Juliana Polachini de Castro

Reavaliação da distribuição, diversidade de seqüências, número de cópias e atividade transcricional de elementos transponíveis no grupo *saltans de Drosophila* com ênfase no elemento P

1910001162

Entregue em 06/05/03

unesp UNIVERSIDADE ESTADUAL PAULISTA
Instituto de Biociências, Letras e Ciências Exatas
Programa de Pós-Graduação em Genética

JULIANA POLACHINI DE CASTRO

Reavaliação da distribuição, diversidade de seqüências, número de cópias e atividade transcricional de elementos transponíveis no grupo *saltans* de *Drosophila*, com ênfase no elemento *P*

Inst. de Biociências, Letras e Ciências Exatas - UNESP
BIBLIOTECA
S. J. Rio Preto
t. 1.162/03

Tese apresentada para obtenção do Título de Doutor em Genética.

Orientador: Profa. Dra. Cláudia Márcia A. Carareto
Auxílio Financeiro: FAPESP

São José do Rio Preto, 2003

†575
C355r

20204000

O Programa de Pós-Graduação em Ciências Biológicas, Área de Concentração em Genética, passou a denominar-se Programa de Pós-Graduação em Genética pela Resolução UNESP n° RU. 10/01 de 08/03/2001, DOE de 09/03/2001.

Castro, Juliana Polachini de.
Reavaliação da distribuição, diversidade de seqüências, número de cópias e atividade transcricional de elementos transponíveis no grupo *saltans* de *Drosophila*, com ênfase no elemento *P* / Juliana Polachini de Castro. – São José do Rio Preto : [s.n.], 2003
147 f. : il. ; 30 cm.

Orientador: Cláudia Márcia Aparecida Carareto
Tese (doutorado) – Universidade Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas

1. Drosofila. 2. Drosofila - Elemento transponível. I. Carareto, Cláudia Márcia Aparecida . II. Universidade Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas. III. Título.

CDU – 575

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

JULIANA POLACHINI DE CASTRO

REAVALIAÇÃO DA DISTRIBUIÇÃO, DIVERSIDADE DE SEQÜÊNCIAS, NÚMERO DE CÓPIAS E ATIVIDADE TRANSCRICIONAL DE ELEMENTOS TRANSPONÍVEIS NO GRUPO *saltans* DE *Drosophila*, COM ÊNFASE NO ELEMENTO *P*

COMISSÃO JULGADORA

TESE PARA OBTENÇÃO DO TÍTULO DE DOUTOR

Presidente e Orientador .....

2° Examinador .....

3° Examinador .....

4° Examinador .....

5° Examinador .....

São José do Rio Preto, 23 de abril de 2003

À minha orientadora Cláudia Carareto, que uma vez me presenteou com os seguintes dizeres: "Na Ciência, como na vida, o fruto vem sempre depois do amor". Cláudia, aqui estão alguns dos frutos que nasceram da nossa dedicação e amor à Ciência. Dedico-os a você, minha querida mestra. Que estes sejam apenas os primeiros...

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

AGRADECIMENTOS

Ao término de um trabalho, sempre nos lembramos dos momentos bons, bem como dos momentos difíceis pelos quais passamos durante a sua realização. Em todos estes momentos, algumas pessoas nos ajudaram, apoiaram, torceram por nós e até mesmo comemoraram conosco cada conquista. A todas estas pessoas, expresso a minha gratidão. E em especial agradeço:

À minha admirada orientadora, Cláudia Carareto, pelo seu entusiasmo, dedicação e amor à Ciência, pelo exemplo de seriedade, ética e profissionalismo, por confiar no meu trabalho e investir tanto na minha formação. Além da honra de ter sido sua orientada, você ainda me presenteou com a sua amizade, e eu te agradeço por todos estes anos de harmoniosa convivência.

Ao Instituto de Biociências, Letras e Ciências Exatas, por ter me acolhido nestes últimos dez anos.

À FAPESP, pelo auxílio financeiro concedido e por ter me possibilitado realizar parte deste trabalho nos Estados Unidos.

À Dra. Margaret Kidwell, por ter me recebido de forma tão carinhosa em seu laboratório na Universidade do Arizona - USA.

Ao meu eterno mestre, Andrew Holyoake, que tanto me ensinou em tão pouco tempo, por ter me proporcionado um ambiente de trabalho tão cheio de alegria e cumplicidade. Você me ajudou a entender que depois que nos deixamos contagiar pela Ciência, nem sequer nos questionamos se a forma como vivemos é certa ou errada.

À minha amiga Elizabete Amaral, pelo incentivo constante, pela cumplicidade, pelas longas conversas e pela amizade que me dedicou desde o nosso primeiro encontro.

Aos professores do Programa de Pós-graduação em Genética pelo aprendizado e incentivo constantes, em especial: Lílian Madi, Eloiza Tajara, Paula Rahal, Osmar Gaspar, Carlos Ceron e Hermione Bicudo.

Aos funcionários do Departamento de Biologia, em especial ao nosso Auxiliar Acadêmico Josué, pelo convívio e ajuda diários; à nossa secretária Simone, pela presteza na realização de suas tarefas, e aos cozinheiros preferidos das minhas "filhas", Paulo e Tião, pela confecção dos meios de cultura.

Às companheiras de laboratório Luciane e Flávia por agüentarem meu autoritarismo e tagarelice diários, pela amizade e companheirismo. Gostaria de agradecer em especial à Luciane por ter cuidado das minhas "filhas" quando estive ausente.

A todos os colegas do Programa de Pós-Graduação pela convivência, e em especial aos amigos Felipe, Luciana, Priscila e Marcos, pela amizade que construímos ao longo desses anos.

À minha amiga Silvia C. Nascimento e toda a sua família, por fazerem parte da minha vida e me apoiarem em todos os momentos.

À minha querida irmã Lívia, um presente que Deus me enviou. Minha irmã acima de tudo e de todos e sabe o valor e a importância de se ter em um irmão um grande amigo. Obrigada por tudo que você é, pela sua alegria constante, e pela competência com que "gerenciou" a minha vida enquanto estive fora do Brasil.

Aos meus pais Elza e Raul, por me ensinarem a ser batalhadora e a ter dignidade. Com vocês, aprendi que devemos enfrentar o mundo de cabeça erguida e desejar somente o que podemos obter com o fruto do nosso trabalho. Traduzo aqui, com as palavras de Madre Tereza de Calcutá, tudo o que vocês me ensinaram nestes 28 anos:

"Muitas vezes as pessoas são egocêntricas, ilógicas e insensatas - Perdoe-as assim mesmo. Se você é gentil, as pessoas podem acusá-lo de egoísta, interesseiro - Seja gentil assim mesmo. Se você é um vencedor, terá alguns falsos amigos e alguns inimigos verdadeiros - Vença assim mesmo. Se você é honesto e franco, as pessoas podem enganá-lo - Seja honesto assim mesmo. O que você levou anos para construir, alguém pode destruir de uma hora para outra - Construa assim mesmo. Se você tem paz e é feliz, as pessoas podem sentir inveja - Seja feliz assim mesmo. Dê ao mundo o melhor de você, mas isso pode nunca ser o bastante - Dê o melhor de você assim mesmo. Veja você que no final das contas, é entre você e Deus. Nunca foi entre você e as outras pessoas".

"In research you really have to love and be committed to your work, because things have more of a chance of going wrong than right. But when things go right, there is nothing more exciting".

Michael Smith

Prêmio Nobel de Química em 1993

SUMÁRIO

Capítulo 1 - Introdução 8

I. O grupo *saltans* de *Drosophila* 8
II. Os elementos transponíveis 9
II.1. Origem e dinâmica no genoma hospedeiro 12
II.2. Papel na produção de variabilidade genética 15
II.3. Caracterização dos TEs analisados neste estudo 18
II.4. Distribuição heterogênea dos TEs no grupo *saltans* 25
III. Objetivos 26

Capítulo 2 - Mecanismos de Transposição e Regulação do elemento transponível *P* 28

Capítulo 3 - Elemento *P* no grupo *saltans*: uma reavaliação 56

Capítulo 4 - Caracterização de seqüências *P* completas em *D. sturtevanti* e *D. prosaltans* 77

Capítulo 5 -Transcrição de elementos *P* canônicos no grupo *saltans* 89

Capítulo 6 - Distribuição e número de cópias de TEs no grupo *saltans* 114

Capítulo 7 - Discussão 127

Capítulo 8 - Conclusões 135

Referências Bibliográficas 137

Resumo 146

Abstract 147

Introdução

Capítulo Um

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

I. O grupo *saltans* de *Drosophila*

Os dados existentes na literatura sugerem que o gênero *Drosophila* originou-se durante o Eoceno de uma das várias radiações iniciais dentro da família Drosophilidae. Subseqüentemente, sofreu uma extensa diversificação resultando em cinco radiações sucessivas. Uma destas radiações originou o subgênero *Sophophora*, o qual é constituído por três das maiores linhagens de *Drosophila*: a linhagem *melanogaster*, nos trópicos do Velho Mundo; a linhagem *obscura*, ao Norte da zona temperada; e a linhagem *saltans-willistoni*, nos trópicos do Novo Mundo. Os grupos *melanogaster* e *obscura*, presumivelmente, originaram-se de um ancestral comum *protomelanogaster*. Similarmente, propõe-se que uma única linhagem do subgênero *Sophophora* deu origem aos grupos *saltans* e *willistoni*, estreitamente relacionados (Throckmorton, 1975).

O grupo *saltans* de *Drosophila*, um dos quatro maiores grupos de espécies do subgênero *Sophophora* (Sturtevant, 1942), foi subdividido em cinco subgrupos, com base em uma variedade de características morfológicas (Throckmorton, 1975). Os subgrupos *cordata* (*D. neocordata*) e *elliptica* (*D. emarginata, D. neoelliptica*) foram considerados, por Magalhães (1962), os mais primitivos; e os subgrupos *parasaltans* (*D. parasaltans, D. subsaltans*), *saltans* (*D. saltans, D. lusaltans, D. prosaltans, D. nigrosaltans, D. septentriosaltans, D. austrosaltans, D. pseudosaltans, D. neosaltans*) e *sturtevanti* (*D. rectangularis, D. sturtevanti, D. milleri, D. magalhãensi, D. dacunhai*), os mais derivados.

Com base nos padrões contemporâneos de distribuição geográfica e nas informações geológicas, Throckmorton (1975) propôs que o ancestral das espécies do grupo *saltans* originou-se na América do Norte, onde os subgrupos "primitivos" *cordata* e *elliptica* foram primeiramente encontrados. Este grupo ancestral colonizou o continente Sul Americano, onde os subgrupos *sturtevanti, saltans* e *parasaltans* então se diversificaram (O'Grady *et al.*, 1998). Entretanto, Rodríguez-Trelles *et al.* (1999), estudando seqüências do gene *Xdh* nas espécies do grupo *saltans*, sugeriram que o subgrupo *parasaltans* seria o mais antigo seguido, em progressão sucessiva, pelos subgrupos *sturtevanti, elliptica, cordata* e *saltans*. Estas relações são

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

concordantes com a hipótese filogenética proposta por Throckmorton & Magalhães (1962) somente na posição do subgrupo *saltans* como o último clado derivado e que os subgrupos *elliptica* e *cordata* são adjacentes um ao outro.

Estudos evolutivos ou da diferenciação genética das espécies do grupo *saltans* têm sido baseados predominantemente em análises morfológicas (Throckmorton & Magalhães, 1962), cromossômicas (Bicudo, 1973a), bioquímicas (Nascimento & de Campos Bicudo, 2002) e do isolamento reprodutivo (Bicudo, 1973b). Com o desenvolvimento das análises moleculares, novas abordagens têm sido utilizadas, como, por exemplo, a comparação de seqüências de genes estruturais (O'Grady *et al.*, 1998) ou de elementos transponíveis (Clark *et al.*, 1995; Clark & Kidwell, 1997; Silva & Kidwell, 2000).

II. Os elementos transponíveis

Os elementos transponíveis (TEs) são seqüências particulares de DNA, moderadamente repetidas, geralmente com 10^3 a 10^4 pb de comprimento, que têm a habilidade de mudar de posição dentro e entre seqüências gênicas (Hartl *et al.*, 1992). Esta capacidade pode ser, ou não, intrínseca ao elemento. Outras peculiaridades distinguem os TEs das demais seqüências do DNA medianamente repetido: causam uma duplicação de alguns pares de base em cada extremidade do sítio de inserção, apresentam polimorfismo de sítios de inserção e variabilidade no número de cópias dentro e entre espécies (Capy *et al.*, 1998).

Os elementos transponíveis que codificam as enzimas necessárias para seu próprio movimento, ou transposição, são denominados autônomos. Em contraste, os elementos não-autônomos são dependentes dos autônomos para sua transposição. No processo de transposição podem ocorrer quebras dos TEs, que se tornam mais curtos, e por isto são denominados de elementos incompletos. O tamanho e a localização dessas deleções podem ser altamente variáveis, mas não envolvem as repetições terminais. Como qualquer outro parasita molecular, muitos TEs estão sujeitos a altas

taxas de mutação (Nee & Maynard Smith, 1990), o que leva a uma alta freqüência de elementos incompletos não-autônomos.

Desde que foram descobertos por Barbara McClintock na década de 40, os TEs têm sido encontrados na maioria dos organismos, constituindo o maior componente do DNA medianamente repetitivo do genoma de animais e plantas. Seu número de cópias nos genomas varia de poucos até dezenas, centenas ou milhares. Neste caso, TEs podem representar a maior fração do genoma, especialmente em algumas plantas (Kidwell & Lisch, 1997). Recentemente, estimou-se que mais de 50% do genoma do milho; 15% do genoma de *Drosophila* e 45% do genoma humano são compostos por TEs (San Miguel *et al.*, 1996; Biémont & Cizeron, 1999; International Human Genome Sequencing Consortium, 2001).

A classificação dos TEs baseia-se, principalmente, no modo como o elemento se transpõe e nas suas características estruturais e moleculares. Os TEs são agrupados em classes, subclasses, superfamílias, famílias e subfamílias, segundo alguns critérios. As classes são subdivididas com base no mecanismo de transposição do elemento, e as subclasses com base na estrutura molecular dos seus produtos gênicos, ou seja, presença ou ausência de domínios e assinaturas. As superfamílias se diferenciam pela existência de domínios e assinaturas similares, enquanto que as famílias se caracterizam por elementos com similaridade de 30 a 50% em seus motivos e seqüência de aminoácidos. Finalmente, as subfamílias agrupam os elementos que possuem similaridade entre 50 e 70% na seqüência de aminoácidos (Capy *et al.*, 1998).

Duas classes principais são diferenciadas com base nos modos de transposição dos TEs. A Classe I contém os retroelementos que utilizam uma transcriptase reversa para se transporem via um intermediário de RNA, enquanto que a Classe II compreende os elementos que se transpõem diretamente de DNA a DNA, usando uma enzima por eles codificada, chamada transposase. A integração dos elementos transponíveis resulta na duplicação de uma pequena seqüência de DNA no sítio alvo, sendo o tamanho desta duplicação específico e constante para cada elemento. Elementos da Classe II têm sido encontrados tanto em procariotos quanto em eucariotos, enquanto que os retrotransposons parecem estar restritos aos

unesp

eucariotos (Capy *et al.*, 1998). Uma característica comum aos elementos das Classes I e II, e que merece ênfase, é que elementos totalmente funcionais tendem a estar em minoria em muitos genomas, sendo que a grande maioria das cópias é, geralmente, de elementos defectivos ou não-autônomos.

Os elementos da Classe I, podem ser diferenciados em duas subclasses de acordo com a presença ou não de longas repetições terminais (*Long Terminal Repeats* - LTRs). Os elementos com LTRs, ou retrotransposons constituem pelo menos 2% do genoma de *Drosophila* (Capy *et al.*, 1998) e são subdivididos em duas grandes superfamílias, *copia-Ty1* e *gypsy-Ty3*. Em *Drosophila*, entre os componentes da família *copia-Ty1* estão os elementos *copia, roo, H.M.S. Beagle, opus, blood, mdg1, mdg3, springer* e *1731*; e dentre os elementos da família *gypsy-Ty3* podemos citar os elementos *gypsy, Tom, 17.6, 297, 412, micropia* e *Ulisses*.

Os retrotransposons sem LTRs, ou retroposons, são subdivididos nas famílias *SINE* (*Short Interspersed Nucleotide Elements*) e *LINE* (*Long Interspersed Nucleotide Elements*). Os *SINE*s variam de 75 a 500 pb, são formados por transcrição reversa de transcritos produzidos pela RNA polimerase II ou III e não codificam as enzimas necessárias para sua transposição, sendo estas provavelmente fornecidas *in trans* por outro elemento presente no genoma. Estes elementos são encontrados em alguns mamíferos e insetos, mas ainda não foram encontrados em *Drosophila*. Os *LINE*s apresentam de 5 a 9 kb, usam promotores internos para a sua transcrição e codificam uma transcriptase reversa e uma RNAse. Os representantes dessa classe em *Drosophila* são os elementos *D, F, G, I, Doc, jockey, Het-A* e *TART*.

Todos os elementos da Classe II contêm repetições invertidas de 10 a 500 pb em suas extremidades (TIRs = do inglês *Terminal Inverted Repeats*). Estes transposons de DNA, como também são chamados, são subdivididos em três principais superfamílias: *Ac/hobo, Tc1-mariner* e *P*. A família *Ac/hobo* é constituída por elementos amplamente distribuídos nos eucariotos, e em *Drosophila* é representada pelo elemento *hobo*. A família *Tc1/mariner* é formada por elementos de tamanho pequeno, aproximadamente 1,3 kb, e como representantes desse grupo em *Drosophila* encontram-se os elementos *Tc1, Tc3, mariner, Bari-1, Pogo, S, Minos*, e *Uhu*. Finalmente, a família *P* é constituída pelo elemento *P*.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Além destas duas classes bem documentadas, existem ainda os elementos do tipo *MITE* e *Foldback*, que muitos autores consideram como uma terceira classe, cujo mecanismo de transposição permanece desconhecido. Os elementos do tipo *Foldback* contêm amplas repetições terminais invertidas que formam estruturas do tipo grampo. A estrutura interna dos elementos *MITEs* segue a da Classe II, mas seu alto número de cópias e sua preferência insercional lembram os elementos da Classe I (Capy *et al.*, 1998).

Os elementos transponíveis podem se transpor de forma conservativa ou replicativa. Na primeira, ocorre a manutenção do número de cópias do elemento, sendo este o mecanismo usual dos elementos da Classe II. Por este mecanismo, o elemento se excisa do sítio original e se re-insere em um novo sítio alvo. O modo de transposição replicativo leva ao aumento do número de cópias, sendo um mecanismo típico dos retrotransposons. De modo geral, é feita uma cópia do elemento original que é então inserida em um novo sítio. Um aumento do número de cópias também é observado entre os elementos da Classe II por meio de eventos de recombinação e reparo.

II.1. Origem e dinâmica dos TEs no genoma hospedeiro

Após mais de três décadas de intensivo trabalho experimental e teórico, várias questões sobre o significado evolutivo dos elementos transponíveis no genoma hospedeiro ainda permanecem sem resposta. Entender a dinâmica de um TE dentro de um genoma hospedeiro implica compreender como estas seqüências surgem, permanecem, são reguladas e evoluem em meio às pressões seletivas a que são submetidas.

Uma família de elementos móveis pode se originar em uma espécie *de novo* por eventos de mutação e de recombinação, que alteram ou reorganizam seqüências já presentes no genoma (Finnegan, 1985); por meio de transmissão vertical; e, alternativamente, por transmissão horizontal de outra espécie, mediada por um ou mais vetores. O ciclo de vida de um elemento em qualquer linhagem filogenética

pode aparentemente prolongar-se por muitas centenas ou milhões de anos, e pode ser considerado como uma sucessão de três fases: replicação dinâmica; inativação; e degradação (revisão em Kidwell & Lisch, 1997). A partir do momento que uma destas seqüências se insere em um novo genoma, estabelece-se uma fase invasora na qual o número de cópias deste elemento aumenta. A taxa com a qual este fenômeno ocorre depende de vários fatores que regulam a transposição do elemento, como a sua capacidade de invasão, as características do genoma hospedeiro, fatores ambientais e a interação de todos estes fatores. A segunda fase da vida de um TE em um genoma, caracterizada por sua inativação ou imobilização, depende, em grande parte, do impacto do elemento sobre o valor adaptativo do hospedeiro. Finalmente, um estágio de senescência se extende possivelmente por milhões de anos. Neste estágio, todos os elementos autônomos foram perdidos, não ocorre aumento no número de cópias, as seqüências não-autônomas são perdidas, deletadas ou divergem (Pinsker *et al.*, 2001).

Se, durante a evolução de uma espécie, a autonomia do TE for perdida, por transmissão vertical, os descendentes desta espécie não terão cópias ativas deste elemento em seus genomas. Mas se a transmissão ocorrer horizontalmente entre espécies reprodutivamente isoladas, essa espécie poderá receber cópias de elementos ativos e o ciclo reiniciar-se (Kidwell & Lisch, 1997). Através deste mecanismo, o elemento tem maior chance de permanecer no genoma de uma dada espécie. Dessa forma, a transferência horizontal se contrapõe à perda do elemento autônomo e, por este mecanismo, o elemento pode ser transmitido entre as espécies, apesar das barreiras reprodutivas (Daniels *et al.*, 1990a).

Análises de vários TEs têm mostrado que suas distribuições entre espécies não seguem a filogenia da espécie hospedeira; a filogenia da espécie baseada nos seus TEs nem sempre é a mesma daquela baseada em caracteres clássicos, tais como a morfologia. A distribuição de um dado TE entre várias espécies freqüentemente apresenta lacunas e, algumas vezes, uma espécie mais distantemente relacionada contém o elemento, enquanto outra espécie proximamente relacionada não o possui. Além disso, elementos encontrados em espécies que pertencem a diferentes gêneros, famílias, ordens e até mesmo classes, podem ser algumas vezes muito similares. Esta



perda de consistência é freqüentemente devida à transferência horizontal, hibridação entre espécies ou polimorfismo ancestral, sendo que estas explicações não são necessariamente mutuamente exclusivas (Kidwell & Lish, 1997).

Além disso, algumas inconsistências filogenéticas podem refletir o polimorfismo entre as cópias de um elemento no ancestral comum. As cópias de um TE podem ser autônomas; não-autônomas, mas capazes de serem mobilizadas por elementos autônomos; ou não-autônomas, e sem capacidade de mobilização. Neste caso, as diferentes cópias de um elemento podem ser distribuídas ao acaso durante o processo de especiação e cada uma das novas espécies herdar uma ou várias cópias ancestrais. Estas cópias podem então divergir de acordo com as suas taxas evolutivas (Capy *et al.*, 1994). Entretanto, hipóteses alternativas para estas similaridades podem ser consideradas. Primeiro, deve-se assumir que alguns elementos podem ser perdidos por processos estocásticos. Segundo, o alto grau de similaridade entre partes de um elemento em diferentes espécies pode ser o resultado de forte pressão seletiva. Adicionalmente, é possível que estas regiões exerçam uma função importante na mobilidade do elemento e evoluam mais lentamente que outras (Capy *et al.*, 1994).

Alguns modelos para explicar a manutenção do número de cópias de um TE em uma dada população têm sido propostos com base em análises estatísticas de dados populacionais. Um deles, chamado de modelo neutro, propõe que a taxa de transposição por elemento declina com o aumento do número de cópias, e que a inserção é independente da localização cromossômica. Por este modelo, espera-se uma equiparidade entre o número de elementos no cromossomo X e nos autossomos. Outro modelo, chamado de transposição/seleção, propõe que os elementos seriam eliminados pela seleção natural, assumindo que os efeitos negativos das inserções são iguais para todos os tipos de inserção. Assim, as inserções seriam principalmente heterozigotas e um número estável de cópias seria mantido. Uma alternativa para o modelo descrito acima seria que a seleção poderia agir contra mutações cromossômicas letais dominantes produzidas por trocas desiguais entre inserções de TEs em regiões não-homólogas (Biémont, 1992).



II.2. Papel dos TEs na produção de variabilidade genética

No final da década de 70, os elementos transponíveis foram descritos como seqüências parasitas por usarem a maquinaria hospedeira para sua transposição (Doolittle & Sapienza, 1980; Orgel & Crick, 1980). Desde o início dos anos 80, têm sido realizadas análises detalhadas sobre o polimorfismo de DNA e proteínas codificadas pelos TEs, número de cópias, localização cromossômica, ativação, regulação e taxas de transposição em populações naturais e de laboratório e entre espécies estreitamente relacionadas.

Devido à sua capacidade de movimentação no genoma, propõe-se que os elementos transponíveis desempenhem um importante papel na produção de variabilidade genética. Alguns TEs se inserem preferencialmente em regiões distantes das seqüências gênicas do hospedeiro, tais como regiões intergênicas, heterocromatina ou, freqüentemente, dentro de outro elemento. Outros TEs se inserem preferencialmente dentro ou próximo de seqüências de cópia única. A extensão das alterações induzidas pela atividade dos TEs vai desde modificações no tamanho e arranjo de todo o genoma, até substituições, deleções e inserções de um único nucleotídeo. Eles podem produzir efeitos maiores nas características fenotípicas ou pequenas mudanças detectadas somente na seqüência de DNA. Existem evidências crescentes de que os elementos transponíveis possuam um importante papel na re-estruturação da arquitetura genômica, podendo alterar a estrutura e a função do genoma ao promoverem rearranjos cromossômicos, tais como deleções, duplicações, inversões e translocações recíprocas; afetarem a regulação gênica ou codificarem produtos gênicos que podem interagir com o metabolismo celular; além da sua habilidade em aumentar o tamanho do genoma pela transposição. É importante notar que os TEs também produzem efeitos mutagênicos quando se excisam e deixam pequenas marcas, ou pegadas, da sua presença prévia ou deletam seqüências hospedeiras que flanqueavam o sítio de inserção (Kidwell & Lish, 1997, 2001; Kidwell, 2002).

A maioria dos TEs que se inserem nos exons resultam em mutações nulas devido à sensibilidade destas regiões a mutações e à falta de tolerância das regiões



altamente conservadas para qualquer tipo de mutação. Em contraste, os TEs que se inserem em introns geralmente têm maior chance de serem mantidos no genoma, uma vez que estas inserções são menos afetadas pela seleção natural. Muitos deles são provavelmente retirados com sucesso durante o processamento do mRNA e não possuem efeitos óbvios na função do gene. Entretanto, como podem conter sítios de seqüências reguladoras, inserções de TEs em introns podem afetar a regulação gênica (Kidwell & Lish, 1997).

O papel dos TEs no rearranjo de seqüências genômicas tem sido bem documentado e vários estudos mostram a relação entre pontos de quebra para inversões cromossômicas em *Drosophila* e a presença de elementos transponíveis nestes locais (Lim, 1988; Tsubota *et al.*, 1989; Regner *et al.*, 1996; Cáceres *et al.*, 1999; 2001). Inversões podem levar à ruptura de genes, alterações no padrão de segregação de genes localizados dentro da inversão e mudanças na expressão devido ao efeito de posição. A habilidade dos TEs em promover rearranjos cromossômicos nas regiões de eucromatina foi documentada recentemente por Dimitri & Junakovic (1999), e evidências mostram que os retroelementos em *D. melanogaster* também devem reorganizar a heterocromatina através das deleções, inversões e amplificações. Este fato sugere que a instabilidade inerente dos elementos transponíveis contribui para as alterações na quantidade e distribuição da heterocromatina, o que caracteriza a evolução de animais e plantas.

O movimento dos elementos transponíveis durante as divisões celulares da linhagem germinativa de híbridos produz uma variedade de efeitos deletérios, e a presença destes fenótipos levou muitos pesquisadores a fazerem uma conexão entre os elementos genéticos móveis e a origem de isolamento reprodutivo e especiação (Hurst & Schilthuizen, 1998). Por exemplo, tem sido especulado que a disgenesia do híbrido poderia contribuir para o isolamento reprodutivo (Ginzburg *et al.*, 1999). Segundo Syvanen (1984), a indução da disgenesia do híbrido pelos elementos *P* e *I* em *D. melanogaster* pode ser considerada um evento de sub-especiação e poderia fornecer um mecanismo direto de especiação simpátrica. Entretanto, outros autores como Hurst & Schilthuizen (1998) consideraram que, embora os elementos transponíveis reduzam o valor adaptativo do híbrido, comumente não produzem



isolamento reprodutivo completo. Então, a menos que o isolamento seja estabelecido por outros fatores, o fluxo gênico irá levar os TEs para a outra população, espalhando-os, e a incompatibilidade dos híbridos será perdida. Então, a disfunção híbrida causada por esses elementos teria pouca importância na especiação. Outro problema com esta idéia é que depois de cruzar as fronteiras da espécie, alguns TEs podem rapidamente ser estabelecidos e regulados na nova espécie, resultando na eliminação da maioria dos efeitos disgênicos (Kidwell & Lish, 1997). Ao contrário, os mesmos autores consideram que os elementos transponíveis poderiam ter um papel no isolamento reprodutivo através do aumento na taxa pela qual novos rearranjos cromossômicos surgem e pela fixação de diferentes rearranjos cromossômicos em diferentes populações.

Além de todos os efeitos acima citados, os elementos transponíveis podem ainda ser domesticados pelo hospedeiro (Miller *et al.*, 1997; Nouaud *et al.*, 1999; 2003). Em *Drosophila*, a interação entre o genoma hospedeiro e retroelementos tais como o *Het-A* e *TART* resulta na realização de uma função celular básica. Estes elementos se inserem preferencialmente nas extremidades dos cromossomos e acabam por reparar as perdas nas pontas dos mesmos, uma função que é realizada nos outros organismos pela telomerase (Greider, 1996; Pardue *et al.*, 1996).

Outra forma de produção de variabilidade genética pelos TEs foi descrita recentemente por Wilder & Hollocher (2001). Estes autores mostraram que uma nova classe de retroposons de dípteros chamados de *mini-me*, ou elementos móveis iniciadores de microsatélites, atuam tanto como agentes dispersores quanto como sítios para a origem de DNA microsatélite. Além de originarem seqüências microsatélites, estes TEs estão envolvidos em eventos de duplicação gênica, presentes em DNAs satélites e considerados agentes relacionados ao envelhecimento por causarem instabilidade genética (detalhes em Nikitin & Shmookler Reis, 1997 e na revisão de Kidwell, 2002).

Recentemente, grande número de evidências têm sugerido que o silenciamento mediado pela cromatina, a metilação e outros mecanismos globais de repressão tais como as mutações em ponto repetidamente induzidas e a metilação pré-meioticamente induzida em fungos, e o mecanismo de interferência por RNA

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

surgiram como respostas adaptativas ao aumento do número de cópias dos TEs nos genomas hospedeiros (McDonald, 1998; Hurst & Werren, 2001). Embora estes mecanismos reduzam a atividade dos TEs, deve-se ressaltar que isto não significa necessariamente que eles evoluíram como forma de prevenir a atividade dos elementos transponíveis.

Vários autores também têm especulado que os TEs tiveram um papel importante durante a evolução da vida. Kidwell & Lish (2001) ilustram esta hipótese em três exemplos: (1) é proposto que nos primeiros estágios da evolução biológica os pequenos módulos ou unidades de duplicação (primeiros exons) eram altamente móveis e que a transposição e processamento destes módulos levou à formação de proteínas complexas e famílias multigênicas de proteínas (Dwyer, 1998); (2) postula-se que elementos parasitas, tais como os transposons e os introns possam ter suprido os genes responsáveis pelo estabelecimento do ciclo sexual nos eucariotos (Bell, 1993); e (3) que os TEs devem ter tido um papel crítico em dois eventos macroevolutivos importantes: a origem dos eucariotos e a origem dos vertebrados por terem dirigido a evolução dos mecanismos epigenéticos de formação da cromatina e metilação (McDonald, 1998).

II.3. Caracterização dos TEs analisados neste estudo

Retrotransposons com LTRs

Os elementos com LTRs apresentam uma alta similaridade de estrutura com os retrovírus. Suas LTRs, que possuem as seqüências regulatórias para o início da transcrição e poliadenilação do RNA, flanqueiam uma região central de 4,5 a 9,0 kb, composta de uma a três fases de leitura (ORFs). A primeira ORF corresponde ao gene *gag*, que codifica para uma poliproteína, que é processada em três proteínas maduras: matriz, capsídeo e nucleocapsídeo, que correspondem aos componentes capsidiais dos vírus. A segunda ORF é correspondente ao gene *pol*, que codifica as proteínas necessárias para a transposição, incluindo a protease, a transcriptase reversa, a RNAse H e a integrase. A terceira ORF, presente em apenas algumas

famílias deste grupo, corresponde ao gene *env*, que codifica a proteína do envelope viral (Capy *et al.*, 1998). Dentre os elementos deste grupo, os elementos *copia* e *gypsy* apresentam uma ampla distribuição no gênero *Drosophila*, enquanto que o elemento *micropia* tem sido muito estudado apenas em *D. hydei*.

O elemento *copia* é um retrotransposon que possui 5,1 kb de comprimento, com repetições terminais de 276 pares de bases na mesma orientação do elemento e uma única ORF capaz de codificar uma poliproteína de 1.409 aminoácidos (Emori *et al.*, 1985; Mount & Rubin, 1985). Esta poliproteína gera produtos similares aos genes *gag* e *pol* dos retrovírus. Este TE é encontrado em representantes de todas as principais radiações de *Drosophila*, o que sugere que o elemento *copia* apareceu cedo na história dos drosofilídeos (Martin *et al.*, 1983). No genoma de *D. melanogaster*, este elemento é encontrado, na forma ativa, em um número de cópias que varia de 20 a 60, sendo que a maioria delas está localizada nas regiões cromocêntricas, e poucas distribuídas ao longo dos braços cromossômicos (Csinki & McDonald, 1995). Recentemente, foi sugerido que o elemento *copia* teria sido transmitido horizontalmente de *D. melanogaster* para *D. willistoni* (Jordan *et al.*, 1999).

O elemento *gypsy* é um retrotransposon de 7,5 kb com LTRs de 482 pb (Bayev *et al.*, 1984), que ocorre no genoma de quase todos os grupos do gênero *Drosophila* (Vázquez-Manrique *et al.*, 2000). Recentemente, tem sido demonstrado que este elemento possui propriedades de infecção (Kim *et al.*, 1994). Além de expressar os genes *gag* e *pol*, comuns aos demais retrotransposons, *gypsy* possui ainda a terceira ORF, que produz uma proteína *env-like* completamente funcional, que realiza as funções típicas de uma proteína de envelope retroviral. Conseqüentemente, muitos autores consideram que o elemento *gypsy* é o primeiro retrovírus descrito em inseto (Pélisson *et al.*, 1994; Song *et al.*, 1994). Além disso, tem sido demonstrado que seqüências *gypsy* podem ser horizontalmente transferidas através da alimentação (Kim *et al.*, 1994). Este elemento pode ser considerado um retrovírus endógeno por codificar a proteína do envelope ativa, no entanto, as análises filogenéticas, baseadas em seqüências da transcriptase reversa e integrases, o agrupam junto aos demais retrotransposons com LTRs. É por esta razão que os elementos da família *gyspsy* são considerados intermediários entre os retrovírus e os



retrotransposons com LTRs (Lerat & Capy, 1999). De todas as espécies estudadas, *gypsy* somente não foi encontrado no subgrupo *cordata* do grupo *saltans* e no grupo *immigrans* (Stacey *et al.*, 1986; Herédia, 2002). Análises de hibridizações mostram que elementos *gypsy* inativos estão presentes em *D. melanogaster* em um número de aproximadamente 20 cópias por genoma haplóide e encontram-se localizados na heterocromatina pericentromérica. Além dessas cópias, existem de três a cinco sítios de inserção de elementos completos nos braços cromossômicos (Lambertsson *et al.*, 1989).

