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

**NATHALIA DE SETTA**

# ELEMENTOS DE TRANSPOSIÇÃO NO GÊNERO *ZAPRIONUS* (DIPTERA, DROSOPHILIDAE): ESTUDOS GENÔMICOS E EVOLUTIVOS COM ÊNFASE NOS RETROTRANSPOSONS *COPIA*, *GYPSY* E *MICROPIA*

Tese apresentada para obtenção do Título de Doutor em Genética

São José do Rio Preto – SP
2009

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

**NATHALIA DE SETTA**

# ELEMENTOS DE TRANSPOSIÇÃO NO GÊNERO *ZAPRIONUS* (DIPTERA, DROSOPHILIDAE): ESTUDOS GENÔMICOS E EVOLUTIVOS COM ÊNFASE NOS RETROTRANSPOSONS *COPIA*, *GYPSY* E *MICROPIA*

Tese apresentada para obtenção do Título de Doutor em Genética

Orientadora: Profa. Dra. Claudia Marcia Aparecida Carareto

São José do Rio Preto – SP
2009

Setta, Nathalia de.
Elementos de transposição no gênero *Zaprionus* (Diptera, Drosophilidae): estudos genômicos e evolutivos com ênfase nos retrotransposons *copia*, *gypsy* e *micropia* / Nathalia de Setta - São José do Rio Preto : [s.n.], 2009.

179 f. : il. ; 30 cm.

Orientador: Claudia Marcia Aparecida Carareto
Co-orientador: Marie Anne Van Sluys
Tese (doutorado) – Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas

1. Genética. 2. Evolução Molecular – Elementos transponíveis. 3. Drosophila. 4. Zaprionus. I. Carareto, Claudia Marcia Aparecida. II. Sluys, Marie Anne Van. III. Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas. IV. Título.

CDU - 575

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

**NATHALIA DE SETTA**

**Elementos de transposição no gênero *Zaprionus* (Diptera, Drosophilidae): estudos genômicos e evolutivos com ênfase nos retrotransposons *copia*, *gypsy* e *micropia***

COMISSÃO JULGADORA

TESE PARA OBTENÇÃO DO GRAU DE DOUTOR

Presidente e Orientador: .....
2º Examinador: .....
3º Examinador: .....
4° Examinador:.....
5° Examinador:.....

São José do Rio Preto, ____/____/_____.

O presente trabalho foi realizado no Departamento de Biologia do Instituto de Biociências, Letras e Ciências Exatas de São José do Rio Preto, da Universidade Estadual Paulista, sob orientação da Profa. Dra. Claudia Marcia Aparecida Carareto, com bolsa de Doutorado no País do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e bolsa de Estágio de Doutoramento no Exterior da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

*Dedico este estudo à minha mãe Maria Elisa; pelo esforço, dedicação e compreensão, em todos os momentos desta e de outras caminhadas; ao meu companheiro Marcus, pelo amor e por entender minha ausência durante a realização deste trabalho; à minha orientadora Profa. Claudia Carareto, por me ouvir e ensinar outros caminhos.*
*Obrigada!!!*

***DURANTE ESTE TRABALHO...***

*As dificuldades não foram poucas...*
*Os desafios foram muitos...*
*Os obstáculos, muitas vezes, pareciam intransponíveis.*
*O desânimo quis nos contagiar, porém, a garra e a tenacidade foram mais fortes, sobrepondo esse sentimento, fazendo seguir a caminhada, apesar da sinuosidade do caminho.*
*Agora, ao olhar para trás, a sensação do dever cumprido se faz presente e podemos constatar que as noites de sono perdidas, as viagens e visitas realizadas, os longos tempos de leitura, a discussão, a ansiedade em querer fazer e a angústia de muitas vezes não o conseguir, não foram em vão.*
*Aqui estamos, como sobreviventes de uma longa batalha, porém, muito mais fortes e hábeis, com coragem suficiente para mudar a nossa postura, apesar de todos os percalços...*
*Como dizia Antoine de Saint-Exupéry em sua obra prima "O Pequeno Príncipe":*

***"Foi o tempo que dedicaste à tua rosa que fez a tua rosa tão importante"***

*Autor desconhecido*

*Se queres prever o futuro, estuda o passado.*

***Confúcio***

**Agradecimentos**

À minha orientadora, **Profa. Dra. Claudia Carareto**, pela sua dedicação à ciência, pelo suporte, persistência, confiança e paciência, um exemplo de educadora e pesquisadora a ser seguido;

Às **Profas. Dras. Marie Anne Van Sluys** e **Françoise Lemeunier**, bem como ao **Prof. Dr. Pierre Capy,** pela co-orientação, pelas discussões sempre proveitosas e pelo carinho com que me acolheram em seus laboratórios;

Aos funcionários do Departamento de Biologia e da sessão de Pós-graduação, especialmente aos técnicos **Sebastião**, **Paulo**, **Cristiane** e **Rinneu** pelo "papá" de minhas mosquinhas;

À técnica do GateLab **Sílvia Regina Blanco** pela pronta ajuda na realização das reações de seqüênciamento;

Aos Profs. Drs. **Hermione Bicudo**, **Magdalena Rossi**, **Galina Ananina** e **Elgion Loreto**, pela disposição em avaliar esta Tese e contribuir com a finalização dos nossos trabalhos;

Aos amigos e colegas que ainda estão ou já passaram pelo Laboratório de Evolução Molecular de Insetos **Adriana**, **Ana Cristina**, **Elaine**, **Elias**, **Fabrício**, **Flávia Cal**, **Julcimary**, **Juliana**, **Leliane**, **Lilian**, **Luciane** e **Luis Gustavo**, obrigada por sua amizade, pela ajuda constante e por agüentarem meu humor imprevisível;

Aos amigos do GateLab (USP – São Paulo) **Alessandro**, **Andrés**, **Breno**, **Dani K.**, **Dani Q.**, **Douglas**, **Leonor**, **Egdar**, **Fabiana**, **Guilherme**, **Jonas**, **Luisa**, **Marcelo**, **Maria Elisa**, **Marisa**, **Mayra**, **Obberdan**, **Robson** e **Sílvia** por me receberem de braços abertos, por sua companhia, atenção, amizade e pelos cafés de meio de tarde, que nos dão tanta força para continuar o trabalho;

Aos colegas do LEGS (CNRS – Gif-sur-Yvette) **Amir**, **Anna**, **Claire**, **Cushla**, **Jean-Luc**, **Lucie**, **Nicole** e **Thibaud**, bem como aos **Profs. Drs. Sylvie Aulard**, **Jean David**, **Catherine Montchamp-Moreau** e **Marie-Louise Cariou** pelas discussões científicas, pela

amizade, pela paciência que tiveram comigo nos primeiros tempos no exterior e por me mostrarem o quanto o Brasil é querido e valorizado por outros povos;

Aos meus queridos amigos que trago desde antes do Doutorado **Ana Flávia**, **Carolina Panin**, **Caroline Vieira**, **Daniela Freitas**, **Isabella**, **Ligia**, **Márcia**, **Rodrigo**, **Karina** e **Paula Renata** pela amizade e por todos os momentos MARAVILHOSOS que passamos juntos;

Aos queridos amigos que fiz durante o doutorado sanduíche **Aya**, **Cinthia**, **Delia, Felipe**, **Leonardo**, **Marina**, **Priscila**, **Rogério**, **Sabrina** e **Telma**, por todos os jantares depois de longos dias de trabalho, por todas as maravilhas que vimos pela primeira vez juntos e por serem minha família enquanto estive muito longe do ninho;

Aos avôs mais amados do mundo, **Deanira** e **Francisco**, aos meus primos **Tatiana**, **Rodrigo** e **José Roberto** e, agora sim, a minha tia **Claudia**, pelo amor e respeito que recebo de minha família, vocês são meu segundo porto seguro;

À todos os membros da minha segunda família, especialmente ao **Arnaldo**, **Letícia**, **Luís Fernando**, **Marcelo**, **Noemi**, **Vó Isabel** e **Vô Hélio** por me receberem tão carinhosamente e compartilharem comigo o seu afeto;

Ao menino mais amado do mundo, **Marcus** obrigada por você ser do jeitinho que você é, amigo, carinhoso e dedicado, saiba que eu te amo;

**À minha mãe, quero agradecer novamente por tudo que você representa para mim, exemplo de vida e de professora, obrigada por você estar sempre comigo, incondicionalmente, mas sempre me mostrando o caminho mais correto e a forma mais humana de se agir!**

*A percepção do desconhecido é a mais fascinante das experiências. O homem que não tem os olhos abertos para o misterioso passará pela vida sem ver nada.*

*Albert Einstein*

## RESUMO

O gênero *Zaprionus* tem sido eleito como um bom modelo biológico para estudos genético-comparativos com as espécies do subgrupo *melanogaster* do gênero *Drosophila*, embora seu posicionamento filogenético dentro da família Drosophilidae ainda seja controverso. Na presente Tese foi investigada a presença de 10 elementos de transposição (TEs) em *Zaprionus indianus* e *Drosophila malerkotliana*, bem como a distribuição, a atividade transcricional e as relações evolutivas de três retrotransposons (*copia*, *gypsy* e *micropia*) em sete espécies do gênero *Zaprionus*. Para isso, foram empregadas as técnicas de *Dot blot*, PCR, RT-PCR e seqüenciamento. As seqüências obtidas foram comparadas às dos respectivos elementos das demais espécies de drosofilídeos disponíveis nas bases de dados genômicas. Os resultados indicam que *Z. indianus* e *D. malerkotliana* apresentam em seus genomas todos os TEs de *D. melanogaster* investigados. O retrotransposon *copia* foi seqüenciado e está transcricionalmente ativo nas sete espécies do gênero *Zaprionus* e constitui uma nova subfamília relacionada aos elementos do subgrupo *melanogaster*, que foi denominada subfamília *GBFDouble-gap*. Por outro lado, os retrotransposons *gypsy* e *micropia* foram identificados nas espécies do subgênero *Zaprionus*, onde também estão transcricionalmente ativos, e pertencem às subfamílias já descritas para as espécies do subgrupo *melanogaster*. As análises evolutivas sugeriram que esses três retrotransposons devem ter participado de eventos de transferência horizontal com as espécies do subgrupo *melanogaster* e com pelo menos um doador desconhecido, no caso do retrotransposon *micropia*. Além disso, o cálculo dos tempos de divergência dos elementos sugere que eles passaram por ondas de transferências horizontais, mais antigas para o retrotransposon *copia*, e mais recentes para *gypsy* e *micropia*. Esses resultados podem indicar que o fenômeno de transferência horizontal pontuou a evolução das espécies do gênero *Zaprionus* e do subgrupo *melanogaster* algumas vezes durante sua diversificação no continente africano.



**Palavras-chave:** elementos de transposição, retrotransposon *copia*, retrotransposon *gypsy*, retrotransposon *micropia*, *Drosophila*, *Zaprionus*, subgrupo *melanogaster*, transferência horizontal, relações evolutivas.

## ABSTRACT

The *Zaprionus* genus has been elected as a good biological model for comparative analyses with the *melanogaster* subgroup of *Drosophila* genus, though its phylogenetic positioning within the Drosophilidae family is still controversial. This study aiming at investigating the occurrence of 10 transposable elements (TEs) in *Zaprionus indianus* and *Drosophila malerkotliana* species, as well the distribution, transcriptional activity and evolutionary relationships of three retrotransposons (*copy*, *gypsy* and *micropia*) in seven species of *Zaprionus* genus. To do so, Dot blot, PCR, RT-PCR and sequencing methods were employed. The *Zaprionus* sequences obtained were compared with the drosophilid sequences available in genomic databases. The results indicated that *Z. indianus* and *D. malerkotliana* harbor all *D. melanogaster* TEs investigated. The *copia* retrotransposon is present and transcriptionally active in seven species of the *Zaprionus* genus and represents a new subfamily related to that of the *melanogaster* subgroup, named as *GBFDouble-gap* subfamily. Additionally, *gypsy* and *micropia* retrotransposons were identified in the *Zaprionus* species subgenus, which are transcriptionally active and belong to the *melanogaster* subgroup subfamilies. The evolutionary analysis showed the three retrotransposons could have been involved in horizontal transfer events with species of the *melanogaster* subgroup for the three retrotransposons and at least one unknown donor regarding to *micropia* retrotransposon. Moreover, the time of divergence seems to indicate that the retrotransposons experienced horizontal transfer waves, the oldest involving the *copia* element followed by the *gypsy* and *micropia* retrotransposons in more recent times. These results suggest that the horizontal transfer phenomenon has happened repeatedly during the *Zaprionus* genus and *melanogaster* subgroup evolution in the Afrotropical region.



**Keywords:** transposable elements, *copia* retrotransposon, *gypsy* retrotransposon, *micropia* retrotransposon, *Drosophila*, *Zaprionus*, *melanogaster* subgroup, horizontal transfer, evolutionary relationships.

## SUMÁRIO

# 1 INTRODUÇÃO GERAL

## 1 INTRODUÇÃO GERAL

A genética de insetos tem sido amplamente estudada nas últimas décadas, tanto por esses organismos causarem prejuízos à vida humana, apresentando-se como vetores de doenças ou pragas agropecuárias, quanto por comporem um rico objeto de estudo biológico pelo alto número de espécies, pela riqueza de habitats que ocupam e pelos comportamentos complexos que exibem. Insetos também se apresentam como bons organismos-modelo, devido à conservação de mecanismos genéticos básicos, que possibilitam extrapolar os conhecimentos obtidos para animais mais complexos e pela facilidade na realização de estudos populacionais, pois muitos deles não requerem complexo manejo durante cruzamentos e produção de linhagens.

A espécie *Drosophila melanogaster* destaca-se como um dos organismos-modelo mais estudados, utilizada amplamente em análises de Genética Clássica e Molecular e, de Biologia do Desenvolvimento, dentre outras áreas da Biologia, o que a tornou um dos organismos invertebrados chave para o entendimento da herança e transmissão da informação genética. Adicionalmente, *D. melanogaster* e as espécies do subgrupo *melanogaster* têm sido usadas como modelo para estudos de processos ecológicos, comportamentais e filogenéticos visando melhor compreender os processos que regem a evolução dos organismos eucarióticos, e comprovando, dessa forma, a importância do estudo de espécies relacionadas para a elucidação de mecanismos evolutivos.

A amostragem e o estudo de espécies filogeneticamente menos relacionadas a *D. melanogaster* também consiste em uma abordagem fundamental em estudos genético-evolutivos. Dentre os drosofilídeos estudados geneticamente, não pertencentes ao subgrupo *melanogaster*, deve-se citar *Drosophila pseudoobscura*, espécie pertencente ao grupo *obscura*, grupo irmão de *melanogaster*, bem como *Drosophila hydei* e *Drosophila virilis*, pertencentes à radiação *virilis-repleta* do subgênero *Drosophila*. Em relação aos estudos desenvolvidos com as espécies da família Drosophilidae, não pertencentes ao gênero *Drosophila*, destacam-se aqueles que visam o entendimento da evolução orgânica e genômica das espécies dessa família, como, por exemplo, os que utilizam as espécies dos gêneros *Scaptodrosophila*, *Scaptomyza* e *Zaprionus*. Essas análises fornecem informações específicas sobre as relações evolutivas da família Drosophilidae e servem para o embasamento de estudos genéticos mais amplos, focando processos genéticos mais conservados.

A presente tese reafirma a importância da inclusão de espécies de diferentes gêneros, neste caso o gênero *Zaprionus*, como complemento na amostragem de estudos de evolução

molecular em Drosophilidae, mais especificamente, a evolução dos elementos de transposição (TEs), evidenciando a riqueza de informações que podem ser obtidas pela ampliação do número de espécies focadas durante a análise científica.

### 1.1 História do gênero *Zaprionus*

O gênero *Zaprionus* pertence à família Drosophilidae, que é uma das mais numerosas e amplamente distribuídas da ordem Diptera. Com provável origem tropical (THROCKMORTON, 1975), Drosophilidae é composta por cerca de 4.000 espécies classificadas em 69 gêneros, dentre os quais, o gênero *Drosophila* é o mais diverso, com mais de 2.000 espécies (MARKOW; O'GRADY, 2005).

Segundo a classificação dos drosofilídeos proposta por Throckmorton (1975), a radiação Steganinae é a mais basal, incluindo as espécies dos gêneros *Steganina* e *Leucophenga*. Eventos de diversificação posteriores resultaram na radiação Scaptodrosophila, bem como nas demais radiações que originaram os grupos de espécies de Drosophilidae. Dentre esses grupos, está a radiação Sophophora, que inclui as espécies do gênero *Chymomyza* e do subgênero *Sophophora*, dentre elas, as do grupo *melanogaster*. Por último, aparece a radiação Drosophila, dividida em três principais grupos de espécies: a radiação que contém as espécies do grupo *funebris*, como *Drosophila funebris*, espécie-tipo do gênero *Drosophila*; a radiação *virilis-repleta*, incluindo o grupo *repleta* e a radiação *immigrans-Hirtodrosophila*, que agrupa as espécies dos grupos *trinpunctata* e *cardini*, bem como os gêneros *Idyomia* (drosófilas havaianas), *Scaptomyza* e *Zaprionus*.

Críticas foram feitas à classificação de Throckmorton, pois seus estudos não consideraram o conceito de monofilia, nem explicitaram os métodos filogenéticos nos quais as análises foram baseadas, impedindo sua repetibilidade (MARKOW; O'GRADY, 2005). Dessa forma, análises posteriores propuseram novas relações entre os grupos de espécies de Drosophilidae. Okada (1990) sugeriu que o gênero *Drosophila* deveria ser considerado monofilético e as espécies dos outros gêneros, como por exemplo, *Zaprionus*, *Idiomya* e *Hirtodrosophila* deveriam ser colocadas fora dele. Após esse trabalho, Grimaldi (1990) construiu uma árvore filogenética para espécies de *Drosophila* e seus gêneros relacionados, usando análise cladística de 217 caracteres morfológicos de 120 espécies representativas. O principal objetivo de sua análise foi construir uma classificação filogenética da família Drosophilidae e avaliar as hipóteses de Throckmorton (1975). A análise de Grimaldi (1990) propôs que os subgêneros *Drosophila* e *Sophophora* seriam clados irmãos e excluiu do gênero

*Drosophila* os subgêneros *Hirtodrosophila*, *Scaptomyza*, *Idiomya* e *Zaprionus*. Essa proposta contrasta com a de Throckmorton (1975), que propõe uma divergência precoce do subgênero *Sophophora*, que seria grupo irmão de um táxon complexo que incluiria os subgêneros *Drosophila, Hirtodrosophila, Dorsilopha, Scaptomyza, Idyomia* e *Zaprionus*.

Diversos estudos utilizando marcadores moleculares têm reavaliado as relações filogenéticas de Drosophilidae (DE SALLE, 1992; PÉLANDAKIS; SOLIGNAC, 1993; THOMAS; HUNT, 1993; KWIATOWSKI et al. 1994; REMSEN; DESALLE, 1998; RUSSO et al., 1995; TATARENKOV et al. 1999; KWIATOWSKI; AYALA, 1999; ROBE et al., 2005; DA LAGE et al., 2007; YASSIN et al., 2008a). Tendo como foco o gênero *Zaprionus* e os subgêneros *Sophophora* e *Drosophila*, diferentes relações são propostas de acordo com os marcadores moleculares analisados. Apesar das análises de DNA mitocondrial de De Salle (1992) corroborarem a topologia proposta por Grimaldi (1990) e colocarem *Zaprionus* como um grupo basal, que divergiu anteriormente à radiação *Sophophora*, do gênero *Drosophila*, a maior parte das análises moleculares conflitam com as propostas de Grimaldi (1990). Thomas e Hunt (1993) construíram uma árvore filogenética usando 11 seqüências do gene álcool desidrogenase (*Adh*) de diferentes gêneros de drosofilídeos e, nessa árvore, *Zaprionus* é colocado no interior de um ramo, entre os subgêneros *Drosophila* e *Sophophora*. A mesma relação filogenética foi obtida por Kwiatowski et al. (1994), Kwiatowski e Ayala (1999) e Remsen e De Salle (1998) analisando os genes *Adh*, glicerol 3-fosfato desidrogenase (*Gpdh*) e *Cu* e *Zn* superóxido dismutase (*Sod*). No primeiro trabalho foram utilizados como marcadores genéticos seqüências do gene *Sod*, no segundo, seqüências de *Adh, Gpdh* e *Sod* e, no terceiro, uma combinação de informações morfológicas, marcadores nucleares e mitocondriais.

Um padrão um pouco diferente, mas que também coloca *Zaprionus* dentro do gênero *Drosophila*, foi encontrado por Pélandakis e Solignac (1993), Russo et al. (1995) e Tatarenkov et al. (1999), em cujas árvores, as espécies do gênero são colocadas como um ramo basal dentro do subgênero *Drosophila*, intimamente relacionado com os grupos desse subgênero. Pélandakis e Solignac (1993), baseados em análises de seqüências de RNA ribossomal, propuseram que as espécies de *Zaprionus* seriam mais relacionadas às dos grupos *immigrans* e *repleta*. Por outro lado, Russo et al. (1995) e Tatarenkov et al. (1999) não foram capazes de especificar qual grupo do subgênero *Drosophila* seria mais relacionado a *Zaprionus*, usando vários marcadores combinados, entre eles, genes nucleares e mitocondriais. Os dados de Robe et al. (2005) também corroboram uma relação mais íntima entre as espécies de *Zaprionus* e do subgênero *Drosophila*, analisando os genes alfa metil-dopa (*Amd*) e citocromo oxidase II (*COII*).

Em dois estudos recentes o gene *Amyrel* foi utilizado como marcador para a reconstrução da história da família Drosophilidae. O primeiro incluiu mais uma vez o gênero *Zaprionus* como constituinte basal do subgênero *Drosophila* (DA LAGE et al., 2007) e, o segundo, adicionando à análise o gene *COII* e caracteres morfológicos, propôs que *Zaprionus* seria mais intimamente relacionado com *Drosophila repletoides*, do grupo *tumiditarsus*, que é classificado basalmente no subgênero *Drosophila* (YASSIN et al., 2008a).

Os resultados obtidos em todos esses estudos não apontam um posicionamento exato para as espécies do gênero *Zaprionus*, indicando que ainda são necessárias análises mais refinadas e um número maior de marcadores moleculares aliados aos dados morfológicos, para que a reconstrução da história evolutiva desse gênero possa ser mais claramente compreendida. Entretanto, parece estar surgindo um consenso de que a diversificação de *Zaprionus* aconteceu após a origem do subgênero *Sophophora*, tornado-o mais relacionado ao subgênero *Drosophila*. Contudo, a relação filogenética exata entre *Zaprionus* e o subgênero *Drosophila* ainda é motivo de especulações. A adequação da nomenclatura, elegendo o subgênero *Sophophora* ao nível de gênero e incluindo espécies de outros gêneros, como *Zaprionus* e *Hirtodrosophila*, no gênero *Drosophila*, já foi proposta (DALTON, 2009). No entanto, esse assunto ainda é matéria de amplo debate, por duas razões. A primeira se deve ao fato que, dependendo do gene utilizado, *Zaprionus* se agrupa com diferentes grupos de espécies do subgênero *Drosophila*. A segunda refere-se às dificuldades práticas dessa mudança, como a substituição do nome específico de mais de 1.000 espécies do subgênero *Sophophora*, dentre elas *D. melanogaster*, uma espécie-modelo em vários ramos da Biologia, que passaria a ser denominada *Sophophora melanogaster*.

### 1.2 Relações filogenéticas entre as espécies do gênero *Zaprionus*

O gênero *Zaprionus* foi descrito por Coquillet (1901), sendo que as espécies que o compõem são caracterizadas pela presença de faixas brancas longitudinais na cabeça e no mesonotum. As espécies desse gênero são divididas em dois subgêneros, de acordo com suas origens geográficas. O subgênero *Anaprionus* é original da região biogeográfica Oriental e é composto por 11 espécies, já o subgênero *Zaprionus* apresenta 49 espécies e deve ter sido originado na região Afrotropical (OKADA; CARSON, 1983; YASSIN et al., 2008ab).

As relações filogenéticas entre as espécies do subgênero *Anaprionus* não têm sido estudadas e apenas dois estudos focaram as relações evolutivas dentro do subgênero *Zaprionus* (CHASSAGNARD; TSACAS, 1993; YASSIN et al., 2008a). A classificação

proposta por Chassagnard e Tsacas (1993) foi baseada em um único caráter morfológico, a ornamentação e posicionamento das cerdas na região do mesonotum. Os resultados obtidos permitiram a proposição dos grupos *inermis* e *armatus*, esse último subdividindo-se em três subgrupos: *armatus, tuberculatus* e *vittiger*.

No estudo filogenético mais recente, de Yassin et al. (2008a), um conjunto de caracteres morfológicos (21 caracteres morfológicos binários) e moleculares (2164 caracteres nucleotídicos provenientes do gene nuclear *Amyrel* e do gene mitocondrial *COII*) foi analisado visando propor as relações evolutivas de 42 espécies do subgênero *Zaprionus*. Os resultados obtidos corroboraram a proposta de monofilia do gênero *Zaprionus* e os grupos *inermis* e *armatus*, bem como possibilitaram sugerir um arranjo de complexos dentro desses dois grupos (Figura 1). O grupo *inermis* foi dividido em quatro complexos de espécies: *sexvittatus* (duas espécies), *ghesquierei* (uma espécie), *inermis* (duas espécies) e *tuberculatus* (cinco espécies). Adicionalmente, o grupo *armatus* foi dividido em nove complexos: *litos* (uma espécie), *montanus* (duas espécies), *armatus* (oito espécies), *megalorchis* (duas espécies), *indianus* (três espécies), *davidi* (três espécies), *spinosus* (cinco espécies), *vittiger* (três espécies) e *lachaisei* (sete espécies). A árvore apresentada na Figura 1 mostra apenas a espécie *Zaprionus indianus* compondo o complexo *indianus*. As duas outras espécies que compreendem o complexo *indianus* foram descritas em um estudo molecular que permitiu a identificação de *Zaprionus africanus* e *Zaprionus gabonicus*, espécies crípticas e altamente relacionadas à *Z. indianus*, dados confirmados por análises de isolamento reprodutivo e de morfometria (YASSIN et al., 2008b).

Yassin et al. (2008a) também propuseram a data de origem do gênero *Zaprionus*, equivalente aos períodos médio e tardio do Mioceno (cerca de 10 milhões de anos atrás), e corroboraram sua origem na região Oriental e posterior diversificação do subgênero *Zaprionus* na região Afrotropical.

Os estudos sobre a origem e as relações filogenéticas das espécies do gênero *Zaprionus* indicam que sua diversificação aconteceu contemporaneamente com a evolução das espécies do subgrupo *melanogaster* de *Drosophila* (YASSIN et al., 2008a). Esses dados, aliados ao conhecimento da diversidade ecológica, da riqueza de espécies e a facilidade de coleta e manutenção das espécies em culturas laboratoriais, permitiram a esses autores elegerem o gênero *Zaprionus* como um bom modelo para estudos genéticos comparativos com as espécies do subgrupo *melanogaster*.

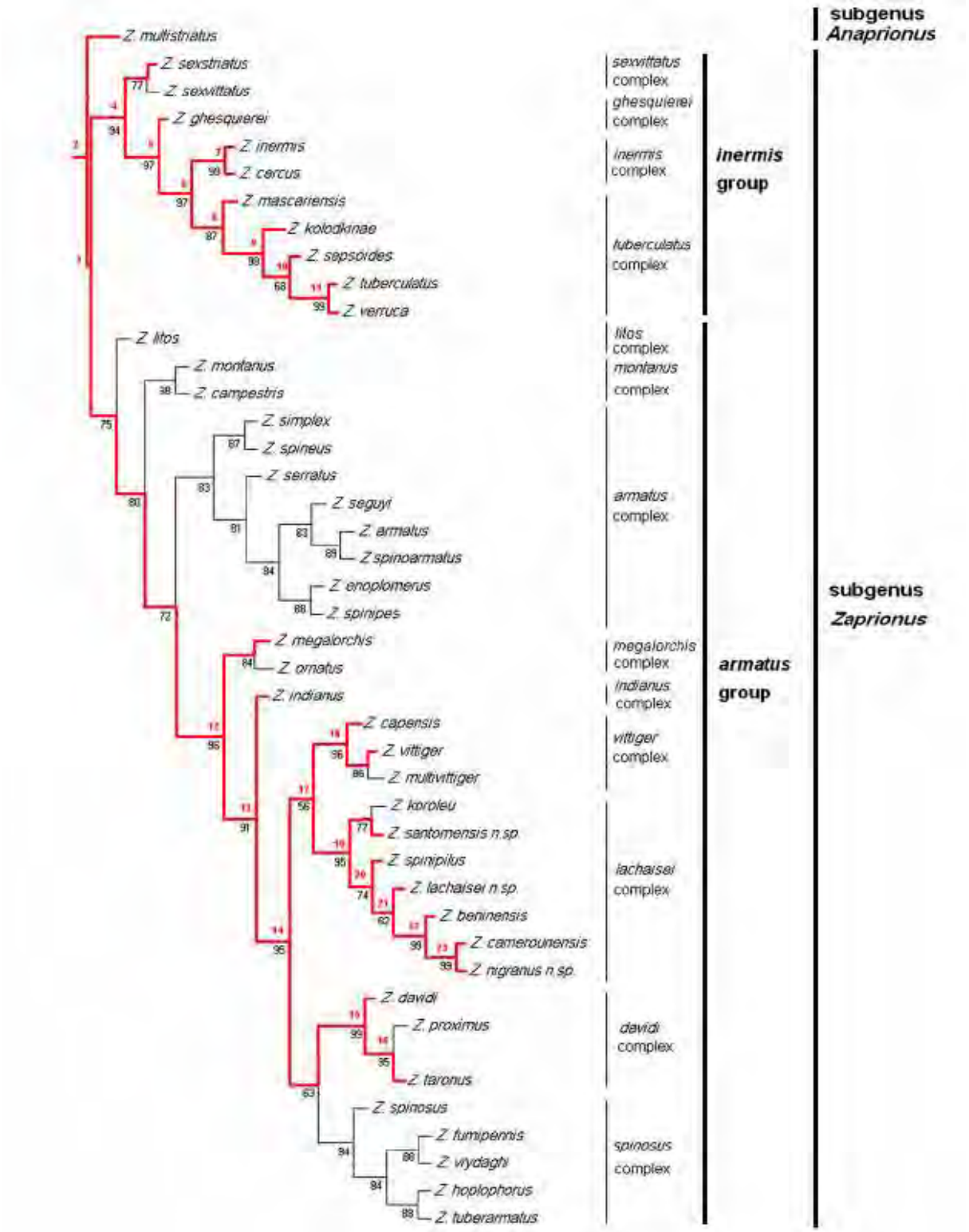



**Figura 1.** Filogenia do gênero *Zaprionus* proposta por Yassin et al. (2008a) baseada em caracteres morfológicos e moleculares construída pelo método "*morphological grafting*". Caracteres morfológicos foram analisados para todas as espécies amostradas. Os ramos em vermelho indicam as espécies estudadas também com dados moleculares.

### 1.3 Os elementos de transposição do gênero *Zaprionus*

Os elementos de transposição são componentes significativos do genoma de todos os seres vivos, e apresentam extrema diversidade nos modos de transposição e na variabilidade de seqüências (WICKER et al., 2007). Sua importância biológica está relacionada à produção de variabilidade genética nos genomas hospedeiros, por poderem gerar mutações gênicas (O'HARE et al., 1984; AQUADRO et al., 1986; JORDAN et al., 2003; ALMEIDA et al., 2008; LOPES et al., 2008) e alterações cromossômicas (KIDWELL et al., 1977; LANGLEY et al., 1988; MONTGOMERY et al., 1991), desempenhando assim, um papel chave como mediadores da evolução molecular (MCDONALD, 1990; KIDWELL; LISCH, 1997).

A classificação mais recente dos elementos de transposição se baseia no sistema hierárquico de agrupamento, levando em consideração características estruturais como a presença de domínios enzimáticos e a diversidade de seqüências (WICKER et al., 2007). Esse sistema inclui classes, subclasses, ordens, superfamílias, famílias e subfamílias. Dessa forma, os TEs são divididos em duas classes: a Classe I, abrigando os elementos que se replicam via um RNA intermediário (também denominados retrotransposons); e, a Classe II, onde estão inclusos os elementos que se mobilizam via DNA, diretamente pela excisão e posterior inserção em novas posições nos genomas (transposons). A classificação em subclasses se refere ao tipo de enzimas que realizam a transposição, colocando os retrotransposons em uma única subclasse, devido à transposição mediada pela transcriptase reversa, e os transposons em duas subclasses diferenciadas, os elementos que se transpõem com auxílio da enzima transposase e os *Helitrons* e *Mavericks*, elementos que se mobilizam com auxílio de enzimas como a DNA helicase e a tirosina recombinase.

As superfamílias de elementos de transposição são caracterizadas pela estratégia de replicação e pelo compartilhamento de proteínas e regiões não codificantes; como por exemplo, a superfamília *copia*, que se caracteriza pelo arranjo das enzimas integrase-transcriptase reversa-RNAseH, no domínio poliproteíco. Por fim, as classificações de famílias e subfamílias são dadas pelo nível de conservação das seqüências de DNA, promovendo a identificação das relações entre seqüências que compartilham ancestrais mais recentes. Elementos de uma mesma família apresentam ao menos 80% de similaridade em seqüências codificantes ou reguladoras maiores que 80 pb, como por exemplo, as famílias *copia* e *gypsy* de Drosophilidae; por sua vez, as subfamílias são compostas por subpopulações de seqüências, claramente segregadas em análises filogenéticas, como demonstrado para a

família *gypsy*, amplamente distribuída em espécies da família Drosophilidade, que é formada por 10 subfamílias (HEREDIA et al., 2004; LUDWIG et al., 2008).

As espécies da família Drosophilidae têm sido alvo de inúmeros estudos sobre conteúdo, diversidade, regulação e mecanismos de transposição de seus TEs. Contudo, a maior parte do conhecimento adquirido sobre essas seqüências foi obtida por meio de estudos com *D. melanogaster* e outras espécies do gênero *Drosophila*, fornecendo informações sobre a dinâmica transposicional e os mecanismos evolutivos que os mantêm nos genomas. Além disso, o seqüenciamento e a anotação do genoma de 12 espécies de *Drosophila* (http://rana.lbl.gov/drosophila), aumentaram significativamente a quantidade de informação e a gama de possibilidades para o estudo dos TEs nessas espécies, principalmente estudos comparativos, que permitem analisar a evolução, a atividade e a dinâmica de manutenção dessas seqüências nos genomas hospedeiros.

Embora muitos trabalhos se refiram aos elementos transponíveis de *D. melanogaster* e de outras espécies do mesmo gênero, praticamente não existem informações sobre a diversidade, a estrutura e a dinâmica dos elementos transponíveis nos demais gêneros de Drosophilidae, como por exemplo, no gênero *Zaprionus*. Estima-se que existam cerca de 96 famílias de TEs apenas no genoma de *D. melanogaster* (KAMINKER et al., 2002), no entanto, foram identificados e superficialmente estudados apenas cinco elementos de transposição em todo o gênero *Zaprionus*, que é composto por pelo menos 60 espécies (OKADA; CARSON, 1983; YASSIN et al., 2008ab). Dentre os TEs descritos encontram-se quatro retrotransposons (*1731*, *412*, *gypsy* e *copia*) e um transposon (*mariner*), estudados em uma ou poucas espécies, em cada trabalho.

O retrotransposon *copia* é um dos elementos melhor caracterizados na família Drosophilidae. Existem informações sobre sua distribuição (BIÈMONT; CIZERON, 1999), regulação mediada pelo hospedeiro (CAVAREC; HEIDMANN, 1993; BHADRA et al., 1999; CAVAREC et al., 1994; TULIN et al., 2002; ARAVIN et al., 2003; KAWAMURA et al., 2008) e seu impacto nos genomas onde está inserido (O'HARE et al., 1984; AQUADRO et al., 1986; STRAND; MCDONALD, 1989; MÉVEL-NINIO et al., 1989). McDonald et al. (1997) publicaram uma seqüência da região LTR-ULR (do Inglês, *Long Terminal Repeat*) do retrotransposon *copia* de *Zaprionus tuberculatus* (subgênero *Zaprionus*, grupo *armatus*, complexo *tuberculatus*), demonstrando a conservação de seus motivos reguladores quando comparados com os do elemento de *D. melanogaster*. Posteriormente, Almeida e Carareto (2006) utilizaram essa seqüência numa análise da história evolutiva de *copia* envolvendo as espécies dos grupos *melanogaster* e *repleta*, *Drosophila willistoni* e *Z. tuberculatus*. As

relações filogenéticas e a alta similaridade entre as seqüências de *Z. tuberculatus*, de *D. willistoni* e do grupo *melanogaster*, foram utilizadas na inferência da ocorrência de um evento de transferência horizontal entre essas espécies.

Outro representante da superfamília *copia*, o retrotransposon *1731*, também foi estudado no gênero *Zaprionus*. Sua presença foi identificada em *Zaprionus ornatus* (subgênero *Zaprionus*, grupo *armatus*, complexo *megalorchis*) por meio da marcação em *Squash blot*, com o objetivo de demonstrar a ampla distribuição de *1731* em Drosophilidae, corroborando a hipótese de origem ancestral nessas espécies (MONTCHAMP-MOREAU et al., 1993).

O retrotransposon *412* é amplamente distribuído em Drosophilidae, indicando uma longa história evolutiva nessas espécies (CIZERON et al., 1998). Também tem sido sugerido que a transposição desse elemento sofre controle do hospedeiro (MUGNIER et al., 2005). Sua identificação em *Z. tuberculatus* foi realizada por meio das técnicas de *Southern blot* e Hibridação *in situ*, que indicaram altos números de inserções e conservação com a sonda do elemento *412* de *D. melanogaster* (CIZERON et al., 1998).

*Gypsy* foi classificado como um retrotransposon pertencente à superfamília dos *errantivirus*, devido à presença de um gene que codifica uma proteína similar à do *envelope* viral, proporcionando características infecciosas a esses TEs (MEJLUMIAN et al., 2002; HEREDIA et al., 2004). Esse elemento foi isolado em *Z. indianus* (subgênero *Zaprionus*, grupo *armatus*, complexo *indianus*) por Heredia et al. (2004) a partir de uma seqüência de cerca de 500 pb do *envelope*. Neste trabalho foram sugeridos vários casos de transferência horizontal entre as espécies da família Drosophilidae, entre eles, um evento de transferência de *Drosophila simulans* para *Z. indianus*, que teria ocorrido a cerca de 2,2 milhões de anos.

Todos os estudos citados acima identificaram a presença apenas de retrotransposons nas espécies do gênero *Zaprionus*. Por sua vez, dentre os representantes da classe II da família Drosophilidae, apenas um TE foi estudado em *Zaprionus*, o transposon *mariner* (MARUYAMA; HARTL, 1991; BRUNET et al., 1999). A presença desse elemento foi verificada por meio de *Southern blot* em *Zaprionus sepsoides* (subgênero *Zaprionus*, grupo *armatus*, complexo *tuberculatus*), *Zaprionus ghesquierei* (subgênero *Zaprionus*, grupo *inermis*, complexo *ghesquierei*) e *Z. indianus* (MARUYAMA; HARTL, 1991). No mesmo estudo, esse elemento foi seqüenciado em *Z. tuberculatus*, sendo que sua seqüência demonstrou alta similaridade (97%) com a de *Drosophila mauritiana*, sugerindo que esses transposons compartilham um ancestral comum mais recente do que as espécies e que podem ter sido envolvidos em um evento de transferência horizontal (MARUYAMA; HARTL,

1991). Brunet et al. (1999) demonstraram também que *Zaprionus verruca* (subgênero *Zaprionus*, grupo *inermis*, complexo *tuberculatus*) possui o transposon *mariner* em seu genoma e que essa seqüência pertence ao mesmo grupo do elemento de *Z. tuberculatus*; além disso, o sequenciamento do *mariner* de *Z. indianus* demonstrou que ele é mais relacionado aos elementos de *Drosophila fima* do que aos elementos das outras espécies do gênero *Zaprionus*.

Embora os dados existentes sobre a presença e a evolução de elementos de transposição nos genomas do gênero *Zaprionus* sejam escassos e pontuais, os resultados demonstram uma ampla distribuição de TEs e sugerem que mais de um tipo de mecanismo evolutivo tenha moldado sua presença e diversificação nessas espécies. Esse padrão sugere ainda que a ampliação desses estudos pode gerar resultados importantes para o entendimento e a proposição de mecanismos que regem a evolução dos TEs em Drosophilidae.

# 2 OBJETIVOS

## 2 OBJETIVOS

O objetivo geral desta Tese foi estudar a distribuição e inferir a história da transmissão de elementos de transposição no gênero *Zaprionus*, um dos grupos de espécies da família Drosophilidae pouco estudados até o momento, visando contribuir para o entendimento da dinâmica evolutiva dessas seqüências repetitivas nos genomas hospedeiros eucarióticos. A busca pelo objetivo geral foi mediada pelos seguintes objetivos específicos:

1. Identificar elementos de transposição representativos das classes dos retrotransposons e dos transposons em *Z. indianus*;
2. Selecionar três elementos de transposição dentre os identificados em *Z. indianus*, para avaliar a distribuição e a atividade transcricional em sete espécies do gênero *Zaprionus*;
3. Seqüenciar parcialmente os três elementos selecionados em todas as espécies nas quais eles foram identificados, e;
4. Avaliar as relações filogenéticas entre as seqüências obtidas e as disponíveis nas bases de dados genômicas.

Os resultados obtidos foram organizados na forma de capítulos que correspondem a uma publicação e três manuscritos, os quais focam a identificação de elementos de transposição de *Z. indianus* e *Drosophila malerkotliana* (Capítulo I), uma revisão sobre o retrotransposon *copia* na família Drosophilidae (Capítulo II), o estudo do retrotransposon *copia* no gênero *Zaprionus* (Capítulo III) e uma análise da transmissão dos retrotransposons *gypsy* e *micropia* no gênero *Zaprionus* (Capítulo IV), sendo que os dois últimos capítulos se referem a estudos comparativos com outras espécies da família Drosophilidae.

# 3 CAPÍTULO I

*terrestres* (Brandão, C.R., and E.M. Cancello, eds.), pp. 245-261; Tosi, D., M. Martins, C.R. Vilela, and M.A.Q.R. Pereira 1990, Braz. J. Genet. 13: 19-31; Val, F.C., and M.D. Marques 1996, Pap. Avul. Zool. 39: 223-230; Val, F.C., C.R. Vilela, and M.D. Marques 1981, In: *The Genetics and Biology of* Drosophila (Ashburner, M., H.L. Carson, and J.N. Thompson, Jr., eds.), pp. 123-168; Valente, C.R., 2006, In: *Natureza viva cerrado: caracterização e conservação* (Guimarães, L.D., M.A.D. Da Silva, and T.C. Anacleto, eds.) pp. 20-44; Vilela, C.R., 1983, Rev. bras. Ent. 27: 1-114; Wheeler, M.R., and M.P. Kambysellis 1966, Univ. Texas Publs. 6615: 533-565.

**Screening for transposable elements in South America invasive species *Zaprionus indianus* and *Drosophila malerkotliana*.**

**de Setta, N. and C.M.A. Carareto.** UNESP – Universidade Estadual Paulista, São José do Rio Preto, SP, Brazil; e-mail: carareto@ibilce.unesp.br

Transposable elements (TEs) have usually been studied focusing on distribution, structure, activity and evolution. Such analyses primarily aim at understanding the impact of TEs in the genetic dynamics and evolution of the host genomes. A kind of impact analysis that has not been enough explored is to look at genomic dynamics of TEs under stressful conditions, like environment invasion and colonization. According to the "wake-up" hypothesis (Vieira *et al.*, 1999), there is a tendency of mobilization intensification in colonizing populations resulting in an increasing of copy number. An example is given by the comparison between *D. melanogaster* and *D. simulans*, two cosmopolitan sibling species. *D. melanogaster* has dispersed around the world some centuries before *D. simulans* (Capy and Gilbert, 2004) and among 34 transposable elements analyzed, *D. melanogaster* appears to have higher insertion site numbers than *D. simulans* for 29 TEs when populations of America, Asia, Europe, Australia and Africa, the last one being the native place of both species, are compared (Vieira *et al.*, 1999).

Besides *D. melanogaster* and *D. simulans*, numerous other Drosophilidae were able to invade new habitats and had recently colonized the Brazilian territory mainly by antropic activity; among them, *Zaprionus indianus* (from Africa) and *D. malerkotliana* (from Asia). *Zaprionus indianus* was first collected in 1998, in a fig culture at Valinhos city, Sao Paulo (Vilela, 1999), and now it can be collected in several regions of Brazil (Vilela, 1999; Castro and Valente, 2001; Kato *et al.*, 2004), Uruguay (Gõni *et al.*, 2002), and southern North America (van der Linde *et al.*, 2006). *D. malerkotliana* has invaded South America in the 1970's decade (Val and Sene, 1980) and nowadays it is frequently collected in southern Brazil. The recent invasion of a continent by a species provides a useful tool for studying the dynamics of TEs during colonizing stress, because we can monitor populations since the first steps of introduction. To do that, it is necessary to accumulate information about distribution, history, and dynamics of its transposable elements. The goal of this study was then to make an initial search for transposable elements in *Z. indianus* and *D. malerkotliana* genomes to guide further studies on these invasive species.

Seven retrotransposons were searched (*copia*, *mdg-1*, *412*, *gypsy*, *297*, *micropia*, and *roo/B104*), two non-LTR retrotransposons (*jockey* and *doc*), and one transposon (*bari-1*) in a Brazilian population of *Z. indianus* (Mirassol, SP) and *D. malerkotliana* (Onda Verde, SP) by the Dot blotting method. The probes were prepared by two different methods: (1) PCR reactions were performed in 25 μl using approximately 50 ng of each TE plasmid, 0.4 μM of each specific primer

(Table 1), 200 mM of dNTPs, 1.5 mM of $MgCl_2$ and 1U of Taq Platinum Polymerase (Invitrogen) in 1× PCR buffer; for amplification we used an initial denaturation step of 3 min at 94°C, 30 cycles consisting of 30 sec at 94°C, 30 sec at 56°C and 1 min at 72°C, finally a additional extension step of 10 min at 72°C were performed; (2) 5 μg of each plasmid was digested with appropriate restriction enzymes (Table 1). The fragments were isolated from agarose 1% gels, purified using GFX PCR DNA and Gel Band Purification (GE Healthcare), and probed using Gene Images Random Prime Labelling Module (GE Healthcare) according to the manufacturer's instructions. For screening the TEs, 8 mg of genomic DNA from each strain was dropped in the Hybond N+ nylon membranes (GE Healthcare). For hybridization and detection we used the chemioluminescent hybridization system Gene Images (GE Healthcare) at high stringency (60°C) according to the manufacturer's instructions. Ultra pure water and 1 mg of each plasmid were used as negative and positive controls, respectively.

Table 1. Plasmids, primers and restriction enzymes used to prepare probes for detection of transposable elements in *Z. indianus* and *D. malerkotliana*.

| Element Plasmid [1] | Probe type | Primers / Restriction enzymes | Probe extension[2] |
|---|---|---|---|
| *copia* - p77E4 | PCR | LTR-5´CTATTCAACCTACAA AAATAACG3´<br>PCS-5´ATTACGTTTAGCCTTGTCCAT3´ | 439 bp |
| *micropia* - dhMiF$_2$ | PCR | 2813-5´TTAACTCCTAGAGTTCATCGCTGG3´<br>2814-5´CATGTACCTGGTTAACTACTGACC3´ | 387 bp |
| *gypsy* - pGGHS | PCR | GM003-5´GTACTGAACATTATCAGAATC3´<br>GM004-5´TCTAAGGAGTCCTCTGCAAGG3´ | 542 bp |
| *412 - pBR322* | Restriction fragment | Hind III | ~850 bp |
| *297 - pBR322* | Restriction fragment | Eco RI | ~2,3 kb |
| *mdg-1 - pBR322* | Restriction fragment | Eco RI / Hind III | ~1,2 kb |
| *roo/ B104 - pBR322* | Restriction fragment | Eco RI / Hind III | ~2,3 kb |
| doc - *pBspt Kst* | Restriction fragment | Eco RI / Hind III | ~800 kb |
| *jockey - puC19* | Restriction fragment | Eco RI / Hind III | ~1,5 kb |
| *bari-1 - puC8* | PCR | Br1 – 5' ATTCGTCGCAGGCTAAAAGA 3'<br>Br2 – 5' TTGTAACACCACCTTTGGCA 3' | 703 bp |

[1] Plasmid source: *copia* and *412* – E. Loreto (UFSM, Santa Maria, RS, Brazil); *gypsy* - D. Dorsett (Memorial Sloan-Kettering Cancer Center, USA); *micropia* – D.H Lankenau (University of Heidelberg, Heidelberg, Germany); *297*, *mdg-1*, *roo/B104* and *doc* – C. Vieira (Université Lyon I, Lyon, France); *jockey* – D.J. Begun (University of California, Davis, USA); *bari-1* – R. Caizzi (Universidade de Bari, Bari, Itália). [2] TE sequences inserted into plasmids were derived from:*D. melanogaster* (p77E4, pGGHS, *pBR322*, *pBspt Kst*, *puC19* and *puC8*) and *D. hydei* (dhMiF$_2$).

The 10 TEs investigated were identified in both *Z. indianus* and *D. malerkotliana*; however, the distribution of sequences homologous to these elements showed different hybridization signals, varying from very strong to weak signals (Figure 1).

The distribution of transposable elements of *Z. indianus* and *D. malerkotliana* is in agreement with the few data previously reported. Among the TEs here analyzed, only *gypsy* had been searched in *Z. indianus* (Heredia *et al.*, 2004), and the authors proposed that this retrotransposon was received

from *D. simulans* through an event of horizontal transfer. Likewise, our result showed a strong homology between *gypsy* probe from *D. melanogaster* and *Z. indianus* sequences, reinforcing the previous data, since *D. melanogaster* and *D. simulans* are sibling species and their *gypsy* sequences are highly similar (Heredia *et al.*, 2004). *D. malerkotliana* had just been studied for *copia* and *412* occurrences. *Copia* had been identified by weak *in situ* hybridization signal on the chromocenter (Biémont and Cizeron, 1999), and *412* had also been shown on the chromocenter both by *in situ* hybridization and Southern blotting (Cizeron *et al.*, 1998). Our analysis showed that *copia* and *412* are components of *D. malerkotliana* genome, but the hybridization signals obtained are also very weak, suggesting that both TEs of *D. malerkotliana* have low homology with those of *D. melanogaster*.

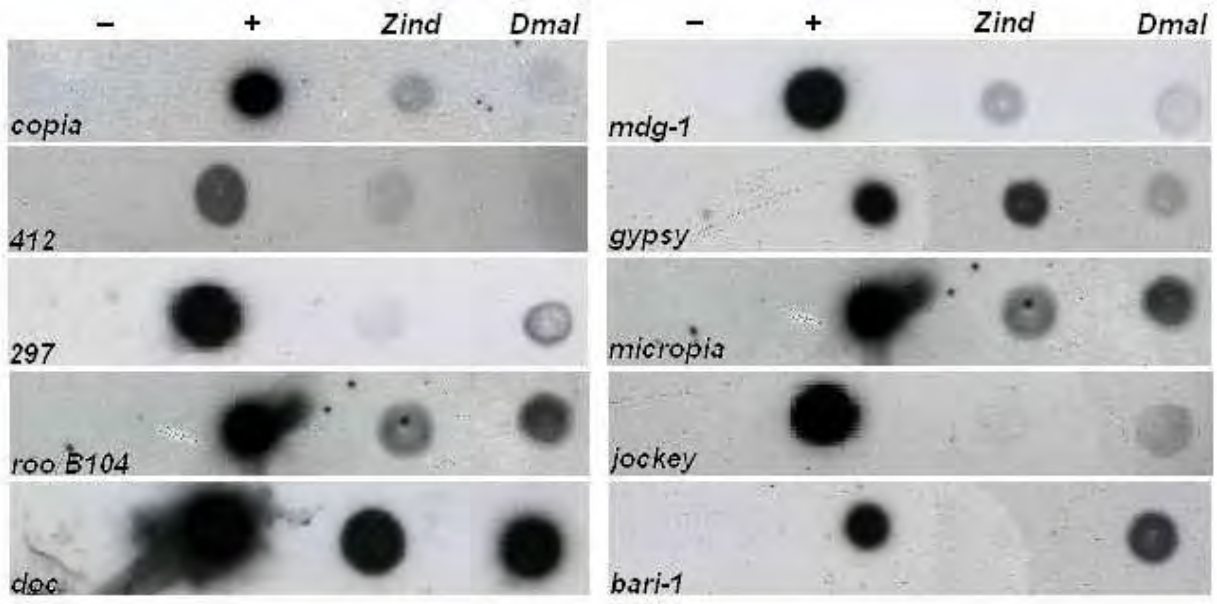



Figure 1. Dot blotting analysis of the transposable elements *copia*, *mdg-1*, *412*, *gypsy*, *297*, *micropia*, *roo/B104*, *jockey*, *doc* and *bari-1* in *Z. indianus* (*Zind*) and *D. malerkotliana* (*Dmal*). (–, negative control; +, positive control).

The results here presented show for the first time the occurrence of several transposable elements in two species scarcely studied in this respect. Except for *doc* and *gypsy* in *Z. indianus* and *doc*, *micropia* and *bari-1* in *D. malerkotliana*, all the analyses produced weak hybridization signals (Figure 1), which suggest that these TEs are divergent from those of *D. melanogaster* and *D. hydei* (for *micropia*) and they should have evolved vertically, accumulating nucleotide variation since the last common ancestor. On the other hand, *doc* and *gypsy* of *Z. indianus* and, *doc*, *micropia* and *bari-1* of *D. malerkotliana* have shown remarkably strong hybridization signals (Figure 1). Further analysis could test which is the most parsimonious hypothesis to explain that high homology between *Z. indianus* / *D. malerkotliana* and *D. melanogaster* / *D. hydei* sequences, if horizontal transfer of these elements between these species or any other alternative factor that can explain high similarity between elements of distantly related species (for example, Almeida and Carareto, 2005; Setta *et al.*, 2007). This report is the first step of a study regarding the dynamics of transposable elements in the

invasive species *Z. indianus* and *D. malerkotliana* and will be used for selecting candidate TEs for more detailed studies.

Acknowledgments: The authors are grateful to CNPq and COFECUB-CAPES for fellowships to NS and to CNPq by fellowship to CMAC.

References: Biémont, C., and G. Cizeron 1999, Genetica 105: 43-62; Capy, P., and P. Gilbert 2004, 120: 5-16; Castro, F.L., and V.L.S. Valente 2001, Dros. Inf. Serv. 84: 15-17; Cizeron, G., F. Lemeunier, C. Loevenbruck, A. Brehm, and C. Biémont 1998, 15: 1589-1599; de Almeida, L.M., and C.M.A. Carareto 2005, Mol. Phylogenet. Evol. 35: 583-594; de Setta, N., E.L. Loreto, and C.M.A. Carareto 2007, J. Mol. Evol. doi: 10.1007/s00239-007-9051-7; García Guerreiro, M.P., and A. Fontdevila 2001, Genet. Res. 77: 227-238; Gõni, B., M.E. Martinez, G. Techera, and P. Fresia 2002, Dros. Inf. Serv. 85: 75-80; Heredia, F., E.L. Loreto, and V.L. Valente 2004, Mol. Biol. Evol. 21: 1831-1842; Kato, C.M., L.V. Foureaux, R.A. Cesar, and M.P. Torres 2004, Ciênc. Agrotec. 28: 454-455; Labrador, M., M.C. Seleme, and A. Fontdevila 1998, Mol. Biol. Evol. 15: 1532-1547; Val, F.C., and F.M. Sene 1980, Papéis Avulsos de Zoologia 33: 293-298; van der Linde, K., G.J. Steck, K. Hibbard, J.S. Birdsley, L.M. Alonso, and D. Houle 2006, Fla. Entomol. 89: 402-404; Vieira, C., D. Lepetit, S. Dumont, and C. Biémont 1999, Mol. Biol. Evol. 16: 1251-1255; Vilela, C.R., 1999, Dros. Inf. Serv. 82: 37-39.

## Effects of Phloxine B and Hematoporphyrin IX on immature stages of *Drosophila melanogaster*.

**Massaldi-Fuchs, A.M.M.[1*], L.M. Pujol-Lereis[1*], A. Rabossi[1], A. Filiberti[2], C.E. Argaraña[2], and L.A. Quesada-Allué[1**].** [1]FCEyN- University of Buenos Aires, CONICET and Fundación Instituto Leloir, Patricias Argentinas 435, (1405) Buenos Aires, Argentina; [2]CIQUIBIC- University of Córdoba, CONICET. Ciudad Universitaria, Córdoba, Argentina; * These authors equally contributed to this work; ** corresponding author (E-mail: lualque@iib.uba.ar).

### Introduction

Photosensitizers such as xanthene derivatives (*e.g.*, phloxine B) and porphyrins (*e.g.*, hematoporphyrin IX) are endowed with photoinsecticidal properties. Xanthene derivatives showed acute photo-toxicity against several dipterans (Ben Amor and Jori, 2000, Berni *et al*., 2002, 2003). Pujol-Lereis (2006) compared the photo-toxic effect of phloxine B (PhB) during the postembryonic development of *D. melanogaster* (*Dm*), *Haematobia irritans*, and *Ceratitis capitata*, and determined that *Dm* is affected during larval development. Low concentrations of hematoporphyrin (HP) rapidly decrease adult survival rates of *Ceratitis capitata*, *Bactrocera (Dacus) oleae*, and *Stomoxys calcitrans* (Ben Amor *et al.*, 1998, 2000).

Photosensitizer molecules react upon absorption of visible radiation with the subsequent formation of reactive oxygen species, mediating signaling cascades which either fortify antioxidant defenses of cells or switch to apoptotic death if oxidative pressure is too great (Girotti, 1998). There are two mechanisms by which the photosensitizer can react with biomolecules: type I reactions produce highly reactive oxygen species (*e.g.*, the superoxide and the peroxide anions) which usually activate enzymatic antioxidant defense, and type II reactions result in the formation of singlet oxygen, leading mainly to lipid peroxidation. Studies of the effects in immature stages were carried

# 4 CAPÍTULO II

# O retrotransposon *copia* de *Drosophila* e sua importância para o entendimento da regulação e do impacto dos elementos de transposição nos genomas eucarióticos

**Nathalia de Setta e Claudia Marcia Aparecida Carareto**

UNESP – Universidade Estadual Paulista, Departamento de Biologia, Rua Cristóvão Colombo, 2265, Jardim Nazareth, São José do Rio Preto, SP, Brasil.

## Resumo

O retrotransposon *copia* foi um dos primeiros elementos de transposição identificados em *Drosophila melanogaster* e, desde sua descoberta, tem sido alvo de diversos estudos que objetivaram compreender sua história evolutiva, o impacto de sua presença e os fatores que determinam a sua dinâmica transposicional e manutenção nos genomas de *Drosophila*. Amplamente distribuído na família Drosophilidae, o retrotransposon *copia* produz variabilidade genética observável em nível fenotípico e é regulado por sistemas que controlam explosões transposicionais, promotoras de efeitos deletérios, entre eles, o mecanismo de RNA de interferência. Desse modo, esta revisão tem como objetivo sintetizar o conhecimento acumulado sobre esse elemento e discutir novos caminhos para a busca do entendimento do papel desse retrotransposon nos genomas hospedeiros.

## Palavras-chave

Retrotransposon *copia* – *Drosophila* – elementos de transposição – *Drosophila melanogaster*

## 4.1 Introdução

O retrotransposon *copia* tem sido amplamente estudado ao longo das últimas três décadas com o objetivo de se compreender sua estrutura, funcionamento e evolução em espécies do gênero *Drosophila*, mais especificamente em *Drosophila melanogaster*. Dentre os estudos pioneiros destaca-se o de Finnegan et al. (1978) que descreve a história da descoberta dos "genes" *copia* a partir de um plasmídeo (*Dm1142*), integrante de uma biblioteca de cDNAs, que apresentava 30 a 40 sítios de hibridação em cromossomos politênicos de *D. melanogaster*. Os autores reportaram a presença de seqüências "terminais redundantes" de cerca de 300 pb localizadas nas extremidades dos mRNAs produzidos. Essas repetições corresponderiam às repetições terminais longas (LTRs) tão bem caracterizadas atualmente e de extrema importância para a regulação desse retrotransposon. De acordo com os autores, as LTRs propiciariam a mobilização por meio de eventos de recombinação e/ou da excisão do complexo, seguidos pela replicação e reintegração em outra posição cromossômica. No entanto, várias ressalvas foram feitas, pois os autores consideraram que aquela discussão era de natureza especulativa, pois existiam "somente evidências indiretas da existência de elementos de transposição nos genomas eucarióticos" fazendo referência aos artigos de McClintock (1957) e Homuzi e Tonegawa (1976), que relatavam a presença de elementos com capacidade de mobilização em milho e camundongo, respectivamente.

Desde então, tem-se demonstrado que o retrotransposon *copia* é um componente medianamente repetido do genoma de *D. melanogaster* (BOWEN; MCDONALD, 2001; MUGNIER et al., 2008) e que se mobiliza em tecidos e estágios específicos do desenvolvimento podendo produzir variabilidade genética perceptível fenotipicamente. Essa mobilização é regulada pelo hospedeiro por meio de mecanismos que ainda não são completamente compreendidos. Nesta revisão, apresentamos uma síntese do conhecimento sobre a estrutura, dinâmica funcional, impacto e evolução do retrotransposon *copia* em espécies de *Drosophila*, como também, procuramos encontrar possíveis respostas para questões ainda não solucionadas, tais como: (i) qual seria sua origem no gênero *Drosophila*? (ii) que mecanismo confere ao *copia* uma expressão estádio-específica e qual seria sua importância na relação retrotransposon-hospedeiro? e, (iii) como ocorre a regulação da transposição pelos organismos hospedeiros, impedindo altas taxas de transposição? Essas questões serão discutidas focando a importância da regulação mediada pelo hospedeiro para a dinâmica transposicional do retrotransposon *copia* em espécies do gênero *Drosophila*.

## 4.2 Classificação e Distribuição

A mais recente classificação dos elementos de transposição (WICKER et al., 2007) inclui o retrotransposon *copia* na Classe I e Ordem LTR, entre os elementos que se mobilizam via um RNA intermediário e apresentam LTRs, respectivamente. Mais especificamente, esse retrotransposon nomeia a superfamília *copia*, caracterizada pelo arranjo das enzimas do domínio poliprotéico na seqüência Integrase-Transcriptase reversa-RNAseH. Além do elemento *copia*, a superfamília agrupa outros retrotransposons presentes nos principais ramos dos eucariotos, como por exemplo, *1731* de *Drosophila* (PERONNET et al., 1986), *Ty1* de levedura (CAMERON et al., 1979), *Tpa1* do anfíbio *Pyxicephalus adpersa* (FLAVELL et al., 1995), *Tcs1* de iguana (FLAVELL et al., 1995), *Tnt1* de tabaco (GRANDBASTIEN et al., 1989), *Retrosol* de batata (MANETTI et al., 2007) e *Retrolyc1* de tomate (COSTA et al., 1999).

Vários estudos foram realizados para delimitar a presença de *copia* na família Drosophilidae (Tabela 1, Figura 1). Os resultados demonstraram que *copia* é amplamente distribuído no gênero *Drosophila*, no entanto, foi impossível identificá-lo nas drosófilas havaianas (gênero *Idyomia*), no grupo *virilis* e *Chymomyza procnenis*. Essa distribuição heterogênea pode ser mais ampla, pois existem diversos grupos de *Drosophila* onde a ocorrência desse retrotransposon não foi ainda investigada. Adicionalmente, *copia* está ausente em outras espécies de dípteros analisadas, tais como em *Musca domestica*, *Sarcophaga bullata* e *Ephydra cineria* (MARTIN et al., 1983), bem como, nas mariposas *Bombyx mori* e *Antheraea yamamai* e nas abelhas *Apis mellifera* e *Apis cerana* (KIMURA et al., 1993).

Três hipóteses têm sido formuladas para explicar a descontinuidade na distribuição de *copia* em Drosophilidae: (1) falha na detecção devido à divergência da seqüência do retrotransposon, (2) presença do elemento no ancestral seguida por perda estocástica em algumas espécies e, (3) ausência do retrotransposon no ancestral, e introdução horizontal nas espécies onde foi possível sua identificação (MARTIN et al., 1983). Se esse retrotransposon estivesse ausente no ancestral da família, duas hipóteses alternativas ainda poderiam ser propostas: primeiro, *copia* poderia ter sido inserido em Drosophilidae por um evento de transferência horizontal (HT) a partir de um doador externo, e se disseminado por transferência vertical, horizontal ou introgressão e; segundo, sua introdução poderia ter acontecido por repetidos eventos de HT com um ou vários doadores externos. Os estudos realizados até o momento não permitiram inferir quais dessas hipóteses se aplicam

preferencialmente aos passos iniciais da sua diversificação na família Drosophilidae, entretanto, há evidências de que a história desse elemento foi marcada tanto por eventos de transmissão vertical como de HT para os grupos melhor caracterizados, isto é, os grupos *melanogaster* e *repleta* de *Drosophila* e o gênero *Zaprionus* (JORDAN; MCDONALD, 1998; JORDAN et al., 1999; ALMEIDA; CARARETO, 2006; DE SETTA et al., dados não publicados).

### 4.3 Estrutura

A seqüência completa do retrotransposon *copia* apresenta cerca de 5 kb e foi inicialmente isolada em *D. melanogaster* por Emori et al. (1985). É similar aos retrovírus, sendo composta por LTRs em suas extremidades, uma região transcrita não-traduzida (ULR, do Inglês, *Untranslated Leader Region*) e uma única ORF (do Inglês, *Open Reading Frame*) que codifica todas as proteínas necessárias à transposição (Figura 2A). Sua inserção é aleatória e gera uma duplicação de 5 pb no sítio alvo do genoma hospedeiro (DUNSMUIR et al., 1980). A região codificante pode ser dividida em dois domínios protéicos diferenciados: *gag* e *pol*. O domínio *gag*, localizado na região 5' da ORF, codifica uma proteína semelhante aos capsídeos virais (VLPs, do Inglês, *virus-like particles*) e uma protease aspártica (ATHAUDA et al., 2006), enquanto o domínio poliprotéico (*pol*) codifica três sítios enzimáticos responsáveis pela replicação: a integrase (DNA endonuclease), a transcriptase reversa e a ribonuclease H.

As LTRs do elemento canônico de *D. melanogaster* estão caracterizadas na Figura 2B e na Tabela 2. Essas seqüências apresentam 276 pb de comprimento e são subdivididas em três regiões U3, R e U5 (Figura 2B). Tanto a ULR, quanto as LTRs, abrigam seqüências reguladoras típicas de genes eucarióticos que intensificam ou reprimem a expressão de maneira estádio e tecido-específica. Dentre os motivos reguladores presentes na região 5' LTR-ULR, estão sítios básicos da transcrição e tradução, como o sítio de iniciação da transcrição, o sinal de poliadenilação e a região de ligação do iniciador da tradução ($RNAt^{Met}$), que delimitam as regiões de produção de RNAm e polipeptídeos (MOUNT; RUBIN, 1985). Outros sítios de ligação de fatores reguladores também já foram identificados na região da ULR, como por exemplo, a díade simétrica intensificadora homóloga ao vírus SV40 (*Simiam vírus 40*), sítios de ligação das proteínas reguladoras *DmC/EBP - CBF-1*, *9.2.1AB* e *BBF-2* e homeodomínios de ligação de proteínas *Antennapedia* e *Engrailed* (CAVAREC; HEIDMANN, 1993; WILSON et al., 1998). Adicionalmente, na região U3

existem duas repetições palindrômicas imperfeitas, que são 60 a 70% homólogas a um sítio de ligação de um fator de transcrição da RNA polimerase II presente na região promotora do gene *hsp-70* (*heat shock*), os quais podem estar relacionados à regulação de *copia* em condições de estresse ambiental (STRAND; MCDONALD 1985).

A estrutura básica de *copia* tem sido observada nos elementos isolados até o momento, no entanto, existe variabilidade inter- e intra-específica, principalmente na região LTR-ULR (JORDAN; McDONALD, 1998; ALMEIDA; CARARETO 2006). Por exemplo, existem três subfamílias de *copia* em oito espécies grupo *melanogaster* - *Full-length* (*copia* canônico), *ULR-gap* e *Double-gap*. As três subfamílias podem ser diferenciadas estruturalmente pela presença de duas duplicaçãoes de 39 pb e 28pb, na LTR e ULR, respectivamente, sendo que a subfamília *Full-length* apresenta as duas duplicações, a *ULR-gap* possui apenas a duplicação na LTR e a *Double-gap* não possui qualquer das duas seqüências duplicadas (MCDONALD et al., 1997; JORDAN; McDONALD, 1998). Tal variabilidade promove a duplicação da repetição imperfeita e da díade simétrica homóloga ao SV40 (Figura 2B). O processo que originou essas variantes ainda não foi claramente entendido, mas mecanismos dependentes de homologia, como derrapagem da polimerase durante a replicação, recombinação desigual entre cromossomos homólogos ou ainda erros na incorporação de nucleotídeos durante a transcrição reversa podem ter sido os responsáveis pelas duplicações (JORDAN; McDONALD, 1998). Adicionalmente, as relações filogenéticas e a divergência nucleotídica entre as três subfamílias indicam que *Double-gap* é a mais ancestral, seguida pela *ULR-gap* e *Full length*. Análises da capacidade de intensificação da expressão pelas diferentes variantes indicam que as subfamílias mais recentes possuem promotores mais eficazes (MATYUNINA et al., 1996).

### 4.4 Expressão e Mobilização

A replicação do elemento *copia* envolve a produção de uma única fita de RNAm que sofre *splicing* alternativo e codifica todos os polipeptídios necessários para sua mobilização (MILLER et al., 1989; YOSHIOKA et al., 1990). A maior parte dos transcritos isolados equivale a duas moléculas de 5 e 2 kb de comprimento (Figura 2A; FLAVELL et al., 1981) que correspondem a cerca de 3% do RNA poliadenilado de culturas celulares de *D. melanogaster* (YOUNG; SCHWARTZ, 1980). A molécula de RNAm de 5 kb corresponde ao trecho entre as regiões R das LTRs 5’ e 3’ e a segunda molécula, de 2 kb, equivale ao RNAm completo, que sofre *splicing* de uma região de cerca de 3 kb correspondente ao domínio *pol*

(MILLER et al., 1989; YOSHIOKA et al., 1990; Figura 2A). Adicionalmente, dois tipos de transcritos (com 1,3 kb e 0,8 kb) foram identificados em baixas concentrações (YOUNG; SCHWARTZ, 1980), mas ainda não se sabe se são produtos de eventos alternativos de processamento do RNAm ou se desempenham alguma função durante o mecanismo de transposição.

As moléculas de RNAm têm a capacidade de codificar no mínimo 8 polipeptídeos de 18 a 51 kDa (FLAVELL et al., 1980). Contudo, a função de apenas algumas dessas proteínas puderam ser caracterizadas. Estudos demonstraram que o RNAm de 2 kb expressa um polipeptídio de 48 kDa correspondente à *gag* (YOSHIOKA et al., 1990), que se auto-catalisa gerando duas proteínas com cerca de 33 e 23 kDa (ATHAUDA et al., 2006). A proteína de 33 kDa corresponde à região N-terminal do polipeptídeo precursor e tem função de VLP (Figura 3). Essas partículas foram primeiramente isoladas a partir do núcleo de linhagens celulares de *D. melanogaster* (SHIBA; SAIGO, 1983; MIYAKE et al., 1987). A proteína de 23 kDa é uma protease aspártica, com sítio ativo Asp-Ser-Gly, correspondente à região C-terminal do precursor, que desempenha a função auto-catalisadora do polipeptídeo (ATHAUDA et al., 2006). Tal enzima pertence à família das proteases retrovirais e é mais estável que outras proteases comuns, como a pepsina A humana e a protease 1 do HIV. A VLP e a protease são expressas em maior concentração do que as proteínas replicativas, provavelmente devido à complementação gerada pela tradução do RNAm de 5 kb, que apesar de menos eficientemente, também parece expressar os produtos da *gag*, além da produção das enzimas do domínio *pol* (BRIERLEY; FLAVELL, 1990).

A conservação estrutural de *copia* em relação aos retrovírus indica que os mecanismos de transcrição reversa e integração devem seguir o mesmo modelo geral. A mobilização depende da geração de uma molécula de DNA linear a partir do RNAm de 5 kb produzido pela transcriptase reversa. Após a produção da fita de DNA linear, a integrase será responsável pela inserção no genoma hospedeiro. Nos retrovírus, esse estágio é denominado DNA proviral, pois não apresenta características infecciosas. Estudos com linhagens celulares de *D. melanogaster* permitiram identificar no retrotransposon *copia* a presença de formas extracromossômicas lineares (FLAVELL, 1984) e circulares com uma ou duas LTRs (FLAVELL; ISH-HOROWICZ, 1981; ILYIN et al., 1984), como as produzidas pelos retrovírus. No entanto, os resultados indicam que a transcrição reversa circular produz produtos incompletos, apresentando apenas uma das LTRs (FLAVELL; ISH-HOROWICZ, 1981). Adicionalmente, DNA extracromossômico circular pode ser originado por eventos de recombinação homóloga entre LTRs (ILYIN et al., 1984).

### 4.5 Impacto no genoma

Os elementos de transposição podem promover alterações nos genomas hospedeiros, promovendo os clássicos exemplos de rearranjos cromossômicos (KIDWELL et al., 1977; LANGLEY et al., 1988; MONTGOMERY et al., 1991) ou alterando os genes do hospedeiro ao nível estrutural (O'HARE et al., 1984; ALMEIDA et al., 2008; SAKAI et al., 2007; LOPES et al., 2008) e regulador (AQUADRO et al., 1986; JORDAN et al., 2003). O retrotransposon *copia* já foi identificado participando de pelo menos três eventos de mutação no genoma de *D. melanogaster*.

A primeira descrição de uma mutação causada pela inserção de um elemento transponível em *Drosophila* referiu-se a uma linhagem de *D. melanogaster* com o fenótipo *white-apricot*, que produzia indivíduos com diminuição da pigmentação nos omatídeos, variando de vermelho para selvagens a amarelado para mutantes (BINGHAM; JUDD, 1981; GOLDBERG et al., 1982), e redução da pigmentação nos testículos e túbulos de Malphighi (RABINOW et al., 1993). O genótipo mutante decorria da inserção de *copia* no segundo íntron do gene *white* (Figura 4A; O'HARE et al., 1984). Essa inserção levou à produção de transcritos com comprimentos alterados, provavelmente por causa dos sinais reguladores nas LTRs do retrotransposon (Figura 4A; LEVIS et al., 1984). Adicionalmente, foi observada a excisão dessa inserção por meio de recombinação homóloga intracromossômica entre as LTRs, levando à reversão da mutação. No entanto, nem a excisão do retrotransposon, nem a reversão do fenótipo foram completas, pois uma das LTRs permaneceu no sítio original e os indivíduos revertentes apresentaram omatídeos mais claros que os selvagens. A manutenção de uma das LTRs durante o processo de excisão resulta na manutenção de seqüências reguladoras do TE no gene hospedeiro, que podem interferir no funcionamento gênico e serem domesticadas pelo hospedeiro produzindo variabilidade genética (CARBONARE; GEHRING, 1985).

O gene da *álcool-desidrogenase* (*Adh*) também pode ser diferencialmente expresso no genoma de *D. melanogaster* devido à presença do retrotransposon *copia* (AQUADRO et al., 1986). Uma linhagem mutante com inserção do *copia* a cerca de 250 pb antes do sítio de iniciação da transcrição, teve a expressão da *Adh* reduzida de 2 a 10 vezes, causando um leve efeito deletério nos portadores da inserção, principalmente nos estágios de desenvolvimento em que o retrotransposon é altamente expresso, isto é, em suas larvas do segundo e terceiro estádio, pupas jovens e adultos maduros (Figura 4B; STRAND; MCDONALD, 1989).

Outro exemplo de modificação dos genótipos hospedeiros em decorrência da inserção dos retrotransposons *copia* e *gypsy* foi observado no loco *ovo* de *D. melanogaster* (MÉVEL-NINIO et al., 1989). Esse loco codifica um fator de transcrição que, entre outras funções, está envolvido na manutenção da linhagem germinativa da fêmea. Fêmeas mutantes *ovo* exibiam uma degeneração de células germinativas no início do desenvolvimento apresentando ovários atrofiados e incapacidade de produção de ovos (OLIVER et al., 1987). A inserção de *copia*, mas principalmente do retrotransposon *gypsy*, nesse sítio citológico induziu à reversão dos mutantes, levando à produção de alelos com perda de função, isto é, apresentaram esterilidade da fêmea apenas de modo recessivo (MÉVEL-NINIO et al., 1989). Alelos revertentes também ocorreram de maneira espontânea para essa mutação. Apenas o mapa de restrição desse loco está disponível para as linhagens onde *copia* foi identificado (Figura 4C), pois este estudo foi realizado com experimentos de hibridação *in situ*. Desse modo, a posição exata e a estrutura da inserção para o loco *ovo* não é conhecida.

### 4.6 Regulação

Existe uma relação positiva quando se compara o número de inserções e a taxa de transposição de *copia* para mais de uma linhagem de laboratório (NUZHDIN et al., 1996; PASYUKOVA et al., 1998). O aumento da taxa de mobilização, concomitante ao alto número de inserções, poderia produzir efeitos deletérios e redução do valor adaptativo do hospedeiro, o que sugere que a auto-regulação não é fenômeno predominante na dinâmica transposicional de *copia*, pois, nesse caso, seria esperado um equilíbrio entre número de inserções e taxa de transposição (CHARLESWORTH et al., 1992). Dessa forma, três hipóteses foram propostas para explicar relação positiva entre taxa / número de inserções em *D. melanogaster*. A primeira hipótese postula a presença de inserções mutantes com transposição mais eficiente, presentes em diferentes freqüências nas linhagens analisadas, alterando o equilíbrio entre a taxa de transposição e o número de inserções. Nesse caso, o aumento da taxa de transposição acarretaria diretamente num efeito aditivo do número de inserções ao longo do tempo. A segunda propõe que algum mecanismo da maquinaria do hospedeiro seria responsável pelos diferentes níveis de regulação da transposição e a variabilidade genética no hospedeiro proporcionaria respostas diferenciadas causando variação na taxa de transposição. Finalmente, a terceira, considera que o controle da transposição seria mediado pela freqüência de inserções, e estas influenciariam o processo transposicional, sendo que linhagens com mais

cópias produziriam uma maior quantidade de proteínas necessárias à transposição, aumentando a probabilidade de mobilização (NUZHDIN et al., 1996; PASYUKOVA et al., 1998).

