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

OCTAVIO MANUEL PALACIOS-GIMENEZ

**Padrões de evolução de sistemas de cromossomos sexuais em
grilos: uma abordagem integrada entre citogenética e
genômica**

Dezembro - 2017

OCTAVIO MANUEL PALACIOS-GIEMENEZ

**Padrões de evolução de sistemas de cromossomos sexuais em
grilos: uma abordagem integrada entre citogenética e
genômica**

Orientador: Prof. Dr. Diogo Cavalcanti Cabral-de-Mello

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

RIO CLARO

2017

591.15 Gimenez, Octavio Manuel Palacios
G491p Padrões de evolução de sistemas de cromossomos sexuais
em grilos: uma abordagem integrada entre citogenética e
genômica / Octavio Manuel Palacios Gimenez. - Rio Claro,
2018

240 f. : il., figs., tabs.

Tese (doutorado) - Universidade Estadual Paulista,
Instituto de Biociências de Rio Claro

Orientador: Diogo Cavalcanti Cabral de Mello

1. Genética animal. 2. Citogenética e genômica de grilos.
3. Cromossomos sexuais. 4. DNA satélite. 5. DNA- e
RNA-seq. 6. FISH. 7. Rearranjos cromossômicos. I. Título.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: Padrões de evolução de sistemas de cromossomos sexuais em grilos: uma abordagem integrada entre citogenética e genômica

AUTOR: OCTAVIO MANUEL PALACIOS GIMÉNEZ

ORIENTADOR: DIOGO CAVALCANTI CABRAL DE MELLO

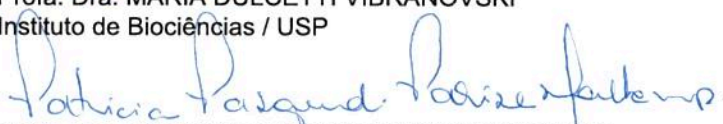
Aprovado como parte das exigências para obtenção do Título de Doutor em CIÊNCIAS BIOLÓGICAS (BIOLOGIA CELULAR E MOLECULAR), pela Comissão Examinadora:



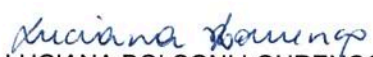
Prof. Dr. DIOGO CAVALCANTI CABRAL DE MELLO
Departamento de Biologia / IB Rio Claro



Profa. Dra. MARIA DULCETTI VIBRANOVSKI
Instituto de Biociências / USP



Profa. Dra. PATRICIA PASQUALI PARISE MALTEMPI
Departamento de Biologia / IB Rio Claro



Profa. Dra. LUCIANA BOLSONI LOURENÇO
Departamento de Biologia Celular / UNIVERSIDADE ESTADUAL DE CAMPINAS



Prof. Dr. ANDRE LUIS LAFORGA VANZELA
Departamento de Biologia Geral / UNIVERSIDADE ESTADUAL DE LONDRINA

Rio Claro, 12 de dezembro de 2017

*Este pequeno livro, a minha
família, com amor dedico.*

AGRADECIMENTOS

Eu gosto de observar e analisar, e muitas vezes me encontro tentando entender e descrever padrões e processos que ocorrem na natureza. Tentando ver o simples no aparente complexo. E estou muito agradecido por ter a oportunidade de fazer o que eu gosto do meu trabalho. Portanto, um agradecimento especial ao meu orientador Diogo. Obrigado por me recrutar como estudante de doutorado! E agradeço-lhe por me fazer parte de uma equipe de pesquisa variada e estimulante. Foi e é ótimo trabalhar em um ambiente tão animado, com uma boa mistura de pessoas. Muito obrigado pelas muitas discussões encorajadoras, pela sua calma, precisão e resistência no seu trabalho. Agradeço-lhe por me dar liberdade no meu trabalho e por me ajudar a me desenvolver como pesquisador. Obrigado por ser um bom orientador, alguém que agradeço e quem eu gosto.

Eu também quero agradecer ao meu supervisor durante o estágio no exterior, Bernardo Lemos (Harvard T.H. Chan School of Public Health, Boston, MA, USA). Gostei muito da nossa colaboração e da nossa persistente tentativa de encontrar algum equilíbrio em biologia. Agradeço-lhe por muitas discussões estimulantes e por ser alguém que questiona e tenta chegar ao fundo.

Agradeço aos atuais e antigos membros do laboratório do Diogo, aos de Lemos Lab, aos do Lab de Gustavo Kuhn por compartilhar seus pensamentos e ideais. Agradeço a todos os Professores do Departamento de Biologia e aos colegas e amigos do Lab pela ótima convivência durante os anos de trabalho. A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (2014/02038-8) pelo apoio financeiro cedido para o desenvolvimento do projeto no país, e pela bolsa de estágio de pesquisa no exterior (BEPE, 2016/01506-3). Finalmente, um agradecimento especial a Universidade Estadual Paulista (UNESP), Campos Rio Claro e a Universidade de Harvard (Harvard T.H. Chan School of Public Health, Boston, MA, USA) pela grande oportunidade. A todos, muito obrigado de verdade...!!

Mas não há estudos de doutorado sem estudos de graduação, e quero agradecer ao meu ex-orientador da graduação Dardo Martí por me apresentar na ciência. Agradeço-me por me ter escolhido em um grupo de alunos, pelo seu apoio e por me encorajar a continuar com a pesquisa. Estou feliz por ter feito isso.

Mas há vida além da pesquisa. E fiquei feliz por ter encontrado bons amigos aqui no Brasil: Thiago, Paixão, Miagui, Ama, Beto, Iata, Eliana, Tati, Vivi, Milani, Vavá. Alguns deles me disseram: "Você nunca tem tempo, apenas para a pesquisa. Arrume um tempo e vamos embora beber uma cerveja, você tem tempo??!!!". Mas às vezes encontrei tempo e não me arrependi. Sempre apreciei nossas discussões, sim, mesmo com gritarias.

E nenhuma distância pode quebrar uma verdadeira amizade. Quero agradecer aos meus amigos fora do Brasil, aqueles novos que fiz nos EUA (Melvin, Meng, Katie, Eduardo, Fernanda, Younfeng, Peter, Diego, Chan) e aqueles velhos que ainda estão na minha terra natal. Obrigado por serem meus amigos. Muitas vezes vocês estão comigo em meus pensamentos e eu estou grato por ter você lá.

E MINHA FAMÍLIA, MESMO QUE LONGE, SEMPRE PERTO DE MIM, O MAIS IMPORTANTE PARA MIM. Pai e Mãe, muitas vezes vocês se preocupam com a maneira como eu ando pelo mundo, mas ainda assim vocês me apoiam em muitos dos meus passos. Agradeço por estar aqui para mim, por acreditar em mim e por aceitar-me do jeito que sou. Isso significa muito para mim, e vocês significam muito para mim. Pai e Mãe, irmãos (Romi e Max), avó Ana e sobrinhos ..., gostaria de poder agradecer-lhes, mas não consigo encontrar as palavras certas. Vocês estão comigo em cada passo que eu levo, em minhas lágrimas e em meu coração ...

*People will not have time for you if you are always
angry or complaining.*

Stephen Hawking

RESUMO

Os cromossomos sexuais se originam independentemente de um par de homólogos autossômicos e em várias linhagens apresentam características comuns, tais como acúmulo de vários tipos de DNA repetitivo, restrição da recombinação e perda ou ganho de genes devido à diferenciação morfológica e genética entre os cromossomos sexuais X e Y ou Z e W. Estas características representam um exemplo fascinante de convergência evolutiva. Em Orthoptera, o sistema cromossômico sexual comumente encontrado na maioria das espécies estudadas é do tipo $X0^{\sigma}/XX^{\eta}$. Entretanto, sistemas cromossômicos sexuais derivados dos tipos neo-XY $^{\sigma}/XX^{\eta}$ e neo- $X_1X_2Y^{\sigma}/X_1X_1X_2X_2^{\eta}$ são também observados, surgindo repetidamente por fusões cêntricas e em tandem, inversões e dissociações envolvendo cromossomos sexuais ancestrais e autossomos. O presente trabalho teve três objetivos. Primeiro, entender o possível papel dos DNAs repetitivos na estrutura/diversificação dos cromossomos sexuais simples e derivados, a partir do isolamento e mapeamento físico de sequências, tais como, famílias multigênicas, DNA satélite (DNAsat) e microssatélites, nas espécies *Gryllus assimilis*, *Cycloptiloides americanus* e *Eneoptera surinamensis*. Segundo, testar e comparar transcrição diferencial de DNAsat entre diferentes tecidos, sexos e espécies a partir de transcriptomas de *Gryllus assimilis*, *G. bimaculatus*, *G. firmus* e *G. rubens*, com o objetivo de entender os possíveis papéis funcionais destas sequências na regulação gênica, modulação da cromatina e como componentes funcionais de importantes estruturas como telômeros, centrômeros e cromossomos sexuais. Terceiro, a partir de transcriptomas de espécies de grilos (*Gryllus assimilis*, *G. bimaculatus* e *G. firmus*) prospectar genes codificadores de proteínas relacionados com a determinação sexual, envolvidos com o *fitness* reprodutivo e genes enviesados do sexo, responsáveis pelas diferenças fenotípicas entre machos e fêmeas, e tentar elucidar de uma maneira comparativa os fatores evolutivos atuando nestes loci. Origem *de novo* de cromossomos sexuais mediante rearranjos cromossômicos, assim como acúmulo de DNA repetitivo que levaram a diferenciação entre cromossomos sexuais são relatados em *C. americanus* (X_1X_20) e *E. surianmensis* (neo- X_1X_2Y). Estas características observadas em grilos representam outro caso notável de convergência evolutiva devido os cromossomos sexuais não relacionados compartilharem muitas propriedades entre táxons distantes. Acúmulo surpreendente de loci de DNAsat foi encontrado no neo-Y altamente diferenciado de *E. surinamensis*, incluindo 39 DNAsat representados em excesso neste cromossomo, que é a maior diversidade de DNAsat até agora relatada para cromossomos sexuais. Foi documentado que, particularmente os DNAsat, contribuíram grandemente para o aumento de tamanho genômico entre *G. assimilis* e *E. surinamensis*. Um achado interessante foi a identificação de DNAsat conservados entre espécies de grilos (*Gryllus assimilis*, *G. bimaculatus* e *G. firmus*), mas transcritos diferencialmente. Os dados relativos à presença de DNAsat no genoma de *G. assimilis* foram discutidos em um contexto evolutivo, com dados transcricionais permitindo comparações entre os sexos e entre os tecidos quando possível. Foram discutidas hipóteses para a conservação e transcrição de DNAsat em *Gryllus*, que podem resultar

do seu papel na diferenciação sexual no nível da cromatina, na formação da heterocromatina e na função centromérica. Outra descoberta foi a identificação de genes determinantes do sexo e outros genes relacionados ao *fitness* reprodutivo, como a biossíntese de hormônios de insetos e ritmo circadiano entre espécies de *Gryllus*. Os efetores e os alvos *downstream* das vias de determinação do sexo foram previamente identificados em outros insetos, mas nunca em Orthoptera. Usando *G. assimilis* como modelo para estudar genes enviesados do sexo foi possível identificar um conjunto de genes altamente expressos que podem explicar diferenças fenotípicas entre os sexos. Estimou-se que os genes codificadores de proteínas relacionadas com a diferenciação sexual e com o *fitness* reprodutivo evoluem mais rapidamente do que os genes não reprodutivos (genes *housekeeping*) como resultado de uma forte seleção positiva nos primeiros. Além disso, foi encontrado que as espécies estudadas apresentam níveis excepcionalmente elevados de duplicações gênicas. As descobertas sugerem que as duplicações gênicas podem desempenhar um papel na expressão de genes enviesados do sexo no grilo de campo *G. assimilis*, uma espécie que no futuro provavelmente irá fornecer informações sobre genômica funcional e epigenética da determinação do sexo.

Palavras-chave: cromossomos sexuais, DNA satélite, DNA- e RNA-seq, FISH, rearranjos cromossômicos, genes enviesados do sexo.

ABSTRACT

Sex chromosomes have arisen independently from an ordinary autosomal pair and in several lineages they present common characteristics, such as accumulation of distinct classes of repetitive DNAs, restriction of the recombination and loss or gain of genes due to the morphological and genetic differentiation between the sexual chromosomes X and Y or Z and W. These characteristics represent a fascinating example of evolutionary convergence. In Orthoptera, the $X_0\text{♂}/XX\text{♀}$ sex-determining system is considered modal but eventually, diverse sex chromosome systems evolved several times, such as neo- $XY\text{♂}/XX\text{♀}$, $X_1X_2\text{♂}/X_1X_1X_2X_2\text{♀}$ and even neo- $X_1X_2Y\text{♂}/X_1X_1X_2X_2\text{♀}$. It was found that particularly centric fusions (i.e., Robertsonian translocations) and tandem fusions with autosomes, dissociations and inversions contributed to the formation of neo-sex chromosomes in Orthoptera. The present work had three objectives. First, get insights of the role of repetitive DNAs in the structure/diversification of simple and derivative sex-chromosomes by isolation and physical mapping of repetitive DNA sequences, such as multigene families, satellite DNA (satDNA) and microsatellites using *Gryllus assimilis*, *Cycloptiloides americanus* e *Eneoptera surinamensis*, as models. Second, looking at differential satDNA transcription between different tissues, sexes, and species from transcriptomes of *Gryllus assimilis*, *G. bimaculatus*, *G. firmus* and *G. rubens*, I tried to understand the possible functional roles of these sequences in gene regulation, chromatin modulation and as functional components of important structures such as telomeres, centromeres and sex chromosomes. Third, using transcriptomes from cricket species (*Gryllus assimilis*, *G. bimaculatus* and *G. firmus*), I searched for genes encoding proteins related to sexual determination, reproductive fitness and sex-biased genes which are responsible for the phenotypic differences between males and females. I also tried to elucidate in a comparative way the evolutionary factors acting at these loci. *De novo* origin of sex chromosomes by chromosomal rearrangements, as well as repetitive DNA accumulation that led to the differentiation between sex chromosomes are reported for *C. americanus* ($X_1X_2\text{♂}$) e *E. surinamensis* (neo- X_1X_2Y). These features observed in crickets represent another remarkable case of evolutionary convergence because unrelated sex chromosomes share many common properties among distant taxa. Especially astonishing accumulation of satDNAs loci was found in the highly differentiated neo-Y, including 39 satDNAs over-represented in this chromosome, which is the greatest satDNAs diversity yet reported for sex chromosomes. It has been documented that, particularly the satDNA, contributed greatly to the increase in genomic size between *G. assimilis* and *E. surinamensis*. An interesting finding was the identification of satDNA conserved among species of crickets (*Gryllus assimilis*, *G. bimaculatus* and *G. firmus*), but differentially transcribed. The data regarding satDNA presence in *G. assimilis* genome was discussed in an evolutionary context, with transcriptional data enabling comparisons between sexes and across tissues when possible. I discussed hypotheses for the conservation and transcription of satDNAs in *Gryllus*, which might result from their role in sexual differentiation at the chromatin

level, heterochromatin formation, and centromeric function. Another finding was the identification of sex-determining genes and other genes related to reproductive fitness, such as biosynthesis of insect hormones and circadian rhythm among *Gryllus* species. The effectors as well as downstream targets of sex-determination pathways have been previously identified in other insects but never in Orthoptera. Using *G. assimilis* to study sex-biased genes I identified a set of highly expressed genes that might account for phenotypic differences between sexes. Furthermore, I estimated that protein-encoding reproductive genes evolve faster than non-reproductive genes as result of strong positive selection at those loci. It was documented that the species studied harbor exceptionally high levels of gene duplications. The findings suggest that gene duplications may play a role in sex-biased genes expression in the field cricket *G. assimilis*, a species likely to yield insights into the functional genomics and epigenetics of sex determination.

Key-words: chromosomal rearrangements, sex-chromosomes, satellite DNA, FISH, DNA- and RNA-seq, sex-biased genes.

Breve histórico profissional

Possuo graduação (2011) em Genética (Bacharelado) pela Universidad Nacional de Misiones (UNaM, Argentina), sob a orientação do Prof. Dr. Dardo Andrea Martí. Obtive o título de Mestre (2014) em Ciências Biológicas (Biologia Celular e Molecular) pela Universidade Estadual Paulista campus de Rio Claro, sob a orientação do Prof. Dr. Diogo Cavalcanti Cabral-de-Mello. Durante o mestrado realizei estágio no Laboratório do Prof. Dr. Frantisek Marec (Institute of Entomology, Biology Centre CAS, České Budějovice, Czech Republic). O presente trabalho tem como objetivo obter o título de Doutor em Ciências Biológicas (Biologia Celular e Molecular). A maior parte dos estudos de doutorado (três anos) foram desenvolvidos no Instituto de Biociências, Departamento de Biologia da Universidade Estadual Paulista campus de Rio Claro, sob a supervisão do Prof. Dr. Diogo Cavalcanti Cabral-de-Mello. Durante um ano, o doutorado foi desenvolvido na Universidade de Harvard (Harvard T.H. Chan School of Public Health-HSPH, Boston, US), sob a supervisão do Prof. Dr. Bernardo Lemos. Tenho experiência na área de Genética com ênfase em citogenética, genômica e transcriptômica de insetos (gafanhotos e grilos), atuando principalmente nos seguintes temas: DNAs repetitivos, expressão genica, genes enviesados do sexo, cromossomos sexuais, cromossomos Bs e evolução cariotípica.

Além dos trabalhos apresentados nesta tese, sou autor ou co-autor dos seguintes artigos científicos:

Octavio Manuel Palacios-Gimenez, Diogo Milani, Bernardo Lemos, Elio R. Castillo, Dardo A. Martí, Erica Ramos, Cesar Martins, Diogo C. Cabral-de-Mello. Uncovering the evolutionary history of neo-XY sex chromosomes in the grasshopper *Ronderosia bergii* (Orthoptera, Melanoplinae) through satellite DNA analysis. BMC Evolutionary Biology. *In press*.

Bernardino ACS, Cabral-de-Mello DC, Machado CB, **Palacios-Gimenez OM**, Santos N, Loreto V. 2017. Populational variants of B chromosomes from the grasshopper *Xyleus discoideus angulatus* (Romaleidae) indicate possible polymorphism originated from the pericentromeric portion of A chromosomes. Cytogenetic and Genome Research. doi: 10.1159/000480036

Castillo ERD, Taffarel A, Maronna MM, Cigliano MM, **Palacios-Gimenez OM**, Cabral-de-Mello DC, Martí DA. 2017. Phylogeny and chromosomal diversification in the *Dichroplus elongatus* species group (Orthoptera, Melanoplinae). PLoS ONE 12(2): e0172352, 2017.

Milani D, **Palacios-Gimenez OM**, Cabral-de-Mello, DC. 2017. The U2 snDNA is a useful marker for B chromosome detection and frequency estimation in the grasshopper *Abracris flavolineata*. Cytogenetic and Genome Research 150: 1-4.

Bueno D, **Palacios-Gimenez OM**, Martí DA, Mariguela TC, Cabral-de-Mello DC. 2016. The 5S rDNA in two *Abracris* grasshoppers (Ommatolampidinae: Acrididae): molecular and chromosomal organization. Molecular Genetics and Genomics. 291:1607-13.

Palacios-Gimenez OM, Martí DA, Cabral-de-Mello DC. 2015. Neo-sex chromosomes of *Ronderosia bergi*: insight into the evolution of sex chromosomes in grasshoppers. Chromosoma 124: 353-365.

Menezes-de-Carvalho NZ, **Palacios-Gimenez OM**, Milani D, Cabral-de-Mello DC. 2015. High similarity of U2 snDNA sequence between A and B chromosomes in the grasshopper *Abracris flavolineata*. Molecular Genetics and Genomics 290: 1787-92.

Palacios-Gimenez OM, Bueno D, Cabral-de-Mello DC. 2014. Chromosomal mapping of two Mariner-like elements in the grasshopper *Abracris flavolineata* (Orthoptera: Acrididae) reveals enrichment in euchromatin. European Journal of Entomology 111: 329-334.

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 Evolutionary Biology 13: 167.

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

SUMÁRIO

1. INTRODUÇÃO	1
1.1- A ordem Orthoptera	1
1.2- A superfamília Grylloidea e Gryllotalpoidea: Uma visão geral.....	1
1.3- DNA repetitivo em eucariotas.....	3
1.3.1- Famílias multigênicas	4
1.3.2- DNA satélites.....	6
1.3.3- Elementos transponíveis (TE).....	9
1.4- Mecanismos de determinação sexual em insetos.....	11
1.5- Origem e evolução de cromossomos sexuais.....	14
1.6- Evolução de cromossomos sexuais em Orthoptera.....	17
1.7- Evolução de genes enviesados do sexo.....	20
1.8- Cromossomos sexuais em grilos: <i>Gryllus assimilis</i> , <i>Cycloptiloides americanus</i> e <i>Eneoptera surinamensis</i> como modelos evolutivos para estudos de cromossomos sexuais em Orthoptera.....	23
2. OBJETIVOS	26
2.1- Objetivos gerais.....	26
2.2- Objetivo específicos.....	27
2.2.1- Estrutural.....	27
2.2.2- Funcional.....	27
3. DESENHO EXPERIMENTAL	28
3.1- Amostras, análises de cromossomos e bandeamentos cromossômicos clássicos.....	28
3.2- Estimação de tamanho genômico.....	28
3.3- Isolamento de DNA repetitivo por de PCR	29
3.4- Isolamento de DNA alta e moderadamente repetitivo: fração de DNA <i>C_{ot}</i>	30
3.5- Sequenciamento de Illumina e clusterização baseado em gráfico de <i>reads</i> sequenciados.....	30
3.6- Isolamento e análises de sequencias de DNAsat.....	31
3.7- Hibridação <i>in situ</i> fluorescente (FISH)	32

3.8- Análise quantitativa de DNAsat em <i>E. surinamensia</i>	33
3.9- Análise de correlação estatística de DNAsat em <i>E. surinamensis</i>	34
3.10- Tecidos de <i>G. assimilis</i> para extração de RNA e sequenciamento.....	34
3.11- Transcrição de DNAsat nas espécies de <i>Gryllus</i>	35
3.12- Montagem e anotação de transcriptomas em espécies de <i>Gryllus</i>	36
3.13- Expressão gênica diferencial.....	37
3.14- Estimativa de taxas evolutivas.....	38
4- RESULTADOS E DISCUSSÃO	39
4.1- Capítulo 1: Repetitive DNA chromosomal organization in the cricket <i>Cycloptiloides americanus</i> : a case of the unusual X ₁ X ₂ 0 sex chromosome system in Orthoptera	40
4.2- Capítulo 2: Contrasting the chromosomal organization of repetitive DNAs in two <i>Gryllidae</i> crickets with highly divergent karyotypes	41
4.3- Capítulo 3: High-throughput analysis of the satellitome revealed enormous diversity of satellite DNAs in the neo-Y chromosome of the cricket <i>Eneoptera surinamensis</i>	42
4.4- Capítulo 4: Satellite DNAs are conserved and differentially transcribed among <i>Gryllus</i> cricket species.....	43
4.5- Capítulo 5: Insights on sex-determining genes and sex-biased genes expression through <i>de novo</i> transcriptome assembly in field crickets.....	44
5- CONCLUSÕES GERAIS	45
6- PERSPECTIVAS FUTURAS	50
6- REFERÊNCIAS BIBLIOGRÁFICAS	52

1. INTRODUÇÃO

1.1. A ordem Orthoptera

Os ortópteros estão representados atualmente por mais de 27.000 espécies reconhecidas principalmente por caracteres morfológicos, capacidade de salto e por seus aparelhos estridulatórios (SONG et al., 2015). O táxon faz parte de uma das faunas mais importantes de insetos terrestres, sendo em sua maioria espécies fitófagas, ocupando uma grande variedade de habitats e confinadas especialmente na região tropical, com poucas espécies ocupando latitudes frias (HEWITT, 1979; CIGLIANO e LANGE, 1998, SONG et al., 2015). A ordem inclui duas subordens: Ensifera e Caelifera (SONG et al., 2015; ZHOU et al., 2015).

Orthoptera originou-se há cerca de 300 milhões de anos durante o Permiano, Triássico e Jurássico com a radiação da subordem Caelifera, posterior a Ensifera (SONG et al., 2015). A característica mais importante durante o Terciário e Quaternário foi a expansão da superfamília Acridoidea que aconteceu há 50 milhões de anos (SONG et al., 2015). A subordem Ensifera, considerada basal, inclui os ortópteros de antenas longas (grilos, katydideos, wetas) pertencentes as superfamílias Tettigonioidea (7.104 espécies), Stenopelmatoidea (913 espécies), Hagloidea (8 espécies), Rhaphidophoroidea (645 espécies), Grylloidea (4.886 espécies), Gryllotalpoidea (547 espécies) Schizodatyloidea (15 espécies), contendo um total de 14.118 espécies existentes distribuídas em aproximadamente 2.000 gêneros (SONG et al., 2015; ZHOU et al., 2017). Caelifera, com aproximadamente 11.680 espécies, inclui os gafanhotos de antenas curtas na superfamília Acridoidea com 7.976 espécies de gafanhotos, Tetrigoidea com 1.763 espécies, Eumastacoidea com 1.008 espécies, Pyrgomorphoidea com 473 espécies, Procospioidea com 214 espécies, Trydactiloidea com cerca de 205 espécies, Trigonopterygoidea com 21 espécies, Pneumoroidea com 17 espécies e Tanaoceroidea com apenas 3 espécies conhecidas (SONG et al., 2015).

1.2. A superfamília Grylloidea e Gryllotalpoidea: uma visão geral

Recentes estudos filogenéticos reconhecem dois clados divergentes dentro da infraordem Gryllidea, com provável origem no Triássico: as superfamílias Grylloidea e Gryllotalpoidea (SONG et al., 2015). Grylloidea está representada por uma única família, Gryllidae (grilos), que representa a terceira linhagem mais diversa dentro de Orthoptera, contendo 4.886 espécies existentes (SONG

et al., 2015). As linhagens de Gryllidae são muito divergentes, reflexo do tempo de origem do grupo (SONG et al., 2015; CHINTAUAN-MARQUIER et al., 2017). Gryllidae foi proposto como grupo irmão ao clado contendo as famílias Gryllotalpidae (grilos moles; 107 espécies), Mogoplistidae (grilos escamosos, 369 espécies) e Myrmecophilidae (grilos formigas; 71 espécies). Estas últimas famílias pertencem a superfamília Gryllotalpoidea (SONG et al., 2015). Cada uma destas quatro famílias mencionadas são consideradas monofiléticas.

O gênero de maior diversidade dentro de Gryllidae está representado por *Gryllus* (verdadeiros grilos), com mais de 90 espécies (<http://orthoptera.speciesfile.org/Common/basic/Taxa.aspx?TaxonNameID=1122353>) distribuídas mundialmente. *Gryllus* tem sido usado como organismo modelo para explorar importantes tópicos em biologia evolutiva, desde estudos em comportamento ecológico a estudos fisiológicos e genéticos (HOWARD et al., 2002; FEDORKA et al., 2004; BUSSIÈRE et al., 2006; ZERA et al., 2006; ANDRÉS et al., 2006, 2008, 2013; SHAW et al., 2007, 2009; MAROJA et al., 2008; HARTFELDER e EMLÉN, 2011; DANBARA et al., 2010; LYNCH et al., 2010; ELLISON et al., 2011; MITO et al., 2011; TONIOKA et al., 2012; ZENG et al., 2013). Por exemplo, o som acústico (chamada de acasalamento) de *Gryllus* tem sido usado no contexto de comportamento sexual intraespecífico, sugerindo que a acústica de grilos são alvos para seleção sexual (OTTE, 1992; GREEN-FIELD, 1997). De fato, espécies relacionadas de grilos são identificadas usando o som acústico do macho. Caracteres sexualmente selecionados, tendem a evoluir mais rapidamente em comparação com aqueles não relacionados ao sexo (SWANSON et al., 2001, 2004; CLARK e SWANSON, 2005; KERN et al., 2004; ANDRÉS et al., 2006; LANDE, 1981; WEST-EBERHARD, 1983) e, de fato, seleção sexual dos cantos em machos foi postulada como uma causa provável para especiação extremamente rápida em um gênero havaiano, *Laupala* (MENDELSON e SHAW, 2005). O som acústico, ao mesmo tempo, pode atrair predadores e parasitas, o que pode proporcionar uma forte pressão seletiva para que os grilos passem rapidamente ao silêncio (PASCOAL et al., 2014), ou podendo em alguns casos serem selecionados cantos (sinais) modificados a ultrassom para que sejam menos conspícuos aos predadores (ROBILLARD et al. 2007). Portanto, é concebível que tanto a seleção sexual do som acústico, bem como a seleção natural, tenham desempenhado um papel importante na diversificação dos grilos nos últimos 200 milhões de anos, o que pode ter contribuído para a diversidade atual.

Gryllidae também contém o gênero *Eneoptera* (Eneopterinae). O gênero é endêmico na

América Central e do Sul, sendo representado por sete espécies (<http://orthoptera.speciesfile.org/Common/basic/Taxa.aspx?TaxonNameID=1127982>).

Eneoptera representa um modelo interessante para estudar evolução da comunicação acústica, que é uma característica importante para separar espécies de grilos. Os aparelhos estridulatórios destas espécies são diversos, e como resultado, esta diversidade é refletida também no som acústico, que mostram características acústicas únicas, como amplitude de modulação de frequência assim como frequências muito altas (ROBILLARD e DESUTTER-GRANDCOLAS, 2006).

Mogoplistidae contém o gênero *Cycloptiloides*, representado por 14 espécies (<http://orthoptera.speciesfile.org/Common/basic/Taxa.aspx?TaxonNameID=1128369>).

Cycloptiloides americanus (uma das espécies estudadas na presente tese) foi coletado pela primeira vez em Cuba, sendo encontrada posteriormente na América do Norte e no Brasil, sempre associada a habitações humanas (MESA et al. 2002). A espécie apresenta tamanho pequeno (5 a 7 mm), hábitos furtivos, e raramente é visto durante o dia em que permanece escondido. Os indivíduos só se movem para se alimentar durante a noite, momento em que o som acústico dos machos de baixa intensidade é ouvido.

1.3. DNA repetitivo em eucariotas

Sequências de DNA repetitivo representam uma parte substancial dos genomas de eucariotos, incluindo arranjos em tandem e repetições dispersas. Repetições em tandem são representadas por DNA satélites (DNAsat, sequências altamente repetitivas), microsatélites e minissatélites (sequências moderadamente repetitivas) e algumas famílias multigênicas (sequências alta ou moderadamente repetitivas), enquanto que repetições dispersas são representadas por transposons e retrotransposons (sequências altamente repetitivas) (CHARLESWORTH et al., 1994; LEWIN, 2004; BIÉMONT e VIERA, 2006; RICHARD et al., 2008). Em muitos genomas eucariotos o DNA repetitivo representa mais da metade do conteúdo de DNA nuclear (GREGORY, 2005; LYNCH, 2007), fato que dificulta os projetos focados em montagens de genomas. Assim, os DNA repetitivos apresentam um desafio e por tanto são excluídos nas análises gênicas ou na reconstrução de longos *contig* com baixo número de cópias. Neste tópico serão apresentadas informações quanto a natureza, comportamento e papel dos DNAs repetitivos no genoma, incluindo sequências em tandem e dispersas e seus transcritos. Será dada especial ênfase nas sequências mais abundantes, tais como, as famílias multigênicas, elementos

transponíveis e DNAsat. Além disso, será abordado o papel destas frações genômicas na organização estrutural de cromossomos e os mecanismos envolvidos na evolução das sequências.

1.3.1. Famílias multigênicas

As famílias multigênicas pertencem a distintos grupos de genes originados de um ancestral comum com similaridade estrutural e funcional (NEI e ROONEY, 2005). Em eucariotos, as famílias multigênicas mais estudadas são os genes de RNA ribossômico (RNAr) que estão representados por duas distintas famílias organizadas em tandem e repetidas centenas de milhares de vezes, ocupando um ou vários loci genômicos. O DNAr 45S ou DNAr maior, está estruturalmente constituído por três diferentes genes (18S, 5.8S e 28S), separados entre si por espaçadores transcritos internos (ITS, *internal transcribed spacer*). O gene codificante de RNAr 45S inteiro tem aproximadamente 9 Kilobases (Kb) de comprimento. Este *cluster* é transcrito pela Polimerase I. Por sua vez, o DNAr 5S ou DNAr menor é formado por múltiplas repetições de uma sequência altamente conservada de 120 pares de bases (pb), separados entre si por espaçadores não transcritos chamados NTS (*non-transcribed spacer*) de tamanho variável. Este gene é transcrito pela Polimerase III (LONG e DAWID 1980; SINGER e BERG, 1991; LEWIN, 2004; NEI e ROONEY, 2005).

Em eucariotos superiores, os transcritos codificantes de proteínas contém múltiplos íntrons que devem ser removidos de RNAs nascentes mediante *splicing* pela maquinaria do spliciossomo, sendo isto um processo essencial para a maturação do RNA mensageiro (RNAm) (NEI e ROONEY, 2005; WEST, 2012). O spliciossomo consiste de um complexo RNA-proteína, composto de ~200 proteínas associadas com cinco tipos diferentes de famílias gênicas não codificantes, nomeados RNAsn U (*small nuclear RNA*): U1, U2, U4, U5, U6 (BRINGMANN e LÜHRMANN, 1986; NILSEN, 2003; VALADKHAN, 2005). Genes de RNAsn U foram isolados em vários eucariotos, apresentando múltiplas cópias organizadas em tandem, em locais específicos ou dispersas no genoma (WISE e WEINER, 1980; MARZLUFF et al., 1983; MATTAJ e ZELLER, 1983; WATANABE-NAGASU et al., 1983; VAN ARSDELL e WEINER, 1984; CABRAL-DE-MELLO et al., 2012; ANJOS et al., 2014; BARZOTTI et al., 2003; MANCHADO et al., 2006; PELLICCIA et al., 2001). Entre espécies diferentes, regiões codificadoras de genes RNAsn U exibem muita similaridade uma com outras, mas as regiões intergênicas mostram heterogeneidade dentro de cada espécie (LIAO et al., 1998; PAVELITZ et al., 1995, 1999;

CABRAL-DE-MELLO et al., 2012; ANJOS et al., 2014), sugerindo seleção negativa atuando eficientemente nas regiões codificantes. Por outro lado, em organismos tão distantes como moluscos, gafanhotos e peixes, foi mostrada uma associação de genes snRNA U com RNA 5S, bem como elementos transponíveis, resultando assim em uma complexa organização genômica em alguns grupos de espécies (BARZOTTI et al., 2003; MANCHADO et al., 2006; PELLICCIA et al., 2001; VIERNA et al., 2011, 2013; CABRAL-DE-MELLO et al., 2012; ANJOS et al., 2014). Desse modo, os elementos transponíveis associados aos genes repetitivos poderiam servir com o uma fonte para a dispersão genômica (ver mecanismos de dispersão de elementos transponíveis no tópico 1.3.3).

As famílias gênicas codificadoras de histonas, produzem proteínas que são encontradas em núcleos de células eucarióticas empacotando e ordenando o DNA em unidades estruturais chamadas nucleossomos (LEWIN, 2004). Os nucleossomos consistem de 147 bp de DNA envolvido em torno de um disco proteico, constituído por quatro pares de proteínas de histonas nucleares (H2A, H2B, H3 e H4), enquanto a histona H1 une os nucleossomos adjacentes, empacotando-os (LEWIN, 2004; LUGER et al., 2012). Sem histonas, o DNA desencadeado nos cromossomos seria muito longo (uma relação comprimento a largura de mais de 10 milhões para 1 em DNA humano). Por exemplo, cada célula humana tem cerca de 1.8 metros de DNA, mas enrolada nas histonas tem cerca de 90 micrômetros de cromatina, que, quando duplicada e condensada durante a mitose, resulta em cerca de 120 micrômetros de cromossomos (LEWIN, 2004; LUGER et al., 2012). Geralmente, genes de histona estão organizados em tandem separados um do outro por espaçadores intergênicos (IGS, *intergenic spacer*) (NEI e ROONEY, 2005). Porém, em algumas espécies, como por exemplo no ouriço-do-mar (*Lytechinus pictus*) são observados arranjos dispersos em baixo número de cópias (CHILDS et al., 1982; MANDL, 1997). Usando DNA heteroduplex e análises de mapeamento de restrição foi mostrado que as regiões codificadoras de H3 são altamente conservadas. Entretanto, uma considerável quantidade de variação foi relatada nos espaçadores intergênicos, similarmente ao observado para genes de RNAr e RNAsn U, (COHN e KEDES, 1979a, 1979b; HOLT e CHILDS, 1984).

Três diferentes modelos de evolução foram propostos para famílias multigênicas: evolução divergente, evolução em concerto, e evolução por nascimento e morte (NEI e ROONEY, 2005). No modelo de evolução em concerto, famílias gênicas evoluem em concerto. Mutações que ocorrem nas repetições se espalham para outras repetições mediante i) *crossing-over* desigual que

pode ocorrer dentro de cromátides, entre cromátides irmãs ou entre cromossomas homólogos, ou por ii) conversão gênica que é a recombinação não-recíproca em que um segmento de DNA de um gene receptor é copiado de um gene doador (CHARLESWORTH et al., 1994; NEI e ROONEY, 2005). Um dos exemplos mais conhecidos de evolução em concerto vem de estudos de genes de DNAr 45S e genes de histona, embora nesta última família evolução por nascimento e morte também foi relatado (NEI e ROONEY, 2005). Na evolução por nascimento e morte, novos genes são gerados por duplicações de uma família gênica. Alguns genes duplicados são mantidos no genoma por um período longo de tempo, enquanto outros são “pseudogenizados” por deleção ou inativação (NEI e ROONEY, 2005). Este modelo de evolução foi proposto para genes de RNAr 5S, embora uma mistura de evolução em concerto e nascimento e morte foi também relatado para este *cluster* (ROONEY e WARD, 2005; ÚBEDA-MANZANARO et al., 2010; PERINA et al., 2011; PINHAL et al., 2011; VIERNA et al., 2011, 2013; VIZOSO et al., 2011; MERLO et al., 2013; BUENO et al., 2016). Finalmente, o modelo de evolução divergente foi proposto para famílias gênicas das hemoglobinas (α , β , γ) e mioglobinas (INGRAM, 1961). Estes genes filogeneticamente relacionados divergiram gradualmente por duplicações gênicas adquirindo novas funções –“neo-funcionalização”– (INGRAM, 1961; NEI e ROONEY, 2005).

1.3.2. DNA satélites

As frações de DNAs repetitivos correspondente aos DNAsat foram identificadas inicialmente por centrifugação em gradiente de densidade de alta velocidade da massa de DNA. Devido a esta fração ser altamente abundante e ter um conteúdo de AT/GC diferenciado na massa de DNA, eles formavam uma banda tipo satélite, dando surgimento o termo “*satellite DNA*”, mais conhecidos como DNAsat (LEWIN, 2004). Nas décadas 60 e 70, essas bandas foram recuperadas e usadas pela primeira vez em experimentos de hibridação *in situ* fluorescente (FISH, fluorescent *in situ* hybridization) com o objetivo de localizar sua posição cromossômica (PADUE, 1969; JOHN et al., 1969). Após o desenvolvimento da tecnologia de enzimas de restrição na década do 70, as repetições de DNAsat foram identificadas por digestão enzimática de DNA por meio de endonuclease de restrição, seguido de separação de tamanho pela eletroforese em gel de agarose. O sequenciamento de fragmentos DNA tratados com enzimas de restrição mostrou que muitas destas sequências eram altamente abundantes no genoma. Além disso, experimentos de FISH usando DNAsat como sondas mostrou que DNAsat estavam localizados em locais discretos, como

centrômeros, telômeros e regiões não recombinantes de cromossomos sexuais (CHARLESWORTH et al., 1994; RICHARD et al., 2008, PLOHL et al., 2008; LÓPEZ-FLORES e GARRIDO-RAMOS, 2012). Porém, em alguns casos, DNAsat dispersos na eucromatina foram igualmente relatados (CHARLESWORTH et al., 1994; PLOHL et al., 2008; LÓPEZ-FLORES e GARRIDO-RAMOS, 2012). Em muitos organismos, DNAsat são compostos de centenas a milhares de sequências não codificadoras arranjadas em tandem, com replicação tardia e orientada de uma maneira cabeça-cauda, mas com um comprimento de unidade de repetição muito mais curto que o *cluster* de DNAr 45S (CHARLESWORTH et al., 1994; LÓPEZ-FLORES e GARRIDO-RAMOS, 2012; SCHMIDT e HESLOP-HARRISON, 1998; PALOMEQUE e LORITE, 2008; RICHARD et al., 2008). Muitas vezes, o comprimento da repetição de DNAsat está relacionado ao *folding* do DNA em torno de nucleossomos, 150 pb mais o DNA de ligação, dando um comprimento de 160 a 180 pb (RICHARD et al., 2008; LÓPEZ-FLORES e GARRIDO-RAMOS, 2012).

Com o desenvolvimento de sequenciamento de alta resolução (*high-throughput DNA sequencing*), existem atualmente muitas abordagens bioinformáticas para identificar DNAsat abundantes. Em particular, o uso de programas para encontrar motivos de *k-mer* (sequências de nucleotídeos *k* longas, onde *k* é tipicamente de 16 a 128 pb), onde alguns motivos são muito superiores ao número aleatório esperado, podem ser usados para encontrar sequências de DNAs repetitivos. O *pipeline* RepeatExplorer, um método de clusterização de *reads* de sequências baseado em grafos, tem sido uma ferramenta valiosa para identificar e caracterizar elementos repetitivos com médio e alto número de cópias no genoma, particularmente DNAsat (NOVÁK et al., 2013). Entretanto, dada a homogeneidade das sequências de DNAsat, a montagem dessas sequências em longos pb ainda representa um desafio. Recentemente, novas tecnologias, como por exemplo PacBio ou Nanopore Minion, permitem a leituras de *reads* com vários Kb, mas ainda não conseguem ler trechos contínuos de milhões de bases de DNA. Por exemplo, o principal componente não montado do genoma humano, um dos mais estudados, é correspondente a DNAsat. A mais recente revisão de Miga (2015) destaca os pontos fortes e fracos do uso de variantes de sequência de baixa cópia como marcadores para ajudar no procedimento de montagem do genoma humano; as estratégias para integrar sequências de satélite em estudos genômicos também são discutidas por Miga (2015).

Em geral, famílias de DNAsat diferem em identidade de sequências, número de cópias e distribuição cromossômica (PLOHL et al., 2008; KING et al., 1995; WANG et al., 1995; VERSHININ e HESLOP-HARRISON, 1998). DNAsat estão sujeitos a evolução em concerto intragenômica, resultando em homogeneização mais eficiente de repetições dentro de espécies que entre espécies, e também entre repetições localizadas no mesmo arranjo/cromossomo que entre diferentes arranjos/cromossomos (PLOHL et al., 2008; LÓPEZ-FLORES e GARRIDO-RAMOS, 2012; KUHN et al., 2012). Mudanças dentro e entre sequências repetidas em tandem ocorrem por mutações, causando contração ou expansão de variantes repetitivas que podem resultar em mudanças nos perfis de DNAsat (CHARLESWORTH et al., 1994; STEPHAN e CHO, 1994), resultando assim, no aumento ou diminuição na abundância e diversidade de DNAsat. Alternativamente, a alta diversidade de DNAsat *per se* pode facilitar ou provocar rearranjos cromossômicos complexos. É bem-sabido que rearranjos cromossômicos podem envolver sequências de DNA altamente repetitivo, uma vez que estes fornecem locais para a reestruturação do cariótipo sem efeitos prejudiciais sobre a integridade das sequências codificadoras (HSU et al. 1978; HOLMQUIST e DANCIS, 1979; SINGER e DOWNHOWE, 1979). Evidências em relação ao mencionado anteriormente podem ser encontradas em espécies de hamsters do gênero *Phodopus*, onde a alta dinâmica de sequências repetitivas foi responsável pela instabilidade cromossômica que levou a rearranjos cromossômicos complexos, resultando em uma extensiva divergência cariotípica destas espécies de roedores (PAÇO et al., 2015). Conseqüentemente, estes rearranjos cromossômicos contribuíram para o potencial adaptativo destas espécies em distintos ambientes (PAÇO et al., 2015).

Embora a maioria dos DNAsat estejam presentes na heterocromatina densamente condensada, a qual é considerada transcricionalmente silenciosa e inerte, as evidências de transcrição deste tipo de elementos foram documentadas em insetos, vertebrados e plantas (PEZER et al., 2011; UGARKOVIĆ, 2005; BISCOTTI et al., 2015). Foi demonstrado também que DNAsat poderiam ser processados em pequenos RNAs, como siRNAs (small interfering RNA) e piwiRNAs (Piwi-interacting RNA), envolvidos, por exemplo, em processos epigenéticos de formação da heterocromatina em grupos de espécies tão distantes como *Schizosaccharomyces pombe*, *Drosophila*, nemátodos e plantas (VOLPE et al., 2002; PAL-BHADRA et al., 2004; GREWAL e ELGIN, 2007; FAGEGALTIER et al., 2009). Em geral, transcritos de DNAsat estão envolvidos na i) formação e manutenção da heterocromatina centromérica e telomérica, com

consequente impacto na integridade cromossômica e estabilidade genômica, ii) na identidade de centrômeros, iii) e no alongamento e replicação de telômeros (BISCOTTI et al. 2015). Recentemente foi relatado que TCAST1, o satDNA mais abundante em *Tribolium castaneum*, desempenha um papel importante na modulação da expressão gene-proteína (FELICIELLO et al., 2015). DNAsat são capazes de adotar estruturas de *folding* específicas, conhecidas como estruturas repetitivas altamente ordenadas (HORs, *high-order repeats structures*), na qual um bloco de unidades repetitivas múltiplas forma uma unidade dobrada maior que são repetidas em tandem e com capacidade de atrair proteínas nucleares (PODGORNAYA et al., 2013). Este *folding* faz que os DNAsat sejam portadores potenciais de um *barcode* de cromatina, possivelmente contribuindo para a identidade celular bem como para a especificidade de territórios cromossômicos (PODGORNAYA et al., 2013).

1.3.3. Elementos transponíveis (TEs)

TEs são ubíquos na maior parte dos eucariotas e podem compreender uma grande porção do DNA. Por exemplo, em humanos e milho, TEs representam mais de 50% do genoma, na espécie mostarda preta ~ 40% e em *Drosophila* ~ 20% (PIMPINELLI et al., 1995; KIDWELL e LISCH, 1997; LYNCH, 2007; MILLS et al., 2007; WANG et al., 2017). Uma das características mais notáveis dos TEs é a capacidade de movimento entre locais não homólogos dentro dos genomas. TEs podem mudar seus locais específicos e esse movimento pode influenciar sua acumulação e a geração de polimorfismos. Devido a estas características, TEs foram primeiro considerados elementos genéticos egoístas ou parasitas genômicos (DOOLITTLE e SAPIENZA, 1980; ORGEL e CRICK, 1980; LEWIN, 2004). No entanto, atualmente acredita-se que eles são importantes para arquitetura e função dos genomas influenciando trajetórias evolutivas e adaptações ecológicas (KAZAZIAN, 2004; BIÉMONT e VIEIRA, 2006; FESCHOTTE e PRITHAM, 2007; BÖHNE et al., 2008; BIÉMONT, 2010; HUA-VAN et al., 2011).

