* utilizadas neste trabalho, na situação controle, e também encontraram 1,12% dos genes produzindo transcritos quiméricos. Portanto, concluímos que genes produtores de transcritos quiméricos são raros no genoma e não atingem 2%, independente da metodologia utilizada para a sua identificação.

Transcritos quiméricos exonizados configuraram cerca de 92% do total produzido em *D. melanogaster* e 81% em *D. simulans*. Em *D. melanogaster* já foi relatado que 28% das inserções de TEs eucromáticas ocorrem dentro de genes (Kaminker et al., 2002; Lipatov et al., 2005). Ou seja, há maior ocorrência de TEs dentro de genes passíveis de transcrição, o que pode justificar o predomínio de transcritos TE-*exonized* encontrados. Oliveira e

colaboradores (2023) encontram 89% de transcritos quiméricos *TE-exonized*. Os autores atribuíram esse resultado à má classificação de transcritos iniciados e terminados por TEs pelo ChimeraTE, que inclui apenas inserções que estão totalmente fora das regiões gênicas na identificação de transcritos dos tipos *TE-initiated* e *TE-terminated*. Os nossos resultados podem também ser fruto dessa circunstância, pois uma vez que a inserção não esteja totalmente independente do gene, mesmo se o transcrito de fato iniciar ou terminar no TE, este será considerado como *TE-exonized* pela ferramenta.

Antigos, Ativos e Invasores: papel de elementos das famílias *INE-1* e *roo* e do elemento *P* na produção de transcritos quiméricos e polimorfismos de inserção

Os elementos de Classe I são os mais abundantes no genoma de ambas espécies, e produziram mais transcritos quiméricos. Os elementos de Classe I são os maiores responsáveis pelo aumento do número de cópias TE no genoma devido ao seu mecanismo de mobilização “copia e cola” (Wicker et al., 2007), sendo os LTRs os maiores contribuintes na produção de transcritos quiméricos tanto em *D. melanogaster* quanto em *D. simulans*. Além de pertencerem à ordem de TEs mais representada em *Drosophila*, os retrotransposons com LTRs também são os mais ativados por fatores estressantes. Por serem mais propensos a manter suas atividades *cis* reguladoras ancestrais (Chuong; Elde; Feschotte, 2017), parecidas com os motivos necessários para ativação de genes responsivos ao estresse (Grandbastien et al., 2005), é provável que os retrotransposons com LTRs também sejam ativados por estarem próximos de genes responsivos ao estresse. De fato, a contribuição dos retrotransposons com LTRs na produção de transcritos quiméricos parece ter sido maior, no entanto, não parece ser resultado de ativação por fatores estressantes, uma vez que não identificamos discrepância no número de transcritos quiméricos entre tratamentos e controle. Portanto, a predominância de retrotransposons com LTRs em transcritos quiméricos parece ter relação com a maior disponibilidade de cópias dessa ordem no genoma.

A família *INE-1* (Helitron/Transposons de DNA) foi a mais abundante nos genomas das duas espécies, e também a maior contribuinte em transcritos quiméricos nas linhagens de *D. melanogaster*. Os elementos da família *INE-1* participaram da produção de 37 transcritos quiméricos polimórficos em *D. simulans*, e somente uma inserção polimórfica, exclusiva da linhagem Dssj27; enquanto que em *D. melanogaster* foram encontrados 22 transcritos quiméricos polimórficos com essa família, mas nenhuma inserção polimórfica. Esses TEs são antigos (Kapitonov; Jurka, 2003; Kofler et al.; 2015), e ao longo de sua história evolutiva se

espalharam pelos genomas, aumentando o número de cópias; contudo, correspondem, a inserções imóveis, que acumularam mutações e perderam sua capacidade de mobilização. Estes elementos parecem ter invadido o genoma de *D. simulans* mais recentemente que em *D. melanogaster* (Chakraborty et al., 2021); portanto, não é surpreendente terem sido encontradas inserções polimórficas dessa família em *D. simulans* já que, nessa espécie, essa família é mais jovem, podendo existir elementos que ainda possuam capacidade de transposição.

A segunda família mais abundante no genoma das espécies irmãs e a mais frequente nos transcritos quiméricos nas linhagens de *D. simulans* é a família *roo*. Essa família é a mais abundante da eucromatina (Vieira; Biémont, 2004; Quesneville et al., 2005; Rahman et al., 2015), e já foi registrada em alguns casos de eventos de domesticação em *Drosophila* (Malik; Henikoff, 2005; Volff, 2006; Nefedova et al., 2014; Merenciano et al., 2023). Sendo uma das famílias mais ativas do genoma, suas altas taxas de transposição (Díaz-Gonzalez 2010, 2011), explicadas por exibir poucas marcas de histonas repressivas (Yasuhara; Wakimoto, 2008; Rebollo et al., 2012), acabam gerando muitos polimorfismos genômicos entre linhagens. Em nossas análises, *roo* foi responsável pela produção de 52 transcritos quiméricos exclusivos e 26 inserções exclusivas entre as linhagens de *D. melanogaster* e 53 transcritos quiméricos exclusivos e 23 inserções exclusivas entre as linhagens de *D. simulans*. Nota-se que *roo* continua sendo a família com o maior número de polimorfismos de inserção em transcritos quiméricos, intra e interpopulacionais de ambas espécies, sendo todas inseridas dentro de genes. A maior disponibilidade de inserções polimórficas *roo* no genoma parece propiciar um maior recrutamento dessas inserções em isoformas de transcritos. Nossos resultados sugerem que as inserções *roo* mantidas dentro desses transcritos apresentam principalmente efeitos neutros em resposta ao estresse, mas também podem resultar em alguma vantagem adaptativa.

Nossos resultados demonstraram que, em *D. melanogaster*, quase metade das inserções de TEs dentro de genes que produziram transcritos quiméricos estão localizadas em CDSs, e cerca de 71% em *D. simulans*. A tolerância a essas inserções que interrompem o quadro de leitura, possivelmente comprometendo a funcionalidade da proteína, pode ter relação com as famílias mais frequentes em cada região. A família *roo* foi a mais encontrada em CDSs em ambas as espécies, com uma média de tamanho de 109bp, enquanto que nas regiões 3'UTR, especialmente em *D. melanogaster*, a família *INE-1* foi a mais frequente (~138bp). Em ambas famílias, a média de tamanho dessas inserções não chega a 200bp, podendo indicar a perda de porções da sequência que permitiam sua mobilização e ativação.

Por *roo* ser uma família muito ativa, que se mobiliza mais regularmente no genoma, não é surpreendente encontrá-la em maior frequência. Contudo, a média de tamanho desses TEs parece indicar que, sequências pequenas são normalmente mantidas em regiões codificantes por, provavelmente, estarem imobilizadas e sua tradução não ser prejudicial ao hospedeiro. Interessante citar que já foi encontrado, em *Arabidopsis thaliana* (Kabelitz; Baurle; 2015), uma correlação entre redução do silenciamento e a quantidade de inserções em 3'UTR, onde foi argumentado que talvez exista uma redução nos mecanismos de silenciamento nessa região, a fim de evitar o silenciamento do gene adjacente; como também, o processamento da extremidade 3' na transcrição, pode indiretamente manter essas inserções protegidas. Assim, *D. melanogaster* possui muitas inserções de TEs em 3'UTR, especialmente *INE-1*, pois ao longo da evolução essas sequências foram protegidas de silenciamento e degradação. Por *INE-1* ser uma família antiga, esta deve ter se mobilizado para as regiões regulatórias a mais tempo. O mesmo pode ainda não ter sido observado em *D. simulans*, por apresentar um histórico de invasão de TE mais recente.

