unesp

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
INSTITUTO DE BIOCIÊNCIAS – RIO CLARO

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS
(Biologia Celular e Molecular)

ALLISON KLEITON DOS ANJOS

DINÂMICA EVOLUTIVA DOS CLUSTERS DE PEQUENOS RNAs NUCLEARES (RNAsn) NOS CARIÓTIPOS DE ESPÉCIES DE GAFANHOTOS COM ÊNFASE EM ACRIDIDAE

Dissertação apresentada ao Instituto de Biociências do Campus de Rio Claro, Universidade Estadual Paulista, como parte dos requisitos para obtenção do título de Mestre em Ciências Biológicas. Área de Concentração: Biologia Celular e Molecular.

2014

ALLISON KLEITON DOS ANJOS

DINÂMICA EVOLUTIVA DOS CLUSTERS DE PEQUENOS RNAs NUCLEARES (RNAsn) NOS CARIÓTIPOS DE ESPÉCIES DE GAFANHOTOS COM ÊNFASE EM ACRIDIDAE

Orientador: Diogo Cavalcanti Cabral-de-Mello

Dissertação apresentada ao Instituto de Biociências do Campus de Rio Claro, Universidade Estadual Paulista, como parte dos requisitos para obtenção do título de Mestre em Ciências Biológicas. Área de Concentração: Biologia Celular e Molecular.

Rio Claro

2014

595.7 Anjos, Allison
A597d Dinâmica evolutiva dos clusters de pequenos RNAs nucleares (RNAsn) nos cariótipos de espécies de gafanhotos com ênfase em Acrididae / Allison Anjos. - Rio Claro, 2014
73 f. : il., figs., gráfs.

Dissertação (mestrado) - Universidade Estadual Paulista, Instituto de Biociências de Rio Claro
Orientador: Diogo Cavalcanti Cabral de Mello
Coorientador: Vilma Loreto da Silva

1. Inseto. 2. Clusters de U1 DNAsn. 3. DNA repetitivo. 4. Evolução genômica. 5. Família multigênica. 6. FISH. 7. Spliceossomo. I. Título.

Ficha Catalográfica elaborada pela STATI - Biblioteca da UNESP
Campus de Rio Claro/SP

unesp **UNIVERSIDADE ESTADUAL PAULISTA**

CAMPUS DE RIO CLARO

INSTITUTO DE BIOCIÊNCIAS DE RIO CLARO

CERTIFICADO DE APROVAÇÃO

TÍTULO: DINÂMICA EVOLUTIVA DOS CLUSTERS DE PEQUENOS RNAs NUCLEARES (RNAsn) NOS CARIÓTIPOS DE ESPÉCIES DE GAFANHOTOS COM ÊNFASE EM ACRIDIDAE

AUTOR: ALLISON KLEITON DOS ANJOS

ORIENTADOR: Prof. Dr. DIOGO CAVALCANTI CABRAL DE MELLO
CO-ORIENTADORA: Profa. Dra. VILMA LORETO DA SILVA

Aprovado como parte das exigências para obtenção do Título de MESTRE EM CIENCIAS BIOLÓGICAS (BIOLOGIA CELULAR E MOLECULAR) , pela Comissão Examinadora:

Prof. Dr. DIOGO CAVALCANTI CABRAL DE MELLO
Departamento de Biologia / Instituto de Biociências de Rio Claro

Prof. Dr. DARDO ANDREA MARTÍ
Instituto de Biología Subtropical / Universidade Nacional de Misiones

Prof. Dr. CESAR MARTINS
Departamento de Morfologia / Instituto de Biociências de Botucatu

Data da realização: 17 de fevereiro de 2014.

Dedico este trabalho a família Anjos, Antônio, Silvânia e Andreza, que apesar da distância sempre encontram uma forma de me ouvir e de me apoiar em todas as minhas decisões.

Agradecimentos

Aos meus pais, Antônio e Silvânia, por todo o amor, carinho e compreensão, bem como pela amizade, afinal, além de pais dedicados eles são meus grandes amigos com os quais eu posso contar em todos os momentos da minha vida. Obrigado pela preocupação com meu bem estar e por fazerem o possível para que eu conseguisse passar por mais uma etapa da minha vida.

A minha irmã e melhor amiga da vida inteira, Andreza, pelo amor verdadeiro, pelas conversas de madrugada chorando bêbado ao telefone, pela saudade interminável, pelas brigas, pelas risadas e pela confiança mesmo com a distância.

Agradeço ao Prof. Diogo Cavalcanti Cabral-de-Mello pela confiança em mim depositada, pela excelente orientação, essencial para o desenvolvimento deste trabalho. E acima de tudo agradeço pela amizade, pela preocupação não só com o andamento do projeto, pelos churrascos, pelas cervejas e pelos conselhos e conversas na mesa de bar.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro cedido para realização do projeto.

A quatro amigos muito especiais que tive o prazer de conhecer durante meus dois anos de mestrado. Primeiro, Matraca, pelas muitas risadas, pelas tantas conversas e desabafos, pelas caronas, pela simples companhia, pelas lambidas desnecessárias e por todo amor conquistado durante esse tempo. Segundo, Vivi, que me ouviu em tempo integral, que sempre está disposta a me dar conselhos e falar alguma coisa pra me animar e que me acompanhou em tantas festas e cervejadas. Terceiro, Mão, pela amizade dedicada, assim como pelas festas, pelos closes nas baladas e pelos beijinhos no ombro pra espantar o recalque das inimigas. Quarto Kety Maravilha, que sempre me faz rolar de rir e é mais uma que está sempre disposta a ajudar seja em que momento for mesmo ela sendo de Diadema.

Á Thiago Leão pelo apoio, dedicação, carinho, amor e paciência durante todos os dias, me fazendo ser um homem melhor e mais feliz.

A família Heleodoro, Cris, Aguinaldo, Jr e Jõao, que sempre me receberam muito bem e principalmente Cris, que sempre se mostrou disposta a dar bons conselhos e sempre se importou comigo, não só como amiga, mas também como uma verdadeira mãe.

Aos amigos de laboratório, Amália, Ana, Dioguinho, Gazoni, Gulosa, Nahanna, Mara, Oct, Roberta e Simone, pela ótima convivência durante as semanas de trabalho e pelas risadas que tornam o ambiente de trabalho muito mais agradável e fazem as horas passarem mais rápido. Aos amigos piauienses Edson, Tamaris e Tati, pela amizade verdadeira, pelos conselhos e pela presença sejam com um telefonema ou um recado no facebook.

Às amigas, Tailine e Nati, que apesar de nossa falta de tempo e da distância, se mostram tão presentes na minha vida me ajudando a esquecer um pouco os problemas e lembrar dos momentos de alegria vividos nos tempos de convívio diário.

Não são nossos talentos que mostram aquilo que realmente somos, mas, sim, as nossas escolhas.

J.K. Rowling

RESUMO

O spliceossomo é responsável pela maturação do RNAm através da remoção dos íntrons. Essa maquinaria é formada por um conjunto de proteínas associadas aos pequenos RNAs nucleares (RNAsn). Entre os genes de RNAsn, o U1 RNAsn é uma sequência conservada de 165 pb repetidas em tandem. A fim de contribuir para o entendimento da dinâmica cromossômica e genômica das famílias multigênicas em gafanhotos os genes de RNAsn U1 foram mapeados através da hibridização *in situ* fluorescente (FISH) em 71 espécies de gafanhotos pertencentes as famílias Proscopiidae, Pyrgomorphidae, Ommexechidae, Romaleidae e Acrididae. Além disso, a organização genômica dessa sequencia foi analisada usando como referência o genoma de *Eyprepocnemis plorans* sequenciado através do método 454. Foi observada uma grande conservação de clusters de DNAsn U1 localizados principalmente em pares de autossomos (nº 3 ou 4) nas primeiras quatro famílias. Em contraste, extensiva variação foi observada nas espécies de Acrididae, com um único par de cromossomos carregando DNAsn U1 a todos os pares de cromossomos portando a sequência, com a ocorrência de dois ou múltiplos clusters no mesmo cromossomo. No genoma de *E. plorans* cinco distintas linhagens foram observadas com distintos padrões de variação além da associação de DNAsn U1 com elementos de transposição e DNAr 5S. Esses resultados são discutidos focando os possíveis mecanismos de dispersão dos clusters desse gene, que aparentemente seguiu distintos caminhos de dispersão nas várias famílias e subfamílias analisadas. Este é o estudo mais abrangente sobe mapeamento por FISH até então realizado em gafanhotos e outros organismos, estudando 71 espécies pertencentes a cinco famílias, fornecendo assim importantes informações lançando luz na evolução cromossômica/genômica dessa família gênica pelo uso combinado de dados cromossômicos e genômicos.

Palavras chave: DNA repetitivo, evolução genômica, família multigênica, FISH, spliceossomo

ABSTRACT

The spliceosome is responsible for mRNA maturation through intron removal. This machinery is formed by a protein set associated to small nuclear RNA (snRNA). Among the snRNA genes, the U1 snRNA is, in general, a conserved 165 bp tandemly arrayed repetitive sequence. Aiming to contribute to the understanding of chromosome and genome dynamics of multigene families in grasshoppers we mapped the U1 snRNA genes by Fluorescent *in situ* Hybridization (FISH) in 71 grasshopper species belonging to the families Proscopiidae, Pyrgomorphidae, Ommexechidae, Romaleidae and Acrididae. Moreover we analyzed the genomic organization for this sequence using as reference the sequenced genome through 454 of *Eyprepocnemis plorans*. High conservation of snDNA clusters mainly located on autosome pairs (no. 3 or 4) was observed in the first four families. In contrast, extensive variation was observed in Acrididae species, from a single chromosome pair carrying U1 snDNA to all chromosome pairs carrying them, with occurrence of two or multiple clusters in the same chromosomes. In the genome of *E. plorans* five distinct lineages were observed with distinct patterns of variability and association of U1 snDNA with transposable elements and 5S rDNA was also noticed. These results are discussed focusing the possible mechanisms of spread of this gene cluster, which apparently seems to have followed different ways of dispersion in the several families and subfamilies analyzed in here. This is the most comprehensive study on FISH mapping hitherto performed in grasshoppers and other organisms by studying 71 species from five families and has thus provided valuable information shedding light in the chromosomal/genomic evolution of this gene family by combined use of chromosomal and genomic data.

Key words: FISH, genome evolution, multigene family, repetitive DNAs, spliceosome

SUMÁRIO

1.0 INTRODUÇÃO 7

1.1 ESTRUTURA CARIOTÍPICA EM GAFANHOTOS 7
1.2 DNAS REPETITIVOS NO GENOMA COM ÊNFASE NA FMÍLIA MULTIGÊNICA DE DNASN U 9
1.3 MAPEAMENTO FÍSICO CROMOSSÔMICO DE FAMÍLIAS MULTIGÊNICAS EM GAFANHOTOS E SUA CONTRIBUIÇÃO NO ENTENDIMENTO DA ORGANIZAÇÃO GENÔMICA 10

2.0 OBJETIVOS 13

2.1 OBJETIVO GERAL 13
2.2 OBJETIVOS ESPECÍFICOS 13

3.0 MATERIAL E MÉTODOS 14

3.1 ANIMAIS, EXTRAÇÃO DE DNA E PREPARAÇÕES CROMOSSÔMICAS 14
3.2 ISOLAMENTO DAS SEQUENCIAS DE DNAS U1 14
3.3 HIBRIDIZAÇÃO *IN SITU* FLUORESCENTE (FISH) 14
3.4 ANÁLISE DA ORGANIZAÇÃO GENÔMICA DO DNASN U1 NO GENOMA DE *EYPREPOCNEMIS PLORANS* 15

4.0 RESULTADOS E DISCUSSÃO 16

4.1 CAPÍTULO 1: U1 SNDNA CLUSTERS IN GRASSHOPPERS: CHROMOSOMAL DYNAMICS AND GENOMIC ORGANIZATION 17

5.0 CONCLUSÕES FINAIS 66

6.0 REFERÊNCIAS BIBLIOGRÁFICAS 67

1.0 INTRODUÇÃO

1.1 Estrutura cariotípica em gafanhotos

Estudos citogenéticos com objetivo de caracterizar cromossomicamente espécies da ordem Orthoptera são bastante extensos, principalmente envolvendo gafanhotos da região Neotropical. No estudo realizado por Mesa et al. (1982), que reuniu cariótipos de 289 espécies pertencentes a distintas famílias de Acridoidea, foi observado que o cariótipo modal para este grupo é 2n=23, X0 para machos e 2n=24, XX para fêmeas, além da ocorrência de cromossomos com morfologia acrocêntrica. Porém, diversos cariótipos derivados têm sido descritos com a observação, por exemplo, de extensas reduções do número diploide causadas por fusões Robertsonianas (Rb), modificação na morfologia cromossômica a partir de inversões, ocorrência de cromossomos supranumerários, além da origem de sistemas sexuais derivados por fusões/translocações entre cromossomos autossômicos e sexuais (MESA e FERREIRA 1977; HEWITT 1979; CASTILLO et al. 2010a,b; CARVALHO et al. 2011). A variação no número diploide é de 2n=8 em *Dichroplus silveiraguidoi* (Acrididae) a 2n=25 em *Conometopus sulcaticollis* (Ommexechidae) e os mecanismos cromossômicos de determinação sexual podem ser do tipo X0/XX, neo-XY/XX e neo-$X_1X_2Y/X_1X_1X_2X_2$, sendo X0 considerado ancestral (MESA et al. 1982; FERREIRA e MESA 2010; CASTILLO et al. 2010a,b).

Dentre as famílias de gafanhotos mais estudados cromossomicamente merecem destaque, Proscopiidae, Ommexechidae, Pyrgomorphidae, Romaleidae e Acrididae, devido à ampla diversidade de espécies estudadas e ampla variabilidade cromossômica em alguns de seus grupos. Dentre as espécies de Proscopiidae, a presença de números diploides inferiores a 2n=23 tem sido bastante comum, sendo este grupo caracterizado por 2n=15, 17, e 19, além da presença de mecanismos sexual do tipo X0 e cromossomos acrocêntricos (MESA e FERREIRA 1981; MOURA et al. 1996; SOUZA e MOURA 2000). Segundo Mesa e Ferreira (1981), fusões, fissões e inversões são os principais mecanismos responsáveis pela variabilidade cromossômica em Proscopiidae. De forma semelhante, na família Ommexechidae dezenove espécies tiveram seus cariótipos descritos, apresentando variações de 2n=21 (*Ommexecha germari*) a 2n=25 (*Conometopus sulcaticollis*), sendo o cariótipo mais frequente 2n=23, X0 (MESA e FERREIRA 1977; CARVALHO et al. 2011). Vale ressaltar a ocorrência de 2n=22, neo-XY em *Spathalium helios* e a presença de um par autossômico submetacêntrico (par 1) que ocorre em 16 das 19 espécies estudadas para a família, possivelmente originado a

partir de inversões pericêntricas envolvendo cromossomos acrocêntricos ancestrais (MESA e FERREIRA 1977; MESA et al. 1982; MESA et al. 1990; CARVALHO et al. 2011). Para a família Pyrgomorphidae, os poucos representantes estudados até o momento apresentam cariótipos com número diploides variando de 2n=11 até 2n=19 e cromossomos acrocêntricos (MESA e FERREIRA et al. 1982; JOHN e KING 1983; SEINO et al. 2013). Segundo Seino et al. (2013), o cariótipo mais frequente é 2n=23,X0 e a variação no número cromossômico deriva de mecanismos como fusões cêntricas do tipo autossomo-autossomo ou X-autossomo.

As famílias Romaleidae e Acrididae são caracterizadas pela marcante presença de 2n=23, X0 (machos)/24, XX (fêmeas) e cromossomos acrocêntricos na maioria das suas espécies (MESA et al. 1982). Dentre os romaleídeos estudados por Mesa et al. (1982), três deles (*Diponthus díspar*, *D. electus* e *D. maculiferus*) apresentaram número diploide reduzido para 21, 22 X0/XX, resultante de fusão entre autossomos que originaram um par de metacêntricos e, além disso, em outras três espécies (*Diponthus communis*, *Xyleus laevipes* e *Zoniopoda iheringi*) a redução para 2n=22 também foi observada, porém a fusão foi do tipo X-autossomo, que resultou em um mecanismo sexual do tipo neo-XY/neo-XX.

Dentre todos os ortópteros, a família Acrididae é a mais estudada e apresenta extensa conservação cariotípica, embora em distintas subfamílias ampla variabilidade tenha sido reportada. A subfamília Gomphocerinae, por exemplo, apresenta-se amplamente distribuída mundialmente, e a macro estrutura cromossômica é bastante diferenciada entre representantes da região Neotropical e das regiões Neárticas e Paleárticas. Enquanto a maioria das espécies ocorrentes nos neotrópicos apresentam 2n= 23, X0 (macho)/ 24, XX (fêmea) e cromossomos acrocêntricos, as espécies da fauna Neártica e Paleártica possuem frequentemente 2n=17, X0 (macho)/ 18, XX (fêmea) com três pares cromossômicos com dois braços provavelmente originados a partir de fusões cêntricas (HEWITT 1979; MESA et al. 1982; CABRERO E CAMACHO 1986; LORETO et al. 2008). Além das reduções do número diploide, a origem de sistemas sexuais derivados parece ter ocorrido diversas vezes durante a evolução de Acrididae, dentre os representantes da subfamília Melanoplinae. Nesta subfamília, por exemplo, aproximadamente 50% apresentaram cariótipos derivados, com frequente ocorrência de sistemas sexuais derivados, tais como neo-XY e neo-X_1X_2Y, merecendo destaque as espécies do gênero *Dichroplus*, o qual das 34 espécies analisadas 20 apresentaram cariótipos derivados (MESA et al. 1982; BIDAU E MARTÍ 2001; MESA et al. 2001;

COLOMBO et al. 2005; CASTILLO et al. 2010b). Além disso, em Acrididae várias espécies (cerca de 15%) apresentam elementos genéticos adicionais conhecidos como cromossomos B, caracterizados pela ausência de homologia com os demais membros do cariótipo padrão e por seu comportamento não-mendeliano de herança (CAMACHO et al. 2000).