Os TEs são classificados em duas classes de acordo com sua estrutura e mecanismo de transposição. Retrotransposons, ou classe I, se amplificam mediante um RNA intermediário; já os DNA transposons, ou classe II, se movimentam e amplificam usando DNA (WICKER et al., 2007; KAPITONOV e JURKA, 2008). Retrotransposons (considerados os TEs eucarióticos mais abundantes) incluem elementos com repetições terminais longas (LTR) e outras ausentes delas (non-LTR), como por exemplo, LINEs (*long interspersed nuclear elements*) e SINEs (*short*

interspersed elements) (VITTE e BENNETZEN, 2006; KRAMEROV e VASSETZKY, 2011). Entre DNA transposons, os elementos *Mariner* constituem uma superfamília amplamente propagada em vários organismos (LANGIN et al., 1995; AUGÉ-GOUILLOU et al., 1995; PLASTERK et al., 1999; FESCHOTTE e WESSLER, 2002; WITHERSPOON e ROBERTSON, 2003; JACOBS et al., 2004; SINZELLE et al., 2006). Um segundo tipo de DNA transposons ainda pouco caracterizado são os chamados *MITEs* (*miniature inverted repeat transposable elements*) (WESSLER et al. 1995; FESCHOTTE e PRITHAM, 2007).

Devido a seu modo de amplificação, TEs estão frequentemente dispersos no genoma, embora alguns possam mostrar uma concentração ou depleção em regiões cromossômicas particulares, tais como centrômeros ou locais subteloméricos (REBUZZINI et al., 2009; MONTIEL et al., 2012; PALACIOS-GIMENEZ, et al., 2014). Em vários genomas, a natureza dos elementos presentes difere. Algumas classes predominam em uma espécie, enquanto outras classes podem ser muito mais abundantes em outras espécies. Em plantas, o conteúdo de TEs varia influenciando o cariótipo e tamanho genômico, que finalmente contribuem para diferenciar genomas entre espécies relacionadas. Essas diferenças genômicas podem também contribuir para a formação de novas espécies poliploides (SANTOS et al., 2015). Recentemente foi investigado o papel das expansões e contrações de retrotransposons na evolução dos genomas do gênero de planta *Brachiaria*, mediante a identificação de motivos característicos em dados transcriptômicos, bem como a distribuição cromossômica e abundância dessas sequências em cromossomos de espécies diploides e poliploides (SANTOS et al., 2015). Os resultados indicam que essas sequências parecem ter evoluído de forma específica e com a variação característica do número de cópias entre genomas levando a especiação no gênero (SANTOS et al., 2015).

Embora satDNAs e TEs apresentem diferenças em relação a estrutura, organização genômica, mecanismos de propagação e dinâmica evolutiva, alguns recentes estudos ressaltam que esses dos tipos de sequências repetitivas estão mutuamente relacionadas. Baseado na homologia entre DNAsat e TEs, uma convincente hipótese postula que repetições em tandem poderiam derivar de transposons ou retrotransposon mediante recombinação ectópica, *crossing-over* entre loci não homólogos, ou inclusive por amplificação de repetições internas preexistentes (MEŠTROVIĆ et al., 2015). Transição inversa pode ocorrer, já que fragmentos de repetições em tandem podem ser encontrados como parte de TEs, mais usualmente em DNA transposons que em retrotransposons, tanto em animais quanto em plantas (DIAS et al., 2014, 2015; MEŠTROVIĆ et

al., 2015).

Usando ferramentas bioinformáticas, filogenéticas e citogenômicas, Dias e colaboradores (2014) descreveram um novo DNA transposon nomeado TETRIS que tem contribuído para moldar o genoma de espécies de *Drosophila*, fornecendo repetições internas em tandem que serviram como *hotspot* para construção e amplificação de satDNA. Mais recentemente, Dias e colaboradores (2015) reportaram o surgimento de uma sequência tipo satélite a partir de uma repetição interna de um *Helitron* (DINE-TR1) em distantes linhagens de *Drosophila*. Este evento ocorreu independentemente em pelo menos três linhagens, sugerindo que a amplificação está em um estágio inicial, ou que outros mecanismos estão operando para evitar o aumento em tamanho do DNAsat (DIAS et al., 2015, 2016). Mediante experimentos de FISH, os autores demonstraram que *Tetris* e *Helitron* localizam em regiões cromossômicas correspondentes com a β -heterocromatina de *Drosophila*, indicando inserção preferencial nestas regiões (DIAS et al., 2014, 2015). Pequenos transcritos derivados de DINE-TR1 são expressos em gônadas de *D. virilis*, possibilitando seu envolvimento na via de piwiRNA e, garantindo assim, uma gametogênese fiel, bem com a prevenção de transposições deletérias em células germinativas (DIAS et al., 2015, 2016). No genoma como um todo, ou em sítios específicos, estes pequenos transcritos que interagem com as proteínas PIWI poderiam ter papéis regulatórios afetando o estado da heterocromatina (DIAS et al., 2015, 2016).

1.4. Mecanismos de determinação sexual em insetos

A determinação sexual é um processo biológico fundamental, afetando não apenas a diferenciação sexual das gônadas, mas também o desenvolvimento da maioria dos órgãos, o qual finalmente leva a diferenças sexuais específicas de comportamento, fisiologia e morfologia. Entre eucariotos, o sexo é determinado por uma ampla gama de mecanismos (MARIN e BAKER, 1998) que podem ser divididos em duas categorias: determinação sexual genética (GSD, *genetic sex determination*) e determinação sexual ambiental (ESD, *environmental sex determination*) (BULL, 1983; GRAVES, 2008). GSD é atribuído à segregação genética de genes, os quais frequentemente residem em cromossomos sexuais que iniciam vias alternativas de desenvolvimento determinantes do sexo, enquanto que ESD é iniciado por estímulos ambientais que presumivelmente desencadeiam sinais genéticos alternativos, que regulam os genes determinantes do sexo masculino ou feminino (BULL, 1983; GRAVES, 2008). Embora o GSD seja um sistema mais

prevalente em animais, a ESD ocorrem em distintos táxons como rotíferos, nematóides, crustáceos, insetos, peixes e répteis (BULL, 1983; KORPELAINEN, 1990; GRAVES, 2008). Entre alguns anfíbios e répteis, por exemplo, o sexo é determinado ambientalmente depois da fertilização. A temperatura do ovo durante um certo período do desenvolvimento é um fator decisivo na determinação do sexo. Muitas vezes, ovos incubados a baixa temperatura (22-27°C) produzem um sexo, enquanto que ovos incubados a altas temperaturas (30°C e acima deste) produzem outro (BULL, 1983; GRAVES, 2008). Por exemplo, na tartaruga de lagoa europeia, *Emys obicularis*, ovos incubados a temperaturas acima de 30°C produzem fêmeas, enquanto que temperaturas abaixo de 25°C produzem machos. A temperatura limite na qual a proporção de sexo é 1:1 é 28.5°C (PIEAU et al., 1994). Em contraste com os organismos modelo que utilizam o sistema GSD, os organismos que usam ESD são geneticamente pouco estudados. Por consequência, poucos genes (como *sox9*) ou hormônios (como estrógeno) aparentemente envolvidos na ESD já foram encontrados em vertebrados (SPOTILA et al., 1998; MORENO-MENDOZA et al., 1999; GILBERT, 2000; GRAVES, 2008). Com isso, as funções permanecem em grande parte desconhecidas, prejudicando a nossa compreensão da ESD e dificultando a comparação de genes determinantes do sexo entre ambos os sistemas. Por este motivo, o objetivo neste tópico é revisar brevemente as bases genéticas da GSD que são responsáveis pela produção do fenótipo macho e fêmea, focados principalmente em insetos.

GSD tem sido amplamente documentado em uma variedade de grupos de insetos como Diptera, Hymenoptera, Coleoptera, Lepidoptera e Hemiptera, sendo o díptero *D. melanogaster* (*fruitfly*) um dos modelos principais de estudo (SÁCHEZ, 2008; VERHULST et al., 2010; MULLON et al., 2012). Em *fruitfly* a proporção X:A (X:A=1) ou a dose de cromossomo X atua como sinal primário para iniciar a cascata de determinação sexual. Nesta espécie, GSD é controlado pelo gene *sex-lethal* (*Sxl*) que contém duas regiões promotoras nomeadas P_{early} e $P_{\text{maintenance}}$, sendo o último o promotor tardio (SÁCHEZ, 2008; VERHULST et al., 2010). O sinal primário da GSD depende dos elementos sinalizadores ligados ao cromossomo X (XSE, X-linked signal elements) que ativam o P_{early} em indivíduos portadores do genótipo fêmea XX. Assim, *Sxl* é ativo somente em fêmeas e não em machos (SÁCHEZ, 2008; VERHULST et al., 2010; MULLON et al., 2012). A transcrição da região P_{early} do *Sxl* resulta num transcrito primário que passa por um processo de *splicing* alternativo resultando na proteína funcional SXL, em fêmeas. Esta proteína primária permite produção de uma proteína tardia SXL funcional, que mantém o

splicing alternativo *Sxl* específico em fêmeas por auto-regulação (SÁCHEZ, 2008; VERHULST et al., 2010; MULLON et al., 2012). A proteína SXL controla o *splicing downstream* do gene *transformer (tra)* resultando na proteína funcional TRA, que interage com a proteína não específica do sexo chamada transformer2 (TRA2) (AMREIN et al., 1988; SÁCHEZ, 2008; VERHULST et al., 2010). Heterodímeros de TRA e TRA2 se ligam ao transcrito *doblexsex (dsx)* no meio do éxon 4, numa região nomeada elemento de repetição *dsx (dsxRE)*. O *dsxRE* está formado por sequências repetidas de 13 nucleotídeos TC(T/A) (T/A)C(A/G)ATCAACA (TIAN e MANIATIS, 1993), e entre os elementos cinco e seis do *dsxRE* está localizado um elemento acentuador rico em purina chamado PRE, o qual é necessário para a união específica de TRA2 ao *dsxRE* (SÁCHEZ, 2008; VERHULST et al., 2010). A ligação de heterodímeros TRA/TRA2 aos sites *dsxRE* e PRE retém o éxon 4 do pré-mRNA *dsx* resultando em um *splicing* fêmea-específico de *dsx* na cascata final de determinação sexual, e gerando a proteína DSX específica para fêmeas (SÁCHEZ, 2008; VERHULST et al., 2010; MULLON et al., 2012).

Em machos heterogaméticos (XY), os níveis XSEs são insuficientes para produzir a transcrição de P_{early} , e conseqüentemente, a proteína SXL não é sintetizada. Assim, o pré-RNA *Sxl* do $P_{\text{maintenance}}$ passa por *splicing* alternativo padrão específico no macho, produzindo uma proteína SXL truncada não funcional (SÁCHEZ, 2008; VERHULST et al., 2010; MULLON et al., 2012). A ausência de SXL leva ao *splicing* alternativo padrão do pré-mRNA de *tra* e, por tanto, a uma proteína TRA não funcional. Sem TRA, o pré-mRNA de *dsx* passa por *splicing* alternativo padrão gerando uma proteína DSX específica do sexo masculino (SÁCHEZ, 2008; VERHULST et al., 2010; MULLON et al., 2012).

O complexo TRA/TRA2 também controla o *splicing* alternativo sexo-específico do gene *fruitless (fru)*. A ausência de TRA nos machos induz o *splicing* alternativo do pré-RNA *fru* produzindo uma proteína FRU não funcional (Ryner et al., 1996; SÁCHEZ, 2008; VERHULST et al., 2010; MULLON et al., 2012). O gene *fru*, bem como o *dsx*, está envolvido no desenvolvimento sexual masculino do sistema nervoso central (RIDEOUT et al., 2007; MULLON et al., 2012), o que é necessário para o comportamento de acasalamento no macho (RYNER et al., 1996; SHIRANGI et al., 2006).

Ortólogos de genes (*Sxl*, *tra*, *tra2*, *dsx* e *fru*) envolvidos na cascata de determinação sexual tem sido descrito em vários grupos de insetos, como por exemplo, Diptera, Hymenoptera, Coleoptera e Hemiptera (SÁCHEZ, 2008; VERHULST et al., 2010; MULLON et al., 2012),

indicando mecanismos de determinação sexual e função gênica conservada entre alguns grupos de insetos. No entanto, em insetos com sistema de cromossomos sexuais X0/XX (exemplo, Orthoptera, Blattodea, Isoptera) foi proposto que o sinal primário para determinação sexual poderia depender da dose de cromossomo X (VERHULST et al., 2010). Se o gene *tra* está presente e como ocorre o mecanismo de regulação do *tra* nestes grupos de insetos continua sendo uma questão aberta e interessante para ser respondida. Certamente, a exploração das tecnologias de sequenciamento de nova geração (NGS, *next-generation sequencing*) poderiam contribuir com estes estudos.

1.5. Origem e evolução de cromossomos sexuais

Os cromossomos sexuais se originaram diversas vezes de forma independente de um par de homólogos autossômicos (OHNO, 1967; BULL, 1983), representando um caso notável de convergência genômica devido os cromossomos sexuais não relacionados compartilharem muitas propriedades entre táxons distantes (BACHTROG et al., 2011, 2014; BEUKEBOON e PERRIN, 2014; PEICHEL, 2017).

Em princípio, a aquisição de um gene determinante do sexo inicia a evolução de cromossomos sexuais, e, em muitas linhagens, cromossomos sexuais originalmente homomórficos se diferenciam em cromossomos sexuais heteromórficos (BULL, 1983; CHARLESWORTH, 1996). Um pré-requisito para que os cromossomos sexuais se diferenciem e evoluam independentemente é que a recombinação seja suprimida entre os proto-cromossomos sexuais homomórficos. Neste contexto, alelos antagônicos sexuais benéficos para um sexo, mas prejudiciais para o outro, poderiam se acumular nos proto-cromossomos sexuais e seriam selecionados para não recombinar (CHARLESWORTH, 1996, KAISER e BACHTROG, 2010; WRIGTH et al., 2016). Deste modo, a ausência de recombinação em parte ou toda a extensão dos cromossomos sexuais limitados a um sexo (Y o W) resultaria na degeneração da região não recombinante. Em outras palavras, a região não recombinante dos cromossomos Y o W pseudogeniza ou perde a maior parte dos seus genes ancestralmente presentes no proto-Y ou proto-W (CHARLESWORTH et al., 1994; RICE, 1996; STEINEMANN e STEINEMANN, 2005; KAISER e BACHTROG, 2010; WRIGTH et al., 2016). A região não recombinante pode inclusive passar por um processo de expansão mediante a aquisição adicional de alelos antagônicos sexuais, levando a formação dos conhecidos “estratos” ou “aglomerados espaciais” (*clusters* de genes) de

ortólogos nos cromossomos X-Y e Z-W com estimativas de divergência similares em mamíferos (LAHN e PAGE, 1999), pássaros (MOGHADAM et al., 2012; WRIGTH et al., 2014), peixes (ROESTI et al., 2013; WHITE et al., 2015; TOMASZKIEWICZ et al., 2017) e plantas (BERGERO et al., 2017), que também sofrem perda de função gênica (pseudogenização) e heterocromatinização.

Em alguns grupos, como por exemplo humanos (CARVALHO, 2002) e *Drosophila* (SKALESTKY et al., 2003), os cromossomos sexuais exibem poucos sinais de sua história evolutiva, enquanto outros táxons como peixes (EZAZ et al., 2006) e cobras (EZAZ et al., 2006; VICOSO et al., 2013a) apresentam vários estados de transição de cromossomos sexuais morfologicamente idênticos até completamente diferenciados. Atualmente, ainda não está claro por que algumas espécies suprimem a recombinação ao longo dos cromossomos sexuais e adquirem cromossomos sexuais heteromórficos XY o ZW, enquanto que outros conservam cromossomos sexuais homomórficos. Interpretações plausíveis poderiam ser devido a uma falta de mutações antagônicas sexuais em alguns grupos, ou pela resolução de conflitos impostos por mutações antagônicas sexuais para evolução sexo específico ou também por causa da expressão enviesada do sexo (VICOSO et al., 2013b; WRIGTH et al., 2016).

Em outros casos, cromossomos sexuais podem evoluir e divergir mediante rearranjos cromossômicos complexos. Por exemplo, uma inversão cromossômica abrangendo proximidades de um locus determinante do sexo ou inclusive um loci antagonista sexual poderia interromper a recombinação por falta de homologia entre as regiões envolvidas no rearranjo, e portanto, levando a divergência entre cromossomos sexuais (CHARLESWORTH et al., 2005). Por exemplo, cromossomos sexuais em muitos animais e plantas mostram evidências de *clusters* espaciais (estratos) de ortólogos X-Y ou Z-W com estimativas de divergência semelhantes (LAHN e PAGE, 1999; MOGHADAM et al., 2012; WRIGTH et al., 2014; ROESTI et al., 2013; WHITE et al., 2015; BERGERO et al., 2017; TOMASZKIEWICZ et al., 2017). Esses estratos são consistentes com eventos de inversões instantaneamente interrompendo a recombinação para todos os loci envolvidos. Recentemente, combinando bioinformática e citogenética, Palacios-Gimenez e colaboradores (2017) observaram cinco variantes citológicas bem definidas para o neo-Y do gafanhoto *R. bergii*, originados mais parcimoniosamente por uma inversão cêntrica seguido de múltiplas inversões paracêntricas e amplificação de DNAsat, que contribuiram para a supressão de recombinação e divergência entre o neo-X e neo-Y. No entanto, relatos de cromossomos sexuais

nascentes sugerem que a supressão da recombinação é inicialmente heterogênea em todos os cromossomos sexuais (BERGERO et al., 2013; NATRI et al., 2013), implicando que a supressão da recombinação evolui inicialmente por outro mecanismo irregular, inconsistente com as inversões em grande escala.

A recombinação é um processo dinâmico e heterogêneo com taxas variando extensivamente entre locus e dentro de loci genômicos, assim como entre sexos (LENORMAND e DUTHEIL, 2005; KONG et al., 2010). Por exemplo, nas espécies onde a recombinação de cromossomos sexuais ocorre nos dois sexos, as taxas locais de recombinação específicas para um sexo podem ser inicialmente importantes para a divergência dos cromossomos sexuais, embora ainda não se conheça um mecanismo de recombinação sexo-específico (heteroquiasmose). Independentemente do mecanismo, uma vez que a recombinação foi interrompida no sexo heterogamético, a seleção para manter a ordem do gene é abolida (FLOT et al., 2013) e as inversões são menos propensas a serem selecionadas contra neste sexo. A seleção relaxada contra inversões sugere que essas podem acompanhar a supressão de recombinação. Portanto, não está claro se as inversões catalisam ou são uma consequência da interrupção da recombinação entre os cromossomos sexuais (JAROLA et al., 1998; LAHN e PAGE, 1999; KIRKPATRICK, 2005; ROSS et al., 2005). Assim, o papel dos rearranjos cromossômicos e particularmente das inversões na evolução de cromossomos sexuais não foram ainda bem explorados, mas eles oferecem uma alternativa plausível para a evolução de supressão de recombinação.

Uma consequência comum da supressão de recombinação entre cromossomos sexuais é o acúmulo gradual de DNA repetitivo, o que pode levar a um aumento de curto prazo no tamanho do Y ou W, mas que normalmente resulta em deleções em larga escala, redução no tamanho físico do cromossomo limitado ao sexo (Y ou W) e cromossomos sexuais altamente heteromórficos. Entre estes DNA repetitivos que se acumulam em cromossomos sexuais se encontram, por exemplo, elementos de transposição, DNAsat, microssatélites e famílias multigênicas (CHARLESWORTH et al., 1994; STEINEMANN et al., 1992, 1993; KEJNOVSKY et al., 2009; 2013; MATSUNAGA, 2009; STEFLOVA et al., 2013; PALACIOS-GIMENEZ et al. 2013; PALACIOS-GIMENEZ et al., 2015, 2017). Alta densidade de transposons foi observada, por exemplo, nas regiões não codificadoras do neo-cromossomo Y de *Drosophila miranda*, comparado com suas regiões homólogas do cromossomo X (CHARLESWORTH et al., 1994; STEINEMANN et al.; 1992, 1993). Um exemplo de alta densidade de TEs (~50%) foi recentemente descoberta na

região não recombinante do cromossomo W do pássaro “*flycatcher*” (SMEDS et al., 2015). Esta proporção é muito maior do que no cromossomo Z e nos autossomos (~9% e ~6%, respectivamente), ilustrando o acúmulo de repetições em regiões não recombinantes de cromossomos sexuais limitados ao sexo (SMEDS et al., 2015; TOMASZKIEWICZ et al., 2017). Casos de acúmulo diferencial de famílias multigênicas, microsatélites e satDNAs foram observados por exemplo na espécie vegetal *Rumex acetosa* (KEJNOVSKY et al., 2013; STEFLOVA et al., 2013), no gafanhoto *Ronderosia bergii* (PALACIOS-GIEMENEZ et al., 2015a, 2017b). Este acúmulo de sequências repetitivas explicaria porque os cromossomos Y ou W das espécies citadas tem um padrão de tamanho maior que os cromossomos X, embora possua menor quantidade de genes. Por sua vez, o acúmulo de elementos repetitivos poderia contribuir com a degeneração do cromossomo Y, afetando seu modo de expressão gênica (STEINEMANN e STEINEMANN, 1992; STEINEMANN et al., 1993; BERGERO et al., 2008). Esta “degeneração genética” seria a causa da divergência morfológica e, conseqüentemente, genética dos neo-cromossomos sexuais (CHARLESWORTH et al., 2005, KAISER e BACHTROG, 2010). Assim, o neo-Y é selecionado para não recombinar talvez em resposta ao acúmulo de sequências que são benéficas para o macho, mas não para as fêmeas (RICE, 1996; CHARLESWORTH et al., 2005).

1.6. Evolução de cromossomos sexuais em Orthoptera

O sistema cromossômico de determinação sexual comumente encontrado em Orthoptera Saltatoria é do tipo $X0^{\text{♂}}/XX^{\text{♀}}$, entretanto, supõe-se que o sistema sexual $X0$ evoluiu a partir do sistema sexual ancestral do tipo $XY^{\text{♂}}/XX^{\text{♀}}$, sugerindo que a reversão para o estado $X0$ ocorreu em um antepassado comum desta linhagem (WHITE, 1973, HEWITT, 1979). O sistema $X0^{\text{♂}}/XX^{\text{♀}}$ de Orthoptera parece ser evolutivamente bastante conservado em muitas espécies quando comparado com as variedades de mecanismos cromossômicos de determinação sexual observado em outros insetos, como por exemplo, Diptera, Coleoptera e Lepidoptera (HEWITT, 1979; YOSHIDO et al., 2013; NGUYEN et al., 2013). O sistema sexual $X0^{\text{♂}}/XX^{\text{♀}}$ de Orthoptera envolve um equilíbrio entre o X e autossomos (WHITE, 1973; HEWITT, 1979; CHARLESWORTH, 1996), como foi descoberto por Bridges em seus estudos com *Drosophila melanogaster* (CHARLESWORTH, 1996). Isto sugere que os genes presentes no cromossomo X controlam o desenvolvimento feminino e os genes presentes em autossomos controlam o desenvolvimento masculino (CHARLESWORTH, 1996; AYLING e GRIFFIN, 2002).

Na meiose masculina de grilos, tetigonídeos e gafanhotos, o cromossomo X representa um elemento único não pareado que exhibe heteropicnose ao longo de toda sua extensão (WHITE, 1973; HEWITT, 1979). Aparentemente, a heteropicnose do X indica uma condição genética inerte, sugerindo que o X carregue um número pequeno de genes ou regiões determinantes do sexo intercaladas ao longo de sua extensão. Entretanto, permanece o mistério se o cromossomo X de Orthoptera apresenta genes determinantes do sexo ou alguns genes relacionados com a fertilidade ou fecundidade dos indivíduos.

Além do sistema sexual $X_0^{\text{♂}}/XX^{\text{♀}}$, existem relatos de várias espécies com sistemas sexuais derivados do tipo neo-XY $^{\text{♂}}/XX^{\text{♀}}$, neo- $X_1X_2Y^{\text{♂}}/X_1X_1X_2X_2^{\text{♀}}$ e inclusive o neo- $X_1X_20^{\text{♂}}/X_1X_1X_2X_2^{\text{♀}}$. Estes sistemas foram originados por fusões cêntricas ou em tandem entre o cromossomo X ancestral e autossomos, além de inversões (WHITE, 1973; HEWITT, 1979; BIDAU e MARTÍ, 2001, MESA e GARCÍA-NOVO, 2001; MESA et al., 2001, 2002; FERREIRA e CELLA 2006; CASTILLO et al., 2010a, 2010b, 2014, 2016; PALACIOS-GIMENEZ et al., 2013, 2015, 2017). Em Orthoptera, sistemas neo-XY são formados quando dois cromossomos acrocêntricos, o X e um autossomo, quebram na região do centrômero e se fusionam. O produto da reestruturação resulta em um neo-X metacêntrico e um neo-Y acrocêntrico (WHITE, 1973; HEWITT, 1979). Com o objetivo de diferenciar os braços dos neo-cromossomos sexuais, White (1973) definiu uma terminologia a qual é ainda utilizada nos trabalhos atuais; o neo-X é formado pelos braços XL e XR, onde o XL corresponde ao cromossomo X do antigo sistema sexual X_0 e o XR é resultado do autossomo envolvido na fusão. Assim, o cromossomo homólogo ao braço XR passa a denominar-se neo-Y e está restrito a linhagem masculina (WHITE, 1973; HEWITT, 1979).

Evidências citológicas consideradas para um neo-XY de origem evolutiva recente são: sinapses entre o neo-Y e o braço XR durante paquíteno, ausência de heterocromatinização e quiasmas intersticiais entre ambos os neo-X e neo-Y. Este comportamento foi observado por exemplo em *Boliviacris noroestensis*, *Mariacris viridipes* e *Neuquenina fictor*. Em contraste, sistemas neo-XY antigos ou de origem avançada se caracterizam por: quiasmas/contatos restritos nas regiões distais, heterocromatinização, e menos frequentemente rearranjos estruturais adicionais, como por exemplo inversões. Estas características foram evidenciadas, por exemplo, em *Dichroplus vittigerum* e em espécies de *Zygoclistron*. Além disso, em cerca de 75 espécies com neo-XY, há pelo menos dez rearranjos estruturais envolvendo distintas regiões do neo-X ou neo-Y, como por exemplo, em *Ronderosia ommexehoides* e em quatro espécies de *Aleuas* com

inversão pericêntrica curta no neo-Y, *Spathalium audouini* com inversão pericêntrica alterando a morfologia do neo-Y para metacêntrico e três *Dichroplus*, *D. silveiraguidoi*, *D. vittatus* e *D. maculipennis* com inversão pericêntrica no neo-X. Entre as duas condições extremas (origem recente e antiga), casos de neo-XY com uma mistura das duas condições foram também observados (MESA e de MESA, 1967; DÍAZ e SÁEZ, 1968; WHITE, 1973; CARDOSO e DUTRA, 1979; CASTILLO et al., 2010a, 2010b, 2014; BIDAU et al., 2011). Um neo-XY com uma origem extraordinariamente complexa foi recentemente documentado em *R. bergii*. Além da fusão cêntrica X-A que deu origem ao neo-XY o cromossomo Y sofreu uma inversão pericêntrica envolvendo mais do 90% de sua extensão (PALACIOS-GIMENEZ et al., 2015). Em seguida, o neo-Y sofreu múltiplas inversões paracêntricas e acúmulo de DNA repetitivo. Isto resultou em um neo-XY altamente divergente assim como cinco bem definidas variantes citológicas do neo-Y (PALACIOS-GIMENEZ et al., 2017).

Como mencionado anteriormente, em Orthoptera também são observados sistemas múltiplos, como por exemplo neo- $X_1X_2Y^{\text{♂}}/X_1X_1X_2X_2^{\text{♀}}$ o neo- $X_1X_20^{\text{♂}}/X_1X_1X_2X_2^{\text{♀}}$. No primeiro caso, sistemas múltiplos provavelmente surgem a partir de um ancestral neo-XY, mediante uma segunda fusão cêntrica entre o neo-Y e um autossomo (WHITE, 1973; HEWITT, 1979; CASTILLO et al., 2010b), uma vez que a probabilidade de dois eventos mutacionais independentes é extremamente baixa. Assim, os braços do metacêntrico neo- X_1 são chamados como descrito anteriormente para o neo-XY; enquanto que o neo-Y agora metacêntrico está formado pelos braços YL e YR, os quais compartilham homologia com os braços XR e o neo- X_2 , respectivamente. Estes sistemas foram observados, por exemplo, em quatro espécies de *Dichromatos* (PALACIOS-GIMENEZ et al., 2013), entre outras espécies (CASTILLO et al., 2010b). No segundo caso (neo- $X_1X_20^{\text{♂}}/X_1X_1X_2X_2^{\text{♀}}$), o sistema múltiplo é formado pela fissão cêntrica do original cromossoma X metacêntrico. Este sistema foi observado unicamente até agora em duas espécies de grilos, *Cycloptiloides americanus* (MESA et al., 2002) e *Endocous ubajarensis* (ZEFA et al., 2014).

Comparado com outras regiões do mundo, os ortópteros Neotropais são ricos em espécies com mecanismos sexuais derivados, e as modificações morfológicas que eles sofrem foram discutidas por vários autores (ver por exemplo, WHITE, 1973; HEWITT, 1979; BIDAU e MARTÍ, 2001; MESA e GARCÍA-NOVO, 2001; MESA et al., 2001, 2002; FERREIRA e CELLA 2006; CASTILLO et al., 2010a, 2010b, 2014, 2016; PALACIOS-GIMENEZ et al., 2013, 2015, 2017).

Porém, diferente de outros organismos como humanos (SKALESTKY et al., 2003; CARVALHO e CLARK, 2013), *Drosophila* (CARVALHO, 2002; STEINEMANN e STEINEMANN, 2005) e plantas (NICOLAS et al., 2003; NAVAJAZ-PÉREZ et al., 2005; MATSUNAGA, 2009; KEJNOVSKY et al., 2009), as mudanças moleculares que levam a diferenciação molecular do neo-Y de Orthoptera são ainda pouco compreendidas (ver por exemplo PALACIOS-GIMENEZ et al., 2013, 2015, 2017). Isto se deve pelas análises iniciais em relação aos cromossomos sexuais de Orthoptera serem realizadas somente do ponto de vista citogenético clássico (ver por exemplo WHITE, 1973; HEWITT, 1979; BIDAU e MARTÍ, 2001; MESA e GARCÍA-NOVO, 2001; MESA et al., 2001, 2002; FERREIRA e CELLA 2006; CASTILLO et al., 2010a, 2010b, 2014, 2016).

1.7. Evolução de genes enviesados do sexo

Machos e fêmeas de diferentes espécies de animais e plantas frequentemente diferem fenotipicamente enquanto compartilham o mesmo genoma. Muitas destas diferenças podem ser atribuídas a genes enviesados do sexo (SBG, sex-biased genes), quer dizer, genes expressos diferencialmente entre machos e fêmeas (ELLEGREN e PARSCH, 2007; PARSCH e ELLEGREN, 2013). Em espécies com GSD (vários insetos incluindo Orthoptera) o sexo dos indivíduos é determinado pela proporção X:A, ou seja, pelo número de cromossomos X relativo ao número de autossomos (SÁCHEZ, 2008; KAISER e BACHTROG, 2010). Neste contexto, a perda do cromossomo Y em Orthoptera representaria um problema para o sexo heterogamético (machos X0) pelo simples fato de que a dosagem gênica está correlacionada com o número de cópias gênicas. A hemicigose do X em machos pode causar desequilíbrios em redes gênicas que envolvem tanto genes ligados ao sexo como genes ligados aos autossomos (JEONG et al., 2001; MANK, 2009).

Estudos genômicos recentes mostraram que genes expressos diferencialmente entre machos e fêmeas (SBG) frequentemente apresentam uma distribuição genômica não aleatória, particularmente no que se refere aos autossomos e ao cromossomo X (ELLEGREN e PARSCH, 2007; PARSCH e ELLEGREN, 2013). Por exemplo, estudos transcricionais em *D. melanogaster* revelaram que há uma quantidade escassa e, porém, significativa de genes com expressão enviesada no macho (MBG, *male-biased genes*), enquanto que há um leve excesso de genes com expressão enviesada na fêmea (FBG, *female-biased genes*) no cromossomo X em relação aos

autossomos (PARISI et al., 2003; RANZ et al., 2003). O fenômeno foi chamado de “desmasculinização” e “feminização” dos cromossomos X, respectivamente. No entanto, estudos mais recentes em várias espécies de *Drosophila* mostraram que a desmasculinização do X não é um fenômeno universal, não existindo assim, uma escassez geral de MBG ligados ao X quando são comparados diferentes tecidos (MEISEL et al., 2012). Por exemplo, os autores citados encontraram um leve excesso de MBG na cabeça de *D. melanogaster* e *D. mojavensis*, enquanto que outros autores reportaram um enriquecimento significativo de MBG ligados ao X na cabeça e cérebro de *D. melanogaster* (CHANG et al., 2011; CATALÁN et al., 2012; HUYLMANS e PARSH, 2015), suportando a não universalidade da desmasculinização do X.

Várias hipóteses foram propostas para explicar as diferenças observadas no conteúdo gênico de SBG entre o cromossomo X e autossomos. Uma possível explicação resulta de antagonismo sexual, situação na qual há conflito entre os sexos em relação ao nível ótimo de expressão gênica (ELLEGREN e PARSCH, 2007; PARSCH e ELLEGREN, 2013). Sob este cenário, um alelo influente na expressão de um gene antagonista sexual pode diferir dependendo da dominância e posição, sendo assim, ligado a um cromossomo X ou a autossomos (RICE, 1984; CHARLESWORTH, 1987; ELLEGREN e PARSCH, 2007; PARSCH e ELLEGREN, 2013). Teoricamente, o cromossomo X é um *hotspot* para alelos antagonistas sexuais que favorecem o acúmulo de alelos recessivos benéficos para o macho e alelos dominantes benéficos para fêmeas (ELLEGREN e PARSCH, 2007). Relatos sobre genes com expressão antagonista sexual enriquecidos no cromossomo X foram documentados em *D. melanogaster* (INOCENTI e MORROW, 2010). Entretanto, ainda não está clara a relação entre a expressão enviesada do sexo e o antagonismo sexual (INOCENTI e MORROW, 2010; PARSCH e ELLEGREN, 2013). Os tipos de alelos antagonistas sexuais que esperamos que se acumulem no cromossomo X dependem de outros parâmetros cruciais, como o grau de dominância e magnitude dos seus efeitos sobre o *fitness* entre os dois sexos, que são em parte ainda desconhecidos (FRY, 2010).

Outro fator que pode influenciar a expressão de SBG são as duplicações gênicas. Neste contexto, a expressão do gene ancestral pode permanecer inalterada enquanto a nova cópia pode ter um papel crucial na expressão enviesada do sexo ou inclusive na expressão sexo específica, regulando uma cascata de genes *downstream*. Outro fator importante que poderia influenciar a distribuição de SBG resulta das diferenças em conteúdo gênico entre o cromossomo X e autossomos. Genes de origem recente, e restritos para um táxon, incluindo retrogenes e novos

genes, são enriquecidos no cromossomo X, e esses genes usualmente podem se expressar de modo enviesado em um sexo (BETRÁN et al., 2002; LEVINE et al., 2006; ZHANG et al., 2010; PALMIERI et al., 2014). O cromossomo X também é pobre em genes com expressão tecido-específica, como por exemplo gônadas (MIKHAYLOVA e NURMINSKY, 2011; MEISEL et al., 2012; VIBRANOVSKI et al., 2012; HUYLEMANS e PARSH, 2015). Particularmente, MBG expressos em testículos e nas glândulas acessórias do macho de *D. melanogaster* tendem a ter uma expressão altamente específica de tecido (ANDRÉS et al., 2006, 2008, 2013; MIKHAYLOVA e NURMINSKY, 2011; MEISEL et al., 2012; HUYLEMANS e PARSH, 2015) o que explicaria a escassez de genes no cromossomo X e por sua vez suportando a desmasculinização do X em amostras que incluem tecidos reprodutivos.

A distribuição de SBG no cromossomo X e nos autossomos poderia ser influenciada por mecanismos regulatórios específicos do X. Por exemplo, o X de *Drosophila* é transcricionalmente silenciado na linha germinal masculina (meioses) por um processo análogo à inativação cromossômica sexual meiótica (MSCI, *meiotic sex chromosome inactivation*) que ocorre em mamíferos (LIFSCHYTZ e LINDSLEY 1972; BETRÁN et al., 2002; GRAVES, 2008; VIBRANOVSKI et al., 2009). Embora exista o debate sobre o alcance de MSCI em *Drosophila* e, ainda, se tal mecanismo está ou não limitado à meiose (MEIKLEJOHN et al., 2011; MIKHAYLOVA e NURMINSKY, 2011, 2012; VIBRANOVSKI et al., 2012), as evidências experimentais mostraram que a expressão de genes repórter específicos do testículo é amplamente suprimida quando eles são ligados ao X (HENSE et al., 2007; KEMKEMER et al., 2011, 2014; MEIKLEJOHN et al., 2011). Isso indicaria que existe um mecanismo ainda desconhecido e, possivelmente, distinto da MSCI que limitaria a expressão de genes ligados ao X no testículo (VIBRANOVSKI, 2014).

Outro fator que pode influenciar a expressão de SBG é a compensação de dose (DC, *dose compensation*), que é um mecanismo para equalizar os níveis de expressão de genes localizados em cromossomos sexuais de machos e fêmeas. DC tem sido amplamente estudado em *D. melanogaster*, e foi demonstrado que ocorre por *up-regulation* de genes no cromossomo X do macho mediante recrutamento de um complexo específico de RNA-proteína, denominado *male-specific lethal* (MSL) (KAISER e BACHTROG, 2010; GALLACH e BETRÁN, 2016). Neste mecanismo, a expressão de SBG pode resultar de uma expressão excessiva de genes próximos aos sítios de união ao complexo DC (PARSCH e ELLEGREN, 2013; GALLACH e BETRÁN 2016).

Finalmente, genes distribuídos em cromossomos limitados ao sexo, genes localizados nos cromossomos Y ou W, com efeitos benéficos para o sexo heterogamético, mas prejudiciais para o sexo homogêneo, poderiam fornecer outra explicação para a expressão de SBG (PARSCH e ELLEGREN, 2013). Em *fruitfly*, genes localizados no Y podem ter efeitos secundários na expressão de genes localizados nos autossomos e no X, causando expressão enviesada num sexo. Por exemplo, a variação de haplótipos no Y de *fruitfly* está correlacionada com os níveis de expressão de MBG localizados em outra parte do genoma (LEMOS et al., 2008 2010). Foi demonstrado que a variação na expressão causada pelo cromossomo Y pode ter importantes efeitos fenotípicos que podem ser relevantes para adaptação da espécie (LEMOS et al., 2008, 2010). Outro exemplo é o gene SRY, que codifica para um fator determinante testicular (TDF), também conhecido como gene SRY determinante do sexo em mamíferos, em que apenas a expressão deste gene demonstra como uma cascata de SBG pode ser desencadeada por um único gene limitado ao sexo (KASHIMADA e KOOPMAN, 2010).

1.8. Cromossomos sexuais em grilos: *Gryllus assimilis*, *Cycloptiloides americanus* e *Eneoptera surinamensis* como modelos evolutivos para estudos de cromossomos sexuais em Orthoptera

Comparado com gafanhotos, os grilos vem sendo pouco estudados citogeneticamente (WHITE, 1973; HEWITT, 1979). Por outro lado, os dados existentes revelaram maior variação cromossômica em grilos do que em gafanhotos (HEWITT, 1979). Neste grupo, os mecanismos cromossômicos de determinação sexual são aparentemente mais complexos, com machos de algumas espécies apresentando cromossomos sexuais em distintos estados de diferenciação, tais como X0 (ZEFA, 1999), neo-XY (MESA e GARCÍA-NOVO, 2001), neo-X₁X₂0 (MESA et al., 2002), neo-X₁X₂Y (FERREIRA e CELLA 2006) e até mesmo o extraordinário neo-X₁X₂X₃Y (MESA, dados não publicados). Bem como gafanhotos, os autores acima citados têm proposto que em grilos a origem de neo-cromossomos sexuais estão relacionados com fusões cêntricas, fusões em tandem, translocações recíprocas, fissões e inversões envolvendo autossomos e cromossomos sexuais.

Os gêneros *Gryllus* (Gryllidae), *Eneoptera* (Gryllidae) e *Cycloptiloides* (Mogoplistidae) são representados por 94, sete e 14 espécies, respetivamente (<http://orthoptera.speciesfile.org/Common/basic/Taxa.aspx?TaxonNameID=1122353>). Do ponto de vista cromossômico o gênero *Gryllus* se caracteriza pela presença de 2n=29-X0♂, enquanto

que as únicas espécies estudadas de *Eneoptera* (*E. surinamensis*) e *Cycloptiloides* (*C. americanus*) possuem $2n=9$, neo- X_1X_2Y ♂ e $2n=14$, X_1X_20 ♂, respectivamente (MESA et al., 2002, FERREIRA e CELLA, 2006). O sistema $X0$ observado no gênero *Gryllus*, assim como em gafanhotos, é considerado modal para os grilos, diferindo apenas na morfologia, pois neste último grupo é metacêntrica. O sistema sexual do grilo *C. americanus* (X_1X_20) possui o X_1 metacêntrico e o X_2 acrocêntrico e é incomum em Orthoptera, sendo descrito pela primeira vez nesta espécie (MESA et al., 2002). Em *E. surinamensis* o complexo sistema sexual neo- X_1X_2Y apresenta o X_1 e o Y com dois braços enquanto o X_2 é acrocêntrico, sendo o cromossomo Y o maior elemento do cariótipo. Esse sistema de cromossomos sexuais possivelmente se originou a partir de fusões cêntricas, além do possível envolvimento de translocações (FERREIRA e CELLA, 2006).

Entre os três sistemas de cromossomos sexuais relatados, apenas em dois a localização da heterocromatina constitutiva foi estudada. A mesma apresentou-se com distribuição distinta, com blocos restritos ao centrômero do X em *Gryllus*; enquanto em *E. surinamensis* o Y apresenta múltiplos blocos de heterocromatina, o X_1 apresenta blocos restritos ao centrômero e o X_2 blocos centroméricos e intersticiais. Análises mais específicas relativas a estrutura e evolução de sistemas cromossômicos de determinação do sexo ainda não foram realizadas, e até o momento, nada foi estudado quanto a composição molecular específica destes cromossomos, que apresentam aparentemente distintos mecanismos de diversificação nas diversas espécies. Por outro lado, devido esta ampla variabilidade de sistemas cromossômicos de determinação sexual os grilos podem ser modelos interessantes para estudos evolutivos deste tipo cromossômico em animais.

Baseado nas informações obtidas durante a presente revisão bibliográfica surgem algumas questões evolutivas fundamentais em relação a evolução genômica e de cromossomos sexuais em Orthoptera, que ainda não foram ou foram pobremente exploradas e testadas. Por exemplo, como e por quê os cromossomos sexuais de Orthoptera evoluem? Qual é a composição de DNA repetitivo nos cromossomos sexuais? Há relação entre o DNA repetitivo e o aumento de tamanho dos cromossomas sexuais? Há transcrição diferencial de DNA repetitivo entre sexos e espécies? Tem o DNA repetitivo efeitos na regulação gênica dos cromossomos sexuais? Quais são os genes determinantes do sexo e quais genes estão envolvidos na diferenciação do fenótipo sexual? Estes genes evoluem mais rapidamente e diferencialmente em relação aos genes não envolvidos com a diferenciação sexual? Como e por que os genes enviesados do sexo evoluem? Na tentativa de explorar tais ideias, nessa tese de doutorado há uma combinação de metodologias genômicas e

citogenômicas em uma maneira comparativa. Nos capítulos (ver resultados) da presente tese há descrições, informações (estrutural e funcional) e hipóteses enfatizando aspectos que abordam estes tópicos que são pobremente explorados em Orthoptera.

2. OBJETIVOS

2.1. Objetivos gerais

O principal objetivo desta tese foi investigar a estrutura, composição molecular e processos evolutivos envolvidos na diferenciação de cromossomos sexuais em Orthoptera mediante uma abordagem citogenômica e genômica comparativa, utilizando como modelos espécies de grilos com distintos sistemas sexuais, *Grillus assimilis* (X0), *Ciclotyloides americanus* (X₁X₂0) e *Eneoptera surinamensis* (neo-X₁X₂Y). Outro objetivo desta tese foi investigar mediante uma abordagem genômica funcional abrangente os genes codificadores de proteínas envolvidos com a determinação sexual, genes envolvidos com o *fitness* reprodutivo bem como a prospecção de genes enviesados do sexo, os quais certamente são responsáveis pelas diferenças fenotípicas entre sexos. As possíveis forças evolutivas atuando nestes loci foram também exploradas. A era da genômica comparativa de Orthoptera começou recentemente com a liberação de sequências de genomas de algumas poucas espécies. Portanto, a qualidade e acessibilidade das sequências genômicas de Orthoptera é bastante inferior a muitas das classes de insetos. Genomas de Orthoptera apresentam características peculiares como tamanho gigante e abundância em DNA repetitivo, que certamente representam um desafio para montagens de genomas. Portanto, eu foquei primeiro no mapeamento físico (análises estruturais) das famílias de DNA repetitivo mais abundantes dentro deste grupo de insetos, isolados principalmente mediante técnicas clássicas de biologia celular e molecular e mediante *reads* de genomas sequenciados. A disponibilidade de novas técnicas de sequenciamento genômico têm aumentado rapidamente, o que abriu uma porta para as análises genômicas comparativas refinadas em Orthoptera. A comparação dos resultados em genomas de grilos com achados prévios em genomas de insetos ajudou a responder várias questões no campo da genômica comparativa. Finalmente, além dos estudos genômicos comparativos estruturais aplicados aos dados, é necessário uma sólida compreensão e melhoria dos princípios evolutivos que atuam na diferenciação de cromossomo sexuais e na diferenciação do fenótipo sexual entre machos e fêmeas.

2.2. Objetivos específicos

2.2.1. Estrutural

- Entender o possível papel dos DNAs repetitivos na estrutura/diversificação dos cromossomos sexuais a partir do isolamento e mapeamento físico de sequências, tais como, famílias multigênicas, DNAsat e microssatélites nas espécies de *Gryllus assimilis*, *Ciclotyloides americanus* e *Eneoptera surinamensis*.

2.2.2. Funcional

- Testar e comparar expressão diferencial de DNAsat entre diferentes tecidos, sexos e espécies a partir de transcriptomas de *Gryllus assimilis*, *G. bimaculatus*, *G. firmus* e *G. rubens*, com o objetivo de entender os possíveis papéis funcionais destas sequências na regulação gênica, modulação da cromatina e como componentes funcionais de importantes estruturas como telômeros, centrômeros e cromossomos sexuais;
- A partir de transcriptomas de espécies de grilos (*Gryllus assimilis*, *G. bimaculatus* e *G. firmus*) prospectar genes codificadores de proteínas relacionados com a determinação sexual, envolvidos com o *fitness* reprodutivo, assim com genes enviesados do sexo responsáveis pelas diferenças fenotípicas entre machos e fêmeas, e tentar elucidar de uma maneira comparativa os fatores evolutivos atuando nestes loci.