Curiosamente, o elemento P, com histórico de invasão genômica em ambas espécies, teve uma participação interessante na produção de transcritos quiméricos e inserções exclusivas. Descoberto em *D. melanogaster* na década de 70 (Hiraizumi, 1971; Hiraizumi et al., 1973; Kidwell; Kidwell; Dved, 1977; Engel; Preston, 1979), devido a sua relação com a síndrome da disgenesia híbrida, esse TE invadiu o genoma da espécie a partir de transferência horizontal de *D. willistoni*, na América do Norte (Daniels et al., 1990) originalmente em meados de 1800 (Engels, 1992), e teve uma rápida dispersão nas populações naturais de *D. melanogaster* entre 1950-1990. Em nossas análises, elementos P participaram de sete transcritos quiméricos exclusivos em *D. melanogaster*, todos na população de São José do Rio Preto, dentre os quais, seis eram inserções polimórficas, revelando que essa família parece estar mais ativa nas populações brasileiras do que nas europeias. Essa diferença pode estar relacionada ao histórico de dispersão de *D. melanogaster*, originária do continente Africano, que colonizou a América do Sul após a bem-sucedida ocupação no Hemisfério Norte. Assim, a invasão do elemento P nas populações naturais do Hemisfério Sul ocorreu mais tarde; por isso, as inserções dessa família nas linhagens francesas tiveram mais tempo de serem reguladas, degradadas e inativadas. Nossas análises indicam que os elementos P estão mais ativos em *D. simulans*, pois formaram três vezes mais transcritos quiméricos exclusivos, dentre os quais 10 configuraram inserções exclusivas entre as linhagens europeias. A detecção de elementos P em populações europeias de *D. simulans* é bem mais recente, a partir de 2006,

fruto provavelmente de um único evento de transferência horizontal a partir de *D. melanogaster* (Kofler et al., 2015), que rapidamente se espalhou para populações da América do Norte, África, Japão (Hill et al., 2016; Yoshitake et al., 2018), e só recentemente foram registrados entre populações naturais brasileiras (Nascimento et al., 2020). Portanto, foram identificadas mais inserções do elemento *P* polimórficas e potencialmente ativas nas populações francesas de *D. simulans*, pois a invasão foi mais recente, com tempo menor para dispersão e fixação entre as populações e o começo da regulação desse elemento. Por outro lado, a invasão do elemento *P* nas populações brasileiras é ainda muito recente para se observar inserções exclusivas.

Inserções de TEs potencialmente adaptativas ativadas por estresse

Ao todo foram contabilizados 41 DEGs entre as condições tratados e controle que produziram transcritos quiméricos. Os elementos de Classe I corresponderam a 63% e 56% das inserções referentes a DEGs, em *D. melanogaster* e *D. simulans*, respectivamente. Dentre eles, 78% em *D. melanogaster* e 86% em *D. simulans* eram retrotransposons com LTRs. Dentre os genes candidatos a terem TEs domesticados associados aos tratamentos, 17 correspondem à inserção da família *roo*. Elementos dessa família já foram associados com alterações na expressão de genes responsivos ao estresse devido a presença de TFBSs (Merenciano et al., 2016). O estresse tem a capacidade de inibir o silenciamento de genes e reativar os TEs, sendo que a regulação epigenética é sensível a mudanças ambientais (Biémont, 2010; Slotkin; Martienssen, 2007). Contudo, já que a desestabilização do genoma pode também levar a um mau funcionamento de sistemas genéticos, a reativação dos TEs pode ser resultado desse efeito em vez da ação direta do estresse (Capy et al., 2000). De todos os DEGs quiméricos identificados, somente em seis a desregulação parece estar relacionada com a inserção do TE e/ou responsividade ao estresse. Os dados obtidos mostram que inserções de TE potencialmente adaptativas ativadas por estresse são raras, e que a grande maioria das inserções que produziram transcritos quiméricos devem ter efeito neutro e, por isso, são mantidos no transcriptoma. Nenhum dos genes relatados têm associação documentada com os estresses na literatura.

As análises realizadas entre os DEGs entre o controle e depois da exposição aos estresses, que também produziram transcritos quiméricos, foram conduzidas a fim de encontrar potenciais casos onde a diferença de expressão encontrada fosse resultado da atuação do TE associado, que por sua vez tenha sido ativado pelo estresse. Deles, a expressão

dos genes *Paris*, *Sulf1* em *D. melanogaster* e *GD24271* em *D. simulans* está associada à presença do TE e, portanto, são sugestivos potenciais casos de domesticação de TEs.

Dentre os três genes cujos transcritos quiméricos sugerem domesticação das inserções de TEs, dois exercem funções bem relatadas. O gene *Paris* apresentou uma inserção polimórfica de *Transpac upstream* ao gene, exclusiva das linhagens europeias, mas que só produziu transcritos quiméricos em uma das linhagens após a exposição ao estresse *Cold-Shock*, onde estava DE entre controle e *Cold-Shock*. Esse gene é descrito como sendo substrato de PINK1 e *parkin*, genes esses associados ao desenvolvimento de deficiência do conteúdo de dopamina (DA) cerebral e degeneração seletiva dos neurônios DA da substância negra na doença de Parkinson (DP) em *Drosophila* (Pirooznia, et al., 2020), além de conter motivos de C2H2 proteínas *zinc-finger* (ZFPs) em suas sequências (Chung; Schäfer; Jäckle; Böhm; 2002; Zolotarev; Maksimenko; Georgiev; Bonchuk, 2016; Shapiro-Kulnane; Bautista; Salz, 2021).

O gene *Sulf1* também exibiu uma inserção polimórfica intrônica do elemento *Micropia*, exclusiva das linhagens europeias, mas que produzia transcritos quiméricos apenas em uma das linhagens de Gotheion após exposição a ambos estresses, onde estava DE em relação ao controle. Esse gene está relacionado com a compensação de sulfatação, um sistema regulador da biossíntese de sulfato de heparano (HS) (Dejima et al., 2013); é o único homólogo da família *Sulf1* em *Drosophila*, conhecidos por removerem especificamente grupos 6-O-sulfato dentro das regiões altamente sulfatadas no HS, sendo que *Sulf1* regula negativamente a sinalização do gene *Wingless* (Wg), resultando na redução das asas (Kleinschmit et al., 2013).

Há uma carência na literatura de relatos funcionais dos genes de *D. simulans*; contudo, conhece-se as funções de seus ortólogos em *D. melanogaster*. *GD24271* possui uma inserção do TE *G5A* comum a todas as linhagens, mas a produção de transcritos quiméricos só foi detectada em uma única linhagem, após exposição ao tratamento *Cold-Shock*, onde também estava DE. A expressão de *GD24271* pareceu ter sido modulada pela inserção, tendo em vista que alterações na expressão foram detectadas apenas quando houve a produção de transcrito quimérico. Esse gene é homólogo de *CG18810*, relatado por pertencer a família de genes transferases que controlam o ciclo S-palmitoilação, que contém um domínio de proteína altamente conservado, rico em cisteína, e um mote de Aspartato-Histidina-Histidina-Cisteína (DHHC). S-palmitoilação é uma modificação dinâmica pós-traducional, que altera a função de proteínas, como proteínas de membrana, citoplasmática e integrais (Bannan et al., 2008).