1.2 DNAs repetitivos no genoma com ênfase na família multigênica de U snDNA

Os elementos repetitivos que constituem grande parte do genoma dos eucariotos se encontram organizados em distintos grupos, incluindo diversas famílias multigênicas, os DNAs satélites, microsatélites e minisatélites e elementos de transposição (transposons e retrotransposons). Estas sequências estão envolvidas em importantes processos celulares, como codificação de importantes proteínas e RNAs que atuam na organização e metabolismo celular, organização e funcionalidade de centrômeros e telômeros, perfeita segregação cromossômica, reparo e replicação do DNA, além da diferenciação dos cromossomos sexuais (ANLEITNER E HAYMER 1992; KRAEMER E SCHMID 1993; MESSIER et al. 1996; MARTINS 2007; SHAPIRO E STERNBERG 2005; BIÉMONT E VIEIRA 2006; FESCHOTTE E PRITHAM 2007). As famílias multigênicas, de acordo com Ney e Rooney (2005), são grupos de genes com similaridades funcionais e estruturais, que descendem de um gene ancestral comum e apresentam similaridade estrutural e funcional. Os genes de RNAs ribossomais (RNAr), genes codificadores de proteínas histônicas e genes de pequenos RNAs nucleares (RNAsn), são exemplos de famílias multigênicas bastante conhecidos e estudados do ponto de vista molecular e cromossômico.

Os genes responsáveis pela produção de RNAsn tem grande importância para o funcionamento celular por estarem envolvidos no processo de *splicing*, que consiste em remover os íntrons da molécula precursora do RNA mensageiro (GREEN 1989; RIO 1992; WEST 2012). O complexo macromolecular responsável por esse processo, chamado *spliceossomo*, é composto por um conjunto de mais de 300 proteínas associadas com os RNAsn U. Os genes que codificam os RNAsn que compõem a unidade maior do *spliciossomo* são codificados por uma família multigênica que inclui U1, U2, U4, U5 e U6 (BRINGMANN E LÜHRMANN 1986; NILSEN 2003; VALADKHAN 2005; WEST 2012). Devido a sua grande importância para o metabolismo celular, os genes RNAsn U constituem uma distinta classe de genes conservados evolutivamente e já foram mapeados citogeneticamente em distintos

organismos, incluindo humanos (LUND et al. 1983), camundongos (LUND e NESBITT 1988), crustáceos (BARZOTTI et al. 2003), peixes (MANCHADO et al. 2006; ÚBEDA-MANZANARO et al. 2010; CABRAL-DE-MELLO et al. 2012) e algumas espécies de gafanhotos (BUENO et al. 2013; PALÁCIOS-GIMENEZ et al. 2013), indicando conservação do número de clusters e posição dos clusters com sítios restritos a um ou poucos pares cromossômicos.

Do ponto de vista molecular, os genes de RNAsn U podem apresentar multiplas cópias dispersas no genoma, como observado em humanos (MANSER E GESTELAND 1982) e ratos (MARZLUFF et al. 1983), ou podem estar organizados em tandem, assim como no sapo *Xenopus laevis* (MATTAJ E ZELLER 1983), no ouriço *Strongylocentrotus purpuratus* (YU et al. 1991) e em alguns peixes teleósteos (MARZ et al. 2008). Além disso, estes genes foram encontrados associados com repetições do gene de DNA ribossomal 5S no crustáceo *Asellus aquaticus* (PELLICCIA et al. 2001), em uma espécie de peixe (MANCHADO et al. 2006) e em 10 espécies de moluscos bivalves (VIERNA et al. 2011).

1.3 Mapeamento físico cromossômico de famílias multigênicas em gafanhotos e sua contribuição no entendimento da organização genômica

Segundo diversos autores, apesar da grande estabilidade observada nos cariótipos das espécies de gafanhotos, existem grandes diferenças na quantidade de DNA entre as espécies (GREGORY 2005; HANARAHAN e JOHNSTON 2011). Além disso, dados recentemente obtidos por análises de mapeamento citogenético de DNAs repetitivos, principalmente famílias multigênicas, tem indicado um distinto cenário contrastando com a ampla estabilidade dos cariótipos de gafanhotos (YOASHIMURA et al. 2006; CABRERO e CAMACHO 2008; CABRAL-DE-MELLO et al. 2011a).

Os estudos relacionados à análise de DNAs repetitivos em Orthoptera se concentraram em sua maioria na análise da localização e qualificação dos blocos de heterocromatina constitutiva através de técnicas como bandeamento C e em menor escala coloração com fluorocromos base-específicos. Além disso, em diversas espécies a descrição das regiões organizadoras de nucléolo (RON), associada diretamente a presença/ativação de clusters de DNAr 45S foi realizada. Grande parte destes estudos foram realizados utilizando representantes da família Acrididae e foi observado que essencialmente a heterocromatina constitutiva localiza-se na região pericentromérica de todos os cromossomos do complemento, embora tenham sido descritos blocos

adicionais nas regiões terminais e intersticiais dos cromossomos de distintas espécies (KING e JONH 1980; CABRERO e CAMACHO 1986; LORETO e SOUZA 2000; ROCHA et al. 2004; SOUZA e MELO 2007). Além disso, frequentemente os blocos de heterocromatina tem apresentado riqueza em pares de bases G+C, com algumas variações relativas à ocorrência de blocos ricos em A+T ou neutros (BRIDLE et al. 2002; LORETO e SOUZA 2000; ROCHA et al. 2004; SOUZA e MELO 2007; LORETO et al. 2008; ANJOS et al. 2013). Adicionalmente, RONs em gafanhotos apresentam ampla variabilidade quanto ao número e precisa localização nos cromossomos, embora a presença de duas RONs por célula seja a característica mais frequente (CABRERO e CAMACHO 1986, ROCHA et al. 1997, ROCHA et al. 2004, SOUZA e MELO 2007, LORETO et al. 2005, LORETO et al. 2008).

Embora os estudos utilizando técnicas de citogenética clássica para caracterização dos DNAs repetitivos em gafanhotos tenham produzido grandes contribuições, características mais específicas e padrões de evolução destas sequências permanecem por serem elucidados. Relativo ao entendimento destes padrões utilizando mapeamento físico cromossômico, os estudos tem se restringido a análise da localização e mecanismos de diversificação das famílias multigênicas de DNAr (45S e 5S) e genes de proteínas histônicas (genes de H3 e H4), principalmente em espécies ocorrentes na Europa (CABRERO E CAMACHO 2008; CABRERO et al. 2009, CABRAL-DE-MELLO et al. 2011a). Estes estudos têm revelado padrões bem diferenciados, com grande conservação do número e localização dos genes de histona H3 e H4 e extensiva variabilidade para os genes de DNAr. Estes padrões podem estar diretamente relacionados à ocorrência de seleção purificadora ao longo da diversificação do grupo, no caso das histonas, enquanto que para o DNAr mecanismos moleculares, tais como ocorrência de elementos de transposição e recombinação ectópica, podem ter ocorrido, ocasionando o espalhamento de alguns clusters nos cariótipos (CABRERO e CAMACHO 2008; CABRERO et al. 2009; CABRAL-DE-MELLO et al. 2011a, b). Segundo Cabral-de-Mello et al. (2011a) as variações destas duas famílias multigênicas ocorreram independentemente ao longo da diversificação das espécies de gafanhotos.

Adicionalmente, o mapeamento das sequências de famílias multigênicas em gafanhotos tem auxiliado no entendimento a respeito da origem e evolução de cromossomos B como, por exemplo, em *Eyprepocnemis plorans*, *Locusta migratoria*, *Rammathocerus brasiliensis* e *Abracris flavolineata*. Em *E. plorans* o mapeamento físico cromossômico utilizando DNAr 18S e DNA satélite (180-bp) indicaram uma

origem múltipla e independente do cromossomo B em populações europeias (CABRERO et al. 2003), enquanto que na população do Caucaso, o mapeamento do DNAr 5S indicou que esse cromossomo se originou a partir do menor elemento autossômico (P11) (CABRERO et al. 2003). Além disso, em outras populações foi relatada a origem do cromossomo B a partir do cromossomo X com o uso do DNAr 45S e DNAsat (180 bp) como sonda (López-León et al. 1994). Similarmente, o uso do mapeamento dos genes de histona H3 e H4 na espécie *L. migratoria* possibilitaram traçar a origem do cromossomo B a partir do cromossomo autossômico 8, por compartilhamento de sequências entre os mesmos (TERUEL et al. 2010). Da mesma forma, através do mapeamento do DNAr 5S em *R. brasiliensis*, foi proposto que o cromossomo supranumerário desta espécie teria se originado a partir de um par autossômico (Loreto et al. 2008). Todavia, a hipótese de origem autossômica do cromossomo B foi recentemente refutada por Oliveira et al. (2011) devido a observação de sítios de DNAr 5S em diferentes cromossomos autossômicos e no X desta espécie, de forma que estes marcadores cromossômicos não são informativos a respeito da origem e evolução dos cromossomos B. Por último, através dos resultados do mapeamento dos genes DNAsn U2 em *A. flavolineata*, foi sugerido que o cromossomo B dessa espécie teria surgido a partir do par maior (par 1) e baseado em isocromossomos (Bueno et al. 2013).

Finalmente, o uso do mapeamento de DNAs repetitivos através da hibridização *in situ* fluorescente em gafanhotos tem se mostrado uma excelente ferramenta para entendimento dos padrões de diversificação cariotípica e genômica em distintos grupos de eucariotos, assim como na análise de origem e evolução de cromossomos sexuais e supranumerários. Levando-se em consideração que os gafanhotos apresentam os maiores genomas de Arthrópodes e que isso pode estar relacionado com a grande quantidade de DNA repetitivo encontrado nesses organismos, o isolamento e caracterização de distintas sequências de DNAs repetitivos, tais como o DNAsn, U1, serão úteis nas análises cromossômicas com enfoque em estudos de evolução cromossômica e dos genomas em gafanhotos.

2.0 OBJETIVOS

2.1. Objetivo geral

Entender os padrões de diversificação dos clusters de DNAsn U1 nos cariótipos de espécies de gafanhotos pertencentes a distintas linhagens e utilizar o genoma sequenciado de *Eyprepocnemis plorans* para entender a organização genômica do gene de RNAsn U1.

2.2. Objetivos específicos e metas

- Realizar o mapeamento citogenético dos genes de RNAsn U1 em espécies de gafanhotos pertencentes as famílias Proscopiidae, Pyrgomorphidae, Ommexechidae, Romaleidae e Acrididae buscando elucidar os padrões de diversificação das mesmas nos cariótipos das distintas espécies;
- Testar a possível presença deste marcador em cromossomos B e sexuais, auxiliando no entendimento da origem e evolução destes cromossomos;
- Entender a organização genômica das repetições de DNAsn U1 no genoma de *Eyprepocnemis plorans*, utilizando os dados obtidos por sequenciamento 454;
- Entender as possíveis relações a nível cromossômico e molecular de diferentes famílias multigênicas em gafanhotos, comparando os resultados obtidos para os genes RNAsn U1 com dados previamente publicados.

3.0 MATERIAL E MÉTODOS

3.1. Animais, extração de DNA e preparações cromossômicas

Indivíduos de 71 espécies de gafanhotos pertencentes às famílias Acrididae, Ommexechidae, Proscopiidae, Pyrgomorphidae e Romaleidae foram coletados de distintas localidades da Argentina, Brasil, Espanha e Paraguai (ver tabela do manuscrito 1). Os testículos dos machos foram fixados em solução Carnoy modificada (3:1 etanol absoluto:ácido acético) e armazenados em freezer -20 ºC, enquanto o ceco gástrico de algumas fêmeas foram fixados de acordo com Castillo et al. (2011). Além disso, embriões de *Eyprepocnemis plorans* foram obtidos de acordo com o procedimento descrito por Camacho et al (1991). As preparações cromossômicas foram obtidas por maceração dos diferentes tecidos em uma gota de ácido acético 45% e as lâminas foram secas em uma plataforma aquecida com temperatura entre 40-45 ºC.

3.2. Isolamento das sequências de DNAsn U1

As sequências parciais de DNAsn U1 foram obtidas através de Reação em Cadeia da Polimerase (PCR) do genoma de *Rhammatocerous brasiliensis* utilizando primers universais descritos por Cabral-de-Mello et al. (2012). Os produtos de PCR foram previamente sequenciados para confirmação do isolamento da sequência de interesse e as sequências foram depositadas no GenBank sobre o acesso KC896793.

3.3. Hibridização *in situ* fluorescente (FISH)

A técnica de FISH foi realizada de acordo com o método proposto por Pinkel et al. (1986), com modificações (Cabral-de-Mello et al. 2010), enquanto que a fiber-FISH foi executada seguindo protocolo descrito por Muñoz-Pajares et al. (2011) utilizando simples ou duplas marcações com sondas de DNAsn U1 ou DNAsn U1 e DNAr 5S, respectivamente. As fragmentos de DNA foram marcados com digoxigenin-11-dUTP (Roche) ou diotin-14-dUTP (Invitrogen) através de PCR ou nick-translation e os produtos de PCR foram visualizados por eletroforese em gel de agarose 1% para verificar a amplificação das sequências. Os sinais da FISH foram detectados utilizando anti-digoxigenin-rhodamine (Roche) ou alexa-fluor-488 (Life Technologies) e as preparações foram contra coradas com 4′,6-diamidine-2′-phenylindole dihydrochloride (DAPI) e posteriormente montadas usando Vectashield (Vector). Os cromossomos e os sinais da FISH foram observadas usando microscópio Olympus BX61 equipado com

uma lâmpada fluorescente e filtros apropriados acoplados a uma câmera digital DP70. As imagens tiveram ajustes de brilho/contraste no Adobe Photoshop CS2 software. Finalmente, os sinais observados na FISH foram organizados em cinco categorias: centromeric (c), proximal (pr), intersticial (i), sub-distal (sd) e distal (d), assim como observado no diagrama do "supplementary material S2".

3.4. Análise da organização genômica do DNAsn U1 no genoma de *Eyprepocnemis plorans*

As análises da variação intragenômica do DNAsn U1 no genoma de *E. plorans* foram realizadas utilizando uma biblioteca de sequenciamento do genoma inteiro (WGS) que consistiu de 3/8 de uma placa de "454 GS FLX Plus reads" (número de acesso XXXXX). Foi feito um gráfico básico de agrupamento da leitura 454 a fim de procurar genes associados ao DNAsn U1 no genoma utilizando o software RepeatExplorer (Novak et al. 2013). Os contigs gerados foram separados para busca no RepBase utilizando o programa CENSOR (Kohany et al. 2006) e BRLASTN no banco de dados não redundante do NCBI. Foram considerados apenas os resultados coincidentes para ambos os métodos de busca. Os contigs contendo DNAsn U1 e outros contigs separados foram representados usando o softwere SeqGrapheR (Novak et al. 2010). As sequências usadas foram extraídas e montadas separadamente a fim de alinhar e cortar manualmente a região do DNAsn U1 utilizando o software Geneious v4.8 (Drummond et al. 2009).

A árvore de "minimum-spanning" (MST) para as sequências DNAsn U1 encontradas nos contigs foi construída com Arlequin v3.5 (Excoffier & Lischer, 2010), utilizando como grupo externo a sequência mais similar com as nossas sequências de DNAsn U1 encontradas na busca com o BLASTN (Altschul et al. 1997) no banco de dados não redundante do NCBI. As análises de diversidade de sequência foi conduzida utilizando DNASP (Librado & Rozas, 2009) e o conteúdo de G+C foi calculado com o software MEGA v5 (Tamura et al, 2011). A predição da estrutura secundária e a energia livre de Gibb's (dG) para a distinção dos haplótipos observados foram realizadas com UNAFold (Markham et al., 2008) com as opções padrão, exceto para temperatura de dobramento de 30 ºC.

4.0 RESULTADOS E DISCUSSÃO

Os resultados da presente dissertação estão apresentados no manuscrito a seguir a ser submetido a revista ***Heredity*** (ISSN: 0018-067X).