3. DESENHO EXPERIMENTAL

3.1. Amostras, análises de cromossomos e bandeamento cromossômicos clássicos

Machos e fêmeas de *Eneoptera surinamensis* foram coletados no Parque Estadual Edmundo Navarro de Andrade (Rio Claro, SP, Brasil) entre maio de 2013 e março de 2014 com a autorização da COTEC (processo número 341/2013). Os indivíduos foram mantidos em cativeiro até que a ocorrência da oviposição. As espécies *Gryllus assimilis* e *Cycloptiloides americanus* foram criados no biotério da Universidade Estadual Paulista - UNESP (Rio Claro, SP).

Em cada uma das espécies, os cromossomos mitóticos foram obtidos a partir de neuroblastos embrionários seguindo o protocolo descrito por Webb et al. (1978), com pequenas modificações na fixação dos materiais, da seguinte forma: após a realização da hipotonização, se aplicou o mesmo volume de solução de Carnoy modificado (3:1, etanol absoluto:ácido acético) durante 15 minutos; em seguida os embriões foram transferidos para a solução fresca de Carnoy e armazenado a uma temperatura de -20 °C até o uso. Adicionalmente, testículos de machos adultos de cada espécie foram dissecados e fixados na solução de Carnoy para posterior preparação e visualização de cromossomos meióticos pelo método de esmagamento (squash). Espécimes de adultos de cada espécie foram armazenados em etanol a 100% para posterior extração de DNA baseado no procedimento de Fenol:Clorofórmio descrito em Sambrook e Russel (2001).

Estudos de cariotípicos gerais em cada espécie foram realizados usando coloração convencional com Giemsa 5%. O procedimento de banda C foi realizado de acordo com Sumner (1972). A coloração com fluorocromos base-específicos (CMA₃/DA/DAPI) para identificar regiões ricas em G+C ou A+T seguiu o protocolo descrito por Schweizer et al. (1983).

3.2. Estimativa de tamanho gnômico

Análises de citometria de fluxo foram realizadas conforme descrito por Lopes et al. (2009), no Laboratório de Citogenética e Citometria, Departamento de Biologia Geral, Universidade Federal de Viçosa - UFV (Minas Gerais). O conteúdo de DNA nuclear de fêmeas e machos adultos das espécies *G. assimilis* e *E. surinamensis* foram determinados usando o conteúdo de DNA C (valor C) da fêmea *Scaptotrigona xantotricha* como referência interna, o que foi confirmado em comparação com o padrão internacional de *Drosophila melanogaster*.

Para preparar as suspensões nucleares para citometria de fluxo, os gânglios cerebrais foram

removidos em solução fisiológica salina (0,155 mM NaCl). As amostras consistiram em três machos e três fêmeas de cada espécie e cada indivíduo foi manipulado e analisado separadamente, representando três repetições independentes. Os materiais foram simultaneamente triturados 10-12 vezes em um moedor de tecido usando um pilão (Kontes Glass Company, NJ, EUA) em 100 µL de tampão de lise OTTO-I (OTTO, 1990) contendo ácido cítrico 0,1 M (Merck, NJ, USA), 0,5% Tween 20 (Merck) e 50 µg mL⁻¹ RNase A (Sigma-Aldrich), pH = 2,3. As suspensões foram ajustadas para 1 mL usando o mesmo tampão, filtradas através de uma malha de nylon de 30 µm (Partec, Nuremberg, Germany) e centrifugadas a 100 g em tubos de microcentrífuga por 5 min. Em seguida, os pellets foram incubados durante 10-15 min em 100 µL de tampão de lise OTTO-I e logo corados usando 1,5 mL de solução OTTO-I:OTTO-II (1:2) (LOUREIRO et al., 2006a, 2006b) suplementado com iodeto de propídio 75 uM (PI) (comprimentos de onda de excitação/emissão: 480-575/550-740 nm, SHAPIRO, 2003) e 50 µg de mL⁻¹ RNase A (Sigma-Aldrich), pH = 7,8. As suspensões nucleares foram filtradas através de um filtro de malha de nylon de 20 µm (Partec) e foram mantidas no escuro durante 30 min.

As suspensões foram analisadas usando um citometro de fluxo Partec PAS (Partec) equipado com uma fonte laser (488 nm). A fluorescência de PI emitida pelos núcleos foi coletada usando um filtro de passagem de banda RG 610 nm e convertida em 1024 canais. O equipamento foi calibrado para linearidade e alinhado usando microesferas e soluções standard de acordo com as recomendações do fabricante. O software FlowMax (Partec) foi utilizado para análise de dados. O pico nuclear standard foi estabelecido para o canal 100 e mais de 10.000 núcleos foram analisados. Se realizaram três replicas independentes, e os histogramas com um coeficiente de variação (CV) superior a 5% foram rejeitados.

3.3. Isolamento de DNAs repetitivos por de PCR

A sequência parcial de rRNA 5S e do gene de histona H3 foram obtidas diretamente por PCR usando o DNA genômico de *Abracris flavolineata* e primers descritos por Loreto et al. (2008) e Cabral de Mello et al. (2010) para DNAr 5S, e Colgam et al. (1998) para histona H3. As sequências para DNAsn U1 e U2 foram obtidas usando o DNA genômico de *Rhammatocerus brasiliensis* com os primers descritos por Cabral de Mello et al. (2012) e Bueno et al. (2013). Estes fragmentos foram sequenciados para confirmação da obtenção das sequências de interesse, números de acesso ao GeneBank KC896792 (histona H3), KC936996 (rDNA 5S), KC896793

(snDNA U1) e KC896794 (snDNA U2).

A sequência de DNAr 18S foi obtida a partir de um fragmento previamente clonado e isolado do genoma de *Dichotomius semisquamosus* (número de acesso do GeneBank GQ443313, CABRAL-DE-MELLO et al., 2010). A sonda telomérica foi obtida por PCR usando os primers (TTAGG)₅ e (CCTAA)₅ como descrito por Ijdo et al. (1991).

3.4. Isolamento de DNA alta e moderadamente repetitivo: fração de DNA C₀t-1

Amostras enriquecidas com DNAs repetitivo de cada uma das espécies foram obtidas baseadas na cinética de renaturação da fração de DNA C₀t-1 (DNA enriquecido com sequências alta e moderadamente repetitivas) de acordo com o protocolo descrito por Zwick et al. (1997), com modificações. Brevemente, amostras de DNA (200 µL de 100-500-ng/µL de DNA genômico em 0,3 M de NaCl) foram digeridas com deoxyribonuclease I (Sigma, St Louis, MO, USA) a 0,01 U/µL durante 80 a 105 segundos, dependendo da concentração da amostra. Em seguida, os fragmentos de DNA foram separados em gel de eletroforese em 1% de agarose. Os fragmentos de DNA esperados variaram em tamanho de 100 a 1.000 bp. Para cada uma das espécies, 50 µL da amostra de DNA fragmentado foram desnaturadas a 95°C por 10 min, colocadas em gelo por 10 segundos e transferidas a um termociclador a 65°C para reanelamento por 25 min. Subsequentemente, as amostras foram incubadas a 37°C por 8 min com 1 U de nuclease S1 para digerir o DNA de cadeia única. Os fragmentos foram purificados e extraídos usando o protocolo tradicional de Fenol:Clorofórmio (SAMBROOK e RUSSEL, 2001).

3.5. Sequenciamento de Illumina e clusterização baseado em gráfico de reads sequenciados

Para sequenciamento Illumina somente *G. assimillis* e *E. surinamensis* foram utilizados. Estas espécies foram escolhidas levando em consideração que a primeira (*G. assimillis*) apresenta o sistema cromossômico de determinação sexual basal (X0/XX) para Orthoptera (WHITE, 1973; HEWITT, 1979), e a segunda (*E. surinamensis*) apresenta uns dos cariótipos mais derivados (2n♂=9, neo-X₁X₂Y) dentro de Orthoptera (HEWITT, 1979; FERREIRA e CELLA, 2006).

Para *G. assimillis*, o sequenciamento paired-ends (2x101) foi realizado em bibliotecas construídas como recomendado por Illumina usando Illumina TruSeq DNA PCR-Free kit (Illumina Inc., San Diego, CA, EUA) a partir do DNA genômico de uma fêmea. O sequenciamento dos fragmentos da biblioteca foram realizados mediante a plataforma HiSeq 2000 utilizando os

serviços da Macrogen (Macrogen Inc., Coréia do Sul). Para *E. surinamensis*, o sequenciamento paired-ends (2x300) foi aplicado para bibliotecas preparadas como recomendado por Illumina usando Nextera DNA Library Preparation Kit v3 a partir de um DNA genômico de macho. Os fragmentos de bibliotecas estavam no intervalo de 400 a 600 pb e o sequenciamento foi realizado usando o sistema de sequenciamento MiSeq.

Os *reads* sequenciados (DNA-seq) das duas espécies foram pré-processados para checar a qualidade dos *reads* com FASTQC (ANDREWS, 2012) e filtrados por qualidade usando FASTX-Toolkit (GORDON e HANNON, 2010). Os *reads paired-ends* foram unidos com o software “fastq-join” do FASTX-Toolkit (GORDON e HANNON, 2010) usando as opções padrão. Clusterização baseada em grafos e montagem das sequências foram realizados usando o software RepeatExplorer (NOVAK et al., 2010, 2013) com objetivo de prospectar DNAsat nos genomas de *G. assimilis* e *E. surinamensis*. Para identificar DNAsat foram examinados os clusters que exibiram gráficos circulares no output de RepeatExplorer que é uma característica típica das famílias de satDNA nesta metodologia (GORDON e HANNON, 2010). Para confirmar a organização em tandem de satélites e precisa identificação de DNAsat analisou-se gráficos de *dotplot* implementados em Dotlet (JUNIER e PAGNI, 2000). Todos os clusters contendo *reads* com sequências representadas acima de 0,01% da proporção genômica (sequências de alto números de cópias) foram analisados em detalhes.

3.6. Isolamento e análise de sequências de DNAsat

Clusters com alta densidade de gráfico circulares foram analisados utilizando o algoritmo TRF (*Tandem Repeat Finder*) (BENSON, 1999) para identificar a sequência de DNA que maximizou os escores de alinhamento entre os diferentes monômeros que poderiam ser definidos em tandem. Todos os clusters foram processados com TRF usando parâmetros de alinhamento 2, 3, 5 para *match*, *mismatch* e *indels*, respectivamente, e um escore de alinhamento mínimo de 50. Além disso, utilizou-se a ferramenta de alinhamento gráfico *dotplot* implementada em Dotlet (JUNIER e PAGNI, 2000) para identificar o início e o fim exatos dos monômeros da mesma família e para confirmar a organização em tandem. O monômero com o comprimento máximo foi usado como a cópia representativa para cada família de DNAsat, e também como sequências *query* em BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) e Repbase (<http://www.girinst.org/censor/index.php>) para busca e verificação de similaridade com

sequências publicadas. Alinhamentos de sequências (cópias) de DNAsat foram realizados usando Muscle (EDGAR, 2004) implementado em MEGA5 (TAMURA et al., 2011). O MEGA5 também foi usado para estimar a divergência nucleotídica (distância p), conteúdo de A+T e para realizar as análises de longitude das repetições. Relações evolutivas entre sequências foram inferidas usando a metodologia de clusterização Neighbor-Joining (NJ) e a proporção de diferenças nucleotídicas (distância p) em MEGA5. Para prever a estrutura secundária de satDNAs foi utilizado o software CentroidFold (HAMADA et al., 2009) com inferência McCaskill e parâmetro de peso 2^{-5} para pareamento de bases como opção.

As sequências montadas (consensos) de cada família de DNAsat foram usadas para fabricar primers com direções opostas, usando o software Primer3 (ROZEN e SKALETSKY, 1999) ou manualmente. A presença de famílias de DNAsat foi verificada mediante reações em cadeia da polimerase (PCR). PCRs foram realizadas usando 10×PCR Rxn Buffer, 0,2 mM de MgCl₂, 0,16 mM de dNTPs, 2mM de cada primer, 1 U de Taq Platinum DNA Polymerase (Invitrogen) e 50-100 ng/μl de DNA molde. As condições de PCR incluíram uma desnaturação inicial a 94°C durante 5 min e 30 ciclos a 94°C (30 s), 55°C (30 s) e 72°C (80 s), mais uma extensão final a 72°C por 5 minutos. Os produtos de PCR foram checados em um gel de agarose a 1%. As bandas monoméricas foram isoladas e purificadas usando Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corp., The Epigenetics Company, EUA) de acordo com as recomendações do fabricante, e depois utilizadas como fonte dereamplificação para posterior análise. Para verificar o isolamento de sequências de interesse, os produtos de PCR purificados foram sequenciados (método Sanger) em ambas as direções utilizando o serviço de MacroGen Inc. e, em seguida, comparados com as sequências consenso obtidas por análise genômica.

3.7. Hibridização *in situ* fluorescente (FISH)

As sequências de DNAs repetitivos de DNAr 5S, DNAsn U (U1 e U2) e repetição telomérica foram marcadas com Digoxigenina 11-dUTP (Roche, Mannheim, Germany) através da PCR. O plasmídeo contendo o gene de DNAr 18S, os produtos de PCR do gene de histona H3 e a fração de DNA *C₀t-1* foram marcados com Biotina 14-dATP (Invitrogen) através da Nick translation, seguindo recomendações do fabricante. Os produtos de PCR de cada família de DNAsat foram marcados por Nick-translation usando Biotina 14-dATP (Invitrogen) ou Digoxigenina 11-dUTP (Roche). DNAsat menores que 60 bp foram marcados diretamente com

Biotina 14-dATP (Invitrogen) na própria síntese no extremo 5'. Esta mesma metodologia de marcação foi também aplicada para os seguintes microssatélites (A)₃₀, (C)₃₀, (CA)₁₅, (CG)₁₅, (TA)₁₅, (AG)₁₀, (CAA)₁₀, (CAC)₁₀, (TAA)₁₀, (GAA)₁₀, (CGG)₁₀, (GAC)₁₀, (CAT)₁₀, (GAG)₁₀, (GACA)₄ e (GATA)₈ (Sigma).

A FISH foi basicamente realizada usando o protocolo descrito por Pinkel et al. (1986), com modificações sugeridas por Cabral-de-Mello et al. (2010), usando preparações mitóticas. Experimentos de fiber-FISH foram efetuados seguindo o protocolo descrito em de Barros et al. (2011), usando suspensões de células do testículo. As sondas foram detectadas usando anti-digoxigenin-Rhodamine (Roche) ou streptavidina alexa fluor-488 (Invitrogen). Todas as preparações foram contracoradas com DAPI e montadas com meio de montagem Vectashield (Vector, Burlingame, CA, USA). As fotografias e sinais foram observadas utilizando um microscópio Olympus BX61 equipado com uma lâmpada fluorescente e filtros adequados. As imagens foram fotografadas usando uma câmera digital DP70 em escala de cinza. Adicionalmente, as imagens foram pseudo-coradas em azul ou cinza (cromossomos), vermelho ou verde (sinais), sobrepostas e otimizadas em brilho e contraste usando Adobe Photoshop CS2.

3.8. Análise quantitativa dos DNAsat em *E. surinamensis*

A PCR quantitativa (qPCR) usando DNA genômico de macho e fêmea como molde foi utilizada para verificar as diferenças de número de cópias entre machos e fêmeas de DNAsat selecionados, por exemplo, algumas das famílias repetitivas com tamanho superior a 60 pb. Os DNAsat selecionados foram escolhidos devido à sua presença no cromossomo neo-Y determinado por FISH. A qPCR do DNA genômico de macho e fêmea foi usada para calcular a dose de satDNA por um método ΔC_t de quantificação relativa (NGUYEN et al., 2013). As proporções de dosagem de genes (GDR) de DNAsat alvo foram comparadas com um gene de referência, a família de proteína de choque térmico de 70 kDa (*Hsp-70*, heat shock protein) usando os primers F 5'-GGTGGTATGACCACTCTTATCAA-3' e R 5'-CACTTCATTTTGAGGCACACC-3' que foram desenhados de acordo com o gene *Hsp-70* de *Locusta migratoria* (acesso AY178988). Este gene foi usado como referência porque não houve diferenças detectadas nas taxas de ampliação entre os sexos em *E. surinamensis*, sugerindo que o gene *Hsp-70* é autossômico e tem o mesmo número de cópias em ambos os sexos. Porque não tínhamos certeza de que *Hsp-70* é um gene de cópia única, a análise quantitativa foi uma comparação relativa da dose de gene entre a macho e fêmea.

Análises quantitativas foram realizadas no MicroAmp® Fast Optical 96-Well Reaction Plate com barcode (0,1 mL) (Applied Biosystems, Life Technology™, Carlsbad, CA) coberta por capas de adesivo óptico (Applied Biosystems) usando o termo ciclador StepOne Real-Time PCR. A qPCR em ambos os satDNAs alvo e o gene de referência foram realizados simultaneamente em triplicados de três amostras independentes, por exemplo, 3 machos e 3 fêmeas. Cada mistura de qPCR continham 6,25 µl 2×GoTaq® qPCR master mix (Promega, Madison, WI, USA), 0,25 mM de cada forward e reverse primers e 30 ng de DNA genômico de macho ou fêmea, num volume final de 10 µl. Misturas de qPCR sem DNA serviram como controles negativos. As condições de ciclagem foram de 95°C durante 10 min, 40 ciclos de 95°C durante 15 s e 60°C durante 1 min. A especificidade dos produtos de PCR foi confirmada pela análise da curva de dissociação. Uma análise de correlação de médias de GDR de DNAsat alvo entre machos e fêmeas foi realizada na versão 3.3.3 do software R (R CORE TEAM, 2017) e editada usando o Adobe Photoshop CS2.

3.9. Análise de correlação estatística de DNAsat em *E. surinamensis*

Foi utilizado o pacote PerformanceAnalytics (PETERSON et al., 2014) implementado na versão 3.3.3 do R (R CORE TEAM, 2017) para calcular os coeficientes de correlação de Spearman para o tamanho de monómero de DNAsat, proporção do genoma, conteúdo de A+T, divergência de nucleotídeos e número de loci. Para os DNAsat mostrando os sinais de FISH espalhados, o número de loci contável total foi aumentado em 10 para explicar o seu perfil disperso, mas ainda de forma conservadora para evitar a contagem de possíveis sinais de hibridação não específicos.

3.10. Tecidos de *G. assimilis* para extração de RNA e sequenciamento

Machos e fêmeas de *Gryllus assimilis* foram obtidos de um grupo de indivíduos que haviam sido criados na Univ. Estadual Paulista - UNESP (Rio Claro, SP, Brasil). O RNA dos machos (cabeça e testículos) e das fêmeas (cabeça e ovário) foram extraídos usando o reagente TRIzol (ThermoFisher Scientific, Invitrogen), seguindo o protocolo standard recomendado pelo fabricante. O grilo *G. assimilis* foi tomado como referência porque foi a única espécie que tivemos replicas biológicas para os diferentes sexos e tecidos gerados em nosso laboratório, e também porque a espécie apresenta o sistema cromossômico de determinação sexual (machos X0) considerado basal para Orthoptera (WHITE, 1973; HEWITT, 1979). Para cada tecido foi reunido uma quantidade equivalente de material de três espécimes por sexo com a mesma idade de

maturidade (isto é, machos sexualmente maduros em estágios de canto e fêmeas adultas virgens) para produzir *pools* representativos de machos e fêmeas. Os *pools* de machos e fêmeas de cada tecido foram realizados em triplicatas biológicas. Bibliotecas de RNA-seq foram preparadas usando TruSeq Quick SBS kit ou Truseq SBS Kit v4 (Illumina Inc., San Diego, CA, EUA). As bibliotecas foram sequenciadas (2x126 pb) em uma plataforma HiSeq 2500 utilizando os serviços Macrogen (Macrogen Inc., Coréia do Sul).

3.11. Transcrição de DNAsat nas espécies de *Gryllus*

Para investigar a transcrição de satDNA em *G. assimilis* foram utilizados os *reads* paired-ends (2x126) Illumina (RNA-seq *reads*) de machos (cabeça e testículo) e fêmeas (cabeça e ovários), pertencentes ao projeto de transcriptomas em preparação em nosso laboratório (dados não publicados). Para esta espécie foram utilizadas três réplicas biológicas para cada tecido. Para as propostas comparativas, eu investiguei se os 13 DNAsat de *G. assimilis* identificadas na minha tese doutorado foram transcritos através dos tecidos disponíveis (*reads* RNA-seq) de *G. bimaculatus*, *G. rubens* e *G. firmus*. Utilizei também as famílias de DNAsat GBH535 e GBH542 isoladas de *G. bimaculatus* (YOSHIMURA et al., 2006) para a mesma abordagem. As sequências consenso para cada DNAsat de *G. assimilis* foram usadas para análise e para *G. bimaculatus*, consensos para famílias de DNAsat foi gerado usando clones depositados no NCBI com os seguintes números de acesso: família GBH535, AB204914-AB204938; família GBH542, AB204939-AB204951.

Bibliotecas de RNA-seq pertencentes a outras espécies de grilos foram descarregadas do NCBI, como segue: bibliotecas de RNA-seq de uma mistura de embriões e ovários (acesso SRX023832), bibliotecas de RNA-seq de embriões (acesso SRX0238310) e bibliotecas de RNA-seq de ovários (acesso SRX023831) de *G. bimaculatus*; bibliotecas de RNA-seq de animais inteiros (acesso SRX1596750, SRX1596749, SRX1596748, SRX1596747, SRX1596746, SRX1596745, SRX1596744, SRX1596743, SRX1596742, SRX1596741, SRX1596740, SRX1596739, SRX1596738, SRX1596737, SRX1596736, SRX1596735, SRX1596734, SRX1596733, SRX1596732, SRX1596731, SRX1596730, SRX1596729, SRX1596728, SRX1596727) de *G. rubens*; bibliotecas de RNA-seq de músculo de vôo de fêmeas com asas longas com músculo de vôo histolizado (LWFHFM, flight muscle from long winged female with histolyzed flight muscle) (acesso SRX272161, SRX272160, SRX272159), bibliotecas de RNA-seq de músculo de vôo de

fêmeas com asas longas com músculo de vôo funcional (LWFFM, flight muscle from long winged female with functional flight muscle) (acesso SRX272158, SRX272157, SRX272156, SRX272155), bibliotecas de RNA-seq de corpo gorduroso de fêmeas de asas curtas incapaz de voar (FBSWFIF, fat body from short winged female incapable of flight) (acesso SRX272155, SRX272154, SRX272153, SRX272152, SRX272151, SRX272150, SRX272127, SRX272125), e bibliotecas de RNA-seq de corpo gorduroso de fêmeas com asas longas com músculos de vôo funcionais (FBLWFFM, fat body from long winged female with functional flight muscles) (acesso SRX272124, SRX272122, SRX272120, SRX272119, SRX272117, SRX272111, SRX272106, SRX272104) de *G. firmus*.

Os *reads* brutos de RNA-seq de cada tecido foram mapeados para cada uma das sequências de DNAsat usando Bowtie2 (LANGMEAD e SALZBERG, 2012) com o parâmetro *--sensitive* como opção. Para repetições menores, como por exemplo Gas1 (11 pb), Gas4 (73 pb), Gas9 (82 pb) e Gas11 (10 pb), o mapeamento foi realizado em dímeros ou vários monômeros foram concatenados até atingir 200 bp de tamanho. Os resultados de mapeamento foram convertidos de formato *sorted* em binário usando SAMtools (LI et al., 2009). Os *reads* alinhados foram contados usando as opções de SAMtools para comparação entre sequências e tecidos com o fim de estimar sua proporção no genoma (ou seja, o valor de expressão de números de *reads* que se alinham a um DNAsat dividido pelo número total de *reads* em cada biblioteca de RNA-seq). O output de Bowtie2 foi utilizado para estimar a abundância relativa desses transcritos usando Cufflinks (TRAPNELL et al., 2010). O passo de quantificação inclui contagens de *reads* brutos e contagens de *reads* escalados. O método de escala aplicado foi FPKM (fragmentos por quilo-base de transcritos por milhão de *reads* mapeadas, o valor de expressão obtido após a normalização das contagens de *reads* pela longitude dos transcritos e pelo número de *reads* mapeadas em cada biblioteca de RNA-seq).

3.12. Montagem e anotação de transcriptomas em espécies de *Gryllus*

Reads de RNA-seq de cada espécie de *Gryllus* foram trimados com Trimmomatic (BOLGER et al., 2014) e em seguida montados *de novo* usando Trinity (HAAS et al., 2013). Os transcritos foram anotados com Trinotate (<https://trinotate.github.io>), uma extensão desenhada para anotação funcional automática de transcriptomas. Este usa várias ferramentas bem referenciadas para anotação funcional, incluindo busca de homologia em bases de dados de

sequências (BLAST+/SwissProt), identificação de domínios de proteínas (HMMER/PFAM), e comparação com os bancos de dados curados atualizados (eggNOG e Gene Ontology terms). Além disso, Transdecoder (<https://transdecoder.github.io>) foi usado para identificar ORF (open Reading frame) com sequências codificadoras completas.

Para mapear as vias KEGG (Kyoto Encyclopedia of Genes and Genomes) e atribuir genes ortólogos (MORIYA et al., 2007) nas espécies de *Gryllus* foi usado o Automatic Annotation Server (KAAS, http://www.genome.jp/kaas-bin/kaas_main?mode5est_b). Designação de ortologia KEGG (KO, KEGG orthology) foram realizados baseados no BHH (bi-directional best hit) em BLAST (KANEHISA e GOTO 2000). Este foi usado para prospectar genes envolvidos em vias da biossíntese de hormônios de insetos (IHB, *insect hormone biosynthesis*), ritmo circadiano (CR, *circadian rhythm*), *entrainment* circadiano (CR, *circadian entrainment*) e ritmo circadiano-fly (CRF, *circadian rhythm-fly*). Além disso, os dados foram comparados com o genoma do gafanhoto modelo *Locusta migratoria*, descarregado do NCBI (acesso GCA_000516895.1). Usando as designações de KO, foi contado e comparado o número de genes únicos encontrados para as vias entre o genoma de *L. migratoria* e os transcriptomas de *Gryllus*. Para examinar duplicações gênicas nas vias acima mencionadas, foi contado o número de genes de cópia única e o número putativo de genes multicópia no genoma de *L. migratoria* e nos transcriptomas de *Gryllus*.

3.13. Expressão gênica diferencial

Para analisar a expressão de genes, os *reads* brutos de RNA-seq de *G. assimilis* foram primeiro trimados usando Trimmomatic (BOLGER et al., 2014), seguido de quantificação de expressão de cada transcrição usando RSEM v1.2.16 (LI e DEWEY, 2011) com Bowtie2 (LANGMEAD e SALZBERG, 2012) como mapeador de *reads*. Os dados de genes expressos diferencialmente (DEG, differentially expressed genes) e genes enviesados do sexo (SBG, sex-biased genes) foram comparados entre as gônadas e entre as cabeças de machos e fêmeas de *G. asssimilis*. Esta espécie foi tomada como referência porque foi a única que tivemos réplicas biológicas (*reads* RNA-seq) para diferentes sexos e tecidos gerados em nosso laboratório. A análise estatística de DEG foi realizada utilizando o DESeq2 (versão 2.1.14.1; LOVE et al., 2014), implementado no pacote Bioconductor (versão 3.0; GENTLEMAN et al., 2004) em R (versão 3.3.3; R CORE TEAM, 2017). Todos os valores de *p* foram ajustados usando o teste de Wald implementado em DESeq2. Para reduzir testes múltiplos, DESeq2 apenas considera genes que

possuem um conjunto mínimo de *reads*. Isso ocorre porque os valores de p ajustados não podem ser calculados para genes com alta variação entre repetições ou baixa expressão média entre bibliotecas (LOVE et al., 2014). O mínimo foi de três *reads* mapeados entre todas as bibliotecas. Um gene foi considerado enviesado se a comparação para o fator condição (amostras) tinha valores de p ajustados $< 0,05$. O grau de viés foi determinado pela diferença de \log_2 *fold-change* (\log_2FC) entre as condições calculadas em DESeq2, ou seja, aqueles genes com $\log_2FC > 0$ e $\log_2FC < 0$ com um valor p ajustado $< 0,05$ foram considerados como *up-regulated* e *down-regulated*, respectivamente.

3.14. Estimativa de taxas evolutivas

Foram avaliadas as restrições seletivas para genes codificadores de proteínas envolvidos na determinação do sexo, IHB, CR, CE e CRF entre as espécies de grilos de campo, *G. assimilis*, *G. bimaculatus* e *G. firmus*. Para atingir esse objetivo, apenas os ortólogos compartilhados entre as três espécies foram considerados. As regiões CDS (*coding DNA sequences*) desses genes que codificam proteínas foram alinhadas com o algoritmo ClustalW (THOMPSON, 1994). Para estimar o número de substituições não sinônimas (Ka) e sinônimas (Ks) foi utilizado o método Nei-Gojobori (Jukes-Cantor) implementado em MEGA7 (KUMAR et al., 2016). As estimativas de seleção (ω) foram calculadas usando a relação Ka/Ks considerando um valor $p < 0,05$ (teste exato de Fisher). Se $\omega = 1$, um modelo neutro de evolução não pode ser rejeitado, enquanto que $\omega < 1$ indica seleção negativa e $\omega > 1$ indica seleção positiva. Para comparar as restrições seletivas entre genes que codificam proteínas não reprodutivas (*housekeeping genes*) e aqueles genes que codificam proteínas sexuais e que influenciam o *fitness* reprodutivo, contrastou-se estes últimos subconjuntos de genes com um grupo de cinco genes *housekeeping*, selecionados para representar diferentes proteínas estruturais e funções metabólicas básicas, por exemplo, transcription factor E2F4 (E2F4), X-ray repair cross-complementing protein 5 (XRCC5), interleukin enhancer-binding factor 2 homolog (ILF2), RNA-binding protein EIF1AD (EIF1A) and ribonuclease H2 subunit B (RNH2B).

4. RESULTADOS E DSCUSSÃO

Os resultados da presente tese são apresentados e discutidos em cinco manuscritos publicados, em revisão ou a serem submetidos em revistas listadas abaixo:

Molecular Genetics and Genomics. ISSN 1617-4623. Fator de impacto 2,979. (Publicado)

PLoS One. ISSN sem. Fator de impacto 2,806. (Publicado)

Scientific Reports. ISSN 2045-2322. Fator de impacto 4,259. (Publicado)

DNA Research. ISSN 1340-2838. Fator de impacto 5,404. (Publicado)

BMC Genomics. ISSN 1471-2164. Fator de impacto 3,729. (Submetido).

4.1. Capítulo 1

*Repetitive DNA chromosomal organization in the cricket *Cycloptiloides americanus*: a case of the unusual $X_1X_2\theta$ sex chromosome system in Orthoptera*

Octavio M. Palacios-Gimenez, Diogo C. Cabral-de-Mello

Publicado na revista *Molecular Genetics and Genomics* (2015), 290: 623-631.

Repetitive DNA chromosomal organization in the cricket *Cycloptiloides americanus*: a case of the unusual X_1X_20 sex chromosome system in Orthoptera

Octavio M. Palacios-Gimenez ·
Diogo C. Cabral-de-Mello

Received: 29 July 2014 / Accepted: 23 October 2014 / Published online: 6 November 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract A common placement for most sex chromosomes that is involved in their evolutionary histories is the accumulation of distinct classes of repetitive DNAs. Here, with the aim of understanding the poorly studied repetitive DNA organization in crickets and its possible role in sex chromosome differentiation, we characterized the chromosomes of the cricket species *Cycloptiloides americanus*, a species with the remarkable presence of the unusual sex chromosome system $X_1X_20\delta/X_1X_1X_2X_2\delta$. For these purposes, we used C-banding and mapping through the fluorescence in situ hybridization of some repetitive DNAs. The C-banding and distribution of highly and moderately repetitive DNAs (C_0t-1 DNA) varied depending of the chromosome. The greater accumulation of repetitive DNAs in the X_2 chromosome was evidenced. The microsatellites were spread along entire chromosomes, but $(AG)_{10}$ and $(TAA)_{10}$ were less enriched, mainly in the centromeric areas. Among the multigene families, the 18S rDNA was spread throughout almost all of the chromosomes, except for pair 5 and X_2 , while the U2 snDNA was placed exclusively in the largest chromosome. Finally, the 5S rDNA was exclusively located in the short arms of the sex chromosomes. The obtained data reinforce the importance of chromosomal dissociation and inversion as a primary evolutionary mechanism to generate neo-sex chromosomes in the species studied, followed by the repetitive DNAs accumulation. Moreover the exclusive placement of 5S rDNA

in the sex chromosomes suggests the involvement of this sequence in sex chromosome recognition throughout meiosis and, consequently, their maintenance, in addition to their avoiding degeneration.

Keywords FISH · Microsatellites · Multigene families · Repetitive DNAs · Sex chromosomes

Introduction

Repetitive DNAs comprise a large fraction of the eukaryotic genomes and include tandem arrays (minisatellites, microsatellites, satellite DNA and some multigene families) and dispersed repeats (transposons and retrotransposons) (Charlesworth et al. 1994; Nei and Rooney 2005). One characteristic of this class of DNA is the ability to colonize specific genomic regions with low or absent recombination, as for example, sex chromosomes, centromeres, telomeres and heterochromatin (Charlesworth et al. 1994; Steinemann and Steinemann 2005; Kejnovsky et al. 2009). In particular, sex chromosomes are subjected to distinct evolutionary forces compared to autosomes, in which repetitive DNAs sequences preferentially accumulate. This accumulation represents, in some instances, the early steps of differentiation between sex chromosomes, most likely before gene degeneration (Steinemann et al. 1993; Vítková et al. 2007; Matsunaga 2009; Kaiser and Bachtrog 2010).

The $X_0\delta/XX\delta$ chromosomal sex-determination mechanism is commonly found in most Orthopteran insects, although it is presumed that this condition evolved from the $XY\delta/XX\delta$ ancestral sexual system, which is common among distinct lineages (White 1941, 1973; Hughes-Schrader 1947; Hewitt 1979; del Cerro et al. 1998). However, deviations of the modal $X_0\delta/XX\delta$ sex-determination

Communicated by S. Hohmann.

O. M. Palacios-Gimenez · D. C. Cabral-de-Mello (✉)
Departamento de Biologia, Instituto de Biociências/IB,
UNESP-Univ Estadual Paulista, Rio Claro,
São Paulo CEP 13506-900, Brazil
e-mail: mellodc@rc.unesp.br

system have been observed, with the occurrence of some distinct derived sex-systems, in most cases neo-XY♂/XX♀ and neo-X₁X₂Y♂/X₁X₁X₂X₂♀ and even X₁X₂0♂/X₁X₁X₂X₂♀ in the cricket species *Cycloptiloides americanus* and *Endecous ubajarensis* (Mesa and de Mesa 1967; White 1969, 1973; Hewitt 1979; Mesa et al. 1982, 2001, 2002; Castillo et al. 2010; Palacios-Gimenez et al. 2013; Zefa et al. 2014). The Robertsonian (Rb) and tandem fusions, dissociations and inversions involving autosomes and sex chromosomes seem to be major evolutionary causes in the emergence of neo-sex chromosomes in this group (White 1973; Hewitt 1979; Bidau and Martí 2001; Mesa and García-Novo 2001; Mesa et al. 2002; Castillo et al. 2010).

In Orthoptera, mainly in crickets, cytogenetic studies concerning the chromosomal organization and sex chromosome evolution focusing on repetitive DNA sequences are scarce, and therefore, the possible contribution of this class of DNA in the origin/evolution of derived sex systems remains largely obscure (Palacios-Gimenez et al. 2013). Here, to understand repetitive DNA chromosomal organization and sex chromosome evolution in crickets, we studied the X₁X₂0♂/X₁X₁X₂X₂♀ sex system of *C. americanus*, which is a rare case of this sex system that has been observed in Orthoptera (Mesa et al. 2002), through classical cytogenetic techniques and fluorescence in situ hybridization (FISH) for some repetitive DNAs. Our data are discussed to shed light on repetitive DNA organization in crickets and the origin/evolution of the unusual X₁X₂0 sex chromosome system.

Materials and methods

Some males and females of *Cycloptiloides americanus* were collected in Rio Claro, São Paulo, Brazil, in January 2013 and maintained in adapted boxes for reproduction to obtain eggs for embryo isolation. The meiotic chromosomes were obtained from testis and embryos were cytologically prepared as in Webb et al. (1978) with slight modifications for mitotic chromosome obtaining. For general chromosome analysis, we used conventional staining with 5 % Giemsa for some males and females to confirm the presence of the X₁X₂0♂/X₁X₁X₂X₂♀ sex system as described by Mesa et al. (2002). C-banding was performed according to Sumner (1972), and fluorochrome staining CMA₃/DA/DAPI followed the protocol of Schweizer et al. (1983).

The probes were obtained through Polymerase Chain Reaction (PCR) using as templates the genomic DNA of *Abracris flavolineata* (5S rDNA) and *Rhammatocerus brasiliensis* (U2 snDNA) and primers as described by Cabral-de-Mello et al. (2010), Bueno et al. (2013), respectively. The fragments were previously sequenced and deposited

in GenBank, KC936996 (5S rDNA) and KC896794 (U2 snDNA). For 18S rDNA, the cloned fragment from the beetle *Dichotomius semisquamosus* was used GQ443313 (Cabral-de-Mello et al. 2010). The telomeric motif was obtained using the self-complementary primers (TTAGG)₅ and (CCTAA)₅ through PCR according to Ijdo et al. (1991). The C₀t-1 DNA (DNA fraction that is enriched for highly and moderately repetitive DNA sequences) fraction was obtained according to the protocol of Zwick et al. (1997) with modifications (Cabral-de-Mello et al. 2010) using genomic DNA isolated following the protocol that was described by Sambrook and Russel (2001). This protocol used 25 min of reassociation at 65 °C.

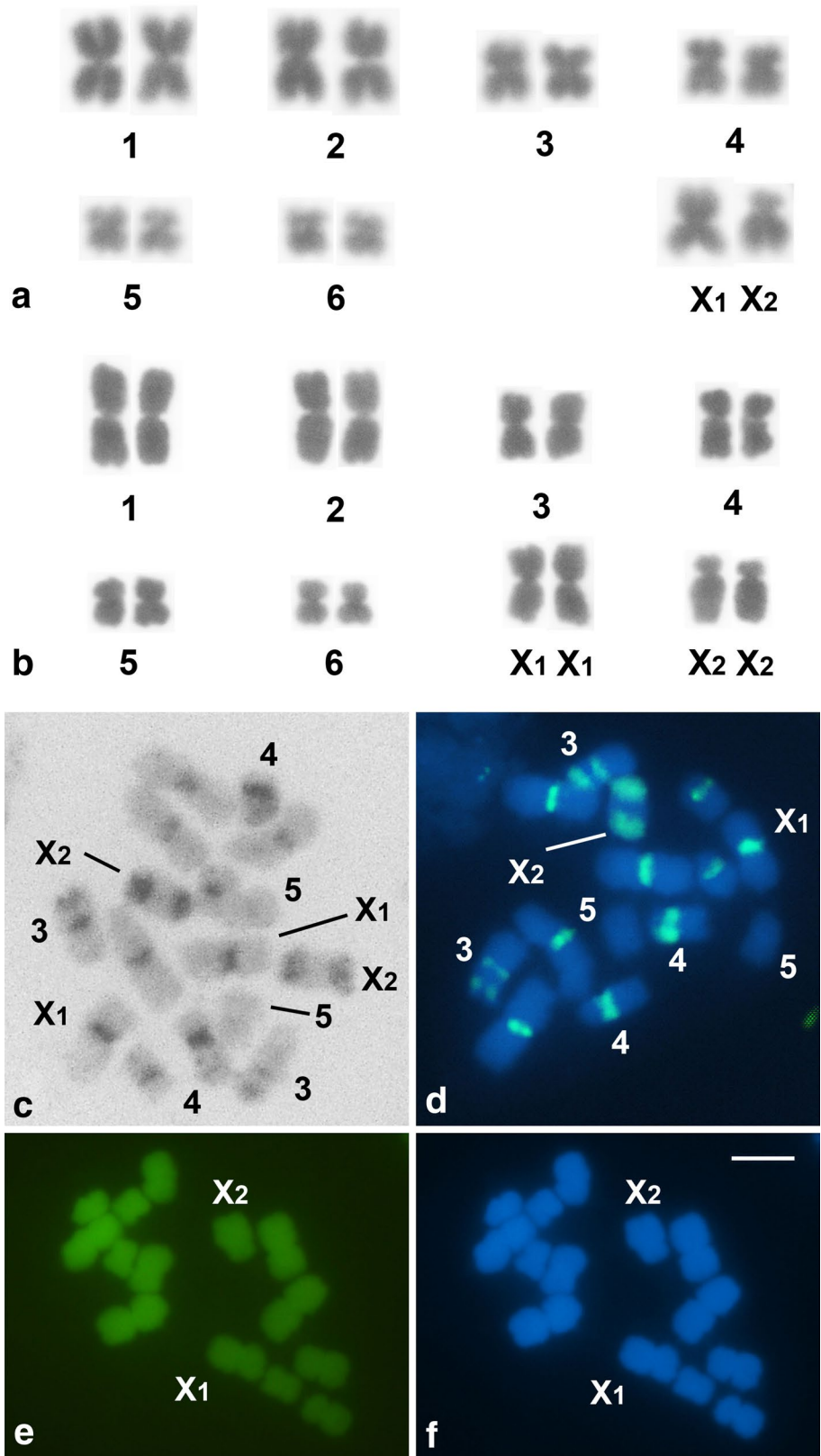
The plasmid containing the fragment of the 18S rRNA gene and the C₀t-1 DNA fraction were labeled through nick translation using biotin-14-dATP (Invitrogen, San Diego, CA, USA). The eight synthetic microsatellites oligonucleotide probes were directly labeled with biotin-14-dATP during their synthesis at the 5' end: (A)₃₀, (C)₃₀, (CA)₁₅, (CG)₁₅, (AG)₁₀, (TAA)₁₀, (GACA)₄ and (GATA)₈ (Sigma, St Louis, MO, USA). The 5S rDNA, U2 snDNA and the telomeric probes were labeled with digoxigenin-11-dUTP (Roche, Mannheim, Germany) by PCR.

Fluorescence in situ hybridization (FISH) was performed according to Pinkel et al. (1986) with modifications (Cabral-de-Mello et al. 2010). The digoxigenin-11-dUTP labeled probes were detected using anti-digoxigenin rhodamine (Roche), and the biotin-14-dATP labeled probes were detected using streptavidin, Alexa Fluor 488 conjugated (Invitrogen). 4',6-Diamidino-2'-phenylindole dihydrochloride (DAPI) was used to counterstain the preparations that were mounted with VECTASHIELD (Vector, Burlingame, CA, USA). The images were obtained using an Olympus microscope BX51 that was equipped with a fluorescence lamp and appropriate filters. The photographs were recorded using a DP70 cooled digital camera in grayscale, and the images were pseudo-colored in blue (chromosomes) and red or green (signals), merged and optimized for brightness and contrast using Adobe Photoshop CS2.

Results

Cycloptiloides americanus presents a diploid number ($2n = 14♂/16♀$) and X₁X₂0♂/X₁X₁X₂X₂♀ sex-determining chromosomal system (Fig. 1a, b) as observed by Mesa et al. (2002). The autosomes are metacentric, and X₁ is metacentric, while X₂ is acrocentric (Fig. 1a, b); differing from the previous description by Mesa et al. (2002) (see “Discussion”). The C-banding and mapping of highly and moderately repetitive DNAs (C₀t-1 DNA) showed a highly variable inter-chromosomal distribution with centromeric (pairs 1–3 and 6), interstitial (pair 4) and terminal (pair 3) blocks,

Fig. 1 Conventional staining of the male (a) and female (b) karyotypes of *Cycloptiloides americanus*. The chromosomes were arranged in descending order of size. C-banding (c), FISH analysis of the C_{0t-1} DNA fraction (d) and CMA₃ (e) and DAPI (f) staining in mitotic cells from female (c) and male (d–f). The sex chromosomes are indicated. Bar 5 μm



in addition to the absence of marks in pair 5 (Fig. 1c, d). Specifically for sex chromosomes, X_1 and X_2 showed distinct patterns, with a centromeric block in X_1 , while the entire short arm and the distal half of the long arm of X_2 were rich in repetitive DNAs. The C-positive blocks were non-A + T or G + C-rich (Fig. 1e, f).

The eight microsatellite arrays that were mapped were spread along entire chromosomes, even both X_1 and X_2 sex-chromosomes; however, $(AG)_{10}$ and $(TAA)_{10}$ were less enriched in the centromeric areas, and their signals were less intense along the chromosome arms (Fig. 2).