Os TEs que foram as quimeras atuam não apenas aumentando, ou diminuindo a expressão de seus genes hospedeiros/adjacentes, podendo estar ou não associados aos estresses. Foram identificados casos em que, apesar da formação do transcrito quimérico, o TE exibia um efeito neutro e não parecia contribuir com a modulação da expressão gênica. A diferença de expressão encontrada entre controle e tratamento era, portanto, referente à responsividade do gene em particular ao estresse, como também à susceptibilidade da linhagem devido seu *background* genético e histórico adaptativo. Assim, os genes *CG15043*, *RhoGEF64C* em *D. melanogaster* e *GD18191* em *D. simulans* pareceram ser genes responsivos ao frio, já que a modulação de sua expressão não pareceu depender da presença do TE e sim do estresse submetido.

Colonização de novos ambientes e domesticação de TEs

Diferenças de expressão em genes quiméricos entre populações que configuraram adaptações relacionadas a colonização de novos ambientes foram ainda mais raras que as ativadas por estresse. Ao todo, 16 DEGs foram identificados em *D. melanogaster* e *D. simulans*. Em ambas espécies houve um predomínio de inserções da família *INE-1*, família já descrita por ser antiga e inativa. Candidatos potenciais de domesticação, neste contexto, idealmente, seriam casos onde a inserção de TE associada à diferença de expressão gênica é, exclusiva de amostras de uma população geográfica, ou comum às populações de diferentes regiões geográficas, mas a produção de transcrito quimérico é exclusiva de populações de apenas uma região. Não foram identificados entre os DEGs das amostras de populações europeias ou sul-americanas casos de polimorfismos de inserção entre essas populações, nem transcritos quiméricos polimórficos sugestivos de domesticação. Contudo, foi possível caracterizar genes DE cuja variação de expressão sugere uma resposta adaptativa associada com a colonização de regiões climáticas diferentes, por estarem super e sub expressos em uma ou outra linhagem de uma região específica, como os genes *CG40498*, *GD10984*, *GD14844*, *GD18191* e *GD24386*. Esses genes exibirão expressões elevadas apenas das linhagens europeias, independente da condição experimental, podendo indicar uma relação com adaptações decorrentes da colonização da zona temperada.

Por outro lado, três genes quiméricos foram considerados potenciais casos de domesticação de TEs, *CG5096*, *CG42598* e *GD24387*. O gene *CG5096*, que é super expresso nas duas linhagens de Gotheron, sugerindo um potencial caso de adaptação ao clima temperado, apresenta uma inserção do retrotransposon *412* em apenas em uma delas. A

produção de transcrito quimérico nessa linhagem parece aumentar ainda mais a expressão desse gene, sugerindo um potencial caso de domesticação. Não há função bem descrita para esse gene, tendo sido relatado apenas seu envolvimento com adesão celular, desenvolvimento e transmissão neural (Riedl et al., 2005). O gene *CG42598* exibe uma inserção comum do retrotransposon *Gypsy4* nas linhagens analisadas, e com exceção de uma linhagem francesa, todas as outras produzem transcritos quiméricos nas três condições, nas quais é observado o aumento de sua expressão. Não há função relatada para esse gene. Já o gene *GD24387* de *D. simulans*, assim como o último gene, também exibe uma inserção do transposon de DNA *INE-1* comum a todas as linhagens, mas a produção de transcrito quimérico em apenas em uma delas (Dssj27), na qual ocorre um aumento significativo de expressão. Como comentado anteriormente, genes em *D. simulans* não são tão bem descritos, seu gene homólogo em *D. melanogaster*, *Nfl*, foi descrito como sendo essencial à fertilidade masculina em *Drosophila* (Flybase (<http://flybase.org/>); MGI database (<http://www.informatics.jax.org/>; Yu et al., 2015).

A domesticação é produto final de um processo evolutivo longo e com muitas etapas

Os genes *DE Paris*, *Sulf1*, *CG5096*, *CG42598*, *GD24271* e *GD24387*, que produziram transcritos quiméricos, se mostraram, portanto, os candidatos mais promissores a revelar eventos de domesticação de inserção de TEs. Esses genes podem ter tido sua atividade transcricional modificada devido ao fornecimento de TFBSs e marcas epigenéticas pelos TEs (Mombach et al., 2022), o que aumenta a acessibilidade da cromatina para a maquinaria transcricional. Contudo, apenas a identificação de transcritos quiméricos não oferece um resultado conclusivo sobre sua domesticação, já que mRNAs podem vir a ser degradados por mecanismos pós-transcricionais, que assegurem a viabilidade do genoma e a funcionalidade do mRNA, ameaçadas pelo TE ou pela transcrição incorreta do gene. Dentre os mecanismos pós-transcricionais que podem silenciar a expressão de quimeras estão metilação do DNA, modificação de histonas e produção de siRNAs (do Inglês, *Small Interfering RNA*). A metilação do DNA consiste na transferência de grupos metil no carbono 5 da citosina em eucariotos, que impede a ligação de fatores de transcrição (Bernstein et al., 2007) ou favorece a ligação de repressores transcricionais (Newell-Price et al., 2000). Contudo, este não é um mecanismo de silenciamento identificado em *D. melanogaster* e *D. simulans*. Já modificações na cauda N-terminal das proteínas histonas (acetilação, fosforilação, metilação, ubiquitinação) podem interferir no estado de condensação da cromatina afetando a acessibilidade da maquinaria transcricional (Sutherland; Costa, 2003). A produção de siRNAs (do Inglês, *Small*

Interfering RNA) em *Drosophila* e humanos, os piRNAs já comentados, está intimamente associado com o tecido germinativo materno e garante o silenciamento de TEs e a viabilidade do zigoto (Kidwell, 1983). Portanto, uma investigação sobre a domesticação dos TEs assinalados a esses genes e o valor adaptativo dessas isoformas deve ser realizada.

7 CONCLUSÕES E CONSIDERAÇÕES FINAIS

Os elementos de transposição (TEs) são capazes de promover variabilidade genética, devido a sua capacidade de mobilização, e podem ser cooptados caso contribuam com a aptidão do organismo, podendo participar na evolução adaptativa orgânica (Sundaram; Wysocka, 2020; Feschotte, 2008). A domesticação desses elementos pode atribuir aos TEs a capacidade de modular a expressão de genes hospedeiros e/ou adjacentes, e será mantida caso seja vantajosa (Capy, 2021). A identificação de eventos de domesticação é de suma importância para mensurar a contribuição de inserções de TEs na evolução adaptativa do organismo. Transcritos quiméricos gene-TE são marcadores da variabilidade gerada por TEs no transcriptoma (Nicolau et al., 2021) e podem ser uma ótima ferramenta para buscar eventos de domesticação dessas sequências genéticas móveis.