O elemento *micropia* foi primeiramente identificado em alças de cromossomos plumosos presentes no cromossomo Y de *D. hydei* durante a prófase da meiose de espermatócitos primários (Hennig *et al.*, 1983; Huijser *et al.*, 1988). Este TE possui 5,5 kb de comprimento, incluindo suas LTRs de 500 pb cada, e apenas uma ORF que codifica produtos com homologia às proteínas do nucleocapsídeo, protease, transcriptase reversa, RNAse H e integrase do retrovírus (Lankenau *et al.*, 1988; 1989). Além deste transcrito, este elemento também codifica um RNA anti-sense que é complementar às regiões codificadoras da transcriptase reversa e RNAse H (Lankenau *et al.*, 1994). Este transcrito anti-sense se acumula nos machos durante a espermatogênese e seu promotor mostra identidade com outros genes de *Drosophila* transcritos especificamente nos testículos. *D. hydei* possui de duas a 11 cópias de *micropia* no cromossomo X e autossomos, além de 50 a 100 cópias no cromossomo Y (Lankenau, 1993).

Retrotransposons sem LTRs

Os retrotransposons sem LTRs, ou retrotransposons, possuem na extremidade 3' uma cauda de poliA, parecendo assim cópias de mRNAs que poderiam ter sido produzidas por transcrição reversa e inseridas no genoma, usando um mecanismo semelhante ao dos retropseudogenes. O elemento *I* é um retrotransposon responsável pelo sistema I-R de disgenesia híbrida em *D. melanogaster*. Esta síndrome surge quando fêmeas reativas (sem o elemento) são cruzadas com machos de linhagens indutoras (com o elemento ativo). As fêmeas resultantes deste cruzamento são pouco férteis e apresentam mutações, rearranjos e não-disjunções cromossômicas na



linhagem germinativa. O elemento *I* possui 5,4 kb de comprimento e é flanqueado por uma duplicação direta da seqüência-alvo, que apresenta comprimento variável. Este elemento possui duas grandes ORFs: uma delas, a ORF1, possui 1.287 pb e codifica um polipeptídeo que mostra similaridade com os polipeptídeos *gag* dos retrovírus; a ORF2, que possui 3.657 pb, codifica proteínas similares a transcriptase reversa e RNAse H. Existem indícios de que ocorrem quebras antes ou durante a integração deste TE no genoma, e que estes elementos tendem a perder sua extremidade 5' durante a transposição (Bucheton *et al.*, 1992). Dados sobre a sua ocorrência em diferentes espécies do grupo *melanogaster* sugerem que os elementos *I* são velhos componentes dos genomas destas espécies e provavelmente estavam presentes antes da radiação do grupo *melanogaster* (Bucheton *et al.*, 1986).

Transposons

Dentre os muitos transposons, os elementos *hobo, mariner, Minos, Bari-1* e *P* apresentam distribuição heterogênea no gênero *Drosophila* e, por este motivo, foram escolhidos para este estudo.

O elemento *hobo* foi molecularmente descrito por McGinnis *et al.* (1983) e caracterizado por Streck *et al.* (1986). Este elemento apresenta 3.016 pb, com repetições terminais invertidas de 12 pb que incluem seqüências TATA e CAAT (Galindo *et al.*, 1995) e uma seqüência de 2,6 kb característica do elemento *hobo* completo. Este elemento apresenta uma ORF de 2 kb que codifica uma enzima de 644 aminoácidos (Karess & Rubin, 1984; Laski *et al.*, 1986) e promove uma duplicação de 8 pb no sítio de inserção. O elemento *hobo* contém na ORF1 uma região S polimórfica que consiste de repetições *in tandem* de uma seqüência de 9 pb 'actccagaa' denominadas TPE (Streck *et al.*, 1986; Calvi *et al.*, 1991; Bazin & Higuet, 1996). Existem diferentes tipos de *hobo* em relação ao número de repetições TPE e isto pode ser um reflexo do cenário de invasão das populações por este elemento (Bonnivard *et al.*, 1997; 2000).

Existem poucas informações sobre a distribuição filogenética das seqüências *hobo*. Até o momento, sua ocorrência era limitada ao grupo *melanogaster* e a *D. willistoni* (Daniels *et al.*, 1990b; Loreto *et al.*, 1998; Rhode, 2000; Klein, 2002). A



atividade do elemento *hobo* foi demonstrada apenas em *D. melanogaster* por causar a síndrome da disgenesia híbrida. A transposição deste TE parece ser específica da linhagem germinativa e controlada transcricionalmente (Calvi & Gelbart, 1994). Além disso, existem evidências de mecanismos de repressão materna e zigótica agindo na mobilização do *hobo* em algumas linhagens (Ho *et al.*, 1993). Análises moleculares revelaram duas classes principais de linhagens, definidas pelos tipos de elementos que possuem. As linhagens H contêm elementos completos (10 a 20 cópias) e numerosos elementos defectivos menores (30 a 75 cópias), enquanto que em linhagens E, não são encontrados elementos completos ou defectivos, mas têm sido detectadas seqüências com menor homologia ao *hobo*, por meio de análises de hibridização.

O elemento *mariner* foi primeiramente descrito em *D. mauritiana* (Jacobson & Hartl, 1985; Haymer & Marsh, 1986). Jacobson *et al.* (1986) sequenciaram a cópia original identificada como uma inserção somaticamente instável no gene *white* e revelaram um elemento de 1.286 pb, com repetições terminais invertidas de 28 pb. *mariner* possui uma única ORF de 1.035 pb (nucleotídeos 172 a 1.209) que codifica uma proteína de 346 amionoácidos e provoca a duplicação de 2 pb da seqüência genômica onde ele se insere. Este TE é capaz de se transpor tanto na linhagem germinativa quanto na somática e possui uma afinidade por regiões de seqüência TA. O movimento somático do *mariner* pode afetar o valor adaptativo de *Drosophila* através da redução do tempo de vida (Nikitin & Woodruff, 1995).

Elementos *mariner*-like são amplamente distribuídos nos metazoários, estando presentes em organismos tão diversos quanto hidras e humanos (Robertson & Zumpano, 1997). Até o momento, existem evidências da existência de 13 subfamílias deste elemento com uma identidade de nucleotídeos de 40 a 56% (Robertson & McLeod, 1993; Robertson, 1995) e, devido a esta ampla distribuição e alta divergência, é proposto que este TE seja um velho componente do genoma dos drosofilídeos (Kidwell, 1993; Robertson, 1993). Apesar de sua ampla distribuição, a filogenia do elemento *mariner* apresenta muitas inconsistências com a filogenia dos hospedeiros, que são atribuídas a perdas estocásticas e/ou transferência horizontal. Em *D. mauritiana*, onde este TE foi primeiramente identificado, existem de 20 a 30

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

cópias de *mariner* por genoma e alguns mecanismos de regulação da transposição têm sido propostos, tais como: inibição por superprodução da transposase, competição entre moléculas mutantes e normais da enzima, envenenamento por multímeros e titulação da transposase (detalhes destes mecanismos em Hartl *et al.*, 1997 e no Capítulo 2 deste trabalho).

O elemento *Bari-1*, que foi isolado de *D. melanogaster* por Caizzi *et al.* (1993), possui 1.728 pb e pode ser encontrado na eucromatina e na heterocromatina como tantos outros elementos. Entretanto, este TE é incomum em muitos aspectos. Primeiro, as cópias heterocromáticas do *Bari-1* constituem um arranjo repetido único na α-heterocromatina e estão estritamente associados ao loco *Responder*, diferentemente de outros elementos que se encontram na β-heterocromatina. Segundo, as cópias heterocromáticas deste elemento apresentam suas ORFs intactas e suas seqüências homogêneas tanto em tamanho quanto em relação ao mapa de restrição (Caggese *et al.*, 1995).

O elemento *Minos* foi identificado como uma seqüência repetitiva dispersa inserida no espaçador transcrito em uma das repetições do loco de rDNA de *D. hydei* (Franz & Savakis, 1991). O elemento é caracterizado por repetições terminais invertidas perfeitas de 255 pb e pela presença de duas ORFs. Franz *et al.* (1994) sugeriram que os membros da família *Minos* estão ativamente se transpondo na linhagem germinativa de *D. hydei*.

O elemento *P* completo possui 2.907 pb com extremidades invertidas repetidas de 31 pb, e uma repetição subterminal invertida de 11-bp (O'Hare & Rubin, 1983). Este elemento apresenta apenas um gene funcional, cujo transcrito de 2,7 kb é processado nas células germinativas com a remoção de três pequenos íntrons, originando um RNAm que codifica uma transposase de 87 kDa. Em tecidos somáticos, o terceiro íntron não é removido e, devido a presença de códons de terminação nessa região, é produzida uma proteína truncada de 66 kDa que atua como repressora da transposição. Várias evidências mostram que os elementos *P* transpõem-se de uma forma não replicativa (Engels *et al.*, 1990), na qual o elemento doador é excisado e reinserido em um local receptor, criando uma duplicação direta de 8 bp no local de inserção (O'Hare & Rubin, 1983).



O elemento *P* completo é o agente causal da disgenesia do híbrido P-M. Esta síndrome, descoberta por Kidwell & Kidwell (1975), ocorre quando machos de linhagens P (possuem o elemento) são cruzados com fêmeas de linhagens M (destituídas do elemento). Os híbridos apresentam esterilidade, altas taxas de mutações, reversão de mutações, rearranjos cromossômicos e recombinação em machos. Em cruzamentos recíprocos, ou entre as próprias linhagens, esses fenômenos não são observados. Ao contrário da disgenesia do híbrido I-R, a esterilidade é uma conseqüência da disgenesia gonadal, em que uma ou ambas as gônadas do híbrido disgênico não se desenvolvem e permanecem rudimentares. Outra diferença é que a disgenesia é dependente da temperatura, sendo que a transposição é maior em temperaturas mais altas (Bregliano & Kidwell, 1983). Os mecanismos que regulam a transposição do elemento *P* são bem estudados e vários modelos de repressão foram propostos. Estes modelos estão reunidos na revisão apresentada no segundo capítulo deste trabalho.

A comparação entre todas as seqüências *P*, encontradas até o presente, sugere que existem várias subfamílias que divergem aproximadamente 30% em suas seqüências de DNA em relação ao elemento *P* canônico de *D. melanogaster*, embora suas estruturas moleculares sejam a mesma que a do elemento canônico. Até o momento, além da subfamília canônica, foram descritas com detalhes outras três subfamílias: tipo-M, tipo-O e tipo-T. Sabe-se que os elementos da subfamília tipo-M apresentam 2.935 pb, os da tipo-O 2.986 pb e os da tipo-T aproximadamente 3 kb. Todos os elementos das três subfamílias são flanqueados por repetições diretas de 8 pb, contém repetições terminais invertidas e uma região codificadora composta de quatro exons (Hagemann *et al.*, 1994; 1998). A subfamília tipo-M é a mais proximamente relacionada ao elemento *P* de *D. melanogaster* (20% de divergência), seus elementos foram primeiramente descritos em espécies do grupo *obscura*, mas também inclui elementos de espécies dos grupos *melanogaster*, *willistoni* e *saltans* e de *Scaptomyza pallida* (Hagemann *et al.*, 1992, 1996a; Haring *et al.*, 1995). A subfamília tipo-O é caracterizada por repetições invertidas, interrompidas por inserções de 12 pb e 14 pb (5' e 3', respectivamente), seus elementos já foram encontrados em *D. bifasciata*, *D. imaii*, *S. pallida*, *D. willistoni* e em algumas



espécies do grupo *saltans* (Hagemann *et al.*, 1994; Haring *et al.*, 1995; Hagemann *et al.*, 1996a; Haring *et al.*, 2000). Finalmente, a subfamília tipo-T compreende, em geral, um grupo de elementos defectivos organizados *in tandem* no genoma de *D. tristis, D. ambigua* e *D. obscura* (Hagemann *et al.*, 1996b). Estes elementos apresentam uma inserção de 300 pb na região 3' não codificadora e os exons 0, 1 e 2 codificam uma proteína semelhante à repressora da transposição. Uma vez que parece não existir elementos ativos nestas espécies, esta proteína pode ter adquirido uma nova função (Miller *et al.*, 1995).

Seqüências *P* estão presentes em muitas espécies dos quatro grupos (*willistoni, saltans, obscura* e *melanogaster*) do subgênero *Sophophora*, sugerindo que estes elementos estavam presentes no ancestral deste subgênero (Daniels *et al.*, 1990a). Até o momento, já foram obtidas mais de 200 seqüências *P* que são agrupadas em 16 clados (ou subfamílias) distintos que caracterizam alguns grupos de espécies sugerindo uma relação de ancestralidade entre as seqüências.

II.4. Distribuição heterogênea dos TEs no grupo *saltans*

Em decorrência de sua capacidade de invadir o genoma e promover mutações, recombinações e conversão gênica, os elementos transponíveis são um sério desafio ao genoma. Embora tenham desempenhado um papel importante na evolução e possam ainda ser úteis para a produção de variabilidade genética nas populações naturais, o modo como invadem os genomas e como seu número de cópias é mantido é ainda um assunto de debates. As informações sobre a ocorrência e a distribuição genômica dos TEs no gênero *Drosophila* foram reunidas em cinco grandes análises (Martin *et al.*, 1983; Stacey *et al.*, 1986; Daniels *et al.*, 1990a,b; Biémont & Cizeron, 1999), sendo *D. melanogaster* a espécie mais estudada.

Em Biémont & Cizeron (1999), a publicação mais recente, são sumarizadas informações sobre 228 espécies do gênero e 43 TEs, incluindo retrotransposons e transposons. Os autores relataram que poderiam ter incluído mais elementos em mais espécies, mas que se restringiram a TEs sobre os quais existiam informações de

unesp

ocorrência em mais de uma espécie, ou para as quais havia sido determinado o número de inserções em sítios eucromáticos. Deste modo, nas revisões apresentadas, nota-se a falta de informações sobre a ocorrência e o número de cópias de muitos TEs em muitas espécies.

Dentre os elementos já estudados no grupo *saltans*, existem vários que foram identificados apenas uma vez, em uma espécie, ou através de uma metodologia e não de outra, ou ainda, que não foram estudados em espécies de todos os subgrupos. Por exemplo, o retrotransposon *copia* foi detectado por *Southern blot* em algumas linhagens de *D. neocordata*, *D. saltans* e *D. sturtevanti*, mas não em outras (Martin *et al.*, 1983). O retrovírus *gypsy*, por sua vez, foi encontrado em *D. elliptica*, *D. saltans*, *D. prosaltans* e *D. sturtevanti*, mas produziu hibridizações fracas em *D. emarginata* e *D. austrosaltans* (Stacey *et al.*, 1986; Herédia, 2002); do mesmo modo, o retroposon *I* foi encontrado apenas em uma linhagem de *D. sturtevanti* (Bucheton *et al.*, 1986). Já, o transposon *mariner* foi detectado em *D. sturtevanti* através de *squash blot* (Brunet *et al.*, 1994), mas não através de *Southern blot* (Maruyama & Hartl, 1991).

III. Objetivos

O objetivo geral deste trabalho foi contribuir para o esclarecimento de questões relacionadas à distribuição, número de cópias e características estruturais de alguns elementos transponíveis no grupo *saltans* de *Drosophila*, com ênfase no estudo do elemento transponível *P*. Dessa forma, foram objetivos específicos:

1- Avaliar a distribuição e o número de cópias das seqüências *P* em espécies do grupo *saltans*;

2- Seqüenciar e caracterizar cópias completas do elemento *P* no genoma de *D. sturtevanti* e *D. prosaltans*;

3- Identificar a presença de diferentes transcritos *P* canônicos nos tecidos somático e germinativo das espécies do grupo *saltans*;

4- Observar a distribuição e o número de cópias dos elementos *copia*, *gypsy*, *micropia*, *I*, *hobo*, *mariner*, *Minos* e *Bari-1* nas espécies do grupo *saltans*.

unesp

Mecanismos de Transposição e Regulação do elemento transponível *P*

Capítulo Dois

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Resumo

Os mecanismos moleculares que controlam a transposição do elemento *P* e determinam sua especificidade tecidual ainda não são completamente entendidos, embora muita informação tenha sido acumulada sobre este elemento na última década. Esta revisão objetiva sumarizar as informações disponíveis sobre a transposição e regulação do elemento *P*. Serão apresentados e discutidos os conceitos atuais sobre a transposição do elemento *P*; o fenômeno da disgenesia do híbrido P-M e as características do citotipo P; os repressores codificados pelo elemento *P* e finalmente, sobre a regulação da transposição deste elemento.

Drosophila melanogaster P transposable element: mechanisms of transposition and regulation

Juliana P. Castro and Claudia M. A. Carareto

Departamento de Biologia, Universidade Estadual Paulista – UNESP, Rua Cristóvão Colombo 2265, CEP 15054-000, São José do Rio Preto – SP, Brazil. Phone: (55) 17 2212382, Fax: (55) 17 2212390. E-mail: carareto@bio.ibilce.unesp.br.

Abstract

The molecular mechanisms that control *P* element transposition and determine its tissue specificity remain not completely understood, although much information has been compiled about this element in the last decade. Due to the great amount of studies that have been done about this subject, this review aims to summarize the available information about *P* element transposition and regulation. This review provides information about *P* element transposition; P-M hybrid dysgenesis and P cytotype features; *P* element codified repressors and regulation of transposition.

Key-words: *Drosophila melanogaster*, *P* element regulation, transposition regulation.

*** Artigo submetido à publicação na Genetica**

Introduction

Transposable elements (TEs), sequences of DNA that move within their host genome from one genomic position to another, were discovered, in maize, during the 1940s by Barbara McClintock. Since then, they have been extensively studied because of their profound effects on gene structure and expression and have been identified in all examined organisms, including prokaryotes and eukaryotes. The fruit fly *Drosophila melanogaster* contains one of the best-characterized eukaryotic transposons, the *P* element family. *P* elements have not only revolutionized the study of *Drosophila* through their use as genetic tools (Kidwell, 1986) and as vectors for germline transformation (Rubin & Spradling, 1982), but they also provide interesting and important system for the study of eukaryotic transposition and gene regulation.

P elements are of two structurally distinct types: completes and incompletes or defectives. The complete 2907 bp elements (Figure 1) contain four protein-coding open reading frames, designated ORF0, ORF1, OFR2 and ORF3, which are transcribed to yield an mRNA of approximately 2.5 kb (O'Hare & Rubin, 1983; Karess & Rubin, 1984). Sequence analysis of several *P* elements revealed that not more than 138 bp at the 5' end and 216 bp at the 3' end are necessary for transposition or excision. Contained within these sequences are perfect terminal inverted repeats (TIR) of 31 bp and an 11-bp sequence found in inverted orientation approximately 130 bp from each end, both needed *in cis* for efficient transposition. Outside the termini of *P* element is found an 8-bp direct repeats, which were present in one copy at the genomic site prior to integration of the transposon (O'Hare & Rubin, 1983). In germline tissues, full-length *P* elements express a *trans*-acting 87 kDa (751 a.a.) transposase protein, which is required for catalyzing forward *P* element transposition as well as precise and imprecise excision events (Rio et al., 1986). In somatic tissues, *P* elements encode a smaller, 66 kDa (576 a.a.) protein as a result of retention of the third ORF2-ORF3 intron (IVS3) in the mature somatic mRNA, which possess a termination codon in the first 9 bp. This 66 kDa protein is known to be a negative regulator transposition in both germline and soma (Misra & Rio, 1990). This germline specificity, which is discussed in details ahead, is regulated at the level of pre-mRNA splicing of the third intron (Rio et al., 1986;



Laski et al., 1986; Laski & Rubin, 1989; Siebel & Rio, 1990), restricting *P* element transposition to this tissue.

In addition to autonomous elements, smaller, non-autonomous *P* elements heterogeneous in size, ranging from about 500-2500 bp, are also common. These elements can be derived from autonomous ones by deletion of some internal sequences. They lack the transposase gene but retain the parts of the sequence required *in cis* for transposition (Engels, 1989). Mobilization of non-autonomous elements occurs only if there is at least one autonomous *P* element present to supply transposase. Most natural populations contain both autonomous and non-autonomous *P* elements and the number of copies per genome ranges from 0 to 30-50 copies (O'Hare et al., 1992).

Although much information has been compiled about *P* elements, all the molecular mechanisms that control their transposition and determine their tissue specificity remain not completely understood, what stimulates new researches and provide us with additional information every year. This review aims to summarize the available information about *P* element transposition and regulation. The first part explains how *P* element transposes in the genome; the second part revises the P-M hybrid dysgenesis and P cytotype features; the third one characterizes the main *P* element codified repressors; and the fourth part focuses on how the *P*-element transposition is regulated.

_P_ element transposition

P element transposition is thought to occur by a DNA-only pathway, in which the transposase enzyme mobilizes *P* elements by a cut-and-paste mechanism (Kaufman & Rio, 1992). Transposase recognizes DNA sites near the *P*-element termini at nucleotides 48-68 and 2855-2871 and makes a staggered cleavage at each end to excise the transposon from the donor site (Kaufman et al., 1989; Beall & Rio, 1997). The 3' cleavage sites are at the ends of the *P* element, and the 5' sites are cleaved 17 bp internal to each end within the 31 bp terminal inverted repeat (Beall & Rio, 1997). The excised element is then inserted into a new target site generating an 8 bp duplication (O'Hare & Rubin, 1983). The double-stranded break at the donor site



is repaired by using the sister chromatid, the homologous chromosome or an ectopically located *P*-element as the template (Gloor et al., 1991). The sister chromatid appears to be used 85% of the time (100% of course, on a male X chromosome), resulting that the sequence copied into the gap is the same as that excised. If transposition is successful, the net effect is the duplication of the TE. If the homologue is used as the donor, then the result will be the precise loss of the element. The repair process, once initiated from a sister chromatid template, can be resolved prematurely before all the sequence information has been transferred, resulting in an internally deleted element (Badge & Brookfield, 1997). There is an apparent preference for chromosomal sites that are likely to be accessible in chromatin: euchromatic sites are favored over heterochromatic sites (Berg & Spradling, 1991), interbands appear to be favored over bands (Semeshin et al., 1986), and there is a marked tendency to integrate at the 5'-end genes (Spradling et al., 1995). Local sequence composition at the site of insertion also appears to play a role (Liao et al., 2000).

Since the high-affinity transposase binding sites are not in the 31-bp terminal inverted repeats, but else adjacent to them (nt 48-68 and 2855-2871), additional factors may be required for *Drosophila P* transposition. One candidate is the non-*P*-element-encoded inverted repeat binding protein, IRBP (Rio & Rubin, 1988). This protein binds to the outer portion (16 bp) of both terminal inverted repeats, directly adjacent to the cleavage sites (Rio & Rubin, 1988; Mullins et al., 1989; Beall et al., 1994) and might serve as an end-aligning factor that helps to properly position the *P*-element ends prior to cleavage by transposase (Beall & Rio, 1997).

Recently, an alternative transposition mechanism has been proposed for several families of TEs, including *P* elements (Svoboda et al., 1995; Gray et al., 1996; Preston et al., 1996; Beall & Rio, 1997; Gray, 2000). According to this proposition, *trans* elements could transpose as hybrid elements in which the left end of one *P* element and the right end of the other move as a unit. Given that alternative transposition events are expected to undergo the same molecular reactions as occur during classical transposition, it is expected that the double-strand breaks created during excision would also undergo repair in the same way. This transposition

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

mechanism results in more complex chromosomal rearrangements than simple insertions (Gray et al., 1996; Gloor & Lankenau, 1998; Lankenau & Gloor, 1998). Chromosomal rearrangements associated with alternative transposition of *P* elements have been shown to be more likely to occur with more complex presentation of functional element ends (Svoboda et al., 1995).

As mentioned before, transposase protein is expressed only in the germline, while in somatic tissues *P* elements encode the 66 kDa protein as a result of retention of the third ORF2-ORF3 intron (IVS3) in the mature somatic mRNA. Genetic and biochemical studies have revealed that the regulation of the tissue-specific splicing of the third *P* element intron in the germline involves inhibition of IVS3 excision in the soma (Laski & Rubin, 1989; Siebel & Rio, 1990; Chain et al., 1991). The sequences responsible for somatic repression of IVS3 splicing are located in the 5'-exon of this intron. The inhibitory element contains two pseudo 5'-splice sites (F1 and F2) that are 20 nucleotides upstream of the accurate 5'-splice site (Siebel & Rio, 1990; Siebel et al., 1992; Chain et al., 1991) and shift U1 snRNP binding from the accurate 5'-splice site. It has been found that a complex containing U1 snRNP and at least four proteins forms on the 5'-exon sequence. Biochemical purification identified two of these proteins: the 97 kDa *P*-element somatic inhibitor protein (PSI), which is highly expressed in the soma, and a 50 kDa hrp48, which is expressed in both the soma and the germline (Siebel et al., 1994, 1995). It was previously thought that PSI bound to sequences outside the F1/F2 motif while hrp48 bound to the pseudo 5'-splice sites (Siebel et al., 1994, 1995). However, recent studies have shown that PSI instead binds to U1 snRNP particles through its AB motif (Labourier et al., 2001; 2002). This interaction modulates U1 snRNP binding on the *P*-element third intron 5'-splice site and its upstream exonic regulatory element in vitro, providing a mechanistic explanation for the soma-specific regulation of *P*-element pre-mRNA splicing (Labourier et al., 2002).

P-M hybrid dysgenesis and P cytotype

In the early 1960s population geneticists first observed deleterious effects during inter-strain crosses between wild and laboratory cultures (Hiraizumi & Crow,

1960). Subsequently, a diverse array of genetic abnormalities, including high rates of mutation, male recombination, sterility resulting from aberrant gonadal development, and a variety of chromosomal rearrangements were observed when males from wild populations were mated to females from laboratory populations (Kidwell et al., 1977; Bregliano & Kidwell, 1983; Engels, 1983). These abnormalities were restricted to the germline of the progeny and these phenotypes were collectively termed hybrid dysgenesis (Kidwell et al., 1977). Molecular experiments showed that *P* elements were the causative agents responsible for these abnormalities (Bingham et al., 1981, 1982; Rubin et al., 1982).

P element often transposes premeiotically in the developing germline cells only in the progeny of a cross in which males from a *P* element-containing strain (P) are mated to females devoid of *P* elements (M). This progeny displays the P-M hybrid dysgenesis syndrome (Kidwell et al., 1977). The absence of the dysgenic traits in the reciprocal cross demonstrates maternal repression of *P*, which is due to a cytoplasmic condition not permissive for *P* element transposition called the P cytotype. The P cytotype state can be maternally inherited for several generations (Kimura & Kidwell, 1994). On the other hand, M strains have "M cytotype", which allows *P* element transposition (Engels, 1983).

D. melanogaster strains differ in their ability to induce the traits of hybrid dysgenesis and to repress them. On the basis of this variation strains are classified phenotypically into the types P, Q and M. A strong P strain has been shown to possess 30-50 *P* elements per genome of which about a third are the complete 2.9 kb element and the rest being deletion derivatives (O'Hare et al., 1992). Some Q strains possess a similar number of genomic elements (Black et al., 1987) and possess the P cytotype, but they do not cause gonadal sterility when out-crossed to M females (Kidwell, 1981). M strains can be divided into true M strains, which possess the M cytotype and no *P* element (Bingham et al., 1982) and M' (or pseudo-M) strains, which contain complete *P* elements together with many deletion derivatives (Anxolabéhère et al., 1984; Black et al., 1987) and are able to partially repress hybrid dysgenesis. This partial repression has been shown to be distinct from P cytotype in that it is inherited through both sexes (Kidwell, 1985; Black et al., 1987).



Surveys of natural populations from several parts of the world have revealed pronounced geographical differences in the distribution of these phenotypes. All strains derived from populations sampled in America, Central Africa, the north-eastern coast of Australia, eastern of Asia and north-western Europe have been determined to be phenotypically P to Q. In the rest of Eurasia, northern Africa and southeastern coast of Australia, populations have been generally found to be M' (see Bonnivard & Higuet, 1999 for a review). In the European populations it seems that the distribution found in the 1980s has been highly stable for the 15 past years (Bonnivard & Higuet, 1999). These authors suggest that this stability could result from the presence of Q populations, which form a buffer zone between P and M' populations in Europe. These Q populations exhibit a particular arrangement of strong regulation properties and thus could slow down or even stop the spread of *P* elements through natural populations. On the other hand, in Australian populations the distribution pattern changed between 1980s and 90s from the P-Q-M North-South cline to a Q-P-Q-M pattern (Boussy, 1987; Boussy & Kidwell, 1987; Boussy et al., 1998).

The prevalent view about *P* element regulation assumes that the P cytotype is consequence of the presence of the 66 kDa protein repressor in the oocytes, which exerts a positive feedback on its own synthesis: individuals produced by a P mother will synthesize more repressor than those that have a similar chromosomal *P* element complement, but which are descended from an M mother (Engels & Preston, 1980; Ronsseray et al., 1993). A model to explain this feedback of the repressor in which the splicing of the IVS3 from *P* pre-mRNA depends on the concentration of this pre-mRNA has been proposed (Lemaitre & Coen, 1991; O'Hare et al., 1992; Lemaitre et al., 1993). Unspliced mRNA will be translated as repressor while spliced mRNA will code for transposase. This will ultimately determine the level of accumulation of repressor in the oocyte, its effect on *P* transcription, and therefore the level of positive feedback and of repression. The low level of transcription with preferential synthesis of the regulator over the transposase could be readily transmitted through succeeding generations. Thus, even in the situation in which transposase-producing *P* elements are paternally derived, the level of maternally inherited regulator is

sufficient to repress these paternally transmitted *P* elements and, hence, to maintain the P cytotype. On the other hand, in the progeny of a dysgenic cross, the paternally transmitted *P* elements are strongly transcribed and the transcripts would be efficiently spliced, leading to the production of high levels of transposase and to no (or a very low level of) 66 kDa repressor in the germ-line. This state of strong transcription will be maintained until there is a sufficient accumulation of repressor to establish a low steady-state level of *P* transcription and therefore switch the cytotype from M to P, which would take several generations. In somatic tissues, irrespective to the direction of the cross, complete *P* elements code for the 66 kDa regulatory protein, due to the lack of splicing of the IVS3 in these tissues (Laski et al., 1986).

Recently, Simmons et al. (2002) proposed a new hypothesis for the maternal transmission of transposase activity. According to these authors transmission of transposase activity through the oocyte poses a problem that does not seem to have been appreciated previously. If either the transposase or fully spliced *P* RNA passed into the oocyte, it could become localized in regions that would give rise to somatic tissues in the fly, thereby subverting the RNA processing mechanism that prevents transposase synthesis (and action) in the soma. Consequently, some mechanism must exclude the transposase and fully spliced *P* RNA from regions of the oocyte that will later form somatic tissues. In their study, a *P* element transgene missing the last intron was unable to transmit transposase activity through the oocytes. This result led them to suggest that the transposase protein cannot enter the oocyte, and the same intron that precludes transposase synthesis in the soma is either required for RNA transport into the oocyte or for its persistence there. Under this hypothesis, only mRNAs that possess the third intron can enter or persist in the oocyte, and those that do so can be fully processed only if they become localized in the region that will become the primordial germline in the embryo after fertilization. Thus, even within the oocyte, the *P* element would preserve a sense of the germline/soma distinction.

These findings have important implications for analyses of *P* element regulation, which have assumed that *P*-encoded polypeptides are responsible for the P cytotype. Simmons et al. (2002) results imply that *P*-encoded polypeptides cannot

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

be the basis of the P cytotype unless they are synthesized in the oocyte or embryo from RNAs that were derived from the mother. Because these RNAs apparently must carry the last *P* intron, polypeptides encoded by *P* elements lacking this intron cannot contribute to the P cytotype, although they could contribute to other forms of *P* element regulation. Transgenes designed by Misra and Rio (1990) to express the 66 kDa protein in the germline showed some regulatory ability, but failed in maternally transmit this ability and were not detected in the oocytes. Simmons et al. (2002) proposed that this failure was due the sequence of the *P* element transgenes. According to these authors, the transgenes were truncated at nucleotide 2078 in the last intron, so missing a nine-nucleotide motif implicated in maternal transmission of developmentally significant *Drosophila* RNAs. These results indicated that *P* elements are adapted to the *Drosophila* germline because they possess an intron that simultaneously precludes transposase synthesis in the soma and excludes transposase mRNA from the oocyte. This bifunctional intron therefore plays an important role in *P* element regulation.

The ability of a strain containing *P* elements to exhibit P cytotype is strongly determined by the genomic position of the repressor-producing elements (Robertson & Engels, 1989; Misra & Rio, 1990; Misra et al., 1993). Biémont et al. (1990) examining 17 inbred *P*-element-containing fly lines for P cytotype over 100 generations of inbreeding identified only three lines that could repress hybrid dysgenesis. All three lines contained at least one complete *P* element that was located at cytological position 1A, at the tip of the X chromosome. Ronsseray et al. (1991) created a strain with only two *P* elements, both located at cytological position 1A, and this strain completely repressed transposition in the germline and did not have a strong ability to induce hybrid dysgenesis even though both elements were complete. Other studies demonstrated that 1A telomere is a hotspot for *P*-element insertion (Ajioka & Eanes, 1989; Biémont et al., 1990; Karpen & Spradling, 1992; Ronsseray et al., 2001), and that it is a region of the genome that strongly promotes P cytotype (Biémont et al., 1990; Ronsseray et al., 1991; Roche & Rio, 1998). It is also known that naturally occurring autonomous *P* elements at 1A show a strong maternal effect for the repression of dysgenic sterility (Ronsseray et al., 1991, 1993, 2001).

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

_P_-element-codified repressors: Type I and Type II

It is known that the *P* element family is regulated by *P*-encoded polypeptides, which repress dysgenesis. There are two nonoverlapping structural classes of repressor-making *P* elements: "type I" repressors, which refer to the larger repressor elements and "type II" for the smaller ones (Gloor et al., 1993).

Type I repressors, like the 66 kDa protein, are produced by elements that have at least the entire sequence of exons 0, 1 and 2 of the transposase gene, and the first 9 base pairs of the 2-3 intron. This protein is collinear with the amino terminal end of the 87-kDa transposase protein for 561 aa., and also shares a hypothetical helix-turn-helix DNA binding motif (Rio et al., 1986). Genetic assays of *P* element excision in the soma and germline using flies containing a single *P* element-encoding 66 kDa protein indicated that such elements repress transposase excision activity in both of these tissues (Rio et al., 1986; Misra & Rio, 1990; Gloor et al., 1993; Handeler et al., 1993; Misra et al., 1993).

Others type I repressors have been found in natural populations. For example, Corish et al. (1996) described, in Q populations, a type I repressor called *SR* (Strong Repressor) that is 2635 bp in length and can be derived from the complete *P* element sequence by a 308 bp deletion extending from position 2341 to 2649. At the site of this breakpoint, there is a 36 bp insertion of a simple AT rich sequence which causes a frame shift resulting in an additional 13 unique amino acids before a premature TAA stop codon. The *SR* element has a coding capacity for a 642 amino acid protein (75 kDa) and as the IVS3 intron is maintained, the *SR* element is also capable of producing the 66 kDa protein. If the germline specific splicing is applied to *SR* RNA, then *SR* will produce a truncated 75 kDa protein in the germline and a 66 kDa repressor protein somatically. The putative 75 kDa *SR* product contains the same protein motifs as the 66 kDa repressor with three protein-binding leucine-zippers and the DNA-binding domain. The authors also showed that the *SR* element effectively reduces transposition levels to zero and operates biparentally.

Type II repressors are deleted for most of exon 1 and all of exon 2 and encode a smaller protein that is collinear with the amino terminal end of transposase for about the first 22 kDa. The coding region for these elements is always truncated near



the middle of exon 1 (Black et al., 1987; Rasmusson et al., 1993). Type II elements are geographically widespread and are not typically associated with the P cytotype, although they can transmit the repressor maternally (Black et al., 1987; Rasmusson et al., 1993).

KP element, the most common type II repressor, is present in high copy number in natural populations and has a deletion comprising the nucleotides 808-2560 (Black et al., 1987). *KP* element produces a 0.8 kb poly $(A)^+$ RNA transcript and encodes a 207 amino acid protein of which the first 199 amino acids are identical to the N-terminus of transposase enzyme (Black et al., 1987). The KP protein also contains two protein-protein interaction regions: a putative leucine-zipper located between nucleotides 101 and 122, and a second region within a segment of 69 carboxy-terminal amino acids. Dimerization of the KP protein is essential for high-affinity DNA binding (Rio, 1990; Andrews & Gloor, 1995; Lee et al., 1996). The KP protein binds to multiple sites on the *P* element termini (11-bp transpositional enhancer and the terminal 31-bp inverted repeats) with higher affinity than the full-length transposase (Lee et al., 1996). It has been suggested that these elements possess some regulatory capacity that protects M' strains from P-induced hybrid dysgenesis, since they have been found in large number in such strains (Rasmusson et al., 1993; Andrews & Gloor, 1995; Simmons et al., 1996).