The telomeric probe revealed signals only in the terminal regions (Fig. 3a). Among the multigene families that were mapped, the U2 snDNA was placed interstitially on the long arm of pair 1 (Fig. 3b), while the FISH for 18S rDNA revealed signals in almost all of the chromosomes, except in pair 5 and the X_2 chromosome. The blocks of 18S rDNA were located as follows: centromeric (pairs 1, 2, 6 and X_1), centromeric and terminal (pair 3) and interstitial (pair 4) (Fig. 3c). The clusters for the 5S rDNA were placed exclusively in the sex chromosomes, proximally in the short arm of both chromosomes (Fig. 3c). The FISH for the 5S rDNA

Fig. 2 Male and female mitotic metaphases of *Cycloptiloides americanus* hybridized with different microsatellite arrays. The sex chromosomes and each type of microsatellite array are shown directly in the images. Note the less-intense signal along the chromosomal arms and the absence of centromeric hybridization for $(AG)_{10}$ and $(TAA)_{10}$. Cells with two X chromosomes (male) and with four X chromosomes (female). Bar 5 μ m

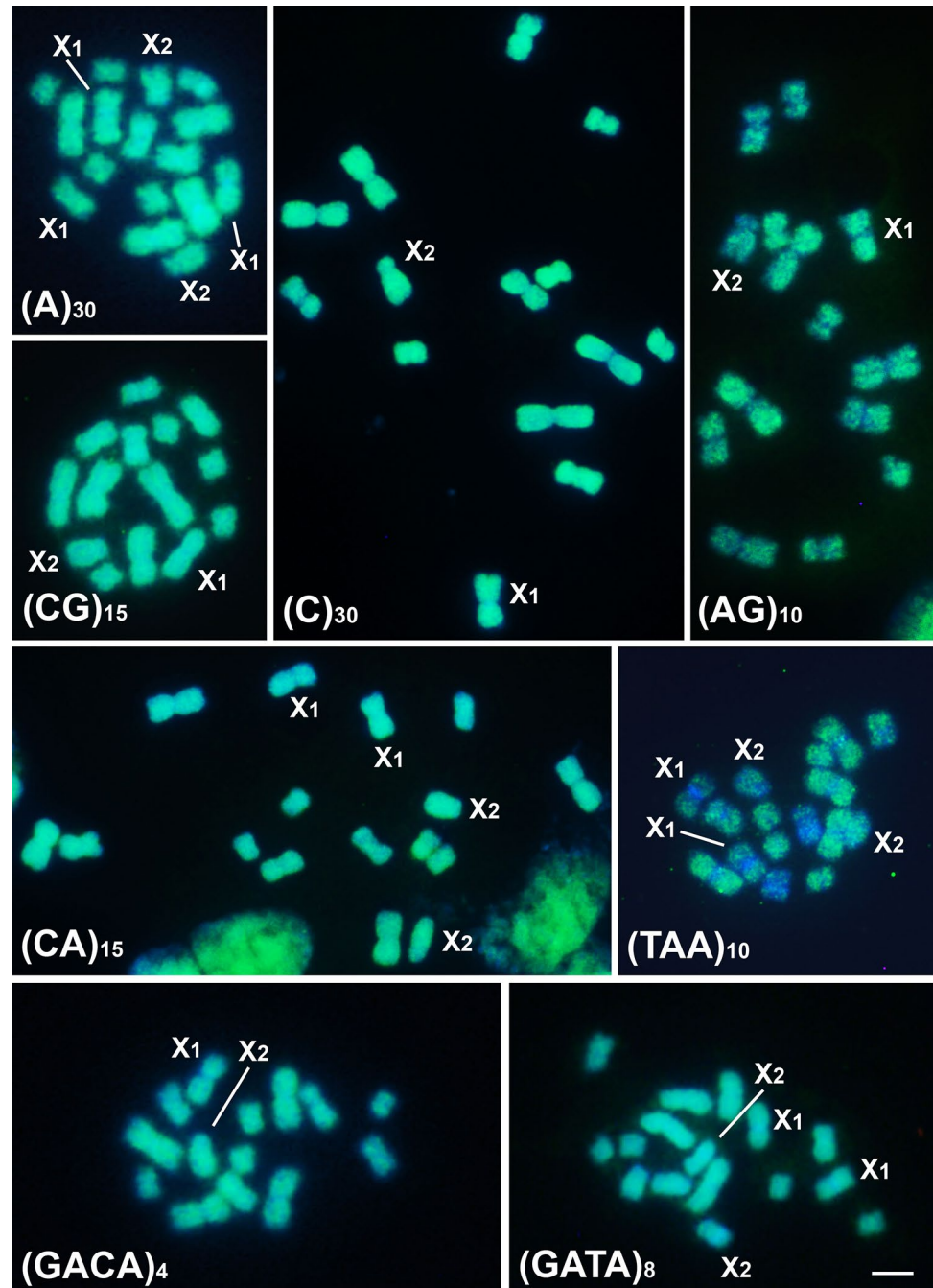
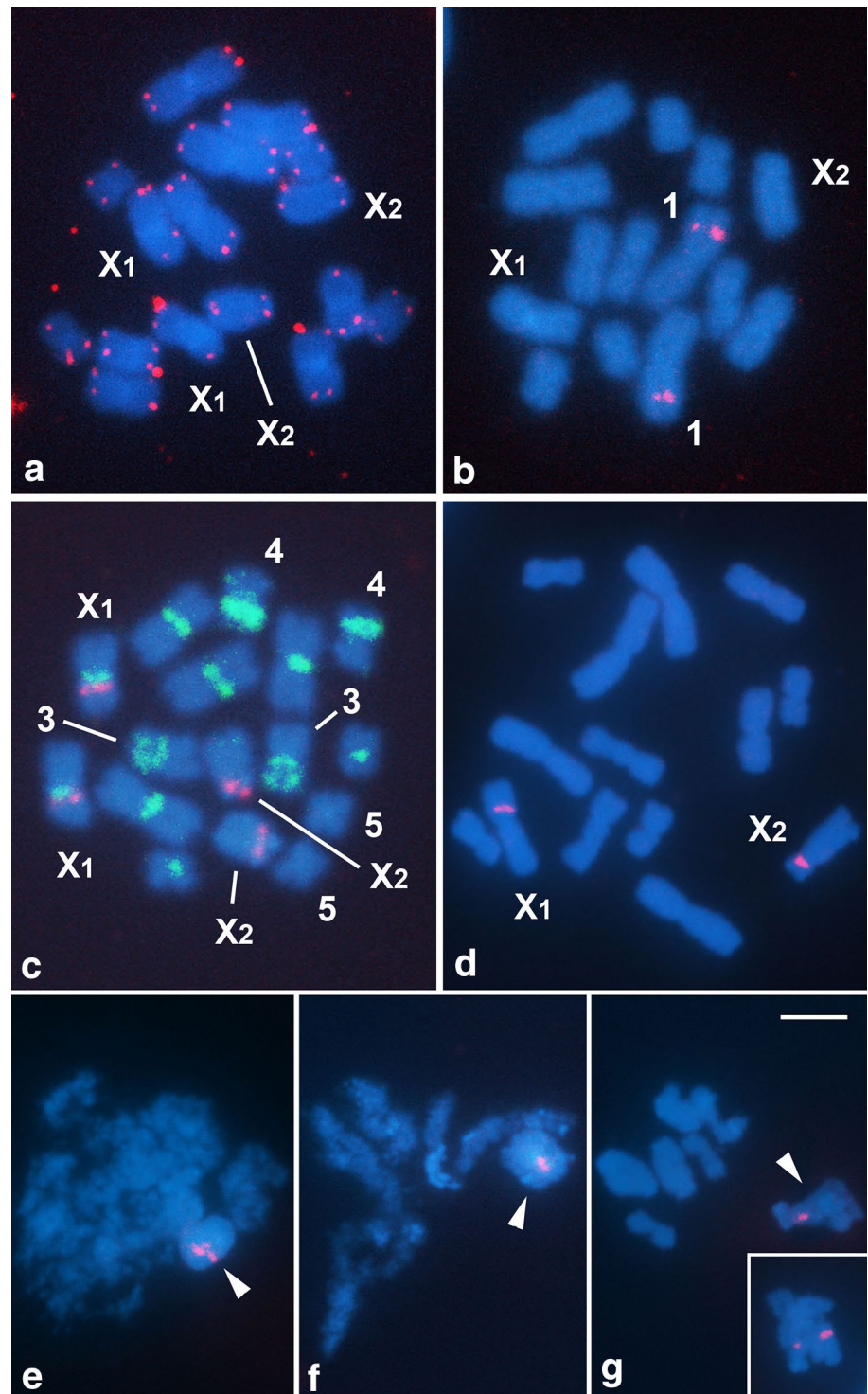


Fig. 3 FISH analysis for telomere repeat (a), U2 snDNA (b), 18S rDNA in green (c) and 5S rDNA in red (c–g) probes in male (b, d) and female (a, c) mitotic metaphases and male meiotic cells, (e) zygotene (f) pachytene and (g) metaphase I. The chromosomes carrying positive signals and the sex chromosomes are indicated. In (e–g) the arrowheads indicate the sex chromosomes associated. Note the multiple sites of 18S rDNA in (c) and the presence of 5S rDNA exclusively in both of the sex chromosomes (c–g), that are associated in some meiotic phases (e–g). The inset in (g) show the side-by-side orientation of 5S rDNA in metaphase I. Bar 5 μ m

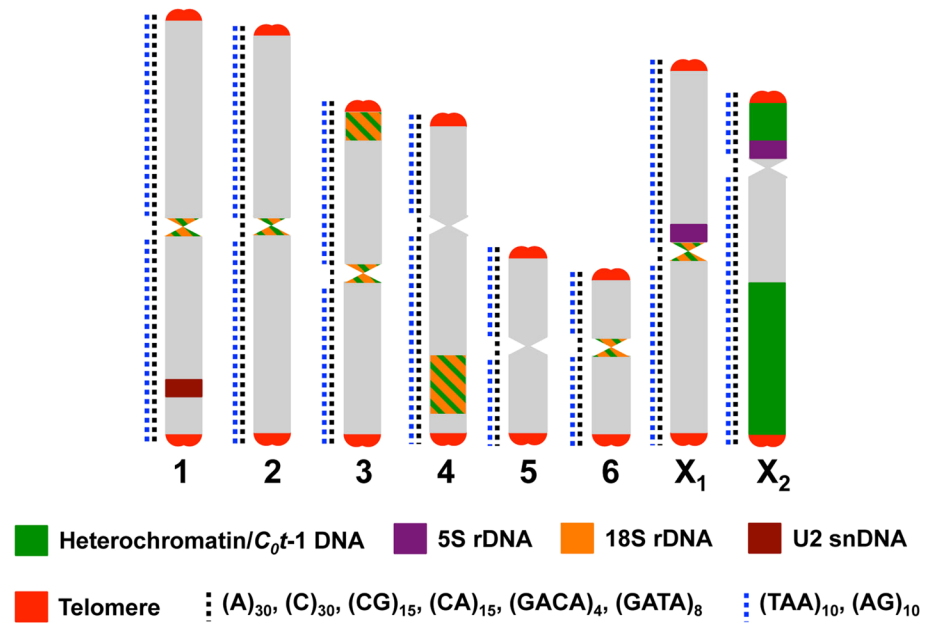


in males confirmed the placement of this sequence in X_1 and X_2 (Fig. 3d). Meiotic analysis using 5S rDNA as probe revealed the two clusters associated in initial cells (Fig. 3e, f) until metaphase (Fig. 3g), although in this phase the two clusters were also oriented side by side (Fig. 3g, inset). In the ideogram (Fig. 4), the markers that were obtained for the *C. americanus* chromosomes are summarized.

Discussion

The diploid number and the sex-determining chromosomal system of *C. americanus* agree with those previously described (Mesa et al. 2002). However, from our analysis, it is evident that the sex chromosomes (X_1 and X_2) were previously defined mistakenly being the X_1

Fig. 4 Ideogram showing the *Cycloptiloides americanus* haploid complement and all of the chromosomal markers that were analyzed. The probes and their relative position on each chromosome are indicated using lines and colors. Note that each chromosome can be differentially recognized by morphology, size, presence/absence and position of the markers (color figure online)



metacentric and the X_2 acrocentric, unlike the acrocentric X_1 and submetacentric X_2 defined by Mesa et al. (2002). The submetacentric X_2 that was described by Mesa et al. (2002) corresponds to autosomal pair 4, which occurs invariably in two copies in both sexes and is easily recognized in some cells due to the occurrence of less-condensed chromatin in the middle of the long arm. The $X_1X_2\sigma/X_1X_1X_2X_2\varphi$ sex-determining system that has been described in the cricket *C. americanus* is a rare case of this type that has been observed in Orthoptera, but it is apparently common in some other invertebrates, such as spiders (White 1973; Maddison and Leduc-Robert 2013; Forman et al. 2013; Zefa et al. 2014) and some species of Heteroptera (Ueshima 1979; Kusnetzova 1988); Homoptera (Morgan 1915; White 1940) and Nematoda (Walton 1924; White 1940), and it is uncommon in eukaryotes as a whole. Concerning the origin of this system, in spiders, for example, the $X_1X_2\sigma$ sex-determining system is considered a plesiomorphic feature, and various hypotheses for its origin have been raised, including the duplication of the original X chromosome, centric fragmentation, fission and the translocation of an X segment to a centric fragment of a supernumerary chromosome, among others (reviewed by Araujo et al. 2012). In *C. americanus*, according to Mesa et al. (2002), the presence of two X chromosomes in males could be explained by the dissociation of the original metacentric X into two acrocentric sex chromosomes that could be followed by pericentric inversion and repetitive DNA expansion.

The inter-chromosomal variability that was observed in the highly and moderately repetitive DNAs indicates differences in the amount of repetitive DNAs for each chromosomal pair and suggests the complex reorganization of

this genome fraction, not following the equilocal model for heterochromatin distribution (Heitz 1935) and differing in relation to other Orthoptera that have been previously studied, with mainly centromeric/proximal C-positive blocks (Webb and Westerman 1978; Santos et al. 1983; John and King 1985; Westerman and Hewitt 1985; Cabrero and Camacho 1986). This variability could be related to chromosomal rearrangement on large scales, considering the reduced chromosome number in *C. americanus*, in addition to the high karyotypic variability in the vast majority of cricket species of the Grylloidea superfamily (Hewitt 1979). For sex chromosomes, the differential accumulation of highly and moderately repetitive DNAs between both the X_1 and X_2 suggests initial steps of evolutionary divergence after the origin through the dissociation of the original metacentric X into two sex-chromosomes (Mesa et al. 2002). These initial steps of evolutionary divergence are supported by the increased accumulation of repetitive DNAs in both of the arms of X_2 and the absence of 18S rDNA clusters in relation to the X_1 chromosome (see below). Apparently, sex chromosomes of *C. americanus* follow similar evolutionary trajectories as those that have been proposed for neo-systems with the Y chromosome, which, due to the absence of recombination and to differential accumulation and expansion of distinct repetitive DNAs, could cause the degenerate fate (Charlesworth et al. 1994, 2005; Steinemann and Steinemann 2005; Kaiser and Bachtrog 2010; Steflava et al. 2013; Kejnovský et al. 2013).

None of the microsatellites that were mapped in *C. americanus* chromosomes were located exclusively or enriched only on sex chromosomes, and apparently, none of these microsatellites has any direct relationship with the differentiation of X_1 and X_2 or contributed

specifically to their origin/evolution. Similar patterns of random microsatellite distribution have been recently reported in other Orthopteran species, i.e., the grasshoppers *Abracris flavolineata* (Milani and Cabral-de-Mello 2014) and *Ronderosia bergi* (Palacios-Gimenez et al. in preparation), although for some repeats, in addition to disperse signals, specific enriched regions were also observed, indicating distinct patterns of accumulation at the intragenomic and interspecific level. Similar dispersal patterns of microsatellite arrays and specific localization in sex chromosomes have been previously reported in other eukaryotes, such as plants (Kubat et al. 2008; Kejnovský et al. 2013), reptiles (Pokorná et al. 2011; Matsubara et al. 2013) and fish (Terencio et al. 2013), and some authors have highlighted a role for this class of sequence in sex chromosome differentiation (see for example Kubat et al. 2008; Pokorná et al. 2011; Kejnovský et al. 2013; Matsubara et al. 2013).

Concerning the multigene families that were mapped, the scattering of major rDNA was reported in other Orthoptera species, such as grasshoppers (Cabrero and Camacho 2008), and the distinct level of dispersal and the position change have also been reported in other insect species, such as Heteroptera (Panzera et al. 2012); Lepidoptera (Nguyen et al. 2010) and Coleoptera (Cabral-de-Mello et al. 2011). According to some authors, the dispersion for major rDNA clusters is favored by an association with transposable elements and the occurrence of ectopic recombination and extrachromosomal circular DNA (see for example Schubert 1984; Schubert and Wobus 1985; Hanson et al. 1996; Cabrero and Camacho 2008; Raskina et al. 2008; Nguyen et al. 2010; Cabral-de-Mello et al. 2011; Panzera et al. 2012). In contrast to this pattern, the U2 snDNA cluster occurring in one chromosomal pair (pair 1 in *C. americanus*) was reported in other animals, including *Gymnotus* fish (Utsunomia et al. 2014) and one grasshopper (Bueno et al. 2013). Interestingly, in *Gymnotus pantanal* and in the grasshopper *Dichromatos lilloanus*, the U2 snDNA was placed on the derivate sex chromosomes (neo- X_1X_2Y) X_1 and Y, respectively, providing insight into the origin/evolution for these chromosomes in both of these species (Palacios-Gimenez et al. 2013; Utsunomia et al. 2014).

Here, the 5S rDNA mapping provides support for discussion concerning the evolution and maintenance of the X_1X_20 system. The presence of this gene in both the X_1 and X_2 sex chromosomes of *C. americanus* supports the hypothesis of the origin of the unusual X_1X_20 through the dissociation of the original X metacentric bearing the 5S rDNA gene into two acrocentric X-chromosomes, followed by a large pericentromeric inversion on the X_1 , turning the latter into a metacentric X. For X_2 , it is likely that the expansion of repetitive DNAs, corroborated by

heterochromatin and C_0t-1 DNA occurrence, occurred in the short arm, causing its expansion. In accordance with Mesa et al. (2002), the X_1 and X_2 chromosomes remain together in a single heterochromatic mass during prophase I and, at metaphase I, remain either connected or independent; however, in every case, both X_1 and X_2 segregate to the same pole during anaphase I. In light of such evidence, we hypothesize that the repetitive DNAs that are present in these chromosomes (such as 5S rDNA) could facilitate the recognition, pairing and synapses of X_1 and X_2 throughout meiosis I, ensuring the passage of both of these chromosomes to the same pole. In addition, repetitive DNAs could account for a major inter-chromosomal recombination (between X_1 and X_2) leading to the homogenization of some repetitive DNA sequences and thereby to the association of both X chromosomes during meiosis I.

Conclusions

Cycloptiloides americanus presents a rare case of the intriguing $X_1X_20\delta/X_1X_1X_2X_2\phi$ sex-determining system in Orthoptera, indicating that the sex-determining systems in crickets are apparently more diverse than in grasshoppers, where only derived variants, such as neo-XY δ /XX ϕ and neo- $X_1X_2Y\delta/X_1X_1X_2X_2\phi$, have been observed (Mesa et al. 1982; Castillo et al. 2010; Palacios-Gimenez et al. 2013), making this group an interesting model for investigating sex chromosome evolution. Our results also reinforce the importance of chromosomal rearrangement (dissociation and inversion) as a primary evolutionary mechanism to generate neo-sex chromosomes in *C. americanus*, followed by the accumulation of repetitive DNAs, which untimely modify the chromatin structures into heterochromatin. Finally, the presence of the 5S rDNA structural gene exclusively on the X_1 and X_2 sex chromosomes could prevent the complete degeneration and loss of these elements, which is a common pattern of derived-sex chromosomes. The obtained data shed light on repetitive DNA organization in crickets and the origin/evolution of the unusual X_1X_20 sex chromosome system, being the starting point for other analysis to elucidate sex chromosome diversification in Orthoptera.

Acknowledgments The authors are grateful to Antonio Sergio Pascon for technical assistance in embryos obtaining. The study was funded by Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP, process number 2014/11763-8), Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and the Programa Primeiros Projetos-PROPE/UNESP. OMPG scholarship was supported by FAPESP (process number 2014/02038-8).

Conflict of interest The authors declare no conflict of interest.

References

- Araujo D, Schneider MC, Paula-Neto E, Cella DM (2012) Sex chromosomes and meiosis in spiders: a review. In: Swan A (ed) Meiosis-molecular mechanisms and cytogenetic diversity. Intech, Rijeka, pp 87–108
- Bidau CJ, Martí DA (2001) Meiosis and the Neo-XY of *Dichroplus vittatus* (Melanoplineae, Acrididae): a comparison between sexes. *Genetica* 110:185–194
- 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
- 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, Oliveira SG, Moura RC, Martins C (2011) 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 Genet* 12:88
- Cabrero J, Camacho JPM (1986) Cytogenetic studies in gomphocerine grasshoppers I. Comparative analysis of chromosome C-banding pattern. *Heredity* 56:365–372
- 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
- Castillo ER, Martí DA, Bidau CJ (2010) Sex and neo-sex chromosomes in Orthoptera: a review. *J Orthopt Res* 19:213–231
- Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371:215–220
- Charlesworth D, Charlesworth B, Marais G (2005) Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95:118–128
- del Cerro AL, Cunãdo N, Santos JL (1998) Synaptonemal complex analysis of the X₁X₂Y trivalent in *Mantis religiosa* L. males: inferences on the origin and maintenance of the sex-determining mechanism. *Chromosome Res* 6:5–11
- Forman M, Nguyen P, Hula V, Král J (2013) Sex chromosome pairing and extensive NOR polymorphism in *Wadicosa fidelis* (Araneae: Lycosidae). *Cytogenet Genome Res* 141:43–49
- Hanson RE, Islam-Faridi MN, Percival EA et al (1996) Distribution of 5S and 18S-28S rDNA loci in a tetraploid cotton (*Gossypium hirsutum* L.) and its putative diploid ancestors. *Chromosoma* 105:55–61
- Heitz E (1935) Die Herkunft der Chromozentren. *Planta* 18:571–635
- Hewitt GM (1979) Grasshoppers and crickets. *Animal cytogenetics: insecta 1 Orthoptera*, vol 3. Gebrüder Borntraeger, Berlin
- Hughes-Schrader S (1947) Reversion of XO to XY sex chromosome mechanism in a phasmid. *Chromosoma* 3:57–65
- Ijdo JW, Wells RA, Baldini A, Reeders ST (1991) Improved telomere detection using a telomere repeat probe (TTAGGG)_n generated by PCR. *Nucleic Acids Res* 19:4780
- John B, King M (1985) The inter-relationship between heterochromatin distribution and chiasma distribution. *Genetica* 66:183–194
- Kaiser VB, Bachtrog D (2010) Evolutions of sex chromosome in insects. *Annu Rev Genet* 44:91–112
- Kejnovsky E, Hobza R, Cermak T, Kubat Z, Vyskot B (2009) The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity* 102:533–541
- Kejnovský E, Michalovova M, Steflava P et al (2013) Expansion of microsatellites on evolutionary young Y chromosome. *PLoS One* 8:e45519
- Kubat Z, Hobza R, Vyskot B, Kejnovsky E (2008) Microsatellite accumulation on the Y chromosome in *Silene latifolia*. *Genome* 51:350–356
- Kusnetzova VG (1988) The unusual mechanism of sex-chromosome determination in the cotton bug *Dysdercus supersticiosus* F (Pyrrhocoridae, Heteroptera). *Dokl Akad Nauk SSSR* 301:456–458
- Maddison WP, Leduc-Robert G (2013) Multiple origins of sex chromosome fusions correlated with chiasma localization in Habronattus jumping spiders (Araneae: Salticidae). *Evolution* 67:2272–2558
- Matsubara K, Knopp T, Sarre SD, Georges A, Ezaz T (2013) Karyotypic analysis and FISH mapping of microsatellite motifs reveal highly differentiated XX/XY sex chromosomes in the pink-tailed worm-lizard (*Aprasia parapulchella*, Pygopodidae, Squamata). *Mol Cytogenet* 6:60
- Matsunaga S (2009) Junk DNA promotes sex chromosome evolution. *Heredity* 102:525–526
- Mesa A, de Mesa RS (1967) Complex sex-determining mechanisms in three species of South American grasshoppers (Orthoptera, Acridoidea). *Chromosoma* 21:163–180
- Mesa A, García-Novo P (2001) *Neometrypus badius* a new species of cricket with an unusual sex determining mechanism (Grylloidea, Eneopteridae, Tafaliscinae, Neometrypini). *J Orthopt Res* 10:81–87
- Mesa A, Ferreira A, Carbonell CS (1982) Cariología de los acridoideos neotropicales: estado actual de su conocimiento y nuevas contribuciones. *Ann de la Société Entomologique de France (NS)* 18:507–526
- Mesa A, Fontanetti CS, García-Novo P (2001) Does an X-autosome centric fusion in Acridoidea condemn the species to extinction? *J Orthopt Res* 10:141–146
- Mesa A, García-novo P, dos Santos D (2002) X₁X₂O (male)-X₁X₁X₂X₂ (female) chromosomal sex determining mechanism in the cricket *Cycloptiloides americanus* (Orthoptera, Grylloidea, Mogoplistidae). *J Orthopt Res* 1:87–90
- Milani D, Cabral-de-Mello DC (2014) Microsatellite organization in the grasshopper *Abracris flavolineata* (Orthoptera: Acrididae) revealed by FISH mapping: remarkable spreading in the A and B chromosomes. *PLoS One* 9:e97956
- Morgan TH (1915) The predetermination of sex in Phylloxerans and Aphids. *J Exp Zool* 19:285–321
- 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
- Palacios-Gimenez OM, Castillo ER, Martí DA, Cabral-de-Mello DC (2013) Tracking the evolution of sex chromosome systems in Melanoplineae grasshoppers through chromosomal mapping of repetitive DNA sequences. *BMC Evol Biol* 13:167
- Panzer Y, Pita S, Ferreira MJ et al (2012) High dynamics of rDNA cluster location in kissing bug holocentric chromosomes (Triatominae, Heteroptera). *Cytogenet Genome Res* 138:56–67
- Pinkel D, Straume T, Gray JW (1986) Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc Natl Acad Sci USA* 83:2934–2938
- Pokorná M, Kratochvíl L, Kejnovský E (2011) Microsatellite distribution on sex chromosomes at different stages of heteromorphism and heterochromatinization in two lizard species (Squamata: Eublepharidae: *Coleonyx elegans* and Lacertidae: *Eremias velox*). *BMC Genet* 12:90
- 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

- Sambrook J, Russel DW (2001) Molecular cloning. A laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, New York
- Santos JL, Arana P, Giráldez L (1983) Chromosome C-banding patterns in Spanish Acridoidea. *Genetica* 61:65–74
- Schubert I (1984) Mobile nucleolus organizing regions (NORs) in *Allium* (Liliaceae S-Lat)-inferences from the specificity of silver staining. *Plant Syst Evol* 144:291–305
- Schubert I, Wobus U (1985) In situ hybridization confirms jumping nucleolus organizing regions in *Allium*. *Chromosoma* 92:143–148
- Schweizer D, Mendelak M, White MJD, Contreras N (1983) Cytogenetics of the parthenogenetic grasshopper *Warramaba virgo* and its bisexual relatives X. Pattern of fluorescent banding. *Chromosoma* 88:227–236
- Stefflova P, Tokan V, Vogel I et al (2013) Contrasting patterns of transposable element and satellite distribution on sex chromosomes (XY₁Y₂) in the dioecious plant *Rumex acetosa*. *Genome Biol Evol* 5:769–782
- Steinemann S, Steinemann M (2005) Retroelements: tools for sex chromosome evolution. *Cytogenet Genome Res* 110:134–143
- Steinemann M, Steinemann S, Lottspeich F (1993) How Y chromosomes become genetically inert. *Proc Natl Acad Sci USA* 90:5737–5741
- Sumner AT (1972) A simple technique for demonstrating centromeric heterochromatin. *Exp Cell Res* 75:304–306
- Terencio ML, Schneider CH, Gross MC et al (2013) Evolutionary dynamics of repetitive DNA in *Semaprochilodus* (Characiformes, Prochilodontidae): a fish model for sex chromosome differentiation. *Sex Dev* 7:325–333
- Ueshima N (1979) Hemiptera II: Heteroptera. *Animal cytogenetics 3: insecta* 6. Gebrueder Borntraeger, Berlin
- Utsunomia R, Scacchetti PC, Pansonato-Alves JC, Oliveira C, Foresti F (2014) Comparative chromosome mapping of U2 snRNA and 5S rRNA genes in *Gymnotus* species (Gymnotiformes, Gymnotidae): evolutionary dynamics and sex chromosome linkage in *G. pantanal*. *Cytogenet Genome Res* 142:286–292
- Vítková M, Fukova I, Kubíčková S, Marec F (2007) Molecular divergence of the W chromosomes in pyralid moths (Lepidoptera). *Chromosome Res* 15:917–930
- Walton AC (1924) Studies on nematode gametogenesis. *Z Zell-u Gewebl* 1:167–239
- Webb GC, Westerman M (1978) G- and C-banding in the Australian grasshopper *Phaulacridium vittatum*. *Heredity* 41:131–136
- Webb GC, White MJD, Contreras N, Cheney J (1978) Cytogenetics of the parthenogenetic grasshopper *Warramaba* (formerly *Moraba*) *virgo* and its bisexual relatives IV. Chromosome banding studies. *Chromosoma* 67:309–339
- Westerman M, Hewitt GM (1985) Chromosome banding in *Podisma pedestris*. *Heredity* 55:157–161
- White MJD (1940) The origin and evolution of multiple sex-chromosome mechanisms. *J Genet* 40:303–336
- White MJD (1941) The evolution of the sex chromosomes I. The XO and X₁X₂Y mechanisms in praying mantids. *J Genet* 42:143–172
- White MJD (1969) Chromosomal rearrangement and speciation in animals. *Ann Rev Genet* 3:75–98
- White MJD (1973) *Animal cytology and evolution*. Cambridge University Press, Cambridge
- Zefa E, Redu DR, Costa MKM, Fontanetti CS, Gottschalk MS, Padilha GB, Silva AF, Martins LP (2014) A new species of *Endecous* Saussure, 1878 (Orthoptera, Gryllidae) from northeast Brazil with the first X₁X₂0 chromosomal sex system in Gryllidae. *Zootaxa* 3847:125–132
- Zwick MS, Hanson RE, McKnight TD, Nurul-Islam-Faridi M, Stelly DM (1997) A rapid procedure for the isolation of C₀t-1 DNA from plants. *Genome* 40:138–142

4.2. Capítulo 2

Contrasting the chromosomal organization of repetitive DNAs in two Gryllidae crickets with highly divergent karyotypes

Octavio M. Palacios-Gimenez, Carlos Roberto Carvalho, Fernanda Aparecida Ferrari Soares, Diogo C. Cabral-de-Mello

Publicado na revista *PLoS ONE* (2015), 10: e0143540.

RESEARCH ARTICLE

Contrasting the Chromosomal Organization of Repetitive DNAs in Two Gryllidae Crickets with Highly Divergent Karyotypes

Octavio M. Palacios-Gimenez¹, Carlos Roberto Carvalho², Fernanda Aparecida Ferrari Soares², Diogo C. Cabral-de-Mello^{1*}

1 UNESP—Univ. Estadual Paulista, Instituto de Biociências/IB, Departamento de Biologia, Rio Claro, SP, Brazil, **2** UFV—Univ. Federal de Viçosa, Centro de Ciências Biológicas, Departamento de Biologia Geral, Viçosa, MG, Brazil

* mellodc@rc.unesp.br



OPEN ACCESS

Citation: Palacios-Gimenez OM, Carvalho CR, Ferrari Soares FA, Cabral-de-Mello DC (2015) Contrasting the Chromosomal Organization of Repetitive DNAs in Two Gryllidae Crickets with Highly Divergent Karyotypes. PLoS ONE 10(12): e0143540. doi:10.1371/journal.pone.0143540

Editor: Igor V Sharakhov, Virginia Tech, UNITED STATES

Received: August 24, 2015

Accepted: November 5, 2015

Published: December 2, 2015

Copyright: © 2015 Palacios-Gimenez et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying the findings described in the paper is freely available to other researchers (1) in the body of the paper and (2) in the supporting information.

Funding: This study was funded by a scholarship obtained from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, process numbers 2014/11763-8, 2014/02038-8) and a research productivity fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (CNPQ, process number 304758/2014-0). The funders had no role in study design, data collection

Abstract

A large percentage of eukaryotic genomes consist of repetitive DNA that plays an important role in the organization, size and evolution. In the case of crickets, chromosomal variability has been found using classical cytogenetics, but almost no information concerning the organization of their repetitive DNAs is available. To better understand the chromosomal organization and diversification of repetitive DNAs in crickets, we studied the chromosomes of two Gryllidae species with highly divergent karyotypes, i.e., $2n(\sigma) = 29,X0$ (*Gryllus assimilis*) and $2n = 9, neo-X_1X_2Y$ (*Eneoptera surinamensis*). The analyses were performed using classical cytogenetic techniques, repetitive DNA mapping and genome-size estimation. Conserved characteristics were observed, such as the occurrence of a small number of clusters of rDNAs and U snDNAs, in contrast to the multiple clusters/dispersal of the H3 histone genes. The positions of U2 snDNA and 18S rDNA are also conserved, being intermingled within the largest autosome. The distribution and base-pair composition of the heterochromatin and repetitive DNA pools of these organisms differed, suggesting reorganization. Although the microsatellite arrays had a similar distribution pattern, being dispersed along entire chromosomes, as has been observed in some grasshopper species, a band-like pattern was also observed in the *E. surinamensis* chromosomes, putatively due to their amplification and clustering. In addition to these differences, the genome of *E. surinamensis* is approximately 2.5 times larger than that of *G. assimilis*, which we hypothesize is due to the amplification of repetitive DNAs. Finally, we discuss the possible involvement of repetitive DNAs in the differentiation of the neo-sex chromosomes of *E. surinamensis*, as has been reported in other eukaryotic groups. This study provided an opportunity to explore the evolutionary dynamics of repetitive DNAs in two non-model species and will contribute to the understanding of chromosomal evolution in a group about which little chromosomal and genomic information is known.

and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: $2n$, diploid number; C_0t -DNA, C_0 is the initial concentration of single-stranded DNA in mol/l and t is the reannealing time in seconds; CMA_3 , Chromomycin A₃; DA, Distamycin; DAPI, 4', 6-Diamidino-2-phenylindole; FCM, Flow cytometry; FISH, Fluorescence in situ hybridization; PCR, Polymerase chain reaction; PI, Propidium iodide; Rb-translocation, Robertsonian translocation; rDNA, Ribosomal DNA; rRNA, Ribosomal RNA; snRNA, Small nuclear RNA.

Introduction

The Orthoptera order comprises more than 25,700 species, which form the most diverse group of polyneopteran insects [1,2,3]. Variability in the diploid number has been documented in distinct lineages, which range in the most highly studied groups from $2n(\sigma) = 15$ to $2n(\sigma) = 35$ in katydids Tettigoniidae [4,5,6], from $2n(\sigma) = 8$ to $2n(\sigma) = 23$ in Acrididae grasshoppers [7,8,9] and from $2n(\sigma) = 7$ to $2n(\sigma) = 29$ in Gryllidae crickets [9,10,11,12,13], although the modal number of chromosomes has been documented, such as $2n(\sigma) = 23$ in grasshoppers and $2n = 21$ in Gryllidae crickets [8,9]. This extreme karyotypic divergence has been attributed to multiple major chromosomal restructuring events, such as centric fusions, tandem fusions, reciprocal translocations, dissociations and inversions involving autosomes and sex chromosomes [8,9]. In addition to these rearrangements being responsible for the variations in the diploid number, they also caused diversification of the sex chromosomes, originating distinct sex systems, such as neo-XY, neo- X_1X_2Y and X_1X_20 , with the origin attributed from the $X0$, which is considered atavistic and modal for the group [8,9,14,15,16,17].

Large portions of eukaryotic genomes are composed of sequences that are repeated hundred to thousand times, which are called repetitive DNAs. These sequences could be tandemly arrayed as microsatellites, minisatellites and satellite DNAs (satDNAs) or occur in a scattered pattern, as transposons and retrotransposons [18,19,20,21,22]. In accordance with the selfish DNA hypothesis [23,24], some of these sequences have been maintained in the genome due to their ability to colonize genomic regions lacking or with a low level of recombination activity. Thus, these repeats tend to be highly abundant and ubiquitous in certain chromosomal compartments, such as telomeric and centromeric heterochromatin and non-recombining regions of the sex chromosomes. Repetitive DNAs generally are subjected to replication and expansion processes through multiple mechanisms, such as unequal cross-over, gene conversion, rolling-circle replication, whole-genome duplication, segmental duplications and transposition [18,21,25,26,27,28,29,30]. Because the size and abundance of the repeats vary greatly within and between species, repetitive DNAs are the major cause of variation in the size of eukaryote genomes, and they have also been involved in genomic evolution [18,31,32,33].

Orthopteran species frequently have large genomes, ranging in size from 1.52 Gb in the cave cricket *Hadenoeus subterraneus* [34] to 16.56 Gb in the grasshopper *Podisma pedestris* [35]. In some Orthopteran species, the genomes are rich in repetitive DNAs, as for example, in *Locusta migratoria* (genome size 6.3 Gb [36]) and *Schistocerca gregaria* grasshoppers (genome size 8.55 Gb [37]). Although the sizes of the genomes of cricket species are generally smaller than those of grasshoppers, they have large genomes compared with those of other insects (see [38,39]); however, very little information concerning the chromosomal organization of the repetitive DNAs in this group is available (see for example [13,17]).

Considering the large genomes rich in repetitive DNAs of some Orthopteran species and the possible role of these sequences in chromosomal evolution, in this study, we aimed to contribute to the understanding of the organization of repetitive DNAs in autosomes and their contribution to the structure of the derived sex chromosomes in crickets. For this purpose, we used two species belonging to the family Gryllidae that have highly divergent karyotypes, i.e., *Gryllus assimilis*, with $2n(\sigma) = 29, X0$ [40], and *Eneoptera surinamensis*, with $2n(\sigma) = 9, neo-X_1X_2Y$ [12]. These genomes were compared using flow cytometric genome-size estimation, classical cytogenetic analysis and fluorescent *in situ* hybridization (FISH) using various repetitive DNAs as probes, such as multigene families (18S and 5S rDNA, U1 and U2 snDNAs and H3 histone), highly and moderately repetitive DNA fractions (C_0t -DNA), the classical insect telomeric repeat (TTAGG) and 16 microsatellite motifs.

Materials and Methods

Samples, classical chromosomal analysis and banding

Five male and seven female *Eneoptera surinamensis* were collected in the Parque Estadual Edmundo Navarro de Andrade (Rio Claro, SP, Brazil) between May 2013 and March 2014 with the authorization of COTEC (process number 341/2013) and were maintained in captivity until ovipositioning occurred, whereas *Gryllus assimilis* animals were obtained from a pool of individuals that had been bred in the biotery of the Univ. Estadual Paulista—UNESP (Rio Claro, SP, Brazil). The embryo preparations for chromosome obtaining were prepared according to Webb et al. [41], with slight modifications of the fixation materials, as follows: After hypotonization was accomplished, the same volume of Carnoy's modified solution (3:1, absolute ethanol:acetic acid) was applied for 15 minutes, then the embryos were transferred to fresh Carnoy's solution and were stored in a -20°C freezer until use. At least 60 embryos of each species were used for cytological preparations. In addition, four adult male testes of each species were dissected and were fixed in Carnoy's modified solution. Adult specimens (ten of each species) were stored in 100% ethanol for subsequent DNA extraction.

Karyological studies were performed using conventional staining with 5% Giemsa solution to confirm the previous karyotypic descriptions. The C-banding procedure was conducted according to Sumner [42], and fluorochrome staining (CMA₃/DA/DAPI) to identify G+C or A+T rich regions was performed as described by Schweizer et al. [43]. Female and male genomic DNA was extracted from femurs using the phenol/chloroform-based procedure described in Sambrook and Russel [44].

Probes for repetitive DNAs

The DNA probes for the 5S rRNA and H3 genes were obtained through polymerase chain reaction (PCR) amplification using genomic DNA from *Abracris flavolineata* as a template and the primers described by Cabral-de-Mello et al. [45] and Colgan et al. [46] for the 5S rRNA gene and the H3 gene, respectively. The sequences of the U snDNAs of *Rhammatocerus brasiliensis* were obtained using the primers described by Cabral-de-Mello et al. [47] for U1 snDNA and those of Bueno et al. [48] for U2 snDNA. These sequences were previously used as probes and are deposited in GenBank under accession numbers KC936996 (5S rDNA), KC896792 (H3 histone gene), KC896793 (U1 snDNA) and KC896794 (U2 snDNA). The 18S rDNA probe was obtained from a cloned fragment previously isolated from the genome of the *Dichotomius semisquamosus* beetle (GenBank accession number GQ443313 [45]).

The repetitive DNA-enriched samples were obtained based on the renaturation kinetics of *C₀t*-DNA (DNA enriched for highly and moderately repetitive DNA sequences), according to the protocol of Zwick et al. [49], with modifications [45]. It was denaturated 150 ng/ μL of fragmented genomic DNA at 95°C . The reassociation time of 25 min was used for both species. The DNA was purified/extracted using a traditional phenol-chloroform procedure [44].

The telomeric probe was obtained through PCR using the self-complementary primers (TTAGG)₅ and (CCTAA)₅ according to Ijdo et al. [50]. Finally, the probes for the microsatellite motifs were directly labeled at the 5' end with biotin-14 dATP (Sigma-Aldrich, St Louis, MO, USA) during their synthesis, including probes for mononucleotides (A)₃₀ and (C)₃₀, dinucleotides, (CA)₁₅, (CG)₁₅, (TA)₁₅ and (AG)₁₀, trinucleotides (CAA)₁₀, (CAC)₁₀, (TAA)₁₀, (GAA)₁₀, (CGG)₁₀, (CAG)₁₀, (TAC)₁₀ and (GAG)₁₀, tetranucleotides (GACA)₄ and (GATA)₈.

Fluorescence *in situ* hybridization (FISH)

The non-cloned 5S rDNA and U snDNAs sequences and telomeric probes were labeled through PCR using digoxigenin-11-dUTP (Roche, Mannheim, Germany). Plasmids containing the 18S rRNA gene or H3 histone gene and the *Cot*-DNA fraction were labeled via nick translation using biotin-14-dATP (Invitrogen, San Diego, CA, USA). Single or two-color FISH of mitotic cells was performed according to Pinkel et al. [51], with modifications [45]. Fiber-FISH was conducted as described in de Barros et al. [52] and Camacho et al. [53] using suspensions of testis cells. Probes labeled with digoxigenin-11-dUTP were detected using rhodamine-conjugated anti-digoxigenin (Roche), and probes labeled with biotin-14-dATP were detected using Streptavidin Alexa Fluor 488-conjugated (Invitrogen).

The preparations were counterstained using 4',6-diamidino-2'-phenylindole dihydrochloride (DAPI) and were mounted in Vectashield (Vector, Burlingame, CA, USA). The chromosomes and signals were observed using an Olympus microscope BX61 equipped with a fluorescence lamp and the appropriate filters. Grey-scale images were captured using a DP70 cooled digital camera and were processed using Adobe Photoshop CS2 software.

Genome size estimation

The flow cytometric (FCM) analyses were conducted as described by Lopes et al. [54], in the Laboratory of Cytogenetics and Cytometry, Department of General Biology, Universidade Federal de Viçosa (UFV). The nuclear DNA contents of adult *G. assimilis* and *E. surinamensis* males and females were determined using the C DNA content of a *Scaptotrigona xantotricha* female as an internal standard, which was confirmed by comparison with that of the international standard, *Drosophila melanogaster*.

To prepare the nuclear suspensions for FCM, brain ganglia were excised from the standard animal and the sample animals in physiological saline solution (0.155 mM NaCl). The samples consisted of three males and three females of each species, and each individual was manipulated and analyzed separately, representing three independent repetitions. The materials were simultaneously crushed 10–12 times in a tissue grinder using a pestle (Kontes Glass Company, NJ, USA) in 100 μ L of OTTO-I lysis buffer [55] containing 0.1 M citric acid (Merck, NJ, USA), 0.5% Tween 20 (Merck) and 50 μ g mL⁻¹ RNase A (Sigma-Aldrich), pH = 2.3. The suspensions were adjusted to 1.0 mL using the same buffer, filtered through a 30- μ m nylon mesh (Partec, Nuremberg, Germany) and centrifuged at 100 g in microcentrifuge tubes for 5 min. The pellets were then incubated for 10–15 min in 100 μ L of OTTO-I lysis buffer and were stained using 1.5 mL of OTTO-I:OTTO-II (1:2) solution [56,57] supplemented with 75 μ M propidium iodide (PI) (excitation/emission wavelengths: 480-575/550-740 nm, [58]) and 50 μ g mL⁻¹ RNase A (Sigma-Aldrich), pH = 7.8. The nuclear suspensions were filtered through a 20- μ m nylon mesh filter (Partec) and were maintained in the dark for 30 min.

The suspensions were analyzed using a Partec PAS flow cytometer (Partec) equipped with a laser source (488 nm). PI fluorescence emitted by the nuclei was collected using an RG 610-nm band-pass filter and converted to 1024 channels. The equipment was calibrated for linearity and was aligned using microbeads and standard solutions according to the manufacturer's recommendations. FlowMax software (Partec) was used for data analysis. The standard nuclear peak was set to channel 100 and more than 10,000 nuclei were analyzed. Three independent replications were conducted, and histograms with a coefficient of variation (CV) of greater than 5% were rejected.

Results

The karyotype of *Gryllus assimilis* is $2n = 29\sigma/30\varphi$, with a $X0\sigma/XX\varphi$ sex-chromosome system (Fig 1A). The autosomes consist of four pairs of metacentric (1, 2, 8 and 10) chromosomes,

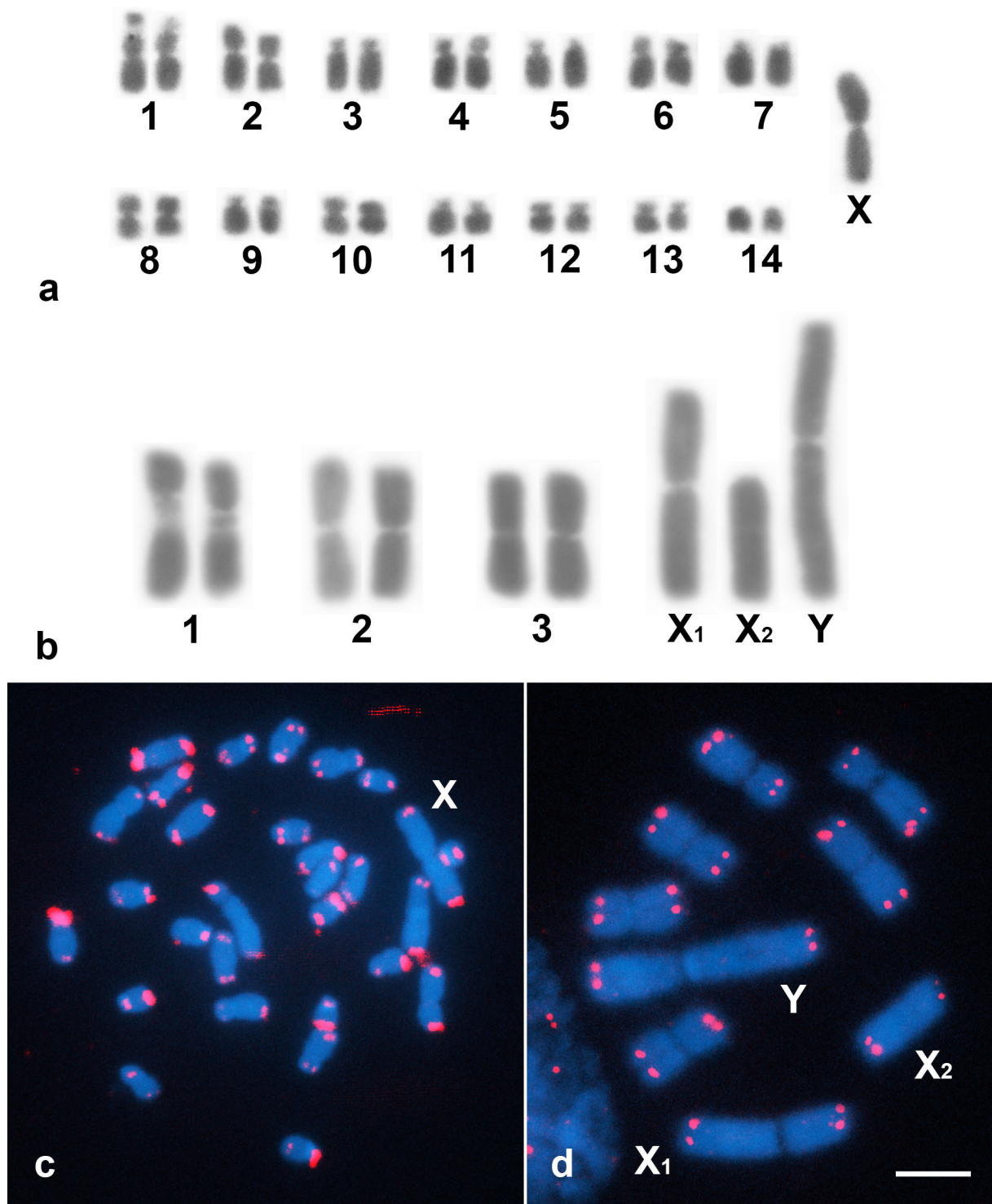


Fig 1. Male karyotypes of *Gryllus assimilis* (a) and *Eneoptera surinamensis* (b) observed using Giemsa staining. The chromosomes of both species were arranged in descending order of size. FISH analysis of the telomeric repeat in the mitotic cells of *G. assimilis* (c) and *E. surinamensis* (d). The sex chromosomes are indicated in the figure. Bar = 5 μ m.

doi:10.1371/journal.pone.0143540.g001

four pairs of submetacentric (4, 6, 12 and 13) chromosomes, five pairs of subtelocentric (3, 5, 7, 9 and 11) chromosomes and one pair of telocentric (14) chromosomes. The X chromosome is metacentric and is the largest chromosome of the karyotype. One evident interstitial secondary constriction was observed in the metacentric pair of chromosome 1 (Fig 1A). Analysis of *E. surinamensis* males and females revealed a karyotype of $2n = 9\sigma/10\text{♀}$ with a neo- $X_1X_2Y\sigma/X_1X_1X_2X_2\text{♀}$ sex-chromosome system. Autosomal pairs 2 and 3 are metacentric chromosomes, whereas pair 1 are submetacentric chromosomes with a conspicuous proximal secondary constriction localized on the long arm. The neo- X_1 chromosome is metacentric, the neo- X_2 chromosome is telocentric and the neo-Y chromosome, which is submetacentric, is the largest sex chromosome (Fig 1B). These results are similar to the previously described results for both species [12,40]. In both species, only terminal sites of the autosomes and sex chromosomes were recognized using the telomeric probe (Fig 1C and 1D).