Em síntese, nosso trabalho permitiu propor:

- 1) Casos potenciais de domesticação de TEs associados aos estresses submetidos destacando-se, dentre eles, os genes quiméricos *Paris*, *Sulf1* e *GD24271*;
- 2) Casos potenciais de domesticação de TEs independente da exposição a estresses, destacando-se os genes *CG5096*, *CG42598* e *GD24387*;
- 3) Casos potenciais de genes cuja variação de expressão seriam respostas aos estresses, como observado em *CG15043*, *RhoGEF64C* e *GD18191*;
- 4) Casos potenciais de genes cuja variação de expressão seriam respostas adaptativas decorrentes da colonização da zona temperada, como observado em *CG40498*, *GD10984*, *GD14844*, *GD18191* e *GD24386*;
- 5) Que mecanismos epigenéticos e outros fatores que constituem o *background* genético das espécies definem a disponibilidade de cópias de TEs e a produção de transcritos quiméricos.

REFERÊNCIAS

- AQUADRO, C.F.; LADO, K.M.; NOON, W.A. The rosy region of *Drosophila melanogaster* and *Drosophila simulans*. I. Contrasting levels of naturally occurring DNA restriction map variation and divergence. **Genetics**. v. 119, n. 4, p. 875–888, 1998.
- ANDERS, S.; PYL, P. T.; HUBER, W. HTSeq--a Python framework to work with high-throughput sequencing data. **Bioinformatics**, v. 31, n. 2, p.166–169, 2015
- AUVINET, J.; GRAÇA, P.; GHIGLIOTTI, L.; PISANO, E.; et al., Insertion Hot Spots of DIRS1 Retrotransposon and Chromosomal Diversifications among the Antarctic *Teleosts* *Nototheniidae*. **International Journal of Molecular Sciences**. v.20, p. 701, 2019.
- AWAD, M.; GAN, X. GALA: Gap-free chromosome-scale assembly with long reads. **bioRxiv**. p. 1– 13, 2020.
- AYALA, F.J; FITCH, W.M. Genetics and the origin of species: an introduction. **Proceedings of the National Academy of Sciences of the United States of America**. v. 94, n. 15, p. 7691-7697, 1997.
- AYRES, M.; AYRES JUNIOR, M.; AYRES, D.L.; SANTOS, A.A.S. BioEstat 5.3: aplicações estatísticas nas áreas das Ciências Biomédicas. Sociedade Civil Mamirauá: Belém, Pará-Brasil. 2007. 324p. Disponível em <https://www.mamiraua.org.br/downloads/programas/> Acesso em 20 set 2022.
- BAILLY-BECHET, M.; HAUDRY, A.; LERAT, E. “One code to find them all”: a perl tool to conveniently parse RepeatMasker output files. **Mobile DNA**, v. 5, n. 13.2014.
- BALZAC, H. A pele de Onagro. 1831. Porto Alegre – RS: L & PM, 2018. DUARTE, N.
- BANNAN, B.A.; VAN ETTEN, J.; KOHLER, J.A, et al. The *Drosophila* protein palmitoylome: characterizing palmitoyl-thioesterases and DHHC palmitoyl-transferases. **Fly (Austin)**. v. 2, n. 4, p. 198-214, 2008.
- BATUR, P.; GINGERAS, T.R. RAMPAGE: promoter activity profiling by paired-end sequencing of 5'-complete cDNAs. **Current Protocols in Molecular Biology**. v. 104, p. 25, 2013.
- BENSON, G. Tandem repeats finder: a program to analyze DNA sequences. **Nucleic Acids Research**, v. 27, p. 573–580, 1999.
- BERGLAND, A. O.; BEHRMAN, E. L.; O'BRIEN, K. R.; SCHMIDT, P. S.; PETROV, D. A. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. **PLoS Genetics**. v. 10, n. 11, 2014.
- BERGMAN, C.M.; BENSASSON, D. Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in *Drosophila melanogaster*. **Proceedings of the National Academy of Sciences of the United States of America** v. 104, n. 27, p. 11340–11345, 2007

BERGMAN, C.M.; QUESNEVILLE, H.; ANXOLABEHERE, D.; ASHBURNER, M. Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. **Genome Biology**, v.7, n. 112, 2006.

BIEMONT, C. A brief history of the status of transposable elements: from junk DNA to major players in evolution. **Genetics**, v. 186, n. 4, p. 1085–1093, 2010.

BIEMONT, C.; VIEIRA, C.; BORIE, N.; LEPETIT, D. Transposable elements and genome evolution: the case of *Drosophila simulans*. **Genetica**, v. 107, p. 113–120, 1999.

BIEMONT, C.; VIEIRA, C. Genetics—Junk DNA as an evolutionary force. **Nature**, v. 443, p. 521–524, 2006.

BOWEN, N.J.; McDONALD, J.F. *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. **Genome Research**, v. 11, n. 9, p. 1527–1540, 2001.

BRENNECKE, J.; MALONE, C.D.; ARAVIN, A.A.; SACHIDANANDAM, R.; STARK, A.; HANNON, G.J. An epigenetic role for maternally inherited piRNAs in transposon silencing. **Science**, v. 322, n. 5906, p. 1387–1392, 2008.

BROSIUS, J. Exaptation at the molecular genetic level. **Science. China Life Science**, v. 62, n. 4, p. 437–452, 2019.

BUTELLI, E. et al. Retrotransposons control fruit specific, cold-dependent accumulation of anthocyanins in blood oranges. **Plant Cell**, v. 24, p. 1242–1255, 2012.

CAPY, P.; GASPERI, G.; BIEMONT, C.; BAZIN, C. Stress and transposable elements: co-evolution or useful parasites? **Heredity (Edinb)**, v. 85, p.101–106, 2000.

CAPY, P.; GIBERT, P. *Drosophila melanogaster*, *Drosophila simulans*: so similar yet so different. **Genetics**, v. 120, n. 1–3, p. 5–16, 2004.

CAPY, P. Taming, Domestication and Exaptation: Trajectories of Transposable Elements in Genomes. **Cells**, v. 10, n. 12, p. 3590, 2021.

CHAKRABORTY, M.; CHANG, C. H.; KHOST, D. E.; VEDANAYAGAM, J.; et al. Evolution of genome structure in the *Drosophila simulans* species complex. **Genome Research**, v. 31, n. 3, p. 380–396, 2021.

CHUNG, H. R.; SCHAFFER, U.; JACLE, H.; BOHM, S. Genomic expansion and clustering of ZAD-containing C2H2 zinc-finger genes in *Drosophila*. **EMBO reports**, v. 3, n. 12, p. 1158–1162, 2002

CHUONG, E.B.; ELDE, N.C.; FESCHOTTE, C. Regulatory activities of transposable elements: from conflicts to benefits. **Nature Reviews Genetics**, v. 18, n. 2, p. 71–86, 2017.

CLARK, A. G.; EISEN, M. B.; SMITH, D. R.; BERGMAN, C. M, et al. *Drosophila* 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. **Nature**, v. 450, n. 7167, p. 203–218, 2007.

DANIELS, S.B.; PETERSON, K.R.; STRAUSBAUGH, L.D.; KIDWELL, M.G.; CHOYNICK, A. Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics*, v. 124, p. 339–355, 1990.

DEAN, M. D.; BALLARD, J. W. O. Linking phylogenetics with population genetics to reconstruct the geographic origin of a species. **Molecular Phylogenetics and Evolution**, v. 32, n. 3, p. 998–1009, 2004.

DEJIMA, K.; KATSUFUMI, A.; TAKEMURA, M.; CHOI, P.; et al. “The role of *Drosophila* heparan sulfate 6-O-endosulfatase in sulfation compensation.” **The Journal of biological chemistry**, v. 288, n. 9, p. 6574–82, 2013.

