UNIVERSIDADE ESTADUAL DE SÃO PAULO

THIAGO FERNANDES

Sequências de DNA da estrutura cromossômica terminal de dípteros da família Sciaridae.

DNA sequences at terminal chromosome structure of Sciaridae flies

São Paulo
2012

THIAGO FERNANDES

Sequências de DNA da estrutura cromossômica terminal de dípteros da família Sciaridae.

DNA sequences at terminal chromosome structure of Sciarid flies

Tese apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Doutor Ciências, na Área de Genética e Biologia Evolutiva

Orientador (a): Eduardo Gorab

São Paulo

2012

Ficha Catalográfica

Fernandes, Thiago

Sequências de DNA da estrutura cromossômica terminal de dípteros da família Sciaridae

158 páginas

Tese (Doutorado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva

1. DNA terminal 2. Sciaridae 3.Microdissecções Cromossômicas - I. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética.

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof. Dr Eduardo Gorab.

Orientador(a)

À minha mãe Márcia Maria Stringheta, à

família e aos amigos, dedico.

... I leave him be, he leaves me be; the paper, of course, is ruined; something has happened, something has soiled my day, sometimes it does not take man or a woman, only something alive; I sit and watch the small one; we are woven together in the air and the living; it is late for both of us.

(2 Flies, Charles Bukowski)

-How many fingers, Winston?
-Four. Four, I suppose there are four. I tried to see five. I wish I could.
-Which do you wish?
(1984, George Orwell)

AGRADECIMENTOS

O doutoramento não pode ser encarado, a meu ver, como um processo decorrente e unicamente vinculado a estes quatro anos de dedicação acadêmica. Mais do que isso, a opção em se tornar "doutor", representa o período de moldar o poético e abstrato projeto de crianças que imaginavam um dia ser "cientistas". É uma etapa de retrospectivas e perspectivas. Etapa de desconstruir, aprimorar, criar ou perpetuar modelos... reticências. Assim, considero-me uma pessoa privilegiada na riqueza e diversidade de modelos e, uma vez finalizada essa tese de doutorado, gostaria de agradecer àqueles que foram agentes fundamentais à minha formação.

Agradeço primeiramente ao Professor Eduardo Gorab por me receber em seu Laboratório mesmo sem me conhecer e por continuar orientando-me mesmo após me conhecer. Indiscutivelmente, a sua postura profissional frente ao que é fazer ciência, os seus diversos "insights" ao decorrer do meu doutoramento e o seu fascínio pela história da ciência preencheram com perfeição as lacunas deixadas pelas orientações anteriores.

Não poderia de deixar de agradecer também à USP e ao Instituto de Biociências por me proporcionarem um ambiente acadêmico cientificamente rico e variado, refletido em termos de qualificação pessoal e pelos inúmeros seminários os quais sempre geravam bons debates e me rendiam grandes indagações.

Agradeço à FAPESP pela bolsa concedida e por me proporcionar a participação tanto em Congressos Nacionais quanto Internações, de grande valia na minha formação acadêmica.

Gostaria de agradecer profundamente aos amigos que me deram suporte, dividiram umas cervejas e fizeram a vida em São Paulo muito menos cinza.

Não poderia deixar de agradecer à Ana, à Maria, ao Zé e ao CA da Biologia por proporcionarem um ambiente sempre alegre e descontraído onde eu podia comemorar os experimentos que deram certos ou chorar os resultados negativos.

Agradeço a todos do laboratório pela convivência nestes quatro anos, pelas trocas de idéias, sugestões e auxílio com os experimentos.

Por fim, gostaria de agradecer à Cidade de São Paulo pela vasta experiência social e cultural proporcionada. Suas inúmeras contradições urbanas e o contraste social sempre visível me deixaram menos Narciso para ser mais Tom Zé.

1 INTRODUÇÃO

1.1 Considerações Gerais Sobre Os Dípteros

Mosquitos, moscas e pernilongos são nomes comuns dados a insetos pertencentes à Ordem Diptera. Embora apenas uma fração minoritária das espécies de dípteros seja de conhecimento público, Diptera é uma das maiores Ordens entre os insetos, sendo composta por mais de 152.000 espécies distribuídas em 10.000 diferentes gêneros, os quais estão organizados em 150 famílias, 22-32 superfamílias, de 8 a 10 infraordens e duas Subordens. Em termos comparativos, o número de espécies de dípteros descritas responde por não menos que um décimo de toda a diversidade encontrada em Metazoa (Wiegman *et al*., 2011).

O grupo fóssil ancestral a todos os dípteros é conhecido do Permiano tardio (há 250 milhões de anos) e uma grande proporção de fósseis de Diptera é encontrada em sedimentos pertencentes ao Período Mesozóico (Labandeira, 2005). As principais linhagens basais de Diptera (Subordem Nematocera) surgiram no final do Triássico, no período entre 25 e 40 milhões de anos, logo após o surgimento da linhagem ancestral à Ordem. Para a Subordem Brachycera, os primeiros fósseis são encontrados em sedimentos relacionados ao início do Jurássico (Wiegman & Yates, 2005).

A longa história evolutiva dos dípteros imprimiu um legado evolutivo marcado por diversificações morfológicas e ecológicas que resultaram na colonização destes insetos nos mais diversos hábitats. Em sua história evolutiva, os dípteros sobreviveram a dois eventos de extinção em massa, apreciaram o surgimento das angiospermas e presenciaram a evolução, predominância e extinção dos dinossauros (Wiegman & Yates, 2005). Assim, a origem remota dos dípteros poderia ser uma justificativa mais do que plausível para a diversificação encontrada na Ordem, pela rápida emergência de novas formas ao longo do tempo.

Recentemente, uma análise filogenética realizada por Wiegman *et al.* (2011) demonstrou que em Diptera não ocorre uma relação linear em que táxons mais antigos são consequentemente os detentores de maior diversidade, o que não corresponde à premissa derivada do que fora postulado no parágrafo anterior. Segundo estes autores, a grande diversidade encontrada em Diptera correlaciona-se a três processos pontuais de grande radiação adaptativa. O primeiro é observado no Triássico com as primeiras linhagens aquáticas de dípteros basais exietentes, originadas no Permiano, apresentando um rápido processo de radiação. Em seguida, no início do Jurássico, as primeiras linhagens terrestres de brachyceros sofreram um processo de grande diversificação concomitantemente ao período hipotisado à origem das primeiras angiospermas. Contudo, a maior radiação adaptativa em Diptera é relatada ao início do Terciário (65-40 milhões de anos atrás) aos dípteros pertencentes à secção Schizophora (Subordem Brachycera). Esse processo de diversificação foi de uma magnitude tal que, juntamente com a ocorrida em Macrolepidoptera (mariposas gigantes), a radiação em Schizophora corresponde à maior radiação ecológica de insetos do Terciário. Para exemplificar tal grandiosidade, a diversidade encontrada em Schizophora responde por

um terço da diversidade encontrada em Diptera e por 3% de toda a diversidade animal (ver referências em Wiegman *et al*., 2011). Desta forma, para estes autores, o motivo pelo qual os dípteros tornaram-se insetos tão diversos deve-se principalmente a processos pontuais, em que algumas linhagens transpuseram eventos de extinção, culminando em uma explosão de radiações ecológicas, e não por emergências recorrentes de novas formas surgidas ao longo do tempo. Tal fato é justificado pela ausência de correlação entre o tamanho e a idade do clados e, também, por mudanças observadas na taxa média de diversificação existente entre diferentes linhagens, a qual coincide com as principais mudanças morfológicas e ecológicas ocorridas nesta Ordem (Figura 1).

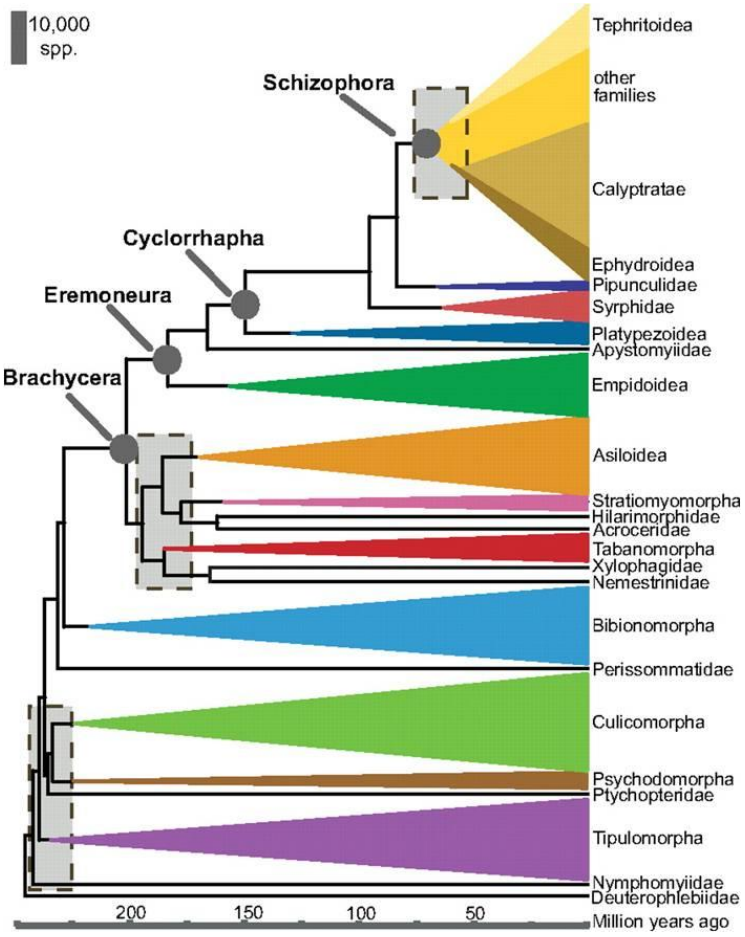


Figura 1: Cronograma representando a filogenia para dípteros e idade estimada de divergências de cada clado. As caixas sombreadas correspondem a áreas de radiação rápida e incerteza filogenética. Estas regiões são relacionadas a períodos de grande e rápida diversificação em Diptera. A altura vertical de cada triângulo representa o número aproximado de espécies descritas em subtipos correspondentes, com a barra de escala que indica 10.000 espécies (Wiegman *et al*., 2011)

Em tempos mais recentes, os dípteros ocuparam, se adaptaram e tornou-se presença marcante em um novo tipo de ambiente, o antrópico. Ao longo do tempo, a relação estabelecida entre dípteros e seres humanos conferiu a estes insetos o status de criaturas impuras e até mesmo satânicas em diferentes civilizações devido ao hábito saprófago de algumas espécies e, principalmente, pela relação direta estabelecida entre dípteros e algumas moléstias (Cornor, 2006). De fato, desde a antiguidade, dípteros são responsáveis por graves problemas de saúde pública, uma vez que algumas espécies são vetores de diferentes agentes etiológicos causadores de doenças como a dengue e a febre amarela (gênero *Aedes*), malária (gênero *Anopheles*) e encefalite (Gênero *Culex*). Embora, estudos com dípteros vetores tenham tido seu início apenas no início do último século, com as espécies *Anopheles gambie e Aedes aegypti*, após mais cem anos de estudos essas duas espécies configuram como principais modelos de estudos relacionados a insetos vetores. Vale ressaltar que a tendência antropofágica, ou antropofílica, de dípteros vetores corresponde apenas a uma pequena fração dos diversos comportamentos alimentares presentes na ordem, sendo muitas espécies componentes importantes do equilíbrio de diversos ecossistemas (Filchak *et al*., 2005).

1.2 Contribuição científica de estudos cromossômicos em dípteros

Estudos científicos nos quais dípteros foram usados como modelo de estudo sobrepõe-se muitas vezes a grandes marcos da recente história da biologia; talvez pela existência de características peculiares encontradas nos representantes desta Ordem, talvez pela obstinação e olhar aguçado de mentes à frente do seu tempo. Indiscutivelmente, a "dipterologia" iniciada no início do século passado definiu a genética clássica e abriu caminho para o desenvolvimento da biologia molecular, tendo

como principal protagonista a espécie *Drosophila melanogaster*. Foi de uma pequena sala localizada na Universidade de Columbia (EUA), vulgarmente conhecida como a "Sala das Moscas", que estudos envolvendo cruzamento entre linhagens selvagens e mutantes naturais de *Drosophila melanogaster* resultaram no célebre livro "*The Mechanism of Mendelian Heredity",* publicado por Morgan, Sturtevant, Muller, e Bridges em 1915. A importância desta publicação foi a de atribuir definitivamente aos cromossomos os padrões de herança Mendelianos. Mais do que compor a teoria cromossômica da herança, as pesquisas realizadas na "Sala das Moscas" definiram ainda: o padrão de herança ligada ao sexo, a ligação entre genes autossomos, os eventos de recombinação cromossômica e o método de mapeamento cromossômico com base na taxa de recombinação entre loci específicos. Embora excepcionais, os mapeamentos genéticos obtidos por esses pesquisadores não eram passíveis de comprovação citológica, devido a uma limitação experimental destes pesquisadores, os quais tinham acesso apenas aos diminutos cromossomos metafásicos de *Drosophila*. Este obstáculo foi transposto com a redescoberta dos trabalhos Balbiani realizados em 1881 com cromossomos presentes em glândulas salivares de larvas do díptero *Chironomus plumosus*, denominados cromossomos politênicos (Sturtevant, 1965).

Embora descritos por Balbiani em 1981, a importância e o significado dos cromossomos politênicos só foram compreendidos quase meio século mais tarde, principalmente pelos estudos de Painter e Bridges realizados em *Drosophila*, de King e Beams com *Chironomus* e o estudo de Heitz e Bauer com *Bibio* (para referências consultar Lacadena, 1996). Posteriormente, os gêneros *Sciara* (Gabrusewycz-García, 1964) e *Rhynchosciara* (Pavan & Breuer, 1952) tornaram-se gêneros empregados em pesquisas envolvendo a biologia dos cromossomos politênicos. A politenia é um caso

especial no qual o DNA em núcleos interfásicos sofre sucessivas rodadas de replicação (5, 10 ou mais) dando origem a cromossomos que estão constituídos por 2^n cromátides (512, 1024, ou, até mesmo, 8192 cromátides). Como consequência, os cromossomos politênicos são geralmente tratados por "cromossomos gigantes", por superarem os cromossomos metafásicos em 100 vezes o seu comprimento e em até 10 vezes a sua espessura. Além disso, a justaposição das cromátides durante os processos consecutivos de endoreplicação resultam em um padrão característico de bandas e interbandas em cromossomos politênicos, o qual é espécie específico. A partir do momento em que foi demonstrado que os padrões de bandas em cromossomos politênicos eram constantes em diferentes preparações, deu-se origem a construção dos primeiros mapas citológicos (Bridiges, 1935, 1938), possibilitando a integração destes com os mapas genéticos e, por consequência, a possibilidade de mapeamentos físicos de genomas. Em 1930, Sturtevant e Dobzhansky estudando populações naturais de *D. pseudoobscura* sugeriram que diferenças nos padrões de bandas em cromossomos politênicos, devido a grandes mutações cromossômicas (normalmente inversões paracêntricas), poderiam ser utilizadas como marcadores filogenéticos, tanto para estudos populacionais quanto interespecífico (para revisões ver Krimbas & Powell, 1992). Como resultado de sua pesquisa, Dobzhansky publicou em 1937 a primeira edição de seu livro *Genetics and the Origin of Species,* um dos livros mais importantes de evolução do século XX.

Após mais de 70 anos de pesquisas, os cromossomos politênicos configuram ainda como um dos modelos mais importantes não somente para a genética clássica. Tornaram-se ferramentas uteis para estudos populacionais e filogenéticos e àqueles que visam uma melhor compreensão da organização e fisiologia dos cromossomos interfásicos. A colinearidade dos padrões de bandas destes cromossomos

foi foco de estudos filogenéticos e populacionais, principalmente em Chironomidae, Culicidae, Simuliidae e Drosophlidae. Destacam-se ainda os estudos detalhando as alterações nos padrões de bandas, manifestados na forma de *puffs* cromossômicos, ao longo do desenvolvimento larval e início da metamorfose em *Drosophila* e *Chironomus,* os quais propiciaram um detalhado modelo de expressão gênica controlada pelo hormonio ecdisterona nestes insetos (ver Ashburner, 2005). Consequentemente, a existência de *puffs* cromossômicos passou a ser relacionadas a regiões cromossômicas-gênicas com alta atividade transcricional, dando origem ao termo *puffs* de RNA. Contudo, estudos realizados com o sciarideo neotropical *Rhynchosciara americana* pelo cientista brasileiro Crodowaldo Pavan demonstraram que em alguns loci específicos os puffs formados ao longo do desenvolvimento desta espécie não estava relacionado a um comportamento da cromatina frente a uma alta atividade transcricional. Surpreendentemente, a formação destes *puffs* específicos devia-se a ciclos adicionais de replicação nestes sítios, ou seja, eram *puffs* de DNA (Pavan & da Cunha, 1969; Glover *et al*.,1982).

Nesta última década, um grande salto informativo quanto ao genoma dos dípteros foi obtido pelo sequênciamento completo do genoma de três espécies desta Ordem: *Drosophila melanogaster*, *Drosophila pseudoobscura* e *Anopheles gambie*. A razão para escolha do sequeciamento destas três espécies teve como base a extraordinária utilidade da *D. melanogaster* como organismo modelo, tanto para estudos relacionados à ciência básica quanto à biomédica, e a escolha de *D. pseudoobscura* pela necessidade comparativa dentro do gênero, uma vez que uma divergência de 25-55 milhões de anos é estipulada para estes dois drosophilídeos. Já a espécie *A. gambie,* foi escolhida por ser o mosquito transmissor da malária. Embora cientificamente

entusiasmante e de importância inquestionável, as informações resultantes destes três projetos revelam a organização de apenas três dentre mais de 150.000 genomas existentes na Ordem. Adicionalmente, estudos prévios ressaltam uma grande variação quanto ao tamanho, composição e organização dos genomas em Diptera, mesmo entre espécies do mesmo gênero. Por exemplo, o tamanho do genoma haplóide nesta Ordem pode variar de 0,09 pg (90 Mb), no cecidomídeo *Mayetiola destrucdor* à 1.9 pg (1,9 Gb) em *Aedes zoosophus*. Em populações de *Aedes albopictus* tem sido reportada uma variação de até três vezes no conteúdo de DNA, de 0,62 pg a 1,66 pg. Além do mais, estas variações quanto ao tamanho do genoma representa um subproduto da amplificação e dispersão diferencial das mais variadas categorias de DNAs repetitivos: DNAs satélites, transposons e retrotransposon (referências em Ashburne 2005). Dessa forma, o sequênciamento completo do genoma das três espécies mencionadas acima pouco esclarecem questões importantes, principalmente quanto ao padrão diferencial de amplificação e dispersão de diferentes classes de DNAs repetitivos compondo os genomas em Diptera. Um caso que merece atenção especial quanto ao padrão diferencial de amplificação e dispersão de diferentes DNAs repetitivos em Diptera refere-se ao seu DNA cromossômico terminal, o qual escapa ao modelo geral encontrado em eucariontes.

1.3 O DNA terminal e a estrutura telomérica.

A evolução dos cromossomos circulares em estruturas lineares teve como principais obstáculos: *i*) a incapacidade da replicação do DNA terminal pelo complexo da DNA polimerase, *ii*) a proteção das extremidades geradas do ataque de exonucleases e *iii*) evitar o reconhecimento das extremidades por enzimas envolvidas no sitema de

reparo de DNA. Independente da estratégia utilizada para estabilizar as extremidades resultantes da conformação cromossômica linear, a região terminal dos cromossomos eucarióticos estruturas nucleoprotéicas especializadas denominadas *telômeros*. A principal consequência do surgimento dos cromossomos lineares foi a possibilidade de estruturas meioticas estáveis, ou seja, passíveis de pareamento cromossômico, recombinação e segregrações pós-sinápticas balanceadas. Por consequência, o surgimento de estruturas cromossômicas lineares está diretamente relacionado à origem da reprodução sexual (Ishikawa & Naito 1999, Nosek *et al*., 2006).

O termo telômero, do grego *telos* (final) e *meros* (parte), foi empregado pela primeira vez por Muller (1938) ao notar que as regiões terminais dos cromossomos de *Drosophila* nunca eram perdidas após rearranjos cromossômicos induzidos por raios-X. Nos anos seguintes, estudos meióticos em milho mostraram que as extremidades geradas pela quebra de cromossomos dicêntricos pareciam ser "cicatrizadas", impedindo fusões cromossômicas (MacClintock, 1941, 1942). Postulou-se então que os telômeros seriam estruturas protetoras dos cromossomos, evitando possíveis degradações do material genético, rearranjos cromossômicos e fusões intercromossômicas.

As observações Muller e MacClintock levaram a uma intensificação de estudos explorando a organização da estrutura telomérica. Nas últimas décadas, os telômeros deixaram de ser vistos apenas como uma mera região protetora dos cromossomos. Mais do que isso, os telômeros, e os mecanismos de regulação telomérica, têm sido alvo de estudos clínicos e da indústria farmacêutica devido a sua relação com os processos de envelhecimento celular, imortalização celular e câncer

(Campisi, Kim, Lim, & Rubio, 2001; Cech, 2004; McEachern, Krauskopf, & Blackburn, 2000). Outros eventos, como o controle do ciclo celular, a associação dos cromossomos à lâmina nuclear, e o pareamento dos cromossomos homólogos na meiose são também de alguma forma influenciados pela estrutura telomérica (Zakian, 1995). Além do mais, recentemente foi postulado que os telômeros originaram os centrômeros durante a evolução dos cromossomos eucarióticos (Villasante, Abad, & Méndez-Lago, 2007). Desse modo, o DNA telomérico representa um componente importantíssimo para a fisiologia cromossômica e celular, além de um excelente substrato para estudos concernentes ao surgimento e à evolução dos cromossomos eucarióticos.

A caracterização telomérica de alguns táxons revelou que o DNA telomérico da maioria dos eucariontes é composto por sequências curtas organizadas em *tandem*, seguindo a fórmula geral 5'-[T(1-4) A (0-1)G(1-6)]n- 3' (revisado em Zakian, 1995). A conservação dessa seqüência consenso entre os eucariontes é tão alta que o termo ***telômeros canônicos*** é muitas vezes usado para se referir aos telômeros com essa constituição. Dentro de alguns grupos filogenéticos, a conservação das repetições teloméricas é mais precisa. Por exemplo, todos os vertebrados possuem a repetição hexamérica (TTAGGG)n (Meyne *et al*., 1989) e os telômeros da maioria dos vegetais superiores possuem um motivo heptamérico (TTTAGGG)n (revisado em Zellinger & Riha, 2007). Em insetos, uma seqüência pentamérica (TTAGG)n é altamente compartilhada, ainda que não seja o único motivo telomérico (Frydrychová *et al*., 2004; Vítková *et al*., 2005).

A replicação do DNA telomérico em organismos que apresentam repetições simples e curtas, como as descritas acima, é realizada por um mecanismo de

transcrição reversa efetuado por uma ribonucleoproteína sugestivamente chamada de *telomerase*. Essa enzima evita o encurtamento dos telômeros utilizando o seu componente de RNA como molde para a síntese de seqüências teloméricas que são adicionadas a uma protusão da fita 3`OH rica em G pelo seu conjunto protéico, possibilitando assim a extensão da fita complementar pelo complexo DNA polimerase (Chan & Blackburn, 2004). Em humanos, trabalhos demonstraram que embora mais de 30 proteínas estejam relacionadas com o complexo enzimático da telomerase, apenas duas proteínas desempenham a função catalítica: TART e *disquerina* (ver Mergny *et al*., 2007).

O sucesso do sistema de replicação via telomerase é refletido pelo seu produto, o DNA telomérico, o qual é uma das sequências de DNA satélite mais abundante encontrada na natureza (de Lange, 2004). No entanto, exceções a esse modelo geral foram relatadas em organismos e tipos celulares em que a função catalítica da telomerase foi perdida e, principalmente, em seres vivos nos quais o DNA telomérico divergiu significantemente das sequências canônicas descritas anteriormente, sendo consequentemente chamados de ***telômeros não canônicos***. Em ambos os casos, mecanismos alternativos para elongação telomérica como a retrotransposição, recombinações cromossômicas, a replicação em círculo rolante e até mesmo uma replicação via DNA extracromossomal têm sido recrutados para a replicação telomérica na ausência da telomerase (Biessmann & Mason, 1997, 2003; Louis & Vershinin, 2005).

1.1.1 Telômeros não canônicos

Telômeros não canônicos têm sido caracterizados principalmente em dois grandes grupos, plantas e insetos. Independentemente do grupo estudado, diferentes famílias de retroelementos ou de sequências complexas de DNAs satélites (>300 pb) parecem substituir as sequências curtas de DNA não codificante encontradas nos telômeros canônicos (Biessmann & Mason, 1997, 2003; Louis & Vershinin, 2005; Zellinger & Riha, 2007). Em plantas, telômeros não canônicos foram encontrados em algumas monocotiledôneas do gênero *Allium* (Asparagalles) (Pich *et al*., 1996; Pich & Shubert, 1998) e nos três gêneros que formam a tribo Cestreae (Solanaceae): *Cestrum*, *Sessea* e *Vestia*, dicotiledôneas (Sýryková *et al*., 2003). Por outro lado, em insetos a adição de seqüências teloméricas por mecanismos independentes da ação da telomerase parece ser um evento muito mais fortuito e freqüente se comparado às plantas. A seqüência consenso (TTAGG)n, encontrada na maioria dos artrópodes e insetos, foi perdida em algumas espécies da ordem Coleoptera e em todas as espécies das Ordens Heteroptera, Dermaptera e Diptera até o momento estudadas (Frydrychová *et al*., 2004; Vítiková *et al*., 2005). Além disso, a atividade da telomerase não foi detectada em *Drosophila* e *Sarcophaga* (dípteros) e nem no lepidóptero *Bombyx mori*, o qual possui telômero canônico (Okasaki *et al*., 1993; Sasaki & Fujiwara, 2000). Surpreendentemente, têm-se sugerido que dois retrotransposons (TRAS e SART) sejam os responsáveis pela manutenção do motivo (TTAGG) nos telômeros de *B. mori* (Fujiwara *et al*., 2005). Já em Diptera, *Drosophila melanogaster* é o modelo para o entendimento da manutenção dos telômeros nâo-canônicos, uma vez que nesta espécie esse processo tem sido estudado em detalhe.

1.4 Estrutura Telomérica em Diptera

O gênero *Drosophila* é, entre os dípteros, o que possui maiores informações quanto à composição e organização da sua estrutura telomérica. Mesmo em espécies evolutivamente muito distantes, como *Drosophila melanogaster* e *D. virilis*¸ o DNA telomérico é composto principalmente por arranjos de duas classes de retroelementos, chamados *HET*-A e TART. Esses retroelementos evitam o encurtamento cromossômico a cada divisão celular através de transposições sítio-específicas na região terminal dos cromossomos, o que é de se estranhar tratando-se de segmentos de DNA ditos egoístas. A composição e o tamanho desses arranjos podem variar entre diferentes extremidades cromossômicas, sendo as seqüências *HET*-A sempre as mais abundantes (revisões em Melinikova & Geogiev, 2005; Pardue *et al*., 2005).

Em *D. melanogaster*, outro retroelemento, intitulado TAHRE, foi descrito (Abad *et al*., 2004).Vale lembrar, que, inicialmente, uma repetição aparentemente telomérica fora identificada em *D. virilis* e intrigantemente era mais semelhante às encontradas em *Chironomus* quando comparada a de *D. melanogaster* e *D. yakuba* (Biessmann *et al*., 2000). A posterior caracterização dos retrotransposons da família TART e *HET*-A como os componentes teloméricos de *D. virilis* (Casacuberta & Pardue, 2003) mostram que a exploração de regiões terminais dos cromossomos de dípteros não é trivial e pode trazer algumas dificuldades e inúmeras surpresas mesmo quando organismos bem conhecidos como *Drosophila* são estudados.

Um apoio adicional ao fato de que os telômeros de dípteros escapavam ao modelo conservado de organização visto em outros eucariontes veio com a caracterização de sequências teloméricas em espécies do gênero *Chironomus*. O DNA telomérico nestas espécies encontra-se estruturado em repetições complexas em tandem que variam de 170 a 350 pares de bases (Saiga & Edström, 1985; Carmona *et al*., 1985; Nielsen *et al*., 1990; Rosén & Edström, 2000; Martinez *et al*., 2001). Em espécies de *Chironomus*, DNA telomérico hibrida na maioria, mas não em todos os extremos cromossômicos. Nestas espécies, extremidades cromossômicas coincidentes com os centrômeros (telocêntricas) perderam as seqüências teloméricas e o DNA centromérico é a seqüência terminal (Rovira *et al*., 1993). Outra característica interessante observada em *Chironomus* é a existência de RNA complementar ao DNA telomérico (Saiga & Edström, 1985; Carmona *et al*., 1985) e de transcriptase reversa associada transitoriamente às extremidades cromossômicas (López *et al*., 1999). Esses eventos foram posteriormente relacionados à geração de complexos de DNA-RNA teloméricos extracromossomais, os quais poderiam intermediar a manutenção telomérica (Rosén *et al*., 2002).

Unidades de repetição complexas aparentemente teloméricas foram também caracterizadas em *Anopheles gambiae* (Biessmann *et al*., 1998). Foi demonstrado que a extensão do DNA terminal de um cromossomo de *A. gambiae* era possível através de um processo de recombinação (Roth *et al*., 1997), sugerindo assim uma via alternativa de manutenção telomérica em dípteros cujo DNA terminal seja composto de repetições complexas. Porém, anos mais tarde, mostrou-se que esse processo tratava-se de um evento raro, e que a conversão gênica entre as seqüências

satélites terminais deva ser o principal mecanismo de elongação telomérica em *A. gambiae* (Walter *et al*., 2001).

1.5 Estrutura cromossômica terminal na família Sciaridae

Extremidades cromossômicas de duas espécies do gênero *Rhynchosciara* hibridam com poli-rA (Stocker *et al*., 1993). Resultados posteriores sugeriram que o DNA genômico detectado com a sonda de poli-rA nestas espécies seja composto de homopolímeros dA/dT com extensão não usualmente observada em genomas de eucariontes (Stocker & Gorab, 2003). Com base no que foi observado em *Chironomus*, decidiu-se empreender uma busca sistemática de proteínas antigenicamente relacionadas a transcriptases reversas em cromossomos de dípteros estudados pelo laboratório. Os resultados obtidos mostraram que estas proteínas aparecem associadas de forma estável a extremidades não centroméricas de *Rhynchosciara americana* ao longo de todo o desenvolvimento larval (Gorab, 2003). A partir deste dado, iniciou-se a caracterização de elementos de DNA componentes de sua estrutura terminal. Além de um satélite subtelomérico de 414 pares de bases (Madalena & Gorab, 2005), outra seqüência de 22 pares de bases foi identificada, denominada M-22 (Rossato *et al*., 2007). Resultados posteriores sugerem que outro elemento repetitivo de 16 pares de bases possa, juntamente com o M-22, atingir as extremidades cromossômicas desta espécie (Madalena *et al*., 2010). A colocalização citológica destes dois motivos com regiões terminais enriquecidas com homopolímeros dA/dT reforçam a possibilidade de que estruturas cromossômicas não convencionais possam compor extremidades cromossômicas em Sciaridae. É importante salientar que elementos repetitivos curtos compondo a estrutura cromossômica terminal não eram esperados e nunca haviam sido

caracterizados em dípteros. Além disso, um retroelemento denominado "RaTART" foi descrito em extremidades cromossômicas de *Rhynchosciara americana* (Rezende-Teixeira *et al*., 2008). Embora contrastante com os resultados obtidos até o momento, a possibilidade de elementos móveis estarem compondo extremidades cromossômicas dípteros basais representaria uma boa oportunidade para estudos de diversificação telomérica na Ordem.

Embora seja um membro da família Sciaridae empregado há tempo em pesquisa que forneceram, sobretudo, informações importantes sobre aspectos relacionados à fisiologia dos cromossomos politênicos (Amabis 1974; Amabis & Janczur, 1978; Busen *et al*., 1982; Amabis, 1990; Andrioli *et al*., 2008), estudos especificamente voltados para a organização da heterocromatina de *Trichomegalosphys pubescens* nunca foram realizados. Contrastando com *Rhynchosciara* e *Chironomus*, *T. pubescens* (2n=8) não possui cromossomos telocêntricos (com extremidades coincidentes com a região centromérica). Ao contrário, possui em seu lote básico três cromossomos metacêntricos (A, C e X) e um submetacêntrico, B (Amabis, 1974). A ausência de cromossomos não telocêntricos em *T. pubescens* é uma indicativa de divergência quanto ao padrão de amplificação e dispersão da heterocromatina telomérica quando comparada a *Rhynchosciara*.

Quanto à estrutura cromossômica terminal, *T. pubescens*, diferentemente de *Rhynchosciara*, não apresenta enriquecimento de DNA homopolimérico dA/dT ou proteínas relacionadas à transcriptase reversa (Stocker & Gorab, 2003; Gorab, 2003). Estes dados também são sugestivos de divergência em sua estrutura telomérica quando comparada à de *Rhynchosciara*. Além disto, RNA na forma híbrida DNA/RNA foi

imunologicamente detectado em telômeros deste organismo (Büsen *et al*., 1982). Diante dos recentes resultados com relação à organização terminal dos cromossomos de *Rhynchosciara*, que escapam ao que tem sido observado em outros dípteros, chamam atenção para a necessidade de um estudo comparativo dentro de Sciaridae. Neste sentido, *T. pubescens* representa a oportunidade única de realizar um estudo comparativo entre dois gêneros pertencentes à mesma família, o que não foi feito em *Drosophila*, *Chironomus* e *Anopheles*.

2 OBJETIVOS

O objetivo geral deste trabalho foi o de ampliar a informação disponível sobre telômeros não canônicos, com ênfase na caracterização de sequências componentes da estrutura cromossômica terminal de dípteros da família Sciaridae. Quanto aos objetivos específicos, investigamos inicialmente a possibilidade de que uma das espécies, *R. americana*, apresentasse retrotransposons teloméricos a exemplo de *Drosophila*, tendo como ponto de partida um trabalho que havia sido publicado sobre este tema. Em seguida, prosseguindo com estudos na mesma espécie, o foco da pesquisa tomou outra direção e uma nova sequência repetitiva localizada em extremidades cromossômicas de *R. americana* foi descoberta a partir de indícios de sua existência que datam do início da década de 1970. A exploração da estrutura cromossômica terminal de *T. pubescens*, espécie que nunca havia sido estudada sob tal ponto de vista, constitui a última parte deste trabalho. Resultados não esperados obtidos nesta espécie são de interesse considerável para um aprofundamento no futuro.

REFERÊNCIAS

Abad JP, Pablos B, Osoegawa K, Jong PJ, Martín-Gallardo A, Villasante A (2004) TAHRE, a Novel Telomeric Retrotransposon from *Drosophila melanogaster*, Reveals the Origin of *Drosophila* Telomeres. Mol Biol Evol 21: 1620–1624.

Amabis JM. (1974) Estudo dos Cromossomos de *Trichomegalosphys pubescens*. Tese de doutorado. Universidade de São Paulo, Brasil.

Amabis JM, Janczur C (1978) Experimental Induction of Gene Activity in the Salivary Gland Chromosomes of Trichomegalosphys pubescens (Diptera: Sciaridae). J Cell Bio 78:1-7.

Amabis JM (1990) The presence of an antigen reactive wuth a human autoantibody in Trichomegalosphys pubescens (Diptera:Sciaridae) and its association with certain transcriptionally active regions of the genome. Chromosoma 99:102-110.

Andrioli LP, Gorab E, Amabis JM (2008) The DNA Puff 4C Expresses a Salivary Secretion Protein in *Trichomegalosphys pubescens* (Diptera; Sciaridae). Archives of Insect Biochemistry and Physiology 67:76–86.

Ashburne (2005) "The genomes of Diptera" IN: B. Wiegmann and D. Yeates, (eds.) pp. 217-273". pp. 66-94 in The Evolutionary Biology of Flies (Wiegmann, B. and Yeates, D.). Columbia University Press

Biessmann H, Mason JM (1997) Telomere maintenance without telomerase. Chromosoma 106: 63-69.

Biessmann H, Kobeski F, Walter MF, Kasravi A, Roth CW (1998) DNA organization and length polymorphism at the 2L telomeric region of *Anopheles gambiae*. Insect Mol Biol 7: 83-93.

Biessmann H, Zurovcova M, Yao JG, Lozovskaya E, Walter MF (2000) A telomeric satellite in *Drosophila virilis* and its sibling species. Chromosoma 109: 372-380.

Biessmann H, Mason JM (2003) Telomerase-independent mechanisms of telomere elongation. Cell Mol Life Sci 60: 2325-2333.

Bridges, C. B. (1935). Salivary chromosome maps. *J. Hered.*, 26: 60–64.

Bridges, C. B. (1938). A revised map of the salivary gland X-chromosome. *J Hered*, 29: 11–13

Büsen W, Amabis JM, Leoncini O, Stollar BD, Lara FJS (1982) Immunofluorescent characterization of DNA-RNA Hybrids on Polytene Chromosomes of *Trichomegalosphys pubescens* (Diptera, Sciaridae). Chromosoma 87: 247-262.

Campisi J, Sahn-Ho K, Chang-su L, Rubio M (2001) Cellular Senescence, cancer and aging: the telomere conection. Exp Geront 36: 1619-1637.

Carmona MJ, Morcillo G, Galler R, Martinez-Salas E, de la Campa AG, Diez JL, Edström JE (1985) Cloning and molecular characterization of a telomeric sequence from a temperature induced Balbiani ring. Chromosoma 92: 108-115.

Casacuberta E, Pardue ML (2003) Transposon Telomeres Are Widely Distributed in the *Drosophila* Genus: TART Elements in the virilis Group. Proc Natl Acad Sci 100: 3363-3368.

Cech TR (2004) Beginning to understand the end of the chromosome. Cell 116: 273-279.

Chan SRWL, Blackburn EH (2004) Telomeres and telomerase. Phil Trans R Soc Lond 359: 109-121.

Connor, Steven, (1955) Fly (Animal) British Library Cataloguing in Publication Data.

de Lange T (2004) T-loops and the origin of telomeres. Nat Rev Mol Cell Biol 5: 323-329.

Filchak, K. E., W. J. Etges, N. J. Besansky, and J. L. Feder. (2005). Ecological genetics of host use in diptera. pp. 340–370 *in* DK. Yeates and BM. Wiegman, eds. The evolutionary biology of flies. Columbia Univ. Press, New York.

Frydrychová R, Grossmann P, Truba P, Vítková M, Marec F (2004) Phylogenetic distribution of TTAGG telomeric repeats in insects. Genome 47: 163-178.

Fujiwara H, Osanai M, Matsumoto T, Kojima KK (2005) Telomere-specific non-LTR retrotransposons and telomere maintenance in the silkworm, *Bombyx mori*. Chromosome Res 13: 455-467.

Gabrusewycz-Garcia, N (1964) Cytological and autoradiography of Sciara coprofila salivary gland chromosome; Chromosoma 15, 312-344.