Other such *P* derivative element with regulatory properties is the *SP* element (Rasmusson et al., 1993). This element is only 517 bp long and its deletion spanned nucleotides 186 to 2577 and apparently arose within a 3-bp direct repeat. The putative polypeptide of the *SP* element is only 14 amino acids long, 11 of which would be identical with the N-terminal sequence of the *P* transposase. Another type II repressor element is the *D50*, which has similar repressor properties, but slightly different deletion endpoints. This element has a deletion from nucleotides 819 to 2527 plus a 21-bp insertion of non-*P* DNA at the deletion breakpoints and codifies for a 204 amino acids polypeptide. Elements with more extreme deletions did not make detectable repressor.



Regulation of *P* element transposition

Several mechanisms of *P* element regulation have been proposed: transcriptional regulation, binding site competition and multimer poisoning by polypeptides, titration of transposase by non-autonomous elements and regulation via an antisense RNA. They are not mutually exclusive and are probably independent. Here, we present the mechanisms previously discussed in other reviews (for example, Rio, 1990), adding the new findings related to each one of them, and summarize the new proposed mechanisms.

Regulation by polypeptides

The transposase, 66 kDa repressor and KP proteins may have similar DNA binding characteristics, and might conceivably have similar effects on transcription when bound to *P* element DNA (Kaufman et al., 1989). Based on several molecular and genetic studies, Rio (1990) summarized the models by which truncated forms of transposase can act as negative regulators of transposase activity. If the 66 kDa repressor or the KP protein bind adjacent or identical DNA sequences, perhaps with different affinity or specificity, they could prevent access of transposase molecules to the *P* element DNA (Figure 2A). As they possess the transposase DNA binding domain, both might block transposase action by competing for the same or nearby sites on *P* element DNA. It is also possible that these polypeptides interfere with protein-DNA interactions by binding other regions of the *P* element, such as the outer half of the inverted repeat, which is recognized by IRBP (Rio & Rubin, 1988). The KP repressor binds to the transposase-binding sites near the 5'- and 3'-ends, to the 31 bp inverted repeats and to the 11 bp internal transpositional enhancer sequences (Lee et al., 1996), while transposase binds only to its single recognition site at each end of the *P* element. This suggests that the repressor protein may block transposase complex assembly by preventing a transient interaction of transposase with the enhancer sequence (Lee et al., 1998).

The 66 kDa and KP proteins might regulate transposition not binding to DNA, but altering the activity of proteins required for transposition. For example, these repressors might indirectly inactivate the transposase protein, for example, by

affecting posttranslational modifications (Figure 2B). It is interesting that the transposase-binding site overlaps the TATA box promoter element of the transposase open reading frame (Kaufman et al., 1989) and that binding of transposase to this site repress *P* element transcription *in vitro* (Kaufman & Rio, 1991). Thus, the 66 kDa and KP proteins could regulate *P* element transcription from binding to the *P* promoter, though they might not be able to repress expression from all *P* element insertions, particularly if the elements were also transcribed under the control of adjacent genomic promoters or enhancers (Misra & Rio, 1990; Andrews & Gloor, 1995; Lee et al., 1996, 1998).

Alternatively, the 66 kDa and KP proteins could directly interact with the 87 kDa transposase to form a complex that is not capable of binding *P* element DNA (Figure 3A and B). This model is specially compelling because there are three sequences resembling the leucine-zipper oligomerization motif (Landshultz et al., 1988) in the transposase, KP and 66 kDa proteins. At this point, it is not known whether the transposase protein functions as an oligomer, but it is possible that the repressor proteins might inhibit transposase activity by forming inactive heteromultimers with transposase (Misra & Rio, 1990; Lee et al., 1998). The model in which a leucine-zipper motif mediates protein-protein interactions would hold for proteins encoded by other internally deleted *P* elements that may lack the transposase DNA binding domain (Rio, 1990).

The main evidence for the *KP* regulation by multimer poisoning is the loss of the repressor function when a single amino acid is changed or deleted from the *KP* leucine-zipper domain (Andrews & Gloor, 1995). These changes could affect the heteromultimerization of the *KP* and transposase proteins preventing the repression. However, prediction of the putative secondary structure of the mutated region showed that none of the mutations disrupted the predicted secondary structure (Lee et al., 1996). Additionally, Lee et al. (1996) showed that there is a second protein-protein interaction domain near the C-terminus of the protein that is sufficient for oligomerization *in vitro.* The same authors also showed that the *KP* protein represses both the donor DNA cleavage and the strand transfer steps of the *P*-element transposition reaction. If the transposition reaction is inhibited at the donor DNA

cleavage step, then the formation of double-stranded breaks generated by excision of the *P* element by transposase is prevented. It is also possible that a second level of repression occurs at the strand transfer step to prevent integration of an excised *P* element. The ability of the *KP* repressor protein to function at both catalytic steps of the *P*-element transposition reaction may serve as a fail-safe mechanism to prevent integration of excised *P* elements that escaped repression by *KP* at the donor cleavage step of the reaction (Lee et al., 1998).

Regulation by titration of transposase

An alternative model of *P* element regulation was proposed which does not require a *P*-encoded regulator. Simmons & Bucholz (1985) suggested that nonautonomous *P* elements could repress *P* mobility by binding available transposase, thereby lowering the concentration of transposase available for interaction with functional *P* elements. This multimer-poisoning model depends on the number of subunits of transposase required for accurate DNA binding. If more than one subunit is required, then truncated transposase molecules (repressors) could inhibit transposase by preventing either DNA binding or transposase enzymatic activity (Engels, 1989). Unfortunately, the number of transposase subunits required for function is unknown, but it is reasonable to assume from studies on bacterial transposons that *P* transposase is acting as part of a multimeric protein-DNA complex (Bainton et al., 1991).

Regulation by antisense RNA repression

Several studies have reported evidences for regulation by *P*-encoded polypeptides, but there have been few published reports of regulation by antisense *P* RNA. Rasmusson et al. (1993), studying the regulatory abilities of individual *KP, DP* and *SP* elements on X chromosomes proposed that, in some cases, repression might be mediated by an antisense *P* RNA rather than by a *P*-encoded polypeptide. These authors suggested that the repression by *SP* elements, for example, could not be caused by a polypeptide, because the 14 amino acid polypeptide encoded by these elements is too small to be of any significance. One explanation might be that



repression by *SP* is mediated by an antisense RNA that interferes with the processing or translation of complete *P* element transcripts. This antisense RNA could be produced by transcription backwards through the *SP* element from an external promoter. Once formed, the antisense RNA molecules could bind to sense RNA molecules produced by complete *P* elements, blocking their processing or translation, or even causing their degradation by RNAses. Any of these events would impair the synthesis of transposase polypeptides and ultimately reduce the incidence and severity of hybrid dysgenesis. The authors pointed out that any kind of *P* element could insert backwards and downstream of a genomic promoter, leading to the synthesis of antisense *P* RNA. This mechanism of repression could therefore be quite general, and could even apply to complete *P* elements. This RNA could then interfere with the expression of sense RNA from other copies of the transposons, thereby blocking the synthesis of the protein needed for transposition. Under this hypothesis, the evolution of the regulatory state depends on the ability of genomic promoters to mediate the synthesis of antisense transposons RNA. Such promoters would therefore provide the genome with general and effective instruments to combat the accumulation of a multitude of parasitic transposable elements.

Simmons et al. (1996) tested the ability of sets of sense and antisense *P*-element constructs, controlled by a heat-shock-inducible promoter, to repress manifestations of *P*-element activity *in vivo*. Their results showed that some sets of antisense *P*-element constructs were able to repress pupal lethality and gonadal dysgenesis. Although, many different types of *P* elements could, in principle, regulate *P* activity by producing antisense *P* RNAs. So, the authors suggest that small elements may be more likely to produce complete antisense transcripts, *i.e.*, transcripts that include the translational initiation sequences at the element's 5' end; thus, they might be more effective as antisense regulators of *P* activity.

Concluding Remarks

Complete *P* elements in *Drosophila melanogaster* catalyze their own transposition via a self-encoded 87 kDa transposase, and auto-regulate their transposition via a truncated 66 kDa repressor protein produced by incomplete

splicing of the third intron (Engels, 1989; Rio, 1990). Questions still remain as whether and how differential splicing takes place in germline and somatic tissues and how the maternal inheritance of cytotype is brought about. As pointed out by Corish et al. (1996), it is probable that the balance between the production of transposase and repressor proteins determines the rate at which the P cytotype becomes established.

The current state in *D. melanogaster*, having P, Q and M' natural populations, may reflect the recent invasion of the *P* family into this species and the initial demand on populations to co-evolve a repression system, in whatever form, in order to survive the damaging consequences of hybrid dysgenesis. As long as some degree of protection is achieved, then there is a short term, but possibly transitory, benefit and the final resting place for *D. melanogaster* will depend on the long-term selective differences between the type I and type II systems (Corish et al., 1996).

The model proposed by Corish et al. (1996) seems to us a valuable way to explain the possible course of evolution of repression of *P*-induced hybrid dysgenesis. The authors wondered that if the production of the 66 kDa or a "66 kDa-like" protein switches from being dependent on intron splicing efficiencies to being the only product of a permanently altered type I repressor element, then a strong, biparentally transmitted repression might ensue. This has the added advantage that no significant levels of hybrid dysgenesis would be induced by autonomous *P* elements inherited through the males. This would explain the evolution of the Q phenotype; that is, populations with strong repressor activity but less than 10% induction activity in gonadal dysgenesis tests despite the presence of autonomous elements (Kidwell, 1983). A schematic outline of this model is shown on Figure 4. *P* elements from other *Drosophila* species invade M populations of *D. melanogaster* and cause hybrid dysgenesis and some populations go extinct. Survival depends on evolution of repression systems, of which there are two types defined by their mode of transmission (maternal or biparental). Maternal transmission characterizes P cytotype repression in P populations. Biparental transmission characterizes M' populations with the *de novo* origin of type II elements. According to the model, evolution of four known classes of Q populations has occurred: (1) Qmat –*KP* with strong, exclusively



maternally transmitted repression and without *KP*; (2) Qmat +*KP* with predominantly maternally transmitted repression, despite the presence of *KP* elements which give a moderate level of paternal transmission (this class probably evolved from crosses between P and M' populations); (3) Qbip –*KP* +*SR*-like with biparentally transmitted repression but no *KP* elements; (4) Qbip +*KP* +*SR*-like with biparentally transmitted, strong repression due to its *SR* and *KP* elements (this class probably evolved from an M' population). Evolution of Qmat populations into Qbip, with permanent, biparentally inherited strong type I repression is a consequence of the introgression or *de novo* origins of *SR*-like elements producing 66 kDa repressors. By this model, Corish et al. (1996) believe that the *SR* element would be strongly favored by selection, given its strong repression ability.

This model of repression evolution is strongly supported by Bonnivard and Higuet (1999) proposition of a buffer zone in the European populations. According to this authors this buffer zone is made up of Q populations that separate western French P populations from M' populations. Then, migration of individuals from P or M' populations is limited to Q populations, which can repress transposition and thus maintain their Q phenotype. Reciprocally, P and M' populations receive migrants coming only from Q populations, and then keep their P or M' status. The population distribution pattern found in Australia (Boussy et al., 1998; French et al., 1999; Itoh et al., 1999) also supports the proposition that Q populations would be the last stage of *P* element repression evolution.

This review showed that the regulation of *P* element transposition is complex and probably involves several types of mechanisms that depend on the TEs themselves and/or interactions with the host genome, including transcriptional repression, DNA binding of truncated transposase-related molecules, RNA-RNA interaction and protein-protein interactions. Thus, despite the great amount of knowledge summarized here, much more investigation must be done to elucidate the mechanisms that regulate *P* element transposition. So, one would not be surprised if in the next few years a new review at this subject will come up.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

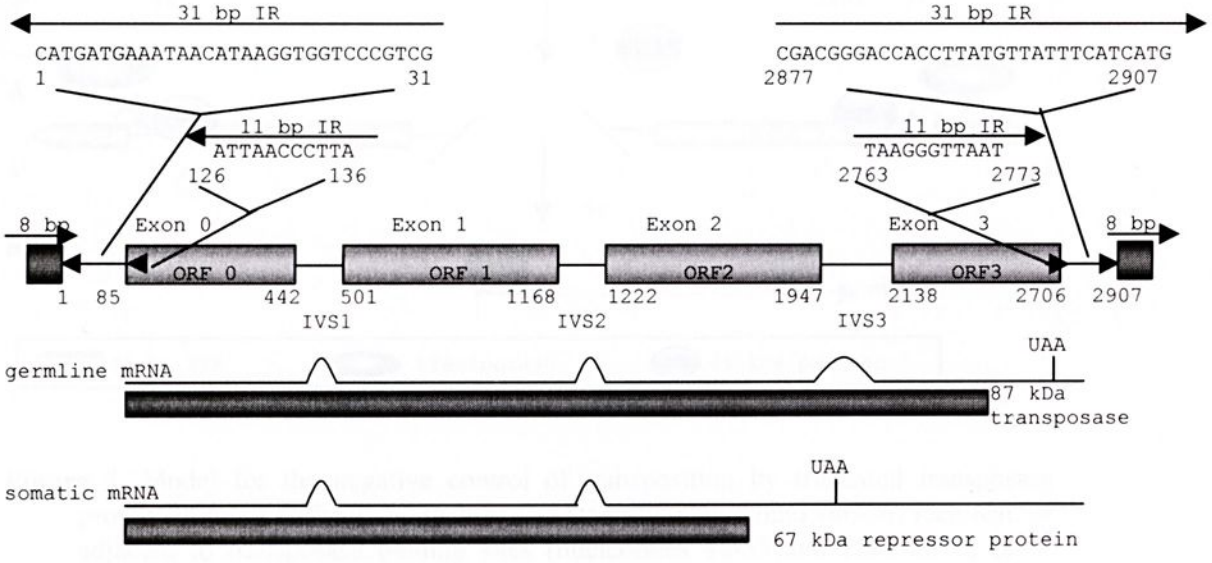


Figure 1. Structure of the 2.9 kb *P* element, *P* element RNA and proteins. Location of ORFs, exons, introns (IVS), and terminal 31 bp and internal 11 bp inverted repeats. There are two major RNA transcripts, a germline-specific mRNA that encodes the 87 kDa transposase protein, and a mRNA made in somatic cells that encodes a 66 kDa repressor protein (Redrawn from Rio, 1990).

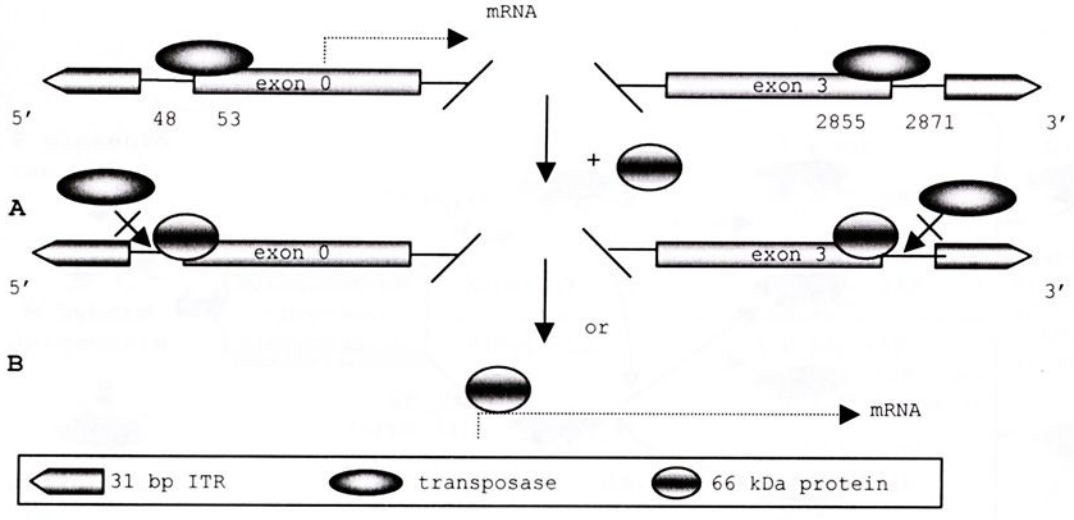


Figure 2. Model for the negative control of transposition by truncated transposase proteins. Larger derivatives, such as 66 kDa protein may bind similar, identical, or adjacent to transposase binding sites (nucleotides 48-53 and 2855-2871) on *P* element DNA, thereby blocking transposase action **(A)**, or interact with *P* element mRNA to block translation or processing of the transposase RNA **(B)** (Redrawn from Rio, 1990).

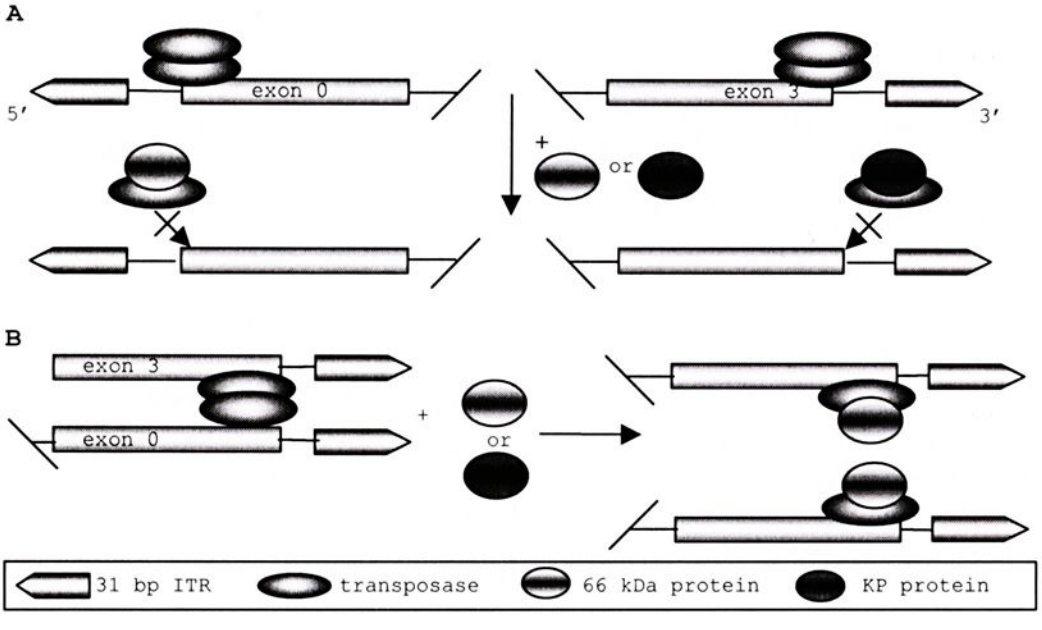


Figure 3. Model for the negative control of transposition by truncated transposase proteins. Transposase is depicted binding to DNA as a dimmer or higher oligomer. Transposase multimer is inactivated for DNA binding by interaction with 66 kDa or KP protein **(A)** or the interaction of transposases during excision is disrupted **(B)** (Redrawn from Rio, 1990).

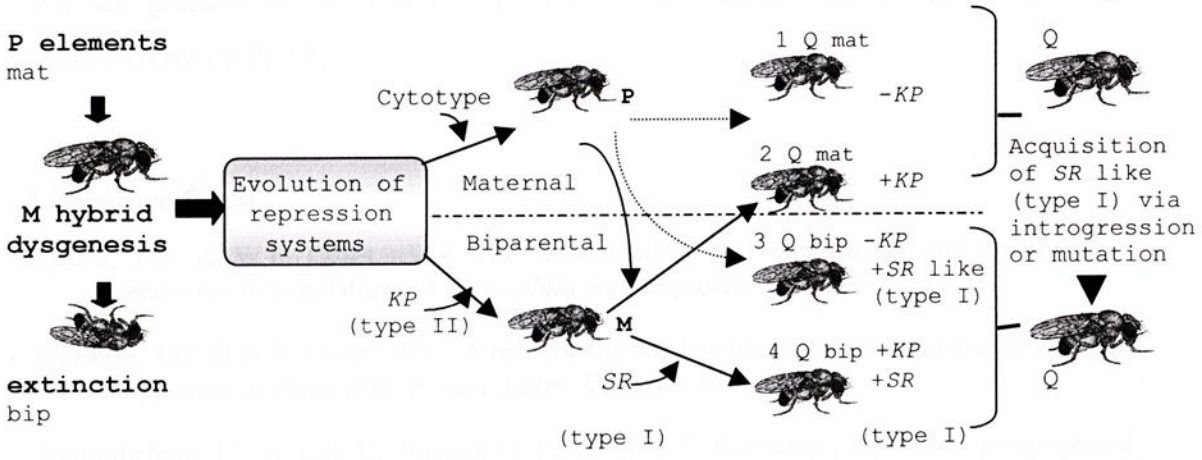


Figure 4. Schematic outline of the possible course of evolution of repression of *P*-induced hybrid dysgenesis, according to Corish *et al.* (1996).

Acknowledgements

We are grateful to Dr. Pierre Capy for his critical comments. This work was supported by FAPESP.

Literature Cited

Ajioka, J.M. & W.F. Eanes, 1989. The accumulation of *P*-elements on the tip of the X chromosome in populations of *Drosophila melanogaster*. Genet. Res. 53: 1-6.

Andrews, J.D. & G.B. Gloor, 1995. A role for the KP leucine zipper in regulating *P* element transposition in *Drosophila melanogaster*. Genetics 141: 587-594.

Anxolabéhère, D., H. Kai, D. Nouard, G. Periquet & S. Ronserray, 1984. The geographical distribution of P-M hybrid dysgenesis in *Drosophila melanogaster*. Genet. Sel. Evol. 16: 15-26.

Badge, R.M. & J.F.Y. Brookfield, 1997. The role of host factors in the population dynamics of selfish transposable elements. J. Theor. Biol. 187: 261-271.

Bainton, R., P. Gamas & N.L. Craig, 1991. Tn7 transposition *in vitro* proceeds through an excised transposons intermediate generated by staggered breaks in DNA. Cell 65: 805-816.

Beall, E.L. & D.C. Rio, 1997. *Drosophila P*-element transposase is a novel site-specific endonuclease. Genes Dev. 11: 2137-2151.

Beall, E.L., A. Admon & D.C. Rio, 1994. A *Drosophila* protein homologous to the human p70 Ku autoimmune antigen interacts with the *P* transposable element inverted repeats. Proc. Natl. Acad. Sci. USA 91: 12681-12685.

Berg, C.A. & A.C. Spradling, 1991. Studies on the rate and site-specificity of *P*-element transposition. Genetics 127: 515-524.

Biémont, C., S. Ronsseray, D. Anxolabéhère, H. Izaabel, & C. Gautier, 1990. Localization of *P* elements, copy-number regulation, and cytotype determination in *Drosophila melanogaster*. Genet. Res. 56: 3-14.

Bingham, P.M., R. Levis, & G.M. Rubin, 1981. The cloning of the DNA sequences from the *white* locus of *Drosophila melanogaster* using a novel and general method. Cell 25: 693-704.

Bingham, P.M., M.G. Kidwell & G.M. Rubin, 1982. The molecular basis of P-M hybrid dysgenesis: the role of the *P* element, a P strain specific transposable element family. Cell 29: 995-1004.

Black, D.M., M.S. Jackson, M.G. Kidwell & G.A. Dover, 1987. KP elements repress P-induced hybrid dysgenesis in *D. melanogaster*. EMBO J. 6: 4125-4135.

Bonnivard, E. & D. Higuet, 1999. Stability of European natural populations of *Drosophila melanogaster* with regard to the P-M system: a buffer zone made up of Q populations. J. Evol. Biol. 12: 633-647.

Boussy, I.A., 1987. A latitudinal cline in P-M gonadal dysgenesis potential in Australian *Drosophila melanogaster* populations. Genet. Res. 49: 11-18.

Boussy, I.A. & M.G. Kidwell, 1987. The P-M hybrid dysgenesis cline in Eastern Australian *Drosophila melanogaster*: discrete P, Q and M regions are nearly contiguous. Genetica 115: 737-745.

Boussy, I.A., M. Itoh, D. Rand & R.C. Woodruff, 1998. Origin and decay of the *P* element-associated latitudinal cline in Australian *Drosophila melanogaster*. Genetics 104, 45-57.

Bregliano, J.C. & M.G. Kidwell, 1983. Hybrid dysgenesis determinants. In: *Mobile Genetic Elements*, ed. J. A. Shapiro, pp. 363-410. London Academic Press.

Chain, A.C., S. Zollman, J.C. Tseng & F.A. Laski, 1991. Identification of a *cis*-acting sequence required for germ-line-specific splicing of the P element ORF2-ORF3 intron. Mol. Cell. Biol. 11: 1538-1546.

Corish, P., D.M. Black, D.W. Featherston, J. Merriam & G.A. Dover, 1996. Natural repressors of *P*-induced hybrid dysgenesis in *Drosophila melanogaster*: a model for repressor evolution. Genet. Res. 67: 109-121.

Engels, W.R., 1983. The *P* family of transposable elements in *Drosophila.* Ann. Rev. Genet. 17: 315-344.

Engels, W. R., 1989. *P* elements in *Drosophila*, pp. 437-484 In *Mobile DNA*. D. Berg and M. Howe (ed.). American Society of Microbiology, Washington, D.C.

Engels, W.R. & C.R. Preston, 1980. Components of hybrid dysgenesis in wild population of *Drosophila melanogaster.* Genetics 95: 111-128.

French, D.J., P. Corish, M. Shi & G.A. Dover, 1999. Evolutionary flux of *P* element regulation in *Drosophila melanogaster* hybrid dysgenesis cline. Genet. Res. 73: 205-216.

Gloor, G.B. & D.H. Lankenau, 1998. Gene conversion in mitotically dividing cells: a view from *Drosophila.* Trends Genet. 14: 43-46.

Gloor, G.B., D.M. Nassie, D.M. Johnson-Schlitz, C.R. Preston & W.R. Engels, 1991. Targeted gene replacement in *Drosophila* via *P* element-induced gap repair. Science 253: 1110-1117.

Gloor, G.B., C.R. Preston, D.M. Johnson-Schlitz, N.A. Nassif & R.W. Phillis, 1993. Type I repressors of *P* element mobility. Genetics 135: 81-95.

Gray, Y.H., 2000. It takes to transposons to tango: transposable-element-mediated chromosomal rearrangements. Trends Genet. 16:461-467.

Gray, Y.H., M.M Tanaka & J.A. Sved, 1996. *P*-element-induced recombination in *Drosophila melanogaster*: hybrid element insertion. Genetics 144: 1601-1610.

Handeler, A.M., S.P. Gomez & D.A. O'Brochta, 1993. Negative regulation of *P* element excision by the somatic product and terminal sequences of *P* in *Drosophila melanogaster*. Mol. Gen. Genet. 237: 145-151.

Hiraizumi, Y. & J.F. Crow, 1960. Heterozygous effects on viability, fertility, rate of development and longevity of *Drosophila* chromosomes that are lethal when homozygous. Genetics 45: 1071-1083.

Karess, R.E. & G.M. Rubin, 1984. Analysis of *P* transposable element functions in *Drosophila*. Cell 38: 135-146.

Karpen, G.H. & A.C. Spradling, 1992. Analysis of subtelomeric heterochromatin in the *Drosophila* minichromosome *Dp1187* by single *P* element insertional mutagenesis. Genetics 132: 737-753.

Kaufman, P.D. & D.C. Rio, 1991. *Drosophila P* element transposase is a transcriptional repressor *in vitro*. Proc. Natl. Acad. Sci. USA 88: 2613-2617.

Kaufman, P.D. & D.C. Rio, 1992. P element transposition in vitro proceeds by a cut-and-paste mechanism and uses GTP as a cofactor. Cell 69: 27-39.

Kaufman, P.D., R.F. Doll & D.C. Rio, 1989. *Drosophila P* element transposase recognizes internal *P* element DNA sequence. Cell 59: 359-371.

Kidwell, M.G., 1981. Hybrid dysgenesis in *Drosophila melanogaster*: the genetics of cytotype determination in a neutral strain. Genetics 98: 275-290.

Kidwell, M.G., 1985. Hybrid dysgenesis in *Drosophila melanogaster*: nature and inheritance of *P* element regulation. Genetics 111: 337-350.

Kidwell, M.G., 1986. P-M mutagenesis. *In: Drosophila*: A practical approach. D. B. Roberts, ed. (Oxford: IRL Press), pp. 59-82, 1986.

Kidwell, M.G., J.F. Kidwell & J.A. Sved, 1977. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics 86: 813-833.

Kimura, K. & M.G. Kidwell, 1994. Differences in *P* element population dynamics between the sibling species *Drosophila melanogaster* and *Drosophila simulans*. Genet. Res. 63: 27-38.

Labourier, E., M.D. Adams & D.C. Rio, 2001. Modulation of *P*-element pre-mRNA splicing by a direct interaction between PSI and U1 snRNP 70K protein. Mol. Cell 8: 363-373.

Labourier, E., M. Blanchette, J.W. Feiger, M.D. Adams & D.C. Rio, 2002. The KH-type RNA binding protein PSI is required for *Drosophila* viability, male fertility, and cellular mRNA processing. Genes & Dev. 16: 72-84.

Landshultz, W.H., P.F. Johnson & S.C. McKnight, 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240: 1759-1764.

Lankenau, D.H. & G.B. Gloor, 1998. *In vivo* gap repair in *Drosophila*: a one-way street with many destinations. BioEssays 20: 317-327.



Laski, F.A. & G.M. Rubin, 1989. Analysis of the *cis*-acting requirements for germ-line-specific splicing of the P-element ORF2-ORF3 intron. Genes and Dev. 3: 720-728.

Laski, F.A., D.C. Rio & G.M. Rubin, 1986. Tissue specificity of *Drosophila P* element transposition is regulated at the level of mRNA splicing. Cell 44: 7-19.

Lee, C.C., Y.M. Mul & D. C. Rio, 1996. The *Drosophila P*-element *KP* repressor protein dimerizes and interacts with multiple sites on *P*-element DNA. Mol. Cell. Biol. 16: 5616-5622.

Lee, C.C., E.L. Beall & D.C. Rio, 1998. DNA binding by the KP repressor protein inhibits P-element transposase activity *in vitro.* The EMBO J. 17: 4166-4174.

Lemaitre, B. & D. Coen, 1991. *P* regulatory products repress *in vivo* the *P* promoter activity in *P*-lacZ fusion genes. Proc. Natl. Acad. Sci. USA **88:** 4419-4423.

Lemaitre, B., S. Ronsseray & D. Coen, 1993. Maternal repression of the *P* element promoter in the germline of *Drosophila melanogaster*: a model for the P cytotype. Genetics 135: 149-160.

Liao, G., E.J. Rehm & G.M. Rubin, 2000. Insertion site preferences of the *P* transposable element in *Drosophila melanogaster.* Proc. Natl. Acad. Sci. USA 97: 3347-3351.

McConnel, K. & W.S. Dynan, 1996. The DNA-dependent protein kinase: a matter of life and (cell) death. Curr. Opin. Cell Biol. 8: 325-330.

Misra, S. & D.C. Rio, 1990. Cytotype control of *Drosophila P* element transposition: The 66 kd protein is a repressor of transposase activity. Cell 62: 269-284.

Misra, S., R.M. Buratowski, T. Ohkawa & D.C. Rio, 1993. Cytotype control of *Drosophila melanogaster P* element transposition: genomic position determines maternal repression. Genetics 135: 785-800.

Mullins, M.C., D.C. Rio & G. M. Rubin. 1989. *Cis*-acting DNA sequence requirements for *P* element transposition. Genes Dev. 3: 729-738.

O'Hare, K. & G.M. Rubin, 1983. Structure of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell 34: 25-35.

O'Hare, K., A. Driver, S. McGrath & D.M. Johnson-Schlitz, 1992. Distribution and structure of cloned *P* elements from the *Drosophila melanogaster P* strain π2. Genet. Res. 60: 33-41.

Preston, C.R., J.A. Sved & W.R. Engels, 1996. Flanking duplications and deletions associated with *P*-induced male recombination in *Drosophila.* Genetics 144: 1623-1638.

Rasmusson, K.E., J.D. Raymond & M.J. Simmons, 1993. Repression of hybrid dysgenesis in *Drosophila melanogaster* by individual naturally occurring *P* elements. Genetics 133: 605-622.

Rio, D.C., 1990. Molecular mechanisms regulating *Drosophila P* element transposition. Annu. Rev. Genet. 24: 543-578.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Rio, D.C. & G.M. Rubin, 1988. Identification and purification of a *Drosophila* protein that binds to the 31-base pair inverted repeats of the *P* transposable element. Proc. Natl. Acad. Sci. USA 85: 8929-8933.

Rio, D.C., F.A. Laski & G.M. Rubin, 1986. Identification and immunochemical analysis of biologically active *Drosophila P* element transposase. Cell 44: 21-32.

Robertson, H.M. & W.R. Engels, 1989. Modified P elements that mimic P cytotype in *Drosophila melanogaster*. Genetics 123: 815-824.

Roche, S.E. & D.C. Rio, 1998. *Trans*-silencing by *P* elements inserted in subtelomeric heterochromatin involves the *Drosophila* polycomb group gene, *Enhancer of zeste*. Genetics 149: 1839-1855.

Ronsseray, S., M. Lehmann & D. Anxolabéhère, 1991. The maternally inherited regulation of *P* elements in *Drosophila melanogaster* can be elicited by two *P* copies at cytological site 1A on the X chromosome. Genetics 129: 501-512.

Ronsseray, S., B. Lemaitre & D. Coen, 1993. Maternal inheritance of P cytotype in *Drosophila melanogaster*: a "pre-P cytotype" is strictly extra-chromosomally transmitted. Mol. Gen. Genet. 241: 115-123.

Ronsseray, S., A. Boivin & D.Anxolabéhère, 2001. *P*-element repression in *Drosophila melanogaster* by variegating clusters of *P-lacZ-white* transgenes. Genetics 159: 1631-1642.

Rubin, G.M. & A.C. Spradling, 1982. Genetic transformation of *Drosophila* with transposable element vectors. Science 218: 348-353.

Rubin, G.M., M.G. Kidwell & P.M. Bingham, 1982. The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell 29: 987-994.

Semeshin, V.F., E.S. Belyaeva, I.F. Zhimulev, J.T. Lis, G. Richards & M. Bourouis, 1986. Electron microscopical analysis of Drosophila polytene chromosomes. Chromosoma 93: 461-468.

Siebel, C.W. & D.C. Rio, 1990. Regulated splicing of the *Drosophila P* transposable element third intron *in vitro*: somatic repression. Science 248: 1200-1248.

Siebel, C.W., L.D. Fresco & D.C. Rio, 1992. The mechanism of somatic inhibition: Multiprotein complexes at an exon pseudo-5' splice site control U1 snRNP binding. Genes Dev. 6: 1386-1401.

Siebel, S.W., R. Kanaar & D.C. Rio, 1994. Regulation of tissue-specific *P* element pre-mRNA splicing requires the RNA-binding protein PSI. Genes & Dev. 8: 1713-1725.

Siebel, C.W., A. Admon & D.C. Rio, 1995. Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. Genes Dev. 9: 269-283.

Simmons, M.J. & L.M. Bucholz, 1985. Transposase titration in *Drosophila melanogaster*: A model of cytotype in the P-M system of hybrid dysgenesis. Proc. Natl. Acad. Sci. USA 82: 8119-8123.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Simmons, M.J., J.D. Raymond, C.D. Grimes, C. Belinco, B.C. Haake, M. Jordan, C. Lund, T.A. Ojala & D. Papermaster, 1996. Repression of hybrid dysgenesis in *Drosophila melanogaster* by heat-shock-inducible sense and antisense *P*-element constructs. Genetics 144: 1529-1544.

Simmons, M.J., K.J. Haley & S.J. Thompson, 2002. Maternal transmission of *P* element transposase activity in *Drosophila melanogaster* depends on the last *P* intron. Proc. Natl. Acad. Sci. USA 99: 9306-9309.