Essas três hipóteses têm sido testadas nos últimos anos e os resultados têm indicado que o controle da transposição do retrotransposon *copia* é um processo multifatorial, envolvendo fatores intrínsecos do retrotransposon, como por exemplo, a conservação de sítios de ligação a proteínas e a produção de diferentes formas de fitas de RNAm e, igualmente importantes, os mecanismos de controle do hospedeiro. Assim, fatores como o número efetivo, a conservação estrutural das inserções e os genótipos específicos da linhagem hospedeira seriam importantes nesse processo (NUZHDIN et al., 1998). Dentre os fatores determinantes da mobilização, a maquinaria enzimática do hospedeiro tem sido observada participando das fases reguladoras pré-transcricional e pós-transcricional (Figura 5). Nos próximos dois subitens serão discutidos os mecanismos que levam à intensificação e a repressão da transcrição, conseqüentemente regulando a produção de novas inserções do retrotransposon.

### 4.6.1 Regulação pré-transcricional

De acordo com Nuzhdin et al. (1996) e Pasyukova et al. (1997), o elemento *copia* apresenta altas taxas de transposição apenas em machos em *D. melanogaster* e essa especificidade se deve aos maiores níveis de produção de transcritos nos testículos em relação aos ovários. Esses dados sugerem que a regulação deva ocorrer principalmente em nível transcricional (FILATOV et al., 1998). A razão desse padrão de mobilização ainda não foi claramente determinada, mas é possível que as fêmeas apresentem algum tipo de mecanismo específico de repressão da transposição, ausente em machos.

Sabe-se também que a produção de RNAm é variável entre as diversas fases do desenvolvimento de *D. melanogaster*, sendo que os maiores níveis de produção de transcritos ocorrem a partir do estádio pupal, passando pelos estágios pré-meióticos do desenvolvimento das gônadas, isto é, em espermatócitos com maturação primária e, finalmente, em machos envelhecidos cerca de 25 dias (FLAVELL et al., 1980; PARKHURST; CORCES, 1987; PASYUKOVA et al., 1997; FILATOV et al., 1998). Essa transposição dependente da idade corrobora a proposta de que existem fatores do hospedeiro regulando a transposição do retrotransposon *copia* (FILATOV et al., 1998).

Apesar do conhecimento de que a regulação de *copia* acontece de modo tecido e idade específicos, somado à identificação de sítios de ligação de fatores de transcrição na região da LTR-ULR 5', ainda não está claro como ocorre e quais fatores realmente contribuem para a regulação da produção de transcritos em *D. melanogaster*. Contudo, alguns genes de *Drosophila* têm sido associados à regulação por meio de estudos com o sistema *white-apricot* e de transformações em linhagens celulares, principalmente genes relacionados à regulação gênica (CAVAREC; HEIDMANN, 1993; BHADRA et al., 1999; PAL-BHADRA et al., 1998; CAVAREC et al., 1994; TULIN et al., 2002).

As homeoproteínas reguladoras *engrailed*, *even-skipped* e *zerknüllt* regulam negativamente a expressão de *copia* em linhagens celulares de *Drosophila hydei*, que não possuem esse retrotransposon. Por outro lado, as homeoproteínas *zerknullt* e *fushi tarazu* regulam positivamente a sua expressão em linhagens celulares de *D. melanogaster*. Dessa forma, a regulação do retrotransposon *copia* por esta classe de reguladores é dependente do tipo de homeoproteína, do *background* celular e da concentração dos reguladores (CAVAREC; HEIDMANN, 1993). Adicionalmente, os sítios de ligação das homeoproteínas estão próximos à díade simétrica que compartilha similaridade com o intensificador do SV40, indicando que as diferentes subfamílias *copia* podem interagir de modo diferenciado com esses reguladores.

O loco *mls* (*male specific lethal*), associado à compensação de dose do cromossomo X de machos, também deve participar do mecanismo de controle transcricional do *copia*, desde que mutantes para esse loco apresentam um significativo aumento na expressão desse retrotransposon (BHADRA et al., 1999). Já os locos reguladores *Inr-a*, *Wow* e *Mow* intensificam o padrão de expressão da mutação *white-apricot*, de modo diferenciado, de acordo com a combinação de genótipos de cada um destes três genes (PAL-BHADRA et al., 1998). A influência exercida por pelo menos esses três fatores reguladores ocorre do mesmo modo em genes do hospedeiro, indicando que esses locos são responsáveis pela produção de fatores reguladores epistáticos e que trabalham para a manutenção dos mecanismos moleculares básicos da transcrição (PAL-BHADRA et al., 1998).

Cavarec et al. (1994) isolaram um fator de transcrição, ainda não-caracterizado, de aproximadamente 50 kD, que ativa intensamente a transcrição de *copia*. Essa proteína se liga às regiões reguladoras de ULR de linhagens celulares de *D. hydei*, mas está ausente em linhagens de *D. melanogaster*. Os autores discutem que a presença dessa proteína poderia ser um fator evolutivamente limitante para a manutenção do retrotransposon em espécies do gênero *Drosophila*, como em *D. hydei*, onde o *copia* canônico está naturalmente ausente, se o

elemento fosse introduzido por transferência horizontal. A intensificação da transcrição poderia levar ao aumento excessivo de retrotransposições e conseqüentemente a restrições funcionais devido ao potencial mutacional dos elementos de transposição.

Outra proteína que influencia a expressão do retrotransposon *copia* é *PARP* (do Inglês, *Poly ADP-ribose polymerase*), que tem como função geral a organização estrutural da cromatina durante o desenvolvimento de *Drosophila*, mantendo a integridade do genoma (TULIN et al., 2002). Mais especificamente, *PARP* influencia a expressão e silenciamento de blocos de eucromatina e heterocromatina apresentando atividade mais pronunciada no início do desenvolvimento. Mutantes *Parp* hiperexpressam o RNAm *copia* de 5 kb, indicando que esta proteína possui alguma função repressora no nível estrutural, isto é, na conformação do DNA, inibindo a expressão desse retrotransposon.

Fatores ambientais como estresse térmico, químico e exposição à radiação também podem ativar o retrotransposon *copia* em *D. melanogaster* (STRAND; MCDONALD, 1985). Após 30 minutos de estresse térmico a 37 °C, a expressão do RNAm de 5 kb foi até 13 vezes mais alta que nos controles, num padrão similar aos genes *heat shock*. Os responsáveis por essa intensificação podem ser os elementos *heat shock* na região U3 da LTR (Tabela 2, Figura 2B). Além disso, esse trabalho demonstrou que o estresse químico pela exposição a peróxido de hidrogênio, e azida de sódio, também induzem a expressão de RNAm *copia*, sobretudo a molécula de 2 kb de comprimento, contudo o mecanismo relacionado à indução não é bem compreendido.

### 4.6.2 Regulação pós-transcricional

O mecanismo de interferência de RNA (RNAi, do Inglês, *RNA interference*) tem fundamental importância na regulação pós-transcricional, promovendo o silenciamento gênico e a defesa do hospedeiro contra os efeitos deletérios da presença de vírus e elementos de transposição por meio da degradação de RNAm, metilação e acetilação do DNA. Em *Drosophila* têm sido descritos três mecanismos básicos de RNAi com maquinaria e alvos da regulação diferenciados: (i) o mecanismo de microRNA, relacionado à repressão da tradução de genes por meio da degradação de RNAs maduros; (ii) o sistema de degradação de mRNAs antivirais, que é direcionado por RNAs dupla-fita pequenos (siRNAs) produzidos a partir de RNAs virais e de retrotransposons, e; (iii) o mecanismo de silenciamento de elementos de transposição da linhagem germinativa mediado pelos piRNAs, RNAs pequenos que interagem com as proteínas da subfamília *piwi* (OBBARD et al., 2009).

Dentre os fatores chave responsáveis pelo funcionamento dos mecanismos de RNAi estão as famílias multigênicas *Dicer* e *Argonauta* (HAMMOND et al., 2001; BERNSTEIN et al., 2001; OKAMURA et al., 2004). As proteínas *Dicer* têm como função a maturação dos RNAs pequenos de fita dupla que atuam no silenciamento e as proteínas *Argonauta* se ligam às fitas de RNA picotadas por *Dicer* desempenhando atividade endonucleolítica supressora degradando os RNAs alvo (atividade *Slicer*) ou, ainda, funcionam como uma plataforma para a reunião de proteínas do complexo de silenciamento (HAMMOND, 2005).

Os pequenos RNAs podem ser classificados em diversos tipos, de acordo com seu tamanho, com as proteínas que participam da sua produção e com o tipo de RNA alvo para o silenciamento. Dentre os diferentes pequenos RNAs, estão os rasiRNAs (de 24 e 25 pb; do Inglês, repeat *associated small interfering RNAs*), também denominados como piRNAs (do Inglês, *piwi-interacting RNA*) ou esiRNAs (do Inglês, *endogenous short interfering RNA*), relacionados ao silenciamento de elementos de transposição, entre eles, o retrotransposon c*opia* (ARAVIN et al., 2003; KAWAMURA et al., 2008). Em *Drosophila*, a degradação dos RNAs, e posterior inativação dos TEs, ocorre por vias diferenciadas nas células da linhagem somática e germinativa (Figura 5B).

Na linhagem germinativa de *Drosophila*, o processo de silenciamento de RNA é mediado pelas proteínas *Argonautas* da subfamília *Piwi* (*Ago-3*, *Aubergine* e *piwi*). A subfamília *piwi* é independente de *Dicer* agindo por meio da interação com piRNAs. Esse processo ocorre durante estágios iniciais de desenvolvimento do embrião (0-2 horas), em testículos (ARAVIN et al., 2003) e ovários de adultos (SAITO et al., 2006) e se inicia com um aumento na expressão de *Ago-3*, *Aub* e *piwi* (WILLIANS; RUBIN, 2002), que por sua vez, utilizarão as seqüências dos retrotransposons, inclusive *copia*, como molde para a produção de piRNAs e posterior degradação dos RNAm (ARAVIN et al., 2003). Análises de expressão *in situ* têm demonstrado que mutantes *piwi* inibem a repressão do retrotransposon *copia* em células da linhagem germinativa de machos (KALMYKOVA et al., 2005). Além disso, foi demonstrado que a LTR é a região alvo da maquinaria de silenciamento, devido a sua superexpressão nos mutantes *piwi*, quando ligada a um gene repórter.

Outro gene da família *Argonauta* envolvido no mecanismo de RNAi na linhagem germinativa é *spindle-E* (VAGIN et al., 2006). O gene *spindle-E* codifica uma RNA helicase DexH-box (GILLESPIE; BERG, 1995) que participa do desdobramento e/ou anelamento do RNA do complexo de RNAi (STAPLETON et al., 2001). Foi demonstrado que em células germinativas de fêmeas (VAGIN et al., 2004) e machos (STAPLETON et al., 2001) esse gene

participa do controle da expressão do retrotransposon *copia*, desde que a expressão desse TE aumenta até 5,7 vezes em mutantes *spindle-E*.

Adicionalmente, existe uma via de RNAi silenciadora de elementos de transposição expressos na linhagem somática de *D. melanogaster* (Figura 5B; KAWAMURA et al., 2008). Esse mecanismo envolve a proteína *Ago-2* que interage com os esiRNAs produzidos a partir de dsRNAs (do Inglês, *double-strand RNA*) pela proteína *Dicer-2*. Sugere-se que os esiRNAs guiam *Ago-2* na clivagem de transcritos nas células, incluindo RNAs de elementos de transposição. Dentre os RNAs de TEs utilizados como molde para a formação de esiRNAs, está o retrotransposon *copia*.

### 4.7 Conclusões e Perspectivas

As informações sobre a estrutura e o funcionamento do retrotransposon *copia* na família Drosophilidae, sobretudo em *D. melanogaster*, apresentados nesta revisão, ilustram o papel fundamental dos elementos transponíveis como promotores de variabilidade genética, reforçando a importância do seu estudo para seu próprio conhecimento, bem como para o entendimento do funcionamento e evolução dos genomas em que são abrigados. Não obstante o grande volume de dados produzido, ainda existem lacunas a serem preenchidas para a compreensão da evolução do elemento *copia*, em especial sobre sua origem no gênero *Drosophila* e o papel da maquinaria celular dos hospedeiros na regulação de sua mobilização.

Uma dessas lacunas refere-se à descontinuidade da distribuição do retrotransposon *copia* em *Drosophila*, que poderia ser decorrente de falhas na detecção do elemento em algumas espécies, devido à divergência entre as seqüências, transferência horizontal ou perda estocástica, ou talvez, devido a todos esses eventos acontecendo simultaneamente e/ou ocasionalmente, durante sua evolução. Desde que seqüências dos grupos *melanogaster* e *repleta*, divergentes em cerca de 70% no nível nucleotídico, puderam ser isoladas com os iniciadores disponíveis na literatura (ALMEIDA; CARARETO, 2006), e que métodos de hibridação, amplamente empregados na busca por esse TE, foram utilizados para identificar seqüências com relativa divergência (MARTIN et al., 1983; STACEY et al., 1986; DE FRUTOS et al., 1992; FRANCINO et al., 1993), a hipótese de falha na detecção pode não ser realmente importante para explicar sua ausência em algumas espécies.

As análises das seqüências de diversos elementos de transposição, entre eles o retrotransposon *gypsy*, *mariner* e o elemento *P*, entre outros (LORETO et al., 2008), demonstraram que eventos de HT são freqüentes em *Drosophila*. Para *copia*, foram propostos

ao menos dois eventos pontuais de HT, entre *D. melanogaster* e *Drosophila willistoni* (JORDAN et al., 1999) e, entre *Zaprionus tuberculatus* e uma espécie desconhecida do grupo *melanogaster* ou *D. willistoni* (ALMEIDA; CARARETO, 2006). No entanto, não existem hipóteses da ocorrência de eventos de HT envolvendo os ancestrais dos grupos atuais para explicar a ampla distribuição desse retrotransposon, seguida pela ausência em alguns grupos. Devido à identificação de *copia* em pelo menos uma espécie de cada grande clado da família Drosophilidae, e da falta de evidência sobre a ocorrência de eventos de HT nos ancestrais dos grupos de espécies, parece-nos que a hipótese mais parcimoniosa para a evolução de *copia* seria a presença do retrotransposon no ancestral dessa família. Adicionalmente, ao menos três eventos de perda poderiam ter ocorrido, sugeridos pela ausência no grupo *virilis*, nas drosófilas havaianas e no gênero *Chymomyza*. O seqüenciamento e o estudo dos fatores que regulam a transposição em espécies ainda não estudadas podem oferecer indicativos dos processos que levaram à perda nesses grupos de *Drosophila*. Vários processos podem ser vislumbrados, tais como: (1) eventos randômicos; (2) altas taxas de mutação, com conseqüente degradação de seqüências codificantes e reguladoras, levando à produção de proteínas truncadas e/ou inviabilizando a ligação de fatores de transcrição; (3) perda ou diferenciação de proteínas reguladoras do hospedeiro que participam da dinâmica de mobilização de *copia*, ou ainda; (4) aquisição de proteínas que alterariam o equilíbrio número de inserções / taxa de transposição, levando ao aumento do potencial mutacional de *copia* e posterior eliminação por seleção natural. Essa última hipótese foi sugerida baseada no trabalho experimental de Cavarec et al. (1994), devido a identificação do ativador de 50 kDa de *D. hydei*, que está ausente em *D. melanogaster*. Alternativamente, mais de um mecanismo poderia estar agindo nos diferentes grupos de *Drosophila*, individualmente, ou mesmo simultaneamente. Portanto, o estudo da distribuição e dos mecanismos evolutivos que agem nesse retrotransposon é fundamental para que a hipótese de ancestralidade pontuada por eventos de HT e perda estocástica possa ser corroborada.

Os mecanismos de regulação do hospedeiro são essenciais para o controle da transposição dos TEs em *Drosophila*, pois afetam sua manutenção ou perda, sobretudo no caso do retrotransposon *copia*. O controle da transposição pelo hospedeiro pode acontecer pela intensificação, inibição ou mesmo silenciamento da expressão e acarreta num controle do número de inserções, impedindo os efeitos deletérios de uma explosão transposicional, o que geraria um desequilíbrio na relação TE - hospedeiro. Ao mesmo tempo, acontece uma compensação por parte do retrotransposon, levando à maximização de sua atividade, seja pela especificidade de mobilização em tecidos germinativos, o que é indispensável para que as

novas inserções sejam herdáveis e passíveis de fixação nas populações, como pela seleção de subfamílias que sejam transcricionalmente mais eficazes, como aconteceu com as espécies do grupo *melanogaster*. Um modo de aprofundar o conhecimento sobre os fatores que modulam especificamente o retrotransposon *copia* seria analisar os genes e as proteínas envolvidas na regulação pelo hospedeiro em linhagens que apresentam altas taxas de mobilização desse elemento em particular, utilizadas anteriormente para investigar a instabilidade e variabilidade genética entre linhagens de *D. melanogaster* (BIÉMONT et al., 1987). O estudo deveria focar tanto os fatores de transcrição que interferem na produção de transcritos, em diferentes níveis, diferentes tecidos e estádios de desenvolvimento, quanto as proteínas responsáveis pelo mecanismo de RNAi dos tecidos somáticos e germinativos, o que permitiria investigar a ocorrência de fatores reguladores relacionados à manutenção do equilíbrio TE – hospedeiro, que controlariam a instabilidade genômica gerada pelos elementos de transposição.

O conhecimento relatado nesta revisão, e as questões que continuam em aberto sobre sua dinâmica evolutiva e formas de regulação, permitem eleger o elemento *copia* como um bom modelo para a investigação das relações TE – hospedeiro, de modo a se melhor compreender a dinâmica responsável pela fixação e eliminação desses promotores de variabilidade genética dos genomas procarióticos e eucarióticos.

**Referências**

ALMEIDA, L. M.; CARARETO, C. M. Sequence heterogeneity and phylogenetic relationships between the *copia* retrotransposon in *Drosophila* species of the *repleta* and *melanogaster* groups. **Genetics, Selection, Evolution**, França, v. 38, p. 535-550, 2006.

ALMEIDA, L. M. et al. Report of a chimeric origin of transposable elements in a bovine-coding gene. **Genetics and Molecular Research**, Brasil, v. 1, p. 107-116, 2008.

ALONSO-GONZÁLEZ, L.; DOMÍNGUEZ, A.; ALBORNOZ, J. Structural heterogeneity and genomic distribution of *Drosophila melanogaster* LTR-retrotransposons. **Molecular Biology and Evolution**, EUA, v. 20, p. 401-409, 2003.

AQUADRO, C. F. et al. Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. **Genetics**, EUA, v. 114, p. 1165-1190, 1986.

ARAVIN, A. A. et al. The small RNA profile during *Drosophila melanogaster* development. **Developmental Cell**, EUA, v. 5, p. 337-350, 2003.

ATHAUDA, S. B. et al. Isolation and characterization of recombinant *Drosophila copia* aspartic proteinase. **The Biochemical Journal**, Inglaterra, v. 399, p. 535-542, 2006.

BACHMANN, A. S. et al. A simple method for the rapid purification of *copia* virus-like particles from *Drosophila* Schneider 2 cells. **Journal of Virological Methods**, Holanda, v. 115, p. 159-165, 2004.

BERNSTEIN, E.; DENLI, A. M.; HANNON, G. J. The rest is silence. **RNA**, EUA, v. 7, p. 1509-1521, 2001.

BHADRA, U.; PAL-BHADRA, M.; BIRCHLER, J. A. Role of the *male specific lethal* (*msl*) genes in modifying the effects of sex chromosomal dosage in *Drosophila*. **Genetics**, EUA, v. 152, p. 249-268, 1999.

BIÉMONT, C.; AOUAR, A.; ARNAULT, C. Genome reshuffling of the *copia* element in an inbred line of *Drosophila melanogaster*. **Nature**, Inglaterra, v. 329, p. 742-744, 1987.

BIÉMONT, C.; GAUTIER, C. Localisation polymorphism of *mdg-1*, *copia*, *I* and *P* mobile elements in genomes of *Drosophila melanogaster*, from data of inbred lines. **Heredity**, Inglaterra, v. 60, p. 335-346, 1988.

BINGHAM, P. M.; JUDD, B. H. A copy of the *copia* transposable element is very tightly linked to the *Wa* allele at the *white* locus of *D. melanogaster*. **Cell**, EUA, v. 25, p. 705-711, 1981.

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

BRIERLEY, C.; FLAVELL, A. J. The retrotransposon *copia* controls the relative levels of its gene products post-transcriptionally by differential expression from its two major mRNAs. **Nucleic Acids Research**, Inglaterra, v. 18, p. 2947-2951, 1990.

BROOKFIELD, J. F.; MONTGOMERY, E.; LANGLEY, C. H. Apparent absence of transposable elements related to the *P* elements of *Drosophila melanogaster* in other species of *Drosophila*. **Nature**, Inglaterra, v. 310, p. 330-332, 1984.

CAMERON, J. R.; LOH, E. Y. ; DAVIS, R. W. Evidence for transposition of dispersed repetitive DNA families in yeast. **Cell**, EUA, v. 16, p. 739-751, 1979.

CARBONARE, B. D.; GEHRING, W. J. Excision of *copia* element in a revertant of the *white-apricot* mutation of *Drosophila melanogaster* leaves behind one long-terminal repeat. **Molecular & General Genetics**, Suíça, v. 199, p. 1-6, 1985.

CASTRO, J. P. de; de SETTA, N.; CARARETO, C. M. A. Distribution and insertion numbers of transposable elements in species of the *Drosophila saltans* group. **Genetics and Molecular Biology**, Brasil, v. 29, p. 384-390, 2006.

CAVAREC, L.; HEIDMANN, T. The *Drosophila copia* retrotransposon contains binding sites for transcriptional regulation by homeoproteins. **Nucleic Acids Research**, Inglaterra, v. 21, p. 5041-5049, 1993.

CAVAREC, L.; JENSEN, S.; HEIDMANN, T. Identification of a strong transcriptional activator for the *copia* retrotransposon responsible for its differential expression in *Drosophila hydei* and *melanogaster* cell lines. **Biochemical and Biophysical Research Communications**, EUA, v. 203, p. 392-399, 1994.

CAVAREC, L. et al. Molecular cloning and characterization of a transcription factor for the *copia* retrotransposon with homology to the BTB-containing lola neurogenic factor. **Molecular and Cellular Biology**, EUA, v. 17, p. 482-494, 1997.

CHARLESWORTH, B.; LAPID, A.; CANADA, D. The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. 2. inferences on the nature of selection against elements. **Genetical Research**, Inglaterra, v. 60, p. 115-130, 1992.

COSTA, A. P. et al. *Retrolycl-1*, a member of the *Tntl* retrotransposon super-family in the *Lycopersicon peruvianum* genome. **Genetica**, Holanda, v. 107, p. 65-72, 1999.

CSINK, A.K.; LINSK, R.; BIRCHLER, J. A. Mosaic suppressor, a gene in *Drosophila* that modifies retrotransposon expression and interacts with *zeste*. **Genetics**, EUA, v. 136, p. 573-583, 1994.

DE FRUTOS, R.; PETERSON, K. R.; KIDWELL, M. G. Distribution of *Drosophila melanogaster* transposable element sequences in species of the *obscura* group. **Chromosoma**, Alemanha, v. 101, p. 293-300, 1992.

DUNSMUIR, P. et al. Insertion of the *Drosophila* transposable element *copia* generates a 5 base pair duplication. **Cell**, EUA, v. 21, p. 575-579, 1980.

EMORI, Y. et al. The nucleotide sequences of *copia* and *copia*-related RNA in *Drosophila* virus-like particles. **Nature**, Inglaterra, v. 315, p. 773-776, 1985.

FILATOV, D. A.; MOROZOVA, T. V.; PASYUKOVA, E. G. Age dependence of the *copia* transposition rate is positively associated with *copia* transcript abundance in a *Drosophila melanogaster* isogenic line. **Molecular & General Genetics**, Alemanha, v. 258, p. 646-654, 1998.

FINNEGAN, D. J. et al. Repeated gene families in *Drosophila melanogaster*. **Cold Spring Harbor Symposia on Quantitative Biology**, EUA, v. 42, p. 1053-1063, 1978.

FLAVELL, A. J. Role of reverse transcription in the generation of extrachromosomal *copia* mobile genetic elements. **Nature**, Inglaterra, v. 310, p. 514-516, 1984.

FLAVELL, A. J.; ISH-HOROWICZ, D. Extrachromosomal circular copies of the eukaryotic transposable element *copia* in cultured *Drosophila* cells. **Nature**, Inglaterra, v. 292, p. 591-595, 1981.

FLAVELL, A. J. et al. Translation and developmental regulation of RNA encoded by the eukaryotic transposable element *copia*. **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 77, p. 7107-7111, 1980.

FLAVELL, A. J. et al. The 5' termini of RNAs encoded by the transposable element *copia*. **Nucleic Acids Research**, Inglaterra, v. 9, p. 6279-6291, 1981.

FLAVELL, A. J. et al. *Ty1-copia* group retrotransposon sequences in amphibia and reptilia. **Molecular & General Genetics**, Alemanha, v. 246, p. 65-71, 1995.

FRANCINO, O.; CABRE, O.; FONTDEVILA, A. Distribution of the *copia* transposable element in the *repleta* group of *Drosophila*. **Genetics, Selection, Evolution**, França, v. 25, p. 501-516, 1993.

GILLESPIE, D. E.; BERG, C. A. *homeless* is required for RNA localization in *Drosophila* oogenesis and encodes a new member of the DE-H family of RNA-dependent ATPases. **Genes & Development**, EUA, v. 9, p. 2495-2508, 1995.

GOLDBERG, M. L.; PARO, R.; GEHRING, W. J. Molecular cloning of the *white* locus region of *Drosophila melanogaster* using a large transposable element. **The EMBO Journal**, Inglaterra, v. 1, p. 93-98, 1982.

GRANDBASTIEN, M. A.; SPIELMANN, A.; CABOCHE, M. *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. **Nature**, Inglaterra, v. 337, p.376-380, 1989.

GREWAL, S.I.; ELGIN, S. C. Transcription and RNA interference in the formation of heterochromatin. **Nature**, Inglaterra, v. 447, p. 399-406, 2007.

HAMMOND, S. M. et al. *Argonaute2*, a link between genetic and biochemical analyses of RNAi. **Science**, EUA, v. 293, p. 1146-1150, 2001.

HAMMOND, S. M. Dicing and slicing: the core machinery of the RNA interference pathway. **FEBS Letters**, Holanda, v. 579, p. 5822-5829, 2005.

HOMUZI, N.; TONEGAWA, S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 73, p. 3628-3632, 1976.

ILYIN, Y. V. et al. Circular copies of mobile dispersed genetic elements in cultured *Drosophila melanogaster* cells. **Nucleic Acids Research**, Inglaterra, v. 12, p. 7517-7531, 1984.

JORDAN, I. K.; MCDONALD, J. F. *Evolution of the copia retrotransposon in the Drosophila melanogaster* species subgroup. **Molecular Biology and Evolution**, EUA, v. 15, p. 1160-1171, 1998.

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

JORDAN, I. K. et al. Origin of a substantial fraction of human regulatory sequences from transposable elements. **Trends in Genetics**, Inglaterra, v. 19, p. 68-72, 2003.

KALMYKOVA, A. I.; KLENOV, M. S.; GVOZDEV, V. A. Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. **Nucleic Acids Research**, Inglaterra, v. 33, p. 2052-2059, 2005.

KAWAMURA, Y. et al. *Drosophila* endogenous small RNAs bind to *Argonaute 2* in somatic cells. **Nature**, Inglaterra, v. 453, p. 793-797, 2008.

KIDWELL, M. G.; KIDWELL, J. F.; SVED, J. A. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including, sterility and male recombination. **Genetics**, EUA, v. 86, p. 813-833, 1977.

KIKUCHI, Y.; ANDO, Y.; SHIBA, T. Unusual priming mechanism of RNA-directed DNA synthesis in *copia* retrovirus-like particles of *Drosophila*. **Nature**, Inglaterra, v. 323, p. 824-826, 1986.

KIMURA, K. et al. Transposable elements in commercially useful insects: I. Southern hybridization study of silkworms and honeybees using *Drosophila* probes. **Japanese Journal of Genetics**, v. 68, p. 63-71, 1993.

LANGLEY, C.H. et al. On the role of unequal exchange in the containment of transposable element copy number. **Genetical Research**, Inglaterra, v. 52, p. 223-235, 1988.

LEVIS, R.; O'HARE, K.; RUBIN, G. M. Effects of transposable element insertions on RNA encoded by the *white* gene of *Drosophila*. **Cell**, EUA, v. 38, p. 471-481, 1984.

LOPES, F. R. et al. Transposable elements in *Coffea* (Gentianales: Rubiacea) transcripts and their role in the origin of protein diversity in flowering plants. **Molecular Genetics and Genomics**, Alemanha, v. 279, p. 385-401, 2008.

LORETO, E. L.; CARARETO, C. M.; CAPY, P. Revisiting horizontal transfer of transposable elements in *Drosophila*. **Heredity**, Inglaterra, v. 100, p. 545-54, 2008.

MANETTI, M. E. et al. Radiation of the *Tnt1* retrotransposon superfamily in three *Solanaceae* genera. **BMC Evolutionary Biology**, Inglaterra, v. 7, p. 34, 2007.

MARKOW, T. A.; O'GRADY, P. M. ***Drosophila*: A Guide to Species Identification and Use.** Londres: Academic Press, 2005.

MARTIN, G.; WIERNASZ, D.; SCHEDL, P. Evolution of *Drosophila* repetitive-dispersed DNA. **Journal of Molecular Evolution**, Alemanha, v. 19, p. 203-213, 1983.

MASIDE, X. et al. Rates of movement and distribution of transposable elements in *Drosophila melanogaster*: *in situ* hybridization vs. Southern blotting data. **Genetical Research**, Inglaterra, v. 78, p. 121-136, 2001.

MATYUNINA, L. V.; JORDAN, I. K.; MCDONALD, J. F. Naturally occurring variation in *copia* expression is due to both element (cis) and host (trans) regulatory variation. **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 93, p. 7097-7102, 1996.

MCCLINTOCK, B. Controlling elements and the gene. **Cold Spring Harbor Symposia on Quantitative Biology**, EUA, n. 21, p. 197-216, 1957.

MCDONALD, J. F. et al. LTR retrotransposons and the evolution of eukaryotic enhancers. **Genetica**, Holanda, v. 100, p. 3-13, 1997.

MÉVEL-NINIO, M.; MARIOL, M. C.; GANS, M. Mobilization of the *gypsy* and *copia* retrotransposons in *Drosophila melanogaster* induces reversion of the *ovo* dominant female-sterile mutations: molecular analysis of revertant alleles. **The EMBO Journal**, Inglaterra, v. 8, p. 1549-1558, 1989.

MILLER, K. et al. The nucleotide sequence of *Drosophila melanogaster copia*-specific 2.1-kb mRNA. **Nucleic Acids Research**, Inglaterra, v. 17, p. 2134, 1989.

MIYAKE, T. et al. Production of virus-like particles by the transposable genetic element, *copia*, of *Drosophila melanogaster*. **Molecular & General Genetics**, Alemanha, v. 207, p. 29-37, 1987.

MONTGOMERY, E. A. et al. Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. **Genetics**, EUA, v. 129, p. 1085-1098, 1991.

MOUNT, S. M.; RUBIN, G. M. Complete nucleotide sequence of the *Drosophila* transposable element *copia*: homology between *copia* and retroviral proteins. **Molecular and Cellular Biology**, EUA, v. 5, p. 1630-1638, 1985.

MUGNIER, N. et al. The heterochromatic copies of the LTR retrotransposons as a record of the genomic events that have shaped the *Drosophila melanogaster* genome. **Gene**, Holanda, v. 411, p. 87-93, 2008.

NUZHDIN, S. V.; PASYUKOVA, E. G.; MACKAY, T. F. Positive association between *copia* transposition rate and copy number in *Drosophila melanogaster*. **Proceedings of the Royal Society of London**, Inglaterra, v. 263, p. 823-831, 1996.

NUZHDIN, S. V. et al. Quantitative genetic analysis of *copia* retrotransposon activity in inbred *Drosophila melanogaster* lines. **Genetics**, EUA, v. 150, p. 755-766, 1998.

OBBARD, D. J. et al. The evolution of RNAi as a defense against viruses and transposable elements. **Philosophical transactions of Royal Society B**, Inglaterra, v. 364, p. 99-115, 2009.

O'HARE, K. et al. DNA sequence of the *white* locus of *Drosophila melanogaster*. **Journal of Molecular Biology**, Inglaterra, v. 180, p. 437-455, 1984.

OKAMURA, K. et al. Distinct roles for *Argonaute* proteins in small RNA-directed RNA cleavage pathways. **Genes & Development**, EUA, v. 18, p. 1655-1666, 2004.

OLIVER, B.; PERRIMON, N.; MAHOWALD, A. P. The *ovo* locus is required for sex-specific germ line maintenance in *Drosophila*. **Genes & Development**, EUA, v. 1, p. 913-923, 1987.

PAL-BHADRA, M.; BHADRA, U.; BIRCHLER, J. A. Role of multiple *trans-acting* regulators in modifying the effect of the retrotransposon *copia* on host gene expression in *Drosophila*. **Molecular & General Genetics**, Alemanha, v. 259, p. 198-206, 1998.

PARKHURST, S. M.; CORCES, V. G. Developmental expression of *Drosophila melanogaster* retrovirus-like transposable elements. **The EMBO Journal**, Inglaterra, v. 6, p. 419-24, 1987.

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

PASYUKOVA, E. G.; NUZHDIN, S. V.; FILATOV, D. A. The relationship between the rate of transposition and transposable element copy number for *copia* and *Doc* retrotransposons of *Drosophila melanogaster*. **Genetical Research**, Inglaterra, v. 72, p. 1-11, 1998.

PERONNET, F. et al. *1731*, a new retrotransposon with hormone modulated expression. **Nucleic Acids Research**, Inglaterra, v. 14, p. 9017-9033, 1986.

RABINOW, L.; CHIANG, S. L.; BIRCHLER, J. A. Mutations at the *Darkener of apricot* locus modulate transcript levels of *copia* and *copia*-induced mutations in *Drosophila melanogaster*. **Genetics**, EUA, v. 134, 1175-1185, 1993.

SAITO, K. et al. Specific association of *Piwi* with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. **Genes & Development**, EUA, v. 20, p. 2214-2222, 2006.

SAKAI, H.; TANAKA, T.; ITOH, T. Birth and death of genes promoted by transposable elements in *Oryza sativa*. **Gene**, Holanda, v. 392, p. 59-63, 2007.

SHIBA, T.; SAIGO, K. Retrovirus-like particles containing RNA homologous to the transposable element *copia* in *Drosophila melanogaster*. **Nature**, Inglaterra, v. 302, p. 119-124, 1983.

STACEY, S. N. et al. Distribution and conservation of móbile elements in the genus *Drosophila*. **Molecular Biology and Evolution**, v. 3, p. 522-534, 1986.

STAPLETON, W.; DAS, S.; MCKEE B. D. A role of the *Drosophila* homeless gene in repression of *Stellate* in male meiosis. **Chromosoma**, Alemanha, v. 110, p. 228-40, 2001.

STRAND, D. J.; MCDONALD, J. F. *Copia* is transcriptionally responsive to environmental stress. **Nucleic Acids Research**, Inglaterra, v. 13, p. 4401-4410, 1985.

STRAND, D. J.; MCDONALD, J. F. Insertion of a *copia* element 5' to the *Drosophila melanogaster* alcohol dehydrogenase gene (*adh*) is associated with altered developmental and tissue-specific patterns of expression. **Genetics**, EUA, v. 121, p. 787-794, 1989.

STROBEL, E.; DUNSMUIR, P.; RUBIN, G. M. Polymorphisms in the chromosomal locations of elements of the *412*, *copia* and *297* dispersed repeated gene families in *Drosophila*. **Cell**, EUA, v. 17, p. 429-439, 1979.

THROCKMORTON, L. H. The phylogeny, ecology and geography of *Drosophila*. In: KING, R.C. (Org.) **Handbook of genetics**. USA: Plenum, p. 421-469, 1975.

TULIN, A.; STEWART, D.; SPRADLING, A. C. The *Drosophila* heterochromatic gene encoding poly(ADP-ribose) polymerase (*PARP*) is required to modulate chromatin structure during development. **Genes & Development**, EUA, v. 16, p. 2108-2119, 2002.

VAGIN, V. V. et al. The RNA interference proteins and *vasa* locus are involved in the silencing of retrotransposons in the female germline of *Drosophila melanogaster*. **RNA Biology**, EUA, v. 1, p. 54-58, 2004.

VAGIN, V. V. et al. A distinct small RNA pathway silences selfish genetic elements in the germline. **Science**, EUA, v. 313, p. 320-324, 2006.

VIEIRA, C.; BIÉMONT, C. *Selection against transposable elements in D. simulans and D. melanogaster*. **Genetical Research**, Inglaterra, v. 68, p. 9-15, 1996.

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

WILLIAMS, R. W.; RUBIN, G. M. ARGONAUTE1 is required for efficient RNA interference in *Drosophila* embryos. **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 99, p. 6889-6894, 2002.

WILSON, S.; MATYUNINA, L. V.; MCDONALD, J. F. An enhancer region within the *copia* untranslated leader contains binding sites for *Drosophila* regulatory proteins. **Gene**, Holanda, v. 209, p. 239-246, 1998.

YAMAGUCHI, O. et al. Distributions of three transposable elements, *P*, *297*, and *copia*, in natural populations of *Drosophila melanogaster*. **The Japanese Journal of Genetics**, Japão, v. 62, p. 205-216, 1987.

YOSHIOKA, K. et al. Virus-like particle formation of *Drosophila copia* through autocatalytic processing. **The EMBO Journal**, Inglaterra, v. 9, p. 535-541, 1990.

YOUNG, M. W.; SCHWARTZ, H. E. Nomadic gene families in *Drosophila*. **Cold Spring Harbor Symposia on Quantitative Biology**, EUA, n. 45, p. 629-640, 1980.

**Tabela 1.** Distribuição do retrotransposon *copia* na família Drosophilidae (+: presente; -: ausente).

| Espécie | Status | Metodologia* | Referências** |
|---|---|---|---|
| Gênero *Chymomyza* | | | |
| *C. procnenis* | + | SB | 1 |
| Gênero *Drosophila* | | | |
| Subgênero *Drosophila* | | | |
| Grupo *guarani* | | | |
| *D. subbadia* | + | SB | 1 |
| Grupo *immigrans* | | | |
| *D. stonei* | + | SB | 1 |
| Grupo *pinicola* | | | |
| *D. pinicola* | + | SB | 1 |
| Grupo *repleta* | | | |
| *D. aldrichi* | + | SB / PCR | 2, 3 |
| *D. anceps* | - | PCR | 3 |
| *D. antonietae* | + | SQ | 3 |
| *D. arizonae* | - /+ | SB / PCR | 2, 3 |
| *D. borborema* | + | SB | 2 |
| *D. buzzatii* | + | SB / SQ | 2, 3 |
| *D. eremophila* | - | PCR | 3 |
| *D. gouveai* | + | SQ | 3 |
| *D. hexastigma* | + | PCR | 3 |
| *D. hydei* | - | SB / PCR | 1, 2, 3 |
| *D. hydeoides* | - | SB | 2 |
| *D. huaylasi* | + | SB | 2 |
| *D. koepferae* | + | SB / SQ | 2, 3 |
| *D. longicornis* | - | PCR | 3 |
| *D. martensis* | + | SB | 2 |
| *D. mercatorum* | + | SB / PCR | 1, 2, 3 |
| *D. mettleri* | - | PCR | 3 |
| *D. mojavensis* | + | SB / SQ | 2, 3 |
| *D. mulleri* | + | SB / PCR | 2, 3 |
| *D. navojoa* | + | PCR | 3 |
| *D. nigrodumosa* | + | SB | 2 |
| *D. pachuca* | + | SQ | 3 |
| *D. paranaensis* | + | PCR | 3 |
| *D. propachuca* | + | PCR | 3 |
| *D. ritae* | - | PCR | 3 |
| *D. serido* | + | SB / SQ | 2, 3 |
| *D. seriema* | + | SQ | 3 |
| *D. spenceri* | + | PCR | 3 |
| *D. starmeri* | + | SB | 2 |
| *D. uniseta* | + | SB | 2 |
| *D. venezolana* | + | SB | 2 |
| *D. wheeleri* | + | SB / PCR | 2, 3 |
| Grupo *virilis* | | | |
| *D. littoralis* | - | SB | 1 |
| *D. montana* | - | SB | 1 |
| *D. virilis* | - | SB | 1 |
| Subgênero *Sophophora* | | | |
| Subgrupo *melanogaster* | | | |
| *D. ananassae* | + | SB | 1 |
| *D. bipectinata* | + | SB | 1 |
| *D. erecta* | + | SQ | 4 |
| *D. mauritiana* | + | SB / SQ | 1, 4, 5 |
| *D. melanogaster* | 17-42 cópias | HIS / SB / SQ | 4, 6, 7, 8, 9, 10, 11 |
| *D. sechellia* | + | SQ | 4 |
| *D. simulans* | 1-4 cópias | HIS / SB / SQ | 1, 4, 5, 9 |
| *D. takahashi* | + | SB | 1 |

**Continuação Tabela 1.**

| Espécie | Status | Metodologia* | Referências** |
|---|---|---|---|
| *D. teissieri* | + | SQ | 4 |
| *D. yakuba* | + | SB / SQ | 1, 4, 5 |
| Subgrupo *obscura* | | | |
| *D. affinis* | + | SB | 1, 14 |
| *D. algoquin* | + | SB | 12 |
| *D. azteca* | + | SB | 12 |
| *D. bifasciata* | + | SB | 12 |
| *D. guanche* | + | SB | 12 |
| *D. miranda* | + | SB | 12 |
| *D. narragansett* | + | SB | 12 |
| *D. obscura* | + | SB | 12 |
| *D. persimilis* | + | SB | 12 |
| *D. pseudoobscura* | >30 cópias | HIS / SB | 1, 5, 12 |
| *D. subobscura* | - | SB | 1 |
| *D. subsilvestris* | + | SB | 12 |
| *D. tristis* | + | SB | 12 |
| *D. tolteca* | + | SB | 12 |
| Subgrupo *saltans* | | | |
| *D. austrosaltans* | 9 cópias | SB | 13 |
| *D. dacunhai* | 12 cópias | SB | 13 |
| *D. emarginata* | 5 cópias | SB | 13 |
| *D. milleri* | 13 cópias | SB | 13 |
| *D. neocordata* | 3 cópias | SB | 1, 13 |
| *D. parasaltans* | 5 cópias | SB | 13 |
| *D. prosaltans* | 9 cópias | SB | 13 |
| *D. saltans* | 10 cópias | SB | 13 |
| *D. sturtevanti* | 6 cópias | SB | 13 |
| *D. subsaltans* | 8 cópias | SB | 13 |
| Subgrupo *willistoni* | | | |
| *D. paulistorum* | + | SB | 1 |
| *D. willistoni* | + | SQ | 14 |
| Gênero *Dorsilopha* | | | |
| *D. busckii* | + | SB | 1 |
| Gênero *Idyomia* | | SB | |
| *D. grimshawi* | - | SB | 1 |
| *D. hawaiiensis* | - | SB | 1 |
| *D. hirtipalpus* | - | SB | 1 |
| Gênero *Hirtodrosophila* | | | |
| *D. duncani* | + | SB | 1 |
| Gênero *Scaptodrosophila* | | | |
| *S. dimorpha* | 2 cópias | HIS | 1 |
| *S. latifasciaeformis* | + | SB | 1 |
| Gênero *Zaprionus* | | | |
| *Z. africanus* | + | SQ | 15 |
| *Z. camerounensis* | + | SQ | 15 |
| *Z. davidi* | + | SQ | 15 |
| *Z. gabonicus* | + | SQ | 15 |
| *Z. indianus* | + | SQ | 15 |
| *Z. tuberculatus* | + | SQ | 15 |
| *Z. vittiger* | Cromocentro | HIS | 1 |

* Metodologias utilizadas na identificação do retrotransposon *copia*: Hibridação *in situ* (HIS), *Southern blot* (SB), PCR e seqüenciamento (SQ). **1: Martin et al. (1983), 2: Francino et al. (1993), 3: Almeida e Carareto (2006), 4: Jordan e McDonald (1998), 5: Brookfield et al. (1984), 6: Strobel et al. (1979), 7: Yamaguchi et al. (1987), 8: Biémont e Gautier (1988), 9: Vieira e Biémont (1996), 10: Maside et al. (2001), 11: Alonso-Gonzáles et al. (2003), 12: de Frutos et al. (1992), 13: Castro et al. (2006), 14: Jordan et al. (1999) e, 15: De Setta et al., dados não-publicados.

**Tabela 2.** Sítios reguladores presentes nas regiões LTR 5' e ULR do retrotransposon *copia* de *D. melanogaster*.

| **Elementos reguladores** | **Região** | **Seqüência (5' → 3')** | **Referencia*** |
|---|---|---|---|
| Elementos *heat shock* | LTR (U3) | CTACAAAAATAACG<br>TTGGAATATACTAT | 1 |
| Repetição imperfeita | LTR (U3) | CACCTTTTAT | 2 |
| Início da transcrição | LTR (U3/U5) | CTTTCCTTCTTGTACGTTTTT | 3 |
| Elemento *downstream* | LTR (R) | CGTG | 3 |
| Sinal de poliadenilação | LTR (R/U5) | AAATATAAATC | 4 |
| Sitio de iniciação da tradução ($tRNA^{Met}$) | ULR | GGTTATGGGCCCAG | 5 |
| Sítio de ligação de *antennapedia* | ULR | ATTA | 6 |
| Sítio de ligação de *engrailed* | ULR | DVAAWTAAAK | 6 |
| Sítios de ligação de DmC/EBP e CBF-1 | ULR | TGTGAAW | 7 |
| Díade simétrica intensificadora (*core SV40*) | ULR | TTTCACA / TGTGAAW | 8 |
| Sítio de ligação de 9.2.1AB | ULR | GAAATAGCTTTTTTTTC | 9 |
| Fator 2 de ligação da caixa B (BBF-2) | ULR | TMNACGTANKC | 7 |

* 1: Strand e McDonald (1985), 2: McDonald et al. (1997), 3: Flavell et al. (1981), 4: Emori et al. (1985), 5: Kikuchi et al. (1986), 6: Cavarec e Heidmann (1993), 7: Wilson et al. (1998), 8: Mount e Rubin (1985), 9: Cavarec et al. (1997).

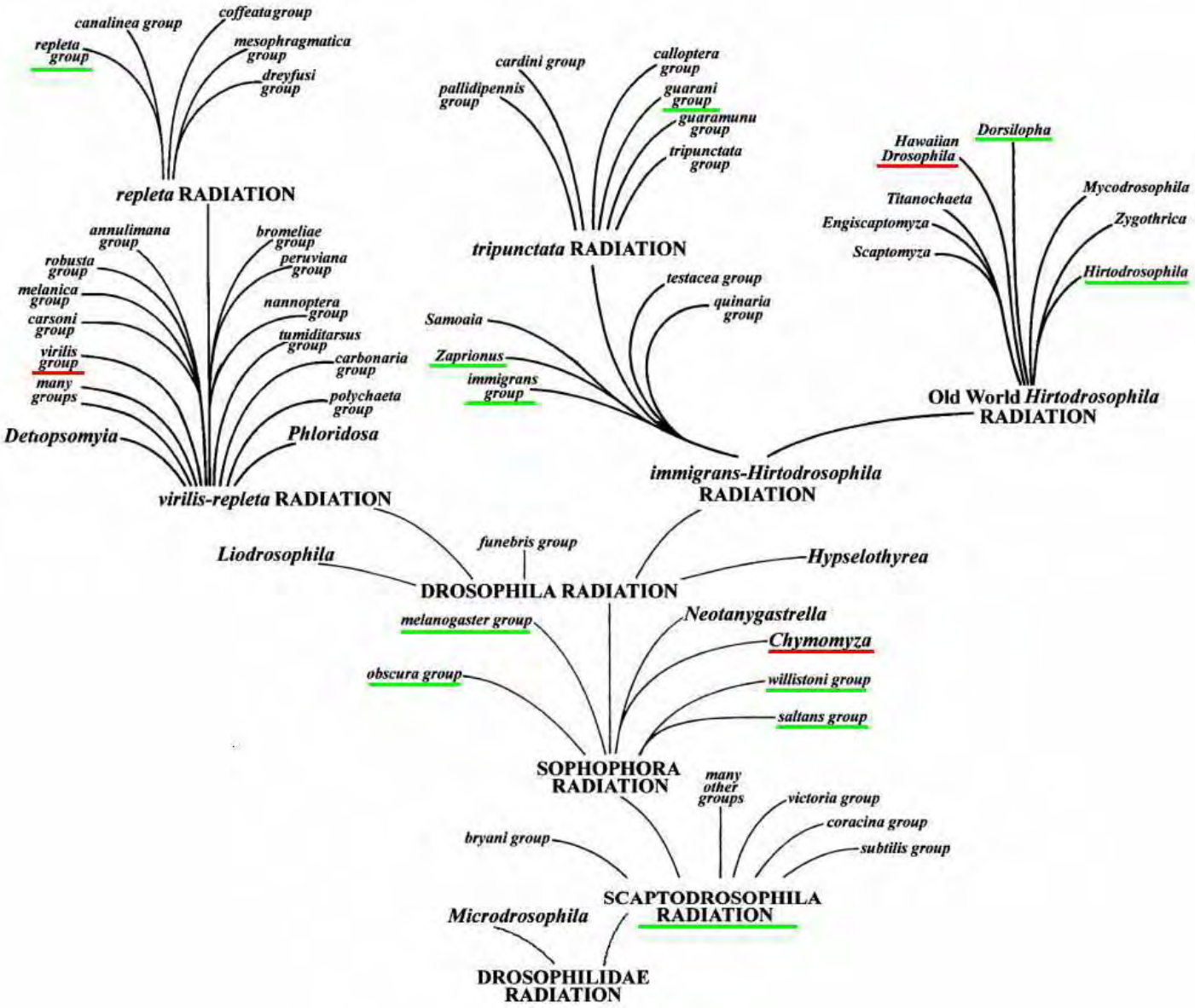



**Figura 1.** Distribuição do retrotransposon *copia* nos principais ramos taxonômicos da família Drosophilidae. Os grupos grifados em verde e vermelho apresentam pelo menos uma espécie onde o retrotransposon *copia* foi estudado e identificado ou não encontrado, respectivamente. Diagrama modificado de Markow e O'Grady (2005) baseado na classificação proposta por Throckmorton (1975).

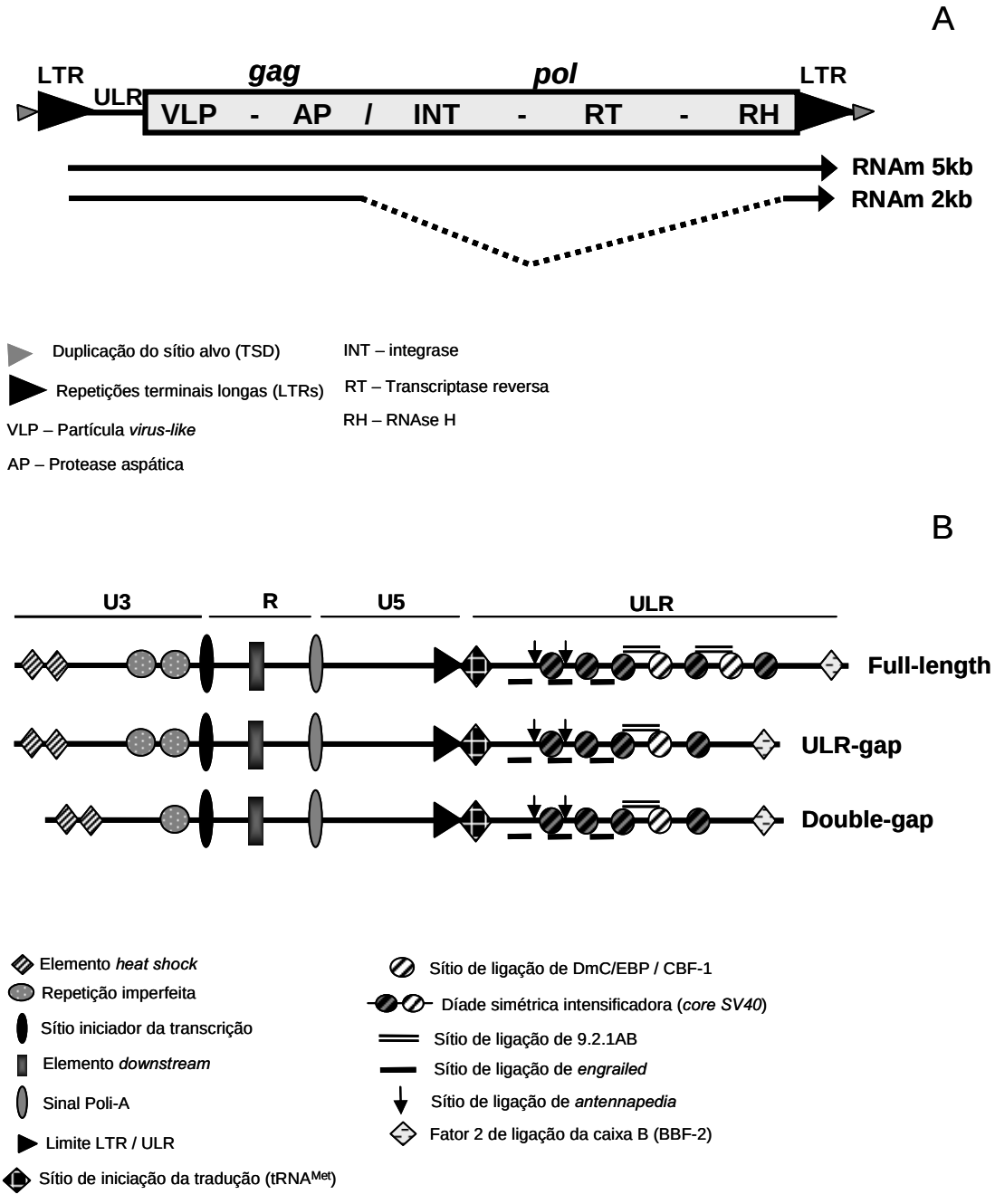



**Figura 2.** Características estruturais do retrotransposon *copia*. A: Diagrama do elemento canônico completo com as regiões reguladoras (LTR e ULR), domínios codificantes e os dois tipos de mRNAs produzidos por meio do *splicing* alternativo (EMORI et al., 1985; MOUNT; RUBIN, 1985). B: Diagrama detalhado da distribuição e localização dos sítios reguladores presentes na LTR-ULR 5' das três subfamílias *copia* do grupo *melanogaster* (FLAVELL et al., 1981; STRAND; MCDONALD, 1985; KIKICHI et al., 1986; CAVAREC; HEIDMANN, 1993; CAVAREC et al., 1997; MCDONALD et al., 1997; WILSON et al., 1998).

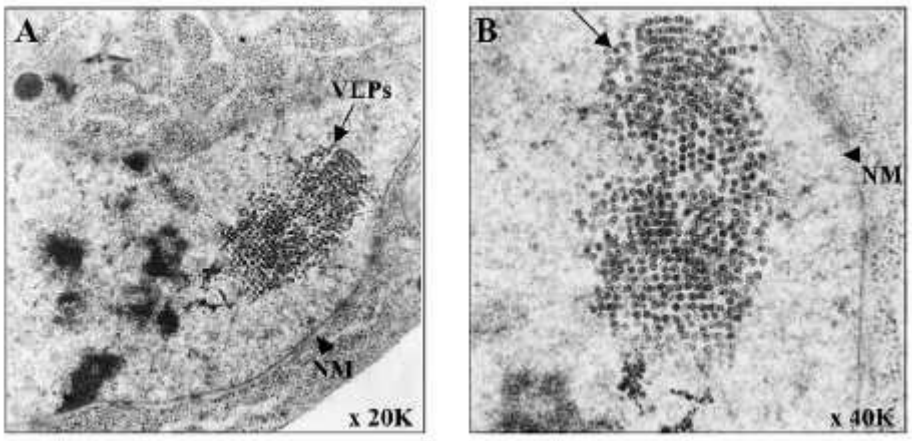



**Figura 3.** Imagem gerada por microscopia eletrônica de transmissão mostrando as VLPs do retrotransposon *copia* em cultura de células da linhagem Schneider 2 (S2) de *D. melanogaster*. As VLPs se apresentam em *clusters* no núcleo das células (seta). NM: membrana nuclear; K: milhares de vezes. Imagem obtida no trabalho de Bachmann et al. (2004).

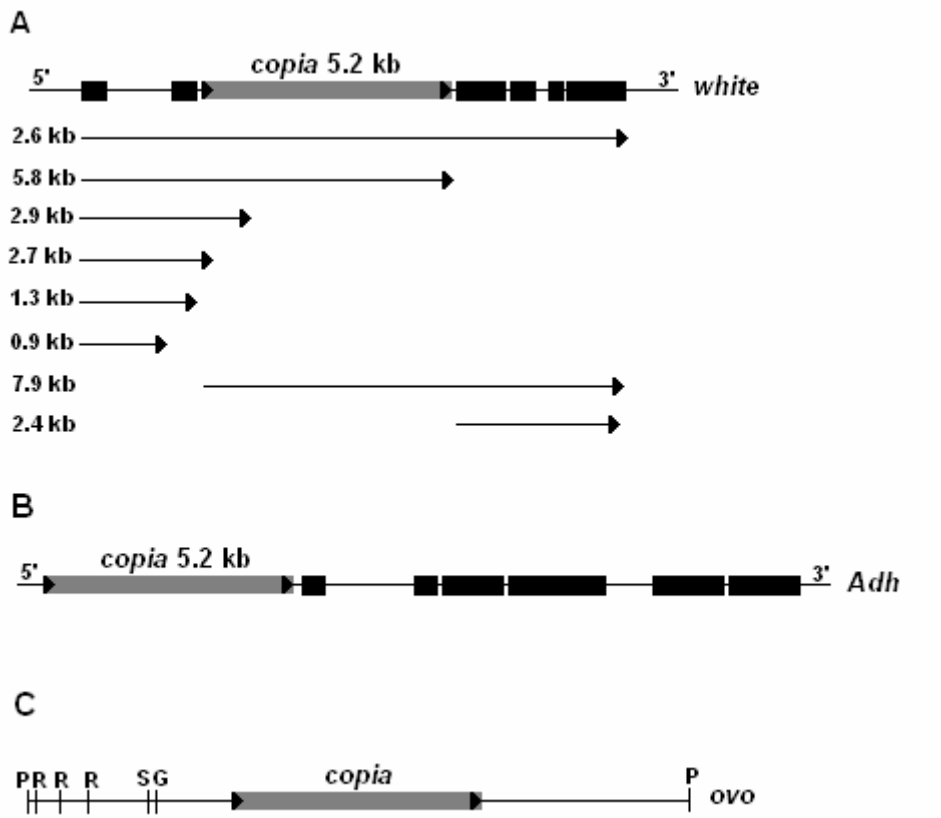



**Figura 4.** Exemplos de genes de *D. melanogaster* com funcionamento alterado devido inserções do retrotransposon *copia*. A: Gene *white* com inserção no segundo intron e conseqüente produção de RNAms alternativos. B: Gene *Adh* com inserção à montante do códon de iniciação da transcrição. C: Inserção de *copia* no loco *ovo*. Caixas pretas, cinzas, setas e traços verticais correspondem aos exons, inserções *copia*, RNAm e sítios de restrição, respectivamente. P: *Pst*I, R: *Eco*RI, S: *Sal*I, G: *Bgl*II. Os esquemas A, B e C foram baseados nos trabalhos de Csink et al. (1994), Strand e McDonald (1989) e Mével-Ninio et al. (1989), respectivamente.

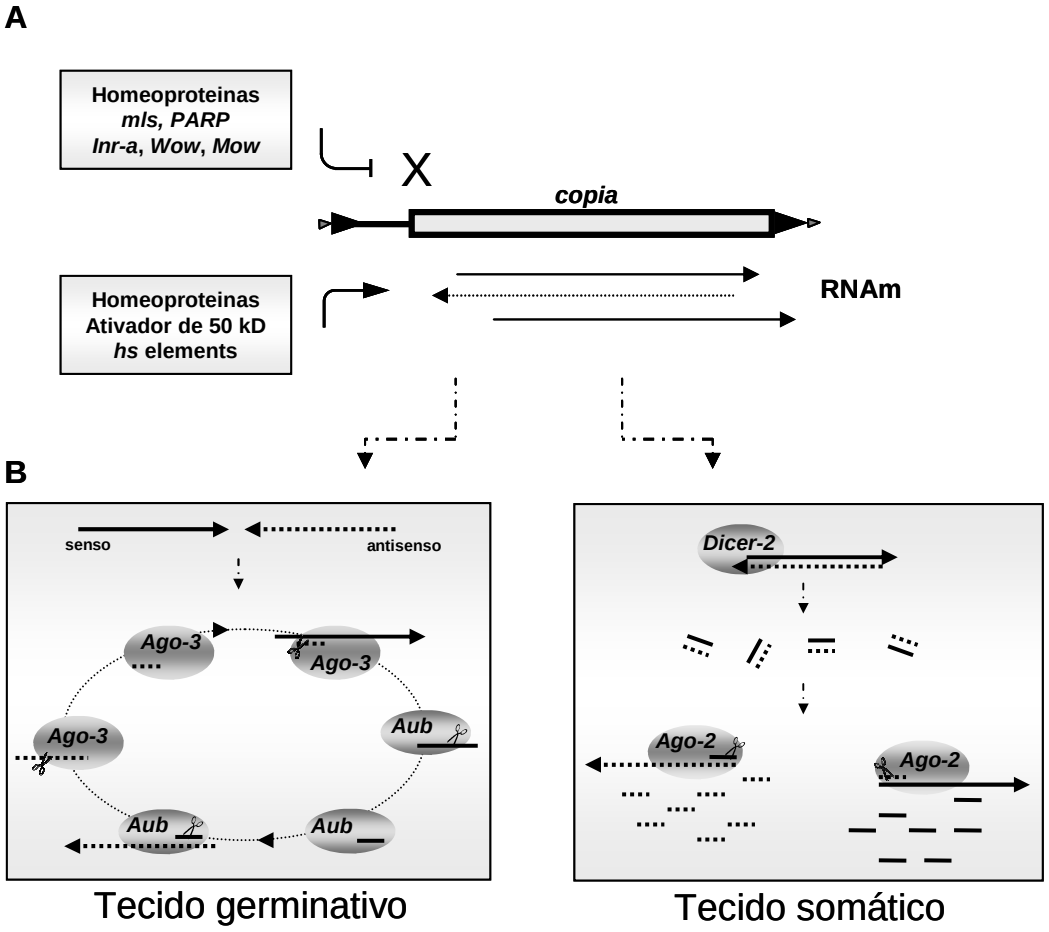



**Figura 5.** Regulação do retrotransposon *copia* pelo hospedeiro. (A) Regulação pré-transcricional: a repressão da transcrição pode ser modulada por homeoproteínas, pelos genes *mls*, *Parp*, *Inr-a*, *Wow* e *Mow*; a ativação da transcrição pode ser regulada também por homeoproteínas, por proteínas *heat shock* e, exclusivamente em *D. hydei*, pelo ativador de 50kD. (B) Regulação pós-transcricional: o retrotransposon *copia* pode ser silenciado pelo mecanismo de RNAi, tanto pelo sistema *piwi* independente de *Dicer* na linhagem germinativa, quanto por *Dicer-2* e *Ago-2* na linhagem somática. Na linhagem germinativa, as fitas simples senso e antisenso de RNAm são picotadas pelas proteínas *Aub* e *Ago-3* e servem como molde para o reconhecimento de outros RNAm que serão degradados promovendo o silenciamento. A função da proteína *piwi* nesse mecanismo seria a ligação aos transcritos nascentes para recrutar fatores da heterocromatina (GREWAL; ELGIN, 2007). Na linhagem somática, os RNAm de fita dupla são reconhecidos e picotados por *Dicer-2* e posteriormente também servem de molde para a degradação de outros RNAms ligados à *Ago-2*.

# 5 CAPÍTULO III

# *Copia* retrotransposon evolutionary history in Drosophilidae family, depicted by the *Zaprionus* genus analysis

**Nathalia de Setta[1], Marie Anne Van Sluys[2], Pierre Capy[3], Claudia Marcia Aparecida Carareto[1]**

[1]UNESP - São Paulo State University, Department of Biology, São José do Rio Preto, SP, Brazil; [2]USP - São Paulo University, Department of Botany, SP, Brazil; [3]Laboratoire Evolution, Génomes et Spéciation UPR9034, CNRS, Gif-sur-Yvette, France.

## Abstract

*Copia* is a retrotransposon largely distributed among the Drosophilidae family but sequences analyses concentrated only on the *Drosophila* genus *repleta* group and *melanogaster* subgroup. In this study, seven species of genus *Zaprionus* were analyzed by RT-PCR and sequencing of the LTR-ULR and reverse transcriptase regions and the results were compared with those reported previously. It was shown that the *copia* element occurs and is transcriptionally active in all species investigated, and a specific *Zaprionus* subfamily, closer related to the most ancient subfamily of the *melanogaster* subgroup, was identified. One ancient horizontal transfer event was suggested between *melanogaster* and *Zaprionus* species, indicating that the *copia* of the latter is a derivative of the *melanogaster* subgroup *copia* retrotransposon.

## Keywords

*copia* retrotransposon, *Drosophila*, *Zaprionus*, horizontal transfer, transposable elements

## 5.1 Introduction

The *copia* retrotransposon is broadly although patchy distributed in the Drosophilidae family (BIÉMONT; CIZERON, 1999). Despite its wide distribution, the evolutionary reports using nucleotide sequences have focused on the LTR-ULR region of the *melanogaster* subgroup and the *repleta* species group of *Drosophila*. The LTR-ULR region contains the transcriptional promoters and enhancers, such as, TATA box, dyad symmetric enhancer and homeoproteins-binding sites, as well the traduction start and stop sites (MOUNT; RUBIN, 1985; CAVAREC; HEIDMANN, 1993; WILSON et al., 1998; ALMEIDA; CARARETO, 2006).

The variability in the LTR-ULR region allowed identifying families and subfamilies of the *copia* element that are dispersing among species through vertical or horizontal inheritance (JORDAN; MCDONALD, 1998a; ALMEIDA; CARARETO, 2006). Within the *repleta* group, *copia* shows low nucleotide distances, however, it is highly divergent compared to the *melanogaster* subgroup sequences. This pattern indicates a *copia* family specific to the *repleta* group, which has been subjected to selective constraints that promoted conservation of regulatory sites in the eight species studied (ALMEIDA; CARARETO, 2006). In the *melanogaster* subgroup, horizontal transfer (HT), as well as positive and purifying selection have promoted the diversification and variable fixation of the Full-length (the most recent), the ULR-gap and the Double-gap (the oldest) subfamilies, differentiated by the duplication of 39 and 28 nt in the LTR and ULR regions, which generate an imperfect repeat and a dyad symmetric enhancer, respectively (JORDAN; MCDONALD, 1998a, b). These subfamilies are discontinuously distributed among seven species of the *melanogaster* subgroup (*Drosophila melanogaster*, *Drosophila simulans*, *Drosophila sechellia*, *Drosophila mauritiana*, *Drosophila yakuba*, *Drosophila teissieri* and *Drosophila erecta*) and each species encompass more than one, as for example, *D. simulans* has the three subfamilies, *D. melanogaster*, the Full-length and the ULR-gap, and *D. sechellia* with only the Double-gap subfamily (JORDAN; MCDONALD, 1998a, b). Additionally, the high levels of conservation and phylogenetic incongruence have been used to infer HT within the *melanogaster* species subgroup (JORDAN; MCDONALD, 1998a; BOWEN; MCDONALD, 2001; SÁNCHEZ-GRACIA et al., 2005), between *D. melanogaster* and *Drosophila willistoni* (JORDAN et al., 1999), and between *Zaprionus tuberculatus* and an unknown species of the *melanogaster* subgroup or *D. willistoni* (ALMEIDA; CARARETO, 2006).

The genus *Zaprionus* Coquillet, 1991 (Drosophilidae) is composed by two monophyletic subgenera, *Anaprionus* s.s. (11 species) originated in the Oriental biogeographic region, and *Zaprionus* s.s. (49 species) from the Afrotropical region (OKADA; CARSON, 1983; YASSIN et al., 2008ab). The relationships within the subgenus *Anaprionus* are so far poorly understood. On the other hand, based on morphological as well as molecular data, a proposition was recently presented regarding the relationships within the subgenus *Zaprionus*, clustering its species into two groups: the group *inermis*, with four complexes (*sexvittatus*, *ghesquieri*, *inermis* and *tuberculatus*), and the group *armatus*, with nine complexes (*litos*, *montanus*, *armatus*, *megalorchis*, *indianus*, *vittiger*, *lachaisei*, *davidi* and *spinous*) (YASSIN et al., 2008a).

The phylogenetic relationships of the *Zaprionus* species within the Drosophilidae family are still a matter of discussion. The genus status outside the genus *Drosophila* has been maintained this far (GRIMALDI, 1990; DE SALLE, 1992), although several studies have clustered *Zaprionus* as a subgenus within the genus *Drosophila*, as originally proposed by Throckmorton (1975). One of the difficulties for including *Zaprionus* in the genus *Drosophila* is the discordance about its exact phylogenetic relationship with the subgenera *Drosophila* and *Sophophora*. Although most molecular marker studies reinforce the idea that *Zaprionus* species are closer related to the *Drosophila* subgenus (PÉLANDAKIS; SOLIGNAC, 1993; KWIATOWSKI et al., 1994; RUSSO et al., 1995; REMSEN; DE SALLE, 1998; KWIATOWSKI; AYALA, 1999; TATARENKOV et al., 1999; DA LAGE et al., 2007; YASSIN et al., 2008a), there are alternative phylogenies clustering them to a clade between *Drosophila* and the *Sophophora* (THOMAS; HUNT, 1993), or closer to the subgenus *Sophophora* (DE SETTA et al., 2008).

As opposed to *Drosophila*, transposable elements have been only occasionally studied in *Zaprionus*, mainly regarding their occurrence, using Southern and *in situ* hybridization analyses (MARUYAMA; HARTL, 1991; MONTCHAMP-MOREAU et al., 1993; CIZERON et al., 1998; DE SETTA; CARARETO, 2007). Few sequencing studies have been performed and were restricted to partial sequences of the *copia* (MCDONALD et al., 1997) and the *mariner* elements (MARUYAMA; HARTL, 1991) in *Z. tuberculatus*, and the *gypsy* retrotransposon in *Zaprionus indianus* (HEREDIA et al., 2004). Interestingly, these three transposable elements have been proposed to be involved in HT events (MARUYAMA; HARTL, 1991; HEREDIA et al., 2004; ALMEIDA; CARARETO, 2006). The species richness and the ecological diversity of genus *Zaprionus* make it stand out as an important model for comparative studies with the *Drosophila* species, aiming to amplify the knowledge

about the evolutionary dynamics of transposable elements in Drosophilidae. Here, a survey was conducted on seven *Zaprionus* species, for analyzing the distribution and evolution of the *copia* retrotransposon and evaluating the inheritance mechanisms responsible for its distribution within Drosophilidae, by comparing the *copia* sequences of genus *Zaprionus* and the *melanogaster* and *repleta* species groups.

## 5.2 Materials and methods

### 5.2.1 Species

Seven species of genus *Zaprionus* were investigated in this study, whose strains were kindly provided by Drs. Jean David and Amir Yassin from LEGS, CNRS, France. The taxonomical classification and geographical origin are shown in Table 1.

### 5.2.2 DNA and PCR reaction

Genomic DNA was extracted from 10 individuals of each strain, using the phenol-chloroform method as described by Jowett (1986). Two pairs of primers were used to amplify and sequence two different regions of the *copia* retrotransposon: primers COP-LTR (5' CTA TTC AAC CTA CAA AAA TAA CG 3') and COP-PCS (5' ATT ACG TTT AGC CTT GTC CAT 3') that amplify 421 bp of LTR-ULR region, and primers ZCopRTF (5' GTT GCA CGA GGA TTC ACT CA 3') and ZCopRTR (5' GCT TGA GTC CGT AAA TTG CC 3') which anneal to region 3306-3558 of the *copia* reverse transcriptase domain (RT), producing a 253 bp fragment in *D. melanogaster* (X04456). The PCR reaction conditions were the following: 200 ng of genomic DNA, 0.4 mM of each dNTP, 7.5 mM $MgCl_2$, 0.4 μM of each primer, 1.5 U of Taq Platinum polymerase (Invitrogen) in 1X PCR buffer. The reactions were heated to 94 °C for 3 min and then submitted to 40 cycles of 30 s at 94 °C, a 1 min step at 55 °C, a 1 min step at 72 °C, and an additional extension step of 10 min at 72 °C. DNA from *D. melanogaster* and ultrapure water were used as positive and negative controls, respectively.

### 5.2.3 Cloning and sequencing

PCR fragments obtained with the COP-LTR / COP-PCS and ZCopRTF / ZCopRTR primers were extracted from agarose gels, purified using the GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) and cloned with the TOPO TA Cloning Kit (Invitrogen). For each primer, five clones chosen randomly were automatically sequenced in an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems/Hitashi).

### 5.2.4 Sequence analysis and phylogenetic relationships

The sequences obtained were manipulated into the BioEdit Sequence Alignment Editor (HALL, 1999) and aligned with ClustalW 1.81 (THOMPSON et al., 1994). The reverse transcriptase sequences were assembled in consensus sequences, in view of their high conservation levels (>95%). Novel regulatory motifs and repetitions were searched using the Alibaba 2.1 (GRABE, 2002) and Tandem Repeats Finder (BENSON, 1999) programs, respectively. The phylogenetic relationships were inferred using the maximum likelihood (ML), neighbor-joining (NJ) and maximum parsimony (MP) methods, as implemented in PhyML 3.0 (GUINDON; GASCUEL, 2003), MEGA 4.1 (TAMURA et al., 2007) and PAUP v.4.0b10 (SWOFFORD, 1997), respectively. Branch support was calculated by bootstrap analysis with 1000 replicates (FELSENSTEIN, 1985). In the NJ and ML analyses, the maximum composite likelihood (TAMURA, 2004) and the HKY85 distances (HASEGAWA et al., 1985) were used to construct the distance matrices and trees. A heuristic search method was used for MP reconstruction. The sequences produced are registered in the GenBank database (Table 1). The *Z. indianus*1, *Z. tuberculatus*1 and *Zaprionus camerounensis*1 clones, which were the most divergent (>25%) LTR-ULR, *D. melanogaster* Full-length sequence (X02599), *D. melanogaster* ULR-gap sequence (U60292), *D. simulans* Double-gap sequence (AF063880) and *Z. indianus* sequence for reverse transcriptase, were selected as queries for searching 12 *Drosophila* genomes available in Flybase (flybase.org/blast/), using the BLASTn tool with cut-off parameters of $e^{-50}$ and 90% coverage, in order to get only highly related subjects of the *Zaprionus* genus. Redundant genomic subjects (100% identical) were not included in the phylogenetic analyses (Supplementary Table 1). *Drosophila koepferae* (AY655745, X96971), *Drosophila buzzatii* (AY655746, X96972), *Drosophila serido* (AY655747), *Drosophila gouveai* (AY655748), *Drosophila seriema* (AY655750), *Drosophila pachuca* (DQ494345), *Drosophila mojavensis* (DQ494346) sequences from

GenBank and the *Z. tuberculatus* sequence published in McDonald et al. (1997) (hereafter *Z. tuberculatus*MD) were used for phylogenetic analysis.

### 5.2.5 RNA extraction and RT-PCR reactions

Heads and gonads from 10 individuals of each sex were dissected in Testis Buffer (183 mM KCl, 47 mM NaCl and 10 mM Tris-HCl pH 6.8). Total RNA was isolated from dissected tissues using TRIZOL (Invitrogen) and genomic DNA was eliminated from the samples by RQ1 RNase-Free DNase (Promega) treatment, according to the manufacturer's instructions. The cDNA pool was generated from total RNAs using the High Capacity cDNA Archive Kit (Applied Biosystems) and random primers at low stringent temperature (37 °C) according to the standart protocol. The primers which amplify the RT fragment (ZCopRTF and ZCopRTR) were used to investigate if *copia* is transcriptionally active in genus *Zaprionus*, using the same conditions applied to the PCR reactions with genomic DNA. Primers ZapGPDHF (5' GTT CGG CAA TTG AAC CAA TG 3') and ZapGPDHR (5' AGA GAG TCC GTG TGC ATG TG 3') that amplify a 337 bp sequence of the *Z. tuberculatus Gpdh* gene (L37039) were used to verify genomic DNA contamination in the total RNA extraction and for the cDNA quality control. These PCR reactions were carried out using 200 ng of total cDNA, 0.1 mM of each dNTP, 0.4 μM of each primer, 2 mM $MgCl_2$ and 1 U Taq Platinum polymerase (Invitrogen) in 1X PCR buffer. The cycling parameters were: 94 °C for 2 min for initial denaturation, 35 cycles of 94 °C for 1 min, 60 °C for 1 min, and 72 °C for 1 min, and an additional extension step at 72 °C for 10 min. Additionally, the *Gpdh* amplicons were directly sequenced, used in the evolutionary analyses and submitted to the GenBank database under the following accession numbers: FJ705445 (*Z. multistriatus*), L37039 (*Z. tuberculatus*), FJ705446 (*Z. camerounensis*), FJ705447 (*Z. davidi*), FJ705448 (*Z. gabonicus*), FJ705449 (*Z. africanus*), FJ705450 (*Z. indianus*).

### 5.2.6 Evolutionary analyses

To test the models of *copia* sequences inheritance in the Drosophilidae, three approaches were utilized: (1) analysis of selective constraints in the RT and LTR-ULR regions, using Tajima's D test to compare the number of segregating sites and the average number of nucleotide differences (TAJIMA, 1989), as implemented by DnaSP 4.50 (ROZAS et al., 2003); (2) inference of the selection model shaping the *copia* RT sequences by

evaluating the number of nonsynonymous substitutions per nonsynonymous sites (dN), the number of synonymous substitutions per synonymous sites (dS), the dN/dS ratios (ω) and the Codon Based Z-test, assuming purifying (dN < dS), or positive (dN > dS) selection as alternative hypotheses, using MEGA 4.1 (TAMURA et al., 2007); and (3) the comparison of dS distances of *copia* and *Gpdh* sequences. The codon-based tests assume that synonymous mutations are under almost strictly neutral evolution and that $\omega < 1$, $\omega = 1$ and $\omega > 1$ represents purifying selection, neutral evolution, and positive Darwinian selection, respectively. Additionally, the *copia* element divergence time was estimated according to the equation $T = k/2r$ (GRAUR; LI, 2000), where $T$ is the divergence time between species, $k$ is the dS divergence between *copia* sequences, and $r$ is the evolutionary rate, using the rate synonymous substitutions with low codon bias in *Drosophila* genes (1.6% substitutions per site per million years, SHARP; LI, 1989).

## 5.3 Results

### 5.3.1 Distribution and transcriptional activity of *copia* retrotransposon in *Zaprionus* species

To study the presence, distribution and transcriptional activity of the *copia* retrotransposon in *Zaprionus* species, RT-PCR reactions were performed using primers that amplify about 400 bp and 250 bp of the 5' LTR-ULR region and the RT domain, respectively. Although it was not possible to amplify the LTR-ULR regulatory region of *Zaprionus multistriatus* and *Zaprionus davidi* (data not shown), the results for the RT domain indicate that transcriptionally active *copia* sequences occur in all *Zaprionus* species (Figure 1). The lack of amplification in those two species may be due to the high divergence of that region, at least at the primers annealing sites, once the LTR-ULR sequences are more variable than the coding sequences for *copia*-like retrotransposons (COSTA et al., 1999).