Gorab E (2003) Reverse transcriptase-related proteins in telomeres and in certain chromosomal loci of *Rhynchosciara* (Diptera: Sciaridae). Chromosoma 111: 445-454.

Glover, D.M., Zaha, A., Stocker, A.J. *et al*. (1982) Gene amplification in *Rhynchosciara* salivary gland chromosomes. *Proceedings of the National Academy of Sciences of the USA* **79**, 2947–2951.

Ishikawa F, Naito T (1999) Why do we have linear chromosomes? A matter of Adam and Eve Mutat Res 434: 99-107.

Krimbas, C. B., & J. R. Poweij (1992) *Lhosophila* Inversion *Polymm*phism. CRC Press, Boca Raton, FL.

Lacadena, J. (1996) Citogenética. Madrid: Editorial Complutense. pp 329-344

Labandeira, C.C. (2005) "Fossil history of the Diptera and their associations with plants. IN: B. Wiegmann and D. Yeates, (eds.) pp. 217-273". Pp. 217-273 in The Evolutionary Biology of Flies (Wiegmann, B. and Yeates, D.). Columbia University Press

López C, Rodriguez E, Díez J, Edström JE, Morcillo G (1999) Histochemical localization of reverse transcriptase in polytene chromosomes of chironomids. Chromosoma 108: 302-307.

Louis E, Vershinin A (2005) Telomeres: a diversity of solutions to the problem of chromosome ends. Chromosome Res 13: 425-429.

Madalena CRG, Gorab E (2005) A chromosome end satellite of *Rhynchosciara americana* (Diptera: Sciaridae) resembling nematoceran telomeric repeats. Insect Mol Biol 14: 255-262.

Madalena CRG, Amabis JM & Gorab E. (2010). Unusually short tandem repeats appear to reach chromosome ends of Rhynchosciara americana (Diptera: Sciaridae). *Chromosoma*, *119*(6), 613-23. doi:10.1007/s00412-010-0282-9

Martinez JL, Sanches-Elsner T, Morcillo G, Diez JL (2001) Heat Shock regulatory elements are presents in telomeric repeats of *Chironomus thummi*. Nucleic Acids Res 29: 4760-4766.

McClintock B (1941) The stability of broken ends of chromosome in *Zea mayz*. Genetics 26: 234-82.

McClintock B (1942) The fusion of broken ends of chromosomes following nuclear fusion. Proc Natl Acad Sci 28: 458-463.

McEachern MJ, Krauskopf A, Blackburn EH (2000) Telomeres and their control. Annu Rev Genet 34: 331-358.

Melinikova L, Georgiev P (2005) *Drosophila* telomeres: the non-telomerase alternative Chromosome Res 13: 431-441.

Mergny JL, Lacroix L, Riou JF (2007) Telomerase: A Dimeric Ménage à Trois. Chem BioChem 8: 1483-1485.

Meyne J, Ratliff RL, Moyzis RK (1989) Conservation of the human telomere sequence $(TTAGGG)_n$ among vertebrates. Proc Natl Acad Sci 86: 7049-7053.

Muller HJ (1938) The remaking of chromosomes. The Collecting Net 8: 182-195.

Nielsen L, Schmidt ER, Edström JE (1990) Subrepeats result from regional DNA sequence conservation in tandem repeats in *Chironomus* telomeres. J Mol Biol 216: 577-584.

Nosek J, Kosa P, Tomaska L (2006) On the origin of telomeres: a glimpse at the pretelomerase World. BioEssays 28: 182-190.

Okazaki S, Tsuchida K, Maekawa H, Ishikawa H, Fujiwara H (1993) Identification of a pentanucleotide telomeric sequence (TTAGG)n in the silkworm *Bombyx mori* and in other insects. Mol Cell Biol 13: 1424-1432.

Pardue ML, Rashkova S, Casacuberta E, DeBaryshe PG, George JA, Traverse KL (2005) Two retrotransposons maintain telomeres in *Drosophila*, Chromosome Res 13: 443-453.

Pavan, C., and M. E. Breuer. 1952. Polytene chromosomes in different tissues of Rhynchosciara. *J. Hered.*, 43: 151–157.

Pavan, C. & da Cunha, A.B. (1969) Chromosomal activities in *Rhynchosciara* and other Sciaridae. *Annual Review of Genetics* **3**, 425–450.

Pich U, Fuchs J, Schubert I (1996) How do Alliaceae stabilize their chromosome ends in the absence of TTTAGGG sequences? Chromosome Res 4: 207-213.

Pich U, Schubert I (1998) Terminal heterochromatin and alternative telomeric sequences in *Allium cepa*. Chromosome Res 6: 315-321.

Rezende-Teixeira, P., Siviero, F., Brandão, A. S., Santelli, R. V., & Machado-Santelli, G. M. (2008). Molecular characterization of a retrotransposon in the Rhynchosciara americana genome and its association with telomere. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, *16*(5), 729-42. doi:10.1007/s10577-008-1223-9

Rosén M, Edström JE (2000) DNA structures common for chironomid telomeres terminating with complex repeats. Insect Mol Biol 9: 341-347.

Rosén M, Kamnert I, Edström JE (2002) Extrachromosomal DNA/RNA complex containing long telomeric repeats in chironomids. Insect Mol Biol 11: 167-174.

Rossato RM, Madalena CRG, Gorab E (2007) Unusually short tandem repeats in the chromosome end structure of *Rhynchosciara* (Diptera: Sciaridae). Genetica 131:109-116.

Roth CW, Kobeski F, Walter MF, Biessmann H (1997) Chromosome end elongation by recombination in the mosquito *Anopheles gambiae*. Mol Cell Biol 17: 5176-5183.

Rovira C, Beermann W, Edström JE (1993) A repetitive DNA sequence associated with the centromeres of *Chironomus pallidivittatus*. Nucleic Acids Res 21: 1775-1781.

Saiga H, Edström JE (1985) Long tandem arrays of complex repeat units in *Chironomus* telomeres. EMBO J 4: 799-804.

Sasaki T, Fujiwara H (2000) Detection and distribution patterns of telomerase activity in insects. Eur J Biochem 267: 3025-3031.

Stocker AJ, Gorab E, Amabis JM, Lara FJS (1993) A molecular cytogenetic comparison between *Rhynchosciara americana* and *Rhynchosciara hollaenderi* (Diptera: Sciaridae). Genome 36: 831-843.

Stocker AJ, Gorab E (2003) Local enrichment with homopolymeric (dA/dT) DNA in genomes of some lower dipterans and *Drosophila melanogaster*. Chromosoma 111: 455-460.

Sturtevant AH (1965) The Hystory of Genetics, New York: Harper & Row, pp 9-58.

Sýkorová E, Lim KY, Chase MW, Knapp S, Leitch IJ, Leitch AR, Fajkus J (2003) The absence of *Arabidopsis*-type telomeres in *Cestrum* and closely related genera *Vestia* and *Sessea* (Solanaceae), first evidence from eudicots. Plant J 34: 283-291.

Villasante A, Abad JP, Méndez-Lago M (2007) Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. Proc Natl Acad Sci 104: 10542-10547.

Vítková M, Král J, Traut W, Zrzavý J, Frantisek M (2005) The evolutionary origin of insect telomeric repeats, $(TTAGG)_n$. Chromosome Res 13: 145-156.

Walter MF, Bozorgnia L, Maheshwari A, Biessmann H (2001) The rate of terminal nucleotide loss from a telomere of the mosquito *Anopheles gambiae*. Insect Mol Biol 10: 105-110.

Wiegmann BM, Trautwein MD, Winkler IS; Barra NB, KIM JW, Lambkin C, Bertone MA, Cassel BK, Bayless KM.; Heimberg AM, Wheeler BM, Peterson KJ, Pape T, Sinclair BJ, Skevington JH, Blagoderov V, Caravas J, Kutty SN, Schmidt-OTT U, Kampmeier GE, Thompson FC, Grimaldi DA, Beckenbach AT, Courtney GW, Friedrich M, Meier R. & Yeates DK (2011) Episodic radiations in the fly tree of life. *Proceedings of the National Academy of Sciences,* 108:5690-5695

Wiegmann B &. Yeates D (2005) The Evolutionary Biology of Flies (Wiegmann B & Yeates D) Columbia University Press

Zakian VA (1995) Telomeres: beginning to understand the end. Science 270: 1601-1607.

Zellinger B, Riha K (2007) Composition of plant telomeres. Biochim Biophys Acta. 1769: 399-409.

CAPÍTULO 1

Curiously composite structures of a retrotransposon and a complex repeat associated with chromosome ends of *Rhynchosciara americana* (Diptera: Sciaridae)

Christiane Rodriguez Gutierrez Madalena.
Thiago Fernandes
Alfredo Villasante[#] and Eduardo Gorab*

Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo. Rua do Matão 277, Cidade Universitária, São Paulo, SP,

Centro de Biología Molecular "Severo Ochoa". Universidad Autónoma de Madrid, Cantoblanco, 28049, Madrid, Spain

Abbreviations

bp: base pair
DAPI: 4',6-diamidino-2-phenylindole
dNTP: deoxy-ribonucleotide 5'-triphosphate
dUTP: 2'-deoxy-uridine 5'-triphosphate
FITC: fluorescein isothiocyanate
LTR: long terminal repeat
Kb: kilo base
Kbp: kilobase pair
ORF: open reading frame
PCR: polymerase chain reaction
rDNA: ribosomal DNA
RT: reverse transcriptase
SDS: sodium dodecyl sulfate
SSC: sodium saline citrate
UTR: untranslated region

Acknowledgements: The authors are indebted to Dr. A.J. Stocker for critical reading of the manuscript. This work was supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo).

RESUMO

Em *Drosophila*, a perda do DNA telomérico é contrabalanceada por um mecanismo de retrotransposição realizado por dois retroelementos telômeros-específicos. O mecanismo excepcional de recuperação dos telômeros caracterizado em *Drosophila* não tem sido encontrado em dípteros inferiores (Nematocera) em que sequências complexas de DNAs satélites (>170pb) desempenham a função de manutenção telomérica (*Anopheles* e *Chironomus*). Adicionalmente, sequências curtas (16pb e 22pb) tem sido localizadas citologicamente em extremidades cromossômicas não-centroméricas do sciarídeo *Rhynchosciara americana* (Nematocera). No entanto, um retroelemento denominado "RaTART" foi descrito em extremidades cromossômicas de *R. americana*. A possibilidade de um mecanismo de manutenção telomérica pela ação de elementos móveis em dípteros basais representa uma boa oportunidade para estudos de diversificação telomérica em Diptera, além de possibilitar um melhor entendimento da manutenção telomérica por intermédio de elementos móveis. Por estas razões, nós resolvemos realizar análises adicionais do elemento denominado "RaTART". Análises *in silico* foram primeiramente realizadas com a sequência "RaTART" depositada no GenBank para um melhor detalhamento da sequência em si e de suas regiões codificantes. Posteriormente, *primers* foram desenhados para a amplificação de trechos significativos das regiões 5'UTR e 3'UTR deste elemento e também da região deseguinada como codificante para a transcriptase reversa (RT). Os produtos obtidos da amplificação de DNA de glândulas salivares foram clonados e sequênciados, apresentando alta similaridade com as sequências esperadas. Intrigantemente, quando avaliados por hibridação *in situ*, dos três clones obtidos apenas um produziu sinal nas cinco extremidades não telocêntricas de *R. americana* (clone 3'UTR), enquanto marcações nas heterocromatina pericentromérica e em regiões eucromáticas foram observadas para os dois clones restantes. Coincidentemente, o clone que hibridou em extremidades não-centroméricas possui em sua região terminal uma sequência satélite anteriormente descrita como subterminal em *R. americama* e que está presente na porção final da sequência "RaTART" publicada (3'UTR). Os resultados obtidos, indicam que o

retroelemento denominado “RaTART” corresponde a um clone genômico quimérico, composto por distintos elementos repetitivos dos quais pelo menos um é reponsável pelo aparente enriquecimento terminal. Em adição, o estudo despendido com o “RaTART” nos permitiu acessar uma nova repetição complexa relacionada à região subtelomérica de *R. americana*. Obtida, por triagem de biblioteca genômica, esta nova sequência está localizada próxima às sequências curtas organizadas em tandem presentes nas extremidades não centroméricas desta espécie. Uma análise quantitativa demonstrou um número inferior de cópias desta sequência em relação à existente de DNAr em glandulas salivares de *R. americana*. Adicionalmente, um trecho curto do “RaTART” foi identificado neste segmento clonado o qual hibrida preferencialmente em regiões subteloméricas. Assim como o "RaTART", este fragmento exibe sequências truncadas relacionadas a retrotransposons distintos, e uma similaridade significativa com uma endonuclease de um retrotransposon de lepidópteros é encontrada após a tradução de uma destas sequências. A estrutura complexa deste trecho de DNA reflete provavelmente a actividade de elementos móveis em regiões subteloméricas analisadas neste trabalho

Palavras-chaves: *Rhynchosciara*, *Drosophila*, heterocromatina, retrotransposons, subtelômeros e telômeros.

ABSTRACT

In *Drosophila*, telomere retrotransposons counterbalance the loss of telomeric DNA. The exceptional mechanism of telomere recovery characterized in *Drosophila* has not been found in lower dipterans (Nematocera). However, a retroelement resembling a telomere transposon and termed "RaTART" has been described in the nematoceran *Rhynchosciara americana*. In this work, DNA and protein sequence analyses, DNA cloning and chromosomal localization of probes obtained either by PCR or by screening a genomic library were carried out in order to examine additional features of this retroelement. The analyses performed raise the possibility that "RaTART" represents a genomic clone composed of distinct repetitive elements, one of which is likely to be responsible for its apparent enrichment at chromosome ends. "RaTART" sequence in addition allowed to assess a novel subtelomeric region of *R. americana* chromosomes that was analyzed in this work after subcloning a DNA fragment from a phage insert. It contains a complex repeat that is located in the vicinity of simple and complex tandem repeats characterized previously. Quantification data suggest that the copy number of the repeat is significantly lower than that observed for the rDNA in the salivary gland of *R. americana*. A short insertion of the "RaTART" was identified in the cloned segment which hybridized preferentially to subtelomeres. Like "RaTART", it displays truncated sequences related to distinct retrotransposons, one of which has a conceptual translation product with significant identity with an endonuclease from a lepidopteran retrotransposon. The composite structure of this DNA stretch probably reflects mobile element activity in the subtelomeric region analyzed in this work.

Key words: *Rhynchosciara*, *Drosophila*, heterochromatin, retrotransposons, subtelomeres, telomeres.

1 INTRODUCTION

Telomere regeneration based on short repeats and telomerase has not been identified in dipterans. Telomeres in *Drosophila melanogaster* subgroup species are composed of *Het-A*, *TART* and *TAHRE* retrotransposons (reviewed in Mason et al. 2008; Villasante et al. 2008). Among lower dipterans, suborder Nematocera, tandem arrays of complex telomeric satellites reach the termini of *Chironomus* (Diptera: Chironomidae) species (Carmona et al. 1985; Saiga and Edström 1985; López et al. 1996; Rosén and Edström 2000; Martinez et al. 2001) and *Anopheles gambiae* (Biessmann et al. 1998).

Except for *Drosophila*, it is not known how telomeres are maintained in dipterans. Recombination processes at chromosome ends composed of complex repeats might contribute to telomeric DNA extension as suggested for the nematoceran *Anopheles gambiae* (Roth et al. 1997). On the other hand, reverse transcription, the most conserved mechanism for telomere regeneration, could play some role in telomere maintenance of lower dipterans as argued in the two following examples.

Telomeric repeats of the lower dipteran *Chironomus* are transcribed (Carmona et al. 1985; Saiga and Edström 1985) and Hsp90 and Ku-related proteins are present in telomere puffs (Morcillo et al. 1993; Gorab et al. 1996). In addition, reverse transcriptase was detected at telomeres and in nucleoli of distantly related nematocerans (López et al. 1999; Gorab 2003; Díez et al. 2006). Hsp90 and Ku are both involved in

telomeric DNA extension based on the telomerase complex (Holt et al. 1999; reviewed in Bryan and Cech 1999). It is also worth mentioning that the catalytic component of the telomerase has nucleolar localization in lower and higher eukaryotes (Etheridge et al. 2002; Yang et al. 2002; Teixeira et al. 2002). Furthermore, since telomerase-specific repeats are transcribed (Azzalin et al. 2007; Schoefter and Blasco 2008), the picture appearing in *Chironomus* has some indications of a canonical telomere regeneration process. However, the functional link between these telomeric components of dipterans still remains conjectural.

A non–long terminal repeat (LTR) retrotransposon was identified in the nematoceran *Rhynchosciara americana*. Since its open reading frames (ORFs) display similarities with those of *Drosophila* telomeric elements and it has an apparent specificity for chromosome ends in addition to an unusually long 3' untranslated region (UTR), it was termed "RaTART" (Resende-Teixeira et al. 2008). Telomere transposons in lower dipterans could be attractive for the study of diversification of the telomere structure in Diptera and might also provide insights into the evolutionary origin of telomere maintenance based on mobile elements. For this reason, further analysis of the "RaTART" was carried out in this work. The data obtained in this work suggest that the genomic distribution of the "RaTART" is mostly due to one of the repeat types within the cloned DNA which contains stretches representing more than a single transposon. In addition, a novel complex repeat located at subtelomeres of *R. americana* was characterized in this work, a few kilobases away from arrays of simple and complex tandem repeats which were characterized previously. Sequence analyses showed that this

genomic region has a composite structure comprising a stretch of the ”RaTART” and also sequences related to lepidopteran retrotransposons.

2 MATERIALS AND METHODS

2.1 Animals

Rhynchosciara americana larvae were collected in the coastal regions of Mongaguá, State of São Paulo, and maintained in the laboratory at 18-22oC.

2.2 Genomic library

The library was prepared with DNA from *R. americana* salivary glands using Lambda Dash II vector and Lambda Dash II/*Bam*HI cloning kit following the instructions of the manufacturer (Stratagene). Procedures for screening, phage DNA extraction, phage DNA map and subcloning were described in usual protocols (Sambrook et al. 1989). The whole sequence of the p*Ra*1-20 insert was achieved with primer 20a: 5'AGCACAATTCAAAGTGC3'.

2.3 Polymerase Chain Reactions

Salivary gland DNA was extracted according to usual protocols (Sambrook et al. 1989) and amplified in a Perkin Elmer 2400 thermocycler. The following reaction mixture was added to the PCR tube: salivary gland DNA (200 ng); 0.2 mM of each dNTP; 1X Taq buffer; primers (p1a/p1b or p2a/p2b), 2.5 U Taq DNA polymerase (Fermentas); 1 mM $MgCl_2$ and H_2O to a final volume of 50 μl. Cycle conditions: 93^{o}C for 2 min followed by 26 cycles of 94^{o}C for 10 s, 56^{o}C for 20 s and 64^{o}C for 3 min and a

final extension either for 2 min (for probe 1) or 5 min (for probe 2) at 68^{o}C. p2a: 5'-GCATAAGACGGGTGTAGGG-3'; p2b: 5'-CGGATTCCACTCATCGCGC-3'; p1a: 5'-AGGTATTCTGGAGCCGTGC-3'; p1b: 5'-GATATTGCACTCATTTCGG-3'. Standard procedures for cloning of PCR products using pGEM T-Easy vector (Promega) and sequencing were described previously (Madalena and Gorab 2005). DNA and protein searches were done using BLAST (Altschul et al. 1990). Sequence alignments were done using the "Align two highly similar sequences" tool available within BLAST.

2.4 Preparation of chromosome spreads and in situ hybridization

The probes were labeled by nick translation either with digoxigenin 11-dUTP or with biotin 11-dUTP using "Dig-nick" or "Bio-nick" kits according to manufacturer´s protocol (Roche). Preparation of chromosome spreads, probe mixture, hybridization and detection were done as already described (Madalena and Gorab 2005) except for washes that were done either in 0.2X SSC, 0.1% SDS at 40°C (high stringent washes) or in 2X SSC, 0.1% SDS at 30°C (low stringent washes).

2.5 Simultaneous hybridization/reverse transcriptase detection

The procedure was described previously (Madalena and Gorab 2005). Briefly, in situ hybridization using formaldehyde-fixed chromosomes was initiated with slides rehydrated in 1X TBS avoiding air-drying. DNA denaturation by steam heating was done at 70^{o}C for 2 min and the hybridization time was reduced to 2 h. High stringency washes and hybridization detection of digoxigenin-labeled probes were done as described above. Reverse transcriptase detection was made with rabbit IgG raised to a 208 amino acid

peptide containing five domains of reverse transcriptase from a *Chironomus* non-LTR retrotransposon (López et al. 1999). Goat IgG anti-rabbit IgG conjugated with FITC (Sigma) was used as the secondary antibody.

2.6 Southern blot hybridizations and phage insert map

DNA from gels was alkaline transferred to Hybond N+ membranes according to standard procedures (GE Healthcare). Hybridization was carried out overnight at 65^{o}C in 0.5 M Na_2HPO_4, 2% SDS. Plasmid inserts and total salivary gland DNA of *R. americana* were labeled by random priming with α^{32}P-dATP following current protocols (Invitrogen). The membranes were washed twice at 65^{o}C for 30 min in 40 mM Na_2HPO_4, 2% SDS. Restriction enzymes and DNA size markers were purchased from Invitrogen and GE Healthcare. A restriction map of the λ1 phage insert was done on the basis of partial (data not shown) and complete digestions of λ1 DNA using *Eco*RI and *Sal*I performed separately. Products from phage DNA digestion were transferred to membrane replicas and subsequently hybridized to distinct probes: M-22 repeats (Rossato et al. 2007); 414 bp satellite (Madalena and Gorab 2005); probe 2, representing the RT coding region of the "RaTART" (this work) and p*Ra*1-20 insert (containing a novel complex repeat characterized in this work). Quantification was carried out with 1-D Image Analysis Software (Kodak Digital Science^{TM}, Eastman Kodak Co.).

3 RESULTS

3.1 "RaTART" 5'UTR analysis and chromosomal localization

"RaTART" is 8268 base pairs (bp) long and has, like *DmTART*, an unusually long (> 2000 bp) 3'UTR (Resende-Teixeira et al. 2008). In addition, sequences encoding *gag* and reverse transcriptase (RT) proteins were found in the regions identified as ORF1 and ORF2, respectively (Fig. 1a). Sequence analysis showed that DNA stretches in the 5'UTR (as deduced by its position relative to ORFs) align at the coding regions of "RaTART" *gag* and RT (Fig. 1b; SM Fig. S1). Intriguingly, this segment is the non-coding strand of the "RaTART" in comparison with sequences encoding *gag* and RT. Since these are unexpected features of the 5'UTR from retroelements, a probe for this region (probe 1) was synthesized using specific primers (p1a and p1b) and salivary gland DNA. A DNA stretch comprising nucleotides 104-1203 in the "RaTART" sequence (accession number EU 363480) is expected after PCR. PCR products were then cloned and sequenced, displaying very high similarity (95%) to the sequence previously deposited (SM Fig. S2). Probe 1 hybridized consistently to many polytene chromosomal regions, particularly to the pericentric heterochromatin (chromosome sections A10/12, B15, C11, X12) and several euchromatic sites. No enrichment at chromosome ends was observed even when low stringency hybridization and washes were performed (SM Fig. S3).

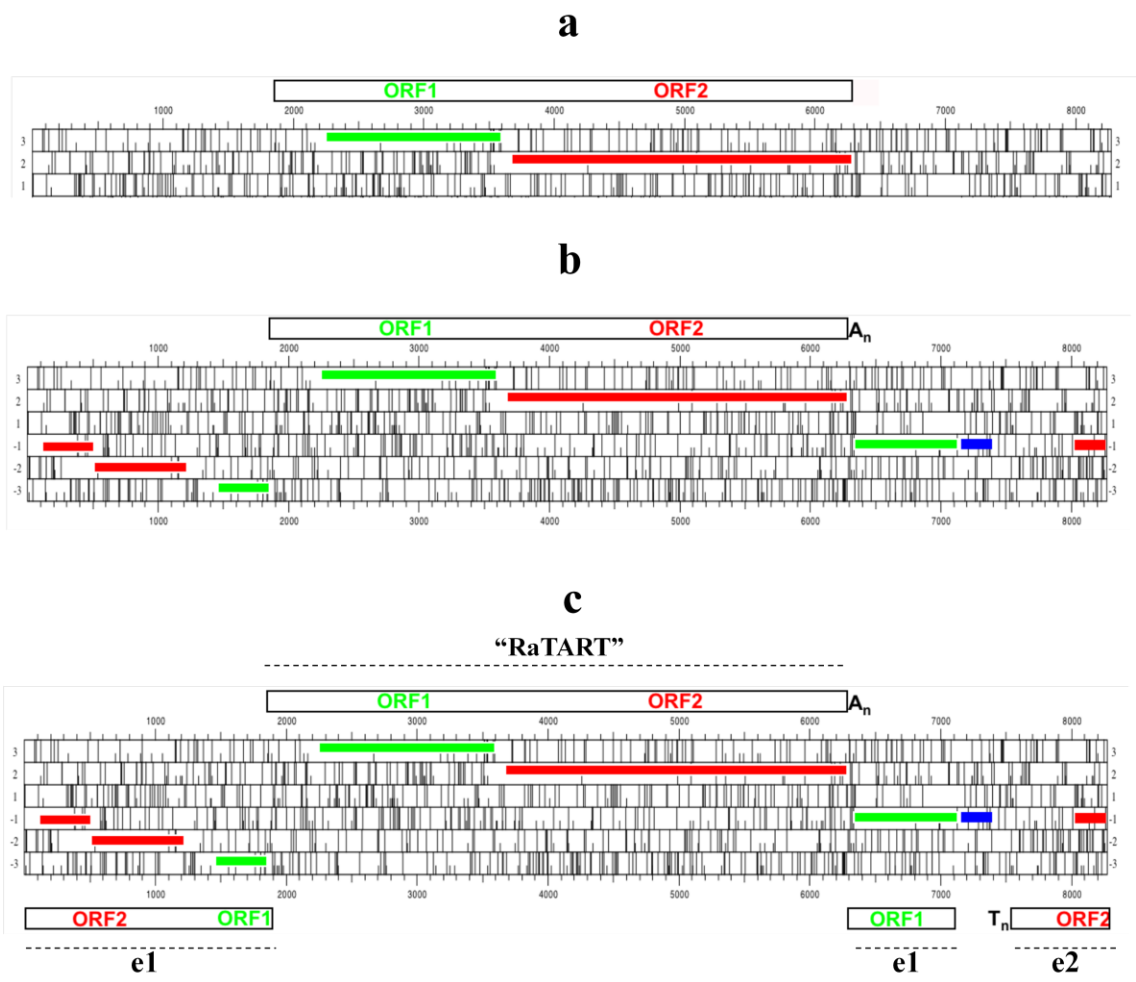


Figure 1: (a) Schematic representation of the "RaTART" (based on Resende-Teixeira et al. 2008) showing positions of *gag* (contained in *ORF*1), endonuclease and reverse transcriptase (RT, contained in *ORF2*) protein coding regions. **(b)** Schematic representation of the "RaTART" genomic clone after sequence analyses performed in this work. The largest green and red bars indicate the position of the coding regions (*ORF1* and *ORF2*, according to Resende-Teixeira et al. 2008). The shorter green and red bars designate the regions of either sequence or protein similarity with ORF1 and ORF2 on the basis of the alignments and conceptual translation showed in SM Fig. S1, SM Fig. S6a and SM Fig. S6c. The position of the nearly complete 414 bp, subtelomeric satellite of *R. americana* is indicated (*blue bar*); sequence similarity is shown in SM Fig. S6b. **(c)** Hypothetical representation of events that may have occurred in the genomic region comprised by the whole "RaTART" sequence on the basis of sequences that were identified throughout its length. Truncated ORFs of an incomplete retrotransposon (*e1*) were interrupted by the insertion of a (*"RaTART"*) non-LTR retrotransposon. A truncated copy of another retrotransposon (*e2*) is suggested with its limits defined by the position of its (*Tn*) (dT)12 stretch.

3.2 Chromosomal localization of sequences comprising “RaTART” endonuclease and RT coding regions

Probes used in the previous report to study the localization of the genomic clone named “RaTART” hybridized consistently to polytene chromosome tips and to several interstitial bands (Resende-Teixeira et al. 2008). Given the localization data of sequences composing the 5’UTR, a second probe was prepared using primers (p2a and p2b) in order to study its chromosomal distribution. These primers were designed to amplify a specific segment (nucleotides 4015-6254 in the sequence under accession number EU 363480) comprising endonuclease and RT coding regions of the “RaTART”. Electrophoresis and sequencing of the cloned PCR product showed that the primers amplified a DNA stretch not only with the expected size for the PCR products but also with very high sequence similarity to the corresponding stretch of the “RaTART” (SM Fig. S4). This probe hybridized to many chromosome locations and, similar to the probe 1, pericentric and euchromatic regions were predominantly stained. Weak labeling was detected at chromosome ends, but no evidence for signal enrichment at these sites was found even using low stringency hybridization and washes (SM Fig. S5).

3.3 “RaTART” 3’UTR analysis and chromosomal localization of related sequences

The unusual characteristics of the region defined as 5’UTR of the “RaTART” led us to extend the sequence analysis to its 3’UTR. The 3’UTR of the “RaTART”, as deduced by its position relative to the ORFs, displays sequences that align at the *gag* coding region of the “RaTART” (positions 6369-6957 in the sequence under

accession number EU 363480; SM Fig. S6a). Surprisingly, a stretch significantly similar to the 414 bp repeat of *R. americana* (Madalena and Gorab 2005) was also identified within its sequence, whose length (364 nucleotides) is close to the complete copy of the repeat (Fig. 1b; SM Fig. S6b). The third probe used in this work therefore consisted of the subclone obtained in a previous work (p*Ra*-38) whose insert (268 bp) contains part of the 414 bp repetitive element (Madalena and Gorab 2005). Probe hybridization showed that this repeat appears clearly enriched at five non centromeric ends and is poorly represented in pericentric regions (Fig. 2). One of the intriguing features of *R. americana* polytene chromosomes is the presence of reverse transcriptase regularly associated with its non centromeric end subset (Gorab 2003). Although its function still remains to be characterized, telomeric reverse transcriptase of this species is useful as a marker to distinguish the position of fluorescent signals in close proximity to the chromosome end. The 414 bp repeat does not reach chromosome termini as deduced by simultaneous detection of probe hybridization and telomeric reverse transcriptase (Fig. 2). Sequence analysis also showed that a stretch at the very end of the "RaTART" 3'UTR (positions 7826-8268) starts with (dT)12, reminiscent of a poly(A) tail. As observed for the 5'UTR, the 3'UTR of the retroelement containing the sequences described above is the non-coding strand of the "RaTART" in relation to the ORF sequences encoding *gag* and RT. While no significant similarity was found with nucleotide sequences deposited in data bases, conceptual translation of part of this DNA segment resulted in a polypeptide displaying significant identity to *Drosophila* reverse transcriptases (Fig. 1b; SM Fig. S6c).

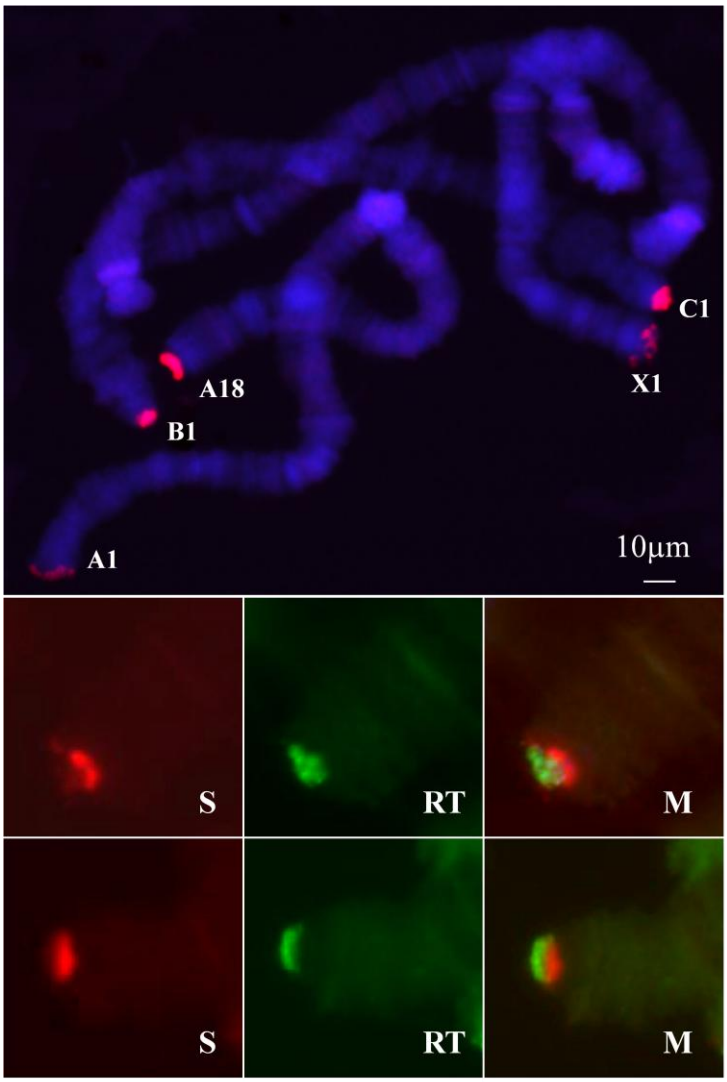


Figure 2: Localization of the p*Ra*-38 probe containing part of the 414 bp satellite in polytene chromosomes of *R. americana* counterstained with DAPI (blue) showing probe enrichment at non-centromeric ends (A1, A18, B1, C1, X1). (S, RT, M) Simultaneous detection of 414 bp satellite probe hybridization (S) and telomeric RT (RT) in polytene chromosomes of R. americana showing the merged signals (M). Chromosome ends were identified according to Madalena and Gorab (2005). Bar value except for S, RT and M.

3.4 Characterization of a subtelomeric region of *R. americana* chromosomes containing a novel complex repeat

Data on the chromosome end structure of *Rhynchosciara* that has been currently studied in the laboratory provided further details about the sequences present in the "RaTART" 3'UTR. Since the results of in situ hybridization suggested non preferential localization of the "RaTART" in the distal heterochromatin (see SM Fig. S5), a search for chromosome end sequences related to this retroelement was carried out in an attempt to assess genomic regions in which the retrotransposon could be inserted. Phages from a genomic library were isolated using a probe made of short repeats of *R. americana* specific for chromosome ends (Rossato et al. 2007; Madalena et al. 2008). Phage DNA extraction followed by labeling and in situ hybridization showed that the inserts contained sequences specific for all non centromeric ends of *R. americana* (data not shown). Given the repetitive nature of the DNA in the region covered by the genomic clone analyzed in more detail, it has not been possible to construct a complete map of the phage insert. Despite this limitation, a partial map of the insert named λ1 was made on the basis of Southern-blot hybridization results and thus allowed us to determine the approximate position of repeat types throughout its length (Fig. 3, Fig. 4a). Among the *Sal*I and *Eco*RI restriction fragments derived from λ1 DNA, two digestion products deserved particular attention. Southern-blot hybridizations showed initially that a *Sal*I restriction fragment migrating as approximately 4 Kbp did not hybridize to probes for 414 bp and M-22 repeats characterized previously (Madalena and Gorab 2005; Rossato et al. 2007; Madalena et al. 2008). However it did hybridize to probe 2 which contains the RT coding region and the 3'UTR of the "RaTART" (Fig. 3).

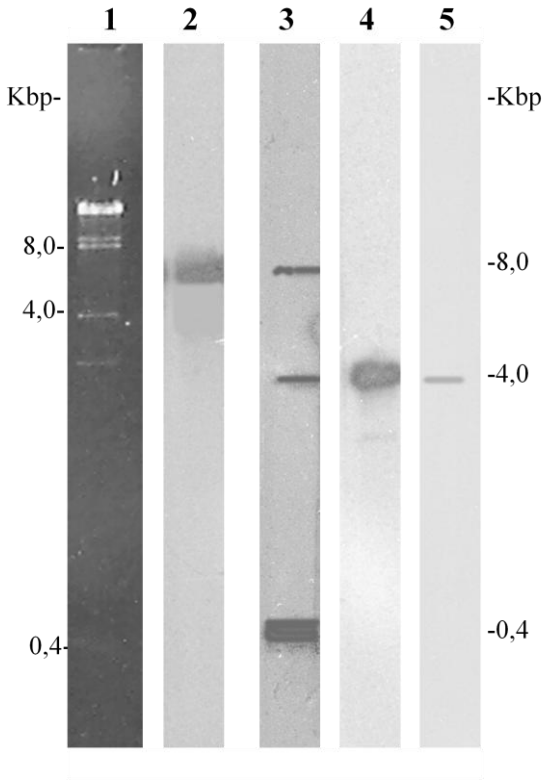


Figure 3: (1) Phage DNA named λ1 was digested to completion with *Sal*I, analyzed in agarose gel stained with ethidium bromide. Restriction fragments were subsequently transferred to membranes and hybridized to the following probes: **(2**) M-22 repeats; **(3)** 414 bp repeat; **(4)** p*Ra*1-20 insert; **(5)** probe 2, comprising endonuclease and RT coding regions of the "RaTART". Southern-blot results are presented with membrane replicas. Specific probes for M-22 and 414 bp repeats came from p*Ra*-M47.33 (Rossato et al. 2007) and p*Ra*-38 (Madalena and Gorab 2005) inserts respectively.