Spradling, A., D. Stern, I. Kiss, J. Rootc, T. Laverty & G.M. Rubin, 1995. Gene disruptions using *P* transposable elements: an integral component of the *Drosophila* genome project. Proc. Natl. Acad. Sci. USA 92: 10824-10830.

Svoboda, Y.H.M., M.K. Robson & J.A. Sved, 1995. *P*-element-induced male recombination can be produced in *Drosophila melanogaster* by combining end-deficient elements *in trans*. Genetics 139: 1601-1610.

Capítulo Três

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Elemento *P* no grupo *saltans*: uma reavaliação

Capítulo Três

unesp

Resumo

Dos poucos estudos realizados sobre elementos *P* no grupo *saltans*, a grande maioria foi restrita a análises de parte do exon 2. Com base nestes estudos, não eram conhecidas quaisquer seqüências homólogas ao elemento *P* nas espécies dos subgrupos *cordata* e *elliptica*, enquanto que em todas as espécies dos subgrupos *parasaltans*, *sturtevanti* e *saltans* já haviam sido identificadas tais seqüências. Este estudo mostrou a presença de uma a 15 cópias do elemento *P* em espécies do grupo *saltans*, incluindo *D. neocordata* e *D. emarginata* (espécies dos subgrupos *cordata* e *elliptica*, respectivamente). Destas espécies, somente aquelas dos subgrupos *sturtevanti* e *saltans* parecem conter elementos potencialmente completos. Devido à baixa homologia das seqüências encontradas em *D. neocordata* e *D. emarginata* com as demais seqüências *P* conhecidas, sugere-se neste trabalho que estas seqüências devam representar seqüências rudimentares derivadas de elementos *P* que estavam presentes no ancestral do subgênero *Sophophora*.

P elements in the *saltans* group of *Drosophila*: a re-evaluation of their distribution and copy numbers

Juliana Polachini de Castro and Claudia M. A. Carareto

Departamento de Biologia, Universidade Estadual Paulista – UNESP, Rua Cristóvão Colombo 2265, CEP 15054-000, São José do Rio Preto – SP, Brazil. Phone: (55) 17 2212382, Fax: (55) 17 2212390. E-mail: carareto@bio.ibilce.unesp.br.

Abstract

From the few studies on *P* elements that have addressed the *saltans* group, most of them were restricted to the analysis of part of exon 2. These studies had shown that species from the *cordata* and *elliptica* subgroups were devoid of any discernible *P* homologous sequences, while species from the *parasaltans*, *sturtevanti* and *saltans* subgroups all contain *P* element sequences. Our analyses showed the presence of one to 15 *P* element copies in species of the *saltans* group, including *D. neocordata* and *D. emarginata* (*cordata* and *elliptica* subgroups, respectively). From these species, only those from the *sturtevanti* and *saltans* subgroups seem to harbor putative full-sized elements. Due to the low homology of the sequences found in *D. neocordata* and *D. emarginata* to those early described, we suggest that those sequences might be rudimental *P* element derivatives that were present in the ancestral of the subgenus *Sophophora*.

Key-words: copy number, *D. neocordata*, *D. emarginata*, *P* element, *saltans* group, sequence divergence.

*** Artigo a ser submetido para publicação na Genetica**

Introduction

The complete canonical *P* element was first isolated from *D. melanogaster* and classified as a class II element (O'Hare & Rubin, 1983). *P* elements are of two structurally distinct types: large full-sized or autonomous elements 2.907 bp in length, and smaller non-autonomous elements that are heterogeneous in size, ranging from about 500-2.500 bp. Non-autonomous elements on the whole are formed by internal deletions of larger elements, probably due imprecise excision during transposition (O'Hare & Rubin, 1983). Some specific 5' and 3' *P* element sequences are required for transposition and excision, including perfect 31-bp inverted terminal repeats and an 11 bp sequence found in opposite orientation approximately 130 bp from each end (O'Hare & Rubin, 1983).

The taxonomic distribution of the *P* family of elements is patchy and relatively restricted to the subgenus *Sophophora* of the genus *Drosophila*. Besides the canonical subfamily, found in *D. melanogaster* and in the *saltans-willistoni* groups, the *P* elements from some species of the subgenus *Sophophora* belong to more than one different subfamily, and the diversity of sequences within subfamilies can be as great as 40% (Clark & Kidwell, 1997).

Three distinct *P* element subfamilies (M-, O-, and T-type) have been described and well characterized (Hagemann et al., 1992, 1994, 1996). The M-type subfamily is most closely related to the *D. melanogaster P* element (20% divergence); its sequence is 2935-bp long and has been found in *D. bifasciata* and *D. imaii* from the *obscura* group, *Scaptomyza pallida,* and in species from the *willistoni* and *saltans* groups (Hagemann et al., 1992, 1996; Haring et al., 1995). The O-type is a 2986-bp element characterized by interrupted inverted repeats described in *D. bifasciata, D. imaii, S. pallida* and some species of the *saltans* and *willistoni* groups (Hagemann et al., 1994; Haring et al., 1995; Hagemann et al., 1996; Haring et al., 2000). Finally, the T-type subfamily comprises a group of mostly defective elements restricted to the *obscura* lineage.

The distribution of the elements, their copy numbers, and the extent to which potentially active copies are present are part of the information required to understand the behavior of *P* element in various species. Few studies have addressed



the *saltans* group, and most of them were restricted to the analysis of a sequence amplified from part of exon 2 (Clark et al., 1995, Clark & Kidwell, 1997; Silva & Kidwell, 2000). From these studies it is known that species from the *elliptica* and *cordata* subgroups are devoid of any discernible *P* elements while species from the *parasaltans, sturtevanti* and *saltans* subgroups all contain *P*-homologous sequences (Daniels et al., 1990; Clark et al., 1995). Daniels and Strausbaugh (1986) and Daniels et al. (1990) analyzed, by Southern blot, some species from *saltans* group aiming to determine the distribution of *P* elements in the group and to estimate the number of insertions in each species. However, these authors estimated the number of *P* element insertions comparing the intensity of an internal 0.9 kb fragment. Once it is expected that most of transposable element copies in a genome are internally deleted or possess a great amount of mutations on its sequence, we believe that such analysis might have missed information.

It is believed that a genome recently invaded by a transposable element harbors a high copy number of this element, while a genome invaded long ago might have lost some or most copies during the evolutionary time. Then, the knowledge of the number of insertions of such element in different related species can help to understand the invasion scenery as well as the distribution pattern of these sequences can reflect their evolution in species and populations. Based on these assumptions, this study aimed to investigate the total genomic copy number of sequences homologous to the *D. melanogaster P* transposable element and whether putative full-length copies were present in species from the five subgroups of *saltans* group.

Materials and Methods

Fly stocks

All the species and strains used in the present study are listed in Table 1. All strains were derived from a single, randomly selected female from a mass culture. *D. melanogaster* – Harwich, a strong P line, was used as positive control.

PCR reactions and Sequencing

Polymerase chain reactions were carried out with 200 ng of genomic DNA in a solution of 2 mM $MgCl_2$, 0.16 mM of each dNTP, 0.5 μM M-IR (CATAAGGTGGTCCCGTCG; Haring et al., 1995) and 1 unit of Taq DNA polymerase in 1x buffer, in a total volume of 25 μl. Temperature cycling was performed with the following profile: the solutions were heated to 94°C for 7 min, and then put through 30 cycles of 94°C for 45 sec, 57°C for 45 sec, 72°C for 1 min and 30 sec, followed by a final extension at 72°C for 10 min.

For amplification using primers p2684/p2687 4 mM $MgCl_2$ and 0.5 μM of each primer were used and the primer annealing temperature was 50°C. Primer 2684 (GCTATTTGNYTNCAYACCGC) is 256-fold degenerate and primer 2687 (CCCAATGNATWGCANCGTCTKAT) is 64-fold degenerate (Lee et al., 1999). They correspond to two regions of the most conserved sequences located between nucleotides 703 (exon 1) and 1530 (exon 2) of the *P* element. The products of amplification were visualized in 1% agarose gel stained with ethidium bromide. The PCR products were cloned into a TA cloning vector (Invitrogen) and both strands of five clones, chosen at random, were sequenced for each species.

Southern blot analyses

To estimate the overall amount of *P* elements in each species, genomic DNA was digested with *XbaI* and probed with a 2.8 kb fragment amplified by PCR from the *P* element contained in the pπ25.1 plasmid (primer M-IR). This restriction enzyme has no restriction sites in the canonical *P* element, so each hybridized fragment may correspond to one genomic insertion. To evaluate the presence of putative complete *P* element copies, genomic DNA was digested with *AvaII* and also probed with the 2.8 kb fragment. This digestion produces fragments of 1.8 kb, 0.54 kb and 0.48 kb in the complete canonical *P* element. For the first analysis the chemioluminescent hybridization system Gene Images (Amersham Biosciences) was used according to manufacturer's instructions. In the second analysis the ECL™ Direct Nucleic Acid Labelling and Detection Systems (Amersham Biosciences) was used.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Evolutionary Analysis

Alignments of sequences were done by the program CLUSTAL W 1.8 (Thompson *et al.*, 1994) and improved by hand. Consensus sequences were assumed when the clone sequences were very similar. Relationships among the *P* element sequences of each strain were inferred according to the absence (0) or the presence (1) of restriction fragments. The phylogenetic relationships among *P* sequences and the presence or absence of specific fragments were constructed using the maximum parsimony method, using an heuristic search algorithm as implemented in PAUP v.4.0b10 (Swofford, 1997). Branch support was calculated by bootstrap analysis consisting of 1000 replicates. The neighbor-joining method (Saitou & Nei, 1987) was also used to construct the distance matrix according to the Kimura two-parameter model of nucleotide substitution (Kimura, 1980). In this divergence analysis, the primer binding regions were excluded from the sequences.

Results

Copy number

Sequences homologous to *P* elements were found in all species analyzed. Southern blot analysis showed that the total number of insertion sites of *P* elements on entire genomes differ greatly, ranging from one to 15 (Figure 1). It is possible to notice that after a long exposure time there is a faint hybridization signal in both *D. neocordata* and *D. emarginata* (showed by arrows). Our analysis shows one *P* element insertion in both *D. neocordata* and *D. emarginata*; two in *D. parasaltans*; three in *D. subsaltans*; nine in *D. milleri*; 10 in *D. dacunhai*; 11 in *D. sturtevanti* and *D. saltans*; 12 in *D. austrosaltans* and 15 in *D. prosaltans.*

The presence or absence of each fragment was analyzed and the relationships among the *P* element insertion sites in each species was inferred using the maximum parsimony and neighbor-joining methods (Figure 2). Both methods produced the same basic arrangements. Of 25 characters, 6 were uninformative and 19 parsimony informative. It is possible to note two clades; one composed by the species of the subgroups *cordata* (*D. neocordata*), *elliptica* (*D. emarginata*) and *parasaltans* (*D.*

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

subsaltans and *D. parasaltans*), and the other composed by the species of the subgroups *sturtevanti* (*D. milleri, D. dacunhai* and *D. sturtevanti*) and *saltans* (*D. prosaltans, D. austrosaltans* and *D. saltans*).

Identification of putative full-sized P element

Figure 3 shows the *Ava* II digests hybridized with the 2.8 kb fragment amplified from pπ 25.1. This restriction enzyme cuts a full-sized *P* element at four sites (positions 23, 500, 1044 and 2883) producing internal fragments of 478 bp, 544 bp and 1838 bp (showed by arrows). The absence of one of those fragments, or variation in their sizes, can indicate internal deletions or even lack of the restriction sites due to point mutations. As the probe used hybridizes with the whole *P* element, we expected to find three bands in strains that carry complete elements, plus smaller or bigger bands produced by the digestion of defective elements.

In this analysis, *D. neocordata* and *D. emarginata* did not show any hybridization signal; *D. parasaltans* and *D. subsaltans* showed very weak signal. The other species presented the DNA fragments of the expected sizes for the full-sized canonical *P* element and also fragments smaller and/or bigger than 1.8 kb, indicating polymorphisms. Some of these larger bands are shared by more than one species (some examples showed by arrowheads) and might be product of sequences that have been present in their ancestral.

The 0.48 kb fragment is not visualized in this blot because in 1% agarose gels it runs together with the 0.54 kb fragment. As the species of the *saltans* group have fewer *P* elements than *D. melanogaster*, it seems that only one fragment is present in the former and two adjacent fragments are present in the later. To be sure of its presence, we performed a radioactive hybridization using the same probe used before, but double digesting the genomic DNA with *AvaII* and *HaeIII* (data not shown). The digestion of canonical *P* element with these two enzymes produced the same 1.8 kb and 0.48 kb resulting from the *AvaII* digestion and two smaller fragments of 0.32 kb and 0.22 kb produced by the digestion of the 0.54 kb fragment by the *HaeIII* enzyme.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Detection of the P element by PCR and Sequencing

In order to explain the presence of *P* homologous sequences detected by *Southern blot* in *D. neocordata* and *D. emarginata*, we performed two PCR surveys followed by transference to a nylon membrane and hybridization with a *P* element probe. The first approach used a primer complementary to *P* element inverted repeats and was performed in all species. The second one used degenerated primers complementary to nucleotides 703 to 725 and 1508 to 1530 of canonical *P* element, and was performed only in *D. neocordata*, *D. emarginata*, *D. parasaltans* and *D. subsaltans*.

The amplification of a fragment of approximately 2.9 kb in *D. dacunhai*, *D. sturtevanti*, *D. austrosaltans*, *D. saltans* and *D. prosaltans* confirmed the presence of at least one potentially full-length element in these species (Figure 4A). *D. neocordata*, *D. parasaltans* and *D. subsaltans* did not show any visible amplified fragment. Fragments smaller than 2.9 kb were observed in *D. emarginata* and *D. milleri*, suggesting the presence of internally deleted elements. Internally deleted elements were also present in strains of *D. dacunhai* and *D. saltans*. The products were transferred to a nylon membrane and hybridized with the fragment of 2.9 kb amplified from the pπ25.1 plasmid (Figure 4B). The hybridization confirmed the homology of these fragments to the *P* element sequence and allowed us to visualize amplified fragments of 2.9 kb also in *D. neocordata*, *D. emarginata*, *D. parasaltans* and *D. subsaltans*, corresponding to those of the other species, plus some internally deleted elements in the other species that could not be seen in the agarose gel.

The second analysis showed that in agarose gels it is not possible to visualize the amplified fragments (Figure 5). However, when the products were submitted to hybridization with a *P* element specific probe, the generated signal was very strong.

The products obtained in this last survey were cloned, and five clones of each species were sequenced. The sequences were aligned with *P* element sequences of *D. melanogaster* (X06779), *D. mediopunctata* (AF313770), *D. bifasciata* (X60990), *D. helvetica* (AF313771), *D. sturtevanti* (AF530052), *D. prosaltans* (AF530053), *S. pallida* (M63342) and *Musca domestica* (AF183396). The phylogenetic analysis was inferred using the maximum parsimony and neighbor-joining methods (Figure 6). Both



methods produced the same basic arrangements. Of 1129 characters, 314 were uninformative, and 584 parsimony informative. All remnant sequences were grouped in the same clade in 79% of replicates. This clade also includes the sequence found in *Musca domestica.* The sequence distance among these degenerated sequences and the canonical ones is between 45% (*D. neocordata* 4) and 54% (*D. parasaltans* 1).

We also performed the sequencing of about 600 nucleotides from both ends of the *P* element derivative sequences found in *D. emarginata* and *D. neocordata*. *D. neocordata* sequences did not show a satisfactory alignment to any *P* element derivative sequence. The consensus sequence found in *D. emarginata* showed a divergence of 72% and 48% at its 5' and 3' ends, respectively when compared with the canonical *P* element of *D. melanogaster*. Nevertheless, both ends of this consensus sequence showed lower divergence when compared with *P* element sequences previously described for us in *D. sturtevanti* (GenBank AF530052) and in *D. prosaltans* (GenBank AF530053). The divergence found among these sequences when compared with *D. sturtevanti* sequence was of 52% and 21% at its 5' and 3' ends, respectively, and of 51% and 11% (5' and 3' ends, respectively) when compared with *D. prosaltans* sequence. Part of the alignments obtained for these sequences are shown in Figure 7. Even though the TIRs were not completely sequenced, PCR amplification with primers specific to that region indicates that at least the second half of the TIRs are present and well conserved. However, the transposase binding sites are not well conserved: there is a 6-nt deletion including the TATA box in the 5' end, and several nucleotide substitutions and 1-nt deletion in the 3' end. The 5' and 3' 11-bp subterminal IRs partially conserved.

Discussion

The *cordata* and *elliptica* subgroups, thought to be the most anciently derived within the *saltans* group, are believed to have diversified in North America. Meanwhile, and prior to the formation of the Isthmus of Panama, an ancestor of the *sturtevanti*, *parasaltans* and *saltans* subgroups crossed to South America, where these subgroups later diversified (Throckmorton, 1975). Once any *P* homologous

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

sequence had been found in species of the *cordata* and *elliptica* subgroups (Daniels et al., 1990; Clark et al., 1995; Clark & Kidwell, 1997), Daniels et al. (1990) suggested that these transposable elements could have spread throughout the *saltans* and *willistoni* groups in South America, at a time when the *elliptica* and *cordata* subgroups were still geographically isolated in North America. Alternatively, Clark et al. (1995) proposed that *P* elements could have been lost from the *elliptica* and *cordata* subgroups early in the history of the *saltans* group, when those two subgroups were isolated from the lineages leading to the more recently derived *saltans* subgroups. Additionally, Silva and Kidwell (2000) suggested that there might be host factors in the *elliptica* and *cordata* lineages that confer resistance to *P*-element invasion.

However, now we know that the current distributions of the *elliptica* and *cordata* subgroups throughout Central and South America overlap those of the other *saltans* and *willistoni* subgroups (Patterson & Mainland, 1944; Magalhães, 1962; Spassky et al., 1971) and it is estimated that the spread of canonical *P* elements within the group took place within the last 3 Myr, at a time when North and South America were not longer disconnected (Silva & Kidwell, 2000).

In our analyses we used a probe that overlaps the entire sequence of *P* element, allowing hybridization with any defective element. By this survey, we detected sequences homologous to the *P* element in *D. neocordata* and *D. emarginata.* Daniels and Strausbaugh (1986) admitted that their approach would not detect elements containing deletions that span the probe interval. We also detected up to 15 different genomic insertions in the other species of the *saltans* group. In addition, putative full-size *P* element copies were detected in *D. milleri, D. dacunhai, D. sturtevanti, D. austrosaltans, D. prosaltans* and *D. saltans.*

Clark et al. (1995) and Clark and Kidwell (1997) using degenerated oligonucleotide primers homologous to the region within exon 2 (nt. 703 to 1530) of the *P* element were also unable to detect *P* homologous sequences in species of the *cordata* and *elliptica* subgroups. Our data showed the presence of a poorly amplified *P* homologous fragment of 2.9 kb in *D. neocordata* and *D. emarginata* and also in *D. parasaltans* and *D. subsaltans.* We believe that these fragments might be rearranged

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

and rudimental *P* elements as those found in *L. cuprina* (Perkins & Howells, 1992) and *M. domestica* (Lee et al., 1999). Sequencing of the internal fragments obtained in *D. neocordata* and *D. emarginata* estimated that their divergence to other *P* elements found in the *saltans* group and other species is about 50%. Sequencing of both ends of the relictual sequences found in *D. neocordata* and *D. emarginata* showed that for the former the ends are so divergent that even an alignment was impossible to infer, and for the last, that the 3' end is more conserved than the 5' end, and. Also, the ends of this *P* element remnant sequence present in *D. emarginata* are more similar to those described in *P* element derivative sequences of *D. sturtevanti* and *D. prosaltans* than to the characteristic ends of *D. melanogaster* canonical *P* element.

There are two basic ways to account for the current distribution of *P* homologous sequences in the *saltans* group. If a single introduction of *P* elements has occurred into the group, with homologous sequences being transmitted vertically between species, some sequences have been gradually drifting to extinction within the lineage. This could be the case of the sequences found in *D. neocordata* and *D. emarginata*. Alternatively, it is proposed that *P* elements may have entered in this group more than once and that certain features of the present distribution may be the result of several events of horizontal transfer (Silva & Kidwell, 2000). Haring et al. (2000) suggested that an ancestral M-type was present already in the common ancestor of the *melanogaster*, *willistoni*, *saltans* and *obscura* groups. Within the *saltans* and *willistoni* groups the M-type related canonical *P* elements evolved, which recently invaded the *D. melanogaster* lineage. The same authors also proposed that O-type elements were present in the common ancestor of the *saltans* and *willistoni* groups and have been transmitted vertically within both lineages.

It is possible that the elements present in species of the *cordata*, *elliptica* and *parasaltans* subgroups are remnants of ancestral sequences while the elements present in species of the *sturtevanti* and *saltans* subgroups could be representatives of an earlier invasion. This hypothesis is supported by the fragment polymorphism phylogenetic analysis that clearly separates the species of the subgroups *cordata*, *elliptica* and *parasaltans* from the species of the subgroups *sturtevanti* and *saltans*. The *P* element sequences of these last two subgroups are probably those derived



from more recent invasions. This assumption is corroborated by the fact that only the species of the *sturtevanti* and *saltans* subgroups seems to harbor putative full-sized elements.

Acknowledgements

This investigation was supported by FAPESP (Grants n° 95/7192-2, 98/08734-1 and 00/11313-0 and a fellowship to J.P.C.) and CNPq.

Table 1. Species and strains of *Drosophila* used in this study. Local and year of collection are also indicated.

Subgroup	Species	Strain	Location	Collection Year
cordata	*D. neocordata*	**NEO**	Minas Gerais – Brazil (SC – 14041-0831.0)	1959
elliptica	*D. emarginata*	**EVC**	Vera Cruz – Mexico (SC – 14042-0841.6)	1962
parasaltans	*D. parasaltans*	**PAT**	Tapuruquara (AM) – Brazil *	1962
	D. subsaltans	**SUB**	Belém (PA) – Brazil (SC – 14044-0872.0)	1959
sturtevanti	*D. milleri*	**MEY**	El Yunque – Puerto Rico (SC – 14043-0861.0)	1962
	D. dacunhai	**DAP**	Pentionville – Haiti (SC – 14043-0854.0)	1962
	D. sturtevanti	**SMX**	Matlapa (San Luis Potosi) - Mexico **	1998
saltans	*D. austrosaltans*	**API**	Pirassununga (SP) – Brazil (SC – 14045-0881.0)	1959
	D. saltans	**SAC**	Chilpancingo – Mexico *	1962
	D. prosaltans	**PTT**	Sangre Grande –Trinidad Tobago *	1962

Notes: SC indicates Stock Center ordering number; * provided by H.E.M.C. Bicudo; ** collected by J.C. Silva.



Figure 1. Southern blot hybridization of genomic DNA from species of the *saltans* group. DNA was digested with *XbaI* and probed with the 2.8 kb fragment amplified from pπ25.1. The species analyzed were: *D. neocordata* (NEO); *D. emarginata* (EVC); *D. parasaltans* (PAT); *D. subsaltans* (SUB); *D. milleri* (MEY); *D. dacunhai* (DAP); *D. sturtevanti* (SMX); *D. austrosaltans* (API); *D. saltans* (SAC); *D. prosaltans* (PTT) and *D. melanogaster* (HAR). Each fragment corresponds to one genomic insertion site.

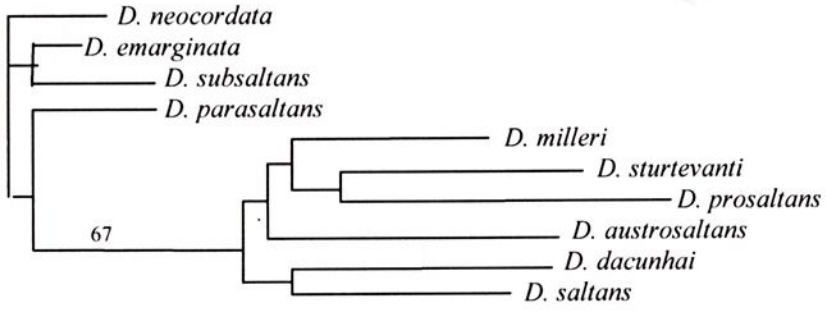


Figure 2. Relationships among the *P* element different insertion sites in species of the *saltans* group. This cladogram was generated by neighbor-joining analysis as implemented by PAUP 4.0b10 (Swofford, 1997) and required 44 steps. The consistency index is 0.57 and the retention index 0.5. Bootstrap analysis of 1000 replicates was performed on the data.

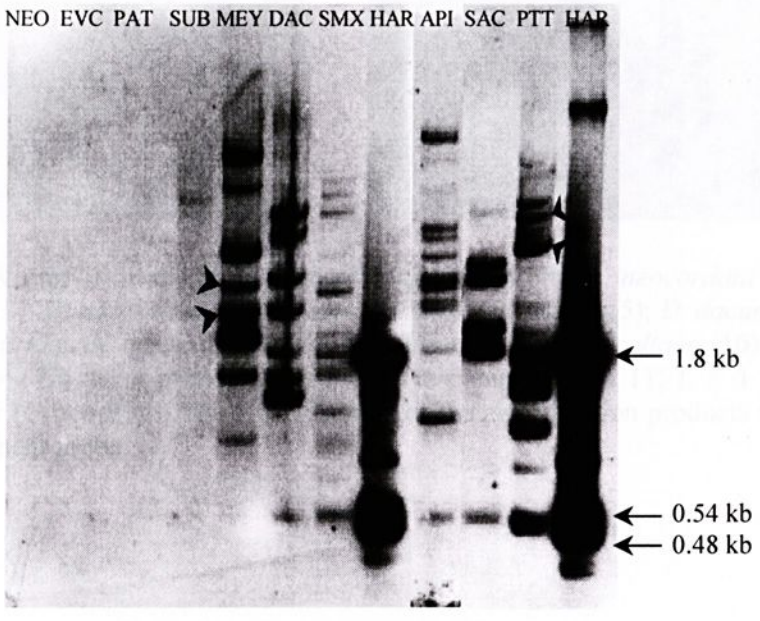


Figure 3. Southern blot hybridization of genomic DNA from species of the *saltans* group. DNA was digested with *AvaII* and probed with the 2.8 kb fragment amplified from pπ25.1. The species analyzed were: *D. neocordata* (NEO); *D. emarginata* (EVC); *D. parasaltans* (PAT); *D. subsaltans* (SUB); *D. milleri* (MEY); *D. dacunhai* (DAP); *D. sturtevanti* (SMX); *D. austrosaltans* (API); *D. saltans* (SAC); *D. prosaltans* (PTT) and *D. melanogaster* (HAR). The presence of fragments of 1.8 kb, 0.54 kb and 0.48 kb (showed by arrows) is an indication of the presence of putative complete *P* elements. The arrowheads show fragments shared by more than one species.

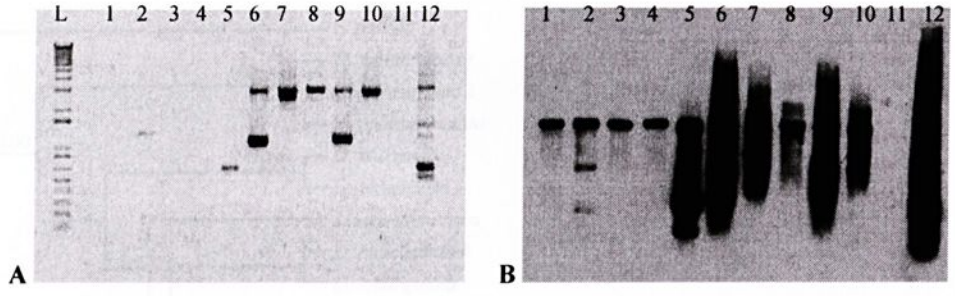


Figure 4. (A) PCR amplification of *P* element sequences from *D. neocordata* (1); *D. emarginata* (2); *D. parasaltans* (3); *D. subsaltans* (4); *D. milleri* (5); *D. dacunhai* (6); *D. sturtevanti* (7); *D. austrosaltans* (8); *D. saltans* (9); *D. prosaltans* (10) and *D. melanogaster* (12) using primer M-IR. Negative control = line 11; L = 1 kb plus DNA Ladder (Gibco-BRL). **(B)** Hybridization of the amplification products with the 2.9 kb *P* element probe.

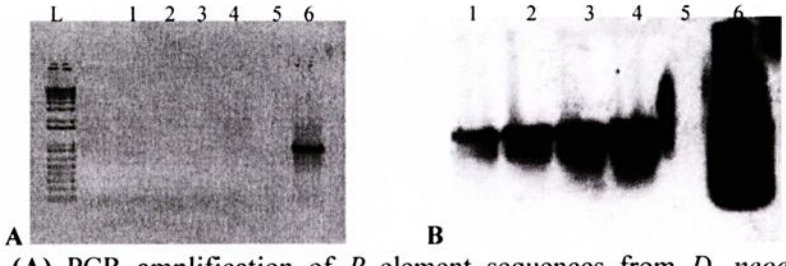


Figure 5. (A) PCR amplification of *P* element sequences from *D. neocordata* (1); *D. emarginata* (2); *D. parasaltans* (3) *D. subsaltans* (4) *D. melanogaster* (6) using primers #2684 / #2687. Negative control = line 5; L = 1 kb plus DNA Ladder (Gibco-BRL). **(B)** Hybridization of the amplification products with the 2.9 kb *P* element probe.

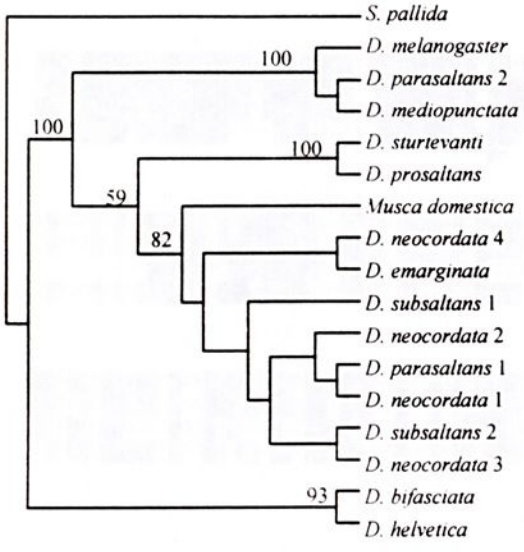


Figure 6. Cladogram of *P* element sequences and those generated with degenerate primers for *D. neocordata, D. emarginata, D. parasaltans* and *D. subsaltans*. This cladogram was generated by parsimony analysis as implemented by PAUP 4.0b10 (Swofford, 1997) and required 2627 steps. The consistency index is 0.65 and the retention index 0.46. Bootstrap analysis of 1000 replicates was performed on the data.

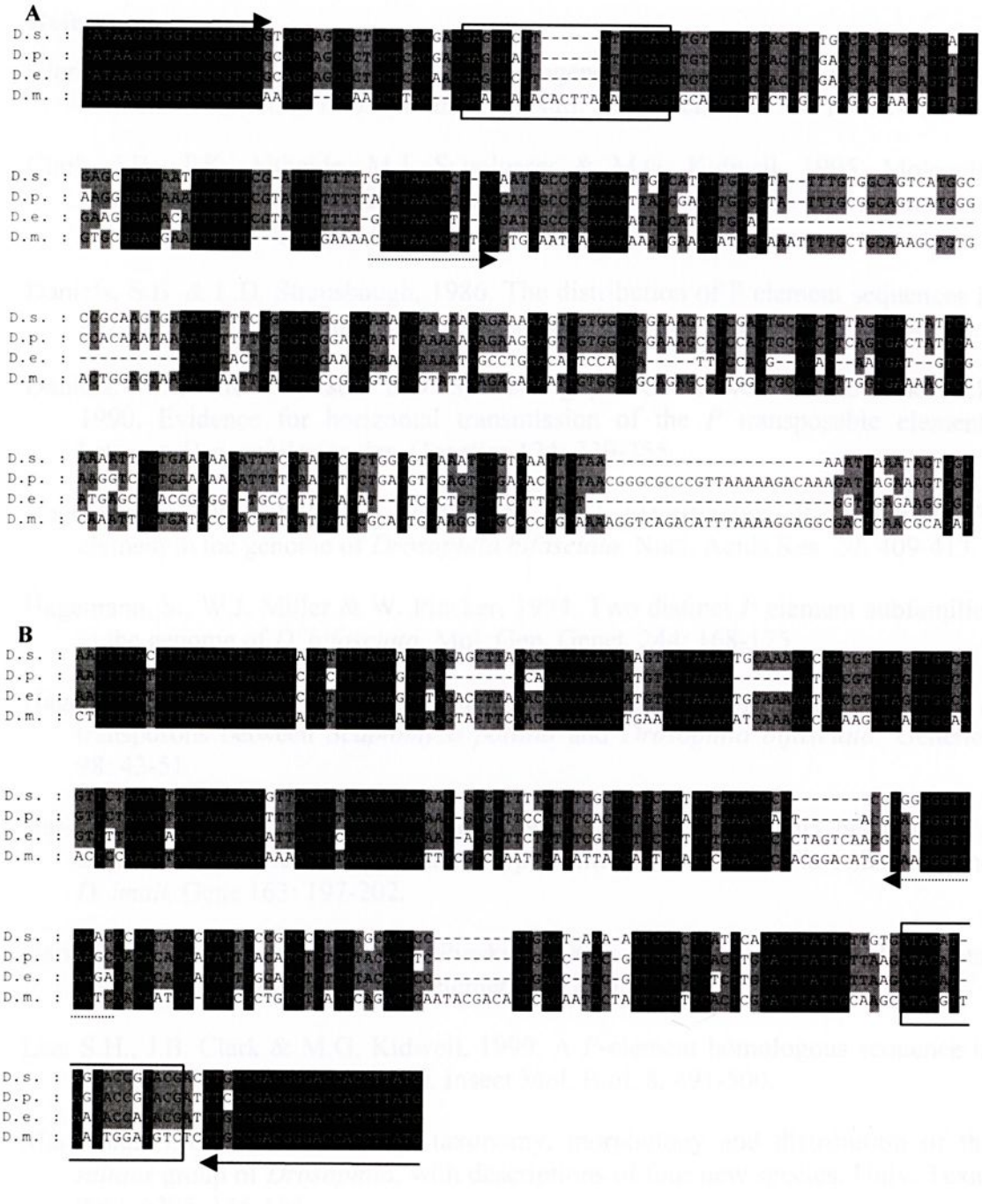


Figure 7. Partial alignment of nucleotide sequences from the 5' (**A**) and 3' (**B**) ends of different *P* homologues: D.s. = *D. sturtevanti* (AF530052), D.p. = *D. prosaltans* (AF530053), D.m. = *D. melanogaster* (pπ25.1) and D.e. = *D. emarginata*. Gaps are shown as dashes; background shade corresponds to the nucleotide conservation; terminal inverted repeats are marked by solid arrows; the 11-bp subterminal inverted repeats are marked by dotted arrows; and transposase binding sites are marked by solid boxes.

References

Clark, J.B. & M.G. Kidwell, 1997. A phylogenetic perspective on *P* transposable element evolution in *Drosophila.* Proc. Natl. Acad. Sci. USA 94: 11428-11433.

Clark, J.B., T.K. Altheide, M.J. Scholosser & M.G. Kidwell, 1995. Molecular evolution of *P* transposable elements in genus *Drosophila.* I. The *saltans* and *willistoni* species groups. Mol. Biol. Evol. 12: 902-913.

Daniels, S.B. & L.D. Strausbaugh, 1986. The distribution of P element sequences in *Drosophila*: the *willistoni* and *saltans* species group. J. Mol. Evol. 23: 138-148.

Daniels, S.B., K.R. Petterson, L.D. Strausbaugh, M.G. Kidwell & A.C. Chovnick, 1990. Evidence for horizontal transmission of the *P* transposable elements between *Drosophila* species. Genetics 124: 339-355.

Hagemann, S., W.J. Miller & W. Pinsker, 1992. Identification of a complete *P*-element in the genome of *Drosophila bifasciata.* Nucl. Acids Res. 20: 409-413.