The C-positive regions were observed in distinct areas for the distinct chromosomes of *G. assimilis*, with more frequent terminal blocks on the short chromosomal arms and in the pericentromeric regions. The X chromosome harbored only C-positive terminal blocks in both arms. The secondary constriction of chromosome 1 was intensely stained via C-banding (Fig 2A). In *E. surinamensis*, all of the chromosomes except the neo- X_2 chromosome exhibited pericentromeric C-positive bands. The neo- X_2 chromosome had a faint interstitial C-positive band, and the pair 1 chromosomes exhibited a heterochromatic block coincident with the secondary constriction. Additionally, some dispersed C-positive blocks were observed in the neo-Y chromosomes (Fig 2D). Most of the C-positive blocks in the *G. assimilis* chromosomes were neutral for G+C or A+T base pairs. A remarkable number of G+C positive blocks (CMA_3^+) were observed only in the secondary constriction of the chromosome pair 1 and in the terminal region of the X chromosome, although faint CMA_3^+ signals were also observed in a few other chromosomes, in terminal or interstitial positions (Fig 2B). In the *E. surinamensis* chromosomes, all of the C-positive blocks were CMA_3^+ and the neo- X_1 and neo- X_2 chromosomes exhibited C-negative blocks rich in G+C base pairs, which were interstitial in the short arm and in the pericentromeric region, respectively (Fig 2E). No DAPI positive blocks were observed (result not shown). In both species, C-positive regions were also identified in the C_{0t} -DNA fractions (Fig 2C and 2F). Additionally, in the *E. surinamensis* cells, faint signals were observed distributed along some of the chromosomes in the C-negative regions, which include the pericentromeric area of the neo- X_2 chromosome (Fig 2F).

Labeling of the mono-, di-, tri and tetra-nucleotide arrays in *G. assimilis* cells revealed a relatively uniform dispersed distribution of each on all of the chromosomes (Fig 3A–3F). Only the $(\text{CGG})_{10}$ signal was less intense in the centromeric regions (Fig 3D). Although scattered signals for these arrays were generally observed in *E. surinamensis* cells, some of the signals formed band-like patterns in distinct chromosomes and in distinct positions, depending on the repeat mapped, with the exception of the $(\text{CAA})_{10}$ and $(\text{GATA})_8$ arrays, which showed only a scattered distribution (Fig 3G–3J, Table 1).

The FISH analyses revealed 18S rDNA and U2 snDNA signals on the pair 1 of *G. assimilis*, which were co-localized in the arm containing the secondary constriction (Fig 4A and 4B), whereas signals for 5S rDNA and U1 snDNA clusters were observed in the pericentromeric and proximal areas, respectively, of small subtelocentric chromosome pairs; however, due to slight differences in these chromosomes, it was not possible to precisely identify them (Fig 4C and 4D). The H3 histone signal occurred in evident bands in the terminal regions of some of the autosomes, whereas no H3 histone signal was observed on the X chromosome (Fig 4E). In *E. surinamensis* cells, labeling for all of the multigene families revealed clusters in the pair 1, and the 18S rDNA, U2 snDNA and H3 histone signals were coincident with the secondary constriction (Fig 4F, 4G and 4I), whereas the 5S rDNA and U1 snDNA signals were observed

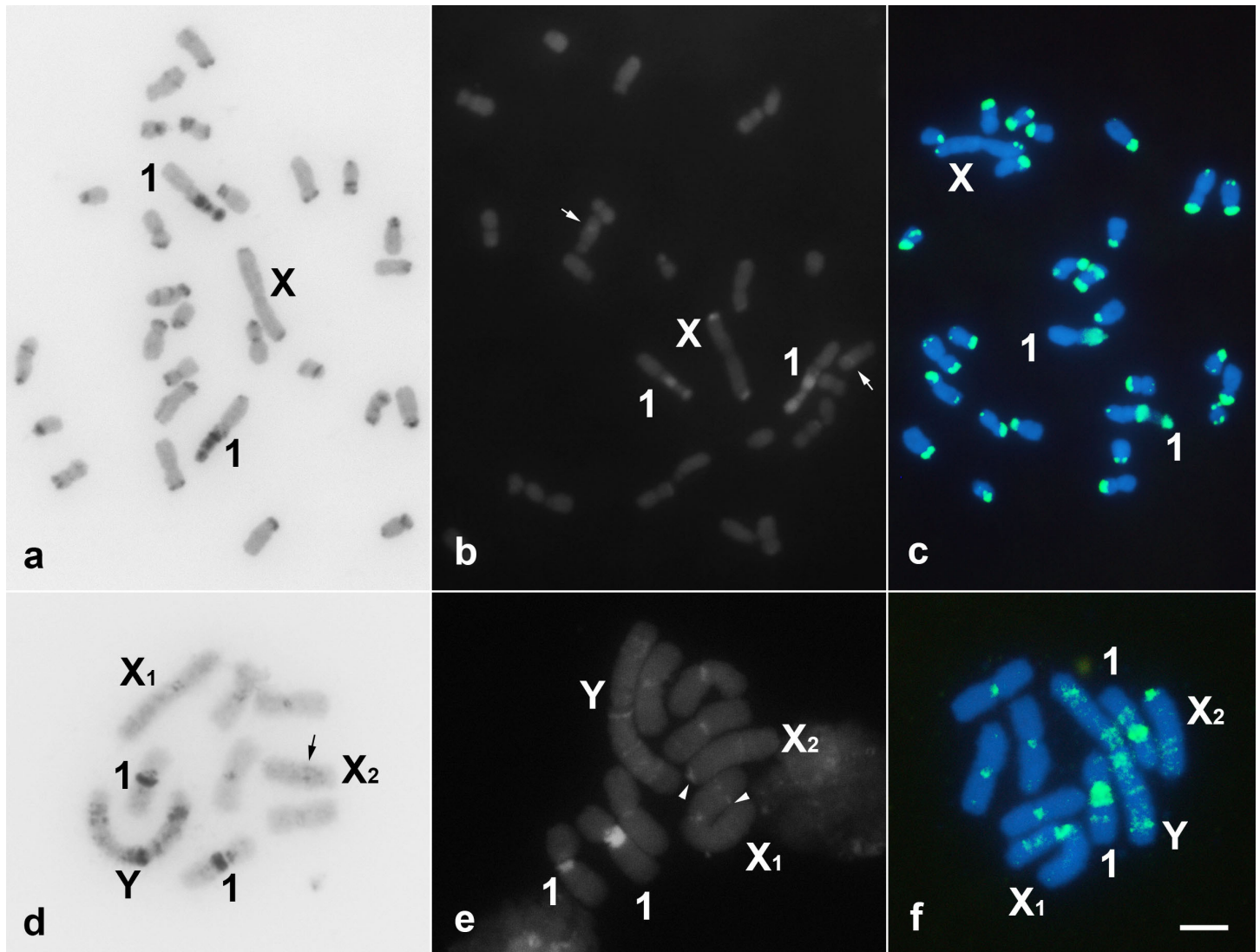


Fig 2. C-banding (a, d), CMA₃ fluorochrome staining (b, e) and FISH of the C₀t-DNA (c, f) on mitotic chromosomes of *G. assimilis* (a, b, c) and *E. surinamensis* (d, e, f) cells. The sex chromosomes and the chromosomes bearing the secondary constrictions (pair 1) are indicated in each of the panels. Note the predominance of terminal C-positive blocks in *G. assimilis* and *E. surinamensis* chromosomes, the neo-Y chromosome with some heterochromatic blocks and CMA₃⁺ signals along its entire length, as well as the C₀t-DNA signals. In (b), the white arrow indicates the interstitial CMA₃⁺ block, and in (d), the black arrow indicates the interstitial C-positive block in the neo-X₂ chromosome. In (e) arrowheads indicate the CMA₃⁺ signals in C-negative regions of the neo-X₁ and neo-X₂. One chromosome is missing in (b). Bar = 5 μm.

doi:10.1371/journal.pone.0143540.g002

on the short arm (Fig 4F and 4H). However, an additional cluster of 5S rDNA labeling was observed in the interstitial region of the long arm of the neo-Y chromosome (Fig 4H), and a faint H3 histone signal was observed along all of the chromosomes, which was more evident in the neo-Y chromosome (Fig 4I). Fiber-FISH using probes for 18S rDNA and U2 snDNA, which were co-localized in the secondary constriction of the pair 1 chromosomes in metaphasic chromosomes in both species, revealed that these two sequences are interspersed in similar patterns (Fig 5).

The FCM analyses revealed that the mean genome size (1C) of *G. assimilis* cells is 2.13 pg (male: 2.06 pg and female: 2.21 pg), whereas *E. surinamensis* presents approximately 2.5 times greater DNA content, with a mean value of 5.54 pg (male: 5.44 pg and female: 5.65 pg)

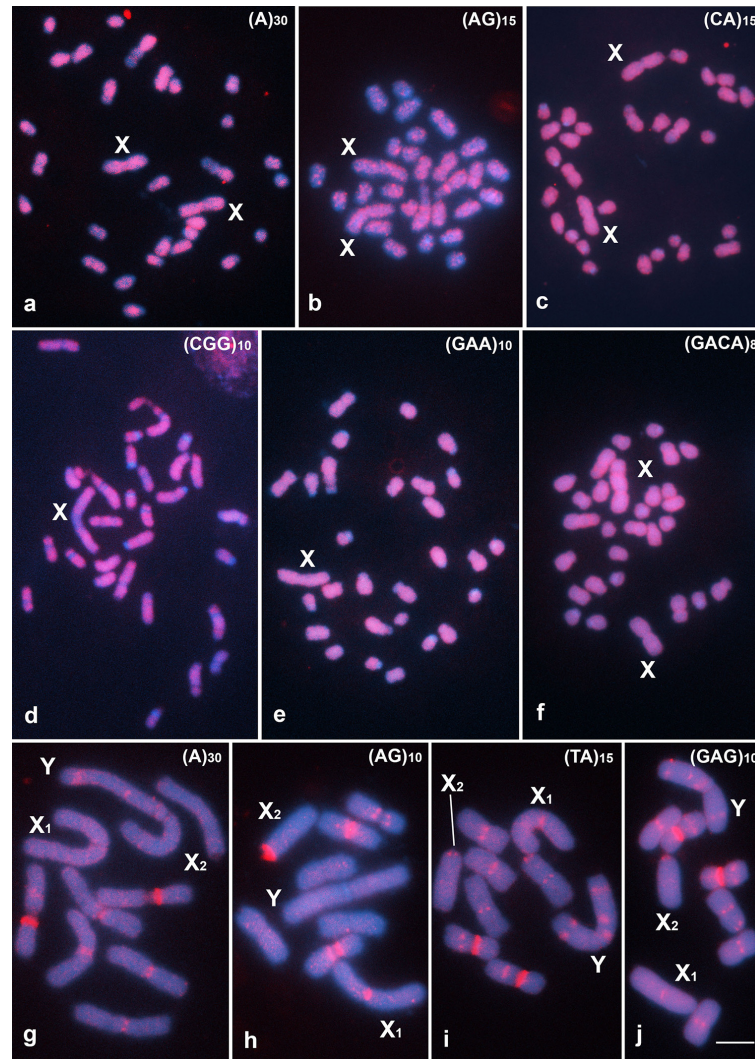


Fig 3. Mitotic metaphase cells of *G. assimilis* (a-f) and *E. surinamensis* (g-j), demonstrating examples of the hybridization patterns of different microsatellite-motif probes. The sex chromosomes and the signals for each type of microsatellite motif are shown in the images. Note the intense signals along the chromosomal arms of both species, although band-like signals are also evident in distinct chromosomes and in distinct positions in *E. surinamensis* cells. One chromosome is missing in (b). Bar = 5 μ m.

doi:10.1371/journal.pone.0143540.g003

(see [S1 Fig](#)). Thus, the mean genome size of *G. assimilis* is 2.08 Gb and that of *E. surinamensis* is 5.42 Gb. The greater size of the genomes of the females of both species is certainly due to the presence of two copies of the X chromosome in female *G. assimilis* cells and the presence of two copies for the X₁ and X₂ chromosomes in female *E. surinamensis* cells.

Discussion

Karyotypes and general organization of repetitive DNA

The two cricket species examined in this study have highly divergent karyotypes. Considering that fusions are more frequent in Orthopteran chromosomal evolutionary history (for example, see [\[8,9,59\]](#)), leading to a reduced diploid number, the karyotype of *E. surinamensis* could be considered more evolved relative to that of *G. assimilis*. Repetitive DNAs can be relevant

Table 1. Chromosomal positions of the band-like signals for microsatellite arrays in the cricket *Eneoptera surinamensis*.

Microsatellite motif	Pair 1	Pair 2	Pair 3	X ₁	X ₂	Y
(A) ₃₀	S: p	c	c			L: i, sd
(C) ₃₀	S: p			c	c	S: p, sd
(AG) ₁₀	S: p; L: p			c	c	
(CA) ₁₅	S: p			c	c	L: i, sd
(CG) ₁₅	c; L: 2i			c	c	c; L: 2p, i, sd; S: 2p, sd
(TA) ₁₅	c; L: i	c, d	c	c	c	L: p, i, sd; S: sd
(CAC) ₁₀					S: sd	L: 2p
(CAG) ₁₀	c	c	c	c	c	S: p; S: sd
(CGG) ₁₀	c			c	c	c; L: p
(TAC) ₁₀	c; L: 2i			c	c	c; L: 2p, sd; S: p, sd
(GAG) ₁₀	c; L: i	c	c, d	c	c	L: p, i, sd; S: sd
(GAA) ₁₀	c			c	c	S: p, sd
(TAA) ₁₀	c			c	c	
(GACA) ₄	c; L: p			c	c	c; L: 2p, sd; S: p, sd

L = long arm; S = short arm; c = centromeric; p = proximal; i = interstitial; sd = subdistal; d = distal.

doi:10.1371/journal.pone.0143540.t001

markers for tracing the evolution of the karyotypes and genomes of eukaryotes, but there is little information concerning the sizes of cricket genomes or the organization of their repetitive DNAs and their possible role in chromosomal evolution [13,17,38,39].

The diploid number and the sex system of *G. assimilis* observed in this study is identical to that previously described [40] and resembles that of most of the other *Gryllus* species that have been analyzed [10,13,60,61,62]. However, reduction of the diploid number has been observed in this genus [10], which includes the polymorphic condition of *G. assimilis* [11]. The main difference observed here in relation to other *Gryllus* karyotypes is the chromosomal morphology and the chromosome bearing the evident secondary constriction (see above references), suggesting that although *Gryllus* exhibits great karyotypic stability, small structural rearrangements may have occurred during the evolution of the genus. The diploid number observed in *E. surinamensis* (2n = 9) is well established in this species, having been observed in certain populations [63], and considering that a higher number of diploid chromosomes is more common in Gryllidae crickets [9,12,64], the karyotype of this species appears to be highly rearranged. Autosomal and autosomal/sex chromosome reshuffling most likely caused the significant reduction in the diploid number and the origin of a multiple sex system (neo-X₁X₂Y) with large chromosomes. It is difficult to specify the evolutionary causes or the order of the specific chromosomal rearrangements that led to the origin of the extreme karyotype of *E. surinamensis*, due to its extensive reorganization and the scarce phylogenetic and chromosomal information available for crickets. However, Robertsonian translocations (Rb-translocation), tandem fusions and pericentric inversions are most likely involved. In all cases of Rb-translocations or tandem fusions, the telomeres appeared to be either lost during the chromosomal rearrangement or eliminated during chromosome differentiation and no internal telomere sites are observed, as indicated by our FISH-based mapping of the telomere motif, which revealed signals only in the actual telomeres. A similar condition was reported in a small number of studied grasshoppers with derived karyotypes in which centric fusion or Rb-translocation occurred [15,16]. However, due to the limited sensitivity of the classical FISH technique, the possibility of small internal telomeric sites cannot be completely ruled out. In *Gryllus*, if the karyotype experienced inversions, as justified by distinct chromosomal morphologies relative to those of

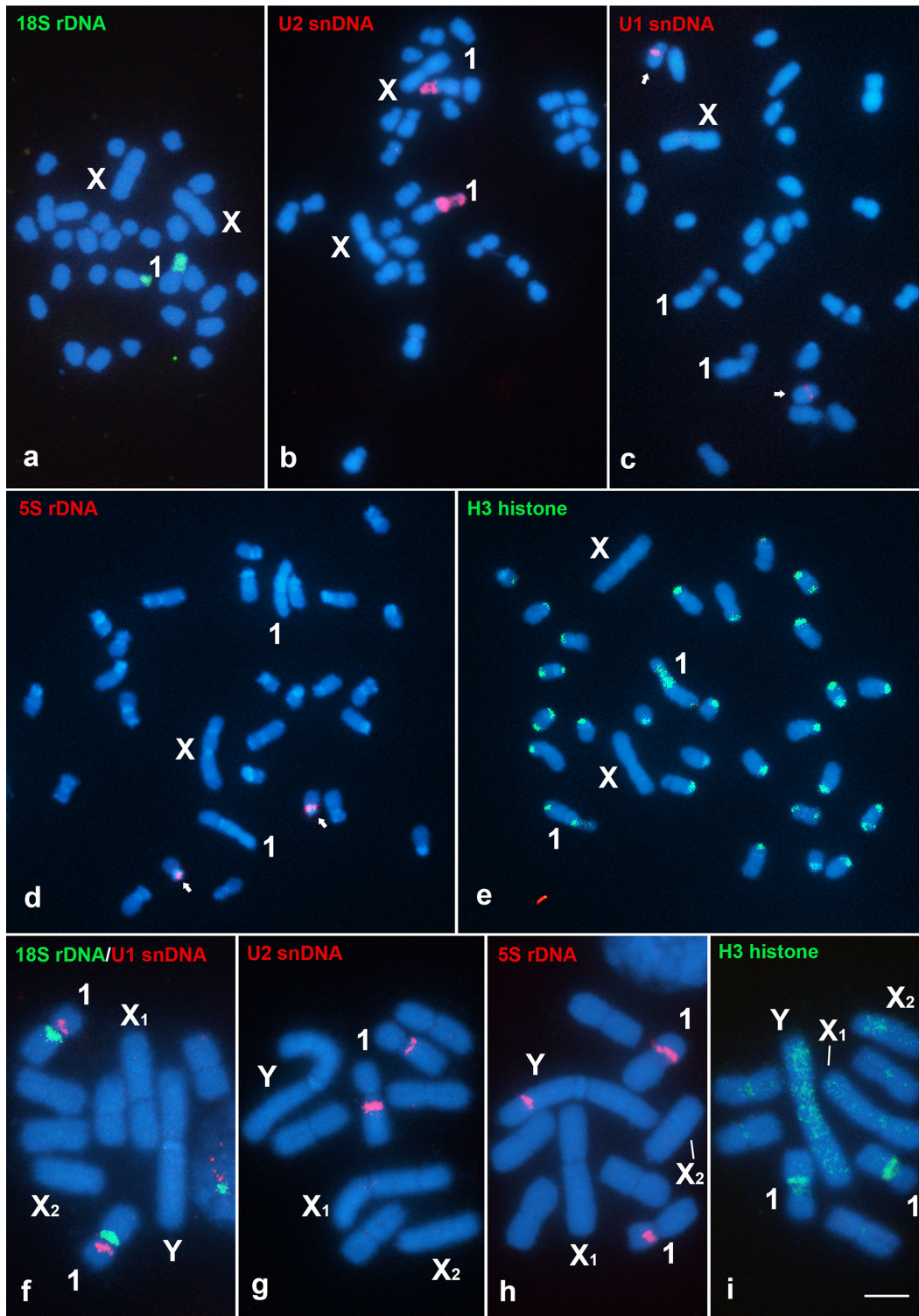


Fig 4. Chromosomal mapping of multigene families in mitotic cells of *G. assimilis* (a-e) and *E. surinamensis* (f-i). The sex chromosomes and the signals for each type of probe are shown in the images. Note the same chromosomal localization (pair 1) of 18S rDNA and U2 snDNA in both species and the signals for all of the multigene families in the pair 1 of *E. surinamensis*. Also, note the multiple and scattered signals of the H3 histone gene (e, i). The arrows in (c, d) indicate the chromosomes bearing signals. Bar = 5 μ m.

doi:10.1371/journal.pone.0143540.g004

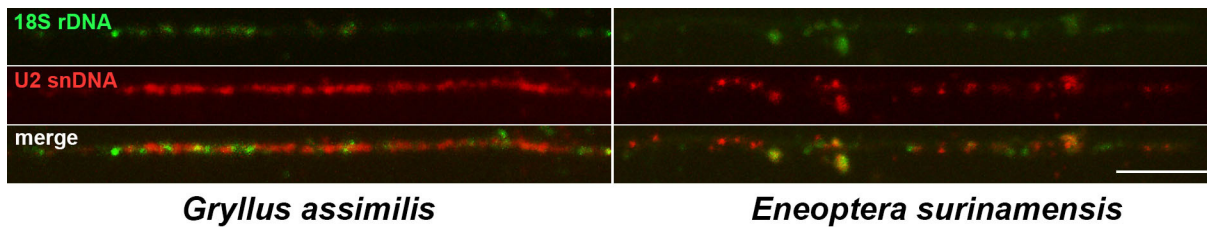


Fig 5. Fiber-FISH of the 18S rDNA and U2 snDNA probes in *G. assimilis* (a) and *E. surinamensis* (b). Note that these probes are co-localized and are interspersed in both species.

doi:10.1371/journal.pone.0143540.g005

other species, these rearrangements occurred without the involvement of telomeres or the telomeres were lost. An alternative explanation for this condition is that *G. assimilis* has a non-rearranged karyotype compared with those of other species of the genus, which should be evaluated using a better species sampling.

There is little information concerning heterochromatin distribution in crickets, but in the genus *Gryllus*, the occurrence of mainly centromeric and terminal heterochromatic blocks has been observed in certain species, i.e., *G. bimaculatus*, *G. argentinus* and *Gryllus* sp. [13,60,62], as observed in this study in *G. assimilis*. In contrast, in *E. surinamensis*, the C-positive blocks were most organized in the centromeric region of the autosomes, as well as in other chromosomal regions, with extensive spreading in the neo-Y chromosome. These patterns, including those observed in *Cycloptiloides americanus* [17], suggest the intense reorganization of the C-positive blocks in the rearranged karyotypes. These data indicate that the heterochromatin chromosomal dynamic is more intense in crickets than in other Orthoptera, such as grasshoppers, in which pericentromeric C-positive blocks are most commonly observed [65,66], although other species should be studied. The data obtained through fluorochrome staining suggests the occurrence of repetitive DNA families with distinct richness in A+T or G+C base pairs, which in *E. surinamensis* are G+C-rich, whereas in *G. assimilis*, they are mainly neutral for A+T or G+C, suggesting the restructuring of repetitive DNA families in their two karyotypes. This data was confirmed in *E. surinamensis* through genomic analysis that revealed the occurrence of mainly G+C-rich satDNAs (Palacios-Gimenez et al., unpublished data). In another *Gryllus* species, i.e., *G. bimaculatus*, some terminal heterochromatic blocks were enriched for two A+T-rich satDNAs families, one of which is also present in other *Gryllus* species, including *G. rubens* and *Gryllus* sp [13]. The occurrence of A+T-rich repetitive families in the genome of *G. assimilis* could not be ruled out, and DAPI⁺ blocks could not be observed in this organism due to the limits of resolution.

The chromosomal organization of repetitive DNAs obtained through C-banding was corroborated by *C₀t*-DNA mapping that revealed a similar pattern, as well as additional regions enriched in repetitive sequences, such as the centromere of the neo-X₁ chromosome of *E. surinamensis* and some faintly labelled dispersed areas. In other Orthopteran species, the *C₀t*-DNA fraction is also enriched in C-positive autosomal blocks, but divergent patterns have been observed in the sex chromosomes, including the derivate neo-sex chromosomes, as observed in *E. surinamensis*. For example, enrichment of repetitive DNAs in the neo-Y element of the grasshopper *Ronderosia bergi* [16] and the X₂ element of the cricket *C. americanus* [17] have been reported, suggesting a role of this genomic fraction in sex chromosome differentiation, causing chromatin to change into heterochromatin. The accumulation of repetitive DNAs is a common feature during the differentiation of the Y or W sex chromosomes, and it has been documented in various animal and plant species, such as *Rumex* species [27,67], *Silene latifolia* [30], lizard species [68], *Schistosoma mansoni* [69] and *Drosophila miranda* [70,71]. However,

in some grasshopper species, the C_{0t} -DNA fraction is restricted to the centromeres of the sex chromosomes [15].

The FISH mapping results regarding the microsatellites suggest that a major strong differentiation in the chromosomal organization of the karyotypes of the two species occurred. Although the repeats are mainly dispersed in both euchromatin and heterochromatin, clustering was observed, with the repeats forming a band-like pattern, particularly those in the heterochromatic regions of the pair 1 and sex chromosomes of *E. surinamensis*. This result suggests that the reduction in the diploid number was followed by the “compartmentalization” of the microsatellites in the genome, with these sequences being involved in the heterochromatin, including those of the sex chromosomes. The accumulation of some of the microsatellite arrays in the neo-Y chromosomes of *E. surinamensis* was expected, considering that they are non-recombining elements, as suggested by the lack of contact of these three chromosomes during meiosis [8], and this data is consistent with that observed in other species [72,73,74,75]. As previously observed in the grasshopper *R. bergi*, the microsatellites of *E. surinamensis* are clearly involved in the differentiation between sex chromosomes [16]. Although poorly studied in Orthoptera, the dispersal or specific chromosomal distribution of the microsatellite arrays in band-like patterns was recently reported in other species, such as the cricket *C. americanus* [17] and the grasshoppers *Abracris flavolineata* [76], *R. bergi* [16], *Locusta migratoria* and *Eyrepocnemis plorans* [77]. However, next generation sequencing analysis revealed uneven and nonrandom localization of microsatellites in *L. migratoria* and *E. plorans*, with the dinucleotide motifs predominantly associated with other repetitive DNAs, such as the histone gene spacer, rDNA intergenic spacers (IGSs) and transposable elements [77]. A similar pattern was observed in *E. surinamensis*, in which some of the microsatellite arrays that are distributed in a band-like pattern are co-localized with some of the multigene families and the anonymous C_{0t} -DNA fraction.

The increased size of the genome of *E. surinamensis* compared with that of *G. assimilis* is notable, and considering the little information available concerning the size of the genomes of crickets [38,39], this is apparently a derivative characteristic. The increased size of this genome could be attributed to the amplification, spreading and accumulation of repetitive DNAs in specific chromosomal regions, such as observed for C_{0t} -DNA and some satellite repeats (Palacios-Gimenez et al., unpublished data), besides some microsatellites. Although, there is no evidence that the reduction in the diploid number contributed to this process. This process is more evident in the derivative sex chromosomes, most likely due to the non-recombinant state of the neo-Y chromosome and the restricted recombination of the neo-X₁ and neo-X₂ chromosomes in the female germline. Additionally, the structural mutation rates of the neo-Y chromosomes of some species are more rapid than those of other genomic regions, suggesting the reduced efficacy of natural selection in this chromosome [18,78,79,80], which is likely to occur in the genome of *E. surinamensis*.

The multigene families

No variation in the location and number of clusters of 18S rDNA and U2 snDNA was observed, both of which are conserved on the pair 1. In grasshoppers [81,82] and other insects, such as Lepidoptera [83], Coleoptera [84] and Heteroptera [85], extremely variable patterns of the major rDNA was observed contrasting with this study. Remarkably, the pattern of U2 snDNA on the pair 1 is highly conserved in some grasshopper species [15,48] and in the cricket *C. americanus* (Mogoplistidae) [17]. Conservation of the number of U2 snDNA clusters is observed also in other groups, such as *Gymnotus* species [86], in contrast to the case in some of the other fish species of the Batrachoididae family, in which the U2 snDNA signals appear to

be widely scattered [87]. The 18S rDNA and U2 snDNA signals were located in the same sites and exhibited the same association in the two species, which was supported by the fiber-FISH results, which may indicate an ancestral characteristic in Gryllidae that was maintained despite the exceptional chromosomal divergence. However, other species in the family should be studied to confirm this hypothesis. The association or interspersion of distinct multigene families has been reported in distinct groups, but the significance of these patterns is not clear and they have no apparent selective advantage [45,82,84,88,89,90,91].

The presence of 5S rDNA and U1 snDNA in the small chromosomes of *G. assimilis* in contrast to their presence in the pair 1 of *E. surinamensis* may be the consequence of large-scale chromosomal rearrangements, as mentioned above, which translocated these genes to the same chromosome. However, the localization of 5S rDNA in the neo-Y chromosome of *E. surinamensis* suggests both amplification and transposition or even further chromosomal rearrangement involving sex chromosomes and ancestral autosomes bearing this gene, as has been suggested for other Orthopteran species that have neo-sex chromosomes [15,16,17]. The presence of this gene in the neo-Y chromosomes and absence in the neo-X chromosomes reinforce the differentiation between the neo-sex chromosomes of *E. surinamensis* that was established after the chromosomal rearrangements occurred.

Contrary to what was observed for rDNAs and U snDNAs, significant dispersion of the H3 histone gene occurred in *G. assimilis* and *E. surinamensis*, with differentially spread patterns observed in the two species, with these DNA elements clearly in blocks in *G. assimilis* and faint signals observed in *E. surinamensis*. These results suggest the independent amplification/dispersion and dynamism of the H3 histone genes at the chromosomal level. Multiple H3 histone sites, in blocks such as observed in *G. assimilis*, have rarely been reported in grasshoppers, such as *Abracris flavolineata* [48], *Dichromatos lilloanus* [15] and *Ronderosia bergi* [16]. Although, this pattern is less common, being commonly observed an interstitial cluster in the pair 8 a highly conserved character [92]. Our data suggest that the organization and evolution of the H3 histone genes may be more dynamic than was previously reported and that this dispersal in *G. assimilis* and *E. surinamensis* could be due to other mechanisms, such as association with other repetitive DNAs (transposons or satDNAs), ectopic recombination or extrachromosomal circular DNA (eccDNA), as has been proposed for rDNAs [19,81,93]. An alternative hypothesis for the dispersal of H3 histone genes is the similarity of the sequences of these genes with those of unknown repetitive DNAs. Remarkably, a main cluster of H3 histone genes occurs in the pair 1, as is the case for the other four multigene families investigated in this study. This arrangement of certain multigene families in one unique bivalent chromosome has not yet been described and could be result of the large-scale rearrangements observed in the *E. surinamensis* karyotype.

In conclusion, our study provided an opportunity to explore the evolutionary dynamics of repetitive DNAs in two non-model species and its possible involvement in karyotypic organization and genome increasing. As we suggested some of the differences between the two evaluated karyotypes were most likely a consequence of the evolution and distribution of repetitive DNA that were accumulated in specific chromosomal regions. Whereas certain characteristics that these species shared could indicate their ancestral relationship. Moreover, the distinct patterns of the classes of repetitive sequences that were mapped in this study suggests that the differential amplification and accumulation of this class of DNA occurred mainly in the neo-Y chromosomes, highlighting the differentiation of the neo-sex chromosomes of *E. surinamensis*.

Supporting Information

S1 Fig. (a,b) Examples of histograms showing the relative nuclear DNA content from cerebral ganglion cells stained with PI for *S. xantotricha* (internal standard), *G. assimilis*

and *E. surinamensis*. (c) Genome size estimation in three replicates form each sex for species studied in this paper and for the female of *S. xantotricha* used as internal standard. The values are in pictogram (pg).

(PDF)

Acknowledgments

The authors are grateful to the “Parque Estadual Edmundo Navarro de Andrade” administration for sample collecting authorization, to Antonio Sergio Pascon for technical assistance in obtaining embryos and to Edison Zefa and Luciano Martins for the taxonomic identification of *Gryllus assimilis*. OMPG acknowledges the scholarship obtained from FAPESP (process number 2014/02038-8). DCCM was the recipient of a research productivity fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (process number 304758/2014-0).

Author Contributions

Conceived and designed the experiments: DCCM OMPG CRC FAFS. Performed the experiments: OMPG FAFS. Analyzed the data: DCCM OMPG CRC FAFS. Contributed reagents/materials/analysis tools: DCCM CRC. Wrote the paper: DCCM OMPG CRC.

References

1. Grimaldi D, Engel MS. Evolution of the insects. Cambridge University Press, New York; 2005.
2. Eades DC, Otte D, Cigliano MM, Braun H. Orthoptera species file. Version 5.0/5.0. [21/6/2015]. Available: <http://Orthoptera.speciesfile.org>.
3. Song H, Amédégno C, Cigliano MM, Desutter-Grandcolas L, Heads SW, Huang Y, et al. 300 million years of diversification: elucidating the patterns of orthopteran evolution based on comprehensive taxon and gene sampling. Cladistics. 2015 Feb 4. doi: [10.1111/cla.12116](https://doi.org/10.1111/cla.12116)
4. Warchalowska- liwa E. Karyotype characteristics of katydid Orthopterans (Ensifera, Tettigoniidae) and remarks on their evolution at different taxonomic levels. Folia Biol. 1998; 4: 143–176.
5. Ferreira A, Mesa A. Cytogenetics studies in Brazilian species of Pseudophyllinae (Orthoptera: Tettigoniidae): $2n(\sigma) = 35$ and $FN = 35$ the probable basic and ancestral karyotype of the family Tettigoniidae. Neotr Entomol. 2010; 39: 590–594.
6. Grzywacz B, Heller KG, Lehmann AW, Warchalowska- liwa E, Lehmann GUC. Chromosomal diversification in the flightless Western Mediterranean bush cricket genus *Odontura* (Orthoptera: Tettigoniidae: Phaneropterinae) inferred from molecular data. J Zoolog Syst Evol Res. 2014; 52: 109–118.
7. Sáez FA. An extreme karyotype in an orthopteran insect. Amer Nat. 1957; 91: 259–264.
8. White MD. Animal cytology and evolution. 3re ed. Cambridge Univ. Press, London; 1973.
9. Hewitt GM. Grasshoppers and crickets. Animal cytogenetics. vol 3: Insecta 1. Orthoptera. Gebrüder Borntraeger, Berlin; 1979.
10. Handa SM, Mittal OP, Sehgal S. Cytology of ten species of crickets from Chandigarh (India). Cytologia. 1985; 50: 711–72.
11. Zefa E. Autosomal rearrangement in *Gryllus assimilis* Fabricius, 1775 (Orthoptera, Gryllidae). Genet Mol Biol. 1999; 22: 333–336.
12. Ferreira A, Cella DM. Chromosome structure of *Eneoptera surinamensis* (Orthoptera, Grilloidea, Eneopterinae) as revealed by C, NOR and N-banding techniques. Chromosome Sci. 2006; 9: 47–51.
13. Yoshimura A, Nakata A, Mito T, Noji S. The characteristics of karyotype and telomeric satellite DNA sequences in the cricket, *Gryllus bimaculatus* (Orthoptera, Gryllidae). Cytogenet. Genome Res. 2006; 112: 329–336. PMID: [16484791](https://pubmed.ncbi.nlm.nih.gov/16484791/)
14. Castillo ER, Martí DA, Bidau CJ. Sex and neo-sex chromosomes in Orthoptera: a review. J Orthopt Res. 2010; 19: 213–231.
15. Palacios-Gimenez OM, Castillo ER, Martí DA, Cabral-de-Mello DC. Tracking the evolution of sex chromosome systems in Melanoplinae grasshoppers through chromosomal mapping of repetitive DNA sequences. BMC Evol Biol. 2013; 13: 167. doi: [10.1186/1471-2148-13-167](https://doi.org/10.1186/1471-2148-13-167) PMID: [23937327](https://pubmed.ncbi.nlm.nih.gov/23937327/)

16. Palacios-Gimenez OM, Marti DA, Cabral-de-Mello DC. Neo-sex chromosomes of *Ronderosia bergi*: insight into the evolution of sex chromosomes in grasshoppers. *Chromosoma*. 2015 Jan 21. doi: [10.1007/s00412-015-0505-1](https://doi.org/10.1007/s00412-015-0505-1)
17. Palacios-Gimenez OM, Cabral-de-Mello DC. Repetitive DNA chromosomal organization in the cricket *Cycloptiloides americanus*: a case of the unusual X₁X₂0 sex chromosome system in Orthoptera. *Mol Genet Genomics*. 2015; 290: 623–631. doi: [10.1007/s00438-014-0947-9](https://doi.org/10.1007/s00438-014-0947-9) PMID: [25373534](https://pubmed.ncbi.nlm.nih.gov/25373534/)
18. Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*. 1994; 371: 215–220. PMID: [8078581](https://pubmed.ncbi.nlm.nih.gov/8078581/)
19. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet*. 2005; 39: 121–52. PMID: [16285855](https://pubmed.ncbi.nlm.nih.gov/16285855/)
20. Biémont C, Vieira C. Genetics: junk DNA as an evolutionary force. *Nature*. 2006; 443: 521–524. PMID: [17024082](https://pubmed.ncbi.nlm.nih.gov/17024082/)
21. Richard GF, Kerrest A, Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev*. 2008; 72: 686–727. doi: [10.1128/MMBR.00011-08](https://doi.org/10.1128/MMBR.00011-08) PMID: [19052325](https://pubmed.ncbi.nlm.nih.gov/19052325/)
22. López-Flores I, Garrido-Ramos MA. The repetitive DNA content of eukaryotic genomes. In: Garrido-Ramos MA editor. *Repetitive DNA*. Genome Dyn. Basel, Karger, vol 7; 2012. pp. 1–28. doi: [10.1159/000337118](https://doi.org/10.1159/000337118) PMID: [22759811](https://pubmed.ncbi.nlm.nih.gov/22759811/)
23. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. *Nature*. 1990; 346: 601–603. PMID: [6245369](https://pubmed.ncbi.nlm.nih.gov/6245369/)
24. Orgel LE, Crick FHC. Selfish DNA: the ultimate parasite. *Nature*. 1980; 284: 604–607. PMID: [7366731](https://pubmed.ncbi.nlm.nih.gov/7366731/)
25. Charlesworth B. The evolution of chromosomal sex determination and dosage compensation. *Curr Biol*. 1996; 6: 149–162. PMID: [8673462](https://pubmed.ncbi.nlm.nih.gov/8673462/)
26. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*. 2003; 423: 825–837. PMID: [12815422](https://pubmed.ncbi.nlm.nih.gov/12815422/)
27. Navajas-Pérez R, de la Herrán R, Jamilena M, Lozano R, Ruiz Rejón C, Ruiz Rejón M, Garrido-Ramos MA (2005). Reduced rates of sequence evolution of Y-linked satellite DNA in *Rumex* (Polygonaceae). *J Mol Evol*. 2005; 60: 391–399. PMID: [15871049](https://pubmed.ncbi.nlm.nih.gov/15871049/)
28. Eickbush TH, Eickbush DG. Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics*. 2007; 175: 477–485. PMID: [17322354](https://pubmed.ncbi.nlm.nih.gov/17322354/)
29. Matsunaga S. Junk DNA promotes sex chromosome evolution. *Heredity*. 2009; 102: 525–526. doi: [10.1038/hdy.2009.36](https://doi.org/10.1038/hdy.2009.36) PMID: [19337304](https://pubmed.ncbi.nlm.nih.gov/19337304/)
30. Kejnovsky E, Hobza R, Cermak T, Kubat Z, Vyskot B. The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity*. 2009; 102: 533–541. doi: [10.1038/hdy.2009.17](https://doi.org/10.1038/hdy.2009.17) PMID: [19277056](https://pubmed.ncbi.nlm.nih.gov/19277056/)
31. Kidwell MG, Lisch D. Transposable elements as sources of variation in animals and plants. *Proc Natl Acad Sci USA*. 1997; 94: 7704–7711. PMID: [9223252](https://pubmed.ncbi.nlm.nih.gov/9223252/)
32. Gregory TR. *The evolution of the genome*. Elsevier Academic Press, San Diego CA USA; 2005.
33. Lynch M. *The origins of genome architecture*. Sinauer Associates, Inc. Publishers, Sunderland, MA, USA; 2007.
34. Rasch EM, Rasch RW. Cytophotometric determination of genome size for two species of cave crickets (Orthoptera, Rhaphidophoridae). *J Histochem Cytochem*. 1991; 29: 885.
35. Westerman M, Barton NH, Hewitt GM. Differences in DNA content between two chromosomal races of the grasshopper *Podisma pedestris*. *Heredity*. 1987; 58: 221–228.
36. Wang X, Fang X, Yang P, Jiang X, Jiang F, Zhao D, et al. The locust genome provides insight into swarm formation and long-distance flight. *Nature Commun*. 2014; 5: 2957.
37. Camacho JPM, Ruiz-Ruano FJ, Martín-Blázquez R, López-León MD, Cabrero J, Lorite P, et al. A step to the gigantic genome of the desert locust: chromosome sizes and repeated DNAs. *Chromosoma*. 2015; 124: 263–275. doi: [10.1007/s00412-014-0499-0](https://doi.org/10.1007/s00412-014-0499-0) PMID: [25472934](https://pubmed.ncbi.nlm.nih.gov/25472934/)
38. Hanrahan SJ, Johnston S. New genome size estimates of 134 species of arthropods. *Chromosome Res*. 2011; 19: 809–823. doi: [10.1007/s10577-011-9231-6](https://doi.org/10.1007/s10577-011-9231-6) PMID: [21877225](https://pubmed.ncbi.nlm.nih.gov/21877225/)
39. Gregory TR. *Animal Genome Size Database*. 2015. Available: <http://www.genomesize.com>.
40. Baumgartner WJ. Some evidence for the individuality of the chromosomes. *Biol Bull*. 1904; 8: 1–23.
41. Webb GC, White MJD, Contreras N, Cheney J. Cytogenetics of the parthenogenetic grasshopper *Warrawaba* (formerly *Moraba*) *virgo* and its bisexual relatives. IV. Chromosome banding studies. *Chromosoma*. 1978; 67: 309–339.

42. Sumner AT. A simple technique for demonstrating centromeric heterochromatin. *Exp Cell Res.* 1972; 75: 304–306. PMID: [4117921](#)
43. Schweizer D, Mendelak M, White MJD, Contreras N. Cytogenetics of the parthenogenetic grasshopper *Warramaba virgo* and its bisexual relatives. X. Pattern of fluorescent banding. *Chromosoma.* 1983; 88: 227–236.
44. Sambrook J, Russel DW. *Molecular cloning: A laboratory manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2001.
45. Cabral-de-Mello DC, Moura RC, 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.* 2010; 104: 393–400. doi: [10.1038/hdy.2009.126](#) PMID: [19756039](#)
46. Colgan DJ, McLauchlan A, Wilson GDF, Livingston SP, Edgecombe GD, Macaranas J, et al. Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Austral J Zool.* 1998; 46: 419–437.
47. Cabral-de-Mello DC, Valente GT, Nakajima RT, Martins C. Genomic organization and comparative chromosome mapping of the U1 snRNA gene in cichlid fish, with an emphasis in *Oreochromis niloticus*. *Chromosome Res.* 2012; 20: 279–292. doi: [10.1007/s10577-011-9271-y](#) PMID: [22234547](#)
48. Bueno D, Palacios-Gimenez OM, Cabral-de-Mello DC. Chromosomal mapping of repetitive DNAs in *Abracris flavolineata* reveal possible ancestry for the B chromosome and surprisingly H3 histone spreading. *PLoS ONE.* 2103; 8: e66532.
49. Zwick MS, Hanson RE, McKnight TD, Nurul-Islam-Faridi M, Stelly DM. A rapid procedure for the isolation of *C₀t*-1 DNA from plants. *Genome.* 1997; 40: 138–142. PMID: [18464813](#)
50. Ijdo JW, Wells RA, Baldini A, Reeders ST. Improved telomere detection using a telomere repeat probe (TTAGGG)_n generated by PCR. *Nucleic Acids Res.* 1991; 19: 4780. PMID: [1891373](#)
51. Pinkel D, Straume T, Gray JW. Cytogenetic analysis using quantitative, high sensitivity, fluorescence hybridization. *Proc Natl Acad Sci USA.* 1986; 83: 2934–2938. PMID: [3458254](#)
52. de Barros AV, Sczepanski TS, Cabrero J, Camacho JPM, Vicari MR, Artoni RF. Fiber FISH reveals different patterns of high-resolution physical mapping for repetitive DNA in fish. *Aquaculture.* 2011; 322: 47–50.
53. Camacho JPM, Cabrero J, López-León MA, Cabral-de-Mello DC, Ruiz-Ruano FJ. Grasshoppers (Orthoptera). In: Sharakhov V editor. *Protocols for cytogenetic mapping of arthropod genomes.* Boca Raton, FL, USA: CRC Press; 2015. pp. 381–438.
54. Lopes DM, Carvalho CR, Clarindo WR, Praça MM, Tavares MG. Genome size estimation of three stingless bee species (Hymenoptera, Meliponinae) by flow cytometry. *Apidologie.* 2009; 40: 517–523.
55. Otto F. DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. In: Darzynkiewicz Z, Crissman HA editors. *Methods in cell biology.* Vol 33; 1990. pp. 105–110.
56. Loureiro J, Rodriguez E, Doležel J, Santos C. Comparison of four nuclear isolation buffers for plant DNA flow cytometry. *Ann Bot.* 2006; 98: 679–689. PMID: [16820407](#)
57. Loureiro J, Rodriguez E, Doležel J, Santos C. Flow cytometric and microscopic analysis of the effect of tannic acid on plant nuclei and estimation of DNA content. *Ann Bot.* 2006; 98: 515–527. PMID: [16820406](#)
58. Shapiro HM. *Practical flow cytometry,* 4th ed., John Wiley & Sons, New Jersey; 2003.
59. Mesa A, Ferreira A, Carbonell CS. Cariología de los acridios neotropicales: estado actual de su conocimiento y nuevas contribuciones. *Ann Soc Entomol Fr (NS).* 1982; 18: 507–526.
60. Drets ME, Stoll M. C-banding and non-homologous associations in *Gryllus argentinus*. *Chromosoma.* 1974; 48: 367–390. PMID: [4448110](#)
61. Lambrot M, Alvarez-Sarret E. The chromosomes of *Gryllus* field cricket populations in central Chile (Insecta: Grylloptera: Gryllidae). *Can J Zool.* 1985; 63: 2626–2631.
62. Yoshimura A. Karyotypes of two American field crickets, *Gryllus rubens* and *Gryllus* sp. (Orthoptera: Gryllidae). *Entomol Sci.* 2005; 8: 219–222.
63. Zefa E, Cordeiro J, Blauth M, Piumbini M, Silva AF, Costa MKM, Martins LP. Expanding the geographic cytogenetic studies in the bush crickets *Eneoptera surinamensis* (De Geer, 1773) (Orthoptera, Gryllidae, Eneopterinae) from Brazilian Atlantic and Amazon Forest. *Zootaxa.* 2014; 3860: 396–400. doi: [10.11646/zootaxa.3860.4.7](#) PMID: [25283215](#)
64. Mesa A, Bran EJ. Acerca de los cromosomas de *Eneoptera surinamensis*. *An II Cong Latino-Americ Zol S Paulo.* 1964; 1: 9–16.
65. Santos JL, Arana P, Giráldez L. Chromosome C-banding patterns in Spanish Acridoidea. *Genetica.* 1983; 61: 65–74.