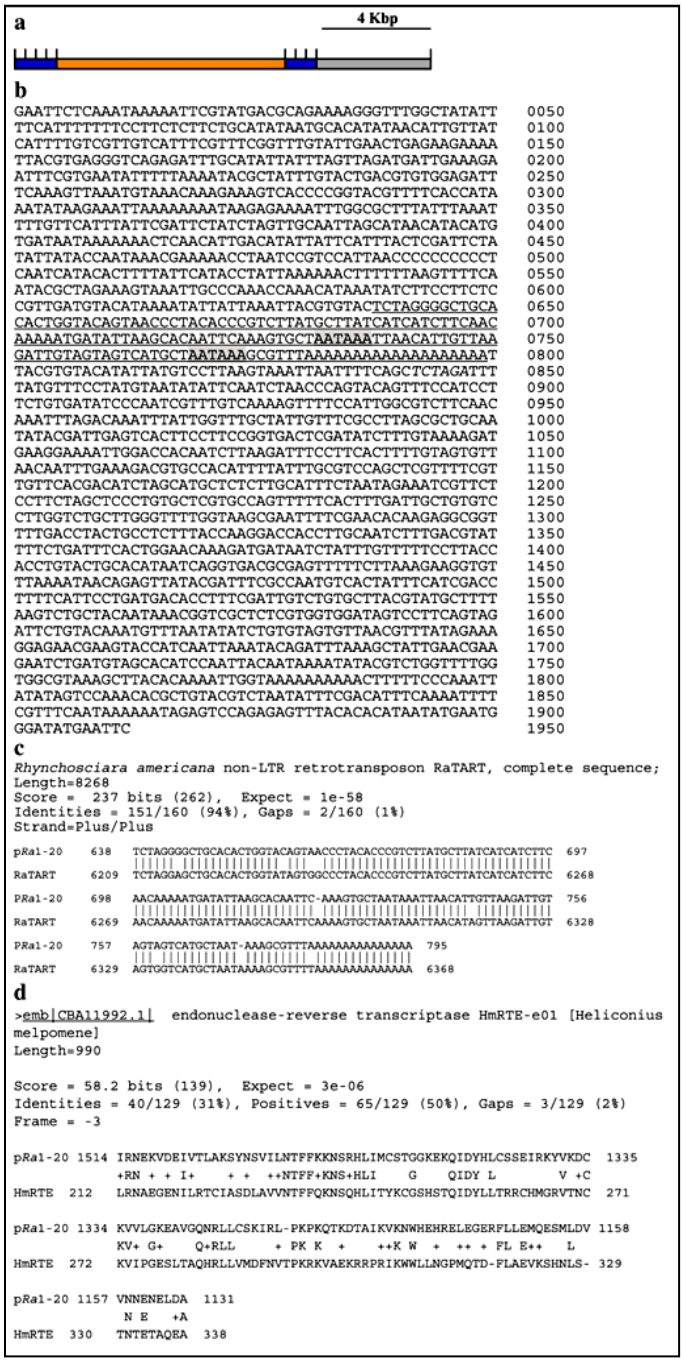


Figure 4: (**a**) Partial restriction map of the λ1 phage insert. The vertical lines indicate the position of *Sal*I restriction sites. *Blue*, *orange* and *gray* areas in the map contain respectively 414 bp satellites, M-22 repeats and the p*Ra*1-20 insert as deduced by Southern-blot hybridizations (see SM, Figure S5). (**b**) p*Ra*1-20 insert sequence. Underlined nucleotides aligned with part of the "RaTART" 3'UTR; putative polyadenylation signals are shadowed; italicized bases indicate the *Xba*I restriction site. (**c**) Sequence alignment with part of the "RaTART" 3'UTR (accession number EU 363480) and the underlined nucleotides of the p*Ra*1-20 insert sequence. (**d**) Protein alignment using conceptual translation from part (nucleotides 1131-1514) of the p*Ra*1-20 insert sequence.

3.5 p*Ra*1-20 sequence analysis and chromosomal localization

Unexpectedly, *Sal*I restriction fragments from λ1 DNA could not be cloned and the problem is still under investigation. Southern-blot hybridization using probe 2 was then performed with λ1 DNA digested with *Eco*RI (data not shown). For this reason, the subclone named p*Ra*1-20 was obtained by digesting λ1 DNA with *Eco*RI. Sequencing confirmed that the p*Ra*1-20 insert contains a short segment that was found to be significantly similar to one stretch of the "RaTART". Consensus poly(A) addition signals are present in this region which terminates in oligo(dA), strongly resembling a true 3'UTR of a non-LTR retrotransposon (Fig. 4b-c). Flanking sequences of the two aligned DNA segments display no sequence similarity. The p*Ra*1-20 insert sequence showed neither internal repetition nor significant sequence similarity with those deposited in databases. However, the conceptual translation product of part of its sequence revealed significant identity with an endonuclease (Fig. 4d) found to be associated with a reverse transcriptase of *Heliconius melpomene* (Lepidoptera: Nymphalidae). Other features of the subcloned sequence are worthy of mention. In situ hybridization signals with the p*Ra*1-20 probe were mainly present at A-18, B-1 and C-1 chromosome ends (Fig. 5). Weak labeling was also observed in a few interstitial bands and in the pericentric heterochromatin (chromosome sections A10/12, B15, C11 and X12). Simultaneous detection of p*Ra*1-20 probe hybridization and telomeric reverse transcriptase was then performed in order to visualize more precisely the position of the cloned sequence. The results showed that the insert of the p*Ra*1-20, as already observed with the 414 bp satellite, does not reach the chromosome ends of *R. americana* (Fig. 6). The occurrence

of bridges connecting non homologous polytene chromosome ends of *R. americana* was also exploited in order to give further information on the location of the p*Ra*1-20 insert at chromosome termini. The probe did not reach inter-terminal bridges while displaying strong signals close to the chromosome end, thus reinforcing the sub-terminal localization of the cloned sequence (Fig. 6).

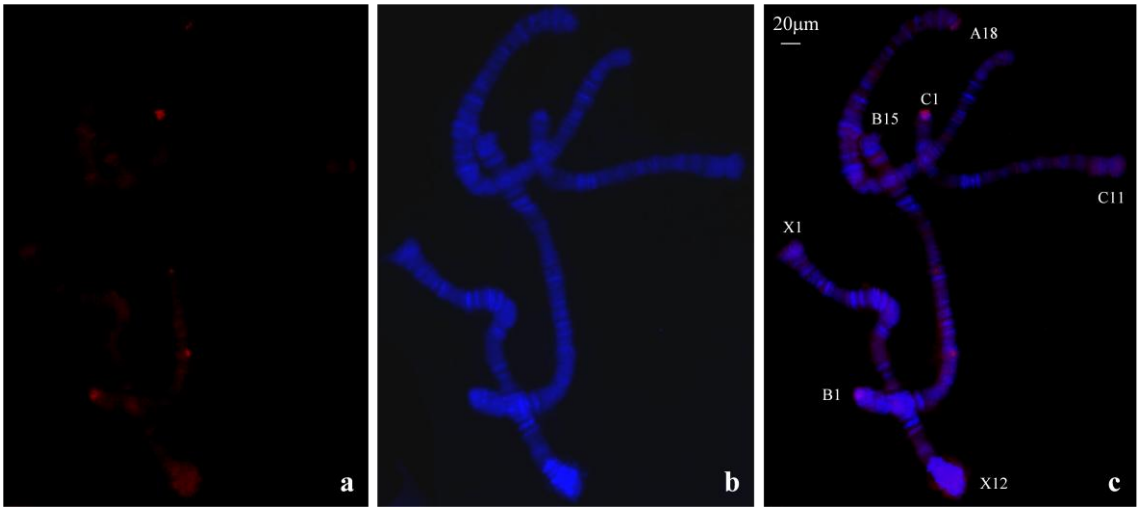


Figure 5: (a) Chromosomal localization of the p*Ra*1-20 probe hybridization showing **(b)** DAPI counterstaning and **(c)** the merged signals.

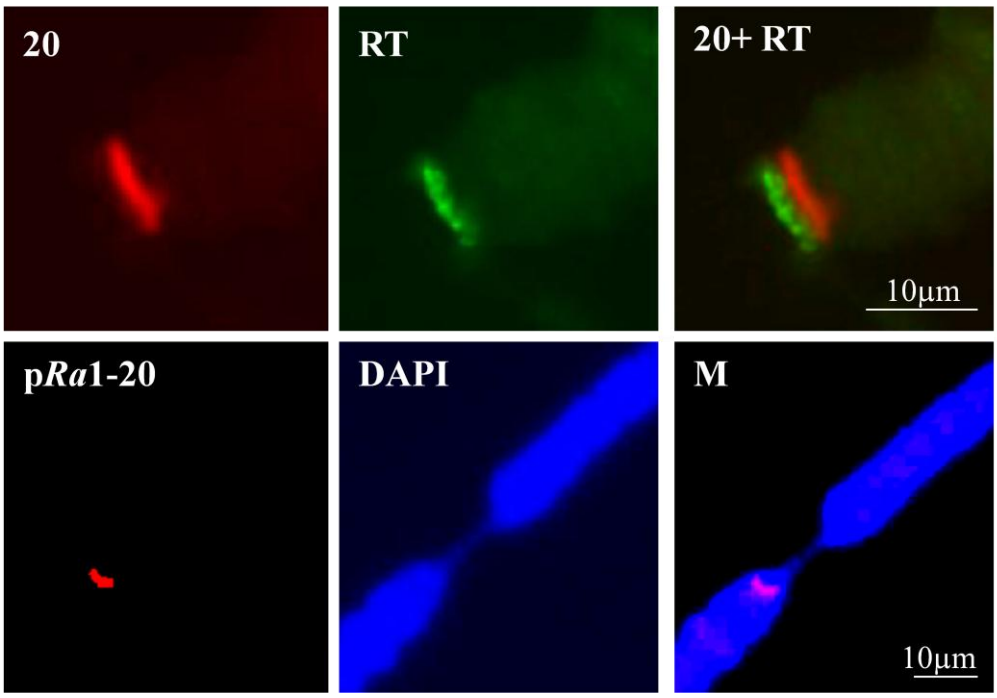


Figure 6: (*1.9, RT, M*) Simultaneous detection of the p*Ra*1-20 probe hybridization (*20*) and telomeric reverse transcriptase (*RT*) at the polytene chromosome.

3.6 Copy number of the complex repeat

The p*Ra*1-20 insert together with two reference sequences were hybridized to a probe composed of genomic DNA from the salivary gland of *R. americana*. Repeated sequences, with significantly higher representation in the genome, will be differentially labeled in this peculiar probe so that the strength of the hybridization signals observed will be related to the repetitiveness of the cloned DNA. Although it cannot accurately quantify genomic sequences, this assay is useful for giving a measure of the degree of reiteration of cloned sequences using appropriate controls. As controls, insect ribosomal

DNA (rDNA; Gorab et al. 1995) and M-22 repeats (Rossato et al. 2007; Madalena et al. 2008) were used as references for middle and highly repetitive sequences respectively. Digestion of the p*Ra*1-20 plasmid with *Eco*RI/*Xba*I allowed us to split its insert into two bands, with the fast migrating segment containing the "RaTART" stretch characterized in this work (Fig. 7a2). As expected, the strongest hybridization signal was observed in the insert composed of M-22 repeats (Fig. 7b1). Densitometric analyses performed showed that the intensity of the slow migrating band of the p*Ra*1-20 insert in the gel was approximately twice (1.96X) as intense as that observed for the rDNA insert band (Fig. 7a2-3). On the other hand, densitometry on the Southern-blot hybridization results showed that the signal at the rDNA band is nearly three times (2.82X) stronger than that of the slow migrating band of the p*Ra*1-20 insert (Fig. 7b2-3). The results indicated that the number of copies of the complex repeat is lower than that observed for the rDNA in the salivary gland of this species.

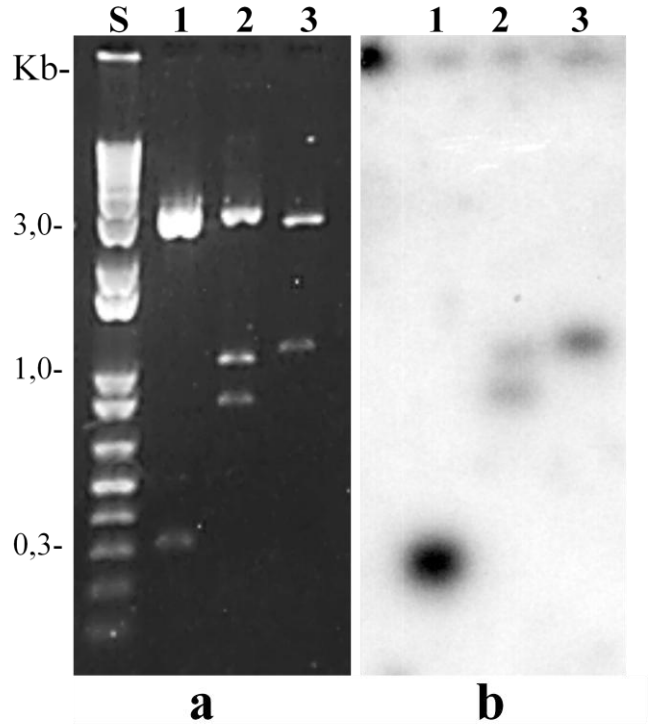


Figure 7: (a) Plasmid DNAs were cut with restriction enzymes and analyzed in an agarose gel stained with ethidium bromide. Except for p*Ra*1-20, plasmids were cut with *Eco*RI (p*Ra*M47.33) and *Eco*RI/*Hind*III (p*Ctt*26S). S: DNA size markers; **1**: p*Ra*M47.33 (the insert contains M-22 repeats; further details in Rossato et al. 2007); **2**: p*Ra*1-20 cut with *Eco*RI/*Xba*I; **3**: p*Ctt*26S (the insert contains fragment of 26S rDNA from *Chironomus*; further details in Gorab et al. 1995). **(b)** DNA from the gel was transferred to a nylon membrane and hybridized to salivary gland DNA of *R. Americana.*

4 DISCUSSION

Predicted polypeptides encoded in ORFs of *Drosophila* telomere transposons are phylogenetically related to those found in retrotransposons such as *jockey* (Priimägi et al. 1988; Villasante et al. 2007). Telomere retrotransposons are also distinguished by their specific telomeric localization and unusually long 3'UTRs (reviewed in Mason et al. 2008; Villasante et al. 2008). In addition to RT and *gag* identities shared with *jockey* and other retroelements, signatures of a telomere

retrotransposon were found in the genomic clone termed “RaTART” (Resende-Teixeira et al. 2008). "RaTART" thus represented the first evidence for telomere transposons in the Sub-Order Nematocera suggesting that mobile elements might be more broadly used to maintain telomeres in Diptera.

Except for the 414 bp repeat probe (Madalena and Gorab 2005), probes containing sequences related to both 5’UTR and RT coding regions of the “RaTART” hybridized to pericentric and euchromatic regions of *R. americana* chromosomes. In situ hybridizations suggest a broad chromosomal distribution of sequences present in the genomic clone named “RaTART”. The presence of a segment containing part of the 414 bp satellite in the “RaTART” sequence might offer a simple explanation for the significant hybridization signals at chromosome ends that were shown previously (Resende-Teixeira et al. 2008).

In a hypothetical reconstruction of events that took place in this genomic region (Fig. 1c), a possibly complete copy of a non-LTR retrotransposon (“RaTART”) was inserted into a second, partially represented retrotransposon (e1) both showing evolutionarily related ORFs as inferred by sequence comparison. The existence of part of another mobile element in the “RaTART” is also suggested (e2, Fig. 1c). These events might have occurred in the subtelomeric region as supported by the presence of part of a 414 bp subtelomeric satellite within the “RaTART” genomic clone. In connection with the latter, the short “RaTART” insertion in the site covered by the p*Ra*1-20 insert characterized in this work lies in the subtelomeric region next to 414 bp repeat arrays.

However, other results argue for the pericentric origin of this genomic clone. Probes derived from coding regions of the “RaTART” hybridized to pericentromeric regions of all chromosomes. In addition, the 414 bp repeat probe also hybridized weakly to the pericentric heterochromatin, indicating its possible presence in these sites as a dispersed sequence.

Research directed toward the characterization of sequences and proteins specific for chromosome ends of *R. americana* has been carried out in our laboratory. In the course of these experiments, we isolated part of a genomic clone (p*Ra*1-20) containing a short sequence similar to a stretch in the 3'UTR region of the "RaTART". Since flanking sequences of the two aligned DNA segments display no similarity, part of the p*Ra*1-20 insert sequence exhibited in the alignment suggests an insertion event of an extremely truncated copy of the "RaTART". Using telomeric reverse transcriptase of *R. americana* (Gorab 2003; Madalena and Gorab 2005; Rossato et al. 2007) and inter-telomeric bridges as markers for chromosome ends, the results showed that the p*Ra*1-20 insert is localized in the subtelomeric region rather than at the very end of the *Rhynchosciara* chromosomes. This finding probably represents an attempt of "RaTART" transposition to subtelomeres and, in this sense, could explain the weak hybridization signals detected in the distal heterochromatin using the "RaTART" probes synthesized in this work by PCR (see SM Fig. S5).

Data on the subtelomeric structure of *R. americana* presented in this work extend previous findings on the characterization of repeats near chromosome ends of this

species. The sequence cloned in the p*Ra*1-20 lies in the vicinity of 414 bp repeats (Madalena and Gorab 2005) and M-22 repeats (Rossato et al. 2007). It can be classified as complex repeat because of its length, its apparent lack of internal repeats and also its occurrence in distinct subtelomeres. Quantification data suggest a copy number significantly lower (approximately six times) than that observed for the rDNA. If the salivary gland of *R. americana* has one hundred copies of rDNA (Gambarini and Lara 1974), approximately sixteen copies of the complex subtelomeric repeat are likely to coexist in the salivary gland of this species.

Like "RaTART", this complex repeat also displays a composite structure. One of its components is an extremely truncated segment of the "RaTART" terminating with an oligo(dA) tail. This observation, together with sequence analyses performed in this work, reinforces the idea that this short stretch must represent the 3'end of the "RaTART". Whether this insertion type is present within all the genomic copies of this complex repeat, the answer will require further investigation. In relation to the structure of this complex subtelomeric repeat, its partial identity with the lepidopteran endonuclease is intriguing. One might suspect that this DNA segment represents a very incomplete copy of a retrotransposon since the *Heliconius* endonuclease is also a conceptual product that appeared in a sequence which also encodes a reverse transcriptase. However, protein identity is restricted to this segment within the p*Ra*1-20 sequence and *Heliconius* data on the subject are limited to one sequence deposited in the data bank. Hence, the existence of *Rhynchosciara* retrotransposons related to the mobile

elements of lepidopterans localized at chromosome ends (reviewed in Fujiwara et al. 2005) is conjectural at present.

In summary, the results of this work have shown that sequences related to ORFs present in the genomic clone named "RaTART" have a distribution similar to that described for a number of previously studied retroelements. However, given the diversity of nematocerans and the few species whose telomere structure has been analyzed, the existence of telomere transposons in lower dipterans is still an open question. On the other hand, data obtained in this study on the subtelomeres of *R. americana* show that these regions are composed of apparently dispersed complex sequences a few kilobases away from the arrays of simple and complex tandem repeats which were characterized previously. This composite structure can be seen throughout the approximately 15 Kb contained in the λ1 phage insert and also observed within the p*Ra*1-20 insert. It is possible that subtelomeric regions in the vicinity of 414 bp tandem repeats are preferential sites for the insertion of mobile elements. This assumption is supported by the short "RaTART" sequence identified in the p*Ra*1-20 insert and could be reinforced if the whole "RaTART", which also contains the 414 bp subtelomeric repeat, had a subtelomeric origin as discussed above. However, further evidence for relatives of lepidopteran retrotransposons in *R. americana* subtelomeres, as suggested by the p*Ra*1-20 insert sequence and its location, will require further experimental work.

REFERENCES

Altschul SF, Gisg W. Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410

Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J (2007) Telomeric repeat-containing RNA and RNA surveillance factors at mammalian chromosome ends. Science 318: 798-801

Biessmann H, Kobeski F, Walter MF, Kasravi A, Roth CW (1998) DNA organization and length polymorphism at the 2L telomeric region of *Anopheles gambiae*. Insect Mol Biol 7: 83-93

Bryan TM, Cech TR (1999) Telomerase and the maintenance of chromosome ends. Curr Opin Cell Biol 11: 318-324

Carmona MJ, Morcillo G, Galler R, Martinez-Salas E, de la Campa A, Díez JL, Edström JE (1985) Cloning and molecular characterization of a telomeric sequence from a temperature-induced Balbiani ring. Chromosoma 92: 108-115

Díez JL, Vilariño VR, Medina FJ, Morcillo G (2006) Nucleolar localization of reverse transcriptase related to telomere maintenance of *Chironomus* (Diptera). Histochem Cell Biol 126: 445-452

Etheridge KT, Banik SSR, Armbruster BN, Zhu Y, Terns RM, Terns MP, Counter CM (2002) The nucleolar localization domain of the catalytic subunit of human telomerase. J Biol Chem 277: 24746–24770

Fujiwara H, Osanai M, Matsumoto T, Kojima KK (2005) Telomere-specific non–LTR retrotransposons and telomere maintenance in the silkworm, *Bombyx mori*. Chromosome Res 13: 455-467

Gambarini AG, Lara FJS (1974) Under-replication of ribosomal cistrons in polytene chromosomes of *Rhynchosciara*. J Cell Biol 62: 215-222

Gorab E (2003) Reverse transcriptase-related proteins in telomeres and in certain chromosomal *loci* of *Rhynchosciara* (Diptera: Sciaridae). Chromosoma 111: 445-454

Gorab E, de Lacoba MG, Botella LM (1995) Structural constraints in expansion segments from a midge 26 S rDNA. J Mol Evol 41: 1016-1021

Gorab E, Botella LM, Quinn JP, Amabis JM, Díez JL (1996) Ku-related antigens are associated with transcriptionally active loci in *Chironomus* polytene chromosomes. Chromosoma 105: 150-157

Holt SE, Aisner DL, Baur J, Tesmer VM, Dy M, Ouellette M, Trager JB, Morin GB, Toft DO, Shay JW, Wright WE, White MA (1999) Functional requirement of p23 and hsp90 in telomerase complexes. Genes Dev 13: 817-826

López CC, Nielsen L, Edström J-E (1996) Terminal Long Tandem Repeats in Chromosomes from *Chironomus pallidivittatus*. Mol Cell Biol 16: 3285-3290

López CC, Rodriguez E, Díez JL, Edström J-E, Morcillo G (1999) Histochemical localization of reverse transcriptase in polytene chromosomes of chironomids. Chromosoma 10: 302-307

Madalena CRG, Gorab E (2005) A chromosome end satellite of *Rhynchosciara americana* (Diptera: Sciaridae) resembling nematoceran telomeric repeats. Insect Mol Biol 14: 255-262

Madalena CRG, Andrioli LPM, Gorab E (2008) Ribosomal RNA gene insertions in the R2 site of *Rhynchosciara* (Diptera: Sciaridae). Chromosome Res 16: 1233-1241

Martinez JL, Edström JE, Morcillo G, Díez JL (2001) Telomeres in *Chironomus thummi* are characterized by different subfamilies of complex DNA repeats. Chromosoma 110: 221-227

Mason JM, Frydrychova RC, Biessmann H (2008) *Drosophila* telomeres: an exception providing new insights. BioEssays 30: 25-37

Morcillo G, Díez JL, Carbajal ME, Tanguay RM (1993) HSP90 associates with specific heat shock puffs (hsrw) in polytene chromosomes of *Drosophila* and *Chironomus*. Chromosoma 102: 648-659

Priimägi AF, Mizrokhi LJ, Ilyin YV (1988) The *Drosophila* mobile element *jockey* belongs to LINEs and contains coding sequences homologous to some retroviral proteins. Gene 70: 253-262

Resende-Teixeira P, Siviero F, Brandão AS, Santelli RV, Machado-Santelli GM (2008) Molecular characterization of a retrotransposon in the *Rhynchosciara americana* genome and its association with telomere. Chromosome Res 16: 729-742

Rosén M, Edström JE (2000) DNA structures common for chironomid telomeres terminating with complex repeats. Insect Mol Biol 9: 341-347

Rossato RM, Madalena CRG, Gorab E (2007) Unusually short tandem repeats in the chromosome end structure of *Rhynchosciara* (Diptera: Nematocera). Genetica 131: 109-116

Roth CW, Kobeski F, Walter MF, Biessmann H (1997) Chromosome end elongation by recombination in the mosquito *Anopheles gambiae*. Mol Cell Biol 17: 5176-5183

Saiga H, Edström JE (1985) Long tandem arrays of complex repeat units in *Chironomus* telomeres. EMBO J 4: 799-804

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schoeftner S, Blasco MA (2008) Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. Nat Cell Biol 10: 228-236

Teixeira MT, Förstemann K, Gasser SM, Lingner J (2002) Intracellular trafficking of yeast telomerase components. EMBO Rep 3: 652–659

Villasante A, Abad JP, Planelló R, Méndez-Lago M, Celniker SE, de Pablos B (2007) *Drosophila* telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. Genome Res 17: 1909-1918

Villasante A, de Pablos B, Méndez-Lago M, Abad JP (2008) Telomere maintenance in *Drosophila*: rapid transposon evolution to chromosome ends. Cell Cycle 7: 2134-2138

Yang Y, Chen Y, Zhang C, Huang H, Weissman SM (2002) Nucleolar localization of hTERT protein is associated with telomerase function. Exp Cell Res 277: 201–209

SUPPLEMENTARY MATERIALS

SM Fig. S1: (**a**) The alignment was done between "5'UTR" and RT coding regions of the "RaTART"; (**b**) The alignment was done between "5'UTR" and ***gag*** coding regions of the "RaTART" (accession number EU 363480).

a)

```
Score =  475 bits (247),  Expect = 2e-130
Identities = 638/831 (76%), Gaps = 4/831 (0%)
Strand=Plus/Minus

5'UTR   8     ATTTCATCAAATCCTGGCGCTTTCTTGGAATTCAGGTTTTTGTCGATTTCCTGTGCCACT  67
              ||||| || | |||||| | |||||| |  || |||||||||||| ||||| ||||| |||
RT      4872  ATTTCGTCGATTCCTGGTGATTTCTTTGGGTTGAGGTTTTTGTCTATTTCTTGTGCTACT  4813

5'UTR   68    TCGTACGGGGTGATTGGTTTAATCGTTGCTTCAGGTAGGTATTCTGGAGCCGTGCTGATG  127
              ||||| ||||||| ||| ||||||||| ||||||||  ||| |||  || |    | ||
RT      4812  TCGTAGGGGGTGAATGGCTTAATCGTTTCTTCAGGTGAGTAATCTAAAGTCACTTTAATA  4753

5'UTR   128   TTCGAGTTTATTTCATGTGGTTGAAATACACTCGAAAGATACTGAGCAAAGAGTTCAGCT  187
              ||||| |||||||||||||||||||||||   ||| || | ||| ||||| |||||||||
RT      4752  TTCGAATTTATTTCATGTGGTTGAAATACTTGCGATAGGTGCTGGGCAAAAAGTTCAGCT  4693

5'UTR   188   TTTTCTTTATCCTTTCTAATCCATTGGCCTTGTGCATTTTTGATTGGCGGCACTTGGACT  247
              ||||||| |||||| || ||||| ||   ||||| ||||||||||||||| ||||||| |
RT      4692  TTTTCTTGATCCTTCCTGATCCACTGATGTTGTGAATTTTTGATTGGCGGTACTTGGATT  4633

5'UTR   248   GTTGGTCTTTTAAGTCGTCTTGTGTCGTCTTGTCCATAGTGAATATTCTTTTTCGGCTCC  307
               |||||| ||| |  ||||||||    |  | |||||||||||| || || ||  ||||
RT      4632  ATTGGTCGTTTGAACCGTCTTGT----TGCTTTCCATAGTGAATAGTCCTTCTCTCCTCC  4577

5'UTR   308   TGGTCCCAATCCGTTGAGGTATTCACTAATGCAGCTCTGCTTGAATTCTTTGATAAGTCG  367
              |   |||| |||||| |||||||||| |||||||||||| |||||||| || |||  | |
RT      4576  TCCCCCCAGTCCGTTCAGGTATTCACCAATGCAGCTCTGTTTGAATTCCTTAATAGCTTG  4517

5'UTR   368   GTTAACACGATTGCTCACCTCATTGAAGTTCTTTTTATCTGAAGGATGTCTGGATCCGTG  427
              | ||||| |||| ||||||| ||| || ||| ||||||||   || ||||| ||||||||
RT      4516  GCTAACAAGATTACTCACCTGATTAAATTTCCTTTTATCTTCTGGGTGTCTTGATCCGTG  4457

5'UTR   428   CCATATTCGTCTTAATTTGCGCCGTTCCTTGATTAGCTCTCTAATTTCAGTTGGGTAATT  487
              |||| |||| ||   ||| |||| ||| ||||||| |||||| ||||| | ||| |||||
RT      4456  CCATGTTCGCCTCCGTTTACGCCTTTCTTTGATTAACTCTCTGATTTCGGCTGGATAATT  4397

5'UTR   488   GGTTTCGTATTCTGGTTTGTCACATGGAGCTGGAGTTGCCTTTTTGGCAGCCATTTGTAT  547
              | |||| | ||||||||| || | ||| |||||||| || | | | |||||||  || |
RT      4396  GATTTCTTGTTCTGGTTTATCCCCTGGTGCTGGAGTCGCTTCTCTTGCAGCCAGATGAAG  4337

5'UTR   548   CCAGTTGATAAATATTTGAGCAGCACTCTCTAGCTCATCTTTGTCTCTCAGTTTAACGTT  607
                  ||||| |   ||||   ||||  ||| |  || || |||| ||| | |||||||||
RT      4336  TTGATTGATGAGGTTTTGGAGAGCAAACTCCAAGTCTTCCTTGTTTCTTAATTTAACGTT  4277

5'UTR   608   CAGATCAATTAGTTCATCCAAATATTCTCTGAAAGTTTCCCAGTTTGTACGTTTGGTTGT  667
               || |||||    |  ||||  | ||| ||||| || ||||| || ||  |||| |||||
RT      4276  TAGTTCAATCGAATTGTCCATTTTTTCCCTGAATGTCTCCCAATTCGTGAGTTTAGTTGT  4217

5'UTR   668   GAGATTTTGTTTCTTCACTTTTTTGATGACTTTACTACTAATAATCAGGAGTACTGGGGT  727
              ||||||||| || || | |||||||||||||  |  | || | |   ||||| || || ||
RT      4216  GAGATTTTGCTTTTTTATTTTTTTGATGACTGTTGTGCTGAGAGAGAGGAGAACCGGAGT  4157
```

```
5'UTR   728   GTGATCTGACGATAGATCTTCTATTCCCTCTGTTGAGCAGCAATTATTATTGATTCCTTT  787
               |||||||| |||||||| |||||||| |||||      ||||||    || ||||||||
RT      4156  ATGATCTGATGATAGATCGTCTATTCCTTCTGTGACAGCGCAATTCGGGTTTATTCCTTT  4097

5'UTR   788   AGAGATGAAGAAGTCGATCAAATCCGGAACTTTTTTGTTGTCAGTTGGCCA  838
               || |||||||||| |||| |||||||| ||||||||||||||||||||
RT      4096  GGAAATGAAGAAGTCAATCAGATCCGGAATTTTTTTGTTGTCAGTTGGCCA  4046
```

b)

```
Score =  114 bits (59),  Expect = 2e-21
Identities = 142/181 (78%), Gaps = 3/181 (1%)
Strand=Plus/Minus

5'UTR  1543  TGCTTATCGTTTTTCTTCATATCCTGTAGTCGGGAAGTGACTGGAGATTTTTTGGCTTGT  1602
             ||||||||||||||| ||   |||||||||||||| ||||| |||| |||| | | |||
gag    3250  TGCTTATCGTTTTTCATC---TCCTGTAGTCGGGATGTGACCGGAGCTTTTGTTGTTTGA  3194

5'UTR  1603  AGCGGCTTCAGCATTTTTTGATATTCCAAACAACCTTTGTAGTTAGCTGGGTGTCCGAAT  1662
             | |||||||||||||||||| |||||   ||| |||||||||||| ||||| |||||   |
gag    3193  AACGGCTTCAGCATTTTTTGGTATTCTGTACAGCCTTTGTAGTTCGCTGGATGTCCAGTT  3134

5'UTR  1663  TGATCACATAACACGCACTTAACAGGGCTGTTTGGCGTTTTTGTGCAATCATTTGTCAGG  1722
             || |  |||| |||| ||||  | ||||| ||  | |||||||| |||   ||||| |||
gag    3133  TGTTTGCATAGCACGTACTTGGCGGGGCTATTAAGTGTTTTTGTACAAGATTTTGTTAGG  3074

5'UTR  1723  T  1723
             |
gag    3073  T  3073
```

SM Fig. S2: (a) Sequence obtained by genomic PCR (probe 1) using primers (*underlined*) designed from sequences within the "5'UTR" of the "RaTART". **(b)** The alignment was done between probe 1 sequence and the corresponding stretch in the "5'UTR" of the "RaTART".

a)

```
GATATTGCACTCATTTCGGAGACGCATTTCTCGAACAAGAACTTCATTAA  0050
ATTCAAAGGCTACTGCACCTATTGGACTACCCACCCGAGTGAGCGAGCCC  0100
GCGGAGGTACATCAATACTGATCAAACAAAATATTGAACATCACCTGCAA  0150
GACGACATACGGGAAACATATATTCAGGCTACAATTGTTTGCATTCAACA  0200
TAGTGGGACTGAATTAAATGTTGGGGCAGTATACTGCCCTCCTAGACACA  0250
CAATAACTAAAACACAATATCTTAGTATTTTCAACAAATTGGGATCATCC  0300
TGGGGGGAGACTTCAACGTAAAACACACAGCTTGGGGATCAAGACTGATT  0350
ACTCCAAATAAAGGCAGCCAGTTATTATCGGCCATTAATGAGGAAAAATG  0400
CGGATTCCACTCATCAGGAAAGCCTACGTACTGGCCAACTGACAACAAAA  0450
AAGTTTCGGATTTGATCGACTTCTTCATCTCTAAAGGAATCAATAATAAT  0500
TGCTGCTCACAGAGGGAATAGAAGATCTATCGTCAGATCATACCCCAGTA  0550
CTCCTGATTATTACTAGTGAAGTCATCAAAAAAGTGAAGAAACAAAATCT  0600
CACAACCAAACGTACCAACTGGGAAACTTTCAAGAATATCTGGATGAACT  0650
AATTGATCTGACGTTAACTGAGAGACAAAGATGAGCTAGAGAGTGCTGCT  0700
```

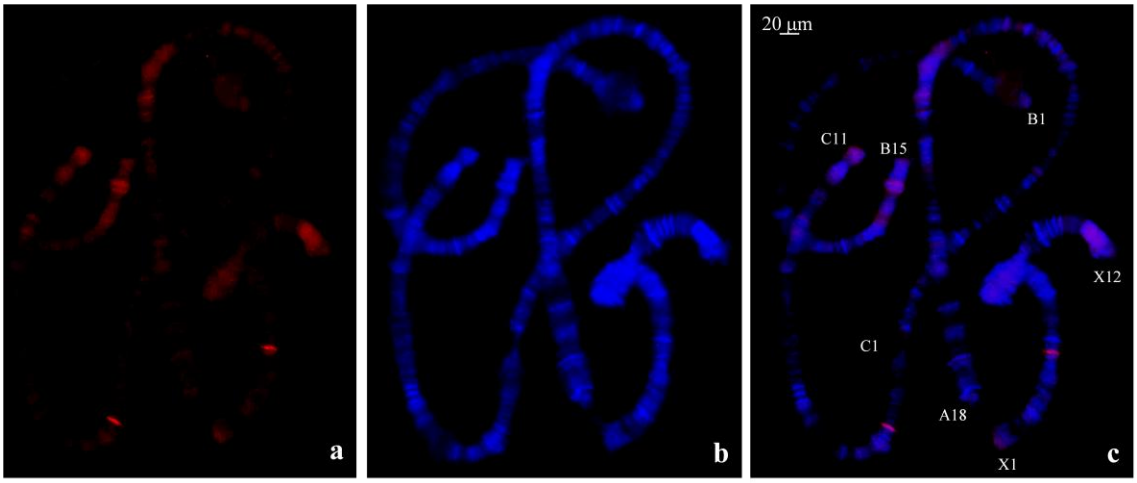
```
CAATATTTATCAACTGGATACAAATGGCTGCAAAAAAGGCAACTCCAGCT  0750
CCATGTGACAAACCAGAATACGAAACCAATTACCCAACTGAAATTAGAGA  0800
GCTAATCAAGGAACGGCGCAAATTATGACGAATATGGCACGGATCCAGAC  0850
ATCCTTCAGATAAAAGGAACTTCAATGTGGTGAGCAATCGTGTTAACTGA  0900
CTTATCAAAGAATTCAAGCAGAGCTGCATTAGTGAATACCTCAACGGATT  0950
GGGACCAGGAGCCGAAAAAGAATATTCTCTATGGAAAGCCACAAGACGAC  1000
TTAAAAGACCAACAGTCCAAGTGCCGCCAATCAAAATATGCACAAGGCCA  1050
ATGGATTAGAAAGGATAAAGAAAAAGCTGAACTCTTTGCCCAGTATCTTT  1100
CGAGTGTATTTCAACCACATGAAATAAACTCGAACATCAGCACGGCTCCA  1150
GAATACCT
```

b)

```
Rhynchosciara americana non-LTR retrotransposon RaTART, complete
sequence  Length=8268
Score = 1509 bits (817),  Expect = 0.0
Identities = 907/947 (95%), Gaps = 19/947 (2%)
Strand=Plus/Minus


Probe 1  229   GTATACTGCCCTCCTAGACACACAATAACTAAAACACAATATCTTAG-TATTTTCAACAA  287
               |||||||| |||||||||| |||||||||||||||||||||||| | ||||||||||||
RaTART   1048  GTATACTGTCCTCCTAGACGCACAATAACTAAAACACAATATCTT-GATATTTTCAACAA  990

Probe 1  288   A-TTGG-----GA-TCATCCTGGGGGGGAGACTTCAACGTAAAACACACAGCTTGGGGATC  340
               | ||||      | |||||||||||| |||||||||||||||||||||||||||||| |||
RaTART   989   ATTTGGATCAAAATTCATCCTGGGGGCAGACTTCAACGTAAAACACACAGCTTGGGAATC  930

Probe 1  341   AAGACTGATTACTCCAAATAAAGGCAGCCAGTTATTATCGGCCATTAATGAGGAAAAATG  400
               |||||||||||||||||||||||||||| || |||||||||||||||||||||| ||||||
RaTART   929   AAGACTGATTACTCCAAATAAAGGCAGTCAATTATTATCGGCCATTAATGAGGGAAAATG  870

Probe 1  401   CGGATTCCACTCATCAGGAAAGCCTACGTACTGGCCAACTGACAACAAAAAAGTTTCGGA  460
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
RaTART   869   CGGATTCCACTCATCAGGAAAGCCTACGTACTGGCCAACTGACAACAAAAAAGTTCCGGA  810

Probe 1  461   TTTGATCGACTTCTTCATCTCTAAAGGAATCAATAATAATTGCTGCTC-ACAGAGGGAAT  519
               |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
RaTART   809   TTTGATCGACTTCTTCATCTCTAAAGGAATCAATAATAATTGCTGCTCAACAGAGGGAAT  750

Probe 1  520   AGAAGATCTATCGTCAGATCATACCCCAGTACTCCTGATTATTACTAGTGAAGTCATCAA  579
               |||||||||||||||||||| |||||||||||||||||||||| |||| ||||||||||
RaTART   749   AGAAGATCTATCGTCAGATCACACCCCAGTACTCCTGATTATTAGTAGTAAAGTCATCAA  690

Probe 1  580   AAAAGTGAAGAAACAAAATCTCACAACCAAACGTACCAACTGGGAAACTTTCA-AGAATA  638
               |||||||||||||||||||||||||||||||||||| ||||||||||||||| ||||||
RaTART   689   AAAAGTGAAGAAACAAAATCTCACAACCAAACGTACAAACTGGGAAACTTTCAGAGAATA  630

Probe 1  639   TCTGGATGAACTAATTGATCTG-ACGTT-AACTGAGAGACAAAGATGAGCTAGAGAGTGC  696
               | |||||||||||||||||||| ||||| ||||||||||||||||||||||||||||||
RaTART   629   TTTGGATGAACTAATTGATCTGAACGTTAAACTGAGAGACAAAGATGAGCTAGAGAGTGC  570
```

```
Probe 1  697   TGCTC-AATATTTATCAACTGGATACAAATGGCTGCAAAAAAGGCAACTCCAGCTCCATG  755
               ||||| |||||||||||||||||||||||||||| ||||||||||||||||||||||||
RaTART   569   TGCTCAAATATTTATCAACTGGATACAAATGGCTGCCAAAAAGGCAACTCCAGCTCCATG  510

Probe 1  756   TGACAAACCAGAATACGAAACCAATTACCCAACTGAAATTAGAGAGCTAATCAAGGAACG  815
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART   509   TGACAAACCAGAATACGAAACCAATTACCCAACTGAAATTAGAGAGCTAATCAAGGAACG  450

Probe 1  816   GCGCAAATTATGACGAATATGGCACGGATCCAGACATCCTTCAGATAAAAGGAACTTCAA  875
               |||||||||| ||||||||||||||||||||||||||||||||||||| |||||||||
RaTART   449   GCGCAAATTAAGACGAATATGGCACGGATCCAGACATCCTTCAGATAAAAAGAACTTCAA  390

Probe 1  876   TGTGGTGAGCAATCGTGTTAACTGACTTATCAAAGAATTCAAGCAGAGCTGCATTAGTGA  935
               || ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
RaTART   389   TGAGGTGAGCAATCGTGTTAACCGACTTATCAAAGAATTCAAGCAGAGCTGCATTAGTGA  330

Probe 1  936   ATACCTCAACGGATTGGGACCAGGAGCCGAAAAAGAATATTCTCTATGGA-AAG-C--CA  991
               |||||||||||||||||||||||||||||||||||||||||||| ||||||| ||| |  ||
RaTART   329   ATACCTCAACGGATTGGGACCAGGAGCCGAAAAAGAATATTCACTATGGACAAGACGACA  270

Probe 1  992   CAAGACGACTTAAAAGACCAACAGTCCAAGTGCCGCCAATCAAAATATGCACAAGGCCAA  1051
               |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
RaTART   269   CAAGACGACTTAAAAGACCAACAGTCCAAGTGCCGCCAATCAAAA-ATGCACAAGGCCAA  211

Probe 1  1052  TGGATTAGAAAGGATAAAGAAAAAGCTGAACTCTTTGCCCAGTATCTTTCGAGTGTATTT  1111
               |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
RaTART   210   TGGATTAGAAAGGATAAAGAAAAAGCTGAACTCTTTGCTCAGTATCTTTCGAGTGTATTT  151

Probe 1  1112  CAACCACATGAAATAAACTCGAACATCAGCACGGCTCCAGAATACCT  1158
               |||||||||||||||||||||||||||||||||||||||||||||||
RaTART   150   CAACCACATGAAATAAACTCGAACATCAGCACGGCTCCAGAATACCT  104

 Score =  110 bits (59),  Expect = 2e-20
 Identities = 66/69 (95%), Gaps = 1/69 (1%)
 Strand=Plus/Minus

Probe 1  1     GATATTGCACTCATTTCGGAGACGCATTTCTCGAACAAGAACTTCATTAAATTCAAAGGC  60
               |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
RaTART   1203  GATATTGCACTCATTTCGGAGACGCATTTCTCGAACAAGAACTTCATTAAATTTAAAGGC  1144

Probe 1  61    TACTGCACC  69
                || |||||
RaTART   1143  -ACGGCACC  1136
```

SM Fig. S3: (a) Localization of probe 1 (comprising the 5'UTR of the "RaTART") in polytene chromosomes of *R. americana*. **(b)** chromosome counterstaining with 4',6-diamidino-2-phenylindo-l e and **(c)** the two merged signals.