DÍAZ-GONZALEZ, J.; DOMINGUEZ, A.; ALBORNOZ, .F. Genomic distribution of retrotransposons 297, 1731, copia, mdg1 and roo in the *Drosophila melanogaster* species subgroup. **Genetica**, v. 138, p. 579–586, 2010.

DÍAZ-GONZALEZ, J.; VAZQUEZ, J.; ALBORNOZ, .F.; DOMINGUEZ, A. Long-term evolution of the roo transposable element copy number in mutation accumulation lines of *Drosophila melanogaster*. **Genetics Research.**, v. 93, p. 181–187, 2011.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* (Oxford, England), 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>

ENGELS, W.R., The origin of P elements in *Drosophila melanogaster*. **BioEssays**, v. 14, p. 681–686, 1992.

ENGEL, W.R.; PRESTON, C.R. Hybrid dysgenesis in *Drosophila melanogaster*: The biology of female and male sterility. **Genetics**, v. 92, p. 161–174, 1979.

FABLET, M.; SALCEZ-ORTIZ, J.; JACQUET, A.; MENEZES, B.F.; et al. A quantitative, genome-wide analysis in *Drosophila* reveals transposable elements’ influence on gene expression is species-specific. **bioRxiv**. 2022.

FARRÉ, D., ENGEL, P.; ANGULO, A. Novel role of 3’UTR-embedded alu elements as facilitators of processed pseudogene genesis and host gene capture by viral genomes. **In PLoS ONE**, v. 11, n. 12, 2016.

FAULKNER, G. J; KIMURA, Y; et al. The regulated retrotransposon transcriptome of mammalian cells. **Nature Genetics**, v. 41, n. 5, p. 563–571, 2009.

FESCHOTTE, C. Transposable elements and the evolution of regulatory networks. **Nature Reviews Genetics**, v. 9, n. 5, p. 397–405, 2008.

FLUTRE, T.; DUPRAT, E.; FEUILLET, C.; QUESNEVILLE, H. Considering Transposable Element Diversification in De Novo Annotation Approaches. **PLoS ONE**, v. 6, 2011.

FONTANILLAS, P.; HARTL, D. L.; REUTER, M. Genome organization and gene expression shape the transposable element distribution in the *Drosophila melanogaster* euchromatin. **PLoS Genetics** v. 3, n. 11, p. 2256–2267, 2007.

GARCÍA GUERREIRO, M.P.; CHAVEZ-SANDOVAL, B.E.; BALANYA, J.; SERRA, L.; FONTDEVILA, A. Distribution of the transposable elements bilbo and gypsy in original and colonizing populations of *Drosophila subobscura*. **BMC Evolutionary Biology**, v. 8, p. 234, 2008

GARCÍA GUERREIRO, M.P.; FONTDEVILA, A. Chromosomal distribution of the transposable elements *Oswaldo* and *blanco* in original and colonizer populations of *Drosophila buzzatii*. **Genetics Research**, v. 77, p. 227–238, 2001.

GARCÍA GUERREIRO, M.P.; FONTDEVILA, A. *Oswaldo* and *Isis* retrotransposons as markers of the *Drosophila buzzatii* colonization in Australia. **BMC Evolutionary Biology**, v. 11, p. 111, 2011.

GARCÍA GUERREIRO, M.P. What makes transposable elements move in the *Drosophila* genome? **Heredity** v. 108, p. 461–468, 2012.

GONZÁLEZ, J.; KARASOV, T. L.; MESSER, P. W.; PETROV, D. A. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. **PLoS Genetics**, v. 6, n. 4, p. 33–35, 2010.

GONZÁLEZ, J.; LENKOV, K.; LIPATOV, M.; MACPHERSON, J. M.; PETROV, D. A. High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. **PLoS Biology**, v. 6, n. 10, p. 2109–2129, 2008.

GOUBERT, C.; MODOLO, L.; VIEIRA, C.; MORO, C. V.; et al, M. De novo assembly and annotation of the Asian tiger mosquito (*Aedes Albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). **Genome Biology and Evolution**. v. 7, n. 4, p. 1192–1205, 2015.

GRANDBASTIEN, M.A; AUDEON, C.; BONNIVARD, E.; CASACUBERTA, J.M.; et al. A-PP et al. Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenetic and Genome Research*, v. 110, n.1-4, p. 229–241, 2005.

HIRAIZUMI, Y.; SLATKO, B.; LANGLEY, C.; NILL, A. Recombination in *Drosophila melanogaster* male. **Genetics**, v. 73, p. 439–444, 1973.

HIRAIZUMI, Y. Spontaneous recombination in *Drosophila melanogaster* males. **Proceedings of the National Academy of Sciences of the United States of America**, v. 68, p. 268–270, 1971.

HORVÁTH, V.; MERENCIANO, M.; GONZÁLEZ, J. Revisiting the relationship between transposable elements and the eukaryotic stress response. **Trends Genetics**. v. 33, n. 11, p. 832–841, 2017.

HYTTIA, P.; CAPY, P.; DAVID, J.R., Singh, R.S. Enzymatic and quantitative variation in European and African populations of *Drosophila simulans*. **Heredity**, v. 54, p. 209–217, 1985.

IWASAKI, Y. W.; SIOMI, M. C.; SIOMI, H. PIWI-interacting RNA: Its biogenesis and functions. **Annual Review of Biochemistry**. v. 84, p. 405–433, 2015.

KABELITZ, T.; BAURLE, I. Get the jump - Do 3'UTRs protect transposable elements from silencing? **Mob Genet Elements**. v. 5, n. 4, p. 51-54, 2015.

KAMINKER, J.S.; BERGMAN, C.M.; KRONMILLER, B.; CARLSON, J.; SVIRSKAS, R.; et al. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. **Genome Biology**. v.3, n. 84, 2002.

KANG, L.; RASHKOVETSKY, E.; MICHALAK, K., et al. Genomic divergence and adaptive convergence in *Drosophila simulans* from Evolution Canyon, Israel. **Proceedings of the National Academy of Sciences of the United States of America**, v. 116, n. 24, p. 11839-11844, 2019.

KAPITNOV, V. V.; JURKA, J. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, n. 11, p. 6569–6574, 2003.

KAPITONOV, V.V.; JURKA, J. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, p. 6569-6574, 2003.

KAVI, H. H.; FERNANDEZ, H. R.; XIE, W.; BIRCHLER, J. A. RNA silencing in *Drosophila*. **FEBS Letters**. v. 579, n. 26, p. 5940–5949, 2005.

KELLEHER, E.S.; BARBASHG, D.A. Analysis of piRNA-mediated silencing of active TEs in *Drosophila melanogaster* suggests limits on the evolution of host genome defense. **Molecular Biology and Evolution**.v. 30, n. 8, p. 1816–1829, 2013

KIDWELL, M.G; KIDWELL, J.F.; DVED, J.A. Hybrid dysgenesis in *Drosophila melanogaster*: A syndrome of aberrant including mutations, sterility and male recombination. **Genetics**, v. 86, p. 813–833, 1977.

KIDWELL, M.G. Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 80, n. 6, p. 1655-1659, 1983.