Hagemann, S., W.J. Miller & W. Pinsker, 1994. Two distinct *P* element subfamilies in the genome of *D. bifasciata*. Mol. Gen. Genet. 244: 168-175.

Hagemann, S., E. Haring & W. Pinsker, 1996. Repeated horizontal transfer of *P* transposons between *Scaptomyza pallida* and *Drosophila bifasciata.* Genetica 98: 43-51.

Haring, E., S. Hagemann & W. Pinsker, 1995. Different evolutionary behavior of *P* element subfamilies: M-type and O-type elements in *Drosophila bifasciata* and *D. imaii.* Gene 163: 197-202.

Haring, E., S. Hagemann & W. Pinsker, 2000. Ancient and recent horizontal invasions of Drosophilids by *P* elements. J. Mol. Evol. 51: 577-586.

Lee, S.H., J.B. Clark & M.G. Kidwell, 1999. A *P*-element homologous sequence in the house fly, *Musca domestica.* Insect Mol. Biol. 8: 491-500.

Magalhães, L.E., 1962. Notes on taxonomy, morphology and distribution of the *saltans* group of *Drosophila*, with descriptions of four new species. Univ. Texas Publ. 6205: 135-154.

O'Hare, K. & G.M. Rubin, 1983. Structure of *P* element transposable and their sites of insertion and excision in *Drosophila melanogaster* genome. Cell 25: 24-35.

Patterson, J.T. & G.B. Mainland, 1944. The Drosophilidae of Mexico. Univ. Texas Publ. 4445: 1-101.

Perkins, H.D. & A.J. Howells, 1992. Genomic sequences with homology to the *P* element of *Drosophila melanogaster* occur in the blowfly *Lucilia cuprina.* Proc. Natl. Acad. Sci. USA 89: 10753-10757.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Silva, J.C. & M.G. Kidwell, 2000. Horizontal transfer and selection in the evolution of *P* elements. Mol. Biol. Evol. 17: 1542-1557.

Spassky, B.S., R.C. Richmond, S. Peréz-Salas, O. Pavlovsky, C.A. Mourão, A.S. Hunter, H. Hoenigsberg, T. Dobzhansky & F.J. Ayala, 1971. Geography of sibling species related to *Drosophila willistoni* and semispecies of the *Drosophila paulistorum* complex. Evolution 25: 129-143.

Saitou, N. & M. Nei, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406-425.

Swofford, D., 1997. PAUP: Phylogenetic analysis using parsimony. Version 4.0b10. Smithsonian Institution, Washington, D.C.

Thompson, J.D., D.G. Higgins & T.J. Gibson, 1994. CLUSTAL W: improving the sensitivity of progressive multiple alignment through sequence weighting, positions-specific gap penalties and weigh matrix choice. Nucl. Acids Res. 22: 4673-4680.

Throckmorton, L.H., 1975. The phylogeny, ecology and geography of *Drosophila*. Pp. 421-469 *in* R. C. King, ed. Handbook of genetics. Plenum Press, New York.

Caracterização de seqüências *P* completas em *D. sturtevanti* e *D. prosaltans*

Capítulo Quatro

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Resumo

Até o momento, somente seqüências parciais do elemento *P* haviam sido descritas para o grupo *saltans* de *Drosophila*. Neste trabalho, relatamos duas seqüências completas do elemento *P* em *D. sturtevanti* e *D. prosaltans*. A divergência nucleotídica destas seqüências em relação ao elemento *P* canônico de *D. melanogaster* é de aproximadamente 31%. Análises filogenéticas revelaram que ambos elementos pertencem a um clado de seqüências divergentes dos grupos *saltans* e *willistoni*, previamente descrito por outros autores.

Characterization of two full-sized *P* elements from *Drosophila sturtevanti* and *D. prosaltans*.

Juliana Polachini de Castro and Claudia M. A. Carareto[1]

Department of Biology, Universidade Estadual Paulista - UNESP, São José do Rio Preto – SP, Brazil.

Abstract

To date, only partial *P* element sequences have been reported in the *saltans* group of *Drosophila*. We report here two complete *P* element sequences from *D. sturtevanti* and *D. prosaltans*. The divergence of these sequences from the canonical *P* element of *D. melanogaster* is about 31% at the nucleotide level. Phylogenetic analysis revealed that both elements belong to a clade of divergent sequences from the *saltans* and *willistoni* groups previously described by other authors.

Running title: Complete *P* elements in *D. sturtevanti* and *D. prosaltans*

Key-words: *D. sturtevanti*, *D. prosaltans*, full-size *P* element, phylogeny

*** Artigo a ser submetido para publicação no Genetics and Molecular Biology**

[1] Present Address: Dr. Claudia M. A. Carareto, Departamento de Biologia, Universidade Estadual Paulista – UNESP, Rua Cristóvão Colombo 2265, CEP 15054-000, São José do Rio Preto – SP, Brazil. (55) 17 2212382, * (55) 17 2212390. E-mail: carareto@bio.ibilce.unesp.br.

Introduction

P elements were first discovered in *Drosophila melanogaster* because of their ability to induce hybrid dysgenesis (Kidwell, Kidwell and Sved, 1977). Autonomous *P* elements are 2.9 kb in length and have four open reading frames, which encode two polypeptides, an 87 kDa transposase enzyme necessary for transposition (Rio, Laski and Rubin, 1986) and a 66 kDa repressor protein (Robertson and Engels, 1989). Also required for transposition are the element termini, which include flanking 31-bp perfect inverted repeats (O'Hare and Rubin, 1983), 11-bp subterminal repeats, and unique terminal sequences comprising approximately 150 bp (see Engels, 1989 for a review).

Sequences belonging to the *P* family are particularly common in the four principal species groups comprising the subgenus *Sophophora*: *melanogaster*, *obscura*, *saltans* and *willistoni* groups (Daniels *et al.*, 1990), but have also been described in a few species outside of the subgenus *Sophophora* (Loreto *et al.*, 2001) and the genus *Drosophila* (Anxolabéhère and Périquet, 1987; Simonelig and Anxolabéhère, 1991, 1994; Perkins and Howells, 1992; Lee, Clark and Kidwell, 1999). Phylogenetic studies based on nucleotide sequences (Clark and Kidwell, 1997; Hagemann, Miller and Pinsker, 1994, 1996; Silva and Kidwell, 2000) indicated that the more than 200 *P* element sequences obtained to date fall into 16 distinct clades or subfamilies.

Four distinct *P* element subfamilies have been described and well characterized. The canonical subfamily appears to be restricted to the sophophoran New World species groups *saltans* and *willistoni* (Clark *et al.*, 1995), with a notable exception explained by horizontal transfer (Loreto *et al.*, 2001). The M-, O- and T-type subfamilies are found in the Old World *obscura* species group (Hagemann, Miller and Pinsker, 1992, 1994, 1996). The T-type seems to be restricted to the *obscura* lineage (Hagemann, Haring and Pinsker, 1998), while the M- and O-types have also been found in the *saltans* and *willistoni* groups.

The *P* element subfamilies in the *saltans* and *willistoni* species groups have been described so far based mainly on partial sequences (Clark *et al.*, 1995; Clark and Kidwell, 1997; Haring, Hagemann and Pinsker, 2000; Silva and Kidwell, 2000).



This work aimed to compare two complete sequences obtained from two different subgroups of *saltans* group to some of the complete sequences from different *P* element subfamilies, and to a data set consisting of 80 partial consensus sequences (provided by Clark and Kidwell, 1997) in order to establish their phylogenetic relationships.

Materials and Methods

Drosophila Stocks

The strains of *D. sturtevanti* from Matlapa, Mexico (Joana C. Silva) and *D. prosaltans* from Eldorado, State of Rio Grande do Sul, Brazil (Vera L. S. Valente) were recently derived from natural populations and subsequently maintained in the laboratory, under standard conditions.

DNA Amplification and Sequencing

Polymerase chain reactions were carried out with 100 ng of genomic DNA from both strains in a solution of 1 mM $MgCl_2$, 400 µM of each deoxynucleotide, 0.5 µM of primer M-IR (5'CATAAGGTGGTCCCGTCG3', corresponding to nt. 14-31 and 2877-2894, within the ITRs of the canonical *P* element; Haring, Hagemann and Pinsker, 1995), and 2 units of *Taq* polymerase in 1x polymerase buffer in a total volume of 50 µl. PCRs consisted of 30 cycles of 45 s of denaturation at 94°C, 45 s of primer annealing at 57°C, and 1.5 m of primer extension at 72°C. In the very first cycle there was a denaturation period of 7 min at 94°C, and in the very last, an elongation step of 10 min at 72°C. The PCR products were cloned into a TA cloning vector (Invitrogen) and both strands of a single clone, chosen at random, were sequenced for each species. The primers $Stu1_{369}$ (5'GTTCCGTATCGAGACCCGAC), $Stu2_{2402}$ (5'AATGACGAAGACTCGTCGCG), $Stu5_{1091}$ (5' GGAAGCAACCAGTTTTCTTT) and $Stu6_{1598}$ (5' CACATCAAACCAATCATTTA), designed based on the clone sequences were used to complete the element sequencing.

Sequence Alignment and Phylogenetic Analysis

A data set of 80 consensus partial *P*-element sequences of 429 bp long, mapped to positions 1328-1757 in ORF2 of the canonical *P* element was obtained from Clark and Kidwell (1997). Some full-length *P* element nucleotide sequences were obtained from literature: *D. melanogaster* pπ25.1 (O'Hare and Rubin, 1983); *D. nebulosa* N10 (Lansman *et al.*, 1987); *D. willistoni* (Daniels *et al.*, 1990); *D. guanche*, *D. bifasciata* and *D. bifasciata* jbifM3 (Hagemann, Miller and Pinsker, 1992); *D. ambigua* T-type (Hagemann, Haring and Pinsker, 1998); *D. helvetica* M-type (Haring, Hagemann and Pinsker, 2000); *D. mediopunctata* (Loreto *et al.*, 2001); *Scaptomyza pallida* (Simonelig and Anxolabéhère, 1991) and *Musca domestica* (Lee, Clark and Kidwell, 1999). The DNA sequences of the *D. sturtevanti* and *D. prosaltans P* elements obtained in this study have been deposited in GenBank (accession numbers AF530052 and AF530053, respectively).

Alignments of sequences were done by the program CLUSTAL W (Thompson, Higgins and Gibson, 1994) and improved by hand. The phylogenetic relationships among *P* sequences were constructed using the maximum parsimony method, using an heuristic search algorithm as implemented in PAUP v.4.0b10 (Swofford, 1997). Branch support was calculated by bootstrap analysis consisted of 500 replicates. The neighbor-joining method (Saitou and Nei, 1987) was also used to construct the distance matrix according to the Kimura two-parameter model of nucleotide substitution (Kimura, 1980).

Results and Discussion

Sequence features

The M-IR primer corresponds to positions 14-31 and 2877-2894 of the canonical *P* element, and so do not amplify the first and last 13 bp of a complete *P* element. If the TIRs of these elements are complete, the *D. sturtevanti* and *D. prosaltans P* elements are 2829 bp and 2828 bp in length, respectively, about 50 bp less than the canonical *P* element. The alignment of the two sequences against the *D. melanogaster* one showed that *D. sturtevanti* and *D. prosaltans P* elements are similar in structure and sequence to each other (12% different), but strongly



divergent from the canonical *P* element (31% different). Table 1 shows the main differences obtained from this alignment. The *D. sturtevanti P* element sequence has a total of 866 nucleotide substitutions, 113 deletions and 58 insertions. The *D. prosaltans P* element has a total of 971 nucleotide substitutions, 99 deletions and 35 insertions. In general, this species has the same deletions and insertions as *D. sturtevanti*.

Even though the TIRs were not completely sequenced, PCR amplification with specific primers to that region indicates that at least the second half of the TIRs are present and well conserved in *D. sturtevanti* and *D. prosaltans*. However, the transposase binding sites, located at positions 48-68 and 2855-2871 in *D. melanogaster*, the TATA box and the 11-bp subterminal IRs are not well conserved. In all four exons the translational reading frame is interrupted by stop codons and frameshift mutations, suggesting that these sequences do not encode a functional protein. Indeed, leucine-zipper and helix-turn-helix motifs were not detected, corroborating the suggestion that these sequences might be immobile in the genome.

The nucleotide differentiation and genetic difference matrix based on Kimura's two-parameter method were calculated among the full-length *P* element sequences from literature and the two sequences described here (Table 2). *D. sturtevanti* and *D. prosaltans* sequences present an overall divergence of 31% compared to the canonical sequences described in *D. melanogaster*, *D. willistoni*, *D. mediopunctata* and *D. nebulosa*.

Phylogenetic Analysis

The results of phylogenetic analyses of *P* elements in the subgenus *Sophophora* (Clark *et al.*, 1995; Clark, Kim and Kidwell, 1998; Clark and Kidwell, 1997; Silva and Kidwell, 2000) have provided evidence for the existence of multiple *P* elements subfamilies in lineages of single species that apparently must have entered the genome at different times during the past (Lee, Clark and Kidwell, 1999; Haring, Hagemann and Pinsker, 2000). The aim of the phylogenetic analysis was to determine if the sequences obtained of *D. sturtevanti* and *D. prosaltans* belonged to some of the well-characterized *P* element subfamilies. Figure 1 summarizes the

results of phylogenetic analysis of *P* element sequences using parsimony. Neighbor-joining and parsimony analyses gave identical branching patterns for the major groups (data not shown) and clustered the *D. sturtevanti* and *D. prosaltans* sequences in the clade called F by Clark and Kidwell (1997) in 100% of bootstrap replicates. This clade comprises *P* sequences of some other species of the *saltans* group and some species of the *willistoni* group.

Clark *et al.* (1995) and Clark and Kidwell (1997), based on an internal portion of the *P* element exon 2, grouped the *P* elements of *saltans* group in four different clades or subfamilies. These authors found that the divergence within the other three *saltans-willistoni* clades, excluding the canonical *P* element clade, ranges from 17% to 30% to 46%. Our sequences belong to the Clark and Kidwell's (1997) clade F, with a nucleotide divergence varying from 6% to 17% within the clade. Once these authors found only few sequences belonging to this clade they wondered that this frequency could be due to the scarce presence of this subfamily in these flies genome but probably by the sampling done with the PCR primers. Nevertheless, our data suggest that the subfamily grouped into the F clade might be more widely distributed in the *saltans* group than previously believed because, notwithstanding sequencing only one clone, this element was sampled in two different species in spite of the existence of the canonical sequence in these genomes.

Acknowledgements

We are indebted to Drs. M.G. Kidwell and A.J.Holyoake who kindly provided advice and laboratory facilities at the University of Arizona and for their critical comments on earlier versions of the manuscript. This work was supported by FAPESP (Grants n° 95/7192-2, 98/08734-1 and 00/11313-0 and a fellowship to J.P.C.) and CNPq (fellowship to CMAC).

Table 1. Differences observed between *Drosophila sturtevanti* and *D. prosaltans P* element sequences and that of the canonical *P* element of *D. melanogaster*.

Characteristics		*D. sturtevanti*	*D. prosaltans*
Sequence length		2829 bp	2828 bp
Total of substitutions		866	971
Total of deletions		113	99
Total of insertions		58	35
Main indels	**E0**	+2; +2; -6; +3; -1; -2; -22; -2; -6; -5	+2; +2; -6; +4; -1; -2; -2; -6; -5
	E1	+2; -1; +1; +1; -1	+2; -1; +2; -1; +2; -1
	E2	-13	
	E3	-2; -1; +3; -1; -2; -3; +1; +2; -1; +2; +13; +1; +4; -3; -1; -5; -3; +1; -10; -1	-2; -1; +3; -1; -2; -3; +1; +2; -2; +1; +4; -1; -3; -1; -7; -6; -1; -3; -3; +1; -10; -1
	I1	+3; +5; +1; +5; -1	+3; +5; +1; -1
	I2	+5	+2
	I3	+1; -4; -1; -1; -10; -2	-2; -4; -1; -1; -10; -2

Notes: E0= exon 0; E1= exon 1; E2= exon 2; E3= exon 3; I1= intron 1; I2= intron 2 e I3= intron 3. The size and nature of the insertion (+) or deletion (-) are denoted separately when the indels are noncontiguous.

Table 2. Proportion of differences in complete *P* element nucleotide sequences (above) and genetic distances calculated using Kimura's two-parameter method (below) among full-length *P* elements described here and those from other *Drosophila* species, *Scaptomyza pallida* and *Musca domestica*.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 *D. melanogaster*		0.00	0.03	0.03	0.29	0.19	0.21	0.32	0.37	**0.31**	**0.31**	0.20	0.48
2 *D. willistoni*	0.00		0.03	0.03	0.29	0.19	0.21	0.32	0.37	**0.31**	**0.31**	0.20	0.48
3 *D. nebulosa*	0.03	0.03		0.04	0.30	0.20	0.21	0.33	0.38	**0.31**	**0.31**	0.21	0.49
4 *D. mediopunctata*	0.03	0.03	0.04		0.30	0.20	0.21	0.33	0.37	**0.31**	**0.31**	0.21	0.48
5 *D. bifasciata*	0.37	0.37	0.38	0.38		0.28	0.28	0.33	0.38	**0.32**	**0.33**	0.29	0.51
6 *D. bifasciata* M	0.23	0.23	0.24	0.23	0.35		0.06	0.32	0.35	**0.28**	**0.29**	0.07	0.47
7 *D. helvetica*	0.24	0.24	0.25	0.25	0.36	0.07		0.32	0.35	**0.30**	**0.30**	0.09	0.47
8 *D. ambigua*	0.42	0.42	0.44	0.43	0.44	0.41	0.42		0.35	**0.34**	**0.34**	0.33	0.51
9 *D. guanche*	0.51	0.51	0.53	0.50	0.52	0.48	0.48	0.48		**0.38**	**0.39**	0.29	0.54
10 *D. prosaltans*	**0.40**	**0.40**	**0.41**	**0.41**	**0.42**	**0.37**	**0.39**	**0.45**	**0.54**		**0.12**	0.30	0.51
11 *D. sturtevanti*	**0.40**	**0.40**	**0.41**	**0.40**	**0.43**	**0.36**	**0.39**	**0.46**	**0.56**	**0.39**		**0.29**	**0.51**
12 *S. pallida*	0.24	0.24	0.25	0.25	0.36	0.07	0.10	0.43	0.50	**0.39**	**0.13**		**0.48**
13 *M. domestica*	0.77	0.77	0.78	0.76	0.85	0.75	0.75	0.84	0.96	**0.86**	**0.86**	0.76	

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

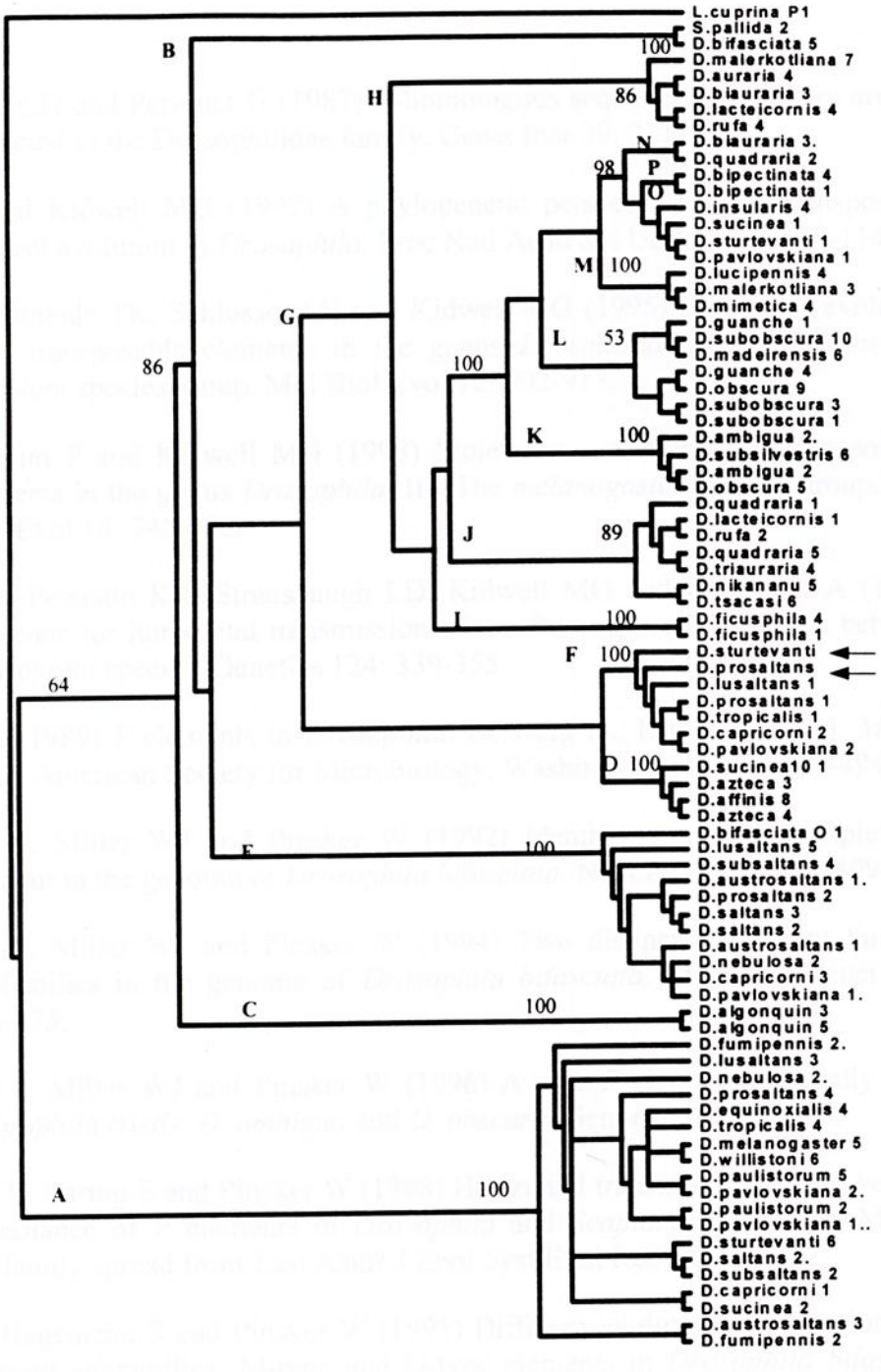


Figure 1. Cladogram of *P* element nucleotide sequences obtained in this study and those obtained from Clark and Kidwell (1997). Species names are given followed by the total number of clones used to obtain a consensus sequence or, for the sequences obtained in this study, followed by an arrow. This is a consensus tree of 1,000 equally parsimonious trees, each requiring 2,193 steps. Of 449 characters, 34 were constant, 60 uninformative, and 355 parsimony informative. The consistency index is 0.40 and the retention index 0.815. Bootstrap analysis of 500 replicates was performed on the data.

References

Anxolabéhère D and Périquet G (1987) *P*-homologous sequences in Diptera are not restricted to the Drosophilidae family. Genet Iber 39: 211-222.

Clark JB and Kidwell MG (1997) A phylogenetic perspective on *P* transposable element evolution in *Drosophila.* Proc Natl Acad Sci USA 94: 11428-11433.

Clark JB, Altheide TK, Schlosser MJ and Kidwell MG (1995) Molecular evolution of *P* transposable elements in the genus *Drosophila*. I. The *saltans* and *willistoni* species group. Mol Biol Evol 12: 902-913.

Clark JB, Kim P and Kidwell MG (1998) Molecular evolution of *P* transposable elements in the genus *Drosophila*. III. The *melanogaster* species group. Mol Biol Evol 15: 746-755.

Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG and Chovnick A (1990) Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. Genetics 124: 339-355.

Engels WR (1989) *P* elements in *Drosophila*. *In:* Berg D., Howe M. (eds). *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 437-484.

Hagemann S, Miller WJ and Pinsker W (1992) Identification of a complete *P*-element in the genome of *Drosophila bifasciata.* Nucl Acids Res 20: 409-413.

Hagemann S, Miller WJ and Pinsker W (1994) Two distinct *P* element families subfamilies in the genome of *Drosophila bifasciata.* Mol Gen Genet 244: 168-175.

Hagemann S, Miller WJ and Pinsker W (1996) A new *P* element subfamily from *Drosophila tristis, D. ambigua* and *D. obscura.* Genome 39: 978-985.

Hagemann S, Haring E and Pinsker W (1998) Horizontal transmission versus vertical inheritance of *P* elements in *Drosophila* and *Scaptomyza*: has the M-type subfamily spread from East Asia? J Zool Syst Evol Res 36: 75-83.

Haring E, Hagemann S and Pinsker W (1995) Different evolutionary behavior of *P* element subfamilies: M-type and O-type elements in *Drosophila bifasciata* and *D. imaii.* Gene 163: 197-202.

Haring E, Hagemann S and Pinsker W (2000) Ancient and recent horizontal invasions of drosophilids by *P* elements. J Mol Evol 51: 577-586.

Kidwell MG, Kidwell JF and Sved JA (1977) Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics 86: 813-833.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative analysis of nucleotide sequences. J Mol Evol 16: 111-120.

Lansman RA, Shade RO, Grigliatti TA and Brock HW (1987) Evolution of *P* transposable elements: Sequences of *Drosophila nebulosa P* elements. Proc Natl Acad Sci USA 84: 6491-6495.

Lee SH, Clark JB and Kidwell MG (1999) A *P* element-homologous sequence in the house fly, *Musca domestica.* Insect Mol Biol 8: 491-500.

Loreto ELS, Valente VLS, Zaha A, Silva JC and Kidwell MG (2001) *Drosophila mediopunctata P* elements: a new example of horizontal transfer. The J Hered 92: 375-381.

O'Hare K and Rubin GM (1983) Structure of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell 34: 25-35.

Perkins HD and Howells AJ (1992) Genomic sequences with homology to the *P* element of *Drosophila melanogaster* occur in the blowfly *Lucilia cuprina.* Proc Natl Acad Sci USA 89: 10753-10757.

Rio DC, Laski FA and Rubin GM (1986) Identification and immunochemical analysis of biologically active *Drosophila P* element transposase. Cell 44: 21-32.

Robertson HM and Engels WR (1989) Modified P elements that mimic P cytotype in *Drosophila melanogaster*. Genetics 123: 815-824.

Saitou N and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425.

Silva JC and Kidwell MG (2000) Horizontal transfer and selection in the evolution of *P* elements. Mol Biol Evol 17: 1542-1557.

Simonelig M and Anxolabéhère D (1991) A *P* element of *Scaptomyza pallida* is active in *Drosophila melanogaster*. Proc Natl Acad Sci USA 88: 6102-6106.

Simonelig M and Anxolabéhère D (1994) *P* elements are old components of the *Scaptomyza pallida* genome. J Mol Evol 38: 232-240.

Swofford D (1997) PAUP: Phylogenetic analysis using parsimony. Version 4.0b10. Smithsonian Institution, Washington, D.C.

Thompson JD, Higgins DG and Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple alignment through sequence weighting, positions-specific gap penalties and weigh matrix choice. Nucl Acids Res 22: 4673-4680.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Transcrição de elementos *P* canônicos no grupo *saltans*

Capítulo Cinco

unesp

Resumo

Até o momento, as investigações sobre a expressão e a regulação do elemento transponível *P* têm sido quase que exclusivamente realizadas com o elemento *P* canônico de *Drosophila melanogaster*. Analisando oito espécies diferentes de moscas do grupo *saltans*, foram detectados mRNAs para transposase em tecidos germinativos de *D. saltans* e *D. prosaltans*, e mRNAs para proteína repressora em tecidos somáticos de *D. saltans* e *D. sturtevanti*. As análises de seqüenciamento sugerem que estes transcritos devam pertencer à subfamília canônica e que eles possam ser transposicionalmente ativos somente em *D. saltans*. Os valores de d_N e d_S para *Adh* e para o elemento *P* sugerem que as seqüências encontradas em *D. saltans* e em *D. prosaltans* poderiam estar presentes no ancestral do subgrupo *saltans* e que a seqüência encontrada em *D. sturtevanti* pode ter sido transmitida horizontalmente para esta espécie por *D. saltans*.

Canonical *P* elements are transcriptionally active in the *saltans* group of *Drosophila*

Juliana Polachini de Castro and Cláudia Márcia A. Carareto[1]

Department of Biology, Universidade Estadual Paulista - UNESP, São José do Rio Preto –SP, Brazil.

Abstract

Up to now, investigations of expression and regulation of *P* transposable element have been almost exclusively carried out with the *Drosophila melanogaster* canonical *P* element. Analyzing eight different species of the *saltans* group, we detected transposase mRNA in germline tissues of *D. saltans* and *D. prosaltans*, and repressor mRNA in somatic tissues of *D. saltans* and *D. sturtevanti*. Sequencing analysis suggested that these transcripts might belong to the canonical subfamily and that they can be transpositionally active only in *D. saltans*. d_N and d_S values of *Adh* and *P* element suggested that the sequences found in *D. saltans* and *D. prosaltans* might have been present in the ancestor of the *saltans* subgroup and that the sequence found in *D. sturtevanti* might have been horizontally transferred from *D. saltans*.

Key words: transcription - *Drosophila saltans* - *P* element – mRNA – transposase – repressor – horizontal transfer.

*** Artigo submetido à publicação no Journal of Molecular Evolution**

[1] Correspondence to: Dr. Cláudia M. A. Carareto, Departamento de Biologia, Universidade Estadual Paulista – UNESP, Rua Cristóvão Colombo 2265, CEP 15054-000, São José do Rio Preto – SP, Brazil. (55) 17 2212382, * (55) 17 2212390. E-mail: carareto@bio.ibilce.unesp.br.

Introduction

The *P* element is a transposable element with short inverted terminal repeats which transposes by a DNA-only pathway. The canonical *P* element was initially discovered in *Drosophila melanogaster* as the causative agent responsible for the P-M hybrid dysgenesis, a phenomenon characterized by high rates of mutations, chromosomal rearrangements, male recombination and a thermosensitive agametic sterility (GD sterility), as a consequence of *P* element mobilization (Kidwell *et al.*, 1977). These traits are found in the progeny of a cross between males from a *P* element-containing strain (P strain) and females devoid of *P* elements (M strain). The progeny of the reciprocal cross does not exhibit any of the dysgenic traits.

P elements are of two structurally distinct types: large full-sized or autonomous elements 2,907 bp in length, and smaller non-autonomous elements that are heterogeneous in size, ranging from about 500-2,500 bp. Non-autonomous elements on the whole are formed by internal deletions of larger elements, probably due imprecise excision during transposition (O'Hare and Rubin, 1983). Some specific 5' and 3' *P* element sequences are required for transposition and excision, including perfect 31 bp inverted terminal repeats and an 11 bp sequence found in inverted orientation approximately 130 bp from each end (O'Hare and Rubin, 1983; Mullins *et al.*, 1989).

D. melanogaster full-sized 2.9 kb *P* element has four exons that encode for at least two different proteins produced by differential splicing of the primary transcript (Misra and Rio, 1990). The 87 kDa transposase is made exclusively in germline cells and mediates transposition via a cut-and-paste mechanism and, as recently shown, via "alternative transposition" (Gray, 2000). The second protein is translated from an mRNA that still retains the third intron. Owing to a stop codon in intron 3, a shorter, 66 kDa, protein is formed that acts as a repressor of transposition, inhibiting the genomic mobility of *P* elements in somatic cells (Rio, 1990). In somatic tissues, the host-encoded factor PSI (*P*-element somatic inhibitor protein) prevents splicing of the *P* element intron 3. The mechanism of differential splicing restricts transpositional activity to the germline, minimizing deleterious effects on the host (Laski *et al.*, 1986; Laski and Rubin, 1989). Alternative double-spliced variants, in

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

which two small introns are removed within the IVS3 region, have been also described in *D. melanogaster*, *D. bifasciata* and *D. helvetica* (Chain *et al.*, 1991; Haring *et al.*, 1998).

The *saltans* group is one of the four major *Drosophila* species groups placed in the subgenus *Sophophora* (Sturtevant, 1942). Throckmorton (1975) considered the neotropical *saltans* and *willistoni* species groups to be distinct diverged lineages within *Sophophora*, clearly separated from the Old World *melanogaster* and *obscura* species groups. The *saltans* species group consists of 21 species, which are divided into the *cordata*, *elliptica*, *parasaltans*, *sturtevanti* and *saltans* subgroups on the basis of a variety of morphological characters (Spassky, 1957; Magalhães and Bjornberg, 1957; Magalhães, 1962; Mourão and Bicudo, 1967; Throckmorton and Magalhães, 1962; Throckmorton, 1975). Throckmorton (1975) proposed that the ancestor of the *saltans* species group originated in tropical North America, where the so-called "primitive" *cordata* and *elliptica* subgroups are found. This ancestral group colonized the South American continent and the *sturtevanti*, *saltans* and *parasaltans* subgroups then diversified prior to formation of the Isthmus of Panama. According to this author, some members of the *saltans* subgroup, such as *D. saltans* and *D. prosaltans*, have recently diffused back into North America, probably within the past 4.5 Myr.

In the *elliptica* and *cordata* subgroups all species surveyed are devoid of any discernible *P* elements; however, species from the more derived *parasaltans*, *sturtevanti* and *saltans* subgroups all contain *P*-homologous sequences (Daniels *et al.*, 1990; Clark *et al.*, 1995). Daniels *et al.* (1990) suggested that *P* elements could have spread throughout the *saltans* and *willistoni* groups in South America, at a time when the *elliptica* and *cordata* subgroups were still geographically isolated in North America. Alternatively, Clark *et al.* (1995) proposed that *P* elements could have been lost from the *elliptica* and *cordata* subgroups early in the history of the *saltans* group, when those two subgroups were isolated from the lineages leading to the more recently derived *saltans* subgroups. However, we now know that the spread of canonical *P* elements within the group took place within the last 3 Myr, at a time when North and South America were no longer disconnected (Silva and Kidwell,



2000). The current distributions of the *elliptica* and *cordata* subgroups throughout Central and South America overlap those of the other *saltans* and *willistoni* subgroups (Patterson and Mainland, 1944; Magalhães, 1962; Spassky *et al.*, 1971) many or all of which contain canonical *P* elements (Clark *et al.*, 1995). Therefore, the absence of canonical *P* elements in the *elliptica* and *cordata* subgroups does not seem to be due to their geographical isolation. An alternative hypothesis is that there might be host factors in the *elliptica* and *cordata* lineages that confer resistance to *P*-element invasion (Silva and Kidwell, 2000).

Presently, little is known about the expression and repression of *P* transposable elements in species other than *D. melanogaster* and some species of the *obscura* group. The *P* elements of *Drosophila saltans* group has been shown previously to belong to at least four different clades or subfamilies: the canonical subfamily also found in *D. melanogaster*, a second and third well-characterized subfamilies termed O-type and M-type, respectively, and a fourth clade or subfamily that is still under study (Clark *et al.*, 1995; Haring *et al.*, 1995; Clark and Kidwell, 1997; Haring *et al.*, 2000; Castro and Carareto, unpublished data). In spite of the 30% divergence at the sequence level among different subfamilies, the basic molecular structure of full-sized elements is the same in all three types. The *termini* are formed by inverted repeats and the coding region comprises four exons.

With the exception of *D. melanogaster* (Daniels *et al.*, 1990) and *D. mediopuntacta* (Loreto *et al.*, 2001), the canonical subfamily of *P* elements appears to be restricted to the sophophoran New World *saltans* and *willistoni* species groups. Up to now, investigations of expression and regulation have been almost exclusively carried out with the *D. melanogaster* canonical *P* element, and nothing is known about the regulation of these elements in the *saltans* and *willistoni* species groups. This work aimed to investigate whether the canonical *P* elements of the *saltans* species group are transcribed and/or repressed, in a similar manner to those in *D. melanogaster*. Here we also describe the sequences obtained in this study and compare them with the canonical ones.



Materials and Methods

Fly Stocks

All the species and strains used in the present study are listed in Table 1. All fly strains were derived from a single, randomly selected female from a mass culture. *D. melanogaster* – Harwich, a strong P line, was used as positive control.

RNA isolation and reverse transcription

Total RNA was isolated from 16 ovaries (somatic and germline tissues) and 16 heads (somatic tissues) using TRIZOL Reagent (GIBCO – BRL Technologies) according to the manufacturer's instructions. Canonical *P* element specific cDNA was generated from total RNA using Superscript II RT (GIBCO – BRL Technologies) with an antisense primer (Mel e3-) binding within canonical *P* exon 3 (Figure 1 and Table 2) at low stringency (42°C).