SM Fig. S4 (a) Sequences obtained by genomic PCR (probe 2) using primers (*underlined*) designed from sequences comprising endonuclease and RT coding regions of the "RaTART". **(b)** The alignments were done between probe 2 sequences and the corresponding stretches of the "RaTART".

a)

```
CGGATTCCACTCATCGCGCAAGCCAACTCATTGGCCAACTGACAACAAAA    50
AAATTCCGGATCTGATTGACTTCTTCATTTCCAAAGGAATAAACCCGAAT   100
TGCGCTGTCACAGAAGGAATAGACGATCTATCATCAGATCATACTCCGGT   150
TCTCCTCTCTCTCAGCACAACAGTCATCAAAAAAATAAAAAAGCAAAATC   200
TCACAACTAAACTCACGAATTGGGAGACATTCAGGGAAAAAATGGACAAT   250
TCGATTGAACTAAACGTTAAATTAAGAAACAAGGAAGACTTGGAATTTGC   300
TCTCCAAAACCTCATCAATCAACTTCATCTGGCTGCAAGAGAAGCGACTC   350
CAGCACCAGGGGATAAACCAGAACAAGAAATCAATTATCCAGCCGAAATC   400
AGAGAGTTAATCAAAGAAAGGCGTAAACGGAGGCGAACATGGCACGGATC   450
AAGACACCCAGAAGATAAAAGGAAATTTAATCAGGTGAGTAATCTTGTTA   500
GCCAAGCTATTAAGGAATTCAAACAGAGCTGCATTGGTGAATACCTGAAC   550
GGACTGGGGGGAGGAGCAGAGAAGGACTATTCACTATGGAAAGCAACAAG   600
ACGGTTCAACGACCAATAATCCAGTACGCCATCAAAATTCCAAGATCAGT   650
GGATCAGGAGGATCAGAAAAGCTGA                            700

CAATCGAGGCATCGCAAGGCACGGCAGTTGAAGGACCACGGAAATGGTGC    50
AACGAACAATAGACAAAGTTTACCAATGGAGCTCGGACTGGGAGATCAAA   100
```

```
TCAATGATAAGCAAAATGCACTCATGGCACATTTGCAATGCGAAAAACAT   150
CGAGGCAATCATGAGGGTTATCTCAACGGAACTCCTGTTCCGCAAGCAGA   200
TACAGTTAAACTTCTTGGCTTGCATCTTGACAAGAGGCCAAGCTGGAAAC   250
ACCACGTTCAACAAAAAGTCAAACAAATAAAGTTCAAAACACGCAAGAAT   300
GGATGGTTGGACACCAATCTCAACTAAGCCTGTATAACAAAAAGCTGATT   350
TATCAAACCATCTTTAAACCGATCTGGACGTACGGAATCCAACTCTGGGG   400
ATGTACTAAAAAATCCAATCGTGAAATCATCCAAAGATGTCAAAACAAGT   450
TTCTGCGAATGATCACAAACGCCTATCGATTTGAGACCAACGAAGAAATC   500
CACAACGATCTCAACATCAAACAAATTTCCGAGGTTATCCAGGAATACGC   550
CATCAAACATGAAAAAAGGCTACTGCACCATACCAACATAGAAGCCATCC   600
GACTCTTGGATATCACCTACGAATTGAGAAGGCTGAAACGAACTAAGCCT   650
CATGAGCTAACGTGTTAAATTAATAAACTCAACGAACACTAATCATTCTT   700
ATCTCGTTCTAGGGGCTGCACACTGGTACAGTAACCCTACACCCGTCTTA   750
TGC                                                  800
```

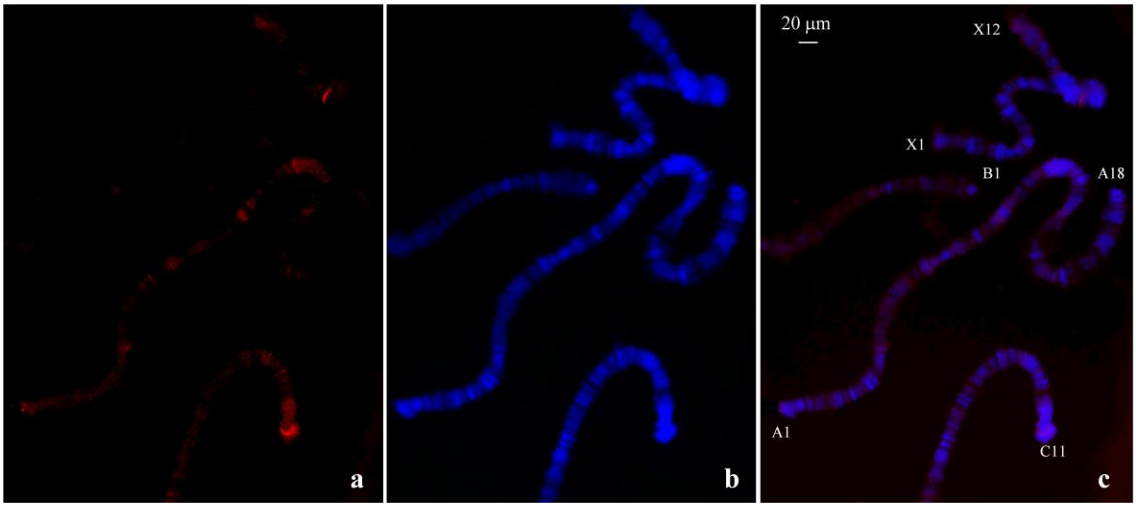
b)

```
Rhynchosciara americana non-LTR retrotransposon RaTART, complete sequence
Length=8268
Score = 1188 bits (643),  Expect = 0.0
Identities = 672/684 (98%), Gaps = 9/684 (1%)
Strand=Plus/Plus

Probe 2  1     CGGATTCCACTCATCGCGCAAGCCAACTCATTGGCCAACTGACAACaaaaaaaTTCCGGA  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART   4015  CGGATTCCACTCATCGCGCAAGCCAACTCATTGGCCAACTGACAACAAAAAAATTCCGGA  4074

Probe 2  61    TCTGATTGACTTCTTCATTTCCAAAGGAATAAACCCGAATTGCGCTGTCACAGAAGGAAT  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART   4075  TCTGATTGACTTCTTCATTTCCAAAGGAATAAACCCGAATTGCGCTGTCACAGAAGGAAT  4134

Probe 2  121   AGACGATCTATCATCAGATCATACTCCGGTTCTCCTCTCTCTCAGCACAACAGTCATCaa  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART   4135  AGACGATCTATCATCAGATCATACTCCGGTTCTCCTCTCTCTCAGCACAACAGTCATCAA  4194

Probe 2  181   aaaaataaaaaagcaaaaTCTCACAACTAAACTCACGAATTGGGAGACATTCAGGGAAAA  240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART  4195   AAAAATAAAAAAGCAAAATCTCACAACTAAACTCACGAATTGGGAGACATTCAGGGAAAA  4254

Probe 2  241   AATGGACAATTCGATTGAACTAAACGTTAAATTAAGAAACAAGGAAGACTTGGAATTTGC  300
               |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
RaTART   4255  AATGGACAATTCGATTGAACTAAACGTTAAATTAAGAAACAAGGAAGACTTGGAGTTTGC  4314

Probe 2  301   TCTCCAAAACCTCATCAATCAACTTCATCTGGCTGCAAGAGAAGCGACTCCAGCACCAGG  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART   4315  TCTCCAAAACCTCATCAATCAACTTCATCTGGCTGCAAGAGAAGCGACTCCAGCACCAGG  4374
```

```
Probe 2  361   GGATAAACCAGAACAAGAAATCAATTATCCAGCCGAAATCAGAGAGTTAATCAAAGAAAG  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART  4375   GGATAAACCAGAACAAGAAATCAATTATCCAGCCGAAATCAGAGAGTTAATCAAAGAAAG  4434

Probe 2  421   GCGTAAACGGAGGCGAACATGGCACGGATCAAGACACCCAGAAGATAAAAGGAAATTTAA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART   4435  GCGTAAACGGAGGCGAACATGGCACGGATCAAGACACCCAGAAGATAAAAGGAAATTTAA  4494

Probe 2  481   TCAGGTGAGTAATCTTGTTAGCCAAGCTATTAAGGAATTCAAACAGAGCTGCATTGGTGA  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART   4495  TCAGGTGAGTAATCTTGTTAGCCAAGCTATTAAGGAATTCAAACAGAGCTGCATTGGTGA  4554

Probe 2  541   ATACCTGAACGGACTGGGGGGGAGGAGCAGAGAAGGACTATTCACTATGGAAAGCAACAAG  600
               |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
RaTART   4555  ATACCTGAACGGACTGGGGGGGAGGAGGAGAGAAGGACTATTCACTATGGAAAGCAACAAG  4614

Probe 2  601   ACGGTTCAA-CGACCAATAATCCA-GTAC-GCCA-TCAAAA-TTC-CAAGATCAGTGGAT  654
               ||||||||| |||||||||||||| |||| |||| |||||| ||| ||| ||||||||||
RaTART   4615  ACGGTTCAAACGACCAATAATCCAAGTACCGCCAATCAAAAATTCACAACATCAGTGGAT  4674

Probe 2  655   CAGGA-GGATCA-GAAAA-GCTGA  675
               ||||| |||||| ||||| |||||
RaTART   4675  CAGGAAGGATCAAGAAAAAGCTGA  4698


Rhynchosciara americana non-LTR retrotransposon RaTART, complete sequence
Length=8268
Score = 1142 bits (1266),  Expect = 0.0
Identities = 711/758 (93%), Gaps = 10/758 (1%)
Strand=Plus/Plus

Probe 2  1     CAATCGAGGCATCGCAAGGCACGGCAGTTGAAGGACCACGGAAATGGTGCAACGAACAAT  60
               |||||  |||| |   || ||||  | |||| |  |||| |||||  | ||||||||| |
RaTART   5502  CAATCATGGCAACTGCAGACACGCAATTTGAGG--CCACAGAAATACTACAACGAACAGT  5559

Probe 2  61    AGACAAAGTTTACCAATGGAGCTCGGACTGGGAGATCAAA-TCAATGATAAGCAAAATGC  119
               |||||||||||||||||||| | | ||||||| |||||||| |||||||||  | |||| |
RaTART   5560  AGACAAAGTTTACCAATGGACCACTGACTGGAAGATCAAAATCAATGATA--CCAAATCC  5617

Probe 2  120   ACTCATGGCACATTTGCAATGCGAAAAACATCGAGGCAATCATGAGGGTTATCTCAACGG  179
               ||||||| |||||||||| |||||||||||||||| ||||||| ||| ||||||||||||
RaTART   5618  ACTCATGTCACATTTGCATTGCGAAAAACATCGAG-CAATCATAAGGTTTATCTCAACGG  5676

Probe 2  180   AACTCCTGTTCCGCAAGCAGATACAGTTAAACTTCTTGGCTTGCATCTTGACAAGAGGCC  239
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART   5677  AACTCCTGTTCCGCAAGCAGATACAGTTAAACTTCTTGGCTTGCATCTTGACAAGAGGCT  5736

Probe 2  240   AAGCTGGAAACACCACGTTCAACAAAAAGTCAAACAAATAAAGTTCAAAACACGCAAGAA  299
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
RaTART   5737  AAGCTGGAAACACCACGTTCAACAAAAAGTCAAACAAATAAAGTTCAAAACACGCAAAAT  5796
```

```
Probe 2  300   ----TGGATGGTTGGACACCAATCTCAACTAAGCCTGTATAACAAAAAGCTGATTTATCA  355
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART   5797  GTATTGGATGGTTGGACACCAATCTCAACTAAGCCTGTATAACAAAAAGCTGATTTATCA  5856

Probe 2  356   AACCATCTTTAAACCGATCTGGACGTACGGAATCCAACTCTGGGGATGTACTAAAAAATC  415
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART   5857  AACCATCTTTAAACCGATCTGGACGTACGGAATCCAACTCTGGGGATGTACTAAAAAATC  5916

Probe 2  416   CAATCGTGAAATCATCCAAAGATGTCAAAACAAGTTTCTGCGAATGATCACAAACGCCTA  475
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART   5917  CAATCGTGAAATCATCCAAAGATGTCAAAACAAGTTTCTGCGAATGATCACAAACGCCTA  5976

Probe 2  476   TCGATTTGAGACCAACGAAGAAATCCACAACGATCTCAACATCAAACAAATTTCCGAGGT  535
               || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART   5977  TCAATTTGAGACCAACGAAGAAATCCACAACGATCTCAACATCAAACAAATTTCCGAGGT  6036

Probe 2  536   TATCCAGGAATACGCCATCAAACATGAAAAAAGGCTACTGCACCATACCAACATAGAAGC  595
               |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
RaTART   6037  TATCCAGGAATACGCCATCAAACATGAAAAAAGGCTACTACACCATACCAACATAGAAGC  6096

Probe 2  596   CATCCGACTCTTGGATATCACCTACGAATTGAGAAGGCTGAAACGAACTAAGCCTCATGA  655
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART   6097  CATCCGACTCTTGGATATCACCTACGAATTGAGAAGGCTGAAACGAACTAAGCCTCATGA  6156

Probe 2  656   GCTAACGTGTTAAATTAATAAACTCAACGAACACTAATCATTCTTATCTCGTTCTAGGGG  715
               |||||||||||||||||||||||| |||||||||||||||| |||||||||||| |
RaTART   6157  GCTAACGTGTTAAATTAATAAACTCAGCGAACACTAATCATTCTTGTCTCGTTCTAGGAG  6216

Probe 2  716   CTGCACACTGGTACAGTAACCCTACACCCGTCTTATGC  753
               |||||||||||||| |||  ||||||||||||||||||
RaTART   6217  CTGCACACTGGTATAGTGGCCCTACACCCGTCTTATGC  6254
```

SM Fig. S5 (**a**) Localization of probe for the endonuclease and RT coding regions of the "RaTART" (probe 2) in polytene chromosomes of *R. americana* showing DAPI counterstaining (**b**) and the merged signals (**c**).



SM Fig. S6 (**a**) Sequence alignment using *gag* and part of the 3'UTR of the "RaTART" (accession number EU 363480); (**b**) Sequence alignment using part of the 3'UTR of the "RaTART" and the subtelomeric satellite (*414*) of *R. americana* (accession number AY 728357); (**c**) Protein alignment using part of the 3'UTR of the "RaTART" (position 7826-8268 in the sequence under ac.

a)

```
Score =  387 bits (201),  Expect = 1e-103
Identities = 467/595 (78%), Gaps = 18/595 (3%)
Strand=Plus/Minus

gag    2499  TTGCCACCTCCCACTTTTGTTAGCAATGTTAAGGACTTTAGTCTTTTCCAAAAGGAAATA  2558
             ||||||||||||||||| || || ||  | ||||| |||   |  || ||||||   ||
3'UTR  6957  TTGCCACCTCCCACTTTCGTGAGTAACATCAAGGATTTTTTCCCGTTTCAAAAGAGGATT  6898

gag    2559  TTAAACAAAGCACCAAACGCTCAGTTCAAAGCTCTTTCCAATACAGATATAAAAATAACA  2618
              | ||||||||||| |  || ||||||||||||||||| || ||||| || ||| | ||
3'UTR  6897  CTGAACAAAGCACCGAGTGCACAGTTCAAAGCTCTTTCAAACACAGACATCAAAGTGACT  6838

gag    2619  GTAAAAACTGAAGAAGAATATAGGAATCTGAAAGAGTTGCTGAAGAGAATGAAAACGGAT  2678
             ||  |||||||| |||||||||| |  |||||||  |||||||||| || |||||| ||
3'UTR  6837  GTTCAAACTGAAAAAGAATATAGGGACATGAAAGAACTGCTGAAGAGTATCAAAACGAAT  6778
```

```
gag     2679  AA------------AAATGAAATCCAATATCACACTCATCAGCTCAAAAGCGAAAAATTA  2726
               |            |||  |||||||||  |||||||||||||| |||||||| | ||||
3'UTR   6777  GATGGAGAGCTAGAAAACAAAATCCAATTCCACACTCATCAGCTGAAAAGCGATAGATTA  6718

gag     2727  TTCAGGGTAGTACTCAGAGGACTACCTTATTCAATCCCACATAACGATATAAAAACTGCC  2786
              || ||||||||| |||||||||||||||| ||||||| || ||||| || ||||||||
3'UTR   6717  TTTAGGGTAGTATTCAGAGGACTACCTTATACAATCCCGCACAACGACATTAAAACTGCT  6658

gag     2787  GTCGATGCTAAAGGACACGATGTAGCTGCAATTATAAATGTCAAAAGAAGCACGAAGATT  2846
              || ||||||      ||||| || |||||| ||| |||| ||| || ||    ||| | |
3'UTR   6657  GTAGATGCT------CACGAAGTTGCTGCAGTTACAAATATCAGAAAAACTGTGAAAAAT  6604

gag     2847  AATGGAGAGAAAATTATCAAAAACTTTCCACTCTTCTACGTAGATTTATTCCCGAAAGAG  2906
              || || |||||| |||||| |||||||||||| |||||||| || |||  |||||||||
3'UTR   6603  AACGGCGAGAAAGTTATCATAAACTTTCCACTGTTCTACGTGGACTTACACCCGAAAGAC  6544

gag     2907  AACAATAAGGAGATCTATAACATTAAAGATATCCTTCACTGCAAAATCACAGTGGAACCT  2966
              |||||||||||| |||| ||||||||||  | ||||| |||||||||| |||||||||
3'UTR   6543  AACAATAAGGAGGTCTACAACATTAAAGACCTACTTCATTGCAAAATCACCGTGGAACCT  6484

gag     2967  CCTAAGAAGGTTAAAGGAATTCCACAATGTATAAATTGTCAGAGACTTGGACACACTAAA  3026
              || |||||  ||||||  |||||||| || |  | ||| ||||||||||||||||||||
3'UTR   6483  CCAAAGAAAATTAAAGTGATTCCACAGTGCACCAGTTGCCAGAGACTTGGACACACTAAG  6424

gag     3027  GCTTTTTGTAGCCGTGAACCGAGATGTGTGAAATGTGCTGGTAATCACCTAACAA  3081
              ||| | ||| || | ||||||| |||||| || |||||||| |||||||| ||||
3'UTR   6423  GCTATCTGTTGCAGGGAACCGAAATGTGTAAAGTGTGCTGGAAATCACCTGACAA  6369
```

b)

```
Rhynchosciara americana non-LTR retrotransposon RaTART, complete
sequence; Length=8268
Score =  527 bits (584),  Expect = 1e-146
Identities = 336/364 (92%), Gaps = 16/364 (4%)
Strand=Plus/Plus

414     64    AAGACATTTGTGTGCCGTTGGCAACAACTTGATTGAATTGCATTATGCTGACTAATATTT  123
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
RaTART  7281  AAGACATTTGTGTGCCGTTGGCAACAACTTGATTGAATTGCATTATGCTGGCTAATATTT  7340

414     124   GTACACGTGGTTATAATATCACACAGCACATAAACTTTTATGGTTTCTT-----ATAAAA  178
               |||||||||||||||||||||||||||||| |||||||||||||||     ||||||
RaTART  7341  TTACACGTGGTTATAATATCACACAGCACATAACCTTTTATGGTTTCTTATAAAATAAAA  7400

414     179   TAAAATAAACACAGTCCGTACACAGTGTTGTTTTTGCACGGTGTTATATTAATAGTTTGT  238
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RaTART  7401  TAAAATAAACACAGTCCGTACACAGTGTTGTTTTTGCACGGTGTTATATTAATAGTTTGT  7460

414     239   ATTTGATAATTGTTTTGATAATTGTGTTTTT---AATTAAA------AGaaaaaaa--aC  287
              |||||||||||||||||||||||| |||||    ||||||      |||||||||  ||
RaTART  7461  ATTTGATAATTGTTTTGATAATTGTTTTTTTTTATATTAAAAAAAATAGAAAAAAAAGAC  7520

414     288   AATAAACTTCTAGAATGTTATCATGTAGCACATGGTGTGTTTTGTTGCACTTCTATTATT  347
              |||||||||||||||||| ||||| |||||||||||| |||||||||||||||||| |||
RaTART  7521  AATAAACTTCTAGAATGTTCTCATGAAGCACATGGTGTCTTTTGTTGCACTTCTATAATT  7580

414     348   TCAATTGTTCACGAACACAGTTGTTAGCACAGATGTTCTCTGTTTGAAAGCTACGCATCG  407
              |||||||||| |||||||||| |||||||||||||||||||||||||||||| |||||
RaTART  7581  TCAATTGTTCGCGAACACAGTTATTAGCACAGATGTTCTCTGTTTGAAAGCTACACATCG  7640

414     408   ACTG  411
              ||||
RaTART  7641  ACTG  7644
```

c)

```
>ref|XP_001973883.1| Gene info GG21430 [Drosophila erecta]
 gb|EDV54283.1| Gene info GG21430 [Drosophila erecta]
Length=149

 GENE ID: 6548623 Dere\GG21430 | GG21430 gene product from transcript GG21430-
RA
[Drosophila erecta] (10 or fewer PubMed links)

 Score = 67.0 bits (162),  Expect = 6e-10
 Identities = 32/73 (43%), Positives = 51/73 (69%), Gaps = 0/73 (0%)
 Frame = -1

Query  428  IQL*GCTSKSNILKIQRFQNKSLRTIISAP*YARNIDIHRDLKVPMVAEVIKQIAAKHED  249
            +Q+ G  S SNI+++QR QN++LRTI +AP YARN+DI  +LK+ +V + I + ++++ +
Sbjct  47   VQIWGTASDSNIMRVQRVQNRALRTITNAPWYARNVDIANELKILIVRQQINRHSSRYNE  106

Query  248  RLLQHENPEALEL  210
            RL  H N  A  L
Sbjct  107  RLQAHTNHLAASL  119



>ref|XP_002075902.1| Gene info GK24003 [Drosophila willistoni]
 gb|EDW86888.1| Gene info GK24003 [Drosophila willistoni]
Length=179

 GENE ID: 6653561 Dwil\GK24003 | GK24003 gene product from transcript GK24003-
RA
[Drosophila willistoni] (10 or fewer PubMed links)

 Score = 65.5 bits (158),  Expect = 2e-09
 Identities = 34/74 (45%), Positives = 44/74 (59%), Gaps = 0/74 (0%)
 Frame = -1

Query  428  IQL*GCTSKSNILKIQRFQNKSLRTIISAP*YARNIDIHRDLKVPMVAEVIKQIAAKHED  249
            IQL G  S SN+  +QRFQNK LR ++ AP Y RN  +HRDL +P V   I +   +H
Sbjct  82   IQLWGTASHSNVEILQRFQNKLLRKLVRAPFYVRNSVLHRDLNIPSVHREIPRYGVRHLG  141

Query  248  RLLQHENPEALELL  207
            RL  H N  A +L+
Sbjct  142  RLTSHTNRLAADLI  155
>ref|XP_002072471.1| Gene info GK18829 [Drosophila willistoni]
 gb|EDW83457.1| Gene info GK18829 [Drosophila willistoni]
Length=220

 GENE ID: 6649934 Dwil\GK18829 | GK18829 gene product from transcript GK18829-
RA
[Drosophila willistoni] (10 or fewer PubMed links)

 Score = 65.1 bits (157),  Expect = 2e-09
 Identities = 35/74 (47%), Positives = 45/74 (60%), Gaps = 0/74 (0%)
 Frame = -1

Query  428  IQL*GCTSKSNILKIQRFQNKSLRTIISAP*YARNIDIHRDLKVPMVAEVIKQIAAKHED  249
            IQL G  S SN   +QRFQ+K LR I+ AP + RN  IHRDL +P +   I++   ++
Sbjct  122  IQLWGTASDSNFEILQRFQSKLLRNIVEAPPFVRNSIIHRDLNIPSIKSEIQRYGTRYCS  181
```

```
Query  248  RLLQHENPEALELL  207
            RL  HEN  AL LL
Sbjct  182  RLETHENTLALNLL  195


>gb|AAS13459.1|  putative reverse transcriptase [Drosophila simulans]
Length=888

 Score = 65.1 bits (157),  Expect = 2e-09
 Identities = 35/73 (47%), Positives = 48/73 (65%), Gaps = 0/73 (0%)
 Frame = -1

Query  425  QL*GCTSKSNILKIQRFQNKSLRTIISAP*YARNIDIHRDLKVPMVAEVIKQIAAKHEDR  246
            +L G  S+SNI  IQR Q++ LR I  AP Y RN +IHRDLK+ +V EVI +   K+ ++
Sbjct  799  ELWGNASRSNIDIIQRAQSRILRIITGAPWYLRNENIHRDLKIKLVIEVIAEKKTKYNEK  858

Query  245  LLQHENPEALELL  207
            L  H NP A +L+
Sbjct  859  LTTHTNPLARKLI  871
>gb|AAK77263.1|  GH03753p [Drosophila melanogaster]
Length=888

 Score = 65.1 bits (157),  Expect = 2e-09
 Identities = 35/73 (47%), Positives = 48/73 (65%), Gaps = 0/73 (0%)
 Frame = -1

Query  425  QL*GCTSKSNILKIQRFQNKSLRTIISAP*YARNIDIHRDLKVPMVAEVIKQIAAKHEDR  246
            +L G  S+SNI  IQR Q++ LR I  AP Y RN +IHRDLK+ +V EVI +   K+ ++
Sbjct  799  ELWGNASRSNIDIIQRAQSRILRIITGAPWYLRNENIHRDLKIKLVIEVIAEKKTKYNEK  858

Query  245  LLQHENPEALELL  207
            L  H NP A +L+
Sbjct  859  LTTHTNPLARKLI  871
```

ANEXO

CAPÍTULO 2

Cloning and characterisation of a novel chromosome end repeat enriched with homopolymeric dA/dT DNA in *Rhynchosciara americana* (Diptera: Sciaridae)

Thiago Fernandes
Christiane Rodriguez Gutierrez Madalena
Eduardo Gorab*

Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, Rua do Matão 277, Cidade Universitária, São Paulo, SP, Brazil. CEP: 05508-090. Phone: +55-11-30918061; FAX: +55-11-30917553. Email: egorab@usp.br
* Corresponding author

Abbreviations

bp: base pairs
DAPI: 4',6-diamidino-2-phenylindole
DOP- PCR: degenerate oligonucleotide primed PCR
dNTP: deoxy-ribonucleotide 5'-triphosphate
dUTP: 2'-deoxy-uridine 5'-triphosphate
FITC: fluorescein isothiocyanate
Kbp: kilobase pairs
PCR: polymerase chain reaction
SDS: sodium dodecyl sulfate
SSC: sodium saline citrate
TBS: Tris-buffered saline
TBST: Tris buffered saline, 0,1% Triton X-100

Acknowledgements: The authors thank Dr. A. J. Stocker for critical reading of the manuscript and Dr. B. D. Stollar for gift of anti-hybrid antibodies. This work was supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo).

RESUMO

A grande maioria dos organismos eucariontes possuem repetições curtas em *tandem* e telomerase como componentes estruturais de suas estruturas teloméricas. Contudo, em dípteros essa conservação estrutural não tem sido encontrada. Enquanto em *Drosophila* telômeros são compostos por específicos retrotransposons, repetições complexas em *tandem* são encontradas em *Anopheles* e em espécies do gênero *Chironomus*. No sciarídeo *Rhynchosciara americana*, repetições curtas (16 e 22 pares de bases) arranjadas em *tandem* são observadas em regiões cromossômicas terminais desta espécie. Além do mais, resultados de hibridação *in situ* usando sondas de RNA hopolimérico sugerem a existência de uma terceira repetição enriquecida com homopolimero (dA)/(dT). Neste trabalho, microdissecções da extremidade cromossômica C1 de *R. americana* foram realizadas e os fragmentos obtidos amplificados por PCR utilizando um *primer* homopolimérico $(dA)_{20}$ como iniciador. Do conjunto de DNAs amplificados, uma microbiblioteca em vetores plasmídeais foi construída e posteriormente triada para a presença de clones contendo insertos de natureza repetitiva. Após triagem e sequenciamento, uma sequência homopolimérica não pura foi indentificada. Denominada T-14, essa sequência é 93% rica em dA/dT e contem uma unidade de repetição de 14 pb organizada em *tandem*. Resultados de hibridação *in situ* demonstraram que essa repetição está presente nas cinco extremidades não centroméricas e em quatro sítios cromossômicos intersticiais de *R. americana*. Até o momento, T-14 é a menor repetição que tem sido caracterizada em regiões cromossômicas terminais de dípteros, e, em adição, uma região de sua sequência compartilha certa similaridade com repetições teloméricas canônicas. Como observado anteriormente com as sequências curtas M16 e M22, hibridações *in situ* realizadas com o T-14 também geraram sinais em pontes intertelomércas conectando regiões terminais de cromossomos politênicos não homólogos. Os resultados obtidos neste trabalho sugerem que *R. americana* representa um exemplo adicional de um organismo em que mais de uma sequência de DNA possa se estender em direção a regiões cromossômicas terminais.

Palavras-chaves: Diptera, *Rhynchosciara*, repetições, sub-telômeros, telômeros.

ABSTRACT

Short tandem DNA repeats and telomerase compose the telomere structure in the vast majority of eukaryotic organisms. However, such a conserved organisation has not been found in dipterans. While telomeric DNA in *Drosophila* is composed of specific retrotransposons, complex terminal tandem repeats are present in chromosomes of *Anopheles* and chironomid species. In the sciarid *Rhynchosciara americana*, short repeats (16 and 22 base pairs long) tandemly arrayed seem to reach chromosome ends. Moreover, in situ hybridisation data using homopolymeric RNA probes suggested in this species the existence of a third putative chromosome end repeat enriched with (dA).(dT) homopolymers. In this work, chromosome microdissection and PCR primed by homopolymeric primers were employed to clone these repeats. Named T-14 and 93% AT-rich, the repetitive unit is 14 base pairs long and appears organised in tandem arrays. It is localised in five non-centromeric ends and in four interstitial bands of *R. americana* chromosomes. To date, T-14 is the shortest repeat that has been characterised in chromosome ends of dipterans and, in addition, part of its sequence shares some similarity with canonical telomere repeats. As observed for short tandem repeats identified previously in chromosome ends of *R. americana*, the T-14 probe hybridised to bridges connecting non-homologous polytene chromosome ends, indicative of close association of T-14 repeats with the very end of the chromosomes. The results of this work suggest that *R. americana* represents an additional example of organism provided with more than one DNA sequence that is able to reach chromosome termini.

Key words Diptera, *Rhynchosciara*, repeats, sub-telomeres, telomeres.

1 INTRODUCTION

Telomere research data place organisms into three main classes. Most eukaryotes belong to the group characterised by short tandem DNA repeats specified by telomerase. Dipterans are present into two other classes, one has *Drosophila* species whose telomere structure is composed of specific retrotransposons (reviewed in Mason et al. 2008; Villasante et al. 2008). Long tandem arrays of complex (≥170 bp) terminal repeats are the main feature of the third chromosome end class. It includes some plants (reviewed in McKnight and Shippen 2004), lower dipterans (Sub-order Nematocera) such as *Chironomus* (Carmona et al. 1985; Saiga and Edström 1985; Nielsen et al. 1990; López et al. 1996; Rosén and Edström 2000; Martínez et al. 2001) and *Anopheles gambiae* (Biessmann et al. 1996; Biessmann et al. 1998). Eukaryotes provided with telomerase and *Drosophila*, despite huge structural differences with regard to telomere organisation, share reverse transcription as the mechanism responsible for the recovery of telomere DNA.

In contrast to *Chironomus* and *Anopheles*, another nematoceran, *Rhynchosciara americana*, has reverse transcriptase and RNA as stable components of non-centromeric chromosome ends throughout the whole larval development (Gorab 2003). Although the function of the telomeric reverse transcriptase in *R. americana* still remains speculative, its occurrence led to start the characterisation of sequences

composing chromosome termini of this species. A plasmid library made from a microdissected chromosome end allowed the characterisation of a 414 base pairs (bp) repeat resembling nematoceran telomeric repeats (Madalena and Gorab 2005). However, its sub-telomeric location prompted us to continue the search for sequences composing chromosome ends of *R. americana*. A second repeat type was eventually identified in the chromosome tips of *Rhynchosciara*. Named M-22, it is unusually short for a tandem repeat at chromosome ends of dipterans. These regions showed an overlap of labelling by M-22 probe hybridisation and immunodetection of reverse transcriptase (Rossato et al. 2007), suggesting that M-22 tandem repeats lie distal to the 414 bp repeat arrays.

Continued study of *Rhynchosciara* on this subject led to the characterisation of a second unusually short repeat composed of 16 nucleotides, named M-16, tandemly arrayed and intermingled with M-22 tandem repeats. Data analyses of several cloned segments containing M-16 and M-22 repeats suggested segmental duplication events as a probable consequence of unequal recombination. Such a feature, as proposed previously for *Anopheles* (Roth et al. 1997), might be related to chromosome end elongation in this species (Madalena et al. 2010b). Despite of this mechanism possibly operating in *R. americana*, a missing link between canonical and dipteran telomeres has been hypothesised from the short repeats, RNA and reverse transcriptase at the chromosome end structure of this species.

This work focused on the cloning of a terminal repeat of this species that has remained to be identified since its first identification by means of in situ hybridisation

using poly-(rU) as a probe (Jones et al. 1973). Unusually enriched with (dA).(dT) homopolymeric DNA and localised at chromosome ends, pericentric regions and specific interstitial sites (Stocker et al. 1993; Stocker and Gorab 2003), it also has the potential to form three-stranded structures with exogenous poly-(rU) (Gorab et al. 2009). The repetitive sequence characterised in this work represents the shortest tandem repeat that has been identified in the chromosome end structure of dipterans up to now. Possible implications of this novel genomic component in the chromosome end structure of *R. americana* are discussed.

2 MATERIALS AND METHODS

2.1 Animals

Rhynchosciara americana, *R. hollaenderi*, *R. baschanti*, and *R. milleri* larvae were collected in the coastal regions of Mongaguá and Ubatuba, State of São Paulo, and maintained in the laboratory at 18-22^{o}C.

2.2 Preparation of chromosome spreads

The salivary glands were dissected in Ringer's buffer and fixed briefly in ethanol-acetic acid (3:1). After squashing the glands in 50% acetic acid, the slides were frozen for 10

min on dry ice. The coverslips were flipped off with a razor blade and the slides were then kept in absolute ethanol at -20°C until use.

2.3 Microdissection and polymerase chain reactions

Microdissection of a chromosome C-1 tip of *R. americana* was described previously (Madalena and Gorab 2005) with a single modification. All the microdissection steps were performed with dried slides. The following reaction mixture was added to the PCR tube: 0.2 mM of each dNTP; 1X Taq buffer; 2.5 U Taq DNA polymerase (Fermentas); 1 mM $MgCl_2$; 1 μg $(dA)_{20}$ primer, H_2O to a final volume of 50 μl. Prior to the amplification rounds, the mixture was briefly centrifuged. DNA in the chromosome fragment was amplified using two distinct routines. Cycle conditions for the first routine were: 93^{o}C for 4 min followed by 8 cycles of 94^{o}C for 1 min, 30^{o}C for 90 s and 72^{o}C for 3 min. The amplification was completed with additional 28 cycles of 94^{o}C for 1 min, 56^{o}C for 1 min, 72^{o}C for 3 min and a final extension at 72^{o}C for 5 min. Cycle conditions for the second routine were: 93^{o}C for 4 min followed by 28 cycles of 94^{o}C for 1 min, 42^{o}C for 50 s, 72^{o}C for 3 min and a final extension at 72^{o}C for 5 min. After sequencing p*Ra*-55 and p*Ra*-59 inserts, specific primers T19.1 and T19.2 were designed to clone more repeats. T19.1: 5'TTATGTTTATTTTTTATG3'. T19.2:5'AATACAAATAAAAAAATAC3' Cycle conditions were the same as those described above for the second amplification round in the first routine or for the second routine.