The analyses of sex and tissue transcriptional specificity showed differences in amplification intensity only in *Z. tuberculatus* and *Z. davidi*, which could indicate lower transcription levels in females than in males. Whilst the RT-PCR technique is not convenient to quantitative analyses, those differences might reflect true sex-specific expression levels of *copia*, since in *D. melanogaster* this expression pattern leads to a male-specific mobilization (NUZHDIN et al., 1996; PASYUKOVA et al., 1997).

### 5.3.2 Characterization and structure of the *copia* regulatory regions of *Zaprionus* species

The LTR-ULR sequences of *Zaprionus* species were compared with the three subfamilies of the *melanogaster* species (Full-length, ULR-gap and Double-gap) and the *repleta* group family. The regulatory signals of the LTR-ULR region of the *repleta* species were not identified in the *Zaprionus* sequences (data not show). The alignment with *melanogaster* sequences showed that most of the basic *copia* regulatory signals are present in the *Zaprionus* sequences, but with some variation that makes them closer to the Double-gap subfamily (Figure 2, Table 2 and Supplementary Figure 1). At least one copy of the heat shock element, the TATA box, the transcriptional start, the downstream element, the poly-A signal, the PBS (primer binding site) and the BBF-2 (B-box binding factor-2) sites are present in all *Zaprionus* species, despite the occurrence of variable sites in some of them, particularly in *Z. tuberculatus*. The imperfect repeat sequences located at the end of the U3 region are absent in all *Zaprionus* species. There are only two repetitions of the *DmC/EBP* and *engrailed* sites, and the dyad symmetric enhancer homologous to SV40 is not present in the *Zaprionus copia* retrotransposons. Moreover, the analyses performed with the Tandem Repeat Finder and Alibaba 2.1 programs allowed identifying a novel regulatory motif in the *Zaprionus* ULR sequences, the G-box binding factor-1 site (GBF-1; consensus sequence: NNGMCACGTS), a leucine zipper described in plants (XIANG et al., 1997), which is, 90% and 70% similar to the signals of *Arabdopsis thaliana* (KLIMCZAK et al., 1992) and *Triticum aestivem* (TABATA et al., 1991), respectively. Only a single motif is present in the ULR of *Z. tuberculatus* (*tuberculatus* complex) and *Z. camerounensis* (*lachaisei* complex), but it is duplicated in *Zaprionus gabonicus*, *Zaprionus africanus* and *Z. indianus* (*indianus* complex). Based on the closer structure of the *Zaprionus copia* retrotransposon to the Double-gap subfamily, without the diagnostic repetitions (imperfect repeat and dyad symmetric), and the presence of the GBF-1 binding site, we propose the existence of a specific *Zaprionus copia* subfamily, the GBFDouble-gap.

### 5.3.3 Phylogenetic analyses

The *in silico* search performed in the 12 *Drosophila* genomes recovered *copia* sequences for LTR-ULR regions only in *D. melanogaster*, *D. simulans* and *D. sechellia*, and for RT regions in *D. melanogaster*, *D. simulans*, *D. sechellia* and *D. yakuba* species

(Supplementary Table 1). Although Jordan and McDonald (1998a) identified LTR-ULR *copia* sequences of the Full-length, ULR-gap and Double-gap subfamilies in seven species of the *melanogaster* subgroup (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D. mauritiana*, *D. yakuba*, *D. teissieri* and *D. erecta*), the stringent search did not retrieve LTR-ULR sequences in the *D. yakuba* and *D. erecta* genomes nor RT sequences in the *D. erecta* genome. This was the reason why we chose to analyze only the LTR-ULR sequences of the *melanogaster* subgroup genome database.

Two phylogenetic reconstructions were made with the LTR-ULR and RT sequences, using the genomic *copia* sequences of the *melanogaster* subgroup, the *repleta* group sequences available in the GenBank database, the *Z. tuberculatus*MD sequence and the *Zaprionus* species sequences. The reconstructions were achieved using the NJ, ML and MP methods, which produced similar results. Figures 3 and 4 show the unrooted trees obtained with the ML method.

The LTR-ULR phylogeny (Figure 3) clustered the *copia* sequences in three well-supported and monophyletic clades hereafter named as Group A (*Zaprionus* species), Group B (*melanogaster* species and *Z. tuberculatus*MD), and Group C (*repleta* species). The average divergences (maximum composite likelihood distance) between Groups A and B was 0.325 (Table 3 and Supplementary Table 2), about two times smaller than those of Groups C *versus* B (0.562), and A *versus* C (0.647). The topology within the *repleta* and the *melanogaster* clades corroborates previous reconstructions (JORDAN; MCDONALD, 1998a; ALMEIDA; CARARETO, 2006). The *Z. tuberculatus* and the *Z. camerounensis* LTR-ULR sequences were grouped into species-specific clades, but those of *Z. indianus*, *Z. africanus* and *Z. gabonicus* were clustered together (Figure 3). The distances within Group A varied from zero (*Z. camerounensis5* / *Z. camerounensis3* and *Z. africanus4* / *Z. africanus5*) to 0.209 (*Z. tuberculatus3* / *Z. camerounensis1*). The lack of resolution in the *Z. indianus* / *Z. africanus* / *Z. gabonicus* clade may be due to the recent diversification of these species, identified only by DNA barcoding (YASSIN et al., 2008b). The *Z. tuberculatus*MD sequence did not cluster together with the other *Z. tuberculatus* sequences, but with the *melanogaster* species (Group B).

The phylogenetic reconstruction using the RT sequences corroborates the LTR-ULR tree, with the presence of Groups A, B and C; however, due to the inclusion of three new species (*Z. davidi*, *Z. multistriatus* and *D. yakuba*), additional information and a new group (Group D) were obtained (Figure 4). The *Z. davidi copia* sequence was included in Group A, together with the other *Zaprionus* species, despite the absence of support inside the clade; on

the other hand, *Z. multistriatus* clustered with *D. yakuba* sequences in Group D. The divergence between *Z. multistriatus* and *D. yakuba* is similar to that of Group A and B (0.133 for *Z. multistriatus* / *D. yakuba*, 0.145 for *Z. multistriatus* / Group A and 0.149 for *D. yakuba* / Group B) (Table 3 and Supplementary Table 3).

The striking length of the branches in the *Z. multistriatus / D. yakuba* clade could indicate that these sequences were clustered by the long-branch attraction effect, a phylogenetic methodological artifact originated by convergent evolution (for review, see BERGSTEN, 2005). A strategy to prevent long-branch attraction is the reconstruction of the phylogenetic tree excluding the faster evolving third codon positions (BERGSTEN, 2005), which was applied (Figure 5) and again Group D was formed, showing an decrease of branch lengths, but the branch of *Z. multistriatus* remains the longest in the tree. Hence, Group D was not considered and *Z. multistriatus* and *D. yakuba* were each one included in its species group for the evolutionary analyses.

The structural and sequence similarity and the phylogenetic incongruences regarding the relationships of the *copia* sequences in *Zaprionus* (*Zaprionus* genus or probably *Drosophila* subgenus), *melanogaster* (*Sophophora* subgenus) and *repleta* (*Drosophila* subgenus) species suggest the occurrence of *copia* HT between *Zaprionus* and the *melanogaster* species.

### 5.3.4 Evolutionary analyses

To test the HT hypotheses, two approaches were used. The strength of selective constraints on the *copia* retrotransposons sequences was investigated by calculating Tajima´s D values for LTR-ULR (LTR-ULR, LTR and ULR regions) and the RT sequences and by comparing the dN/dS ratios (ω) and the dS distances among the RT sequences of *Zaprionus* and *melanogaster* species, with those obtained for the *Gpdh* host genes, as it has usually been done for this purpose (for review, see LORETO et al., 2008).

Tajima's D values for the LTR-ULR (-0.05513), LTR (-0.28917) and ULR (-0.48867) comparisons were nonsignificant ($P > 0.10$); however, for the RT region, the value was significant, indicating that this domain is subjected to selective constraints (-1.93812, $P < 0.05$). The ω values were also computed in order to evaluate the occurrence of selection acting on the *copia* RT sequences (Table 4, Supplementary Tables 5 and 6). The results indicate that purifying selection is acting in sequence conservation between *melanogaster*

and *Zaprionus* sequences (ω: 0.156). On the other hand, the mean values show relaxed purifying selection in *melanogaster* (ω: 0.375) and *Zaprionus* sequences (ω: 0.829). The Codon Based Z-test corroborates these results (Table 4). For the *Gpdh* gene, the ω values varied from zero (*melanogaster* species subgroup) to 0.044 (*Zaprionus* species), demonstrating that, although the RT *copia* sequences are under purifying selection, since it is an active coding gene, selection on the TE is about 3 to 19 times more relaxed than on the gene. These comparisons suggest that the higher similarity between *Zaprionus* and *melanogaster* species cannot be due to differential selective constraints acting in the different species groups.

Since the selective test showed that selection constraints are lower in the *Gpdh* gene than in RT *copia* sequences, the pairwise dS distances between *melanogaster* and *Zaprionus* sequences were compared in order to infer HT, because dS values offer a measurement of neutral evolution in the absence of a strong codon usage bias (CBI *copia*: 0.505, varying from 0.463 to 0.644; CBI *Gpdh*: 0.600, varying from 0.492 to 0.700). Therefore, when two species are compared, a similar or higher proportion of synonymous substitutions should be expected in transposable elements (TEs) than in the host gene. Figure 6 and Table 4 show that all pairwise *copia* dS comparisons were lower than those of *Gpdh*, with proportions varying from 1.8 (*Z. davidi vs. D. yakuba*) to 9.9 (*Z. camerounensis vs. D. simulans*) times lower.

**5.4 Discussion**

In this study, we analyzed the *copia* retrotransposon in Drosophilidae species groups, focusing the genus *Zaprionus* in order to understand its distribution, activity and evolutionary history, once it had previously been identified only in *Z. tuberculatus*, as a LTR-ULR sequence available only in a published study (MCDONALD et al., 1997). The data obtained here show that *copia* is widely distributed in genus *Zaprionus* and that, during its evolution in the Drosophilidae family, it has experienced ancestral HT and vertical inheritance combined with subfamilies diversification and occasional extinction in the most recent species groups. Horizontal transfer has been suggested for *copia* evolution, but only within the *melanogaster* species subgroup (JORDAN; MCDONALD, 1998a; BOWEN; MCDONALD, 2001; SÁNCHEZ-GRACIA et al., 2005), between *D. melanogaster* and *D. willistoni* (JORDAN et al., 1999), and between an unknown species of the *melanogaster* subgroup and *Z.*

*tuberculatus* (ALMEIDA; CARARETO, 2006). We were unable to identify any *Z. tuberculatus* sequence closely related to the *melanogaster* species subgroup such as Z. *tuberculatus*MD; however, the absence of sequences homologous to *Z. tuberculatus*MD could be due to inter-population variability in the *copia* retrotransposon subfamilies.

Our results show *copia* as a transcriptional active component of the genomes of all *Zaprionus* species, although it was not possible to isolate the LTR-ULR region in *Z. multistriatus* and *Z. davidi*. Structural analysis of the regulatory region of the *Zaprionus* subgenus species revealed that *copia* sequences are closer to the most ancient subfamily of the *melanogaster* species subgroup, the Double-gap subfamily (MATYUNINA et al., 1996; JORDAN; MCDONALD et al., 1998a). This similarity, along with the lower nucleotide divergence and the closer phylogenetic relationships between the *Zaprionus* and the *melanogaster* sequences, suggests that the GBFDouble-gap subfamily belongs to the *melanogaster* subgroup *copia* family. Additionally, the high nucleotide and structural divergence from the *repleta* sequences suggests that at least one type of the *copia* retrotransposon was harbored by the ancestor of the Drosophilidae family, which could have diversified giving rise to the *repleta* group *copia* and the *melanogaster* / *Zaprionus copia* families. Later on, the latter family originated the three subfamilies of the *melanogaster* subgroup (Full-length, ULR-gap and Double-gap) and the GBFDouble-gap subfamily of genus *Zaprionus*.

The discordance about the taxonomic position of genus *Zaprionus* within the Drosophilidae family generates uncertainty about the initial steps of *copia* diversification. The majority of reports agree that *Zaprionus* species are closer related to those of the *Drosophila* subgenus, i.e., more related to the *repleta* than to the *melanogaster* species subgroup, a relationship we assumed here to propose a scenario for the evolutionary history of *copia*. Hence, the *copia* family sharing by *Zaprionus* and *melanogaster* species, as well as the higher sequence similarity, the closer phylogenetic relationships and the low levels of dS divergence, when compared to the *Gpdh* gene, allow us to envisage one ancient HT transfer from the ancestor of the *Zaprionus* genus to the ancestor of *D. melanogaster* / *D. simulans* / *D. sechellia* / *D. yakuba* species. The ancient status of the Double-gap subfamily plus the closer relationship with the GBFDouble-gap subfamily corroborate the latter proposition, although the possibility of a more ancient or even an extra HT event event having occurred cannot be ruled out, once Jordan and McDonald (1998a) identified the Double-gap subfamily in the *D. yakuba*, *D. erecta* and *D. teissieri* strains studied by them, but that was absent in the genome database. Although the grouping between *D. yakuba* and *Z. multistriatus copia* sequences

could suggest an extra HT event, the lack of geographic sharing between these species, as well as the similar distances in Group D regarding to the Groups A and B, indicate that grouping was produced by the long-branches attraction phenomenon.

An alternative hypothesis to explain the conservation between the *Zaprionus* and *melanogaster copia* sequences could be vertical transmission under high selective sequence constraints including also the synonymous site, as evidenced by the *copia* dS conservation regarding the *Gpdh* dS values. It has been suggested that a substantial fraction of ancestral repeats, at least 10,000 TE fragments in the human genome, have evolved under strong purifying selection throughout the eutherian radiation (LOWE et al., 2007), which could have acquired beneficial functions for their host (SILVA et al., 2003). The punctual distribution of *copia* in the *melanogaster* and *Zaprionus* species and the lack of any evidence that justifies the domestication hypothesis do not support such proposal in that case.

The time of divergence between *Zaprionus* and the *melanogaster* species sequences, calculated using the divergence rate of synonymous sites in *Drosophila* (1.6 % per million years, SHARP; LI, 1989), is congruent with above proposed scenario. If the mean dS value between *melanogaster* and *Zaprionus* sequences is 0.248 (0.099-0.509), the time of divergence should be about 8 (3-16) MYA, the period of *Zaprionus* subgenus divergence (7-9 MYA) and the splitting of *D. yakuba* and *D. melanogaster* / *D. simulans* / *D. sechellia* ancestors (8-15 MYA).

Finally, the negative selection estimated for the RT sequences between *melanogaster* and *Zaprionus* sequences is an intriguing issue, since at the first glance it seems unlikely to envisage any adaptive advantage represented by conserving the *copia* coding sequences. Nevertheless, as proposed by Jordan and McDonald (1998b), selection on the *copia* retrotransposon could occur at the genomic rather than the organismic level, which might have enabled the TE to succeed in the “arms race”, escaping the host pressures during the initial steps of the Drosophilidae evolution. On the other hand, a relaxed or even positive selection on the RT sequences inside *Zaprionus* subgenus and *melanogaster* subgroup *copia* elements, combined with neutral evolution in the LTR-ULR region, might have allowed the diversification of the regulatory region in the *indianus* complex and the *melanogaster* subgroup subfamilies more recently. These changes in the LTR-ULR could have resulted in the *copia* specificities in *Zaprionus* subgenus, such as the G-box duplication in the *Z. indianus* complex (*Z. indianus*, *Z. africanus* and *Z. gabonicus*), as well as the duplications in the Full-length and ULR-gap subfamilies, which enhance the transcription levels (MATYUNINA et al, 1996) ensuing the overcoming of the host regulation mechanisms that

control transposition, such as the transcription regulation by the action of host regulatory proteins (CAVAREC; HEIDMANN, 1993; PAL-BHADRA et al., 1998; BHADRA et al., 1999), or in the RNAi pathways (KALMYKOVA et al., 2005; KAWAMURA et al., 2008).

In conclusion, the data provided by this survey of the *copia* retrotransposon in genus *Zaprionus* reinforces the relevance of enlarging the range of species studied within a species group, as well as in less-related groups, in order to demonstrate the diversity and complexity of evolutionary forces acting on transposable elements and to infer their evolutionary history.

**Acknowledgments**

We gratefully acknowledge funding from the CAPES-COFECUB International Cooperation Program (to NS, CMAC and PC), FAPESP (MAVS), CNPq (CMAC and MAVS). NS was receipt of a CNPq fellowship.

**References**

ALMEIDA, L. M.; CARARETO, C. M. Sequence heterogeneity and phylogenetic relationships between the *copia* retrotransposon in *Drosophila* species of the *repleta* and *melanogaster* groups. **Genetics, Selection, Evolution**, França, v. 38, p. 535-550, 2006.

BENSON, G. Tandem repeats finder: a program to analyze DNA sequences. **Nucleic Acids Research**, Inglaterra, v. 27, 573-80, 1999.

BERGSTEN, J. A review of long-branch attraction. **Cladistics**, Inglaterra, v. 21, p. 163-193, 2005.

BHADRA, U.; PAL-BHADRA, M.; BIRCHLER, J. A. Role of the *male specific lethal* (*msl*) genes in modifying the effects of sex chromosomal dosage in *Drosophila*. **Genetics**, EUA, v. 152, p. 249-268, 1999.

BIÉMONT, C.; CIZERON, G. Distribution of transposable elements in *Drosophila* species. **Genetica**, Holanda, v. 105, p. 43-62, 1999.

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

CAVAREC, L.; HEIDMANN, T. The *Drosophila copia* retrotransposon contains binding sites for transcriptional regulation by homeoproteins. **Nucleic Acids Research**, Inglaterra, v. 21, p. 5041-5049, 1993.

CIZERON, G. et al. Distribution of the retrotransposable element *412* in *Drosophila* species. **Molecular Biology and Evolution**, EUA, v.15, p.1589-1599, 1998.

COSTA, A. P. et al. *Retrolycl-1*, a member of the *Tntl* retrotransposon super-family in the *Lycopersicon peruvianum* genome. **Genetica**, Holanda, v. 107, p. 65-72, 1999.

DA LAGE, J-L. et al. A phylogeny of Drosophilidae using the *amyrel* gene: questioning the *Drosophila melanogaster* species group boundaries. **Journal of Zoological Systematics and Evolutionary Research**, Alemanha, v. 45, p. 47-63, 2007.

DE SALLE, R. The origin and possible time of divergence of the Hawaiian Drosophilidae: evidence from DNA sequences. **Molecular Biology and Evolution**, EUA, v. 9, p. 905-916, 1992.

DE SETTA, N.; CARARETO, C. M. A. Screening for transposable elements in South America invasive species *Zaprionus indianus* and *Drosophila malerkotliana*. **Drosophila Information Service**, EUA, v. 90, p. 96-99, 2007.

DE SETTA, N. et al. Insights about the phylogenetic relationship between *Zaprionus* and *Drosophila* from molecular and cytogenetic data of *α-esterase7* gene of *Zaprionus indianus*. **Drosophila Information Service**, EUA, in press, 2008.

FELSENSTEIN, J. Confidence limits on phylogenies: an approach using the bootstrap. **Evolution**, EUA, v. 39, p. 783-791, 1985.

GRABE, N. AliBaba2: context specific identification of transcription factor binding sites. **In Silico Biology**, Holanda, v. 2, p. S1-S15, 2002.

GRAUR, D.; LI, W-H. **Fundamentals of Molecular Evolution**. 2ª Ed. Sunderland, Massachussets: Sinauer Associates, 2000.

GRIMALDI, D. A. A phylogenetic, revised classification of genera in the Drosophilidae (Diptera). **Bulletin of the American Museum of Natural History**, EUA, v. 197, p. 123-128, 1990.

GUINDON, S.; GASCUEL, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. **Systematic Biology**, Inglaterra, v. 52, p. 696-704, 2003.

HALL, T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symposium Series**, Inglatera, v. 41, p. 95-98, 1999.

HASEGAWA, M.; KISHINO, H.; YANO T. Dating of the human-ape splitting by molecular clock of mitochondrial DNA. **Journal of Molecular Evolution**, Alemanha, v. 21, p. 160-174, 1985.

HEREDIA, F. et al. Complex evolution of *gypsy* in drosophilid species. **Molecular Biology and Evolution**, EUA, v. 21, p. 1831-1842, 2004.

JORDAN, I. K.; MCDONALD, J. F. *Evolution of the copia retrotransposon in the Drosophila melanogaster* species subgroup. **Molecular Biology and Evolution**, EUA, v. 15, p. 1160-1171, 1998a.

JORDAN, I. K.; MCDONALD, J. F. Interelement selection in the regulatory region of the *copia* retrotransposon. **Journal of Molecular Evolution**, Alemanha, v. 47, p. 670–676, 1998b.

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

JOWETT, T. Preparation of nucleic acids. In: ROBERTS, D. B. **Drosophila: A Practical Approach.** Oxford, IRL Press, pp. 275-277, 1986.

KALMYKOVA, A. I.; KLENOV, M. S.; GVOZDEV, V. A. Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. **Nucleic Acids Research**, Inglaterra, v. 33, p. 2052-2059, 2005.

KAWAMURA, Y. et al. *Drosophila* endogenous small RNAs bind to *Argonaute 2* in somatic cells. **Nature**, Inglaterra, v. 453, p. 793-797, 2008.

KLIMCZAK, L. J.; SCHINDLER, U.; CASHMORE, A. R. DNA binding activity of the *Arabidopsis* G-box binding factor GBF1 is stimulated by phosphorylation by casein kinase II from broccoli. **The Plant Cell**, EUA, v. 4, p. 87-98, 1992.

KWIATOWSKI, J; AYALA, F. J. Phylogeny of *Drosophila* and related genera: conflict between molecular and anatomical analyses. **Molecular Phylogenetics and Evolution**, v. 13, p. 319-328, 1999.

KWIATOWSKI, J. et al. Phylogeny of *Drosophila* and related genera inferred from the nucleotide sequence of the Cu, Zn Sod gene. **Journal of Molecular Evolution**, Alemanha, v. 38, p. 443-454, 1994.

LACHAISE, D.; SILVAIN, J. F. How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster-D. simulans* palaeogeographic riddle. **Genetica**, Holanda, v. 120, p. 17-39, 2004.

LORETO, E. L.; CARARETO, C. M.; CAPY, P. Revisiting horizontal transfer of transposable elements in *Drosophila*. **Heredity**, Inglaterra, v. 100, p. 545-54, 2008.

LOWE, C. B.; BEJERANO, G.; HAUSSLER, D. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 104, p. 8005-8010, 2007.

MARUYAMA, K.; HARTL, D. L. Evolution of the transposable element *mariner* in *Drosophila* species. **Genetics**, EUA, v. 128, p. 319-329, 1991.

MATYUNINA, L. V.; JORDAN, I. K.; MCDONALD, J. F. Naturally occurring variation in *copia* expression is due to both element (cis) and host (trans) regulatory variation. **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 93, p. 7097-7102, 1996.

MCDONALD, J. F. et al. LTR retrotransposons and the evolution of eukaryotic enhancers. **Genetica**, Holanda, v. 100, p. 3-13, 1997.

MONTCHAMP-MOREAU, C. et al. Distribution and conservation of sequences homologous to the *1731* retrotransposon in *Drosophila*. **Molecular Biology and Evolution**, EUA, v. 10, p. 791-803, 1993.

MOUNT, S. M.; RUBIN, G. M. Complete nucleotide sequence of the *Drosophila* transposable element *copia*: homology between *copia* and retroviral proteins. **Molecular and Cellular Biology**, EUA, v. 5, p. 1630-1638, 1985.

NUZHDIN, S. V.; PASYUKOVA, E. G.; MACKAY, T. F. Positive association between *copia* transposition rate and copy number in *Drosophila melanogaster*. **Proceedings of the Royal Society of London**, Inglaterra, v. 263, p. 823-831, 1996.

OKADA, T.; CARSON, H. L. The genera *Phorticella* Duda and *Zaprionus* Coquillett (Diptera, Drosophilidae) of the Oriental region and New Guinea. **Kontyu**, Japão, v. 51, p. 539-553, 1983.

PAL-BHADRA, M.; BHADRA, U.; BIRCHLER, J. A. Role of multiple trans-acting regulators in modifying the effect of the retrotransposon *copia* on host gene expression in *Drosophila*. **Molecular & General Genetics**, Alemanha, v. 259, p. 198-206, 1998.

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

PÉLANDAKIS, M.; SOLIGNAC, M. Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. **Journal of Molecular Evolution**, Alemanha, v. 37, p. 525-543, 1993.

REMSEN, J.; DE SALLE, R. Character congruence of multiple data partitions and the origin of the Hawaiian Drosophilidae. **Molecular Phylogenetics and Evolution**, EUA, v. 9, p. 225-235, 1998.

ROZAS, J. et al. DnaSP, DNA polymorphism analyses by the coalescent and other methods. **Bioinformatics**, Inglaterra, v. 19, p. 2496-2497, 2003.

RUSSO, C. A. M. TAKEZAKI, N.; NEI, M. Molecular phylogeny and divergence times of drosophilid species. **Molecular Biology and Evolution**, EUA, v. 12, p. 391-404, 1995.

SÁNCHEZ-GRACIA, A.; MASIDE, X.; CHARLESWORTH, B. High rate of horizontal transfer of transposable elements in *Drosophila*. **Trends in Genetics**, Inglaterra, v. 21, p. 200-203, 2005.

SHARP, P. M.; LI, W. H. On the rate of DNA sequence evolution in *Drosophila*. **Journal of Molecular Evolution**, Alemanha, v. 28, p. 398-402, 1989.

SILVA, J. C. et al. Conserved fragments of transposable elements in intergenic regions: evidence for widespread recruitment of MIR- and L2-derived sequences within the mouse and human genomes. **Genetical Research**, Inglaterra, v. 82, p. 1-18, 2003.

SWOFFORD, D. **PAUP: phylogenetic analysis using parsimony:** Version 4.0b10. Washington, DC: Smithsonian Institution, 1997.

TABATA, T. et al. HBP-1a and HBP-1b: leucine zipper-type transcription factors of wheat. **The EMBO Journal**, Inglaterra, v. 10, p. 1459-1467, 1991.

TAJIMA, F. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. **Genetics**, EUA, v. 123, p. 585-595, 1989.

TAMURA, K. et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. (1993) v. 24, p. 1596-1599, 2007.

TAMURA, K.; NEI, M.; KUMAR, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. **Proceedings of the National Academy of Sciences (USA)**, EUA, v. 101, p. 11030-11035, 2004.

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

TATARENKOV, A. et al. On the evolution of *Dopa decarboxylase* (*Ddc*) and *Drosophila* systematics. **Journal of Molecular Evolution**, Alemanha, v. 48, p. 445-462, 1999.

THOMPSON, J. D. et al. CLUSTAL W: improving the sensitivity of progressive multiple sequence aligment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, Inglaterra, v. 22, p. 4673-4680, 1994.

THOMAS, R. H.; HUNT, J. A. Phylogenetic relationships in *Drosophila*: a conflict between molecular and morphological data. **Molecular Biology and Evolution**, EUA, v. 10, p. 362-374, 1993.

THROCKMORTON, L. H. The phylogeny, ecology and geography of *Drosophila*. In: KING, R.C. (Org.) **Handbook of genetics**. Nova Iorque: Plenum, pp. 421-469, 1975.

WILSON, S.; MATYUNINA, L. V.; MCDONALD, J. F. An enhancer region within the copia untranslated leader contains binding sites for *Drosophila* regulatory proteins. **Gene**, Holanda, v. 209, p. 239-246, 1998.

XIANG, C.; MIAO, Z.; LAM, E. DNA-binding properties, genomic organization and expression pattern of TGA6, a new member of the TGA family of bZIP transcription factors in *Arabidopsis thaliana*. **Plant Molecular Biology**, Holanda, v. 34, p. 403-415, 1997.

YASSIN, A. et al. Grafting the molecular phylogenetic tree with morphological branches to reconstruct the evolutionary history of the genus *Zaprionus* (Diptera: Drosophilidae). **Molecular Phylogenetics and Evolution**, EUA, v. 47, p. 903-15, 2008a.

YASSIN, A. et al. DNA barcode discovers two cryptic species and two geographical radiations in the invasive drosophilid *Zaprionus indianus*. **Molecular Ecology Notes**, Inglaterra, v. 8, p. 491-501, 2008b.

**Table 1.** Species used, taxonomic classification according to Yassin et al. (2008a), geographic origin and Genbank accession numbers for *copia* retrotransposon sequences obtained in this work.

| Genus *Zaprionus* | Geographic origin | GenBank accessions | |
|---|---|---|---|
| | | LTR – ULR | RT |
| Subgenus *Anaprionus* | - | - | - |
| *Z. multistriatus* | Bangalore (India) | - | FJ715493 |
| Subgenus *Zaprionus* | - | - | - |
| Group *inermis* | - | - | - |
| Complex *tuberculatus* | - | - | - |
| *Z. tuberculatus* | Ithala (South Africa) | FJ755275 to FJ755279 | FJ715494 |
| Group *armatus* | - | - | - |
| Complex *lachaisei* | - | - | - |
| *Z. camerounensis* | Amani (Tanzania) | FJ755280 to FJ755284 | FJ715495 |
| Complex *davidi* | - | - | - |
| *Z. davidi* | São Tomé (São Tomé e Principe) | - | FJ715496 |
| Complex *indianus* | - | - | - |
| *Z. gabonicus* | Makokou (Gabon) | FJ755285 to FJ755289 | FJ715497 |
| *Z. africanus* | Kibale (Uganda) | FJ755290 to FJ755294 | FJ715498 |
| *Z. indianus* | Brasília (Brazil) | FJ755295 to FJ755299 | FJ715499 |

**Table 2.** Regulatory motif sequences and location of *copia* 5' LTR-ULR in *Zaprionus* species and *D. melanogaster* Full-length variant. Nucleotides in parenthesis mean *indels* in the alignment.

| Motif | *D. melanogaster* Sequence | *Zaprionus* species consensus | Alignment position |
|---|---|---|---|
| Heat shock element | CTACAAAAATAACG | CTAMAAAdvkAAmG | 10-23 |
| TATA box | CTTTATATTT | yTTTATATTT | 37-46 |
| Imperfect repeat | CAC(T)CTTTT(T)AT | - | 61-70 ; 98-109 |
| Transcriptional start | CTTTCCTTCTTGTACGTTTTT | bCCTTCTTyTwddACkyTTyC | 115-135 |
| Downstream element | CGTG | CGTG | 149-152 |
| Poly-A signal | AAATATAAATC | Highly variable | 172-193 |
| Primer binding site | GGTTATGGGCCCAG | GG(TTATGGG)CCCAG | 267-280 |
| *Engrailed* site | DVAAwTAAAk | wmAAwTArAT | 295-304; 307-316; 322-331 |
| DmC/EBP site | TTGTGAAw | TTGTGAAA | 305-311; 318-325; 331-338; 357-365; 395-403 |
| Dyad symmetric | TTTCACA / TTGTGAAw | - | 347-365; 376-403 |
| G-box binding factor 1 | - | GTGCCACGTG | 357-365; 374-392 |
| B-box binding factor 2 | AAGGCTAAACGT | AWK(K)Y(T)WAACG(T) | 440-451 |

**Table 3.** Summary of genetic divergence (Maximum composite likelihood distance) of *copia* sequences used in the phylogenetic analysis. *Z. tuberculatus* LTR-ULR sequence from McDonald et al. (1997) was included with the *melanogaster* subgroup sequences. NS: number of sequences; NC: number of pairwise comparisons.

| Groups | NS | NC | Mean (Min-Max) | |
|---|---|---|---|---|
| | | | LTR-ULR | RT |
| *Zaprionus* genus | 7 | 25 | 0.084 (0-0.220) | 0.065 (0.012-0.168) |
| *melanogaster sub*group | 27 | 15 | 0.027 (0-0.062) | 0.024 (0-0.158) |
| *repleta* group | 2 | 8 | 0.056 (0.002-0.098) | 0.053 |
| *Zaprionus* genus x *melanogaster* subgroup | 34 | 40 | 0.305 (0.238-0.447) | 0.09 (0.041-0.016) |
| *Zaprionus* genus x *repleta* group | 9 | 33 | 0.671 (0.606-0.807) | 0.251 (0.227-0.285) |
| *melanogaster* subgroup x *repleta* group | 29 | 23 | 0.672 (0.603-0.737) | 0.252 (0.215-0.288) |

**Table 4.** Comparative dN, dS, dN/dS ratios and Z-test for genus *Zaprionus* and *melanogaster* subgroup *copia* RT sequences. The pairwise dN and dS values are shown in detail in the Supplementary Table 4. * P < 0.05.

| Species | dN mean (min-max) | dS mean (min-max) | dN/dS mean (min-max) | Z-test * (%) | |
|---|---|---|---|---|---|
| | | | | dN < dS | dN > dS |
| *Copia* | | | | | |
| *melanogaster* | 0.014 (0-0.058) | 0.060 (0-0.598) | 0.375 (0-1.263) | 54 (15%) | 27 (8%) |
| *Zaprionus* | 0.046 (0.009-0.106) | 0.116 (0-0.391) | 0.820 (0.202-5.000) | 6 (29%) | 3 (14%) |
| *melanogaster vs. Zaprionus* | 0.038 (0.012-0.082) | 0.248 (0.099-0.509) | 0.156 (0.046-0.462) | 186 (98%) | 0 |
| *Gpdh* | | | | | |
| *melanogaster* | 0 | 0.175 (0.317-0.058) | 0 | 100% | 0 |
| *Zaprionus* | 0.012 (0-0.039) | 0.286 (0.019-0.728) | 0.044 (0-0.132) | 93% | 0 |
| *melanogaster vs. Zaprionus* | 0.010 (0.006-0.018) | 0.866 (0.806-1.360) | 0.011 (0.004-0.027) | 100% | 0 |

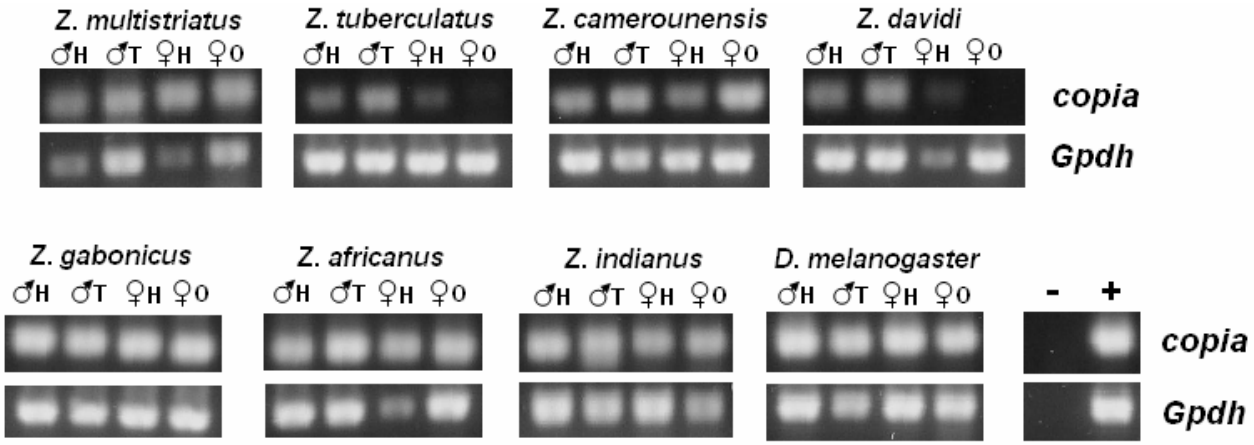



**Figure 1.** RT-PCR reactions of the RT *copia* retrotransposon from zero to 10 days old germline (testes and ovaries) and somatic (heads) tissues of *Zaprionus* species and *D. melanogaster*. Ultrapure water and *D. melanogaster* DNA (cDNA and genomic DNA) were used as negative and positive controls, respectively. The *Gpdh* RT-PCR was used as control of cDNA quality. ♂: male; ♀: female; H: head; T: testis; O: ovary.

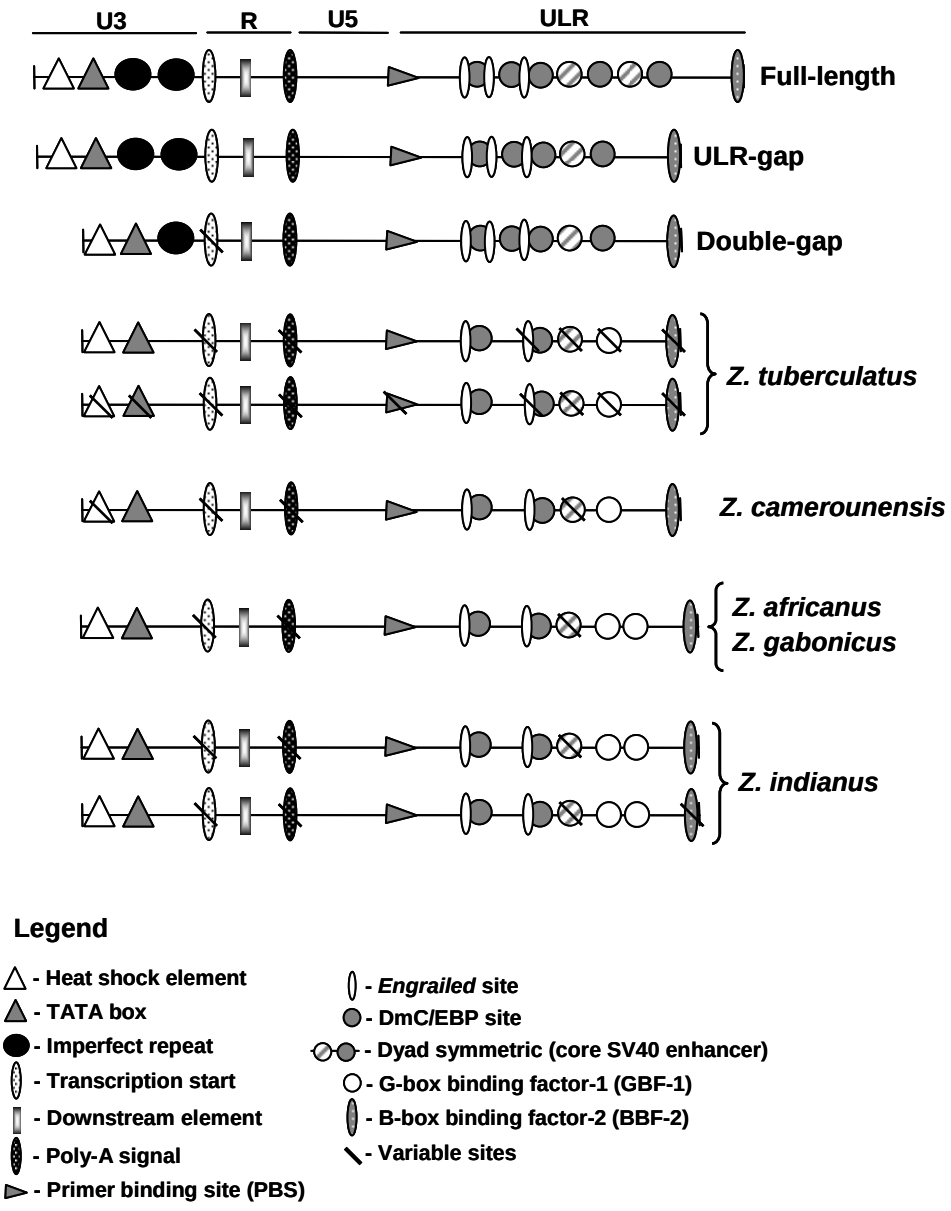



**Figure 2.** Structure of the 5' LTR-ULR region of the *copia* retrotransposon of *Zaprionus* species. The comparative analysis was based on the *copia* elements of *D. melanogaster* (Full-length and ULR-gap subfamilies) and *D. simulans* (Double-gap subfamily).

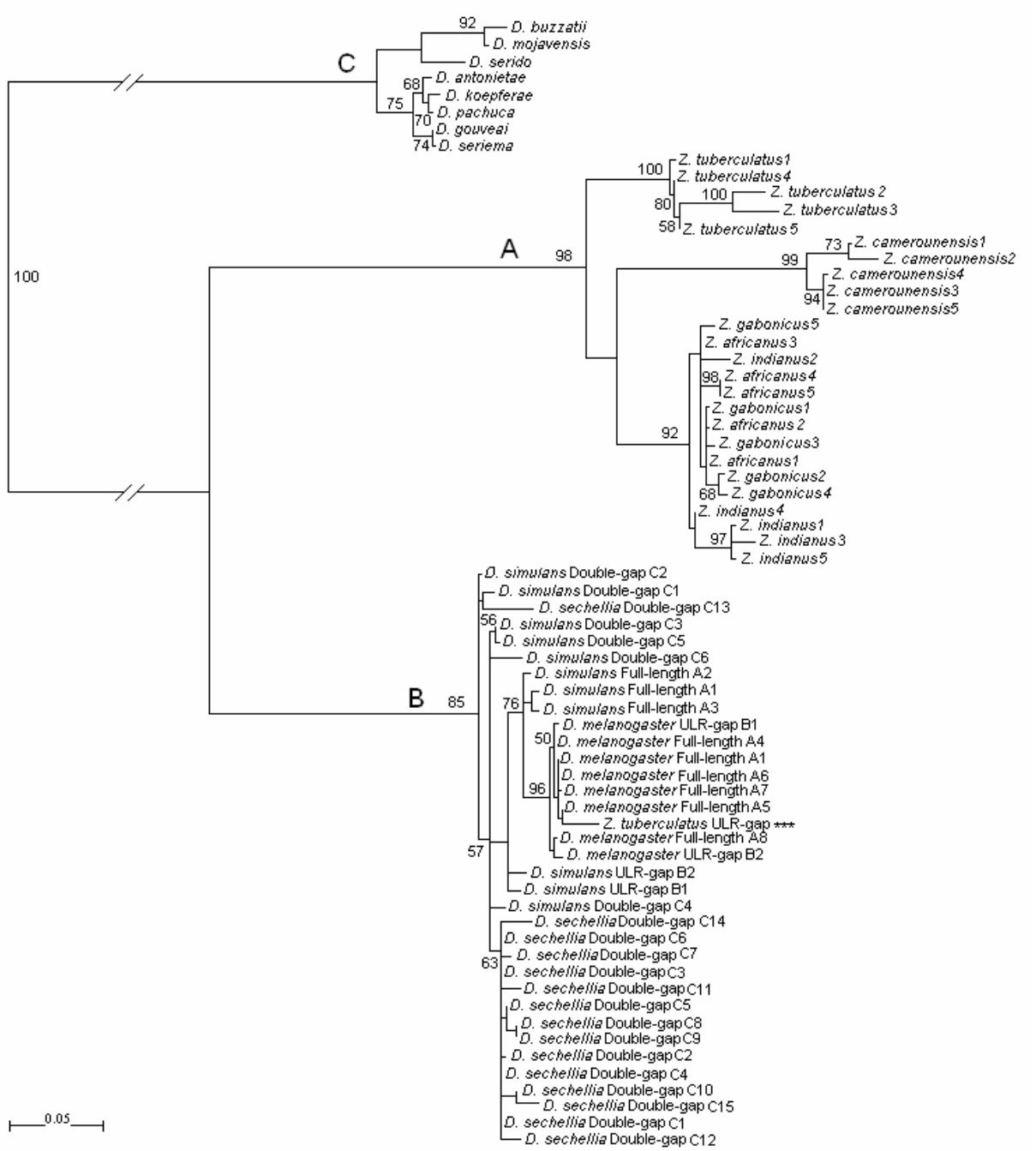



**Figure 3.** Phylogenetic relationships of the *copia* retrotransposon of genus *Zaprionus*, group *repleta* and subgroup *melanogaster*, generated by ML analysis (HKY85 distance) as implemented by PhyML. Branch consistencies were evaluated by the bootstrap method (1,000 replications). NJ and MP reconstructions produce the same basic topology shown in that tree, with minor differences in the *melanogaster* subgroup and *Zaprionus* clades. ***: *Z. tuberculatus*MD

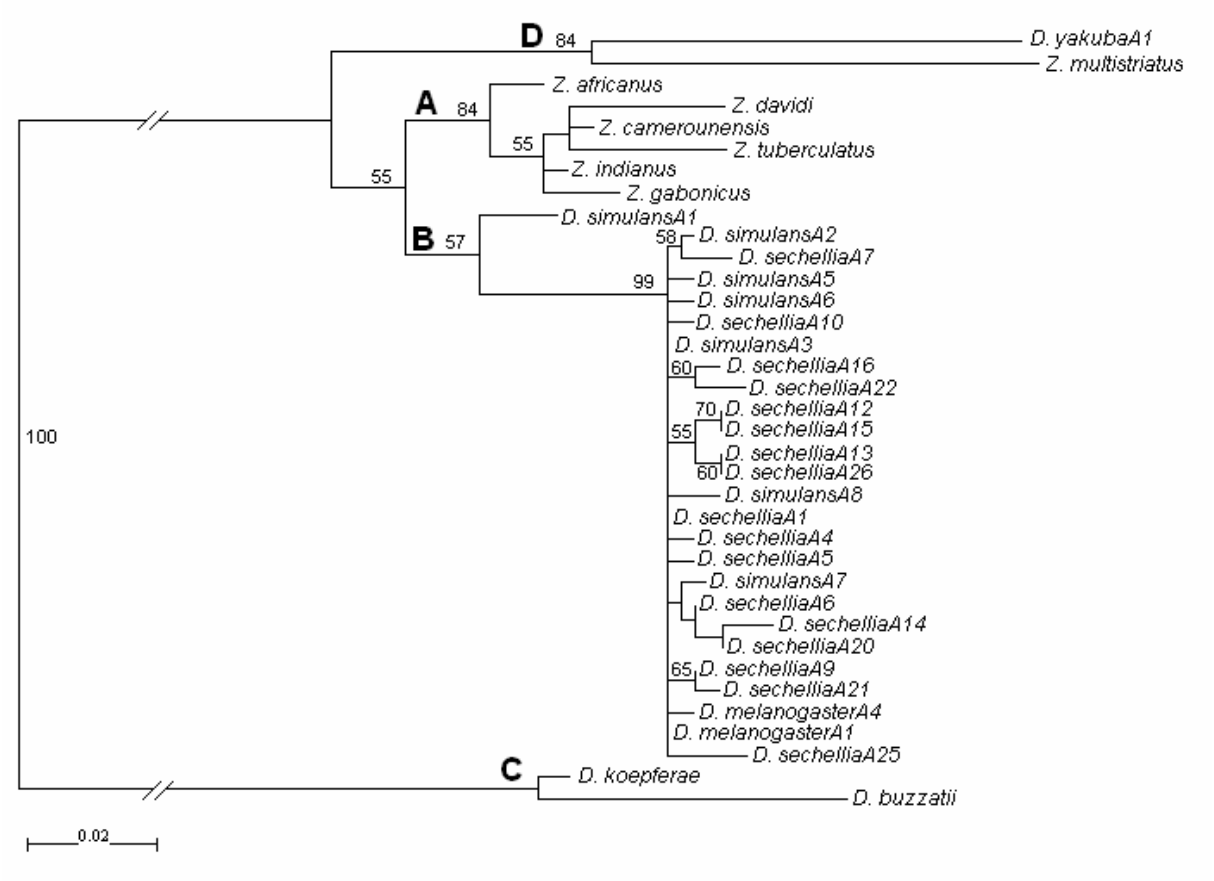



**Figure 4.** Maximum likelihood (HKY85 distance) tree based on *copia* RT sequences of *Zaprionus*, *melanogaster* and *repleta* species. Bootstrap support was calculated in 1,000 replications.

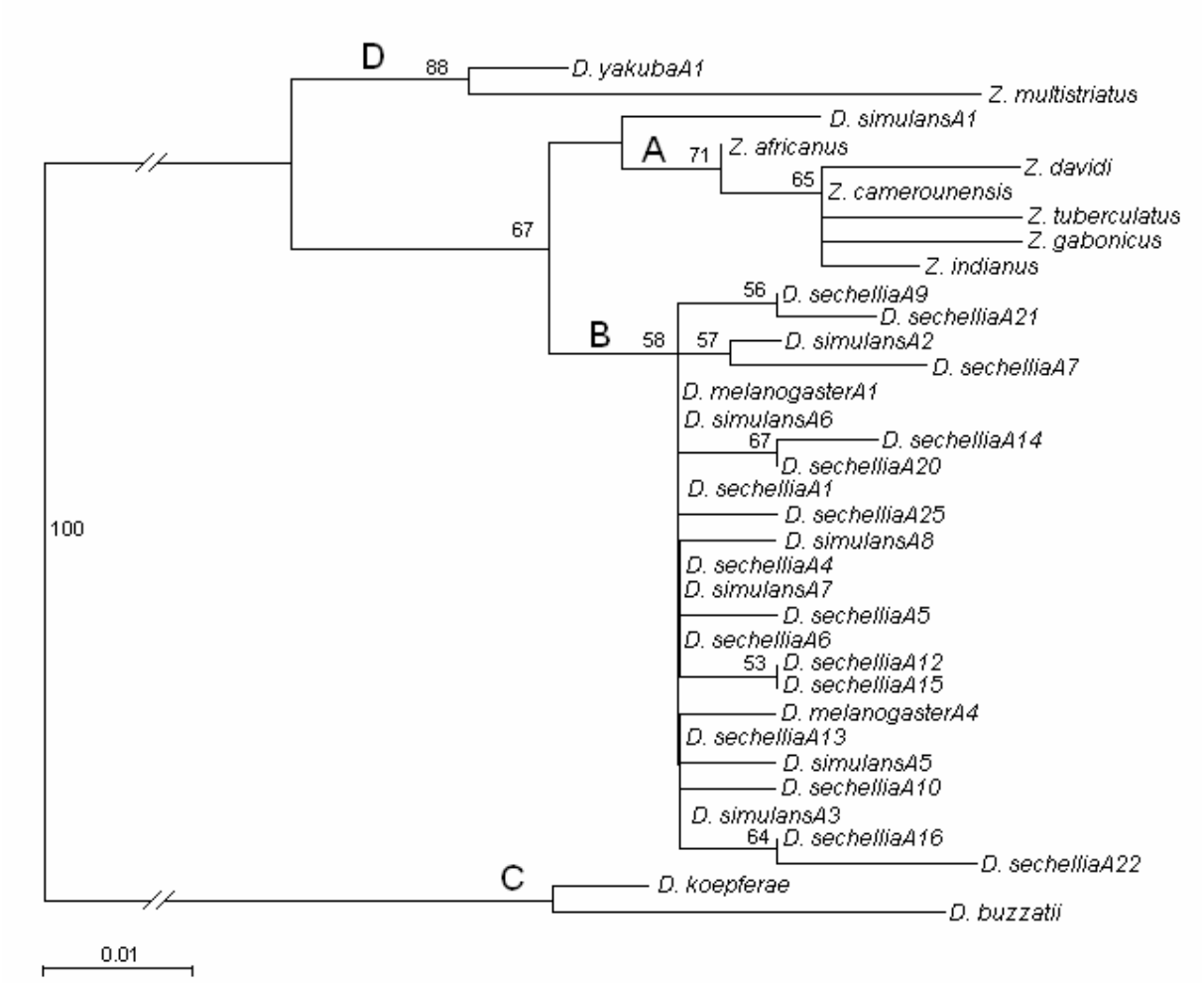



**Figure 5.** Phylogenetic relationships of the RT region of *Zaprionus*, *melanogaster* and *repleta* species reconstructed by ML method (HKY85 distance) using the first and second codon position. Bootstrap analysis was computed with 1,000 replication. The elimination of the third codons position minimizes the long-branch attraction effect in the Group D.

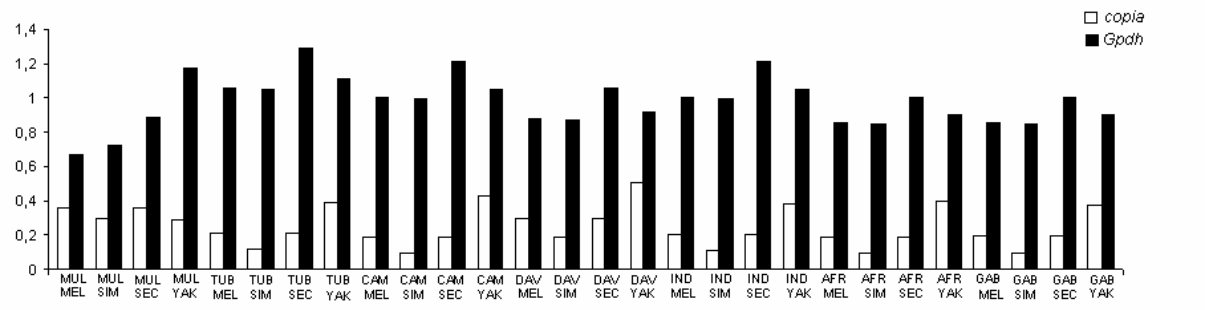



**Figure 6.** Pairwise comparisons of the mean numbers of synonymous substitutions per synonymous sites (dS) for *copia* RT sequences and *Gpdh* in species of genus *Zaprionus* and the *melanogaster* species subgroup. MUL: *Z. multistriatus*; TUB: *Z. tuberculatus*; CAM: *Z. camerounensis*; DAV: *Z. davidi*; IND: *Z. indianus*; AFR: *Z. africanus*; GAB: *Z. gabonicus*; MEL: *D. melanogaster*; SIM: *D. simulans*; SEC: *D. sechellia*; YAK: *D. yakuba*.

**Supplementary Table 1.** Genomic sequences of *D. melanogaster*, *D. simulans*, *D. sechellia* and *D. yakuba* obtained from 12 *Drosophila* genome searches. * Sequences eliminated from the phylogenetic analysis due to 100% identity among LTR-ULR *D. melanogaster* A1/A2/A3 sequences and RT *D. melanogaster* A1/A2/A3, *D. melanogaster* A4/A5/A6/A7, *D. simulans* A3/A4, *D. sechellia* A1/A2/A3, *D. sechellia* A5/A8, *D. sechellia* A10/A11, *D. sechellia* A15/A17/A18, *D. sechellia* A13/A19, *D. sechellia* A22/A23/A24 sequences.

| Species | Symbol | Region | Position |
|---|---|---|---|
| LTR-ULR | | | |
| *D. melanogaster* | A1 | 3R | 15940011-15940450 |
| | A2* | 2R | 1105270-1105709 |
| | A3* | 2L | 3073099-3073538 |
| | A4 | U | 9085950-9086389 |
| | A5 | 3L | 2113000-2113439 |
| | A6 | X | 5825876-5826316 |
| | A7 | 3RHet | 2448871-2449311 |
| | A8 | 3LHet | 666615-667055 |
| | B1 | U | 6802618-6803031 |
| | B2 | 3L | 12954939-12955353 |
| *D. simulans* | A1 | 2R | 1083313-1083753 |
| | A2 | chr2h_Mrandom_039 | 25750-26187 |
| | A3 | chrU_M_1494 | 1250-1689 |
| | B1 | chrU_M_1901 | 530-942 |
| | B2 | chr2h_Mrandom_038 | 43713-44129 |
| | C1 | chrU_M_1028 | 1205-1567 |
| | C2 | chrU_M_5031 | 789-1163 |
| | C3 | chr3h_Mrandom_003 | 48505-48879 |
| | C4 | chrU_M_2295 | 7581-7956 |
| | C5 | chr2h_Mrandom_039 | 21097-21471 |
| | C6 | 2R | 1205642-1206017 |
| *D. sechellia* | C1 | scaffold_67 | 99798-100173 |
| | C2 | scaffold_479 | 3583-3958 |
| | C3 | scaffold_80 | 12906-13283 |
| | C4 | scaffold_1185 | 4664-5041 |
| | C5 | scaffold_2788 | 1600-1976 |
| | C6 | scaffold_26 | 331633-332012 |
| | C7 | scaffold_18 | 747254-747630 |
| | C8 | scaffold_10813 | 906-1282 |
| | C9 | scaffold_4520 | 1763-2139 |
| | C10 | scaffold_94 | 3841-4217 |
| | C11 | scaffold_878 | 1195-1571 |
| | C12 | scaffold_1731 | 3115-3487 |
| | C13 | scaffold_1639 | 1871-2253 |
| | C14 | scaffold_467 | 7125-7503 |
| | C15 | scaffold_351 | 1-358 |
| Reverse transcriptase | | | |
| *D. melanogaster* | A1 | 3R | 27159518-27159770 |
| | A2* | 3L | 12952529-12952781 |
| | A3* | 2L | 16144526-16144778 |
| | A4 | X | 4180135-4180387 |
| | A5* | 3RHet | 2452164-2452416 |
| | A6* | 3LHet | 669909-670161 |
| | A7* | 2R | 7836-8088 |

**Continuation Supplementary Table 1.**

| Species | Symbol | Region | Position |
|---|---|---|---|
| *D. simulans* | A1 | chrU_M_804 | 1933-2185 |
| | A2 | chr2h_Mrandom_039 | 24207-24459 |
| | A3 | chr2h_Mrandom_038 | 50551-50803 |
| | A4* | chrU_M_356 | 2893-3145 |
| | A5 | chrU_M_2608 | 1115-1367 |
| | A6 | chrU_M_226 | 238-490 |
| | A7 | chrU_M_2295 | 4483-4735 |
| | A8 | chrU_M_2579 | 795-1047 |
| *D. sechellia* | A1 | scaffold_116 | 88475-88727 |
| | A2* | scaffold_57 | 255787-256039 |
| | A3* | scaffold_932 | 369-621 |
| | A4 | scaffold_209 | 17995-18247 |
| | A5 | scaffold_94 | 743-995 |
| | A6 | scaffold_7595 | 370-622 |
| | A7 | scaffold_7475 | 248-500 |
| | A8* | scaffold_2586 | 849-1101 |
| | A9 | scaffold_1833 | 1943-2195 |
| | A10 | scaffold_4601 | 1027-1277 |
| | A11* | scaffold_467 | 4123-4373 |
| | A12 | scaffold_8980 | 1204-1456 |
| | A13 | scaffold_18 | 744205-744457 |
| | A14 | scaffold_5484 | 586-838 |
| | A15 | scaffold_1731 | 133-385 |
| | A16 | scaffold_1501 | 2322-2574 |
| | A17* | scaffold_1185 | 1666-1918 |
| | A18* | scaffold_2607 | 599-851 |
| | A19* | scaffold_1296 | 3705-3957 |
| | A20 | scaffold_878 | 4331-4583 |
| | A21 | scaffold_479 | 8707-8959 |
| | A22 | scaffold_80 | 16133-16385 |
| | A23* | scaffold_26 | 334753-335005 |
| | A24* | scaffold_2195 | 437-689 |
| | A25 | scaffold_1581 | 3005-3257 |
| | A26 | scaffold_7249 | 17-260 |
| *D. yakuba* | A1 | v2_chrUn_141 | 5733-5983 |

**Supplementary Table 2.** Pairwise genetic distance (maximum composite likelihood distance) among the 5' LTR-ULR *copia* regions of *Zaprionus*, *melanogaster* and *repleta* species.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Dkoep | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2. Dbuzz | .098 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3. Dseri | .075 | .065 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4. Dgouv | .022 | .095 | .073 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5. Danto | .012 | .089 | .068 | .019 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6. Dserm | .022 | .098 | .071 | .002 | .019 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7. Dpach | .007 | .095 | .076 | .019 | .010 | .022 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8. Dmoja | .091 | .015 | .057 | .087 | .080 | .089 | .086 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 9. DmelA1 | .593 | .584 | .563 | .567 | .577 | .566 | .594 | .589 | | | | | | | | | | | | | | | | | | | | | | | | | |
| 10. DmelA4 | .593 | .584 | .563 | .567 | .577 | .566 | .594 | .589 | .002 | | | | | | | | | | | | | | | | | | | | | | | | |
| 11. DmelA5 | .598 | .589 | .567 | .572 | .582 | .571 | .599 | .594 | .002 | .005 | | | | | | | | | | | | | | | | | | | | | | | |
| 12. DmelA6 | .595 | .587 | .565 | .570 | .580 | .569 | .596 | .592 | .000 | .002 | .002 | | | | | | | | | | | | | | | | | | | | | | |
| 13. DmelA7 | .595 | .587 | .570 | .570 | .580 | .569 | .596 | .592 | .002 | .005 | .005 | .002 | | | | | | | | | | | | | | | | | | | | | |
| 14. DmelA8 | .591 | .582 | .561 | .566 | .575 | .564 | .592 | .587 | .005 | .007 | .007 | .005 | .007 | | | | | | | | | | | | | | | | | | | | |
| 15. DmelB1 | .562 | .553 | .533 | .536 | .546 | .535 | .562 | .558 | .005 | .002 | .005 | .005 | .007 | .007 | | | | | | | | | | | | | | | | | | | |
| 16. DmelB2 | .585 | .575 | .555 | .558 | .568 | .557 | .585 | .581 | .012 | .010 | .012 | .012 | .015 | .007 | .010 | | | | | | | | | | | | | | | | | | |
| 17. DsimA1 | .580 | .577 | .547 | .556 | .577 | .555 | .587 | .589 | .021 | .018 | .023 | .021 | .023 | .023 | .022 | .027 | | | | | | | | | | | | | | | | | |
| 18. DsimA2 | .580 | .582 | .561 | .560 | .581 | .559 | .587 | .593 | .021 | .019 | .023 | .021 | .023 | .023 | .022 | .027 | .014 | | | | | | | | | | | | | | | | |
| 19. DsimA3 | .574 | .576 | .545 | .567 | .576 | .566 | .581 | .587 | .028 | .026 | .030 | .028 | .030 | .028 | .027 | .032 | .009 | .014 | | | | | | | | | | | | | | | |
| 20. DsimB1 | .553 | .560 | .548 | .533 | .555 | .532 | .560 | .578 | .035 | .032 | .035 | .035 | .035 | .035 | .032 | .037 | .025 | .020 | .025 | | | | | | | | | | | | | | |
| 21. DsimB2 | .585 | .587 | .569 | .560 | .581 | .558 | .592 | .604 | .037 | .034 | .037 | .037 | .037 | .037 | .035 | .040 | .027 | .022 | .027 | .017 | | | | | | | | | | | | | |
| 22. DsimC1 | .524 | .521 | .500 | .505 | .521 | .504 | .527 | .534 | .034 | .034 | .034 | .034 | .031 | .034 | .034 | .040 | .020 | .022 | .025 | .025 | .028 | | | | | | | | | | | | |
| 23. DsimC2 | .531 | .538 | .520 | .518 | .533 | .516 | .534 | .556 | .038 | .038 | .038 | .038 | .038 | .038 | .038 | .044 | .024 | .024 | .030 | .019 | .022 | .011 | | | | | | | | | | | |
| 24. DsimC3 | .543 | .550 | .532 | .529 | .544 | .528 | .546 | .567 | .035 | .035 | .035 | .035 | .035 | .036 | .036 | .041 | .022 | .022 | .027 | .016 | .019 | .017 | .011 | | | | | | | | | | |
| 25. DsimC4 | .557 | .552 | .546 | .543 | .559 | .542 | .560 | .570 | .041 | .041 | .041 | .041 | .041 | .041 | .041 | .047 | .027 | .027 | .033 | .022 | .024 | .022 | .016 | .011 | | | | | | | | | |
| 26. DsimC5 | .538 | .545 | .527 | .524 | .539 | .523 | .541 | .562 | .038 | .038 | .038 | .038 | .038 | .038 | .038 | .044 | .024 | .024 | .030 | .019 | .022 | .020 | .013 | .003 | .013 | | | | | | | | |
| 27. DsimC6 | .571 | .572 | .552 | .557 | .573 | .556 | .575 | .590 | .047 | .047 | .047 | .047 | .044 | .047 | .047 | .053 | .033 | .033 | .038 | .030 | .033 | .025 | .024 | .019 | .024 | .022 | | | | | | | |
| 28. DsecC1 | .548 | .555 | .537 | .534 | .549 | .532 | .551 | .572 | .038 | .038 | .038 | .038 | .038 | .038 | .038 | .044 | .024 | .024 | .030 | .019 | .022 | .020 | .014 | .008 | .014 | .011 | .022 | | | | | | |
| 29. DsecC2 | .553 | .560 | .542 | .539 | .554 | .538 | .556 | .578 | .041 | .041 | .041 | .041 | .041 | .041 | .041 | .047 | .027 | .027 | .033 | .022 | .024 | .023 | .016 | .011 | .016 | .013 | .025 | .003 | | | | | |
| 30. DsecC3 | .553 | .555 | .537 | .534 | .550 | .533 | .556 | .573 | .038 | .038 | .038 | .038 | .038 | .038 | .038 | .044 | .024 | .025 | .030 | .019 | .022 | .020 | .014 | .008 | .014 | .011 | .022 | .000 | .003 | | | | |
| 31. DsecC4 | .553 | .555 | .537 | .534 | .550 | .533 | .556 | .573 | .038 | .038 | .038 | .038 | .038 | .038 | .038 | .044 | .024 | .024 | .030 | .019 | .022 | .020 | .014 | .008 | .013 | .011 | .022 | .000 | .003 | .000 | | | |
| 32. DsecC5 | .557 | .559 | .541 | .538 | .553 | .537 | .560 | .576 | .041 | .041 | .041 | .041 | .041 | .041 | .041 | .047 | .027 | .027 | .033 | .022 | .024 | .023 | .016 | .011 | .016 | .014 | .025 | .003 | .005 | .003 | .003 | | |
| 33. DsecC6 | .549 | .551 | .533 | .530 | .545 | .529 | .552 | .568 | .038 | .038 | .038 | .038 | .038 | .038 | .038 | .044 | .024 | .024 | .030 | .019 | .021 | .020 | .013 | .008 | .013 | .011 | .022 | .000 | .003 | .000 | .000 | .003 | |
| 34. DsecC7 | .543 | .545 | .533 | .524 | .540 | .523 | .546 | .563 | .044 | .044 | .044 | .044 | .044 | .044 | .044 | .050 | .030 | .030 | .035 | .019 | .027 | .025 | .019 | .014 | .019 | .016 | .027 | .005 | .008 | .005 | .005 | .008 | .005 |
| 35. DsecC8 | .564 | .567 | .548 | .538 | .561 | .537 | .568 | .584 | .044 | .044 | .044 | .044 | .044 | .044 | .044 | .050 | .033 | .033 | .038 | .027 | .030 | .028 | .022 | .016 | .022 | .019 | .030 | .008 | .011 | .008 | .008 | .005 | .008 |
| 36. DsecC9 | .564 | .567 | .548 | .538 | .561 | .537 | .568 | .584 | .044 | .044 | .044 | .044 | .044 | .044 | .044 | .050 | .033 | .033 | .038 | .027 | .030 | .028 | .022 | .016 | .022 | .019 | .030 | .008 | .011 | .008 | .008 | .005 | .008 |
| 37. DsecC10 | .569 | .571 | .553 | .550 | .565 | .548 | .572 | .589 | .047 | .047 | .047 | .047 | .047 | .047 | .047 | .052 | .033 | .033 | .038 | .027 | .030 | .028 | .022 | .016 | .022 | .019 | .030 | .008 | .011 | .008 | .008 | .011 | .008 |
| 38. DsecC11 | .562 | .576 | .557 | .543 | .558 | .541 | .565 | .593 | .050 | .050 | .050 | .049 | .049 | .047 | .050 | .053 | .036 | .036 | .041 | .030 | .033 | .031 | .025 | .019 | .025 | .022 | .033 | .011 | .013 | .011 | .011 | .013 | .011 |
| 39. DsecC12 | .554 | .561 | .542 | .540 | .555 | .538 | .557 | .579 | .050 | .050 | .050 | .050 | .050 | .050 | .050 | .056 | .036 | .036 | .041 | .030 | .033 | .026 | .025 | .019 | .025 | .022 | .033 | .011 | .014 | .011 | .011 | .014 | .011 |
| 40. DsecC13 | .581 | .573 | .564 | .555 | .577 | .554 | .584 | .585 | .057 | .057 | .057 | .057 | .057 | .057 | .057 | .063 | .043 | .043 | .052 | .043 | .046 | .031 | .030 | .033 | .038 | .035 | .047 | .030 | .033 | .030 | .030 | .033 | .030 |
| 41. DsecC14 | .565 | .561 | .548 | .546 | .561 | .544 | .568 | .578 | .055 | .055 | .055 | .055 | .055 | .055 | .055 | .061 | .041 | .041 | .046 | .035 | .038 | .031 | .024 | .024 | .027 | .027 | .038 | .016 | .019 | .016 | .016 | .019 | .016 |
| 42. DsecC15 | .613 | .622 | .601 | .592 | .615 | .591 | .618 | .640 | .061 | .061 | .061 | .061 | .061 | .061 | .062 | .068 | .046 | .047 | .052 | .038 | .043 | .042 | .035 | .029 | .035 | .032 | .044 | .020 | .023 | .020 | .020 | .023 | .020 |
| 43. ZtubMD | .611 | .602 | .580 | .585 | .595 | .584 | .612 | .607 | .021 | .023 | .019 | .021 | .023 | .023 | .012 | .020 | .040 | .040 | .047 | .042 | .045 | .043 | .047 | .044 | .050 | .047 | .055 | .047 | .050 | .047 | .047 | .049 | .046 |
| 44. Ztub1 | .651 | .665 | .616 | .642 | .638 | .640 | .642 | .663 | .339 | .339 | .339 | .339 | .343 | .338 | .310 | .318 | .332 | .341 | .338 | .306 | .316 | .308 | .302 | .303 | .311 | .303 | .316 | .303 | .307 | .300 | .303 | .308 | .302 |
| 45. Ztub2 | .737 | .771 | .699 | .728 | .723 | .726 | .728 | .766 | .394 | .394 | .394 | .394 | .398 | .393 | .365 | .374 | .387 | .396 | .393 | .357 | .372 | .366 | .357 | .358 | .367 | .358 | .372 | .358 | .363 | .355 | .358 | .364 | .357 |
| 46. Ztub3 | .739 | .757 | .701 | .731 | .725 | .729 | .731 | .752 | .412 | .412 | .412 | .412 | .416 | .411 | .384 | .392 | .405 | .415 | .411 | .380 | .391 | .385 | .375 | .376 | .386 | .376 | .390 | .377 | .381 | .374 | .377 | .382 | .375 |
| 47. Ztub4 | .650 | .664 | .616 | .641 | .638 | .640 | .641 | .662 | .339 | .339 | .339 | .339 | .343 | .338 | .310 | .317 | .332 | .341 | .338 | .306 | .316 | .308 | .302 | .303 | .311 | .303 | .315 | .303 | .307 | .300 | .303 | .308 | .302 |
| 48. Ztub5 | .655 | .669 | .620 | .646 | .642 | .644 | .646 | .667 | .342 | .342 | .342 | .342 | .346 | .342 | .313 | .321 | .335 | .344 | .342 | .310 | .319 | .312 | .306 | .306 | .315 | .306 | .319 | .307 | .311 | .303 | .307 | .311 | .305 |
| 49. Zcam1 | .627 | .622 | .592 | .638 | .621 | .636 | .619 | .621 | .361 | .361 | .361 | .364 | .368 | .360 | .320 | .328 | .350 | .360 | .358 | .314 | .330 | .320 | .306 | .307 | .318 | .303 | .313 | .307 | .311 | .311 | .313 | .315 | .312 |
| 50. Zcam2 | .620 | .628 | .591 | .631 | .614 | .629 | .612 | .627 | .354 | .354 | .354 | .358 | .362 | .354 | .314 | .321 | .344 | .354 | .351 | .307 | .323 | .309 | .300 | .300 | .312 | .296 | .307 | .300 | .304 | .304 | .307 | .308 | .306 |
| 51. Zcam3 | .665 | .655 | .616 | .657 | .653 | .655 | .657 | .653 | .354 | .354 | .354 | .354 | .358 | .353 | .321 | .328 | .351 | .356 | .353 | .317 | .322 | .325 | .317 | .318 | .326 | .314 | .326 | .318 | .322 | .315 | .318 | .323 | .317 |
| 52. Zcam4 | .660 | .653 | .615 | .655 | .648 | .654 | .655 | .651 | .350 | .350 | .350 | .350 | .354 | .349 | .320 | .328 | .347 | .352 | .349 | .316 | .321 | .324 | .317 | .317 | .325 | .313 | .326 | .318 | .322 | .315 | .318 | .322 | .316 |
| 53. Zcam5 | .665 | .659 | .620 | .661 | .653 | .659 | .661 | .656 | .352 | .352 | .352 | .352 | .356 | .352 | .316 | .324 | .349 | .354 | .352 | .313 | .317 | .320 | .313 | .313 | .321 | .309 | .322 | .314 | .318 | .311 | .314 | .318 | .313 |
| 54. Zgab1 | .653 | .660 | .621 | .645 | .641 | .643 | .645 | .663 | .346 | .346 | .346 | .346 | .350 | .345 | .299 | .307 | .335 | .348 | .345 | .304 | .313 | .301 | .296 | .297 | .305 | .293 | .305 | .297 | .301 | .294 | .297 | .301 | .296 |
| 55. Zgab2 | .655 | .667 | .628 | .647 | .643 | .645 | .647 | .671 | .343 | .343 | .343 | .343 | .347 | .343 | .297 | .304 | .333 | .345 | .343 | .301 | .310 | .299 | .294 | .294 | .302 | .290 | .302 | .294 | .298 | .292 | .294 | .299 | .293 |
| 56. Zgab3 | .655 | .667 | .628 | .646 | .643 | .645 | .646 | .670 | .352 | .352 | .352 | .352 | .356 | .351 | .306 | .313 | .341 | .354 | .351 | .310 | .319 | .308 | .302 | .303 | .311 | .299 | .311 | .303 | .307 | .300 | .303 | .308 | .302 |
| 57. Zgab4 | .642 | .659 | .620 | .633 | .635 | .632 | .633 | .663 | .337 | .337 | .337 | .337 | .341 | .336 | .290 | .298 | .326 | .338 | .336 | .290 | .304 | .292 | .287 | .288 | .295 | .284 | .296 | .288 | .292 | .285 | .288 | .292 | .287 |

## Supplementary Table 2, continuation.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 58. Zgab5 | .651 | .658 | .619 | .643 | .639 | .641 | .643 | .661 | .346 | .346 | .346 | .346 | .350 | .345 | .299 | .306 | .335 | .347 | .345 | .303 | .313 | .301 | .296 | .296 | .304 | .292 | .305 | .297 | .301 | .294 | .297 | .301 | .296 |
| 59. Zafr1 | .640 | .652 | .614 | .632 | .629 | .631 | .632 | .656 | .343 | .343 | .343 | .343 | .347 | .342 | .297 | .304 | .333 | .345 | .342 | .301 | .310 | .298 | .293 | .294 | .302 | .290 | .302 | .294 | .298 | .291 | .294 | .299 | .293 |
| 60. Zafr2 | .642 | .649 | .610 | .620 | .630 | .618 | .634 | .652 | .347 | .347 | .347 | .347 | .351 | .346 | .300 | .308 | .332 | .344 | .342 | .300 | .309 | .298 | .293 | .293 | .301 | .289 | .301 | .293 | .297 | .291 | .293 | .298 | .292 |
| 61. Zafr3 | .634 | .645 | .607 | .625 | .622 | .624 | .625 | .649 | .338 | .338 | .338 | .338 | .342 | .338 | .292 | .300 | .328 | .340 | .338 | .297 | .306 | .294 | .289 | .290 | .297 | .286 | .298 | .290 | .294 | .287 | .290 | .294 | .289 |
| 62. Zafr4 | .643 | .648 | .611 | .627 | .624 | .627 | .627 | .651 | .352 | .352 | .352 | .352 | .356 | .352 | .306 | .313 | .342 | .354 | .344 | .310 | .320 | .308 | .303 | .303 | .311 | .299 | .312 | .304 | .308 | .301 | .304 | .308 | .303 |
| 63. Zafr5 | .640 | .652 | .614 | .632 | .628 | .630 | .632 | .656 | .351 | .351 | .351 | .351 | .355 | .351 | .305 | .312 | .341 | .353 | .343 | .309 | .319 | .307 | .302 | .302 | .310 | .298 | .311 | .303 | .307 | .300 | .303 | .307 | .302 |
| 64. Zind1 | .628 | .647 | .622 | .612 | .623 | .611 | .626 | .651 | .356 | .356 | .356 | .356 | .356 | .355 | .309 | .316 | .345 | .357 | .355 | .301 | .310 | .299 | .293 | .297 | .302 | .294 | .302 | .294 | .298 | .291 | .294 | .298 | .293 |
| 65. Zind2 | .659 | .672 | .632 | .636 | .646 | .634 | .650 | .675 | .360 | .360 | .360 | .360 | .364 | .360 | .313 | .321 | .350 | .362 | .360 | .310 | .319 | .307 | .302 | .302 | .311 | .298 | .311 | .303 | .307 | .300 | .303 | .307 | .302 |
| 66. Zind3 | .637 | .656 | .631 | .621 | .631 | .620 | .635 | .659 | .372 | .372 | .372 | .372 | .372 | .372 | .321 | .328 | .362 | .374 | .372 | .308 | .322 | .311 | .305 | .309 | .313 | .305 | .314 | .306 | .310 | .303 | .306 | .310 | .305 |
| 67. Zind4 | .649 | .656 | .617 | .627 | .637 | .625 | .641 | .659 | .344 | .344 | .344 | .344 | .348 | .343 | .297 | .304 | .329 | .341 | .339 | .289 | .298 | .286 | .281 | .282 | .290 | .278 | .290 | .282 | .286 | .279 | .282 | .287 | .281 |
| 68. Zind5 | .632 | .651 | .626 | .617 | .627 | .615 | .630 | .655 | .350 | .350 | .350 | .350 | .350 | .349 | .303 | .311 | .339 | .352 | .349 | .295 | .304 | .293 | .288 | .292 | .296 | .288 | .297 | .288 | .292 | .285 | .288 | .293 | .287 |

## Supplementary Table 2, continuation.

| | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 35. DsecC8 | .013 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 36. DsecC9 | .013 | .000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 37. DsecC10 | .013 | .016 | .016 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 38. DsecC11 | .016 | .019 | .019 | .019 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 39. DsecC12 | .016 | .019 | .019 | .019 | .022 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 40. DsecC13 | .035 | .038 | .038 | .038 | .041 | .041 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 41. DsecC14 | .022 | .024 | .024 | .024 | .027 | .027 | .044 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 42. DsecC15 | .026 | .029 | .029 | .011 | .032 | .032 | .052 | .037 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 43. ZtubMD | .052 | .052 | .052 | .055 | .058 | .059 | .066 | .064 | .071 | | | | | | | | | | | | | | | | | | | | | | | | | |
| 44. Ztub1 | .308 | .308 | .308 | .308 | .315 | .318 | .319 | .320 | .342 | .344 | | | | | | | | | | | | | | | | | | | | | | | | |
| 45. Ztub2 | .364 | .364 | .364 | .364 | .372 | .376 | .375 | .377 | .377 | .396 | .049 | | | | | | | | | | | | | | | | | | | | | | | |
| 46. Ztub3 | .382 | .382 | .382 | .382 | .391 | .394 | .380 | .396 | .387 | .414 | .054 | .040 | | | | | | | | | | | | | | | | | | | | | | |
| 47. Ztub4 | .308 | .308 | .308 | .308 | .315 | .323 | .319 | .320 | .342 | .340 | .005 | .043 | .048 | | | | | | | | | | | | | | | | | | | | | |
| 48. Ztub5 | .311 | .311 | .311 | .311 | .319 | .327 | .323 | .324 | .340 | .343 | .008 | .043 | .048 | .003 | | | | | | | | | | | | | | | | | | | | |
| 49. Zcam1 | .315 | .314 | .314 | .323 | .329 | .326 | .333 | .326 | .361 | .366 | .157 | .202 | .209 | .154 | .157 | | | | | | | | | | | | | | | | | | | |
| 50. Zcam2 | .308 | .307 | .307 | .312 | .322 | .320 | .327 | .319 | .348 | .360 | .152 | .196 | .200 | .149 | .152 | .018 | | | | | | | | | | | | | | | | | | |
| 51. Zcam3 | .323 | .322 | .322 | .327 | .334 | .338 | .338 | .331 | .365 | .351 | .151 | .192 | .199 | .144 | .148 | .027 | .040 | | | | | | | | | | | | | | | | | |
| 52. Zcam4 | .322 | .321 | .321 | .327 | .334 | .337 | .330 | .330 | .365 | .351 | .155 | .196 | .196 | .148 | .151 | .029 | .043 | .003 | | | | | | | | | | | | | | | | |
| 53. Zcam5 | .318 | .317 | .317 | .323 | .330 | .333 | .334 | .326 | .361 | .349 | .151 | .192 | .199 | .144 | .148 | .029 | .043 | .000 | .003 | | | | | | | | | | | | | | | |
| 54. Zgab1 | .301 | .306 | .306 | .301 | .308 | .316 | .317 | .313 | .333 | .347 | .097 | .144 | .150 | .097 | .100 | .153 | .154 | .141 | .141 | .141 | | | | | | | | | | | | | | |
| 55. Zgab2 | .299 | .299 | .299 | .299 | .306 | .313 | .314 | .311 | .329 | .344 | .103 | .157 | .164 | .106 | .109 | .159 | .156 | .150 | .151 | .150 | .013 | | | | | | | | | | | | | |
| 56. Zgab3 | .308 | .312 | .312 | .308 | .315 | .322 | .323 | .319 | .339 | .353 | .106 | .157 | .164 | .106 | .109 | .159 | .156 | .144 | .147 | .147 | .008 | .016 | | | | | | | | | | | | |
| 57. Zgab4 | .292 | .297 | .297 | .292 | .299 | .307 | .307 | .304 | .317 | .338 | .107 | .151 | .165 | .107 | .110 | .156 | .154 | .148 | .148 | .147 | .013 | .008 | .016 | | | | | | | | | | | |
| 58. Zgab5 | .301 | .305 | .305 | .301 | .308 | .316 | .316 | .313 | .333 | .347 | .100 | .146 | .156 | .100 | .103 | .152 | .154 | .141 | .141 | .140 | .008 | .021 | .016 | .021 | | | | | | | | | | |
| 59. Zafr1 | .299 | .303 | .303 | .299 | .306 | .313 | .314 | .310 | .334 | .344 | .100 | .150 | .157 | .100 | .103 | .152 | .150 | .140 | .141 | .140 | .003 | .010 | .005 | .010 | .010 | | | | | | | | | |
| 60. Zafr2 | .298 | .302 | .302 | .298 | .305 | .312 | .313 | .310 | .334 | .348 | .101 | .152 | .158 | .101 | .104 | .153 | .154 | .142 | .142 | .141 | .003 | .013 | .008 | .013 | .011 | .003 | | | | | | | | |
| 61. Zafr3 | .294 | .299 | .299 | .294 | .301 | .309 | .309 | .306 | .329 | .340 | .097 | .147 | .153 | .097 | .100 | .149 | .146 | .137 | .137 | .137 | .005 | .013 | .008 | .013 | .008 | .003 | .005 | | | | | | | |
| 62. Zafr4 | .308 | .312 | .312 | .308 | .315 | .318 | .324 | .320 | .345 | .354 | .107 | .157 | .164 | .106 | .109 | .159 | .157 | .147 | .147 | .147 | .016 | .024 | .018 | .024 | .018 | .013 | .016 | .010 | | | | | | |
| 63. Zafr5 | .307 | .311 | .311 | .307 | .314 | .317 | .322 | .319 | .343 | .352 | .106 | .157 | .163 | .106 | .109 | .158 | .156 | .147 | .147 | .146 | .016 | .024 | .018 | .024 | .018 | .013 | .016 | .010 | .000 | | | | | |
| 64. Zind1 | .298 | .303 | .303 | .298 | .305 | .313 | .314 | .310 | .328 | .357 | .120 | .169 | .175 | .120 | .120 | .167 | .165 | .155 | .155 | .155 | .032 | .040 | .035 | .040 | .035 | .029 | .027 | .026 | .037 | .037 | | | | |
| 65. Zind2 | .307 | .312 | .312 | .307 | .315 | .322 | .323 | .319 | .335 | .362 | .099 | .150 | .156 | .099 | .102 | .155 | .153 | .144 | .144 | .143 | .021 | .030 | .024 | .030 | .024 | .019 | .016 | .016 | .027 | .027 | .030 | | | |
| 66. Zind3 | .301 | .315 | .315 | .310 | .317 | .325 | .326 | .322 | .333 | .364 | .126 | .165 | .178 | .126 | .123 | .173 | .171 | .161 | .165 | .161 | .043 | .051 | .046 | .046 | .046 | .040 | .038 | .037 | .048 | .048 | .013 | .041 | | |
| 67. Zind4 | .287 | .291 | .291 | .287 | .293 | .301 | .302 | .298 | .317 | .345 | .092 | .141 | .148 | .091 | .094 | .147 | .148 | .136 | .136 | .135 | .011 | .021 | .016 | .021 | .013 | .010 | .008 | .008 | .018 | .018 | .021 | .008 | .032 | |
| 68. Zind5 | .293 | .297 | .297 | .293 | .300 | .307 | .308 | .304 | .324 | .351 | .120 | .172 | .178 | .120 | .123 | .163 | .161 | .151 | .151 | .151 | .032 | .040 | .035 | .040 | .035 | .029 | .027 | .026 | .037 | .037 | .005 | .030 | .016 | .021 |

Symbols: A: Full-length subfamily, B: ULR-gap subfamily, C: Double-gap subfamily, *: McDonald et al. (1997) sequence, Dmel: *D. melanogaster*, Dsim: *D. simulans*, Dsec: *D. sechellia*, Ztub: *Z. tuberculatus*, Zcam: *Z. camerounensis*, Zgab: *Z. gabonicus*, Zafr: *Z. africanus*, Zind: *Z. indianus*, Dkoe: *D. koepferae*, Dbuz: *D. buzzatii*, Dser: *D. serido*, Dgou: *D. gouveai*, Dant: *D. antonietae*, Dsem: *D. seriema*, Dpac: *D. pachuca*, Dmoj: *D. mojavensis*.