PCR reactions

The first round of PCR amplification was performed, according to Haring *et al.* (1998), using the exon-3-specific primer (Mel e3-) and a primer spanning intron 0-1 (Mel i1+), generating a PCR product comprising all three *P* splicing junctions (Figure 1 and Table 2). PCR was done in a volume of 50 µl with 2.5 units Taq DNA polymerase in 1X polymerase buffer, 20 nmol of each primer, 200 µM of each dNTP and 2 mM of $MgCl_2$. The solutions were heated to 95°C for 5 min and then cycled through 35 reactions of: 95°C for 30 s, 58°C for 30 s, 72°C for 2.5 min, followed by a final extension at 72°C for 10 min.

With the exception of *D. melanogaster*, another round of PCR amplification with internal primers was necessary in the other species to obtain products enough to be visualized in agarose gels. These re-amplifications were carried out with intron-spanning primers combined as following: Mel i1+ and Mel i3-, Mel i2+ and Mel i3-, and Mel i2+ and Mel e3- (Figure 1 and Table 2). The PCR products obtained in these re-amplifications include the entire IVS3 region, allowing one to differentiate between alternative splicing variants. Since intron-spanning primers are complementary to the 5'and 3' adjacent sequences of the respective intron, they bind



to processed mRNA only but not to occasional contaminating genomic DNA or pre-mRNA. PCR was done in a volume of 25 µl with 0.5 units Taq DNA polymerase in 1X polymerase buffer, 20 nmol of each primer, 200 µM of each dNTP and 2 mM of $MgCl_2$. The solutions were heated to 94°C for 2 min and then cycled through 35 reactions of: 94°C for 1 min, 56°C for 1 min, 72°C for 1 min, followed by a final extension at 72°C for 10 min.

DNA Sequencing and Sequence Analyses

PCR products were separated on 1% agarose gels. DNA was extracted from the gel using the Concert Matrix System (GIBCO–BRL Technologies) and cloned into a TOPO-TA vector (Invitrogen). Both strands of each clone were sequenced for each species, using an ABI 377 automated sequencer (University of Arizona DNA Sequencing Service, Tucson – AZ).

The *P* element data set consisted of 80 *P*-element partial sequences was obtained from Clark and Kidwell (1997). The sequences were 429 nt long and mapped to positions 1328-1757 in ORF2 of the canonical *P* element (O'Hare and Rubin, 1983). These sequences were a consensus of 250 original sequences covering *P* elements across a broad taxonomic range (Lansman *et al.*, 1987; Daniels *et al.*, 1990; Simonelig and Anxolabéhère, 1991; Paricio *et al.*, 1991; Perkins and Howells, 1992; Miller *et al.*, 1992; Hangemann *et al.*, 1992, 1994; Clark and Kidwell, 1997). *Adh* sequences were used to compare the divergence of nucleotide sequence of a host gene with those of the *P* element transcripts. The *Adh* partial sequences were obtained for nine species of the *saltans* group from O'Grady *et al.* (1998) and consisted of 405 nt corresponding to the complete second exon of the gene.

Alignments of sequences were done by the program CLUSTAL W (Thompson *et al.*, 1994) and improved by hand. The protein sequences were predicted using 6-frame translation (http://dot.imgen.bcm.tmc.edu:9331/seq-util/seq-util.html). The phylogenetic relationships among *P* sequences were constructed using the maximum parsimony method, using an heuristic search algorithm as implemented in PAUP v.4.0b10 (Swofford, 1997). Branch support was calculated by bootstrap analysis consisted of 500 replicates.



The number of synonymous substitutions per synonymous site, d_S, and non-synonymous substitutions per non-synonymous site, d_N, were estimated for the sequence of exon 2 using the method of Nei and Gojobori (1986) implemented in MEGA 2 (Kumar *et al.*, 2001). In order to preserve the reading frame of the functional canonical *P* element, the alignment gaps introduced in the data set were deleted prior to estimation of d_S, and d_N. Also, the region comprising the deletion of 268 nt of *D. prosaltans* was excluded in all sequences for this analysis. The effective number of codons (N_c; Wright, 1990) was determined for each sequence using the codonW program (Peden, 1997). N_c varies between 21 for maximum codon bias and 61 for minimum codon bias. The frequency of G and C nucleotides in third positions in synonymous codons was determined as the proportion of C- and G-ending codons, with the exclusion of termination, methionine, and tryptophan codons. The primer binding regions were excluded from the sequences in all proceeded analyses.

Results

The aim of this study was to investigate whether different types of canonical *P* element mature mRNAs are produced in *saltans* species group through tissue-specific splicing, as is the case in *D. melanogaster*. RNA samples from heads and ovaries were prepared and after reverse transcription with the primer Mel e3-. The first PCR amplification was carried out with the primer pair Mel i1+/Mel e3- and re-amplification was done with several primer sets (Figure 1 and Table 2). Based on the splicing pattern found in *D. melanogaster*, transposase mRNA (IVS3 excised) was expected to occur in ovaries only and, as the gonads are composed of both germline and somatic cells, repressor mRNA (containing IVS3) was also expected. On the other hand, head samples should contain repressor mRNA only. Tests with genomic DNA revealed that none of the intron-spanning primers bind nonspecifically to DNA (data not shown).

Products of amplification

The fragments amplified from presumptive transposase and repressor mRNAs (Figure 2) were cloned and sequenced. The results showed that with the exception of

some of the *D. saltans*, *D. prosaltans* and *D. sturtevanti* products, none of the other species have detectable *P* elements transcripts, or transcripts conserved enough to be amplified by the canonical primers. In the *D. saltans* lines both transposase and repressor transcripts were amplified, in the *D. prosaltans* line only a transposase-like mRNA was detected and in one of the two *D. sturtevanti* lines studied (SMX) only the repressor transcript was amplified (Figure 2A-C).

Figure 2A shows the amplification of a 752 bp transposase-related fragment in ovary samples of all *D. saltans* lines (S7, S3, S6 and S10) and a smaller fragment, due to internal deletions, in *D. prosaltans* line (PE2). Figure 2B shows a 1370 bp repressor-related fragment amplified in ovary and head samples of S7 and heads of S6 and S3, a smaller fragment in ovary sample of *D. sturtevanti* SMX and in heads of S3. This difference in size is probably due to internal deletions in the *P* element sequence. Figure 2C shows the presence of a 1419 bp transposase-like fragment in ovaries of S7, and S3, and again a smaller fragment in ovary samples of PE2. The other weak and strong fragments amplified in ovaries and heads (for example, ovary sample of API in Fig. 2B and ovary sample of S6 in Fig. 2C) were also sequenced but did not have any *P* element similarity. These fragments might be due to amplifications of unrelated sequence caused by primer mis-annealing.

Sequences description

The fragments obtained were cloned and sequenced. The sequences of each strain were aligned and combined in order to produce a fragment comprising all amplified regions and are deposited in GenBank (accession numbers AF532567 to AF532577). The main characteristics of each sequence are described in Table 3. *D. saltans* S7 line presented two different clones for transposase fragments (1 and 2) and two different repressor sequences, one in ovary and another in head samples. The other lines presented only one clone sequenced. Only *D. prosaltans* sequence presented deletions, but most of the sequences showed big insertions, especially in exon 3. These insertions vary from eight to 54 nucleotides and are basically derived from duplications of the sequence 'atgatgagcctgtcg' or from part of it. Despite the indels, all transposase sequences showed approximately 97% nucleotide sequence similarity



relative to the canonical *P* element of *D. melanogaster*, and the repressor sequences showed about 90% similarity to canonical *P*. The transposase transcripts of *D. saltans* S7 are only 2.86% divergent from each other and the nucleotide divergence between the repressor transcripts is of 2%, suggesting that they might be products of the same *P* element subfamily in that genome. The same situation is observed between the sequences obtained in ovaries and heads of S6 and S3 lines. Additionally, we also calculated the divergence values among the sequences obtained in this study and those of *D. helvetica* (M-type) and *D. bifasciata* (M- and O-type) described as transcriptionally active by Haring *et al.* (1998). The values varied from 21% to 31%, suggesting that they do not belong to the same subfamily.

The protein sequences were predicted based on the amplified sequences. The protein sequences, in general, are quite conserved (80% to 97% identity with *D. melanogaster* transposase and repressor proteins). In the *D. prosaltans* and S7 (o2) transposase sequences and in the S7 (h), S6 (h) and *D. sturtevanti* repressor sequences some premature termination codons are present, which suggest that these proteins might have lost their function. Some of the protein motifs as leucine-zipper and helix-turn-helix could be detected in some of the sequences and most of them were conserved or partially conserved as shown in Table 3.

There are two possibilities for *P* element sequences to be strongly conserved among different taxa: they are under some kind of strong selection, or they have been transferred relatively recently from one taxon to the other. The strength of selective constraints on *P*-element sequences was investigated by comparing the estimated sequence divergence among species with that obtained for the *Adh* gene. By this analysis, ancestral polymorphism can be assumed if the divergence of TE nucleotide sequences is higher than those observed for host genes. On the other hand, if divergence between TE sequences are significantly lower than that observed for the host genes, one explanation that could be invoked is that a horizontal transfer event has happened. However, there are many possible factors, such as variable evolution rates, leading to a low sequence divergence among TE sequences, and horizontal transfer should be accepted after carefully excluding all other possibilities. Estimates of d_N and d_S for *Adh* were 0.035 ± 0.001 and 0.167 ± 0.012, respectively, with a d_N :

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

d_S ratio of 0.251. N_c value for *Adh* sequences was 46.01 ± 1.27. The number of substitutions per synonymous and nonsynonymous sites was estimated for all pairwise comparisons among the 11 *P* element sequences, generating medium values of 0.035 ± 0.006 for d_N and 0.038 ± 0.048 for d_S. d_N was smaller than d_S in 26 out of the 55 pairwise comparisons between sequences, with a median value of 0.34 for the ratio d_N/d_S. N_c value for *P* element sequences was 56.41 ± 0.68.

We also performed the same analysis comparing the sequences obtained in *D. saltans* and *D. prosaltans* (*saltans* subgroup) with that obtained in *D. sturtevanti* (*sturtevanti* subgroup). It is possible to note in Figure 3 that both d_N and d_S values are lower for *Adh* than for *P* only in the comparison between *D. saltans* and *D. prosaltans*. Additionally, *P* element divergence is similar in the last two comparisons, and twice higher than the comparison between *D. saltans* and *D. sturtevanti* sequences.

Phylogenetic analysis

The transcribed sequences were inserted in the *P*-element data set obtained from Clark and Kidwell (1997) in order to confirm their similarity to canonical sequences described to date. The phylogenetic analysis grouped all the transcripts, even the most divergent, in the canonical clade in 100% of 500 bootstrap replicates (Figure 3).

Discussion

The presence of transcripts

P-homologous sequences have been identified in the *saltans* group in the *parasaltans*, *sturtevanti* and *saltans* subgroups (Daniels *et al.*, 1990; Clark *et al.*, 1995), but, to date, there was no data available about the regulation of *P* element in this species group. This work describes the presence of transcriptionally active canonical elements in the *saltans* species group. Using primers complementary to canonical sequences, we found that transposase mRNA has been transcribed in germline tissues of *D. saltans* and *D. prosaltans*, and repressor mRNA (coding for the 66 kDa protein) has been transcribed in somatic tissues of *D. saltans* and *D. sturtevanti*. Since there were transposase products amplified in *D. saltans* and *D.*

BIBLIOTECA
Inst. de Biociências, Letras e Ciências Exatas - UNESP - S. J. Rio Preto

prosaltans, it can be concluded that the three splicing junctions (IVS1, IVS2 and IVS3) are conserved and located at positions homologous to those described to the *D. melanogaster P* element. In the other species from the *saltans* group any transcript was detected. However, it does not ensure that some *P* elements with very low transcriptional activity, due to positional effect or other factors, could be present in some of these species and were not detected. In addition, it is interesting to note that in *D. melanogaster* only one round of RT-PCR was needed. This could be an indication that the transcription rate in the *saltans* group is low, probably due to the positional effect or whatever epigenetic trans-silencing, or even because the repression mechanism is active. This fact could also be related to a relatively smaller copy number of full-sized *P* element in these species than in *D. melanogaster* (Daniels and Strausbaugh, 1986; Castro and Carareto, unpublished data).

Haring *et al.* (1998), using the same technique employed here, showed that in *D. helvetica* M-type elements are transcribed, and that in *D. bifasciata* and *S. pallida* both O-type and M-type elements are transcribed. They also found an alternative splicing pattern in *D. bifasciata*, *D. helvetica,* and *D. melanogaster,* in which two introns are removed from the boundaries of IVS3, whereby for the second one the usage of two alternative 5' sites results in two double-spliced mRNA variants. One of them retains an open reading frame for a protein 12 aa longer than the predicted transposase. In *D. helvetica* and in *D. melanogaster* double splicing was germline specific, like the complete excision of IVS3. In *D. bifasciata* specific-specific double splicing was observed only with O-type transcripts, whereas double-spliced M-type variants were found in all tissues. In the *saltans* group, this double splicing was not observed for any of the obtained sequences, although the entire region comprising the branch points and splice sites are highly conserved. Our results suggest that the *P* element-encoded repression systems of *D. saltans* are similar to those of *D. melanogaster*, with exception that the double-spliced variant that occurs in *D. melanogaster* was not detected in any species in our study. From an evolutionary point of view, the mechanisms of *P* element regulation (alternative splicing, specific specificity) are similar between the different species of the *saltans* group as well between these species and *D. melanogaster*. The same behavior was found by Haring

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

et al. (1998) for M- and O-type elements, even the sequence divergence between them being high (30%).

The results indicate that *D. saltans* canonical *P* elements are transcriptionally active, with both transposase and repressor mRNA being transcribed, and some of them may also be transpositionally active, since the transposase transcripts of S7 (1), S3, S6 (o), and S10, do not show any premature stop codon in their sequences. Both somatic and specific tissues have alternatively-spliced repressor transcripts and transposase transcripts are limited to specific tissues. However, in this study, in *D. prosaltans* only transposase could be amplified and in *D. sturtevanti* only repressor protein was detected. Haring *et al.* (1998) reported that in *S. pallida* M-type *P* element, the IVS3 was always present, indicating that in this species only repressor proteins are produced. The authors suggested that because IVS3 is never removed in *S. pallida* transcripts, this subfamily is transpositionally inactive. Taking this point of view, and the presence of some premature stop codons in the sequenced transcript, we may suggest that the *D. sturtevanti* canonical *P* element is also transpositionally inactive. The same hypothesis might be true to the *D. prosaltans* canonical *P* element that also has some premature stop codons in the transposase sequence, probably being transpositionally inactive.

From where the sequences come from?

In the phylogenetic analysis, all transcripts were placed into the canonical subfamily clade confirming their high similarity to *D. melanogaster* canonical *P* element. A detailed phylogenetic analysis of *P*-element sequences in the subgenus *Sophophora* places the clade of canonical elements as sister group to all other subfamilies of *Drosophila* and *Scaptomyza P* elements (Clark and Kidwell, 1997). In addition to the canonical subfamily, that study identified three other *P* subfamilies in species of the *saltans* and *willistoni* groups. Silva and Kidwell (2000) reported that the four subfamilies form distantly related clades when elements in different subfamilies are compared, but within *P* subfamilies, elements sampled from different species, and even species groups, are often very closely related. The discreteness of these subfamilies coexisting in the same genome provides a strong indication that the



presence of multiple *P* element subfamilies in the *saltans* and *willistoni* species groups is the result of multiple waves of horizontal transfer events. The sequences here described, due to the low divergence (3% to 10%) regarding to the *D. melanogaster* ones plus the high divergence (21% to 31%) regarding to the *D. helvetica* (M-type) and *D. bifasciata* (M- and O-type) sequences, seem to belong to the canonical *P* element subfamily.

While the *P*-element family, as a whole, seems to be older than 40 Myr, the diversification of the canonical *P* subfamily dates back only 2-3 Myr (Silva and Kidwell, 2000). Since then, canonical *P* elements have spread into nine out of 10 species surveyed in the *willistoni* group and into six out of eight species surveyed in the *saltans* group. Silva and Kidwell (2000) suggested that this spread occurred soon after introduction of the subfamily into these groups and that it was achieved mostly by means of horizontal transfer, with subsequent diversification within species. The presence of a very homologous sequence in both *D. prosaltans* and *D. saltans* species suggests that they might be present in the common ancestor of the *saltans* subgroup.

Our analysis showed that the $d_N : d_S$ ratio for *P* element sequences in *saltans* group as a whole is high when compared with that of *Adh*. Comparison of d_N and d_S values between *P* element sequences and host gene reveals that the average d_N for *P* elements is similar to that observed for the *Adh* while the average value of d_S observed for *P* elements is almost five times smaller than the d_S observed for *Adh*. These results taken together with the low level of codon bias in the *P* element sequences, indicated by a medium N_c value of 51.55, suggest weak selection on nonsynonymous sites. Also, the proportion of G's and C's at synonymous positions in *P* elements was approximately 36%, which is very similar to the frequency expected due to genome mutation bias (Shields *et al.*, 1988). The presence of termination codons in six of the 11 sequences and the presence of deletions and insertions that unambiguously disrupt the reading frame in most of the sequences further support a scenario of weak selection. These results agree with those described for Silva and Kidwell (2000) for canonical *P* sequences in the *saltans* group. The authors obtained an N_c value of 52 and a proportion of G's and C's at synonymous positions of 40%, very similar to our results.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Horizontal transfer has traditionally been inferred when the high degree of similarity between TE sequences is impossible to reconcile with the long divergence time of their respective host species (Daniels *et al.*, 1990). The high similarity found among the transcripts of *D. saltans*, *D. prosaltans* and *D. sturtevanti* and those described for *D. melanogaster*, plus the lack of detectable transcripts related to canonical *P* in the other species of the *saltans* group support the hypothesis of a recent event of horizontal transfer into these species. Silva and Kidwell (2000), analyzing canonical sequences of exon 2 with unknown transcriptional activity in the *saltans* and *willistoni* groups, also proposed this hypothesis. Our results expand these conclusions based on the d_N and d_S values found for *Adh* and *P* element sequences when comparing pairs of taxa: 1) d_N and d_S values are lower for *Adh* than for *P* element in the comparison of *D. saltans* vs. *D. prosaltans*, suggesting that these sequences were present in the ancestor of the *saltans* subgroup; 2) once d_N and d_S values for *P* element are lower in the comparison of *D. sturtevanti* vs. *D. saltans* than in those of *D. sturtevanti* vs. *D. prosaltans* and *D. saltans* vs. *D. prosaltans*, it could be suggested that the sequence found in *D. sturtevanti* might have been horizontally transferred from *D. saltans*.

Our data also reinforce that the principles of *P* element regulation as germline specificity of transposase and alternative splicing of the third intron are ubiquitous irrespective of the *P* element subfamily or host species as previously shown for *D. melanogaster*, *S. pallida* and species of the *obscura* group (Haring *et al.*, 1998). Regardless of the origin of these sequences, the ability to retain their mobility and regulation patterns after horizontal transfer into a new host species seems to be an essential property for *P* element survival.

Acknowledgements

We are indebted to Drs. M.G. Kidwell and A.J.Holyoake who kindly provided advice and laboratory facilities at the University of Arizona and for their critical comments on earlier versions of the manuscript. This work was supported by FAPESP (Grants n° 95/7192-2, 98/08734-1 and 00/11313-0 and a fellowship to J.P.C.) and CNPq.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Table 1. Species and strains of *Drosophila* used in this study. Location and year of collection are also indicated.

Subgroup	Species	Strain	Location	Year
parasaltans	*D. parasaltans*	**PAT**	Tapuruquara (AM) – Brazil *	1962
	D. subsaltans	**SUB**	Belém (PA) – Brazil (SC – 14044-0872.0)	1959
sturtevanti	*D. milleri*	**MEY**	El Yunque – Puerto Rico (SC – 14043-0861.0)	1962
	D. dacunhai	**DAP**	Pentionville – Haiti (SC – 14043-0854.0)	1962
	D. sturtevanti	**SMX**	Matlapa (San Luis Potosi) - Mexico **	1998
		SSR	Santana do Riacho, (MG) – Brazil ***	1995
saltans	*D. prosaltans*	**PE2**	Eldorado (RS) – Brazil ****	1995
	D. austrosaltans	**API**	Pirassununga (SP) – Brazil (SC – 14045-0881.0)	1959
	D. saltans	**S7**	Chilpancingo – Mexico *	1962
		S3	Guatemala*	1960
		S6	Huichihuayan – Mexico*	1962
		S10	Matlapa (San Luis Potosi) – Mexico **	1998

Notes: SC indicates Stock Center ordering number; * provided by H.E.M.C. Bicudo; ** collected by J.C. Silva; *** collected by C.R. Vilela; **** provided by V.L.S. Valente.

Table 2. List of PCR primers used in this study.

Primer	Sequence	Position
Mel e3-	5'-GTTTATCAACATCGACGTTTC	2581 – 2561
Mel i1+*	5'-TACACAAACAGAGTCCTGTT	431 - 442, 501 – 508
Mel i2+*	5'-GTATATCAGAATCAAAAACCTG	1157 - 1168, 1222 - 1231
Mel i3-*	5'-CATTTCTGTATTCCTGGCTATT	2154 - 2138, 1947 - 1943

* Primer sequences were adapted from Haring *et al.* (1998).



Table 3. Main characteristics of the mRNA *P*-element related amplified sequences.

Species and Lines	Sequence (nt)*	Type	Divergence (%)**	Indels				Conserved Motifs			
				E1	E2	E3	I3	LZ1	LZ2	LZ3	HTH
D. saltans S7 (o1)	501-1947	T	3.22	+11	+7			+	±	+	-
D. saltans S7 (o2)	501-1947	T	3.30		+20			+	+	-	+
D. saltans S7 (or)	1222-2581	R	9.72			+44	+56	?	?	+	?
D. saltans S7 (h)	1222-2581	R	11.34			+42		?	?	-	?
D. saltans S6 (o)	1222-1947	T	2.94					?	?	+	?
D. saltans S6 (h)	1222-1581	R	9.73			+44		?	?	-	?
D. saltans S3 (o)	591-1947	T	2.77					?	+	+	+
D. saltans S3 (h)	1222-2581	R	9.58			+54		?	?	+	?
D. saltans S10 (o)	1222-1947	T	3.40					?	?	+	?
D. prosaltans (o)	501-1947	T	4.36	+30 +36	-5 -268			+	-	-	±
D. sturtevanti (h)	1222-2581	R	10.42			+14 +2		?	?	-	?

Notes: *Nucleotide positions corresponding to those of *D. melanogaster*; T= transposase; R= repressor; ** Divergence regarding to *D. melanogaster P* element transposase and repressor sequences; E1= exon 1; E2= exon 2; E3= exon 3; I3= intron 3 (IVS3); LZ1= first leucine-zipper motif; LZ2= second leucine-zipper motif; LZ3= third leucine-zipper motif; HTL= helix-turn-helix motif; + = conserved; ± = partially conserved; - = not conserved; ? = not analyzed. The size and nature of the insertion (+) or deletion (-) are denoted separately when the indels are noncontiguous.



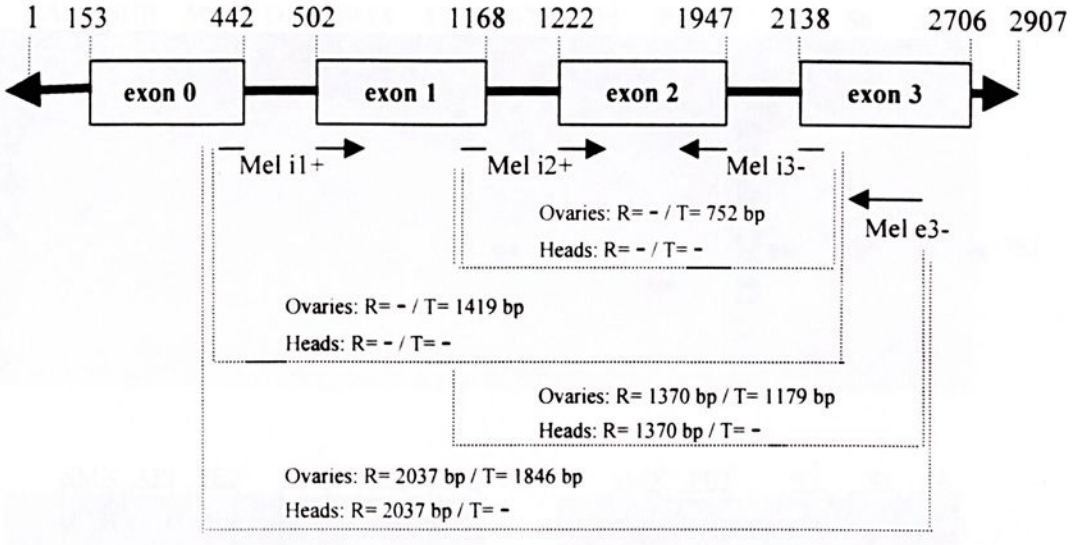


Figure 1. Primer binding sites in the canonical *P* element of *D. melanogaster* and PCR products expected or not (-) for repressor (R) and transposase (T) transcripts in heads and ovaries for each primer combination.

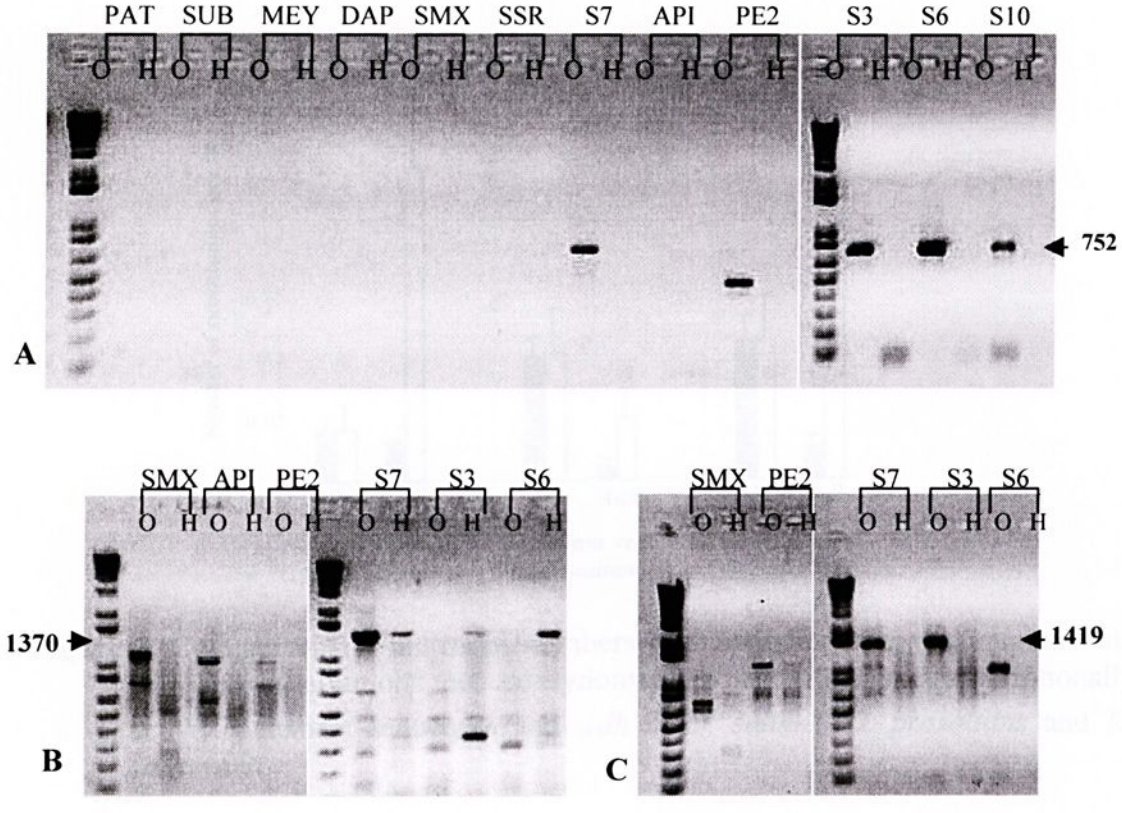


Figure 2. RT-PCR fragments obtained from ovaries (O) and heads (H) of *D. parasaltans* (PAT), *D. subsaltans* (SUB), *D. milleri* (MEY), *D. dacunhai* (DAP), *D. sturtevanti* (SMX and SSR), *D. saltans* (S3, S6, S7 and S10), *D. austrosaltans* (API) and *D. prosaltans* (PE2), with the primer combinations Mel i2+/ Mel i3- **(A)**, Mel i2+/ Mel e3- **(B)** and Mel i1+/ Mel i3- **(C)**. Marker: 1 kb Plus DNA Ladder (GIBCO – BRL Technologies).

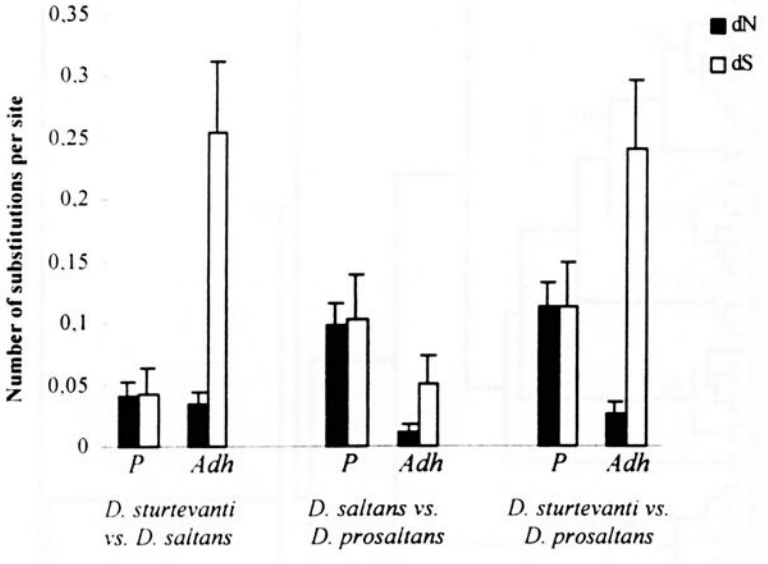


Figure 3. Comparisons of the mean numbers of substitutions per synonymous site (d_S) and substitutions per nonsynonymous site (d_N) for transcriptionally active *P* element sequences and *Adh* in *D. saltans*, *D. prosaltans* and *D. sturtevanti*.

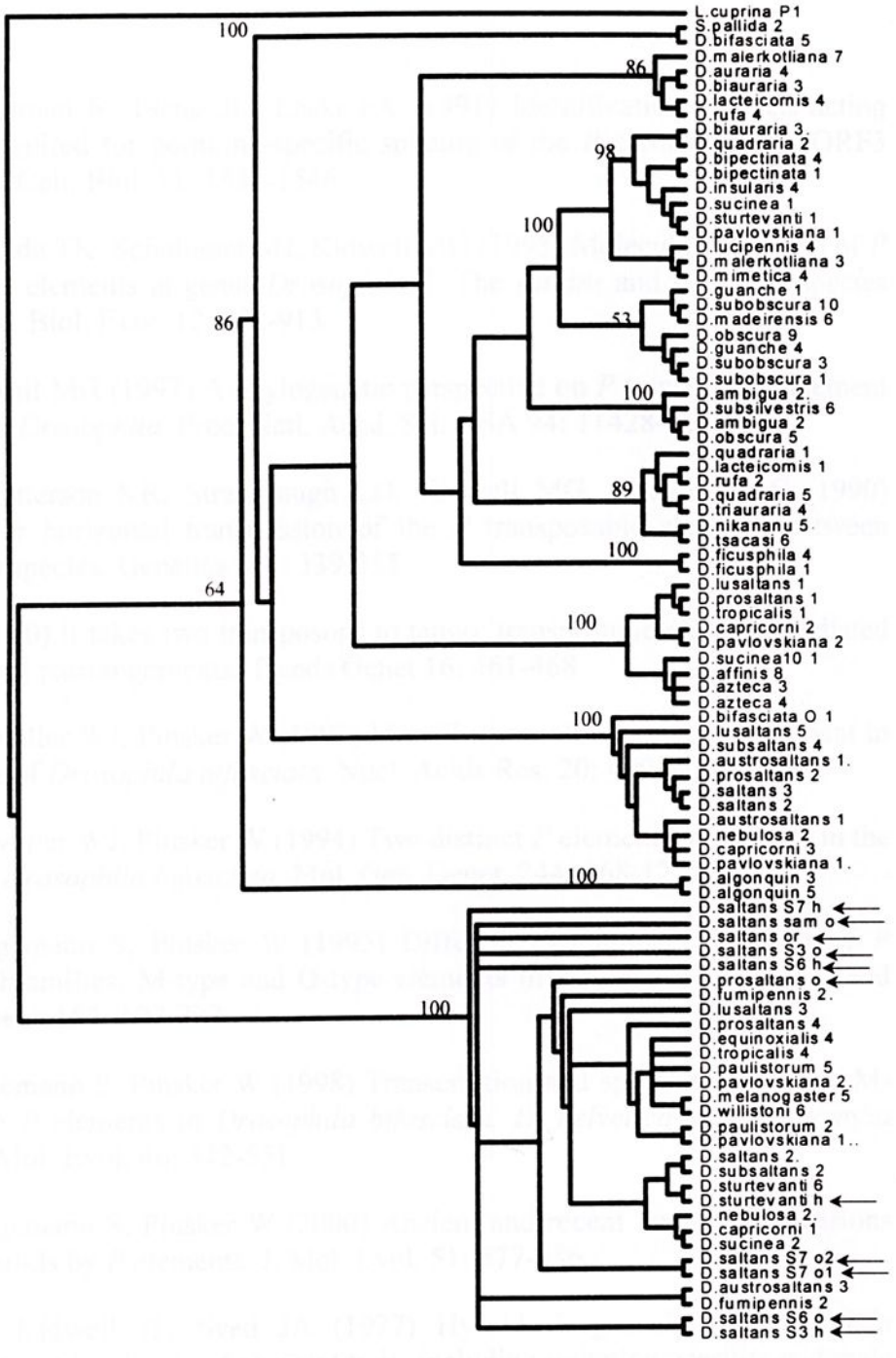


Figure 4. Phylogenetic analysis of *P* element nucleotide sequences obtained in this study and those obtained from Clark and Kidwell (1997). Species names are given followed by the total number of clones used to obtain a consensus sequence or, for the sequences obtained in this study, followed by the strain name and first letter of the tissue in which it was transcribed. This cladogram was generated by parsimony analysis as implemented PAUP 4.0 b10 (Swofford, 1997). This is a consensus tree of 1,000 equally parsimonious trees, each requiring 2,193 steps. The consistency index is 0.40 and the retention index 0.815. Bootstrap analysis of 500 replicates was performed on the data. Arrows identify the transcribed sequences obtained in this study.

unesp

References

Chain AC, Zollman S, Tseng JC, Laski FA (1991) Identification of a *cis*-acting sequence required for germline-specific splicing of the *P* element ORF2-ORF3 intron. Mol. Cell. Biol. 11: 1538-1546

Clark JB, Altheide TK, Scholosser MJ, Kidwell MG (1995) Molecular evolution of *P* transposable elements in genus *Drosophila.* I. The *saltans* and *willistoni* species groups. Mol. Biol. Evol. 12: 902-913

Clark JB, Kidwell MG (1997) A phylogenetic perspective on *P* transposable element evolution in *Drosophila.* Proc. Natl. Acad. Sci. USA 94: 11428-11433

Daniels SB, Petterson KR, Strausbaugh LD, Kidwell MG, Chovnick AC (1990) Evidence for horizontal transmission of the *P* transposable elements between *Drosophila* species. Genetics 124: 339-355

Gray YHM (2000) It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. Trends Genet 16: 461-468

Hagemann S, Miller WJ, Pinsker W (1992) Identification of a complete *P* element in the genome of *Drosophila bifasciata*. Nucl. Acids Res. 20: 409-413

Hagemann S, Miller WJ, Pinsker W (1994) Two distinct *P* element subfamilies in the genome of *Drosophila bifasciata*. Mol. Gen. Genet. 244: 168-175

Haring E, Hagemann S, Pinsker W (1995) Different evolutionary behavior of *P* element subfamilies: M-type and O-type elements in *Drosophila bifasciata* and *D. imaii.* Gene 163: 197-202

Haring E, Hagemann S, Pinsker W (1998) Transcription and splicing patterns of M- and O-type *P* elements in *Drosophila bifasciata, D. helvetica* and *Scaptomyza pallida*. J. Mol. Evol. 46: 542-551

Haring E., Hagemann S, Pinsker W (2000) Ancient and recent horizontal invasions of Drosophilids by *P* elements. J. Mol. Evol. 51: 577-586

Kidwell MG, Kidwell JF, Sved JA (1977) Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics 86: 813-833

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA 2: Molecular Evolutionary Genetic Analysis software, Arizona State University, Tempe, Arizona, USA

Lansman RA, Shade RO, Grigliatti TA, Brock HW (1987) Evolution of *P* transposable elements: sequences of *Drosophila nebulosa P* elements. Proc. Natl. Acad. Sci. USA 84: 6491-6495

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Laski FA, Rubin GM (1989) Analysis of the cis-acting requirements for germ-line-specific splicing of the *P* element ORF2-ORF3 intron. Genes Dev. 3: 720-728

Laski FA, Rio DC, Rubin GM (1986) Tissue specificity of *Drosophila P* element transposition is regulated at the level of mRNA splicing. Cell 44: 7-19

Loreto ELS, Valente VLS, Zaha A, Silva JC and Kidwell MG (2001) *Drosophila mediopunctata P* elements: a new example of horizontal transfer. The J Hered 92: 375-381.