4.1 Capítulo 1:

U1 snDNA clusters in grasshoppers: chromosomal dynamics and genomic organization

Allison Anjos, Francisco J Ruiz-Ruano, Juan Pedro Martinez Camacho, Vilma Loreto, Josefa Cabrero, Maria José de Souza, Diogo Cavalcanti Cabral-de-Mello

Manuscrito a ser submetido a revista Heredity

U1 snDNA clusters in grasshoppers: chromosomal dynamics and genomic organization

Allison Anjos[1], Francisco J Ruiz-Ruano[2], Juan Pedro Martinez Camacho[2], Vilma Loreto[3], Josefa Cabrero[2], Maria José de Souza[3], Diogo Cavalcanti Cabral-de-Mello[1,*]

[1]UNESP - Univ Estadual Paulista, Instituto de Biociências/IB, Departamento de Biologia, Rio Claro, São Paulo, Brazil;

[2]UGR – Univ de Granada, Facultad de Ciencias, Departamento de Genética, Granada, Spain.

[3]UFPE – Univ Federal de Pernambuco, Centro de Ciências Biológicas/CCB, Departamento de Genética, Recife, Pernambuco, Brazil;

Short running title: U1 snDNA clusters in grasshoppers

**Corresponding author:* UNESP - Univ Estadual Paulista, Instituto de Biociências/IB, Departamento de Biologia, CEP 13506-900 Rio Claro, SP, Brazil Phone/Fax: 55 19 35264152. e-mail: mellodc@rc.unesp.br

Abstract

The spliceosome is responsible for mRNA maturation through intron removal. This machinery is formed by a protein set associated to small nuclear RNA (snRNA). Among the snRNA genes, the U1 snRNA is, in general, a conserved 165 bp tandemly arrayed repetitive sequence. Aiming to contribute to the understanding of chromosome and genome dynamics of multigene families in grasshoppers we mapped the U1 snRNA genes by Fluorescent *in situ* Hybridization (FISH) in 71 grasshopper species belonging to the families Proscopiidae, Pyrgomorphidae, Ommexechidae, Romaleidae and Acrididae. Moreover we analyzed the genomic organization for this sequence using as reference the sequenced genome through 454 of *Eyprepocnemis plorans*. High conservation of snDNA clusters mainly located on autosome pairs (no. 3 or 4) was observed in the first four families. In contrast, extensive variation was observed in Acrididae species, from a single chromosome pair carrying U1 snDNA to all chromosome pairs carrying them, with occurrence of two or multiple clusters in the same chromosomes. In the genome of *E. plorans* five distinct lineages were observed with distinct patterns of variability and association of U1 snDNA with transposable elements and 5S rDNA was also noticed. These results are discussed focusing the possible mechanisms of spread of this gene cluster, which apparently seems to have followed different ways of dispersion in the several families and subfamilies analyzed in here. This is the most comprehensive study on FISH mapping hitherto performed in grasshoppers and other organisms by studying 71 species from five families and has thus provided valuable information shedding light in the chromosomal/genomic evolution of this gene family by combined use of chromosomal and genomic data.

Key words: FISH, genome evolution, multigene family, repetitive DNAs, spliceosome

Introduction

In higher eukaryotes, most protein-coding transcripts contain multiple introns which need to be removed from the nascent RNA transcripts through *splicing*, being this essential mechanism involved in messenger RNA (mRNA) maturation (West 2012). The spliceosome consists of a protein set associated with U small nuclear RNAs (snRNAs) that are crucial components of this macromolecular complex. The major spliceosome complex is coded by a multigene family including the U1, U2, U4, U5 and U6 snRNAs genes (Gilbert 1978; Busch et al. 1982; Bringmann and Lührmann 1986; Nilsen 2003; Valadkhan 2005). Among the U snRNA genes, the U1 snDNA is a conserved 165 bp sequence, although distinct U1 snDNA variants have been found in several organisms, like toad (Forbes et al. 1984), fruit fly (Lo and Mount 1990), pea (Hanley and Schuler 1991), silk moth (Sierra-Montes et al. 2003) and human beings (O'Reilly et al. 2012).

From a molecular point of view, the U1 snRNA gene is tandemly arranged in some species, such as the toad *Xenopus laevis* (Zeller et al. 1984), the sea urchin *Strongylocentrotus purpuratus* (Yu et al. 1991), the tapeworm *Echinoccocus multilocularis* (Bretagne et al. 1991), and some teleost fish (Marz et al. 2008), among others. In addition, the U1 snRNA has been found linked to other gene repeats (e.g. 5S rDNA) in the crustacean *Asellus aquaticus* (Pelliccia et al. 2001), the fish *Solea senegalensis* (Manchado et al. 2006) and 10 razor shell species (Vierna et al. 2011). From the chromosomal point of view, the mapping of U1 snRNA genes has been performed only in a few species including, for instance, humans (Lund et al. 1983), mice (Lund and Nesbitt 1988), crustaceans (Barzotti et al. 2003) and fish (Cabral-de-Mello et al. 2012), and revealing that these genes are conserved in location, occupying one or few chromosome pairs.

Chromosomal studies in grasshopper species have demonstrated an extensive karyotype conservation with most species showing 2n=23♂/24♀, with predominantly acrocentric chromosomes and an X0♂/XX♀ sex chromosome system. Furthermore, derived karyotypes showing reduction of diploid number, derived sex chromosomes and supernumerary elements have also been reported (White 1978; Hewitt 1979; Mesa et al. 1982; Castillo et al. 2010a). On the other hand, current knowledge about repetitive DNA organization in grasshopper chromosomes is still scarce and concerning multigene families only the 45S rDNA (Cabrero and Camacho 2008), H3 and H4 histone genes (Cabrero et al. 2009) and 5S rDNA (Cabral-de-Mello et al. 2011a), were mapped mainly in Acrididae species from Europe.

In order to advance in the knowledge of chromosomal evolution and organization of multigene families in grasshopper karyotypes/genomes, we performed here the FISH mapping of the U1 snRNA gene in 71 species of grasshoppers belonging to five distantly related families according to the most recent phylogeny proposed for the group (Leavitt et al. 2013), i.e. Proscopiidae, Ommexechidae, Pyrgomorphidae, Romaleidae and Acrididae (the latter including ten subfamilies). Moreover it was analyzed the data of 454 genome sequencing obtained from *Eyprepocnemis plorans* in order to advance in the knowledge of specific genome organization and variability for U1 snDNA arrays and their relationship with other DNA sequences, like Transposable Elements (TEs) and other multigene familes. Our data are discussed with the aim to shed light on the understanding of diversification patterns for U1 snRNA genes in grasshopper genomes.

Material and methods

Chromosomes, probe obtaining and Fluorescent in situ *hybridization (FISH)*

Adult males and females of 64 species of grasshoppers belonging to the families Proscopiidae (four species), Ommexechidae (two species), Pyrgomorphidae (two species), Romaleidae (11 species) and Acrididae (45 species belonging to ten subfamilies) were sampled from distinct localities in Spain, Paraguay, Argentina and Brazil (see Supplementary Material S1). Testes were fixed in 3:1 absolute ethanol:acetic acid and stored at –20 °C until use. Moreover, embryos of *Eyprepocnemis plorans* were also obtained as described by Camacho et al (1991) and female gastric caeca were removed and fixed following Castillo et al. (2011). The different tissues were macerated in a drop of 50% acetic acid and the slides were dried using a hot plate at 40-45 ºC. In addition, previous results for *Abracris flavolineata* (Bueno et al. 2013) and six other species (Palacios-Gimenez et al. 2013) were also included for analysis.

The DNA probe for the U1 snDNA sequence was obtained through Polymerase Chain Reaction (PCR) from the genome of *Rhammatocerous brasiliensis* using the primers described by Cabral-de-Mello et al. (2012) and the 5S rDNA was obtained from a previously cloned fragment isolated from *E. plorans* genome. The PCR product was previously sequenced and the sequence deposited in GenBank under the accession number KC896793 (Bueno et al. 2013). The U1 snDNA and 5S rDNA fragment were labeled with digoxigenin-11-dUTP (Roche, Mannheim, Germany) or biotin-14-dUTP (Invitrogen, San diego, CA, USA) through PCR or nick-translation. The FISH procedures were performed according to Pinkel et al. (1986), with some modifications (Cabral-de-Mello et al. 2010), and for fiber-FISH it was followed the protocol described by Muñoz-Pajares et al. (2011). The probe was detected using Anti-digoxigenin-rhodamine (Roche, Mannheim, Germany). The chromosomes were counterstained with 4,6-diamidino-2-phenylindole (DAPI) and mounted in Vectashield (Vector, Burlingame, CA, USA).

The chromosomes and signals were observed using an Olympus microscope BX61 equipped with a fluorescence lamp and appropriate filters. The photographs were recorded using a DP70 cooled digital camera and images were merged and optimized for brightness and contrast using Adobe Photoshop CS2. Finally the signals were analyzed and organized into five categories: centromeric (c), proximal (pr), interstitial (i), sub-distal (sd) and distal (d), as shown in the diagram in Supplementary Material S2.

Genomic analysis of U1 and related repetitive elements in the* E. plorans *genome

It was performed an analysis of the intragenomic variation of U1 snDNA in the genome of *E. plorans* using a Whole Genome Sequencing (WGS) library consisting in 3/8 plate of 454 GS FLX Plus reads (accession number XXXXX). We performed a graph based clustering of the 454 reads to search associated genes to U1 snDNA in the genome using the RepeatExplorer software (Novak et al. 2013). The generated contigs were annotated with a search in RepBase with CENSOR program (Kohany et al. 2006) and a BLASTN search (Altschul et al. 1997) in the NCBI NR database. It was considered only those matches resulting with both methods. Then searching for other reads containing U1 snDNA out of the detected cluster was performed.

The contings containing U1 snDNA and other anotated contings were represented using the SeqGrapheR (Novak et al. 2010) software. We extracted the sequences used to assembly each contig separately in order to align them and manually trimming the U1 snDNA region with Geneious v4.8 (Drummond et al. 2009).

A minimum-spanning tree (MST) was built with Arlequin v3.5 (Excoffier & Lischer, 2010) for the U1 snDNA sequences found in the contigs, using as outgroup the sequence showing the highest similarity with our U1 snDNA sequences in a BLASTN search (Altschul et al. 1997) in the NCBI nr database. The analysis for sequence

diversity was conducted using DNASP (Librado & Rozas, 2009). The G+C content was calculated with the MEGA v5 (Tamura et al, 2011) software, and the prediction of secondary structure and Gibb's free energy (dG) for the distinct haplotypes observed was performed with UNAFold (Markham et al., 2008) with default options except a folding temperature of 30ºC.

Results

Chromosomal mapping of U1 snDNA

The diploid chromosome number and sex chromosome system for the 71 species (some give here for the first time) are shown in Supplementary Material S1. A total of 171 U1 snDNA sites were detected by FISH in the 71 species included here, with 2.4, on average, per haploid genome. Most of these sites (165) were located in autosomes whereas only five were in the X chromosome and one in a neo-Y chromosome. Most sites were located proximal to centromere (62 sites, 36.2%), and less of them were interstitial (39, 22.8%), centromeric (35, 20.5%), sub-distal (21, 12.3%) or distal (14, 8.2%) (Table 1). Although most species carried U1 snDNA clusters in only one autosomal pair, variable patterns were also observed, even including two species with clusters in all chromosome pairs. Furthermore, in several species, one (e.g. *Ommexecha virens*, *Agriacris auripennis*, *Chorthippus apicalis*, *Rhammatocerus brasiliensis*, *Eumastusia koebelei koebelei*, *Ronderosia bergi*) or two (e.g. *Omocestus bolivari*, *Omocestus burri*, *Omocestus panteli* and *Stenobothrus festivus*) chromosome pairs carried two or more clusters of U1 snDNA in a same chromosome (Table 1).

The five grasshopper families analyzed showed distinct patterns of distribution for the U1 snDNA clusters. In Proscopiidae, all four species analyzed showed a single interstitial cluster in the autosome pair no. 4 (Figure 1a,b). Likewise, the two

Pyrgomorphidae species analyzed showed a single cluster proximally located in the autosome pair no. 3 (Figure 1c,d). The two Ommexechidae representatives also showed a coincident pattern but, in this case, they showed three sites per species, one of them proximally located in autosome pair no. 1, and the two other at centromeric and interstitial regions in the autosome pair no. 3 (Figure 1e).

Among Romaleidae, 7 out of the 11 species analyzed showed a single cluster of U1 snDNA in the autosome pair no. 3, being proximal in 6 species and sub-distal in 1 species (Figure 2). However, *Agriacris auripennis* showed two clusters (centromeric and proximal) in the autosome pair no. 8 (Figure 2a), the two *Chromacris* species showed clusters in a single pair proximal in autosome no. 5 in *C. nuptialis*, and sub-distal in autosome no. 4 in *C. speciosa* (Figure 2c,d). Finally, *Brasilacris gigas* showed a single interstitial cluster pair no. 7 (Figure 2b).

In the Acrididae family, the most representative sample analyzed with 52 species from 10 subfamilies, we found 147 U1 snDNA clusters, with 2.83 per haploid genome on average (Figure 3, 4, Table 1). This analysis showed that 48 of these sites were proximal (32.6%), 34 were interstitial (23.1%), 32 were centromeric (21.7%), 19 were sub-distal (13%) and 14 were distal (9.5%). About 40% of the species (21) showed U1 snDNA in a single chromosome pair, but some species showed it in two or more pairs, with two species (*Oedipoda fuscocincta* and *Sphingonotus caerulans*) carrying it in all chromosome pairs (see Fig. 3l). In several species, two or more clusters were located in the same chromosome (Figure 4, Table 1).

Among Acrididae, three subfamilies were more represented, i.e. Gomphocerinae (16 species), Leptysminae (7 species) and Melanoplinae (14 species). In Leptysminae, we found 1-3 clusters per species and they were predominantly interstitial (6 clusters) and proximal (4), but rarely centromeric (1) and never subdistal or distal, with two

species (*Cylindrotettix obscurus* and *Stenopola* sp) showing clusters in the X chromosome (Figure 3e, Table 1). In Melanoplinae, almost all species showed only one autosomal pair carrying the genes for U1 snRNA, but they never showed centromeric location (Table 1). However, *Propedies auriculares* and *Ronderosia bergi* carried two clusters in a same chromosome (Figure 3i, Table 1). Finally, Gomphocerinae species showed 1-6 U1 snDNA clusters, always in autosomes, with all types of location but higher frequency in proximal (21 clusters), interstitial (18) and subdistal (14) locations, and lower frequency in centromeric (7) and distal (6) locations (Figure 4, Table 1). Remarkably, all *Chorthippus* and *Omocestus* species carried 2, 3 or more U1 snDNA clusters (see Figure 4) in the largest autosome (no. 1), which could have been present in their common ancestor. The average number of U1 snDNA clusters per haploid genome was 1.6, 1.6 and 4.12 in the subfamilies Leptysminae, Melanoplinae and Gomphocerinae, respectively.

Finally, some acridid species carried B chromosomes, such as *Eyprepocnemis plorans*, *Cylindrotettix attenuatus*, *Eumastusia koebelei koebelei*, *Stenopola dorsalis*, *Rhammatocerus brasiliensis*, *Abracris flavolineata* (Bueno et al. 2013), *Orthoscapheus rufipes* and *Vilerna rugulosa*, but no clusters of U1 snDNA were observed in any of these elements (see for example Figure 3f,j,k).

U1 snDNA and other repetitive elements in the* E. plorans *genome

With the graph based clustering, we got a complex cluster with U1 anDNA in central position and multitude of “loops” and “branches” (Figure 5) connected with it, as expected from its multiple locations in the *E. plorans* genome. This cluster constituted 0.222% of the whole genome. Since we found no more sequences of U1

snDNA out of the cluster we have only considered the reads in this cluster. The assembly gave a total of 63 contigs, five of them carrying U1 snDNA.

On the basis of its coding region, we found complete U1 snRNA genes in two contigs (lineages 1 and 2), but there were several others showing different mutations, such as a deletion in the middle (lineage 3), a 5' truncation (lineage 4), and a deletion and an insertion close to the 5' end (lineage 5) (Figure 6a). We predicted the secondary structure of the five U1 snDNA consensus sequences as predictor of functionality or pseudogenicity (Figure 6b). Lineages 1 and 2 showed the four conserved sites described in Vierna et al. (2013) (Figure 7). However, we observed a difference in folding pattern in respect to other reported secondary structures of U1 snDNA, i.e. the pairing of a part of the Box A with a part of the Sm-binding site. The secondary structure of lineage 1 showed higher stability (ΔG = -66.7 kcal/mol) than lineage 2 (ΔG = -61.4 kcal/mol). On the contrary, lineages 3, 4 and 5 show secondary structures lacking the conserved motives present in the two other lineages, even though lineage 5 shows a dG value being very similar to that of lineage 2 (Figure 6b). The analysis of diversity and possible functionality for each lineage (Table 2) indicates high similarity for lineages 1 and 3 (π = 0) but high heterogeneity for lineages 2 (π = 0.01987), 4 (π= 0.04879) and 5 (π = 0.01242),

We built a MST with all five lineages of U1 (Figure 6b). As NCBI's nucleotide database lacks full-length copies of U1 snDNA from Orthoptera we searched the sequences of lineages 1 and 2 in the NR database of the NCBI and we chose that U1 snDNA sequence showing the lowest e-value for both lineages, which corresponded to that in the mountain pine beetle *Dendroctonus ponderosae* (accession number APGK01030738.1, position from 13746 to 13908). The MST shows the lineage 1 as the

more connected and directly joined to the outgroup, suggesting that it is ancestral in *E. plorans* genome.