**Supplementary Table 3.** Pairwise genetic distance (maximum composite likelihood distance) among the reverse transcriptase *copia* regions of *Zaprionus*, *melanogaster* and *repleta* species.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Dkoe | | | | | | | | | | | | | | | | | |
| 2. Dbuz | 0.053 | | | | | | | | | | | | | | | | |
| 3. DmelA1 | 0.227 | 0.260 | | | | | | | | | | | | | | | |
| 4. DmelA4 | 0.232 | 0.266 | 0.004 | | | | | | | | | | | | | | |
| 5. DsimA1 | 0.215 | 0.248 | 0.041 | 0.045 | | | | | | | | | | | | | |
| 6. DsimA2 | 0.232 | 0.266 | 0.004 | 0.008 | 0.045 | | | | | | | | | | | | |
| 7. DsimA3 | 0.227 | 0.260 | 0.000 | 0.004 | 0.041 | 0.004 | | | | | | | | | | | |
| 8. DsimA5 | 0.232 | 0.266 | 0.004 | 0.008 | 0.045 | 0.008 | 0.004 | | | | | | | | | | |
| 9. DsimA6 | 0.232 | 0.266 | 0.004 | 0.008 | 0.045 | 0.008 | 0.004 | 0.008 | | | | | | | | | |
| 10. DsimA7 | 0.233 | 0.267 | 0.004 | 0.008 | 0.045 | 0.008 | 0.004 | 0.008 | 0.008 | | | | | | | | |
| 11. DsimA8 | 0.239 | 0.273 | 0.008 | 0.012 | 0.049 | 0.012 | 0.008 | 0.012 | 0.012 | 0.012 | | | | | | | |
| 12. DsecA1 | 0.227 | 0.260 | 0.000 | 0.004 | 0.041 | 0.004 | 0.000 | 0.004 | 0.004 | 0.004 | 0.008 | | | | | | |
| 13. DsecA4 | 0.233 | 0.267 | 0.004 | 0.008 | 0.045 | 0.008 | 0.004 | 0.008 | 0.008 | 0.008 | 0.012 | 0.004 | | | | | |
| 14. DsecA5 | 0.233 | 0.267 | 0.004 | 0.008 | 0.045 | 0.008 | 0.004 | 0.008 | 0.008 | 0.008 | 0.012 | 0.004 | 0.008 | | | | |
| 15. DsecA6 | 0.232 | 0.266 | 0.004 | 0.008 | 0.045 | 0.008 | 0.004 | 0.008 | 0.008 | 0.004 | 0.012 | 0.004 | 0.008 | 0.008 | | | |
| 16. DsecA7 | 0.238 | 0.272 | 0.008 | 0.012 | 0.049 | 0.008 | 0.008 | 0.012 | 0.012 | 0.012 | 0.016 | 0.008 | 0.012 | 0.012 | 0.012 | | |
| 17. DsecA9 | 0.233 | 0.267 | 0.004 | 0.008 | 0.045 | 0.008 | 0.004 | 0.008 | 0.008 | 0.008 | 0.012 | 0.004 | 0.008 | 0.008 | 0.008 | 0.012 | |
| 18. DsecA10 | 0.232 | 0.255 | 0.004 | 0.008 | 0.045 | 0.008 | 0.004 | 0.008 | 0.008 | 0.008 | 0.012 | 0.004 | 0.008 | 0.008 | 0.008 | 0.012 | 0.008 |
| 19. DsecA12 | 0.238 | 0.272 | 0.008 | 0.012 | 0.049 | 0.012 | 0.008 | 0.012 | 0.012 | 0.012 | 0.016 | 0.008 | 0.012 | 0.012 | 0.012 | 0.016 | 0.012 |
| 20. DsecA13 | 0.238 | 0.272 | 0.008 | 0.012 | 0.049 | 0.012 | 0.008 | 0.012 | 0.012 | 0.012 | 0.016 | 0.008 | 0.012 | 0.012 | 0.012 | 0.016 | 0.012 |
| 21. DsecA14 | 0.238 | 0.272 | 0.008 | 0.012 | 0.049 | 0.012 | 0.008 | 0.012 | 0.012 | 0.012 | 0.016 | 0.008 | 0.012 | 0.012 | 0.012 | 0.016 | 0.012 |
| 22. DsecA15 | 0.238 | 0.272 | 0.008 | 0.012 | 0.049 | 0.012 | 0.008 | 0.012 | 0.012 | 0.012 | 0.016 | 0.008 | 0.012 | 0.012 | 0.012 | 0.016 | 0.012 |
| 23. DsecA16 | 0.238 | 0.272 | 0.008 | 0.012 | 0.049 | 0.012 | 0.008 | 0.012 | 0.012 | 0.012 | 0.016 | 0.008 | 0.012 | 0.012 | 0.012 | 0.016 | 0.012 |
| 24. DsecA20 | 0.238 | 0.271 | 0.008 | 0.012 | 0.049 | 0.012 | 0.008 | 0.012 | 0.012 | 0.008 | 0.016 | 0.008 | 0.012 | 0.012 | 0.004 | 0.016 | 0.012 |
| 25. DsecA21 | 0.238 | 0.272 | 0.008 | 0.012 | 0.041 | 0.012 | 0.008 | 0.012 | 0.012 | 0.012 | 0.016 | 0.008 | 0.012 | 0.012 | 0.012 | 0.016 | 0.004 |
| 26. DsecA22 | 0.244 | 0.278 | 0.012 | 0.016 | 0.053 | 0.016 | 0.012 | 0.016 | 0.016 | 0.016 | 0.020 | 0.012 | 0.016 | 0.016 | 0.016 | 0.020 | 0.016 |
| 27. DsecA25 | 0.235 | 0.274 | 0.012 | 0.016 | 0.054 | 0.016 | 0.012 | 0.016 | 0.016 | 0.016 | 0.020 | 0.012 | 0.016 | 0.016 | 0.016 | 0.020 | 0.016 |
| 28. DsecA26 | 0.242 | 0.278 | 0.008 | 0.012 | 0.051 | 0.012 | 0.008 | 0.012 | 0.012 | 0.013 | 0.017 | 0.008 | 0.013 | 0.013 | 0.012 | 0.017 | 0.013 |
| 29. DyakA1 | 0.266 | 0.288 | 0.142 | 0.147 | 0.132 | 0.147 | 0.142 | 0.147 | 0.147 | 0.148 | 0.153 | 0.142 | 0.148 | 0.148 | 0.147 | 0.153 | 0.148 |
| 30. Zmul | 0.246 | 0.262 | 0.142 | 0.147 | 0.142 | 0.147 | 0.142 | 0.147 | 0.147 | 0.148 | 0.153 | 0.142 | 0.148 | 0.148 | 0.147 | 0.153 | 0.148 |
| 31. Ztub | 0.262 | 0.285 | 0.089 | 0.093 | 0.071 | 0.089 | 0.089 | 0.093 | 0.093 | 0.093 | 0.098 | 0.089 | 0.093 | 0.093 | 0.093 | 0.093 | 0.093 |
| 32. Zcam | 0.232 | 0.254 | 0.058 | 0.062 | 0.041 | 0.058 | 0.058 | 0.062 | 0.062 | 0.062 | 0.066 | 0.058 | 0.062 | 0.062 | 0.062 | 0.062 | 0.062 |
| 33. Zdav | 0.250 | 0.273 | 0.084 | 0.088 | 0.066 | 0.084 | 0.084 | 0.088 | 0.088 | 0.088 | 0.093 | 0.084 | 0.088 | 0.088 | 0.088 | 0.084 | 0.088 |
| 34. Zgab | 0.240 | 0.263 | 0.075 | 0.080 | 0.058 | 0.075 | 0.075 | 0.080 | 0.080 | 0.080 | 0.085 | 0.075 | 0.080 | 0.080 | 0.080 | 0.080 | 0.080 |
| 35. Zafr | 0.227 | 0.244 | 0.062 | 0.066 | 0.045 | 0.066 | 0.062 | 0.066 | 0.066 | 0.066 | 0.071 | 0.062 | 0.066 | 0.066 | 0.066 | 0.071 | 0.066 |
| 36. Zind | 0.228 | 0.250 | 0.066 | 0.071 | 0.049 | 0.066 | 0.066 | 0.071 | 0.071 | 0.071 | 0.075 | 0.066 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 |

**Supplementary Table 3, continuation.**

| | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 19. DsecA12 | 0.012 | | | | | | | | | | | | | | | | | |
| 20. DsecA13 | 0.012 | 0.008 | | | | | | | | | | | | | | | | |
| 21. DsecA14 | 0.012 | 0.016 | 0.016 | | | | | | | | | | | | | | | |
| 22. DsecA15 | 0.012 | 0.000 | 0.008 | 0.016 | | | | | | | | | | | | | | |
| 23. DsecA16 | 0.012 | 0.016 | 0.016 | 0.016 | 0.016 | | | | | | | | | | | | | |
| 24. DsecA20 | 0.012 | 0.016 | 0.016 | 0.008 | 0.016 | 0.016 | | | | | | | | | | | | |
| 25. DsecA21 | 0.012 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | | | | | | | | | | | |
| 26. DsecA22 | 0.016 | 0.020 | 0.020 | 0.020 | 0.020 | 0.012 | 0.020 | 0.020 | | | | | | | | | | |
| 27. DsecA25 | 0.016 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.024 | | | | | | | | | |
| 28. DsecA26 | 0.012 | 0.008 | 0.000 | 0.017 | 0.008 | 0.012 | 0.017 | 0.017 | 0.021 | 0.021 | | | | | | | | |
| 29. DyakA1 | 0.147 | 0.153 | 0.153 | 0.147 | 0.153 | 0.153 | 0.152 | 0.153 | 0.157 | 0.154 | 0.158 | | | | | | | |
| 30. Zmul | 0.149 | 0.153 | 0.153 | 0.147 | 0.153 | 0.153 | 0.152 | 0.153 | 0.158 | 0.154 | 0.160 | 0.133 | | | | | | |
| 31. Ztub | 0.094 | 0.098 | 0.098 | 0.098 | 0.098 | 0.098 | 0.097 | 0.098 | 0.102 | 0.103 | 0.102 | 0.153 | 0,168 | | | | | |
| 32. Zcam | 0.062 | 0.066 | 0.066 | 0.066 | 0.066 | 0.066 | 0.066 | 0.066 | 0.071 | 0.071 | 0.069 | 0.137 | 0,142 | 0,028 | | | | |
| 33. Zdav | 0.089 | 0.093 | 0.093 | 0.093 | 0.093 | 0.093 | 0.093 | 0.093 | 0.097 | 0.098 | 0.097 | 0.157 | 0,162 | 0,049 | 0,028 | | | |
| 34. Zgab | 0.081 | 0.085 | 0.085 | 0.084 | 0.085 | 0.084 | 0.084 | 0.084 | 0.089 | 0.089 | 0.088 | 0.149 | 0,154 | 0,033 | 0,02 | 0,041 | | |
| 35. Zafr | 0.067 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.071 | 0.075 | 0.075 | 0.074 | 0.138 | 0,138 | 0,045 | 0,024 | 0,041 | 0,028 | |
| 36. Zind | 0.071 | 0.075 | 0.075 | 0.075 | 0.075 | 0.075 | 0.075 | 0.075 | 0.080 | 0.080 | 0.079 | 0.138 | 0,143 | 0,033 | 0,012 | 0,033 | 0,016 | 0,02 |

Symbols: Dmel: *D. melanogaster*, Dsim: *D. simulans*, Dyak: *D. yakuba*, Dere: *D. erecta*, Dwil: *D. willistoni*, Dmau: *D. mauritiana*, Dsec: *D. sechellia*, Zmul: *Z. multistriatus*, Ztub: *Z. tuberculatus*, Zcam: *Z. camerounensis*, Zdav: *Z. davidi*, Zgab: *Z. gabonicus*, Zafr: *Z. africanus*, Zind: *Z. indianus*, Dkoe: *D. koepferae*, Dbuz: *D. buzzatii.*

**Supplementary Table 4.** dN (below) and dS (above) values of pairwise comparisons among *Zaprionus*, *melanogaster* and *repleta* RT *copia* sequences. Distances calculated by Nei-Gojobori method with Jukes-Cantor's correction, as implemented by MEGA 4.1.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Dkoe | - | 0.151 | 0.659 | 0.659 | 0.608 | 0.666 | 0.659 | 0.656 | 0.706 | 0.659 | 0.656 | 0.659 | 0.664 | 0.659 | 0.659 | 0.659 | 0.643 | 0.649 |
| 2. Dbuz | 0.027 | - | 0.824 | 0.824 | 0.76 | 0.833 | 0.824 | 0.819 | 0.882 | 0.824 | 0.838 | 0.824 | 0.831 | 0.824 | 0.824 | 0.824 | 0.806 | 0.838 |
| 3. DmelA1 | 0.123 | 0.134 | - | 0 | 0.122 | 0 | 0 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0 | 0 | 0 | 0 | 0 |
| 4. DmelA4 | 0.13 | 0.141 | 0.006 | - | 0.122 | 0 | 0 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0 | 0 | 0 | 0 | 0 |
| 5. DsimA1 | 0.124 | 0.134 | 0.018 | 0.024 | - | 0.123 | 0.122 | 0.111 | 0.145 | 0.122 | 0.122 | 0.122 | 0.145 | 0.122 | 0.122 | 0.122 | 0.123 | 0.121 |
| 6. DsimA2 | 0.13 | 0.141 | 0.006 | 0.012 | 0.024 | - | 0 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0 | 0 | 0 | 0 | 0 |
| 7. DsimA3 | 0.123 | 0.134 | 0 | 0.006 | 0.018 | 0.006 | - | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0 | 0 | 0 | 0 | 0 |
| 8. DsimA5 | 0.13 | 0.141 | 0.006 | 0.012 | 0.027 | 0.012 | 0.006 | - | 0.019 | 0 | 0 | 0 | 0.019 | 0 | 0 | 0 | 0 | 0 |
| 9. DsimA6 | 0.123 | 0.134 | 0 | 0.006 | 0.018 | 0.006 | 0 | 0.006 | - | 0.019 | 0.019 | 0.019 | 0.039 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
| 10. DsimA7 | 0.123 | 0.134 | 0 | 0.006 | 0.018 | 0.006 | 0 | 0.006 | 0 | - | 0 | 0 | 0.019 | 0 | 0 | 0 | 0 | 0 |
| 11. DsimA8 | 0.13 | 0.139 | 0.006 | 0.012 | 0.024 | 0.012 | 0.006 | 0.012 | 0.006 | 0.006 | - | 0 | 0.019 | 0 | 0 | 0 | 0 | 0 |
| 12. DsecA1 | 0.123 | 0.134 | 0 | 0.006 | 0.018 | 0.006 | 0 | 0.006 | 0 | 0 | 0.006 | - | 0.019 | 0 | 0 | 0 | 0 | 0 |
| 13. DsecA4 | 0.13 | 0.141 | 0 | 0.006 | 0.018 | 0.006 | 0 | 0.006 | 0 | 0 | 0.006 | 0 | - | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
| 14. DsecA5 | 0.13 | 0.141 | 0.006 | 0.012 | 0.024 | 0.012 | 0.006 | 0.012 | 0.006 | 0.006 | 0.012 | 0.006 | 0.006 | - | 0 | 0 | 0 | 0 |
| 15. DsecA6 | 0.123 | 0.134 | 0 | 0.006 | 0.018 | 0.006 | 0 | 0.006 | 0 | 0 | 0.006 | 0 | 0 | 0.006 | - | 0 | 0 | 0 |
| 16. DsecA7 | 0.137 | 0.148 | 0.012 | 0.018 | 0.03 | 0.012 | 0.012 | 0.018 | 0.012 | 0.012 | 0.018 | 0.012 | 0.012 | 0.018 | 0.012 | - | 0 | 0 |
| 17. DsecA9 | 0.133 | 0.144 | 0.006 | 0.012 | 0.024 | 0.012 | 0.006 | 0.012 | 0.006 | 0.006 | 0.012 | 0.006 | 0.006 | 0.012 | 0.006 | 0.018 | - | 0 |
| 18. DsecA10 | 0.131 | 0.124 | 0.006 | 0.012 | 0.024 | 0.012 | 0.006 | 0.012 | 0.006 | 0.006 | 0.012 | 0.006 | 0.006 | 0.012 | 0.006 | 0.018 | 0.012 | - |
| 19. DsecA12 | 0.13 | 0.141 | 0.006 | 0.012 | 0.024 | 0.012 | 0.006 | 0.012 | 0.006 | 0.006 | 0.012 | 0.006 | 0.006 | 0.012 | 0.006 | 0.018 | 0.012 | 0.012 |
| 20. DsecA13 | 0.123 | 0.134 | 0 | 0.006 | 0.018 | 0.006 | 0 | 0.006 | 0 | 0 | 0.006 | 0 | 0 | 0.006 | 0 | 0.012 | 0.006 | 0.006 |
| 21. DsecA14 | 0.133 | 0.147 | 0.012 | 0.018 | 0.03 | 0.018 | 0.012 | 0.018 | 0.012 | 0.012 | 0.018 | 0.012 | 0.012 | 0.018 | 0.012 | 0.024 | 0.018 | 0.018 |
| 22. DsecA15 | 0.13 | 0.141 | 0.006 | 0.012 | 0.024 | 0.012 | 0.006 | 0.012 | 0.006 | 0.006 | 0.012 | 0.006 | 0.006 | 0.012 | 0.006 | 0.018 | 0.012 | 0.012 |
| 23. DsecA16 | 0.137 | 0.148 | 0.012 | 0.018 | 0.03 | 0.018 | 0.012 | 0.018 | 0.012 | 0.012 | 0.018 | 0.012 | 0.012 | 0.018 | 0.012 | 0.024 | 0.018 | 0.018 |
| 24. DsecA20 | 0.13 | 0.141 | 0.006 | 0.012 | 0.024 | 0.012 | 0.006 | 0.012 | 0.006 | 0.006 | 0.012 | 0.006 | 0.006 | 0.012 | 0.006 | 0.018 | 0.012 | 0.012 |
| 25. DsecA21 | 0.14 | 0.151 | 0.012 | 0.018 | 0.018 | 0.018 | 0.012 | 0.018 | 0.012 | 0.012 | 0.018 | 0.012 | 0.012 | 0.018 | 0.012 | 0.024 | 0.006 | 0.018 |
| 26. DsecA22 | 0.138 | 0.149 | 0.012 | 0.018 | 0.03 | 0.018 | 0.012 | 0.018 | 0.012 | 0.012 | 0.018 | 0.012 | 0.012 | 0.018 | 0.012 | 0.024 | 0.018 | 0.018 |
| 27. DsecA25 | 0.141 | 0.152 | 0.012 | 0.018 | 0.03 | 0.018 | 0.012 | 0.018 | 0.012 | 0.012 | 0.018 | 0.012 | 0.012 | 0.018 | 0.012 | 0.024 | 0.018 | 0.018 |
| 28. DsecA26 | 0.124 | 0.135 | 0 | 0.006 | 0.019 | 0.006 | 0 | 0.006 | 0 | 0 | 0.006 | 0 | 0 | 0.006 | 0 | 0.012 | 0.006 | 0.006 |
| 29. DyakA1 | 0.124 | 0.138 | 0.045 | 0.051 | 0.051 | 0.051 | 0.045 | 0.051 | 0.045 | 0.045 | 0.051 | 0.045 | 0.045 | 0.051 | 0.045 | 0.057 | 0.051 | 0.051 |
| 30. Zmul | 0.149 | 0.151 | 0.067 | 0.073 | 0.073 | 0.073 | 0.067 | 0.073 | 0.067 | 0.067 | 0.07 | 0.067 | 0.067 | 0.073 | 0.067 | 0.079 | 0.076 | 0.07 |
| 31. Ztub | 0.16 | 0.171 | 0.048 | 0.055 | 0.055 | 0.048 | 0.048 | 0.058 | 0.048 | 0.048 | 0.055 | 0.048 | 0.048 | 0.055 | 0.048 | 0.055 | 0.054 | 0.055 |
| 32. Zcam | 0.117 | 0.127 | 0.012 | 0.018 | 0.018 | 0.012 | 0.012 | 0.021 | 0.012 | 0.012 | 0.018 | 0.012 | 0.012 | 0.018 | 0.012 | 0.018 | 0.018 | 0.018 |
| 33. Zdav | 0.131 | 0.142 | 0.024 | 0.03 | 0.03 | 0.024 | 0.024 | 0.033 | 0.024 | 0.024 | 0.03 | 0.024 | 0.024 | 0.03 | 0.024 | 0.024 | 0.03 | 0.03 |
| 34. Zgab | 0.13 | 0.141 | 0.036 | 0.042 | 0.042 | 0.036 | 0.036 | 0.045 | 0.036 | 0.036 | 0.042 | 0.036 | 0.036 | 0.042 | 0.036 | 0.042 | 0.042 | 0.042 |
| 35. Zafr | 0.117 | 0.128 | 0.018 | 0.024 | 0.024 | 0.024 | 0.018 | 0.027 | 0.018 | 0.018 | 0.024 | 0.018 | 0.018 | 0.024 | 0.018 | 0.03 | 0.024 | 0.024 |
| 36. Zind | 0.117 | 0.128 | 0.021 | 0.027 | 0.027 | 0.021 | 0.021 | 0.03 | 0.021 | 0.021 | 0.027 | 0.021 | 0.021 | 0.027 | 0.021 | 0.027 | 0.027 | 0.027 |

**Supplementary Table 4, continuation.**

| | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Dkoe | 0.713 | 0.755 | 0.689 | 0.713 | 0.656 | 0.659 | 0.645 | 0.688 | 0.597 | 0.836 | 0.918 | 0.478 | 0.639 | 0.698 | 0.746 | 0.656 | 0.649 | 0.649 |
| 2. Dbuz | 0.892 | 0.944 | 0.843 | 0.892 | 0.819 | 0.824 | 0.808 | 0.857 | 0.801 | 1.065 | 0.981 | 0.534 | 0.795 | 0.871 | 0.933 | 0.819 | 0.758 | 0.81 |
| 3. DmelA1 | 0.019 | 0.039 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0.019 | 0.041 | 0.509 | 0.36 | 0.213 | 0.192 | 0.298 | 0.193 | 0.191 | 0.203 |
| 4. DmelA4 | 0.019 | 0.039 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0.019 | 0.041 | 0.509 | 0.36 | 0.213 | 0.192 | 0.298 | 0.193 | 0.191 | 0.203 |
| 5. DsimA1 | 0.146 | 0.168 | 0.123 | 0.146 | 0.122 | 0.122 | 0.123 | 0.142 | 0.145 | 0.179 | 0.369 | 0.298 | 0.119 | 0.1 | 0.192 | 0.1 | 0.099 | 0.11 |
| 6. DsimA2 | 0.019 | 0.039 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0.019 | 0.042 | 0.514 | 0.363 | 0.215 | 0.193 | 0.301 | 0.194 | 0.193 | 0.205 |
| 7. DsimA3 | 0.019 | 0.039 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0.019 | 0.041 | 0.509 | 0.36 | 0.213 | 0.192 | 0.298 | 0.193 | 0.191 | 0.203 |
| 8. DsimA5 | 0.019 | 0.039 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0.019 | 0.041 | 0.507 | 0.358 | 0.2 | 0.179 | 0.283 | 0.18 | 0.179 | 0.19 |
| 9. DsimA6 | 0.039 | 0.059 | 0.019 | 0.039 | 0.019 | 0.019 | 0.019 | 0.038 | 0.039 | 0.063 | 0.548 | 0.391 | 0.238 | 0.217 | 0.327 | 0.217 | 0.216 | 0.228 |
| 10. DsimA7 | 0.019 | 0.039 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0.019 | 0.041 | 0.509 | 0.36 | 0.213 | 0.192 | 0.298 | 0.193 | 0.191 | 0.203 |
| 11. DsimA8 | 0.019 | 0.039 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0.019 | 0.041 | 0.507 | 0.374 | 0.212 | 0.191 | 0.297 | 0.192 | 0.19 | 0.203 |
| 12. DsecA1 | 0.019 | 0.039 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0.019 | 0.041 | 0.509 | 0.36 | 0.213 | 0.192 | 0.298 | 0.193 | 0.191 | 0.203 |
| 13. DsecA4 | 0.039 | 0.059 | 0.019 | 0.039 | 0.019 | 0.019 | 0.019 | 0.038 | 0.039 | 0.063 | 0.551 | 0.394 | 0.239 | 0.218 | 0.329 | 0.219 | 0.217 | 0.23 |
| 14. DsecA5 | 0.019 | 0.039 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0.019 | 0.041 | 0.509 | 0.36 | 0.213 | 0.192 | 0.298 | 0.193 | 0.191 | 0.203 |
| 15. DsecA6 | 0.019 | 0.039 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0.019 | 0.041 | 0.509 | 0.36 | 0.213 | 0.192 | 0.298 | 0.193 | 0.191 | 0.203 |
| 16. DsecA7 | 0.019 | 0.039 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0.019 | 0.041 | 0.509 | 0.36 | 0.213 | 0.192 | 0.298 | 0.193 | 0.191 | 0.203 |
| 17. DsecA9 | 0.019 | 0.039 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0.019 | 0.042 | 0.514 | 0.347 | 0.215 | 0.193 | 0.301 | 0.194 | 0.193 | 0.205 |
| 18. DsecA10 | 0.019 | 0.038 | 0 | 0.019 | 0 | 0 | 0 | 0.019 | 0.019 | 0.041 | 0.502 | 0.371 | 0.211 | 0.19 | 0.295 | 0.19 | 0.189 | 0.201 |
| 19. DsecA12 | - | 0.019 | 0.019 | 0 | 0.019 | 0.019 | 0.019 | 0.038 | 0.039 | 0.02 | 0.553 | 0.394 | 0.24 | 0.218 | 0.33 | 0.219 | 0.217 | 0.23 |
| 20. DsecA13 | 0.006 | - | 0.039 | 0.019 | 0.039 | 0.039 | 0.039 | 0.058 | 0.059 | 0 | 0.588 | 0.424 | 0.264 | 0.242 | 0.357 | 0.243 | 0.228 | 0.254 |
| 21. DsecA14 | 0.018 | 0.012 | - | 0.019 | 0 | 0 | 0 | 0.019 | 0.019 | 0.042 | 0.495 | 0.347 | 0.215 | 0.193 | 0.301 | 0.194 | 0.193 | 0.205 |
| 22. DsecA15 | 0 | 0.006 | 0.018 | - | 0.019 | 0.019 | 0.019 | 0.038 | 0.039 | 0.02 | 0.553 | 0.394 | 0.24 | 0.218 | 0.33 | 0.219 | 0.217 | 0.23 |
| 23. DsecA16 | 0.018 | 0.012 | 0.024 | 0.018 | - | 0 | 0 | 0.019 | 0.019 | 0.041 | 0.507 | 0.358 | 0.212 | 0.191 | 0.297 | 0.192 | 0.19 | 0.203 |
| 24. DsecA20 | 0.012 | 0.006 | 0.006 | 0.012 | 0.018 | - | 0 | 0.019 | 0.019 | 0.041 | 0.509 | 0.36 | 0.213 | 0.192 | 0.298 | 0.193 | 0.191 | 0.203 |
| 25. DsecA21 | 0.018 | 0.012 | 0.023 | 0.018 | 0.024 | 0.018 | - | 0.019 | 0.019 | 0.042 | 0.515 | 0.348 | 0.215 | 0.194 | 0.301 | 0.194 | 0.193 | 0.205 |
| 26. DsecA22 | 0.018 | 0.012 | 0.024 | 0.018 | 0.012 | 0.018 | 0.024 | - | 0.038 | 0.062 | 0.535 | 0.383 | 0.234 | 0.213 | 0.321 | 0.213 | 0.212 | 0.224 |
| 27. DsecA25 | 0.018 | 0.012 | 0.024 | 0.018 | 0.024 | 0.018 | 0.024 | 0.024 | - | 0.063 | 0.512 | 0.361 | 0.239 | 0.217 | 0.328 | 0.218 | 0.217 | 0.229 |
| 28. DsecA26 | 0.006 | 0 | 0.012 | 0.006 | 0.006 | 0.006 | 0.012 | 0.012 | 0.012 | - | 0.598 | 0.423 | 0.282 | 0.259 | 0.385 | 0.26 | 0.244 | 0.272 |
| 29. DyakA1 | 0.051 | 0.045 | 0.054 | 0.051 | 0.057 | 0.051 | 0.057 | 0.058 | 0.057 | 0.047 | - | 0.287 | 0.394 | 0.434 | 0.509 | 0.371 | 0.4 | 0.384 |
| 30. Zmul | 0.073 | 0.067 | 0.076 | 0.073 | 0.08 | 0.073 | 0.082 | 0.08 | 0.079 | 0.07 | 0.076 | - | 0.293 | 0.327 | 0.391 | 0.272 | 0.256 | 0.283 |
| 31. Ztub | 0.054 | 0.048 | 0.061 | 0.054 | 0.061 | 0.055 | 0.061 | 0.061 | 0.061 | 0.045 | 0.07 | 0.106 | - | 0.019 | 0.058 | 0 | 0.038 | 0.009 |
| 32. Zcam | 0.018 | 0.012 | 0.024 | 0.018 | 0.024 | 0.018 | 0.024 | 0.024 | 0.024 | 0.012 | 0.045 | 0.067 | 0.036 | - | 0.079 | 0.019 | 0.058 | 0.028 |
| 33. Zdav | 0.03 | 0.024 | 0.036 | 0.03 | 0.036 | 0.03 | 0.036 | 0.036 | 0.036 | 0.025 | 0.057 | 0.079 | 0.055 | 0.018 | - | 0.059 | 0.1 | 0.069 |
| 34. Zgab | 0.042 | 0.036 | 0.048 | 0.042 | 0.048 | 0.042 | 0.048 | 0.048 | 0.048 | 0.038 | 0.07 | 0.093 | 0.048 | 0.024 | 0.042 | - | 0.038 | 0 |
| 35. Zafr | 0.024 | 0.021 | 0.03 | 0.024 | 0.03 | 0.024 | 0.03 | 0.03 | 0.03 | 0.022 | 0.051 | 0.076 | 0.055 | 0.018 | 0.03 | 0.03 | - | 0.038 |
| 36. Zind | 0.027 | 0.021 | 0.033 | 0.027 | 0.033 | 0.027 | 0.033 | 0.033 | 0.033 | 0.022 | 0.054 | 0.076 | 0.045 | 0.009 | 0.027 | 0.024 | 0.018 | - |

Symbols: Dmel: *D. melanogaster*, Dsim: *D. simulans*, Dyak: *D. yakuba*, Dere: *D. erecta*, Dwil: *D. willistoni*, Dmau: *D. mauritiana*, Dsec: *D. sechellia*, Zmul: *Z. multistriatus*, Ztub: *Z. tuberculatus*, Zcam: *Z. camerounensis*, Zdav: *Z. davidi*, Zgab: *Z. gabonicus*, Zafr: *Z. africanus*, Zind: *Z. indianus*, Dkoe: *D. koepferae*, Dbuz: *D. buzzatii*.

**Supplementary Table 5.** Z (above) and significance P-values (below) for the Z-test of selection between *Zaprionus*, *melanogaster* and *repleta copia* RT sequences. Test performed using an alternative hypothesis of purifying selection (dN < dS) and the Nei -Gojobori distance (Jukes-Cantor correction). Gray cells correspond to significant (P < 0.05) pairwise comparisons.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Dkoep | - | 2.134 | 3.221 | 3.179 | 3.160 | 3.179 | 3.221 | 3.179 | 3.265 | 3.221 | 3.179 | 3.221 | 3.179 | 3.179 | 3.221 | 3.137 | 3.130 | 3.179 |
| 2. Dbuzz | 0.017 | - | 3.267 | 3.234 | 3.264 | 3.226 | 3.267 | 3.239 | 3.257 | 3.267 | 3.250 | 3.267 | 3.228 | 3.234 | 3.267 | 3.201 | 3.207 | 3.328 |
| 3. DmelA1 | 0.001 | 0.001 | - | -0.999 | 2.008 | -0.999 | 0.000 | -0.999 | 0.997 | 0.000 | -0.999 | 0.000 | 0.997 | -0.999 | 0.000 | -1.411 | -0.999 | -0.999 |
| 4. DmelA4 | 0.001 | 0.001 | 1.000 | - | 1.882 | -1.411 | -0.999 | -1.411 | 0.665 | -0.999 | -1.411 | -0.999 | 0.668 | -1.411 | -0.999 | -1.727 | -1.411 | -1.411 |
| 5. DsimA1 | 0.001 | 0.001 | 0.023 | 0.031 | - | 1.886 | 2.008 | 1.689 | 2.222 | 2.008 | 1.879 | 2.008 | 2.224 | 1.882 | 2.008 | 1.756 | 1.886 | 1.875 |
| 6. DsimA2 | 0.001 | 0.001 | 1.000 | 1.000 | 0.031 | - | -0.999 | -1.411 | 0.668 | -0.999 | -1.411 | -0.999 | 0.671 | -1.411 | -0.999 | -1.411 | -1.411 | -1.411 |
| 7. DsimA3 | 0.001 | 0.001 | 1.000 | 1.000 | 0.023 | 1.000 | - | -0.999 | 0.997 | 0.000 | -0.999 | 0.000 | 0.997 | -0.999 | 0.000 | -1.411 | -0.999 | -0.999 |
| 8. DsimA5 | 0.001 | 0.001 | 1.000 | 1.000 | 0.047 | 1.000 | 1.000 | - | 0.664 | -0.999 | -1.411 | -0.999 | 0.666 | -1.411 | -0.999 | -1.727 | -1.411 | -1.411 |
| 9. DsimA6 | 0.001 | 0.001 | 0.160 | 0.254 | 0.014 | 0.253 | 0.160 | 0.254 | - | 0.997 | 0.664 | 0.997 | 1.404 | 0.665 | 0.997 | 0.359 | 0.668 | 0.661 |
| 10. DsimA7 | 0.001 | 0.001 | 1.000 | 1.000 | 0.023 | 1.000 | 1.000 | 1.000 | 0.160 | - | -0.999 | 0.000 | 0.997 | -0.999 | 0.000 | -1.411 | -0.999 | -0.999 |
| 11. DsimA8 | 0.001 | 0.001 | 1.000 | 1.000 | 0.031 | 1.000 | 1.000 | 1.000 | 0.254 | 1.000 | - | -0.999 | 0.666 | -1.411 | -0.999 | -1.727 | -1.411 | -1.411 |
| 12. DsecA1 | 0.001 | 0.001 | 1.000 | 1.000 | 0.023 | 1.000 | 1.000 | 1.000 | 0.160 | 1.000 | 1.000 | - | 0.997 | -0.999 | 0.000 | -1.411 | -0.999 | -0.999 |
| 13. DsecA4 | 0.001 | 0.001 | 0.160 | 0.253 | 0.014 | 0.252 | 0.160 | 0.253 | 0.081 | 0.160 | 0.253 | 0.160 | - | 0.668 | 0.997 | 0.363 | 0.671 | 0.663 |
| 14. DsecA5 | 0.001 | 0.001 | 1.000 | 1.000 | 0.031 | 1.000 | 1.000 | 1.000 | 0.254 | 1.000 | 1.000 | 1.000 | 0.253 | - | -0.999 | -1.727 | -1.411 | -1.411 |
| 15. DsecA6 | 0.001 | 0.001 | 1.000 | 1.000 | 0.023 | 1.000 | 1.000 | 1.000 | 0.160 | 1.000 | 1.000 | 1.000 | 0.160 | 1.000 | - | -1.411 | -0.999 | -0.999 |
| 16. DsecA7 | 0.001 | 0.001 | 1.000 | 1.000 | 0.041 | 1.000 | 1.000 | 1.000 | 0.360 | 1.000 | 1.000 | 1.000 | 0.358 | 1.000 | 1.000 | - | -1.727 | -1.727 |
| 17. DsecA9 | 0.001 | 0.001 | 1.000 | 1.000 | 0.031 | 1.000 | 1.000 | 1.000 | 0.253 | 1.000 | 1.000 | 1.000 | 0.252 | 1.000 | 1.000 | 1.000 | - | -1.411 |
| 18. DsecA10 | 0.001 | 0.001 | 1.000 | 1.000 | 0.032 | 1.000 | 1.000 | 1.000 | 0.255 | 1.000 | 1.000 | 1.000 | 0.254 | 1.000 | 1.000 | 1.000 | 1.000 | - |
| 19. DsecA12 | 0.001 | 0.001 | 0.253 | 0.358 | 0.018 | 0.356 | 0.253 | 0.359 | 0.122 | 0.253 | 0.359 | 0.253 | 0.122 | 0.358 | 0.253 | 0.467 | 0.356 | 0.361 |
| 20. DsecA13 | 0.001 | 0.001 | 0.081 | 0.122 | 0.009 | 0.122 | 0.081 | 0.123 | 0.045 | 0.081 | 0.123 | 0.081 | 0.045 | 0.122 | 0.081 | 0.174 | 0.122 | 0.123 |
| 21. DsecA14 | 0.001 | 0.001 | 1.000 | 1.000 | 0.040 | 1.000 | 1.000 | 1.000 | 0.358 | 1.000 | 1.000 | 1.000 | 0.356 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 22. DsecA15 | 0.001 | 0.001 | 0.253 | 0.358 | 0.018 | 0.356 | 0.253 | 0.359 | 0.122 | 0.253 | 0.359 | 0.253 | 0.122 | 0.358 | 0.253 | 0.467 | 0.356 | 0.361 |
| 23. DsecA16 | 0.001 | 0.001 | 1.000 | 1.000 | 0.032 | 1.000 | 1.000 | 1.000 | 0.255 | 1.000 | 1.000 | 1.000 | 0.254 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 24. DsecA20 | 0.001 | 0.001 | 1.000 | 1.000 | 0.031 | 1.000 | 1.000 | 1.000 | 0.254 | 1.000 | 1.000 | 1.000 | 0.253 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 25. DsecA21 | 0.001 | 0.001 | 1.000 | 1.000 | 0.023 | 1.000 | 1.000 | 1.000 | 0.357 | 1.000 | 1.000 | 1.000 | 0.356 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 26. DsecA22 | 0.001 | 0.001 | 0.365 | 0.478 | 0.025 | 0.475 | 0.365 | 0.479 | 0.177 | 0.365 | 0.479 | 0.365 | 0.176 | 0.478 | 0.365 | 1.000 | 0.475 | 0.482 |
| 27. DsecA25 | 0.002 | 0.001 | 0.359 | 0.469 | 0.024 | 0.466 | 0.359 | 0.470 | 0.173 | 0.359 | 0.470 | 0.359 | 0.173 | 0.469 | 0.359 | 1.000 | 0.466 | 0.473 |
| 28. DsecA26 | 0.001 | 0.001 | 0.081 | 0.122 | 0.009 | 0.122 | 0.081 | 0.123 | 0.045 | 0.081 | 0.123 | 0.081 | 0.045 | 0.122 | 0.081 | 0.174 | 0.122 | 0.123 |
| 29. DyakA1 | 0.001 | 0.001 | 0.000 | 0.001 | 0.002 | 0.001 | 0.000 | 0.001 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.001 | 0.000 | 0.001 | 0.001 | 0.001 |
| 30. Zmul | 0.007 | 0.004 | 0.004 | 0.005 | 0.011 | 0.005 | 0.004 | 0.005 | 0.003 | 0.004 | 0.004 | 0.004 | 0.003 | 0.005 | 0.004 | 0.006 | 0.006 | 0.004 |
| 31. Ztub | 0.001 | 0.001 | 0.010 | 0.012 | 0.090 | 0.010 | 0.010 | 0.018 | 0.006 | 0.010 | 0.012 | 0.010 | 0.006 | 0.012 | 0.010 | 0.012 | 0.012 | 0.013 |
| 32. Zcam | 0.001 | 0.001 | 0.004 | 0.005 | 0.040 | 0.004 | 0.004 | 0.008 | 0.003 | 0.004 | 0.005 | 0.004 | 0.003 | 0.005 | 0.004 | 0.005 | 0.005 | 0.005 |
| 33. Zdav | 0.001 | 0.001 | 0.001 | 0.002 | 0.009 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 |
| 34. Zgab | 0.001 | 0.001 | 0.012 | 0.015 | 0.115 | 0.012 | 0.012 | 0.021 | 0.007 | 0.012 | 0.015 | 0.012 | 0.007 | 0.015 | 0.012 | 0.015 | 0.015 | 0.015 |
| 35. Zafr | 0.001 | 0.001 | 0.005 | 0.007 | 0.054 | 0.007 | 0.005 | 0.010 | 0.003 | 0.005 | 0.007 | 0.005 | 0.003 | 0.007 | 0.005 | 0.009 | 0.007 | 0.007 |
| 36. Zind | 0.001 | 0.001 | 0.005 | 0.006 | 0.047 | 0.005 | 0.005 | 0.009 | 0.003 | 0.005 | 0.006 | 0.005 | 0.003 | 0.006 | 0.005 | 0.006 | 0.006 | 0.007 |

**Supplementary Table 5, continuation.**

| | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Dkoep | 3.224 | 3.294 | 3.184 | 3.224 | 3.179 | 3.179 | 3.087 | 3.190 | 3.012 | 3.294 | 3.290 | 2.498 | 3.040 | 3.307 | 3.248 | 3.179 | 3.264 | 3.264 |
| 2. Dbuzz | 3.214 | 3.223 | 3.199 | 3.214 | 3.246 | 3.234 | 3.171 | 3.223 | 3.176 | 3.223 | 3.225 | 2.679 | 3.153 | 3.300 | 3.199 | 3.239 | 3.302 | 3.313 |
| 3. DmelA1 | 0.668 | 1.404 | -1.411 | 0.668 | -0.999 | -0.999 | -1.411 | 0.346 | 0.362 | 1.404 | 3.402 | 2.710 | 2.367 | 2.685 | 3.074 | 2.298 | 2.587 | 2.621 |
| 4. DmelA4 | 0.365 | 1.169 | -1.727 | 0.365 | -1.411 | -1.411 | -1.727 | 0.057 | 0.078 | 1.169 | 3.351 | 2.640 | 2.273 | 2.588 | 3.001 | 2.200 | 2.489 | 2.526 |
| 5. DsimA1 | 2.111 | 2.414 | 1.762 | 2.111 | 1.875 | 1.882 | 2.013 | 1.978 | 1.994 | 2.414 | 2.939 | 2.312 | 1.346 | 1.762 | 2.392 | 1.208 | 1.618 | 1.684 |
| 6. DsimA2 | 0.370 | 1.171 | -1.727 | 0.370 | -1.411 | -1.411 | -1.727 | 0.064 | 0.086 | 1.171 | 3.350 | 2.647 | 2.373 | 2.686 | 3.075 | 2.303 | 2.492 | 2.623 |
| 7. DsimA3 | 0.668 | 1.404 | -1.411 | 0.668 | -0.999 | -0.999 | -1.411 | 0.346 | 0.362 | 1.404 | 3.402 | 2.710 | 2.367 | 2.685 | 3.074 | 2.298 | 2.587 | 2.621 |
| 8. DsimA5 | 0.362 | 1.168 | -1.727 | 0.362 | -1.411 | -1.411 | -1.727 | 0.053 | 0.075 | 1.168 | 3.351 | 2.636 | 2.126 | 2.451 | 2.897 | 2.046 | 2.347 | 2.389 |
| 9. DsimA6 | 1.171 | 1.713 | 0.365 | 1.171 | 0.661 | 0.665 | 0.366 | 0.932 | 0.945 | 1.713 | 3.473 | 2.844 | 2.539 | 2.837 | 3.189 | 2.477 | 2.746 | 2.776 |
| 10. DsimA7 | 0.668 | 1.404 | -1.411 | 0.668 | -0.999 | -0.999 | -1.411 | 0.346 | 0.362 | 1.404 | 3.402 | 2.710 | 2.367 | 2.685 | 3.074 | 2.298 | 2.587 | 2.621 |
| 11. DsimA8 | 0.362 | 1.168 | -1.727 | 0.362 | -1.411 | -1.411 | -1.727 | 0.053 | 0.075 | 1.168 | 3.351 | 2.742 | 2.270 | 2.587 | 3.000 | 2.197 | 2.487 | 2.525 |
| 12. DsecA1 | 0.668 | 1.404 | -1.411 | 0.668 | -0.999 | -0.999 | -1.411 | 0.346 | 0.362 | 1.404 | 3.402 | 2.710 | 2.367 | 2.685 | 3.074 | 2.298 | 2.587 | 2.621 |
| 13. DsecA4 | 1.173 | 1.713 | 0.369 | 1.173 | 0.663 | 0.668 | 0.370 | 0.935 | 0.948 | 1.713 | 3.471 | 2.848 | 2.543 | 2.837 | 3.189 | 2.480 | 2.747 | 2.777 |
| 14. DsecA5 | 0.365 | 1.169 | -1.727 | 0.365 | -1.411 | -1.411 | -1.727 | 0.057 | 0.078 | 1.169 | 3.351 | 2.640 | 2.273 | 2.588 | 3.001 | 2.200 | 2.489 | 2.526 |
| 15. DsecA6 | 0.668 | 1.404 | -1.411 | 0.668 | -0.999 | -0.999 | -1.411 | 0.346 | 0.362 | 1.404 | 3.402 | 2.710 | 2.367 | 2.685 | 3.074 | 2.298 | 2.587 | 2.621 |
| 16. DsecA7 | 0.082 | 0.942 | -1.992 | 0.082 | -1.727 | -1.727 | -1.992 | -0.213 | -0.187 | 0.942 | 3.298 | 2.570 | 2.273 | 2.588 | 3.074 | 2.200 | 2.391 | 2.526 |
| 17. DsecA9 | 0.370 | 1.171 | -1.727 | 0.370 | -1.411 | -1.411 | -0.999 | 0.064 | 0.086 | 1.171 | 3.350 | 2.537 | 2.280 | 2.590 | 3.003 | 2.206 | 2.492 | 2.530 |
| 18. DsecA10 | 0.357 | 1.166 | -1.727 | 0.357 | -1.411 | -1.411 | -1.727 | 0.046 | 0.067 | 1.166 | 3.351 | 2.735 | 2.263 | 2.585 | 2.998 | 2.191 | 2.484 | 2.521 |
| 19. DsecA12 | - | 0.668 | 0.089 | 0.000 | 0.357 | 0.365 | 0.091 | 0.714 | 0.732 | 0.668 | 3.423 | 2.785 | 2.457 | 2.748 | 3.122 | 2.392 | 2.657 | 2.691 |
| 20. DsecA13 | 0.253 | - | 0.947 | 0.668 | 1.166 | 1.169 | 0.948 | 1.329 | 1.340 | 0.000 | 3.532 | 2.966 | 2.696 | 2.975 | 3.294 | 2.640 | 2.776 | 2.918 |
| 21. DsecA14 | 0.465 | 0.173 | - | 0.089 | -1.727 | -0.999 | -1.992 | -0.205 | -0.178 | 0.947 | 3.283 | 2.537 | 2.187 | 2.494 | 2.930 | 2.110 | 2.395 | 2.436 |
| 22. DsecA15 | 1.000 | 0.253 | 0.465 | - | 0.357 | 0.365 | 0.091 | 0.714 | 0.732 | 0.668 | 3.423 | 2.785 | 2.457 | 2.748 | 3.122 | 2.392 | 2.657 | 2.691 |
| 23. DsecA16 | 0.361 | 0.123 | 1.000 | 0.361 | - | -1.411 | -1.727 | 0.653 | 0.067 | 1.166 | 3.351 | 2.629 | 2.263 | 2.585 | 2.998 | 2.191 | 2.484 | 2.521 |
| 24. DsecA20 | 0.358 | 0.122 | 1.000 | 0.358 | 1.000 | - | -1.727 | 0.057 | 0.078 | 1.169 | 3.351 | 2.640 | 2.273 | 2.588 | 3.001 | 2.200 | 2.489 | 2.526 |
| 25. DsecA21 | 0.464 | 0.173 | 1.000 | 0.464 | 1.000 | 1.000 | - | -0.202 | -0.176 | 0.948 | 3.299 | 2.468 | 2.189 | 2.495 | 2.931 | 2.112 | 2.396 | 2.437 |
| 26. DsecA22 | 0.238 | 0.093 | 1.000 | 0.238 | 0.257 | 0.478 | 1.000 | - | 0.496 | 1.329 | 3.377 | 2.695 | 2.345 | 2.649 | 3.047 | 2.278 | 2.554 | 2.590 |
| 27. DsecA25 | 0.233 | 0.091 | 1.000 | 0.233 | 0.473 | 0.469 | 1.000 | 0.310 | - | 1.340 | 3.299 | 2.574 | 2.366 | 2.657 | 3.053 | 2.298 | 2.565 | 2.602 |
| 28. DsecA26 | 0.253 | 1.000 | 0.173 | 0.253 | 0.123 | 0.122 | 0.173 | 0.093 | 0.091 | - | 3.532 | 2.966 | 2.696 | 2.975 | 3.294 | 2.640 | 2.776 | 2.918 |
| 29. DyakA1 | 0.000 | 0.000 | 0.001 | 0.000 | 0.001 | 0.001 | 0.001 | 0.000 | 0.001 | 0.000 | - | 2.371 | 2.917 | 3.227 | 3.298 | 2.737 | 3.058 | 2.966 |
| 30. Zmul | 0.003 | 0.002 | 0.006 | 0.003 | 0.005 | 0.005 | 0.008 | 0.004 | 0.006 | 0.002 | 0.010 | - | 1.956 | 2.556 | 2.712 | 1.881 | 1.986 | 2.179 |
| 31. Ztub | 0.008 | 0.004 | 0.015 | 0.008 | 0.013 | 0.012 | 0.015 | 0.010 | 0.010 | 0.004 | 0.002 | 0.026 | - | -0.489 | 0.251 | -2.625 | -0.340 | -1.478 |
| 32. Zcam | 0.003 | 0.002 | 0.007 | 0.003 | 0.005 | 0.005 | 0.007 | 0.005 | 0.004 | 0.002 | 0.001 | 0.006 | 1.000 | - | 1.487 | -0.205 | 1.143 | 0.806 |
| 33. Zdav | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.004 | 0.401 | 0.070 | - | 0.455 | 1.486 | 1.074 |
| 34. Zgab | 0.009 | 0.005 | 0.018 | 0.009 | 0.015 | 0.015 | 0.018 | 0.012 | 0.012 | 0.005 | 0.004 | 0.031 | 1.000 | 1.000 | 0.325 | - | 0.288 | -1.991 |
| 35. Zafr | 0.004 | 0.003 | 0.009 | 0.004 | 0.007 | 0.007 | 0.009 | 0.006 | 0.006 | 0.003 | 0.001 | 0.025 | 1.000 | 0.128 | 0.070 | 0.387 | - | 0.705 |
| 36. Zind | 0.004 | 0.002 | 0.008 | 0.004 | 0.007 | 0.006 | 0.008 | 0.005 | 0.005 | 0.002 | 0.002 | 0.016 | 1.000 | 0.211 | 0.143 | 1.000 | 0.241 | - |

Symbols: Dmel: *D. melanogaster*, Dsim: *D. simulans*, Dyak: *D. yakuba*, Dere: *D. erecta*, Dwil: *D. willistoni*, Dmau: *D. mauritiana*, Dsec: *D. sechellia*, Zmul: *Z. multistriatus*, Ztub: *Z. tuberculatus*, Zcam: *Z. camerounensis*, Zdav: *Z. davidi*, Zgab: *Z. gabonicus*, Zafr: *Z. africanus*, Zind: *Z. indianus*, Dkoe: *D. koepferae*, Dbuz: *D. buzzatii*.

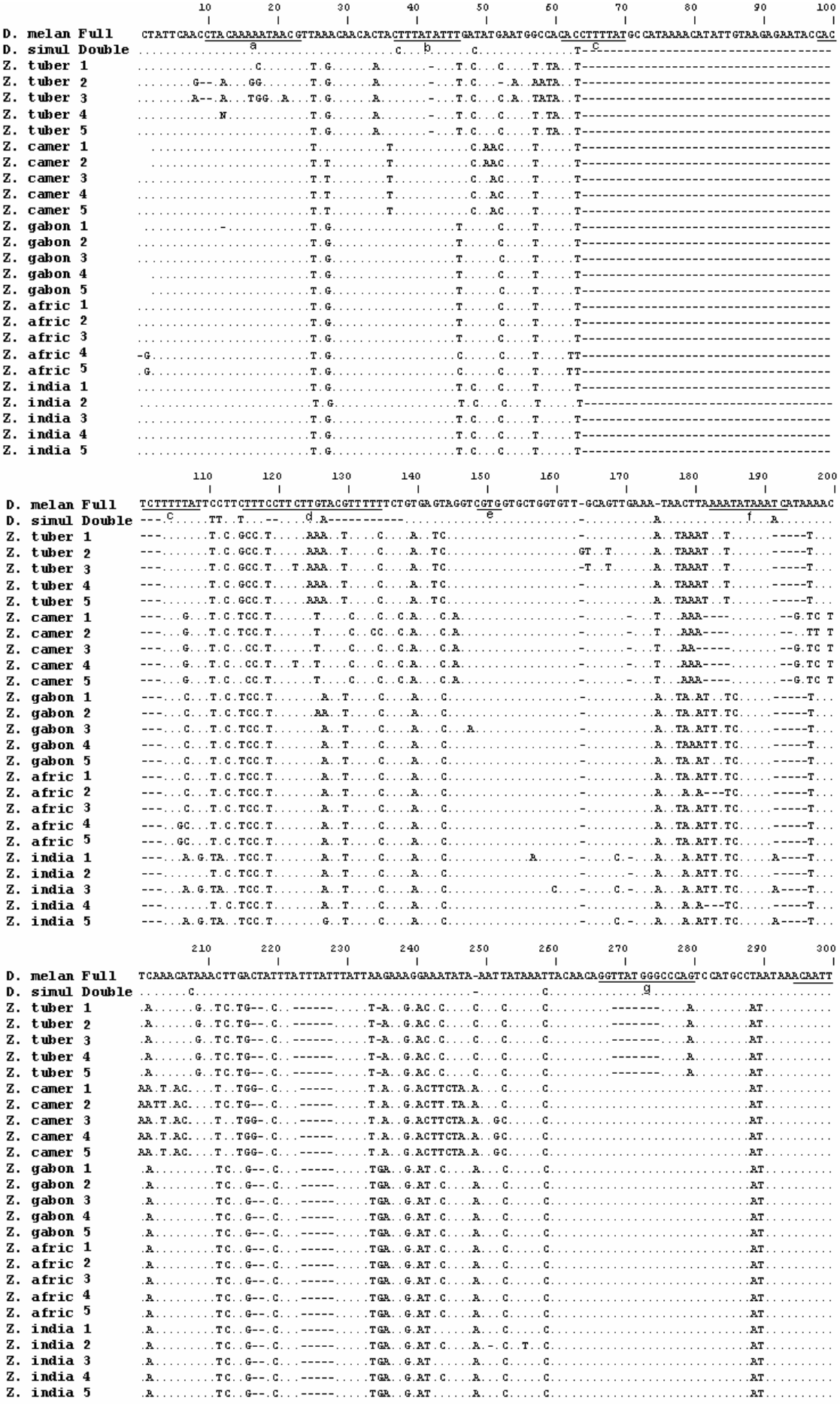
**Supplementary Table 6.** Z (above) and significance P-values (below) for the Z-test of selection between *Zaprionus*, *melanogaster* and *repleta copia* RT sequences. Test performed using an alternative hypothesis of positive selection (dN > dS) and Nei -Gojobori distance (Jukes-Cantor correction). Gray cells correspond to significant (P < 0.05) pairwise comparisons.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Dkoep | - | -2.062 | -2.192 | -2.158 | -2.309 | -2.136 | -2.192 | -2.183 | -2.069 | -2.192 | -2.171 | -2.192 | -2.138 | -2.173 | -2.192 | -2.134 | -2.172 | -2.183 |
| 2. Dbuzz | 1.000 | - | -2.080 | -2.057 | -2.305 | -2.110 | -2.080 | -2.090 | -2.061 | -2.080 | -2.084 | -2.080 | -2.001 | -2.065 | -2.080 | -2.040 | -2.095 | -2.147 |
| 3. DmelA1 | 1.000 | 1.000 | - | 1.051 | -2.079 | 1.019 | 0.000 | 0.969 | -1.060 | 0.000 | 1.064 | 0.000 | -1.015 | 1.024 | 0.000 | 1.516 | 1.039 | 1.060 |
| 4. DmelA4 | 1.000 | 1.000 | 0.148 | - | -1.928 | 1.488 | 1.051 | 1.467 | -0.699 | 1.051 | 1.477 | 1.051 | -0.683 | 1.498 | 1.051 | 1.887 | 1.484 | 1.473 |
| 5. DsimA1 | 1.000 | 1.000 | 1.000 | 1.000 | - | -1.949 | -2.079 | -1.859 | -2.242 | -2.079 | -1.936 | -2.079 | -2.310 | -1.943 | -2.079 | -1.812 | -1.958 | -1.926 |
| 6. DsimA2 | 1.000 | 1.000 | 0.155 | 0.070 | 1.000 | - | 1.019 | 1.350 | -0.708 | 1.019 | 1.494 | 1.019 | -0.692 | 1.486 | 1.019 | 1.521 | 1.373 | 1.490 |
| 7. DsimA3 | 1.000 | 1.000 | 1.000 | 0.148 | 1.000 | 0.155 | - | 0.969 | -1.060 | 0.000 | 1.064 | 0.000 | -1.015 | 1.024 | 0.000 | 1.516 | 1.039 | 1.060 |
| 8. DsimA5 | 1.000 | 1.000 | 0.167 | 0.072 | 1.000 | 0.090 | 0.167 | - | -0.690 | 0.969 | 1.437 | 0.969 | -0.681 | 1.327 | 0.969 | 1.736 | 1.421 | 1.433 |
| 9. DsimA6 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | - | -1.060 | -0.698 | -1.060 | -1.458 | -0.709 | -1.060 | -0.379 | -0.710 | -0.693 |
| 10. DsimA7 | 1.000 | 1.000 | 1.000 | 0.148 | 1.000 | 0.155 | 1.000 | 0.167 | 1.000 | - | 1.064 | 0.000 | -1.015 | 1.024 | 0.000 | 1.516 | 1.039 | 1.060 |
| 11. DsimA8 | 1.000 | 1.000 | 0.145 | 0.071 | 1.000 | 0.069 | 0.145 | 0.077 | 1.000 | 0.145 | - | 1.064 | -0.681 | 1.474 | 1.064 | 1.855 | 1.484 | 1.051 |
| 12. DsecA1 | 1.000 | 1.000 | 1.000 | 0.148 | 1.000 | 0.155 | 1.000 | 0.167 | 1.000 | 1.000 | 0.145 | - | -1.015 | 1.024 | 0.000 | 1.516 | 1.039 | 1.060 |
| 13. DsecA4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | - | -0.686 | -1.015 | -0.374 | -0.687 | -0.677 |
| 14. DsecA5 | 1.000 | 1.000 | 0.154 | 0.068 | 1.000 | 0.070 | 0.154 | 0.094 | 1.000 | 0.154 | 0.071 | 0.154 | 1.000 | - | 1.024 | 1.898 | 1.485 | 1.471 |
| 15. DsecA6 | 1.000 | 1.000 | 1.000 | 0.148 | 1.000 | 0.155 | 1.000 | 0.167 | 1.000 | 1.000 | 0.145 | 1.000 | 1.000 | 0.154 | - | 1.516 | 1.039 | 1.060 |
| 16. DsecA7 | 1.000 | 1.000 | 0.066 | 0.031 | 1.000 | 0.065 | 0.066 | 0.043 | 1.000 | 0.066 | 0.033 | 0.066 | 1.000 | 0.030 | 0.066 | - | 1.771 | 1.851 |
| 17. DsecA9 | 1.000 | 1.000 | 0.150 | 0.070 | 1.000 | 0.086 | 0.150 | 0.079 | 1.000 | 0.150 | 0.070 | 0.150 | 1.000 | 0.070 | 0.150 | 0.040 | - | 1.481 |
| 18. DsecA10 | 1.000 | 1.000 | 0.146 | 0.072 | 1.000 | 0.069 | 0.146 | 0.077 | 1.000 | 0.146 | 0.148 | 0.146 | 1.000 | 0.072 | 0.146 | 0.033 | 0.071 | - |
| 19. DsecA12 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 20. DsecA13 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 21. DsecA14 | 1.000 | 1.000 | 0.082 | 0.040 | 1.000 | 0.042 | 0.082 | 0.046 | 1.000 | 0.082 | 0.043 | 0.082 | 1.000 | 0.043 | 0.082 | 0.021 | 0.045 | 0.043 |
| 22. DsecA15 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 23. DsecA16 | 1.000 | 1.000 | 0.145 | 0.073 | 1.000 | 0.068 | 0.145 | 0.075 | 1.000 | 0.145 | 0.058 | 0.145 | 1.000 | 0.078 | 0.145 | 0.032 | 0.068 | 0.059 |
| 24. DsecA20 | 1.000 | 1.000 | 0.166 | 0.072 | 1.000 | 0.082 | 0.166 | 0.084 | 1.000 | 0.166 | 0.075 | 0.166 | 1.000 | 0.080 | 0.166 | 0.037 | 0.085 | 0.076 |
| 25. DsecA21 | 1.000 | 1.000 | 0.066 | 0.032 | 1.000 | 0.041 | 0.066 | 0.037 | 1.000 | 0.066 | 0.029 | 0.066 | 1.000 | 0.034 | 0.066 | 0.021 | 0.140 | 0.029 |
| 26. DsecA22 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.416 | 1.000 | 1.000 |
| 27. DsecA25 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.427 | 1.000 | 1.000 |
| 28. DsecA26 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 29. DyakA1 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 30. Zmul | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 31. Ztub | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 32. Zcam | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 33. Zdav | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 34. Zgab | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 35. Zafr | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 36. Zind | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

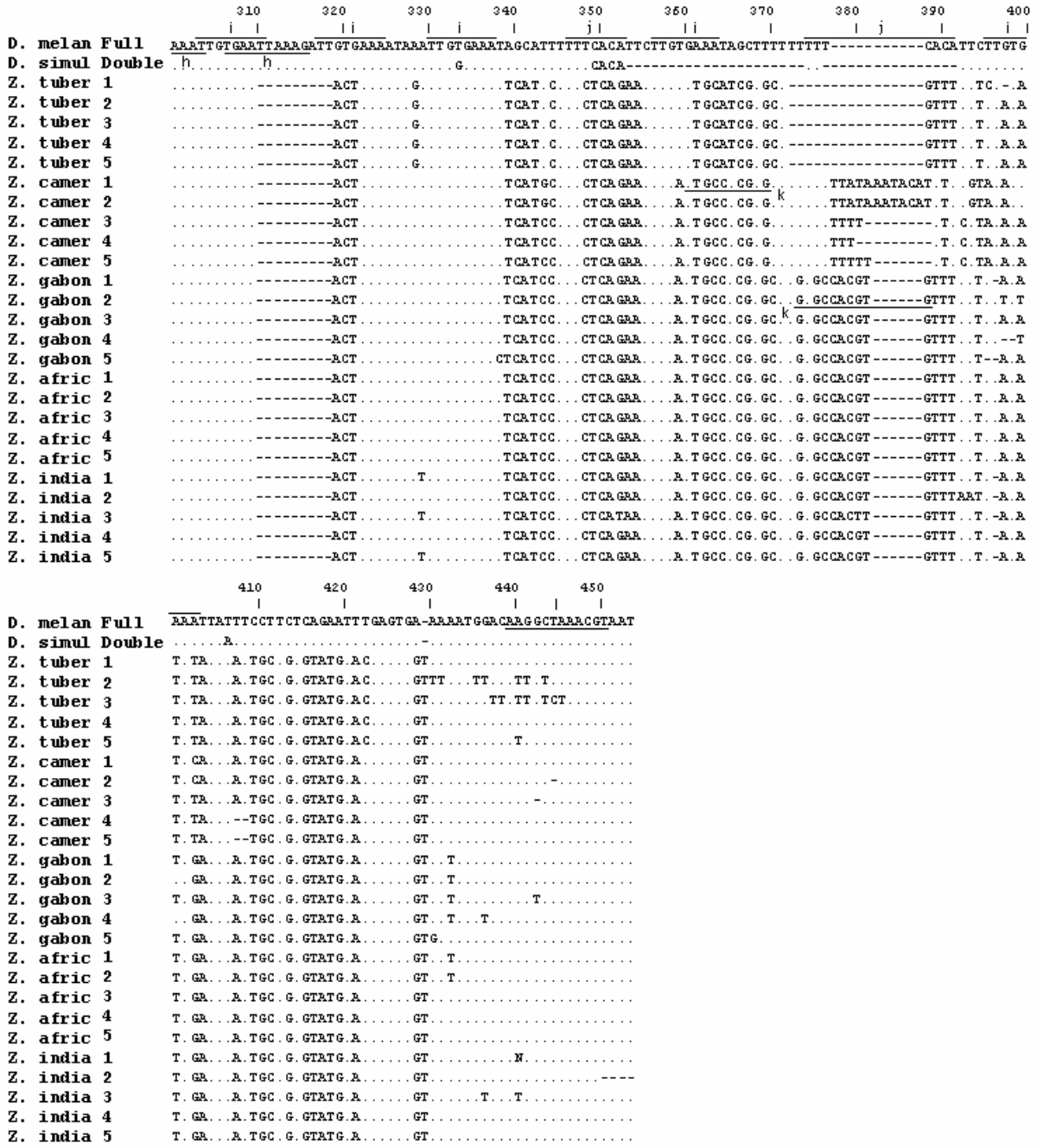
**Supplementary Table 6, continuation.**

| | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Dkoep | -2.064 | -2.121 | -2.153 | -2.064 | -2.054 | -2.159 | -2.135 | -2.048 | -1.941 | -2.121 | -2.037 | -2.368 | -2.235 | -2.316 | -2.149 | -2.336 | -2.528 | -2.405 |
| 2. Dbuzz | -2.010 | -2.085 | -2.059 | -2.010 | -2.046 | -2.057 | -2.063 | -2.077 | -2.149 | -2.085 | -2.063 | -2.411 | -2.000 | -2.131 | -2.093 | -2.071 | -2.392 | -2.070 |
| 3. DmelA1 | -0.678 | -1.343 | 1.398 | -0.678 | 1.064 | 0.974 | 1.517 | -0.351 | -0.359 | -1.343 | -2.934 | -2.410 | -1.972 | -2.241 | -2.550 | -1.883 | -2.242 | -2.226 |
| 4. DmelA4 | -0.368 | -1.110 | 1.762 | -0.368 | 1.463 | 1.472 | 1.875 | -0.057 | -0.078 | -1.110 | -2.880 | -2.334 | -1.881 | -2.142 | -2.474 | -1.789 | -2.141 | -2.128 |
| 5. DsimA1 | -2.142 | -2.420 | -1.789 | -2.142 | -1.941 | -1.924 | -2.094 | -2.046 | -1.985 | -2.420 | -2.470 | -2.043 | -1.297 | -1.695 | -2.412 | -1.182 | -1.639 | -1.760 |
| 6. DsimA2 | -0.524 | -1.295 | 1.746 | -0.524 | 1.504 | 1.398 | 1.756 | -0.064 | -0.084 | -1.295 | -2.876 | -2.349 | -1.971 | -2.236 | -2.539 | -1.884 | -2.160 | -2.223 |
| 7. DsimA3 | -0.678 | -1.343 | 1.398 | -0.678 | 1.064 | 0.974 | 1.517 | -0.351 | -0.359 | -1.343 | -2.934 | -2.410 | -1.972 | -2.241 | -2.550 | -1.883 | -2.242 | -2.226 |
| 8. DsimA5 | -0.374 | -1.122 | 1.697 | -0.374 | 1.446 | 1.386 | 1.798 | -0.053 | -0.075 | -1.122 | -2.931 | -2.381 | -1.998 | -2.344 | -2.674 | -1.907 | -2.334 | -2.323 |
| 9. DsimA6 | -1.205 | -1.697 | -0.385 | -1.205 | -0.714 | -0.708 | -0.381 | -0.959 | -0.937 | -1.697 | -2.948 | -2.482 | -2.123 | -2.402 | -2.654 | -2.043 | -2.379 | -2.368 |
| 10. DsimA7 | -0.678 | -1.343 | 1.398 | -0.678 | 1.064 | 0.974 | 1.517 | -0.351 | -0.359 | -1.343 | -2.934 | -2.410 | -1.972 | -2.241 | -2.550 | -1.883 | -2.242 | -2.226 |
| 11. DsimA8 | -0.366 | -1.115 | 1.735 | -0.366 | 1.581 | 1.446 | 1.913 | -0.055 | -0.075 | -1.115 | -2.888 | -2.476 | -1.885 | -2.155 | -2.479 | -1.795 | -2.153 | -2.135 |
| 12. DsecA1 | -0.678 | -1.343 | 1.398 | -0.678 | 1.064 | 0.974 | 1.517 | -0.351 | -0.359 | -1.343 | -2.934 | -2.410 | -1.972 | -2.241 | -2.550 | -1.883 | -2.242 | -2.226 |
| 13. DsecA4 | -1.153 | -1.610 | -0.385 | -1.153 | -0.692 | -0.696 | -0.377 | -0.999 | -0.905 | -1.610 | -2.989 | -2.544 | -2.170 | -2.464 | -2.695 | -2.086 | -2.414 | -2.421 |
| 14. DsecA5 | -0.364 | -1.105 | 1.729 | -0.364 | 1.431 | 1.412 | 1.842 | -0.057 | -0.077 | -1.105 | -2.881 | -2.348 | -1.905 | -2.171 | -2.498 | -1.811 | -2.163 | -2.155 |
| 15. DsecA6 | -0.678 | -1.343 | 1.398 | -0.678 | 1.064 | 0.974 | 1.517 | -0.351 | -0.359 | -1.343 | -2.934 | -2.410 | -1.972 | -2.241 | -2.550 | -1.883 | -2.242 | -2.226 |
| 16. DsecA7 | -0.116 | -1.055 | 2.065 | -0.116 | 1.870 | 1.808 | 2.062 | 0.213 | 0.184 | -1.055 | -2.866 | -2.289 | -1.898 | -2.166 | -2.544 | -1.810 | -2.077 | -2.152 |
| 17. DsecA9 | -0.387 | -1.128 | 1.711 | -0.387 | 1.498 | 1.382 | 1.087 | -0.064 | -0.084 | -1.128 | -2.872 | -2.321 | -1.897 | -2.159 | -2.483 | -1.808 | -2.160 | -2.146 |
| 18. DsecA10 | -0.360 | -1.111 | 1.731 | -0.360 | 1.577 | 1.443 | 1.908 | -0.047 | -0.068 | -1.111 | -2.886 | -2.475 | -1.875 | -2.148 | -2.470 | -1.784 | -2.147 | -2.126 |
| 19. DsecA12 | | -0.642 | -0.089 | 0.000 | -0.359 | -0.368 | -0.095 | -0.740 | -0.755 | -0.642 | -2.884 | -2.398 | -2.021 | -2.268 | -2.506 | -1.952 | -2.270 | -2.262 |
| 20. DsecA13 | 1.000 | | -0.897 | -0.642 | -1.111 | -1.117 | -0.912 | -1.318 | -1.320 | 0.000 | -2.889 | -2.556 | -2.206 | -2.440 | -2.575 | -2.137 | -2.430 | -2.431 |
| 21. DsecA14 | 1.000 | 1.000 | | -0.089 | 1.796 | 1.012 | 2.001 | 0.208 | 0.175 | -0.897 | -2.842 | -2.256 | -1.818 | -2.079 | -2.418 | -1.722 | -2.069 | -2.060 |
| 22. DsecA15 | 1.000 | 1.000 | 1.000 | | -0.359 | -0.368 | -0.095 | -0.740 | -0.755 | -0.642 | -2.884 | -2.398 | -2.021 | -2.268 | -2.506 | -1.952 | -2.270 | -2.262 |
| 23. DsecA16 | 1.000 | 1.000 | 0.037 | 1.000 | | 1.447 | 1.911 | -0.662 | -0.067 | -1.111 | -2.989 | -2.418 | -1.936 | -2.155 | -2.548 | -1.852 | -2.229 | -2.208 |
| 24. DsecA20 | 1.000 | 1.000 | 0.157 | 1.000 | 0.075 | | 1.754 | -0.056 | -0.077 | -1.117 | -2.869 | -2.337 | -1.891 | -2.158 | -2.484 | -1.799 | -2.155 | -2.140 |
| 25. DsecA21 | 1.000 | 1.000 | 0.024 | 1.000 | 0.029 | 0.041 | | 0.211 | 0.172 | -0.912 | -2.822 | -2.253 | -1.817 | -2.078 | -2.415 | -1.725 | -2.069 | -2.061 |
| 26. DsecA22 | 1.000 | 1.000 | 0.418 | 1.000 | 1.000 | 1.000 | 0.417 | | -0.508 | -1.318 | -2.959 | -2.397 | -1.986 | -2.219 | -2.558 | -1.950 | -2.309 | -2.261 |
| 27. DsecA25 | 1.000 | 1.000 | 0.431 | 1.000 | 1.000 | 1.000 | 0.432 | 1.000 | | -1.320 | -2.832 | -2.378 | -1.962 | -2.218 | -2.525 | -1.892 | -2.228 | -2.213 |
| 28. DsecA26 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | | -2.889 | -2.556 | -2.206 | -2.440 | -2.575 | -2.137 | -2.430 | -2.431 |
| 29. DyakA1 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | | -2.090 | -2.643 | -2.864 | -2.939 | -2.512 | -2.779 | -2.780 |
| 30. Zmul | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | | -1.735 | -2.266 | -2.393 | -1.713 | -1.936 | -2.031 |
| 31. Ztub | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | | 0.498 | -0.246 | 2.788 | 0.336 | 2.233 |
| 32. Zcam | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.310 | | -1.453 | 0.192 | -1.124 | -1.064 |
| 33. Zdav | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | | -0.448 | -1.483 | -1.145 |
| 34. Zgab | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.003 | 0.424 | 1.000 | | -0.279 | 1.673 |
| 35. Zafr | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.369 | 1.000 | 1.000 | 1.000 | | -0.696 |
| 36. Zind | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.014 | 1.000 | 1.000 | 0.048 | 1.000 | |

Symbols: Dmel: *D. melanogaster*, Dsim: *D. simulans*, Dyak: *D. yakuba*, Dere: *D. erecta*, Dwil: *D. willistoni*, Dmau: *D. mauritiana*, Dsec: *D. sechellia*, Zmul: *Z. multistriatus*, Ztub: *Z. tuberculatus*, Zcam: *Z. camerounensis*, Zdav: *Z. davidi*, Zgab: *Z. gabonicus*, Zafr: *Z. africanus*, Zind: *Z. indianus*, Dkoe: *D. koepferae*, Dbuz: *D. buzzatii*.