2.4 Chromosome end micro-library and sequencing

PCR products were ligated in pGEM-T Easy vector according to standard protocols (Promega). Procedures for transformation and plating of *E. coli* DH5α strain were described elsewhere (Sambrook et al. 1989). Individual colonies were randomly chosen to grow in liquid medium for plasmid DNA extraction using Concert Rapid Plasmid Miniprep System (Invitrogen). Sequencing reactions of plasmid DNAs were performed with Big Dye Terminator Cycle Sequencing Ready Reaction as recommended by the manufacturer (Applied Biosystems) using M13 Forward or Reverse primers (Invitrogen) and subsequently run in the ABI PRISM 310 Genetic Analyser (Applied Biosystems). Sequence alignments and GenBank searches were done with BLAST (Altschul et al. 1990).

2.5 Non radioactive labelling of the probes

Probe labelling was done with biotinylated-11 dUTP (Sigma) replacing dTTP in 50 µl of PCR reaction mixture containing either 1 µl of PCR product obtained previously together with $(dA)_{20}$ primer and following the same cycling conditions described above. Cloned sequences were labelled from 5 ng plasmid DNA together with M-13 Forward and Reverse primers (Invitrogen) following the same cycling conditions described for the second amplification round in the first PCR routine. Alternatively, 5'end-biotinylated primers (Invitrogen) representing individual repeats 5'TTTATTTTTTATG3' or 5'AAATAAAAAAATAC3'were used as probes for in situ hybridisation.

2.6 In situ hybridisation

In situ hybridisation and detection procedures were described elsewhere (Madalena and Gorab 2005). When poly-(rA) was used as a probe, the air-dried polytene chromosome squashes were previously treated with RNase A (0.6 mg/ml) diluted in 2X SSC at 37°C for at least 2 h in a moisted chamber. The slides were extensively washed in 2X SSC before being immersed in ethanol. The probe mixture consisted of 50% formamide, 2X SSPE, 0.1% SDS and poly-(rA) (1 μg per slide); 5-10 μl of probe mixture were applied to each air-dried slide and covered with a plastic coverslip. The slides were steam heated for 2 min at 70°C and kept for 1 h in a closed box at 37°C. The slides were washed twice in TBST for 15 min and incubated for 1 h at room temperature with goat IgG anti-DNA/RNA hybrid (Stollar 1970; Kitagawa and Stollar 1982) diluted 1:50 in TBST. The slides were washed twice in TBST for 15 min and incubated for 1 h with rabbit IgG anti-goat labelled with FITC or rhodamine (Sigma) diluted 1:200 as described above. The slides were washed twice in TBST for 15 min and finally once in 1X TBS for 5 min. The slides were mounted in antifading medium and inspected with epifluorescence optics (Nikon).

2.7 Bal 31 assay

High molecular weight DNA extraction used for the analysis of replication intermediates was performed according to Yokosawa et al. (1999) except for the dialysis step that was made in *Bal* 31 buffer (Sambrook et al. 1989). Approximately 35 μg of DNA dissolved in 300 μl *Bal* 31 buffer were treated with *Bal* 31 (15 U, Stratagene). Time

course digestion was interrupted by transferring equal volumes of each sample to another tube containing 150 µl NaI/Geneclean II followed by DNA purification according to the standard protocol (BIO 101 Inc.). DNA samples were then digested overnight with *Bgl*I before Southern-blot analysis.

2.8 Southern-blot hybridisation

DNA from gels was alkaline transferred to Hybond N+ membranes according to standard procedures (GE Healthcare). Hybridisation was carried out overnight at 60^{o}C in 0.5 M Na_2HPO_4, 4% SDS. Plasmid inserts or *Rhynchosciara* total salivary gland DNA were labelled by random priming with α^{32}P-dATP following current protocols (Invitrogen). The membranes were washed twice at 60^{o}C for 30 min in 40 mM Na_2HPO_4, 2% SDS. Restriction enzymes and DNA size markers were purchased from Invitrogen and GE Healthcare.

3 RESULTS

3.1 Chromosomal localization of PCR products primed by $(dA)_{20}$ in *R. americana*

Our previous results obtained by degenerate oligonucleotide primed PCR (DOP-PCR) applied to DNA from microdissected chromosome ends of *R. americana* suggested that sequences responsible for the highly reproducible in situ hybridisation pattern obtained with the poly-(rA) probe (SM, Fig. S1) had not been amplified. Given the enrichment with homopolymeric (dA).(dT) DNA in the sequences of interest, amplification was performed with $(dA)_{20}$ in place of random primers together with DNA from a single microdissected C-1 chromosome tip of *R. americana*.

In the first experiments, the hybridisation temperature of the primer was fixed at 30°C following DOP-PCR conditions described previously (Telenius et al. 1992). PCR products were visualised in gels as a smear ranging from 100 bp up to 3 Kbp (data not shown). The whole amplification product was labelled and, when used for in situ hybridisation, was able to reproduce the same results observed with the poly-(rA) probe (Fig. 1; SM, Fig. S1). Although the data indicated that sequences of interest had been amplified, cloning was not successful probably due to their low representation in the pool of amplified products.

In order to try to increase the amplification specificity of sequences enriched with (dA).(dT) homopolymers, a more restrictive primer annealing was introduced in a shorter PCR routine. When the hybridisation temperature of the primer was raised to 45°C, not only the mass of PCR products decreased significantly but also the size of DNA fragments obtained fell into a narrower range when compared to the first experiments (data not shown). After labelling the total PCR product, in situ hybridisation results with the new probe reproduced partially those observed with probes made of either poly-(rA) (SM, Fig. S1) or microdissected DNA (Fig. 1). Interestingly, hybridisation signals were not detected at pericentric regions and were restricted to interstitial bands in addition to non-centromeric ends (Fig. 2).

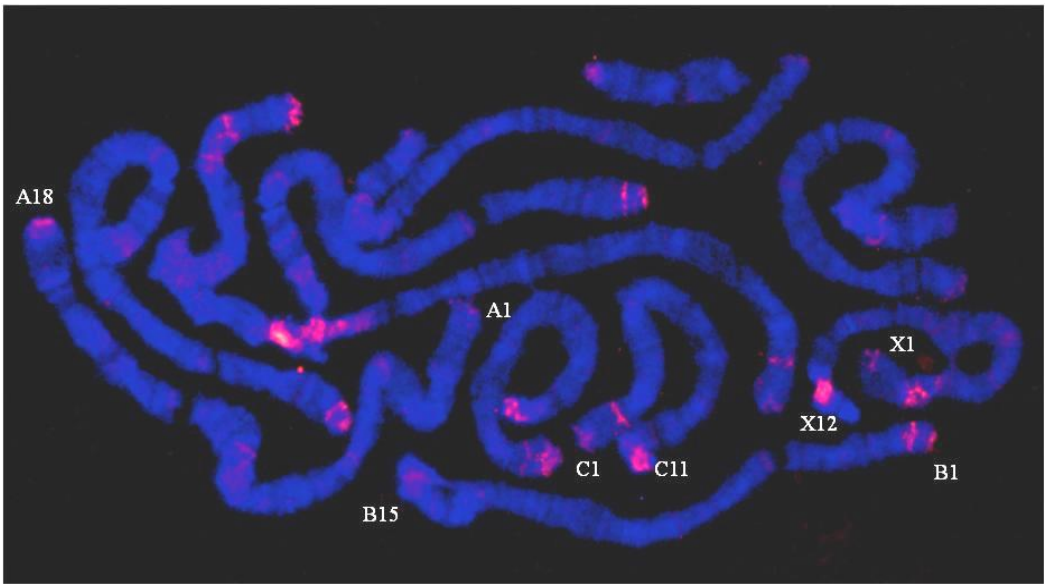


Figure 1: Localisation of PCR products primed by (dA)20 (*red signal*) in polytene chromosomes of *R. americana* counterstained with DAPI (*blue signal*). Probe synthesis was carried out with DNA from a single C-1 chromosome tip and hybridisation temperature of the primer at 30°C. The identification of chromosome ends was made according to Stocker et al. 1993.

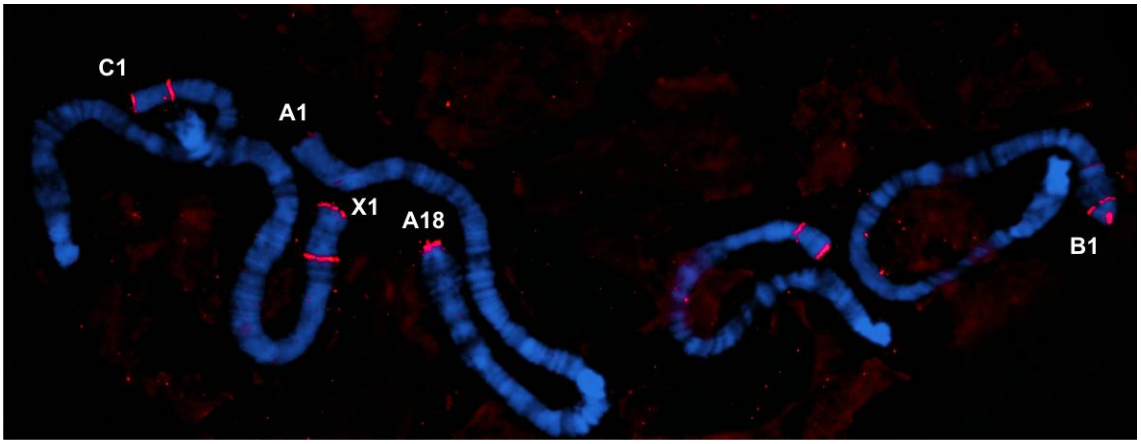


Figure 2: Localisation of PCR products primed by $(dA)_{20}$ (*red signal*) in polytene chromosomes of *R. americana* counterstained with DAPI (*blue signal*). Probe synthesis was carried out with DNA from a single C-1 chromosome tip and hybridisation temperature of the primer at 45°C.

3.2 Cloning and characterisation of terminal (dA).(dT)-rich repeats of *R. americana*

The results described above showed that the primer annealing performed at the higher temperature (45°C) resulted in the synthesis of a probe that did not hybridise to centromeric ends but labelled consistently the expected interstitial bands as well as non-centromeric ends of *R. americana* chromosomes. This PCR product was chosen for the construction of a plasmid microlibrary since the aim of the work was the identification of sequences composing non-centromeric ends of this species.

Plasmid DNA from 36 colonies was digested with *Eco*RI and the inserts were analysed by Southern-blot hybridisation using total salivary gland DNA as a probe. Attention was given to two clones whose inserts hybridised strongly to the probe,

indicating cloned repetitive sequences (data not shown). Plasmids named p*Ra*-55 and p*Ra*-59 were labelled and hybridised to polytene chromosome squashes of *R. americana*. Identical results were obtained with the two probes that reproduced exactly the same hybridisation pattern observed in Fig. 2, namely, fluorescent signals detected at non-centromeric ends and specific interstitial bands (Fig. 3).

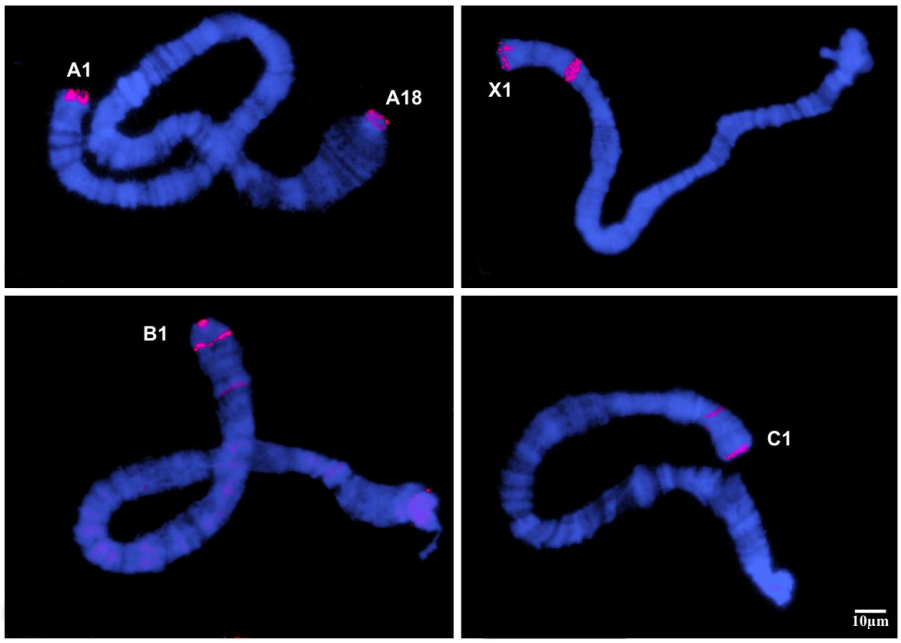


Figure 3: Localisation of p*Ra*-55 probe (*red signal*) in polytene chromosomes of *R. americana* counterstained with DAPI (*blue signal*).

After sequencing p*Ra*-55 and p*Ra*-59 inserts, homopolymeric sequences showing distinct extensions, probably representing primers, were identified flanking identical inserts, one of which is shown (Fig. 4a). A manual alignment of sequences

composing the insert was then carried out and enabled the identification of 14mers, named T-14, 93% AT-rich and organised as a tandem array (Fig. 4b). Specific primers were designed in order to amplify and clone more genomic regions containing T-14 repeats. Although this procedure increased T-14 repeat sampling, very few plasmids provided with short inserts were obtained (SM, Fig. S2). In relation to the sequences deposited in databases, T-14 displayed significant sequence similarity (86%) to specific stretches localised in the chromosome X of *Drosophila melanogaster* (SM, Fig. S3).

a

```
TTTTTTTTTTTTTTTTTTTTATGTTTATTTTTTTATGTTTATTTTTTTA 050
TTTTTATTTTTATATGTTTATTTTTTTTATTTTTATTTTTATTTTTATATG 100
TTTATTTTTTTATGTTTATTTTTTGTTGTTTATTTTTTTATGTTTATTTT 150
TTTATGTTTATTTTTTTATGTTTATTTTATAAAAAAAAAAAAAAAAAAAA 200
AAA
```

b

```
TTTTTTTATGTTTA
TTTTTTTATGTTTA
TTTTTTTATTTTTA
TTTTTATATGTTTA
TTTTTTTATTTTTA
TTTTT--ATTTTTA
------TATGTTTA
TTTTTTTATGTTTA
TTTTTTGTTGTTTA
TTTTTTTATGTTTA
TTTTTTTATGTTTA
TTTTTTTATGTTTA
TTTT-ATA
```

c

```
TTTTTTTATGTTTA
```

Figure 4: a-c p*Ra*55 insert sequence **(a)**; stretches comprising homopolymeric primers are underlined. **(b)** The manual alignment was done with sequences between primers; divergences in relation to the most frequent base within a column are shadowed. Gaps (-) were introduced to maximise the alignment. **(c)** Consensus sequence derived from the alignment and from additional sequencing results (SM, Fig. S2).

3.3 Genomic analysis of T-14 repeats

The hybridisation signals obtained with the cloned sequences can only be explained if long T-14 tandem arrays are present in the genome of *R. americana*. This assumption is also supported by positive in situ hybridisation results using, as probes, biotinylated primers representing a single T-14 repeat.

Southern-blots of *R. americana* genomic DNA cut with restriction enzymes were hybridised to the T-14 repeat probe. Strong hybridisation signals were usually detected in DNA fragments migrating in the range of 10-20 Kbp. However, the probe also hybridised significantly to smaller fragments when genomic DNA was cut with *Dra*I (Fig. 5).

Plasmid inserts containing cloned T-14, M-16 and M-22 repeats were hybridised to total genomic DNA of *R. americana*. This method was already employed in previous reports (Madalena et al. 2010a; Madalena et al. 2010b) and has provided a measure of sequence representation in genomes. The results clearly showed that the reiteration degree of T-14 repeats is lower than that of M-22 repeats in the genome of *R. americana* (Fig. 6).

Simultaneous detection of T-14 and M-22 probe hybridisation was also carried out to try to distinguish the position of each fluorescent signal relative to the chromosome end. Use of this method never showed a separation of the two signals in many slides analysed (Fig. 7). The occurrence of bridges connecting non-homologous

polytene chromosome ends of *R. americana* was then exploited in an attempt to discriminate the location of T-14 repeats in chromosome termini. It was previously shown that M-22 and M-16 probes hybridised to telomeric bridges, suggesting very close proximity to the chromosome ends (Madalena et al. 2010b). Accordingly, the T-14 repeat probe hybridised clearly to telomeric bridges (Fig. 8).

Although the *Bal*-31 assay is not usually recommended for the study of non canonical chromosome ends, T-14 repeat sensitivity to *Bal*-31 was examined in comparison to that of genomic regions containing M-22 and M-16 repeats. The histone gene repeat locus, localised interstitially in the chromosome A (Gorab 2003) was included as a control. The quality of the DNA sample used for the *Bal*-31 assay was readily assessed after *Bgl*I digestion and electrophoresis by examination of the *Bgl*I restriction fragment of the histone gene. This fragment remained visible without apparent signs of degradation (SM, Fig. S4a), indicating that high molecular weight DNA was obtained for the assay. In contrast to the histone gene DNA, M-22 and M-16 repeats were found to be highly sensitive to the exonuclease, showing a comparable digestion profile up to 15 minutes. However, M-16 and M-22 repeats were nearly digested to completion within thirty minutes (SM, Fig. S4b-d) while T-14 digestions, although highly sensitive to *Bal*-31, showed a slower profile (SM, Fig. S4e).

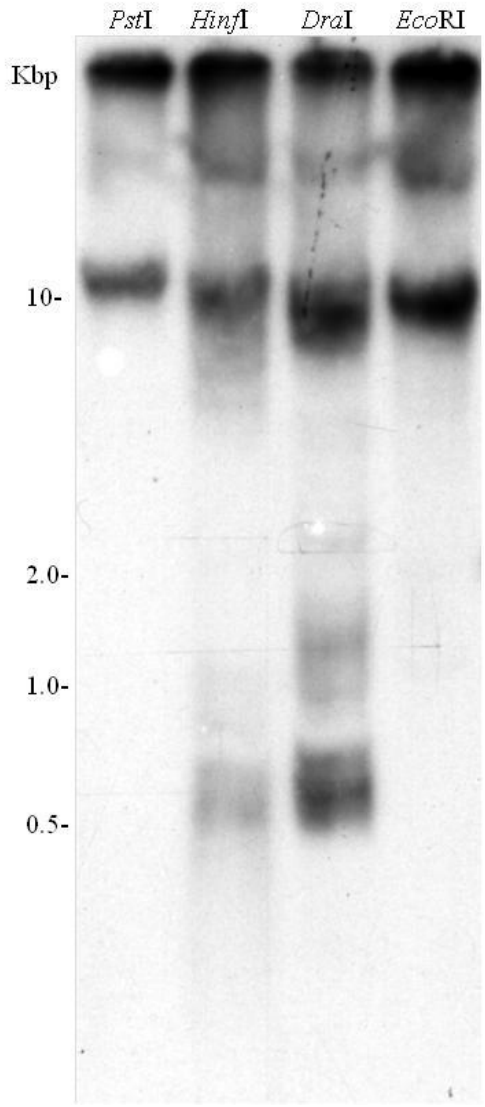


Figure 5: Southern-blot hybridisation of *R. americana* salivary gland DNA (approximately 3 µg per lane) digested with restriction enzymes identified on top and probed with the p*Ra*-55 insert.

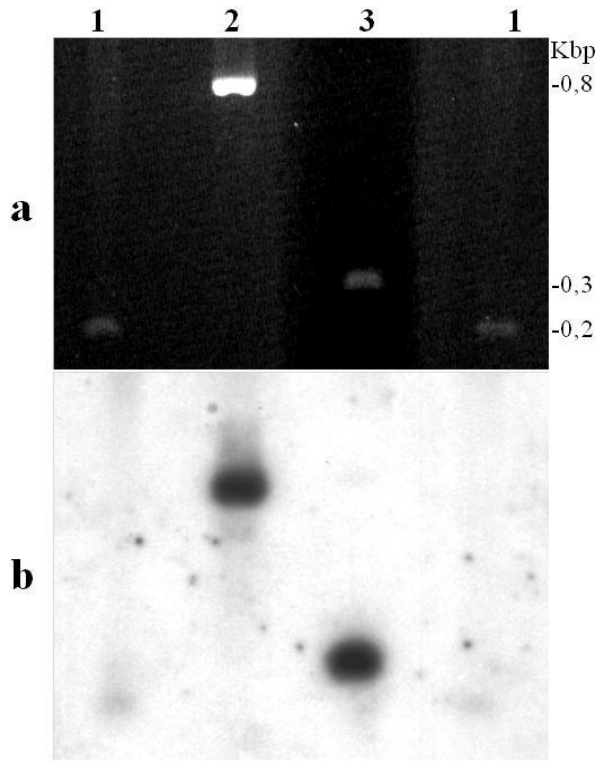


Figure 6: (a) Plasmid DNAs were cut with *Eco*RI and analysed in an agarose gel (only inserts are shown). **1**: p*Ra*-55 (*T-14 repeats*); **2**: p*Ra*11-5.2 (*M-16 repeats*; Madalena et al. 2010); **3**: p*Ra*M47.33 (*M-22 repeats*; Rossato et al. 2007). DNA from the gel was transferred to membrane and **(b)** hybridised to salivary gland DNA of *R. Americana*.

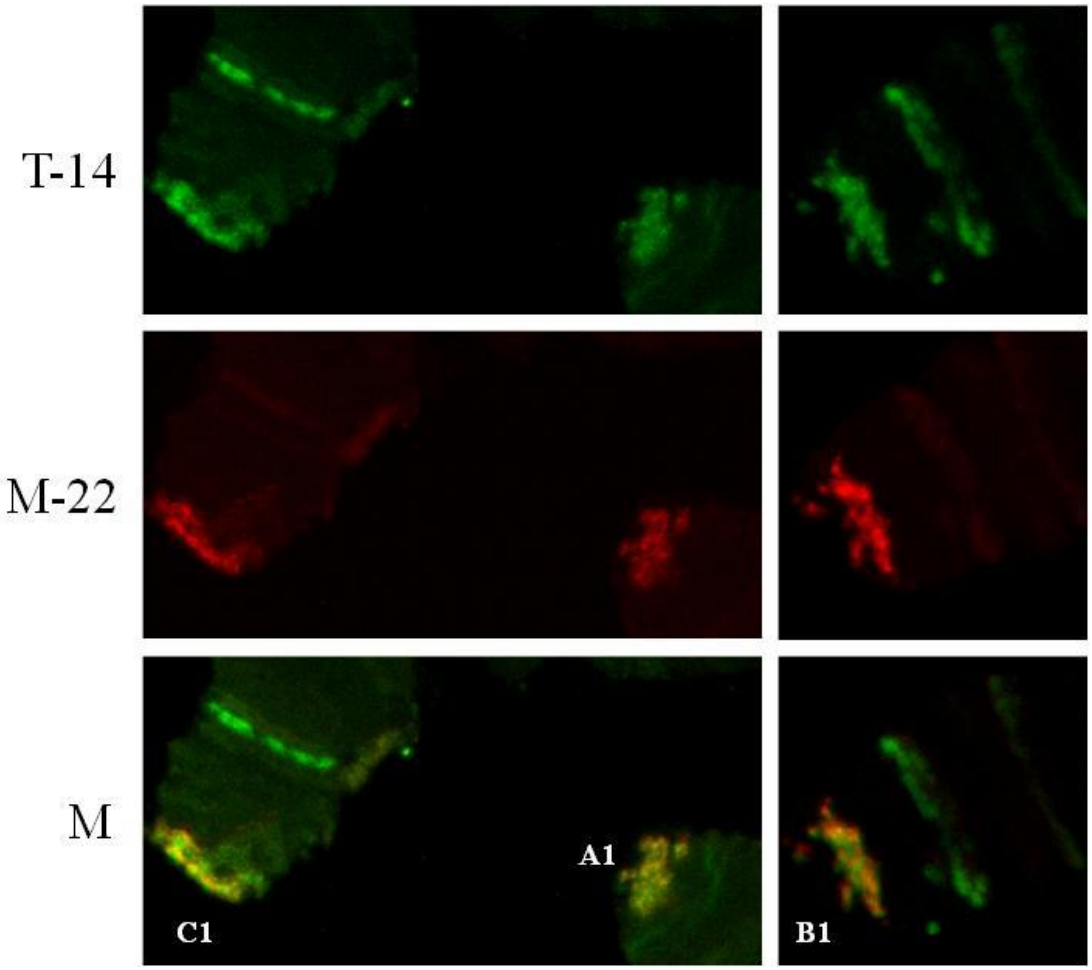


Figure 7: Simultaneous detection of T-14 (*green signal*) and M-22 (*red signal*) probe hybridisation at chromosome end sections of *R. americana* showing (*M*) the merged signals.

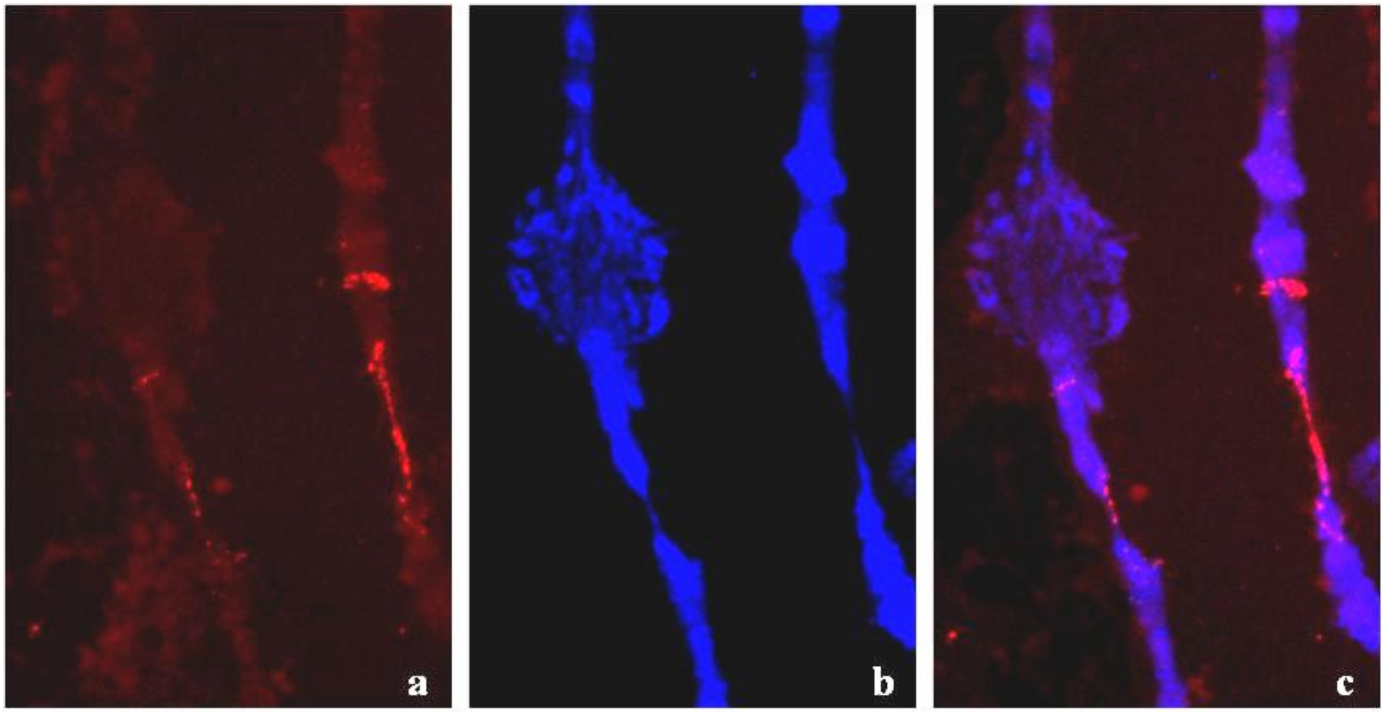


Figure 8: (a) Localisation of p*Ra*-55 probe (*red signal*) in polytene chromosomes of *R. americana* connected by inter-telomeric bridges showing **(b)** DAPI counterstaining and **(c)** the merged signals.

4 DISCUSSION

To our knowledge, in situ hybridisation using exogenous poly-(rU) in *R. americana* polytene chromosomes (Jones et al. 1973) provided the first data of probe localisation at dipteran chromosome ends. Those results could be reproduced with the poly-(rA) probe followed by the immunological detection of DNA.RNA hybrids (Stocker et al. 1993; Stocker and Gorab 2003; Gorab et al. 2009), reinforcing the existence of local enrichment with homopolymeric (dA).(dT) DNA in the genome of this species. A strategy based on PCR using DNA from a single microdissected chromosome tip and a single homopolymeric $(dA)_{20}$ primer was then employed to clone the sequences of interest.

Named T-14, the repetitive sequence can be defined by the formula (T_3AT_7ATG) and represents the shortest tandem repeat unit localised at chromosome ends that has been characterised in dipterans up to now. T-14 repeats are present at five non-centromeric ends which also display reverse transcriptase as a constitutive component of chromosome termini throughout the larval development (Gorab 2003). The PCR routines designed in this work also revealed the existence of a second repeat, localised in pericentric regions of *R. americana* chromosomes, that remains to be cloned. It is probably organised as tandem arrays and enriched with (dA).(dT) homopolymers as inferred by in situ hybridisation using either RNA or DNA probes. Amplification of these sequences, as shown in the results, requires annealing temperature of the $(dA)_{20}$ much lower than that used for T-14 cloning.

The structure of T-14 repeats provides an explanation for the lack of hybridisation results in *R. americana* chromosomes when labelled homopolymeric (dA).(dT) probes replaced the poly-(rA) probe. Many fruitless attempts were done in our laboratory with DNA probes synthesised either by terminal transferase, or by random priming or even by PCR using commercial homopolymeric (dA).(dT) DNA and homopolymeric primers. T-14 tandem arrays can be viewed as islands of (dA).(dT) homopolymers composed of seven and three base pairs per repeat. While the poly-(rA) probe is considerably longer than the T-14 length (Gorab et al. 2009), it is able to hybridise to islands of homopolymeric chromosomal DNA forming DNA.RNA hybrids that are more stable than DNA.DNA duplexes produced by homopolymeric DNA probes. Although DNA.RNA hybrid duplexes are not continuous but rather interrupted in the T-14 array, the extension of the tandem repeat is likely to warrant sufficient antibody density to be visualised by immunofluorescence.

The observation that T-14 repeats are differentially distributed in *Rhynchosciara* genomes is in agreement with previous findings arguing that the evolution of *Rhynchosciara* genomes has proceeded, among other processes, by differential occupation of repetitive sequences. *R. americana* clearly displays more genomic regions containing repeats than other *Rhynchosciara* species that have been studied so far, which can be observed not only at chromosome tips (SM, Fig. S5; Table 2) but also at other chromosomal sites. The increase of repetitive DNA types particularly at non-centromeric ends correlates with the detection of telomeric reverse transcriptase in *R. americana*, a feature restricted to this species in the genus (Gorab 2003).

In relation to the *R. americana* repeats that are common to non-centromeric termini, T-14 is less abundant than M-22 and probably less represented than M-16 at chromosome ends of this species. Although the quantification assay was not conclusive with regard to the reiteration degree of M-16 and T-14 repeats, the genomic distribution of the latter is not restricted to the five non-centromeric ends. There are four interstitial sites that hybridised to the T-14 probe, implying significant tandem repeat extension within those regions. The genomic distribution of T-14 repeats may thus explain the lower sensitivity of T-14 repeats to *Bal*-31 digestion when compared to those from M-22 and M-16 repeats. Southern-blots of genomic DNA of *R. americana* cut with restriction enzymes and probed with T-14 repeats regularly showed hybridisation bands in the range of 10-20 kbp. However, *Dra*I digestion resulted additionally in probe hybridisation of shorter DNA fragments. Considering the *Dra*I restriction site (TTT/AAA), the results might indicate T-14 sequence polymorphism. Alternatively, the data could reflect the occurrence of *Dra*I restriction sites within sequences flanking T-14 tandem repeats.

Concerning the origin of repeats characterised in this work, the T-14 repeat unit can be divided into two partially similar modules reminiscent of a duplication event of a shorter, putative ancestral unit composed of seven nucleotides. The module terminating with G displays variable sequence similarity with canonical telomere repeats (Table 1), which has not been observed when alignments were done using any other repetitive sequence from non-canonical telomeres. Whether T-14 has some evolutionary relationship to canonical terminal repeats, it remains to be elucidated. In any case, T-14 to date is the only dipteran repeat located at chromosome termini that allows some comparison to canonical telomere sequences in terms of similarity as well as repeat size.

On the other hand, T-14 resemblance with DNA stretches within the genome of *D. melanogaster* deserves further investigation. *Drosophila* sequences aligned with T-14 repeats are located in the pericentric heterochromatin of the chromosome X and are apparently organised as a tandem repeat array. At a first glance, *Drosophila* and *Rhynchosciara* repeats seem to be related as deduced by sequence similarity. However, the characterisation of this potentially novel satellite-like component of *Drosophila* heterochromatin is still lacking.

Table 1 Canonical telomere sequences were aligned with part of the T-14 repeat consensus showing base divergences that appear in bold type/shadowed. *Ra*: *Rhynchosciara americana*; 1: *Chlamydomonas reinhardtii*; 2: *Euplotes*, *Oxytricha*, *Stylonychia*; 3: *Arabidopsis*, tomato, *Plasmodium*; 4: Vertebrata, *Trypanosoma brucei*, *Podospora*, *Fusarium*, *Neurospora*, *Didymium iridis*, *Physarum polycephalum*; 5: *Glaucoma*, *Tetrahymena*; 6: most Arthropoda; 7: *Giardia lamblia* and 8 *Tribolium castaneum* (beatle). Data were taken from this work (*Ra*), Vítková et al. 2005 (*Arthropoda*) and Kipling 1995.

Ra	**T**	**T**	**T**	**A**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**A**	**T**	**G**
1							T	T	T	T	**A**	**G**	**G**	G
2							T	T	T	T	**G**	**G**	**G**	G
3								T	T	T	**A**	**G**	**G**	G
4									T	T	**A**	**G**	**G**	G
5									T	T	**G**	**G**	**G**	G
6										T	T	A	**G**	G
7										T	**A**	**G**	**G**	G

Chromatin bridges connecting non-homologous polytene chromosome ends have been documented in distantly related dipterans. Although it is not known if terminal bridges discriminate telomeric and sub-telomeric compartments, the T-14 probe hybridised to inter-telomeric bridges of *R. americana*, as already observed with other probes of this species such as M-22 and M-16 (Madalena et al. 2010b). It is worth mentioning that cloned sequences characterised as telomeric DNA of *D. melanogaster* and *Chironomus* species also hybridised to bridges linking non-homologous chromosome ends (Young et al. 1983; Carmona et al. 1985; Cohn and Edström 1991). The above data support the interpretation of T-14 repeats as DNA sequences that, as for M-16 and M-22, are able to reach chromosome ends of *R. americana*.

Canonical telomeres usually have a basic repeat type specified by telomerase. However, there are examples of organisms provided with more than one telomeric sequence. The first evidence pointing to such a possibility came from the study of chironomid telomeres. *Chironomus* species display complex repeats that are present at seven of eight polytene chromosome ends (references in the Introduction). The telocentric end of chironomid chromosome IV has instead centromeric sequences (Rovira et al. 1993; Rosén et al. 2002). *Drosophila* constitutes the best example studied of an organism with different telomere repeats. While terminal DNA of *D. melanogaster* chromosomes is composed of a single class of repetitive DNA, telomere retrotransposons of this species have distinct sequences (reviewed in Villasante et al. 2008; Mason et al. 2008). *Anopheles gambiae* also has complex repeats apparently close to the telomere 2L that have not been detected at other chromosome ends (Biessmann et al. 1996). The results of this work argue that *R. americana* may represent another example of organism with distinct terminal

sequences. T-14, M-16 and M-22 repeats can be included in the same repetitive DNA class and might be viewed as terminal components of *R. americana* chromosomes.

Finally, it was shown in a previous report that RNA forming hybrid molecules with its template DNA was detected at chromosome ends of *R. americana* with an antibody to RNA.DNA hybrids (Stollar 1970; Kitagawa and Stollar 1982). The hybrid signal was regularly present at five non-centromeric tips reactive with anti-reverse transcriptase throughout the larval development (Gorab 2003). It is not known yet whether transcription of T-14, M-16 and M-22 repeats is related to the results described above. This possibility is under current investigation in the laboratory. Reverse transcriptase, RNA and short repeats at the chromosome ends of *R. americana* make a picture resembling canonical telomeres but the functional connexion of these components is conjectural at present and requires more experimental work.