In addition to U1 snDNA, we found other identifiable genes in the cluster. The contig with the highest coverage corresponded with lineage 4 of U1 snDNA and showed association with the 5S rDNA. Both were present in the same orientation in the 3,191 bp long contig with overlapping ends, a typical assembly of tandem repeat elements in the genome. After joining contig ends, and trimming one of them, the resulting sequence was 2,879 bp long with a spacer 1 of 1,209 bp and a spacer 2 of 1,421 bp. The spacer 1 also showed a tandemly duplicated region of about 350 bp. In addition, we found two more contigs including a region annotated as 5S rDNA (Figure 5). This association of U1 snDNA and 5S rDNA repeats was corroborabed by analysis using two color FISH in mitotic metaphase and in distended fibers (fiber-FISH) (Figure 8). In addition to 5S rDNA, we found some regions in the cluster, being very connected with U1 snDNA, which could be SINE-like elements associated to the U1 snDNA cluster. For instance, we annotated a contig as tRNA-thr/pseudogenic, and another with homology with ALPINE/SINE (Figure 5).

Discussion

U1 snDNA organization in grasshopper chromosomes

Our data revealed clearly two main patterns of organization for the U1 snDNA clusters: (i) A single cluster in an autosomal pair, as observed in most species of Proscopiidae, Pyrgomorphidae, Romaleidae and some species of Acrididae; and (ii) increased numbers of U1 snDNA clusters, ranging from two sites to all chromosomes of the complement (Figure 9), mainly observed in Acrididae.

A comparative analysis of the results in the five families analyzed indicates that it is remarkable that the presence of a single cluster for U1 snDNA in one autosomal pair is the modal and median number since it was found in half of the species analyzed. Therefore, we hypothesize that it is the ancestral pattern in grasshoppers, and the divergent situations observed in specific groups could be result of molecular mechanisms moving some U1 snRNA genes between non-homologous chromosomes, as observed for other multigene families (see discussion below). The ancestrally of a single U1 snDNA cluster per haploid genome is reinforced by the fact that this pattern was invariant in the family Proscopiidae, which is included in the superfamily Eumastacoidea, considered the earliest diverging lineage of Acridomorpha with basal placement in molecular phylogenies (Leavitt et al. 2013). Likewise in grasshoppers, the occurrence of a single site of U1 snDNA per haploid genome has also been reported in mouse (Lund and Nesbitt 1988), crustaceans (Barzotti et al. 2003) and fish (Cabral-de-Mello et al. 2012). Although we suggested the occurrence of one autosomal pair harboring the U1 snRNA as the ancestral condition, it was not possible to determinate precisely the ancestral specific pair and the position of the cluster in this element. However the occurrence of clusters in the pair no. 3 was common in Pyrgomorphidae and in Acridoidea representatives (i.e. Ommexechidae, Romaleidae and Acrididae) that are sister groups and it could represent the ancestral condition for these groups, emerging after the divergence of Eumastacoidea. In the same way, it seems to be a common patterns clusters located in the proximal (36.2%), interstitial (22.8%) and centromeric (12.3%) regions; however we can not determine the exactly ancestral placement for the cluster.

Grasshopper genomes usually consists of a gradual series of morphologically similar chromosomes slightly differing in size, so that it is difficult to identify each

chromosome pair (except the X chromosome and the megameric bivalent, because their heteropycnotic behaviour during meiosis). This makes it impossible to know whether chromosomes occupying the same size order position (e.g. the third) in two different genomes are actually homeologous. This implies that chromosome numbers in Table 1 are only strictly valid per species, but extrapolation among species needs to be made with extreme care. With this problem in mind, we can reach two main conclusions: i) The single U1 snDNA cluster in the ancestor of the grasshopper species analyzed here was not located in the X chromosome, since this location is extremely rare. ii) It is likely that the ancestral location of the U1 snDNA cluster was in an autosome of a size being about the third in the genome, since this was the size order position being found more frequently (in 30 species). In addition, the adjacent positions (2^{nd} and 4^{th}) were also very frequently found (14 and 17 species), suggesting that, bearing in mind the uncertainty in determining homeology, in some species, the autosome 2 or 4 could actually be derived from the ancestor chromosome 3, and its size changed due to some genomic events such as, for instance, a higher proliferation for a given mobile element. On this basis, chromosome 4 in the Proscopiidae could be homeologous of chromosome 3 in Romaleidae. This fact can explain many of the differences among species, although the increase in the number of clusters per haploid genome needs additional explanations, as mobile elements, chromosome rearrangements, etc.

For the family Proscopiidae the conservative organization evidenced for U1 snRNA, i.e. an interstitial cluster located in the pair no. 4, and the occurrence of H3 histone and 5S rRNA genes in this same chromosome (Cabral-de-Mello et al. 2011b) indicate that this chromosome could be maintained without gross rearrangements after the speciation process. In a similar way, at family level, this conservative pattern for distribution of U1 snDNA clusters was also found in Pyrgomorphidae (pair 3),

Ommexechidae (pairs 1 and 3) and Romaleidae (pair 3). Furthermore, at genus level similarity was also observed, like in Ommexechidae (*Ommexecha gracilis* and *O. virens*); Romaleidae (*Xyleus discoideus angulatus* and *X. d. discoideus*); and Acrididae representatives, i.e. *Parascopas* (*P. obesus* and *P. sanguineus*) and *Schistocerca* (*Sh. flavofasciata* and *Sh. pallens*) which may be an indicative for a possible common descendent and conservation of the U1 snRNA genes distribution in these genera.

In Acrididae several species from some subfamilies showed clusters in one or two autosomal bivalents (e.g. Acridinae, Catantopinae, Copiocerinae, Cyrtacanthacridinae and Ommatolampinae), although, no specific pattern of distribution in a specific chromosome emerged in these groups. In general, apparently, the variability for U1 snDNA number and location do not have phylogenetic signals beyond the genus level, being the variation particular for each genus, except in the case of Gomphocerinae in which spreading was observed in most representatives (see discussion below). Examples of this independent variation could be evidently observed, for example, in Oedipodinae with *Aiolopus strepens* and *Locusta migratoria* presenting clusters restricted to one chromosome pair, contrarily with *Oedipoda fuscocincta* and *Sphingonotus caerulans* with clusters in all chromosomes; in Ommatolampidinae each species presented an specific pattern; and in Melanoplinae and Leptysminae, in addition to autosomal variability, sex located clusters were observed in *Cylindrotettix obscurus* and *Stenopola* sp. and *Eurotettix minor*, the latter showing U1 snDNA clusters in both the neo-X and the neo-Y most likely due to their presence in the autosome fused to the X chromosome (Palacios-Gimenez et al. 2013).

The Gomphocerinae constitutes one of the most diverse subfamilies within Acrididae (Contreras and Chapco 2006) and includes a group of species with reduction of diploid number to 2n=17 due to three centric fusion events (Hewitt 1979). Likewise,

we observed the highest number of U1 snDNA clusters per haploid genome (4.12) within the Acrididae, which is comparatively much higher than the 1.6 found in Leptysminae and Melanoplinae. Apparently this capacity of movement and multiplication for U1 snDNA in Gomphocerinae is basal in the group and preceded the occurrence of diploid number reduction, although *O. punctata* retains the modal character with only one site for U1 snDNA. Additionally in the species with 2n=17 this mechanisms were intensified in the pairs 1 and 2, that are resultant from centric fusions, suggesting that some of the acrocentric chromosomes involved in the fusions already carried U1 snRNA genes in the ancestor species.

***Possible causes of U1 snDNA variation and its organization in the genome of* E. plorans**

As general aspect although the occurrence of one U1 snDNA site is common in grasshoppers there is an evident multiplication in number and variability in position for these clusters, mainly observed in Acrididae. This variation at interspecific level resembles the extensive variability for rDNA loci previously reported in Acrididae grasshoppers (Cabrero and Camacho 2008; Cabral-de-Mello et al. 2011a). This, as for rDNAs, could be a consequence of mechanisms such translocations or inversions, ectopic recombination, occurrence of eccDNAs and even action of TEs (Charlesworth et al. 1994; Coyne and Orr 1998; Raskina et al. 2008). These cited mechanism were proposed as also responsible for the wide range in the distribution and position of major rDNA clusters in other insects like bugs (Panzera et al. 2012), moths and butterflies (Nguyen et al. 2010) and beetles (Cabral-de-Mello et al. 2011c).

The possible role of TEs in U1 snDNA spread could be corroborated by the analysis of *E. plorans* genome that revealed association of U1 snDNA with this kind of

repetitive DNA. The transposition mechanism could also be responsible to the linked organization observed for U1 snDNA and 5S rDNA in *E. plorans*, that moves within genomes as reported by Drouin and Moniz de Sá (1995). The association with SINE elements is possibly an explanation for the wide spread of U1 snDNA across the *E. plorans* genome, compared to other species, bearing also in mind that pseudogenic copies of 5S rDNA could acquire SINE behavior (Kapitonov et al., 2003; Gogolevsky et al., 2009). The genomic linkage for these multigene families have also been reported for example in crustaceans (Pelliccia et al. 2001), molluscs (Vierna et al. 2011) and fish (Manchado et al. 2006). According to Vierna et al. (2011) the linkage of distinct multigene families is resultant of stochastic events and occurred several times in distinct lineages and until now no benefits were reported for this condition.

Besides possible transposition and association with 5S rDNA some variability for U1 snDNA in respect to the sequence (with occurrence of pseudogenes) was observed here in the *E. plorans* genome, indicating complex organization and evolution for these repeats. The occurrence of possible pseudogenes (lineages 3-5) is in accordance with the process of birth-and-death evolution (Nei and Rooney 2005), as observed in other insect genomes, as *Aedes aegypti* (Mount et al. 2007). On the other hand, apparently the concerted mode of evolution is also playing role in the U1 snDNA copies in *E. plorans* genome, corroborated by the occurrence of homogenized copies (lineages 1 and 3), moreover the purifying selection could be acting in the lineage 1, that due its characteristics could be functional. The mixed model of evolution have also been proposed for other multigene families like rDNAs in distinct groups (Freire et al. 2010; Pinhal et al. 2011; Merlo et al. 2012; Vierna et al. 2010, 2013).

Conclusions

The analysis of five distinct families and ten subfamilies from most specious grasshopper family, i.e. Acrididae, allowed a more precise picture concerning the organization of U1 snDNA clusters in grasshoppers, permitting a deeper knowledge for karyotype organization of this sequence in the group as a whole. The high genomic dynamism for U1 snDNA clusters contrast with the extensive conservation of macro-chromosomal structure in grasshoppers, indicating that the variability for this gene is not associated with major chromosomal rearrangements, and occurred mainly in a short period of time as evidenced in Acrididae that was originated about 60 Mya (Hewitt 1979). As demonstrated by the analysis of *E. plorans* genome the high dynamism of this gene in Acrididae could be consequence of association with TEs. The next perspective is the analysis of distribution of other U snDNAs in grasshopper karyotypes to a more precise inference of evolutionary fates for this multigene family at chromosomal and molecular point to views.

Acknowledgements

We are grateful to Dr. Carlos Salvador Carbonell, Universidad de Montevidéo, Uruguai, for the taxonomic identification of some specimens used here. This study was supported by the Programa Primeiros Projetos (PROPe, UNESP), Fundação de Amparo a Pesquisa do Estado de São Paulo-FAPESP (2011/19481-3), Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (475308/2011-5), the Spanish Ministerio de Ciencia e Innovación (CGL2009-11917) and Plan Andaluz de Investigación (CVI-6649), and was partially performed by FEDER funds. A. Anjos and F.J. Ruíz-Ruano were supported by scholaships from the Brazilian CNPq and the Spanish Junta de Andalucia, respectively.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Research, 25: 3389-3402.

Barzotti R, Pelliccia F, Rocchi A (2003) Identification and characterization of U1 small nuclear RNA genes from two crustacean isopod species. Chromosome Res 11: 365-373.

Bretagne S, Robert B, Vidaud D, Goossens M, Houin R (1991) Structure of the *Echinococcus multilocularis* U 1 snRNA gene repeat. Mol Biochem Parasitol 46: 285-292.

Bringmann P, Lührmann R (1986) Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. EMBO J 5: 3509–3516.

Bueno D, Palacios-Gimenez OM, Cabral-de-Mello DC (2013) Chromosomal mapping of repetitive DNAs in *Abracris flavolineata* reveal possible ancestry for the B chromosome and surprisingly H3 histone spreading. PLoS ONE 8: e66532.

Bush H, Reddy R, Rothblum L, Choi YC (1982) SnRNAs, SnRNPs, and RNA processing. Ann Rev Bioche 51: 617-54.

Cabral-de-Mello DC, Moura RC, Martins C (2010) Chromosomal mapping of repetitive DNAs in the beetle *Dichotomius geminatus* provides the first evidence for an association of 5S rRNA and histone H3 genes in insects, and repetitive DNA similarity between the B chromosome and A complement. Heredity 104: 393–400.

Cabral-de-Mello DC, Cabrero J, Loópez-Leoón MD, Camacho JPM (2011a) Evolutionary dynamics of 5S rDNA location in acridid grasshoppers and its relationship with H3 histone gene and 45S rDNA location. Genetica 139: 921-931.

Cabral-de-Mello DC, Martins C, Souza MJ, Moura RC (2011b) Cytogenetic mapping of 5S and 18S rRNAs and H3 histone genes in 4 ancient Proscopiidae grasshopper species: contribution to understanding the evolutionary dynamics of multigene families. Cytogenet Genome Res 132: 89-93.

Cabral-de-Mello DC, Oliveira SG, Moura RC, Martins C (2011c) Chromosomal organization of the 18S and 5S rRNAs and histone H3 genes in Scarabaeinae coleopterans: insights into the evolutionary dynamics of multigene families and heterochromatin. BMC Genetics 12: 88.

Cabral-de-Mello DC, Valente GT, Nakajima RT, Martins C (2012) Genomic organization and comparative chromosome mapping of the U1 snRNA gene in cichlid fish, with an emphasis in *Oreochromis niloticus*. Chromosome Res 20: 279–292.

Cabrero J, Camacho JP (2008) Location and expression of ribosomal RNA genes in grasshoppers: abundance of silent and cryptic loci. Chromosome Res 16: 595-607.

Cabrero J, López-Leoón MD, Teruel M, Camacho JP (2009) Chromosome mapping of H3 and H4 histone gene clusters in 35 species of acridid grasshoppers. Chromosome Res 17: 397-404.

Camacho JPM, Cabrero J, Viseras E, Lopez-Leon MD, Navas-Castillo J, Alche JD (1991) G banding in two species of grasshopper and its relationship to C, N, and fluorescence banding techniques. Genome 34: 638-643.

Castillo ERD, Bidau CJ, Martí DA (2010a) Neo-sex chromosome diversity in neotropical melanopline grasshoppers (Melanoplinae, Acrididae). Genetica 138: 775-786.

Castillo ER, Taffarel A, Martí DA (2011) Una técnica alternativa para el cariotipado mitótico en saltamontes: bandeo C y Fluorescente en *Adimantus ornatissimus* (Orthoptera: Acrididae). Rev Cienc Tecnol 16: 31–35.

Charlesworth B, Sniegowski P and StephanW (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371: 215-220.

Contreras D, Chapco W (2006) Molecular phylogenetic evidence for multiple dispersal events in Gomphocerine grasshoppers. J Orthopt Res 15: 91-98.

Coyne JA, Orr HA (1998) The evolutionary genetics of speciation. Phil Trans R Soc Lond 353: 287-305.

Drouin G, Moniz de Sá M (1995) The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. Mol Biol Evol 12(3): 481-493

Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M (2009) Geneious 4.8. Biomatters, Auckland, New Zealand.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10: 564-567.

Forbes DJ, Kirschner MW, Caput D, Dahlberg JE, Lund E (1984) Differential expression of multiple U1 small nuclear RNAs in oocytes and embryos of *Xenopus laevis*. Cell 38: 681-689.

Freire R, Arias A, Insua AM, Méndez J, Eirın-López JM (2010). Evolutionary dynamics of the 5S rDNA gene family in the mussel Mytilus: mixed effects of birth-and-death and concerted evolution. J Mol Evol 70: 413–426.

Gilbert W (1978) Why genes in pieces? Nature 271: 501.

Gogolevsky KP, Vassetzky NS, Kramerov DA (2009) 5S rRNA-derived and tRNA-derived SINEs in fruit bats. Genomics 93(5): 494-500.

Hanley BA, Schule MA (1991) Developmental expression of plant snRNAs. Nucleic Acids Res 22: 6319-6325.

Hewitt GM (1979) Grasshoppers and crickets. Animal Cytogenetics. vol 3: Insecta 1. Orthoptera. Berlin: Gebrüder Borntraeger.

Kapitonov VV, Jurka J (2003) A novel class of SINE elements derived from 5S rRNA. Mol Biol Evol 20(5): 694-702.

Kohany O, Gentles AJ, Hankus L, Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics, Oct 25, 7: 474.

Leavitt JR, Hiatt KD, Whiting MF, Song H (2013) Searching for the optimal data partitioning strategy in mitochondrial phylogenomics: A phylogeny of Acridoidea (Insecta: Orthoptera: Caelifera) as a case study. Mol Phylogenet Evol 67: 494-508.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25(11): 1451-1452.