KLEINSCHMIT, A; TAKEMURA, M.; DEJIMA, K.; CHOI, P. Y.; NAKATO, H. *Drosophila* heparan sulfate 6-O-endosulfatase Sulf1 facilitates wingless (Wg) protein degradation. *The Journal of biological chemistry*, v. 288, n. 7, p. 5081-9, 2013.

KOFLER, R.; BETANCOURT, A. J.; SCHLÖTTERER, C. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. **PLoS Genetics**. v. 8, n. 1, 2012.

KOFLER, R.; HILL, T.; NOLTE, V.; BETANCOURT, A.J, SCHLOTTERER, C. The recent invasion of natural *Drosophila simulans* populations by the P-element. **Proceedings of the National Academy of Sciences of the United States of America**, v. 112, n. 21, p .6659-6663, 2015.

KOFLER, R.; NOLTE, V.; SCHLOTTERER, C. Tempo and mode of transposable element activity in *Drosophila*. **PLoS Genet**, v. 11, n. 1005406, 2015.

KOLACZKOWSKI, B.; KERN, A. D.; HOLLOWAY, A. K.; BEGUN, D. J. Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. **Genetics**, v. 187, n. 1, p. 245–260, 2011.

LABRADOR, M.; SELEME, M.C.; FONTDEVILA, A. The evolutionary history of *Drosophila buzzatii*. XXXIV. The distribution of the retrotransposon *Oswaldo* in original and colonizing populations. **Molecular Biology and Evolution** v. 15, p. 1532–1547, 1998.

LACHAISE, D.; CARIOU, M.-L.; DAVID, J. R.; LEMEUNIER, F.; TSACAS, L.; ASHBURNER, M. Historical biogeography of the *Drosophila melanogaster* species subgroup. **Evolutionary Biology**, p. 159–225, 1988.

LAMPRECHT, B.; WALTER, K.; et al. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. **Nature Medicine**, v. 16, n. 5, p. 571–579, 2010.

LANDER, E.S.; LINTON, L.M.; BIRREN, B. et al. Initial sequencing and analysis of the human genome. **Nature**, v. 411, n. 6838, p. 720, 2001.

LEE, Y.C. The role of piRNA-Mediated epigenetic silencing in the Population Dynamics of transposable elements in *Drosophila Melanogaster*. **PLoS Genetics** v. 11, n. 6, e1005269, 2015.

LEE, Y. C. G.; KARPEN G. H. Pervasive epigenetic effects of *Drosophila* euchromatic transposable elements impact their evolution. **eLife** v. 6, n. 25762, 2017.

LERAT, E.; BURLET, N.; BIEMONT, C.; VIEIRA, C. Comparative analysis of transposable elements in the *melanogaster* subgroup sequenced genomes. **Gene**, v. 473, n. 2, p. 100–109, 2011.

LERAT, E.; FABLET, M.; MODOLO, L.; LOPEZ-MAESTRE, H.; VIEIRA, C. TEtools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. **Nucleic Acids Research**, v. 45, n. 4, p. e17, 2017.

LIPATOV, M.; LENKOV, K.; PETROV, D.A.; BERGMAB, C.M. Paucity of chimeric gene-transposable element transcripts in the *Drosophila melanogaster* genome. **BMC Biology**. v. 3, n. 24, 2005.

LOVE, M.I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, p. 550, 2014.

MALIK, H.S.; HENIKIFF, S. Positive selection of Iris, a retroviral envelope-Derived host gene in *Drosophila melanogaster*. **PLoS Genetics**, v. 1, n. 44, 2005.

MALONE, C. D.; BRENNECKE, J.; DUS, M.; STARK, A.; et al. Specialized piRNA Pathways Act in Germline and Somatic Tissues of the *Drosophila* Ovary. **Cell**. v. 137, n. 3, p. 522–535, 2009.

MATEO, L.; ULLASTRES, A.; GONZÁLEZ, J. A transposable element insertion confers xenobiotic resistance in *Drosophila*. **PLoS Genet**, v. 10, n. 8, e1004560, 2014.

MAZZUCHELLI, J.; MARTINS, C. Genomic organization of repetitive DNAs in the cichlid fish *Astronotus ocellatus*. **Genetica**. v. 136, p. 461–469, 2009.

McGINNIS, W.; SHERMOEN, A. W.; BECKENDORF, S. K. A transposable element inserted just 5' to a *Drosophila* glue protein gene alters gene expression and chromatin structure. **Cell**, v. 34, n. 1, p. 75–84, 1983.

MENEZES, B.F.; SALCES-ORTIZ, J.; MULLER, H.; BURLET, N. et al. An attempt to select non-genetic variation in resistance to starvation and reduced chill coma recovery time in *Drosophila melanogaster*. **The Journal of Experimental Biology**. v. 23, p. 221, 2018.

MÉREL, V.; BOULESTEIX, M.; FABLET, M. et al. Transposable elements in *Drosophila*. **Mobile DNA**, v. 11, n. 23, 2020.

MERENCIANO, M.; ULLASTRES, A.; de CARA, M. A.; BARRÓN, M. G.; GONZÁLEZ, J. Multiple Independent Retroelement Insertions in the Promoter of a Stress Response Gene Have Variable Molecular and Functional Effects in *Drosophila*. **PLOS Genetics**, v. 12, n. 8, e1006249, 2016.

MOHAMED, M.; DANG, N. T. M.; OGYAMA, Y; BURLET, N., et al. A transposon story: from te content to te dynamic invasion of *Drosophila* genomes using the single-molecule sequencing technology from oxford nanopore. **Cells**, v. 9, n. 8, p. 1–23, 2020.

MODOLO, L.; LERAT, E. UrQt: an efficient software for the Unsupervised Quality trimming of NGS data. **BMC Bioinformatics**. v. 16, p. 137, 2015.

MOMBACH, D. M; DA FONTOURA GOMES, T. M. F.; LORETO, E. L. S. Stress does not induce a general transcription of transposable elements in *Drosophila*. **Molecular Biology Reports**, v. 49, n. 9, p. 9033–9040, 2022.

NASCIMENTO, A.M.L.; SILVA, B.S.M.L.; SVARTMAN, M.; KUHN, G.C.S. Tracking a recent horizontal transfer event: The P-element reaches Brazilian populations of *Drosophila simulans*. **Genetics and Molecular Biology**, v. 43, n. 2, e20190342, 2020.

NEEFEDOVA, L.N.; KUZMIN, I.V.; MAKHNOVSKII, P.A.; KIM, A.I. Domesticated retroviral GAG gene in *Drosophila*: new functions for an old gene. **Virology**, p. 450–451, p. 196–204, 2014.

NEWELL-PRICE, J.; CLARCK, A. J. L.; KING, P. DNA methylation and silencing of gene expression. **Trends in Endocrinology and Metabolism: TEM**, v. 11, n. 4, p. 142–148, 2000.

NICOLAU, M.; PICAULT, N.; MOISSIARD, G. The Evolutionary Volte-Face of Transposable Elements: From Harmful Jumping Genes to Major Drivers of Genetic Innovation. **Cells**, v. 10, n. 11, p. 2952, 2021.

OLIVEIRA, D. S; FABLET, M.; VALLIER, A.N.A; CARARETO, M.A; REBOLLO, R.; VIEIRA, C. ChimeraTE: A pipeline to detect chimeric transcripts derived from genes and transposable elements. **bioRxiv**, 2023.

PARDUE, M.L.; DeBARYSHE, P.G. *Drosophila* telomeres: Two transposable elements with important roles in chromosomes. **Genetica**. v. 107, p. 189–196, 1999.