**Continuation.**



**Supplementary Figure 1.** Multiple alignment between 5' LTR-ULR *copia* of *Zaprionus* species and *melanogaster* species subgroup variants (Full-length and Double gap subfamilies). a: heat shock element; b: TATA box; c: imperfect repeat; d: transcription start; e: downstream element; f: poly-A signal; g: PBS; h: *engrailed* binding site; i: DmC/EBP biding site; j-i: dyad symmetric (core SV40 enhancer); k: G-box binding factor-1 (GBF-1); l: B-box binding factor-2 (BBF-2).

# 6 CAPÍTULO IV

# Comparative analyses of *gypsy* and *micropia* retrotransposons evidence similar evolution patterns in genus *Zaprionus* and the *melanogaster* species subgroup (Drosophilidae)

**Nathalia de Setta[1], Marie Anne Van Sluys[2], Pierre Capy[3], Claudia Marcia Aparecida Carareto[1]**

[1]UNESP - São Paulo State University, Department of Biology, São José do Rio Preto, SP, Brazil; [2]USP - São Paulo University, Department of Botany, SP, Brazil; [3]Laboratoire Evolution, Génomes et Spéciation UPR9034, CNRS, Gif-sur-Yvette, France.

## Abstract

The *Ty3*/*gypsy* retroelements are components of the *Drosophila* genomes with complex evolutionary patterns, alternating vertical and horizontal transfer mechanisms. Among them, the *gypsy* and *micropia* retrotransposons have been studied mainly in the *Drosophila* genus. Specifically in genus *Zaprionus*, *gypsy* has been punctually studied, and there is no information at all about *micropia* elements. In this work, we performed the identification, transcriptional activity evaluation and evolutionary analyses of seven species belonging to the two *Zaprionus* subgenera. *Zaprionus gypsy* and *micropia* retrotransposons seem to be present and transcriptional active only in the *Zaprionus* subgenus species. These sequences are highly related to those of the *melanogaster* species subgroup, and the most parsimonious hypothesis indicates that several horizontal transfer events have shaped their evolution. The results suggest that *gypsy* and *micropia* sequences share similar age and evolutionary patterns in the genus *Zaprionus* and subgroup *melanogaster*.

## Keywords

*gypsy* retrotransposon, *micropia* retrotransposon, horizontal transfer, *Drosophila*, *Zaprionus*

## 6.1 Introduction

The *gypsy* and *micropia* retrotransposons of *Drosophila* are *Ty3/gypsy* LTR retroelements characterized by the reverse transcriptase-RNAseH-integrase arrangement in the *pol* coding region (LLORENS et al., 2008a). The *Ty3/Gypsy* retroelements, which are similar to vertebrate retroviruses in both sequence and genomic structure, are broadly distributed in the eukaryotic organisms (EICKBUSH, 1994). Their main difference from the vertebrates retroviruses is that retroviruses have an additional ORF that encodes an *env* polyprotein necessary for the complete infectious cycle (LLORENS et al. 2008a). In contrast to the typical *Ty3*/*gypsy* structure of the *micropia* retrotransposon, the *gypsy* retrotransposon is an exception, due to the presence of an active *env* gene, potentially capable of maintaining the infectious activity (LLORENS et al., 2008b).

The *micropia* retrotransposon was first identified in the *Drosophila hydei*, associated to fertility genes of the Y chromosome in primary spermatocytes (HUIJSER et al., 1988). Afterwards, homologous elements have been identified in the *melanogaster*, *willistoni* and *saltans* groups of the subgenus *Sophophora* (LANKENAU, 1993; ALMEIDA et al., 2001), and in the *immigrans*, *funebris*, *mellanica*, *repleta* and *cardini* groups of the subgenus *Drosophila* (LANKENAU, 1993; ALMEIDA; CARARETO, 2004; CORDEIRO et al., 2008). However, *micropia* nucleotide sequences are available only for the *Drosophila melanogaster*, *repleta* and *cardini* species groups (GenBank database). Evolutionary analysis of these sequences has demonstrated the existence of at least two *micropia* families, which are about 30% divergent in their nucleotide sequence (ALMEIDA; CARARETO, 2004). The first family is present in the *repleta* group and is composed of two subfamilies, differentiated by nucleotide divergence and antisense RNA expression. Additionally, it seems that the *repleta* species group has transmitted one of these subfamilies to the *Drosophila cardini* group through horizontal transfer (HT) (CORDEIRO et al., 2008). The second *micropia* family is represented by the *D. melanogaster* element that also transcribes antisense RNAs, responsible for protein expression control (LANKENAU et al., 1994). These data suggest that the *micropia* retrotransposon is an ancient component of *Drosophila* genomes (ALMEIDA; CARARETO, 2004).

Differently from *micropia*, the *gypsy* retrotransposon has been analyzed in the 12 *Drosophila* genomes (LUDWIG et al., 2008) and individual sequences of several Drosophilidae species are available, such as the *guarani*, *cardini*, *pallidipennis*, *tripunctata*, *virilis*, *repleta*, *anulimana* and *flavopilosa* groups of the *Drosophila* subgenus; the *willistoni*

and *melanogaster* groups of the *Sophophora* subgenus; and the *Scaptodrosophila* and *Zaprionus* genera (MIZROKHI; MAZO, 1991; ALBEROLA; DE FRUTOS, 1996; HEREDIA et al., 2004, LUDWIG et al., 2008). Two *gypsy* families have been identified, one exclusive of *Drosophila willistoni* and the other widely distributed in Drosophilidae species (HEREDIA et al., 2004). The widely distributed *gypsy* family is divided into at least ten subfamilies and seven of them have been involved in HT events, demonstrating that the *gypsy* retrotransposon inheritance is complex (HEREDIA et al., 2004; LUDWIG et al., 2008). *Gypsy* HT events could be facilitated by the presence of the *env* gene, assumed to confer to the retrotransposon an autonomous infectious capacity, allowing to cross species barriers (MEJLUMIAN et al., 2002; HEREDIA et al., 2004; LUDWIG et al., 2008). Specifically in the *melanogaster* species subgroup, two *gypsy* subfamilies were identified, and seven HTs might be proposed between the *melanogaster* species subgroups or between one species of the *melanogaster* subgroup and *Scaptodrosophila latisfasciaeformis* and *Zaprionus indianus* (HEREDIA et al., 2004; LUDWIG et al., 2008). The occurrence of these inter-genera HTs has been assumed because neutral sites of *gypsy* were found to be three to five times more conserved than those of host genes harbored by species of different genera which diverged longer than 40 MYA (HEREDIA et al., 2004).

In this study, we surveyed the *gypsy* and *micropia* retrotransposons in seven species of genus *Zaprionus*, comprising one species of subgenus *Anaprionus* and six of subgenus *Zaprionus*, and characterized their transcriptional activity and the phylogenetic relationships between sequences of different drosophilid species. The analyses aimed at understanding the distribution and evolutionary history of this scarcely studied genus by comparing the *gypsy* and *micropia* sequences with those of the 12 *Drosophila* genomes and other sequences available in the nucleotide database. The results show that the *Zaprionus gypsy* and *micropia* retrotransposons are present and transcriptionally active only in the subgenus *Zaprionus*, in where they were introduced by more than one horizontal transfer event from different donor species.

## 6.2 Materials and methods

### 6.2.1 Fly stocks and DNA extraction

All *Zaprionus* species strains used in the present study are listed in Table 1. They were derived from a single female selected randomly from a mass culture, kindly provided by Drs.

Jean David and Amir Yassin from LEGS, CNRS, France. The *D. melanogaster* Canton-S strain was used as positive control for the molecular analyses. Genomic DNA was extracted from 10 individuals of each strain, using the phenol-chloroform method as described by Jowett (1986).

### 6.2.2 PCR reactions, cloning and sequencing

Primers 2813 (5' TTA ACT CCT AGA GTT CAT CGC TGG 3') and 2814 (5' CAT GTA CCT GGT TAA CTA CTG ACC 3') were used to amplify the equivalent 386 bp fragment of the *D. hydei micropia* retrotransposon located in a highly conserved region of the RNAseH domain, between nucleotides 2813 and 3198 (LANKENAU et al., 1994). The *gypsy* retrotransposon fragment was obtained using primers GYP3S2F (5' AAA GGC GAY TTG GTT GAC ACT CC 3') and GYP3S2R (5' CAR GTG GCT RGG TTG RGT GTG 3') and corresponds to a 485 bp sequence located in the 6491-6511 region of the *envelope* gene (ORF 3) of the *D. melanogaster gypsy* (HEREDIA et al., 2004). Both PCR reactions were performed in a final volume of 25 µl, using 200 ng of genomic DNA, 0.4 µM of each primer, 160 µM of each dNTP, 2 mM $MgCl_2$ and 1 U Taq Platinum polymerase (Invitrogen) in 1X PCR buffer. The cycling parameters used to amplify the *gypsy* fragment were: 96 °C for 2 min for initial denaturation, 35 cycles of 96 °C for 15 sec, 65 °C for 30 sec, and 72 °C for 1 min, and an additional extension step at 72 °C for 5 min. The *micropia* PCR cycling parameters were: 94 °C for 3 min for initial denaturation, 40 cycles of 94 °C for 30 sec, 58 °C for 1 min, and 72 °C for 1min, followed by a final extension step at 72 °C for 10 min. The fragments obtained were purified directly from the PCR product, using the GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) and cloned with the TOPO TA Cloning Kit (Invitrogen). Three randomly chosen clones were automatically sequenced in an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems/Hitashi), using the primers pair T7 and M13R.

### 6.2.3 RNA extraction and RT-PCR reactions

Heads and gonads from 10 individuals of each sex were dissected in Testis Buffer (183 mM KCl, 47 mM NaCl and 10 mM Tris-HCl pH 6.8). Total RNA was isolated from the dissected tissues using the TRIZOL reagent (Invitrogen) method and genomic DNA was eliminated from the samples by RQ1 RNase-Free DNase (Promega) treatment, according to the manufacturer's instructions. The cDNA pool was generated from total RNAs using

random primers and the High Capacity cDNA Archive Kit (Applied Biosystems) at low stringency conditions (37 °C). The PCR reaction conditions and cycling parameters were performed the same used for genomic DNA amplification of both retrotransposons. Primers ZapGPDHF (5' GTT CGG CAA TTG AAC CAA TG 3') and ZapGPDHR (5' AGA GAG TCC GTG TGC ATG TG 3'), which amplify a 337 bp sequence of the *Zaprionus tuberculatus Gpdh* gene ORF1 (L37039), were used to check genomic DNA contamination in the total RNA extraction and cDNA quality. These PCR reactions were carried out using 200 ng of total cDNA, 0.1 mM of each dNTP, 0.4 μM of each primer, 2 mM $MgCl_2$ and 1 U Taq Platinum polymerase (Invitrogen) in 1X PCR buffer. The cycling parameters were: 94 °C for 2 min for initial denaturation, 35 cycles of 94 °C for 1 min, 60 °C for 1 min and 72 °C for 1 min, and an additional extension step at 72 °C for 10 min.

### 6.2.4 *In silico* and phylogenetic analyses

The *Zaprionus* subgenus sequences were manipulated into the BioEdit Sequence Alignment Editor (HALL, 1999) and aligned with Clustal W 1.81 (THOMPSON et al., 1994). The most divergent clones (>25%) were selected as queries for a search in the 12 *Drosophila* genomes using the flybase BLASTn tool (http://flybase.bio.indiana.edu/blast/) and the genome database released on October 17th, 2008. The queries selected were *Z. tuberculatus*1, *Z. tuberculatus*5 and *Zaprionus gabonicus*1 for *gypsy*, and *Z. tuberculatus*2 and *Zaprionus africanus*1 for *micropia*. In order to obtain only highly related and non-disrupted sequences, the searches were performed with stringency parameters (e-values > $e^{-50}$ and 90% coverage). The redundant genomic sequences (100% identical) were not included in the phylogenetic analyses, in order to minimize polytomies.

Multiple alignments of the *Zaprionus* and the genomic sequences were used to infer the phylogenetic relationships, using the maximum likelihood (ML), neighbor-joining (NJ) and maximum parsimony (MP) methods, as implemented in PhyML 3.0 (GUINDON; GASCUEL, 2003), MEGA 4.1 (TAMURA et al., 2007) and PAUP v.4.0b10 (SWOFFORD, 1997), respectively. Branch support was calculated by bootstrap analysis consisting of 1000 replicates (FELSENSTEIN, 1985). In the NJ and ML analyses, the maximum composite likelihood (TAMURA et al., 2004) and HKY85 (HASEGAWA et al., 1985) distances were used to estimate the divergence matrices and reconstructed trees, respectively. A heuristic search algorithm was used for the MP reconstruction. The sequences produced were registered in the GenBank database (Table 1). *Gypsy* sequences of *S. latisfaciaeformis*

(AF548144, AF548153 and AF548152) and *D. willistoni* (AF548159, AF548176 and AF548143), and *micropia* sequences of *Drosophila buzzatii* (AY522351), *D. hydei* (AY519123), *Drosophila paranaensis* (AY519124), *Drosophila mercatorum* (AY519125), *Drosophila seriema* (AY522346), *Drosophila gouveai* (AY522353), *Drosophila antonietae* (AY522345), *Drosophila serido* (AY522344), *Drosophila spenceri* (AY522347), *Drosophila koepferae* (AY522352) from GenBank were also used in the phylogenetic reconstruction.

The number of synonymous substitutions per synonymous site (dS), non-synonymous substitutions per non-synonymous site (dN), the Codon Based Z-test and dN/dS ratios were estimated for the *Zaprionus* and *melanogaster* sequences using the Nei and Gojobori (1986) distance with the Jukes and Cantor correction as implemented in MEGA 4.1 (TAMURA et al., 2007). The *gypsy* sequences *Drosophila sechellia*A6/B14/B15, *Drosophila yakuba*A6 and *Drosophila erecta*A2/A6/B9, as well the *micropia* sequences *Drosophila simulans*A3/A6, *D. sechellia*A9/A18/A20 and *D. yakubaA3* were excluded from those alignments because of the presence of large indels (>2 bp), disrupting the open reading frames. Punctual (1 bp) alignment gaps were deleted prior to estimation of dN and dS. The codon bias index (CBI, [MORTON, 1993]) was estimated for each sequence, using the DnaSP 4.50 program (ROZAS et al., 2003), where zero indicates no bias, and 1 shows maximum bias. Sequences of exon 4 of the *Gpdh* gene of *Zaprionus* and *melanogaster* subgroup species (FJ705445 to FJ705450, L37039, NM_057218, XM_002078253, XM_002089126, XM_001968825 and *D. sechellia* genomic sequence: scaffold_5 / 4016995-4017372) were used to compare the dS values of *gypsy* and *micropia* sequences.

The time of divergence of *gypsy* and *micropia* retrotransposons was estimated according to the equation $r = k / 2T$, where $r$ is the evolutionary rate (1.6 % per site per millions of years, according to synonymous sites rate for genes with low codon bias in *Drosophila*, [SHARP; LI, 1989]), $k$ is the dS pairwise divergence, and T is the divergence time between species.

### 6.2.5. Median-joining networks

The median-joining network reconstructions were performed using the DNA Aligment 1.3.0.1 and NETWORK 4.5.1.0 softwares (BANDELT, 1999), available at the Fluxus Technology Ltd. website. MP calculation was performed for the networks post-processing, the characters were equally weighted, and the other median-joining parameters followed the software default.

## 6.3 Results

### 6.3.1 Distribution and transcriptional activity of *gypsy* and *micropia* in genus *Zaprionus*

The PCR reactions allowed isolating the *gypsy* and *micropia* sequences in all *Zaprionus* subgenus species investigated, but not in subgenus *Anaprionus*, by screening *Zaprionus multistriatus* (data not show). The *gypsy* and *micropia* transcriptional activity was studied by a RT-PCR approach. Amplicons of the expected size were clearly detected in samples obtained from all *Zaprionus* subgenus species, but again not in *Z. multistriatus* (Figure 1), corroborating the analysis performed with genomic DNA. No sex or tissue specificity in the transcriptional pattern was observed by the RT-PCR analyses. However, amplification of *micropia* was absent in the *Zaprionus camerounensis* female heads and weaker in the *Z. indianus* ovaries.

### 6.3.2 Evolutionary analysis of the *gypsy* retrotransposon

The *in silico* search retrieved genomic sequences only for the *melanogaster* species subgroup, indicating that the genomes of *Drosophila ananassae*, *Drosophila pseudoobscura*, *Drosophila persimilis*, *D. willistoni*, *Drosophila mojavensis*, *Drosophila virilis* and *Drosophila grimshawi* do not harbor *gypsy* sequences highly related to those of subgenus *Zaprionus*. Eight sequences were identified in *D. melanogaster*, nine in *D. simulans*, 19 in *D. sechellia*, six in *D. yakuba*, and 13 in *D. erecta* (Supplementary Table 1). After redundancy exclusion of the genomic sequences, phylogenetic analyses were performed with three sequences of each *Zaprionus* species and 52 sequences of the *melanogaster* species subgroup, totalizing 70 sequences. The *S. latisfasciaeformis* and *D. willistoni* sequences were included in the reconstructions in view of their, respectively, closer and more distant evolutionary relationships to the *melanogaster* species subgroup (HEREDIA et al. 2004). Additionally, a tree was constructed using sequences of 17 *Drosophila* species described by Heredia et al. (2004), but no new information was obtained compared to the *Zaprionus* / *melanogaster* / *D. willistoni* / *S. latisfasciaeformis* species tree (data not shown). The phylogenetic relationships detected using the MP, ML, and NJ methods produced similar patterns, with identical main branches. Figure 2 shows the tree reconstructed by ML analysis. A major clade harbors all sequences of genus *Zaprionus*, the *melanogaster* species subgroup and *S. latisfasciaeformis* in 100% of bootstrap replications. It was divided into three secondary groups of sequences not

congruent with the species phylogeny. The first clade (bootstrap 97 %) contains sequences of *D. melanogaster*, *D. erecta*, *D. simulans* and only one sequence of genus *Zaprionus*, *Z. tuberculatus*1. The second clade (bootstrap 84 %) includes *Z. tuberculatus*, *Zaprionus camerounensis*, *S. latisfasciaeformis*, *D. melanogaster*, *D. erecta* and all sequences of *D. yakuba*, except for *D. yakubaA6* that branches outside the three groups. The clustering of *D. erecta* and S. *latisfasciaeformis* has already been reported and explained as an HT event by Ludwig et al. (2008). Finally, the third clade comprises the *gypsy* sequences of *Zaprionus davidi*, *Z. indianus* and of two cryptic species, *Z. gabonicus* and *Z. africanus*, as well as sequences of *D. melanogaster*, *D. simulans* and all sequences of *D. sechellia*, with no bootstrap support.

The maximum composite likelihood pairwise distances show that the sequences within each clade have relatively low divergence levels, with mean values of 0.063 within clade one, 0.070 within clade 2, and 0.042 within clade 3 (Table 2 and Supplementary Table 2). The distances of the *Zaprionus vs. melanogaster* species within each group are similar to the mean values. The mean distances among the clades show that clade 2 is more similar to clade 3 than to clade 1, with means of 0.119 (clade 2 *vs.* clade 3), 0.239 (clade 1 *vs.* clade 2) and 0.241 (clade 1 *vs.* clade 3). Since clade 1 of sequences is about 0.24 divergent from the clades 2 and 3, it was considered as belonging to subfamily (Subfamily 1); on the other hand, the divergence of 0.12 between clades 2 and 3 led us to classify them as two variants (V1 and V2) of the same subfamily (Subfamily 2).

The *Drosophila gypsy* sequences, with a variation of up to 20% in the nucleotide composition, were classified into subfamilies by Heredia et al. (2004) and Ludwig et al. (2008). These authors classified sequences found in *Z. indianus* and in the *melanogaster* species subgroup into two *gypsy* subfamilies: clade 13, clustering *D. melanogaster*, *D. simulans*, *D. sechellia* and *D. erecta* sequences, and clade 12, comprising *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. erecta*, *D. yakuba* and *Z. indianus* sequences. The addition of 18 sequences belonging to six *Zaprionus* species corroborates those results. The distribution of the *melanogaster* species subgroup sequences in Subfamily 1 indicates that it corresponds to clade 13, except for the absence of any *D. sechellia* sequence. Hence, although the phylogenetic tree topology did not cluster the variants of Subfamily 2 together, the absence of bootstrap clustering of Subfamily 1 and Subfamily 2 – V1, the smaller distances and the species distribution show that they are equivalent to clade 12. Additionally, the inclusion of new *Zaprionus* species sequences allowed identifying two different *gypsy* variants belonging to the clade 12 family.

Three hypotheses could explain the high levels of conservation among sequences within each group (0.04-0.07) and the *gypsy* phylogeny inconsistencies compared to the species tree: ancestral polymorphism promoting differential subfamily fixation in the species, high selective constraints over the sequences, or the occurrence of HT. In order to test which hypothesis better explains the estimated conservation, the dS values and dN/dS ratios within the groups of sequences were compared to those of the host *Gpdh* gene. The mean *gypsy* dN/dS ratios varied from 0.153 to 1.002, while the *Gpdh* mean values did not exceed 0.041 (Table 3). These ratios indicate that, in spite of the dN/dS < 1, purifying selection on the *gypsy env* sequences is relaxed (dN/dS mean: 0.527), as expected for transposable elements. Additionally, the Z-test indicates that most of the pairwise comparisons within the groups (62 %) evidenced neutrality for the *gypsy env* region (Supplementary Table 4). Nevertheless, the neutrality hypothesis was refuted for all pairwise comparisons between the groups, indicating that purifying selection plays a role in the conservation of the *env* coding sequence among the *gypsy* subfamilies.

The dS comparisons also allow rejecting the hypothesis of ancestral polymorphism within groups (Figure 3). By this approach, one clone representing each species of *Zaprionus* and the best Blast subject of the genomic sequences of the different groups were compared. In a vertical transmission scenario, the dS values of the TE should be equivalent to or higher than those of the host gene, evolving in the absence of a strong codon usage bias. The CBI index indicates that the *gypsy* and *Gpdh* sequences did not suffer high codon bias, although the mean value of 0.602 for *Gpdh* suggests that low variation in the dS magnitude between *Gpdh* and *gypsy* should be taken with caution for the HT inferences, since the variation in the *Gpdh* dS distance could reflect the codon usage. Hence, the comparisons between *gypsy* and *Gpdh* dS were used to infer HT only if the dS values of *gypsy* were at least two times lower than those of *Gpdh*. In all comparisons between *Zaprionus* and *melanogaster* species, the dS values of *gypsy* were lower than those of *Gpdh*, varying from three times (*Z. tuberculatus*5 vs. *D. melanogaster*5 and *D. erecta*3) to 20 times (*D. melanogaster*7 vs. *Z. davidi*2) lower. Moreover, the dS comparisons between *Zaprionus* species within each group were not as highly variable as in *Zaprionus vs. melanogaster* species. These results may indicate that the *gypsy* retrotransposon evolved mainly by vertical inheritance within the *Zaprionus* subgenus species, but HT events might have happened between the *Zaprionus* and the *melanogaster* species.

### 6.3.3 Phylogenetic relationships of the *micropia* retrotransposon

The BLASTn tool also allowed identifying *micropia* sequences homologous to the *Zaprionus* species only in genomes of the *melanogaster* species subgroup, except for *D. erecta* (Supplementary Table 1). *D. sechellia* had the highest number of hits (22 sequences), followed by *D. simulans* (six sequences), *D. yakuba* (four sequences) and *D. melanogaster* (three sequences). The tree reconstruction performed with the *micropia* clones of subgenus *Zaprionus*, the genomic sequences of the *melanogaster* species subgroup and the GenBank sequences of the *repleta* species, using the MP, NJ and ML methods, resulted in similar topologies. The *repleta* group sequences were clustered outside of a well-supported clade that harbors all sequences of subgenus *Zaprionus* and the *melanogaster* species subgroup, divided into two clades (Figure 4). The first clade, with low support (48%), grouped the *D. sechellia*A20 sequence and three internal clades, corresponding to the *Z. camerounensis* sequences (bootstrap 98%), *D. yakuba* (no support) and the sequences of *Z. davidi*, *Z. indianus*, *Z. africanus* and *Z. gabonicus* (bootstrap 75%). The last clade did not present a clear internal species-specific clusterization. The second clade grouped together the sequences of the *melanogaster* species group (*D. melanogaster*, *D. simulans* and *D. sechellia*) and *Z. tuberculatus* in 76% of the replications. Despite the clusterization of all *melanogaster* sequences in the second clade, the support is not high (52%). The mean divergence within clade 1 (0.009) is lower that than in clade 2 (0.051) and between the two clades (0.074). Analyzing the species separately, the mean divergence of the *Zaprionus* sequences was 0.0328, within the *melanogaster* species subgroup it was 0.0466, and between *Zaprionus* and the *melanogaster* sequences it was 0.0727 (Table 2 and Supplementary Table 5). The *Z. davidi2* / *Z. africanus5*, *D. melanogasterA1* / *D. simulansA1* / *D. sechelliaA1* and *D. simulansA6* / *D. sechelliaA9* pairs of sequences were identical.

The same rationale and approach used to interpret the results obtained for the *gypsy* retrotransposon were applied to infer the evolutionary history of the *micropia* retrotransposon of *Zaprionus* and *melanogaster* species. The mean dN/dS values of the *micropia* sequences were 14 to 58 times lower than those of the *Gpdh* gene (Table 3). For the *Zaprionus* and *melanogaster* sequences, for example, they were: *micropia*: 0.524; *Gpdh*: 0.009. Moreover, the Z-test of neutrality failed to reject the null hypothesis in about 80% of the pairwise comparisons (Supplementary Table 7). These results do not corroborate the hypothesis that the high similarity between the *micropia* sequences of *Zaprionus* and *melanogaster* due to high selective constraints conserving the nucleotide sequences.

The dS comparisons of *micropia* and *Gpdh* were made between the sequences of subgenus *Zaprionus* within clade 1, between the *melanogaster* sequences (*D. melanogaster*, *D. simulans* and *D. sechellia*) within clade 2, and in the *Zaprionus* vs. *melanogaster* sequences for clades 1 and 2 (Figure 5). All the dS distances for the *micropia* sequences were lower than those of the *Gpdh*; for instance, the dS between *Z. tuberculatus* and the *melanogaster* sequences in clade 1 was 61 (*vs. D. simulans*) to 76 (*vs. D. sechellia*) times lower than that those of the host gene, and those of the *D. yakuba*A1 sequence were 18 (*vs. Z. indianus*2) to 31 (*vs. Z. camerounensis*1) times lower. Additionally, there were pairwise comparisons in which the synonymous sites were invariable, such as those of the *D. melanogaster*, *D. simulans* and *D. sechellia* sequences, and those of *Z. camerounensis vs. Z. davidi*, *Z. africanus* and *Z. gabonicus*. The high conservation in the neutral positions of the RNAseH sequences favor the hypothesis that *micropia* has not been evolving separately since the split of the *Zaprionus* subgenus and the *melanogaster* species subgroup, suggesting that this retrotransposon has been horizontally transmitted in the Drosophilidae family, as already proposed for the *gypsy* retrotransposon.

### 6.3.4 Network reconstructions corroborate the HT inferences

Phylogenetic methods have been widely used to infer the evolutionary history of transposable elements. However, this approach represents only bifurcation processes (in which the sampled sequences necessarily split into descendent nodes), whereas the internal nodes are unsampled, consequently, not reconstructed. On the other hand, network phylogenetic approaches have been developed to reconstruct the history between closely conserved sequences, providing indication of the ancestral nodes and expansion events (CORDAUX et al., 2004). This kind of analysis has been punctually used to infer the dispersion pattern of Class I transposable elements, such as the human *Alu* elements (CORDAUX et al., 2004) and the *Solanacea Tnt-1*-like retrotransposons (MANETTI et al., 2007), or to understand the pattern of nested TEs in the *D. melanogaster* genome (BERGMAN et al. 2006). Originally, the network approach was developed for studying highly related data of intraspecific variation, however, the TE sequences of the *gypsy* and *micropia* retrotransposons can be considered sequence populations that share a common ancestor. Here, network phylogenetic analysis was utilized to identify sequences which could represent possible ancestral nodes and to apply an extra approach to test the hypothesis that horizontal transfer has shaped the evolution of *gypsy* and *micropia* in *Zaprionus* and

*melanogaster* species. The median-joining network phylogenetic analyses were made using all sequences studied in the conventional phylogenetic trees. The results are shown in detail in Figure 6.

The typical phylogenetic tree and the distance values for the *gypsy* retrotransposon demonstrated the presence of two sequence subfamilies, the Subfamily 1 and the Subfamily 2, composed by two variant types. Both subfamilies presented incongruences in the tree, due to the clustering of *Zaprionus* and *melanogaster* sequences. The network corroborates these results, since the same sequence clusters were produced (Figure 6A). Moreover, the Subfamily 1 subfamily was closer to the Subfamily V2 than to V1. In the network analysis, the presence of median vectors connecting the *Zaprionus* and *melanogaster* sequences suggests that the HT events happened in the ancestors of the sampled sequences, indicating that the transmissions were not strictly recent and that the sequences have accumulated mutations after the HT. The sole exception was within Subfamily 2 – V2, where the *Z. davidi*2 sequence was the ancestral node of the *Z. africanus*2 sequence (Figure 6B), suggesting one HT event from *Z. davidi* to *Z. africanus*, that was not evidenced by the dS comparisons of *gypsy* and *Gpdh*. Additionally, the network reconstruction allowed to include the *D. yakuba*A6 sequence as a Subfamily 2 – V2 member.

Network reconstruction for the *micropia* retrotransposon also followed the basic topology of the phylogenetic tree (Figure 6C). The sequences of *D. yakuba, D. melanogaster*, *Z. camerounensis, Z. davidi, Z. indianus* and *Z. africanus* branched with the median vectors, suggesting that high sequence conservation was not due to HT events occurred directly from the sampled sequences. Nevertheless, the *D. melanogaster*A3 sequence connects the median vectors with a large cluster that harbors *D. sechellia, D. simulans* and *D. melanogaster* sequences, which in turn links a median branch to the *Z. tuberculatus* sequences. This arrangement could indicate al least two more recent HT events, one from *D. melanogaster* to *D. simulans* and/or *D. sechellia*, or even the ancestor of these highly related sequences, and a second event from any of these species to *Z. tuberculatus*. Additionally, the branch of the *D. sechellia*A20 sequence with a median vector that links *Z. camerounensis, Z. davidi, Z. indianus* and *Z. africanus* sequences suggests an additional HT.

### 6.3.5 Estimation of the divergence time between TE sequences

To reinforce the suggestion of horizontal transfer occurrence, the divergence time of both retrotransposons was calculated using the substitution rate for non-synonymous sites of

*Drosophila* nuclear genes, estimated at 1.6% per MY (SHARP; LI, 1989). Among the HTs inferred for the *gypsy* retrotransposon, five events seemed to involve *Zaprionus* and *melanogaster* species, three events only species of the *melanogaster* group, and one only *Zaprionus* species. For the *micropia* retrotransposon, eight HTs were suggested: two between the *Zaprionus* and the *melanogaster* species, three among the *melanogaster* species, and three restricted to the *Zaprionus* species.

The estimation of the divergence times for *gypsy* sequences showed that the HTs seem to have occurred around 0.4 to 7.6 MYA (Table 4), i.e., in a period that is more recent than the origin of the *Drosophila* genus (at least 40 MYA, [RUSSO et al. 1995]). For Subfamily 1, two HTs are proposed, one involving *D. melanogaster* and *D. erecta* sequences (0.406-2.125 MYA), and the other between the *D. melanogaster* / *D. simulans* species ancestor and *Z. tuberculatus* (about 2.5 MYA). Within Subfamily 2 - V1, four HTs seem to have occurred, all of them involving *D. erecta*: the first with *D. yakuba* (2.156-2.625 MYA), the second with *D. melanogaster* (1.718-2.656 MYA), the third with *Z. tuberculatus* (3.5 to 7.6 MYA), and the last one with the ancestor of *Z. camerounensis* / *Z. davidi* (0.8 to 2.4 MYA). Finally, for Subfamily 2 - V2, the divergence time suggested three HTs: two involving the *D. melanogaster*/*D. simulans*/*D. sechellia* species ancestor *vs.* the *indianus* complex ancestor (around 0.4-3.2 MYA) and *Z. davidi* (0.4-1.3 MYA), and the third between *Z. davidi* and *Z. africanus* (1.3 MYA).

Regarding the *micropia* retrotransposon, the HTs seem to have occurred very recently, since the times of divergence were very low or impossible to be calculated, due to the invariability in its synonymous positions (Table 4). Seven HTs were proposed, five of which seemed to involve an unknown donor and *D. melanogaster*, *D. yakuba* (1.156-1.781 MYA), the *indianus* complex (up to 1.181 MYA), *Z. davidi* and *Z. camerounensis*. Other two HTs seemingly to have happened between *D. melanogaster* or the *D. simulans*/*D. sechellia* ancestor and *Z. tuberculatus* (0.5-1.0 MYA), and the ancestor of *D. sechellia* and the *indianus* or *davidi* complex ancestors (1.7-2.4 MYA).

### 6.4 Discussion

This survey of the *gypsy* and *micropia* retrotransposons in seven species of genus *Zaprionus* indicates that they are widely distributed and transcriptionally active only in species of subgenus *Zaprionus*. Additionally, the evolutionary analyses demonstrated that these two *Zaprionus* retrotransposons are closely related to those of the *melanogaster* species

subgroup, and that their histories might have been repeatedly marked by events of horizontal transfer. Several pieces of evidence showed that the phenomenon of horizontal transfer is frequent in eukaryotes (LORETO et al., 2008; PACE et al., 2008; ROULIN et al., 2008; CHENG et al., 2009), and in *Drosophila* the number of HTs suspected for class II elements and LTR retrotransposons is fairly similar (LORETO et al., 2008). The studies of *micropia* and *gypsy* retrotransposons in subgenus *Zaprionus* species highlight horizontal transmission as an alternative evolutionary mechanism of the LTR retrotransposon evolution.

### 6.4.1 *Gypsy* evolution in the *Zaprionus* subgenus and *melanogaster* subgroup species

Evolutionary studies focusing the *gypsy* retrotransposon in the Drosophilidae family have shown that this element evolves through a considerable amount of HT events (HEREDIA et al., 2004; LUDWIG et al., 2008). Prior to the present study, sequences of only one species of subgenus *Zaprionus* were found in the database, and their nucleotide composition suggested that *Z. indianus* received its *gypsy* retrotransposon from *D. simulans* (HEREDIA et al., 2004). The data presented here broaden the comprehension of the *gypsy* evolutionary history in Drosophilidae. We were able to show that the *gypsy* sequences of subgenus *Zaprionus* can be grouped into two subfamilies (Subfamily 1 and Subfamily 2) and V2) which diverge by about 20% of their nucleotide sequences. One of these subfamilies is composed of two variants (Subfamily 2 – V1 and V2), with different distribution patterns. The differentiation of these variants was inferred based on the clusterization pattern in the phylogenetic tree, the network analyses and the mean nucleotide divergence of 0.12. The sampling of three *gypsy* sequences for each *Zaprionus* species could underestimate the subfamily variability in the six species studied, leading misunderstanding of the evolutionary patterns. However, a Southern blot analysis has indicated that *gypsy* and *micropia* retrotransposons present low insertion numbers in the subgenus *Zaprionus*, varying from one *gypsy* insertion (*Z. gabonicus*) to seven (*Z. indianus* and *Z. davidi*), and two *micropia* insertions (*Z. camerounensis* and *Z. africanus*) to seven (*Z. indianus*) (Supplementary Figure 1). These copy numbers and the close clustering of sequences of each species in the *gypsy* and *micropia* trees, except for *gypsy* of *Z. tuberculatus*, led us to consider adequate the sequencing of three clones for these phylogenetic studies.

The dN/dS ratios low values between *gypsy* subfamilies and variants could demonstrate that non-synonymous sites of *gypsy env* gene were negatively selected during the retrotransposon evolution, maintaining the structure of this protein. It is difficult to envisage

purifying selection playing a role in TE evolution, but it has been proposed that natural selection could act at the genomic level, conserving the different TE genes, as proposed for the *copia* retrotransposon of *melanogaster* species group. TE protein conservation might have been achieved for active elements which transpose more efficiently and at a higher rate, permitting their survival and propagation (JORDAN; MCDONALD, 1998).

In contrast to the inter-subfamily and variant comparisons, the comparisons made within the sequences in each clade show low selection and sequence variability, as well as a complex distribution, suggesting that the *gypsy* retrotransposon was introduced in the *Zaprionus* species genomes through horizontal transfer, probably from species of the *melanogaster* subgroup, or from their ancestors, to the *Zaprionus* species ancestors. Additionally, our data corroborate previous reports indicating that *gypsy* sequences of Subfamily 1 and 2 crossed the species limits within the *melanogaster* species subgroup (LUDWIG et al., 2008). It seems, however, that, after its introduction into the *Zaprionus* species, *gypsy* has been transmitted only vertically. Although we are unable to explain the reason, it seems that the *Zaprionus* species are no longer as prone to horizontal transfer of *gypsy* as they once were; it is however possible that HT continues to occur among *Zaprionus* species not studied yet. Sampling higher number of *Zaprionus* and other drosophilid species is fundamental to devising a more reliable scenario of *gypsy* evolution in the *Zaprionus* subgenus species.

The Subfamily 1 of *gypsy* sequences seems to be an ancestral component of the *D. melanogaster* and *D. simulans* genomes, once the comparative analyses of the neutral sites show that *gypsy* is more variable than the *Gpdh* gene. Although our *in silico* approach did not retrive Subfamily 1 sequences in *D. sechellia*, Ludwig et al. (2008) reported their presence, corroborating their ancestor status in *D. melanogaster* / *D. simulans* / *D. sechellia* species. On the other hand, the presence of this subfamily in *Z. tuberculatus* and *D. erecta* genomes, as well the *gypsy* and *Gpdh* dS comparisons, suggest their introduction by two independent HT events. A transfer from a *D. melanogaster* / *D. simulans* ancestor to *Z. tuberculatus* is proposed, based on the divergence time of the sequences (around 2.5 MYA), close but prior to the time of *D. melanogaster* / *D. simulans* splitting (2-3 MYA), and the basal positioning of *Z. tuberculatus* sequence in the network. Later on (0.4-2.1 MYA), a transfer between *D. melanogaster* and *D. erecta* may have occurred. Although an alternative hypothesis of HT between *D. simulans* and *D. erecta* could be formulated, based on the phylogenetic tree, the greater difference in the dS comparisons between the pair *D. erecta* / *D. melanogaster* (*gypsy* dS is 20.2 times higher than *Gpdh* dS) than in *D. erecta* / *D. simulans* (2.6 times) reinforces

the first hypothesis. Ludwig et al. (2008) have already proposed such a transfer, but no direction was inferred. We propose that *D. melanogaster* was the donor species, based on the evidence of *gypsy* vertical transmission between *D. melanogaster* / *D. simulans* / *D. erecta* species.

The V1 *gypsy* sequences seem to have an independent evolutionary history from that of Subfamily 1. At least four HT events can be proposed. Based on the basal branching of the *D. erecta gypsy* sequences in the network, the repeated clustering to all the other species of this group in the phylogeny, the TE divergence times and the estimate that the *D. erecta* diverged from *D. melanogaster* and *D. yakuba* about 13-15 MYA (LACHAISE; SILVAIN, 2004), we suggest that *D. erecta* was the donor of V1 sequences in a more ancient transfer to *Z. tuberculatus* (occurred about 3.5-7.6 MYA), followed by the introduction into *D. yakuba* (2.1-2.6 MYA), *D. melanogaster* (1.7-2.6 MYA), and ultimately in the *Z. camerounensis* / *Z. davidi* ancestor (0.8-2.4 MYA). The last HT seems to have occurred in the *Z. camerounensis* / *Z. davidi* ancestor, because these species diverged around 2.2 MYA (YASSIN et al., 2008). Ludwig et al. (2008) proposed three horizontal transfers with no defined direction into this group: *D. melanogaster vs. D. erecta, D. yakuba vs. D. erecta,* and *D. melanogaster vs. D. yakuba*. As for sequences Subfamily 1, our results allow proposing the direction of the two first transfers: from *D. erecta* to *D. melanogaster* and *D. yakuba*. Although the dS comparisons showed that *gypsy* is less variable than *Gpdh* in *D. melanogaster* / *D. yakuba* comparison, the phylogenetic tree does not corroborate the last HT proposition of Ludwig et al. (2008).

The dS comparisons of the V2 sequences of *D. melanogaster*, *D. simulans* and *D. sechellia*, as well as of the *indianus* species complex (*Z. indianus*, *Z. africanus* and *Z. gabonicus*), indicate that these *gypsy* sequences were transferred by vertical transmission within these groups. However, the dS comparisons, the phylogenetic inconsistencies and the times of divergence among them and from *Z. davidi* indicate that two HTs could have occurred in the ancestor of the modern sequences, the first between a *D. melanogaster / D. simulans* / *D. sechellia* ancestor and an *indianus* complex ancestor (around 0.4-3.2 MYA), and the second between a *D. melanogaster / D. simulans* / *D. sechellia* ancestor and *Z. davidi* (0.4-1.3 MYA), or its ancestor, since *Z. davidi* is the only species of the *lachaisei* complex analyzed in this study, and this species complex diverged about 0.8 MYA (YASSIN et al., 2008). A HT between a *D. melanogaster / D. simulans* / *D. sechellia* ancestor and the *indianus* complex ancestor is a wider scenario to the HT inference between *Z. indianus* and *D.*

*simulans* made by HEREDIA et al. (2004), exemplifying the importance of analyzing as many species as possible before inferring horizontal transfers, as pointed out by Loreto et al. (2008).

The last horizontal transfer proposed for Subfamily 2 – V2 exemplifies the typical HT evidence that can be demonstrated by network reconstructions but is impossible to be visualized in a classical phylogenetic tree. In the network tree, the *Z. davidi*2 sequence appears as an ancestral node of the *Z. africanus*2 sequence branch (Figure 6B), suggesting an additional HT between these species. The *gypsy* and *Gpdh* dS comparison corroborate such a transfer, which seems to have happened around 1.3 MYA, subsequent to the divergence of these species (2.6 to 4.1 MYA). The HT between *Z. davidi* and *Z. africanus* demonstrates the relevance of applying the network method as a complementary analysis to reconstruct the history of transposable elements.

### 6.4.2 *Micropia* evolution in the *Zaprionus* subgenus and *melanogaster* species subgroup

The analyses of the m*icropia* retrotransposon demonstrated that the subgenus *Zaprionus* harbors potentially active *micropia* sequences and interestingly, as demonstrated for the *gypsy* retrotransposon, they are closely related to those of the *melanogaster* subgroup, with one exception: the absence of related sequences in the *D. erecta* genome. The phylogenetic and distance estimations of the *Zaprionus*, *melanogaster* and *repleta* sequences show that the *micropia* retrotransposons of the *Zaprionus* and *melanogaster* species belong to the same subfamily, since the divergence between their nucleotide sequences is less than 20%. Taking together the low significance of the purifying selection evidenced by the dN/dS ratios, the low divergence, the comparisons with the *Gpdh* gene, the phylogeny and the network reconstruction, we suggest that HT has also played an important role in the evolution of the *micropia* retrotransposon.

Comparing the divergence times between the *micropia* retrotransposon of the *Zaprionus* and *melanogaster* sequences with those of *gypsy*, it can be inferred that the HT *micropia* wave might have started more recently than that of *gypsy* (*gypsy* HTs: 0.4-7.6 MYA, *micropia* HTs: 0.5-2.4 MYA). The isolated central cluster of median vectors in the network diagram and the assumption that *micropia* divergence took place after the splitting of genus *Zaprionus* and the *melanogaster* group (40 MYA), *Z. camerounensis* and the *indianus* complex (2.2 MYA), and *Z. camerounensis* and *Z. davidi* species (0.8 MYA) suggest that an unknown species transferred *micropia* sequences independently to *D. melanogaster*, *D. yakuba*, *Z. camerounensis*, *Z. davidi* and the *indianus* complex. The invariability of the

synonymous positions makes it difficult to date these events, but the higher dS levels suggest that of the species analyzed in this study, *D. yakuba* and the *indianus* complex species were the first ones to receive *micropia*. The probable donor might have been a non-studied species, and neither the *repleta* nor the *cardini* species groups of *Drosophila* are good candidates, since they harbor *micropia* retrotransposons of a different family (ALMEIDA; CARARETO, 2004; CORDEIRO et al., 2008).

The other three *micropia* HTs were emphasized by the network analysis. *D. melanogaster* could be the donor of *micropia* to the *D. simulans* / *D. sechellia* ancestor (HT estimated at 0.5 MYA), or even to the two species separately, due to their positioning as an ancestral node in the network and the probable species splitting time (0.4 MYA [LACHAISE; SILVAIN, 2004]). In turn, one of these three species was involved in the *Z. tuberculatus micropia* transmission. The divergence time of the *Z. tuberculatus micropia* sequences (0.5-1.0 MYA) could indicate that the HT happened before the *D. simulans*/*D. sechellia* transfer, which was not supported by the network analysis. The divergence time (around 1.7 and 2.4 MYA) of the sequences involved in the last HT proposed, from the *indianus* or the *davidi* complex to a *D. sechellia* ancestor, could be an overestimation, due to a sequence-specific high evolutionary rate, evidenced by the absence of clusterization and the long branch in the phylogenetic tree (BERGSTEN, 2005). All these results indicate that horizontal transfer was significant to the history of *micropia* in these species, but also stress the importance of analyzing a great number of species, in order to make more realistic inferences of the evolutionary steps of transposable elements.

### 6.4.3 Similar evolution patterns of *gypsy* and *micropia* retrotransposons

A high frequency of horizontal transmission of transposable elements in the *melanogaster* species group has been inferred previously (BOWEN; MCDONALD, 2001; HEREDIA et al., 2004; SANCHEZ-GRACIA et al., 2005; LUDWIG et al., 2008). Our results suggest that the *Zaprionus* species have also experienced waves of retrotransposon introductions, particularly in the last seven million of years. Apparently, species of the *melanogaster* subgroup might have donated their *gypsy* sequences to species of the *Zaprionus* subgenus, and these two species groups could have received the *micropia* retrotransposon from one or more unknown donor species. Additionally, after the initial introduction, these species were able to share their elements, a phenomenon evidenced by the posterior horizontal transmissions. Since genus *Zaprionus* and the *melanogaster* subgroup seem to share the same

age of origin and diversification in the tropical Africa (LACHAISE; SILVAIN, 2004; YASSIN et al., 2008), as well as similar oviposition site environments and anthophilic relationships for some species (YASSIN et al., 2008), our data suggest that they passed through a permissive period of transposable element introduction during the diversification period, characterized here for the evolution of the *micropia* and *gypsy* retrotransposons. In a more recent period, *copia* also appears to have been involved in horizontal transmissions from the *Zaprionus* genus ancestor to the *melanogaster* species group (DE SETTA et al., data not published). Although the selection tests did not show high selective constraints conserving the *Zaprionus* and *melanogaster* sequences of both retrotransposons, the alternative hypothesis of some selective constraints conserving these retrotransposon coding sequences, including the synonymous sites cannot be ruled out. High purifying selection has been shown to maintain conserved non-coding elements in the human genome due to TE fragments exaptation (SILVA et al., 2003; LOWE et al., 2007) According to our results, the exaptation hypothesis might not be reasonable for the *gypsy* and *micropia* conservation, since these elements are autonomous and full-length retrotransposons in *D. melanogaster* (MUGNIER et al., 2008).

We have shown here that the *gypsy* and *micropia* retrotransposons had similar evolutionary patterns in the *Zaprionus* subgenus and the *melanogaster* subgroup species, once they seem to have participated of several inter- and intra-subgenus horizontal transfers. A question that can be formulated is whether these elements have been transferred by similar ways. Horizontal transfer of *gypsy* is easier to explain, once the active *gypsy env* gene can enable its dispersion and expansion of the genomic territory, but this mechanism does not fit to *micropia* that does not have an *envelope* gene. Nevertheless, it has been proposed that retrotransposons could have been transferred concomitantly with retroviruses infections, even the *gypsy* retrotransposon, or parasitic infestations, as for example by the *Wolbachia* intracellular bacteria or mites (LORETO et al., 2008). Although our data allowed to propose horizontal transmissions of *gypsy* and *micropia* among sister species such as *D. melanogaster*, *D. simulans* and *D. sechellia*, the occurrence of introgression cannot be ruled out as an alternative explanation for HT.

Another concern is the evolutionary process of fixation of the transposable elements after the horizontal introduction. These species could share similar mechanisms of transposable element regulation, which would permit escaping from the selective forces and allow fixation after the introduction. It has been shown that the control of *Drosophila* retrotransposon mobilization depends on the transcription rate, which is directly related to the presence of specific regulatory proteins (CAVAREC; HEIDMANN, 1993; BHADRA et al.,

1999; PAL-BHADRA et al., 1998; CAVAREC et al., 1994; TULIN et al., 2002) and the rate of RNA degradation, mediated by the RNAi system (BRENNECKE et al., 2007; SIOMI et al., 2008). However, genetic information concerning the species of genus *Zaprionus* is scarce. Further, studies of the *Zaprionus* genes involved in the mobilization control could lead to proposing hypotheses about the mechanisms that permit the sharing of highly related sequences in the *Zaprionus* and *melanogaster* species genomes.

**Acknowledgments**

We gratefully acknowledge funding from the CAPES-COFECUB International Cooperation Program (to NS, CMAC and PC), FAPESP (MAVS), CNPq (CMAC and MAVS). NS was receipt of a CNPq fellowship.

**References**

ALBEROLA, T. M.; DE FRUTOS, R. Molecular structure of a *gypsy* element of *Drosophila subobscura* (*gypsyDs*) constituting a degenerate form of insect retroviruses. **Nucleic Acids Research**, Inglaterra, v. 24, p. 914-923, 1996.

ALMEIDA, L. M.; CASTRO, J. P.; CARARETO, C. M. A. *Micropia* transposable element presence in *Drosophila saltans* group. **Drosophila Information Service**, v. 84, p. 114-117, 2001.

ALMEIDA, L. M.; CARARETO, C. M. A. Identification of two subfamilies of *micropia* transposable element in species of the *repleta* group of *Drosophila*. **Genetica**, Holanda, v. 121, p. 155-164, 2004.

BANDELT, H. J. Median-joining networks for inferring intraespecific phylogenies. **Molecular Biology and Evolution**, EUA, v. 16, p. 37-48, 1999.

BERGMAN, C. M. et al. Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. **Genome Biology**, Inglaterra, v. 7, p. R112, 2006.

BERGSTEN, J. A review of long-branch attraction. **Cladistics**, Inglaterra, v. 21, p. 163-193, 2005.

BHADRA, U.; PAL-BHADRA, M.; BIRCHLER, J. A. Role of the *male specific lethal* (*msl*) genes in modifying the effects of sex chromosomal dosage in *Drosophila*. **Genetics**, EUA, v. 152, p. 249-268, 1999.

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

BRENNECKE, J. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. **Cell**, EUA, v. 128, p. 1089-1103, 2007.

CAVAREC, L.; HEIDMANN, T. The *Drosophila copia* retrotransposon contains binding sites for transcriptional regulation by homeoproteins. **Nucleic Acids Research**, Inglaterra, v. 21, p. 5041-5049, 1993.

CAVAREC, L.; JENSEN, S.; HEIDMANN, T. Identification of a strong transcriptional activator for the *copia* retrotransposon responsible for its differential expression in *Drosophila hydei* and *melanogaster* cell lines. **Biochemical and Biophysical Research Communications**, EUA, v. 203, p. 392-399, 1994.

CHENG, X. et al. A new family of *Ty1-copia*-like retrotransposon originated in the tomato genomes by a recent horizontal transfer event. **Genetics**, EUA, doi. 10.1534/genetics.108.099150, 2009.

CORDAUX, R.; HEDGES, D. J.; BATZER, M. A. Retrotransposition of *Alu* elements: how many sources? **Trends in Genetics**, Inglaterra, v. 20, p. 464-467, 2004.

CORDEIRO, J. et al. The LTR retrotransposon *micropia* in the *cardini* group of *Drosophila* (Diptera: Drosophilidae): a possible case of horizontal transfer. **Genetica**, Holanda, v. 134, p. 335-44, 2008.

EICKBUSH, T. H. Origin and evolutionary relationships of LTR retroelements. In: MORSE, S. S. (Org.) **The evolutionary biology of viruses**. Nova Iorque: Raven, pp. 121-157, 1994.

FELSENSTEIN, J. Confidence limits on phylogenies: an approach using the bootstrap. **Evolution**, EUA, v. 39, p. 783-791, 1985.

GUINDON, S.; GASCUEL, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. **Systematic Biology**, Inglaterra, v. 52, p. 696-704, 2003.

HALL, T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symposium Series**, Inglaterra, v. 41, p. 95-98, 1999.

HASEGAWA, M.; KISHINO, H.; YANO T. Dating of the human-ape splitting by molecular clock of mitochondrial DNA. **Journal of Molecular Evolution**, Alemanha, v. 21, p. 160-174, 1985.

HEREDIA, F. et al. Complex evolution of *gypsy* in drosophilid species. **Molecular Biology and Evolution**, EUA, v. 21, p. 1831-1842, 2004.

HUIJSER, P. et al. Retrotransposon-like sequences are expressed in Y chromosomal lampbrush loops of *Drosophila hydei*. **Journal of Molecular Biology**, Inglaterra, v. 203, p. 689-697, 1988.

JORDAN, I. K.; MCDONALD, J. F. Interelement selection in the regulatory region of the *copia* retrotransposon. **Journal of Molecular Evolution**, Alemanha, v. 47, p. 670-676, 1998.

JOWETT, T. Preparation of nucleic acids. In: ROBERTS, D. B. **Drosophila: A Practical Approach.** Oxford: IRL Press, pp. 275-277, 1986.

LACHAISE, D.; SILVAIN, J. F. How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster-D. simulans* palaeogeographic riddle. **Genetica**, Holanda, v. 120, p. 17-39, 2004.

LANKENAU, D. H. The retrotransposon family *micropia* in *Drosophila* species. In: MCDONALD, J. (Org.) **Transposable Elements and Evolution.** Dordrecht: Kluwer Academic Publisher, pp. 232-241, 1993.

LANKENAU, S.; CORCES, G. V.; LANKENAU, D. H. The *Drosophila micropia* retrotransposon encodes a testis-specific antisense RNA complementary to reverse transcriptase. **Molecular and Cellular Biology**, EUA, v. 14, p. 1764-1775, 1994.

LLORÉNS, C. et al. The *Gypsy* Database (GyDB) of mobile genetic elements. **Nucleic Acids Research**, Inglaterra, v. 36, p. D38-D46, 2008a.

LLORÉNS, J. V. et al. *Gypsy* endogenous retrovirus maintains potential infectivity in several species of Drosophilids. **BMC Evolutionary Biology**, Inglaterra, v. 8, p. 302, 2008b.

LORETO, E. L.; CARARETO, C. M.; CAPY, P. Revisiting horizontal transfer of transposable elements in *Drosophila*. **Heredity**, Inglaterra, v. 100, p. 545-54, 2008.

LOWE, C. B.; BEJERANO, G.; HAUSSLER, D. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 104, p. 8005-8010, 2007.

LUDWIG, A.; VALENTE, V. L.; LORETO, E. L. Multiple invasions of *Errantivirus* in the genus *Drosophila*. **Insect Molecular Biology**, Inglaterra, v. 17, p. 113-24, 2008.

MANETTI, M. E. et al. Radiation of the *Tnt1* retrotransposon superfamily in three Solanaceae genera. **BMC Evolutionary Biology**, Inglaterra, v. 7, p. 34, 2007.

MEJLUMIAN, L. et al. Comparative and functional studies of *Drosophila* species invasion by the *gypsy* endogenous retrovirus. **Genetics**, EUA, v. 160, p. 201-209, 2002.

MIZROKHI, L. J.; MAZO, A. M. Cloning and analysis of the mobile element *gypsy* from *D. virilis*. **Nucleic Acids Research**, Inglaterra, v. 19, p. 913-916, 1991.

MORTON, B. R. Chloroplast DNA codon use: evidence for selection at the *psb* locus based on tRNA availability. **Journal of Molecular Evolution**, Alemanha, v. 37, p. 273-280, 1993.

MUGNIER, N. et al. The heterochromatic copies of the LTR retrotransposons as a record of the genomic events that have shaped the *Drosophila melanogaster* genome. **Gene**, Holanda, v. 411, p. 87-93, 2008.

NEI, M.; GOJOBORI, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. **Molecular Biology and Evolution**, EUA, v. 3, p. 418-426, 1986.

PACE, J. K. 2nd et al. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 105, p. 17023-17028, 2008.

PAL-BHADRA, M.; BHADRA, U.; BIRCHLER, J. A. Role of multiple trans-acting regulators in modifying the effect of the retrotransposon *copia* on host gene expression in *Drosophila*. **Molecular & General Genetics**, Alemanha, v. 259, p. 198-206, 1998.

ROULIN, A. et al. Evidence of multiple horizontal transfers of the long terminal repeat retrotransposon *RIRE1* within the genus Oryza. **The Plant Journal**, Inglaterra, v. 53, p. 950-959, 2008.

ROZAS, J. et al. DnaSP, DNA polymorphism analyses by the coalescent and other methods. **Bioinformatics**, Inglaterra, v. 19, p. 2496-7, 2003.

RUSSO, C. A. M. TAKEZAKI, N.; NEI, M. Molecular phylogeny and divergence times of drosophilid species. **Molecular Biology and Evolution**, EUA, v. 12, p. 391-404, 1995.

SÁNCHEZ-GRACIA, A.; MASIDE, X.; CHARLESWORTH, B. High rate of horizontal transfer of transposable elements in *Drosophila*. **Trends in Genetics**, Inglaterra, v. 21, p. 200-203, 2005.

SHARP, P. M.; LI, W. H. On the rate of DNA sequence evolution in *Drosophila*. **Journal of Molecular Evolution**, Alemanha, v. 28, p. 398-402, 1989.

SILVA, J. C. et al. Conserved fragments of transposable elements in intergenic regions: evidence for widespread recruitment of MIR- and L2-derived sequences within the mouse and human genomes. **Genetical Research**, Inglaterra, v. 82, p. 1-18, 2003.

SIOMI, M. C.; SAITO, K.; SIOMI, H. How selfish retrotransposons are silenced in *Drosophila* germline and somatic cells. **FEBS Letters**, Holanda, v. 582, p. 2473-2478, 2008.

SWOFFORD, D. **PAUP: phylogenetic analysis using parsimony:** Version 4.0b10. Washington, DC: Smithsonian Institution, 1997.

TAMURA, K.; NEI, M.; KUMAR, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. **Proceedings of the National Academy of Sciences (USA)**, EUA, v. 101, p. 11030-11035, 2004.

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

TAMURA, K. et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. (1993) v. 24, p. 1596-1599, 2007.

THOMPSON, J. D. et al. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, Inglaterra, v. 22, p. 4673-4680, 1994.

TULIN, A.; STEWART, D.; SPRADLING, A. C. The *Drosophila* heterochromatic gene encoding poly(ADP-ribose) polymerase (*PARP*) is required to modulate chromatin structure during development. **Genes & Development**, EUA, v. 16, p. 2108-2119, 2002.

YASSIN, A. et al. Grafting the molecular phylogenetic tree with morphological branches to reconstruct the evolutionary history of the genus *Zaprionus* (Diptera: Drosophilidae). **Molecular Phylogenetics and Evolution**, EUA, v. 47, p. 903-15, 2008.

**Table 1.** Species studied, taxonomic classification, divergence time according to Yassin et al. (2008), geographic origin of the *Zaprionus* strains and sequence GenBank accession numbers.

| Species | Time (MYA) | Geographic origin | GenBank accession numbers | |
|---|---|---|---|---|
| | | | *gypsy* | *micropia* |
| Genus *Zaprionus* | 13.81 (10.87-14.90) | - | - | - |
| Subgenus *Anaprionus* | 10.59 (7.68-12.05) | - | - | - |
| *Z. multistriatus* | - | Bangalore (India) | - | - |
| Subgenus *Zaprionus* | 7.37 (6.71-9.01) | - | - | - |
| Group *inermis* | 6.98 | - | - | - |
| *tuberculatus* complex | 2.99 (2.93-18.00) | - | - | - |
| *Z. tuberculatus* | - | Ithala (South Africa) | FJ710406 to 408 | FJ710423 to 425 |
| Group *armatus* | - | - | - | - |
| *lachaisei* complex | 0.85 (0.39-1.13) | - | - | - |
| *Z. camerounensis* | - | Amani (Tanzania) | FJ710409 to 411 | FJ710426 to 428 |
| *davidi* complex | 2.21 (1.39-2.44) | - | - | - |
| *Z. davidi* | - | São Tomé (São Tomé e Príncipe) | FJ710412 to 414 | FJ710429 to 431 |
| *indianus* complex | 3.15 (2.61-4.14) | - | - | - |
| *Z. gabonicus* | - | Makokou (Gabon) | FJ710415 to 417 | FJ710432 to 434 |
| *Z. africanus* | - | Kibale (Uganda) | FJ710418 to 420 | FJ710435 to 437 |
| *Z. indianus* | - | Brasília (Brasil) | FJ710421 to 422 | FJ710438 to 440 |

**Table 2.** Genetic divergence (maximum composite likelihood distance) between the *gypsy* and *micropia* transposable elements for the sequences used in the phylogenetic analysis. n: number of pairwise comparisons.

| Retrotransposon | n | Mean | Minimum | Maximum |
|---|---|---|---|---|
| *gypsy* | | | | |
| Subfamily 1 (S1) | 28 | 0.0628 | 0.0213 | 0.1112 |
| *Zaprionus* x *melanogaster* S1 | 7 | 0.0742 | 0.0529 | 0.1006 |
| Subfamily 2 - V1 | 300 | 0.0704 | 0.0021 | 0.1687 |
| *Zaprionus* x *melanogaster* V1 | 96 | 0.0641 | 0.0211 | 0.1122 |
| Subfamily 2 - V2 | 741 | 0.0422 | 0.0042 | 0.0888 |
| *Zaprionus* x *melanogaster* V2 | 308 | 0.0506 | 0.0086 | 0.0888 |
| S1 x V1 | 200 | 0.2388 | 0.2080 | 0.2910 |
| S1 x V2 | 312 | 0.2406 | 0.2055 | 0.286 |
| V1 x V2 | 975 | 0.1189 | 0.0771 | 0.1814 |
| *micropia* | | | | |
| *Zaprionus* genus x *melanogaster* group | 468 | 0.0727 | 0.0188 | 0.1341 |
| *Zaprionus* genus | 153 | 0.0328 | 0 | 0.0850 |
| *melanogaster* group | 325 | 0.0466 | 0 | 0.1473 |

**Table 3.** Comparative analyses of synonymous substitutions per synonymous site (dS), non-synonymous substitutions per nonsynonymous site (dN), dN/dS ratios and codon bias index of the *gypsy* and *micropia* sequences and *Gpdh* gene. (See details of the pairwise dS and dN distances in Supplementary Tables 3 and 6).

| Sequences | CBI mean (min-max) | dN mean (SE) | dS mean (SE) | dN/dS mean (SE) |
|---|---|---|---|---|
| *gypsy* | | | | |
| Subfamily 1 | | | | |
| *Zaprionus* x *melanogaster* spp. | | 0.045 ($9.70x10^{-8}$) | 0.128 ($5.95x10^{-4}$) | 0.374 (0.004) |
| *melanogaster* spp. | | 0.035 ($3.88x10^{-5}$) | 0.057 ($6.49x10^{-4}$) | 1.002 (0.465) |
| Subfamily 2 – V1 | | | | |
| *Zaprionus* x *melanogaster* spp. | 0.414 (0.344-0.523) | 0.028 ($3.55x10^{-5}$) | 0.176 (0.002) | 0.189 (0.005) |
| *Zaprionus* spp. | | 0.032 ($4.42x10^{-5}$) | 0.177 (0.004) | 0.310 (0.019) |
| *melanogaster* spp. | | 0.019 ($4.85x10^{-5}$) | 0.119 (0.002) | 0.153 (0.002) |
| Subfamily 2 – V2 | | | | |
| *Zaprionus* x *melanogaster* spp. | | 0.028 ($2.65x10^{-5}$) | 0.106 ($2.62x10^{-4}$) | 0.291 (0.004) |
| *Zaprionus* spp. | | 0.027 ($2.50x10^{-5}$) | 0.088 ($2.55x10^{-4}$) | 0.365 (0.011) |
| *melanogaster* spp. | | 0.018 ($1.10x10^{-6}$) | 0.070 ($1.90x10^{-4}$) | 0.477 (0.013) |
| *micropia* | | | | |
| *Zaprionus* x *melanogaster* spp. | 0.467 (0.410-0.527) | 0.045 ($1.44x10^{-5}$) | 0.124 ($1.94x10^{-4}$) | 0.524 (0.015) |
| *Zaprionus* spp. | | 0.026 ($3.80x10^{-5}$) | 0.040 ($2.45x10^{-4}$) | 0.579 (0.019) |
| *melanogaster* spp. | | 0.028 ($6.45x10^{-5}$) | 0.068 ($4.87x10^{-4}$) | 0.388 (0.005) |
| *Gpdh* | | | | |
| *Zaprionus* x *melanogaster* spp | 0.602 (0.492-0.700) | 0.009 ($3.07x10^{-6}$) | 1.012 (0.004) | 0.009 ($4.90x10^{-6}$) |
| *Zaprionus* spp. | | 0.005 ($2.16x10^{-5}$) | 0.150 (0.002) | 0.041 ($5.43x10^{-4}$) |
| *melanogaster* spp. | | 0 | 0.193 (0.003) | 0 |

**Table 4.** Time of divergence and horizontal transfer events identified for the *gypsy* and *micropia* retrotransposons between *Zaprionus* and *melanogaster* species.

| Species 1 | Species 2 | MYA | Inference |
|---|---|---|---|
| *gypsy* | | | |
| Subfamily 1 | | | |
| *D. melanogaster*/*D. simulans** | *Z. tuberculatus*1 | 2.594 | A,B,C |
| *D. melanogaster* | *D. erecta*A1 | 0.406-2.125 | A,B,C,D |
| Subfamily 2 - V1 | | | |
| *D. erecta* | *Z. tuberculatus* | 3.593-7.688 | A,B,C |
| *D. erecta* | *Z. davidi*/*Z. camerounensis** | 0.843-2.424 | A,B,C |
| *D. erecta* | *D. yakuba* | 2.156-2.625 | B,C,D |
| *D. erecta* | *D. melanogaster* | 1.718-2.656 | A,B,C,D |
| Subfamily 2 - V2 | | | |
| *D. melanogaster*/*D. simulans*/*D. sechellia** | *indianus* complex* | 0.437-3.219 | A,B,C |
| *D. melanogaster*/*D. simulans*/*D. sechellia** | *Z. davidi* | 0.437-1.343 | A,B,C |
| *Z. davidi* | *Z. africanus* | 1.343 | A,B,C |
| *micropia* | | | |
| ? | *D. melanogaster* | 0 | C |
| ? | *D. yakuba* | 1.156-1.781 | C |
| ? | *indianus* complex | 0-1.181 | C |
| ? | *Z. davidi* | 0 | C |
| ? | *Z. camerounensis* | 0 | C |
| *D. melanogaster* | *D. simulans/D. sechellia* | 0-0.500 | A,B,C |
| *D. melanogaster* or *D. simulans/D. sechellia* | *Z. tuberculatus* | 0.500-1.000 | A,B,C |
| *indianus* or *davidi* complexes* | *D. sechellia** | 1.781-2.406 | A,B,C |

Symbols - A: phylogenetic tree, B: dS comparison of retrotransposon *vs.* host gene, C: network, D: Ludwig et al. 2008,*: Species ancestor.

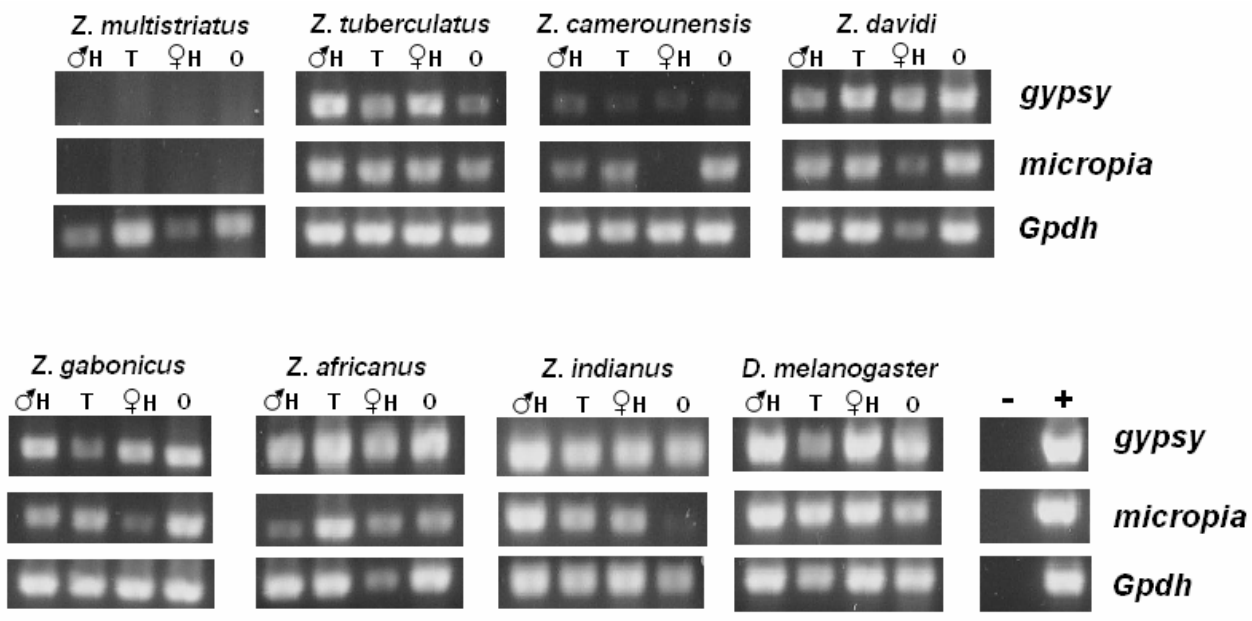



**Figure 1.** RT-PCR fragments obtained from ovaries (O), testes (T) and heads (H) for *gypsy* and *micropia* retrotransposons. -: negative control with ultrapure water; + positive control with *D. melanogaster* genomic DNA. *Gpgh* gene amplification was used as total RNA quality control.

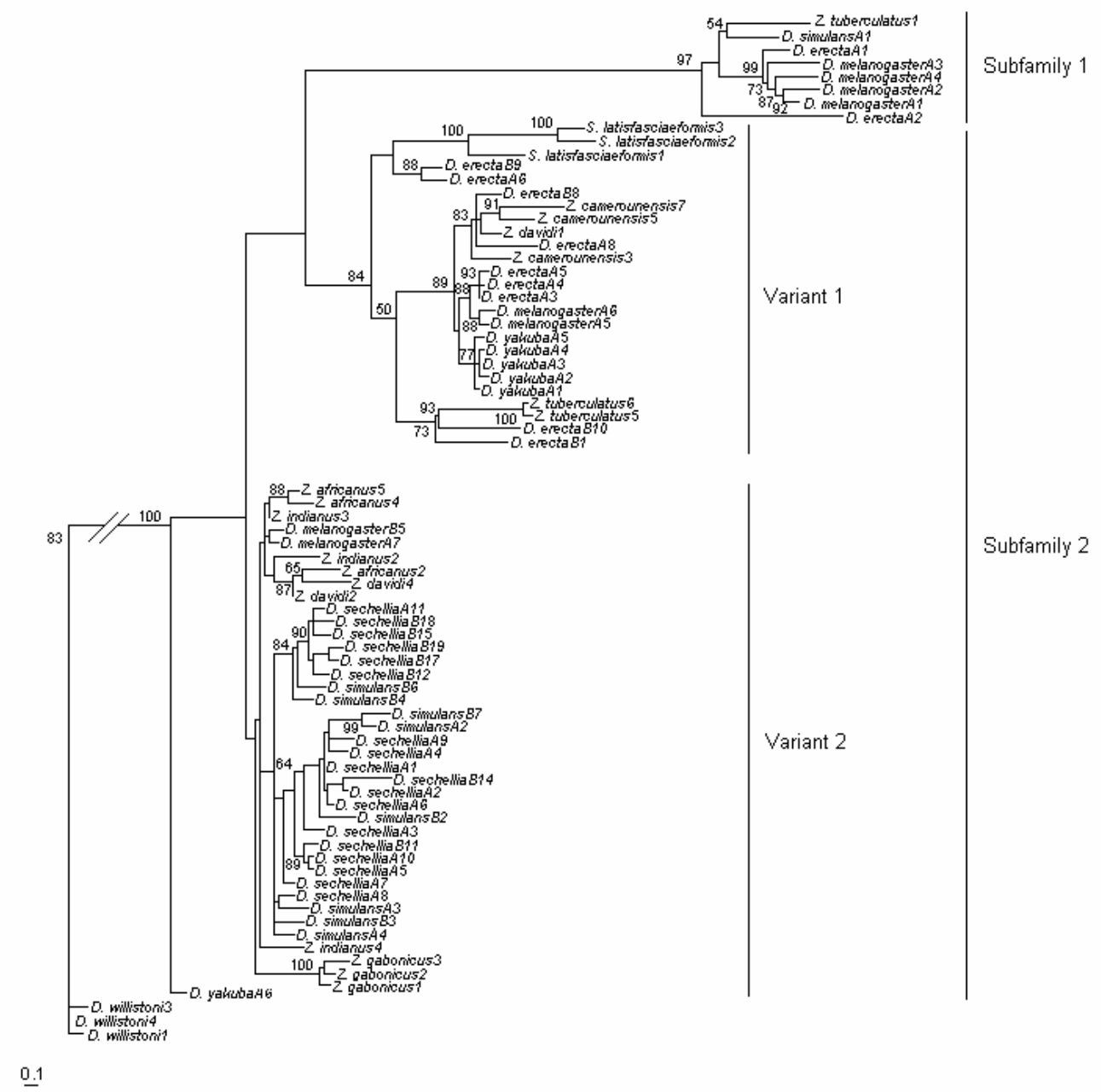



**Figure 2.** Phylogenetic relationships between *gypsy* sequences of genus *Zaprionus*, the *melanogaster* group, *D. willistoni* and *S. latisfasciaeformis*. The tree was constructed using the maximum-likelihood method (HKY85 distance) as implemented in the PhyML program. Numbers on the branches indicate bootstrap values based on 1000 replications. Numbers next to species names indicate the clone or genomic sequences identification.