Lo PC, Mount SM (1990) Drosophila melanogaster genes for U1 snRNA variants and their expression during development. Nucleic Acids Res. 18: 6971–6979.

Lund E, Bostock C, Robertson M, Christie S, Mitchen JL, Dahlberg JE (1983) U1 small nuclear RNA genes are located on human chromosome1 and are expressed in mouse–human hybrid cells. Mol Cell Biol 3: 2211–2220.

Lund E, Nesbitt MN (1988) The embryonic and adult mouse U1 snRNA genes map to different chromosomal loci. Som Cell Mol Genet 14: 143–148.

Manchado M, Zuasti E, Cross I, Merlo A, Infante C, Rebordinos L (2006) Molecular characterization and chromosomal mapping of the 5S rRNA gene in *Solea senegalensis*: a new linkage to the U1, U2, and U5 small nuclear RNA genes. Genome 49: 79–86.

Marz M, Kirsten T, Stadler PF (2008) Evolution of spliceosomal snRNA genes in metazoan animals. J Mol Evol 7: 594–607.

Merlo MA, Pacchiarini T, Portela-Bens S, Cross I, Manchado M, Rebordinos L (2012) Genetic characterization of *Plectorhinchus mediterraneus* yields important clues about genome organization and evolution of multigene families. BMC Geneitcs 13: 33.

Mesa A, Ferreira A, Carbonell CS (1982) Cariología de los acridoideos neotropicales: estado actual de su conocimiento y nuevas contribuciones. Ann Soc Entomol Fr 18: 507-526.

Mount SM, Gotea V, Lin CF, Hernandez K, Makalowski W (2007) Spliceosomal small nuclear RNA genes in 11 insect genomes. RNA 13: 5–14.

Munõz-Pajares AJ, Martínez-Rodríguez L, Teruel M, Cabrero J, Camacho JPM, et al. (2011) A Single, Recent Origin of the Accessory B Chromosome of the Grasshopper *Eyprepocnemis plorans*. Genetics 187: 853–863.

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 39: 121-152.

Nguyen P, Sahara k, Yoshido A, Marec F (2010) Evolutionary dynamics of rDNA clusters on chromosomes of moths and butterflies (Lepidoptera). Genetica 138: 343-354.

Nilsen TW (2003) The spliceosome: the most complex macromolecular machine in the cell? BioEssays 25: 1147–1149.

Novák P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC bioinformatics 11(1): 378.

Novak P, Neumann P, Pech J, Steinhaisl J, Macas, J (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads. Bioinformatics 29: 792–793.

O'Reilly D, Dienstbier M, Cowley SA, Vazquez P, Drozdz M, Taylor S, et al. (2012) Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. Genome Res 23: 281-291.

Palacios-Gimenez OM, Castillo ER, Martí DA, Cabral-de-Mello DC (2013) Tracking the evolution of sex chromosome systems in Melanoplinae grasshoppers through chromosomal mapping of repetitive DNA sequences. BMC Evol Biol 13: 167.

Panzera Y, Pita S, Ferreiro MJ, Ferrandis I, Lages C, Pérez R, et al. (2012) High dynamics of rDNA cluster location in kissing bug holocentric chromosomes (Triatominae, Heteroptera). Cytogenet Genome Res 138: 56–67.

Pelliccia F, Barzotti R, Bucciarelli E, Rocchi A (2001) 5S ribosomal and U1 small nuclear RNA genes: a new linkage type in the genome of a crustacean that has three different tandemly repeated units containing 5S ribosomal DNA sequences. Genome 44: 331–335.

Pinkel D, Straume T, Gray JW (1986) Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. Proc Natl Acad Sci USA 83: 2934–2938.

Pinhal D, Yoshimura TS, Araki CS, Martins C (2011) The 5S rDNA family evolves through concerted and birth-and-death evolution in fish genomes: an example from freshwater stingrays. BMC Evol Biol 11: 151–164.

Raskina O, Barber JC, Nevo E, Belyayev A (2008) Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. Cytogenet Genome Res 120: 351–357.

Sierra-Montes JM, Pereira-Simon S, Freund AV, Ruiz LM, Szmulewicz MN, Herrera RJ (2003) A diversity of U1 small nuclear RNAs in the silk moth *Bombyx mori*. Insect Biochem Mol Biol 33: 29–39.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10): 2731-2739.

Valadkhan S (2005) snRNAs as the catalysts of pre-mRNA splicing. Curr Op Chem Biol 9: 603–608.

Vierna J, Jensen KT, Martínez-Lage1 A, González-Tizón AM (2011) The linked units of 5S rDNA and U1 snDNA of razor shells (Mollusca: Bivalvia: Pharidae). Heredity 107: 127–142.

Vierna J, Wehner S, Zu Siederdissen CH, Martínez-Lage A, Marz M (2013) Systematic analysis and evolution of 5S ribosomal DNA in metazoans. Heredity 111(5): 410-421.

West S (2012) The increasing functional repertoire of U1 snRNA. Biochem Soc Trans 40: 846-849.

White MJD (1978) Modes of Speciation. San Francisco: WH Freeman and Company.

Yu JC, Wendelburg B, Sakallah S, Marzluff WF (1991) The Ul snRNA gene repeat from the sea urchin (*Strongylocentrotus purpuratus*): the 70 kilobase tandem repeat ends directly 3' to a Ul gene. Nucleic Acids Res 19: 1093-1098.

Zeller R, Carri MT, Mattaj IW, De Robertis EM (1984) *Xenopus laevis* Ul snRNA genes: characterisation of transcriptionally active genes reveals major and minor repeated gene families. EMBO J 3: 1075-1081.

Tables

Table 1. U1 snDNA chromosomal location in 71 grasshopper species belonging to five families**.** *c* centromeric, *pr* proximal, *i* interstitial, *sd* sub-distal, *d* distal, *s* spread. Grey areas correspond to absent chromosomes in species with less than 23 chromosomes (males). (*) indicate results for species previously published and (#) species where the individuals analyzed carried B chromosomes.

Table 2. Analysis of diversity and functionality within each lineage of U1 snDNA found in the genome of *E. plorans*, showing the number of sequences (n), number of sites (s), number of haplotypes (H), haplotype diversity (Hd), nucleotide diversity bycolumn (π), number of segregating sites per site by means of Watterson's estimator (Θ_w), mean percentage of G+C and mean variation of Gibbs' free energy in kcal/mol (ΔG). Note that all five sequences for lineage 3 showed exactly the same sequence, for which reason π and Θ_w where equal to zero. Lineage 1, however, showed five haplotypes differing for some insertions from the consensus sequence, but not for substitutions, so that the former parameters were also equal to zero, since they are calculated including only those nucleotide positions being present in all DNA sequences.

Figure captions

Figure 1. Fluorescent *in situ* hybridization using U1 snDNA as probe in species of Proscopiidae (a,b), Pyrgomorphidae (c,d) and Ommexechidae (e). (a) *Stiphra robusta* (metaphase I), (b) *Scleratoscopia silvae* (metaphase II), (c) *Algete brunneri* (metaphase I), (d) *Omura congrua* (diplotene), (e) *Ommexecha virens* (metaphase I). Chromosomes harboring signals and the X chromosome are indicated. c = centromere location. The

insert in (e) corresponds to the pair 3 from a metaphase II, showing precisely the U1 snDNA cluster location.

Figure 2. U1 snDNA chromosomal mapping in nine Romaleidae species. (a) *Agriacris auripennis* (metaphase I), (b) *Brasilacris gigas* (metaphase I), (c) *Chromacris nuptialis* (metaphase II), (d) *Chromacris speciosa* (metaphase I), (e) *Helionotus mirabilis* (diplotene), (f) *Phaeopharia megacephala* (metaphase I), (g) *Radacridium nordestinum* (diplotene), (h) *Xestotrachelus robustus* (metaphase I), (i) *Xyleus discoideus angulatus* (metaphase I). Chromosomes harboring signals and the X chromosome are indicated. c = centromere location. The insert in (a) corresponds to the pair 3 from an anaphase I cell, showing precisely the U1 snDNA cluster location.

Figure 3. Chromosome location of U1 snDNA in 16 species of Acrididae grasshoppers, belonging to seven subfamilies. (a) *Metaleptea adspersa* (diplotene), (b) *Adimantus ornatissimus* (mitotic metaphase), (c) *Anacridium aegyptium* (diplotene), (d) *Schistocerca flavofasciata* (diplotene), (e) *Cylindrotettix obscurus* (metaphase I), (f) *Stenopola dorsalis* (diplotene), (g) *Tucayaca parvula* (metaphase I), (h) *Parascopas obesus* (metaphase I), (i) *Propiedies auricularis* (diplotene), (j) *Vilerna rugulosa* (metaphase I), (k) *Locusta migratoria* (metaphase I), (l) *Oedipoda fuscocincta* (diplotene). Chromosomes harboring signals and sex and B chromosomes are indicated. c = centromere location. Chromosomes bearing small U1 snDNA clusters are showed enlarged in the inserts directly in the cells. Note the variability of number and position for the gene mapped.

Figure 4. FISH for U1 snRNA gene in eight Gomphocerinae (Acrididae) grasshoppers. (a) *Chorthippus binotatus* (metaphase I), (b) *Euchorthippus pulvinatus* (metaphase I), (c) *Omocestus panteli* (diplotene), (d) *Stenobothrus festivus* (diplotene), (e,e') selected chromosome no. 1 of *Chorthippus nevadensis*, DAPI (e) U1 snDNA (e'), (f) *Dociostaurus maroccanus* (diplotene), (g) *Orphulella punctata* (metaphase I), (h) *Rhammatocerus brasiliensis* (metaphase I). Chromosomes harboring signals and the X chromosome are indicated. c = centromere location. Note the multiple sites in large chromosomes of species with diploid number reduction (a-e) and restrict distribution in (g).

Figure 5. Schematic representation of the cluster that contains the U1 snDNA lineages in the genome of *E. plorans*. The five contigs with U1 snDNA are marked in different colors. In addition to U1 snDNA, we found homology with several elements such as incomplete 5S rDNA and SINEs (indicated as boxes).

Figure 6. (a) Alignment of the consensus sequence for the five U1 snDNA lineages in *E. plorans*; (b) MST built with the consensus sequence of the five lineages, also depicting their predicted secondary structure. Three kinds of change are represented with different filled forms: circle, triangle and square corresponding to substitutions, deletions and insertions, respectively (the number of involved nucleotides is also indicated for the two latter). Deletions and insertions are considered as single mutational events and the number of involved positions in the alignment is also indicated beside the corresponding symbol. Mean Gibbs' free energy (ΔG) for U1 secondary structure, expressed in kcal/mol, is depicted between brackets.

Figure 7. Secondary structure of the consensus sequence for lineage 1 of U1 snDNA in *E. plorans*. Conserved domains are colored. Only lineage 2 showed these same conserved domains (not shown).

Figure 8. FISH analysis for U1 snDNA (a-c) and 5S rDNA (b, c) in *E. plorans* chromosomes. (a) mitotic metaphase, (b) selected chromosomes (pair 6) from a mitotic metaphase and (c) fiber-FISH. Note the spread distribution for U1 snDNA in the chromosomes (a) and the co-localization (interspersed organization) for U1 snDNA and 5S rDNA (b, c).

Figure 9. Number of U1 snDNA sites observed through FISH in 71 species of grasshoppers belonging to five families. Families are indicated by color in the bars (Red = Proscopiidae; Yellow = Pyrgomorphidae; Green = Ommexechidae; Blue = Romaleidae; Purple = Acrididae) and the symbols indicate the subfamilies of Acrididae (□ = Acridinae; † = Catantopinae; ◊ = Copiocerinae; ψ = Cyrtacanthacridinae; ○ = Eyprepocnemidinae; Δ = Gomphocerinae; ● = Leptysminae; Φ = Melanoplinae; ‡ = Ommatolampinae; ▪ = Oedipodinae).

Supplementary material S1. Grasshopper species studied in this work with their respective sampling place and karyotypes.

Supplementary material S2. Classification adopted for distinct location patterns of U1 snDNA clustres. (a) Schematic representation of mitotic chromosome and (b) Diplotene acrocentric chromosome showing the possible distinct positions for U1 snDNA (distinct colors); (c) schematic representation of the Diplotene acrocentric chromosome

presented in (b) showing the points of contact for this bivalent (arrowheads). c = centromere, t = telomere. In case of biarmed chromosome, the same classification for cluster positions was followed.

Table 1.

Family Subfamily *Species*	Chromosome n° (in order of decreasing size)												Number of U1 snRNA sites					
	1	2	3	4	5	6	7	8	9	10	11	Sex	c	pr	i	sd	d	Total
Proscopiidae																		
Proscopiinae																		
Scleratoscopia protopeirae				i											1			1
Scleratoscopia spinosa				i											1			1
Scleratoscopia silvae				i											1			1
Stiphra robusta				i											1			1
Pyrgomorphidae																		
Pyrgomorphinae																		
Algete brunneri			pr											1				1
Omura congrua			pr											1				1
Ommexechidae																		
Ommexechinae																		
Ommexecha gracilis	pr		c/pr										1	2				3
Ommexecha virens	pr		c/pr										1	2				3
Romaleidae																		
Romaleinae																		
Agriacris auripennis								c/pr					1	1				2
Brasilacris gigas							i								1			1
Chromacris nuptialis					pr									1				1
Chromacris speciosa				sd												1		1
Helionotus mirabilis			pr											1				1
Phaeopharia megacephala			pr											1				1
Radacridium nordestinum			pr											1				1

Tropidacris collaris			pr								1					1
Xestotrachelus robustus			sd										1			1
Xyleus discoideus angulatus			pr								1					1
Xyleus discoideus discoideus			pr								1					1
Acrididae																
Acridinae																
Metaleptea adspersa					pr	pr					2					2
Orphula pagana			d	d										2		2
Catantopinae																
Pezotettix giornae						pr					1					1
Copiocerinae																
Adimantus ornatissimus								pr			1					1
Cyrtacanthacridinae																
Anacridium aegyptium	i											1				1
Schistocerca flavofasciata	i	i										2				2
Schistocerca pallens	i	i										2				2
Eyprepocnemidinae																
Eyprepocnemis plorans[#]	pr			pr	pr	pr		pr	pr	pr		7				7
Gomphocerinae																
Amblytropidia australis			i					pr/i				1	2			3
Amblytropidia sp.	pr	sd			sd		i	i				1	2	2		5
Chorthippus apicalis	pr/d											1			1	2
Chorthippus binotatus	pr/i/sd											1	1	1		3
Chorthippus jacobsi	pr/i/sd		c		pr						1	2	1	1		5
Chorthippus jucundus	pr/i/sd		sd									1	1	2		4
Chorthippus nevadensis	s	d	pr	pr								2			1	3
Chorthippus parallelus	2pr/2i	sd			i							2	3	1		6
Dociostaurus maroccanus			c		c		c		c		4					4
Euchorthippus pulvinatus		pr/2i/sd						pr				2	2	1		5

Omocestus bolivari	2pr/i/sd	pr/d									3	1	1	1	6
Omocestus burri	pr/i/sd	pr/d									2	1	1	1	5
Omocestus panteli	pr/sd	pr/d									2		1	1	4
Orphulella punctata					d									1	1
Rhammatocerus brasiliensis[#]			i	c/sd	i					1		2	1		4
Stenobothrus festivus	2i/sd	c/sd		pr						1	1	2	2		6
Leptysminae															
Cornops frenatum frenatum							i					1			1
Cylindrotettix attenuatus[#]							i					1			1
Cylindrotettix obscurus			i					pr(X)			1	1			2
Eumastusia koebelei koebelei			c/pr							1	1				2
Stenopola sp				i			pr	pr(X)			2	1			3
Stenopola dorsalis[#]			i									1			1
Tucayaca parvula							i					1			1
Melanoplinae															
Baeacris pseudopunctulatus		i										1			1
Baeacris punctulatus				sd									1		1
*Chlorus chiquitensis**			d											1	1
*Chlorus vittatus**				d										1	1
*Dichromatos lilloanus**		i										1			1
*Dichromatos schrottkyi**			d											1	1
Dichroplus fuscus			sd	pr							1			1	2
*Eurotettix brevicerci**				d										1	1
*Eurotettix minor**				d					i(XR); i(Y)			2		1	3
Parascopas obesus						sd							1		1
Parascopas sanguineus						sd							1		1
Propedies auricularis			pr/sd								1		1		2
Propedies viriosus			pr								1				1

Ronderosia bergi			pr	pr/sd									2		1		3
Ommatolampidinae																		
Abracris flavolineata[*#]							pr						1				1
Orthoscapheus rufipes[#]					pr	pr							2				2
Vilerna rugulosa[#]	i		pr										1	1			2
Oedipodinae																		
Aiolopus strepens								pr					1				1
Locusta migratoria[#]			pr										1				1
Oedipoda fuscocincta	c	c	c	c	c	c/pr	c	c	c	c	c	c(X)	12	1				13
Sphingonotus caerulans	c	c	c	c	c	c	c	c	c	c	c	c(X)	12					12
Total													35	62	39	21	14	171
Percent													20.5	36.2	22.8	12.3	8.2	100
Average																		2.4

Table 2.