PARSONS, P.A. Evolutionary rates under environmental stress. *Evol. Biol.* v. 21, p. 311–347, 1987.

PARSONS, P. A. Resistance to cold temperature stress in populations of *Drosophila melanogaster* and *D. simulans*. **Australian Journal of Zoology**, v. 25, p. 693–698, 1977.

PHILLIPS, P. K.; HEATH, J. E. Heat exchange by the pinna of the african elephant (*Loxodonta africana*). **Comparative Biochemistry and Physiology**, v. 101, n. 4, p. 693–699, 1992.

PIMPINELLI, S.; BERLOCO, M.; FANTI, L.; DIMITRI, P.; et al. Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. **Proceedings of the National Academy of Sciences of the United States of America**. v. 92, p. 3804–3808, 1995.

PIROOZANIA, S. K.; YUAN, C.; KHAN, M. R.; et al. PARIS induced defects in mitochondrial biogenesis drive dopamine neuron loss under conditions of parkin or PINK1 deficiency. **Molecular Neurodegeneration**, v. 15, n. 1, p.17, 2020.

QUESNEVILLE, H.; BERGAMAN, C. M.; ANDRIEU, O.; AUTARD, D.; et al. Combined evidence annotation of transposable elements in genome sequences. **PLoS Computational Biology**, v.1, n. 22, 2005.

QUINLAN, A. R.; HALL, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. **Bioinformatics**, v. 26, p. 841–842, 2010.

RAHMAN, R.; CHIRN, G.; KANODIA, A.; SYTNIKOVA, Y. A.; et al. Unique transposon landscapes are pervasive across *Drosophila melanogaster* genomes. **Nucleic Acids Res.**, 43, 10655–10672, 2015.

R CORE TEAM (2023). *_R: A Language and Environment for Statistical Computing_*. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.

REBOLLO, R.; ROMANISH, M. T.; MAGER, D.L. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annual Review of Genetics*. v. 46, p. 21-42, 2012.

REINHARDT, J. A.; KOLACZKOWSKI, B.; JONES, C. D.; BEGUN, D. J.; KERN, A. D. Parallel geographic variation in *Drosophila melanogaster*. **Genetics**. v. 197, n. 1, p. 361–373, 2014.

RIEDL, C. A.; NEAL, S. J.; ROBICHON, A.; WESTWOOD, J. T.; SOKOLOWSKI, M. B. *Drosophila* soluble guanylyl cyclase mutants exhibit increased foraging locomotion: behavioral and genomic investigations. **Behavior genetics**, v. 35, n. 3, p. 231–244, 2005.

RISHISHWAR, L.; TELLEZ VILLA, C. E.; JORDAN, I. K. Transposable element polymorphisms recapitulate human evolution. **Mobile DNA**, v. 6, n. 21, 2015.

SANSEVERINO, W.; HÉNAFF, E.; VIVES, C.; PINOSIO, S.; et al. Transposon Insertions, Structural Variations, and SNPs Contribute to the Evolution of the Melon Genome. **Molecular Biology and Evolution**, v. 32, n. 10, p. 2760–2774, 2015.

SAINT-LEANDRE, B.; CAPY, P.; HUA-VAN, A.; FILÉE, J. piRNA and Transposon Dynamics in *Drosophila*: A Female Story. **Genome Biology and Evolution**, v. 12, n. 6, p. 931-947, 2020.

SCHNABLE, P. S. et al. The B73 maize genome: complexity, diversity, and dynamics. **Science**. v. 326, n. 5956, p. 1112–1115, 2009.

SCHRADER, L.; SCHMITZ, J. The impact of transposable elements in adaptive evolution. **Molecular Ecology**, v. 28, n. 6, p. 1537–1549, 2019.

SEDGHIFAR, A.; SAELAO, P.; BEGUN, D.J. Genomic patterns of geographic differentiation in *Drosophila simulans*. **Genetics**, v. 202, n. 3, p. 1229–1240, 2016.

SHAPIRO-KULNAME, L.; BAUTISTA, O.; SALZ, H. K. An RNA-interference screen in *Drosophila* to identify ZAD-containing C2H2 zinc finger genes that function in female germ cells. **G3 (Bethesda, Md.)**, v. 11, n. 1, 2021.

SHAPIRO, S.S.; WILK, M.B. An Analysis of Variance Test for Normality (Complete Samples). **Biometrika**, v. 52, p. 591-611, 1965.

SHUMATE, A; SALZBERG, S.L. Liftoff: accurate mapping of gene annotations. **Bioinformatics**, v. 37, p. 1639–1643, 2021.

SINGH, R.S.; CHOUDHARY, M.; DAVID, J.R. Contrasting patterns of geographic variation in the cosmopolitan sibling species *Drosophila melanogaster* and *Drosophila simulans*. **Biochemistry Genetics**. v. 25, p. 27–40, 1987.

SINZELLE, L., IZSVÁK, Z. & IVICS, Z. Molecular domestication of transposable elements: From detrimental parasites to useful host genes. **Cellular Molecular Life Sciences**., v. 66, n. 6, p. 1073–1093, 2009.

SLOTKIN, R. K.; MARTIENSSEN, R. Transposable elements and the epigenetic regulation of the genome. **Nature Reviews Genetics**. v. 8, n. 4, p. 272–285, 2007.

SMIT, A.F. RepeatMasker Open 3.0. 2004.

SONG, J.; LIU, J.; SCHNAKENBERG, S.L.; HA H.; et al. Variation in piRNA and transposable element content in strains of *Drosophila melanogaster*. **Genome Biol Evol**. v. 6, n. 10, p. 2786–2798, 2014.

SPRENGELMEYER, Q. D.; MANSOURIAN, S.; LANGE, J. D.; MATUTE, D. R.; et al. A recurrent collection of *drosophila melanogaster* from wild african environments and genomic insights into species history. **Molecular Biology and Evolution**. v. 37, n. 3, p. 627–638, 2020.

STUDER, A.; ZHAO, Q.; ROSS-IBARRA, J.; DOEBLEY, J. Identification of a functional transposon insertion in the maize domestication gene *tb1*. **Nature Genetics**. v. 43, p. 1160–1163, 2011.

SUHRE, K.; MCCARTHY, M.I.; SCHWENK, J.M. Genetics meets proteomics: perspectives for large population-based studies. **Nature Reviews Genetics**.v. 22, n. 1, p. 19-37, 2021.

SUTHERLAND, J. E.; COSTA, M. Epigenetics and the environment. **Annals of the New York Academy of Sciences**, v. 983, n. 1, p. 151-160, 2003.

SUNDARAM, V.; WYSOCKA, J. Transposable elements as a potent source of diverse cis-regulatory sequences in *mammalian* genomes. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 375, n. 1795, 2020.

THORVALDSDOTTIR, H.; ROBNSON, J.T.; MESIROV, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. **Briefings in Bioinformatics**, v. 14, p. 178–192, 2013.

TREIBER, C. D.; WADDELL, L. S. Transposon expression in the *Drosophila* brain is driven by neighboring genes and diversifies the neural transcriptome. **Genome Res.** v. 30, p. 1559–1569, 2020.

ULLASTRES, A.; MERENCIANO, M.; GONZÁLEZ, J. Regulatory regions in natural transposable element insertions drive interindividual differences in response to immune challenges in *Drosophila*. **Genome Biology**, v. 22, n.265, 2021.