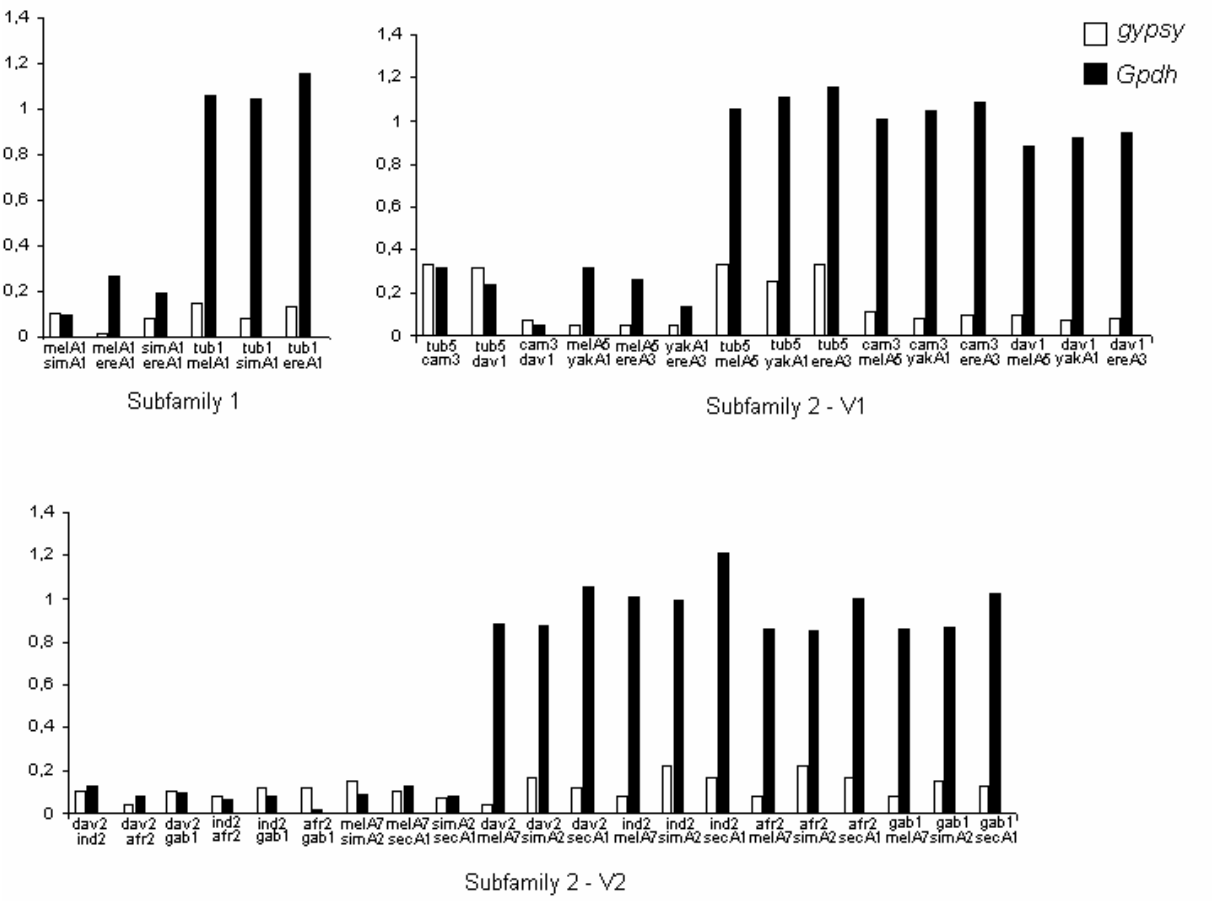



**Figure 3.** Comparative analyses of the dS values between *gypsy* and *Gpdh* sequences of *Zaprionus* and *melanogaster* species within the Subfamily 1, Subfamily 2 - V1 and V2. tub: *Z. tuberculatus*, cam: *Z. camerounensis*, dav: *Z. davidi*, gab: *Z. gabonicus*, afr: *Z. africanus*, ind: *Z. indianus*, mel: *D. melanogaster*, sim: *D. simulans*, sec: *D. sechellia*, yak: *D. yakuba*, ere: *D. erecta*.

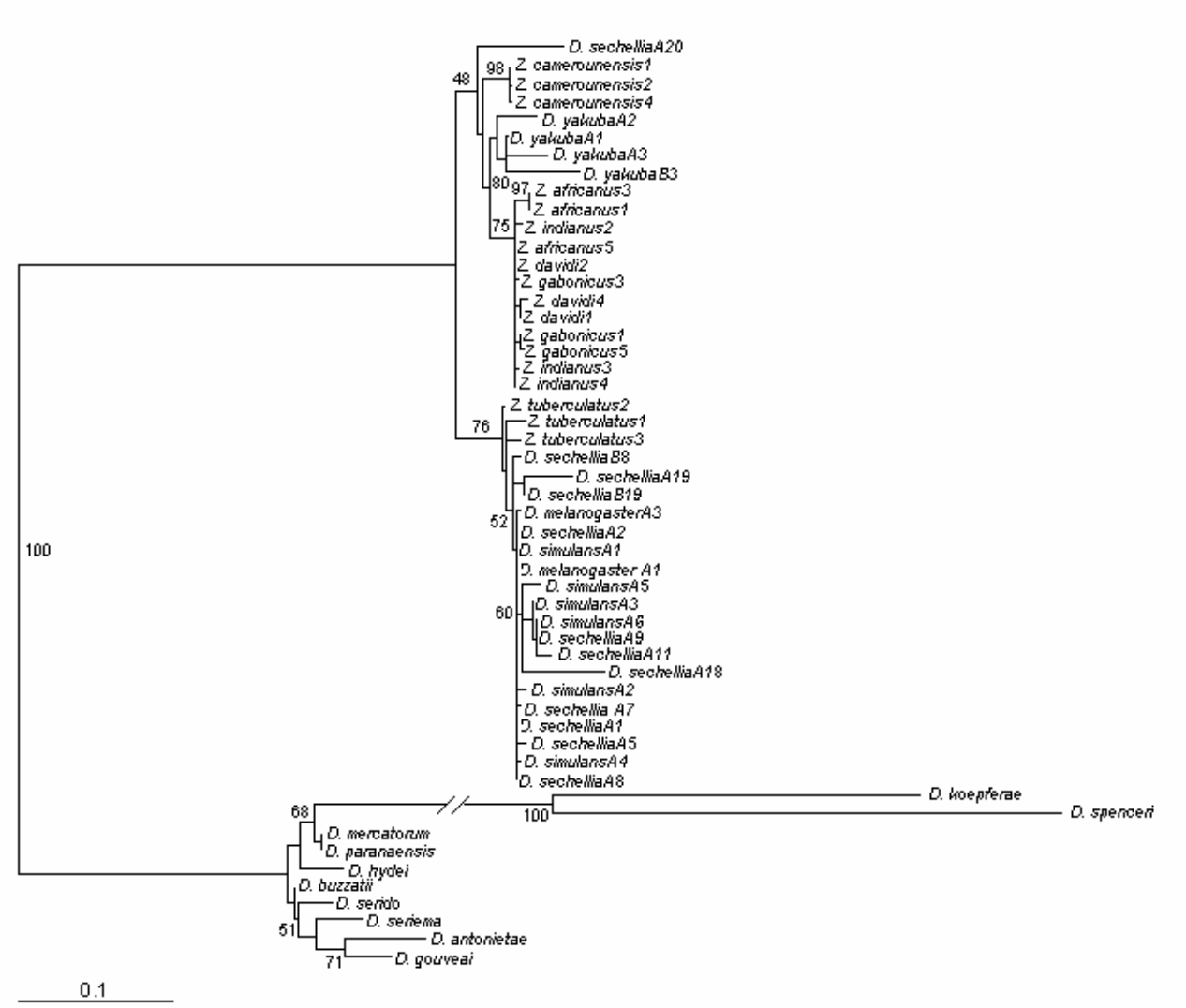



**Figure 4.** Maximum-likelihood (HKY85 distance) tree of *micropia* of genus *Zaprionus* and the *melanogaster* and *repleta* groups. Numbers on the branches indicate bootstrap values based on 1000 replications. Numbers next to species names identify the clone or genomic sequences identification.

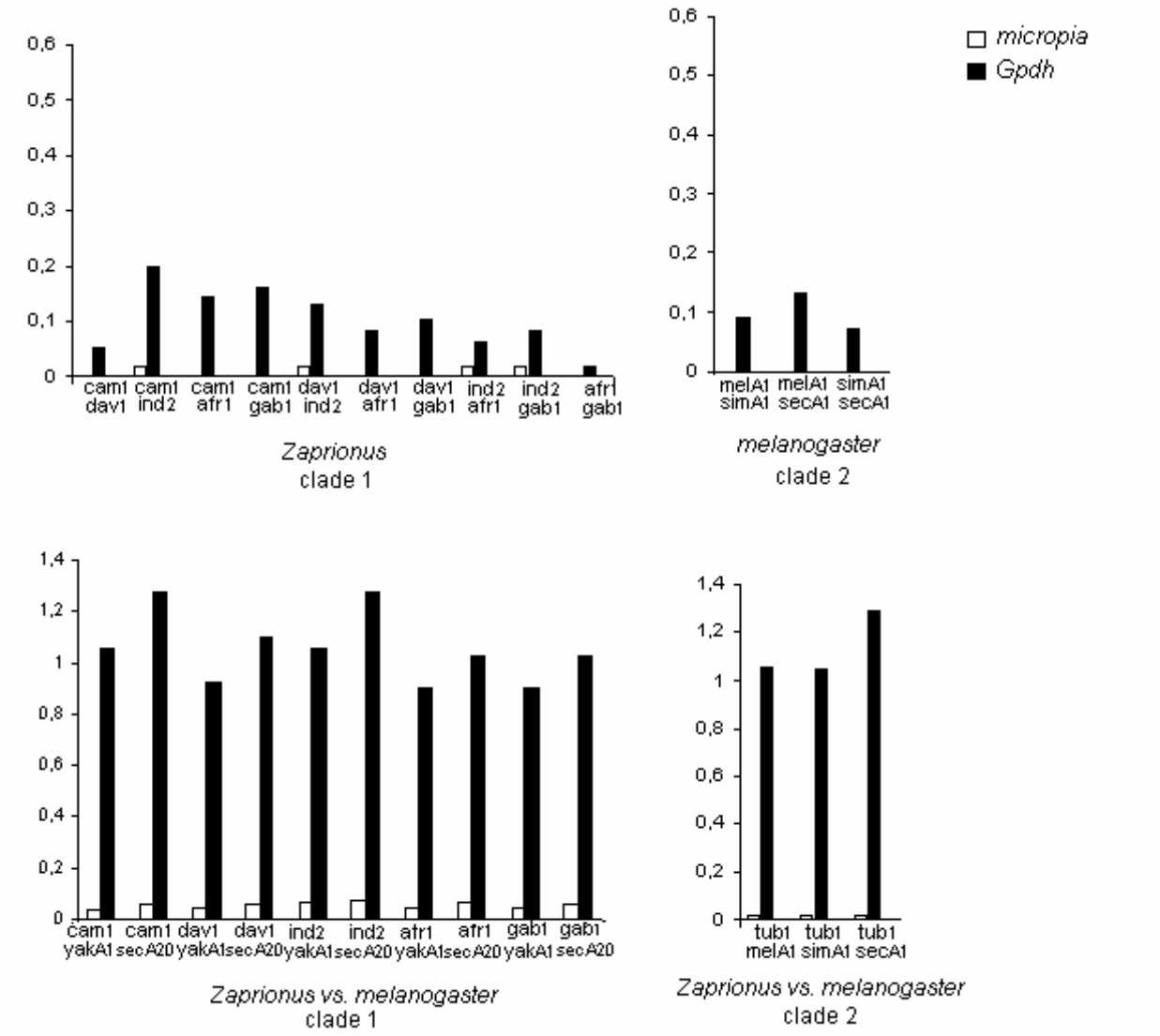



**Figure 5.** Comparative analyses of the dS values between *micropia* and *Gpdh* of *Zaprionus* and *melanogaster* species. Comparison with only black column indicates that the synonymous sites of *micropia* are invariable. tub: *Z. tuberculatus*, cam: *Z. camerounensis*, dav: *Z. davidi*, gab: *Z. gabonicus*, afr: *Z. africanus*, ind: *Z. indianus*, mel: *D. melanogaster*, sim: *D. simulans*, sec: *D. sechellia*, yak: *D. yakuba*.

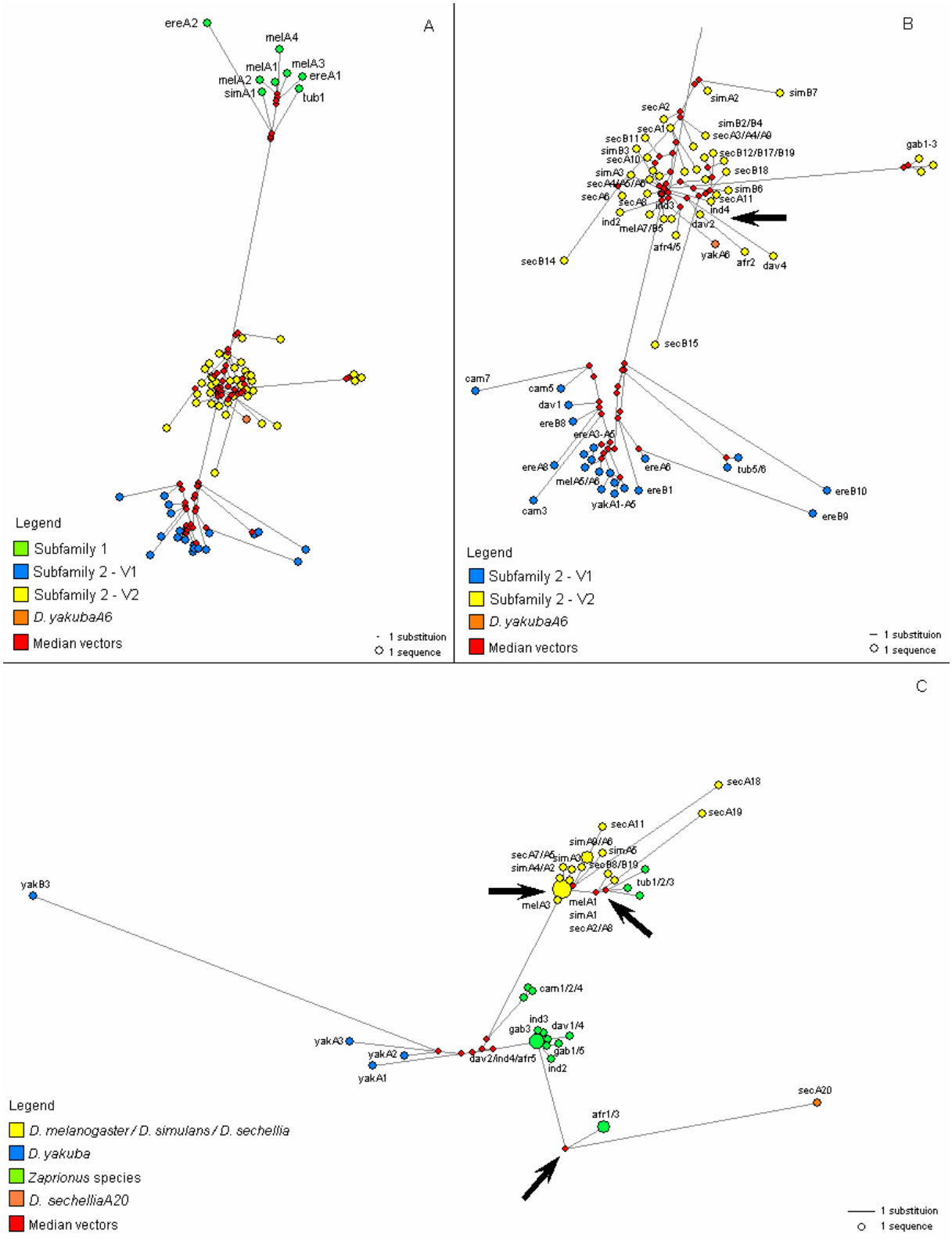



**Figure 6.** Median-joining network analyses for the *Zaprionus* and the *melanogaster* species. The size of circles denotes number of sequences grouped together. (A) Network for the *gypsy* retrotransposon. Each group obtained in the phylogenetic tree is differently colored. (B) Detail of the *gypsy* Subfamily 2. (C) Network for the *micropia* retrotransposon. The sequences are colored according the clustering in the phylogenetic tree. The black arrows show the HT events evidenced by network analysis.

**Supplementary Table 1.** Description of the genomic sequences from the 12 *Drosophila* genome searches. * Sequences excluded from the phylogenetic analysis due to 100% identity between *D. sechellia*B15 / B16, *D. erecta*A6 / A7, *D. erecta*A3 / B3 for *gypsy* and *D. melanogaster*A1 / A2 / A4, *D. sechellia*A2 / A3 / A4, *D. sechellia*A5 / A6, *D. sechellia*A9 / A10 / A12 to A17 and *D. sechellia* B19 / B20 for *micropia*.

| Retrotransposon | Species | Symbol | Region | Position |
|---|---|---|---|---|
| *gypsy* | *D. melanogaster* | A1 | 3RHet | 1969693 - 1970168 |
| | | A2 | 2RHet | 401863 - 402339 |
| | | A3 | 3LHet | 1599041 - 1599517 |
| | | A4 | U | 4045342 - 4045802 |
| | | A5 | 3L | 24462706 - 24463191 |
| | | A6 | 2R | 937680 - 938165 |
| | | A7 | X | 21541979 - 21542452 |
| | | B5 | Uextra | 7342889 - 7343370 |
| | *D. simulans* | A1 | chr3h_Mrandom_053 | 512 - 983 |
| | | A2 | chrX_Mrandom_1054 | 1181 - 1662 |
| | | A3 | chr2h_Mrandom_019 | 37683 - 38164 |
| | | A4 | chrX_Mrandom_1056 | 464 - 945 |
| | | B2 | chr2h_Mrandom_019 | 3424 - 3905 |
| | | B3 | chrX_Mrandom_012 | 2361 - 2840 |
| | | B4 | chrX_Mrandom_1055 | 4088 - 4569 |
| | | B6 | X | 16613272 - 16613752 |
| | | B7 | chrX_Mrandom_1057 | 1701 - 2174 |
| | *D. sechellia* | A1 | scaffold_6727 | 553 - 1034 |
| | | A2 | scaffold_217 | 38501 - 38982 |
| | | A3 | scaffold_2010 | 2401 - 2882 |
| | | A4 | scaffold_157 | 60762 - 61242 |
| | | A5 | scaffold_2705 | 2505 - 2985 |
| | | A6 | scaffold_5941 | 390 - 835 |
| | | A7 | scaffold_3971 | 1615 - 2096 |
| | | A8 | scaffold_1311 | 3875 - 4356 |
| | | A9 | scaffold_2671 | 2966 - 3447 |
| | | A10 | scaffold_7315 | 931 - 1411 |
| | | A11 | scaffold_753 | 1956 - 2436 |
| | | B11 | scaffold_267 | 15475 - 15952 |
| | | B12 | scaffold_1749 | 1881 - 2361 |
| | | B14 | scaffold_865 | 5141 - 5579 |
| | | B15 | scaffold_205 | 11155 - 11595 |
| | | B16* | scaffold_5708 | 673 - 1113 |
| | | B17 | scaffold_1410 | 563 - 1043 |
| | | B18 | scaffold_3135 | 436 - 917 |
| | | B19 | scaffold_905 | 4346 - 4826 |
| | *D. yakuba* | A1 | v2_chrUn_1037 | 262 - 747 |
| | | A2 | v2_chrUn_120 | 95206 - 95691 |
| | | A3 | v2_chrUn_002 | 145016 - 145501 |
| | | A4 | v2_chrUn_838 | 395 - 880 |
| | | A5 | v2_chrX_random_058 | 72970 - 73455 |
| | | A6 | v2_chrUn_884 | 11187 - 11668 |
| | *D. erecta* | A1 | scaffold_4929 | 24204122 - 24204592 |
| | | A2 | scaffold_4845 | 1721676 - 1722133 |
| | | A3 | scaffold_1379 | 4655 - 5140 |
| | | A4 | scaffold_1371 | 5196 - 5681 |
| | | A5 | scaffold_4690 | 18093512 - 18093997 |
| | | A6 | scaffold_3351 | 14812 - 15287 |
| | | A7* | scaffold_4895 | 315532 - 316007 |

**Supplementary Table 1, continuation.**

| Retrotransposon | Species | Symbol | Region | Position |
|---|---|---|---|---|
| *gypsy* | *D. erecta* | A8 | scaffold_603 | 1478 - 1954 |
| | | B1 | scaffold_4784 | 24499320 - 24499797 |
| | | B3* | scaffold_4929 | 21034502 - 21034987 |
| | | B8 | scaffold_3414 | 11250 - 11731 |
| | | B9 | scaffold_1907 | 20931 - 21360 |
| | | B10 | scaffold_3160 | 6841 - 7304 |
| *micropia* | *D. melanogaster* | A1 | Uextra | 9860338 - 9860709 |
| | | A2* | 3RHet | 1083774 - 1084145 |
| | | A3 | 3R | 3177510 - 3177881 |
| | | A4* | U | 9310244 - 9310616 |
| | *D. simulans* | A1 | chr2h_Mrandom_027 | 6515 - 6886 |
| | | A2 | chrU_M_786 | 2638 - 3009 |
| | | A3 | chrU_M_6079 | 7173 - 7536 |
| | | A4 | 2R | 396614 - 396982 |
| | | A5 | chr2h_Mrandom_009 | 176810 - 177181 |
| | | A6 | chrU_M_2288 | 3144 - 3507 |
| | *D. sechellia* | A1 | scaffold_540 | 515 - 886 |
| | | A2 | scaffold_89 | 156 - 527 |
| | | A3* | scaffold_7 | 2883385 - 2883756 |
| | | A4* | scaffold_443 | 40281 - 40652 |
| | | A5 | scaffold_1378 | 3013 - 3384 |
| | | A6* | scaffold_2487 | 559 - 930 |
| | | A7 | scaffold_935 | 1478 - 1849 |
| | | A8 | scaffold_72 | 69278 - 69650 |
| | | A9 | scaffold_28 | 679168 - 679531 |
| | | A10* | scaffold_15 | 715855 - 716218 |
| | | A11 | scaffold_13 | 1014547 - 1014918 |
| | | A12* | scaffold_6 | 3409039 - 3409402 |
| | | A13* | scaffold_0 | 6055065 - 6055428 |
| | | A14* | scaffold_1 | 3884992 - 3885355 |
| | | A15* | scaffold_2 | 3656728 - 3657091 |
| | | A16* | scaffold_2597 | 4417 - 4780 |
| | | A17* | scaffold_1972 | 2457 - 2820 |
| | | A18 | scaffold_2140 | 50 - 417 |
| | | A19 | scaffold_5109 | 119 - 456 |
| | | A20 | scaffold_57 | 128729 - 129077 |
| | | B8 | scaffold_218 | 12547 - 12926 |
| | | B19 | scaffold_62 | 141878 - 142228 |
| | | B20* | scaffold_1078 | 4531 - 4874 |
| | *D. yakuba* | A1 | v2_chr2h_random_005 | 701807 - 702178 |
| | | A2 | v2_chrX_random_046 | 267419 - 267790 |
| | | A3 | v2_chrUn_975 | 14667 - 15015 |
| | | B3 | v2_chrUn_1883 | 5436 - 5787 |

**Supplementary Table 2.** Pairwise genetic distance (maximum composite likelihood distance) among *gypsy* sequences of *Zaprionus*, *melanogaster*, *D. willistoni* and *S. latisfasciaeformis* sequences. The sequences were clustered according the phylogenetic clades.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Dwil | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2. Dwil | .014 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3. Dwil | .006 | .008 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4. Dyak6 | 1.111 | 1.091 | 1.089 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5. Zdav2 | 1.060 | 1.041 | 1.040 | .038 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6. Zdav4 | 1.114 | 1.108 | 1.093 | .059 | .023 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7. Zgab1 | 1.168 | 1.134 | 1.131 | .055 | .047 | .064 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8. Zgab2 | 1.174 | 1.154 | 1.151 | .057 | .040 | .062 | .011 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 9. Zgab3 | 1.190 | 1.185 | 1.166 | .06 | .048 | .064 | .018 | .015 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 10. Zafr2 | 1.147 | 1.128 | 1.126 | .059 | .019 | .036 | .059 | .052 | .064 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 11. Zafr4 | 1.085 | 1.040 | 1.052 | .047 | .029 | .053 | .054 | .057 | .064 | .045 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 12. Zafr5 | 1.087 | 1.068 | 1.053 | .041 | .023 | .042 | .048 | .050 | .050 | .038 | .015 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 13. Zind2 | 1.089 | 1.073 | 1.055 | .047 | .027 | .047 | .050 | .052 | .060 | .038 | .040 | .029 | | | | | | | | | | | | | | | | | | | | | | | | | |
| 14. Zind3 | 1.071 | 1.051 | 1.050 | .028 | .015 | .034 | .038 | .040 | .043 | .029 | .019 | .012 | .025 | | | | | | | | | | | | | | | | | | | | | | | | |
| 15. Zind4 | 1.078 | 1.040 | 1.045 | .043 | .029 | .053 | .045 | .047 | .059 | .043 | .032 | .029 | .036 | .021 | | | | | | | | | | | | | | | | | | | | | | | |
| 16. DmelA7 | 1.120 | 1.107 | 1.098 | .032 | .017 | .035 | .044 | .044 | .047 | .037 | .028 | .019 | .028 | .008 | .028 | | | | | | | | | | | | | | | | | | | | | | |
| 17. DmelB5 | 1.070 | 1.050 | 1.049 | .034 | .021 | .041 | .043 | .046 | .048 | .041 | .030 | .023 | .032 | .010 | .030 | .011 | | | | | | | | | | | | | | | | | | | | | |
| 18. DsimA2 | 1.112 | 1.092 | 1.090 | .061 | .056 | .075 | .067 | .069 | .072 | .075 | .061 | .054 | .068 | .045 | .061 | .050 | .052 | | | | | | | | | | | | | | | | | | | | |
| 19. DsimA3 | 1.128 | 1.109 | 1.106 | .045 | .036 | .056 | .055 | .057 | .060 | .056 | .045 | .039 | .047 | .025 | .041 | .030 | .032 | .041 | | | | | | | | | | | | | | | | | | | |
| 20. DsimA4 | 1.080 | 1.060 | 1.059 | .038 | .025 | .045 | .050 | .053 | .055 | .045 | .038 | .032 | .041 | .019 | .034 | .021 | .025 | .043 | .023 | | | | | | | | | | | | | | | | | | |
| 21. DsimB2 | 1.046 | 1.026 | 1.025 | .065 | .056 | .077 | .077 | .079 | .082 | .077 | .061 | .054 | .066 | .045 | .061 | .048 | .052 | .034 | .041 | .041 | | | | | | | | | | | | | | | | | |
| 22. DsimB3 | 1.060 | 1.040 | 1.039 | .043 | .032 | .052 | .053 | .055 | .058 | .052 | .043 | .036 | .045 | .023 | .038 | .028 | .030 | .048 | .028 | .019 | .048 | | | | | | | | | | | | | | | | |
| 23. DsimB4 | 1.110 | 1.097 | 1.095 | .047 | .034 | .052 | .060 | .062 | .065 | .052 | .043 | .036 | .043 | .023 | .043 | .026 | .025 | .052 | .032 | .025 | .052 | .030 | | | | | | | | | | | | | | | |
| 24. DsimB6 | 1.106 | 1.092 | 1.090 | .049 | .040 | .058 | .062 | .064 | .067 | .058 | .049 | .043 | .049 | .030 | .045 | .035 | .032 | .045 | .034 | .028 | .050 | .032 | .023 | | | | | | | | | | | | | | |
| 25. DsimB7 | 1.087 | 1.067 | 1.066 | .064 | .064 | .083 | .076 | .078 | .081 | .083 | .064 | .062 | .076 | .053 | .069 | .056 | .057 | .017 | .048 | .050 | .039 | .053 | .057 | .057 | | | | | | | | | | | | | |
| 26. DsecA1 | 1.075 | 1.055 | 1.054 | .047 | .038 | .059 | .060 | .062 | .065 | .059 | .043 | .036 | .050 | .028 | .043 | .032 | .034 | .021 | .028 | .030 | .017 | .034 | .039 | .036 | .028 | | | | | | | | | | | | |
| 27. DsecA2 | 1.083 | 1.063 | 1.062 | .052 | .045 | .065 | .067 | .070 | .072 | .065 | .049 | .043 | .056 | .034 | .049 | .039 | .041 | .023 | .034 | .036 | .028 | .041 | .045 | .043 | .030 | .010 | | | | | | | | | | | |
| 28. DsecA3 | 1.100 | 1.080 | 1.078 | .050 | .041 | .059 | .060 | .062 | .065 | .059 | .045 | .038 | .052 | .030 | .045 | .035 | .036 | .032 | .028 | .030 | .030 | .034 | .039 | .039 | .039 | .017 | .023 | | | | | | | | | | |
| 29. DsecA4 | 1.104 | 1.089 | 1.087 | .054 | .045 | .063 | .065 | .067 | .070 | .066 | .050 | .043 | .057 | .034 | .050 | .039 | .039 | .028 | .034 | .036 | .025 | .041 | .041 | .039 | .035 | .010 | .017 | .028 | | | | | | | | | |
| 30. DsecA5 | 1.131 | 1.117 | 1.114 | .041 | .032 | .052 | .053 | .055 | .058 | .052 | .036 | .030 | .043 | .021 | .036 | .026 | .028 | .039 | .032 | .026 | .034 | .030 | .034 | .034 | .046 | .017 | .028 | .017 | .028 | | | | | | | |
| 31. DsecA6 | 1.215 | 1.203 | 1.190 | .049 | .042 | .061 | .063 | .063 | .063 | .059 | .046 | .037 | .054 | .030 | .046 | .037 | .039 | .025 | .035 | .037 | .023 | .042 | .047 | .044 | .033 | .004 | .011 | .016 | .016 | .021 | | | | | | | |
| 32. DsecA7 | 1.113 | 1.093 | 1.091 | .038 | .030 | .050 | .050 | .053 | .055 | .050 | .034 | .028 | .041 | .019 | .034 | .024 | .025 | .034 | .023 | .017 | .034 | .021 | .025 | .025 | .041 | .021 | .028 | .021 | .028 | .017 | .028 | | | | | | |
| 33. DsecA8 | 1.100 | 1.080 | 1.078 | .038 | .030 | .050 | .050 | .053 | .055 | .050 | .038 | .032 | .041 | .019 | .034 | .024 | .025 | .039 | .019 | .017 | .039 | .021 | .025 | .028 | .044 | .025 | .032 | .025 | .032 | .026 | .032 | .017 | | | | | |
| 34. DsecA9 | 1.116 | 1.096 | 1.094 | .045 | .036 | .057 | .058 | .060 | .063 | .057 | .045 | .039 | .048 | .026 | .041 | .030 | .032 | .028 | .026 | .028 | .028 | .032 | .037 | .034 | .035 | .013 | .019 | .023 | .019 | .026 | .016 | .023 | .023 | | | | |
| 35. DsecA10 | 1.119 | 1.104 | 1.102 | .043 | .034 | .052 | .055 | .058 | .060 | .054 | .039 | .032 | .045 | .023 | .039 | .028 | .030 | .039 | .032 | .026 | .034 | .030 | .034 | .034 | .046 | .017 | .028 | .021 | .028 | .004 | .023 | .017 | .026 | .030 | | | |
| 36. DsecA11 | 1.123 | 1.109 | 1.106 | .047 | .038 | .056 | .062 | .065 | .067 | .056 | .047 | .041 | .047 | .028 | .047 | .032 | .030 | .057 | .034 | .030 | .057 | .034 | .021 | .023 | .064 | .043 | .050 | .043 | .045 | .039 | .052 | .030 | .030 | .041 | .039 | | |
| 37. DsecB11 | 1.130 | 1.116 | 1.113 | .046 | .037 | .055 | .058 | .061 | .063 | .057 | .041 | .034 | .048 | .026 | .041 | .030 | .032 | .041 | .035 | .028 | .037 | .032 | .037 | .037 | .046 | .019 | .030 | .024 | .030 | .011 | .023 | .019 | .028 | .032 | .011 | .041 | |
| 38. DsecB12 | 1.116 | 1.102 | 1.099 | .054 | .041 | .059 | .065 | .067 | .070 | .059 | .050 | .043 | .050 | .030 | .050 | .035 | .032 | .059 | .037 | .032 | .059 | .037 | .023 | .025 | .067 | .045 | .052 | .045 | .043 | .041 | .054 | .032 | .032 | .043 | .041 | .015 | .044 |
| 39. DsecB14 | 1.253 | 1.240 | 1.226 | .065 | .065 | .085 | .088 | .088 | .089 | .083 | .070 | .060 | .078 | .052 | .067 | .060 | .062 | .038 | .053 | .055 | .040 | .060 | .065 | .062 | .046 | .026 | .023 | .038 | .038 | .043 | .030 | .045 | .050 | .035 | .045 | .070 | .045 |
| 40. DsecB15 | 1.197 | 1.170 | 1.189 | .049 | .037 | .059 | .061 | .063 | .072 | .054 | .044 | .040 | .040 | .028 | .040 | .033 | .030 | .057 | .038 | .030 | .057 | .035 | .021 | .023 | .063 | .042 | .050 | .045 | .045 | .040 | .051 | .030 | .030 | .040 | .040 | .011 | .040 |
| 41. DsecB17 | 1.119 | 1.104 | 1.102 | .059 | .045 | .063 | .070 | .072 | .075 | .063 | .054 | .047 | .054 | .034 | .054 | .039 | .036 | .064 | .041 | .037 | .061 | .039 | .028 | .030 | .072 | .050 | .057 | .050 | .048 | .046 | .059 | .037 | .037 | .048 | .046 | .015 | .048 |
| 42. DsecB18 | 1.124 | 1.110 | 1.108 | .056 | .043 | .061 | .067 | .069 | .072 | .061 | .051 | .045 | .052 | .032 | .052 | .037 | .034 | .059 | .039 | .034 | .059 | .039 | .025 | .027 | .067 | .045 | .052 | .047 | .048 | .043 | .051 | .034 | .034 | .043 | .043 | .013 | .041 |
| 43. DsecB19 | 1.119 | 1.104 | 1.102 | .061 | .047 | .063 | .067 | .070 | .072 | .063 | .056 | .050 | .057 | .036 | .056 | .041 | .039 | .063 | .043 | .039 | .064 | .043 | .030 | .032 | .071 | .052 | .059 | .048 | .050 | .048 | .062 | .039 | .039 | .050 | .048 | .017 | .050 |
| 44. Ztub1 | 1.136 | 1.119 | 1.115 | .231 | .21 | .225 | .243 | .240 | .250 | .222 | .226 | .218 | .225 | .212 | .229 | .221 | .223 | .225 | .232 | .231 | .235 | .238 | .238 | .242 | .238 | .217 | .223 | .223 | .224 | .224 | .220 | .225 | .225 | .227 | .227 | .230 | .232 |
| 45. DmelA1 | 1.271 | 1.239 | 1.247 | .236 | .227 | .237 | .256 | .260 | .268 | .240 | .241 | .233 | .240 | .224 | .244 | .229 | .233 | .236 | .242 | .241 | .237 | .242 | .248 | .253 | .246 | .227 | .233 | .233 | .228 | .234 | .228 | .236 | .230 | .237 | .237 | .243 | .240 |
| 46. DmelA2 | 1.279 | 1.246 | 1.254 | .236 | .227 | .237 | .260 | .260 | .268 | .240 | .238 | .233 | .240 | .225 | .244 | .229 | .233 | .236 | .242 | .242 | .237 | .249 | .248 | .253 | .244 | .228 | .233 | .233 | .229 | .235 | .231 | .236 | .230 | .237 | .238 | .244 | .240 |
| 47. DmelA3 | 1.249 | 1.231 | 1.225 | .246 | .236 | .247 | .267 | .271 | .278 | .249 | .251 | .242 | .250 | .234 | .254 | .238 | .242 | .248 | .252 | .251 | .249 | .258 | .258 | .263 | .256 | .237 | .243 | .243 | .244 | .244 | .238 | .245 | .239 | .247 | .247 | .253 | .249 |
| 48. DmelA4 | 1.406 | 1.397 | 1.395 | .256 | .246 | .257 | .272 | .279 | .291 | .256 | .258 | .249 | .254 | .243 | .258 | .248 | .252 | .255 | .255 | .261 | .259 | .266 | .268 | .270 | .263 | .246 | .252 | .252 | .253 | .253 | .249 | .255 | .249 | .256 | .256 | .259 | .259 |
| 49. DsimA1 | 1.275 | 1.258 | 1.251 | .237 | .228 | .238 | .255 | .258 | .262 | .241 | .242 | .234 | .241 | .225 | .245 | .227 | .231 | .234 | .240 | .239 | .225 | .250 | .240 | .249 | .244 | .225 | .231 | .231 | .229 | .232 | .222 | .233 | .231 | .235 | .235 | .242 | .237 |
| 50. DereA1 | 1.288 | 1.270 | 1.263 | .225 | .215 | .231 | .244 | .248 | .255 | .228 | .227 | .218 | .228 | .213 | .233 | .217 | .221 | .224 | .230 | .230 | .228 | .237 | .236 | .241 | .234 | .216 | .222 | .222 | .222 | .223 | .216 | .224 | .218 | .225 | .226 | .232 | .228 |
| 51. DereA2 | 1.370 | 1.360 | 1.351 | .251 | .238 | .252 | .263 | .267 | .268 | .248 | .250 | .241 | .255 | .232 | .253 | .238 | .241 | .244 | .251 | .250 | .248 | .258 | .257 | .262 | .249 | .235 | .241 | .241 | .242 | .242 | .235 | .244 | .238 | .245 | .245 | .252 | .239 |
| 52. Ztub5 | 1.170 | 1.130 | 1.148 | .126 | .119 | .138 | .157 | .154 | .163 | .132 | .112 | .114 | .134 | .110 | .122 | .123 | .123 | .135 | .133 | .128 | .130 | .128 | .133 | .138 | .132 | .115 | .125 | .128 | .126 | .123 | .118 | .125 | .130 | .123 | .123 | .138 | .127 |
| 53. Ztub6 | 1.164 | 1.152 | 1.142 | .128 | .121 | .136 | .160 | .157 | .160 | .135 | .119 | .112 | .132 | .112 | .129 | .123 | .125 | .138 | .136 | .130 | .133 | .131 | .136 | .140 | .135 | .118 | .128 | .130 | .128 | .126 | .118 | .128 | .133 | .126 | .126 | .141 | .129 |
| 54. Zcam3 | 1.114 | 1.101 | 1.092 | .106 | .111 | .125 | .137 | .140 | .143 | .129 | .111 | .109 | .119 | .100 | .114 | .107 | .108 | .120 | .118 | .110 | .125 | .115 | .113 | .117 | .117 | .105 | .115 | .120 | .115 | .111 | .112 | .110 | .110 | .110 | .108 | .118 | .111 |
| 55. Zcam5 | 1.100 | 1.099 | 1.091 | .115 | .111 | .130 | .147 | .139 | .148 | .128 | .121 | .113 | .123 | .109 | .118 | .114 | .114 | .129 | .127 | .119 | .134 | .125 | .122 | .124 | .126 | .114 | .124 | .124 | .122 | .117 | .119 | .119 | .119 | .120 | .115 | .125 | .118 |
| 56. Zcam7 | 1.163 | 1.151 | 1.141 | .126 | .119 | .139 | .160 | .152 | .160 | .137 | .132 | .124 | .135 | .119 | .129 | .123 | .125 | .143 | .139 | .128 | .146 | .133 | .133 | .138 | .141 | .125 | .136 | .136 | .136 | .129 | .131 | .131 | .131 | .131 | .126 | .139 | .132 |
| 57. Zdav1 | 1.090 | 1.063 | 1.069 | .098 | .099 | .118 | .130 | .127 | .135 | .117 | .104 | .102 | .112 | .093 | .102 | .100 | .100 | .113 | .111 | .103 | .118 | .108 | .106 | .110 | .110 | .098 | .108 | .108 | .108 | .103 | .104 | .103 | .103 | .103 | .098 | .111 | .104 |
| 58. DmelA5 | 1.112 | 1.099 | 1.090 | .101 | .106 | .120 | .127 | .129 | .132 | .124 | .111 | .104 | .114 | .095 | .109 | .102 | .103 | .115 | .113 | .105 | .120 | .110 | .108 | .112 | .112 | .100 | .110 | .115 | .110 | .106 | .107 | .105 | .105 | .106 | .103 | .113 | .106 |

## Supplementary Table 2, continuation.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 59. DmelA6 | 1.105 | 1.092 | 1.084 | .103 | .109 | .123 | .129 | .132 | .135 | .126 | .114 | .107 | .116 | .097 | .111 | .104 | .105 | .108 | .106 | .103 | .113 | .108 | .106 | .110 | .105 | .098 | .108 | .108 | .108 | .108 | .104 | .103 | .103 | .101 | .105 | .111 | .109 |
| 60. DyakA1 | 1.096 | 1.083 | 1.075 | .094 | .099 | .113 | .122 | .124 | .127 | .112 | .099 | .092 | .107 | .083 | .099 | .095 | .096 | .108 | .106 | .098 | .112 | .103 | .101 | .105 | .110 | .093 | .103 | .108 | .103 | .098 | .096 | .098 | .098 | .098 | .096 | .106 | .099 |
| 61. DyakA2 | 1.120 | 1.107 | 1.098 | .101 | .107 | .121 | .127 | .129 | .133 | .119 | .107 | .100 | .114 | .090 | .107 | .102 | .103 | .115 | .113 | .105 | .120 | .111 | .108 | .113 | .117 | .101 | .110 | .115 | .111 | .106 | .104 | .105 | .106 | .106 | .103 | .113 | .107 |
| 62. DyakA3 | 1.108 | 1.095 | 1.086 | .096 | .102 | .116 | .124 | .127 | .130 | .114 | .102 | .095 | .109 | .086 | .102 | .097 | .098 | .110 | .108 | .101 | .115 | .106 | .103 | .108 | .112 | .096 | .105 | .110 | .106 | .101 | .099 | .101 | .101 | .101 | .098 | .108 | .102 |
| 63. DyakA4 | 1.108 | 1.095 | 1.086 | .096 | .102 | .116 | .124 | .127 | .130 | .114 | .102 | .095 | .109 | .086 | .102 | .097 | .098 | .110 | .108 | .101 | .115 | .106 | .103 | .108 | .112 | .096 | .105 | .110 | .106 | .101 | .099 | .101 | .101 | .101 | .098 | .108 | .102 |
| 64. DyakA5 | 1.090 | 1.077 | 1.069 | .101 | .106 | .120 | .129 | .132 | .135 | .119 | .106 | .099 | .114 | .090 | .106 | .102 | .103 | .115 | .113 | .105 | .120 | .110 | .103 | .112 | .114 | .100 | .110 | .115 | .110 | .106 | .104 | .105 | .105 | .105 | .103 | .113 | .106 |
| 65. DereA3 | 1.112 | 1.099 | 1.090 | .096 | .102 | .116 | .124 | .127 | .130 | .119 | .107 | .100 | .109 | .090 | .104 | .097 | .098 | .110 | .108 | .100 | .115 | .106 | .103 | .108 | .107 | .096 | .105 | .110 | .106 | .101 | .101 | .101 | .101 | .101 | .098 | .108 | .102 |
| 66. DereA4 | 1.124 | 1.112 | 1.102 | .098 | .104 | .118 | .127 | .129 | .127 | .122 | .109 | .102 | .112 | .093 | .107 | .100 | .101 | .113 | .111 | .103 | .117 | .108 | .106 | .110 | .110 | .098 | .108 | .113 | .108 | .103 | .101 | .103 | .103 | .103 | .101 | .111 | .104 |
| 67. DereA5 | 1.120 | 1.107 | 1.098 | .101 | .107 | .121 | .130 | .132 | .135 | .124 | .112 | .105 | .114 | .095 | .109 | .102 | .103 | .115 | .113 | .105 | .120 | .111 | .108 | .113 | .112 | .101 | .110 | .115 | .111 | .106 | .107 | .106 | .106 | .106 | .103 | .113 | .107 |
| 68. DereA6 | 1.164 | 1.144 | 1.141 | .093 | .092 | .099 | .125 | .125 | .128 | .107 | .109 | .100 | .105 | .088 | .102 | .093 | .095 | .107 | .105 | .092 | .114 | .100 | .100 | .104 | .106 | .097 | .102 | .107 | .098 | .098 | .099 | .097 | .097 | .102 | .095 | .105 | .098 |
| 69. DereA8 | 1.173 | 1.154 | 1.149 | .112 | .116 | .123 | .146 | .146 | .149 | .132 | .124 | .114 | .124 | .106 | .121 | .112 | .114 | .126 | .124 | .116 | .134 | .122 | .119 | .121 | .123 | .114 | .121 | .124 | .122 | .119 | .117 | .116 | .116 | .117 | .117 | .124 | .119 |
| 70. DereB1 | 1.239 | 1.233 | 1.236 | .112 | .121 | .128 | .152 | .154 | .157 | .137 | .129 | .121 | .132 | .112 | .126 | .118 | .119 | .126 | .129 | .119 | .134 | .124 | .124 | .129 | .126 | .114 | .122 | .129 | .122 | .119 | .118 | .121 | .122 | .122 | .117 | .129 | .120 |
| 71. DereB8 | 1.112 | 1.099 | 1.090 | .103 | .110 | .124 | .131 | .133 | .136 | .127 | .115 | .108 | .118 | .098 | .112 | .105 | .105 | .117 | .115 | .108 | .122 | .113 | .110 | .115 | .114 | .103 | .112 | .112 | .113 | .108 | .109 | .108 | .108 | .108 | .105 | .115 | .109 |
| 72. DereB9 | 1.354 | 1.334 | 1.325 | .084 | .084 | .100 | .122 | .122 | .123 | .092 | .100 | .089 | .095 | .076 | .092 | .082 | .087 | .100 | .095 | .079 | .097 | .090 | .090 | .095 | .099 | .084 | .092 | .092 | .092 | .082 | .089 | .087 | .087 | .090 | .082 | .095 | .082 |
| 73. DereB10 | 1.173 | 1.145 | 1.150 | .129 | .133 | .143 | .157 | .160 | .164 | .147 | .136 | .128 | .144 | .118 | .136 | .125 | .131 | .144 | .147 | .133 | .147 | .142 | .139 | .144 | .142 | .133 | .144 | .147 | .142 | .137 | .137 | .139 | .139 | .137 | .136 | .142 | .140 |
| 74. Slat1 | 1.180 | 1.140 | 1.144 | .117 | .106 | .133 | .140 | .139 | .150 | .124 | .114 | .111 | .121 | .106 | .116 | .114 | .115 | .129 | .127 | .119 | .124 | .125 | .127 | .129 | .124 | .105 | .110 | .107 | .117 | .105 | .101 | .112 | .120 | .117 | .107 | .127 | .108 |
| 75. Slat2 | 1.230 | 1.203 | 1.191 | .150 | .133 | .162 | .174 | .171 | .180 | .149 | .143 | .141 | .152 | .141 | .151 | .147 | .150 | .163 | .158 | .150 | .160 | .153 | .164 | .166 | .155 | .142 | .150 | .147 | .155 | .140 | .147 | .147 | .155 | .155 | .142 | .156 | .143 |
| 76. Slat3 | 1.204 | 1.171 | 1.164 | .129 | .117 | .138 | .152 | .149 | .158 | .136 | .125 | .120 | .131 | .120 | .130 | .126 | .129 | .139 | .134 | .134 | .136 | .134 | .143 | .145 | .130 | .115 | .126 | .123 | .129 | .118 | .120 | .126 | .134 | .129 | .121 | .135 | .121 |

## Supplementary Table 2, continuation.

| | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 39. DsecB14 | .073 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 40. DsecB15 | .018 | .069 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 41. DsecB17 | .017 | .078 | .023 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 42. DsecB18 | .019 | .070 | .014 | .023 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 43. DsecB19 | .019 | .081 | .026 | .011 | .025 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 44. Ztub1 | .233 | .247 | .228 | .242 | .232 | .242 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 45. DmelA1 | .243 | .246 | .235 | .250 | .243 | .250 | .069 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 46. DmelA2 | .246 | .250 | .236 | .250 | .243 | .250 | .078 | .021 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 47. DmelA3 | .253 | .257 | .246 | .266 | .253 | .266 | .081 | .035 | .044 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 48. DmelA4 | .263 | .265 | .249 | .272 | .262 | .272 | .076 | .027 | .036 | .041 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 49. DsimA1 | .242 | .244 | .237 | .251 | .241 | .251 | .053 | .056 | .065 | .067 | .067 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 50. DereA1 | .232 | .237 | .226 | .244 | .231 | .244 | .065 | .026 | .035 | .033 | .034 | .056 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 51. DereA2 | .252 | .260 | .237 | .265 | .242 | .262 | .102 | .099 | .104 | .112 | .111 | .088 | .092 | | | | | | | | | | | | | | | | | | | | | | | | | |
| 52. Ztub5 | .136 | .134 | .130 | .144 | .141 | .147 | .231 | .247 | .241 | .256 | .263 | .244 | .232 | .252 | | | | | | | | | | | | | | | | | | | | | | | | |
| 53. Ztub6 | .138 | .134 | .133 | .147 | .143 | .149 | .233 | .250 | .244 | .259 | .266 | .247 | .234 | .249 | .006 | | | | | | | | | | | | | | | | | | | | | | | |
| 54. Zcam3 | .121 | .128 | .116 | .123 | .115 | .126 | .234 | .237 | .233 | .248 | .252 | .242 | .232 | .256 | .092 | .094 | | | | | | | | | | | | | | | | | | | | | | |
| 55. Zcam5 | .127 | .135 | .121 | .130 | .122 | .133 | .236 | .247 | .244 | .259 | .260 | .250 | .243 | .267 | .096 | .099 | .038 | | | | | | | | | | | | | | | | | | | | | |
| 56. Zcam7 | .142 | .148 | .130 | .144 | .136 | .147 | .253 | .258 | .254 | .270 | .274 | .268 | .254 | .283 | .099 | .102 | .049 | .040 | | | | | | | | | | | | | | | | | | | | |
| 57. Zdav1 | .113 | .117 | .106 | .116 | .108 | .118 | .231 | .239 | .236 | .251 | .255 | .242 | .232 | .249 | .083 | .090 | .030 | .029 | .041 | | | | | | | | | | | | | | | | | | | |
| 58. DmelA5 | .116 | .122 | .105 | .118 | .110 | .121 | .225 | .229 | .229 | .244 | .248 | .232 | .228 | .239 | .087 | .089 | .038 | .047 | .056 | .034 | | | | | | | | | | | | | | | | | | |
| 59. DmelA6 | .113 | .112 | .105 | .116 | .108 | .118 | .225 | .229 | .229 | .244 | .248 | .232 | .228 | .243 | .085 | .087 | .040 | .049 | .058 | .036 | .010 | | | | | | | | | | | | | | | | | |
| 60. DyakA1 | .108 | .111 | .100 | .111 | .103 | .114 | .208 | .212 | .209 | .224 | .227 | .215 | .208 | .227 | .078 | .080 | .034 | .042 | .051 | .029 | .021 | .023 | | | | | | | | | | | | | | | | |
| 61. DyakA2 | .116 | .120 | .108 | .119 | .111 | .121 | .214 | .219 | .215 | .230 | .234 | .221 | .215 | .231 | .078 | .080 | .038 | .047 | .056 | .034 | .021 | .023 | .008 | | | | | | | | | | | | | | | |
| 62. DyakA3 | .111 | .114 | .103 | .114 | .106 | .116 | .214 | .218 | .215 | .230 | .233 | .221 | .214 | .230 | .078 | .080 | .034 | .042 | .051 | .030 | .017 | .019 | .004 | .004 | | | | | | | | | | | | | | |
| 63. DyakA4 | .111 | .114 | .103 | .114 | .106 | .116 | .216 | .221 | .218 | .233 | .236 | .224 | .217 | .230 | .080 | .082 | .036 | .044 | .054 | .032 | .019 | .021 | .006 | .006 | .002 | | | | | | | | | | | | | |
| 64. DyakA5 | .116 | .119 | .108 | .118 | .110 | .121 | .216 | .221 | .217 | .232 | .235 | .223 | .214 | .233 | .080 | .082 | .036 | .044 | .054 | .027 | .023 | .025 | .006 | .010 | .006 | .008 | | | | | | | | | | | | |
| 65. DereA3 | .111 | .117 | .103 | .114 | .106 | .116 | .219 | .223 | .224 | .238 | .242 | .227 | .223 | .240 | .082 | .085 | .032 | .040 | .047 | .027 | .012 | .015 | .017 | .021 | .017 | .019 | .019 | | | | | | | | | | | |
| 66. DereA4 | .114 | .117 | .105 | .116 | .108 | .119 | .222 | .227 | .227 | .242 | .246 | .226 | .226 | .240 | .085 | .087 | .034 | .042 | .049 | .030 | .015 | .017 | .019 | .023 | .019 | .021 | .021 | .002 | | | | | | | | | | |
| 67. DereA5 | .116 | .123 | .108 | .119 | .111 | .121 | .225 | .230 | .230 | .245 | .249 | .233 | .229 | .246 | .087 | .090 | .036 | .045 | .052 | .032 | .017 | .019 | .021 | .025 | .021 | .023 | .023 | .004 | .006 | | | | | | | | | |
| 68. DereA6 | .108 | .112 | .105 | .105 | .107 | .108 | .227 | .227 | .224 | .247 | .248 | .232 | .229 | .236 | .087 | .087 | .082 | .089 | .094 | .075 | .073 | .075 | .066 | .073 | .068 | .070 | .072 | .070 | .073 | .075 | | | | | | | | |
| 69. DereA8 | .127 | .133 | .120 | .130 | .122 | .132 | .237 | .233 | .236 | .252 | .249 | .240 | .236 | .257 | .111 | .111 | .046 | .048 | .059 | .037 | .050 | .052 | .041 | .050 | .046 | .048 | .048 | .043 | .046 | .048 | .066 | | | | | | | |
| 70. DereB1 | .132 | .132 | .120 | .135 | .132 | .137 | .239 | .243 | .240 | .258 | .258 | .246 | .239 | .245 | .070 | .068 | .086 | .093 | .098 | .082 | .081 | .084 | .072 | .077 | .077 | .079 | .079 | .077 | .079 | .082 | .044 | .071 | | | | | | |
| 71. DereB8 | .118 | .125 | .113 | .121 | .113 | .121 | .242 | .245 | .241 | .257 | .261 | .248 | .241 | .256 | .093 | .095 | .030 | .034 | .045 | .021 | .034 | .036 | .030 | .034 | .030 | .032 | .032 | .028 | .030 | .032 | .080 | .037 | .086 | | | | | |
| 72. DereB9 | .098 | .103 | .094 | .101 | .095 | .104 | .242 | .248 | .241 | .261 | .266 | .238 | .234 | .241 | .079 | .079 | .086 | .094 | .100 | .078 | .076 | .076 | .065 | .073 | .068 | .070 | .073 | .073 | .073 | .078 | .017 | .076 | .051 | .084 | | | | |
| 73. DereB10 | .151 | .149 | .131 | .151 | .145 | .150 | .267 | .262 | .259 | .275 | .276 | .262 | .255 | .272 | .070 | .068 | .092 | .101 | .107 | .087 | .087 | .089 | .075 | .077 | .077 | .079 | .079 | .082 | .084 | .087 | .088 | .107 | .068 | .089 | .090 | | | |
| 74. Slat1 | .128 | .119 | .124 | .135 | .130 | .138 | .236 | .244 | .240 | .255 | .257 | .252 | .234 | .257 | .104 | .111 | .109 | .113 | .117 | .100 | .097 | .099 | .092 | .099 | .095 | .097 | .099 | .097 | .099 | .102 | .078 | .116 | .101 | .108 | .071 | .122 | | |
| 75. Slat2 | .159 | .159 | .161 | .168 | .161 | .168 | .256 | .262 | .255 | .272 | .263 | .264 | .258 | .273 | .146 | .149 | .144 | .149 | .155 | .139 | .138 | .141 | .130 | .138 | .133 | .135 | .138 | .136 | .138 | .141 | .111 | .155 | .144 | .142 | .106 | .166 | .079 | |
| 76. Slat3 | .136 | .129 | .137 | .144 | .140 | .149 | .261 | .274 | .267 | .291 | .275 | .277 | .267 | .288 | .120 | .122 | .130 | .133 | .136 | .123 | .119 | .119 | .114 | .119 | .117 | .119 | .122 | .117 | .119 | .122 | .096 | .144 | .127 | .129 | .090 | .146 | .071 | .028 |

Symbols for species names: Dwil: *D. willistoni*; Ztub1: *Z. tuberculatus*; Zcam: *Z. camerounensis*; Zdav: *Z. davidi*; Zgab: *Z. gabonicus*; Zafr: *Z. africanus*; Zind: *Z. indianus*; Dmel: *D. melanogaster*; Dsim: *D. simulans*; Dsec: *D. sechellia*; Dyak: *D. yakuba*; Dere: *D. erecta*.

**Supplementary Table 3.** dN (below) and dS (above) values of pairwise comparisons among *Zaprionus* and *melanogaster gypsy* sequences. Distances calculated by Nei-Gojobori method with Jukes-Cantor's correction as implemented by MEGA 4.1.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. DmelA7 | - | .014 | .152 | .080 | .057 | .151 | .043 | .028 | .072 | .152 | .103 | .118 | .119 | .087 | .072 | .073 | .088 | .058 | .072 | .042 | .072 | .043 | .028 | .042 | .028 | .043 | .028 | .087 | .028 | .072 | .088 | .057 |
| 2. DmelB5 | .000 | - | .136 | .065 | .042 | .135 | .028 | .028 | .057 | .136 | .087 | .102 | .103 | .072 | .057 | .057 | .072 | .043 | .057 | .028 | .057 | .028 | .014 | .028 | .014 | .028 | .014 | .072 | .014 | .057 | .073 | .042 |
| 3. DsimA2 | .013 | .013 | - | .081 | .119 | .088 | .105 | .136 | .104 | .058 | .072 | .087 | .058 | .057 | .103 | .104 | .120 | .073 | .103 | .136 | .103 | .137 | .120 | .136 | .120 | .171 | .154 | .223 | .152 | .170 | .226 | .151 |
| 4. DsimA3 | .015 | .015 | .020 | - | .050 | .080 | .036 | .065 | .066 | .081 | .065 | .080 | .050 | .050 | .096 | .066 | .050 | .007 | .096 | .065 | .096 | .066 | .050 | .065 | .050 | .097 | .081 | .145 | .080 | .096 | .147 | .111 |
| 5. DsimA4 | .009 | .009 | .013 | .015 | - | .119 | .014 | .042 | .043 | .119 | .103 | .118 | .088 | .087 | .072 | .043 | .057 | .043 | .072 | .042 | .072 | .043 | .028 | .042 | .028 | .072 | .058 | .118 | .057 | .072 | .120 | .087 |
| 6. DsimB2 | .009 | .009 | .013 | .015 | .009 | - | .104 | .144 | .144 | .088 | .042 | .087 | .028 | .057 | .072 | .104 | .120 | .073 | .072 | .143 | .072 | .145 | .127 | .143 | .127 | .170 | .153 | .204 | .151 | .168 | .225 | .151 |
| 7. DsimB3 | .018 | .018 | .022 | .024 | .018 | .018 | - | .028 | .029 | .105 | .088 | .104 | .073 | .073 | .058 | .029 | .043 | .029 | .058 | .028 | .058 | .029 | .014 | .028 | .014 | .058 | .043 | .104 | .043 | .058 | .105 | .072 |
| 8. DsimB4 | .018 | .018 | .022 | .024 | .018 | .015 | .026 | - | .058 | .136 | .119 | .135 | .104 | .103 | .087 | .058 | .073 | .058 | .087 | .028 | .087 | .028 | .014 | .028 | .014 | .058 | .043 | .087 | .042 | .087 | .088 | .072 |
| 9. DsimB6 | .013 | .013 | .018 | .020 | .013 | .011 | .022 | .013 | - | .137 | .119 | .135 | .104 | .103 | .088 | .058 | .073 | .058 | .088 | .057 | .088 | .058 | .043 | .057 | .043 | .088 | .073 | .119 | .072 | .088 | .121 | .103 |
| 10. DsimB7 | .018 | .018 | .009 | .024 | .018 | .018 | .026 | .026 | .022 | - | .072 | .087 | .058 | .057 | .103 | .104 | .101 | .073 | .103 | .136 | .103 | .137 | .120 | .136 | .120 | .171 | .154 | .223 | .152 | .170 | .226 | .151 |
| 11. DsecA1 | .004 | .004 | .009 | .011 | .004 | .004 | .013 | .013 | .009 | .013 | - | .042 | .014 | .014 | .028 | .088 | .103 | .043 | .028 | .119 | .028 | .120 | .103 | .119 | .103 | .119 | .104 | .168 | .103 | .118 | .170 | .102 |
| 12. DsecA2 | .009 | .009 | .013 | .015 | .009 | .009 | .018 | .018 | .013 | .018 | .004 | - | .057 | .028 | .072 | .103 | .119 | .058 | .072 | .135 | .072 | .136 | .118 | .135 | .118 | .135 | .120 | .185 | .118 | .134 | .187 | .117 |
| 13. DsecA3 | .013 | .013 | .018 | .020 | .013 | .013 | .022 | .022 | .018 | .022 | .009 | .004 | - | .028 | .043 | .073 | .088 | .043 | .043 | .104 | .043 | .104 | .088 | .104 | .088 | .137 | .121 | .187 | .119 | .136 | .189 | .119 |
| 14. DsecA4 | .009 | .009 | .013 | .015 | .009 | .009 | .018 | .018 | .013 | .018 | .004 | .009 | .013 | - | .042 | .073 | .088 | .028 | .042 | .103 | .042 | .104 | .087 | .103 | .087 | .103 | .088 | .151 | .087 | .103 | .153 | .087 |
| 15. DsecA5 | .004 | .004 | .009 | .011 | .004 | .004 | .013 | .013 | .009 | .013 | .000 | .004 | .009 | .004 | - | .057 | .103 | .073 | 0 | .087 | 0 | .088 | .072 | .087 | .072 | .088 | .073 | .135 | .072 | .087 | .136 | .071 |
| 16. DsecA7 | .009 | .009 | .013 | .015 | .009 | .009 | .017 | .018 | .013 | .018 | .004 | .009 | .013 | .009 | .004 | - | .073 | .058 | .057 | .058 | .057 | .058 | .043 | .058 | .043 | .089 | .073 | .136 | .073 | .088 | .138 | .072 |
| 17. DsecA8 | .009 | .009 | .013 | .015 | .009 | .009 | .017 | .018 | .013 | .018 | .004 | .009 | .013 | .009 | .004 | .009 | - | .043 | .103 | .073 | .103 | .073 | .057 | .073 | .057 | .104 | .089 | .152 | .088 | .103 | .154 | .119 |
| 18. DsecA9 | .013 | .013 | .017 | .020 | .013 | .013 | .022 | .022 | .017 | .022 | .009 | .013 | .017 | .013 | .009 | .013 | .013 | - | .073 | .058 | .073 | .058 | .043 | .058 | .043 | .073 | .058 | .120 | .058 | .073 | .122 | .088 |
| 19. DsecA10 | .004 | .004 | .009 | .011 | .004 | .004 | .013 | .013 | .009 | .013 | .000 | .004 | .009 | .004 | .000 | .004 | .004 | .009 | - | .087 | 0 | .088 | .072 | .087 | .072 | .088 | .073 | .135 | .072 | .087 | .136 | .071 |
| 20. DsecA11 | .022 | .022 | .026 | .024 | .022 | .020 | .031 | .022 | .009 | .031 | .018 | .022 | .026 | .022 | .018 | .022 | .022 | .026 | .018 | - | .087 | .028 | .014 | 0 | .014 | .057 | .043 | .087 | .042 | .087 | .088 | .072 |
| 21. DsecB11 | .009 | .009 | .013 | .015 | .009 | .009 | .018 | .018 | .013 | .018 | .004 | .009 | .013 | .009 | .004 | .009 | .009 | .013 | .004 | .022 | - | .088 | .072 | .087 | .072 | .088 | .073 | .135 | .072 | .087 | .136 | .071 |
| 22. DsecB12 | .031 | .031 | .035 | .033 | .031 | .029 | .040 | .031 | .018 | .040 | .026 | .031 | .035 | .022 | .026 | .031 | .031 | .035 | .026 | .018 | .031 | - | .014 | .028 | .014 | .058 | .043 | .088 | .043 | .088 | .089 | .072 |
| 23. DsecB17 | .031 | .031 | .036 | .033 | .031 | .029 | .040 | .031 | .018 | .040 | .027 | .031 | .036 | .031 | .027 | .031 | .031 | .035 | .027 | .009 | .031 | .018 | - | .014 | 0 | .043 | .028 | .072 | .028 | .072 | .073 | .057 |
| 24. DsecB18 | .022 | .022 | .026 | .024 | .022 | .020 | .031 | .022 | .009 | .031 | .018 | .022 | .026 | .022 | .018 | .022 | .022 | .026 | .018 | .009 | .022 | .018 | .018 | - | .014 | .057 | .043 | .087 | .042 | .087 | .096 | .072 |
| 25. DsecB19 | .027 | .027 | .031 | .029 | .027 | .024 | .035 | .027 | .013 | .036 | .022 | .027 | .031 | .027 | .022 | .026 | .026 | .031 | .022 | .004 | .027 | .013 | .004 | .013 | - | .043 | .028 | .072 | .028 | .072 | .073 | .057 |
| 26. Zdav2 | .004 | .004 | .018 | .020 | .013 | .013 | .022 | .022 | .018 | .022 | .009 | .013 | .018 | .013 | .009 | .013 | .013 | .017 | .009 | .026 | .013 | .035 | .036 | .026 | .031 | - | .014 | .103 | .043 | .057 | .043 | .042 |
| 27. Zdav4 | .018 | .018 | .031 | .033 | .026 | .026 | .035 | .035 | .031 | .035 | .022 | .026 | .031 | .026 | .022 | .026 | .026 | .031 | .022 | .040 | .026 | .049 | .049 | .040 | .045 | .013 | - | .088 | .028 | .073 | .058 | .057 |
| 28. Zind2 | .013 | .013 | .026 | .029 | .022 | .022 | .031 | .031 | .026 | .031 | .018 | .022 | .026 | .022 | .018 | .022 | .022 | .026 | .018 | .036 | .022 | .045 | .045 | .036 | .040 | .018 | .031 | - | .087 | .103 | .088 | .118 |
| 29. Zind3 | .000 | .000 | .013 | .015 | .009 | .009 | .018 | .018 | .013 | .018 | .004 | .009 | .013 | .009 | .004 | .009 | .009 | .013 | .004 | .022 | .009 | .031 | .031 | .022 | .027 | .004 | .018 | .013 | - | .057 | .058 | .028 |
| 30. Zind4 | .013 | .013 | .026 | .029 | .022 | .022 | .031 | .031 | .026 | .031 | .018 | .022 | .026 | .022 | .018 | .022 | .022 | .026 | .018 | .036 | .022 | .045 | .045 | .036 | .040 | .018 | .031 | .027 | .013 | - | .058 | .057 |
| 31. Zafr2 | .013 | .013 | .026 | .029 | .022 | .022 | .031 | .031 | .026 | .031 | .018 | .022 | .026 | .022 | .018 | .022 | .022 | .026 | .018 | .035 | .022 | .044 | .045 | .033 | .040 | .009 | .022 | .026 | .013 | .026 | - | .057 |
| 32. Zafr4 | .013 | .013 | .027 | .029 | .022 | .022 | .031 | .031 | .027 | .031 | .018 | .022 | .027 | .022 | .018 | .022 | .022 | .026 | .018 | .036 | .022 | .045 | .045 | .036 | .040 | .018 | .031 | .027 | .013 | .027 | .027 | - |
| 33. Zafr5 | .009 | .009 | .022 | .024 | .018 | .018 | .026 | .027 | .022 | .027 | .013 | .018 | .022 | .018 | .013 | .018 | .018 | .022 | .013 | .031 | .018 | .040 | .040 | .031 | .036 | .013 | .026 | .022 | .009 | .022 | .022 | .013 |
| 34. Zgab1 | .022 | .022 | .036 | .038 | .031 | .031 | .040 | .040 | .036 | .040 | .027 | .031 | .036 | .031 | .027 | .031 | .031 | .038 | .027 | .045 | .031 | .054 | .054 | .045 | .049 | .026 | .040 | .036 | .022 | .036 | .035 | .036 |
| 35. Zgab2 | .022 | .022 | .036 | .038 | .031 | .031 | .040 | .040 | .036 | .040 | .027 | .031 | .036 | .031 | .027 | .031 | .031 | .038 | .027 | .045 | .031 | .054 | .054 | .045 | .049 | .026 | .040 | .036 | .022 | .036 | .035 | .036 |
| 36. Zgab3 | .036 | .036 | .049 | .052 | .045 | .045 | .054 | .054 | .049 | .054 | .040 | .045 | .049 | .045 | .040 | .045 | .045 | .051 | .040 | .059 | .045 | .068 | .068 | .059 | .063 | .040 | .054 | .049 | .036 | .049 | .049 | .049 |
| 37. DmelA1 | .076 | .076 | .091 | .093 | .086 | .086 | .088 | .096 | .091 | .096 | .081 | .084 | .088 | .086 | .081 | .086 | .086 | .093 | .081 | .101 | .086 | .111 | .111 | .101 | .106 | .083 | .088 | .091 | .076 | .093 | .096 | .091 |
| 38. DmelA2 | .074 | .074 | .091 | .093 | .086 | .086 | .098 | .096 | .091 | .095 | .081 | .083 | .088 | .086 | .081 | .086 | .086 | .093 | .081 | .101 | .086 | .110 | .111 | .101 | .106 | .081 | .085 | .088 | .074 | .091 | .093 | .088 |
| 39. DmelA3 | .088 | .088 | .103 | .105 | .098 | .098 | .110 | .108 | .103 | .108 | .093 | .096 | .100 | .098 | .093 | .098 | .098 | .105 | .093 | .113 | .098 | .123 | .123 | .113 | .118 | .095 | .100 | .103 | .088 | .106 | .108 | .103 |
| 40. DsimA1 | .064 | .064 | .078 | .081 | .074 | .074 | .085 | .083 | .078 | .083 | .069 | .071 | .076 | .074 | .069 | .073 | .073 | .081 | .069 | .088 | .074 | .098 | .098 | .088 | .093 | .071 | .076 | .078 | .064 | .081 | .083 | .079 |
| 41. DereA1 | .064 | .064 | .078 | .081 | .074 | .074 | .085 | .083 | .078 | .083 | .069 | .071 | .076 | .074 | .069 | .073 | .073 | .081 | .069 | .088 | .074 | .098 | .098 | .088 | .093 | .071 | .085 | .078 | .064 | .081 | .083 | .079 |
| 42. Ztub1 | .078 | .078 | .093 | .095 | .088 | .088 | .100 | .098 | .093 | .098 | .083 | .086 | .090 | .088 | .083 | .088 | .088 | .095 | .083 | .098 | .088 | .113 | .108 | .103 | .103 | .085 | .090 | .093 | .078 | .096 | .098 | .093 |
| 43. DmelA5 | .009 | .009 | .013 | .024 | .018 | .018 | .027 | .027 | .022 | .018 | .013 | .018 | .022 | .018 | .013 | .018 | .018 | .022 | .013 | .031 | .018 | .040 | .040 | .031 | .036 | .015 | .029 | .022 | .009 | .024 | .024 | .013 |
| 44. DmelA6 | .009 | .009 | .013 | .024 | .018 | .018 | .027 | .027 | .022 | .018 | .013 | .018 | .022 | .018 | .013 | .018 | .018 | .022 | .013 | .031 | .018 | .040 | .040 | .031 | .036 | .015 | .029 | .022 | .009 | .024 | .024 | .013 |
| 45. DyakA1 | .013 | .013 | .018 | .029 | .022 | .022 | .031 | .031 | .027 | .022 | .018 | .022 | .027 | .022 | .018 | .022 | .022 | .027 | .018 | .036 | .022 | .045 | .045 | .036 | .040 | .020 | .033 | .027 | .013 | .029 | .029 | .018 |
| 46. DyakA2 | .013 | .013 | .018 | .029 | .022 | .022 | .031 | .031 | .027 | .022 | .018 | .022 | .027 | .022 | .018 | .022 | .022 | .027 | .018 | .036 | .022 | .045 | .045 | .036 | .040 | .020 | .033 | .027 | .013 | .029 | .029 | .018 |
| 47. DyakA3 | .013 | .013 | .018 | .029 | .022 | .022 | .031 | .031 | .027 | .022 | .018 | .022 | .027 | .022 | .018 | .022 | .022 | .027 | .018 | .036 | .022 | .045 | .045 | .036 | .040 | .020 | .033 | .027 | .013 | .029 | .029 | .018 |
| 48. DyakA4 | .013 | .013 | .018 | .029 | .022 | .022 | .031 | .031 | .027 | .022 | .018 | .022 | .027 | .022 | .018 | .022 | .022 | .027 | .018 | .036 | .022 | .045 | .045 | .036 | .040 | .020 | .033 | .027 | .013 | .029 | .029 | .018 |
| 49. DyakA5 | .013 | .013 | .018 | .029 | .022 | .022 | .031 | .031 | .027 | .022 | .018 | .022 | .027 | .022 | .018 | .022 | .022 | .027 | .018 | .036 | .022 | .045 | .045 | .036 | .040 | .020 | .033 | .027 | .013 | .029 | .029 | .018 |
| 50. DereA3 | .009 | .009 | .013 | .024 | .018 | .018 | .027 | .027 | .022 | .018 | .013 | .018 | .022 | .018 | .013 | .018 | .018 | .022 | .013 | .031 | .018 | .040 | .040 | .031 | .036 | .015 | .029 | .022 | .009 | .024 | .024 | .013 |
| 51. DereA4 | .013 | .013 | .018 | .029 | .022 | .022 | .031 | .031 | .027 | .022 | .018 | .022 | .027 | .022 | .018 | .022 | .022 | .027 | .018 | .036 | .022 | .045 | .045 | .036 | .040 | .020 | .033 | .027 | .013 | .029 | .029 | .018 |
| 52. DereA5 | .009 | .009 | .013 | .024 | .018 | .018 | .027 | .027 | .022 | .018 | .013 | .018 | .022 | .018 | .013 | .018 | .018 | .022 | .013 | .031 | .018 | .040 | .040 | .031 | .036 | .015 | .029 | .022 | .009 | .024 | .024 | .013 |
| 53. DereA8 | .036 | .036 | .040 | .052 | .045 | .045 | .054 | .054 | .049 | .045 | .040 | .045 | .040 | .045 | .040 | .045 | .045 | .049 | .040 | .059 | .045 | .068 | .068 | .059 | .064 | .042 | .047 | .050 | .036 | .052 | .052 | .040 |
| 54. DereB1 | .043 | .043 | .056 | .059 | .052 | .052 | .061 | .061 | .056 | .061 | .047 | .052 | .058 | .052 | .047 | .052 | .052 | .056 | .047 | .066 | .052 | .075 | .076 | .066 | .071 | .049 | .054 | .057 | .043 | .059 | .059 | .047 |
| 55. DereB8 | .031 | .031 | .036 | .047 | .040 | .040 | .049 | .050 | .045 | .040 | .036 | .040 | .036 | .040 | .036 | .040 | .040 | .045 | .036 | .054 | .040 | .064 | .064 | .054 | .059 | .038 | .052 | .045 | .031 | .047 | .047 | .036 |
| 56. DereB10 | .045 | .045 | .050 | .061 | .050 | .055 | .064 | .064 | .059 | .050 | .050 | .055 | .059 | .055 | .050 | .054 | .057 | .059 | .050 | .069 | .055 | .078 | .078 | .069 | .074 | .052 | .066 | .059 | .045 | .062 | .061 | .050 |
| 57. Ztub5 | .031 | .031 | .045 | .047 | .040 | .040 | .040 | .049 | .045 | .049 | .036 | .040 | .045 | .040 | .036 | .040 | .040 | .045 | .036 | .054 | .040 | .063 | .064 | .054 | .059 | .038 | .052 | .045 | .031 | .047 | .047 | .036 |
| 58. Ztub6 | .031 | .031 | .045 | .047 | .040 | .040 | .040 | .049 | .045 | .049 | .036 | .040 | .045 | .040 | .036 | .040 | .040 | .045 | .036 | .054 | .040 | .063 | .064 | .054 | .059 | .038 | .052 | .045 | .031 | .047 | .047 | .036 |
| 59. Zcam3 | .031 | .031 | .036 | .047 | .040 | .040 | .049 | .050 | .045 | .040 | .036 | .040 | .045 | .040 | .036 | .040 | .040 | .045 | .036 | .054 | .040 | .064 | .064 | .054 | .059 | .038 | .052 | .045 | .031 | .047 | .047 | .027 |
| 60. Zcam5 | .022 | .022 | .027 | .038 | .031 | .031 | .040 | .040 | .036 | .031 | .027 | .031 | .027 | .031 | .027 | .031 | .031 | .036 | .027 | .045 | .031 | .054 | .054 | .045 | .050 | .029 | .042 | .036 | .022 | .038 | .038 | .027 |
| 61. Zcam7 | .040 | .040 | .049 | .056 | .050 | .049 | .059 | .059 | .054 | .054 | .045 | .050 | .045 | .050 | .045 | .049 | .049 | .054 | .045 | .064 | .050 | .073 | .073 | .064 | .068 | .047 | .061 | .054 | .040 | .057 | .056 | .045 |
| 62. Zdav1 | .022 | .022 | .027 | .038 | .031 | .031 | .040 | .040 | .036 | .031 | .027 | .031 | .027 | .031 | .027 | .031 | .031 | .036 | .027 | .045 | .031 | .054 | .054 | .045 | .050 | .029 | .042 | .036 | .022 | .038 | .038 | .027 |