66. John B, King M. The inter-relationship between heterochromatin distribution and chiasma distribution. *Genetica*. 1985; 66: 183–194.
67. Steflova P, Tokan V, Vogel I, Lexa M, Macas J, Novak P, et al. Contrasting patterns of transposable element and satellite distribution on sex chromosomes (XY₁Y₂) in the dioecious plant *Rumex acetosa*. *Genome Biol Evol*. 2013; 5: 769–782. doi: [10.1093/gbe/evt049](https://doi.org/10.1093/gbe/evt049) PMID: [23542206](https://pubmed.ncbi.nlm.nih.gov/23542206/)
68. Pokorná M, Kratochvíl L, Kejnovský E. Microsatellite distribution on sex chromosomes at different stages of heteromorphism and heterochromatinization in two lizard species (Squamata: Eublepharidae: *Coleonyx elegans* and Lacertidae: *Eremias velox*). *BMC Genet*. 2011; 12: 90. doi: [10.1186/1471-2156-12-90](https://doi.org/10.1186/1471-2156-12-90) PMID: [22013909](https://pubmed.ncbi.nlm.nih.gov/22013909/)
69. Lepesant JM, Cosseau C, Boissier J, Freitag M, Portela J, Climent D, et al. Chromatin structural changes around satellite repeats on the female sex chromosome in *Schistosoma mansoni* and their possible role in sex chromosome emergence. *Genome Biol*. 2012; 13: R14. doi: [10.1186/gb-2012-13-2-r14](https://doi.org/10.1186/gb-2012-13-2-r14) PMID: [22377319](https://pubmed.ncbi.nlm.nih.gov/22377319/)
70. Steinemann S, Steinemann M. The enigma of Y chromosome degeneration: TRAM, a novel retrotransposon preferentially located on the Neo-Y chromosome of *Drosophila miranda*. *Genetics*. 1997; 145: 261–266. PMID: [9071582](https://pubmed.ncbi.nlm.nih.gov/9071582/)
71. Steinemann S, Steinemann M. Retroelements: tools for sex chromosome evolution. *Cytogenet Genome Res*. 2005; 110: 134–143. PMID: [16093665](https://pubmed.ncbi.nlm.nih.gov/16093665/)
72. Kubat Z, Hobza R, Vyskot B, Kejnovsky E. Microsatellite accumulation on the Y chromosome in *Silene latifolia*. *Genome*. 2008; 51: 350–356. doi: [10.1139/G08-024](https://doi.org/10.1139/G08-024) PMID: [18438438](https://pubmed.ncbi.nlm.nih.gov/18438438/)
73. Kejnovský E, Michalovova M, Steflova P, Kejnovska I, Manzano S, Hobza R, et al. Expansion of microsatellites on evolutionary young Y chromosome. *PLoS One*. 2013; 8: e45519. doi: [10.1371/journal.pone.0045519](https://doi.org/10.1371/journal.pone.0045519) PMID: [23341866](https://pubmed.ncbi.nlm.nih.gov/23341866/)
74. Matsubara K, Knopp T, Sarre SD, Georges A, Ezaz T. Karyotypic analysis and FISH mapping of microsatellite motifs reveal highly differentiated XX/XY sex chromosomes in the pink-tailed worm-lizard (*Aprasia parapulchella*, Pygopodidae, Squamata). *Mol Cytogenet*. 2013; 6: 60. doi: [10.1186/1755-8166-6-60](https://doi.org/10.1186/1755-8166-6-60) PMID: [24344753](https://pubmed.ncbi.nlm.nih.gov/24344753/)
75. Matsubara K, O'Meally D, Azad B, Georges A, Sarre SD, Graves JA, et al. Amplification of microsatellite repeat motifs is associated with the evolutionary differentiation and heterochromatinization of sex chromosomes in Sauropsida. *Chromosoma*. 2015 Jul 21. doi: [10.1007/s00412-015-0531-z](https://doi.org/10.1007/s00412-015-0531-z)
76. Milani D, Cabral-de-Mello DC. Microsatellite organization in the grasshopper *Abracris flavolineata* (Orthoptera: Acrididae) revealed by FISH mapping: remarkable spreading in the A and B chromosomes. *PLoS ONE*. 2014; 9: e97956. doi: [10.1371/journal.pone.0097956](https://doi.org/10.1371/journal.pone.0097956) PMID: [24871300](https://pubmed.ncbi.nlm.nih.gov/24871300/)
77. Ruiz-Ruano FJ, Cuadrado Á, Montiel EE, Camacho JPM, López-León MD. Next generation sequencing and FISH reveal uneven and nonrandom microsatellite distribution in two grasshopper genomes. *Chromosoma*. 2015; 124: 221–234. doi: [10.1007/s00412-014-0492-7](https://doi.org/10.1007/s00412-014-0492-7) PMID: [25387401](https://pubmed.ncbi.nlm.nih.gov/25387401/)
78. Clark AG. The vital Y chromosome. *Nature*. 2014; 508: 463–465. doi: [10.1038/508463a](https://doi.org/10.1038/508463a) PMID: [24759407](https://pubmed.ncbi.nlm.nih.gov/24759407/)
79. Toups M, Veltsos P, Pannell JR. Plant sex chromosomes: lost genes with little compensation. *Curr Biol*. 2015; 25: 427–429.
80. Bergero R, Qiu S, Charlesworth D. Gene loss from a plant sex chromosome system. *Curr Biol*. 2015; 25: 1234–1240. doi: [10.1016/j.cub.2015.03.015](https://doi.org/10.1016/j.cub.2015.03.015) PMID: [25913399](https://pubmed.ncbi.nlm.nih.gov/25913399/)
81. Cabrero J, Camacho JP. Location and expression of ribosomal RNA genes in grasshoppers: abundance of silent and cryptic loci. *Chromosome Res*. 2008; 16: 595–607. doi: [10.1007/s10577-008-1214-x](https://doi.org/10.1007/s10577-008-1214-x) PMID: [18431681](https://pubmed.ncbi.nlm.nih.gov/18431681/)
82. Cabral-de-Mello DC, Cabrero J, López-León MD, Camacho JPM. Evolutionary dynamics of 5S rDNA location in acridid grasshoppers and its relationship with H3 histone gene and 45S rDNA location. *Genetica*. 2011; 139: 921–931. doi: [10.1007/s10709-011-9596-7](https://doi.org/10.1007/s10709-011-9596-7) PMID: [21755328](https://pubmed.ncbi.nlm.nih.gov/21755328/)
83. Nguyen P, Sahara K, Yoshido A, Marec F. Evolutionary dynamics of rDNA clusters on chromosomes of moths and butterflies (Lepidoptera). *Genetica*. 2010; 138: 343–354. doi: [10.1007/s10709-009-9424-5](https://doi.org/10.1007/s10709-009-9424-5) PMID: [19921441](https://pubmed.ncbi.nlm.nih.gov/19921441/)
84. Cabral-de-Mello DC, Oliveira SG, Moura RC, Martins C. 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 Genet*. 2011; 12: 88. doi: [10.1186/1471-2156-12-88](https://doi.org/10.1186/1471-2156-12-88) PMID: [21999519](https://pubmed.ncbi.nlm.nih.gov/21999519/)
85. Panzera Y, Pita S, Ferreiro MJ, Ferrandis I, Lages C, Pérez R, et al. High dynamics of rDNA cluster location in kissing bug holocentric chromosomes (Triatominae, Heteroptera). *Cytogenet Genome Res*. 2012; 138: 56–67. doi: [10.1159/000341888](https://doi.org/10.1159/000341888) PMID: [22907389](https://pubmed.ncbi.nlm.nih.gov/22907389/)

86. Utsunomia R, Scacchetti PC, Pansonato-Alves JC, Oliveira C, Foresti F. Comparative chromosome mapping of U2 snRNA and 5S rRNA genes in *Gymnotus* species (Gymnotiformes, Gymnotidae): evolutionary dynamics and sex chromosome linkage in *G. pantanal*. *Cytogenet Genome Res.* 2014; 142: 286–292. doi: [10.1159/000362258](https://doi.org/10.1159/000362258) PMID: [24776647](https://pubmed.ncbi.nlm.nih.gov/24776647/)
87. Úbeda-Manzanaro M, Merlo MA, Palazón JL, 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.* 2010; 138: 787–794. doi: [10.1007/s10709-010-9460-1](https://doi.org/10.1007/s10709-010-9460-1) PMID: [20440541](https://pubmed.ncbi.nlm.nih.gov/20440541/)
88. 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.* 2001; 44: 331–335. PMID: [11444690](https://pubmed.ncbi.nlm.nih.gov/11444690/)
89. Manchado M, Zuasti E, Cross I, Merlo 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.* 2006; 49: 79–86. PMID: [16462904](https://pubmed.ncbi.nlm.nih.gov/16462904/)
90. Vierna J, Jensen KT, Martínez-Lage A, González-Tizón AM. The linked units of 5S rDNA and U1 snDNA of razor shells (Mollusca: Bivalvia: Pharidae). *Heredity.* 2011; 107: 127–142. doi: [10.1038/hdy.2010.174](https://doi.org/10.1038/hdy.2010.174) PMID: [21364693](https://pubmed.ncbi.nlm.nih.gov/21364693/)
91. Anjos A, Ruiz-Ruano FJ, Camacho JPM, Loreto V, Cabrero J, de Souza MJ, Cabral-de-Mello DC. U1 snDNA clusters in grasshoppers: chromosomal dynamics and genomic organization. *Heredity.* 2015; 114: 207–219. doi: [10.1038/hdy.2014.87](https://doi.org/10.1038/hdy.2014.87) PMID: [25248465](https://pubmed.ncbi.nlm.nih.gov/25248465/)
92. Cabrero J, López-León MD, Teruel M, Camacho JP. Chromosome mapping of H3 and H4 histone gene clusters in 35 species of acridid grasshoppers. *Chromosome Res.* 2009; 17: 397–404. doi: [10.1007/s10577-009-9030-5](https://doi.org/10.1007/s10577-009-9030-5) PMID: [19337846](https://pubmed.ncbi.nlm.nih.gov/19337846/)
93. Cohen S, Agmon N, Sobol O, Segal D. Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. *Mobile DNA.* 2010; 1: 11. doi: [10.1186/1759-8753-1-11](https://doi.org/10.1186/1759-8753-1-11) PMID: [20226008](https://pubmed.ncbi.nlm.nih.gov/20226008/)


4.3. Capítulo 3

*High-throughput analysis of the satellitome revealed enormous diversity of satellite DNAs in the neo-Y chromosome of the cricket *Eneoptera surinamensis**

Octavio Manuel Palacios-Gimenez, Guilherme Borges Dias, Leonardo Gomes de Lima, Gustavo Campos e Silva Kuhn, Érica Ramos, Cesar Martins, Diogo Cavalcanti Cabral-de-Mello


Publicado na revista *Scientific Reports* (2017), 7: 6422.

SCIENTIFIC REPORTS



OPEN

High-throughput analysis of the satellitome revealed enormous diversity of satellite DNAs in the neo-Y chromosome of the cricket *Eneoptera surinamensis*

Octavio Manuel Palacios-Gimenez¹, Guilherme Borges Dias², Leonardo Gomes de Lima², Gustavo Campos e Silva Kuhn², Érica Ramos³ , Cesar Martins³ & Diogo Cavalcanti Cabral-de-Mello¹

Satellite DNAs (satDNAs) constitute large portion of eukaryote genomes, comprising non-protein-coding sequences tandemly repeated. They are mostly found in heterochromatic regions of chromosomes such as around centromere or near telomeres, in intercalary heterochromatin, and often in non-recombining segments of sex chromosomes. We examined the satellitome in the cricket *Eneoptera surinamensis* ($2n = 9$, neo- X_1X_2Y , males) to characterize the molecular evolution of its neo-sex chromosomes. To achieve this, we analyzed illumina reads using graph-based clustering and complementary analyses. We found an unusually high number of 45 families of satDNAs, ranging from 4 bp to 517 bp, accounting for about 14% of the genome and showing different modular structures and high diversity of arrays. FISH mapping revealed that satDNAs are located mostly in C-positive pericentromeric regions of the chromosomes. SatDNAs enrichment was also observed in the neo-sex chromosomes in comparison to autosomes. Especially astonishing accumulation of satDNAs loci was found in the highly differentiated neo-Y, including 39 satDNAs over-represented in this chromosome, which is the greatest satDNAs diversity yet reported for sex chromosomes. Our results suggest possible involvement of satDNAs in genome increasing and in molecular differentiation of the neo-sex chromosomes in this species, contributing to the understanding of sex chromosome composition and evolution in Orthoptera.

Among the repetitive sequences in the genomes of eukaryotes, tandem repeats (TRs) are very abundant and are mostly represented by satellite DNAs (satDNAs). SatDNA sequences are mainly located in centromeric, telomeric or intercalary heterochromatin¹⁻³ but, in some cases, also dispersed in euchromatin^{4,5}. This genomic fraction is composed of hundreds to thousands of noncoding tandemly-arrayed sequences with late-replication, and oriented in a head-to-tail fashion^{1,3,6-9}.

SatDNA families, in general, differ in sequence identity, copy number and chromosome distribution^{2,10-12}. These sequences are subject to intragenomic concerted evolution, resulting in more efficient homogenization of repeats within species than between species and also between repeats located in the same array/chromosome than between different ones^{2-4,7}. Concerted evolution is achieved through multiple mechanisms of non-reciprocal transfer such as unequal cross-over, gene conversion, rolling-circle replication and transposition^{13,14}.

Sex chromosomes have arisen independently several times in a wide range of animals and plants from an ordinary autosomal pair^{15,16}, presenting as a recurrent trait the suppression of recombination and accumulation of distinct classes of repetitive DNAs, including satDNAs^{1,17-21}. In Orthoptera, the $X0\sigma/XX\varphi$ sex-determining

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

²Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil. ³UNESP - Univ Estadual Paulista, Instituto de Biociências/IB, Departamento de Morfologia, Botucatu, São Paulo, Brazil. Correspondence and requests for materials should be addressed to D.C.C.-d.M. (email: mellodc@rc.unesp.br)

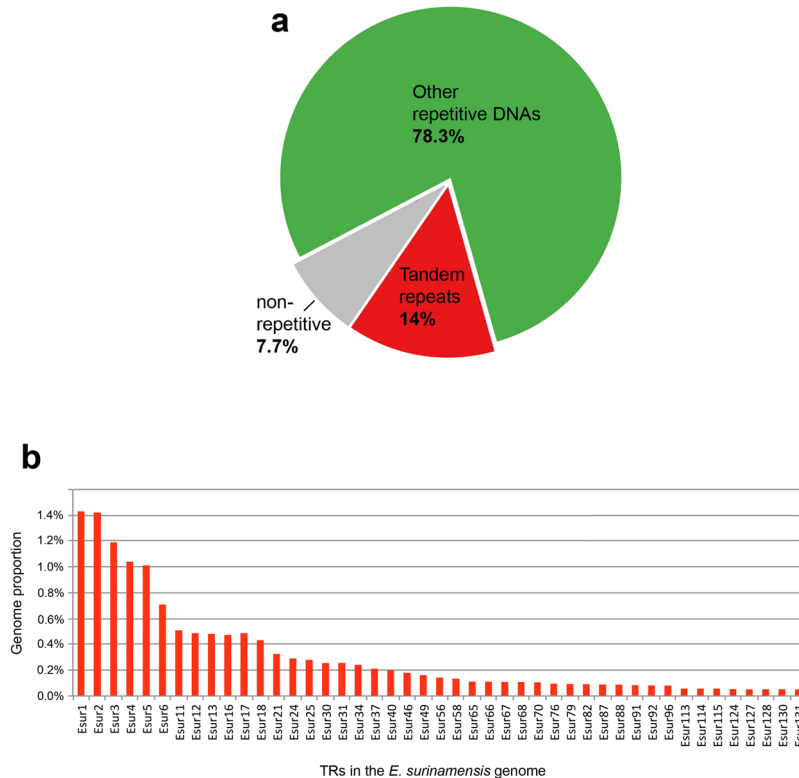


Figure 1. (a) Overview of the repetitive composition of *E. surinamensis* genome based on the output of RepeatExplorer. (b) Genomic proportion for each of the 45 satDNAs isolated and characterized. TRs, tandem repeats.

system is considered modal^{22, 23} but eventually, diverse sex chromosome systems evolved several times, such as neo-XY♂/XX♀^{24–26}, X₁X₂0♂/X₁X₁X₂X₂♀²⁷ and even neo-X₁X₂Y♂/X₁X₁X₂X₂♀^{25, 28}. It was found that particularly centric fusions (i.e. Robertsonian translocations) and tandem fusions with autosomes, dissociations and inversions contributed to the formation of neo-sex chromosomes in Orthoptera^{22, 23, 25–28}. The DNA composition of the orthopteran neo-sex chromosomes was studied only in a few species, mainly focusing on certain types of repetitive DNAs, such as multigene families, C₀t DNA fraction, telomeric repeats and microsatellite arrays^{25–29}.

Chromosomal evolution and repetitive DNA organization was addressed in the cricket *Eneoptera surinamensis*, a species with the genome size of 5.42 Gbp, chromosome number of 2n = 9♂/10♀ and a neo-X₁X₂Y♂/X₁X₁X₂X₂♀ sex-determining system²⁸. During male meiosis the multiple sex chromosomes remain unpaired and do not form chiasmata, suggesting that they do not proceed recombination^{22, 23}. The neo-Y chromosome is the largest element and it exhibits multiple heterochromatic bands, while the neo-X₁ and neo-X₂ are poor in C-positive heterochromatin. Of the repetitive DNAs, two multigene families (5S rDNA and histone genes), C₀t DNA fraction, and diverse microsatellites mapped to the neo-Y²⁸.

In the present study, in order to provide comprehensive information on repetitive DNAs that occur in the genome, and specifically in the neo-sex chromosomes of *E. surinamensis*, we performed a detailed analysis of satDNAs (the satellitome)³⁰, integrating genomic and chromosomal data. Our results revealed accumulation of satDNAs, with the occurrence of unusually high number of 45 distinct families in the *E. surinamensis* genome. In addition, some of these families were enriched in the highly differentiated neo-X₁X₂Y sex chromosomes in comparison to autosomes, mostly over-represented in the neo-Y. To our knowledge, this is the largest diversity of satDNAs yet reported for a Y chromosome in eukaryotes.

Results

Identification of satDNAs and sequence characterization. Illumina sequencing returned 5,872,912 paired-end reads (ranging from 35 to 288 nt, mean reads length of 164.6 nt) totaling 615,581,150 nt that were trimmed to 150 nt. Given the estimated genome size of 5.42 Gbp for *E. surinamensis*²⁸, this represents about 0.16× genome coverage. The clustering analysis through the RepeatExplorer^{31, 32} used as input 1,299,110 illumina previously trimmed paired-ends reads and produced 292,070 clusters (containing 92.3% of reads) differing in size, sequence composition and genomic abundance, including satDNAs and other non-characterized repetitive elements. The singletons should represent the low copy number fraction of the genome, which yielded in 100,002 singletons (containing 7.7% of reads) (Fig. 1a). A set of 131 of the most abundant clusters representing repetitive elements was analyzed in the search for satDNAs. A representation of the genomic proportion of 45 clusters identified as satDNAs is showed in Fig. 1b.

The analysis of dotplots confirmed the occurrence of 45 families of tandem repeats ranging from 4 to 517 bp long, showing different modular structure and array diversity. Among them, 21 satDNAs families were recovered through PCR that revealed a ladder pattern. For the remaining repeat families, with small monomers (less than 60 bp), we were not able to confirm the ladder pattern through PCR due to the very limited window for primer design and amplification. The characteristics for all satDNAs families are presented in the Supplementary Tables S1 and S2. Altogether, the 45 satDNAs families comprised about 14% of the male genome that represent 758,8 Mb of 5,420,00 Mb of species genome size. The nucleotide divergence within the families varied from 1.9 to 30.2%. The nucleotide sequences showed an A + T content ranging from 15 to 67.15%. It can be concluded that most of the satDNAs analyzed here constitute light satDNAs due to the low A + T content (see Supplementary Tables S1 and S2).

For two satDNAs, i.e. Esur17 and Esur18, two subfamilies were identified showing an individual repeat called α and a composite repeat called α/β (see Supplementary Table S2 and Supplementary Fig. S1). There was no significant sequence similarity between α and β sequences. The size of the Esur17 α -repeat was 163 bp while the Esur17 α/β composite repeat was 317 bp, each consisting of an α -unit and additional 154 bp β -unit. The nucleotide divergence between both Esur17- α and Esur17- α/β repeats was 9.4% (see Supplementary Table S2). Regarding Esur18, our sequences analysis showed that this is an array composed of α/β composite units with 517 bp, consisting of a 329 bp α -unit and a 188 bp β -unit (see Supplementary Fig. S1). The Esur18 α/β repeats showed 4.3% nucleotide divergence (see Supplementary Table S2). We also detected sequence similarity between Esur2 and Esur34 repeats (~82.6%) and between Esur3 and Esur58 repeats (80.9%). This result explains the overlapping chromosomal distribution of both satDNA pairs (see below) and indicate that they are representatives of two satDNA families, Esur2/Esur34 and Esur3/Esur58. NJ trees showed Esur2, Esur34, Esur3 and Esur58 allocated in cluster-specific branches, indicating that each subfamily is composed of exclusive repeat-variants (see Supplementary Fig. S2). Both NCBI BLAST and Repbase searches, with the consensus monomer sequence belonging to each repeat family as a query, did not revealed significant similarity with any other previously described sequences.

Considering all satDNAs identified, we found that monomer size and nucleotide divergence displayed a somewhat low but significant negative correlation ($\rho = -0.48$, $P = <0.001$). Also, monomer size and a total number of loci displayed a moderate negative correlation ($\rho = -0.63$, $P = <0.0001$) (Supplementary Fig. S3).

Chromosomal localization of satDNAs. To detect the chromosomal localization of the 45 satDNA families single- or two-color FISH were carried out on male mitotic metaphases. In autosomes, distinct satDNAs were located mostly in the pericentromeric regions that correspond to the C-band positive blocks observed by Palacios-Gimenez *et al.*²⁸. A few satDNA loci were also found in interstitial and distal C-band negative blocks. The patterns were variable depending on the repeat mapped (Fig. 2, Supplementary Tables S1 and S2). Most satDNAs were located in multiples autosomes, while some of them were located exclusively in one chromosome pair, for example, Esur4, Esur6, and Esur18- α/β , were all co-located in the centromere and secondary constriction of pair 1. Two-color fiber-FISH confirmed regions of interspersed between Esur4, Esur6 and Esur18- α/β repeats (see Supplementary Fig. S4). For Esur17, the α and α/β repeats were located in the secondary constriction of pair 1 and in the X₂ chromosome; for Esur18, the α and β repeats were located in secondary constriction of pair 1, but additional multiple loci of the α repeats were detected in the Y chromosome (see Supplementary Table S2 and Supplementary Fig. S5).

Some of satDNAs families were placed in the neo-X₁ and neo-X₂ but the abundance (number of loci) of satDNAs on the neo-Y chromosomes was remarkable, showing either multiple discrete loci or a scattered pattern (Fig. 3). The abundance in relation to the number of loci on the neo-Y chromosome was also noteworthy compared to autosomes and the neo-X₁ and neo-X₂ chromosomes (Fig. 2 and Supplementary Tables S1 and S2). Among the 45 satDNAs families, 39 were located in the neo-Y chromosome and six of them (Esur11, Esur31, Esur37, Esur46, Esur65 and Esur66) through FISH, mapped exclusively to the neo-Y chromosome, however, they were recovered through PCR in both sexes (Fig. 3, Supplementary Tables S1 and S2).

Male versus female satDNAs abundance. The comparative relative abundance of satDNAs between male and female genomes carried out through qPCR clearly show that the satDNAs doses differ significantly between sexes, with males harboring a higher copy number than females for most of repeats studied (Fig. 4, Supplementary Table S3). For example, most remarkable difference were seen in Esur2, Esur11, Esur46, Esur65, Esur66, Esur87 and Esur128, with males showing six to ten times more copies than females (Fig. 4).

Discussion

General organization of satDNAs in the genome of *E. surinamensis*. Through the graph-based clustering of sequencing reads of the *E. surinamensis* genome, followed by complementary analysis, we discovered 45 new satDNAs families. These satDNAs coexist in the *E. surinamensis* genome, with variable monomer sizes and nucleotide divergence, accounting for ~14% of the genome (758.8 Mb). It is well known that diverse satDNA families are commonly found in eukaryotic genomes, e.g., in *Olea europea* with six satDNA families³³, *Tribolium castaneum* with nine³⁴, and *Camellia japonica* with four³⁵, reaching up to 16 satDNAs in *Drosophila melanogaster* (reviewed by ref. 36) and 62 in the grasshopper *L. migratoria*³⁰. Besides the considerable proportion of satDNAs in the genome of *E. surinamensis*, it is also remarkable the presence of so many different families in comparison with most eukaryotes. Another interesting feature of *E. surinamensis* is that no satDNA family is predominant, which contrasts with what has been observed in several other species, in which a few families prevailed^{33–36}.

We could speculate that the accumulation of satDNAs in *E. surinamensis* could have been favored by highly rearranged karyotype ($2n = 9\delta$) in comparison with the modal diploid chromosome number in Gryllidae crickets²³. Alternatively, the high diversity of repetitive DNA itself might facilitate/cause such a complex

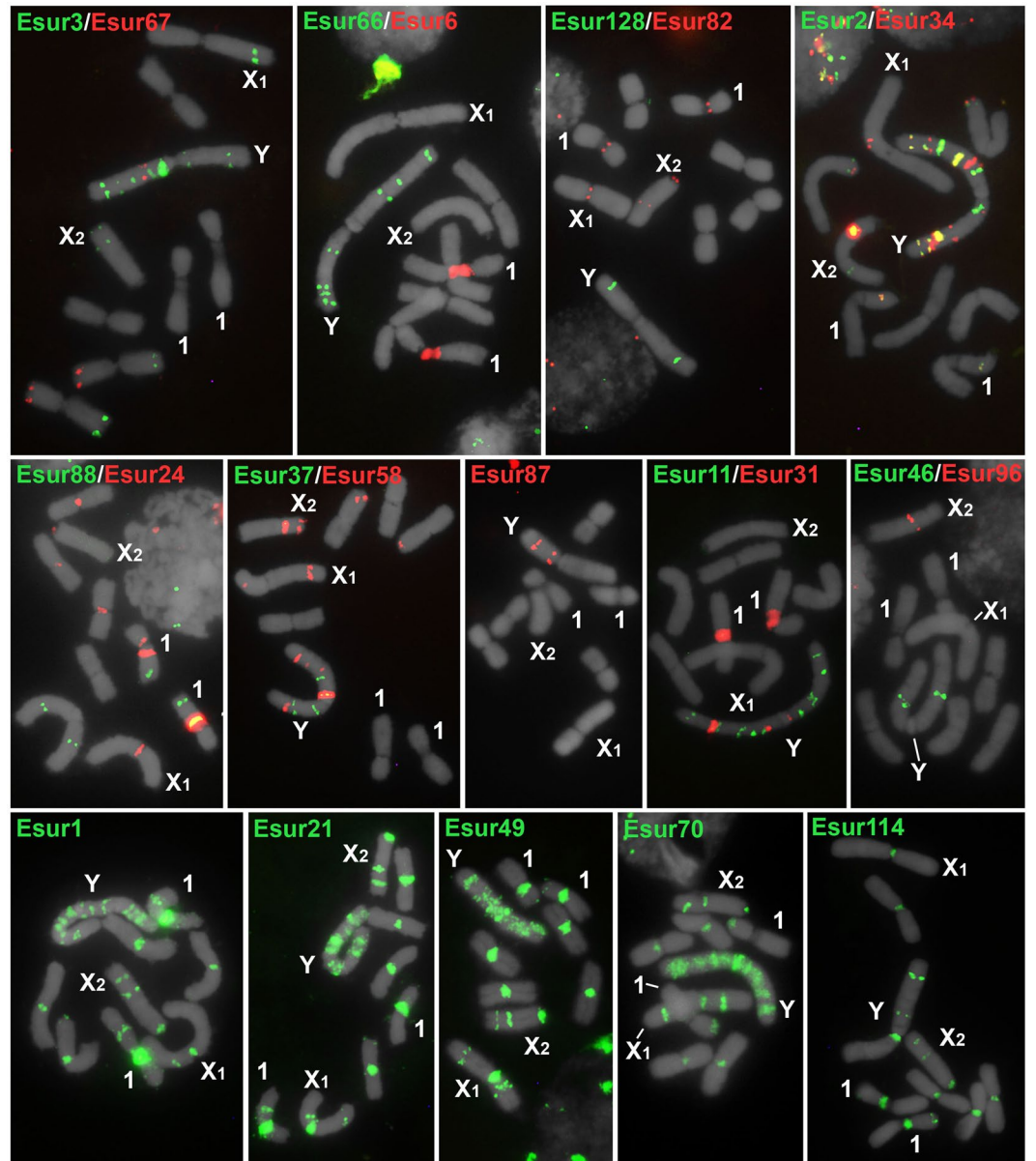


Figure 2. Chromosomal location of 22 satDNAs in mitotic chromosomes of male embryos of *E. surinamensis* by FISH. Upper and middle panels show satDNAs with monomers >60 bp, lower panels satDNAs with monomers <60 bp. The satDNA family names and hybridization signals for each type of the probe are shown in the images by colors. Note the enrichment in the neo-Y (Y) sex chromosome with multiple loci or spread signals for some repeats, while the X₁ and X₂ sex chromosomes show only a few or even none hybridization signals.

karyotype rearrangements. It is well known that chromosomal rearrangements may involve highly repetitive DNA sequences, since these could provide sites for karyotype reshuffling without detrimental effects on the integrity of coding sequences^{37–39}. In any case, the rearranged karyotype of *E. surinamensis* represents a specific environment with limited recombination that could facilitate the rate of homogenization of repetitive DNA, like in multiple sex chromosomes (see below). In addition to chromosome rearrangements, mechanisms involved in the satDNAs evolution could have occurred, such as amplification mediated by rolling-circle replication and reinsertion, unequal crossing-over between DNA repeats from sister chromatids, transposition and gene conversion. These mechanisms have been suggested as possible causes of sequence homogenization within satDNA^{1,8}. The putative involvement of highly rearranged karyotype for satDNA multiplication could be supported by comparative analysis with the related species *Gryllus assimilis* (Gryllidae) with higher diploid number that is ancient for cricket, in which the satDNA analysis revealed the occurrence of only 11 satDNA families (Palacios-Gimenez *et al.*, submitted). Considering the genome size (*G. assimilis* 2.13 Gb and *E. surinamensis* 5.42 Gb) and quantity of satDNA in each species (*G. assimilis* 4% and *E. surinamensis* 14%) the quantity of satDNAs was increased about 8.9 times in the genome of *E. surinamensis*.

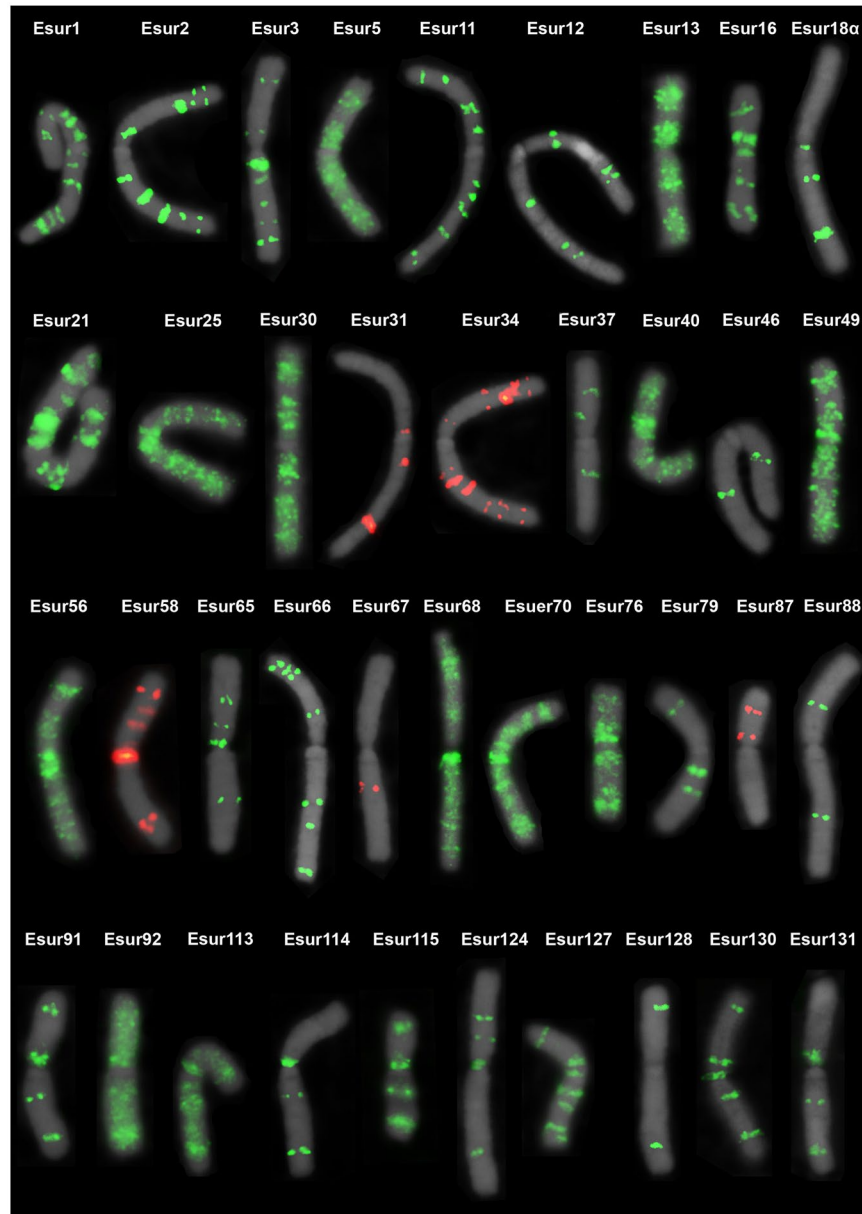


Figure 3. Distribution of the 39 satDNAs in the neo-Y chromosome of *E. surinamensis*.

We found that most satDNAs are non-homologous, suggesting inefficient homogenization between distinct repeats. The intra-family homogenization is evident, even more for larger satDNAs than smaller ones, and could be directly related to the sequence size and specific chromosomal distribution. Comparing the distribution of the large- and small-sized satDNAs it is evident that the smaller, ranging up to 60 bp, repeats are more dispersed than larger ones (Fig. 2), mostly occurring in all chromosomes of the complement, which could facilitate their sequence divergence.

In our data, the satDNA monomer size was negatively correlated with sequence divergence ($\rho = -0.48$, $P < 0.001$) (Supplementary Fig. S3). A similar trend was also observed for three satDNAs of the cave cricket *Dolichopoda schiavazzii*⁴⁰. This finding disagrees with predictions from computer simulations, which anticipated a direct correlation between the monomer size and the divergence due to low recombination frequency⁴¹. However, as noted by Bachmann *et al.*⁴⁰, A + T content and sequence complexity, which were not included in these early models, could play a role in facilitating recombination and homogenization of the repeats. We also found a negative correlation between the satDNA monomer size and the number of loci ($\rho = -0.63$, $P < 0.0001$) (Supplementary Fig. S3). The distribution of small satDNAs in almost all chromosomes, and remarkably in the neo-Y, could be the result of dispersion through extra-chromosomal circular DNAs, formed via intrastrand recombination, and subsequent reintegration elsewhere in the genome by illegitimate or homologous recombination (reviewed in ref. 42). In addition, the movement of otherwise non-transposable DNA fragments via their capture by transposable elements could represent another possible mechanism of the intragenomic dispersion

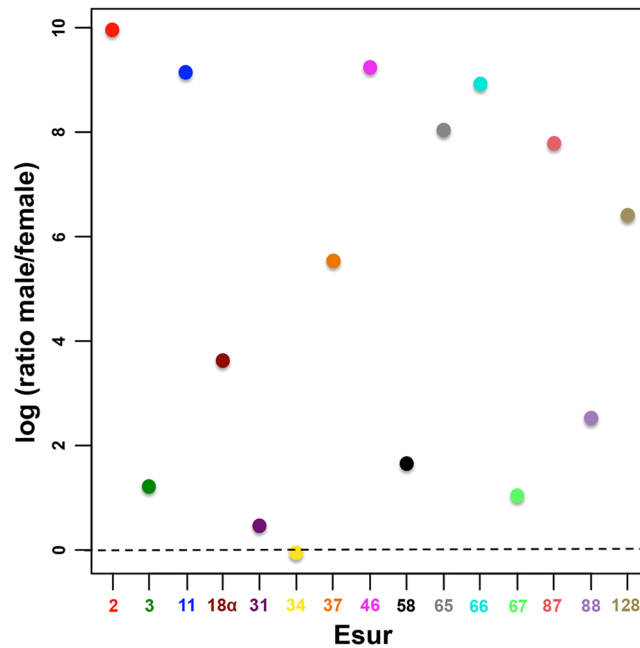


Figure 4. Copy number variation between male vs female genomes of 15 satDNAs with monomers >60 bp in *E. surinamensis*. The male/female ratio of a relative copy number is shown on a logarithmic scale. The qPCR of male and female genomic DNA was used to calculate the satDNA dose by a Δ Ct method of relative quantification (see Supplementary Table S3). Each satDNA family (Esur) is represented by numbers and colors directly in the images. Note that most satDNAs differ significantly between sexes with males having a higher copy number than females for most repeats studied. The statistical significance for difference in copy number between males and females using chi-squared test was highly significant, p -value < 0.01 for each satDNA.

of tandem repeats^{43–47}. What remains is to investigate the exact role played by the monomer size in facilitating satDNA dispersion.

For the large satDNAs, which exhibit in a more chromosome-specific distribution, our combined analysis using cytogenetic and genomic tools revealed a distinct degree of complexity, i.e., simple tandem arrangements and composed units, like α/β repeats for Esur17 and Esur18. satDNAs with composite units have been reported for example in *Chrysolina carnifex*⁴⁸, *Tribolium brevicornis*⁴⁹ and primates⁵⁰. Besides the composite units observed for Esur17 and Esur18 we also found cases of overlapping distribution of satDNAs as demonstrated by fiber-FISH for the satDNAs families exclusively located in chromosome pair 1, Esur4, Esur6 and Esur18. It occurred independently of sequence homology. Moreover, we observed other cases of co-localization depending on the chromosome. Overlapping distribution of satDNAs is usually reported for repeats with a remarkable similarity, such as the satellite I and satellite II subfamilies in *Tribolium madens*⁵¹ the pBuM-1 and pBuM-2 subfamilies in the *Drosophila buzzatii* species complex⁵² and the psr2 and psr18 subfamilies in *Nasonia vitripennis*⁵³. However, overlapping distribution of non-homologous satDNAs as observed here for *E. surinamensis* was reported only in a few cases, e.g., the DBC-150 and pBuM satDNAs family in several *Drosophila* species⁵⁴ the pSc200 and pSc250 families in the rye chromosomes⁵⁵ and the PROsat, PSUchr1sat and PsatDNA satDNAs in two species of hamsters⁵⁶.

The complexity of the evolution of satDNAs identified in the *E. surinamensis* genome is reflected by (i) junctions of monomers forming complex units, as explained above and reinforced (ii) by the occurrence of divergent satDNAs, probably originating from a common ancestor, as indicated by the sequence similarity. For example, we found an apparent similarity of monomer satDNA sequences between Esur2 and Esur34 and between Esur3 and Esur58. However, each of these sequences in the constructed NJ tree constitutes separate branches. This allows us to conclude that these sequences belong to closely related satDNAs families with a common ancestor but they diverged during their evolution, including moving to distinct chromosomal regions. Moreover, these sequences differentially amplified which resulted in variation in their copy number, as demonstrated by qPCR experiments.

Accumulation of satDNAs occurred in the *E. surinamensis* neo-Y chromosome. It is well known that the Y or W sex chromosomes accumulate high quantities of distinct classes of repetitive DNAs due to the low frequency of recombination^{1, 18, 19, 21, 57, 58}. In some species, it was shown that these repetitive sequences are involved in genetic degeneration, chromatin organization and regulation of expression^{36, 57, 59}. This could also be the case in the *E. surinamensis* sex chromosomes that remain either unpaired and achiasmatic during male meiosis, suggesting the absence of recombination between them^{22, 23}. Regarding satDNAs, more than one family harboring sex chromosomes were reported, for example, in *D. melanogaster*⁶⁰, *Muntiacus muntjac*^{61, 62}, *T. castaneum*³⁴, *Rumex acetosa*^{63–65} and *Silene latifolia*⁶⁶. However, in comparison with these species the *E. surinamensis* neo-Y chromosome harbors the highest diversity of satDNAs documented to date, representing 39 distinct families, with seven being exclusive to this chromosome. Moreover, most satDNAs are enriched in this sex

chromosome and constitute multiple loci as shown by FISH. These findings suggest that the neo-Y accumulated satDNAs after its origin, causing its enlargement, in comparison with the neo-X₁ and neo-X₂. A similar pattern was reported for the large Y chromosomes of plants that also accumulated repetitive DNAs^{65,66} in contrast to the small mammalian Y chromosomes^{67,68}.

Our data suggest a complex origin of the neo-sex chromosomes in *E. surinamensis*, involving more than two simple translocations anticipated for the Orthoptera neo-X₁X₂Y sex chromosomes^{22,23,25}. The evolution of *E. surinamensis* sex chromosomes probably included centric and tandem fusions and inversions, like in *M. muntjac*^{61,62}. The multiple interstitial satDNAs loci in the *E. surinamensis* neo-Y could represent remnants of the ancestral centromeric material at chromosome fusions sites, or they are the consequence of massive expansion of DNA repeats that could induce new chromosomal rearrangements.

Although some satDNA families were exclusively mapped by FISH to the neo-Y chromosome, they were also recovered in the female genomic DNA using PCR, indicating their occurrence in other chromosomes, at least in a low copy number. Differential expansion of satDNAs between sexes was corroborated by qPCR. Our results clearly showed that the doses of most satDNAs examined differ significantly between sexes, probably due to differential expansion, with males having on average two to ten times more copies than females. We hypothesize that the differential repeats distribution patterns on the *E. surinamensis* sex chromosomes can be explained by a higher rate of colonization and insertion and lower rate of removal of satDNAs in the neo-Y chromosome, similarly to that suggested for the *R. aestosa* sex chromosomes⁶⁵. Therefore, these satDNAs lead to compartmentalization of the neo-Y chromosome and a chromosome-wide DNA sequence diversity. In addition, the neo-Y chromosome may exhibit a differentiated long-range chromatin structure/composition compared to autosomes and both neo-X chromosomes. Partially-heterochromatic appearance of the neo-Y chromosome in *E. surinamensis* with G + C-rich blocks dispersed in its entire length²⁵ supports this view. The involvement of neo-sex chromosomes in increasing of satDNAs quantity and diversity in *E. surinamensis* emerges by analysis of *G. assimilis* with X0 sex system, which is ancestral for Orthoptera as a whole^{22,23}. In this species only 11 satDNAs families were recovered by the same pipeline used here. Among them only eight populate the sex chromosome with discrete bands mainly in terminal region (Palacios-Gimenez *et al.* submitted).

Our data, together with the analysis of sex chromosome that revealed either unpaired and achiasmatic during meiosis^{22,23} – a common feature for old sex chromosomes⁶⁹ – suggest the occurrence of old neo-sex chromosome system in *E. surinamensis*, which contrasts with the described young sex chromosomes^{21,70,71}. Moreover considerable accumulation of various satDNAs and other repetitive DNA classes in the neo-Y chromosome²⁵ reinforce this view, as reported in other species^{67,69}. This could mean that there are no evolutionary strata on the *E. surinamensis* sex chromosomes, similar to the strata found in human X chromosome⁷². Theoretically, satDNAs should accumulate more intensively in a region of the Y chromosome corresponding to the ancestral X chromosome, in which recombination was ceased earlier, but no such region was identified in our study, supporting this view.

Our study provides important information concerning composition and evolution of neo-sex chromosomes among crickets, suggesting the involvement of neo-sex chromosomes in amplification and sequence divergence for satDNAs, which generated the highest diversity of satDNAs in sex chromosomes. This data with future analysis using other repetitive sequences and sex related single copy genes will be also relevant to understand precise composition of sex chromosomes, helping in the understanding of sex chromosome evolution, sex determination and possible mechanisms involved in dosage compensation, an issue almost completely unknown in Orthoptera. Moreover to shed light in the evolutionary history of sex chromosomes in crickets it will be informative the analysis of other species with divergent sex systems using similar strategies applied here in a comparative manner.

Materials and Methods

Samples, Chromosome Preparations and Genomic DNA Extraction. Males and females of *Eneoptera surinamensis* were collected in the Parque Estadual Edmundo Navarro de Andrade (Rio Claro, SP, Brazil) between May 2013 and March 2014 with the authorization of COTEC (process number 341/2013) and were maintained in captivity until oviposition. Mitotic chromosomes preparations were obtained from embryo neuroblasts using standard procedures described elsewhere⁷³. In addition, adult male testes were dissected and fixed in Carnoy's modified solution (3:1, 100% ethanol: absolute acetic acid). Genomic DNA of adult males and females were extracted from femurs using the phenol/chloroform-based procedure described in Sambrook and Russel⁷⁴.

Illumina sequencing and graph-based clustering of sequencing reads. Paired-ends sequencing (2 × 300) was applied in libraries prepared as recommended by illumina (illumina Inc., San Diego, CA, USA) using Nextera DNA Library Preparation Kit v3 to one male specimen genomic DNA. Library fragments sizes were in the range from 400 to 600 bp and sequencing was performed using Miseq Sequencing System. The obtained reads were preprocessed to check the quality of the reads with FASTQC⁷⁵ and we did a quality filtering with the FASTX-Toolkit suit⁷⁶. The paired-end reads were also trimmed at 150 nt in length, and then were joined using the “fastq-join” software of the FASTX-Toolkit suit⁷⁶ using default options. Based on illumina sequencing we estimated the G + C content of the whole genome using FastQC High Throughput Sequences QC Report version: 0.11.4 (available at <http://www.bioinformatics.babraham.ac.uk/projects/>). To search for satDNAs in the *E. surinamensis* genome, we carried out a graph-based clustering and assembly of these sequences using the RepeatExplorer^{31,32}. Afterward, we searched for clusters that showed repeat graph density in summary output, which is a typical characteristic of satDNAs families in this approach³¹, and refined this search using Dot plot charts implemented in Dotlet⁷⁷.