REFERENCES

Altschul SF, Gisg W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410

Biessmann H, Donath J, Walter MF (1996) Molecular characterization of the *Anopheles gambiae* 2L telomeric region via an integrated transgene. Insect Mol Biol 5:11-20

Biessmann H, Kobeski F, Walter MF, Kasravi A, Roth CW (1998) DNA organization and length polymorphism at the 2L telomeric region of *Anopheles gambiae*. Insect Mol Biol 7: 83-93

Carmona MJ, Morcillo G, Galler R, Martinez-Salas E, de la Campa A, Díez JL, Edström J-E (1985) Cloning and molecular characterization of a telomeric sequence from a temperature-induced Balbiani ring. Chromosoma 92: 108-115

Cohn M, Edström J- E (1991) Evolutionary relations between subtypes of telomere associated repeats in *Chironomus*. J Mol Evol 32: 463-468

Gorab E (2003) Reverse transcriptase-related proteins in telomeres and in certain chromosomal *loci* of *Rhynchosciara* (Diptera: Sciaridae). Chromosoma 111: 445-454

Gorab E, Amabis JM, Stocker AJ, Drummond L, Stollar BD (2009) Potential sites of triple helical nucleic acid formation in chromosomes of *Rhynchosciara americana* (Diptera: Sciaridae) and *Drosophila melanogaster*. Chromosome Res 17: 821-832

Jones KW, Bishop JO, Brito da Cunha A (1973) Complex formation between poly(r)-U and various chromosomal loci in *Rhynchosciara*. Chromosoma 43: 375-390

Kipling D (1995) The Telomere. Oxford University Press, Oxford, New York, Tokio, pp 37

Kitagawa Y, Stollar BD (1982) Comparison of poly (A).poly (dT) and poly (A).poly (dC) as immunogens of the induction of antibodies to RNA.DNA hybrids. Mol Immunol 19: 413-420

López CC, Nielsen L, Edström J-E (1996) Terminal long tandem repeats in chromosomes from *Chironomus pallidivittatus*. Mol Cell Biol 16: 3285-3290

Madalena CRG, Gorab E (2005) A chromosome end satellite of *Rhynchosciara americana* (Diptera: Sciaridae) resembling nematoceran telomeric repeats. Insect Mol Biol 14: 255-262

Madalena CRG, Fernandes T, Villasante A, Gorab E (2010a) Curiously composite structures of a retrotransposon and a complex repeat associated with chromosome

ends of *Rhynchosciara americana* (Diptera: Sciaridae). Chromosome Res 18: 587-598

Madalena CRG, Amabis JM, Gorab E (2010b) Unusually short tandem repeats appear to reach chromosome ends of *Rhynchosciara americana* (Diptera: Sciaridae). Chromosoma 119: 613-623

Martinez JL, Edström J-E, Morcillo G, Díez JL (2001) Telomeres in *Chironomus thummi* are characterized by different families of complex DNA repeats. Chromosoma 110: 221-227

Mason JM, Frydrychova RC, Biessmann H (2008) *Drosophila* telomeres: an exception providing new insights. BioEssays 30: 25-37

McKnight TD, Shippen DE (2004). Plant telomere biology. Plant Cell 16: 794-803

Nielsen L, Schmidt ER, Edström J-E (1990) Subrepeats result from regional DNA sequence conservation in tandem repeats in *Chironomus* telomeres. J Mol Biol 216: 577-584

Rosén M, Edström J-E (2000) DNA structures common for chironomid telomeres terminating with complex repeats. Insect Mol Biol 9: 341-347

Rosén M, López CC, Edström JE (2002) Telomere terminating with centromere specific repeats is closely associated with a transposon derived gene in *Chironomus pallidivittatus*. Chromosoma 110: 532-541

Rossato RM, Madalena CRG, Gorab E (2007) Unusually short tandem repeats in the chromosome end structure of *Rhynchosciara* (Diptera: Nematocera). Genetica 131: 109-116

Roth CW, Kobeski F, Walter MF, Biessmann H (1997) Chromosome end elongation by recombination in the mosquito *Anopheles gambiae*. Mol Cell Biol 17: 5176-5183

Rovira C, Beermann W, Edström JE (1993) A repetitive DNA sequence associated with the centromeres of *Chironomus pallidivittatus*. Nucleic Acids Res 21: 1775-1781

Saiga H, Edström J-E (1985) Long tandem arrays of complex repeat units in *Chironomus* telomeres. EMBO J 4: 799-804

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Stocker AJ, Gorab E, Amabis JM, Lara FJS (1993) A molecular cytogenetic comparison between *Rhynchosciara americana* and *Rhynchosciara hollaenderi* (Diptera: Sciaridae). Genome 36: 831-843

Stocker AJ, Gorab E (2003) Local enrichment with homopolymeric (dA/dT) DNA in genomes of some lower dipterans and *Drosophila melanogaster*. Chromosoma 111: 455-460

Stollar BD (1970) Double helical polynucleotides: immunochemical recognition of conformations. Science 169: 609-611

Telenius H, Carter NB, Bebb CE, Nordenskjöld M, Ponder BAJ, Tunnacliff A (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics 13:718-725

Villasante A, de Pablos B, Méndez-Lago M, Abad JP (2008) Telomere maintenance in *Drosophila*: rapid transposon evolution to chromosome ends. Cell Cycle 7:2134-2138

Vítková M, Král J, Traut W, Zrzavˇ J, Marec F (2005) The evolutionary origin of insect telomeric repeats, (TTAGG)n. Chromosome Res 13:145-156

Yokosawa J, Soares MAM, Dijkwel P, Stocker AJ, Hamlin J, Lara, FJS (1999) DNA replication during amplification of the C3 puff of *Rhynchosciara americana* initiates at multiple sites in a 6 Kb region. Chromosoma 108: 291-301

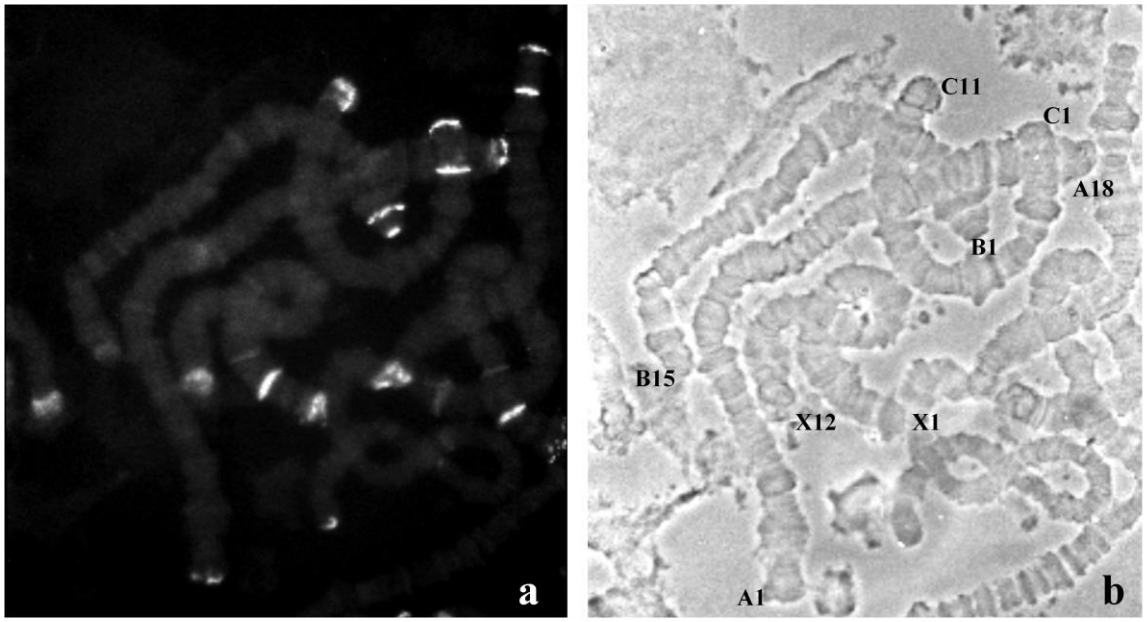
Young BS, Pession A, Traverse KL, French C, Pardue ML (1983). Telomere regions in *Drosophila* share complex DNA sequences with pericentric heterochromatin. Cell 34: 85-94

SUPPLEMENTARY MATERIALS

SM. Table 2 Detection results of repeat probe hybridisation at chromosome ends of *Rhynchosciara* species. Lack of chromosome end labelling is indicated (-). Chromosome ends are identified according Stocker and Gorab 2003. Data were taken from this work (*T-14 repeats*), Madalena and Gorab 2005 (*414 bp repeats*), Rossato et al. 2007 (*M-22 repeats*) and Madalena et al. 2010 (*M-16/M-22 repeats*).

Organism/Probe	T-14 repeats	M-22 repeats	M-16 repeats	414 bp repeats
R. americana	A-1, A-18, B-1, C-1, X-1	A-1, A-18, B-1, C-1, X-1	A-1, A-18, B-1, C-1, X-1	A-1, A-18, B-1, C-1, X-1
R. hollaenderi	A-18, B-1, C-1	A-18, B-1, C-1	A-18	A-18
R. baschanti	-	-	-	-
R. milleri	-	-	-	-

SM Fig. S1 (**a**) Local enrichment with (dA)/(dT) DNA in polytene chromosomes *of R. americana* detected by in situ hybridisation using a poly-(rA) probe. (**b**) The respective phase contrast image



SM Fig. S2 a-c Insert sequences obtained by PCR using DNA from microdissected polytene chromosome tips and primers 19.1 and 19.2 (described in Materials and Methods). Sequences are shown in vertical alignments to facilitate repeat visualisation; divergences in relation to the most frequent base within a column are shadowed. Gaps (-) were introduced to maximise the alignment

a)

```
---------TTATG
TTTANTTTTTTATG
TTTATTTTTTGTTG
TTTATTTTTTTATG
TTTATTTTATTATG
TTTATTTTTTTATG
TTTATTTTTTGTTG
TTTATTTTTTTATG
TTTATTTTTTTATG
TTTATTTTTTTATG
TTTATTTTATATA
```

b)

```
---------TTATG
TTTATTTTTTTATG
TTAATTTTTTTATG
TTTATTTTATTATG
TTTATTTTTTTATG
TTTATTTTTTGTTG
TTTATTTTTTTATG
TTTATTTTTT-ATG
TTTATTTTATATA
```

c)

```
TTAATTTTTTTATG
TTTATTTTTTTATG
TTTATTTTTTTATG
TTTATTTTTTGTTG
TTTATTTTTTTATG
TTTATTTTTT-ATG
TTTATTTTAT
```

SM Fig. S3 BLAST alignments obtained using p*Ra*-55 insert sequences (*Query*) and (*Sbjct*) those within the X chromosome of *D. melanogaster* (accession numbers AE014298.4; CG1462)

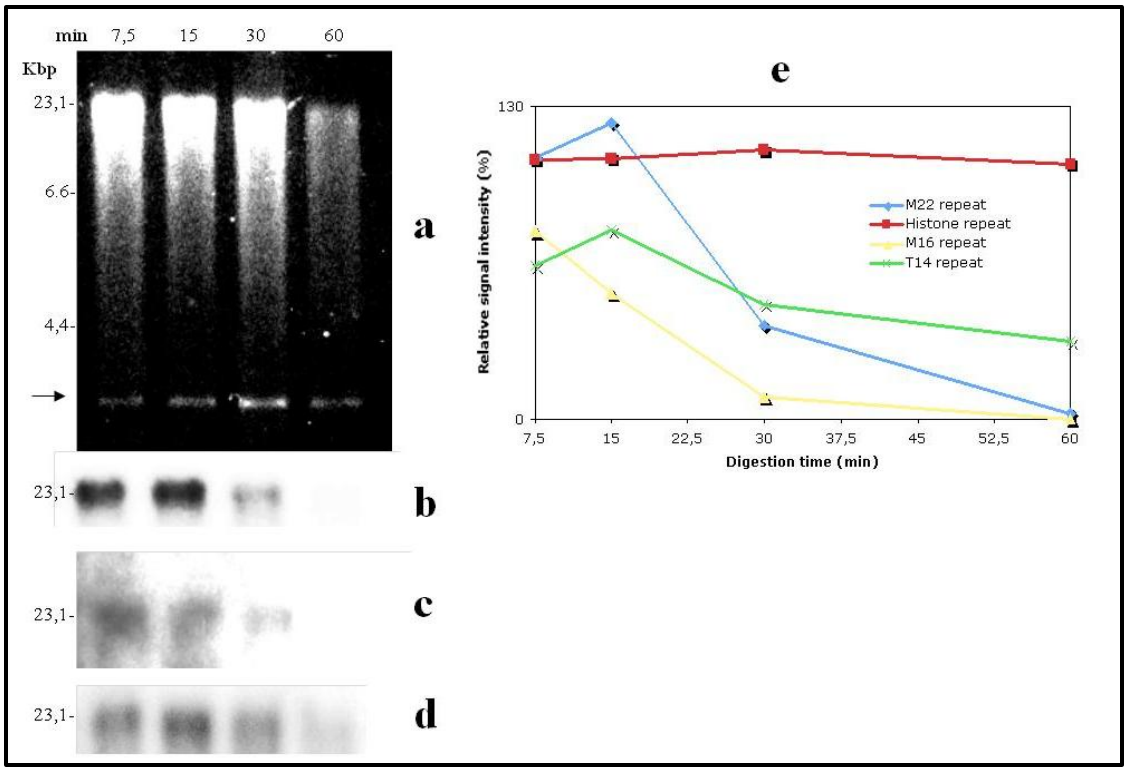
```
gb|AE014298.4| GeoDownload subject sequence spanning the
   HSP Drosophila melanogaster chromosome X, complete sequence
Length=22422827

 Features flanking this part of subject sequence:
   64279 bp at 5' side: DISCO interacting protein 1, isoform D
   281366 bp at 3' side: CG14621

 Score =  127 bits (140),  Expect = 1e-26
 Identities = 95/110 (86%), Gaps = 1/110 (0%)
 Strand=Plus/Minus

Query  2         ACAAATAAAAAAATACAATTAAAAAAATACAAATAAAATAATACAAATAAAAAAATACAA
                 ||||||||||||  |||| |||||||  |||||||||| |  ||||||||||||  ||||
Sbjct  21565451  ACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAA
21565392

Query  62        ATAAAAAACAACAAATAAAAAAATACAAATAAAAAA-TACAAATAAAATA  110
                 |||||||| ||||||||||||  ||||||||||||  ||||||||||||
Sbjct  21565391  ATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAATA  21565342


 Score =  123 bits (136),  Expect = 1e-25
 Identities = 93/108 (86%), Gaps = 1/108 (0%)
 Strand=Plus/Minus

Query  2         ACAAATAAAAAAATACAATTAAAAAAATACAAATAAAATAATACAAATAAAAAAATACAA
                 ||||||||||||  |||| |||||||  |||||||||| |  ||||||||||||  ||||
Sbjct  21565465  ACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAA
21565406

Query  62        ATAAAAAACAACAAATAAAAAAATACAAATAAAAAA-TACAAATAAAA  108
                 |||||||| ||||||||||||  ||||||||||||  ||||||||||
Sbjct  21565405  ATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAA  21565358

 Score =  123 bits (136),  Expect = 1e-25
 Identities = 93/108 (86%), Gaps = 1/108 (0%)
 Strand=Plus/Minus

Query  2         ACAAATAAAAAAATACAATTAAAAAAATACAAATAAAATAATACAAATAAAAAAATACAA
                 ||||||||||||  |||| |||||||  |||||||||| |  ||||||||||||  ||||
Sbjct  21565479  ACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAA
21565420

Query  62        ATAAAAAACAACAAATAAAAAAATACAAATAAAAAA-TACAAATAAAA  108
                 |||||||| ||||||||||||  ||||||||||||  ||||||||||
Sbjct  21565419  ATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAA  21565372


 Score =  123 bits (136),  Expect = 1e-25
 Identities = 93/108 (86%), Gaps = 1/108 (0%)
 Strand=Plus/Minus

Query  2         ACAAATAAAAAAATACAATTAAAAAAATACAAATAAAATAATACAAATAAAAAAATACAA
                 ||||||||||||  |||| |||||||  |||||||||| |  ||||||||||||  ||||
Sbjct  21565493  ACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAA
21565434
```

```
Query  62        ATAAAAAACAACAAATAAAAAAATACAAATAAAAAA-TACAAATAAAA  108
                 ||||||||| ||||||||||  ||||||||||  ||||||||||
Sbjct  21565433  ATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAA  21565386



Score =  123 bits (136),  Expect = 1e-25
 Identities = 93/108 (86%), Gaps = 1/108 (0%)
 Strand=Plus/Minus
Query  2         ACAAATAAAAAAATACAATTAAAAAAATACAAATAAAATAATACAAATAAAAAAATACAA
                 ||||||||||||  |||| |||||||  |||||||||| |  ||||||||||||  ||||
Sbjct  21565507  ACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAA
21565448

Query  62        ATAAAAAACAACAAATAAAAAAATACAAATAAAAAA-TACAAATAAAA  108
                 ||||||||| ||||||||||  ||||||||||  ||||||||||
Sbjct  21565447  ATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAA  21565400


 Score =  123 bits (136),  Expect = 1e-25
 Identities = 93/108 (86%), Gaps = 1/108 (0%)
 Strand=Plus/Minus

Query  2         ACAAATAAAAAAATACAATTAAAAAAATACAAATAAAATAATACAAATAAAAAAATACAA
                 ||||||||||||  |||| |||||||  |||||||||| |  ||||||||||||  ||||
Sbjct  21565521  ACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAA
21565462

Query  62        ATAAAAAACAACAAATAAAAAAATACAAATAAAAAA-TACAAATAAAA  108
                 ||||||||| ||||||||||  ||||||||||  ||||||||||
Sbjct  21565461  ATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAA  21565414


 Score =  123 bits (136),  Expect = 1e-25
 Identities = 93/108 (86%), Gaps = 1/108 (0%)
 Strand=Plus/Minus

Query  2         ACAAATAAAAAAATACAATTAAAAAAATACAAATAAAATAATACAAATAAAAAAATACAA
                 ||||||||||||  |||| |||||||  |||||||||| |  ||||||||||||  ||||
Sbjct  21565535  ACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAA
21565476

Query  62        ATAAAAAACAACAAATAAAAAAATACAAATAAAAAA-TACAAATAAAA  108
                 ||||||||| ||||||||||  ||||||||||  ||||||||||
Sbjct  21565475  ATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAA  21565428


 Score =  118 bits (130),  Expect = 5e-24
 Identities = 92/109 (84%), Gaps = 2/109 (1%)
 Strand=Plus/Plus

Query  2         ACAAATAAAAAAATACAATTAAAAAAATACAAATAAAATAATACAAATAAAAAAATACAA
                 ||||||||||||  |||| |||||||  |||||||||| |  ||||||||||||  ||||
Sbjct  21559983  ACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAA
21560042

Query  62        ATAAAAAACAACAAATAAAAAAATACAAATAAA--AAATACAAATAAAA  108
                 ||||||||| ||||||||||  |||||||||  ||| || |||||||
Sbjct  21560043  ATAAAAAACCACAAATAAAAAACCACAAATAAAATAAAAACCAATAAAA  21560091
```

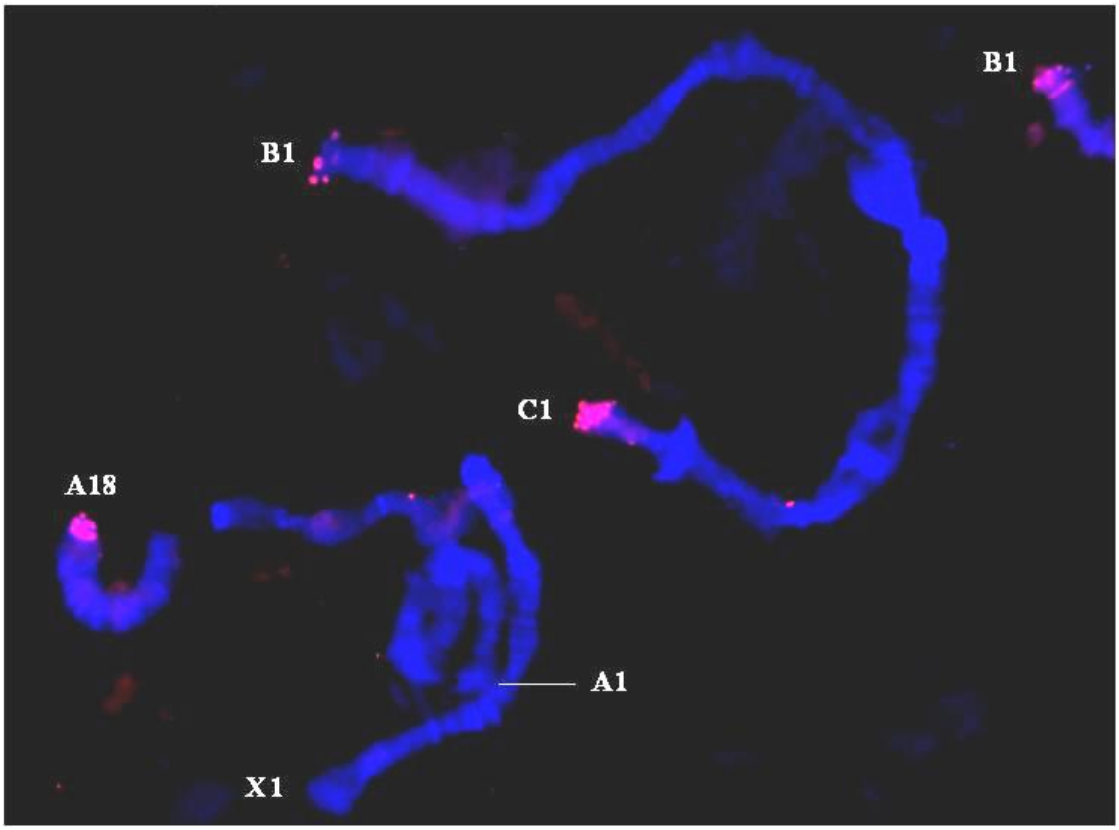
```
 Score =  111 bits (122),  Expect = 8e-22
 Identities = 83/96 (86%), Gaps = 1/96 (1%)
 Strand=Plus/Plus

Query  16        ACAATTAAAAAAATACAAATAAAATAATACAAATAAAAAAATACAAATAAAAAACAACAA
                 |||| |||||||  |||||||||| |  ||||||||||||  |||||||||||| ||||
Sbjct  21559983  ACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAA
21560042


Query  76        ATAAAAAAATACAAATAAAAAAT-ACAAATAAAATA  110
                 ||||||||  ||||||||||||  ||||||||||||
Sbjct  21560043  ATAAAAAACCACAAATAAAAAACCACAAATAAAATA  21560078

 Score =  107 bits (118),  Expect = 1e-20
 Identities = 95/116 (81%), Gaps = 2/116 (1%)
 Strand=Plus/Minus

Query  2         ACAAATAAAAAAATACAATTAAAAAAATACAAATAAAATAATACAAATAAAAAAATACAA
                 ||||||||||||  |||| |||||||  |||||||||| |  ||||||||||||  ||||
Sbjct  21565423  ACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCACAA
21565364

Query  62        ATAAAAAACAACAAATAAAA-AAATACAAATAAAAA-ATACAAATAAAATATTTTT  115
                 ||||||||| |||||||||| ||| || |||||||| ||   ||||| | | ||||
Sbjct  21565363  ATAAAAAACCACAAATAAAATAAAAACCAATAAAAACATTATAATAATAAACTTTT
21565308


 Score =  104 bits (114),  Expect = 1e-19
 Identities = 89/109 (81%), Gaps = 1/109 (0%)
 Strand=Plus/Plus

Query  3         CAAATAAAAAAATACAATTAAAAAAATACAAATAAAATAATACAAATAAAAAAATACAAA
                 ||||||||||   ||| |||||||   ||||||||| |   ||||||||||  |||||
Sbjct  21574921  CAAATAAAAAACCCCAAATAAAAAACCCCAAATAAAAAACCCCAAATAAAAAACCACAAA
21574980

Query  63        TAAAAAACAACAAATAAAAAAATACAAATAAAAAA-TACAAATAAAATA  110
                 |||||||| ||||||||||   ||||||||||   ||||||||||
Sbjct  21574981  TAAAAAACCACAAATAAAAAACCCCAAATAAAAAACCCCAAATAAAATA  21575029


 Score =  100 bits (110),  Expect = 1e-18
 Identities = 87/107 (81%), Gaps = 1/107 (0%)
 Strand=Plus/Plus

Query  3         CAAATAAAAAAATACAATTAAAAAAATACAAATAAAATAATACAAATAAAAAAATACAAA
                 ||||||||||   ||| |||||||   ||||||||| |   ||||||||||   ||||
Sbjct  21574907  CAAATAAAAAACCCCAAATAAAAAACCCCAAATAAAAAACCCCAAATAAAAAACCCCAAA
21574966

Query  63        TAAAAAACAACAAATAAAAAAATACAAATAAAAAA-TACAAATAAAA  108
                 |||||||| ||||||||||||  ||||||||||||   |||||||||
Sbjct  21574967  TAAAAAACCACAAATAAAAAACCACAAATAAAAAACCCCAAATAAAA  21575013
```

```
 Score = 93.3 bits (102),  Expect = 2e-16
 Identities = 91/115 (79%), Gaps = 2/115 (1%)
 Strand=Plus/Plus

Query  3         CAAATAAAAAAATACAATTAAAAAAATACAAATAAAATAATACAAATAAAAAAATACAAA
                 ||||||||||   ||| |||||||  |||||||||| |  ||||||||||   ||||
Sbjct  21574949  CAAATAAAAAACCCCAAATAAAAAACCACAAATAAAAAACCACAAATAAAAAACCCCAAA
21575008

Query  63        TAAAAAACAACAAATAAAA-AAATACAAATAAAAA-ATACAAATAAAATATTTTT  115
                 ||||||||  ||||||||| ||| || |||||||| ||   |||| || | ||||
Sbjct  21575009  TAAAAAACCCCAAATAAAATAAAAACCAATAAAAACATTATAATACAAAACTTTT
21575063
```

SM Fig. S4 (**a**) Electrophoretic analysis of the time course digestion of high molecular weight DNA of *R. americana* with *Bal* 31 exonuclease and subsequent cutting with *Bgl*I; the arrow points to the *Bgl*I restriction fragment of the histone gene (3.344 Kbp, accession number AF 378198). (**b-d**) Southern blot hybridisations to the T-14 (**b**), M-22 (**c**) and M-16 (**d**) repeat probes. (**f**) Densitometric values from *Bal*-31 exonuclease sensitivities of histone, T-14, M-16 and M-22 repeats were plotted in the graph; signal intensities were quantified by scanning the autoradiographs except for the histone gene bands that were taken from the gel image



SM Fig. S5 Localisation of p*Ra*-55 probe (*red signal*) in polytene chromosomes of *R. hollaenderi* counterstained with DAPI (*blue signal*).



CAPÍTULO 3

Macrostructure of chromosome ends in sciarid flies

Thiago Fernandes
Christiane Rodriguez Gutierrez Madalena
Eduardo Gorab*

Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo. Rua do Matão 277, Cidade Universitária, São Paulo, SP, Brazil., CEP: 05508-090. Phone: +55-11-30918061; FAX: +55-11-30917553. Email: egorab@usp.br
*Correspondence

Abbreviations:

bp: base pair
DAPI: 4',6-diamidino-2-phenylindole
dNTP: deoxy-ribonucleotide 5'-triphosphate.
DOP: degenerate oligonucleotide-primed
dUTP: 2'-deoxy-uridine 5'-triphosphate
PCR: polymerase chain reaction
rDNA: ribosomal DNA
SDS: sodium dodecyl sulfate
SSC: sodium saline citrate
TBS: Tris-buffered saline

Acknowledgements: The authors thank Dr. Ann J. Stocker for critical reading of the manuscript and Dr. J.L. Martinez for providing us with chromosome spreads of *C. riparius*. This work was supported by **FAPESP** (Fundação de Amparo à Pesquisa do Estado de São Paulo).

RESUMO

Centrômeros e telômeros são usualmente definidos como regiões heterocromáticas e geralmente compostos por diferentes DNAs repetitivos (satélites e/ou elementos móveis). Visando elucidar aspectos estruturais de conservação e/ou divergência da heterocromatina terminal, três espécies da família Sciaridae foram avaliadas por hibridação *in situ* em cromossomos politênicos: *Rhynchosciara americana*, *Rhynchosciara milleri* e *Trichomegalosphys pubescens*. O DNA utilizado como sonda foi obtido por microdissecção de extremidades cromossômicas não-centroméricas dessas espécies, exceto *R. milleri*, seguido de amplificação e marcação por DOP-PCR. Quando cada sonda foi hibridada nos cromossomos que lhe deram origem dois padrões foram observados: i) hibridação em todas as extremidades não-centroméricas (*R. americana*) e, o mais surpreendente, ii) apenas na extremidade que deu origem ao produto de microdissecção (*T. pubescens*). Sondas obtidas de *R. americana* não produziram sinais de hibridação em cromossomos de *T. pubescens* e *R. milleri*. Resultados inesperados foram observados quando a sonda obtida de *T. pubescens* foi hibridada nas espécies *R. americana* e *R. milleri*. Surpreendentemente, essa sonda que marcou apenas uma extremidade no seu próprio complemento cromossômico, gerou sinais de hibridação em todas as extremidades cromossômicas não-centroméricas de *R. americana* e também na heterocromatina centromérica-pericentromérica desta espécie. Em *R. milleri*, a sonda de *T. pubescens* hibridou claramente com a heterocromatina associada ao centrômero do seu cromossomo C. A análise em conjunto dos resultados sugere que a ausência de sinais de hibridação em certas regiões heterocromáticas não se deva a ausência de sequências complementares à sonda, mas sim à diminuição significativa destas sequências na composição de cada região heterocromática. Evidências de sequências comuns entre a heterocromatina terminal de *T. pubescens* e *R. americana* foram obtida após exploração, por sequenciamento, de uma microbiblioteca da extremidade cromossômica X1 de *T. pubescens*. O inserto de um subclone desta microbiblioteca mostrou alta similaridade com uma sequência de *R. americana* previamente caracterizada e proveniente de microdissecção cromossômica, reforçando uma possível amplificação diferencial de sequências repetitivas na heterocromatina terminal das espécies estudadas. No caso mais extremo, sequências associadas a uma extremidade não-centromérica de *T. pubescens* parecem estar significativamente representadas na heterocromatina associada ao centrômero do cromossomo C de *R. milleri*, embora pareça não estar representada nas extremidades não-centroméricas desta espécie. Os dados obtidos neste trabalho apóiam um processo de divergência cromossômica entre estas três espécies de sciarídeos ocorrido pela amplificação diferencial de sequências componentes da heterocromatina e apontam para

uma estrutura não usual em *T. pubescens* comparada a outros dípteros estudados quanto à estrutura cromossômica terminal.

Palavras-chaves: extremidades cromossômicas, sondas obtidas por microdissecção cromossômica, heterocromatina, sciarídeos.

ABSTRACT

The genomic localisation of sequences composing chromosome ends of sciarid species (suborder Nematocera) was studied by means of in situ hybridisation with homologous and heterologous probes synthesised with DNA from microdissected polytene chromosome tips. Data obtained with homologous probes showed that sequences common to non-centromeric ends, pericentric and interstitial regions were detected in all the species analysed. When non-centromeric end labelling was particularly examined, homologous and heterologous probe detection was observed at all non-centromeric ends of *Rhynchosciara americana* chromosomes. However, in situ hybridisation of either homologous or heterologous probes to the chromosomes of *Trichomegalosphys pubescens* only produced significant labelling at one chromosome end of this species. These results together with data from the exploitation of a chromosome end microlibrary argue for the occurrence of an unusual structure in *T. pubescens* compared to other dipterans studied with a focus on chromosome ends.

Key words: chromosome ends, microdissected polytene chromosome probe, homologous and heterologous probe, sciarid species.

1 INTRODUCTION

The widespread occurrence of telomeres that are regenerated by telomerase suggests, at first glance, a broadly conserved chromosome end structure. However, as has emerged from studies of a few representatives of Diptera, organisms of this Order represents exceptions to the above assumption and also display structural diversity with regard to chromosome ends which comprise telomeres and sub-telomeres. In *Drosophila* species, telomeres are composed of specific retrotransposons (reviewed in Mason et al. 2008; Villasante et al. 2008) and in *D. virilis* complex terminal satellites were identified at chromosome ends (Biessmann et al. 2000). Among lower dipterans (Nematocera), tandem arrays of complex repeats have been characterised at telomeres in *Chironomus* species (Carmona et al. 1985; Saiga and Edström 1985; Nielsen et al. 1990; López et al. 1996; Rosén and Edström 2000; Martínez et al. 2001), and in *Anopheles gambiae* (Biessmann et al. 1996; Biessmann et al. 1998).

In the nematoceran *Rhynchosciara americana*, apparently long homopolymer dA/dT tracts (Jones et al. 1973; Stocker et al. 1993; Stocker and Gorab 2003; Gorab 2003; Gorab et al. 2009) and reverse transcriptase-related proteins are regularly associated with its non-centromeric end subset (Gorab 2003). These chromosome end components are uncommon in Diptera and led us to initiate the characterisation of terminal sequences in *R. americana*. We constructed a plasmid library from a single microdissected chromosome end and subsequently characterised a 414 base pair (bp) repeat, enriched at five non-centromeric ends, resembling nematoceran telomeric DNA (Madalena and Gorab

2005). A second repeat type named M-22 was also eventually identified in the chromosome end structure of *Rhynchosciara*. Unusually short for a repeat tandemly arrayed at the chromosome ends of dipterans, additional data suggested that M-22 tandem repeats lie distal to the 414 bp repeat arrays (Rossato et al. 2007).

Continued investigation of the chromosome ends of *R. americana* identified another short tandem repeat type, named M-16, composed of 16 nucleotides and frequently intermingled within M-22 arrays. The occurrence of segmental duplications, as deduced by sequence analyses involving islands of M-22 and M-16 repeats that appear to reach chromosome ends, might have implications for the process of chromosome end maintenance in this species (Madalena et al. 2010b).

The data that have been obtained up to this time by analysing distinct sciarid species suggest the differential occupation of repetitive sequences at chromosome ends during the evolution of sciarid genomes (Stocker and Gorab 2003; Madalena and Gorab 2005; Rossato et al. 2007; Madalena et al. 2010b). However, comparative studies have previously been restricted to one genus, *Rhynchosciara*. Although useful, such an approach has not provided a general view of the chromosome end structure in other representatives of Sciaridae. This could be important since some sciarid species currently being studied in the laboratory display polytene chromosomes with a structure differing from that of *Rhynchosciara* species. For this reason, the experimental design utilised in situ hybridisation with probes synthesized with DNA from microdissected chromosome tips on polytene chromosomes from distantly related sciarid species. Also a plasmid microlibrary from *Trichomegalosphys pubescens*, studied for the first time in this work, was analysed.

The results obtained with homologous probes in the species analysed showed that non-centromeric ends and pericentric heterochromatin share common sequences. However, probe detection did not occur always at all non-centromeric ends. In addition, the results make *Trichomegalosphys pubescens* a candidate for a divergent structure concerning dipteran chromosome ends. Results of heterologous probe hybridisation in addition allowed the establishment of sequence relationships between pericentric heterochromatin and non-centromeric chromosome ends or even between non-centromeric ends of distantly related sciarid species.

2 MATERIALS AND METHODS

2.1 Animals

Larvae of *Rhynchosciara americana* and *R. milleri* were collected in the region of Mongaguá, state of São Paulo. Larvae were kept in the laboratory at 18-22^{o}C. *Trichomegalosphys pubescens* (Diptera: Sciaridae) larvae were collected either on the campus of the University of São Paulo or in the region of Mogi das Cruzes, state of São Paulo. This species was formerly named *Trichomegalosphys pubescens* (see Amorim 1992). Laboratory cultures were kept at 18-22^{o}C.

2.2 Preparation of chromosome spreads

The salivary glands were dissected in Ringer and briefly fixed in ethanol-acetic acid (3:1). After squashing the salivary glands in 50% acetic acid, the slides were frozen on dry ice for 10 min. The coverslips were pried off with a razor blade and the slides were then kept in absolute ethanol at –20^{o}C until the microdissection or hybridisation procedures. Chromosome spreads of *Chironomus riparius* (Diptera: Chironomidae were prepared as described above and were kindly sent by Dr. J.L. Martinez (UNED, Madrid, Spain). This species was formerly named *Chironomus thummi thummi*; see Michailova 1989).

2.3 Microdissection

Chromosome ends were microdissected with the aid of microneedles made from solid glass tubes with the aid of a microforge (De Fonbrune). The microdissection was carried out on a Zeiss Axiovert S100 inverted microscope coupled

with a manual micromanipulator (Narishige). The microdissected fragment was recovered from air-dried slides with the tip of a microneedle that was subsequently broken into a PCR tube.

2.4 DOP-PCR

The DNA in the chromosome fragment was amplified by polymerase chain reaction (PCR) using degenerate oligonucleotides (DOP-PCR, Telenius et al. 1992) in a thermocycler (Mastercycler, Eppendorf), omitting topoisomerase and proteinase K treatments. The following reaction mixture was added to the PCR tube: 0.2 mM of each dNTP; 1X Taq buffer; 2.5 U Taq DNA polymerase (Invitrogen); 1 mM $MgCl_2$; 75 pM DOP primer, H_2O to a final volume of 50 μl. Prior to the amplification rounds, the mixture was briefly centrifuged. Cycle conditions for the first amplification were 93°C for 4 min followed by 8 cycles of 94°C for 1 min, 30°C for 90 s and 72°C for 3 min. The first amplification was completed with additional 28 cycles of 94°C for 1 min, 56°C for 1 min, 72°C for 3 min. For the second amplification round, 5 μl from the volume of the first amplification reaction were added to 50 μl of a new reaction mixture exactly as described above. Cycle conditions were 93°C for 4 min followed by 26 cycles of 94°C for 1 min, 56°C for 1 min, 72°C for 3 min and a final extension at 72°C for 5 min. The amplification products were checked in agarose gels before cloning or labelling procedures.

2.5 Cloning and sequencing of PCR products

The purified PCR products were ligated in pGEM-T Easy vector according to standard protocols (Promega). Procedures for transformation and plating of *E.*

coli DH5α strain were described elsewhere (Sambrook et al. 1989). Individual colonies were randomly chosen to grow in liquid medium for plasmid DNA extraction using Concert Rapid Plasmid Miniprep System (Invitrogen). Plasmid DNAs were then cut with *Eco*RI, the digests run in agarose gels and analysed in Southern-blot hybridisation with genomic DNA of *T. pubescens* as a probe. Sequencing reactions of both strands of plasmid DNAs were performed with Big Dye Terminator Cycle Sequencing Ready Reaction as recommended by the manufacturer (Applied Biosystems) using M13 Forward or Reverse primers (Stratagene) and subsequently run in the ABI PRISM 310 Genetic Analyser (Applied Biosystems). GenBank searches were done with BLAST (Altschul et al. 1990).

2.6 Non radioactive labelling of the probes

Probe labelling was done either with 0.1 mM biotinylated-11 dUTP (Sigma) replacing dTTP in 50 μl of PCR reaction mixture as described previously for the second amplification round of DOP-PCR using 2 μl of total DOP PCR products. Plasmid inserts were labelled replacing DOPs by M-13 Forward and Reverse primers (Invitrogen).

2.7 In situ hybridisation

The probe mixture consisted of labelling reaction product (50 μl), 40% formamide, 2X SSC and 0.1% SDS to a final volume of 110 μl. The probe mixture (5-10 μl) was applied on each air-dried slide and covered with a plastic coverslip. The slides were steam heated at 75^{o}C for 5-10 min and immediately kept in a closed box. Hybridisation was carried out overnight at either 32^{o}C or 37^{o}C. The slides were then washed in 0.5X SSC, 0.2% SDS at either 32^{o}C or 37^{o}C for 30 min, followed by incubation at room temperature

in 1X TBS, 0.1% Triton X-100 (TBST) for 10-20 min. For fluorescent detection, goat IgG anti-biotin (Sigma) diluted 1:50 in TBST was applied on the slides for 1 h at room temperature followed by washes in TBST and incubation for 45 min with rhodamine-labelled rabbit IgG anti-goat (Sigma) diluted 1:100 in TBST. Counterstaining was done with 4',6-diamidino-2-phenylindole (DAPI). The slides were then washed in TBST for 10 min and finally in 1X TBS for 5 min. The slides were mounted in antifading medium (Vectashield, Vector Labs) and inspected under a Nikon Eclipse 80i microscope equipped with epifluorescence optics. Images were captured using the NIS-Elements software package (Nikon).

3 RESULTS

3.1 Localisation of DNA sequences from the B-1 tip of *Rhynchosciara americana* in *R. americana* chromosomes

Three of the four polytene chromosomes of *R. americana* (B, C and X) display large proximal heterocromatin blocks within sections B-15, C-11 and X-12. The precise location of centromeres within such regions is not yet known. Thus, polytene chromosomes of this species have five non-centromeric ends, corresponding to the sections A-1, A-18, B-1, C-1 and X-1, and three centromeric or pericentric heterochromatic ends (SM, Fig. S1; see also chromosome maps in Stocker et al. 1993; Madalena et al. 2010b). The results were obtained with a probe synthesised with DNA from a single B-1 polytene chromosome tip. Consistent hybridisation signals were seen in all telomeres but with differences in signal intensity at certain chromosome ends. A-18, B-1 and C-1 tips always appeared more intensely labelled than A-1 and X-1 tips. Signals were also observed at centromeric ends, at pericentric regions of all chromosomes and at several interstitial sites (Fig. 1).