	n	s	H	Hd	π	Θ_w	G+C(SD)	ΔG(SD)
Lineage 1	11	168	5	0.8182	0.00000	0.00000	54.8(0.4)	-66.7(0.9)
Lineage 2	16	166	13	0.9667	0.01987	0.02314	53.9(0.5)	-61.4(2.4)
Lineage 3	5	134	1	0.0000	0.00000	0.00000	55.2(0)	-53.5(0)
Lineage 4	21	133	18	0.9857	0.04879	0.08360	55.9(0.5)	-43.7(2.0)
Lineage 5	7	162	6	0.9524	0.01242	0.01775	51.9(0.6)	-62.1(3.0)

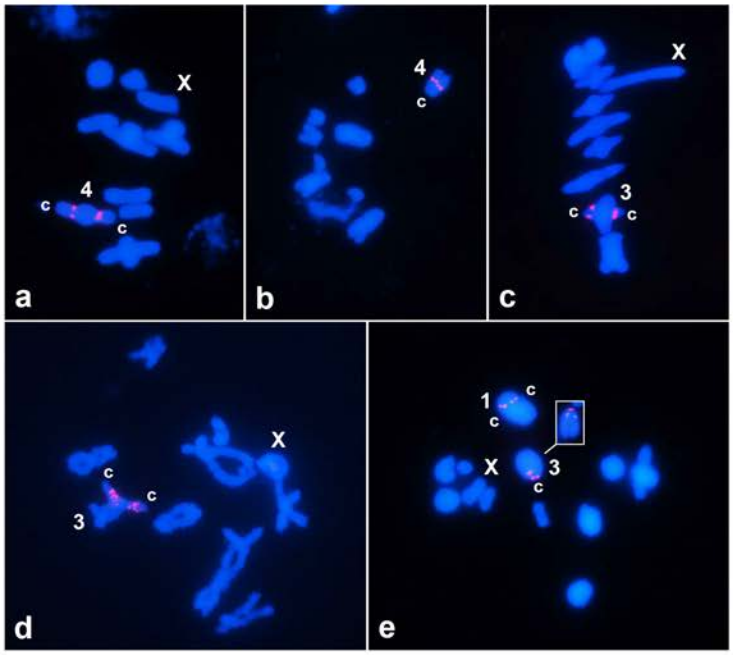


Figure 1

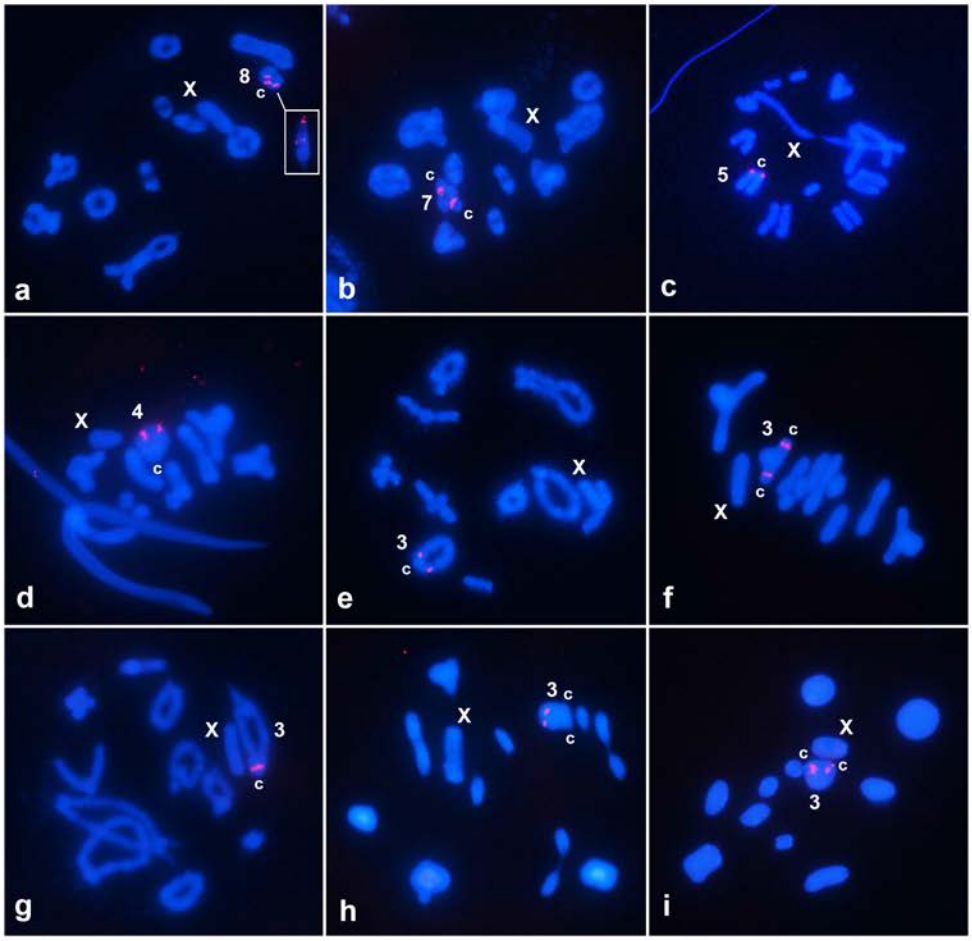


Figure 2

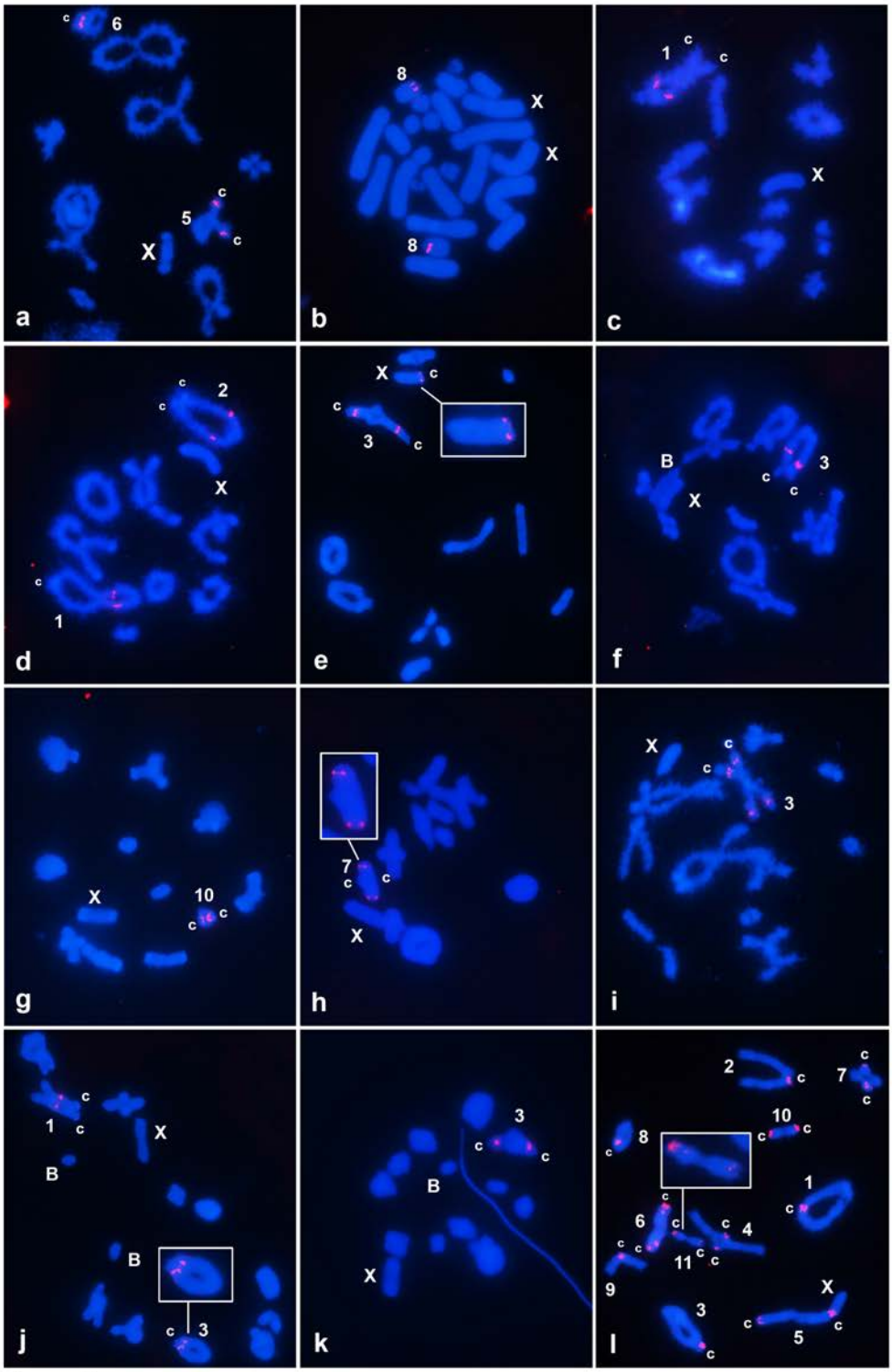

a
6
5
X
b
8
X
X
8
c
1
X
d
2
X
1
e
X
3
f
B
3
X
g
X
10
h
7
X
i
X
3
j
1
X
B
B
3
k
3
B
X
l
2
7
10
8
1
6
4
11
9
X
3
5

Figure 3

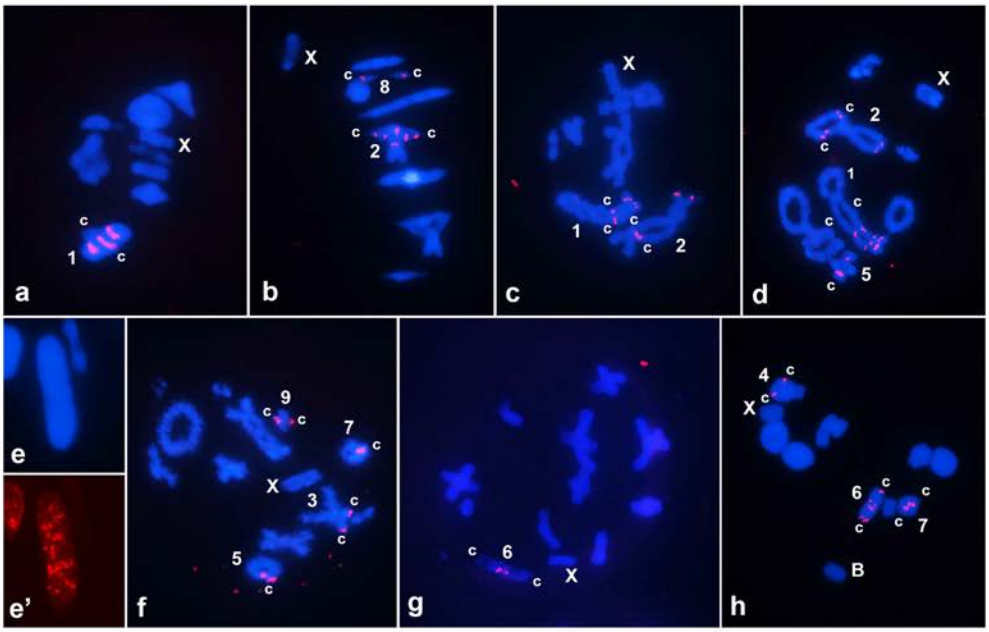
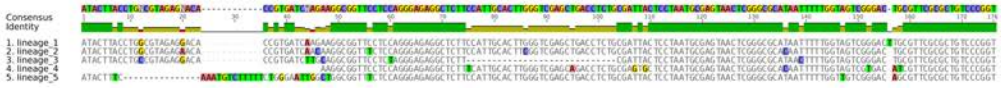

a
X
c
1
c
b
X
c
8
c
c
2
c
c
X
c
1
c
c
c
2
d
X
c
2
c
1
c
c
c
5
e
e'
f
9
c
c
7
c
X
3
c
c
5
c
g
c
6
c
X
h
4
c
X
c
6
c
c
c
7
B

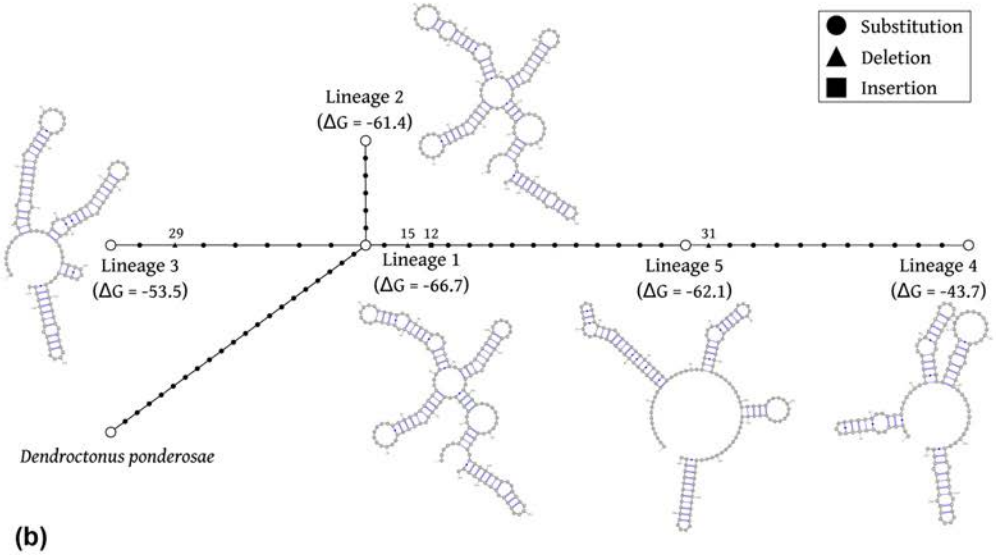
Figure 4


Lineage 1
Lineage 2
Lineage 3
Lineage 4
Lineage 5
Alpine
SINE?
5S
U1
5S
5S
tRNA-Thr
SINE?

Figure 5


Consensus
Identity
1. lineage_1
2. lineage_2
3. lineage_3
4. lineage_4
5. lineage_5

(a)


Substitution
Deletion
Insertion
Lineage 2
(ΔG = -61.4)
29
15
12
31
Lineage 3
(ΔG = -53.5)
Lineage 1
(ΔG = -66.7)
Lineage 5
(ΔG = -62.1)
Lineage 4
(ΔG = -43.7)
Dendroctonus ponderosae

(b)

Figure 6

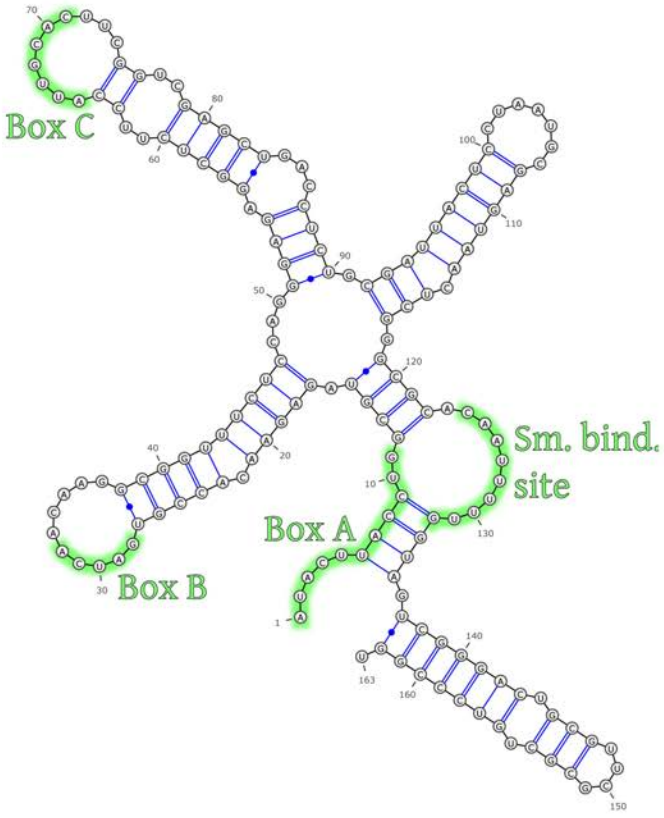


Figure 7

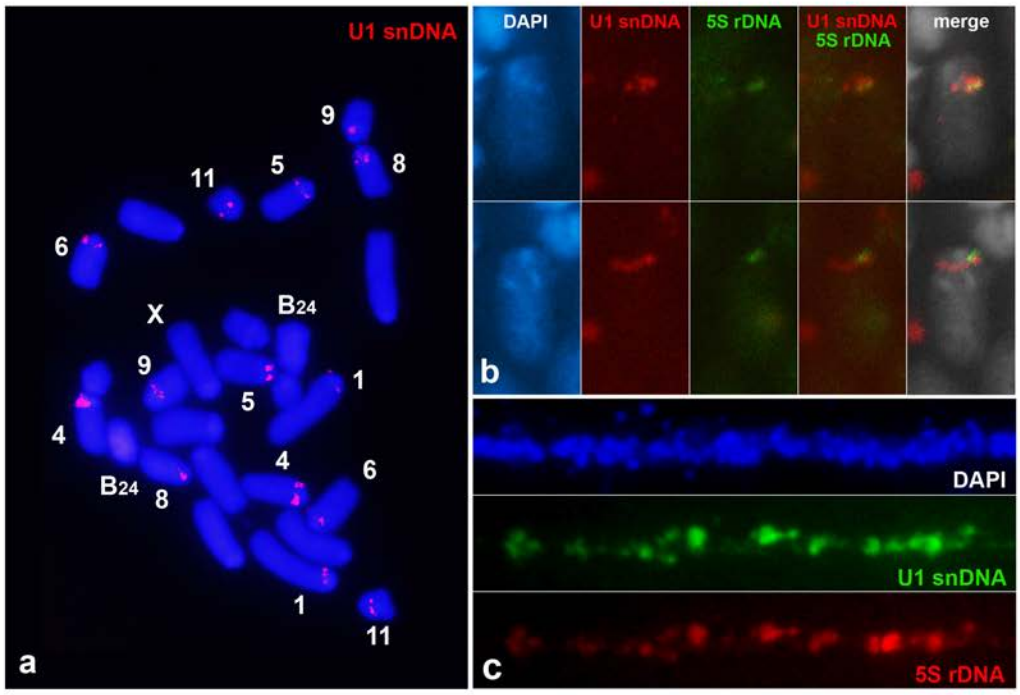

U1 snDNA
9
11
5
8
6
X
B24
9
1
5
4
4
6
B24
8
1
11
a
DAPI
U1 snDNA
5S rDNA
U1 snDNA
5S rDNA
merge
b
DAPI
U1 snDNA
c
5S rDNA