Isolation and sequence analysis of satDNAs. Clusters with high graph density were submitted to the Tandem Repeats Finder (TRF) algorithm⁷⁸ to identify the DNA sequence that maximized the alignment scores between the different monomers that could be defined in tandem. The TRF alignment parameters were 2, 3, 5 for match, mismatch and indels, respectively, and a minimum alignment score of 50 was required for reporting. Additionally, we used the dotplot graphic alignment tool implemented in Dotlet⁷⁷ to identify monomers of the same family and to confirm the tandem organization. The monomers with maximum length were used as the representative copy for each satDNA family, and as the query sequences for further BLAST (<http://www.ncbi.nlm.gov/Blast/>) and Repbase (<http://www.girinst.org/repbase/>) searches to check similarity with published sequences. Sequence alignments of satDNAs copies were performed using Muscle⁷⁹ implemented in MEGA5⁸⁰. MEGA5 was also used to estimate nucleotide divergence (*p* distance), A + T content and perform repeat length analysis. The evolutionary relationships among sequences were inferred by neighbor-joining (NJ) trees using the implemented option in MEGA5 and the proportion of nucleotide differences (*p* distance). Satellite alignments are available upon request to the author.

The consensus sequences of each satDNA family was used to design primers with opposite directions (Supplementary Table S4), using the Primer3 software⁸¹ or manually. In order to verify the presence of satDNAs families in male and female, we performed polymerase chain reactions (PCR). PCRs were carried out using 10× PCR Rxn Buffer, 0.2 mM MgCl₂, 0.16 mM dNTPs, 2 mM of each primer, 1 U of *Taq* Platinum DNA Polymerase (Invitrogen, San Diego, CA, USA) and 50–100 ng/μl of template DNA. The PCR conditions included an initial denaturation at 94 °C for 5 min and 30 cycles at 94 °C (30 s), 55 °C (30 s), and 72 °C (80 s), plus a final extension at 72 °C for 5 min. The PCR products were visualized on a 1% agarose gel. The monomeric bands were isolated and purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corp., The Epigenetics Company, USA) according to the manufacturer's recommendations and then used as source for reamplification.

To check the isolated sequences, the purified PCR products were sequenced in both directions using the service of the Macrogen Inc., and then compared to the consensus sequences obtained by the genomic analysis. The consensus sequences for each satDNAs family can be found in Supplementary Results S1, and sequence alignments are available upon request.

Probes and fluorescence *in situ* hybridization (FISH). PCR products for each satDNA family were labeled by nick translation using biotin-14-dATP (Invitrogen) or digoxigenin-11-dUTP (Roche, Mannheim, Germany). SatDNAs with less than 60 bp were labelled directly at the 5' end with biotin-14 dATP (Sigma-Aldrich, St Louis, MO, USA) during their synthesis. Single or two-color FISH was carried out according to Pinkel *et al.*⁸² with modifications⁸³ using mitotic chromosome preparations. Fiber-FISH experiments were conducted as described in de Barros *et al.*⁸⁴ using suspensions of testis cells. The probes that were labeled with digoxigenin-11-dUTP were detected using anti-digoxigenin-rhodamine (Roche) and the probes labeled with biotin-14-dATP were detected using streptavidin conjugated with Alexa Fluor 488 (Invitrogen).

Following FISH, chromosomal preparations were counterstained using 4',6-diamidino-2'-phenylindole (DAPI) and mounted in VECTASHIELD (Vector, Burlingame, CA, USA). Chromosomes and hybridization signals were observed using an Olympus BX61 fluorescence microscope equipped with appropriate filter sets. Black-and-white images were recorded using a DP71 cooled digital camera. The images were pseudo-colored in blue (chromosomes) and red or green (signals), merged and optimized for brightness and contrast using Adobe Photoshop CS2.

Quantitative analysis of satDNAs. Quantitative PCR (qPCR) using male and female genomic DNA as templates was used to check the copy number differences between males and females of selected satDNAs, i.e., some of the repetitive families larger than 60 bp length (see primers in Supplementary Table S4). The selected satDNAs were chosen due to their presence on the neo-Y chromosome determined by FISH. The qPCR of male and female genomic DNA was used to calculate the satDNA dose by a Δ Ct method of relative quantification⁸⁵. Gene dosage ratios (GDR) of the target satDNAs were compared with a reference gene, 70-kDa heat shock protein (*Hsp-70*) family using as primers F: 5'-GGTGGTATGACCACTCTTATCAA-3' and R: 5'-CACTTCATTTTGAGGCACACC-3' that were designed according to the *Hsp-70* gene from *Locusta migratoria* (accession number AY178988). This gene was used as a reference because there were no differences detected in amplification rates between sexes in *E. surinamensis*, suggesting that the *Hsp-70* gene is autosomal and has equal copy number in both sexes. Because we were not sure whether *Hsp-70* is a single-copy gene, the quantitative analysis was a relative comparison of the gene dose between male and female. Quantitative analyses were carried out in MicroAmp® Fast Optical 96-Well Reaction Plate with Barcode (0.1 mL) (Applied Biosystems, Life Technology™, Carlsbad, CA) covered by Optical Adhesive Covers (Applied Biosystems) using the StepOne Real-Time PCR Systems Thermal cycler. qPCR in both the target satDNAs and the reference gene were performed simultaneously in triplicates of three independent samples, i.e. 3 males and 3 females. Each qPCR mixture contained 6.25 μl 2× GoTaq® qPCR master mix (Promega, Madison, WI, USA), 0.25 mM of each forward and reverse primers and 30 ng of either male or female genomic DNA, in a final volume of 10 μl. qPCR mixtures without DNA served as negative controls. The cycling conditions were 95 °C for 10 min, 40 cycles of 95 °C for 15 s, and 60 °C for 1 min. Specificity of the PCR products was confirmed by analysis of the dissociation curve. A correlation analysis of GDR means of target satDNAs between males and females were carried out with R statistical software version 3.3.1⁸⁶ and edited using Adobe Photoshop CS2.

Statistic correlation analysis. We used the PerformanceAnalytics package⁸⁷ implemented in the R statistical software version 3.3.1⁸⁶ to calculate Spearman's rank correlation coefficients for satDNA monomer size, genome proportion, A + T content, nucleotide divergence and loci number. For the satDNAs showing spread FISH signals (see Supplementary Tables S1 and S2), the total countable loci number was increased by 10 to account for their dispersed profile, but still in a conservative manner to avoid counting the possible existing nonspecific hybridization signals.

References

- Charlesworth, B., Sniegowski, P. & Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**, 215–220 (1994).
- Plohl, M., Luchetti, A., Meštrović, N. & Mantovani, B. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* **409**, 72–82 (2008).
- López-Flores, I. & Garrido-Ramos, M. A. Repetitive DNA in The Repetitive DNA Content of Eukaryotic Genomes (ed. Garrido-Ramos, M. A.) 1–28 (Genome Dyn. Basel, Karger, vol 7 (2012).
- Kuhn, G. C. S., Küttler, H., Moreira-Filho, O. & Heslop-Harrison, J. S. The 1.688 repetitive DNA of *Drosophila*: concerted evolution at different genomic scales and association with genes. *Mol. Biol. Evol.* **29**, 7–11 (2012).
- Larracuente, A. M. The organization and evolution of the Responder satellite in species of the *Drosophila melanogaster* group: dynamic evolution of a target of meiotic drive. *BMC Evol. Biol.* **14**, 233 (2014).
- Schmidt, T. & Heslop-Harrison, J. S. Genomes, genes and junk: the large-scale organization of plant chromosomes. *Trends in Plant Sci.* **3**, 195–199 (1998).
- Palomeque, T. & Lorite, P. Satellite DNA in insects: a review. *Heredity* **100**, 564–573 (2008).
- Richard, G. F., Kerrest, A. & Dujon, B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.* **72**, 686–727 (2008).
- Wang, S., Lorenzen, M. D., Beeman, R. W. & Brown, S. J. Analysis of repetitive DNA distribution patterns in the *Tribolium castaneum* genome. *Genome Biol.* **9**, R61 (2008).
- King, K., Jobst, J. & Hemleben, V. Differential homogenisation and amplification of two satellite DNAs in the genus *Cucurbita* (Cucurbitaceae). *J. Mol. Evol.* **41**, 996–1005 (1995).
- Wang, Z. X., Kurata, N., Saji, S., Katayose, Y. & Minobe, Y. A chromosome 5-specific repetitive DNA-sequence in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **90**, 907–913 (1995).
- Vershinin, A. V. & Heslop-Harrison, J. S. Comparative analysis of the nucleosomal structure of rye, wheat and their relatives. *Plant Mol. Biol.* **36**, 149–161 (1998).
- Dover, G. A. Molecular drive: a cohesive mode of species evolution. *Nature* **299**, 111 (1982).
- Dover, G. Molecular drive. *Trends Genet.* **18**, 587–589 (2002).
- Ohno, S. Sex Chromosomes and sex linked genes, Springer, Berlin (1967).
- Bull, J. J. Evolution of Sex Determining Mechanisms, Menlo Park, CA, Benjamin Cummings (1983).
- Rice, W. R. Evolution of the Y sex chromosome in animals. *BioScience* **46**, 331–343 (1996).
- Steinemann, M. & Steinemann, S. The enigma of Y chromosome degeneration: TRAM, a novel retrotransposon is preferentially located on the neo-Y chromosome of *Drosophila miranda*. *Genetics* **145**, 261–266 (1997).
- Steinemann, S. & Steinemann, M. Retroelements: tools for sex chromosome evolution. *Cytogenet. Genome Res.* **110**, 134–143 (2005).
- Bachtrog, D. A dynamic view of sex chromosome evolution. *Curr. Opin. Genetic Dev.* **16**, 578–85 (2006).
- Kaiser, V. B. & Bachtrog, D. 2010. Evolutions of sex chromosome in insects. *Annu. Rev. Genet.* **44**, 91–112 (2010).
- White, M. J. D. Animal cytology and evolution, Cambridge: Cambridge University Press (1997).
- Hewitt, G. M. Grasshoppers and crickets, Animal Cytogenetics, vol 3, Insecta 1, Orthoptera, Gebrüder Borntraeger, Berlin (1979).
- Mesa, A., Fontanetti, C. S. & García Novo, P. Does an X-autosome centric fusion in Acridoidea condemn the species to extinction? *J. Orthoptera Res.* **10**, 141–146 (2001).
- 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 Evol. Biol.* **13**, 167 (2013).
- Palacios-Gimenez, O. M., Martí, D. A. & Cabral-de-Mello, D. C. Neo-sex chromosomes of *Ronderosia bergi*: insight into the evolution of sex chromosomes in grasshoppers. *Chromosoma* **124**, 353–365 (2015a).
- Palacios-Gimenez, O. M. & Cabral-de-Mello, D. C. Repetitive DNA chromosomal organization in the cricket *Cycloptiloides americanus*: a case of the unusual X1X20 sex chromosome system in Orthoptera. *Mol. Genet. Genomics* **290**, 623–631 (2015).
- Palacios-Gimenez, O. M., Carvalho, C. R., Ferrari Soares, F. A. & Cabral-de-Mello, D. C. Contrasting the chromosomal organization of repetitive DNAs in two Gryllidae crickets with highly divergent karyotypes. *PLoS ONE* **10**, e0143540 (2015b).
- Bugrov, A. G., Jetybayev, I. E., Karagyan, G. H. & Rubtsov, N. B. Sex chromosome diversity in Armenian toad grasshoppers (Orthoptera, Acridoidea, Pamphagidae). *Comp. Cytogen.* **10**, 45–59 (2016).
- Ruiz-Ruano, F. J., López-León, M. D., Cabrero, J. & Camacho, J. P. M. High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci. Reports.* **6**, 28333 (2016).
- Novak, P., Neumann, P. & Macas, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* **11**, 378 (2010).
- Novak, P., Neumann, P., Pech, P., Steinhaisl, J. & Macas, J. RepeatExplorer: a galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **29**, 792–793 (2013).
- Barghini, E. The peculiar landscape of repetitive sequences in the Olive (*Olea europaea* L.) genome. *Genome Biol. Evol.* **6**, 776–791 (2014).
- Pavlek, M., Gelfand, Y., Plohl, M. & Meštrović, N. Genome-wide analysis of tandem repeats in *Tribolium castaneum* genome reveals abundant and highly dynamic tandem repeat families with satellite DNA features in euchromatic chromosomal arms. *DNA Res.* **22**, 387–401 (2015).
- Heitkman, T. et al. Next-generating sequencing reveals differentially amplified tandem repeats as major genome component of Europe's oldest *Camellia japonica*. *Chromosome Res.* **23**, 791–806 (2015).
- Kuhn, G. C. S. Satellite DNA transcripts have diverse biological roles in *Drosophila*. *Heredity* **115**, 1–2 (2015).
- Hsu, T. C., Pathak, S., Basen, B. M. & Stahl, G. J. Induced Robertsonian fusions and tandem translocations in mammalian cell cultures. *Cytogenet. Cell. Genet.* **21**, 86–98 (1978).
- Holmquist, G. P. & Dancis, B. Telomere replication, kinetochore organizers, and satellite DNA evolution. *Proc. Natl. Acad. Sci. USA* **76**, 4566–4570 (1979).
- Singer, D. & Downhower, L. Highly repeated DNA of the baboon: organization of sequences homologous to highly repeated DNA of the African green monkey. *J. Mol. Biol.* **134**, 835–842 (1979).
- Bachmann, L., Venanzetti, F. & Sbordoni, V. Tandemly repeated satellite DNA of *Dolichopoda schiavazzii*: a test for models on the evolution of highly repetitive DNA. *J. Mol. Evol.* **43**, 135–144 (1996).
- Stephan, W. & Cho, S. Possible role of natural selection in the formation of tandem-repetitive noncoding DNA. *Genetics* **156**, 333–341 (1994).
- Cohen, S. & Segal, D. Extrachromosomal circular DNA in Eukaryotes: possible involvement in the plasticity of tandem repeats. *Cytogenet. Genome Res.* **124**, 327–338 (2009).
- Inukai, T. Role of transposable elements in the propagation of minisatellites in the rice genome. *Mol. Genet. Genomics* **271**, 220–227 (2004).
- Lal, S., Oetjens, M. & Hannah, L. C. Helitrons: enigmatic abductors and mobilizers of host genome sequences. *Plant Sci.* **176**, 181–186 (2009).
- Macas, J., Koblížková, A., Navrátilová, A. & Neumann, P. Hypervariable 3' UTR region of plant LTR-retrotransposons as a source of novel satellite repeats. *Gene* **448**, 198–206 (2009).
- Smykal, P., Kalendar, R., Ford, R., Macas, J. & Griga, M. Evolutionary conserved lineage of Angela-family retrotransposons as a genome-wide microsatellite repeat dispersal agent. *Heredity* **103**, 157–167 (2009).

47. Thomas, J., Vadnagara, K. & Pritham, E. J. *DINE-1*, the highest copy number repeats in *Drosophila melanogaster* are non-autonomous endonuclease-encoding rolling-circle transposable elements (*Helentrons*). *Mobile DNA* **5**, 18 (2014).
48. Palomeque, T., Muñoz-Loópez, M., Carrillo, J. A. & Lorite, P. Characterization and evolutionary dynamics of a complex family of satellite DNA in the leaf beetle *Chrysolina carnifex* (Coleoptera, Chrysomelidae). *Chromosome Res.* **13**, 795–807 (2005).
49. Mravinac, B., Ugarković, D., Franjević, D. & Ploh, I. M. Long inversely oriented subunits form a complex monomer of *Tribolium brevicornis* satellite DNA. *J. Mol. Evol.* **60**, 513–525 (2005).
50. Koga, A. *et al.* Evolutionary origin of higher-order repeat structure in alpha-satellite DNA of Primate centromeres. *DNA Res.* 1–9 (2014).
51. Zinic, S. D., Ugarkovic, D., Cornudella, L. & Plohl, M. A novel interspersed type of organization of satellite DNAs in *Tribolium madens* heterochromatin. *Chromosome Res.* **8**, 201–212 (2000).
52. Kuhn, G. C. S., Sene, F. M., Moreira-Filho, O., Schwarzacher, T. & Heslop-Harrison, J. S. Sequence analysis, chromosomal distribution and long-range organization show that rapid turnover of new and old pBuM satellite DNA repeats leads to different patterns of variation in seven species of the *Drosophila buzzatii* cluster. *Chromosome Res.* **16**, 307–324 (2008).
53. Reed, K. M., Beukeboom, L. W., Eickbush, D. G. & Werren, J. H. Junctions between repetitive DNAs on the PSR chromosome of *Nasonia vitripennis*: association of palindromes with recombination. *J. Mol. Evol.* **38**, 352–362 (1994).
54. Kuhn, G. C. S., Teo, C. H., Schwarzacher, T. & Heslop-Harrison, J. S. Evolutionary dynamics and sites of illegitimate recombination revealed in the interspersion and sequence junctions of two nonhomologous satellite DNAs in cactophilic *Drosophila* species. *Heredity* **102**, 453–464 (2009).
55. Alkhimova, O. G. *et al.* Diverse patterns of the tandem repeat organization in rye chromosomes. *Chromosoma* **113**, 42–52 (2004).
56. Paço, A., Adegá, F., Meštrović, N., Plohl, M. & Chaves, R. The puzzling character of repetitive DNA in *Phodopus* genomes (Cricetidae, Rodentia). *Chromosome Res.* **23**, 427–440 (2015).
57. Lepesant, J. M. J. *et al.* Chromatin structural changes around satellite repeats on the female sex chromosome in *Schistosoma mansoni* and their possible role in sex chromosome emergence. *Genome Biol.* **13**, R14 (2012).
58. Matsunaga, S. Junk DNA promotes sex chromosome evolution. *Heredity* **102**, 525–526 (2009).
59. Deshpande, N. & Meller, V. H. Sex chromosome evolution: life, death and repetitive DNA. *Fly* **8**, 197–199 (2015).
60. Lohe, A. R., Hilliker, A. J. & Roberts, P. A. Mapping simple repeated DNA sequences in heterochromatin of *Drosophila melanogaster*. *Genetics* **134**, 1149–1174 (1993).
61. Fröncke, L. & Scherthan, H. Zoo-fluorescence *in situ* hybridization analysis of human and Indian muntjac karyotypes (*Muntiacus muntjak vaginalis*) reveals satellite DNA clusters at the margins of conserved syntenic segments. *Chromosome Res.* **5**, 254–261 (1997).
62. Hartmann, N. & Scherthan, H. Characterization of ancestral chromosome fusion points in the Indian muntjac deer. *Chromosoma* **112**, 213–220 (2004).
63. Navajas-Pérez, R. *et al.* Reduced rates of sequence evolution of Y-linked satellite DNA in *Rumex* (Polygonaceae). *J. Mol. Evol.* **60**, 391–399 (2005).
64. Navajas-Pérez, R. *et al.* The origin and evolution of the variability in a Y-specific satellite-DNA of *Rumex acetosa* and its relatives. *Gene* **368**, 61–71 (2006).
65. Steflava, P. *et al.* Contrasting patterns of transposable element and satellite distribution on sex chromosomes (XY₁Y₂) in the dioecious plant *Rumex acetosa*. *Genome Biol. Evol.* **5**, 769–782 (2013).
66. Kejnovsky, E., Hobza, R., Cermak, T., Kubat, Z. & Vyskot, B. The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity* **102**, 533–541 (2009).
67. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
68. Graves, J. A. M. Evolution of vertebrate sex chromosomes and dosage compensation. *Nature Rev. Genet.* **17**, 33–46 (2016).
69. Sex chromosome evolution: historical insights and future perspectives. *Proc. R. Soc. B.* **284**, 20162806 (2017).
70. Nicolas, M. *et al.* A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biol.* **3**, e4 (2005).
71. Bergero, R., Forrest, A., Kamau, E. & Charlesworth, D. Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes. *Genetics* **175**, 1945–1954 (2007).
72. Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Nature* **286**, 964–967 (1999).
73. Webb, G. C., White, M. J. D., Contreras, N. & Cheney, J. Cytogenetics of the parthenogenetic grasshopper *Warramaba* (formerly *Moraba*) *virgo* and its bisexual relatives, IV, Chromosome banding studies. *Chromosoma* **67**, 309–339 (1978).
74. Sambrook, J. & Russel, D. W. *Molecular cloning: A laboratory manual* (3rd ed.) Cold Spring Harbor Laboratory Press, New York (2001).
75. Andrews, S. FastQC. *A quality control tool for high throughput sequence data* <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2012).
76. Gordon, A. & Hannon, G. J. Fastx-toolkit. *FASTQ/A short-reads pre-processing tools* http://hannonlab.csh.edu/fastx_toolkit (2010).
77. Junier, T. & Pagni, M. Dotlet: diagonal plots in a web browser, *Bioinformatics* **16**, 178–9 (2000).
78. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
79. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
80. Tamura, K. *et al.* MEGA: molecular evolutionary genetics using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
81. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers in *Bioinformatics Methods and Protocols* Humana Press 365–386 (1999).
82. Pinkel, D., Straume, T. & Gray, J. W. Cytogenetic analysis using quantitative, high sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci. USA* **83**, 2934–2938 (1986).
83. 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* **104**, 393–400 (2010).
84. de Barros, A. V. *et al.* Fiber FISH reveals different patterns of high-resolution physical mapping for repetitive DNA in fish. *Aquaculture* **322**, 47–50 (2001).
85. Nguyen, P. *et al.* Neo-sex chromosomes and adaptive potential in tortricid pests. *Proc. Natl. Acad. Sci. USA* **110**, 6931–6936 (2013).
86. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2016).
87. Peterson, B. G. *et al.* PerformanceAnalytics package. *Econometric tools for performance and risk analysis* <http://r-forge.r-project.org/projects/returnanalytics/> (2014).

Acknowledgements

The authors are grateful to “Parque Estadual Edmundo Navarro de Andrade” administration for sample collecting authorization and Antonio Sergio Pascon for technical assistance in obtaining embryos. We thank Dr. Frantisek Marec (Institute of Entomology, Biology Centre CAS, České Budějovice, Czech Republic) for the critical review of a previous version of this manuscript, and Dr. Melvin Bonilla (Harvard T.H. Chan School of Public Health, Boston, MA, USA) for English corrections. To the anonymous reviewers for valuable suggestions. OMPG acknowledges scholarship obtained from Fundação de Amparo a Pesquisa do Estado de São Paulo-FAPESP (process number 2014/02038-8). This study was supported by FAPESP (process number 2014/11763-8) and Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior-CAPES. DCCM was awarded a research productivity fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (process number 304758/2014-0).

Author Contributions

O.M.P.G. designed the experiments, collected specimens and harvest tissues, performed the FISH mapping and the bioinformatics analyses, interpreted the data and drafted the text. G.B.D., L.G.L. and G.C.S.K. performed the bioinformatics analyses and interpreted the data. E.R. and C.M. prepared nucleic acids, sequencing libraries and interpreted the data. D.C.C.M. conceived and designed the experiments, analyzed the data and drafted the text. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-06822-8](https://doi.org/10.1038/s41598-017-06822-8)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017

4.4. Capítulo 4

Satellite DNAs are conserved and differentially transcribed among Gryllus cricket species

Octavio Manuel Palacios-Gimenez, Vanessa Bellini Bardella, Bernardo Lemos, Diogo Cavalcanti Cabral-de-Mello

Publicado na revista *DNA Research* (2017), 0: 1-11

Full Paper

Satellite DNAs are conserved and differentially transcribed among *Gryllus* cricket species

Octavio Manuel Palacios-Gimenez^{1,2}, Vanessa Bellini Bardella¹, Bernardo Lemos², and Diogo Cavalcanti Cabral-de-Mello^{1*}

¹Departamento de Biologia, Instituto de Biociências/IB, UNESP–Univ Estadual Paulista, Rio Claro, São Paulo, Brazil, and ²Program in Molecular and Integrative Physiological Sciences, Department of Environmental Health, Harvard University T. H. Chan School of Public Health, Boston, MA 02115, USA

*To whom correspondence should be addressed. Tel. +55 19 35264152; Fax. +55 19 3534 0009. Email: mellocd@rc.unesp.br

Edited by Dr. Toshihiko Shiroishi

Received 19 June 2017; Editorial decision 13 October 2017; Accepted 19 October 2017

Abstract

Satellite DNA (satDNA) is an abundant class of non-coding repetitive DNA that is preferentially found as tandemly repeated arrays in gene-poor heterochromatin but is also present in gene-rich euchromatin. Here, we used DNA- and RNA-seq from *Gryllus assimilis* to address the content and transcriptional patterns of satDNAs. We also mapped RNA-seq libraries for other *Gryllus* species against the satDNAs found in *G. assimilis* and *G. bimaculatus* genomes to investigate their evolutionary conservation and transcriptional profiles in *Gryllus*. Through DNA-seq read clustering analysis using RepeatExplorer, dotplots analysis and fluorescence *in situ* hybridization mapping, we found that ~4% of the *G. assimilis* genome is represented by 11 well-defined A + T-rich satDNA families. These are mainly located in heterochromatic areas, with some repeats able to form high-order repeat structures. By *in silico* transcriptional analysis we identified satDNAs that are conserved in *Gryllus* but differentially transcribed. The data regarding satDNA presence in *G. assimilis* genome were discussed in an evolutionary context, with transcriptional data enabling comparisons between sexes and across tissues when possible. We discuss hypotheses for the conservation and transcription of satDNAs in *Gryllus*, which might result from their role in sexual differentiation at the chromatin level, heterochromatin formation and centromeric function.

Key words: satellite DNA, FISH mapping, RNA mapping, evolution

1. Introduction

Satellite DNA (satDNA) is an abundant class of non-coding repetitive DNA in most eukaryotic genomes. SatDNAs constitute clustered arrays of tandemly repeated sequences often located in the gene-poor heterochromatin of centromeres and telomeres.^{1–4} SatDNA arrays are also dispersed in eu/heterochromatin of sex chromosomes as for example in the sorrel *Rumex acetosa*,⁵ the flatworm *Schistosoma*

*mansonii*⁶ and the cricket *Eneoptera surinamensis*.⁷ SatDNA can also occur as single or short arrays nearby protein-coding genes within autosomal euchromatin.^{8–13} Clustered and dispersed organizational patterns of satDNA are achieved by multiple mechanisms of non-reciprocal transfer such as unequal crossing-over, intrastrand homologous recombination, gene conversion, rolling-circle replication and transposition.^{14–16} These genetic exchanges lead to

homogenization of arrays over the time in a process called concerted evolution.¹⁴

Even though most satDNA arrays are embedded in tightly condensed heterochromatin that is considered transcriptionally silent and inert, evidence for their transcription has been documented in insects, vertebrates and plants.^{3,17} It has been shown that satDNA could be processed into small RNAs, like siRNAs and piwiRNAs, which are involved, for example, in epigenetic process of heterochromatin formation in organisms as diverse as *Schizosaccharomyces pombe*, *Drosophila*, nematodes and plants.^{18–21} Recently, it has also been shown that the major satDNA TCAST1 plays a role in the modulation of protein-gene expression in the beetle *Tribolium castaneum*.¹² Finally, satDNA arrays adopt specific folding structures, known as high order repeat structures (HORs), in which a block of multiple repeat units form large folding units that are tandemly repeated and attract nuclear proteins.²² This folding makes satellites potential carriers of a ‘chromatin code’ possibly contributing to cell identity and the specificity of chromosome territories.²²

In Orthoptera, satDNAs have been described in about 25 species, including several grasshoppers,^{7,23–28} but also the desert locust *Schistocerca gregaria* in which transcription patterns were investigated²⁵ and the cave cricket genus *Dolichopoda*.²⁹ For crickets, satDNAs were isolated in about 15 species^{7,23,24} and among *Gryllus* two distal satDNAs were isolated from the genome of *G. bimaculatus*³⁰ with one of them (GBH535) observed in other two genus representatives, suggesting conservation.

The genus *Gryllus* is composed of about 94 species (<http://orthoptera.speciesfile.org/Common/basic/Taxa.aspx?TaxonNameID=1122353>) that have been used as models for speciation, behavior and ecological, physiology, developmental biology, population genetics and evolutionary studies over several decades.^{31–47} With respect to chromosomal analysis, the genus presents highly conserved karyotypes with $2n = 29\delta/30\eta$ and $X0\delta/XX$,^{30,48–51} although in some cases diploid number reduction and chromosomal polymorphisms have been reported.^{51,52} In a previous study, we characterized the chromosomes of *G. assimilis* to understand genome organization and evolution of some repetitive DNAs (e.g. 18 S and 5 S rDNA, U1 and U2 snDNA, histone H3 gene, microsatellites arrays and *C₀t*-DNA fractions) and use this species as model for chromosomal and genomic analysis.⁵³ The species presents the typical $2n = 29\delta/30\eta$ karyotype, an $X0\delta/XX\eta$ sex-determining system, C-positive heterochromatin around centromeres and terminal regions that primarily did not reveal A + T or G + C base pair richness. C-positive regions were enriched with the *C₀t*-DNA fraction. The chromosomal localization of 18 S rDNA, 5 S rDNA, U1 and U2 snDNA was shown to be conserved with the occurrence of a small number of clusters, in contrast to the dispersed localization and multiple clusters for histone H3 gene and several microsatellites arrays.⁵³

In the present study, we used DNA- and RNA-seq from a *G. assimilis* inbred line kept in our lab to address the content and transcriptional patterns of satDNAs in the species. Additionally, we mapped the RNA-seq libraries available at NCBI for other *Gryllus* (*G. bimaculatus*, *G. rubens* and *G. firmus*) against the satDNAs found in *G. assimilis* and *G. bimaculatus*³⁰ genomes to investigate the conservation of satDNA and their transcriptional profiles in *Gryllus*. The data regarding satDNAs presence in the genome of *G. assimilis* were discussed in an evolutionary context, with transcriptional data enabling comparisons between the sexes and across tissues when possible. We hypothesize that the functionality of satDNAs in *Gryllus* might result from their role in sexual

differentiation at the chromatin level, heterochromatin formation and centromeric function.

2. Materials and methods

2.1. Samples, chromosome obtaining and genomic DNA extraction

Males and females of *Gryllus assimilis* were obtained from a pool of individuals that had been bred at the Univ. Estadual Paulista–UNESP (Rio Claro, SP, Brazil). Mitotic chromosomes preparations were obtained from embryo neuroblasts using standard procedures described elsewhere⁵⁴ with slight modifications according to Palacios-Gimenez *et al.*⁵³ Genomic DNA of adult males and females were extracted from femurs using the phenol/chloroform-based procedure described in Sambrook and Russel.⁵⁵

2.2. Illumina sequencing and graph-based clustering of sequencing reads

Paired-end sequencing (2x101) were performed in libraries constructed as recommended using Illumina TruSeq DNA PCR-Free kit (Illumina Inc., San Diego, CA, USA) from female genomic DNA. The library fragment sequencing was performed by Macrogen service facility (Macrogen Inc., South Korea) using a HiSeq2000 system. Sequencing reads were preprocessed to check the quality of the reads with FASTQC⁵⁶ and we did a quality filtering with the FASTX-Toolkit suit.⁵⁷ The paired-end reads were joined using the ‘fastq-join’ software of the FASTX-Toolkit suit⁵⁷ using default options. To search for satDNA in *G. assimilis* genome, we performed a graph-based clustering and assembly of these sequences using the standard RepeatExplorer pipeline.^{58,59} Subsequently, we examined those clusters that displayed repeat graph density in the RepeatExplorer summary output to identify satDNAs families.⁵⁸ We refined the identification using dotplot graphics implemented in Dotlet⁶⁰ to confirm their tandem organization. All clusters containing reads with sequences represented above 0.01% of genome proportion (high copy number sequences) were analysed in detail.

2.3. Isolation and sequence analysis of satDNAs

Clusters with high graph density were analysed using Tandem Repeats Finder (TRF) algorithm⁶¹ to identify the DNA sequence that maximized the alignment scores between the different monomers that could be defined in tandem. All clusters have been processed with TRF using alignment parameters 2, 3, 5 for match, mismatch and indels, respectively, and a minimum alignment score of 50. Moreover, we used the dotplot graphic alignment tool implemented in Dotlet⁶⁰ to identify the exact start and end of monomers of the same family and to confirm their tandem organization. The monomer with maximum length was used as the representative copy for each satDNA family, and also as the query sequences in further BLAST (<http://www.ncbi.nlm.gov/Blast/>) and Repbase (<http://www.girinst.org/repbase/>) searches to check similarity with published sequences. Also, these canonical monomers were BLASTed against the satDNAs of the cricket *Eneoptera surinamensis*⁷ and the grasshoppers *Ronderosia bergii*,²⁸ *Locusta migratoria*²⁶ and *Eumigus monticola*.²⁷ Sequence alignments of satDNAs copies were performed using Muscle⁶² implemented in MEGA5.⁶³ MEGA5 was also used to estimate nucleotide divergence (*p* distance), A + T content and to perform repeat length analysis. Evolutionary relationships among sequences were inferred by neighbor-joining (NJ) trees using the

proportion of nucleotide differences (p distance) in MEGA5. To predicted secondary structure of *G. assimilis* satDNAs we used CentroidFold software⁶⁴ with McCaskill inference engine and 2–5 weight for base pairs set parameters as options.

The assembled consensus sequences of each satDNA family was used to design primers with opposite directions (Supplementary File S1), using the Primer3 software⁶⁵ or manually. In order to verify the presence of satDNAs families, we performed polymerase chain reactions (PCR). PCRs were performed using 10× PCR Rxn Buffer, 0.2 mM MgCl₂, 0.16 mM dNTPs, 2 mM of each primer, 1 U of *Taq* Platinum DNA Polymerase (Invitrogen, San Diego, CA, USA) and 50–100 ng/μl of template DNA. The PCR conditions included an initial denaturation at 94°C for 5 min and 30 cycles at 94°C (30 s), 55°C (30 s) and 72°C (80 s), plus a final extension at 72°C for 5 min. The PCR products were visualized on a 1% agarose gel. The monomeric bands were isolated and purified using the ZymoClean™ Gel DNA Recovery Kit (Zymo Research Corp., The Epigenetics Company, USA) according to the manufacturer's recommendations and then used as source for reamplification for subsequent analysis.

To check the isolation of sequences of interest, purified PCR products were Sanger sequenced in both directions using the service of MacroGen Inc., and then compared to consensus sequences obtained by genomic analysis. The monomer consensus sequences belonging to each of the satDNAs families were deposited in the NCBI database under the accession numbers MF991236–MF991248. In addition, the consensus sequences for each satDNAs family can be found in the Supplementary File S2 and sequence alignments are available upon request.

2.4. Probes and fluorescence *in situ* hybridization

PCR products for each satDNA with more than 50 bp were labelled by nick translation using biotin-14-dATP (Invitrogen) or digoxigenin-11-dUTP (Roche, Mannheim, Germany). SatDNAs with less than 50 bp were labelled directly at the 5' end with biotin-14 dATP (Sigma-Aldrich, St Louis, MO, USA) during their synthesis. Single- or two-colour fluorescence *in situ* hybridization (FISH) was performed according to Pinkel et al.⁶⁶ with modifications⁶⁷ using mitotic chromosome preparations. The 18 S rDNA probe from *Dichotomius geminatus*⁶⁷ was used to check the possible overlapping with satDNAs in the secondary constriction of pair 1. The probes labelled with digoxigenin-11-dUTP were detected using anti-digoxigenin-rhodamine (Roche) and the probes labelled with biotin-14-dATP were detected using streptavidin conjugated with Alexa Fluor 488 (Invitrogen).

Following FISH, chromosomal preparations were counterstained using 4', 6-diamidino-2'-phenylindole (DAPI) and mounted in VECTASHIELD (Vector, Burlingame, CA, USA). Chromosomes and hybridization signals were observed using an Olympus BX61 fluorescence microscope equipped with appropriate filter sets. Black-and-white images were recorded using a DP71 cooled digital camera. The images were pseudo-coloured in blue (chromosomes) and red or green (signals), merged and optimized for brightness and contrast using Adobe Photoshop CS2.

2.5. Transcription of satDNAs

We used Illumina RNA-seq reads (2x126) from *G. assimilis* male and female heads, testis and ovary transcriptome projects that are in preparation in our lab (unpublished data) to investigate satDNA transcription in each tissue. For this species, three biological replicates for every tissue were used. For comparative purposes, we

investigated whether the 13 satDNAs from *G. assimilis* identified here were transcribed across available tissues of *G. bimaculatus*, *G. rubens* and *G. firmus* RNA-seq data. We also used the GBH535 and GBH542 satDNAs families isolated from *G. bimaculatus*³⁰ for the same approach. Consensus sequences for each satDNA from *G. assimilis* were used for analysis and for *G. bimaculatus* consensus for satDNA families were generated using clones deposited at NCBI with the follows access numbers: GBH535 family, AB204914–AB204938; GBH542 family, AB204939–AB204951.

We downloaded from NCBI database RNA-seq data from three *Gryllus* species, as follows: mixed ovaries and embryos RNA-seq reads (accession SRX023832), mixed-stage embryos RNA-seq reads (accession SRX0238310) and ovaries RNA-seq reads (accession SRX023831) from *G. bimaculatus*; whole animal samples RNA-seq reads (accession SRX1596750, SRX1596749, SRX1596748, SRX1596747, SRX1596746, SRX1596745, SRX1596744, SRX1596743, SRX1596742, SRX1596741, SRX1596740, SRX1596739, SRX1596738, SRX1596737, SRX1596736, SRX1596735, SRX1596734, SRX1596733, SRX1596732, SRX1596731, SRX1596730, SRX1596729, SRX1596728, SRX1596727) from *G. rubens*; flight muscle from long winged female with histolyzed flight muscle (LWFHFM) samples RNA-seq reads (accession SRX272161, SRX272160, SRX272159), flight muscle from long winged female with functional flight muscle (LWFFM) samples RNA-seq reads (accession SRX272158, SRX272157, SRX272156, SRX272155), fat body from short winged female incapable of flight (FBSWFIF) samples RNA-seq reads (accession SRX272155, SRX272154, SRX272153, SRX272152, SRX272151, SRX272150, SRX272127, SRX272125), and fat body from long winged female with functional flight muscles (FBLWFFM) samples RNA-seq reads (accession SRX272124, SRX272122, SRX272120, SRX272119, SRX272117, SRX272111, SRX272106, SRX272104) from *G. firmus*.

Raw RNA-seq reads from all tissue libraries were mapped to each of the satDNA sequences using Bowtie2⁶⁸ with the parameters *–sensitive* as options. For smaller repeats, as for example Gas1 (11 bp), Gas4 (73 bp), Gas9 (82 bp) and Gas11 (10 bp) the mapping was performed on dimers or several monomers were concatenated until reaching 200 bp length. The mapping results were converted from sorted into binary format using the SAMtools⁶⁹ and the aligned reads were counted using the SAMtools options to compare between sequences and tissues in order to estimate their genome proportion (i.e. the expression value of the number of raw reads that align to a satDNA divided by the total number of raw reads in the sequencing library). The Bowtie2 output was used to estimate the relative abundances of these transcripts with Cufflinks.⁷⁰ The quantification step includes raw read counts and scaled read counts. The scaling method applied was FPKM (fragments per kilo-base of transcript per million mapped reads, the expression value obtained after normalization of read counts by both transcript length and number of mapped reads in each RNA-seq library).

3. Results

3.1. SatDNAs identification and sequence characterization

Illumina DNA sequencing produced 37,297,670 paired-end reads with a total of 3,767,064,670 nucleotides (nt). The G + C content is 38.66% and the ratio of reads that have a phred quality score over 30 (Q30) is 91.3%. Given that the mean genome size of *G. assimilis*

Table 1. Main features of the satDNAs isolated from *G. assimilis* genome

Repeat family	Monomer length (bp)	AT %	Genome proportion %	Nucleotide divergence (\pm SE) %	Reads/Contigs	Total of monomers into clusters	Max. number of tandem arrays per contigs	Total repeat family length (kb)
Gas1	11	54.5	1.35	8.5 (\pm 2.9)	222.81	3, 138	216	34.518
Gas2	205	64.1	1.03	15.5 (\pm 0.9)	95.35	112	4	22.96
Gas3	187	59.4	0.246	9.3 (\pm 1)	90.24	21	2	3.927
Gas4	73	43.8	0.242	30.5 (\pm 2.9)	696.71	51	25	3.723
Gas5	165	63.6	0.212	16.6 (\pm 1.2)	22.08	71	5	11.715
Gas6								
Gas6-1	199	38.7	0.151	7 (\pm 1)	1218.2	8	3	1.592
Gas6-2	200	35.2	0.182	21.7 (\pm 1.7)	915.625	6	3	1.224
Gas7	19	73.7	0.150	6.4 (\pm 1.5)	377.62	157	31	2.983
Gas8								
Gas8-1	179	65.9	0.144	2.8 (\pm 1.3)	724.75	2	2	0.358
Gas8-2	181	62.4	0.127	5.4 (\pm 0.8)	73	12	2	2.172
Gas9	82	46.3	0.049	15.7 (\pm 2.1)	14.25	15	3	1.29
Gas10	161	68.9	0.019	13.2 (\pm 1.6)	21.4	8	2	1.288
Gas11	10	50	0.014	3.9 (\pm 1.1)	279	45	35	0.45

SE, standard deviation.

is 2.13 pg,⁵³ this represents about 1.8x genome coverage. For clustering analysis through RepeatExplorer pipeline^{58,59} we used 4,035,746 Illumina paired-end reads as input. This subset was randomly selected for computational efficiency and returned 277,347 clusters (containing 37% of reads) that corresponded to the most abundant repetitive sequences in *G. assimilis*, including satDNAs and other non-characterized repetitive elements. The number of singletons sequences was 2,540,450 containing 63% of reads. To search for satDNAs, the top 224 most abundant clusters representing repetitive elements in the RepeatExplorer summary output with number of reads above 0.01% of genome proportion were analysed in detail. The abundance for each family was variable ranging from 0.014% to 1.35% of the genome (Table 1).

Through dotplot analysis we confirmed the tandem organization of 13 satDNAs. These were grouped into 11 well-defined satDNA families, named Gas1, Gas2, Gas3, Gas4, Gas5, Gas6-1, Gas6-2, Gas7, Gas8-1, Gas8-2, Gas9, Gas10 and Gas11 according to decreasing abundance. Together, satDNAs comprised about 4% of the female genome, with sequences showing an A + T content ranging from 35.2% to 73.7%. Repetitive monomeric units ranged from 10 to 205 bp long with nucleotide divergence within the families varying from 2.8% to 30.5%. Most satDNAs found here constitute heavy satDNAs due to high A + T content (Table 1). To search for HORs, we recovered and counted the maximum number of tandem arrays per contigs for each satDNA family as possible, using dotplot analysis. Then, we counted the total number of monomers present in each the clusters (Table 1). These results mean that some satDNAs family, e.g. Gas1, Gas4, Gas7 and Gas11, are present in multiples distinct contigs, indicating the presence of distinct sequence subtypes and possibly containing sequences organized into HORs.

For Gas6 satDNAs family we identified two subfamilies, named Gas6-1 and Gas6-2 (Table 1). The size of Gas6-1 is 199 bp while Gas6-2 is 200 bp. Nucleotide divergence between both Gas6-1 and Gas6-2 is 21.4%. Similarly, Gas8 displayed two subfamilies, named Gas8-1 and Gas8-2, with monomer unit ranging from 179 bp and 181 bp, respectively (Table 1). The nucleotide divergence between them is 12.8%. The results of different satDNAs subfamilies in *G. assimilis* genome is supported by NJ trees which showed Gas6-1, Gas6-2, Gas8-1 and Gas8-2 allocated in cluster-specific branches,

indicating that each subfamily is composed of exclusive repeat-variants probably originating from a common ancestor (Supplementary File S3). The search for similarity with other previously described sequences in NCBI BLAST and Repbase for each satDNA showed that Gas9 (82 bp) has 93% of identity with *Gryllus bimaculatus* mRNA of 91 bp length (NCBI access number AK277574.1) and also that Gas10 (161 bp) has 87% of identity with *G. bimaculatus* mRNA of 109 bp length (NCBI access number AK272100.1). The remaining satDNAs did not reveal similarity with any other previously described sequences.

3.2. Chromosomal localization of satDNAs

In most autosomes, satDNAs were located preferentially in pericentromeric regions, extending to the short arm that corresponds to the C-band positive blocks observed by Palacios-Gimenez *et al.*⁵³ (Figs. 1 and 2). This pattern was observed for Gas1 (Fig. 1a), Gas4 (Fig. 1d), Gas7 (Fig. 1f), Gas8 (Fig. 2c and d), Gas9 (Fig. 1g), Gas10 (Fig. 1h), Gas11 (Fig. 1i). Besides pericentromeric signals some small autosomes revealed terminal labelling in both arms, varying from two to six chromosomes, which in some cases extended to pericentromere (Figs. 1a, 1d, 1e-I, 2c and 2d). Moreover, for Gas7 interstitial blocks were noticed, including the pair 1 (Fig. 1f). For Gas1, Gas4, Gas8-1, Gas9, Gas10 and Gas11 no signals were observed in the pair 1 (Figs. 1a, 1d, 1g-I and 2c). Concerning the other four satDNAs, distinct patterns were noticed as follow: Gas2, blocks restrict to pericentromeric region for all chromosomes (Fig. 1b); Gas3, interstitial signals in five pairs (Fig. 1c); Gas5, small terminal blocks in short arm of some chromosomes and occurrence of two small elements with signals in both ends (Fig. 1e); Gas6, the two subfamilies restrict to the secondary constriction of pair 1 (Fig. 2a and b) that also correspond to C-positive heterochromatic band and G + C positive blocks (CMA₃⁺) and location of 18S rDNA (Fig. 2a and b) and U2 snDNA observed by Palacios-Gimenez *et al.*⁵³

For the X chromosome no signals were observed for Gas3 (Fig. 1c), Gas5 (Fig. 1e) and Gas6 (Fig. 2a and b). The other satDNAs occurred at one end of X chromosome (Figs. 1a, 1d, 1f-I, 2c and 2d), except for Gas2 that is pericentromeric (Fig. 1b). Additionally, for Gas8-2 one interstitial block was also noticed (Fig. 2d). The

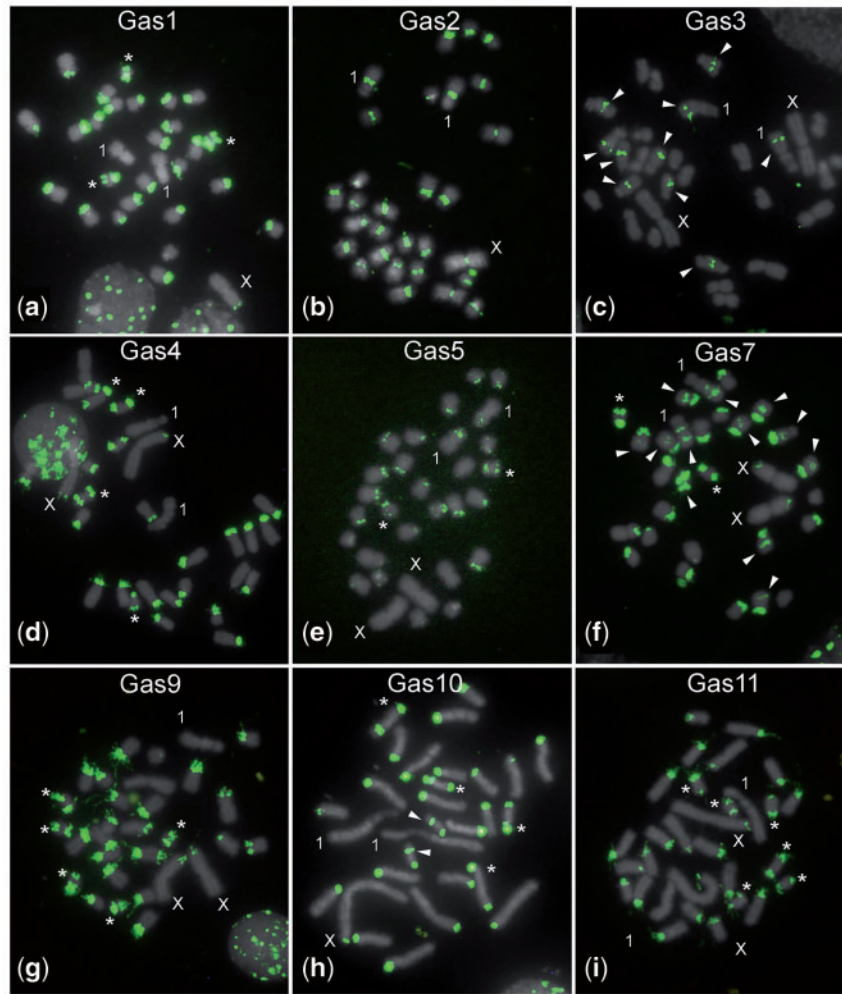


Figure 1. Chromosomal location of nine satDNAs in mitotic chromosomes of embryos of *G. assimilis* by FISH. The satDNA names are shown in the images. Asterisks indicate chromosomes with signals in both termini.

different chromosomal localization of Gas8-1 and Gas8-2 (Fig. 2c and d) highlights that the two subfamilies occupy separate genomic regions.