**Supplementary Table 3, continuation.**

| | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. DmelA7 | .042 | .087 | .103 | .072 | 1.466 | 1.362 | 1.384 | 1.367 | 1.454 | 1.231 | .425 | .477 | .399 | .422 | .399 | .399 | .399 | .425 | .425 | .425 | .453 | .491 | .380 | .418 | .507 | .507 | .422 | .450 | .429 | .381 |
| 2. DmelB5 | .028 | .072 | .087 | .057 | 1.378 | 1.284 | 1.304 | 1.367 | 1.367 | 1.165 | .401 | .452 | .376 | .399 | .376 | .376 | .376 | .401 | .401 | .401 | .429 | .465 | .358 | .442 | .481 | .481 | .399 | .426 | .406 | .359 |
| 3. DsimA2 | .135 | .152 | .170 | .136 | 1.181 | 1.078 | 1.184 | 1.171 | 1.171 | 1.009 | .560 | .508 | .528 | .556 | .528 | .528 | .528 | .560 | .560 | .560 | .595 | .522 | .506 | .551 | .540 | .540 | .556 | .591 | .567 | .535 |
| 4. DsimA3 | .096 | .112 | .128 | .096 | 1.470 | 1.324 | 1.387 | 1.458 | 1.458 | 1.231 | .495 | .446 | .466 | .491 | .466 | .466 | .466 | .495 | .495 | .495 | .527 | .569 | .445 | .540 | .529 | .529 | .491 | .523 | .500 | .472 |
| 5. DsimA4 | .072 | .087 | .103 | .072 | 1.566 | 1.405 | 1.475 | 1.554 | 1.554 | 1.304 | .449 | .452 | .422 | .446 | .422 | .422 | .422 | .449 | .449 | .449 | .479 | .517 | .403 | .492 | .535 | .535 | .446 | .475 | .454 | .427 |
| 6. DsimB2 | .135 | .186 | .204 | .169 | 1.303 | 1.184 | 1.234 | 1.292 | 1.292 | 1.107 | .555 | .504 | .524 | .552 | .524 | .524 | .524 | .555 | .555 | .555 | .621 | .604 | .502 | .604 | .535 | .535 | .552 | .586 | .562 | .531 |
| 7. DsimB3 | .058 | .073 | .089 | .058 | 1.480 | 1.332 | 1.396 | 1.468 | 1.468 | 1.238 | .433 | .435 | .406 | .430 | .406 | .406 | .406 | .433 | .433 | .433 | .462 | .500 | .387 | .475 | .517 | .517 | .430 | .459 | .437 | .411 |
| 8. DsimB4 | .057 | .103 | .119 | .088 | 1.481 | 1.335 | 1.398 | 1.300 | 1.470 | 1.242 | .427 | .429 | .401 | .424 | .401 | .401 | .401 | .427 | .427 | .427 | .455 | .493 | .381 | .444 | .510 | .510 | .424 | .452 | .431 | .383 |
| 9. DsimB6 | .087 | .103 | .120 | .088 | 1.603 | 1.433 | 1.507 | 1.590 | 1.590 | 1.328 | .453 | .456 | .426 | .450 | .426 | .426 | .426 | .453 | .453 | .453 | .457 | .522 | .406 | .497 | .540 | .540 | .450 | .480 | .458 | .431 |
| 10. DsimB7 | .135 | .152 | .170 | .136 | 1.181 | 1.078 | 1.121 | 1.171 | 1.171 | 1.009 | .505 | .456 | .475 | .501 | .475 | .475 | .475 | .505 | .505 | .505 | .537 | .522 | .455 | .551 | .540 | .540 | .501 | .534 | .510 | .482 |
| 11. DsecA1 | .087 | .135 | .152 | .119 | 1.227 | 1.119 | 1.165 | 1.217 | 1.217 | 1.047 | .474 | .477 | .446 | .471 | .446 | .446 | .446 | .474 | .474 | .474 | .532 | .517 | .426 | .518 | .455 | .455 | .471 | .501 | .479 | .452 |
| 12. DsecA2 | .102 | .151 | .168 | .135 | 1.255 | 1.143 | 1.191 | 1.245 | 1.245 | 1.070 | .525 | .528 | .495 | .522 | .495 | .495 | .495 | .525 | .525 | .525 | .558 | .543 | .474 | .572 | .506 | .506 | .522 | .554 | .531 | .501 |
| 13. DsecA3 | .103 | .152 | .170 | .136 | 1.362 | 1.234 | 1.289 | 1.351 | 1.351 | 1.152 | .505 | .456 | .475 | .501 | .475 | .475 | .475 | .505 | .505 | .505 | .566 | .543 | .455 | .551 | .485 | .485 | .501 | .534 | .510 | .482 |
| 14. DsecA4 | .072 | .118 | .135 | .103 | 1.102 | 1.009 | 1.047 | 1.093 | 1.093 | .946 | .500 | .503 | .471 | .497 | .471 | .471 | .471 | .500 | .500 | .500 | .532 | .517 | .450 | .545 | .481 | .481 | .497 | .528 | .506 | .477 |
| 15. DsecA5 | .057 | .103 | .119 | .087 | 1.299 | 1.181 | 1.231 | 1.288 | 1.288 | 1.104 | .425 | .477 | .399 | .422 | .399 | .399 | .399 | .425 | .425 | .425 | .479 | .465 | .380 | .467 | .455 | .455 | .422 | .450 | .429 | .404 |
| 16. DsecA7 | .057 | .104 | .120 | .088 | 1.508 | 1.356 | 1.421 | 1.496 | 1.496 | 1.260 | .480 | .483 | .452 | .477 | .452 | .452 | .452 | .480 | .480 | .480 | .511 | .552 | .431 | .525 | .514 | .514 | .477 | .508 | .485 | .457 |
| 17. DsecA8 | .103 | .119 | .136 | .104 | 1.405 | 1.270 | 1.328 | 1.393 | 1.393 | 1.328 | .505 | .508 | .475 | .501 | .475 | .475 | .475 | .505 | .505 | .505 | .537 | .579 | .455 | .537 | .599 | .599 | .501 | .534 | .510 | .482 |
| 18. DsecA9 | .073 | .081 | .096 | .065 | 1.308 | 1.187 | 1.238 | 1.297 | 1.297 | 1.108 | .457 | .435 | .430 | .455 | .430 | .430 | .430 | .457 | .457 | .457 | .488 | .528 | .410 | .501 | .490 | .490 | .455 | .485 | .463 | .435 |
| 19. DsecA10 | .057 | .103 | .119 | .087 | 1.299 | 1.181 | 1.231 | 1.288 | 1.288 | 1.104 | .425 | .477 | .399 | .422 | .399 | .399 | .399 | .425 | .425 | .425 | .479 | .465 | .380 | .467 | .455 | .455 | .422 | .450 | .429 | .404 |
| 20. DsecA11 | .057 | .103 | .119 | .087 | 1.307 | 1.187 | 1.238 | 1.296 | 1.296 | 1.109 | .426 | .429 | .400 | .424 | .400 | .400 | .400 | .426 | .426 | .426 | .455 | .492 | .381 | .420 | .509 | .509 | .424 | .452 | .431 | .382 |
| 21. DsecB11 | .057 | .103 | .119 | .087 | 1.299 | 1.181 | 1.231 | 1.288 | 1.288 | 1.104 | .425 | .477 | .399 | .422 | .399 | .399 | .399 | .425 | .425 | .425 | .479 | .465 | .380 | .467 | .455 | .455 | .422 | .450 | .429 | .404 |
| 22. DsecB12 | .057 | .104 | .120 | .088 | 1.255 | 1.207 | 1.190 | 1.245 | 1.245 | 1.068 | .430 | .433 | .404 | .427 | .404 | .404 | .404 | .430 | .430 | .430 | .459 | .497 | .384 | .473 | .461 | .461 | .427 | .456 | .435 | .385 |
| 23. DsecB17 | .042 | .087 | .103 | .072 | 1.382 | 1.252 | 1.308 | 1.371 | 1.371 | 1.168 | .402 | .404 | .377 | .400 | .377 | .377 | .377 | .402 | .402 | .402 | .429 | .466 | .358 | .443 | .482 | .482 | .400 | .427 | .406 | .359 |
| 24. DsecB18 | .057 | .103 | .119 | .087 | 1.307 | 1.187 | 1.238 | 1.296 | 1.296 | 1.109 | .426 | .429 | .400 | .424 | .400 | .400 | .400 | .426 | .426 | .426 | .455 | .492 | .381 | .420 | .509 | .509 | .424 | .452 | .431 | .382 |
| 25. DsecB19 | .042 | .087 | .103 | .072 | 1.387 | 1.255 | 1.312 | 1.375 | 1.375 | 1.171 | .403 | .405 | .378 | .400 | .378 | .378 | .378 | .403 | .403 | .403 | .430 | .466 | .359 | .444 | .482 | .482 | .400 | .427 | .407 | .360 |
| 26. Zdav2 | .057 | .103 | .088 | .088 | 1.449 | 1.347 | 1.369 | 1.438 | 1.438 | 1.152 | .441 | .495 | .414 | .438 | .414 | .414 | .414 | .441 | .441 | .441 | .470 | .509 | .395 | .484 | .472 | .526 | .438 | .418 | .398 | .351 |
| 27. Zdav4 | .043 | .088 | .104 | .073 | 1.470 | 1.364 | 1.387 | 1.458 | 1.458 | 1.231 | .419 | .472 | .393 | .417 | .393 | .393 | .393 | .419 | .419 | .419 | .448 | .485 | .374 | .461 | .502 | .502 | .417 | .445 | .424 | .375 |
| 28. Zind2 | .102 | .118 | .135 | .135 | 1.682 | 1.548 | 1.577 | 1.668 | 1.668 | 1.384 | .474 | .530 | .446 | .471 | .446 | .446 | .446 | .474 | .474 | .474 | .505 | .545 | .426 | .518 | .563 | .563 | .471 | .501 | .479 | .427 |
| 29. Zind3 | .014 | .087 | .103 | .072 | 1.466 | 1.362 | 1.384 | 1.454 | 1.454 | 1.231 | .425 | .477 | .355 | .376 | .355 | .355 | .355 | .425 | .425 | .425 | .453 | .491 | .380 | .418 | .455 | .455 | .422 | .450 | .429 | .381 |
| 30. Zind4 | .072 | .072 | .057 | .087 | 1.515 | 1.405 | 1.428 | 1.502 | 1.502 | 1.197 | .461 | .516 | .411 | .434 | .411 | .411 | .411 | .461 | .461 | .461 | .491 | .531 | .414 | .479 | .468 | .521 | .459 | .438 | .417 | .392 |
| 31. Zafr2 | .073 | .120 | .104 | .137 | 1.519 | 1.407 | 1.431 | 1.507 | 1.507 | 1.197 | .469 | .525 | .393 | .417 | .393 | .393 | .393 | .469 | .469 | .469 | .499 | .540 | .421 | .461 | .450 | .502 | .466 | .445 | .424 | .375 |
| 32. Zafr4 | .014 | .118 | .102 | .102 | 1.356 | 1.266 | 1.284 | 1.345 | 1.345 | 1.090 | .470 | .526 | .396 | .419 | .396 | .396 | .396 | .470 | .470 | .470 | .501 | .540 | .423 | .463 | .404 | .452 | .467 | .447 | .426 | .378 |
| 33. Zafr5 | - | .102 | .118 | .087 | 1.369 | 1.277 | 1.296 | 1.358 | 1.358 | 1.158 | .448 | .501 | .375 | .398 | .375 | .375 | .375 | .448 | .448 | .448 | .477 | .516 | .401 | .441 | .429 | .429 | .445 | .474 | .452 | .403 |
| 34. Zgab1 | .031 | - | .014 | .014 | 1.515 | 1.497 | 1.428 | 1.502 | 1.502 | 1.266 | .449 | .503 | .422 | .446 | .422 | .422 | .422 | .449 | .449 | .449 | .479 | .517 | .403 | .492 | .592 | .592 | .446 | .475 | .454 | .427 |
| 35. Zgab2 | .031 | .009 | - | .028 | 1.525 | 1.414 | 1.438 | 1.513 | 1.513 | 1.204 | .475 | .532 | .448 | .473 | .448 | .448 | .448 | .475 | .475 | .475 | .506 | .547 | .427 | .520 | .565 | .625 | .473 | .452 | .431 | .405 |
| 36. Zgab3 | .040 | .022 | .022 | - | 1.430 | 1.331 | 1.351 | 1.419 | 1.419 | 1.204 | .426 | .479 | .400 | .424 | .400 | .400 | .400 | .426 | .426 | .426 | .455 | .492 | .381 | .468 | .565 | .565 | .424 | .452 | .431 | .405 |
| 37. DmelA1 | .086 | .093 | .103 | .119 | - | .054 | .013 | .097 | .013 | .143 | 1.056 | 1.123 | .995 | .995 | 1.046 | 1.101 | .995 | 1.056 | 1.056 | 1.056 | 1.129 | .955 | .984 | .947 | .988 | 1.039 | .948 | 1.009 | 1.021 | 1.067 |
| 38. DmelA2 | .083 | .091 | .101 | .116 | .018 | - | .069 | .143 | .068 | .193 | 1.006 | 1.068 | .949 | .949 | .997 | 1.048 | .949 | 1.006 | 1.006 | 1.006 | 1.194 | .992 | .938 | .892 | .977 | 1.027 | .904 | .961 | .972 | 1.015 |
| 39. DmelA3 | .098 | .106 | .116 | .131 | .036 | .027 | - | .083 | .000 | .129 | 1.128 | 1.204 | 1.061 | 1.061 | 1.117 | 1.178 | 1.061 | 1.128 | 1.128 | 1.128 | 1.210 | 1.084 | 1.049 | 1.072 | 1.013 | 1.067 | 1.008 | 1.076 | 1.090 | 1.140 |
| 40. DsimA1 | .074 | .081 | .091 | .101 | .041 | .036 | .055 | - | .083 | .083 | 1.178 | 1.259 | 1.106 | 1.106 | 1.166 | 1.230 | 1.106 | 1.178 | 1.178 | 1.178 | 1.266 | 1.046 | 1.094 | .985 | .979 | 1.030 | 1.050 | 1.123 | 1.137 | 1.191 |
| 41. DereA1 | .074 | .081 | .091 | .106 | .032 | .022 | .041 | .041 | - | .129 | 1.178 | 1.259 | 1.106 | 1.106 | 1.166 | 1.230 | 1.106 | 1.178 | 1.178 | 1.178 | 1.266 | 1.046 | 1.094 | 1.036 | .979 | 1.030 | 1.050 | 1.123 | 1.137 | 1.191 |
| 42. Ztub1 | .088 | .096 | .105 | .121 | .050 | .045 | .064 | .041 | .050 | - | 1.128 | 1.204 | 1.061 | 1.061 | 1.117 | 1.178 | 1.061 | 1.128 | 1.128 | 1.128 | 1.210 | 1.004 | 1.049 | 1.020 | .894 | .988 | 1.008 | 1.022 | 1.035 | 1.082 |
| 43. DmelA5 | .009 | .031 | .031 | .045 | .088 | .084 | .098 | .074 | .074 | .089 | - | .056 | .055 | .055 | .041 | .055 | .055 | .055 | .055 | .055 | .148 | .281 | .070 | .230 | .334 | .334 | .115 | .131 | .101 | .100 |
| 44. DmelA6 | .009 | .031 | .031 | .045 | .088 | .084 | .098 | .074 | .074 | .088 | .000 | - | .085 | .085 | .070 | .085 | .085 | .085 | .085 | .085 | .182 | .323 | .100 | .268 | .336 | .336 | .147 | .164 | .133 | .132 |
| 45. DyakA1 | .013 | .036 | .036 | .050 | .088 | .079 | .094 | .074 | .069 | .089 | .004 | .004 | - | .027 | .013 | .027 | .000 | .055 | .055 | .055 | .115 | .242 | .041 | .160 | .253 | .253 | .084 | .099 | .070 | .070 |
| 46. DyakA2 | .013 | .036 | .036 | .050 | .088 | .079 | .094 | .074 | .069 | .089 | .004 | .004 | .000 | - | .013 | .027 | .027 | .084 | .084 | .084 | .147 | .242 | .070 | .160 | .253 | .253 | .114 | .130 | .100 | .100 |
| 47. DyakA3 | .013 | .036 | .036 | .050 | .088 | .079 | .094 | .074 | .069 | .089 | .004 | .004 | .000 | .000 | - | .013 | .013 | .069 | .069 | .069 | .131 | .261 | .055 | .177 | .272 | .272 | .099 | .115 | .085 | .085 |
| 48. DyakA4 | .013 | .036 | .036 | .050 | .088 | .079 | .094 | .074 | .069 | .089 | .004 | .004 | .000 | .000 | .000 | - | .027 | .084 | .084 | .084 | .147 | .280 | .070 | .194 | .291 | .291 | .114 | .130 | .100 | .100 |
| 49. DyakA5 | .013 | .036 | .036 | .050 | .088 | .079 | .094 | .074 | .069 | .089 | .004 | .004 | .000 | .000 | .000 | .000 | - | .055 | .055 | .055 | .115 | .242 | .041 | .160 | .253 | .253 | .084 | .099 | .070 | .070 |
| 50. DereA3 | .009 | .031 | .031 | .045 | .088 | .084 | .098 | .074 | .074 | .089 | .000 | .000 | .004 | .004 | .004 | .004 | .004 | - | .000 | .000 | .132 | .281 | .055 | .230 | .334 | .334 | .099 | .115 | .070 | .085 |
| 51. DereA4 | .013 | .036 | .036 | .040 | .093 | .089 | .103 | .074 | .079 | .093 | .004 | .004 | .009 | .009 | .009 | .009 | .009 | .004 | - | .000 | .132 | .281 | .055 | .230 | .334 | .334 | .099 | .115 | .070 | .085 |
| 52. DereA5 | .009 | .031 | .031 | .045 | .088 | .084 | .098 | .074 | .074 | .089 | .000 | .000 | .004 | .004 | .004 | .004 | .004 | .000 | .004 | - | .132 | .281 | .055 | .230 | .334 | .334 | .099 | .115 | .070 | .085 |
| 53. DereA8 | .036 | .059 | .059 | .073 | .097 | .088 | .103 | .079 | .079 | .098 | .027 | .027 | .022 | .022 | .022 | .022 | .022 | .027 | .031 | .027 | - | .204 | .070 | .307 | .427 | .427 | .115 | .132 | .117 | .101 |
| 54. DereB1 | .043 | .066 | .066 | .080 | .109 | .101 | .113 | .096 | .101 | .111 | .043 | .043 | .038 | .038 | .038 | .038 | .038 | .043 | .047 | .043 | .053 | - | .258 | .122 | .246 | .210 | .280 | .297 | .260 | .298 |
| 55. DereB8 | .031 | .054 | .054 | .068 | .110 | .101 | .116 | .096 | .091 | .111 | .022 | .022 | .018 | .018 | .018 | .018 | .018 | .022 | .027 | .022 | .031 | .058 | - | .212 | .314 | .314 | .041 | .056 | .042 | .027 |
| 56. DereB10 | .045 | .069 | .069 | .083 | .122 | .114 | .127 | .109 | .104 | .122 | .036 | .036 | .032 | .032 | .032 | .032 | .032 | .036 | .041 | .036 | .055 | .052 | .050 | - | .147 | .115 | .228 | .248 | .214 | .249 |
| 57. Ztub5 | .031 | .054 | .054 | .068 | .101 | .093 | .106 | .088 | .083 | .103 | .031 | .031 | .027 | .027 | .027 | .027 | .027 | .031 | .036 | .031 | .050 | .038 | .045 | .041 | - | .028 | .332 | .314 | .276 | .315 |
| 58. Ztub6 | .031 | .054 | .054 | .068 | .101 | .093 | .106 | .088 | .083 | .103 | .031 | .031 | .027 | .027 | .027 | .027 | .027 | .031 | .036 | .031 | .050 | .038 | .045 | .041 | .000 | - | .332 | .356 | .316 | .357 |
| 59. Zcam3 | .031 | .054 | .054 | .069 | .108 | .099 | .114 | .094 | .089 | .109 | .022 | .022 | .018 | .018 | .018 | .018 | .018 | .022 | .027 | .022 | .041 | .057 | .036 | .050 | .045 | .045 | - | .070 | .056 | .070 |
| 60. Zcam5 | .022 | .045 | .045 | .059 | .098 | .089 | .103 | .084 | .079 | .098 | .013 | .013 | .009 | .009 | .009 | .009 | .009 | .013 | .018 | .013 | .022 | .049 | .018 | .041 | .036 | .036 | .027 | - | .042 | .056 |
| 61. Zcam7 | .040 | .064 | .064 | .078 | .118 | .108 | .124 | .103 | .098 | .118 | .036 | .036 | .031 | .031 | .031 | .031 | .031 | .036 | .040 | .036 | .045 | .067 | .040 | .064 | .054 | .054 | .050 | .031 | - | .042 |
| 62. Zdav1 | .022 | .045 | .045 | .059 | .097 | .089 | .103 | .084 | .079 | .098 | .013 | .013 | .009 | .009 | .009 | .009 | .009 | .013 | .018 | .013 | .022 | .049 | .018 | .041 | .036 | .036 | .027 | .009 | .031 | - |

Symbols for species names: Ztub1: *Z. tuberculatus*; Zcam: *Z. camerounensis*; Zdav: *Z. davidi*; Zgab: *Z. gabonicus*; Zafr: *Z. africanus*; Zind: *Z. indianus*; Dmel: *D. melanogaster*; Dsim: *D. simulans*; Dsec: *D. sechellia*; Dyak: *D. yakuba*; Dere: *D. erecta*.

**Supplementary Table 4.** Z values (above) and significance P-values (below) for the Z-test of selection between *gypsy* sequences of *Zaprionus* and *melanogaster* species. Test performed using alternative hypothesis of non-neutrality (dN ≠ dS) and Nei-Gojobori distance (Jukes-Cantor's correction). Gray cells correspond to significant ($p < 0.05$) pairwise comparisons.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. DmelA7 | - | -.998 | -2.775 | -1.821 | -1.638 | -2.872 | -.959 | -.479 | -1.760 | -2.676 | -2.471 | -2.536 | -2.425 | -2.138 | -2.055 | -1.908 | -2.140 | -1.481 | -2.055 | -.765 | -1.905 | -.426 | .135 | -.765 | -.064 | -1.520 | -.489 | -2.003 | -1.407 | -1.756 | -2.009 | -1.467 |
| 2. DmelB5 | .320 | - | -2.607 | -1.550 | -1.322 | -2.711 | -.491 | -.479 | -1.476 | -2.501 | -2.274 | -2.346 | -2.226 | -1.905 | -1.808 | -1.643 | -1.908 | -1.142 | -1.808 | -.267 | -1.638 | .121 | .945 | -.267 | .718 | -1.162 | .214 | -1.756 | -.998 | -1.471 | -1.763 | -1.126 |
| 3. DsimA2 | .006 | .010 | - | -1.690 | -2.425 | -2.007 | -1.995 | -2.399 | -2.109 | -1.645 | -1.908 | -2.005 | -1.320 | -1.476 | -2.349 | -2.231 | -2.428 | -1.626 | -2.349 | -2.293 | -2.226 | -2.094 | -1.866 | -2.293 | -1.978 | -2.841 | -2.392 | -3.105 | -2.775 | -2.652 | -3.112 | -2.474 |
| 4. DsimA3 | .071 | .124 | .094 | - | -1.232 | -1.822 | -.473 | -1.253 | -1.406 | -1.555 | -1.704 | -1.819 | -1.070 | -1.232 | -2.185 | -1.557 | -1.238 | .923 | -2.185 | -1.252 | -2.059 | -.970 | -.572 | -1.252 | -.735 | -1.939 | -1.296 | -2.341 | -1.821 | -1.684 | -2.351 | -1.919 |
| 5. DsimA4 | .104 | .189 | .017 | .220 | - | -2.537 | .211 | -.948 | -1.137 | -2.312 | -2.471 | -2.536 | -2.007 | -2.138 | -2.055 | -1.327 | -1.641 | -1.142 | -2.055 | -.765 | -1.905 | -.426 | .135 | -.765 | -.064 | -1.760 | -1.003 | -2.194 | -1.638 | -1.463 | -2.203 | -1.731 |
| 6. DsimB2 | .005 | .008 | .047 | .071 | .012 | - | -2.110 | -2.640 | -2.744 | -1.875 | -1.519 | -2.137 | -.702 | -1.639 | -2.055 | -2.350 | -2.539 | -1.765 | -2.055 | -2.537 | -1.905 | -2.340 | -2.133 | -2.537 | -2.243 | -2.931 | -2.484 | -3.044 | -2.872 | -2.742 | -3.186 | -2.570 |
| 7. DsimB3 | .340 | .624 | .048 | .637 | .833 | .037 | - | -.087 | -.293 | -1.877 | -2.011 | -2.107 | -1.485 | -1.620 | -1.481 | -.502 | -.966 | -.302 | -1.481 | .109 | -1.320 | .465 | 1.335 | .109 | 1.131 | -1.170 | -.286 | -1.751 | -.959 | -.852 | -1.765 | -1.189 |
| 8. DsimB4 | .633 | .633 | .018 | .213 | .345 | .009 | .931 | - | -1.479 | -2.295 | -2.423 | -2.498 | -1.986 | -2.102 | -2.005 | -1.319 | -1.617 | -1.166 | -2.005 | -.272 | -1.872 | .115 | .939 | -.272 | .713 | -1.158 | -.267 | -1.478 | -.948 | -1.478 | -1.493 | -1.174 |
| 9. DsimB6 | .081 | .142 | .037 | .162 | .258 | .007 | .770 | .142 | - | -2.401 | -2.538 | -2.606 | -2.109 | -2.226 | -2.140 | -1.483 | -1.764 | -1.327 | -2.140 | -1.643 | -2.007 | -1.322 | -.953 | -1.643 | -1.139 | -1.879 | -1.203 | -2.087 | -1.760 | -1.612 | -2.098 | -1.857 |
| 10. DsimB7 | .008 | .014 | .103 | .123 | .022 | .063 | .063 | .023 | .018 | - | -1.760 | -1.872 | -1.162 | -1.314 | -2.226 | -2.111 | -2.052 | -1.485 | -2.226 | -2.189 | -2.104 | -1.991 | -1.755 | -2.189 | -1.867 | -2.749 | -2.295 | -3.026 | -2.676 | -2.560 | -3.034 | -2.375 |
| 11. DsecA1 | .015 | .025 | .059 | .091 | .015 | .131 | .047 | .017 | .012 | .081 | - | -1.518 | -.340 | -.651 | -1.407 | -2.276 | -2.472 | -1.330 | -1.407 | -2.309 | -1.160 | -2.089 | -1.857 | -2.309 | -1.979 | -2.538 | -1.987 | -2.835 | -2.471 | -2.307 | -2.840 | -2.095 |
| 12. DsecA2 | .013 | .021 | .047 | .071 | .013 | .035 | .037 | .014 | .010 | .064 | .132 | - | -1.809 | -.921 | -2.055 | -2.349 | -2.538 | -1.480 | -2.055 | -2.391 | -1.904 | -2.187 | -1.967 | -2.391 | -2.081 | -2.606 | -2.089 | -2.896 | -2.536 | -2.389 | -2.902 | -2.187 |
| 13. DsecA3 | .017 | .028 | .189 | .287 | .047 | .484 | .140 | .049 | .037 | .248 | .734 | .073 | - | -.701 | -1.326 | -1.766 | -2.011 | -.966 | -1.326 | -1.865 | -1.137 | -1.636 | -1.356 | -1.865 | -1.486 | -2.505 | -1.988 | -2.814 | -2.425 | -2.291 | -2.821 | -2.082 |
| 14. DsecA4 | .035 | .059 | .142 | .220 | .035 | .104 | .108 | .038 | .028 | .191 | .516 | .359 | .485 | - | -1.518 | -1.908 | -2.140 | -.707 | -1.518 | -1.979 | -1.322 | -1.985 | -1.475 | -1.979 | -1.607 | -2.226 | -1.617 | -2.573 | -2.138 | -1.977 | -2.580 | -1.731 |
| 15. DsecA5 | .042 | .073 | .020 | .031 | .042 | .042 | .141 | .047 | .034 | .028 | .162 | .042 | .187 | .132 | - | -1.810 | -2.472 | -1.910 | .000 | -1.871 | .999 | -1.615 | -1.321 | -1.871 | -1.466 | -2.140 | -1.475 | -2.497 | -2.055 | -1.869 | -2.504 | -1.604 |
| 16. DsecA7 | .059 | .103 | .028 | .122 | .187 | .020 | .616 | .190 | .141 | .037 | .025 | .020 | .080 | .059 | .073 | - | -1.911 | -1.488 | -1.810 | -1.159 | -1.643 | -.859 | -.427 | -1.159 | -.601 | -2.012 | -1.345 | -2.398 | -1.908 | -1.745 | -2.406 | -1.467 |
| 17. DsecA8 | .034 | .059 | .017 | .218 | .103 | .012 | .336 | .108 | .080 | .042 | .015 | .012 | .047 | .034 | .015 | .058 | - | -1.148 | -2.472 | -1.473 | -2.349 | -1.200 | -.844 | -1.473 | -1.000 | -2.230 | -1.625 | -2.577 | -2.140 | -1.983 | -2.585 | -2.195 |
| 18. DsecA9 | .141 | .256 | .107 | .358 | .256 | .080 | .763 | .246 | .187 | .140 | .186 | .142 | .336 | .481 | .058 | .139 | .253 | - | -1.910 | -1.009 | -1.764 | -.716 | -.267 | -1.009 | -.437 | -1.626 | -.869 | -2.094 | -1.481 | -1.336 | -2.105 | -1.613 |
| 19. DsecA10 | .042 | .073 | .020 | .031 | .042 | .042 | .141 | .047 | .034 | .028 | .162 | .042 | .187 | .132 | 1.000 | .073 | .015 | .058 | - | -1.871 | .999 | -1.615 | -1.321 | -1.871 | -1.466 | -2.140 | -1.475 | -2.497 | -2.055 | -1.869 | -2.504 | -1.604 |
| 20. DsecA11 | .446 | .790 | .024 | .213 | .446 | .012 | .914 | .786 | .103 | .031 | .023 | .018 | .065 | .050 | .064 | .249 | .143 | .315 | .064 | - | -1.738 | -.489 | -.337 | 1.412 | -.654 | -1.000 | -.101 | -1.348 | -.765 | -1.348 | -1.365 | -1.032 |
| 21. DsecB11 | .059 | .104 | .028 | .042 | .059 | .059 | .189 | .064 | .047 | .037 | .248 | .059 | .258 | .189 | .320 | .103 | .020 | .080 | .320 | .085 | - | -1.486 | -1.179 | -1.738 | -1.322 | -2.007 | -1.334 | -2.391 | -1.905 | -1.736 | -2.400 | -1.457 |
| 22. DsecB12 | .671 | .904 | .038 | .334 | .671 | .021 | .643 | .909 | .189 | .049 | .039 | .031 | .104 | .049 | .109 | .392 | .233 | .475 | .109 | .626 | .140 | - | .216 | -.489 | -.055 | -.705 | .202 | -1.105 | -.426 | -1.105 | -1.125 | -.767 |
| 23. DsecB17 | .893 | .346 | .065 | .568 | .893 | .035 | .184 | .349 | .343 | .082 | .066 | .052 | .178 | .143 | .189 | .670 | .401 | .790 | .189 | .736 | .241 | .830 | - | .225 | .999 | -.255 | .836 | -.763 | .135 | -.763 | -.785 | -.368 |
| 24. DsecB18 | .446 | .790 | .024 | .213 | .446 | .012 | .914 | .786 | .103 | .031 | .023 | .018 | .065 | .050 | .064 | .249 | .143 | .315 | .064 | .160 | .085 | .626 | .822 | - | -.047 | -1.000 | -.101 | -1.348 | -.765 | -1.348 | -1.564 | -1.032 |
| 25. DsecB19 | .949 | .474 | .050 | .464 | .949 | .027 | .260 | .477 | .257 | .064 | .050 | .040 | .140 | .111 | .145 | .549 | .319 | .663 | .145 | .515 | .189 | .956 | .320 | .962 | - | -.426 | .661 | -.902 | -.064 | -.902 | -.922 | -.520 |
| 26. Zdav2 | .131 | .247 | .005 | .055 | .081 | .004 | .244 | .249 | .063 | .007 | .012 | .010 | .014 | .028 | .034 | .046 | .028 | .107 | .034 | .319 | .047 | .482 | .799 | .319 | .671 | - | -.063 | -2.104 | -1.520 | -1.314 | -1.332 | -.946 |
| 27. Zdav4 | .626 | .831 | .018 | .197 | .318 | .014 | .775 | .790 | .231 | .023 | .049 | .039 | .049 | .108 | .143 | .181 | .107 | .386 | .143 | .920 | .185 | .840 | .405 | .920 | .510 | .950 | - | -1.489 | -.489 | -1.194 | -1.172 | -.840 |
| 28. Zind2 | .047 | .082 | .002 | .021 | .030 | .003 | .082 | .142 | .039 | .003 | .005 | .004 | .006 | .011 | .014 | .018 | .011 | .038 | .014 | .180 | .018 | .271 | .447 | .180 | .369 | .037 | .139 | - | -2.003 | -1.855 | -1.617 | -2.075 |
| 29. Zind3 | .162 | .320 | .006 | .071 | .104 | .005 | .340 | .345 | .081 | .008 | .015 | .013 | .017 | .035 | .042 | .059 | .034 | .141 | .042 | .446 | .059 | .671 | .893 | .446 | .949 | .131 | .626 | .047 | - | -1.471 | -1.480 | -.689 |
| 30. Zind4 | .082 | .144 | .009 | .095 | .146 | .007 | .396 | .142 | .110 | .012 | .023 | .018 | .024 | .050 | .064 | .084 | .050 | .184 | .064 | .180 | .085 | .271 | .447 | .180 | .369 | .191 | .235 | .066 | .144 | - | -1.003 | -.980 |
| 31. Zafr2 | .047 | .080 | .002 | .020 | .030 | .002 | .080 | .138 | .038 | .003 | .005 | .004 | .006 | .011 | .014 | .018 | .011 | .037 | .014 | .175 | .018 | .263 | .434 | .120 | .358 | .185 | .243 | .108 | .142 | .318 | - | -.995 |
| 32. Zafr4 | .145 | .262 | .015 | .057 | .086 | .011 | .237 | .243 | .066 | .019 | .038 | .031 | .039 | .086 | .111 | .145 | .030 | .109 | .111 | .304 | .148 | .444 | .713 | .304 | .604 | .346 | .403 | .040 | .492 | .329 | .322 | - |
| 33. Zafr5 | .189 | .359 | .018 | .073 | .111 | .014 | .318 | .325 | .084 | .024 | .048 | .038 | .050 | .111 | .144 | .191 | .038 | .143 | .144 | .407 | .194 | .591 | .943 | .407 | .810 | .143 | .551 | .051 | .740 | .147 | .143 | .974 |
| 34. Zgab1 | .085 | .146 | .024 | .092 | .143 | .007 | .358 | .136 | .107 | .031 | .024 | .019 | .024 | .051 | .066 | .083 | .050 | .252 | .066 | .170 | .085 | .249 | .403 | .170 | .336 | .065 | .219 | .066 | .085 | .302 | .064 | .067 |
| 35. Zgab2 | .050 | .085 | .015 | .056 | .085 | .005 | .217 | .082 | .064 | .019 | .015 | .012 | .015 | .031 | .039 | .050 | .031 | .151 | .039 | .104 | .051 | .155 | .254 | .104 | .208 | .109 | .132 | .040 | .050 | .497 | .105 | .110 |
| 36. Zgab3 | .300 | .497 | .078 | .282 | .446 | .025 | .903 | .398 | .329 | .097 | .083 | .065 | .078 | .170 | .225 | .270 | .167 | .689 | .225 | .471 | .277 | .621 | .916 | .471 | .812 | .221 | .603 | .080 | .300 | .336 | .077 | .213 |
| 37. DmelA1 | .001 | 0 | 0 | .001 | .001 | 0 | .001 | .001 | .001 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | .001 | .002 | .001 | .001 | .001 | 0 |
| 38. DmelA2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | 0 | 0 | 0 | 0 |
| 39. DmelA3 | 0 | 0 | 0 | 0 | .001 | 0 | .001 | .001 | .001 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | 0 | .001 | .001 | 0 |
| 40. DsimA1 | 0 | 0 | 0 | .001 | .001 | 0 | .001 | 0 | .001 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | .002 | 0 | .001 | .001 | 0 |
| 41. DereA1 | 0 | 0 | 0 | .001 | .001 | 0 | .001 | .001 | .001 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | .002 | 0 | .001 | .001 | 0 |
| 42. Ztub1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 43. DmelA5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 44. DmelA6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 45. DyakA1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 46. DyakA2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 47. DyakA3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 48. DyakA4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 49. DyakA5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 50. DereA3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 51. DereA4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 52. DereA5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 53. DereA8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 54. DereB1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 55. DereB8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | .001 | 0 | .001 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 56. DereB10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 57. Ztub5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 58. Ztub6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 59. Zcam3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 60. Zcam5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 61. Zcam7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | 0 | .001 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 62. Zdav1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Supplementary Table 4, continuation.**

| | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. DmelA7 | -1.321 | -1.736 | -1.979 | -1.041 | -3.563 | -3.745 | -3.654 | -3.764 | -3.603 | -3.931 | -4.330 | -4.444 | -4.212 | -4.281 | -4.212 | -4.212 | -4.212 | -4.330 | -4.279 | -4.330 | -4.097 | -4.128 | -3.912 | -3.915 | -4.271 | -4.271 | -4.074 | -4.244 | -3.971 | -4.024 |
| 2. DmelB5 | -.921 | -1.463 | -1.738 | -.682 | -3.723 | -3.876 | -3.787 | -3.764 | -3.764 | -4.024 | -4.263 | -4.388 | -4.137 | -4.212 | -4.137 | -4.137 | -4.137 | -4.263 | -4.210 | -4.263 | -4.022 | -4.059 | -3.822 | -3.995 | -4.211 | -4.211 | -3.995 | -4.176 | -3.889 | -3.939 |
| 3. DsimA2 | -2.395 | -2.283 | -2.469 | -1.775 | -3.967 | -4.073 | -3.894 | -4.012 | -4.012 | -4.109 | -4.537 | -4.445 | -4.458 | -4.501 | -4.458 | -4.458 | -4.458 | -4.537 | -4.497 | -4.537 | -4.335 | -4.057 | -4.227 | -4.215 | -4.195 | -4.195 | -4.338 | -4.455 | -4.208 | -4.366 |
| 4. DsimA3 | -1.806 | -1.696 | -1.931 | -1.081 | -3.493 | -3.730 | -3.581 | -3.533 | -3.533 | -3.850 | -4.309 | -4.182 | -4.203 | -4.266 | -4.203 | -4.203 | -4.203 | -4.309 | -4.264 | -4.309 | -4.100 | -4.125 | -3.935 | -4.077 | -4.143 | -4.143 | -4.082 | -4.232 | -3.990 | -4.106 |
| 5. DsimA4 | -1.607 | -1.474 | -1.737 | -.765 | -3.350 | -3.636 | -3.469 | -3.389 | -3.389 | -3.787 | -4.294 | -4.292 | -4.178 | -4.247 | -4.178 | -4.178 | -4.178 | -4.294 | -4.245 | -4.294 | -4.071 | -4.102 | -3.888 | -4.092 | -4.240 | -4.240 | -4.047 | -4.213 | -3.949 | -4.074 |
| 6. DsimB2 | -2.497 | -2.721 | -2.878 | -2.277 | -3.817 | -3.979 | -3.855 | -3.859 | -3.859 | -4.053 | -4.503 | -4.406 | -4.420 | -4.466 | -4.420 | -4.420 | -4.420 | -4.503 | -4.462 | -4.503 | -4.341 | -4.264 | -4.183 | -4.276 | -4.239 | -4.239 | -4.299 | -4.421 | -4.210 | -4.328 |
| 7. DsimB3 | -1.003 | -.922 | -1.242 | -.123 | -3.483 | -3.697 | -3.548 | -3.498 | -3.498 | -3.819 | -4.123 | -4.122 | -3.996 | -4.073 | -3.996 | -3.996 | -3.996 | -4.123 | -4.073 | -4.123 | -3.902 | -3.944 | -3.692 | -3.878 | -4.178 | -4.178 | -3.868 | -4.047 | -3.770 | -3.891 |
| 8. DsimB4 | -.989 | -1.500 | -1.752 | -.849 | -3.478 | -3.718 | -3.568 | -3.811 | -3.518 | -3.841 | -4.125 | -4.124 | -3.996 | -4.074 | -3.996 | -3.996 | -3.996 | -4.125 | -4.074 | -4.125 | -3.899 | -3.942 | -3.687 | -3.790 | -4.091 | -4.091 | -3.865 | -4.047 | -3.765 | -3.801 |
| 9. DsimB6 | -1.740 | -1.624 | -1.867 | -.980 | -3.259 | -3.561 | -3.388 | -3.298 | -3.298 | -3.721 | -4.243 | -4.241 | -4.125 | -4.196 | -4.125 | -4.125 | -4.125 | -4.243 | -4.194 | -4.243 | -3.950 | -4.057 | -3.835 | -3.999 | -4.195 | -4.195 | -3.997 | -4.164 | -3.901 | -4.022 |
| 1. DsimB7 | -2.289 | -2.185 | -2.377 | -1.673 | -3.948 | -4.051 | -3.946 | -3.993 | -3.993 | -4.086 | -4.405 | -4.289 | -4.307 | -4.364 | -4.307 | -4.307 | -4.307 | -4.405 | -4.361 | -4.405 | -4.196 | -4.012 | -4.047 | -4.215 | -4.152 | -4.152 | -4.184 | -4.326 | -4.046 | -4.210 |
| 11. DsecA1 | -2.001 | -2.285 | -2.475 | -1.747 | -3.951 | -4.082 | -3.968 | -3.995 | -3.995 | -4.136 | -4.401 | -4.398 | -4.296 | -4.358 | -4.296 | -4.296 | -4.296 | -4.401 | -4.355 | -4.401 | -4.241 | -4.146 | -4.022 | -4.158 | -4.096 | -4.096 | -4.167 | -4.319 | -4.071 | -4.195 |
| 12. DsecA2 | -2.096 | -2.372 | -2.553 | -1.865 | -3.905 | -4.047 | -3.929 | -3.949 | -3.949 | -4.108 | -4.462 | -4.458 | -4.369 | -4.422 | -4.369 | -4.369 | -4.369 | -4.462 | -4.419 | -4.462 | -4.253 | -4.163 | -4.119 | -4.229 | -4.182 | -4.182 | -4.246 | -4.381 | -4.155 | -4.274 |
| 13. DsecA3 | -1.981 | -2.283 | -2.469 | -1.775 | -3.701 | -3.890 | -3.758 | -3.743 | -3.743 | -3.981 | -4.361 | -4.241 | -4.260 | -4.319 | -4.260 | -4.260 | -4.260 | -4.361 | -4.317 | -4.361 | -4.290 | -4.093 | -4.096 | -4.129 | -4.070 | -4.070 | -4.138 | -4.368 | -4.136 | -4.256 |
| 14. DsecA4 | -1.607 | -1.968 | -2.182 | -1.380 | -4.094 | -4.168 | -4.070 | -4.141 | -4.141 | -4.184 | -4.410 | -4.407 | -4.310 | -4.368 | -4.310 | -4.310 | -4.310 | -4.410 | -4.365 | -4.410 | -4.198 | -4.102 | -4.047 | -4.173 | -4.118 | -4.118 | -4.185 | -4.330 | -4.092 | -4.212 |
| 15. DsecA5 | -1.470 | -1.855 | -2.083 | -1.220 | -3.841 | -4.004 | -3.880 | -3.883 | -3.883 | -4.079 | -4.279 | -4.398 | -4.158 | -4.230 | -4.158 | -4.158 | -4.158 | -4.279 | -4.228 | -4.279 | -4.118 | -4.011 | -3.855 | -4.021 | -4.096 | -4.096 | -4.022 | -4.196 | -3.920 | -4.049 |
| 16. DsecA7 | -1.314 | -1.746 | -1.981 | -1.109 | -3.445 | -3.704 | -3.547 | -3.485 | -3.485 | -3.839 | -4.349 | -4.347 | -4.244 | -4.306 | -4.244 | -4.244 | -4.244 | -4.349 | -4.304 | -4.349 | -4.136 | -4.160 | -3.971 | -4.112 | -4.179 | -4.179 | -4.118 | -4.270 | -4.024 | -4.144 |
| 17. DsecA8 | -2.103 | -1.976 | -2.189 | -1.391 | -3.636 | -3.845 | -3.706 | -3.676 | -3.676 | -3.736 | -4.405 | -4.402 | -4.307 | -4.364 | -4.307 | -4.307 | -4.307 | -4.405 | -4.361 | -4.405 | -4.196 | -4.214 | -4.047 | -4.121 | -4.332 | -4.332 | -4.184 | -4.326 | -4.091 | -4.210 |
| 18. DsecA9 | -1.475 | -1.151 | -1.445 | -.401 | -3.763 | -3.925 | -3.802 | -3.805 | -3.805 | -3.997 | -4.240 | -4.172 | -4.124 | -4.194 | -4.124 | -4.124 | -4.124 | -4.240 | -4.192 | -4.240 | -4.024 | -4.057 | -3.838 | -4.000 | -4.070 | -4.070 | -3.998 | -4.163 | -3.903 | -4.022 |
| 19. DsecA10 | -1.470 | -1.855 | -2.083 | -1.220 | -3.841 | -4.004 | -3.880 | -3.883 | -3.883 | -4.079 | -4.279 | -4.398 | -4.158 | -4.230 | -4.158 | -4.158 | -4.158 | -4.279 | -4.228 | -4.279 | -4.118 | -4.011 | -3.855 | -4.021 | -4.096 | -4.096 | -4.022 | -4.196 | -3.920 | -4.049 |
| 20. DsecA11 | -.831 | -1.380 | -1.640 | -.723 | -3.761 | -3.917 | -3.795 | -3.803 | -3.803 | -4.007 | -4.074 | -4.073 | -3.941 | -4.022 | -3.941 | -3.941 | -3.941 | -4.074 | -4.022 | -4.074 | -3.849 | -3.894 | -3.629 | -3.645 | -4.046 | -4.046 | -3.811 | -3.998 | -3.713 | -3.744 |
| 21. DsecB11 | -1.305 | -1.734 | -1.970 | -1.093 | -3.825 | -3.985 | -3.862 | -3.868 | -3.868 | -4.058 | -4.228 | -4.352 | -4.104 | -4.178 | -4.104 | -4.104 | -4.104 | -4.228 | -4.177 | -4.228 | -4.071 | -3.962 | -3.799 | -3.972 | -4.047 | -4.047 | -3.969 | -4.146 | -3.868 | -3.996 |
| 22. DsecB12 | -.539 | -1.159 | -1.433 | -.496 | -3.790 | -3.843 | -3.807 | -3.834 | -3.834 | -3.969 | -3.971 | -3.972 | -3.834 | -3.919 | -3.834 | -3.834 | -3.834 | -3.971 | -3.920 | -3.971 | -3.753 | -3.803 | -3.520 | -3.727 | -3.804 | -3.804 | -3.709 | -3.900 | -3.613 | -3.635 |
| 23. DsecB17 | -.072 | -.840 | -1.147 | -.106 | -3.611 | -3.797 | -3.666 | -3.653 | -3.653 | -3.904 | -3.887 | -3.888 | -3.739 | -3.831 | -3.739 | -3.739 | -3.739 | -3.887 | -3.833 | -3.887 | -3.658 | -3.715 | -3.406 | -3.631 | -3.881 | -3.881 | -3.609 | -3.814 | -3.508 | -3.528 |
| 24. DsecB18 | -.831 | -1.380 | -1.640 | -.723 | -3.761 | -3.917 | -3.795 | -3.803 | -3.803 | -3.986 | -4.074 | -4.073 | -3.941 | -4.022 | -3.941 | -3.941 | -3.941 | -4.074 | -4.022 | -4.074 | -3.849 | -3.894 | -3.629 | -3.645 | -4.046 | -4.046 | -3.811 | -3.998 | -3.713 | -3.744 |
| 25. DsecB19 | -.240 | -.967 | -1.266 | -.238 | -3.617 | -3.808 | -3.675 | -3.659 | -3.659 | -3.918 | -3.942 | -3.942 | -3.797 | -3.886 | -3.797 | -3.797 | -3.797 | -3.942 | -3.888 | -3.942 | -3.712 | -3.766 | -3.467 | -3.685 | -3.929 | -3.929 | -3.665 | -3.867 | -3.564 | -3.588 |
| 26. Zdav2 | -1.475 | -1.863 | -1.615 | -1.231 | -3.562 | -3.737 | -3.647 | -3.603 | -3.603 | -4.000 | -4.284 | -4.398 | -4.166 | -4.236 | -4.166 | -4.166 | -4.166 | -4.284 | -4.234 | -4.284 | -4.059 | -4.092 | -3.873 | -4.034 | -4.106 | -4.231 | -4.035 | -4.061 | -3.766 | -3.807 |
| 27. Zdav4 | -.598 | -1.235 | -1.516 | -.521 | -3.506 | -3.686 | -3.595 | -3.546 | -3.520 | -3.867 | -4.061 | -4.200 | -3.927 | -4.009 | -3.927 | -3.927 | -3.927 | -4.061 | -4.009 | -4.061 | -3.936 | -3.977 | -3.614 | -3.811 | -4.035 | -4.035 | -3.798 | -3.985 | -3.699 | -3.729 |
| 28. Zind2 | -1.975 | -1.856 | -2.077 | -1.764 | -3.117 | -3.368 | -3.270 | -3.156 | -3.156 | -3.640 | -4.308 | -4.414 | -4.197 | -4.263 | -4.197 | -4.197 | -4.197 | -4.308 | -4.261 | -4.308 | -4.092 | -4.120 | -3.918 | -4.066 | -4.251 | -4.251 | -4.071 | -4.229 | -3.976 | -4.022 |
| 29. Zind3 | -.333 | -1.736 | -1.979 | -1.041 | -3.563 | -3.745 | -3.654 | -3.603 | -3.603 | -3.931 | -4.330 | -4.444 | -4.057 | -4.137 | -4.057 | -4.057 | -4.057 | -4.330 | -4.279 | -4.330 | -4.097 | -4.128 | -3.912 | -3.915 | -4.145 | -4.145 | -4.074 | -4.244 | -3.971 | -4.024 |
| 30. Zind4 | -1.461 | -1.037 | -.682 | -.967 | -3.429 | -3.622 | -3.530 | -3.469 | -3.469 | -3.917 | -4.253 | -4.367 | -4.062 | -4.137 | -4.062 | -4.062 | -4.062 | -4.253 | -4.205 | -4.253 | -4.035 | -4.067 | -3.850 | -3.936 | -4.011 | -4.146 | -4.010 | -4.035 | -3.745 | -3.872 |
| 31. Zafr2 | -1.473 | -1.870 | -1.633 | -1.783 | -3.397 | -3.591 | -3.500 | -3.437 | -3.437 | -3.887 | -4.249 | -4.359 | -3.982 | -4.061 | -3.982 | -3.982 | -3.982 | -4.249 | -4.202 | -4.249 | -4.035 | -4.067 | -3.853 | -3.861 | -3.937 | -4.080 | -4.011 | -4.035 | -3.751 | -3.785 |
| 32. Zafr4 | -.033 | -1.848 | -1.610 | -1.252 | -3.725 | -3.866 | -3.778 | -3.767 | -3.767 | -4.063 | -4.405 | -4.503 | -4.159 | -4.232 | -4.159 | -4.159 | -4.159 | -4.405 | -4.358 | -4.405 | -4.184 | -4.207 | -4.022 | -4.020 | -3.942 | -4.097 | -4.265 | -4.197 | -3.918 | -3.968 |
| 33. Zafr5 | - | -1.732 | -1.968 | -1.216 | -3.713 | -3.859 | -3.771 | -3.754 | -3.754 | -3.999 | -4.393 | -4.497 | -4.138 | -4.213 | -4.138 | -4.138 | -4.138 | -4.393 | -4.344 | -4.393 | -4.165 | -4.191 | -3.996 | -3.995 | -4.073 | -4.073 | -4.148 | -4.308 | -4.047 | -4.103 |
| 34. Zgab1 | .086 | - | -.337 | .480 | -3.429 | -3.456 | -3.530 | -3.469 | -3.469 | -3.820 | -4.147 | -4.273 | -4.022 | -4.097 | -4.022 | -4.022 | -4.022 | -4.147 | -4.097 | -4.147 | -3.927 | -3.966 | -3.724 | -3.901 | -4.220 | -4.220 | -3.895 | -4.071 | -3.799 | -3.919 |
| 35. Zgab2 | .051 | .737 | - | -.270 | -3.381 | -3.574 | -3.483 | -3.421 | -3.421 | -3.867 | -4.213 | -4.327 | -4.097 | -4.167 | -4.097 | -4.097 | -4.097 | -4.213 | -4.166 | -4.213 | -4.001 | -4.033 | -3.814 | -3.975 | -4.168 | -4.263 | -3.974 | -3.998 | -3.713 | -3.835 |
| 36. Zgab3 | .226 | .632 | .787 | - | -3.509 | -3.664 | -3.576 | -3.564 | -3.550 | -3.812 | -3.918 | -4.071 | -3.776 | -3.864 | -3.776 | -3.776 | -3.776 | -3.918 | -3.970 | -3.918 | -3.698 | -3.751 | -3.456 | -3.671 | -4.043 | -4.043 | -3.651 | -3.847 | -3.554 | -3.671 |
| 37. DmelA1 | 0 | .001 | .001 | .001 | - | -1.253 | 1.252 | -1.407 | 1.043 | -1.897 | -4.160 | -4.081 | -4.218 | -4.218 | -4.177 | -4.123 | -4.218 | -4.160 | -4.138 | -4.160 | -4.030 | -4.114 | -4.101 | -4.079 | -4.129 | -4.094 | -4.141 | -4.148 | -4.034 | -4.098 |
| 38. DmelA2 | 0 | .001 | .001 | 0 | .213 | - | -1.264 | -2.218 | -1.411 | -2.569 | -4.207 | -4.144 | -4.278 | -4.278 | -4.247 | -4.202 | -4.278 | -4.207 | -4.184 | -4.207 | -3.971 | -4.122 | -4.154 | -4.117 | -4.160 | -4.125 | -4.188 | -4.208 | -4.091 | -4.168 |
| 39. DmelA3 | 0 | .001 | .001 | .001 | .213 | .209 | - | -.746 | 2.979 | -1.368 | -4.015 | -3.915 | -4.113 | -4.113 | -4.057 | -3.984 | -4.113 | -4.015 | -3.995 | -4.015 | -3.884 | -3.986 | -4.009 | -3.969 | -4.066 | -4.026 | -4.060 | -4.043 | -3.933 | -3.974 |
| 40. DsimA1 | 0 | .001 | .001 | .001 | .162 | .028 | .457 | - | -1.137 | -1.144 | -4.059 | -3.931 | -4.160 | -4.160 | -4.084 | -3.991 | -4.160 | -4.059 | -4.059 | -4.059 | -3.899 | -4.104 | -4.067 | -4.117 | -4.181 | -4.144 | -4.131 | -4.087 | -3.979 | -3.996 |
| 41. DereA1 | 0 | .001 | .001 | .001 | .299 | .161 | .004 | .258 | - | -1.688 | -4.059 | -3.931 | -4.180 | -4.180 | -4.103 | -4.008 | -4.180 | -4.059 | -4.040 | -4.059 | -3.899 | -4.082 | -4.088 | -4.109 | -4.205 | -4.167 | -4.153 | -4.107 | -3.999 | -4.014 |
| 42. Ztub1 | 0 | 0 | 0 | 0 | .060 | .011 | .174 | .255 | .094 | - | -4.056 | -3.952 | -4.136 | -4.136 | -4.077 | -4.003 | -4.136 | -4.056 | -4.035 | -4.056 | -3.902 | -4.056 | -4.032 | -4.028 | -4.125 | -4.094 | -4.084 | -4.110 | -3.997 | -4.055 |
| 43. DmelA5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | - | -1.981 | -1.800 | -1.800 | -1.509 | -1.800 | -1.800 | -1.981 | -1.801 | -1.981 | -2.456 | -3.279 | -1.437 | -3.051 | -3.725 | -3.725 | -2.168 | -2.593 | -1.587 | -2.211 |
| 44. DmelA6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .050 | - | -2.269 | -2.269 | -2.050 | -2.269 | -2.269 | -2.412 | -2.270 | -2.412 | -2.797 | -3.515 | -1.959 | -3.316 | -3.726 | -3.726 | -2.554 | -2.922 | -2.056 | -2.596 |
| 45. DyakA1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .074 | .025 | - | -1.408 | -.998 | -1.408 | .000 | -1.800 | -1.622 | -1.800 | -2.173 | -3.090 | -.913 | -2.513 | -3.349 | -3.349 | -1.840 | -2.336 | -1.144 | -1.893 |
| 46. DyakA2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .074 | .025 | .162 | - | -.998 | -1.408 | -1.408 | -2.268 | -2.126 | -2.268 | -2.555 | -3.090 | -1.583 | -2.513 | -3.349 | -3.349 | -2.284 | -2.703 | -1.705 | -2.337 |
| 47. DyakA3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .134 | .043 | .320 | .320 | - | -.998 | -.998 | -2.049 | -1.891 | -2.049 | -2.372 | -3.222 | -1.279 | -2.682 | -3.468 | -3.468 | -2.073 | -2.527 | -1.442 | -2.127 |
| 48. DyakA4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .074 | .025 | .162 | .162 | .320 | - | -1.408 | -2.268 | -2.126 | -2.268 | -2.555 | -3.346 | -1.583 | -2.841 | -3.580 | -3.580 | -2.284 | -2.703 | -1.705 | -2.337 |
| 49. DyakA5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .074 | .025 | 1.000 | .162 | .320 | .162 | - | -1.800 | -1.622 | -1.800 | -2.173 | -3.090 | -.913 | -2.513 | -3.349 | -3.349 | -1.840 | -2.336 | -1.144 | -1.893 |
| 50. DereA3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .050 | .017 | .074 | .025 | .043 | .025 | .074 | - | .999 | .000 | -2.267 | -3.279 | -1.117 | -3.051 | -3.725 | -3.725 | -1.949 | -2.410 | -1.006 | -1.989 |
| 51. DereA4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .074 | .025 | .107 | .036 | .061 | .036 | .107 | .320 | - | .999 | -2.158 | -3.209 | -.954 | -2.970 | -3.662 | -3.662 | -1.821 | -2.291 | -.864 | -1.851 |
| 52. DereA5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .050 | .017 | .074 | .025 | .043 | .025 | .074 | 1.000 | .320 | - | -2.267 | -3.279 | -1.117 | -3.051 | -3.725 | -3.725 | -1.949 | -2.410 | -1.006 | -1.989 |
| 53. DereA8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .015 | .006 | .032 | .012 | .019 | .012 | .032 | .025 | .033 | .025 | - | -2.497 | -1.152 | -3.281 | -3.866 | -3.866 | -1.707 | -2.377 | -1.612 | -1.963 |
| 54. DereB1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | .001 | .002 | .002 | .002 | .001 | .002 | .001 | .002 | .001 | .014 | - | -2.879 | -1.531 | -3.103 | -2.816 | -3.061 | -3.286 | -2.737 | -3.288 |
| 55. DereB8 | 0 | 0 | 0 | .001 | 0 | 0 | 0 | 0 | 0 | 0 | .153 | .052 | .363 | .116 | .203 | .116 | .363 | .266 | .342 | .266 | .252 | .005 | - | -2.655 | -3.428 | -3.428 | -.187 | -1.286 | -.040 | -.451 |
| 56. DereB10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .003 | .001 | .013 | .013 | .008 | .005 | .013 | .003 | .004 | .003 | .001 | .128 | .009 | - | -2.139 | -1.702 | -2.802 | -3.110 | -2.409 | -3.112 |
| 57. Ztub5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | .001 | .001 | 0 | .001 | 0 | 0 | 0 | 0 | .002 | .001 | .034 | - | -1.408 | -3.534 | -3.559 | -3.050 | -3.560 |
| 58. Ztub6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .001 | .001 | .001 | 0 | .001 | 0 | 0 | 0 | 0 | .006 | .001 | .091 | .162 | - | -3.534 | -3.762 | -3.303 | -3.763 |
| 59. Zcam3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .032 | .012 | .068 | .024 | .040 | .024 | .068 | .054 | .071 | .054 | .090 | .003 | .852 | .006 | .001 | .001 | - | -1.283 | -.179 | -1.286 |
| 60. Zcam5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .011 | .004 | .021 | .008 | .013 | .008 | .021 | .017 | .024 | .017 | .019 | .001 | .201 | .002 | .001 | 0 | .202 | - | -.382 | -1.628 |
| 61. Zcam7 | 0 | 0 | 0 | .001 | 0 | 0 | 0 | 0 | 0 | 0 | .115 | .042 | .255 | .091 | .152 | .091 | .255 | .316 | .389 | .316 | .109 | .007 | .968 | .017 | .003 | .001 | .858 | .703 | - | -.386 |
| 62. Zdav1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .029 | .011 | .061 | .021 | .035 | .021 | .061 | .049 | .067 | .049 | .052 | .001 | .653 | .002 | .001 | 0 | .201 | .106 | .700 | - |

Symbols for species names: Ztub1: *Z. tuberculatus*; Zcam: *Z. camerounensis*; Zdav: *Z. davidi*; Zgab: *Z. gabonicus*; Zafr: *Z. africanus*; Zind: *Z. indianus*; Dmel: *D. melanogaster*; Dsim: *D. simulans*; Dsec: *D. sechellia*; Dyak: *D. yakuba*; Dere: *D. erecta*.

**Supplementary Table 5.** Pairwise genetic distance (maximum composite likelihood distance) among *micropia* sequences of *Zaprionus*, *melanogaster* and *repleta* species.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Dser | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2. Dhyd | .065 | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3. Dkoe | .595 | .578 | | | | | | | | | | | | | | | | | | | | | | | | |
| 4. Dmer | .051 | .037 | .557 | | | | | | | | | | | | | | | | | | | | | | | |
| 5. Dspe | .724 | .716 | .592 | .656 | | | | | | | | | | | | | | | | | | | | | | |
| 6. Dsem | .069 | .087 | .717 | .072 | .825 | | | | | | | | | | | | | | | | | | | | | |
| 7. Dant | .105 | .137 | .753 | .123 | .814 | .106 | | | | | | | | | | | | | | | | | | | | |
| 8. Dgou | .071 | .055 | 1.198 | .050 | 1.112 | .050 | .087 | | | | | | | | | | | | | | | | | | | |
| 9. Dbuz | .025 | .018 | .581 | .007 | .702 | .033 | .089 | .050 | | | | | | | | | | | | | | | | | | |
| 10. Dpar | .035 | .013 | .536 | .000 | .657 | .055 | .104 | .050 | .007 | | | | | | | | | | | | | | | | | |
| 11. DmelA1 | .532 | .517 | 1.108 | .503 | 1.037 | .512 | .592 | .548 | .494 | .492 | | | | | | | | | | | | | | | | |
| 12. DmelA3 | .527 | .511 | 1.100 | .497 | 1.047 | .507 | .585 | .539 | .488 | .487 | .003 | | | | | | | | | | | | | | | |
| 13. DsimA1 | .532 | .517 | 1.108 | .503 | 1.037 | .512 | .592 | .548 | .494 | .492 | .000 | .003 | | | | | | | | | | | | | | |
| 14. DsimA2 | .532 | .517 | 1.108 | .503 | 1.037 | .512 | .599 | .548 | .494 | .492 | .006 | .010 | .006 | | | | | | | | | | | | | |
| 15. DsimA3 | .547 | .532 | 1.149 | .517 | 1.071 | .528 | .601 | .560 | .509 | .507 | .010 | .013 | .010 | .016 | | | | | | | | | | | | |
| 16. DsimA4 | .536 | .526 | 1.123 | .511 | 1.062 | .513 | .595 | .550 | .503 | .501 | .003 | .007 | .003 | .010 | .013 | | | | | | | | | | | |
| 17. DsimA5 | .558 | .542 | 1.123 | .527 | 1.059 | .538 | .606 | .587 | .520 | .517 | .016 | .020 | .016 | .023 | .020 | .020 | | | | | | | | | | |
| 18. DsimA6 | .547 | .532 | 1.149 | .517 | 1.071 | .528 | .601 | .560 | .509 | .507 | .013 | .016 | .013 | .019 | .003 | .016 | .023 | | | | | | | | | |
| 19. DsecA1 | .532 | .517 | 1.108 | .503 | 1.037 | .512 | .592 | .548 | .494 | .492 | .000 | .003 | .000 | .006 | .010 | .003 | .016 | .013 | | | | | | | | |
| 20. DsecA2 | .532 | .517 | 1.108 | .503 | 1.037 | .512 | .592 | .548 | .494 | .492 | .000 | .003 | .000 | .006 | .010 | .003 | .016 | .013 | .000 | | | | | | | |
| 21. DsecA5 | .549 | .533 | 1.081 | .518 | 1.057 | .529 | .614 | .573 | .510 | .508 | .006 | .010 | .006 | .013 | .016 | .010 | .023 | .020 | .006 | .006 | | | | | | |
| 22. DsecA7 | .538 | .523 | 1.117 | .508 | 1.047 | .518 | .599 | .548 | .500 | .498 | .003 | .006 | .003 | .010 | .013 | .007 | .020 | .016 | .003 | .003 | .010 | | | | | |
| 23. DsecA8 | .532 | .517 | 1.119 | .503 | 1.043 | .512 | .592 | .548 | .494 | .492 | .000 | .003 | .000 | .006 | .010 | .003 | .016 | .013 | .000 | .000 | .006 | .003 | | | | |
| 24. DsecA9 | .547 | .532 | 1.149 | .517 | 1.071 | .528 | .601 | .560 | .509 | .507 | .013 | .016 | .013 | .019 | .003 | .016 | .023 | .000 | .013 | .013 | .020 | .016 | .013 | | | |
| 25. DsecA11 | .547 | .532 | 1.149 | .517 | 1.071 | .528 | .584 | .560 | .509 | .507 | .016 | .019 | .016 | .023 | .013 | .020 | .026 | .010 | .016 | .016 | .023 | .019 | .016 | .010 | | |
| 26. DsecA18 | .610 | .593 | 1.183 | .595 | 1.062 | .610 | .663 | .671 | .591 | .576 | .058 | .062 | .058 | .065 | .062 | .063 | .069 | .065 | .058 | .058 | .066 | .062 | .058 | .065 | .069 | |
| 27. DsecA19 | .546 | .514 | 1.151 | .505 | 1.135 | .532 | .596 | .549 | .490 | .497 | .043 | .047 | .043 | .050 | .054 | .047 | .061 | .057 | .043 | .043 | .050 | .047 | .043 | .057 | .061 | .107 |
| 28. DsecA20 | .580 | .553 | 1.098 | .546 | 1.043 | .581 | .630 | .620 | .561 | .533 | .105 | .101 | .105 | .112 | .116 | .110 | .121 | .116 | .105 | .105 | .105 | .108 | .105 | .116 | .120 | .156 |
| 29. DsecB8 | .543 | .528 | 1.135 | .513 | 1.066 | .525 | .588 | .548 | .505 | .503 | .010 | .013 | .010 | .016 | .020 | .013 | .026 | .023 | .010 | .010 | .016 | .013 | .010 | .023 | .023 | .070 |
| 30. DsecB19 | .531 | .515 | 1.113 | .501 | 1.062 | .511 | .581 | .539 | .492 | .490 | .010 | .013 | .010 | .016 | .020 | .013 | .026 | .023 | .010 | .010 | .016 | .013 | .010 | .023 | .026 | .069 |
| 31. DyakA1 | .548 | .519 | 1.016 | .505 | 1.027 | .535 | .585 | .547 | .516 | .494 | .078 | .074 | .078 | .085 | .081 | .082 | .092 | .085 | .078 | .078 | .085 | .081 | .078 | .085 | .095 | .132 |
| 32. DyakA2 | .572 | .543 | 1.101 | .528 | 1.089 | .548 | .629 | .560 | .532 | .518 | .088 | .085 | .088 | .095 | .092 | .093 | .103 | .095 | .088 | .088 | .096 | .092 | .088 | .095 | .106 | .144 |
| 33. DyakA3 | .579 | .552 | 1.122 | .534 | 1.097 | .569 | .627 | .568 | .551 | .527 | .095 | .092 | .095 | .103 | .099 | .100 | .111 | .103 | .095 | .095 | .103 | .099 | .095 | .103 | .114 | .160 |
| 34. DyakB3 | .562 | .521 | .979 | .514 | .977 | .547 | .637 | .541 | .527 | .500 | .118 | .114 | .118 | .118 | .122 | .123 | .135 | .122 | .118 | .118 | .118 | .122 | .118 | .122 | .134 | .163 |
| 35. Ztub1 | .558 | .542 | 1.130 | .527 | 1.053 | .538 | .614 | .579 | .519 | .517 | .020 | .023 | .020 | .026 | .030 | .023 | .037 | .033 | .020 | .020 | .026 | .023 | .020 | .033 | .036 | .080 |
| 36. Ztub2 | .522 | .507 | 1.104 | .493 | 1.052 | .508 | .570 | .536 | .489 | .482 | .013 | .016 | .013 | .020 | .023 | .016 | .030 | .026 | .013 | .013 | .020 | .016 | .013 | .026 | .030 | .073 |
| 37. Ztub3 | .539 | .523 | 1.119 | .509 | 1.086 | .519 | .599 | .552 | .500 | .498 | .016 | .016 | .016 | .023 | .026 | .020 | .033 | .030 | .016 | .016 | .023 | .020 | .016 | .030 | .033 | .073 |
| 38. Zcam1 | .544 | .515 | 1.063 | .500 | 1.062 | .530 | .597 | .540 | .511 | .490 | .067 | .064 | .067 | .067 | .078 | .071 | .085 | .081 | .067 | .067 | .074 | .070 | .067 | .081 | .085 | .129 |
| 39. Zcam2 | .551 | .521 | 1.075 | .507 | 1.076 | .537 | .605 | .550 | .518 | .496 | .071 | .067 | .071 | .071 | .081 | .075 | .089 | .085 | .071 | .071 | .078 | .074 | .071 | .085 | .088 | .133 |
| 40. Zcam4 | .551 | .521 | 1.051 | .507 | 1.076 | .537 | .605 | .550 | .518 | .496 | .071 | .067 | .071 | .071 | .081 | .075 | .089 | .085 | .071 | .071 | .078 | .074 | .071 | .085 | .088 | .133 |
| 41. Zdav1 | .540 | .511 | 1.066 | .497 | 1.096 | .526 | .595 | .518 | .507 | .487 | .081 | .078 | .081 | .088 | .092 | .086 | .096 | .095 | .081 | .081 | .089 | .085 | .081 | .095 | .099 | .144 |
| 42. Zdav2 | .518 | .490 | 1.023 | .476 | 1.046 | .504 | .568 | .514 | .485 | .466 | .070 | .067 | .070 | .077 | .081 | .075 | .085 | .084 | .070 | .070 | .078 | .074 | .070 | .084 | .088 | .132 |
| 43. Zdav4 | .524 | .495 | 1.032 | .482 | 1.056 | .510 | .575 | .506 | .491 | .471 | .074 | .070 | .074 | .081 | .084 | .078 | .088 | .088 | .074 | .074 | .081 | .077 | .074 | .088 | .091 | .136 |
| 44. Zgab1 | .526 | .498 | 1.040 | .484 | 1.066 | .504 | .568 | .514 | .493 | .473 | .074 | .070 | .074 | .081 | .085 | .078 | .088 | .088 | .074 | .074 | .081 | .077 | .074 | .088 | .092 | .136 |
| 45. Zgab3 | .510 | .498 | 1.007 | .484 | 1.046 | .512 | .578 | .526 | .493 | .473 | .074 | .070 | .074 | .081 | .085 | .078 | .088 | .088 | .074 | .074 | .081 | .077 | .074 | .088 | .092 | .128 |
| 46. Zgab5 | .520 | .503 | 1.051 | .478 | 1.073 | .498 | .561 | .514 | .487 | .468 | .077 | .074 | .077 | .084 | .088 | .082 | .092 | .092 | .077 | .077 | .085 | .081 | .077 | .092 | .095 | .140 |
| 47. Zafr1 | .540 | .502 | 1.083 | .495 | 1.064 | .517 | .570 | .512 | .506 | .482 | .077 | .073 | .077 | .084 | .088 | .081 | .092 | .092 | .077 | .077 | .084 | .073 | .077 | .092 | .095 | .133 |
| 48. Zafr3 | .540 | .502 | 1.083 | .495 | 1.064 | .517 | .570 | .512 | .506 | .482 | .077 | .073 | .077 | .084 | .088 | .081 | .092 | .092 | .077 | .077 | .084 | .073 | .077 | .092 | .095 | .133 |
| 49. Zafr5 | .518 | .490 | 1.023 | .476 | 1.046 | .504 | .568 | .514 | .485 | .466 | .070 | .067 | .070 | .077 | .081 | .075 | .085 | .084 | .070 | .070 | .078 | .074 | .070 | .084 | .088 | .132 |
| 50. Zind2 | .519 | .491 | 1.028 | .477 | 1.052 | .505 | .570 | .514 | .486 | .467 | .070 | .067 | .070 | .077 | .081 | .075 | .085 | .085 | .070 | .070 | .078 | .074 | .070 | .085 | .088 | .133 |
| 51. Zind3 | .526 | .498 | 1.040 | .484 | 1.066 | .504 | .568 | .514 | .493 | .473 | .074 | .070 | .074 | .081 | .085 | .078 | .088 | .088 | .074 | .074 | .081 | .077 | .074 | .088 | .092 | .136 |
| 52. Zind4 | .518 | .490 | 1.023 | .476 | 1.046 | .504 | .568 | .514 | .485 | .466 | .070 | .067 | .070 | .077 | .081 | .075 | .085 | .084 | .070 | .070 | .078 | .074 | .070 | .084 | .088 | .132 |

**Supplementary Table 5, continuation.**

| | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 28. DsecA20 | .150 | | | | | | | | | | | | | | | | | | | | | | | | |
| 29. DsecB8 | .043 | .117 | | | | | | | | | | | | | | | | | | | | | | | |
| 30. DsecB19 | .030 | .117 | .013 | | | | | | | | | | | | | | | | | | | | | | |
| 31. DyakA1 | .115 | .083 | .085 | .085 | | | | | | | | | | | | | | | | | | | | | |
| 32. DyakA2 | .134 | .094 | .100 | .099 | .043 | | | | | | | | | | | | | | | | | | | | |
| 33. DyakA3 | .134 | .098 | .104 | .103 | .036 | .061 | | | | | | | | | | | | | | | | | | | |
| 34. DyakB3 | .157 | .104 | .127 | .122 | .059 | .086 | .079 | | | | | | | | | | | | | | | | | | |
| 35. Ztub1 | .053 | .121 | .023 | .023 | .089 | .104 | .107 | .131 | | | | | | | | | | | | | | | | | |
| 36. Ztub2 | .047 | .106 | .016 | .016 | .075 | .089 | .093 | .115 | .020 | | | | | | | | | | | | | | | | |
| 37. Ztub3 | .050 | .110 | .020 | .019 | .082 | .096 | .100 | .119 | .023 | .016 | | | | | | | | | | | | | | | |
| 38. Zcam1 | .111 | .075 | .078 | .078 | .043 | .054 | .061 | .059 | .082 | .068 | .075 | | | | | | | | | | | | | | |
| 39. Zcam2 | .115 | .079 | .082 | .081 | .047 | .057 | .064 | .063 | .085 | .071 | .078 | .003 | | | | | | | | | | | | | |
| 40. Zcam4 | .115 | .079 | .082 | .081 | .047 | .057 | .064 | .063 | .085 | .071 | .071 | .003 | .006 | | | | | | | | | | | | |
| 41. Zdav1 | .118 | .078 | .093 | .092 | .043 | .057 | .064 | .082 | .096 | .082 | .089 | .047 | .050 | .050 | | | | | | | | | | | |
| 42. Zdav2 | .107 | .075 | .082 | .081 | .033 | .047 | .053 | .074 | .085 | .071 | .078 | .036 | .040 | .040 | .010 | | | | | | | | | | |
| 43. Zdav4 | .110 | .078 | .085 | .085 | .036 | .050 | .057 | .078 | .089 | .074 | .082 | .040 | .043 | .043 | .006 | .003 | | | | | | | | | |
| 44. Zgab1 | .111 | .078 | .085 | .085 | .036 | .050 | .057 | .078 | .089 | .075 | .082 | .040 | .043 | .043 | .013 | .003 | .006 | | | | | | | | |
| 45. Zgab3 | .111 | .078 | .085 | .085 | .036 | .050 | .057 | .078 | .089 | .075 | .082 | .040 | .043 | .043 | .013 | .003 | .006 | .006 | | | | | | | |
| 46. Zgab5 | .114 | .082 | .089 | .088 | .040 | .054 | .060 | .082 | .092 | .078 | .085 | .043 | .047 | .047 | .016 | .006 | .010 | .003 | .010 | | | | | | |
| 47. Zafr1 | .115 | .086 | .089 | .088 | .041 | .056 | .059 | .088 | .092 | .077 | .081 | .038 | .041 | .041 | .020 | .010 | .013 | .013 | .013 | .017 | | | | | |
| 48. Zafr3 | .115 | .086 | .089 | .088 | .041 | .056 | .059 | .088 | .092 | .077 | .081 | .038 | .041 | .041 | .020 | .010 | .013 | .013 | .013 | .017 | .000 | | | | |
| 49. Zafr5 | .107 | .075 | .082 | .081 | .033 | .047 | .053 | .074 | .085 | .071 | .078 | .036 | .040 | .040 | .010 | .000 | .003 | .003 | .003 | .006 | .010 | .010 | | | |
| 50. Zind2 | .107 | .082 | .082 | .081 | .040 | .054 | .060 | .082 | .092 | .078 | .085 | .043 | .047 | .047 | .016 | .006 | .010 | .010 | .010 | .013 | .017 | .017 | .006 | | |
| 51. Zind3 | .111 | .078 | .085 | .085 | .036 | .050 | .057 | .078 | .089 | .075 | .082 | .040 | .043 | .043 | .013 | .003 | .006 | .006 | .006 | .010 | .013 | .013 | .003 | .010 | |
| 52. Zind4 | .107 | .075 | .082 | .081 | .033 | .047 | .053 | .074 | .085 | .071 | .078 | .036 | .040 | .040 | .010 | .000 | .003 | .003 | .003 | .006 | .010 | .010 | .000 | .006 | .003 |

Symbols for species names: Dser: *D. serido*; Dhyd: *D. hydei*; Dkoe: *D. koepferae*; Dmer: *D. mercatorum*; Dspe: *D. spenceri*; Dsem: *D. seriema*; Dant: *D. antonietae*; Dgou: *D. gouveai*; Dbuz: *D. buzzatii*; Dpar: *D. paranaensis*; Dmel: *D. melanogaster*; Dsim: *D. simulans*; Dsec: *D. sechellia*; Dyak: *D. yakuba*; Ztub: *Z. tuberculatus*; Zcam: *Z. camerounensis*; Zdav: *Z. davidi*; Zgab: *Z. gabonicus*; Zafr: *Z. africanus*; Zind: *Z. indianus*.