Figure 8

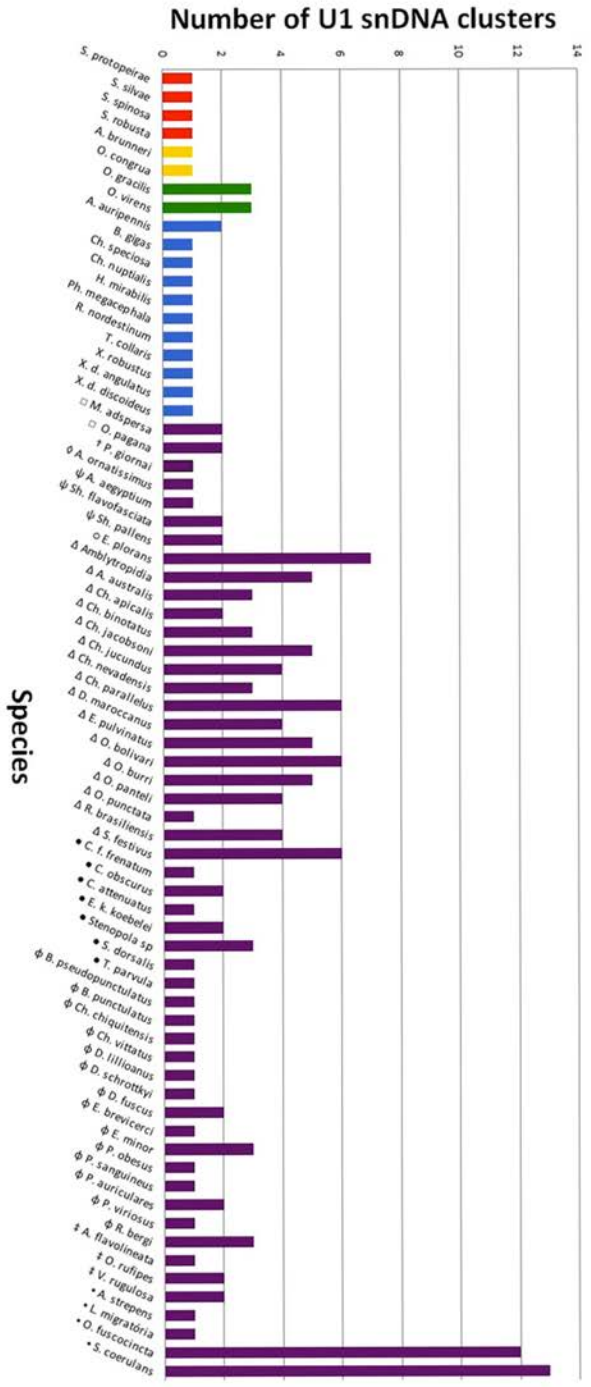
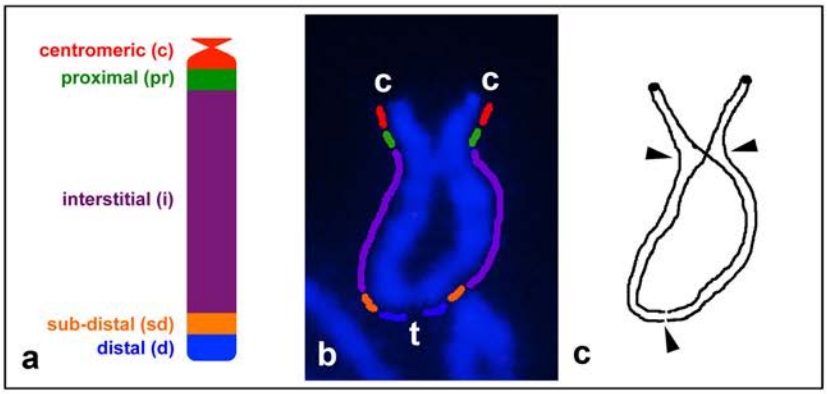

Number of U1 snDNA clusters
0
2
4
6
8
10
12
14
Species
S. protopeirae
S. silvae
S. spinosa
S. robusta
A. brunneri
O. congrua
O. gracilis
O. virens
A. auripennis
B. gigas
Ch. speciosa
Ch. nuptialis
H. mirabilis
Ph. megacephala
R. nordestinum
T. collaris
X. robustus
X. d. angulatus
X. d. discoideus
M. adspersa
O. pagana
P. giornai
A. ornatissimus
A. aegyptium
Sh. flavofasciata
Sh. pallens
E. plorans
Amblytropidia
A. australis
Ch. apicalis
Ch. binotatus
Ch. jacobsoni
Ch. jucundus
Ch. nevadensis
Ch. parallelus
D. maroccanus
E. pulvinatus
O. bolivari
O. burri
O. panteli
O. punctata
R. brasiliensis
S. festivus
C. f. frenatum
C. obscurus
C. attenuatus
E. k. koebelei
Stenopola sp
S. dorsalis
T. parvula
B. pseudopunctulatus
B. punctulatus
Ch. chiquitensis
Ch. vittatus
D. lilloanus
D. schrottkyi
D. fuscus
E. brevicerci
E. minor
P. obesus
P. sanguineus
P. auriculares
P. viriosus
R. bergi
A. flavolineata
O. rufipes
V. rugulosa
A. strepens
L. migratoria
O. fuscocincta
S. coerulans

Figure 9

Supplementary material S2



Supplementary material S1.

Family Subfamily *Species*	Locality	Karyotype (male)
Proscopiidae		
Proscopiinae		
Scleratoscopia protopeirae	Gravatá, Pernambuco, BR	19, X0
Scleratoscopia silvae	Surubim, Pernambuco, BR	19, X0
Scleratoscopia spinosa	Araripina, Pernambuco, BR	19, X0
Stiphra robusta	Araripina, Pernambuco, BR	19, X0
Pyrgomorphidae		
Pyrgomorphinae		
Algete brunneri	Rio Formoso, Pernambuco, BR	20, X0
Omura congrua	Goiana, Pernambuco, BR	20, X0
Ommexechidae		
Ommexechinae		
Ommexecha gracilis	Barra do Garça, Mato Grosso, BR	23, X0
Ommexecha virens	Rio das Contas, Bahia, BR	23, X0
Romaleidae		
Romaleinae		
Agriacris auripennis [#]	Cabo de Santo Agostinho, Pernambuco, BR	23, X0
Brasilacris gigas	Arco Verde, Pernambuco, BR	23, X0
Chromacris speciosa	Bonito, Pernambuco, BR	23, X0
Chromacris nuptialis	Vitória de Santo Antão, Pernambuco, BR	23, X0
Helionotus mirabilis	Bezerros, Pernambuco, BR	23, X0
Phaeopharia megacephala	Igarassu, Pernambuco, BR	23, X0
Radacridium nordestinum	Surubim, Pernambuco, BR	23, X0
Tropidacris collaris [#]	Goiana, Pernambuco, BR	23, X0
Xestotrachelus robustus	Buique, Pernambuco, BR	23, X0
Xyleus discoideus angulatus	Surubim, Pernambuco, BR	23, X0
Xyleus discoideus discoideus	Rio Claro, Pernambuco, BR	23, X0
Acrididae		
Acridinae		
Metaleptea adspersa	Bezerros, Pernambuco, BR	23, X0
Orphula pagana	Cabo de Santo Agostinho, Pernambuco, BR	23, X0
Cantatopinae		
Pezotettix giornae	Granada, ES	23, X0
Copiocerinae		
Adimantus ornatissimus	Rio Claro, São Paulo, BR	23, X0
Cyrtacanthacridinae		
Anacridium aegyptium	Granada, ES	23, X0
Schistocerca flavofasciata	Rio Claro, São Paulo, BR	23, X0
Schistocerca pallens	Goiana, Pernambuco, BR	23, X0
Eyprepocnemidinae		
Eyprepocnemis plorans	Granada, ES	23, X0
Gomphocerinae		
Amblytropidia australis	Rio Claro, São Paulo, BR	23, X0

Amblytropidia sp.	São Lourenço da Mata, Pernambuco, BR	23, X0
Chorthippus apicalis	Granada, ES	17, X0
Chorthippus binotatus	Granada, ES	17, X0
Chorthippus jacobsi	Granada, ES	17, X0
Chorthippus jucundus	Granada, ES	17, X0
Chorthippus nevadensis	Granada, ES	17, X0
Chorthippus parallelus	Granada, ES	17, X0
Dociostaurus maroccanus	Granada, ES	23, X0
Euchorthippus pulvinatus	Granada, ES	17, X0
Omocestus bolivari	Granada, ES	17, X0
Omocestus burri	Granada, ES	17, X0
Omocestus panteli	Granada, ES	17, X0
Orphulella punctata	Araripina, Pernambuco, BR	23, X0
Rhammatocerus brasiliensis	Bezerros, Pernambuco, BR	23, X0
Stenobothrus festivus	Granada, Pernambuco, ES	17, X0
Leptysminae		
Cornops frenatum frenatum	Cabo de Santo Agostinho, Pernambuco, BR	23, X0
Cylindrotettix attenuatus [#]	Rio Claro, São Paulo, BR	23, X0
Cylindrotettix obscurus [#]	Goiana, Pernambuco, BR	23, X0
Eumastusia koebelei koebelei [#]	Serrolândia, Pernambuco, BR	23, X0
Stenopola sp	Rio Claro, São Paulo, BR	23, X0
Stenopola dorsalis	Cabo de Santo Agostinho, Pernambuco, BR	23, X0
Tucayaca parvula	Caruaru, Pernambuco, BR	23, X0
Melanoplinae		
Beacris pseudopunctulatus [#]	Rio Claro, São Paulo, BR	23, X0
Beacris punctulatus [#]	Goiana, Pernambuco, BR	23, X0
Chlorus chiquitensis	Corumba, Mato Grosso, BR	19, X0
Chlorus vittatus	Ybycuí, PY	23, X0
Dichromatos lillioanus	Eldorado, AR	21, X_1X_2Y
Dichromatos schrottkyi	Eldorado, AR	21, X_1X_2Y
Dichroplus fuscus	Ouricuri, Pernambuco, BR	23, X0
Eurotettix brevicerci	Botucatu, São Paulo, BR	23, X0
Eurotettix minor	Paraguarí, PY	22, XY
Parascopas obesus [#]	Rio Claro, São Paulo, BR	23, X0
Parascopas sanguineus	Rio Claro, São Paulo, BR	23, X0
Propiedies auriculares [#]	Corumbá, Mato Grosso, BR	23, X0
Propedies viriosus [#]	Colina, São Paulo, BR	23, X0
Ronderosia bergi	Rio Claro, São Paulo, BR	22, XY
Ommatolampidinae		
Abracris flavolineata	Rio Claro, São Paulo, BR	23, X0
Orthoscapheus rufipes	Toritama, Pernambuco, BR	23, X0
Vilerna rugulosa [#]	Rio Claro, São Paulo, BR	23, X0
Oedipodinae		
Aiolopus strepens	Granada, ES	23, X0
Locusta migratoria	Granada, ES	23, X0
Oedipoda fuscocincta	Granada, ES	23, X0
Sphingonotus caerulans	Granada, ES	23, X0

(#) Species with karyotypes described for the first time in this work.

5.0 CONCLUSÕES FINAIS

1. Apesar da conservada estrutura macro cromossômica observada em gafanhotos, um enorme dinamismo genômico envolve a distribuição dos genes de DNAsn U1 nas diferentes famílias de gafanhotos, principalmente em Acrididae. Esta ampla variabilidade é similar ao reportado para os DNAr 5S e 45S.

2. Os dados obtidos para as distintas famílias multigênicas em gafanhotos apontam uma grande reorganização genômica (microgenômica) nas diferentes espécies, sem, em geral, modificações na macro estrutura dos cariótipos.

3. A grande variação em número e posição dos sítios de DNAsn U1 nos cariótipos de gafanhotos aparentemente não está envolvida com macro rearranjos cromossômicos, sendo mais plausível a ocorrência de recombinações ectópicas e transposições.

4. A ausência de clusters de DNAsn U1 em cromossomos B sugere que este gene não está envolvido com os processos evolutivos que levaram ao surgimento desses elementos no genoma de gafanhotos, ao menos nas espécies estudas neste trabalho.

5. A variação na distribuição dos clusters de DNAsn U1 em *Eyprepocnemis plorans* pode estar relacionada com a associação deste marcador a elementos transponíveis e este poderia ser o mecanismo evolutivo que possibilitou o espalhamento dessa sequência no genoma de *E. plorans*, gerando a ampla diversidade observada (cinco linhagens). Além disso, as análises genômicas indicam que os mecanismos de "birth-and-death" e evolução em concerto atuam na evolução dos clusters de DNAsn U1.

6. A associação de dados cromossômicos e de sequenciamento genômico permitiram um melhor entendimento dos padrões evolutivos dos genes de DNAsn U1 em gafanhotos.

6.0 REFERÊNCIAS

ALTSCHUL, S.F.; MADDEN, T.L.; SCHÄFFER, A.A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, **Nucleic Acids Research**, v. 25, p. 3389-3402, 1997.

ANJOS, A.; LORETO, V.; SOUZA, M.J. Chromosome mapping of ribosomal genes and histone H4 in the genus *Radacridium* (Romaleidae). **Genetics and Molecular Biology**, v. 36, p. 336-340, 2013.

ANLEITNER, J.; HAYMER, D. Y enriched and Y-specific DNA sequences from the genome of the Mediterranean fruit fly, *Ceratitis capitata*. **Chromosoma**, v. 101, p. 271-278, 1992.

BARZOTTI, R.; PELLICCIA, F.; ROCCHI, A. Identification and characterization of U1 small nuclear RNA genes from two crustacean isopod species. **Chromosome Research**, v. 11, p. 365-373, 2003.

BIDAU, C.J.; MARTÍ, D.A. Meiosis and the Neo-XY system of *Dichroplus vittatus* (Melanoplinae, Acrididae): a comparison between sexes. **Genetica**, v. 110, p. 185–194, 2001.

BIÉMONT, C.; VIEIRA, C. Genetics: Junk DNA as an evolutionary force. **Nature**, v. 443, p. 521-524, 2006.

BRIDLE, J.R.; DE LA TORRE, J.; BELLA, J.L.; BUTLIN, R.K.; GONSALVEZ, J. Low levels of chromosomal differentiation between the grasshoppers *Chorthippus brunneus* and *Chorthippus jacobsi* (Orthoptera-Acrididae) in northoren Spain. **Genetica**, v. 114, p. 121-127, 2002.

BRINGMANN, P.; LÜHRMANN, R. Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. **EMBO Journal**, v. 5, p. 3509-3516, 1986.

BUENO, D.; PALACIOS-GIMENEZ, O.M.; CABRAL-DE-MELLO, D.C. Chromosomal mapping of repetitive DNAs in *Abracris flavolineata* reveal possible ancestry for the B chromosome and surprisingly H3 histone spreading. **PLoS ONE**, v. 8:e66532, 2013.

CABRAL-DE-MELLO, D.C.; CABRERO, J.; LÓPEZ-LEÓN, M.D.; CAMACHO, J.P.M. Evolutionary dynamics of 5S rDNA location in acridid grasshoppers and its relationship with H3 histone gene and 45S rDNA location. **Genetica**, v. 139, p. 921-931, 2011a.

CABRAL-DE-MELLO, D.C.; MARTINS, C.; SOUZA, M.J.; MOURA, R.C. Cytogenetic mapping of 5S and 18S rRNAs and H3 histone genes in four ancient Proscopiidae grasshopper species: contribution to understanding the evolutionary dynamics of multigene families. **Cytogenetic and Genome Research**, v. 132, p. 89–93, 2011b.

CABRAL-DE-MELLO, D.C.; MOURA, R.C.; MARTINS, C. Chromosomal mapping of repetitive DNAs in the beetle *Dichotomius geminatus* provides the first evidence for an association of 5S rRNA and histone H3 genes in insects, and repetitive DNA

similarity between the B chromosome and A complement. **Heredity**, v. 104, p. 393–400, 2010.