Magalhães LE (1962) Notes on the taxonomy, morphology, and distribution of the *saltans* group of *Drosophila*, which descriptions of four new species. Univ. Tex. Publ. 6205: 135-154

Magalhães LE, Björnberg AJS (1957) Estudo da genitália masculina de *Drosophila* (Diptera). Rev. Bras. Biol. 16: 273-280

Miller WJ, Hagemann S, Reiter E, Pinsker W (1992) P-element homologous sequences are tandemly repeated in the genome of *Drosophila guanche*. Proc. Natl. Acad. Sci USA 89: 4018-4022

Misra S, Rio DC (1990) Cytotype control of *Drosophila P* element transposition: the 66 kd protein is a repressor of transposase activity. Cell 62: 269-284

Mourão CA, Bicudo HEMC (1967) Duas novas espécies de *Drosophila* do grupo *saltans* (*Drosophilidae*, Diptera). Pap. Dept. Zool. S. Paulo 20: 123-134

Mullins MC, Rio DC, Rubin GM (1989) Cis-acting DNA sequence requirements for *P*-element transposition. Genes Dev. 3: 729-738

Nei M, Gojobori T (1996) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3: 418-426

O'Grady PM, Clark JB, Kidwell MG (1998) Phylogeny of the *Drosophila saltans* species group based on combined analysis of nuclear and mitochondrial DNA sequences. Mol. Biol. Evol. 15: 656-664

O'Hare K, Rubin GM (1983) Structure of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell 34: 25-35

Paricio N, Perez-Alonso M, Martínez-Sebastián MJ, de Frutos R (1991) *P* sequences of *Drosophila subobscura* lack exon 3 and may encode a 66 kd repressor-like protein. Nucl. Acids Res. 19: 6713-6718

Patterson JT, Mainland GB (1944) The Drosophilidae of Mexico. Univ. Tex. Publ. 4445: 1-101

Peden J (1997) CodonW software. http://www.molbiol.ox.ac.uk/cu/codonW.html.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Perkins HD, Howells AJ (1992) Genomic sequences with homology to the *P* element of *Drosophila melanogaster* occur in the blowfly *Lucilia cuprina.* Proc. Natl. Acad. Sci. USA 89: 10753-10757

Rio DC (1990) Molecular mechanisms regulating *Drosophila P* element transposition. Ann. Rev. Genet. 24: 543-578

Shields DC, Sharp PM, Higgins DG, Wright F (1988) Silent sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol. Biol. Evol. 5: 704-716

Silva JC, Kidwell MG (2000) Horizontal transfer and selection in the evolution of *P* elements. Mol. Biol. Evol. 17: 1542-1557

Simonelig M, Anxolabéhère D (1991) A *P* element of *Scaptomyza pallida* is active in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 88: 6102-6106

Spassky BS (1957) Morphological differences between siblings species of *Drosophila.* Univ. Tex. Publ. 5721: 48-61

Spassky BS, Richmond RC, Peréz-Salas S, Pavlovsky O, Mourão CA, Hunter AS, Hoenigsberg H, Dobzhansky T, Ayala FJ (1971) Geography of sibling species related to *Drosophila willistoni* and semispecies of the *Drosophila paulistorum* complex. Evolution 25: 129-143

Sturtevant AH (1942) The classification of the genus *Drosophila*, with descriptions of nine new species. Univ. Tex. Publ., 4113: 6-51

Swofford D (1997) PAUP: Phylogenetic analysis using parsimony. Version 4.0b10. Smithsonian Institution, Washington, D.C.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple alignment through sequence weighting, positions-specific gap penalties and weigh matrix choice. Nucl. Acids Res. 22: 4673-4680

Throckmorton LH (1975) The phylogeny, ecology and geography of *Drosophila.* In: King RC (ed) Handbook of Genetics. Plenum, New York. pp. 421-436

Throckmorton LH, Magalhães LE (1962) XVI. Changes with evolution of pteridine accumulations in species of the *saltans* group of the genus *Drosophila.* Univ. Tex. Publ. 6205: 489-505

Wright F (1990) The 'effective number of codons' used in a gene. Gene 87: 23-29.

Distribuição e número de cópias de TEs no grupo *saltans*

Capítulo Seis

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Resumo

As informações sobre a distribuição e o número de cópias de muitos elementos transponíveis que ocorrem em *Drosophila* são restritas a poucas espécies, embora estes elementos sejam amplamente distribuídos no gênero. Este trabalho teve como objetivo descrever a distribuição e o número de cópias dos retrotransposons *copia*, *gypsy*, *micropia*, *Minos*, *I* e *Bari* e dos transposons *hobo* e *mariner*, no *grupo saltans* de *Drosophila*. De acordo com os resultados obtidos, com exceção do elemento *mariner*, todos os outros elementos são amplamente distribuídos dentro do grupo *saltans* e mostram variabilidade de número de inserções, que atinge até 24 cópias.

Distribution and copy number of transposable elements in species of the *Drosophila saltans* group

Juliana Polachini de Castro and Cláudia Márcia A. Carareto[1]

Department of Biology, Universidade Estadual Paulista - UNESP, São José do Rio Preto –SP, Brazil.

Abstract

Information about the distribution and copy number of many transposable elements is restricted to few species of *Drosophila*, although the wide distribution of these elements throughout the genus. This work aimed to describe the distribution and copy number of *copia*, *gypsy*, *micropia*, *I*, *hobo*, *mariner*, *Minos* and *Bari-1* transposable elements in the *saltans* group of *Drosophila*. According to our data, excepting for *mariner*, all other elements are widespread within the *saltans* group and show variable insertion numbers reaching up to 24 copies.

Running title: Copy number of TEs in the *saltans* group.

Key-words: copy number, *saltans*, transposable elements, distribution

*** Artigo a ser submetido para publicação no Genetics and Molecular Biology**

[1] Correspondence to: Dr. Cláudia M. A. Carareto, Departamento de Biologia, Universidade Estadual Paulista – UNESP, Rua Cristóvão Colombo 2265, CEP 15054-000, São José do Rio Preto – SP, Brazil. (55) 17 2212382, * (55) 17 2212390. E-mail: carareto@bio.ibilce.unesp.br.

Introduction

Transposable elements (TEs) are a major component of the genomes of most species and are widespread throughout the genus *Drosophila*. Although it is well established that TEs have played a major role in evolution and may still be useful in maintaining the genetic variability of natural populations, the way they invade genomes and how their copy number is contained in genomes and populations are still matters of debate.

The intrinsically parasitic nature of active TEs accounts for their undisputed ability to invade new species, increase in copy number, and survive over long periods of evolutionary time. The life cycle of a transposable element in an organism can be visualized as having three main phases (Kidwell and Lisch, 1997). Initially there is a rapid invasion phase in which amplification of copy number occurs accompanied by mutation which renders some elements inactive, followed by a stage of maturity of variable length, in which amplification and loss of copies are roughly in balance. Finally, a stage of senescence extends for possibly millions of years. At this stage, all autonomous elements have been lost, no amplification occurs, and nonautonomous sequences are either lost, become deleted, or otherwise diverge. In conclusion, in a given genome it is possible to find copies of a TE that are autonomous, nonautonomous but capable of being mobilized by autonomous elements, or nonautonomous and not capable of being mobilized (Capy, Anxolabéhère and Langin, 1994). It is also possible that the host has domesticated parts of their sequences (Pinsker *et al.*, 2001). Thus, to understand TE dynamics and estimate their possible effects on evolution, we need data on their copy number and genomic localization in different species.

Two large analyses (Martin, Wiernasz and Schedl, 1983; Stacey *et al.*, 1986) and one review (Biémont and Cizeron, 1999) of the distribution of mobile elements in the genus *Drosophila* were previously published. The first two studies concerned the elements *copia*, *412*, *297* and three other not-well-known middle repetitive sequences in 32 species in Martin Wiernasz and Schedl (1983) and the elements *P*, *I*, *gypsy*, *copia* and *F* in 34 species in Stacey *et al.* (1986). The review (Biémont and Cizeron, 1999) summarizes data of individual works in 228 species and 43 TEs

including *copia*, *gypsy*, *I* and *P* elements. These analyses showed that many TEs are present in many species of the genus, but there are some major differences among them. For example, *copia*, *412* and *gypsy* are detected in almost all species (Martin Wiernasz and Schedl, 1983; Stacey *et al.*, 1986), while the retrotransposon *I* is restricted to the *melanogaster* subgroup (de Frutos, Peterson and Kidwell, 1992). Although much data has been accumulated, much information about the occurrence and copy number of many TEs in many species is still missing. In the *saltans* group, for example, only *P* element was studied in species of the five subgroups.

Aiming to broaden the scenario of TE distribution and copy number, in this study we describe the distribution and copy number of eight transposable elements in the *saltans* group of *Drosophila*. From the elements belonging to Class I we studied *copia*, *gypsy*, *micropia*, and *I*; and from Class II we studied *hobo*, *mariner*, *Minos* and *Bari-1*.

Materials and Methods

Fly stocks

All the species and strains used in the present study are listed in Table 1. All fly strains were derived from a single, randomly selected female from a mass culture. *D. melanogaster* – Harwich and *D. mauritiana* were used as positive control.

Southern blot analyses

Genomic DNA was prepared from 25-30 adult flies, according to Jowett (1986). To estimate the overall amount of each transposable element in the analyzed species, 10 µg of genomic DNA were digested with restriction enzymes, submitted to electrophoresis on 0.8% agarose gels and transferred to nylon membranes (Hybond N+/Amersham Biosciences). The probes and enzymes used are described in Table 2. For the hybridization and detection, the chemioluminescent hybridization system Gene Images (Amersham Biosciences) was used according to manufacturer's instructions. The plasmids used in this investigation were kindly provided by Drs. D. Hartl (Mos1), D. Dorsett (pGGHS), A. Bucheton (pI407), B. Arcà (*Minos*), R. Caizzi (p28/47D), D-H Lankenau (dhMiF2) and E. Loreto (p77E4 and pHx4).



Our analysis is predicated on knowledge of the complete element sequences described for *D. melanogaster* or for the species from which the element was firstly sequenced. The enzymes chosen do not cut the element sequence, then, each fragment observed is assumed as a product of one genomic insertion.

Results and Discussion

Aiming to contribute to the understanding of the distribution pattern of some TEs, we looked for the presence and copy number of sequences homologous to eight transposable elements in 10 species of the *saltans* group by means of Southern analyses. The status of the *saltans* group species in relation to the presence of *copia*, *gypsy*, *micropia*, *I*, *hobo*, *mariner*, *Minos* and *Bari-1* elements is presented in Table 3. This overview was set up taking together all data available about these elements, plus the data produced in this study. The distribution of sequences homologous to these elements showed strong to weak or no hybridization signals among the species of the *saltans* group (Figure 1). All analyses were repeated several times and some of the blots still show unsatisfactory hybridization signals, certifying the difficulty of obtain good hybridizations using probes of *D. melanogaster* elements.

Hybridization to *copia*, *gypsy*, *micropia* and *I* homologous sequences were observed in all species (Figure 1, A - D). Despite the weak signals for *copia* and *I* elements (obtained after two hours of exposition; for the other sequences 30 minutes of exposition were sufficient), our results indicate a broad distribution of both retrotransposons among species of the *saltans* group. de Frutos, Peterson and Kidwell (1992) reported the same difficulty on observing *copia* hybridization blots. Previous reports have provided conflicting data about the distribution of *copia* homologous sequences in other *Drosophila* species. Brookfield, Montgomery and Langley (1984) suggested that *copia* sequences are restricted to *D. melanogaster* and its related species, but a broader distribution within the *Sophophora* subgenus was later observed (Stacey *et al.*, 1986). These same authors also reported that *gypsy* is widely distributed among Sophophoran species including *D. emarginata*, *D. sturtevanti*, *D. austrosaltans* and *D. prosaltans*. However, they do not found *gypsy* homologous sequences in *D. neocordata*. Our results are in agreement with the proposal of wide

distribution of *gypsy*, and extend its presence to representatives of the five subgroups of the *saltans* group, including in *D. neocordata*. The fact that *gypsy* has been considered as a retrovirus (Pélisson *et al.*, 1994; Song *et al.*, 1994) may explain its wide distribution. *Micropia* sequences were previously described for the *saltans* group (Lankenau, 1993; Almeida, Castro and Carareto, 2001) but the number of copies has not been estimated yet. The wide distribution of *I* element in the *saltans* group does not agree with results reported by Stacey *et al.* (1986), who found *I* homologous sequences only in members of the *melanogaster* group; but partially agrees with Bucheton *et al.* (1986) who described a more widespread distribution to *I* element, including in *D. sturtevanti*. These inconsistent results could be accounted for by differences in stringency conditions used by different investigators.

The transposons distribution also conflicted with data previously reported by other investigators. With exception for *mariner*, for which only few signals could be observed in some species after the longer exposure time (three hours), the other DNA elements were widely distributed in the *saltans* group (Figure 1, E – H). Daniels, Chovinick and Boussy (1990) carried out a wide distribution screening of *hobo* transposable elements in the genus *Drosophila* and found homologous sequences only in the *melanogaster* group. After that, Loreto *et al.* (1998) reported the presence of *hobo* homologous sequences in *D. willistoni*. Our results extend the presence of *hobo* homologous sequences to the *saltans* group. Previous surveys report the absence of hybridization signals for *mariner* elements in the *saltans* group (Capy, David and Hartl, 1992; Brunet *et al.*, 1994), and outside the *melanogaster* species group (Maruyama and Hartl, 1991). In our approach, sequences homologous to *mariner* were not observed in *D. dacunhai*, *D. sturtevanti*, *D. saltans* and *D. prosaltans*, and faint hybridization signals were observed in the other species. However, it does not mean that this element is absent in these species. It could simply reflect a high level of divergence between *mariner* elements in different species. This proposal is corroborated by the difficulty of hybridization between the sequences of the *saltans* group and that of the *D. mauritiana* probe, as shown in Figure 1F. *Minos* and *Bari-1* sequences have been reported in some species of the *saltans* group (Arcà and Savakis, 2000; Moschetti *et al.*, 1998). Our study shows that



these sequences are also present in representatives of the five subgroups of *saltans* group.

Previous studies reported that retrotransposons are much more abundant than transposons. For example, Bartolomé, Maside and Charlesworth (2002) found that retrotransposons were responsible for 64% of transposable elements present in *D. melanogaster* genome. In the *saltans* group, at least for the elements studied here, these numbers are not corroborated (Table 4). Among the retrotransposons the number of insertions varied between two (*micropia* element in *D. dacunhai*) and 24 (*I* element in *D. parasaltans*), while for transposons the copy number was between two (*Minos* element in *D. milleri*) and 15 (*Bari-1* element in *D. parasaltans*). Excepting for the *I* element, the insertion means were between seven and nine different insertions per species. However, the copy numbers are very variable among different species, probably due to drift, once the strains used have been maintained in laboratory for a long time.

Although sequences homologous to most of the *D. melanogaster* TEs under study were present in species of the *saltans* group, in general they did not show close sequence similarity to their respective homologous once the hybridization signals generated were not strong enough. Also, the heterogeneity of the copy number of each element in different species does not follow a pattern that could be a reflection of the colonization scenario. But it is possible to hypothesize that all eight elements studied in this work were present in the ancestral of the *saltans* group and considering the transposable element life cycle, it is expected that most of the copies detected in this work are nonautonomous and present high divergence when compared to *D. melanogaster* ones, once these two groups are phylogenetically separated for about 50 million years. As it is known, more divergent the sequence, more difficult is to detect it using the canonical sequence as probe. Thus, this work points out for the necessity of more complete information about the occurrence and molecular characteristics of the elements among different *Drosophila* species groups in order to understand the evolutionary history of these and other transposable elements.



Table 1. Species and strains of *Drosophila* used in this study. Local and year of collection are also indicated.

Subgroup	Species	Strain	Location	Collection Year
cordata	*D. neocordata*	**NEO**	Minas Gerais – Brazil (SC – 14041-0831.0)	1959
elliptica	*D. emarginata*	**EVC**	Vera Cruz – Mexico (SC – 14042-0841.6)	1962
parasaltans	*D. parasaltans*	**PAT**	Tapuruquara (AM) – Brazil *	1962
	D. subsaltans	**SUB**	Belém (PA) – Brazil (SC – 14044-0872.0)	1959
sturtevanti	*D. milleri*	**MEY**	El Yunque – Puerto Rico (SC – 14043-0861.0)	1962
	D. dacunhai	**DAP**	Pentionville – Haiti (SC – 14043-0854.0)	1962
	D. sturtevanti	**SMX**	Matlapa (San Luis Potosi) - Mexico **	1998
saltans	*D. austrosaltans*	**API**	Pirassununga (SP) – Brazil (SC – 14045-0881.0)	1959
	D. saltans	**SAC**	Chilpancingo – Mexico *	1962
	D. prosaltans	**PTT**	Sangre Grande –Trinidad Tobago *	1962

Notes: SC indicates Stock Center ordering number; * collected by H.E.M.C. Bicudo; ** collected by J.C. Silva; *** collected by C.R. Vilela; **** collected by W.J. Tadei; ***** collected by V.L.S. Valente.

Table 2. Probes and restriction enzymes used for the copy number detection of each transposable element in species of the *saltans* group.

Element	Probe	Digestion with
copia	Fragment of 440 bp (nt. 33 to 472) amplified from p77E4plasmid	*Xho I*
gypsy	Fragment of 545 bp (nt. 2155 to 2697) amplified from pGGHS plasmid	*Bam HI*
micropia	Fragment of 386 bp (nt. 2813 to 3198) amplified from dhMiF$_2$plasmid	*Bam HI*
I	Fragment of 635 bp (nt. 2178 to 2813) amplified from pI407plasmid	*Eco RI*
hobo	Fragment of 963 bp (nt. 651 to 1614) amplified from pHX4 plasmid.	*Bam HI*
mariner	*NheI / PvuII* fragment of 800 bp from Mos1 plasmid	*Pst I*
minos	Fragment of 1081 bp (nt 1 to 1081) amplified from pBCKSP minos plasmid	*Xho I*
Bari-1	Fragment of 703 bp (nt. 453 to 1196) amplified from p28/47D plasmid	*Eco RI*



Table 3. Distribution of transposable elements in species of the *saltans* group.

Subgroup	Species	Retrotransposons								Transposons							
		copia (a)	*	*gypsy* (b)	*	*micropia* (c)	*	*I* (d)	*	*hobo* (e)	*	*mariner* (f)	*	*Minos* (g)	*	*Bari-1* (h)	*
cordata	*cordata*	+		-				-									
	neocordata		+	-	+	+	+	-	+	-	+	-	+		+	+	+
elliptica	*elliptica*	+		+				-									
	emarginata		+	+	+	+	+	-	+	-	+		+		+	+	+
	neoelliptica									-							
parasaltans	*parasaltans*		+		+	+	+		+		+		+		+		+
	subsaltans		+		+	+	+		+		+		+		+	+	+
sturtevanti	*sturtevanti*	+	+	+	+	+	+	+	+	-	+	-	-		+	+	+
	milleri		+		+	+	+		+	-	+		+		+		+
	dacunhai		+		+	+	+		+		+		-		+		+
saltans	*saltans*	+	+	+	+	+	+		+	-	+	-	-	+	+	+	+
	lusaltans									-						+	
	prosaltans	+	+	+	+	+	+	-	+	-	+		-		+		+
	austrosaltans	+	+	+	+	+	+	-	+	-	+		+		+		+

(a) Martin *et al.* (1983), Stacey *et al.* (1986), Jordan and McDonald (1998); (b) Stacey *et al.* (1986); (c) Lankenau (1993), Almeida *et al.* (2001); (d) Bucheton *et al.* (1986), Stacey *et al.* (1986); (e) Daniels *et al.* (1990); (f) Maruyama and Hartl (1991), Capy *et al.* (1992), Brunet *et al.* (1994); (g) Arcà and Savakis (2000); (h) Moschetti *et al.* (1998). + and -: presence and absence of the element, respectively as determined by Southern blots. * Data produced in this study.

Table 4. Number of copies of each transposable element found in species of the *saltans* group.

Species	Elements						
	copia	*gypsy*	*micropia*	*I*	*hobo*	*Minos*	*Bari-1*
D. neocordata	3	10	7	10	8	5	10
D. emarginata	5	7	9	14	12	9	6
D. parasaltans	5	6	14	24	8	4	15
D. subsaltans	8	9	2	19	11	5	8
D. milleri	13	10	12	17	11	2	5
D. dacunhai	12	8	2	7	7	14	5
D. sturtevanti	6	9	4	15	8	5	7
D. austrosaltans	9	9	12	5	7	8	12
D. saltans	10	7	4	13	10	13	9
D. prosaltans	9	7	6	16	11	6	7
Mean	8.0 ± 3.2	8.2 ± 1.4	7.2 ± 4.4	14.0 ± 5.6	9.3 ± 1.9	7.1 ± 3.9	8.5 ± 3.2



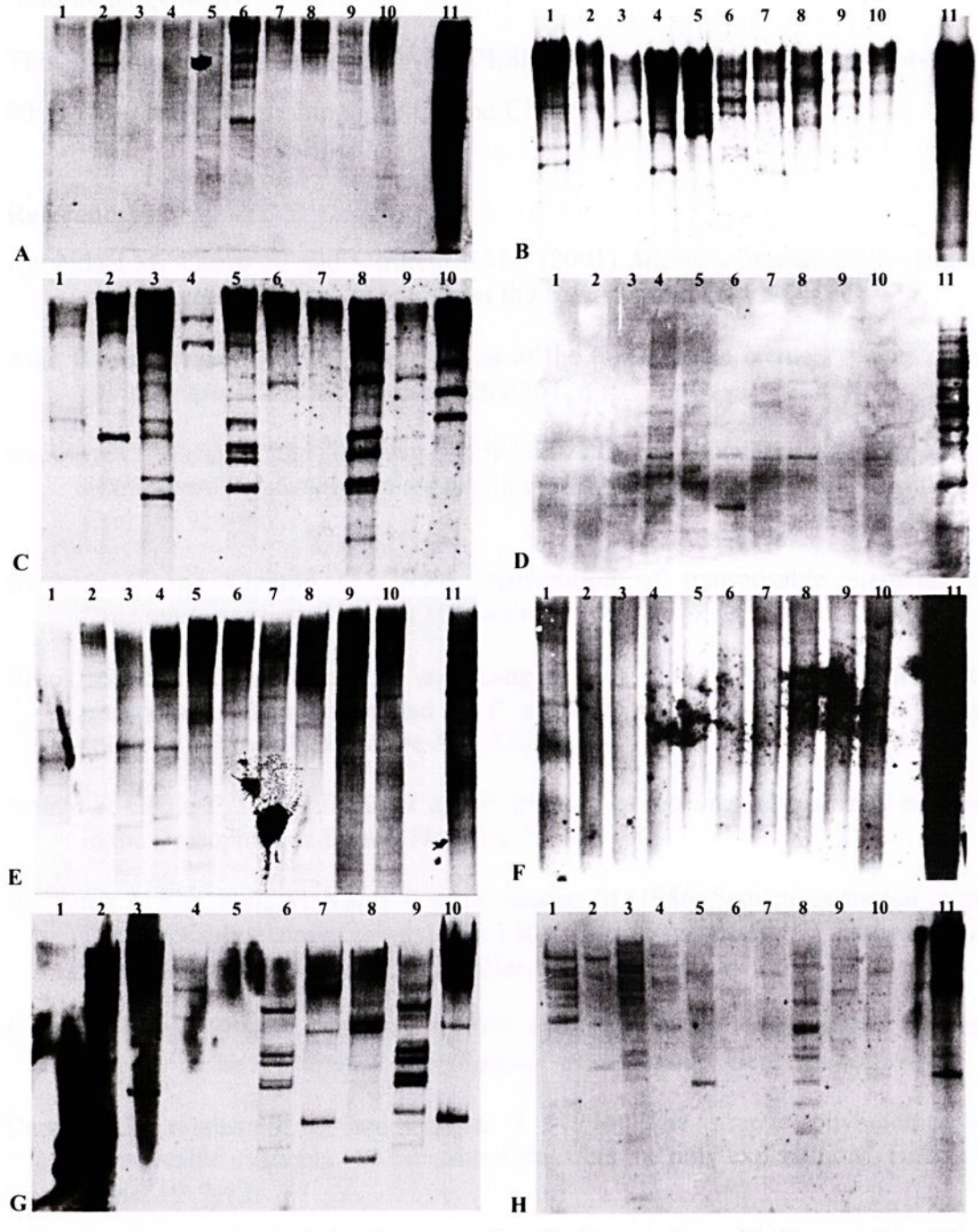


Figure 1. *Southern blot* analysis of *D. neocordata* (1); *D. emarginata* (2); *D. parasaltans* (3); *D. subsaltans* (4); *D. milleri* (5); *D. dacunhai* (6); *D. sturtevanti* (7); *D. austrosaltans* (8); *D. saltans* (9); *D. prosaltans* (10) and *D. melanogaster* (11) or *D. mauritiana* (11G) for *copia* (**A**); *gypsy* (**B**); *micropia* (**C**); *I* (**D**); *hobo* (**E**); *mariner* (**F**); *Minos* (**G**) and *Bari-1* (**H**) transposable elements.

Acknowledgements

This investigation was supported by FAPESP (Grants n° 95/7192-2, 98/08734-1 and 00/11313-0 and a fellowship to J.P.C.) and CNPq.

References

Almeida LM, Castro JP and Carareto CMA (2001) *Micropia* transposable element occurrence in *Drosophila* species of the *saltans* group. D I S 84: 114-117.

Arcà B and Savakis C (2000) Distribution of the transposable element *Minos* in the genus *Drosophila*. Genetica 108: 263-267.

Bartlomé C, Maside X and Charlesworth B (2002) On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. Mol Biol Evol 19: 926-937.

Biémont C and Cizeron G (1999) Distribution of transposable elements in *Drosophila* species. Genetica 105: 43-62.

Brookfield JJY, Montgomery E and Langley CH (1984) Apparent absence of transposable elements related to *P* elements of *D. melanogaster* in other species of *Drosophila*. Nature 310: 330-332.

Brunet F, Godin F, David JR and Capy P (1994) The *mariner* transposable element in the Drosophilidae family. Heredity 73: 377-385.

Bucheton A, Simonelig M, Vaury C and Crozatier M (1986) Sequences similar to the I transposable element involved in I-R hybrid dysgenesis in *D. melanogaster* occur in other *Drosophila* species. Nature 322: 650-652.

Capy P, David JR and Hartl DL (1992) evolution of the transposable element *mariner* in the *Drosophila melanogaster* species group. Genetica 86: 37-46.

Capy P, Anxolabéhère D and Langin T (1994) The strange phylogenies of transposable elements: are horizontal transfers the only explanation? Trends in Genet 10: 7-12.

Daniels SB, Chovinick A and Boussy IA (1990) Distribution of *hobo* transposable elements in the genus *Drosophila*. Mol Biol Evol 7: 589-606.

de Frutos R, Peterson KR and Kidwell MG (1992) Distribution of *Drosophila melanogaster* transposable element sequences in species of the *obscura* group. Chromosoma 101: 293-300.

Jordan IK, and McDonald JF (1998) Evolution of the *copia* retrotransposon in the *Drosophila melanogaster* species subgroup. Mol Biol Evol 15: 1160-1171.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Jowett T (1986) Preparation of nucleic acids. In: Roberts DB (ed) *Drosophila*: a practical approach. Oxford, IRL Press.

Kidwell MG and Lisch D (1997) Transposable elements as sources of variation in animals and plants. Proc Natl Acad Sci USA 94: 7704-7711.

Lankenau DH (1993) In: McDonald JF (ed) Transposable elements and Evolution. Kluwer Academic Publisher, pp. 232-241.

Loreto ELS, Basso da Silva L, Zaha A and Valente VLS (1998) Distribution of transposable elements in neotropical species of *Drosophila*. Genetica 101: 153-165.

Martin G, Wiernasz D and Schedl P (1983) Evolution of *Drosophila* repetitive-dispersed DNA. J Mol Evol 19: 203-213.

Maruyama K and Hartl DL (1991) Evolution of the transposable element *mariner* in *Drosophila* species. Genetics 128: 319-29.

Moschetti R, Caggese C, Barsanti P and Caizzi R (1998) Intra- and interspecies variataion among *Bari-1* elements of the *melanogaster* species group. Genetics 150: 239-250.

Pélisson A, Song SU, Prud'homme N, Smith PA, Bucheton A and Corces VG (1994) *Gyspy* transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila* flamenco gene. EMBO J 13: 4401-4411.

Pinsker W, Haring E, Hagemann S and Miller WJ (2001) The evolutionary life history of *P* transposons: from horizontal invaders to domesticated neogenes. Chromosoma 110: 148-158.

Song SU, Gerasimova T, Kurkulos M, Boeke JD and Corces VG (1994) An env-like protein encoded by a *Drosophila* retroelement: evidence that *gypsy* is an infectious retrovirus. Genes Dev 8: 2046-2057.

Stacey SN, Lansman RA, Brock HW and Grigliatti TA (1986) Distribution and conservation of mobile elements in the genus *Drosophila*. Mol Biol Evol 3: 522-534.

Discussão

Capítulo Sete

unesp

O conceito de hereditariedade fundamenta-se na estabilidade dos genomas. Entretanto, sabe-se que grande parte do genoma de todos os organismos é instável e que uma parcela desta instabilidade gerada pode garantir uma rápida adaptação às mudanças do ambiente. Entre os responsáveis por esta instabilidade estão os elementos transponíveis (TEs), seqüências que meio século atrás foram considerados como uma excentricidade genética. Hoje, estes mesmos elementos são reconhecidos como um dos principais fatores envolvidos na evolução do genoma.

Como a presença de TEs em um genoma hospedeiro muitas vezes não produz efeitos fenotípicos óbvios, foi apenas depois do seqüenciamento de DNA que os geneticistas perceberam que os TEs são ubíquos e abundantes nos genomas eucariotos. O reconhecimento da grande quantidade de DNA aparentemente egoísta originou novas perspectivas sobre os genomas eucariotos e levantou várias questões sobre a interação destes elementos e os genomas hospedeiros. Alguns exemplos destas questões são: Existe uma "guerra de braços" entre TEs e seus hospedeiros? Quanto da evolução do genoma hospedeiro é devida à interação com os TEs? Quanto da evolução é um sub-produto destes elementos? Sendo os organismos extremamente oportunistas, teriam os hospedeiros conseguido manipular alguns destes elementos para usá-los para as suas próprias necessidades? De onde vêm os TEs? Eles são relíquias do passado ou ainda estão sendo criados? Estas e outras questões levam os pesquisadores cada vez mais a se interessarem pelo estudo dos elementos transponíveis em diversos hospedeiros.

Foi com o objetivo de contribuir para a compreensão destas questões que foram realizados os trabalhos descritos nesta Tese de Doutorado. Por ser o elemento transponível mais estudado, e sobre o qual existe uma grande quantidade de informação, o elemento *P* foi escolhido como tema central dos nossos estudos.

Apesar da grande quantidade de informação disponível sobre o elemento *P*, a maioria dos trabalhos teve *Drosophila melanogaster* como protagonista, uma vez que este elemento foi descoberto nesta espécie como causador da síndrome da disgenesia do híbrido (Kidwell & Kidwell, 1975). Dessa forma, todas as informações disponíveis sobre os mecanismos de transposição e regulação do elemento *P* são decorrentes de estudos realizados em *D. melanogaster*. Tendo em vista que os

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

mecanismos de regulação da mobilidade do *P* ainda não foram elucidados por completo, muitos são os trabalhos a este respeito. A mobilização do elemento *P* é restrita a vários fatores determinados pelo próprio elemento bem como pelo hospedeiro, tais como, expressão de proteínas repressoras, acúmulo de cópias defectivas, repressão pós-transcricional via RNA anti-sense e silenciamento epigenético via modificação na cromatina. Com o objetivo de facilitar o acesso aos dados obtidos neste campo de estudo nas últimas décadas, realizamos uma ampla busca na literatura e tentamos reunir as informações mais relevantes na revisão apresentada no segundo capítulo desta Tese.

Embora as espécies do grupo *saltans* tenham recebido alguma atenção nos estudos referentes ao elemento *P*, algumas questões ainda não são plenamente compreendidas, tais como: Será que as seqüências *P* estão realmente ausentes nas espécies dos subgrupos *cordata* e *emarginata*, ou será que elas nunca foram identificadas? Como são as seqüências das cópias completas do *P* no grupo *saltans*? Existem outras subfamílias, além das já descritas para este grupo? As seqüências *P* canônicas presentes no grupo *saltans* são ainda transcricionalmente ativas? Tentando elucidar estas questões, realizamos três diferentes abordagens que foram descritas nos capítulos três a cinco desta Tese, e cujos resultados serão sucintamente discutidos neste capítulo.

Elementos *P* semelhantes àquele de *D. melanogaster* são encontrados em várias espécies dos grupos *saltans* e *willistoni* (Clark & Kidwell, 1997). Estes elementos, denominados canônicos, formam uma subfamília coesa com uma diferença máxima entre suas seqüências de 10%. Além desta subfamília, cópias mais divergentes são detectadas em uma grande variedade de espécies. Algumas destas cópias divergentes foram bem caracterizadas como pertencentes às subfamílias tipo-O, -M e –T (Hagemann *et al.*, 1996b), enquanto outras ainda se encontram em estudo. A diferença de seqüências ao nível de DNA entre estas três subfamílias está em torno de 30%, mas a estrutura básica de quatro exons e repetições terminais invertidas é sempre mantida.

O segundo artigo apresentado nesta Tese, buscou resolver algumas das questões ainda a serem esclarecidas sobre a distribuição de seqüências homólogas ao

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

P no grupo *saltans*. Este trabalho mostrou que, ao contrário do que se pensava, existem seqüências homólogas ao *P* em todos os subgrupos do grupo *saltans*. Entretanto, estas seqüências diferem tanto quanto à porcentagem de similaridade com a seqüência canônica, bem como quanto ao número de cópias presente em cada espécie. Com base nos resultados obtidos, podemos sugerir que as seqüências presentes nos subgrupos *cordata*, *elliptica* e *parasaltans* representam polimorfismos ancestrais e que provavelmente as seqüências descritas nos dois primeiros subgrupos estivessem presentes no ancestral do subgênero *Sophophora*, o que justificaria a alta divergência destas seqüências em relação à canônica e a dificuldade encontrada por outros pesquisadores em detectá-las. As demais espécies do grupo *saltans*, além de possuírem seqüências canônicas, e talvez também seqüências dos tipos-M e -O, resultantes de invasões mais recentes, provavelmente também possuam estas seqüências rudimentares em seus genomas. Várias ondas de transmissão horizontal têm sido propostas para o grupo *saltans* (Silva & Kidwell, 2000), o que torna nossa hipótese plausível.