**Supplementary Table 6.** dN (below) and dS (above) values of pairwise comparisons between *micropia* sequences of *Zaprionus* and *melanogaster* species. Distances calculated by the Nei-Gojobori method with Jukes-Cantor's correction as implemented by MEGA 4.1.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. DmelA1 | - | .000 | .000 | .000 | .000 | .000 | .016 | .000 | .000 | .000 | .000 | .000 | .065 | .099 | .209 | .000 | .032 | .155 | .154 | .187 |
| 2. DmelA3 | .005 | - | .000 | .000 | .000 | .000 | .016 | .000 | .000 | .000 | .000 | .000 | .065 | .099 | .209 | .000 | .032 | .155 | .154 | .187 |
| 3. DdmelA4 | .000 | .005 | - | .000 | .000 | .000 | .016 | .000 | .000 | .000 | .000 | .000 | .065 | .099 | .209 | .000 | .032 | .155 | .154 | .187 |
| 4. DsimA1 | .000 | .005 | .000 | - | .000 | .000 | .016 | .000 | .000 | .000 | .000 | .000 | .065 | .099 | .209 | .000 | .032 | .155 | .154 | .187 |
| 5. DsimA2 | .005 | .009 | .005 | .005 | - | .000 | .016 | .000 | .000 | .000 | .000 | .000 | .065 | .099 | .209 | .000 | .032 | .155 | .154 | .187 |
| 6. DsimA4 | .000 | .005 | .000 | .000 | .005 | - | .016 | .000 | .000 | .000 | .000 | .000 | .065 | .099 | .209 | .000 | .032 | .155 | .154 | .187 |
| 7. DsimA5 | .009 | .014 | .009 | .009 | .014 | .009 | - | .016 | .016 | .016 | .016 | .016 | .048 | .116 | .234 | .016 | .048 | .175 | .174 | .209 |
| 8. DsecA1 | .000 | .005 | .000 | .000 | .005 | .000 | .009 | - | .000 | .000 | .000 | .000 | .065 | .099 | .209 | .000 | .032 | .155 | .154 | .187 |
| 9. DsecA2 | .000 | .005 | .000 | .000 | .005 | .000 | .009 | .000 | - | .000 | .000 | .000 | .065 | .099 | .209 | .000 | .032 | .155 | .154 | .187 |
| 10. DsecA5 | .009 | .014 | .009 | .009 | .014 | .009 | .019 | .009 | .009 | - | .000 | .000 | .065 | .099 | .209 | .000 | .032 | .156 | .155 | .198 |
| 11. DsecA7 | .005 | .009 | .005 | .005 | .009 | .005 | .014 | .005 | .005 | .014 | - | .000 | .065 | .099 | .210 | .000 | .032 | .156 | .155 | .188 |
| 12. DsecA8 | .005 | .009 | .005 | .005 | .009 | .005 | .014 | .005 | .005 | .014 | .009 | - | .066 | .099 | .209 | .000 | .032 | .156 | .155 | .188 |
| 13. DsecA11 | .005 | .009 | .005 | .005 | .009 | .005 | .014 | .005 | .005 | .014 | .009 | .009 | - | .173 | .287 | .066 | .100 | .238 | .236 | .255 |
| 14. DsecA19 | .019 | .024 | .019 | .019 | .024 | .019 | .029 | .019 | .019 | .029 | .024 | .024 | .024 | - | .287 | .099 | .064 | .255 | .253 | .274 |
| 15. DsecA20 | .097 | .090 | .097 | .097 | .103 | .097 | .110 | .097 | .097 | .090 | .103 | .097 | .103 | .124 | - | .210 | .233 | .077 | .086 | .130 |
| 16. DsecB8 | .009 | .014 | .009 | .009 | .014 | .009 | .019 | .009 | .009 | .019 | .014 | .014 | .009 | .029 | .110 | - | .032 | .156 | .155 | .188 |
| 17. DsecB19 | .000 | .005 | .000 | .000 | .005 | .000 | .009 | .000 | .000 | .009 | .005 | .005 | .005 | .019 | .097 | .009 | - | .195 | .193 | .209 |
| 18. DyakA1 | .053 | .048 | .053 | .053 | .058 | .053 | .063 | .053 | .053 | .063 | .058 | .058 | .058 | .074 | .090 | .063 | .053 | - | .048 | .070 |
| 19. DyakA2 | .058 | .053 | .058 | .058 | .063 | .058 | .068 | .058 | .058 | .068 | .063 | .063 | .063 | .079 | .107 | .068 | .058 | .048 | - | .051 |
| 20. DyakB3 | .074 | .069 | .074 | .074 | .08 | .074 | .086 | .074 | .074 | .072 | .080 | .080 | .080 | .097 | .087 | .086 | .074 | .052 | .075 | - |
| 21. Ztub1 | .014 | .019 | .014 | .014 | .019 | .014 | .024 | .014 | .014 | .024 | .019 | .019 | .019 | .033 | .116 | .024 | .014 | .068 | .073 | .091 |
| 22. Ztub2 | .009 | .014 | .009 | .009 | .014 | .009 | .019 | .009 | .009 | .019 | .014 | .014 | .014 | .029 | .103 | .019 | .009 | .063 | .068 | .086 |
| 23. Ztub3 | .014 | .014 | .014 | .014 | .019 | .014 | .024 | .014 | .014 | .024 | .019 | .019 | .019 | .036 | .100 | .024 | .014 | .063 | .068 | .078 |
| 24. Zcam1 | .033 | .029 | .033 | .033 | .038 | .033 | .043 | .033 | .033 | .043 | .038 | .038 | .038 | .053 | .071 | .043 | .033 | .024 | .033 | .042 |
| 25. Zcam2 | .038 | .033 | .038 | .038 | .043 | .038 | .048 | .038 | .038 | .048 | .043 | .043 | .043 | .058 | .078 | .048 | .038 | .028 | .038 | .047 |
| 26. Zcam4 | .038 | .033 | .038 | .038 | .043 | .038 | .048 | .038 | .038 | .048 | .043 | .043 | .043 | .058 | .075 | .048 | .038 | .029 | .038 | .045 |
| 27. Zdav1 | .053 | .048 | .053 | .053 | .058 | .053 | .063 | .053 | .053 | .063 | .058 | .058 | .058 | .073 | .090 | .063 | .053 | .036 | .053 | .055 |
| 28. Zdav2 | .038 | .033 | .038 | .038 | .043 | .038 | .048 | .038 | .038 | .048 | .043 | .043 | .043 | .058 | .084 | .048 | .038 | .021 | .038 | .050 |
| 29. Zdav4 | .043 | .038 | .043 | .043 | .048 | .043 | .053 | .043 | .043 | .053 | .048 | .048 | .048 | .063 | .090 | .053 | .043 | .026 | .043 | .055 |
| 30. Zgab1 | .038 | .033 | .038 | .038 | .043 | .038 | .048 | .038 | .038 | .048 | .043 | .043 | .043 | .058 | .084 | .048 | .038 | .021 | .038 | .050 |
| 31. Zgab3 | .038 | .033 | .038 | .038 | .043 | .038 | .048 | .038 | .038 | .048 | .043 | .043 | .043 | .058 | .084 | .048 | .038 | .021 | .038 | .050 |
| 32. Zgab5 | .041 | .036 | .041 | .041 | .046 | .041 | .051 | .041 | .041 | .051 | .046 | .046 | .046 | .062 | .087 | .05 | .041 | .026 | .043 | .055 |
| 33. Zafr1 | .059 | .054 | .059 | .059 | .064 | .059 | .07 | .059 | .059 | .07 | .054 | .064 | .064 | .080 | .100 | .070 | .059 | .041 | .059 | .072 |
| 34. Zafr3 | .054 | .049 | .054 | .054 | .059 | .054 | .065 | .054 | .054 | .064 | .049 | .059 | .059 | .075 | .100 | .064 | .054 | .037 | .054 | .067 |
| 35. Zafr5 | .038 | .033 | .038 | .038 | .043 | .038 | .048 | .038 | .038 | .048 | .043 | .043 | .043 | .058 | .084 | .048 | .038 | .021 | .038 | .050 |
| 36. Zind2 | .043 | .038 | .043 | .043 | .048 | .043 | .053 | .043 | .043 | .053 | .048 | .048 | .048 | .063 | .090 | .053 | .043 | .026 | .043 | .055 |
| 37. Zind3 | .038 | .033 | .038 | .038 | .043 | .038 | .048 | .038 | .038 | .048 | .043 | .043 | .043 | .058 | .084 | .048 | .038 | .021 | .038 | .050 |
| 38. Zind4 | .038 | .033 | .038 | .038 | .043 | .038 | .048 | .038 | .038 | .048 | .043 | .043 | .043 | .058 | .084 | .048 | .038 | .021 | .038 | .050 |

**Supplementary Table 6, continuation.**

| | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. DmelA1 | .016 | .032 | .016 | .154 | .155 | .154 | .137 | .137 | .137 | .137 | .156 | .146 | .141 | .140 | .137 | .118 | .137 | .137 |
| 2. DmelA3 | .016 | .032 | .016 | .154 | .155 | .154 | .137 | .137 | .137 | .137 | .156 | .146 | .141 | .140 | .137 | .118 | .137 | .137 |
| 3. DdmelA4 | .016 | .032 | .016 | .154 | .155 | .154 | .137 | .137 | .137 | .137 | .156 | .146 | .141 | .140 | .137 | .118 | .137 | .137 |
| 4. DsimA1 | .016 | .032 | .016 | .154 | .155 | .154 | .137 | .137 | .137 | .137 | .156 | .146 | .141 | .140 | .137 | .118 | .137 | .137 |
| 5. DsimA2 | .016 | .032 | .016 | .154 | .155 | .154 | .137 | .137 | .137 | .137 | .156 | .146 | .141 | .140 | .137 | .118 | .137 | .137 |
| 6. DsimA4 | .016 | .032 | .016 | .154 | .155 | .154 | .137 | .137 | .137 | .137 | .156 | .146 | .141 | .140 | .137 | .118 | .137 | .137 |
| 7. DsimA5 | .032 | .048 | .032 | .174 | .175 | .173 | .156 | .156 | .156 | .156 | .175 | .165 | .160 | .159 | .156 | .137 | .156 | .156 |
| 8. DsecA1 | .016 | .032 | .016 | .154 | .155 | .154 | .137 | .137 | .137 | .137 | .156 | .146 | .141 | .140 | .137 | .118 | .137 | .137 |
| 9. DsecA2 | .016 | .032 | .016 | .154 | .155 | .154 | .137 | .137 | .137 | .137 | .156 | .146 | .141 | .140 | .137 | .118 | .137 | .137 |
| 10. DsecA5 | .016 | .032 | .016 | .155 | .156 | .154 | .137 | .137 | .137 | .137 | .156 | .146 | .141 | .140 | .137 | .118 | .137 | .137 |
| 11. DsecA7 | .016 | .032 | .016 | .155 | .156 | .154 | .137 | .137 | .137 | .137 | .156 | .146 | .141 | .140 | .137 | .118 | .137 | .137 |
| 12. DsecA8 | .016 | .032 | .016 | .155 | .156 | .154 | .138 | .137 | .138 | .137 | .156 | .146 | .141 | .140 | .137 | .119 | .137 | .137 |
| 13. DsecA11 | .084 | .100 | .083 | .236 | .238 | .235 | .217 | .217 | .217 | .217 | .238 | .227 | .224 | .222 | .217 | .196 | .217 | .217 |
| 14. DsecA19 | .118 | .134 | .107 | .253 | .255 | .252 | .234 | .234 | .234 | .234 | .255 | .236 | .241 | .239 | .234 | .213 | .234 | .234 |
| 15. DsecA20 | .188 | .162 | .197 | .057 | .057 | .066 | .057 | .057 | .057 | .057 | .077 | .067 | .067 | .067 | .057 | .077 | .057 | .057 |
| 16. DsecB8 | .016 | .032 | .016 | .155 | .156 | .154 | .138 | .137 | .138 | .137 | .156 | .146 | .141 | .140 | .137 | .119 | .137 | .137 |
| 17. DsecB19 | .049 | .065 | .048 | .193 | .195 | .192 | .175 | .175 | .175 | .175 | .195 | .184 | .180 | .179 | .175 | .156 | .175 | .175 |
| 18. DyakA1 | .138 | .117 | .136 | .048 | .048 | .048 | .040 | .040 | .040 | .040 | .057 | .040 | .041 | .041 | .040 | .057 | .040 | .040 |
| 19. DyakA2 | .137 | .117 | .135 | .032 | .032 | .031 | .016 | .016 | .016 | .016 | .032 | .016 | .016 | .016 | .016 | .032 | .016 | .016 |
| 20. DyakB3 | .169 | .145 | .173 | .034 | .034 | .039 | .061 | .043 | .043 | .043 | .061 | .043 | .043 | .043 | .043 | .061 | .043 | .043 |
| 21. Ztub1 | - | .016 | .000 | .138 | .138 | .137 | .120 | .120 | .120 | .12 | .139 | .129 | .123 | .123 | .120 | .139 | .120 | .120 |
| 22. Ztub2 | .024 | - | .016 | .117 | .117 | .116 | .100 | .100 | .100 | .100 | .118 | .108 | .102 | .102 | .100 | .118 | .100 | .100 |
| 23. Ztub3 | .028 | .024 | - | .136 | .136 | .135 | .128 | .118 | .118 | .118 | .137 | .127 | .121 | .121 | .118 | .137 | .118 | .118 |
| 24. Zcam1 | .048 | .043 | .043 | - | .000 | .000 | .016 | .016 | .016 | .016 | .032 | .016 | .000 | .000 | .016 | .032 | .016 | .016 |
| 25. Zcam2 | .053 | .048 | .048 | .005 | - | .000 | .016 | .016 | .016 | .016 | .032 | .016 | .000 | .000 | .016 | .032 | .016 | .016 |
| 26. Zcam4 | .053 | .048 | .038 | .005 | .009 | - | .024 | .016 | .016 | .016 | .032 | .016 | .000 | .000 | .016 | .032 | .016 | .016 |
| 27. Zdav1 | .068 | .063 | .06 | .028 | .033 | .031 | - | .000 | .000 | .000 | .016 | .000 | .000 | .000 | .000 | .016 | .000 | .000 |
| 28. Zdav2 | .053 | .048 | .048 | .014 | .019 | .019 | .014 | - | .000 | .000 | .016 | .000 | .000 | .000 | .000 | .016 | .000 | .000 |
| 29. Zdav4 | .058 | .053 | .053 | .019 | .024 | .024 | .009 | .005 | - | .000 | .016 | .000 | .000 | .000 | .000 | .016 | .000 | .000 |
| 30. Zgab1 | .053 | .048 | .048 | .014 | .019 | .019 | .014 | .000 | .005 | - | .016 | .000 | .000 | .000 | .000 | .016 | .000 | .000 |
| 31. Zgab3 | .053 | .048 | .048 | .014 | .019 | .019 | .014 | .000 | .005 | .000 | - | .016 | .016 | .016 | .016 | .032 | .016 | .016 |
| 32. Zgab5 | .055 | .051 | .051 | .019 | .024 | .024 | .019 | .005 | .009 | .005 | .005 | - | .000 | .000 | .000 | .016 | .000 | .000 |
| 33. Zafr1 | .075 | .07 | .064 | .034 | .039 | .039 | .034 | .019 | .024 | .019 | .019 | .024 | - | .000 | .000 | .016 | .000 | .000 |
| 34. Zafr3 | .069 | .065 | .059 | .029 | .034 | .034 | .029 | .014 | .019 | .014 | .014 | .019 | .005 | - | .000 | .016 | .000 | .000 |
| 35. Zafr5 | .053 | .048 | .048 | .014 | .019 | .019 | .014 | .000 | .005 | .000 | .000 | .005 | .019 | .014 | - | .016 | .000 | .000 |
| 36. Zind2 | .058 | .053 | .053 | .019 | .024 | .024 | .019 | .005 | .009 | .005 | .005 | .009 | .024 | .019 | .005 | - | .016 | .016 |
| 37. Zind3 | .053 | .048 | .048 | .014 | .019 | .019 | .014 | .000 | .005 | .000 | .000 | .005 | .019 | .014 | .000 | .005 | - | .000 |
| 38. Zind4 | .053 | .048 | .048 | .014 | .019 | .019 | .014 | .000 | .005 | .000 | .000 | .005 | .019 | .014 | .000 | .005 | .000 | - |

Symbols for species names: Ztub1: *Z. tuberculatus*; Zcam: *Z. camerounensis*; Zdav: *Z. davidi*; Zgab: *Z. gabonicus*; Zafr: *Z. africanus*; Zind: *Z. indianus*; Dmel: *D. melanogaster*; Dsim: *D. simulans*; Dsec: *D. sechellia*; Dyak: *D. yakuba*.

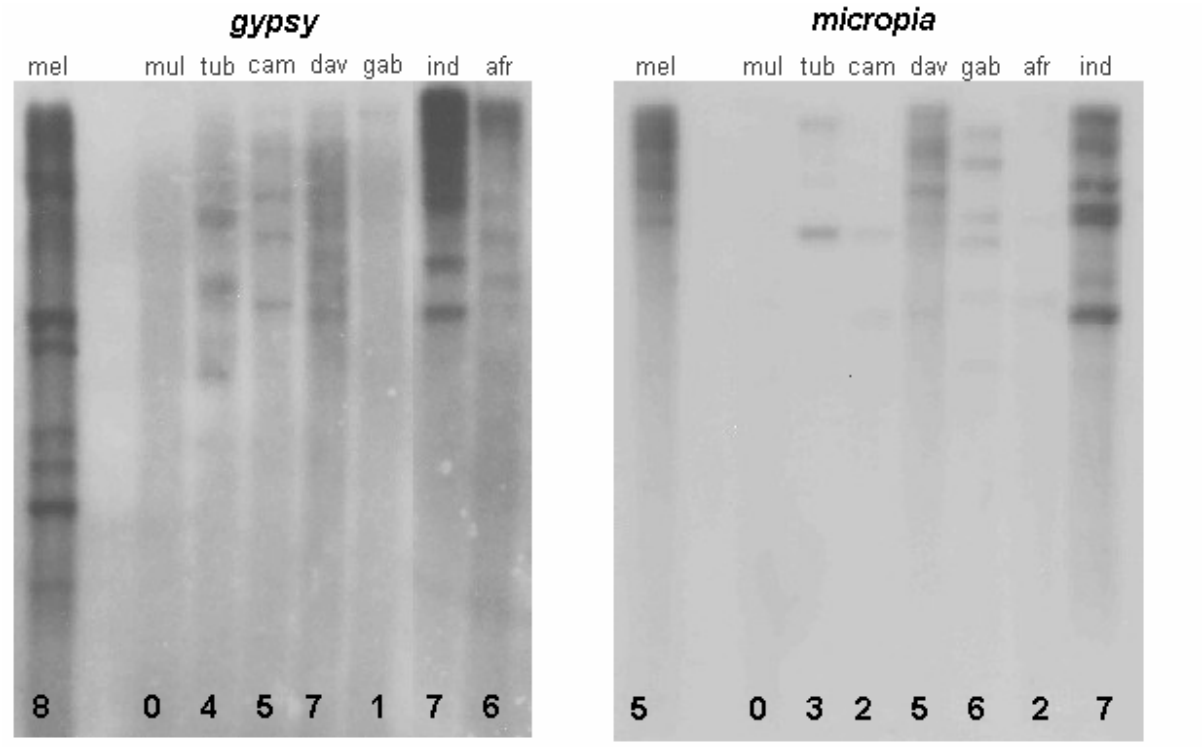
**Supplementary Table 7.** Z values (above) and significance P-values (below) for the Z-test of selection between *micropia* sequences of *Zaprionus* and *melanogaster* species. Test performed using alternative hypothesis of non-neutrality (dN ≠ dS) and Nei-Gojobori distance (Jukes-Cantor's correction). Gray cells correspond to significant ($p < 0.05$) pairwise comparisons.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. DmelA1 | - | .999 | .000 | .000 | .999 | .000 | -.372 | .000 | .000 | 1.412 | .999 | .999 | -1.818 | -1.893 | -1.533 | 1.412 | -1.407 | -1.835 | -1.727 | -1.743 |
| 2. DmelA3 | .320 | - | .999 | .999 | 1.412 | .999 | -.092 | .999 | .999 | 1.728 | 1.412 | 1.412 | -1.660 | -1.767 | -1.627 | 1.728 | -1.175 | -1.933 | -1.825 | -1.837 |
| 3. DdmelA4 | 1.000 | .320 | - | .000 | .999 | .000 | -.372 | .000 | .000 | 1.412 | .999 | .999 | -1.818 | -1.893 | -1.533 | 1.412 | -1.407 | -1.835 | -1.727 | -1.743 |
| 4. DsimA1 | 1.000 | .320 | 1.000 | - | .999 | .000 | -.372 | .000 | .000 | 1.412 | .999 | .999 | -1.818 | -1.893 | -1.533 | 1.412 | -1.407 | -1.835 | -1.727 | -1.743 |
| 5. DsimA2 | .320 | .161 | .320 | .320 | - | .999 | -.092 | .999 | .999 | 1.728 | 1.412 | 1.412 | -1.660 | -1.767 | -1.438 | 1.728 | -1.175 | -1.738 | -1.629 | -1.650 |
| 6. DsimA4 | 1.000 | .320 | 1.000 | 1.000 | .320 | - | -.372 | .000 | .000 | 1.412 | .999 | .999 | -1.818 | -1.893 | -1.533 | 1.412 | -1.407 | -1.835 | -1.727 | -1.743 |
| 7. DsimA5 | .711 | .927 | .711 | .711 | .927 | .711 | - | -.372 | -.372 | .167 | -.094 | -.096 | -1.171 | -1.891 | -1.580 | .165 | -1.347 | -1.867 | -1.764 | -1.781 |
| 8. DsecA1 | 1.000 | .320 | 1.000 | 1.000 | .320 | 1.000 | .711 | - | .000 | 1.412 | .999 | .999 | -1.818 | -1.893 | -1.533 | 1.412 | -1.407 | -1.835 | -1.727 | -1.743 |
| 9. DsecA2 | 1.000 | .320 | 1.000 | 1.000 | .320 | 1.000 | .711 | 1.000 | - | 1.412 | .999 | .999 | -1.818 | -1.893 | -1.533 | 1.412 | -1.407 | -1.835 | -1.727 | -1.743 |
| 10. DsecA5 | .161 | .087 | .161 | .161 | .087 | .161 | .868 | .161 | .161 | - | 1.728 | 1.728 | -1.507 | -1.645 | -1.627 | 1.994 | -.953 | -1.645 | -1.536 | -1.902 |
| 11. DsecA7 | .320 | .161 | .320 | .320 | .161 | .320 | .925 | .320 | .320 | .087 | - | 1.412 | -1.661 | -1.770 | -1.445 | 1.728 | -1.175 | -1.742 | -1.634 | -1.655 |
| 12. DsecA8 | .320 | .161 | .320 | .320 | .161 | .320 | .924 | .320 | .320 | .087 | .161 | - | -1.662 | -1.771 | -1.533 | 1.728 | -1.176 | -1.744 | -1.636 | -1.658 |
| 13. DsecA11 | .072 | .099 | .072 | .072 | .099 | .072 | .244 | .072 | .072 | .134 | .099 | .099 | - | -2.594 | -2.081 | -1.662 | -2.279 | -2.522 | -2.436 | -2.262 |
| 14. DsecA19 | .061 | .080 | .061 | .061 | .080 | .061 | .061 | .061 | .061 | .103 | .079 | .079 | .011 | - | -1.799 | -1.646 | -1.338 | -2.438 | -2.352 | -2.176 |
| 15. DsecA20 | .128 | .106 | .128 | .128 | .153 | .128 | .117 | .128 | .128 | .106 | .151 | .128 | .040 | .074 | - | -1.354 | -1.756 | .291 | .414 | -.754 |
| 16. DsecB8 | .161 | .087 | .161 | .161 | .087 | .161 | .869 | .161 | .161 | .048 | .087 | .087 | .099 | .102 | .178 | - | -.953 | -1.648 | -1.539 | -1.565 |
| 17. DsecB19 | .162 | .242 | .162 | .162 | .242 | .162 | .181 | .162 | .162 | .343 | .242 | .242 | .024 | .183 | .082 | .342 | - | -2.244 | -2.148 | -1.955 |
| 18. DyakA1 | .069 | .056 | .069 | .069 | .085 | .069 | .064 | .069 | .069 | .102 | .084 | .084 | .013 | .016 | .771 | .102 | .027 | - | .001 | -.441 |
| 19. DyakA2 | .087 | .070 | .087 | .087 | .106 | .087 | .080 | .087 | .087 | .127 | .105 | .104 | .016 | .020 | .680 | .126 | .034 | .999 | - | .649 |
| 20. DyakB3 | .084 | .069 | .084 | .084 | .102 | .084 | .077 | .084 | .084 | .060 | .100 | .100 | .026 | .032 | .452 | .120 | .053 | .660 | .517 | - |
| 21. Ztub1 | .917 | .877 | .917 | .917 | .877 | .917 | .732 | .917 | .917 | .693 | .879 | .881 | .100 | .076 | .306 | .694 | .242 | .190 | .231 | .215 |
| 22. Ztub2 | .345 | .466 | .345 | .345 | .466 | .345 | .326 | .345 | .345 | .601 | .464 | .464 | .045 | .037 | .362 | .600 | .100 | .268 | .324 | .301 |
| 23. Ztub3 | .928 | .928 | .928 | .928 | .864 | .928 | .745 | .928 | .928 | .679 | .866 | .868 | .102 | .114 | .175 | .680 | .246 | .166 | .204 | .129 |
| 24. Zcam1 | .028 | .022 | .028 | .028 | .036 | .028 | .028 | .028 | .028 | .045 | .036 | .036 | .006 | .007 | .705 | .045 | .011 | .416 | .942 | .780 |
| 25. Zcam2 | .035 | .028 | .035 | .035 | .045 | .035 | .034 | .035 | .035 | .055 | .044 | .044 | .007 | .009 | .605 | .055 | .014 | .514 | .805 | .653 |
| 26. Zcam4 | .036 | .029 | .036 | .036 | .046 | .036 | .035 | .036 | .036 | .057 | .046 | .045 | .007 | .009 | .835 | .057 | .014 | .523 | .790 | .843 |
| 27. Zdav1 | .111 | .090 | .111 | .111 | .135 | .111 | .102 | .111 | .111 | .162 | .134 | .133 | .020 | .025 | .423 | .161 | .042 | .873 | .101 | .877 |
| 28. Zdav2 | .058 | .046 | .058 | .058 | .073 | .058 | .055 | .058 | .058 | .090 | .072 | .072 | .011 | .014 | .508 | .089 | .022 | .492 | .281 | .829 |
| 29. Zdav4 | .072 | .057 | .072 | .072 | .090 | .072 | .068 | .072 | .072 | .110 | .089 | .089 | .013 | .017 | .423 | .109 | .027 | .611 | .204 | .707 |
| 30. Zgab1 | .058 | .046 | .058 | .058 | .073 | .058 | .055 | .058 | .058 | .090 | .072 | .072 | .011 | .014 | .508 | .089 | .022 | .492 | .281 | .829 |
| 31. Zgab3 | .035 | .028 | .035 | .035 | .044 | .035 | .034 | .035 | .035 | .055 | .044 | .044 | .007 | .009 | .881 | .055 | .014 | .272 | .806 | .765 |
| 32. Zgab5 | .051 | .040 | .051 | .051 | .064 | .051 | .049 | .051 | .051 | .079 | .063 | .063 | .010 | .015 | .637 | .078 | .019 | .615 | .201 | .699 |
| 33. Zafr1 | .134 | .110 | .134 | .134 | .162 | .134 | .122 | .134 | .134 | .193 | .109 | .160 | .024 | .030 | .457 | .192 | .052 | .995 | .071 | .393 |
| 34. Zafr3 | .112 | .091 | .112 | .112 | .136 | .112 | .103 | .112 | .112 | .164 | .090 | .135 | .020 | .025 | .457 | .163 | .043 | .878 | .100 | .475 |
| 35. Zafr5 | .058 | .046 | .058 | .058 | .073 | .058 | .055 | .058 | .058 | .090 | .072 | .072 | .011 | .014 | .508 | .089 | .022 | .492 | .281 | .829 |
| 36. Zind2 | .120 | .096 | .120 | .120 | .148 | .120 | .111 | .120 | .120 | .180 | .147 | .147 | .021 | .026 | .774 | .179 | .044 | .347 | .672 | .882 |
| 37. Zind3 | .058 | .046 | .058 | .058 | .073 | .058 | .055 | .058 | .058 | .090 | .072 | .072 | .011 | .014 | .508 | .089 | .022 | .492 | .281 | .829 |
| 38. Zind4 | .058 | .046 | .058 | .058 | .073 | .058 | .055 | .058 | .058 | .090 | .072 | .072 | .011 | .014 | .508 | .089 | .022 | .492 | .281 | .829 |

**Supplementary Table 7, continuation.**

| | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. DmelA1 | -.104 | -.949 | -.090 | -2.221 | -2.128 | -2.116 | -1.608 | -1.915 | -1.814 | -1.915 | -2.129 | -1.971 | -1.509 | -1.603 | -1.915 | -1.567 | -1.915 | -1.915 |
| 2. DmelA3 | .155 | -.732 | -.090 | -2.319 | -2.225 | -2.216 | -1.711 | -2.019 | -1.918 | -2.019 | -2.227 | -2.072 | -1.611 | -1.706 | -2.019 | -1.678 | -2.019 | -2.019 |
| 3. DdmelA4 | -.104 | -.949 | -.090 | -2.221 | -2.128 | -2.116 | -1.608 | -1.915 | -1.814 | -1.915 | -2.129 | -1.971 | -1.509 | -1.603 | -1.915 | -1.567 | -1.915 | -1.915 |
| 4. DsimA1 | -.104 | -.949 | -.090 | -2.221 | -2.128 | -2.116 | -1.608 | -1.915 | -1.814 | -1.915 | -2.129 | -1.971 | -1.509 | -1.603 | -1.915 | -1.567 | -1.915 | -1.915 |
| 5. DsimA2 | .155 | -.732 | .172 | -2.122 | -2.030 | -2.017 | -1.505 | -1.811 | -1.711 | -1.811 | -2.032 | -1.870 | -1.406 | -1.499 | -1.811 | -1.455 | -1.811 | -1.811 |
| 6. DsimA4 | -.104 | -.949 | -.090 | -2.221 | -2.128 | -2.116 | -1.608 | -1.915 | -1.814 | -1.915 | -2.129 | -1.971 | -1.509 | -1.603 | -1.915 | -1.567 | -1.915 | -1.915 |
| 7. DsimA5 | -.343 | -.987 | -.326 | -2.227 | -2.142 | -2.128 | -1.650 | -1.936 | -1.843 | -1.936 | -2.143 | -1.991 | -1.558 | -1.645 | -1.936 | -1.606 | -1.936 | -1.936 |
| 8. DsecA1 | -.104 | -.949 | -.090 | -2.221 | -2.128 | -2.116 | -1.608 | -1.915 | -1.814 | -1.915 | -2.129 | -1.971 | -1.509 | -1.603 | -1.915 | -1.567 | -1.915 | -1.915 |
| 9. DsecA2 | -.104 | -.949 | -.090 | -2.221 | -2.128 | -2.116 | -1.608 | -1.915 | -1.814 | -1.915 | -2.129 | -1.971 | -1.509 | -1.603 | -1.915 | -1.567 | -1.915 | -1.915 |
| 10. DsecA5 | .396 | -.525 | .415 | -2.027 | -1.936 | -1.922 | -1.407 | -1.711 | -1.612 | -1.711 | -1.938 | -1.773 | -1.309 | -1.401 | -1.711 | -1.349 | -1.711 | -1.711 |
| 11. DsecA7 | .152 | -.734 | .169 | -2.125 | -2.033 | -2.020 | -1.509 | -1.814 | -1.715 | -1.814 | -2.035 | -1.874 | -1.616 | -1.710 | -1.814 | -1.459 | -1.814 | -1.814 |
| 12. DsecA8 | .150 | -.735 | .167 | -2.126 | -2.035 | -2.022 | -1.511 | -1.816 | -1.717 | -1.816 | -2.037 | -1.875 | -1.414 | -1.506 | -1.816 | -1.461 | -1.816 | -1.816 |
| 13. DsecA11 | -1.658 | -2.026 | -1.648 | -2.825 | -2.753 | -2.743 | -2.356 | -2.595 | -2.517 | -2.595 | -2.754 | -2.634 | -2.278 | -2.351 | -2.595 | -2.339 | -2.595 | -2.595 |
| 14. DsecA19 | -1.789 | -2.108 | -1.590 | -2.733 | -2.664 | -2.651 | -2.272 | -2.504 | -2.429 | -2.504 | -2.665 | -2.463 | -2.196 | -2.266 | -2.504 | -2.248 | -2.504 | -2.504 |
| 15. DsecA20 | -1.028 | -.915 | -1.365 | .830 | .518 | .208 | .804 | .665 | .804 | .665 | .150 | .473 | .746 | .746 | .665 | .287 | .665 | .665 |
| 16. DsecB8 | .394 | -.526 | .413 | -2.029 | -1.938 | -1.924 | -1.409 | -1.713 | -1.614 | -1.713 | -1.940 | -1.775 | -1.312 | -1.404 | -1.713 | -1.351 | -1.713 | -1.713 |
| 17. DsecB19 | -1.176 | -1.656 | -1.166 | -2.586 | -2.504 | -2.494 | -2.054 | -2.325 | -2.236 | -2.325 | -2.505 | -2.370 | -1.966 | -2.050 | -2.325 | -2.032 | -2.325 | -2.325 |
| 18. DyakA1 | -1.318 | -1.114 | -1.395 | -.817 | -.655 | -.640 | -.160 | -.690 | -.511 | -.690 | -1.104 | -.504 | .006 | -.154 | -.690 | -.944 | -.690 | -.690 |
| 19. DyakA2 | -1.203 | -.990 | -1.278 | .072 | .248 | .267 | 1.651 | 1.082 | 1.277 | 1.082 | .246 | 1.287 | 1.824 | 1.659 | 1.082 | .425 | 1.082 | 1.082 |
| 20. DyakB3 | -1.248 | -1.038 | -1.527 | .280 | .451 | .199 | -.155 | .217 | .376 | .217 | -.299 | .388 | .858 | .717 | .217 | -.149 | .217 | .217 |
| 21. Ztub1 | - | .405 | 2.438 | -1.717 | -1.622 | -1.606 | -1.047 | -1.367 | -1.263 | -1.367 | -1.624 | -1.442 | -.946 | -1.042 | -1.367 | -1.523 | -1.367 | -1.367 |
| 22. Ztub2 | .686 | - | .424 | -1.552 | -1.447 | -1.431 | -.815 | -1.166 | -1.051 | -1.166 | -1.449 | -1.250 | -.702 | -.808 | -1.166 | -1.338 | -1.166 | -1.166 |
| 23. Ztub3 | .016 | .672 | - | -1.802 | -1.705 | -1.901 | -1.322 | -1.455 | -1.349 | -1.455 | -1.707 | -1.527 | -1.134 | -1.233 | -1.455 | -1.603 | -1.455 | -1.455 |
| 24. Zcam1 | .089 | .123 | .074 | - | .999 | .999 | .645 | -.092 | .167 | -.092 | -.736 | .174 | 2.630 | 2.437 | -.092 | -.528 | -.092 | -.092 |
| 25. Zcam2 | .107 | .150 | .091 | .320 | - | 1.412 | .857 | .164 | .405 | .164 | -.534 | .413 | 2.809 | 2.630 | .164 | -.333 | .164 | .164 |
| 26. Zcam4 | .111 | .155 | .060 | .320 | .161 | - | .317 | .177 | .421 | .177 | -.522 | .428 | 2.809 | 2.630 | .177 | -.319 | .177 | .177 |
| 27. Zdav1 | .297 | .417 | .189 | .520 | .393 | .752 | - | 1.728 | 1.412 | 1.728 | -.101 | 1.994 | 2.630 | 2.437 | 1.728 | .159 | 1.728 | 1.728 |
| 28. Zdav2 | .174 | .246 | .148 | .927 | .870 | .860 | .087 | - | .999 | .000 | -.997 | .999 | 1.993 | 1.728 | .000 | -.675 | .000 | .000 |
| 29. Zdav4 | .209 | .295 | .180 | .868 | .686 | .675 | .161 | .320 | - | .999 | -.676 | 1.412 | 2.227 | 1.993 | .999 | -.379 | .999 | .999 |
| 30. Zgab1 | .174 | .246 | .148 | .927 | .870 | .860 | .087 | 1.000 | .320 | - | -.997 | .999 | 1.993 | 1.728 | .000 | -.675 | .000 | .000 |
| 31. Zgab3 | .107 | .150 | .090 | .463 | .595 | .603 | .919 | .321 | .500 | .321 | - | -.674 | .156 | -.098 | -.997 | -1.177 | -.997 | -.997 |
| 32. Zgab5 | .152 | .214 | .129 | .862 | .680 | .669 | .048 | .320 | .161 | .320 | .502 | - | 2.227 | 1.993 | .999 | -.375 | .999 | .999 |
| 33. Zafr1 | .346 | .484 | .259 | .010 | .006 | .006 | .010 | .048 | .028 | .048 | .876 | .028 | - | .999 | 1.993 | .400 | 1.993 | 1.993 |
| 34. Zafr3 | .300 | .421 | .220 | .016 | .010 | .010 | .016 | .087 | .048 | .087 | .922 | .048 | .320 | - | 1.728 | .163 | 1.728 | 1.728 |
| 35. Zafr5 | .174 | .246 | .148 | .927 | .870 | .860 | .087 | 1.000 | .320 | 1.000 | .321 | .320 | .048 | .087 | - | -.675 | .000 | .000 |
| 36. Zind2 | .130 | .183 | .111 | .599 | .740 | .750 | .874 | .501 | .706 | .501 | .241 | .709 | .690 | .871 | .501 | - | -.675 | -.675 |
| 37. Zind3 | .174 | .246 | .148 | .927 | .870 | .860 | .087 | 1.000 | .320 | 1.000 | .321 | .320 | .048 | .087 | 1.000 | .501 | - | .000 |
| 38. Zind4 | .174 | .246 | .148 | .927 | .870 | .860 | .087 | 1.000 | .320 | 1.000 | .321 | .320 | .048 | .087 | 1.000 | .501 | 1.000 | - |

Symbols for species names: Ztub1: *Z. tuberculatus*; Zcam: *Z. camerounensis*; Zdav: *Z. davidi*; Zgab: *Z. gabonicus*; Zafr: *Z. africanus*; Zind: *Z. indianus*; Dmel: *D. melanogaster*; Dsim: *D. simulans*; Dsec: *D. sechellia*; Dyak: *D. yakuba*.




**Supplementary Figure 1.** Insertion numbers of *gypsy* and *micropia* retrotransposons in *Zaprionus* species. The insertion numbers were indicated in the bottom of the figure. Southern blot analyses were performed using 5 µg of genomic DNA double digested with *BamHI* and *MluI* restriction endonucleases and the random primer [α-P32]dCTP labelled probes derivate from the fragments of *env* and *RNAseH* of *Z. indianus*. The nylon membranes were hybridized at 65 °C and washed twice with 2X SSC and 0.1% SDS and once with 0.2X SSC and 0,1% SDS for 20 minutes at the hybridization temperature. mel: *D. melanogaster*, mul: *Z. multistriatus*, tub: *Z. tuberculatus*, cam: *Z. camerounensis*, dav: *Z. davidi*, gab: *Z. gabonicus*, afr: *Z. africanus*, ind: *Z. indianus*.

# 7 DISCUSSÃO GERAL

## 7 DISCUSSÃO GERAL

Os elementos de transposição têm contribuído para a evolução dos eucariotos por meio da sua atividade nos genomas onde estão inseridos, pela promoção de recombinações cromossômicas e por sua contribuição com a homeostase celular, tanto por fornecerem seqüências codificantes e regulatórias (GOTEA; MAKALOWSKI, 2006; ALMEIDA et al., 2008; LOPES et al., 2008; FESCHOTTE et al., 2008), quanto por participarem da manutenção estrutural dos genomas, devido a sua função telomérica (VILLASANTE et al., 2008), contribuírem com a formação da heterocromatina (KAVI et al., 2008) e exercerem um papel na inativação do cromossomo X, em humanos (LYON et al., 2008; SLOTKIN; MARTIENSSEN, 2007), dentre outros papéis igualmente relevantes.

Além de contribuírem com a evolução dos genomas hospedeiros, os elementos de transposição também estão submetidos a processos evolutivos, como rearranjos mediados pela inserção de outros TEs (BERGMAN et al., 2006), mutações pontuais, deleções e inserções, gerando diversificação nucleotídica (JORDAN; MCDONALD, et al. 1998a, b) e deterioração de seqüências (PETROV; HARTL, 1998; REBOLLO et al., 2008). As forças que moldam a diversificação dos TEs podem ser estocásticas, como também, seletivas (HICKEY, 1982; BROOKFIELD, 1982; LE ROUZIC; CAPY, 2005; LE ROUZIC et al., 2007). Todos esses processos estão relacionados ao seu ciclo de vida nos genomas hospedeiros (PINSKER et al., 2001; ALMEIDA; CARARETO, 2005).

Resumidamente, os elementos de transposição são originados nos genomas hospedeiros por meio de recombinações de seqüências pré-existentes (origem *de novo*) ou pela sua transferência a partir de uma espécie doadora, por introgressão ou mesmo com auxílio de vetores. Após sua origem, essas seqüências podem se espalhar por transmissão vertical e se propagarem por meio da produção de novas inserções, aumentando sua representatividade nos genomas. Nesta fase, ocorre o desenvolvimento de estratégias que possibilitam a coevolução TE – hospedeiro e, a duração desse período depende diretamente da habilidade dos TEs de “escaparem” dos mecanismos de regulação da transposição. Finalmente, inicia a fase de senescência, caracterizada pela redução da mobilidade devido ao aumento do número de inserções defectivas, como pelo estabelecimento de mecanismos regulatórios e epigenéticos dos hospedeiros. Essa última fase pode durar milhões de anos e culminar com a degradação da seqüência do TE, seguida pela eliminação por eventos randômicos.

Os resultados apresentados neste trabalho indicam a presença de elementos de transposição jovens nos genomas das espécies do gênero *Zaprionus*. A recente origem desses elementos foi inferida em face da conservação das seqüências nucleotídicas entre as espécies do gênero, bem como, em relação aos elementos do subgrupo *melanogaster*, pela capacidade de produção de transcritos e o baixo número de inserções. No primeiro capítulo desta Tese foram apresentados os resultados de uma busca por elementos de transposição em *Z. indianus*, bem como em *Drosophila malerkotliana*, espécies de drosofilídeos que invadiram recentemente o território brasileiro (VAL; SENE, 1980; VILELA, 1999). Este trabalho constituiu uma abordagem inicial para servir como guia em estudos posteriores, como aqueles apresentados nos demais capítulos da Tese. Dentre os TEs analisados em *Z. indianus*, apenas o retrotransposon *gypsy* havia sido previamente identificado (HEREDIA et al., 2004), sendo que a identificação dos retrotransposons *copia*, *412*, *297*, *roo/B104*, *mdg-1* e *micropia*, dos retroposons *doc* e *jockey* e, do transposon *Bari-1* constitui dados inéditos. Quanto à análise de *D. malerkotliana*, a presença dos retroelementos *copia* e *412* corroboraram estudos anteriores (CIZERON et al., 1998), mas a ocorrência dos retrotransposons *297*, *roo/B104*, *mdg-1*, *gypsy* e *micropia*, dos retroposons *doc* e *jockey* e, do transposon *Bari-1* corresponde ao primeiro relato da presença desses TEs na espécie.

Dentre os TEs identificados em *Z. indianus*, os retrotransposons *copia*, *gypsy* e *micropia* foram selecionados para análises mais detalhadas com outras espécies pertencentes ao gênero *Zaprionus*, devido à facilidade de amplificação de suas seqüências com os oligonucleotídeos iniciadores disponíveis na literatura. Dessa forma, foram isoladas seqüências que posteriormente foram analisadas quanto à sua diversidade e atividade transcricional. Durante o estudo do retrotransposon *copia* no gênero *Zaprionus*, nós nos deparamos com uma significativa quantidade de informações relacionadas à distribuição, dinâmica de mobilização e evolução desse retrotransposon nas espécies de *Drosophila*, mais especificamente em *D. melanogaster*, que ainda não haviam sido compilados. Assim, consideramos útil reuni-los em um artigo de revisão, que foi apresentado no Capítulo 2.

O retrotransposon *copia* foi um dos primeiros elementos de transposição estudados em *D. melanogaster* (FINNEGAN et al., 1978). Sua recorrente análise em diversos estudos posteriores foi estimulada principalmente pela observação de que ele causava mutações observáveis em nível fenotípico (O'HARE et al., 1984; LEVIS et al., 1984). Atualmente sabe-se que *copia* é regulado por mecanismos multifatoriais, intrínsecos de suas seqüências, como número de inserções, conservação de seqüências codificantes e diferenciação de sítios regulatórios mais eficazes (JORDAN et al., 1998a, b) e expressão tecido e idade-específica

(NUZHDIN et al., 1996; PASYUKOVA et al., 1997), bem como, por mecanismos dependentes do hospedeiro, dentre eles a utilização de homeoproteínas para a regulação da transcrição (CAVAREC; HEIDMANN, 1993; BHADRA et al., 1999; CAVAREC et al., 1994; TULIN et al., 2002) e as vias de RNA de interferência, que promovem um equilíbrio na quantidade de transcritos produzidos (STAPLETON et al., 2001; ARAVIN et al., 2003; KALMYKOVA et al., 2005; VAGIN et al., 2006; KAWAMURA et al., 2008). Uma dinâmica de transposição regulada por fatores do TE e do hospedeiro pode indicar uma relação de competição ou mesmo de co-evolução (LE ROUZIC et al., 2007), que consistiria da diversificação de mecanismos de regulação por parte do hospedeiro e, conseqüente diferenciação de inserções com transposição mais eficaz por parte do TE. Parece que *copia* está na dianteira nessa corrida, desde que ainda se mantém ativo e capaz de produzir novas inserções em linhagens de *D. melanogaster* (NUZHDIN et al., 1996; PASYUKOVA et al., 1997).

Outro aspecto que foi discutido na revisão sobre o retrotransposon *copia* foi sua idade na família Drosophilidae. De acordo com os dados disponíveis, *copia* deveria estar presente no ancestral dessa família devido a sua ampla distribuição (MARTIN et al., 1983; DE FRUTOS et al., 1992; FRANCINO et al., 1993; JORDAN; MCDONALD, 1998a; JORDAN et al., 1999; ALMEIDA; CARARETO, 2006; CASTRO et al., 2006). Adicionalmente, sua ausência em determinados grupos de espécies, como por exemplo, nos gêneros *Idyomia* e *Chymomyza* (MARTIN et al., 1983), poderia ser um reflexo de sua perda nos ancestrais dessas espécies. Processos como a aquisição ou perda de proteínas regulatórias do hospedeiro, altas taxas de mutação ou ainda, eventos randômicos acontecendo individual ou simultaneamente, poderiam ter promovido a eliminação de *copia*. Até o momento foi identificado apenas um fator que poderia limitar a manutenção de *copia* canônico nos genomas hospedeiros, a existência de uma proteína identificada em *D. hydei* que tem a capacidade de amplificar extremamente seus níveis de expressão, o que poderia levar a efeitos deletérios devido a produção descontrolada de novas inserções (CAVAREC et al., 1994). Assim, a revisão propõe que estudos evolutivos sobre *copia* ainda devam ser realizados, focando tanto os mecanismos que promovem sua distribuição e diferenciação em Drosophilidae, como ampliando o espectro de espécies estudadas.

O terceiro artigo apresentado nesta Tese consistiu de um estudo sobre a identificação, atividade transcricional, caracterização da região regulatória e análise filogenética do retrotransposon *copia* em sete espécies do gênero *Zaprionus*. O retrotransposon *copia* foi identificado em todas as espécies analisadas por meio do sequenciamento de duas regiões

diferenciadas, a região 5’ LTR-ULR, de cerca de 400 pb, que abriga os motivos regulatórios promotores da transcrição e da tradução, e um fragmento do gene da transcriptase reversa de cerca de 250 pb. Adicionalmente, a análise do perfil transcricional do gene da transcriptase reversa permitiu sugerir que *copia* é transcricionalmente ativo em todas as espécies estudadas.

Além da identificação de *copia* em *Zaprionus*, o sequenciamento da região 5’ LTR-ULR permitiu a caracterização dos motivos regulatórios e a classificação dos retrotransposons *copia* do gênero *Zaprionus* como uma nova subfamília, denominada *GBFDouble-gap*. Basicamente, essa região é estruturalmente similar à região homóloga da subfamília *copia* mais antiga do grupo *melanogaster*, *Double-gap*. Além disso, foi possível identificar um novo fator de ligação de proteínas regulatórias (*GBF-1*) que se apresenta como cópia única nas espécies *Z. tuberculatus* e *Zaprionus camerounensis*, mas está duplicado nas espécies do complexo *indianus* (*Z. africanus, Z. gabonicus* e *Z. indianus*). A duplicação de sítios de ligação de proteínas regulatórias é um fenômeno que já havia sido observado para o retrotransposon *copia* do subgrupo *melanogaster*, devido à duplicação de duas seqüências que juntas, se assemelham ao centro da região intensificadora da transcrição do vírus SV40 (MATYUNINA et al., 1996). Essa duplicação é uma característica derivada da subfamília *copia Full-length* do subgrupo *melanogaster* e propicia um aumento da capacidade transcricional, em detrimento das outras duas subfamílias mais antigas.

As análises evolutivas com as seqüências *copia* do gênero *Zaprionus* produziram resultados inéditos. Até o presente estudo havia sido disponibilizada apenas uma seqüência *copia* do gênero *Zaprionus*, com cerca de 400 pb da região 5’ LTR-ULR de *Z. tuberculatus*, que pertencia a uma das subfamílias do subgrupo *melanogaster*, *ULR-gap* (MCDONALD et al., 1997). Essa seqüência havia sido utilizada na proposição de um evento de transferência horizontal entre *Z. tuberculatus* e uma espécie desconhecida do subgrupo *melanogaster* ou *D. willistoni* (ALMEIDA; CARARETO, 2006). Nossas análises não permitiram identificar qualquer elemento similar ao publicado anteriormente, desde que todas as seqüências produzidas neste estudo são pertencentes à subfamília *GBFDouble-gap*. Todavia, como Jordan e McDonald (1998b) não descrevem a linhagem analisada em seu estudo, foi impossível replicar biologicamente aquela análise.

A busca por seqüências homólogas ao retrotransposon *copia* do gênero *Zaprionus* na base de dados dos 12 genomas de *Drosophila* resultou em “*matches*” apenas para as espécies do subgrupo *melanogaster*. Além dos dados genômicos, as seqüências do grupo *repleta* depositadas no GenBank por Almeida e Carareto (2006) também foram incluídas nas análises filogenéticas e de distância nucleotídica. Os resultados obtidos permitiram sugerir que os

elementos de *Zaprionus* são mais relacionados às espécies do subgrupo *melanogaster*, do que às espécies do grupo *repleta*. Esse resultado, aliado às análises de restrições seletivas, a comparação de sítios neutros do retrotransposon com um gene do hospedeiro (*Gpdh*), os cálculos de tempo de divergência entre os elementos e os dados de que as espécies do gênero *Zaprionus* são mais relacionadas às do subgênero *Drosophila* (grupo *repleta*) do que às do subgênero *Sophophora* (subgrupo *melanogaster*) (PÉLANDAKIS; SOLIGNAC, 1993; THOMAS; HUNT, 1993; KWIATOWSKI et al. 1994; REMSEN; DE SALLE, 1998; RUSSO et al., 1995; TATARENKOV et al. 1999; KWIATOWSKI; AYALA, 1999; ROBE et al., 2005; DA LAGE et al., 2007; YASSIN et al., 2008a) possibilitaram propor a ocorrência de um evento de transferência horizontal a partir de um ancestral do gênero *Zaprionus* para o ancestral das espécies *D. melanogaster*, *D. simulans*, *D. sechellia* e *D. yakuba*, há cerca de oito milhões de anos.

O último capítulo apresentado nesta Tese corresponde a um estudo comparativo da distribuição, atividade transcricional e relações filogenéticas dos retrotransposons *gypsy* e *micropia* em sete espécies do gênero *Zaprionus*. *Gypsy* e *micropia* se apresentaram amplamente distribuídos e transcricionalmente ativos nas seis espécies do subgênero *Zaprionus* e ausentes na única espécie do subgênero *Anaprionus* estudada, *Z. multistriatus*. As análises das seqüências do gene *envelope* de *gypsy* e da *RNAseH* de *micropia* permitiram estabelecer as relações filogenéticas entre essas seqüências e as do subgrupo *melanogaster*, único grupo de espécies que apresentaram seqüências altamente relacionadas para ambos retrotransposons, dentre as 12 espécies de *Drosophila* com o genoma seqüenciado. A alta similaridade das seqüências dos dois retrotransposons entre as espécies do subgênero *Zaprionus* e subgrupo *melanogaster*, aliada às diversas incongruências obtidas durante as reconstruções filogenéticas, indicaram a necessidade de análises complementares como testes de seleção e reconstrução filogenética por *Network* para formular uma hipótese de transmissão entre as espécies.

A análise do retrotransposon *gypsy* confirmou a proposição de que duas subfamílias estão distribuídas entre as espécies do subgrupo *melanogaster* (HEREDIA et al., 2004; LUDWIG et al., 2008). Além disso, permitiu incluir os elementos do subgênero *Zaprionus* nessas duas subfamílias e propor que a Subfamília 2 pode ser subdividida em duas variantes (V1 e V2). As seqüências de *Z. tuberculatus* foram colocadas nas Subfamília 1 e Subfamília 2, V1; todas as seqüências de *Z. camerounensis* pertencem à V1; os elementos de *Zaprionus davidi* correspondem às duas variantes da Subfamília 2, e; todas as seqüências do complexo *indianus* foram classificadas como constituintes da Subfamília 2 – V2. Juntamente com as

reconstruções filogenéticas, as demais análises evolutivas permitiram corroborar três eventos de transferência horizontal propostos por Ludwig et al. (2008) e inferir mais seis eventos para explicar a atual distribuição de *gypsy* no subgênero *Zaprionus* e no subgrupo *melanogaster*. Além disso, a proposição da transferência entre *Z. indianus* e *D. simulans* (HEREDIA et al., 2004) pôde ser reformulada, abrangendo o ancestral das espécies do complexo *indianus* e o ancestral das espécies *D. melanogaster*, *D. simulans* e *D. sechellia*, reafirmando a proposta de Loreto et al. (2008), sobre a importância da análise de um maior número possível de espécies para a validação de transferências horizontais.

Os resultados das análises filogenéticas e de distância referentes ao retrotransposon *micropia* demonstraram que todas as seqüências do subgênero *Zaprionus* e subgrupo *melanogaster* pertencem a uma mesma subfamília, diferente daquela na qual estão classificados os elementos do grupo *repleta*. As análises evolutivas, dentre elas, a reconstrução filogenética por *Network*, permitiram sugerir a ocorrência de oito eventos de transferência horizontal, entre um doador desconhecido, as espécies do subgênero *Zaprionus* e as do subgrupo *melanogaster*. A estimativa dos tempos de divergência das seqüências *gypsy* e *micropia* demonstraram que esses retrotransposons cruzaram a barreira das espécies em um período entre 500 mil e sete milhões de anos, que pode ter sido caracterizado por ondas de invasão de TEs. No caso de *gypsy*, esses eventos devem ter ocorrido principalmente no sentido subgrupo *melanogaster* para gênero *Zaprionus* e, no caso do *micropia*, principalmente de um doador desconhecido para as espécies do subgrupo *melanogaster* e do gênero *Zaprionus*, mas também do subgrupo *melanogaster* para o gênero *Zaprionus* e entre espécies do gênero *Zaprionus*.

Nos últimos anos, vários trabalhos têm sido publicados demonstrando a ocorrência de transferência horizontal em invertebrados (LORETO et al., 2008), em vertebrados (PACE et al., 2008) e até em plantas (ROULIN et al., 2008; CHENG et al., 2009), organismos em que até pouco tempo, a ocorrência de transferência horizontal não tinha sido demonstrada. Loreto et al. (2008) discutiram os requisitos necessários para a inferência de eventos de transferência horizontal. Para que duas espécies possam trocar seqüências de DNA, no mínimo, elas têm que compartilhar a mesma distribuição geográfica, temporal e ecológica. Além disso, é necessário que as hipóteses alternativas, como por exemplo, polimorfismo ancestral e altas taxas de conservação propiciadas por seleção purificadora, sejam descartadas, ou pelo menos constituam alternativas menos parcimoniosas. Outro fator que poderia confundir uma análise de distância seria a proposição de transferência horizontal usando como base de comparação seqüências de diferentes famílias de TEs, o que levaria a uma superestimativa das distâncias

entre elementos da mesma família. Finalmente, esses autores ressaltam que o número de espécies amostrado é muito importante, pois algumas inferências robustas podem ser refutadas com o aumento do número de espécies analisadas.

Os eventos de transferência horizontal propostos nesta Tese só foram assumidos quando mais de uma análise corroborava sua proposição. As espécies do gênero *Zaprionus* e do subgrupo *melanogaster* se originaram na região biogeográfica Oriental e se diversificaram na região Afrotropical na mesma época (LACHAISE et al., 2004; YASSIN et al., 2008a). Adicionalmente, embora apenas sete das 55 espécies do gênero *Zaprionus* tenham sido estudadas, elas representam seus dois subgêneros e os dois grupos de espécies do subgênero *Zaprionus*; além disso, das nove espécies do subgrupo *melanogaster*, cinco tiveram seus genomas seqüenciados, o que torna nossa amostragem ao menos razoável.

Levando em consideração os requisitos mínimos para a proposição de transferência horizontal, nossos resultados permitiram inferir de 18 eventos. A hipótese alternativa de polimorfismo ancestral foi considerada menos provável nesses casos devido à demonstração de que os sítios sinônimos dos TEs, aparentemente neutros, são mais conservados dos que os mesmos em genes hospedeiros, fundamentais para o metabolismo, e conseqüentemente submetidos a altas restrições seletivas. Por sua vez, a ausência de forte seleção purificadora nos TEs foi demonstrada pela análise da variabilidade nos sítios sinônimos e não-sinônimos. No entanto, a ocorrência de algum tipo de pressão seletiva que estivesse conservando as seqüências dos TEs como um todo, incluindo os sítios sinônimos, aparentemente neutros, poderia justificar a alta conservação dessas seqüências.

Diversos estudos têm demonstrado a importância dos elementos de transposição para a regulação do genoma hospedeiro (FESCHOTTE et al., 2008), para a conservação estrutural da heterocromatina (KAVI et al., 2008), dos telômeros (VILLASANTE et al., 2008) e mesmo para sua auto-regulação, como demonstrado para os locos de piRNAs, formados por blocos de TEs defectivos, presentes na heterocromatina de *D. melanogaster*, que funcionam como molde para a produção de RNAs pequenos, mediando a inativação dos elementos transponíveis completos localizados na eucromatina (BRENNECK et al., 2007). Existem dois estudos que descrevem a ocorrência de seleção purificadora mantendo seqüências de TEs em eucariotos. No primeiro, a análise de seqüências ortólogas de homem e camundongo sugeriu que fragmentos dos elementos MIR e L2 foram submetidos a altas taxas de seleção purificadora, provavelmente devido à aquisição de funções regulatórias nos genomas hospedeiros (SILVA et al., 2003). No segundo, foi demonstrado que cerca de 10 mil fragmentos de TEs não-exônicos estão sofrendo seleção purificadora no genoma hospedeiro.

Desde que essas seqüências estão localizadas em desertos gênicos próximos e apresentam uma preferência por regiões próximas à genes de adesão celular, regulação da transcrição ou desenvolvimento, os autores propõem que elas podem estar desempenhando um papel na especialização dessas vias durante a evolução dos mamíferos (LOWE et al., 2007).

Todos os exemplos apresentados confirmam a capacidade do genoma hospedeiro em seqüestrar os elementos de transposição levando ao desempenho de funções específicas, no entanto, em grande parte destes exemplos, a domesticação acontece com não-autônomas (BRENNECK et al., 2007), elementos transponíveis específicos, como TEs teloméricos *HeT-A* e *TART* de *Drosophila* (CASACUBERTA; PARDUE, 2005) ou em cassetes de TEs (GOTEA; MAKALOWSKI, 2006; LOPES et al., 2008; ALMEIDA et al., 2008). Diferente do exemplo apresentado para os fragmentos de TEs em mamíferos, os retrotransposons *HeT-A* e *TART* evoluem a taxas mais aceleradas do que os genes eucromáticos de *D. melanogaster*, *D. yakuba* e *D. virilis* (CASACUBERTA; PARDUE, 2005). Adicionalmente, ainda não existem quaisquer evidências de que os retrotransposons *copia*, *gypsy* e *micropia* do gênero *Zaprionus* e do subgrupo *melanogaster* estejam relacionados a eventos de domesticação. Além disso, se esses elementos participassem de algum mecanismo evolutivamente fixado, seria esperado que ocorresse sua conservação em outros grupos da família Drosophilidae, no entanto, sua conservação é aparentemente pontual, demonstrada por sua variabilidade nas espécies do grupo *repleta* (*copia* e *micropia*), em *D. willistoni* (*gypsy*) e por sua ausência nos demais genomas de *Drosophila* seqüenciados.

Dessa forma, elegemos a hipótese de transferência horizontal como a mais parcimoniosa para explicar os resultados relatados nesta Tese. A pesquisa científica se caracteriza pela formulação de hipóteses e posterior experimentação para refutação ou corroboração das propostas inicialmente formuladas. Como acreditamos que a evolução dos elementos de transposição seja um tema profícuo, que continuará a ser amplamente investigado, a idealização de novas estratégias de pesquisa e o crescimento do corpo de dados poderão revelar se as hipóteses aqui sugeridas realmente refletem a história da diversificação dos retrotransposons *copia*, *micropia* e *gypsy* nos genomas das espécies analisadas.

# 8 CONCLUSÕES

## 8 CONCLUSÕES

O presente trabalho permitiu estabelecer as seguintes conclusões:

1. Os genomas de *Z. indianus* e *D. malerkotliana* contém os dez TEs característicos de *D. melanogaster* investigados por meio da técnica de *Dot blot* (*copia*, *412*, *297*, *roo/B104*, *mdg-1*, *gypsy*, *micropia*, *doc*, *jockey* e *Bari-1*);
2. Foi identificada uma nova subfamília do retrotransposon *copia*, *GBFDouble-gap*, que está transcricionalmente ativa e amplamente distribuída em sete espécies do gênero *Zaprionus* e deve pertencer a família dos elementos do subgrupo *melanogaster*;
3. O mecanismo de duplicação de motivos de ligação de proteínas regulatórias tem repetidamente acontecido durante a evolução da região regulatória (5’ LTR-ULR) do retrotransposon *copia*. Esse mecanismo já havia sido evidenciado nos elementos do subgrupo *melanogaster* de *Drosophila* e foi novamente verificado durante o estudo das seqüências do complexo *indianus* do gênero *Zaprionus*;
4. De acordo com as análises evolutivas, dois eventos de transferência horizontal podem ter moldado a evolução do retrotransposon *copia* entre as espécies do subgênero *Zaprionus* e do subgrupo *melanogaster*;
5. Os retrotransposons *gypsy* e *micropia* podem ser componentes recentes dos genomas das espécies do subgênero *Zaprionus*, devido à sua atividade transcricional, o baixo número de inserções e as relações evolutivas próximas com as espécies do subgrupo *melanogaster*;
6. Durante os últimos sete milhões de anos, ondas de transferência horizontal envolvendo as espécies do subgênero *Zaprionus*, do subgrupo *melanogaster* e provavelmente, doadores desconhecidos, podem ter feito com que os retrotransposons *gypsy* e *micropia* cruzassem a barreira das espécies aos quais estavam confinados;
7. A reconstrução filogenética por meio da análise de *Network* constitui um bom método alternativo para o estudo das relações de transmissão de elementos de transposição altamente similares em sua seqüência nucleotídica;
8. As espécies do gênero *Zaprionus* propiciam uma rica fonte de informações sobre os processos evolutivos aos quais os elementos de transposição estão submetidos.

# REFERÊNCIAS BIBLIOGRÁFICAS

## REFERÊNCIAS BIBLIOGRÁFICAS


ALMEIDA, L. M.; CARARETO, C. M. A. **Origem, Proliferação e Extinção de Elementos Transponíveis: Qual seria a Importância da Transferência Horizontal na Manutenção desses Ciclos?** Ribeirão Preto: Sociedade Brasileira de Genética, 2005.

ALMEIDA, L. M.; CARARETO, C. M. Sequence heterogeneity and phylogenetic relationships between the *copia* retrotransposon in *Drosophila* species of the *repleta* and *melanogaster* groups. **Genetics, Selection, Evolution**, França, v. 38, p. 535-550, 2006.

ALMEIDA, L. M. et al. Report of a chimeric origin of transposable elements in a bovine-coding gene. **Genetics and Molecular Research**, Brasil, v. 1, p. 107-116, 2008.

AQUADRO, C. F. et al. Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. **Genetics**, EUA, v. 114, p. 1165-1190, 1986.

ARAVIN, A. A. et al. The small RNA profile during *Drosophila melanogaster* development. **Developmental Cell**, EUA, v. 5, p. 337-350, 2003.

BERGMAN, C. M. et al. Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. **Genome Biology**, Inglaterra, v. 7, p. R112, 2006.

BHADRA, U.; PAL-BHADRA, M.; BIRCHLER, J. A. Role of the *male specific lethal* (*msl*) genes in modifying the effects of sex chromosomal dosage in *Drosophila*. **Genetics**, EUA, v. 152, p. 249-268, 1999.

BIÉMONT, C.; CIZERON, G. Distribution of transposable elements in *Drosophila* species. **Genetica**, Holanda, v. 105, p. 43-62, 1999.

BRENNECKE, J. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. **Cell**, EUA, v. 128, p. 1089-1103, 2007.

BROOKFIELD, J. F. Interspersed repetitive DNA sequences are unlikely to be parasitic. **Journal of Theoretical Biology**, Inglaterra, v. 94, p. 281-299, 1982.

BRUNET, F. et al. Phylogenetic analysis of *Mos1*-like transposable elements in the Drosophilidae. **Journal of Molecular Evolution**, Alemanha, v. 49, p. 760-768, 1999.

CASACUBERTA, E.; PARDUE, M. L. *HeT-A* and *TART*, two Drosophila retrotransposons with a bona fide role in chromosome structure for more than 60 million years. Cytogenetic and **Genome Research**, Suíça, v. 110, p. 152-159, 2005.

CASTRO, J. P. DE; DE SETTA, N.; CARARETO, C. M. A. Distribution and insertion numbers of transposable elements in species of the *Drosophila saltans* group. **Genetics and Molecular Biology**, Brasil, v. 29, p. 384-390, 2006.

CAVAREC, L.; HEIDMANN, T. The *Drosophila copia* retrotransposon contains binding sites for transcriptional regulation by homeoproteins. **Nucleic Acids Research**, Inglaterra, v. 21, p. 5041-5049, 1993.

CAVAREC, L.; JENSEN, S.; HEIDMANN, T. Identification of a strong transcriptional activator for the *copia* retrotransposon responsible for its differential expression in *Drosophila hydei* and *melanogaster* cell lines. **Biochemical and Biophysical Research Communications**, EUA, v. 203, p. 392-399, 1994.

CHASSAGNARD, M-T.; TSACAS, L. Le sous-genre *Zaprionus* s. str. Definition de groupes d'espèces et revision du sous-groupe vittiger (Diptera: Drosophilidae). **Annales de la Société entomologique de France**, França, v. 29, p. 173-194, 1993.

CHENG, X. A new family of *Ty1-copia*-like retrotransposon originated in the tomato genomes by a recent horizontal transfer event. **Genetics**, EUA, doi. 10.1534/genetics.108.099150, 2009.

CIZERON, G. et al. Distribution of the retrotransposable element *412* in *Drosophila* species. **Molecular Biology and Evolution**, EUA, v. 15, p. 1589-1599, 1998.

DA LAGE, J-L. et al. A phylogeny of Drosophilidae using the *amyrel* gene: questioning the *Drosophila melanogaster* species group boundaries. **Journal of Zoological Systematics and Evolutionary Research**, Alemanha, v. 45, p. 47-63, 2007.

DALTON, R. A fly by any other name. **Nature**, Inglaterra, v. 457, p. 368, 2009.

DE FRUTOS, R.; PETERSON, K. R.; KIDWELL, M. G. Distribution of *Drosophila melanogaster* transposable element sequences in species of the *obscura* group. **Chromosoma**, Alemanha, v. 101, p. 293-300, 1992.

DE SALLE, R. The origin and possible time of divergence of the Hawaiian Drosophilidae: evidence from DNA sequences. **Molecular Biology and Evolution**, EUA, v. 9, p. 905–916, 1992.

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

FINNEGAN, D. J. et al. Repeated gene families in *Drosophila melanogaster*. **Cold Spring Harbor Symposia on Quantitative Biology**, EUA, v. 42, p. 1053-1063, 1978.

FRANCINO, O.; CABRE, O.; FONTDEVILA, A. Distribution of the *copia* transposable element in the *repleta* group of *Drosophila*. **Genetics, Selection, Evolution**, França, v. 25, p. 501-516, 1993.

GOTEA, V.; MAKALOWSKI, W. Do transposable elements really contribute to proteomes? **Trends in Genetics**, Inglaterra, v. 22, p. 260-267, 2006.

GRIMALDI, D. A. A phylogenetic, revised classification of genera in the Drosophilidae (Diptera). **Bulletin of the American Museum of Natural History**, EUA, v. 197, p. 123–128, 1990.

HEREDIA, F. et al. Complex evolution of *gypsy* in drosophilid species. **Molecular Biology and Evolution**, EUA, v. 21, p.1831-1842, 2004.

HICKEY, D. A. Selfish DNA: a sexually-transmitted nuclear parasite. **Genetics**, EUA, v. 101, p. 519-531, 1982.

JORDAN, I. K. et al. Origin of a substantial fraction of human regulatory sequences from transposable elements. **Trends in Genetics**, Inglaterra, v. 19, p. 68-72, 2003.

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

JORDAN, I. K.; MCDONALD, J. F. Evolution of the *copia* retrotransposon in the *Drosophila melanogaster* species subgroup. **Molecular Biology and Evolution**, EUA, v. 15, p. 1160-1171, 1998a.

JORDAN, I. K.; MCDONALD, J. F. Interelement selection in the regulatory region of the *copia* retrotransposon. **Journal of Molecular Evolution**, Alemanha, v. 47, p. 670–676, 1998b.

KALMYKOVA, A. I.; KLENOV, M. S.; GVOZDEV, V. A. *Argonaute* protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. **Nucleic Acids Research**, Inglaterra, v. 33, p. 2052-2059, 2005.

KAMINKER, J. S. et al. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. **Genome Biology**, Inglaterra, v. 3, n. 12, research0084, 2002.

KAVI, H. H. et al. Genetics and biochemistry of RNAi in *Drosophila*. **Current Topics in Microbiology and Immunology**, Alemanha, v. 320, p. 37-75, 2008.

KAWAMURA, Y. et al. *Drosophila* endogenous small RNAs bind to *Argonaute 2* in somatic cells. **Nature**, Inglaterra, v. 453, p. 793-797, 2008.

KIDWELL, M. G.; KIDWELL, J. F.; SVED, J. A. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including, sterility and male recombination. **Genetics**, EUA, v. 86, p. 813-833, 1977.

KIDWELL, M. G.; LISCH, D. Transposable elements as sources of variation in animals and plants. **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 94, p. 7704-7711, 1997.

KWIATOWSKI, J; AYALA, F. J. Phylogeny of *Drosophila* and related genera: conflict between molecular and anatomical analyses. **Molecular Phylogenetics and Evolution**, v. 13, p. 319-328, 1999.

KWIATOWSKI, J. et al. Phylogeny of *Drosophila* and related genera inferred from the nucleotide sequence of the Cu, Zn Sod gene. **Journal of Molecular Evolution**, Alemanha, v. 38, p. 443-454, 1994.

LACHAISE, D. et al. Nine relatives from one African ancestor: population biology and evolution of the *Drosophila melanogaster* subgroup species. In: SINGH, R. S.; UYENOYAMA, M. K. **The Evolution of Population Biology**. Nova Iorque: Cambridge University Press, 2004, p. 315-344.

LACHAISE, D.; SILVAIN, J. F. How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster-D. simulans* palaeogeographic riddle. **Genetica**, Holanda, v. 120, p. 17-39, 2004.

LANGLEY, C.H. et al. On the role of unequal exchange in the containment of transposable element copy number. **Genetical Research**, Inglaterra, v. 52, p. 223-235, 1988.

LE ROUZIC, A.; BOUTIN, T. S.; CAPY, P. Long-term evolution of transposable elements. **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 104, p. 19375-19380, 2007.

LE ROUZIC, A.; CAPY, P. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. **Genetics**, EUA, v. 169, p. 1033-1043, 2005.

LEVIS, R.; O'HARE, K.; RUBIN, G. M. Effects of transposable element insertions on RNA encoded by the *white* gene of *Drosophila*. **Cell**, EUA, v. 38, p. 471-481, 1984.

LOPES, F. R. et al. Transposable elements in *Coffea* (Gentianales: Rubiacea) transcripts and their role in the origin of protein diversity in flowering plants. **Molecular Genetics and Genomics**, Alemanha, v. 279, p. 385-401, 2008.

LORETO, E. L.; CARARETO, C. M.; CAPY, P. Revisiting horizontal transfer of transposable elements in *Drosophila*. **Heredity**, Inglaterra, v. 100, p. 545-54, 2008.

LOWE, C. B.; BEJERANO, G.; HAUSSLER, D. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 104, p. 8005-8010, 2007.

LUDWIG, A.; VALENTE, V. L.; LORETO, E. L. Multiple invasions of Errantivirus in the genus *Drosophila*. **Insect Molecular Biology**, Inglaterra, v. 17, p. 113-24, 2008.

LYON, M. F. *LINE-1* elements and X chromosome inactivation: A function for ‘‘junk’’ DNA? **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 97, p. 6248-6249, 2000.

MARKOW, T. A.; O’GRADY, P. M. ***Drosophila*: A Guide to Species Identification and Use.** Londres: Academic Press, 2005.

MARTIN, G.; WIERNASZ, D.; SCHEDL, P. Evolution of *Drosophila* repetitive-dispersed DNA. **Journal of Molecular Evolution**, Alemanha, v. 19, p. 203-213, 1983.

MARUYAMA, K.; HARTL, D. L. Evolution of the transposable element *mariner* in *Drosophila* species. **Genetics**, EUA, v. 128, p. 319-329, 1991.

MATYUNINA, L. V.; JORDAN, I. K.; MCDONALD, J. F. Naturally occurring variation in *copia* expression is due to both element (cis) and host (trans) regulatory variation. **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 93, p. 7097-7102, 1996.

MCDONALD, J. F. Macroevolution, gene regulation and retroviral-like elements. **Bioscience**, EUA, v. 40**,** p. 183-191, 1990.

MCDONALD, J. F. et al. LTR retrotransposons and the evolution of eukaryotic enhancers. **Genetica**, Holanda, v. 100, p. 3-13, 1997.

MEJLUMIAN, L. et al. Comparative and functional studies of *Drosophila* species invasion by the *gypsy* endogenous retrovirus. **Genetics**, EUA, v. 160, p. 201-209, 2002.

MÉVEL-NINIO, M.; MARIOL, M. C.; GANS, M. Mobilization of the *gypsy* and *copia* retrotransposons in *Drosophila melanogaster* induces reversion of the *ovo* dominant female-sterile mutations: molecular analysis of revertant alleles. **The EMBO Journal**, Inglaterra, v. 8, p. 1549-1558, 1989.

MONTCHAMP-MOREAU, C. et al. Distribution and conservation of sequences homologous to the *1731* retrotransposon in *Drosophila*. **Molecular Biology and Evolution**, EUA, v. 10, p. 791-803, 1993.

MONTGOMERY, E. A. et al. Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. **Genetics**, EUA, v. 129, p. 1085-1098, 1991.

MUGNIER, N.; BIEMONT, C.; VIEIRA, C. New regulatory regions of *Drosophila 412* retrotransposable element generated by recombination. **Molecular Biology and Evolution**, EUA, v. 22, p. 747-757, 2005.

NUZHDIN, S. V.; PASYUKOVA, E. G.; MACKAY, T. F. Positive association between *copia* transposition rate and copy number in *Drosophila melanogaster*. **Proceedings of the Royal Society of London**, Inglaterra, v. 263, p. 823-831, 1996.

O'HARE, K. et al. DNA sequence of the *white* locus of *Drosophila melanogaster*. **Journal of Molecular Biology**, Inglaterra, v. 180, p. 437-455, 1984.

OKADA, T. New taxonomic changes in the family Drosophilidae (Diptera). **Japanese Journal of Entomology**, Japão, v. 58, p. 154, 1990.

OKADA, T.; CARSON, H. L. The genera *Phorticella* Duda and *Zaprionus* Coquillett (Diptera, Drosophilidae) of the Oriental region and New Guinea. **Kontyu**, Japão, v. 51, p. 539-553, 1983.

PACE, J. K. 2nd et al. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 105, p. 17023-17028, 2008.

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

PELANDAKIS, M.; SOLIGNAC, M. Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. **Journal of Molecular Evolution**, Alemanha, v. 37, p. 525-543, 1993.

PETROV, D. A.; HARTL, D. L. High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. **Molecular Biology and Evolution**, EUA, v. 15, p. 293-302, 1998.

PINSKER, W. et al. The evolutionary life history of *P* transposons: from horizontal invaders to domestical neogenes. **Chromosoma**, Alemanha, v. 110, p. 148-158, 2001.

REBOLLO, R. et al. Losing *helena:* The extinction of a *Drosophila LINE*-like element. **BMC Genomics**, Inglaterra, v. 9, p. 149, 2008.

REMSEN, J.; DE SALLE, R. Character congruence of multiple data partitions and the origin of the Hawaiian Drosophilidae. **Molecular Phylogenetics and Evolution**, EUA, v. 9, p. 225-235, 1998.

ROBE, L. J. et al. Molecular phylogeny of the subgenus *Drosophila* (Diptera, Drosophilidae) with an emphasis on Neotropical species and groups: a nuclear versus mitochondrial gene approach. **Molecular Phylogenetics and Evolution**, EUA, v. 36, p. 623-640, 2005.

ROULIN, A. Evidence of multiple horizontal transfers of the long terminal repeat retrotransposon *RIRE1* within the genus Oryza. **The Plant Journal**, Inglaterra, v. 53, p. 950-959, 2008.

RUSSO, C. A. M.; TAKEZAKI, N.; NEI, M. Molecular phylogeny and divergence times of drosophilid species. **Molecular Biology and Evolution**, EUA, v. 12, p. 391-404, 1995.

SILVA, J. C. et al. Conserved fragments of transposable elements in intergenic regions: evidence for widespread recruitment of MIR- and L2-derived sequences within the mouse and human genomes. **Genetical Research**, Inglaterra, v. 82, p. 1-18, 2003.

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

STAPLETON, W.; DAS, S.; MCKEE B. D. A role of the *Drosophila* homeless gene in repression of *Stellate* in male meiosis. **Chromosoma**, Alemanha, v. 110, p. 228-40, 2001.

STRAND, D. J.; MCDONALD, J. F. Insertion of a *copia* element 5' to the *Drosophila melanogaster* alcohol dehydrogenase gene (*adh*) is associated with altered developmental and tissue-specific patterns of expression. **Genetics**, EUA, v. 121, p. 787-794, 1989.

TATARENKOV, A. et al. On the evolution of *Dopa decarboxylase* (*Ddc*) and *Drosophila* systematics. **Journal of Molecular Evolution**, Alemanha, v. 48, p. 445-462, 1999.

THOMAS, R. H.; HUNT, J. A. Phylogenetic relationships in *Drosophila*: a conflict between molecular and morphological data. **Molecular Biology and Evolution**, EUA, v. 10, p. 362-374, 1993.

THROCKMORTON, L. H. The phylogeny, ecology and geography of *Drosophila*. In: KING, R.C. (Org.) **Handbook of genetics**. USA: Plenum, p. 421-469, 1975.

TULIN, A.; STEWART, D.; SPRADLING, A. C. The *Drosophila* heterochromatic gene encoding poly(ADP-ribose) polymerase (*PARP*) is required to modulate chromatin structure during development. **Genes & Development**, EUA, v. 16, p. 2108-2119, 2002.

VAGIN, V. V. et al. A distinct small RNA pathway silences selfish genetic elements in the germline. **Science**, EUA, v. 313, p. 320-324, 2006.

VAL, F. C.; SENE, F. M. A newly introduced *Drosophila* species in Brazil (Diptera, Drosophilidae). **Papéis Avulsos de Zoologia, Museu Zoologia USP – São Paulo**, v. 33, p. 293-298, 1980.

VILELA, C. R. Is *Zaprionus indianus* Gupta, 1970 (Diptera, Drosophilidae) currently colonizing the Neotropical region? **Drosophila Information Service**, EUA, v. 82, p. 37-39, 1999.

VILLASANTE, A. et al. Telomere maintenance in *Drosophila*: rapid transposon evolution at chromosome ends. **Cell Cycle**, EUA, v. 7, p. 2134-2138, 2008.

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

YASSIN, A. et al. Grafting the molecular phylogenetic tree with morphological branches to reconstruct the evolutionary history of the genus *Zaprionus* (Diptera: Drosophilidae). **Molecular Phylogenetics and Evolution**, EUA, v. 47, p. 903-15, 2008a.

YASSIN, A. et al. DNA barcode discovers two cryptic species and two geographical radiations in the invasive drosophilid *Zaprionus indianus*. **Molecular Ecology Notes**, Inglaterra, v. 8, p. 491-501, 2008b.