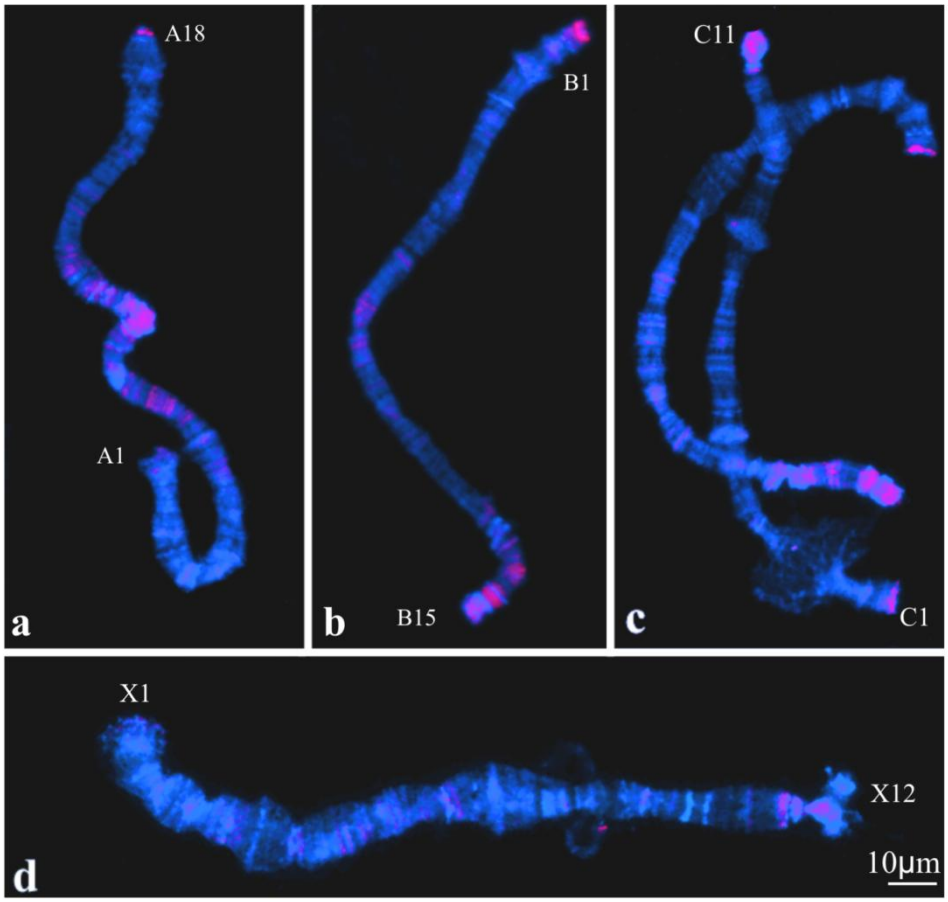


Figure 1: a-d Localisation of the probe (*red signal*) synthesised with DNA from a microdissected B-1 chromosome end of *R. americana* in polytene chromosomes **a** A, **b** B, **c** C and **d** X of *R. americana* counterstained with DAPI (*blue signal*). Chromosome end sections were identified according to Stocker et al. 1993.

3.2 Localisation of DNA sequences from chromosome tips of *Trichomegalosphys pubescens* in *T. pubescens* chromosomes

Although this sciarid species also has four polytene chromosomes, significant differences in their structure can be observed in relation to *Rhynchosciara*. Chromosomes of *T. pubescens* lack centromeric or pericentric ends as well as the conspicuous heterochromatic blocks associated with these regions as seen in *Rhynchosciara* species. Polytene pericentric regions of *T. pubescens* are identified by chromosome breakpoints as a probable result of DNA under-replication (SM, Fig. S2; see also mitotic chromosomes and chromosome maps in Amabis 1974; Amabis 1983). In *T. pubescens*, ribosomal DNA (rDNA) probes hybridised at the mid portion of the chromosome X, very close to the breakpoint of the polytene X chromosome, while in *Rhynchosciara* species hybridisation signals are restricted to the proximal end of the mitotic X chromosome (Madalena et al. 2007). These data showed that the pericentric heterochromatin occupies distinct positions in the chromosome X of *Rhynchosciara* and *T. pubescens*.

A probe synthesised from a single X-1 tip of *T. pubescens* was used first for in situ hybridisation. In contrast to the results obtained in *R. americana* chromosomes with the B-1 probe, no hybridisation signal was seen at tips other than X-1. Labelling was also observed in regions coinciding with chromosome breakpoints and at interstitial sites (Fig. 2). The results led us to repeat the procedures assuming that technical problems might have occurred. Five additional microdissections of different chromosome tips followed by DOP-PCR and in situ hybridisation were carried out. The results obtained with probes from

distinct chromosome ends reproduced the same hybridisation pattern as shown in Fig. 2, namely, there is a main signal corresponding to the microdissected end but no significant labelling was seen at other chromosome termini (SM, Fig. S3). The possible requirement of DNA from more than one microdissected chromosome end was considered as all the *T. pubescens* probes were made with a single chromosome tip. Use of a probe made with two microdissected ends of *T. pubescens* chromosomes resulted in labelling at chromosome ends restricted to the microdissected tips (data not shown).

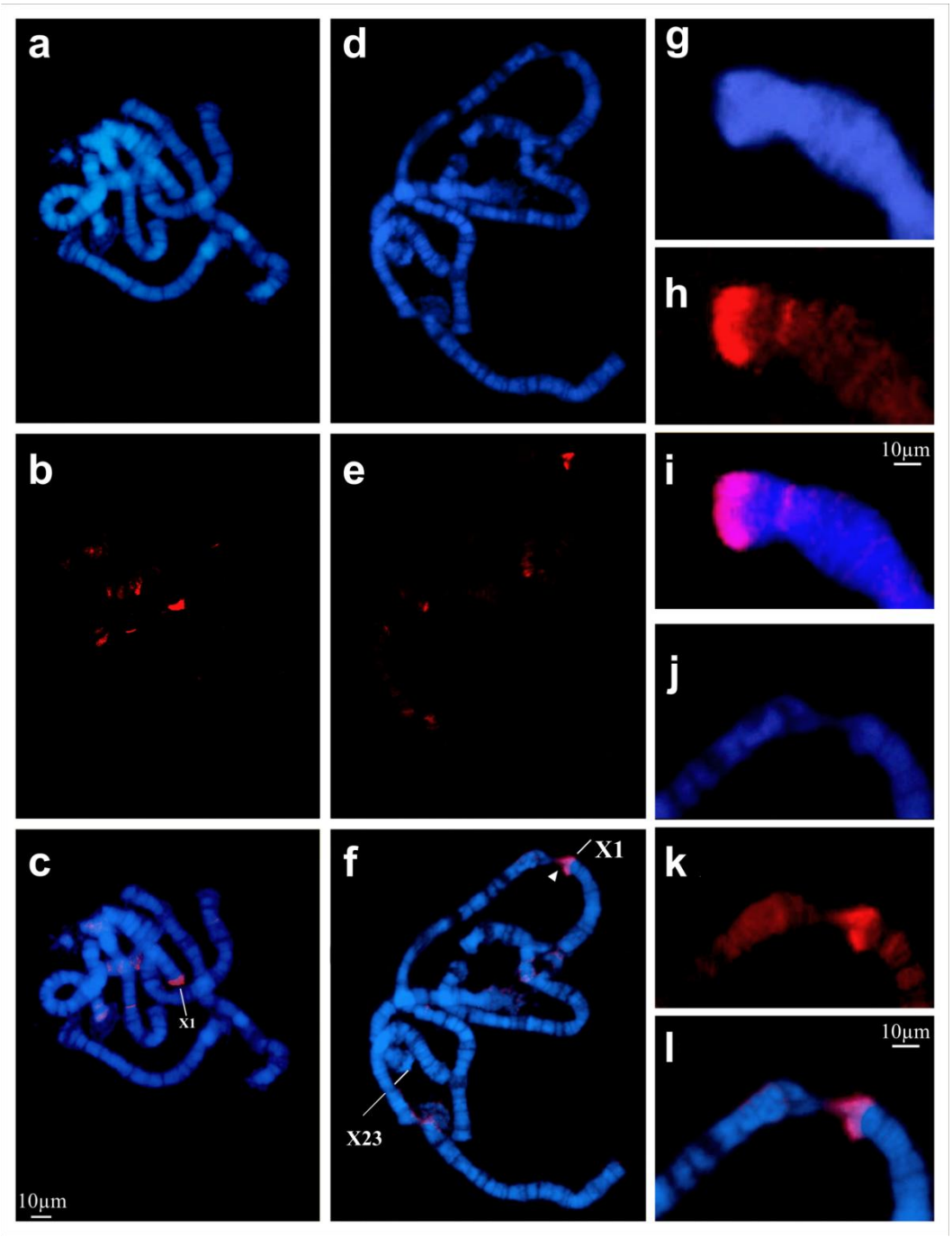


Figure 2: a-d Polytene chromosomes of *T. pubescens* counterstained with DAPI (*blue signal*) after in situ hybridisation (*red signal*) with the probe made with DNA from the microdissected X-1 end of *T. pubescens* showing **c, f** the merged signals; **g-i** detail of the X-1 chromosome end of *T. pubescens* **g** counterstained with DAPI showing **h** hybridisation signals encompassing sub-telomeric sections and **i** the merged signals; **j-l** detail of the X-1 end showed in d-f with hybridisation signals in a telomere bridge; the identification of chromosome ends was done according to Amabis 1983.

3.3 Localisation of heterologous tip probes in *T. pubescens* and *R. americana* chromosomes

The B-1 tip probe of *R. americana* did not produce labelling in polytene chromosomes of *T. pubescens* when hybridisation and washes were performed at 37°C. However, when the temperature was lowered to 32°C, consistent signals were detected, particularly at the B-25 ends. Less intense staining was also observed at interstitial regions of the chromosomes of this species (Fig. 3). These results raised the possibility that cloned repeats from *R. americana* chromosome ends could produce hybridisation signals in *T. pubescens* chromosomes under low stringent hybridisation and washings. Nevertheless, *R. americana* probes representing subtelomeric DNA (Madalena et al. 2005) and terminal repeats localised distally (Rossato et al. 2007; Madalena et al. 2010b) did not produce hybridisation signals in *T. pubescens* chromosomes.

The preferential localisation of sequences from the B-1 chromosome terminus of *R. americana* at the B-25 end of *T. pubescens* suggested on the other hand that probes from chromosome tips of *T. pubescens* could produce hybridisation results in *R. americana* chromosomes. This assumption was confirmed and the probe made with A-1 tip DNA of *T. pubescens* was clearly localised at all non-centromeric ends of *R. americana*. Labelling was also observed in the pericentric heterochromatin and interstitial bands of the two species in all the experiments performed (Fig. 4). Results obtained with the probe from the X-1 end of *T. pubescens* on chromosomes of *R. americana* resembled those seen with the A-1 tip probe of *T. pubescens*. Differences were found in labelling intensity at some chromosome tips, in the pericentric heterochromatin and also for the absence of hybridisation signals at certain interstitial regions (SM, Fig. S4).

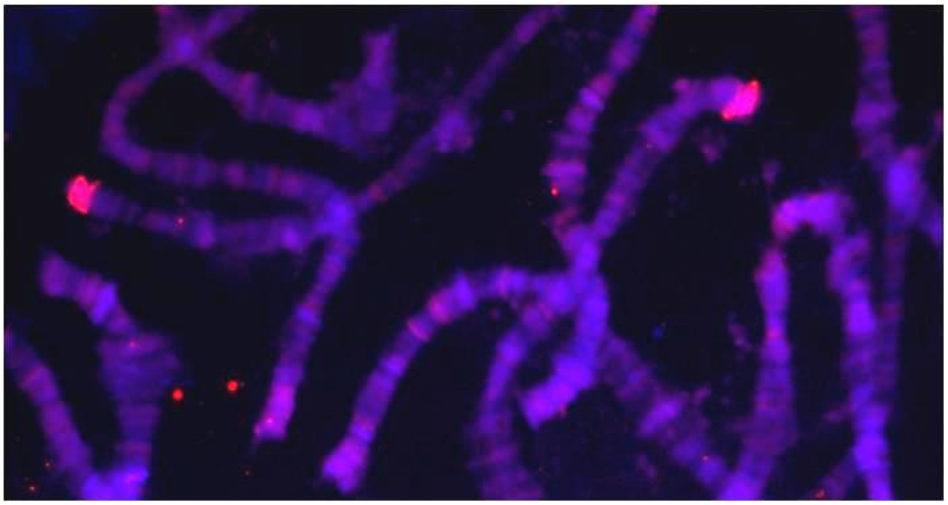


Figure 3: Localisation of the probe (*red signal*) synthesised with DNA from the microdissected B-1 tip of *R. americana* in chromosomes of *T. pubescens* counterstained with DAPI (*blue signal*).

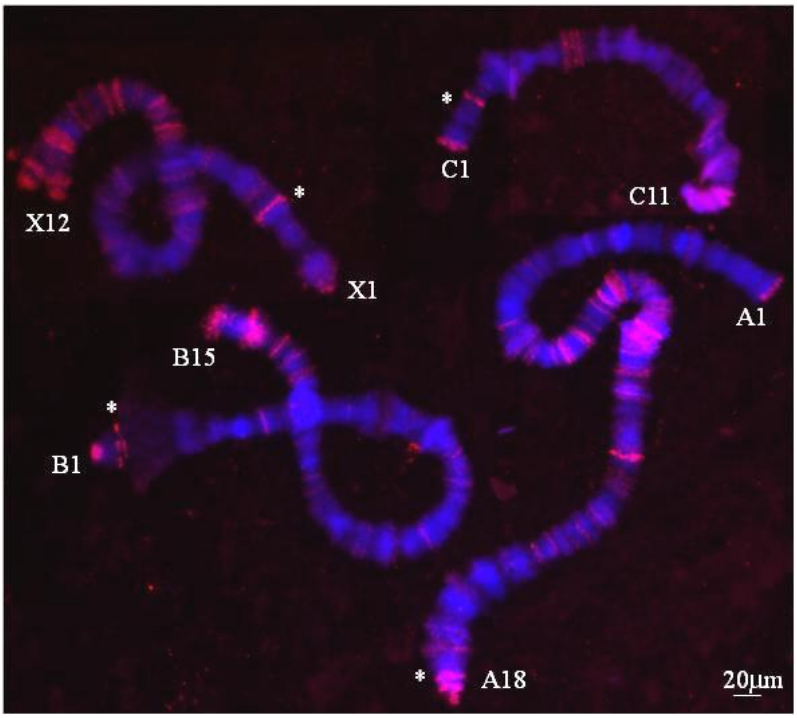


Figure 4: Localisation of the probe (*red signal*) synthesised with DNA from a microdissected A-1 chromosome end of *T. pubescens* in polytene chromosomes of *R. americana* counterstained with DAPI (*blue signal*). The *asterisks* indicate labelling of certain chromosome regions that did not appear stained with the probe made with DNA from a microdissected X-1 tip of *T. pubescens* (see SM Fig. S4).

3.4 Localisation of sequences from *R. americana* and *T. pubescens* chromosome tips in *Rhynchosciara milleri* chromosomes

Studies of adult representatives of Sciaridae place *Rhynchosciara milleri* in the *milleri* group while *R. americana* belong to the *americana* group (Breuer 1969). Polytene and mitotic chromosomes of *R. milleri* have been described (Pavan and Breuer 1955) and further polytene band examination (Stocker, unpublished data) indicated that this species is distantly related to those included in the *americana* group. Salivary gland chromosomes of *R. milleri* usually form a chromocentre and, as observed in *R. americana*, heterochromatic blocks are found in centromeric ends of the *R. milleri* chromosomes B, C and X (SM, Fig. S5). The rDNA in this species is localised in the pericentric heterochromatin and in polytene chromosome ends but restricted to the proximal end of the mitotic X chromosome (Madalena et al. 2007). Chromosome tip probes of *R. milleri* would obviously be important for this work. However, micronucleoli containing rDNA are usually attached to all the non-centromeric polytene ends of this species (Madalena et al. 2007) and would contaminate the probe with extrachromosomal rDNA that is not of telomeric origin. For this reason, probes synthesised from chromosome tips of other sciarid species were employed for in situ hybridisation in chromosomes of *R. milleri*.

Hybridisation results on chromosomes of *R. milleri* were observed with probes synthesised from chromosome ends of *T. pubescens* and *R. americana*. When the temperature was lowered to 32^{o}C (for hybridisation and washes), both probes produced labelling over many regions. Some bands close to the pericentric heterochromatin appeared

more intensely labelled. The B-1 probe from *R. americana* also hybridised to the pericentric heterochromatin as well as to the telomeres although no significantly strong labelling was seen at these regions when compared with the fluorescence observed at other chromosomal sites (SM, Fig. S6). However when in situ hybridisation and washes were performed at 37^{o}C, significant signals were only obtained with the X-1 probe of *T. pubescens* and restricted to the pericentric heterochromatin of *R. milleri* (Fig. 5).

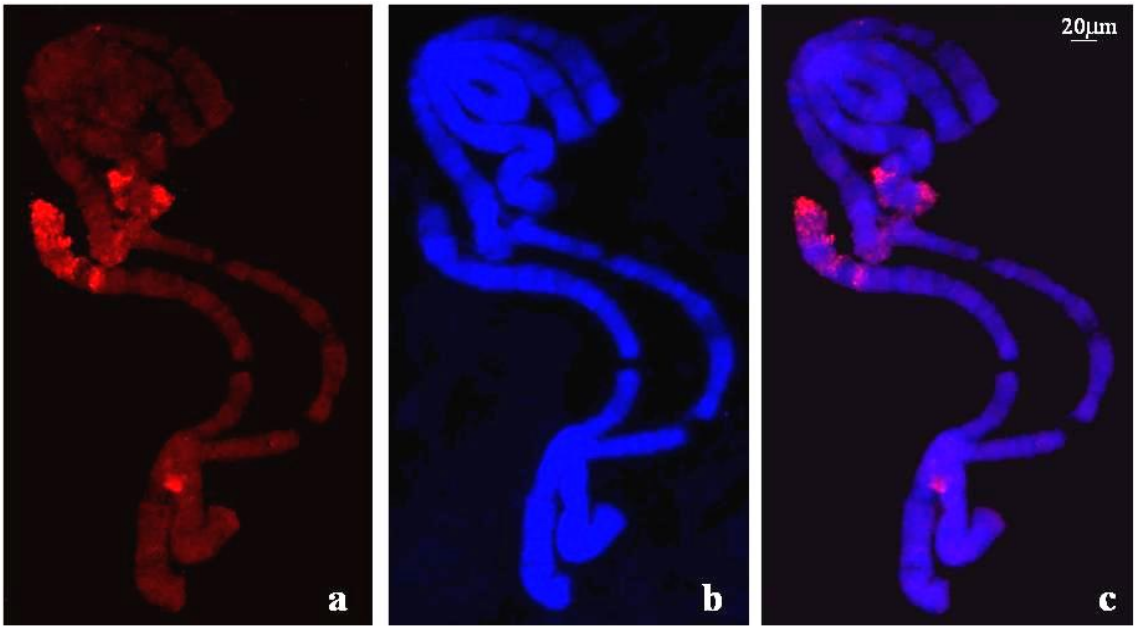


Figure 6: a-c Localisation of the probe (**a,** *red signal*) synthesised with DNA from the microdissected X-1 end of *T. pubescens* in chromosomes of *R. milleri* showing **b** DAPI counterstaining (*blue signal*) and **c** the merged signals; hybridisation and washings were performed at 37^{o}C.

Cloned repeats from *R. americana* chromosome ends used as probes for in situ hybridisation were not detected in chromosomes of *R. milleri* even under low stringent hybridisation and washings. Also, sequences enriched with homopolymeric dA/dT DNA at chromosome ends of *R. americana* (Stocker et al. 1993; Stocker and Gorab 2003; Gorab 2003; Gorab et al. 2009) were not detected by in situ hybridisation in terminal regions of *R. milleri* chromosomes.

3.5 Controls

Microdissection procedures involving chromosomes of distinct species were separated by long time intervals in order to minimise the possibility of contamination and always performed with sterilised material. When the results led us to suspect that either contamination or another methodological problem had occurred, additional microdissections were carried out using a new complete reagent set. The results of in situ hybridisation presented in this work were reproduced with a large number of slides.

A control was introduced in this work by performing the above method with chromosomes of *Chironomus riparius*, a nematoceran which has been extensively studied with a focus on the telomere. Microdissection of the largest telomere puff (TBRIII) followed by direct DNA cloning allowed the identification of telomeric repeats in *C. riparius*, localised in seven of the eight polytene chromosome ends (reviewed in Díez et al. 2009). The efficacy of the method employed in this work was therefore assessed by microdissecting a single TBRIII of *C. riparius* for use in probe synthesis by DOP-PCR. Probe hybridisation to seven non-centromeric ends of this species is expected in the case of

successful amplification of telomeric repeats. The ability of the procedures to amplify sequences composing chromosome ends was confirmed since the probe synthesised from a single TBRIII DNA hybridised to the seven non-centromeric polytene ends of *C. riparius*. In addition, interstitial bands and also some centromeric regions appeared labelled, probably indicating the localisation of subtelomeric sequences microdissected together with TBRIII DNA (Fig. 6).

3.6 Screening of a plasmid library constructed with X-1 tip DNA of *T. pubescens*

DOP-PCR performed with DNA from the X-1 tip of *T. pubescens* was used for the construction of a plasmid library. Results from the screening were expected to provide information toward explaining the in situ hybridisation results observed in this species with probes from their own chromosomes. Plasmid DNA from 104 colonies was digested with *Eco*RI and analysed by Southern-blot hybridisation using salivary gland DNA as the probe. Controls included rDNA inserts as references for moderately repetitive sequences. No signal indicating the existence of cloned inserts with highly repetitive sequences was seen. Inserts were randomly chosen for sequencing and most results showed no significant similarity with sequences deposited in data banks except for two inserts (SM, Fig. S7). The p*Tp*-29 insert is 67% similar to a stretch identified in the genomic clone named "RaTART" analysed in detail previously (Madalena et al. 2010a), which in turn displays high identity with sequences encoding reverse transcriptases of insect mobile elements. The second insert (p*Tp*-36) is 97% similar to that found in the p*Ra*-43 insert from the chromosome end of *R. americana* characterised in a previous work (Rossato et al. 2007). Although the size and sequences of the two above inserts are not exactly identical,

the similarity observed could suggest microlibrary contamination. However, this possibility is unlikely as the two libraries were constructed with a time separation of five years and have never been manipulated simultaneously.

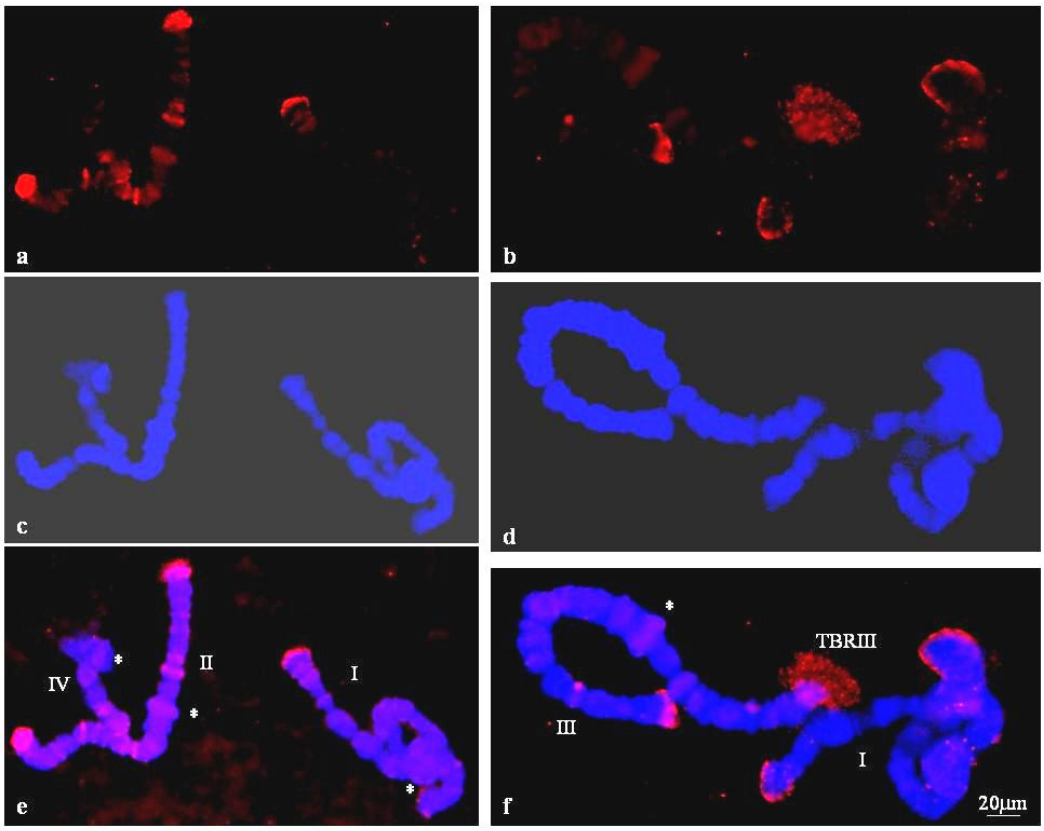


Figure 6: a, b Localisation of the probe (*red signal*) synthesised with DNA from a microdissected telomeric puff (*TBRIII*) of *Chironomus riparius* in polytene chromosomes of this species from a larva that underwent heat shock. **c**, **d** Chromosomes were counterstained with DAPI (*blue signal*) showing **e**, **f** the merged hybridisation and DAPI signals. The *asterisks* indicate the centromeric regions of the chromosomes identified individually (*I, II, III, IV*)

4 DISCUSSION

Results obtained with homologous and heterologous tip probes in *R. americana* chromosomes showed that sequences are present at non-centromeric ends and also in pericentric and interstitial regions. Although cloned terminal and sub-terminal repeats of *R. americana* are not enriched at centromeric ends (Madalena and Gorab 2005; Rossato et al. 2007; Madalena et al. 2010b), these results can be expected if, in addition to tandem repeats, chromosome ends are composed of dispersed repeats and mobile elements. Transposable elements were found to be localised in subtelomeres, pericentric heterochromatin and interstitial regions of *R. americana* chromosomes (Madalena et al. 2010a; Madalena et al. 2010b). Detection of heterologous probe hybridisation in chromosome ends of *R. americana* argues for sequence conservation within these regions of *R. americana* and *T. pubescens* chromosomes. These sequences appear amplified in the genome of *R. americana* since consistent probe labelling occurred in the non-centromeric ends this species while, in *T. pubescens*, signals from the same probe were only seen in a single chromosome tip.

Microdissection followed by DOP-PCR has not been currently used in dipteran telomere research. We employed this method in the past and it allowed the characterisation of *R. americana* repeats enriched at five non-centromeric ends of this species (Madalena and Gorab 2005; Rossato et al. 2007). The little amount of data that has been obtained with this technique might lead to a suspicion that it cannot amplify microdissected terminal DNA of *T. pubescens*. This is unlikely since the TBRIII control

probe, made for the first time by DOP-PCR from a single terminal puff of *C. riparius*, hybridised to all the non-centromeric ends of this species. It is worth mentioning that the first evidence for telomeres with distinct molecular structure within the chromosome complement of a given species came from research in *Chironomus*. Telomeric puffing induced by heat shock is a rare phenomenon that has been observed in polytene chromosomes of *C. riparius* and triggered the study of *Chironomus* telomeres. Data obtained with cloned repeats from chironomid species showed that centromeric ends have centromeric sequences instead telomeric repeats (Carmona et al. 1985; Saiga and Edström 1985; Nielsen et al. 1990; López et al. 1996; Rosén and Edström 2000; Martínez et al. 2001; Rosén et al. 2002).

Probes from *T. pubescens* chromosome tips produced significant hybridisation signals restricted to a single chromosome end of this species and provided the first indication of an unusual structure at chromosome termini of this species. In addition, hybridisation results were also obtained with the B-1 tip probe of a distantly related sciarid, *R. americana*, at the B-25 end of *T. pubescens*. The data represented further support for the peculiar structure of chromosome ends in *T. pubescens*, since homologous and heterologous probes used separately labelled one chromosome end of this species. If chromosome ends of *T. pubescens* have the same telomeric sequences as expected for any eukaryotic organism, one may conclude that they are not reflected in the results shown in this work. If the hybridisation signals do not come from telomeric DNA, they would be representative of sub-telomeric DNA specific for each chromosome end of *T. pubescens* and conserved at a significant degree in one sub-telomere (B-1) of *R. americana*. These features are not expected considering the evolutionary distance between *R. americana* and

T. pubescens and also the rapid evolutionary changes usually observed in telomeric and sub-terminal chromosome regions.

The screening of the plasmid microlibrary constructed with DNA from the X-1 end of *T. pubescens* did not show inserts representing highly repetitive sequences. A comparison with data from the screening of the *R. americana* library made in the past with B-1 tip DNA may give clues about differences in the chromosome end structure of the two sciarid species. Three inserts displaying a high degree of reiteration in the genome of *R. americana* plus one representing moderately repetitive DNA (p*Ra*-43) were found within 54 inserts chosen at random from the B-1 microlibrary (Madalena and Gorab 2005; our unpublished data). On the other hand, more than 100 inserts chosen by chance from the *T. pubescens* microlibrary were screened and only two representing middle repetitive DNA were identified. The data from the two screenings argue that the representation of highly repetitive DNA at chromosome ends of *T. pubescens* is significantly lower than that in terminal regions of *R. americana* chromosomes. Such an assumption is also in agreement with in situ hybridisation results obtained in this work and gives support to the idea of a different telomere/sub-telomere structure in *T. pubescens* compared to those that have been characterised in dipterans up to now.

BLAST searches identified two inserts in the *T. pubescens* microlibrary displaying significant similarity with *R. americana* sequences deposited in databases. The p*Tp*- 36 insert is 97% similar to the p*Ra*-43 (Rossato et al. 2007) which also came from a microlibrary constructed with DNA from microdissected chromosome tips. Sequencing and in situ hybridisation data thus support the occurrence of common genetic elements in

chromosome ends of the two species. The p*Tp*-29 insert aligns at a specific region of the composite genomic clone named "RaTART" analysed previously in detail (Madalena et al. 2010a) showing 67% sequence similarity. Conceptual translation of p*Tp*-29 insert indicated significant identity to reverse transcriptases from insect retrotransposons. Sequences related to retrotranposons inserted in chromosome ends could be involved in telomere maintenance, as extensively characterised in *Drosophila* (reviewed in Mason et al. 2008; Villasante et al. 2008). This hypothesis is unlikely in *T. pubescens* as other sequence stretches from this retrotransposon may have been amplified during DOP-PCR, assuming that at least a copy of the retroelement was part of the microdissected chromosome tip. Whether retrotransposons were actually present at those terminal sites, such a putative template would have been able to generate enough probe to label other chromosome ends of *T. pubescens*, rather than a single tip.

Unless *T. pubescens* has exceptional telomeric sequences that are unable to be amplified by DOP-PCR, possible chromosome end structures of this species can be considered on the basis of the results presented in this work. Tandem repeats would be present in chromosome termini of *T. pubescens* but their extension is unusually short to be visualised by in situ hybridisation. In this case, telomere sequences present in the microlibrary would not be identified as highly repetitive DNA and the results showed would reflect labelling of sub-terminal DNA specific for each chromosome ends of this species. Alternatively, there could be an exceptionally rare structure in *T. pubescens* characterised by the absence of repeats shared by telomeres.

Probes from *T. pubescens* chromosome tips hybridised to chromosomes of *R. milleri* even at more stringent conditions. Significant signals were restricted to the pericentric heterochromatin, indicating sequence conservation in non-centromeric ends and pericentric compartments of distantly related sciarids. Interestingly, the telomeric origin of centromeres has been hypothesised (Villasante et al. 2007a; Villasante et al. 2007b). Results obtained with the chromosome end probe of *R. americana* showing labelling at many chromosome sites of *R. milleri* were obtained by lowering hybridisation and washing temperatures. The data suggest considerable divergence between sequences composing the *R. americana* probe and those of *R. milleri* chromosomes. As sequences present in chromosome ends of *R. americana* failed to hybridise to *R. milleri* chromosomes, signals detected with the *R. americana* tip probe at chromosome ends of *R. milleri* are likely to reflect sequence conservation of sub-telomeric DNA instead of telomere repeats.

In summary, the results suggest the existence of an unusual structure in *T. pubescens* when compared to other organisms displaying non-canonical chromosome ends. The data also argue that the terminal structure of two distantly related *Rhynchosciara* species, *R. americana* and *R. milleri*, is also divergent although sequence data still remain to be obtained in the latter species. This work reinforces the view that dipterans might have other chromosome end structures in addition to those that have already been characterised in *Drosophila*, *Chironomus*, *Anopheles* and *Rhynchosciara americana*.

REFERENCES

Altschul SF, Gisg W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410

Amabis JM (1974) Estudo dos cromossomos de *Trichomegalosphys pubescens*. PhD Thesis, Instituto de Biociências, Universidade de São Paulo

Amabis JM (1983) The polytene chromosomes of the salivary gland of *Trichomegalosphys pubescens* (Diptera: Sciaridae) Brazil J Genetics VI: 415-424

Amorim DS (1992) A catalogue of the family Sciaridae (Diptera) of the Americas south of the United States. Revta Bras Ent 36: 55-77

Biessmann H, Donath J, Walter MF (1996) Molecular characterization of the *Anopheles gambiae* 2L telomeric region via an integrated transgene. Insect Mol Biol 5:11–20

Biessmann H, Kobeski F, Walter MF, Kasravi A, Roth CW (1998) DNA organization and length polymorphism at the 2L telomeric region of *Anopheles gambiae*. Insect Mol Biol 7: 83-93

Biessmann H, Zurovcova M, Yao GY, Lozovskaya E, Walter MF (2000) A telomeric satellite in *Drosophila virilis* and its sibling species. Chromosoma 109: 372-380

Breuer ME (1969) Revision of the genus *Rhynchosciara* Rübsaamen (Diptera: Sciaridae) in the neotropical region. Arq Zool S Paulo 17: 167-198

Carmona MJ, Morcillo G, Galler R, Martinez-Salas E, de la Campa A, Díez JL, Edström J-E (1985) Cloning and molecular characterization of a telomeric sequence from a temperature-induced Balbiani ring. Chromosoma 92: 108-115

Díez JL, Martínez-Guitarte JL, Gorab E, Morcillo G (2009) Transcriptional activity at the telomeres of *Chironomus* (Diptera). Its possible role in the lengthening of the telomeres. In: (Mancini L ed) Telomeres; Function, Shortening and Lengthening. Nova Publishers NY pp 315-331

Gorab E (2003) Reverse transcriptase-related proteins in telomeres and in certain chromosomal *loci* of *Rhynchosciara* (Diptera: Sciaridae). Chromosoma 111: 445-454

Gorab E, Amabis JM, Stocker AJ, Drummond L, Stollar BD (2009) Potential sites of triple helical nucleic acid formation in chromosomes of *Rhynchosciara americana* (Diptera: Sciaridae) and *Drosophila melanogaster*. Chromosome Res 17: 821-832

Jones KW, Bishop JO, Brito da Cunha A (1973) Complex formation between poly(r)-U and various chromosomal loci in *Rhynchosciara*. Chromosoma 43: 375-390

López CC, Nielsen L, Edström J-E (1996) Terminal long tandem repeats in chromosomes from *Chironomus pallidivittatus*. Mol Cell Biol 16: 3285-3290

Madalena CRG, Gorab E (2005) A chromosome end satellite of *Rhynchosciara americana* (Diptera: Sciaridae) resembling nematoceran telomeric repeats. Insect Mol Biol 14: 255-262

Madalena CRG, Fernandes T, Villasante A, Gorab E (2010a) Curiously composite structures of a retrotransposon and a complex repeat associated with chromosome ends of *Rhynchosciara americana* (Diptera: Sciaridae). Chromosome Res 18: 587-598

Madalena CRG, Amabis JM, Gorab E (2010b) Unusually short tandem repeats appear to reach chromosome ends of *Rhynchosciara americana* (Diptera: Sciaridae). Chromosoma 119: 613-623

Martinez JL, Edström J-E, Morcillo G, Díez JL (2001) Telomeres in *Chironomus thummi* are characterized by different families of complex DNA repeats. Chromosoma 110: 221-227

Mason JM, Frydrychova RC, Biessmann H (2008) *Drosophila* telomeres: an exception providing new insights. BioEssays 30: 25-37

Michailova P (1989) The polytene chromosomes and their significance to the systematics of the family Chironomidae, Diptera. Acta Zool Fennica 186: 1-107

Nielsen L, Schmidt ER, Edström J-E (1990) Subrepeats result from regional DNA sequence conservation in tandem repeats in *Chironomus* telomeres. J Mol Biol 216: 577-584

Pavan C, Breuer ME (1955) Polytene chromosomes of *Rhynchosciara milleri* (Diptera: Mycetophilidae). Rev Bras Biol 15: 329-339

Rosén M, Edström J-E (2000) DNA structures common for chironomid telomeres terminating with complex repeats. Insect Mol Biol 9: 341-347

Rosén M, López CC, Edström JE (2002) Telomere terminating with centromere specific repeats is closely associated with a transposon derived gene in *Chironomus pallidivittatus*. Chromosoma 110: 532-541

Rossato RM, Madalena CRG, Gorab E (2007) Unusually short tandem repeats in the chromosome end structure of *Rhynchosciara* (Diptera: Nematocera). Genetica 131: 109-116

Saiga H, Edström J-E (1985) Long tandem arrays of complex repeat units in *Chironomus* telomeres. EMBO J 4: 799-804

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Stocker AJ, Gorab E (2003) Local enrichment with homopolymeric (dA/dT) DNA in genomes of some lower dipterans and *Drosophila melanogaster*. Chromosoma 111: 455-460

Stocker AJ, Gorab E, Amabis JM, Lara FJS (1993) A molecular cytogenetic comparison between *Rhynchosciara americana* and *Rhynchosciara hollaenderi* (Diptera: Sciaridae). Genome 36: 831-843

Telenius H, Carter NB, Bebb CE, Nordenskjöld M, Ponder BAJ, Tunnacliff A (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics 13: 718-725