CABRAL-DE-MELLO, D.C.; VALENTE, G.T.; NAKAJIMA, R.T.; MARTINS, C. Genomic organization and comparative chromosome mapping of the U1 snRNA gene in cichlid fish, with an emphasis in *Oreochromis niloticus*. **Chromosome Research**, v. 20, p. 279–292, 2012.

CABRERO, J.; CAMACHO, J.P.M. Cytogenetic studies in gomphocerine grasshoppers. 1. Comparative analysis of chromosome C-banding pattern. **Heredity**, v. 56, p. 365–372, 1986.

CABRERO, J.; CAMACHO, J.P.M. Location and expression of ribosomal RNA genes in grasshoppers: Abundance of silent and cryptic loci. **Chromosome Research**, v. 16, p. 595–607, 2008.

CABRERO, J.; LÓPEZ-LEÓN, M.D.; TERUEL, M.; CAMACHO, J.P.M. Chromossome mapping of H3 and H4 histone gene clusters in 35 species of acridid grasshoppers. **Chromosome Research**, v. 17, p. 397-404, 2009.

CAMACHO, J.P.M.; CABRERO, J.; LÓPEZ-LEÓN, M.D.; BAKKALI M.; PERFECTTI, F. The B chromosomes of the grasshopper *Eyprepocnemis plorans* and the intragenomic conflict. **Genetica**, v. 117, p. 77–84, 2003.

CAMACHO, J.P.M.; CABRERO, J.; VISERAS, E.; LOPEZ-LEON, M.D.; NAVAS-CASTILLO, J.; ALCHE, J.D. G banding in two species of grasshopper and its relationship to C, N, and fluorescence banding techniques. **Genome**, v. 34, p. 638-643, 1991.

CAMACHO, J.P.M.; SHARBEL, T.F.; BEUKEBOOM, L.W. B-chromosome evolution. **Philosophical Transactions of the Royal Society of London**, v. 355, p. 163-178, 2000.

CARVALHO, D.B.; ROCHA, M.F.; LORETO, V.; SILVA, A.E.B.; SOUZA, M.J. *Ommexecha virens* (Thunberg, 1824) and *Descampsacris serrulatum* (Serville, 1831) (Orthoptera, Ommexechidae): karyotypes, constitutive heterochromatin and nucleous organizing regions. **Comparative Cytogenetics**, v. 5, p. 123-132, 2011.

CASTILLO, E.R.; TAFFAREL, A.; MARTÍ, D.A. Una técnica alternativa para el cariotipado mitótico en saltamontes: bandeo C y Fluorescente en *Adimantus ornatissimus* (Orthoptera: Acrididae). **Revista de Ciência & Tecnologia**, v. 16, p. 31–35, 2011.

CASTILLO, E.R.D.; BIDAU, C.J.; MARTÍ, D.A. Neo sex chromosome diversity in Neotropical melanopline grasshopper (Melanoplinae, Acrididae). **Genetica**, v. 138, p. 775-786, 2010a.

CASTILLO, E.R.D.; MARTÍ, D.A.; BIDAU, C.J. Sex and neo-sex chromossomes in Orthoptera: a review. **Journal of Orthoptera Research**, v. 19, p. 213-231, 2010b.

COLOMBO, P.; CIGLIANO, M.M.; SEQUEIRA, A.S.; LANGE, C.E.; VILARDI, J.C.; CONFALONIERI, V.A. Phylogenetic relationships in *Dichroplus* Stal (Orthoptera: Acrididae: Melanoplinae) inferred from molecular and morphological data: testing karyotype diversification. **Cladistics**, v. 21, p. 375–389, 2005.

DRUMMOND, A.J.; ASHTON, B.; CHEUNG, M.; HELED, J.; KEARSE, M. Geneious 4.8. **Biomatters**, Auckland, New Zealand, 2009.

EXCOFFIER, L.; LISCHER, H.E.L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. **Molecular Ecology Resources**, v. 10, p. 564-567, 2010.

FERREIRA, A.; MESA, A. Cytotaxonomy in the genus *Dichromatos* Cigliano 2007 (Orthoptera, Acridoidea, Melanoplinae). **Journal of Orthoptera Research**, v. 19, p. 233-237, 2010.

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

GOGOLEVSKY, K.P.; VASSETZKY, N.S.; KRAMEROV, D.A. 5S rRNA-derived and tRNA-derived SINEs in fruit bats. **Genomics**, v. 93(5), p. 494-500, 2009.

GREEN, M.R. Pre-mRNA processing and mRNA nuclear export. **Current Opinion in Cell Biology**, v. 1, p. 519-525, 1989.

GREGORY, T.R. Genome size evolution in animals. In Gregory TR (Ed.). The Evolution of the Genome, **Elsevier**, San Diego, pp. 3-87, 2005.

HANARAHAN, S.J.; JOHNSTON, J.S. New genome size estimates of 134 species of arthropods. **Chromosome Research**, v. 6, p. 809-823, 2011.

HEWITT, G.H. Orthoptera. Grasshoppers and Crickets, in *Animal Citogenetics 3. Insecta 1*, edited by B. John. **Gerbrüder Borntraeger**, Berlín-Suttgart, pp 1-170, 1979.

JOHN, B.; KING, M. Population cytogenetics of A tractomorpha similis. **Chromosoma**, v. 88, p. 57-68, 1983.

KAPITONOV, V.V.; JURKA, J. A novel class of SINE elements derived from 5S rRNA. **Molecular Biology and Evolution**, v. 20(5), p. 694-702, 2003.

KING, M.; JOHN, B. Regularities and restrictions governing C-band variation in acridoid grasshoppers. **Chromosoma**, v. 76, p. 123-150, 1980.

KOHANY, O.; GENTLES, A.J.; HANKUS, L.; JURKA, J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. **BMC Bioinformatics**, Oct 25;7:474, 2006.

KRAEMER, C.; SCHMIDT, E.R. The sex determination region of *Chironomus thummi* is associated with highly repetitive DNA and transposable elements. **Chromosoma**, v. 102, p. 553-562, 1993.

LIBRADO, P.; ROZAS, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. **Bioinformatics**, v. 25(11), p. 1451-1452, 2009.

LÓPEZ-LEÓN, M.D.; NEVES, N.; SCHWARZACHER, T.; HESLOP-HARRISON, J.S.; HEWITT, G.M.; CAMACHO, J.P.M. Possible origin of a B chromosome deduced from its DNA composition using double FISH technique. **Chromosome Research**, v. 2, p. 87-92, 1994.

LORETO, V.; CABRERO, J.; LÓPEZ-LEÓN, M.D.; CAMACHO, J.P.M.; SOUZA, M.J. Possible autosomal origin of macro B chromosomes in two grasshopper species. **Chromosome Research**, v. 16, p. 233-241, 2008.

LORETO, V.; SOUZA, M.J. Karyotype, constitutive heterochromatin and nucleolar organizer regions (NORs) in *Belosacris coccineipes* (Acrididae-Leptysminae). **Genetics and Molecular Biology**, v. 23, p. 575-579, 2000.

LORETO, V.; STADTLER, E.; MELO, N.F.; SOUZA, M.J. A comparative cytogenetic analysis between the grasshopper species *Chromacris nuptialis* and *C. speciosa* (Romaleidae): constitutive heterochromatin variability and rDNA sites. **Genetica**, v. 125, p. 253-260, 2005.

LUND E.; NESBITT M.N. The embryonic and adult mouse U1 snRNA genes map to different chromosomal loci. **Somatic Cell and Molecular Genetics**, v. 14, p. 143–148, 1988.

LUND, E.; BOSTOCK, C.; ROBERTSON, M.; CHRISTIE S.; MITCHEN, J.L.; DAHLBERG, J.E. U1 small nuclear RNA genes are located on human chromosome 1 and are expressed in mouse–human hybrid cells. **Molecular And Cellular Biology**, v. 3, p. 2211–2220, 1983.

MANCHADO, M.; ZUASTI, E.; CROSS, I.; MERLO, M.A.; INFANTE, C.; REBORDINOS, L. Molecular characterization and chromosomal mapping of the 5S rRNA gene in *Solea senegalensis*: a new linkage to the U1, U2, and U5 small nuclear RNA genes. **Genome**, v. 49, p. 79–86, 2006.

MANSER, T.; GESTELAND, R.F. Human U1 Loci: Genes for human U1 RNA have dramatically similar genomic environments. **Cell**, v. 29(1), p. 257-264, 1982.

MARKHAM, N.R.; ZUKER, M. UNAFold. In Bioinformatics. **Humana Press**, p. 3-31, 2008.

MARTINS, C. Chromosomes and repetitive DNAs: a contribution to the knowledge of fish genome. In: Pisano E, Ozouf-Costaz C, Foresti F, Kapoor BG (Eds.). **Fish Cytogenetics**. Enfield, New Hampshire: Science Publisher, pp. 421-453, 2007.

MARZ, M.; KIRSTEN, T.; STADLER, P.F. Evolution of spliceosomal snRNA genes in metazoan animals. **Journal of Molecular Evolution**, v. 7, p. 594–607, 2008.

MARZLUFF, W.F.; BROWN, D.T.; LOBO, S.; WANG, S. Isolation and characterization of two mouse U1b small nuclear RNA genes. **Nucleic Acids Research**, v. 11, p. 6255-6270, 1983.

MATTAJ, I.W.; ZELLER, R. *Xenopus laevis* U2 snRNA genes: tandemly repeated transcription units sharing 5' and 3' flanking homology with other RNA polymerase II transcribed genes. **EMBO Journal**, v. 2, p. 1883-1891, 1983.

MESA, A.; FERREIRA, A. Cytological studies in the family Ommexechidae (Orthoptera, Ommexechidae). **Acrida**, v. 6, p. 261-271, 1977.

MESA, A.; FERREIRA, A. Have the Australian Morabinae and the neotropical Proscopiidae envolved from a common ancestor? A cytogenetical point of view (Orthoptera-Caelfera-Eumastacoidea). **Acrida**, v. 10, p. 205-217, 1981.

MESA, A.; FERREIRA, A.; CARBONELL, C.S. Cariología de los Acridoideos Neotropicales: estado actual de su conocimiento y nuevas contribuciones. **Annales de la Société Entomologique de France (N.S)**, v. 18, p. 507-526, 1982.

MESA, A.; FONTANETTI, C.S.; CARCÍA-NOVO, P. Does an X-autosome centric fusion in Acridoidea condemn the species to extinction? **Journal of Orthoptera Research**, v. 10, p. 141–146, 2001.

MESA, A.; FONTANETTI, C.S.; COSTA, H. The karyotype of the grasshoper *Spathalium helios* Rehn 1918 (Orthoptera-Acridoidea-Ommexechidae). **Revista Brasileira de Genética**, v. 13, p. 704-710, 1990.

MESSIER, W.; LI, S.H.; STEWART, C.B. The birth of microsatellites. **Nature**, v. 381, p. 483, 1996.

MOURA, R.C.; SOUZA, M.J.; TASHIRO, T. Cytogenetics characterization of the genera *Scleratoscopia* and *Tetanorhyncus* (Orthoptera, Proscopiidae). **Cytologia**, v. 61, p. 169-178, 1996.

MUNÕZ-PAJARES, A.J.; MARTÍNEZ-RODRÍGUEZ, L.; TERUEL, M.; CABRERO, J.; CAMACHO, J.P.M.; PERFECTTI, F. A Single, Recent Origin of the Accessory B Chromosome of the Grasshopper *Eyprepocnemis plorans*. **Genetics**, v. 187, p. 853–863, 2011.

NEI, M.; ROONEY, A.P. Concerted and birth-and-death evolution of multigene families. **Annual Review of Genetics**, v. 39, p. 121-152, 2005.

NILSEN, T.W. The spliceosome: the most complex macromolecular machine in the cell? **BioEssays**, v. 25, p. 1147–1149, 2003.

NOVÁK, P.; NEUMANN, P.; MACAS, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. **BMC bioinformatics**, v. 11(1), p. 378, 2010.

NOVAK, P.; NEUMANN, P.; PECH, J.; STEINHAISL, J.; MACAS, J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads. **Bioinformatics**, v. 29, p. 792–793, 2013.

OLIVEIRA, N.L.; CABRAL-DE-MELLO, D.C.; ROCHA, M.F.; LORETO, V.; MARTINS, C.; MOURA, R.C. Chromosomal mapping of rDNAs and H3 histone sequences in the grasshopper *Rhammatocerus brasiliensis* (Acrididae, Gomphocerinae): extensive chromosomal dispersion and co-localization of 5S rDNA/H3 histone clusters in the A complement and B chromosome. **Molecular Cytogenetics**, v. 4, p. 24, 2011.

PALACIOS-GIMENEZ, O.M.; CASTILLO, E.R.; MARTÍ, D.A.; CABRAL-DE-MELLO, D.C. Tracking the evolution of sex chromosome systems in Melanoplinae grasshoppers through chromosomal mapping of repetitive DNA sequences. **BMC Evolutionary Biology**, v. 13, p. 167, 2013.

PELLICCIA, F.; BARZOTTI, R.; BUCCIARELLI, E.; ROCCHI, A. 5S ribosomal and U1 small nuclear RNA genes: a new linkage type in the genome of a crustacean that has three different tandemly repeated units containing 5S ribosomal DNA sequences. **Genome**, v. 44, p. 331–335, 2001.

PINKEL. D.; STRAUME, T.; GRAY, J.W. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. **Proceedings of the National Academy of Sciences of the United States of America**, v. 83, p. 2934–2938, 1986.

RIO, D.C. RNA processing. **Current Opinion in Cell Biology**, v.4, p. 444-452, 1992.

ROCHA, M.F.; SOUZA, M.J.; MOURA, R.C. Karyotypic analysis, constitutive heterochromatin and NOR distribution in five grasshopper species of the subfamily Leptysminae (Acrididae). **Caryologia**, v. 1, p. 107-116, 2004.

ROCHA, M.F.; SOUZA, M.J.; TASHIRO, T. Karyotype Variability in the Genus *Radacridium* (Orthoptera, Romaleidae, Romaleinae). **Cytologia**, v. 62, p. 53-60, 1997.

SEINO, R.A.; DONGMO, A.; DONGMO, T.I.; MANJELI, Y. Karyotype and meiosis analysis of four species of Cameroonian Pyrgomorphidae (Orthoptera). **International Journal of Genetics and Molecular Biology**, v. 5, p. 13-19, 2013.

SHAPIRO, J.A.; STERNBERG, R. Why repetitive DNA is essential to genome function. **Biological Reviews**, v. 80, p. 1-24, 2005.

SMIT, A.F.A.; HUBLEY, R.; GREEN, P. RepeatMasker Open-3.0. Available http://www.repeatmasker.org.

SOUZA, M.J.; MELO, N.F. Chromosome study in *Schistocerca* (Orthoptera, Acrididae, Cyrtacanthacridinae): Karyotypes and distribution patterns of constitutive heterochromatin and nucleolus regions (NORs). **Genetics and Molecular Biology**, v. 30, p. 54-59, 2007.

SOUZA, M.J.; MOURA, R.C. Karyotypic characteriza- tion and constitutive heterochromatin in the grasshopper *Stiphra robusta* (Orthoptera, Proscopiidae). **Cytobios**, v. 101, p. 137-144, 2000.

TAMURA, K.; PETERSON, D.; PETERSON, N.; STECHER, G.; NEI, M.; KUMAR, S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. **Molecular Biology and Evolution**, v. 28(10), p. 2731-2739, 2011.

TERUEL, M.; CABRERO, J.; PERFECTTI, F.; CAMACHO, J.P.M. B chromosome ancestry revealed by histone genes in the migratory locust. **Chromosoma**, 119:217-225, 2010.

ÚBEDA-MANZANARO, M.; MERLO, M.A.; PALAZÓN, J.L.; CROSS, I.; SARASQUETE, C.; REBORDINOS, L. Chromosomal mapping of the major and minor ribosomal genes, (GATA)n and U2 snRNA gene by double-colour FISH in species of the Batrachoididae family. **Genetica**, v. 138, p. 787–794, 2010.

VALADKHAN, S. snRNAs as the catalysts of pre-mRNA splicing. **Current Opinion in Chemical Biology**, v. 9, p. 603-608, 2005.

VIERNA, J.; JENSEN, K.T.; MARTÍNEZ-LAGE, A.; GONZÁLEZ-TIZÓN, A.M. The linked units of 5S rDNA and U1 snDNA of razor shells (Mollusca: Bivalvia: Pharidae). **Heredity**, v. 107, p. 127–142, 2011.

WEST, S. The increasing functional repertoire of U1 snRNA. **Biochemical Society Transactions**, v. 40, p. 846–849, 2012.

YOSHIMURA, A.; NAKATA, A.; KURO-O, M.; OBARA, Y.; ANDO, Y. Molecular cytogenetic characterization and chromosomal distribution of the satellite DNA in the genome of *Oxya hyla intricata* (Orthoptera: Catantopidae). **Cytogenetic and Genome Research**, v. 112, p. 160-165, 2006.

YU, J.C.; WENDELBURG, B.; SAKALLAH, S.; MARZLUFF, W.F. The Ul snRNA gene repeat from the sea urchin (*Strongylocentrotus purpuratus*): the 70 kilobase tandem repeat ends directly 3' to a Ul gene. **Nucleic Acids Research**, v. 19, p. 1093-1098, 1991.