VAGIN, V. V; KLENOV, M. S.; STOLYARENKO, A. D.; KOTELNIKOV, R. N. The rna interference proteins and vasa locus are involved in the. **RNA Biology**. v. 1, n. 1, p. 54–58, 2004.

VAN'T HOF, A. E. et al. The industrial melanism mutation in British peppered moths is a transposable element. **Nature**, v. 534, p. 102–105, 2016.

VIEIRA, C.; BIÉMONT, C. Transposable element dynamics in two sibling species: *Drosophila melanogaster* and *Drosophila simulans*. **Genetica**. v. 120, n. 1-3, p. 115-123, 2004.

VIEIRA, C.; NARDON, C.; ARPIN, C.; LEPETIT, D.; BIÉMONT, C. Evolution of genome size in *Drosophila*. Is the invader's genome being invaded by transposable elements? **Molecular Biology and Evolution** v. 19, p. 1154–1161, 2002.

VOLFF, J.N. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. **Bioessays**, v. 28, p. 913–922, 2006.

WESSLER, S.R. Turned on by stress. Plant retrotransposons. **Current Biology**. v. 6, p. 959–61, 1996.

WICKER, T.; SABOT, F.; HUA-VAN, A.; BENNETZEN, J. L.; et al. A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics**. v. 8, p. 973–982, 2007.

WISOTZKEY, R.G.; FELGER, I.; HUNT, J.A. Biogeographic analysis of the UHU and LOA elements in the Hawaiian *Drosophila*. **Chromosoma** 106: 465–477, 1997.

YASUHARA, J.C; WAKIMOTO, B.T. Molecular landscape of modified histones in *Drosophila* heterochromatic genes and euchromatin-heterochromatin transition zones. **PLoS Genetics**., v. 4, n. 16, 2008.

YOSHITAKE, Y.; INOMATA, N.; SANO, M.; KATO, Y.; ITOH, M. The P element invaded rapidly and caused hybrid dysgenesis in natural populations of *Drosophila simulans* in Japan. **Ecology and Evolution**. v. 19, p. 9590–9599, 2018.

YU, J.; WU, H.; WEN, Y.; LIU, Y.; et al. Identification of seven genes essential for male fertility through a genome-wide association study of non-obstructive azoospermia and RNA interference-mediated large-scale functional screening in *Drosophila*. **Human Molecular Genetics**, v. 24, n. 5, p. 1493–1503, 2015.

ZHAO, L.; WIT, J.; SVETEC, N.; BEGUN, D. J. Parallel gene expression differences between low and high latitude populations of *Drosophila melanogaster* and *D. simulans*. **PLoS Genetics**. v. 11, n. 5, p. 1–25, 2015.

ZOLOTAREV, N. A.; MAKSIMENKO, O. G.; GEORGIEV, P. G.; BONCHUK, A. N.. ZAD-Domain Is Essential for Nuclear Localization of Insulator Proteins in *Drosophila melanogaster*. **Acta natural**, v. 8, n. 3, p. 97–102, 2016.

ANEXOS

Em decorrência do número e tamanho as Figuras (arquivos pdf) e Tabelas Suplementares (arquivo EXCEL) estão disponíveis sob solicitação.

ANEXOS A – Figuras Suplementares

Figuras Suplementares 1 e 2. Famílias de TEs mais frequentes nas classes e subclasses de transcritos das linhagens de *D. melanogaster*, após a retirada das duas famílias mais frequentes em transcritos quiméricos (INE-1 e roo).

Figuras Suplementares 3 e 4. Famílias de TEs mais frequentes nas classes e subclasses de transcritos das linhagens de *D. simulans*, após a retirada das duas famílias mais frequentes em transcritos quiméricos.

Figuras Suplementares 5 a 9. Correlação de Pearson entre o conteúdo genômico das famílias de TEs e a frequência das famílias que mais formaram quimeras em *D. melanogaster*.

Figuras Suplementares 10 a 14. Correlação de Pearson entre o conteúdo genômico das famílias de TEs e a frequência das famílias que mais formaram quimeras em *D. simulans*.

Figuras Suplementares 15 a 34. Ilustração da expressão diferencial de genes candidatos com inserções de TE entre os tratamentos (Controle, *Cold-Shock* e *Starvation*) no genoma das quatro linhagens de *D. melanogaster*.

Figuras Suplementares 35 a 56. Ilustração da expressão diferencial de genes candidatos com inserções de TE entre os tratamentos (Controle, *Cold-Shock* e *Starvation*) no genoma das quatro linhagens de *D. simulans*.

Figuras Suplementares 57 a 65. Ilustração da expressão diferencial de genes candidatos com inserções de TE entre as linhagens das populações brasileiras e francesas de *D. melanogaster*.

Figuras Suplementares 66 a 72. Ilustração da expressão diferencial de genes candidatos com inserções de TE entre as linhagens das populações brasileiras e francesas de *D. simulans*.

ANEXOS B – Tabelas Suplementares

Tabelas Suplementares 1 e 2. Totais de transcritos quiméricos nas linhagens de *D. melanogaster* e de *D. simulans*, respectivamente.

Tabelas Suplementares 12 e 13. Dados de Expressão Diferencial dos genes candidatos entre os tratamentos (Controle, *Cold-Shock* e *Starvation*) em *D. melanogaster* e em *D. simulans*, respectivamente.

Tabelas Suplementares 15 e 16. Dados de Expressão Diferencial dos genes candidatos entre as linhagens das populações brasileiras e francesas em *D. melanogaster* e em *D. simulans*, respectivamente.

ANEXOS C – Tabelas Suplementares

Tabela Suplementar 3. Transcritos quiméricos comuns entre as linhagens de *D. melanogaster* - nomes e funções assinaladas a eles a) entre Dmgoth101 e Dmgoth63; b) entre Dmsj23 e Dmsj7; c) entre as quatro linhagens.

Tabela Suplementar 4. Transcritos quiméricos comuns entre as linhagens de *D. simulans* - nomes e funções assinaladas a eles a) entre Dsgoth31 e Dsgoth61; b) entre Dssj27 e Dssj9; c) entre as quatro linhagens.

Tabelas Suplementares 5 e 6. Identificação de genes que formam quimeras entre as linhagens - nomes e funções assinaladas a eles. de *D. melanogaster* e *D. simulans*, respectivamente.

Tabelas Suplementares 7 e 8. Distribuição do número de transcritos quiméricos encontrados entre as classes, superfamílias e famílias de TE dispostos entre as linhagens e condições experimentais analisadas em *D. melanogaster* e *D. simulans*, respectivamente, adicionalmente entre as classes de transcritos definidas.

Tabelas Suplementares 9 e 10. Distribuição de genes quiméricos, total de transcritos quiméricos e os exclusivos entre as linhagens de *D. melanogaster* e *D. simulans*, respectivamente, dos quais puderam ser identificadas inserções exclusivas potencialmente adaptativas.

Tabela Suplementar 11. Expressão normalizada dos genes candidatos entre controle e tratado das réplicas das linhagens de a) *D. melanogaster* e b) *D. simulans*.

Tabela Suplementar 14. Expressão normalizada dos genes candidatos interpopulacionais entre as réplicas das linhagens de a) *D. melanogaster* e b) *D. simulans*.