Kaplan *et al.* (1985) propuseram que no curso de sua coexistência com o respectivo genoma hospedeiro, os TEs estão sujeitos a curtas explosões de atividade seguidas por longos períodos de inativação. Dessa forma, os elementos que não se expressam há muito tempo estão fadados à extinção genômica. Como resultado de recombinação e deriva genética, o número de cópias irá diminuir e finalmente os elementos imóveis serão eliminados do conjunto gênico. Mas, mesmo aquelas cópias transmitidas por muitas gerações irão sofrer degradação do seu conteúdo informativo. Sem pressões seletivas mantendo a funcionalidade da seqüência, um elemento imóvel irá se degradar por acúmulo de mutações, um processo que finalmente leva à inativação vertical (Miller *et al.*, 1996). Tendo como premissa o ciclo de vida do elemento transponível, sabemos que em um determinado genoma, podem coexistir diferentes seqüências em diferentes fases de degradação.

Este trabalho também mostrou a presença de cópias do elemento *P* potencialmente completas apenas em espécies dos subgrupos *sturtevanti* e *saltans*, o que corrobora a proposta de existência de seqüências provenientes de múltiplos eventos de transmissão horizontal neste grupo (Silva & Kidwell, 2000). Outro fato

que corrobora esta afirmação é o número de cópias estimado de seqüências homólogas ao elemento *P* nas espécies do grupo. Foi possível observar um maior número de inserções nas espécies dos subgrupos *sturtevanti* e *saltans*, que variou entre nove e 15 do que nas espécies dos subgrupos *cordata*, *elliptica* e *parasaltans*, que variou entre uma e três cópias.

Ainda procurando caracterizar as seqüências *P* presentes no grupo *saltans*, o terceiro trabalho desta Tese apresenta a seqüência de dois elementos *P* completos em *D. sturtevanti* e *D. prosaltans*. O seqüenciamento destes elementos em quase sua totalidade sugere que a estrutura básica do elemento, caracterizada pela presença de invertidos terminais, quatro exons e três introns, parece ser mantida como no caso das demais subfamílias que já foram bem caracterizadas (canônica, tipo-M, tipo-O e tipo-T). Entretanto, várias alterações na matriz de leitura, incluindo a presença de vários codons de terminação, sugerem que estas seqüências não são mais capazes de codificar uma proteína funcional. As alterações de nucleotídeos presentes ao longo das seqüências, que produzem uma divergência em torno de 31% em relação ao elemento canônico, não permitem sua inclusão em nenhuma das subfamílias anteriormente citadas. Porém, as seqüências presentes em *D. sturtevanti* e *D. prosaltans* possuem uma divergência de apenas 12% entre si, o que permite considerarmos que elas pertençam à mesma subfamília. Além disso, estas seqüências formam um clado consistente com algumas seqüências de espécies do grupo *saltans* e *willistoni* descritas anteriormente por Clark & Kidwell (1997). Estes autores encontraram apenas poucas seqüências pertencentes a este clado e especularam se esta freqüência poderia ser devida a uma escassa presença desta subfamília no genoma destas espécies ou devida à amostragem. Nossos dados sugerem que esta subfamília deve ser mais abundante do que se pensava, uma vez que mesmo seqüenciando apenas um clone de cada espécie, esta seqüência foi encontrada em duas espécies diferentes, apesar da presença de elementos canônicos nestes genomas. Dessa forma, sugerimos que as seqüências agrupadas neste clado possam fazer parte de uma mesma subfamília, cujas características ainda não haviam sido documentadas.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

O terceiro trabalho apresentado nesta Tese teve como objetivo averiguar a atividade transcricional das cópias canônicas presentes nos genomas das espécies do grupo *saltans*. Esta abordagem teve como objetivo colaborar com o entendimento do cenário de invasão destas seqüências no grupo em estudo. Este trabalho mostrou que *D. saltans* parece ser a única espécie do grupo que ainda apresenta uma seqüência *P* que poder ser considerada transcricionalmente ativa e provavelmente ainda apresente elementos transposicionalmente ativos. Esta espécie foi a única que apresentou mRNAs que codificam para transposase e para a proteína repressora de 66 kDa. *D. prosaltans* e *D. sturtevanti* apresentaram apenas transcritos de transposase e repressor, respectivamente. Estes transcritos, ao contrário da maioria daqueles encontrados em *D. saltans*, apresentaram deleções, inserções e várias alterações na matriz de leitura, o que sugere que eles não produzam proteínas funcionais. Nas demais espécies do grupo não foram detectados mRNAs transcritos a partir do elemento *P*. A não detecção destes transcritos não significa, necessariamente, que não existam elementos com atividade transcricional nas demais espécies do grupo, pois é possível que existam cópias com baixa atividade, cujos produtos não eram suficientes para visualização em gel de agarose. Esta hipótese pode ser reforçada pela diferença de quantidade de transcritos observada entre as espécies do grupo *saltans* e *D. melanogaster*. Para esta última espécie, foi necessária apenas uma série de amplificações para a visualização dos produtos, enquanto que para as demais espécies, duas séries foram necessárias.

Outro resultado interessante mostrado por este trabalho foi em relação à antiguidade das seqüências obtidas nos respectivos genomas. Esta análise foi realizada comparando-se o número de substituições sinônimas por sítios sinônimos (d_S) e não sinônimas por sítios não-sinônimos (d_N) entre as seqüências *P* obtidas e seqüências do gene *Adh* de cada espécie analisada. Estas comparações nos levaram a três sugestões: 1) como os valores de d_N e d_S são menores para *Adh* que para o elemento *P* na comparação entre *D. saltans* e *D. prosaltans*, estas seqüências provavelmente estavam presentes no ancestral do subgrupo *saltans*; 2) uma vez que os valores de d_N e d_S para o elemento *P* foram menores na comparação entre *D. sturtevanti* e *D. saltans* que na comparação entre *D. sturtevanti* e *D. prosaltans* e

entre *D. saltans* e *D. prosaltans*, a seqüência encontrada em *D. sturtevanti* pode ter sido transmitida horizontalmente a partir de *D. saltans*. Estes resultados novamente corroboram a hipótese de múltiplos eventos de transmissão horizontal no grupo *saltans* sugerida por Silva & Kidwell (2000).

Adicionalmente, os nossos dados reforçam que os princípios envolvidos na regulação do elemento *P*, tais como a especificidade de produção da transposase na linhagem germinativa e a retirada alternativa do terceiro intron para produção diferencial da transposase e da proteína repressora são ubíquos, independente da subfamília a que o elemento *P* pertença ou da espécie hospedeira. Independentemente da origem destas seqüências, a habilidade de retenção da mobilidade e dos padrões básicos da sua regulação após a transferência horizontal em uma nova espécie hospedeira parece ser uma propriedade essencial para a sobrevivência do elemento *P*.

Finalmente, o último trabalho apresentado nesta Tese se propôs a traçar um panorama geral da presença e do número de cópia dos elementos transponíveis *copia*, *gypsy*, *micropia*, *I*, *hobo*, *mariner*, *Minos* e *Bari-1* nas espécies do grupo *saltans*. Os dados divulgados até o momento eram contraditórios sobre a presença destes elementos neste grupo de espécies, e alguns deles não haviam sido estudados em muitas espécies. Nosso trabalho mostrou que, com exceção do elemento *mariner*, que sabemos ser diferenciado em várias subfamílias, os demais elementos investigados estão presentes nas espécies dos cinco subgrupos do grupo *saltans*. A variabilidade do número de inserções nas diferentes espécies é grande, com números que vão de dois a 24 para os diferentes elementos. Apesar da ampla distribuição destes elementos no grupo *saltans*, de maneira geral eles não mostraram uma alta homologia com as sondas de *D. melanogaster*. Adicionalmente, a heterogeneidade do número de cópias de cada elemento nas diferentes espécies não segue um padrão que poderia refletir o cenário de colonização. Entretanto, é possível sugerir que os oito elementos estudados estavam presentes no ancestral do grupo *saltans*. Considerando-se o ciclo de vida de um elemento transponível, já era esperado que muitas das cópias detectadas neste estudo fossem não-autônomas e apresentassem alta divergência quando comparadas com os mesmos elementos caracterizados em *D. melanogaster*,



já que os dois grupos estão filogenéticamente separados por aproximadamente 50 milhões de anos.

Os dados apresentados mostram que a variabilidade de seqüências de elementos transponíveis existentes nos genomas das espécies do grupo *saltans* de *Drosophila* é ainda maior do que se havia suposto anteriormente. Após a perda da autonomia, as seqüências evoluem diferencialmente umas das outras e a variabilidade resultante deste processo ao longo do tempo é imensa. Dessa forma, fica cada vez mais evidente que seria inviável uma caracterização acurada dos elementos transponíveis nos diferentes hospedeiros. Ao longo do tempo evolutivo, muitas seqüências se deterioram a ponto de não serem mais reconhecidas como elementos transponíveis, enquanto que outras podem ser incorporadas à seqüências gênicas e se tornarem domesticadas pelo genoma hospedeiro.

Concluímos este trabalho, da mesma forma com que começamos, com várias questões ainda para serem respondidas. Qual seria o papel dos elementos transponíveis na evolução dos genomas hospedeiros? Após ser superada a fase do aumento do número de cópias e as seqüências móveis terem perdido sua habilidade de transposição, teria o elemento transponível desempenhado seu papel na produção da variabilidade genética? O destino final deste elemento seria a perda estocástica e deterioração de sua seqüência a ponto de seus resquícios serem considerados lixo genético? Muitas das respostas para estas questões ainda são hipotéticas, mas na medida em que os dados continuem a ser acumulados, provavelmente teremos melhores condições para avaliar o papel dos TEs na regulação da evolução. Torna-se cada vez mais evidente que a crença de que os TEs sejam simplesmente DNA lixo e que não tenham qualquer conseqüência evolutiva não pode mais ser aceita. Ao contrário, parece que estas seqüências de DNA são criticamente importantes para o surgimento inesperado de novidades evolutivas ao longo do tempo.

unesp

Conclusões

Capítulo Oito

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

O presente trabalho permitiu concluir que:

1- ao contrário do que se pensava, existem seqüências homólogas ao elemento *P* em todos os subgrupos do grupo *saltans*;

2- existem cópias do elemento *P* potencialmente completas apenas em espécies dos subgrupos *sturtevanti* e *saltans*;

3- o número de inserções do elemento *P* foi maior nas espécies dos subgrupos *sturtevanti* e *saltans* que nas espécies dos subgrupos *cordata*, *elliptica* e *parasaltans*;

4- as seqüências homólogas ao elemento *P* presentes nas espécies dos subgrupos *cordata* e *elliptica* possivelmente estavam presentes no ancestral do subgênero *Sophophora*;

5- os elementos *P* completos seqüenciados em *D. sturtevanti* e em *D. prosaltans* pertencem a uma mesma subfamília, que ainda não havia sido bem caracterizada e que parece ser mais amplamente distribuída do que se pensava;

6- *D. saltans* parece ser a única espécie do grupo que ainda apresenta seqüências do elemento *P* que podem ser consideradas transcricionalmente ativas e provavelmente ainda apresente elementos transposicionalmente ativos;

7- as seqüências do elemento *P* transcritas em *D. saltans* e *D. prosaltans* provavelmente estavam presentes no ancestral do subgrupo *saltans*;

8- as seqüências do elemento *P* transcritas em *D. sturtevanti* podem ter sido transmitidas horizontalmente a partir de *D. saltans*;

9- o elemento *mariner* detectado em apenas algumas espécies do grupo *saltans* (*D. neocordata*, *D. emarginata*, *D. parasaltans*, *D. subsaltans*, *D. milleri* e *D. austrosaltans*) é possivelmente muito divergente do elemento de *D. mauritiana*;

10- os elementos transponíveis *copia*, *gypsy*, *micropia*, *I*, *hobo*, *Minos* e *Bari-1* estão presentes em todas as espécies dos cinco subgrupos do grupo *saltans*.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Referências Bibliográficas

REFERÊNCIAS BIBLIOGRÁFICAS

BAYEV Jr., A. A., *et al.* Structural organization of transposable element *mdg4* from *Drosophila melanogaster* and a nucleotide sequence of its long terminal repeats. *Nucl. Acids Res.*, v. 12, p. 3707-3723.

BAZIN, C.; HIGUET, D. Lack of correlation between dysgenic traits in the *hobo* system of hybrid dysgenesis in *Drosophila melanogaster*. *Genet. Res.*, v. 67, p. 219-226, 1996.

BELL, G. The sexual nature of the eukaryote genome. *J. Hered.*, v. 84, p. 351-359, 1993.

BICUDO, H. E. M. C. Reproductive isolation in the *saltans* group of *Drosophila*. *Genetica*, v. 44, p. 313-329, 1973a.

BICUDO, H. E. M. C. Chromosomal polymorphism in the *saltans* group of *Drosophila*. *Genetica*, v. 44, p. 520-552, 1973b.

BIÉMONT, C. Population genetics of transposable DNA elements: A *Drosophila* point of view. *Genetica*, v. 86, p. 67-84, 1992.

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

BONNIVARD, E.; HIGUET, D.; BAZIN, C. Characterization of natural populations of *Drosophila melanogaster* with regard to the *hobo* system: a new hypothesis on the invasion. *Genet. Res.*, v. 69, p. 197-208, 1997.

BONNIVARD, E., *et al.* A scenario for the *hobo* transposable element invasion, deduced from the structure of natural populations of *Drosophila melanogaster* using tandem TPE repeats. *Genet. Res.*, v. 75, p. 13-23, 2000.

BREGLIANO, J. C.; KIDWELL, M. G. Hybrid dysgenesis determinants. In: Shapiro J. A. (ed), *Mobile Genetic Elements*, Acad. Press, N.Y., 1983.

BRUNET, F., *et al.* The mariner transposable element in the Drosophilidae family. *Heredity*, v. 73, p. 377-385, 1994.

BUCHETON, A., *et al.* Sequences similar to the I transposable element involved in I-R hybrid dysgenesis in *D. melanogaster* occur in other *Drosophila* species. *Nature*, v. 322, p. 650-652, 1986.

_____ I elements and the *Drosophila* genome. *Genetica*, v. 86, p. 175-190, 1992.

CÁCERES, M., *et al.* Generation of a widespread *Drosophila* inversion by a transposable element. *Science*, v. 285, p. 415-418, 1999.

CÁCERES, M.; PUIG, M.; RUIZ, A. Molecular characterization of two natural hotspots in *Drosophila buzzatii* genome induced by transposon insertions. *Genome Res.*, v. 11, p. 1353-1364, 2001.

CAGGESE, C., *et al.* The distribution of the transposable element *Bari-1* in the *Drosophila melanogaster* and *Drosophila simulans* genomes. *Genetica*, v. 96, p. 269-283, 1995.

CAIZZI, R.; CAGGESE, C.; PIMPINELLI, S. *Bari-1*, a new transposon-like family in *Drosophila melanogaster* with a unique heterochromatic organization. *Genetics*, v. 133, p. 335-345, 1993.

CALVI, B. R.; GELBART, W. M. The basis for germline specificity of the *hobo* transposable element in *Drosophila melanogaster*. *EMBO J.*, v. 13, p. 1636-1644, 1994.

CALVI, B. R., *et al.* Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: *hobo*, *Activator*, and *Tam3*. *Cell*, v. 66, p. 465-471, 1991.

CAPY, P.; ANXOLABÉHÈRE, D.; LANGIN, T. The strange phylogenies of transposable elements: are horizontal transfers the only explanation? *Trends Genet.*, v. 10, p. 7-12, 1994.

CAPY, P., *et al. Dynamics and Evolution of Transposable Elements*. 1st. France: Landes Bioscience, 1998. 197p.

CLARK, J. B., *et al.* Molecular evolution of *P* transposable elements in genus *Drosophila*. I. The *saltans* and *willistoni* species groups. *Mol. Biol. Evol.*, v. 12, p. 902-913, 1995.

CSINKI, A. K.; McDONALD, J. F. Analysis of *copia* sequence variation within and between *Drosophila* species. *Mol. Biol. Evol.*, v. 12, p. 83-93, 1995.

DANIELS, S. B., *et al.* Evidence for horizontal transmission of the *P* transposable elements between *Drosophila* species. *Genetics*, v. 124, p. 339-355, 1990a.

_____ Distribution of *hobo* transposable elements in the genus *Drosophila*. *Mol. Biol. Evol.*, v. 7, p. 589-606, 1990b.

DIMITRI, P.; JUNAKOVIC, N. Revising the selfish DNA hypothesis: new evidence on accumulation of transposable elements in heterochromatin. *Trends Genet.*, v. 15, p. 123-124, 1999.

DOOLITTLE, W. F.; SAPIENZA, C. Selfish genes, the phenotype paradigm and genome evolution. *Nature*, v. 284, p. 601-603, 1980.

DWYER, D. S. Assembly of exons from unitary transposable genetic elements: implications for the evolution of protein-protein interactions. *J. Theor. Biol.*, v. 194, p. 11-27, 1998.

EMORI, Y., *et al.* Determination of the nucleotide sequences of *copia* and *copia*-related RNA in *Drosophila* VLP. *Nature*, v. 315, p. 773-776, 1985.

ENGELS, W.R., *et al.*. High-frequency *P* element loss in *Drosophila* is homolog dependent. *Cell*, v. 62, p. 515-525, 1990.

FINNEGAN, D. J. Transposable elements in eukaryotes. *Int. Ver. Cytol.*, v. 93, p. 281-326, 1985.

FRANZ, G.; SAVAKIS, C. *Minos*, a new transposable element from *Drosophila hydei*, is a member of the *Tc1*-like family of transposons. *Nucl. Acids Res.*, v. 19, p. 6646, 1991.

FRANZ, G., *et al.* Mobile *Minos* elements from *Drosophila hydei* encode a two-exon transposase with similarity to the paired DNA-binding domain. *Proc. Natl. Acad. Sci. USA*, v. 91, p. 4746-4750, 1994.

GALINDO, M. I., *et al.* Spread of the autonomous transposable element *hobo* in the genome of *Drosophila melanogaster*. *Mol. Biol. Evol.*, v. 12, p. 723-734, 1995.

GINZBURG, L.R., *et al.* On the theory of speciation induced by transposable elements. *Genetics*, v. 107, p. 331-341, 1999.

GREIDER, C. W. Telomere length regulation. *Annu. Ver. Biochem.*, v. 65, p. 337-365, 1996.

HAGEMANN, S.; MILLER, W. J.; PINSKER, W. Identification of a complete *P* element in the genome of *Drosophila bifasciata*. *Nucl. Acids. Res.*, v. 20, p. 409-413, 1992.

_____ Two distinct *P* element subfamilies in the genome of *Drosophila bifasciata*. *Mol. Gen. Genet.*, v. 244, p. 168-175, 1994.

HAGEMANN, S.; HARING, E.; PINSKER, W. Repeated horizontal transfer of *P* transposons between *Scaptomyza pallida* and *Drosophila bifasciata*. *Genetica*, v. 98, p. 43-51, 1996a.

_____ A new *P* element subfamily from *Drosophila tristis*, *D. ambigua* and *D. obscura*. *Genome*, v. 39, p. 978-985, 1996b.

_____ Horizontal transmission versus vertical inheritance of *P* elements in *Drosophila* and *Scaptomyza*: has the M-type subfamily spread from East Asia? *J. Zool. Syst. Evol. Research*, v. 36, p. 75-83, 1998.

HARING, E.; HAGEMANN, S.; PINSKER, W. Different evolutionary behavior of *P* element subfamilies: M-type and O-type elements in *Drosophila bifasciata* and *D. imaii*. *Gene*, v. 163, p. 197-202, 1995.

_____ Ancient and recent horizontal invasions of Drosophilids by *P* elements. *J. Mol. Evol.*, v. 51, p. 577-586, 2000.

HARTL, D. L.; LOZOVSKAYA, E. R.; LAWRENCE, J. G. Nonautonomous transposable elements in prokaryotes and eukaryotes. *Genetica*, v. 86, p. 47-53, 1992.

HARTL, D. L.; LOHE, A. R.; LOZOVSKAYA E. R. Regulation of the transposable element *mariner*. *Genetica*. v. 100, p. 177-184, 1997.

HAYMER, D. S.; MARSH, J. L. Germ line and somatic instability of a *white* mutation in *Drosophila mauritiana* due to a transposable element. *Dev. Genet.*, v. 6, p. 281-291, 1986.

HENNING, W., *et al.* Molecular cloning of microdissected lampbrush loop DNA sequences of *Drosophila hydei*. *EMBO J.*, v. 2, p. 1741-1746, 1983.

HERÉDIA, F. Evolução do retroelemento *gypsy* em espécies de *Drosophila* e *Zaprionus indianus*: uma abordagem filogenética. Porto Alegre, 2002. Tese (Doutorado). Universidade Federal do Rio Grande do Sul.

HO, Y. T.; WEBER, S. M.; LIM, J. K. Interacting *hobo* transposon in an inbred strain and interaction regulation in hybrids of *Drosophila melanogaster*. *Genetics*, v. 134, p. 895-908, 1993.

HUIJSER, P., *et al.* Retrotransposon sequences are expressed in Y chromosomal lampbrush loops of *Drosophila hydei*. *J. Mol. Biol.*, v. 203, p. 689-697, 1988.

HURST, G. D. D.; SCHILTHUIZEN, M. Selfish genetic elements and speciation. *Heredity*, v. 80, p. 2-8, 1998.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

HURST, G. D. D.; WERREN, J. H. The role of selfish genetic elements in eukaryotic evolution. *Nature Reviews*, v. 2, p. 597-606, 2001.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. Initial sequencing and analysis of the human genome. *Nature*, v. 409, p. 860-921, 2001.

JACOBSON, J. W.; HARTL, D. L. Coupled instability of tow X-linked genes in *Drosophila mauritiana*: germinal and somatic mutability. *Genetics*, v. 111, p. 57-65, 1985.

JACOBSON, J. W.; MEDHORA, M. M.; HARTL, D. L. Molecular structure of a somatically unstable transposable element in *Drosophila*. *Proc. Natl. Acad. Sci. USA*, v. 83, p. 8684-8688, 1986.

JORDAN, I. K.; MATYUNINA, L. V.; McDONALD, J. F. Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. *Proc. Natl. Acad. Sci. USA*, v. 96, p. 12621-12625, 1999.

KAPLAN, N.; DARDEN, T.; LANGLEY, C. Evolution and extinction of transposable elements in Mendelian populations. *Genetics*, v. 109, p. 459-480, 1985.

KARESS, R. E.; RUBIN, G. M. Analysis of *P* transposable element functions in *Drosophila*. *Cell*, v. 38, p. 135-146, 1984.

KIDWELL, M. G. Voyage of an ancient *mariner*. *Nature*, v. 362, p. 202, 1993.

KIDWELL, M. G. Transposable elements and the evolution of genome size in eukaryotes. *Genetica*, v. 115, p. 49-63, 2002.

KIDWELL, M. G.; KIDWELL, J. F. Cytoplasm chromosome interaction in *Drosophila melanogaster*. *Nature*, v. 253, p. 755-776, 1975.

KIDWELL, M. G.; LISCH, D. Transposable elements as sources of variation in animals and plants. *Proc. Natl. Acad. Sci. USA*, v. 94, p. 7704-7711, 1997.

_____ Transposable elements, parasitc DNA, and genome evolution. *Evolution*, v. 55, p. 1-24, 2001.

KIM, A. L., *et al*. Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA*, v. 91, p. 1285-1289, 1994.

KLEIN, C. Estudo de fenômenos possivelmente relacionados com a mobilidade de elementos transponíveis e à presença de endoparasitas em populaces de *Drosophila*

willistoni. Porto Alegre, 2002. Dissertação (Mestrado). Universidade Federal do Rio Grande do Sul.

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

LAMBERTSSON, A.; ANDERSSON, S.; JOHANSSON, T. Cloning and characterization of a variable-sized *gypsy* mobile elements in *Drosophila melanogaster*. *Plasmid*, v. 22, p. 22-31, 1989.

LANKENAU, D. H. The retrotransposon family *micropia* in *Drosophila* species. *In: Transposable elements and evolution.* J. F. McDonald (ed.), Kluwer Academic Publishers, 1993.

LANKENAU, D. H., *et al. Micropia*: a retrotransposon of *Drosophila* combining structural features of DNA viruses, retroviruses and non-viral transposable elements. *J. Mol. Biol.*, v. 204, p. 233-246, 1988.

LANKENAU, D. H.; HUIJSER, P.; HENNING, W. Characterization of the long terminal repeats of *micropia* elements dissected from the Y chromosomal lampbrush loops "Threads" of *Drosophila hydei*. *J. Mol. Biol.*, v. 209, p. 493-497, 1989.

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

LASKI, F. A.; RIO, D. C.; RUBIN, G. M. Tissue specificity of *Drosophila P* element transposition is regulated at the level of mRNA splicing. *Cell*, v. 44, p. 7-19, 1986.

LERAT, E.; CAPY, P. Retrotransposons and retroviruses: analysis of the envelope gene. *Mol. Biol. Evol.*, v. 16, p. 1198-1207, 1999.

LIM, J. K. Intrachromosomal rearrangements mediated by *hobo* transposons in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA*, v. 85, p. 9153-9157, 1988.

LORETO, E. L. S., *et al.* Distribution of transposable elements in neotropical species of *Drosophila*. *Genetica*, v. 101, n. 3, p. 153-165, 1998.

MAGALHÃES, L. E. Notes on the taxonomy, morphology and distribution of *saltans* group of *Drosophila*, with description of four new species. *Univ. of Texas Publ.*, v. 6205, p. 135-154, 1962.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

MARTIN, G.; WIERNASZ, D.; SCHEDL, P. Evolution of *Drosophila* repetitive-dispersed DNA. *J. Mol. Evol.*, v. 19, p. 203-213, 1983.

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

McDONALD, J. F. Transposable elements, gene silencing and macroevolution. *TREE*, v. 13, p. 94-95, 1998.

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

MILLER, W. J., *et al.* Structure and expression of clustered *P* element homologues in *Drosophila subobscura* and *Drosophila guanche*. *Gene*, v. 156, p.167-174, 1995.

MILLER, W. J.; MCDONALD, J. F.; PINSKER, W. Molecular domestication of mobile elements. *Genetica*, v. 100, p. 261-270, 1997.

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

NASCIMENTO, A. P.; DE CAMPOS BICUDO, H. E. M. Esterase patterns and phylogenetic relationships of *Drosophila* species in the *saltans* subgroup (*saltans* group). *Genetica*, v. 114, p. 41-51, 2002.

NEE, S.; MAYNARD SMITH, J. The evolutionary biology of molecular parasites. *Parasitology*, v. 100, p. 5-18, 1990.

NIKITIN, A. G.; WOODRUFF, R. C. Somatic movement of the *mariner* transposable element and lifespan of *Drosophila* species. *Mutation Res.*, v. 338, p. 43-49, 1995.

NIKITIN, A. G.; SHMOOKLER REIS, R. J. Role of transposable elements in age-related genomic instability. *Genet. Res.*, v. 69, p. 183-195, 1997.

NOUAUD, D., *et al.* A *P* element has induced intron formation in *Drosophila*. *Mol. Biol. Evol.*, v. 16, p. 1503-1510, 1999.

NOUAUD, D.; QUESNEVILLE, H.; ANXOLABÉHÈRE, D. Recurrent exon shuffling between distant *P*-element families. *Mol. Biol. Evol.*, v. 20, p. 190-199, 2003.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

O'GRADY, P. M.; CLARK, J. B.; KIDWELL, M. G. Phylogeny of the *Drosophila saltans* species group based on combined analysis of nuclear and mitochondrial DNA sequences. *Mol. Biol. Evol.*, v. 15, n. 6, p. 656-664, 1998.

O'HARE, K.; RUBIN, G. M. Structure of *P* element transposable and their sites of insertion and excision in *Drosophila melanogaster* genome. *Cell*, v. 25, p. 24-35, 1983.

ORGEL, L. E.; CRICK, F. H. C. Selfish DNA: the ultimate parasite. *Nature*, v. 284, p. 604-607, 1980.

PARDUE, M. L., *et al. Drosophila* telomeres: new views on chromosome evolution. *Trends Genet.*, v. 12, p. 48-52, 1996.

PÉLISSON, A., *et al. Gyspy* transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila* flamenco gene. *EMBO J.*, v. 13, p. 4401-4411, 1994.

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

REGNER, L.P., *et al.* Genomic distribution of *P* elements in *Drosophila willistoni* and a search for their relationship with chromosomal inversions. *J. Hered.*, v. 87, p. 191-198, 1996.

RHODE, C. Polimorfismo cromossômico e elementos transponíveis em *Drosophila willistoni*. Porto Alegre, 2000. Tese (Doutorado). Universidade Federal do Rio Grande do Sul.

ROBERTSON, H. M. The *mariner* transposable element is widespread in insects. *Nature*, v. 362, p. 241-245, 1993.

_____ The *Tc1-mariner* superfamily of transposons in animals. *J. Insect Physiol.*, v. 41, p. 99-105, 1995.

ROBERTSON, H. M.; McLEOD, E. G. Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. *Insect Mol. Biol.*, v. 2, p. 125-139, 1993.

ROBERTSON, H. M.; ZUMPANO, K. L. Molecular evolution of an ancient *mariner* transposons, *Hsmar1*, in the human genome. *Gene*, v. 205, p. 203-217, 1997.

RODRÍGUEZ-TRELLES, F.; TARRÍO, R.; AYALA, F. J. Molecular evolution and phylogeny of the *Drosophila saltans* species group inferred from the *Xdh* gene. *Mol. Phylogen. Evol.*, v.13, p. 110-121, 1999.

SANMIGUEL, P., *et al.* Nested retrotransposons in the intergenic regions of the maize genome. *Science*, v. 274, p. 765-768, 1996.

SONG, S. U., *et al.* An env-like protein encoded by a *Drosophila* retroelement: evidence that *gypsy* is an infectious retrovirus. *Genes Dev.*, v. 8, p. 2046-2057, 1994.

STACEY, S. N., *et al.* Distribution and conservation of mobile elements in the genus *Drosophila*. *Mol. Biol. Evol.*, v. 3, p. 522-534, 1986.

STRECK, R. D.; MACGAFFEY, J. E.; BECKENDORF, S. K. The structure of *hobo* transposable elements and their insertion sites. *EMBO J.*, v. 5, p. 3615-3623, 1986.

STURTEVANT, A. H. The classification of the genus *Drosophila*, with descriptions of nine new species. *Univ. Tex. Publ.*, v. 4213, p. 6-51, 1942.

SYVANEN, M. The evolutionary implications of mobile genetic elements. *Ann. Rev. Genet.*, v. 18, p. 271-293, 1984.

THROCKMORTON, L. H. The phylogeny, ecology and geography of *Drosophila*. In: King , R. C. (ed.) *Handbook of genetics*. Plenum, NY, v. 3. 1975, p. 421-469.

THROCKMORTON, L. H.; MAGALHÃES, L. E. Changes with evolution of pteridine accumulations in species of the *saltans* group of the genus *Drosophila*. *Univ. Texas Publs.*, v. 2, n. 6205, p. 489-505, 1962.

TSUBOTA, S. I., *et al.* The cloning of the *Bar* region and the B breakpoint in *Drosophila melanogaster:* evidence for a transposons-induced rearrangement. *Genetics*, v. 122, p. 881-890, 1989.

VÁZQUEZ-MANRIQUE, R. P., *et al.* Evolution of *gypsy* endogenous retrovirus in the Drosophila obscura species group. *Mol. Biol. Evol.*, v. 17, p. 1185-1193, 2000.

WILDER, J.; HOLLOCHER, H. Mobile elements and the genesis of microsatellites in Dipterans. *Mol. Biol.Evol.*, v. 18, p. 384-392, 2001.

CASTRO, J. P. Reavaliação da distribuição, diversidade de seqüências, número de cópias, e atividade transcricional de elementos transponíveis no grupo *saltans* de *Drosophila*, com ênfase no elemento *P*. São José do Rio Preto, 2003. 132p. Tese (Doutorado em Genética) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista "Júlio de Mesquita Filho".

No presente estudo investigou-se a distribuição; diversidade de seqüências; número de cópias; e atividade transcricional de elementos transponíveis no grupo *saltans* de *Drosophila*. Para isto, foram empregadas as técnicas de *Southern blot*, PCR, RT-PCR de tecidos germinativo e somático, e seqüenciamento. Os resultados mostraram que ao contrário do que se pensava, existem seqüências homólogas ao elemento *P* em todos os subgrupos do grupo *saltans*, entretanto, apenas as espécies dos subgrupos *sturtevanti* e *saltans* parecem possuir cópias potencialmente completas. Elementos *P* completos foram seqüenciados em *D. sturtevanti* e em *D. prosaltans* e a análise filogenética realizada mostrou que estas seqüências pertencem a uma mesma subfamília, para a qual se acreditava existirem poucas seqüências. *D. saltans* parece ser a única espécie do grupo que ainda apresenta uma seqüência *P* que pode ser considerada transcricionalmente ativa, e provavelmente transposicionalmente ativa, por possuir transcritos que codificam para a transposase e para a proteína repressora da transposição. A comparação do número de substituições ocorridas ao longo do tempo nas seqüências *P* transcritas, e em seqüências de *Adh*, sugere que as seqüências *P* transcritas em *D. saltans* e *D. prosaltans* provavelmente estavam presentes no ancestral do subgrupo *saltans*; sugere também que a seqüência *P* transcrita em *D. sturtevanti* pode ter sido transmitida horizontalmente a partir de *D. saltans*. Além do estudo do elemento *P*, outros elementos transponíveis foram avaliados quanto a sua distribuição e o seu número de cópias em espécies do grupo *saltans*. O elemento *mariner* não parece ser amplamente distribuído nas espécies do grupo, enquanto que os elementos *copia*, *gypsy*, *micropia*, *I*, *hobo*, *Minos* e *Bari-1* estão presentes nas espécies dos cinco subgrupos deste grupo de *Drosophila*.

Palavras-chave: *Drosophila*, grupo *saltans*, elementos transponíveis, divergência de seqüências, número de cópias, transposase.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

CASTRO, J. P. Re-evaluation of transposable elements distribution, sequence divergence, copy number, and transcriptional activity in the *saltans* group of *Drosophila*, with emphasis on the *P* element. São José do Rio Preto, 2003. 132p. Tese (Doutorado em Genética) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista "Júlio de Mesquita Filho".

In this study we investigated the transposable elements distribution, sequence divergence, copy number, and transcriptional activity in the *saltans* group of *Drosophila*. To do so, we employed *Southern blot*, PCR, RT-PCR from somatic and germline tissues, and sequencing techniques. The results showed that *P* element homologous sequences are present in all *saltans* subgroups. However, only the species of the *sturtevanti* and *saltans* subgroups seem to harbor putative full-sized *P* element copies. Full-sized *P* elements were sequenced in *D. sturtevanti* and *D. prosaltans;* and the phylogenetic analysis showed that these sequences belong to the same subfamily. From the *saltans* group, we detected both transposase and repressor mRNAs only in *D. saltans*. So, this species seems to be the only species in the *saltans* group that possess transcriptionally active, and probably transpositionally active *P* element sequences. The comparison of the number of found in the *P* element transcribed sequences and those found in *Adh* sequences suggests that the *P* element sequences transcribed in *D. saltans* and *D. prosaltans* were present in the ancestral of the *saltans* subgroup. This analysis also suggests that the *P* element sequence transcribed in *D. sturtevanti* might have been horizontally transferred from *D. saltans*. Additionally to the *P* element analyses, we also studied the distribution and copy number of other transposable elements in *saltans* group. From this survey, we concluded that *mariner* element was not widely distributed in the *saltans* group. Nevertheless, *copia, gypsy, micropia, I, hobo, Minos* and *Bari-1* elements were present in all the species analyzed.

Key words: *Drosophila, saltans* group, transposable elements, sequence divergence, copy number, transposase.

cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18

Autorizo a reprodução deste trabalho
São José do Rio Preto, 15/04/2003

_____

Juliana Polachini de Castro

unesp