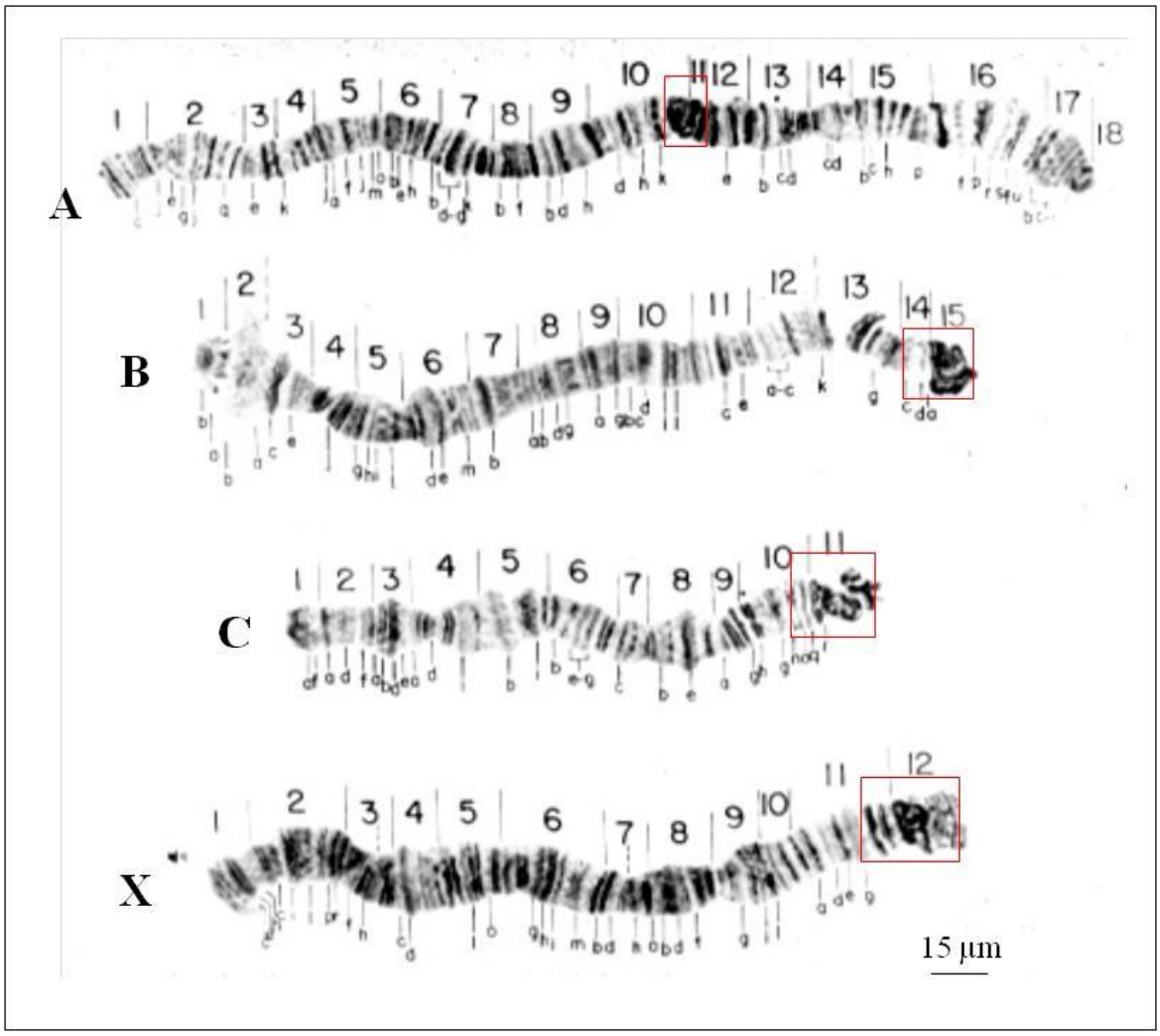
Villasante A, Abad JP, Méndez-Lago M (2007a) Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. Proc Natl Acad Sci USA 104: 10542-10547

Villasante A, Méndez-Lago M, Abad JP, de Garcíni EM (2007b) The birth of the centromere. Cell Cycle 6: 2872-2876

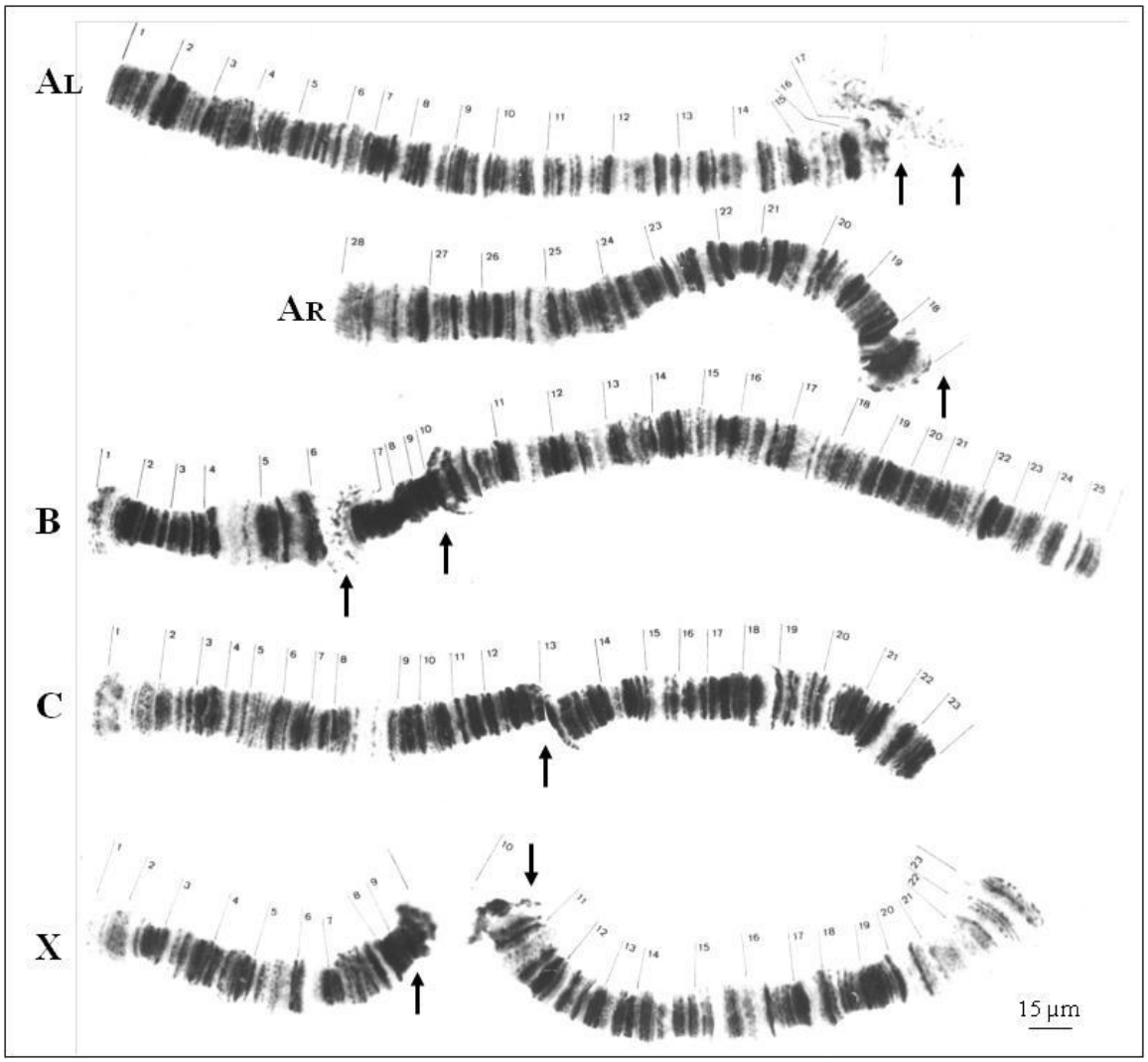
Villasante A, de Pablos B, Méndez-Lago M, Abad JP (2008) Telomere maintenance in *Drosophila*: rapid transposon evolution to chromosome ends. Cell Cycle 7: 2134-2138

SUPPLEMENTARY MATERIALS

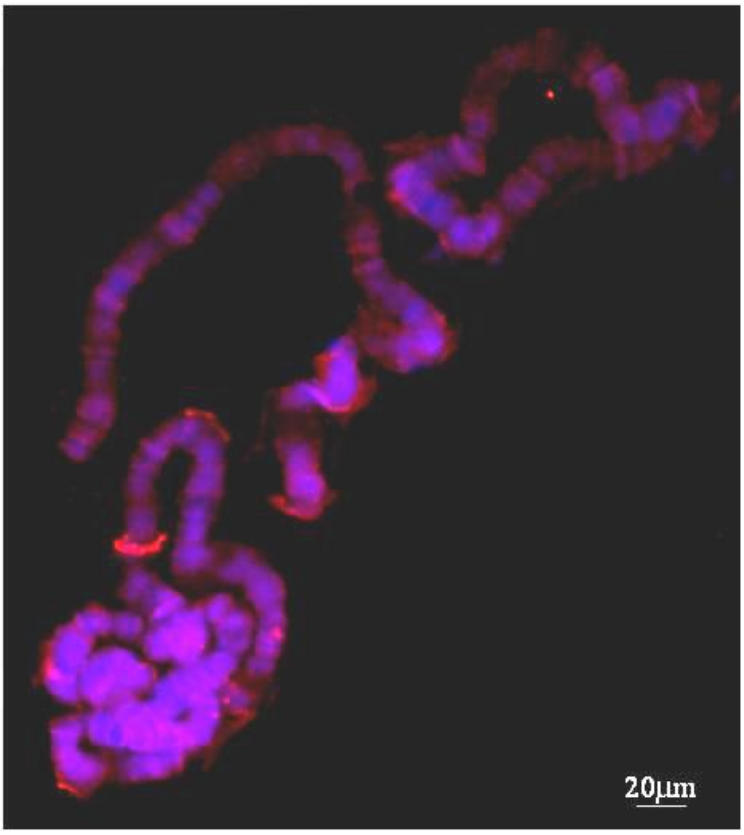
SM Fig. S1: Polytene chromosomes (*A, B, C and X*) of *R. americana*. *Boxes* are signaling chromosome sections comprising centromeric and pericentric regions. Chromosome section numbering was described previously (Stocker et al. 1993).



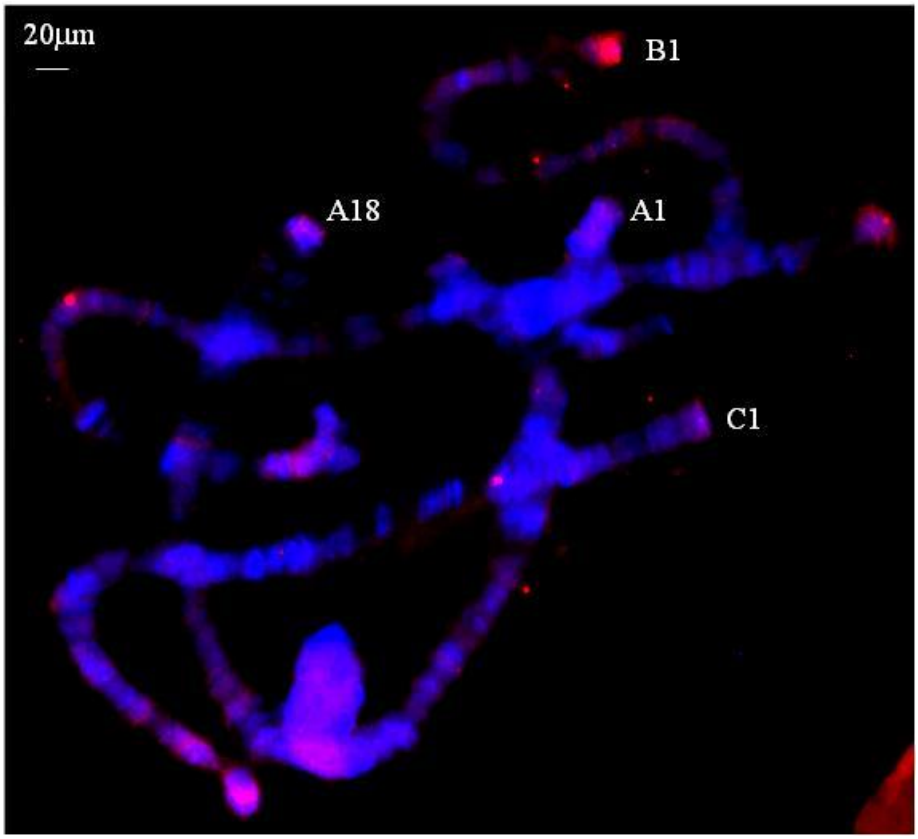
SM Fig. S2: Polytene chromosomes (*A, B, C and X*) of *T. pubescens* (modified from Amabis 1983). *AL* and *AR* correspond to long and short arms of chromosome A respectively. The *arrows* indicate most frequent weakpoint regions in each chromosome.



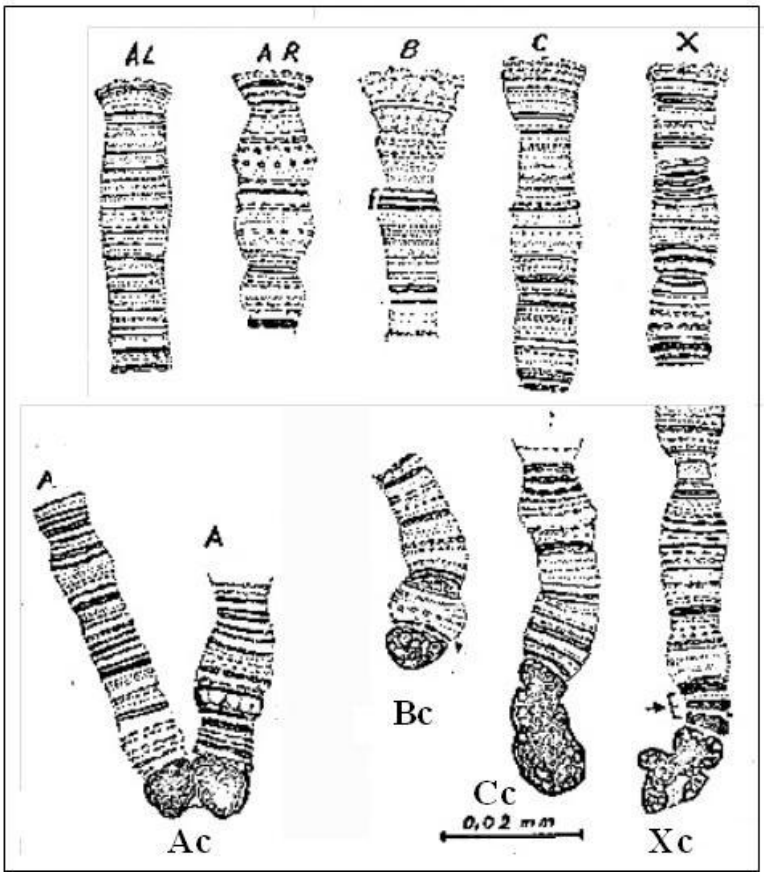
SM Fig. S3: Localisation of the probe (*red signal*) synthesised with DNA from the microdissected A-1 tip of *T. pubescens* in chromosomes of this species counterstained with DAPI (*blue signal*).



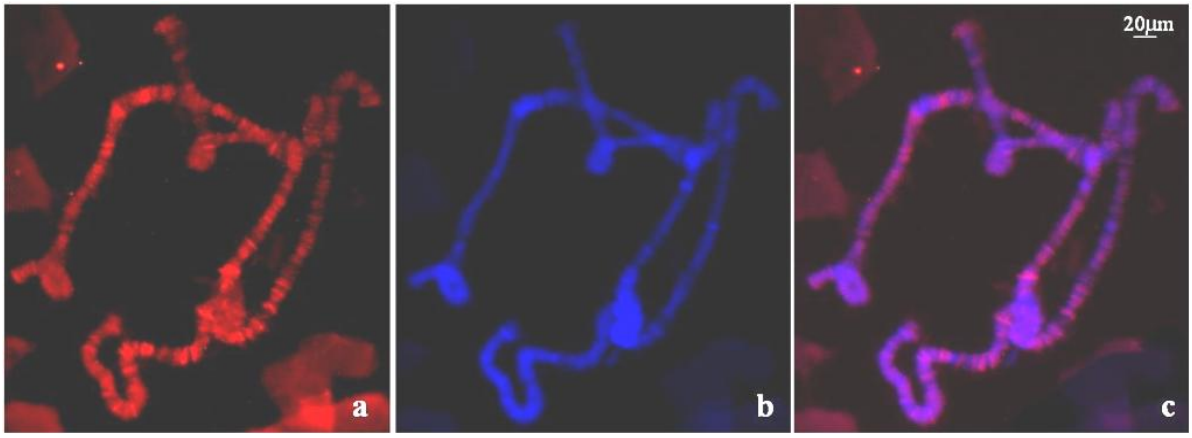
SM Fig. S4: Localisation of the probe (*red signal*) synthesised with DNA from the microdissected X-1 tip of *T. pubescens* in chromosomes of *R. americana* counterstained with DAPI (*blue signal*).



SM Fig. S5: Drawing of polytene chromosomes of *R. milleri* (modified from Pavan and Breuer 1955). Telomeric ends (*AL*, *AR*, *B*, *C*, *X*), centromeric ends (*Bc*, *Cc*, *Xc*) and the pericentric region of chromosome A (*Ac*) are identified.



SM Fig. S6: a-c Localisation of the probe **a** (*red signal*) synthesised with DNA from the microdissected B-1 tip of *R. americana* in chromosomes of *R. milleri*; **b** DAPI counterstaining (*blue signal*) showing **c** the merged signals; hybridisation and washings were performed at 32°C.



SM Fig. S7: Sequencing results of inserts from the plasmid library made with DNA from the microdissected X-1 chromosome end of *T. Pubescens.*

p*Tp*-1
GCGCATCCTAAAAAGATCCAACCTAATAACACCCCAACCCCAAAAACGAC - 50
CCAAAAACTCCTACCACCGATTCCAAGCAGACCTACCCAACCAAACCTGG – 100
CAATCAGACTTCACCCACGTACAACTAAAAAACGGCACCGACACAGAAGT - 150
AATCACCTGGCTAGACGACCACTCCCGCTACCTCCTACACGTAAGCGCAC - 200
ACCCC

p*Tp*-4
GATAAAATTGTAGCAGGTCTGCCAAACGAGCAGGCCCTTGAGAATATCGC - 50
ACTTGTTCGAGATGGTGAGAAACTGCTTCTTGTTCGATCGCAGCCATGTC – 100
CATGTATTACGTGCCCAGCGAATCCATGCCCATACCGAAACGCAAGGGAA - 150
GGATGGTTACCATGTACAAGAAGGATTCCAGGCTCACCAGCGTCCCGAT

p*Tp*-5
AAGACGTAGAAATTATCAACCCGGGGCAAAGCTGAAGCTGACAAAATGTA - 50
CTAGCGCAAAATTCCAATTTAGGTACAACGTAGGAACAAGATTTTAATGT – 100
GCTACTCGTATAGCATTTTACACTTTCATCGTATTAATGCATTACAGTAC - 150
GTAATAGTTAATACGGCTAGTCGCTTTATAGGTCGTTTTGCGCACCAAAT - 200
GTAAAGATTGTGAATAAGGGCCGAAGTCTAGTTTCGCAACGCATCGTGTA - 250
CGCAAACCACTATTAGGCAACGTGTTTTATACA

p*Tp*-7

GGACTACAAGGGGTCCATTAAAAATATTAAAAGAAGTTTTTCAATCGCTT - 50
TTATATTTTTGTGAAAGTGCAATAAAGCGAATGGCATGCATTATTATTTA – 100
CTAATTATTTATGTTAATTAACACGCATATGAAACTGTAAACTGAAACGT - 150
CACACACAATTAAGTGCACATTTTTGCCCTGTTCACATAACATGATACTT - 200
GATGTCGTATTCACACAATTGATTTTAGTTTCCCAAGAAAAGGTAATTAA - 250
ACTTGTCGTTGACATTAATGAAAGATTTTTCCCTAC

p*Tp*-14

GGGAGAATGGATTGAATGGCAAGCAAGACCAGAAACAAAGAGAAAGAGTA - 50
GGAGGCCCTGGCCGTAGTCCAGGCAGGAGATGAGTTTGCATGAGGCAGAT – 100
CATATTGAATGCCTCCCCCAAAATATTCCTTCCCCACTCCACCTCTCCAA - 150
TCCAACCTTCGTCCTGGCCAGCAGAACCTGCTT

p*Tp*-15
GACATGCTGGAGGCAATACTGCTCTTTAAGGCACTGAGATGTTTATCGTAT - 50
GTGCACCTCAAAAGCACAGCATTTTTTTCTTTACCTTGTTTATGATGCGAG – 100
ACATTTGTTCACTTGTTTTCCTGCTGACCCTCTCTCCATTACTACCCTATT - 150
GTCCTG

p*Tp*-16
GGAGCAAGACGAAGAATGACAGGGACTAAAGTAAAGTAGCATGATTCACAAG - 50
CATTATGTTCAGTTTAGAATTGTTTTTCCTAAATGCGAAGTGTGTCTATTAA – 100
AGGGTTTTAAGCATGGGTGCATCATGATCTGCTTTTTTTAAATTATTATACT - 150
TTAAGTTTTAGGGTACATGTGCACAACGTGCAGGTTGGTTACATATGTATAC - 200
ATGTGCCATGTTGGTGTGCTGCACCCAGTAACTCATCATTTAACATTAGTTA - 250
TAACTCCTAATGCTATCCCTCCCCCCTCCCCCCACCC

p*Tp*-21
AGGTATACTGAGAGTTTGTAGCCAGTGTAGACATTTCTCTAAATAAAATGAC - 50
ATTTTTGAAAATTTATGGTTAAAATAGGCTACTAGTAGGCTACAATAGGCTA – 100
CAATTCATGGTAATTTGCATTAAAAGGGTCGAAAGTAAAAGATTTATGCGAA - 150
GTTTGTATCTAGGTTAATATTTTTTAAATG

p*Tp*-22
GAGAGCTCAAAGATAAATCGGTAAGACGTCCAGACGGGTTCTCGTGTCATTA - 50
TCATTTTACTATACAGATCATTATATACCTTTCCCTCCGCGTAATTATATCA – 100
ACTGCCAGATAAATGGTTTTGTCGTCTAAAGAAAGTACCTAACTTCTAGATC - 150
TGAGTTAATTTAATCTCAAAATATTCTTCCTGGCCAACCCCTAGGCCGGAAC - 200
CTATGTACACAAGATACTCTTCGGCAGCCTGTTTATA

p*Tp*-25

```
GGGAAATTAAGCTGTCTTTCTGAAGAAATGGAATTATTGAAGTTTTCAATGC - 50
CTTTCTCCTATGCTATGTGACAGTGCTCCCTGCCTCCTGGGAAACTATAACA - 100
GTTTGGCTCACAGTCTATGTGAAGATGCATATGTTTCCTTCCCTTTTTTTTT - 150
CATAGAGACATTAAAAATTACTTTACTTTGGTGTATTAAGAATCCTGCCCAC - 200
ACAAGAATTACAACCTCAGCCCCCAGGGACCCATCCTTCCTACAGGGTAAAT - 250
GGAGCTTCTTTGTTCCTTATATCCATTCTGCTCATTCTTTTTG
```

p*Tp*-28

```
GAGGTATACTCGAGAGTTTGTAGCCAGTGTAGACATTTCTCTAAATAAAATGA - 50
CATTTTTGAAAATTTATCGTTAAAATAGGCTACTAGTAGGCTACAATAGGCTA - 100
CAATTCATGGTAATTTGCATTAAAAGGGTCGAAAGTAAAAGATTTATGCGAAG - 150
TTTGTATCTAGGTTAATATTTTTTAAATGC
```

p*Tp*-29

```
GCTTGGTTCTCTTGAGTCGTCTTGAGTTCTGAGCGAAGTTTAGCAGTAGGAGG - 50
GCCTCTGGATTTTCATGCTGGCTGAGTCTGTTTTCATGCTTAGCTGCAGTCGT - 100
TTTGATAACATCGGAGACCATAGGGACGTTGAGGTCGCGATGTAGATCACTGT - 150
TTCTGATGTACCAGGGGGCGTTGACGATAGTTCGTAGCACTTTATTCTGAAAT - 200
AATTGTATTGTCTTGATATGGCCTTTGCTCGCACATCCC
```

```
pTp-29 40    TTAGCAGTAGGAGGGCCTCTGGATTTTCATGCTGGCTGAGTCTGTTTTCATGCTTAGCTG  99
             |||| |||   || || || |||||||||||| ||    |||| || |||||| || || |
RaTART 8046  TTAGAAGTTCAAGTGCTTCAGGATTTTCATGTTGTAATAGTCGGTCTTCATGTTTTGCAG  8105
pTp-29 100   CAGTCGTTTTGATAACATCGGAGACCATAGGGACGTTGAGGTCGCGATGTAGATCACTGT  159
             || ||  ||| || || || |  |||||||| || || || || ||||| |  ||| |||
RaTART 8106  CAATCTGTTTAATGACCTCAGCTACCATAGGAACCTTAAGATCTCGATGGATGTCAATGT  8165
pTp-29 160   TTCTGATGTACCAGGGGGCGTTGACGATAGTTCGTAGCACTTTATTCTGAAATAATTGTA  219
             | | |   ||| | || ||| |||  || |||| ||    ||| || ||||||   ||||
RaTART 8166  TACGGGCATACTAAGGTGCGCTGATTATTGTTCTTAATGATTTGTTTTGAAATCTCTGTA  8225
pTp-29 220   TTGTCTTGATATGGCCTTTGCTCGCACATCC  250
             || |    ||||  | ||| || | ||||||
RaTART 8226  TTTTTAATATATTACTTTTACTAGTACATCC  8256
```

p*Tp*-30

```
GGGGGAGACAGGTGTGCAAGTAACAGTGATAAAGTGATATCAACGCTCAT - 50
AGTAAAGGCATCTTTAAACTGCTGTTGATTTGCAGAAGAGACAAACAGAT - 100
TCTATCAGGAGATTGAAAAATATTTTACAGAGGAGGTGATAAATGAGTTG - 150
GACCTTGAAGGCAGTAAGATTTTGTTAGTGGGGGAAGGAGAGGCATTCAG - 200
ATAAAGGTAACAGCATGAGC
```

p*Tp*-31

```
GTGGGGACGCTGCAAATGCTCCTTCAGTTTGAGGTTAAGCATCTCGCGGA - 50
GATAGTTGACGCTCGGGCAAGCTCGACTTGAAAGATCTCAAGGCCCAGTG - 100
GAAGAATGCCTCTCTCTCTAGTTGCTCAACTCCCTTGCGGTTTTGATCAA - 150
TTGCCATTCTACCATCGTCGGCGAGTCTTTGCTCCTCTCACGATTGGAAT - 200
GTTTCCATGATTCAATACCCAAGGTACAGATGGGTTTTGATTGCGACCAT - 250
CTTCCCAATTGATGGACGTTATTTCCACCAGATATTTTACGATTTTTC
```

p*Tp*-32

```
GGACCGACAATTTTTAGATTTTCCATTGTGATAACGCGCTATCGCACCGAG - 50
CCCGAATTATTTAACTAAACATCAAGTAAATACCATCGAGTAAGCACCGTA - 100
TTCACCTCGATTTTTTTTGTTCCC
```

p*Tp*-33

```
GAGAAACAGGAACTCCTGTTCACTATTGGTGGGAATCTAAAATGGTACAGCC - 50
ACTATGGAAAACAGTATGGTAGTTCCTCAAAAAATTAAACATGAAATTACTA - 100
TATGATCCAGAAAGTCTGATTCTGGGGAATATGCCCCAAATAATTGAAAACA - 150
GGTACTCAAAGAAATATTTGTATACTCATGTTCAAAGCAGCATTAATCACAA - 200
AGTCTAAGAGTTGGACACAACCAACGTGTCCATCAACAGATGAATGGATAAA - 250
CAAAATGTCATTTATCCATACAATGGACTATTACTCAACCTTAAAAAGGAGG - 300
GAAATTCTGATACATGCTAC
```

p*Tp*-34

```
AACATGGTCACTTATCAAAAAAGGCAAGTTCCAATGCACTTCAGGGCTTTGT - 50
CCCTATATGGCCCGCTAGCCTTACCAAACCCAAGTTCCAAGCTGCCAACAAG - 100
GCAGGCCAAATTGTCAACAAAGCCCATTCCCCAAGACACACTCTCTCTCTGG - 150
GGAACTAAGACAATGACAAAGCCACGTATCACGTGCATCATGTGCCTGGCAT - 200
TCCCAGAGGTTCTGGTATGTACCACATGGACACCAAGATATCTCAATGAGTT - 250
ATCACCTCCAAACCCCACCATCTG
```

p*Tp*-36

```
GACAGACAGATGGACTGGACACGGACAGAGGGACAGACGGACAGAGGAACAG - 50
ACGGACAGACAGACAGACGGACGGACGGACGGACGGACGGACGGACGGGGAG - 100
AAAAGGTGCTTCGATCATTTGATCAACTTCTCCGGGCTCTGACTAAGACCAG - 150
TCAGGCAGTTAGACATGGCTTGAGTTCCAAATGGCAGCCTATTCCCGGTATA - 200
GTGCACTACTTTTATGGTCCTAGTCAACAGTAGTTGAGGCATAAAGAGACAT - 250
AGCGGACACACACACACACACACACACACACCCCAAC
```

```
pTp-36   2   acagacagatggactggac--acggacagagggacagacggacagaggaacagacggaca  59
              ||||||||||||||| ||||  ||||||||||||||||||||||||||||||||||||||
pRa-43   1   ACAGACAGATGGAC-GGACGGACGGACAGAGGGACAGACGGACAGAGGAACAGACGGACA  59

pTp-36  60   gacagacagacggacggacggacggacggacggacggacggGGAGAAAAGGTGCTTCGAT  119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
pRa-43  60   GACAGACAGACGGACGGACGGACGGACGGACGGACGGACGGGGAGAAAAGGTGCTTCGAT  119

pTp-36  120  CATTTGATCAACTTCTCCGGGCTCTGACTAAGACCAGTCAGGCAGTTAGACATGGCTTGA  179
              |||||||||||||||||||||||||||||||||||||||| || ||||||||||||||||
pRa-43  120  CATTTGATCAACTTCTCCGGGCTCTGACTAAGACCAGTCATGCGGTTAGACATGGCTTGA  179

pTp-36  180  GTTCCAAATGGCAGCCTATTCCCGGTATAGTGCACTACTTTTATGGTCCTAGTCAACAGT  239
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
pRa-43  180  GTTCCAAATGGCAGCCTATTCCCTGTATAGTGCACTACTTTTATGGTCCTAGTCAACAGT  239

pTp-36  240  AGTTGAGGCATAAAGAGACATAGCGGacacacacacacacacacacacacCCCAAC
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
pRa-43  240  AGTTGAGGCATAAAGAGACATAGCGGACACACACACACACACACACACACACCCCAAC
```

p*Tp*-38

```
CATACAAAACTGAAATAAAAGTGAGAAATCGCAGTAACAGTAATGAAAAA - 50
TGGACGTTCTCCCAGCCCAGATGAAATACCCACTGAGCTACGGAGATCAG - 100
CCGGAGCAATCGGAATTTTTGCTACCTGATGATTGGC
```

p*Tp*-40

```
GAGAATGGAACCAAGTTGGAAAACACTCTGCAGGATATTATCCAGGAGAA - 50
CTTCCCCAATCTAGCAAGGCAGGCCAACATTCAGATTCAGGAAATACAGA - 100
GAACGC
```

p*Tp*-41

```
GTCAGGCCAAAAGGAATTTTAATCCTAGCTTTCTTACTATTTGCATGAAT - 50
TTTGGTGAGTTGTATAATCCCCCTCTCTGTATCTTATTCAATCCTTGTAA - 100
AAGTCCATTGAAGTAGGCTTA
```

p*Tp*-42

```
GTCACTTCTCCACACTGTAGCTAAACAAGGCTGTGAAAGTGAGTTTCTAG - 50
GCATTTTCAGCTGTTGGTGTGGGAAGCAGGTTCTGCTGGCCAGGACGAAG - 100
GTTGGATTGGAGAGGTGGAGTGGGGAAGGAATATTTTGGGGGAGGCATTG - 150
CATATGATCTG
```

p*Tp*-47

```
GCAGATGGTGGGGTTTGGAGGTGATAACTCATTGAGATATCTTGGTGTCC - 50
ATGTGGTACATACCAGAACCTCTGGGAATGCCAGGCACATGATGCACGTG - 100
ATACGTGGCTTTGTCATTGTCTTAGTTCCCCATGAGAGAGAGTGTGTCTT - 150
GGGGAATGGGCTTTGTTGGCAATTTGGCCTGCCTTGTTGGCAGCTTGGAA - 200
CTTGGGTTTGGTAAGGCTAGCGGGCCATATAGGGACAAAGCCCTGAAGTG - 250
CATTGGAACTTGCCTTTTTTGATAAGTGACCATGTTTCCTCCTCCAGCAC - 300
TTAAAATGTGCCTTTCTCC
```

p*Tp*-55

```
GAGCAACTAGAAATCTCACATGTTTCTGTTGGGAATTGGGAATGGAAAAT - 50
GGTACTGCCATTTTGGAACAGTATGTCATGTTATTATATAGTTAAACATC - 100
CTCTTATCATATGATCACTCAATTTCATTTCTAGGTCCATACTCAAGAGA - 150
AATAAAAACACATGTCTACAAAAACACTGGCATGCAATGTTCATAGTAAC - 250
ATTATTCACTATAACCCTTAACTAGAAACAGCC
```

p*Tp*-56

```
GGTTTTTCCCCGTCCGAAGGCGAGCTTCACAAGGAACTAGAAAGCTATGT - 50
TACGCTTCTGGAACCGCTGGACTGGGTGCTCTAACTCTTCATCCCTATGA - 100
AGCAGCTTGTTGCGTTGGAAAGGATCCTTCCGAGCTTGGAGGATCCTGCT - 150
ATCTGCATCGACCACTTTGGCTGGCCAACCCTACCACCGTATGACCCTTC - 200
ACGTCAGCTTGTTCCATATTCGTTGGCCGGCTTTGAATCTCTCACTCGTC - 250
TCTTGAAGAGAAATGCTTGGGTTAAACTATCAGCTCCATACCGACTCTCA - 300
AAGGATTCCCAAATGAGTGATCTTGGTGCCATCGGACGAGAGCTCATCAA - 350
GAACCACCGAGAGAGAGTCATCTACGCTACTGACTGG
```

p*Tp*-58

```
CGCGATACCGGGCTACGCTACACCCCGAAAAGGTATCACCTTTGTTTTGT - 50
GTGGCACAAATATACGCTTCTTACAAATTATTTGCAAATGAAAAACGCAA - 100
AATAAAACTAACAAGTTGTTACATTGCGACATACTGTTTTCTCATATGAC - 150
GTAGTACCCTCCTTTACAAAATCAAACCCAATGTTATTCTCCATATTCCT - 200
TAC
```

p*Tp*-59

```
CCGACTCGAGCGGCGCATGTGGAAGTGCAACGAAAATAAAATACGATTAA - 50
AATAGAGGAAAAACTAGGCAGAAGATGGGAATACTCTAACTGAGAAGTGC - 100
ATAAGGTAGAGGTTTTTATTGTTAGCATTTTATGATGATGGTCTTGATCC - 150
TGTTTTTGTACGCTAATGATCAGACTGATACCACGTCTGGCCGAACTCTT - 200
AAGATGCGAGATTCCAACGCATTTAATGCCAACAAACGTCGCCTAAAACC - 250
TGCAATACCCTCCATTGGTTGGCC
```

p*Tp*-758

```
AAAAGACGGCATAAGTTGGTTCGGGCACGTGCAGAAAATGAGGTATAAAC - 50
AGGCTGCCGAAGAGTATCTTGTGTACATAGGTTCCGGCCTAGGGGTTGGC - 100
CAGGAAGAATATTTTGAGATTAAATTAAATCAGATCTAGAAGTTAGGTAC - 150
TTTCTTTAGACGACAAAACCATTTATCT
```

CONCLUSÃO

A Ordem Diptera é constituída de milhares de espécies, cujo tempo de divergência pode chegar a 250 milhões de anos entre representantes das Sub-Ordens Brachycera e Nematocera. Esta janela temporal, no entanto, não é suficiente para explicar o aparecimento de estruturas cromossômicas terminais alternativas aos telômeros canônicos, conservados desde eucariontes unicelulares até mamíferos. Alguns autores admitem que estruturas não canônicas tenham sido geradas e selecionadas a partir de eventos mutacionais que resultaram na perda da telomerase em espécies ancestrais. Especulações deste tipo, embora pertinentes, não levam em conta que o número de espécies de dípteros estudados até aqui quanto a telômeros e sub-telômeros está longe de representar a diversidade na Ordem. Como evidência negativa não constitui prova, a possível existência de dípteros dotados de telômeros canônicos não pode ser descartada. Também neste sentido, a possível ocorrência de retrotransposons especificamente terminais, como vistos em *Drosophila*, em espécies ainda não estudadas pode ser vista como possibilidade em aberto apesar da evidência negativa documentada no presente trabalho em *R. americana* e em *T. pubescens*.

Outra idéia que tem ganhado corpo com dados procedentes de telômeros não canônicos refere-se à possibilidade de que um organismo apresente mais de uma sequência de DNA terminal. Pouco comentada, esta noção teve início em meados da década de 1980 quando foram publicados os primeiros trabalhos sobre o DNA telomérico de espécies da família Chironomidae. O aparecimento destes dados na literatura praticamente coincidiu com a publicação da descoberta da telomerase. Mais tarde, estudos em *Drosophila* têm

mostrado que três retrotransposons com sequências diferentes podem compor o DNA telomérico nesta espécie. Além disto, repetições aparentemente terminais em *Anopheles* hibridam em um único telômero, sugerindo que sequências desconhecidas estejam presentes em outros telômeros. Dados obtidos neste trabalho mostram em *R. americana* uma nova repetição em tandem que apresenta características de sequência telomérica, a exemplo de duas outras caracterizadas com antecedência nesta espécie. Assim, *R. americana* poderia ser vista como exemplo adicional de organismo dotado de mais de uma sequência de DNA terminal.

No final da presente exploração, não foi possível a identificação de sequências comuns às extremidades cromossômicas de *T. pubescens*. Os resultados obtidos sugerem que esta espécie apresenta uma estrutura cromossômica terminal distinta se comparada àquelas de dípteros estudados até então. Uma das hipóteses levantadas a partir dos resultados observados é a de que sequências teloméricas estão presentes em todas as extremidades cromossômicas de *T. pubescens*; o problema estaria na impossibilidade de visualizá-las através de hibridação in situ em função do comprimento do DNA telomérico que estaria abaixo do limite da técnica de detecção. Isto implicaria deixar de lado os métodos usualmente empregados pelo laboratório na busca de DNA repetitivo terminal. Caso esta espécie apresentasse repetições terminais curtas como aquelas caracterizadas em *R. americana*, uma das alternativas seria a de sequênciar massivamente clones obtidos por microdissecção e DOP-PCR até encontrarmos sequências com estas características. Em seguida, o ensaio com *Bal-31* seria decisivo na determinação da posição cromossômica das mesmas.

Finalmente, os dados sintetizados nos três capítulos deste trabalho reforçam a afirmação de que a Ordem Diptera é uma fornte privilegiada de diversidade quanto a estruturas cromossômicas terminais. Apesar das dificuldades inerentes à exploração de organismos não modelares, a continuidade dos estudos sobre telômeros e sub-telômeros nestas espécies certamente ampliará o conhecimento sobre alternativas de manutenção da integridade cromossômica a partir de suas extremidades.