3.3. Transcription of satDNAs

To explore the possibility of satDNA transcription we used RNA-seq data from four *Gryllus* species: Illumina paired-end reads from *G. assimilis* male heads (143,521,774 reads) and female (156,505,578 reads) heads, testis (133,834,504 reads) and ovary (135,002,056 reads); Illumina paired-end reads from *G. rubens* whole animal (792,280,896 reads); Illumina single line reads from *G. firmus* FBLWFFM (56,930,863), FBSWFIF (40,216,469), LWFFM (25,270,610), LWFHFM (18,443,958) and Titanium 454 GS FLX reads from *G. bimaculatus* mixed ovaries and embryos (1,542,093 reads) and mixed-stage embryos (9,867 reads) and ovaries (8,421 reads). Raw RNA-seq reads from all tissue libraries were mapped to each of the satDNA sequences. The quantification step includes raw read counts and scaled read counts. The scaling method applied was FPKM (fragments per kilo-base of transcript per million mapped reads). The mapping of transcriptomic libraries of *G. assimilis*, *G. bimaculatus*, *G. firmus* and *G. rubens* against the 13 *G. assimilis* satDNAs detected here and the two *G. bimaculatus* satDNAs

(GBH535 and GBH542) published by Yoshimura et al.³⁰ revealed the transcription of satDNAs in several species and tissues.

We found evidence that seven satDNA families (Gas1, Gas3, Gas5, Gas6-1, Gas8-2, Gas11 and GBH535) were transcribed in at least one of the three species (*G. assimilis*, *G. firmus* and *G. rubens*). The degree of expression was variable depending on the repeat mapped, the species, tissue and sex (Fig. 3, Tables 2–4). We did not find evidence for satDNA transcription in *G. bimaculatus* even when we examined satDNA isolated from its own genome (i.e. GBH535 and GBH542) by Yoshimura et al.,³⁰ suggesting either no transcription of satellites and absence of Gas repeats in this species when comparing with the congeneric species.

The repeat Gas1 is transcribed in *G. assimilis*, Gas11 is transcribed in fat body tissues (FBLWFFM and FBSWFIF) of *G. firmus* and Gas6-1 is transcribed in *G. rubens* (Fig. 3a–c). The repeat GBH535 is transcribed only in *G. firmus* and *G. rubens*, but, in the former species, this repeat was transcribed only in FBSWFIF (Fig. 3b and c). Three repeats, Gas3, Gas5 and Gas8-2 are transcribed in the three species, but Gas3 is not expressed in the *G. assimilis* ovary and Gas5 is absent in *G. firmus* LWFFM (Fig. 3a and b); Gas8-2 is transcribed in *G. firmus* FBLWFFM and FBSWFIF, while in *G. assimilis* it is observed only in testis, but in *G. rubens* this repeat is transcribed

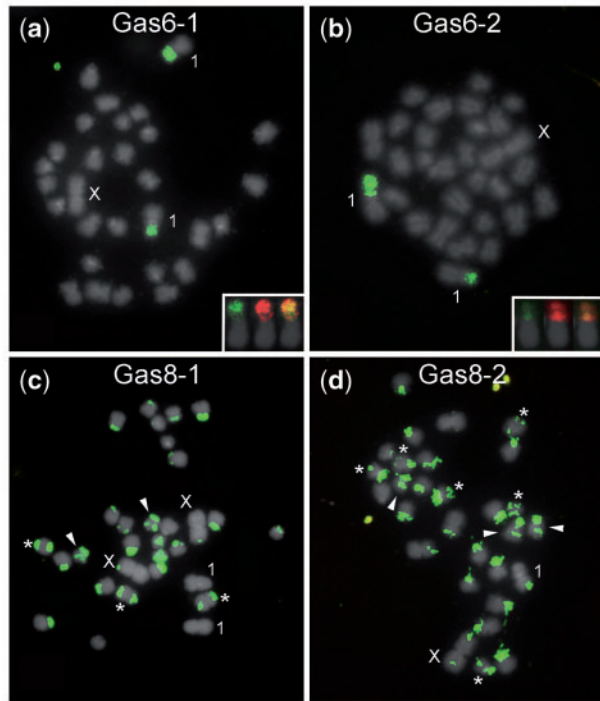


Figure 2. Chromosomal location of the Gas6 (a and b) and Gas8 (c and d) satDNAs subfamilies in mitotic chromosomes of embryos of *G. assimilis* by FISH. The satDNA names are shown in the images. Note the specific chromosomal localization on the pair 1 of the Gas6 subfamilies (a and b) contrasting with the scattered clusters for Gas8 subfamilies (c and d). The insets in (a) and (b) shows the pair 1 with overlapped hybridization signals for the satDNAs and 18S rDNA. Asterisks indicate chromosomes with signals in both termini.

in the whole animal sample (Fig. 3a, b and c). Because the tissues used for transcriptional analysis were all different among the three species we could not assess the species-specific or tissue-specific transcription.

3.4. Secondary structure prediction

Of the 13 satDNAs identified in *G. assimilis*, 7 of them are able to adopt secondary structures with well-defined helices, i.e. Gas2, Gas4, Gas5, Gas6-1, Gas8-1, Gas8-2 and Gas10. The base pairing probability of the secondary structures showed that the abovementioned repeats are able to form short and long helices ranging from 2 bp to 7 bp in length (Supplementary File S4). Interestingly, both transcription and adoption of secondary structure were observed for the repeats Gas5, Gas6-1 and Gas8-2 (Table 2, Fig. 3, Supplementary File S4).

4. Discussion

4.1. Structural organization of satDNAs in *G. assimilis*

Gryllus species have been used as model in a variety of fields, from studies in behavioral ecology, to physiology and genetics. However, information concerning their genome organization and chromosomal aspects are poorly known. A more complete picture concerning chromosomal and genomic organization of distinct DNA classes are important issues that could help in future genomic studies. Our analysis revealed that repetitive DNAs represent a large portion of *G. assimilis* genome (about 40% from 2.13 pg of genome size). This

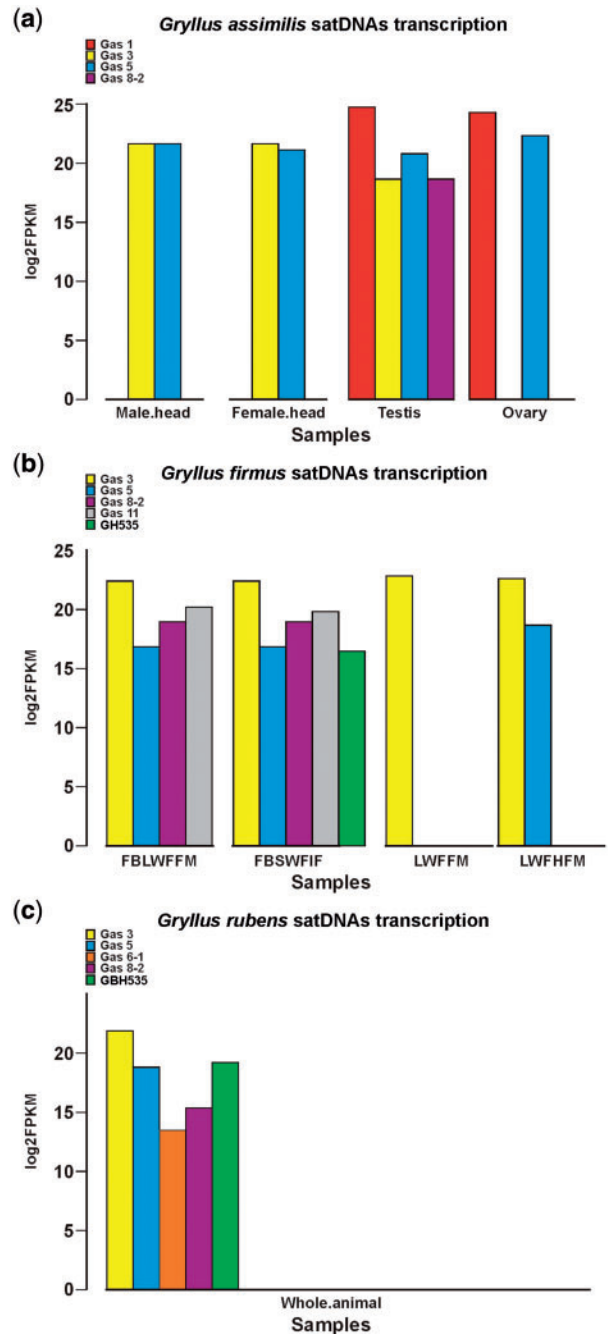


Figure 3. Differential expression of satDNAs between male and female in distinct body parts in the *G. assimilis* genome (a), between different tissues in *G. firmus* female (b) and whole animal samples in *G. rubens* (c). The quantification method applied is FPKM (fragments per kilo-base of transcript per million mapped reads, the expression value obtained after normalization of read counts by both transcript length and number of mapped reads in each RNA-seq library). FBLWFFM, fat body from long winged female with functional flight muscle; FBSWFIF, fat body from short winged female incapable flight; LWFFM, long winged female with functional flight muscle; LWFHFM, long winged female with histolyzed flight muscle.

high abundance of repeats could create gaps and other problems in future genome assembly project in this species, and putatively in other species of the genus should they have similarly high repeat content. These assembly gaps are expected to be particularly problematic

Table 2. Number of raw reads from each tissue sequencing library that align to each *G. assimilis* satDNAs studied and their proportion with respect to the total number of reads (i.e. number of raw reads that align to a satDNA divided by the total number of raw reads in the sequencing library) in the gDNA and in each of the four different tissue transcriptomes obtained by Illumina sequencing

SatDNA	Genome proportion	Transcriptome										
		Male head		Female head		Testis		Ovary				
		Reads	Proportion	FPKM	Reads	Proportion	FPKM	Reads	Proportion	FPKM		
Gas1	0.0135	0	0	0	0	0	0	714	5.3E-06	2.7E+07	2.1E-07	1.9E+07
Gas3	0.00246	62	4.3E-07	3.2E+06	134	8.6E-07	3.1E+06	106	7.9E-07	392,449	0	0
Gas5	0.00212	26	1.8E-07	3.1E+06	70	4.5E-07	2.3E+06	318	2.4E-06	11.8E+06	32	4.7E+06
Gas8-2	0.00127	0	0	0	0	0	0	66	4.9E-07	385,392	0	0
Total reads		143,521,774			156,505,578			133,834,504			135,002,056	

The quantification scaling method applied is FPKM (fragments per kilo-base of transcript per million mapped reads, the expression value obtained after normalization of read counts by both transcript length and number of mapped reads in each RNA-seq library).

in

Table 3. Table showing the number of raw reads from each tissue sequencing library from *G. firmus* female that align to each of the *G. assimilis* (Gas) and *G. bimaculatus* (GB) satDNAs studied and their proportion with respect to the total number of reads (i.e. number of raw reads that align to a satDNA divided by the total number of raw reads in the sequencing library) in each of the four different tissue transcriptomes obtained by Illumina sequencing

SatDNA	Transcriptome	FBSWFIF				LWFFM				LWFFHM			
		Reads	Proportion	FPKM	Reads	Proportion	FPKM	Reads	Proportion	FPKM	Reads	Proportion	FPKM
Gas3	6,200	1.09E-04	5.7E+06	4,690	1.2E-04	5.82E+06	453	1.8E-05	7.7E+06	451	2.4E-05	6.2E+06	
Gas5	86	1.5E-06	117,117	68	1.7E-06	134,350	0	0	0	22	1.2E-06	443,217	
Gas8-2	33	5.8E-07	546,893	12	3.0E-07	475,428	0	0	0	0	0	0	
Gas11	197	3.5E-06	1.23E+06	113	2.8E-06	949,536	0	0	0	0	0	0	
GBH535	0	0	0	17	4.2E-07	91,237.1	0	0	0	0	0	0	
Total reads	56,930,863			40,216,469			25,270,610			18,443,958			

The quantification scaling method applied is FPKM (fragments per kilo-base of transcript per million mapped reads, the expression value obtained after normalization of read counts by both transcript length and number of mapped reads in each RNA-seq library).

FBLWFFM, fat body from long winged female with functional flight muscle; FBSWFIF, fat body from short winged female incapable flight; LWFFM, long winged female with functional flight muscle; LWFFHM, long winged female with hisrolyzed flight muscle.

Table 4. Table showing the number of raw reads from whole animal samples sequencing library from *G. rubens* that align to each of the *G. assimilis* (Gas) and *G. bimaculatus* (GB) satDNAs studied and their proportion with respect to the total number of reads (i.e. number of raw reads that align to a satDNA divided by the total number of raw reads in the sequencing library) in the transcriptome obtained by Illumina sequencing

satDNA	Reads	Proportion	FPKM
Gas3	14,144	1.8E-5	4.1E + 06
Gas5	1,206	1.5E-6	476,399.7
Gas6-1	46	5.8E-08	11,844.3
Gas8-2	142	1.8E-07	44,825.5
GBH535	4,348	5.5E-06	656,004.45
Total reads	792,280,896		

The quantification scaling method applied is FPKM (fragments per kilobase of transcript per million mapped reads, the expression value obtained after normalization of read counts by both transcript length and number of mapped reads in the RNA-seq library).

peri/centromeric chromosomal areas and in short arms of acrocentric chromosomes that as noticed through chromosomal mapping contain large arrays of satDNAs. A prior knowledge of the composition, localization and chromosomal distribution of repeated sequences, including satDNAs as reported here, is therefore needed.

Among crickets, satDNAs have been studied in *Dolichopoda* species,^{24,29,71} *Gryllus bimaculatus*, *G. rubens*, *Gryllus* sp.³⁰ and *Eneoptera surinamensis*.⁷ Although only in *E. surinamensis* the same strategy used here, that allow characterizing the satellites, was used.⁷ Comparative analysis from both satDNAs libraries, i.e. *G. assimilis* and *E. surinamensis*, revealed the absence of similarity between them. The satDNAs of *G. assimilis* were clustered primarily in heterochromatin, coincident with the C₀-DNA fraction.⁵³ Moreover, the chromosomal distribution of *G. assimilis* satDNAs is similar to those A + T-rich GBH535 and GBH542 families described for *G. bimaculatus*,³⁰ but it is distinct from the clustered, dispersed and intermingled pattern seen in the eu/heterochromatin of *E. surinamensis*.⁷ Considering the data from *G. assimilis* and *G. bimaculatus* three features regarding satDNA in this genus are worth highlighting: (i) the predominance of A + T-rich satDNAs families, (ii) the trends of these arrays to occupy more restrict chromosomal localization (i.e. constitutive heterochromatin), like in other Orthoptera, but distinct from *E. surinamensis* and (iii) also the possibility that the predominance of A + T-rich satDNAs could be a common feature in *Gryllus* genomes. This is in contrast with *E. surinamensis*, a species with highly rearranged karyotype,⁷² in which G + C-rich satDNAs predominates, with some of them scattered across multiple genomic locations.⁷

Sequences organized as HORs in which a block of multiple basic repeat units are able to form a larger array unit and larger units are repeated tandemly as observed for Gas1, Gas4, Gas7 and Gas11 is known to be present in centromeres from primates,⁷³ mouse⁷⁴ and insects.²³ The occurrence of such structures at the peri/centromeres levels, their sequence conservation (low nucleotide divergence) and wide evolutionary distribution suggests their involvement in structural and functional organization of centromeres^{17,75} with putative structural function for *G. assimilis* chromosomes.

A few features of the chromosomal distribution of satDNA repeats in *Gryllus* are evident. Some satDNAs are located in multiple chromosomes, while other are located in fewer chromosomes (Gas3) or even exclusive in specific chromosomal elements (Gas6). This

suggests differential dynamics for expansion and clusters dispersion of distinct repeats. Considering the observed satDNA distribution some explanations could be addressed. First, some satellites are coincident with C-positive heterochromatin (i.e. Gas1, Gas4, Gas7, Gas8, Gas9, Gas10 and Gas11), revealing the contribution of this class of repetitive DNA for the amount and complexity of heterochromatin in *Gryllus*. Second, the overlapping of Gas6-1 and Gas6-2 along with 18 S rDNA and U2 snDNA in the secondary constriction of pair,⁵³ reveals the diverse constitution of this region. Third, the localization of Gas2 in all centromeres of *G. assimilis* suggests the possible involvement of this repeat for centromere function. Fourth, interstitial satDNAs placed in C-negative areas indicate the possible occurrence of heterochromatin, not revealed by C-banding, which has low complexity. Fifth, chromosomal distribution of satDNAs (i.e. Gas1, Gas4, Gas5, Gas9, Gas10 and Gas11) in both centromeric and telomeric regions of some acrocentric chromosomes suggests putative inversions involving repeats near centromeres leading to intrachromosomal dispersion of satDNAs. The putative inversions are reinforced by occurrence of karyotypes variable in morphology among *Gryllus* species.^{30,53} Finally, we noted that such differential chromosomal distribution of repeats is not strictly necessary for the emergence of new repetitive variants, as is evident when the subfamilies of Gas6 and Gas8 are compared. For example, Gas6-1 and Gas6-2 are restricted to the pair 1 suggesting common origin within the same chromosome. These repeats diverged by accumulation of mutations in each array followed by amplification and spreading involving well-known molecular mechanisms, like unequal crossing-over, intrastrand homologous recombination, gene conversion, rolling-circle replication and transposition.¹⁴⁻¹⁶ In contrast, Gas8-1 and Gas8-2 are placed and distributed on several different chromosomes suggesting the possibility of emergence of new repertoires from multiple chromosomes.

As expected due to the euchromatic nature of the *G. assimilis* X chromosome—with heterochromatin restrict only to terminal regions—the satDNAs were placed primarily at the ends of this chromosome, and only Gas8-2 presented one interstitial block on the X. Low accumulation of satDNA in the X chromosome of *G. bimaculatus* was also noticed, with signals restrict to interstitial heterochromatin.³⁰ These data contrast with the high accumulation of repeats in sex chromosomes of *E. surinamensis*, with a highly differentiated neo-X₁X₂Y.⁷

4.2. Transcription activity of satDNA in cricket species reveals wide evolutionary conservation of sequences and putative functionality

SatDNAs transcription has been detected in several organisms, with growing evidence pointing to the importance of this kind of non-coding sequence as global genome regulators.^{3,17,76} For instance, in chicken, zebrafish and in the newts *Triturus cristaceus carnifex* and *Notophthalmus viridescens*, satDNAs are transcribed throughout embryogenesis.¹⁷ In many insect species, satDNAs are also expressed throughout the development, and display expression differences between tissues and between sexes.^{3,17,25,76} Here, we found evidence that seven satDNAs which includes those isolated here from *G. assimilis* (i.e. Gas1, Gas3, Gas5, Gas6-1, Gas8-2, Gas11) and those isolated from *G. bimaculatus* (i.e. GBH535),³⁰ are shared among crickets species but they are differentially transcribed in different body parts as well as between sexes. It means wide evolutionary conservation of satDNAs among cricket species after satDNA library divergence possibly as a consequence of its functionality.

The seven transcribed satDNAs are presents as mRNAs polyadenylated, though we are not sure if such transcripts are exported to the cytoplasm or kept in the nucleus. Three repeats (Gas3, Gas5 and Gas8-2) are commonly transcribed in the genome of *G. assimilis*, *G. firmus* and *G. rubens* and considering the chromosomal localization of these repeats in the constitutive heterochromatin of *G. assimilis*, such transcripts can be important to heterochromatin formation. In this way, satDNA transcripts are processed into small interfering RNAs (siRNAs) participating in the heterochromatin formation and control of gene expression.¹⁷

In other cases, we observed that Gas satDNAs show differential transcription for different tissues among species. The library sizes and construction protocol could account for these differences. Because of the tissues studied were all different among the three species we could not speculate about species-specific or tissue-specific transcription. The Gas1 was transcribed in testis and ovary and no transcription of this element is seen in male or female heads. For Gas8-2 it is highlighted differential transcription in gonads probably due to the highly specificity of the tissues studied. These findings suggest possible involvement of satDNAs (Gas1 and Gas8-2) during male and female meiosis or gonad maturation and function. However, further experimental evidences are needed to test this possibility. In other cases when the same tissues were compared, we observed that GBH535 is transcribed in the FBSWFIF but not in the FBLWFFM. Furthermore, Gas5 is transcribed in the LWFHFM but not in the LWFFM. This finding seems to indicate some relation between transcription of satDNAs and flying ability in *G. firmus*, although it deserves experimental validation. Bearing in mind the high diversity of satDNAs and their transcripts, several sequence-specific regulatory signals might reside within them acting as bar code allowing the cell to identify specific chromosome territories. For example, these signals can involve DNAs, RNAs or proteins as well as secondary or tertiary structures of RNA-mediated catalysis, as noticed for example in cave cricket²⁹ and beetle.¹²

Through *in silico* analysis of *G. assimilis* satDNAs we found that seven of them are able to adopt secondary structures with well-defined stem-loops of double-stranded RNA stretches. Moreover, three of them showed to be transcribed adopting stem-loops of double-stranded RNA stretches. Such secondary structures that they adopt could determine RNA-protein interactions, suggesting functional roles. Transcripts of satDNAs able to adopt hummer-head like structures have also been detected, for example, in salamanders, schistosomes and *Dolichopoda* cave cricket species.^{3,17,29} It has been demonstrated that these hummer-head like structures can function as rybozymes with self-cleavage activities, though the physiological role of them is unclear and intriguing.^{17,29} In the light of such evidence, it is also possible that the folding helps to satDNA dispersion along the genome by rolling-circle replication mechanism, in which circular monomer result from secondary structure RNA processing into linear monomers and subsequently circularization by a host-specific RNA ligase.^{77,78}

Our finding prospects raise that the high repertoire of satDNAs and its transcripts might have relevance in organization and regulation of genomes, which will set the stage for further functional genome analyses in *Gryllus*. Bearing in mind the transcription of satellites it is plausible that the transcripts could be responsible for epigenetic chromatin modification as well as might have effect on the gene expression. Although further research in this area is needed, our structural and functional study provides an important step to understand the biology of satDNAs in Orthoptera, highlighting the

importance of crickets as classical model organisms for evolutionary studies.

Acknowledgements

The authors are grateful to Antonio Sergio Pascon for technical assistance in animal breeding. They also thank Dr Meng Wang (Harvard T.H. Chan School of Public Health, Boston, MA, USA) for many valuable suggestions in bioinformatics and help. They also thank the anonymous reviewers for valuable comments. The computations in this article were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University.

Funding

O.M.P.G. was supported by Research Internships Abroad (BEPE) fellowships from Fundação de Amparo a Pesquisa do Estado de São Paulo-FAPESP (process number 2016/01506-3). This study was supported by FAPESP (process numbers 2014/11763-8, 2015/16661-1) and Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior-CAPES. DCCM was awarded a research productivity fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (process number 304758/2014-0).

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at *DNARES* online

References

- Charlesworth, B., Sniegowski, P. and Stephan, W. 1994, The evolutionary dynamics of repetitive DNA in eukaryotes, *Nature*, **371**, 215–20.
- Ugarkovic, D. and Pohl, M. 2002, Variation in satellite DNA profiles-causes and effects, *EMBO J.*, **21**, 5955–9.
- Pezer, Z., Brajković, J., Feliciello, I. and Ugarković, D. 2011, Transcription of satellite DNAs in insects, *Prog. Mol. Subcell. Biol.*, **51**, 161–78.
- Khost, D.E., Eickbush, D.G. and Larracuent, A.M. 2017, Single-molecule sequencing resolves the detailed structure of complex satellite DNA loci in *Drosophila melanogaster*, *Genome Res.*, **27**, 709–21.
- Steflová, P., Tokan, V., Vogel, I., Lexa, M., Macas, J. and Novak, P. 2013, Contrasting patterns of transposable element and satellite distribution on sex chromosomes (XY₁Y₂) in the dioecious plant *Rumex acetosa*, *Genome Biol. Evol.*, **5**, 769–82.
- Lepesant, J.M.J., Cosseau, C. and Boissier, J. 2012, Chromatin structural changes around satellite repeats on the female sex chromosome in *Schistosoma mansoni* and their possible role in sex chromosome emergence, *Genome Biol.*, **13**, R14.
- Palacios-Gimenez, O.M., Dias, G.B., Gomes de Lima, L., et al. 2017, High-throughput analysis of the satellitome revealed enormous diversity of satellite DNAs in the neo-Y chromosome of the cricket *Eneoptera surinamensis*, *Sci. Rep.*, **7**, 6422.
- Kuhn, G.C.S., Küttler, H., Moreira-Filho, O. and Heslop-Harrison, J.S. 2012, The 1.688 repetitive DNA of *Drosophila*: concerted evolution at different genomic scales and association with genes, *Mol. Biol. Evol.*, **29**, 7–11.
- Brajković, J., Feliciello, I., Bruvo-Madarić, B. and Ugarković, Đ. 2012, Satellite DNA-like elements associated with genes within euchromatin of the beetle *Tribolium castaneum*, *G3*, **2**, 9310–941.

10. Larracunte, A.M. 2014, The organization and evolution of the Responder satellite in species of the *Drosophila melanogaster* group: dynamic evolution of a target of meiotic drive, *BMC Evol. Biol.*, **14**, 233.
11. Feliciello, L., Akrap, I., Brajković, J., Zlatar, I. and Ugarković, D. 2015, Satellite DNA as a driver of population divergent in the red flour beetle *Tribolium castaneum*, *Genome Biol. Evol.*, **7**, 228–39.
12. Feliciello, L., Akrap, I., Ugarković, D. and Miska, E.A. 2015, Satellite DNA modulates gene expression in the beetle *Tribolium castaneum* after heat stress, *PLoS Genet.*, **11**, e1005466–547.
13. Pavlek, M., Gelfand, Y., Plohl, M. and Meštrović, N. 2015, Genome-wide analysis of tandem repeats in *Tribolium castaneum* genome reveals abundant and highly dynamic tandem repeat families with satellite DNA features in euchromatic chromosomal arms, *DNA Res.*, **22**, 387–401.
14. Dover, G. 1982, Molecular drive: a cohesive mode of species evolution, *Nature*, **299**, 111–7.
15. Dover, G. 2002, Molecular drive, *Trends Genet.*, **18**, 587–9.
16. Walsh, J.B. 1987, Persistence of tandem arrays: implications for satellite and simple-sequence DNAs, *Genetics*, **115**, 553Y567.
17. Ugarković, D. 2005, Functional elements residing within satellite DNAs, *EMBO Rep.*, **6**, 1035–9.
18. Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I. and Martienssen, R.A. 2002, Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi, *Science*, **297**, 1833–7.
19. Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Chikka, M.R., Bhadra, U. and Birchler, J.A. 2004, Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery, *Science*, **303**, 669–72.
20. Grewal, S.I. and Elgin, S.C. 2007, Transcription and RNA interference in the formation of heterochromatin, *Nature*, **447**, 399–406.
21. Fagegaltier, D., Bougé, A.-L., Berry, B., et al. 2009, The endogenous siRNA pathway is involved in heterochromatin formation in *Drosophila*, *Pnas*, **106**, 21258–63.
22. Podgornaya, O., Gavrilova, E., Stephanova, V., Demin, S. and Komissarov, A. 2013, Large tandem repeats make up the chromosome bar code: a hypothesis, *Adv. Protein Chem. Struct. Biol.*, **90**, 1–30.
23. Palomeque, T. and Lorite, P. 2008, Satellite DNA in insects: a review, *Heredity (Edinb)*, **100**, 564–73.
24. Martinsen, L., Venanzetti, F., Johnsen, A., Sbordoni, V. and Bachmann, L. 2009, Molecular evolution of the pDo500 satellite DNA family in *Dolichopoda* cave crickets (Rhopidophoridae), *BMC Evol. Biol.*, **9**, 301.
25. Camacho, J.P.M., Ruiz-Ruano, F.J., Martín-Blázquez, R., et al. 2015, A step to the gigantic genome of the desert locust: chromosome sizes and repeated DNAs, *Chromosoma*, **124**, 263–75.
26. Ruiz-Ruano, F.J., López-León, M.D., Cabrero, J. and Camacho, J.P.M. 2016, High-throughput analysis of the satellitome illuminates satellite DNA evolution, *Sci. Rep.*, **6**, 28333.
27. Ruiz-Ruano, F.J., Cabrero, J., López-León, M.D. and Camacho, J.P.M. 2017, Satellite DNA content illuminates the ancestry of a supernumerary (B) chromosome, *Chromosoma*, **126**, 487–500.
28. Palacios-Gimenez, O.M., Milani, D., Lemos, B., Castillo, E.R.D., Martí, D.A., Ramos, E., et al. 2017, Uncovering the evolutionary history of neo-XY sex chromosomes in the grasshopper *Ronderosia bergii* (Orthoptera, Melanoplinae) through satellite DNA analysis. *BMC Evol. Biol.*, in press.
29. Rojas, A.A., Vazquez-Tello, A., Ferbeyre, G., Venanzetti, F., Bachmann, L., Paquin, B., et al. 2000, Hammerhead-mediated processing of satellite pDo500 family transcripts from *Dolichopoda* cave crickets, *Nucl. Acid. Res.*, **28**, 4037–43.
30. Yoshimura, A., Nakata, A., Mito, T. and Noji, S. 2006, The characteristics of karyotype and telomeric satellite DNA sequences in the cricket, *Gryllus bimaculatus* (Orthoptera, Gryllidae), *Cytogenet. Genome Res.*, **112**, 329–36.
31. Howard, D.J., Marshall, J.L., Hampton, D.D., et al. 2002, The genetics of reproductive isolation: a retrospective and prospective look with comments on ground crickets, *Am. Nat.*, **3**, S8–S21.
32. Fedorka, K.M. and Mousseau, T.A. 2004, Female mating bias results in conflicting sex-specific offspring fitness, *Nature*, **429**, 65–7.
33. Bussiere, L.F., Hunt, J., Jennions, M.D. and Brooks, R. 2006, Sexual conflict and cryptic female choice in the black field cricket, *Teleogryllus commodus*, *Evolution*, **60**, 792–800.
34. Zera, A.J. 2006, Evolutionary genetics of juvenile hormone and ecdysteroid regulation in *Gryllus*: a case study in the microevolution of endocrine regulation, *Comp. Biochem. Physiol. A Mol. Integr. Physiol.*, **144**, 365–79.
35. Andres, J.A., Maroja, L.S. and Harrison, R.G. 2008, Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets, *Proc. Biol. Sci.*, **275**, 1975–83.
36. Andres, J.A., Maroja, L.S., Bogdanowicz, S.M., Swanson, W.J. and Harrison, R.G. 2006, Molecular evolution of seminal proteins in field crickets, *Mol. Biol. Evol.*, **23**, 1574–84.
37. Andres, J.A., Larson, E.R., Bogdanowicz, S.M. and Harrison, R.G. 2013, Patterns of transcriptome divergence in the male accessory gland of two closely related species of field crickets, *Genetics*, **193**, 501–13.
38. Shaw, K.L., Parsons, Y.M. and Lesnick, S.C. 2007, QTL analysis of a rapidly evolving speciation phenotype in the Hawaiian cricket *Laupala*, *Mol. Ecol.*, **16**, 2879–92.
39. Shaw, K.L. and Lesnick, S.C. 2009, Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation, *Proc. Natl. Acad. Sci USA*, **106**, 9737–42.
40. Maroja, L.S., Clark, M.E. and Harrison, R.G. 2008, Wolbachia plays no role in the one-way reproductive incompatibility between the hybridizing field crickets *Gryllus firmus* and *G. pennsylvanicus*, *Heredity*, **101**, 435–44.
41. Hartfelder, K. and Emlen, D.J. 2012, Endocrine control of insect polyphenism, In: Gilbert, L.I., ed., *Insect Endocrinology*, Elsevier: Oxford, pp. 464–522.
42. Danbara, Y., Sakamoto, T., Uryu, O. and Tomioka, K. 2010, RNA interference of timeless gene does not disrupt circadian locomotor rhythms in the cricket *Gryllus bimaculatus*, *J. Insect Physiol.*, **56**, 1738–45.
43. Lynch, J.A., Peel, A.D., Drechsler, A., Averof, M. and Roth, S. 2010, EGF signaling and the origin of axial polarity among the insects, *Curr. Biol.*, **20**, 1042–7.
44. Ellison, C.K., Wiley, C. and Shaw, K.L. 2011, The genetics of speciation: genes of small effect underlie sexual isolation in the Hawaiian cricket *Laupala*, *J. Evol. Biol.*, **24**, 1110–9.
45. Mito, T., Shinmyo, Y., Kurita, K., et al. 2011, Ancestral functions of Delta/Notch signaling in the formation of body and leg segments in the cricket *Gryllus bimaculatus*, *Development*, **138**, 3823–33.
46. Tomioka, K., Uryu, O., Kamae, Y., Umezaki, Y. and Yoshii, T. 2012, Peripheral circadian rhythms and their regulatory mechanism in insects and some other arthropods: a review, *J. Comp. Physiol. B*, **182**, 729–40.
47. Zeng, V., Ewen-Campen, B., Horch, H. W., Roth, S., Mito, T., et al. 2013, Developmental gene discovery in a hemimetabolous insect: de novo assembly and annotation of a transcriptome for the cricket *Gryllus bimaculatus*, *PLoS ONE*, **8**, e61479.
48. Drets, M.E. and Stoll, M. 1974, C-banding and non-homologous associations in *Gryllus argentinus*, *Chromosoma*, **48**, 367–90.
49. Lamborot, M. and Alvarez-Sarret, E. 1985, The chromosomes of *Gryllus* field cricket populations in central Chile (Insecta: Grylloptera: Gryllidae), *Can. J. Zool.*, **63**, 2626–31.
50. Yoshimura, A. 2005, Karyotypes of two american field crickets, *Gryllus rubens* and *Gryllus* sp. (Orthoptera: Gryllidae), *Entomol. Sci.*, **8**, 219–22.
51. Handa, S.M., Mittal, O.P. and Sehgal, S. 1985, Cytology of ten species of crickets from Chandigarh (India), *Cytologia*, **50**, 711–24.
52. Zefa, E. 1999, Autosomal rearrangement in *Gryllus assimilis* Fabricius, 1775 (Orthoptera, Gryllidae), *Genet. Mol. Biol.*, **22**, 333–6.
53. Palacios-Gimenez, O.M., Carvalho, C.R., Ferrari Soares, F.A., Cabral-de-Mello, D.C. and Sharakhov, I.V. 2015, Contrasting the chromosomal organization of repetitive DNAs in two Gryllidae crickets with highly divergent karyotypes, *PLoS One*, **10**, e0143540.
54. Webb, G.C., White, M.J.D., Contreras, N. and Cheney, J. 1978, Cytogenetics of the parthenogenetic grasshopper *Warramaba* (formely

- Moraba*) *virgo* and its bisexual relatives, IV, Chromosome banding studies, *Chromosoma*, **67**, 309–39.
55. Sambrook, J. and Russel, D.W. 2001, *Molecular Cloning: A Laboratory Manual*, 3rd edition. Cold Spring Harbor Laboratory Press: New York.
 56. Andrews, S. 2012, FastQC, A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
 57. Gordon, A. and Hannon, G.J. 2010, Fastx-toolkit, FASTQ/A short-reads pre-processing tools. Unpublished http://hannonlab.cshl.edu/fastx_toolkit
 58. Novak, P., Neumann, P. and Macas, J. 2010, Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data, *BMC Bioinform.*, **11**, 378.
 59. Novak, P., Neumann, P., Pech, P., Steinhaisl, J. and Macas, J. 2013, RepeatExplorer: a galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads, *Bioinformatics*, **29**, 792–3.
 60. Junier, T. and Pagni, M. 2000, Dotlet: diagonal plots in a web browser, *Bioinformatics*, **16**, 178–9.
 61. Benson, G. 1999, Tandem repeats finder: a program to analyze DNA sequences, *Nucleic Acids Res.*, **27**, 573–80.
 62. Edgar, R.C. 2004, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.*, **32**, 1792–7.
 63. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011, MEGA: molecular evolutionary genetics using maximum likelihood, evolutionary distance, and maximum parsimony methods, *Mol. Biol. Evol.*, **28**, 2731–9.
 64. Hamada, M., Kiryu, H., Sato, K., Mituyama, T. and Asai, K. 2009, Prediction of RNA secondary structure using generalized centroid estimators, *Bioinformatics*, **25**, 465–73.
 65. Rozen, S. and Skaletsky, H. 1999, Primer3 on the WWW for general users and for biologist programmers, In: *Bioinformatics Methods and Protocols*, Humana Press: Totowa, NJ, pp. 365–86.
 66. Pinkel, D., Straume, T. and Gray, J.W. 1986, Cytogenetic analysis using quantitative, high sensitivity, fluorescence hybridization, *Proc. Nat. Acad. Sci. USA*, **83**, 2934–8.
 67. Cabral-de-Mello, D.C., Moura, R.C. and 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.
 68. Langmead, B. and Salzberg, S.L. 2012, Fast gapped-read alignment with Bowtie2, *Nat. Methods*, **9**, 357–9.
 69. Li, H., Handsaker, B., Wysoker, A., et al. 2009, The sequence alignment/map format and SAMtools, *Bioinformatics*, **25**, 2078–9.
 70. Trapnell, C., Williams, B.A., Pertea, G., et al. 2010, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, *Nat. Biotechnol.*, **28**, 511–5.
 71. Bachmann, L., Venanzetti, F. and Sbordoni, V. 1996, Tandemly repeated satellite DNA of *Dolichopoda schiavazzii*: a test for models on the evolution of highly repetitive DNA, *J. Mol. Evol.*, **43**, 135–44.
 72. Hewitt, G.M. 1979, Grasshoppers and crickets, *Animal Cytogenetics*, vol 3, Insecta 1, Orthoptera, Gebrüder Borntraeger, Berlin.
 73. Koga, A., Hirai, Y., Terada, S., et al. 2014, Evolutionary origin of higher-order repeat structure in alpha-satellite DNA of Primate centromeres, *DNA Res.*, **21**, 407–15.
 74. Pertile, M.D., Graham, A.N., Choo, K.H.A. and Kalitsis, P. 2009, Rapid evolution of mouse Y centromere repeat DNA belies recent sequence stability, *Genome Res.*, **19**, 2202–13.
 75. Spence, J.M., Critcher, R., Ebersole, T.A., Valdivia, M.M., Earnshaw, W.C., Fukagawa, T. and Farr, C.J. 2002, Co-localization of centromere activity, proteins and topoisomerase II within a subdomain of the major human X α -satellite array, *Embo J.*, **21**, 5269–80.
 76. Dalíková, M., Zrzavá, M., Kubícková, S. and Marec, F. 2017, W-enriched satellite sequence in the Indian meal moth, *Plodia interpunctella* (Lepidoptera, Pyralidae), *Chromosome Res.*, doi 10.1007/s10577-017-9558-8.
 77. Van Tol, H., Buzayan, J.M. and Bruening, G. 1991, Evidence for spontaneous circle formation in the replication of the satellite RNA of tobacco ringspot virus, *Virology*, **180**, 23–30.
 78. Reid, C.E. and Lazinski, D.W. 2000, A host-specific function is required for ligation of a wide variety of ribozyme-processed RNAs, *Proc. Natl. Acad. Sci. USA*, **97**, 424–9.

4.5. Capítulo 5

Insights on sex-determining genes and sex-biased genes expression through de novo transcriptome assembly in field crickets

Octavio Manuel Palacios-Gimenez, Bernardo Lemos, Diogo Cavalcanti Cabral-de-Mello

Submetido na revista *BMC Genomics*

1 **Insights on sex-determining genes and sex-biased genes expression through *de novo***
2 **transcriptome assembly in field crickets**

3

4 Octavio Manuel Palacios-Gimenez^{1,2}, Bernardo Lemos², Diogo Cavalcanti Cabral-de-Mello^{1*}

5

6 ¹UNESP - Univ Estadual Paulista, Instituto de Biociências/IB, Departamento de Biologia, Rio
7 Claro, São Paulo, Brazil

8 ²Program in Molecular and Integrative Physiological Sciences, Department of Environmental
9 Health, Harvard University T. H. Chan School of Public Health, Boston, Massachusetts 02115

10

11 *Running title:* sex-determining genes and sex-biased genes in crickets

12

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

16

17

18

19

20

21

22

23

24 **Abstract**

25 Field crickets species have become models for behavioral ecology, physiological and genetic
26 studies and exhibit interesting features, such as pronounced sexual dimorphism in body size as
27 well as shape and size of reproductive structures. They also have the X0/XX sex-determining
28 system considered ancestral within Orthoptera, plus a huge genome with high density of
29 repetitive DNA. Here, we present a *de novo* transcriptome assembly of the cricket *G. assimilis*
30 and use it to investigate gene expression between males and females. The transcriptomes of two
31 other XX/X0 cricket's species (*G. bimaculatus* and *G. firmus*) and the genome assembly of the
32 desert locust model *Locusta migratoria* were also analyzed and compared. We identified a set of
33 highly expressed genes that might account for phenotypic differences between sexes. Our work
34 enabled comparison of sex-determining genes, and other genes related to sexual fitness, like
35 insect hormone biosynthesis, circadian rhythm, circadian entrainment and circadian rhythm-fly
36 pathway among species of *Gryllus*. The effectors as well as downstream targets of sex-
37 determination pathways have been previously identified in other insects but never in Orthoptera.
38 We estimated that protein-encoding reproductive genes evolve faster than non-reproductive
39 genes as result of strong positive selection at those loci. We document that the species studied
40 harbors exceptionally high levels of gene duplications. Our findings suggest that gene
41 duplications may play a role in sex-biased genes expression in the field cricket *G. assimilis*, a
42 species likely to yield insights into the functional genomics and epigenetics of sex determination.

43

44

45

46

47 **Author Summary**

48 Sexual differences between male and female is one of the most intriguing aspects in evolutionary
49 biology. Field crickets have become models for behavioral ecology, physiological and genetic
50 studies and exhibit interesting features between sexes, such for example in body size as well as
51 reproductive structures. They also show difference in relation to genetic sex determination with
52 males being the heterogametic sex (X0) and females the homogametic ones (XX). By
53 transcriptome analysis we observed a high number of genes showing differential expression
54 between sexes and tissues which might have effects for differences in sexual phenotypes. Sex-
55 determining genes are documented for the first time in Orthoptera in this paper. Our analysis
56 enabled comparisons for some genes related to sexual fitness, like insect hormone biosynthesis,
57 circadian rhythm, circadian entrainment and circadian rhythm-fly pathway. The reproductive
58 genes in *Gryllus* are under strong positive selection and they suffered high levels of duplications
59 that could account for sex-biased genes expression.

60

61

62

63

64

65

66

67

68

69

70 **Introduction**

71 Males and females often show phenotypically visible differences while sharing almost
72 the same genome. Typically, these sexual differences result from the presence of sex-biased
73 genes (SBG), i.e., differentially expressed genes (DEG) between males and females [1]. Four
74 possible scenarios explaining the observed differences in SBG content between the male and
75 female have been anticipated: sexual conflict, gene duplication, dosage compensation (DC), and
76 sex-limited chromosomes expression [2]. Under sexual conflict –in which either sexes have
77 conflicting optimal fitness strategies concerning gene expression– an allele influencing the
78 expression of a sexually antagonistic gene may differ depending on either dominance and
79 position, i.e., X-linked or autosomal [1,2]. Conflict between the two sexes commonly arises in
80 one of two possible situations: intra-locus sexual conflict or inter-locus sexual conflict. The first
81 case occurs when phenotypic traits that are determined by a common set of genes have different
82 optima in male and female, e.g., body size, diet, development time, longevity, and locomotory
83 activity have been suggested to underlie intra-locus sexual conflict [3-5]. These traits may be
84 subjected to antagonistic patterns of selection. The second case emerges when members of one
85 sex acquire characteristics that benefit their ability to successfully reproduce, while negatively
86 impacting the fitness of the opposite sex. This process occurs at different loci in each sexes [5,6].
87 X chromosome linkage favors the accumulation of recessive male-beneficial alleles and
88 dominant female-beneficial alleles [1]. As expected, genes with sexually antagonistic expression
89 enriched on the X chromosomes were documented in *Drosophila melanogaster* (fruitfly) [7].

90 Gene duplication –in which the expression of the ancestral copy remains unchanged
91 while the new copy evolves a crucial role on sex-biased or sex-specific expression by regulating
92 a cascade of downstream genes– provides other possible explanation for SBG expression.

93 Indirect evidence for gene duplication affecting SBG expression were reported (see [8-10]).
94 Another factor that could influence SBG expression is DC, a mechanism to equalize gene
95 expression levels from genes located on sex chromosomes between males and females. DC has
96 been extensively studied in fruitfly, and it has been demonstrated that occurs through up-
97 regulation levels of genes on the male X chromosome by recruiting of RNA-protein specific
98 complex, called male-specific lethal (MLS) [11-13]. In this mechanism, the expression of SBG
99 could result from over-expression of genes nearby DC complex binding sites [2,13]. Finally,
100 genes on sex-limited chromosomes (e.g., genes allocated on the Y or W chromosomes) having
101 beneficial effect to the heterogametic sex but detrimental one to the homogametic sex could
102 provide another explanation for SBG expression [2]. Evidence of this can be found for example
103 in mammals, where the presence of the *SRY* gene limited to the Y chromosome triggers a
104 cascade of sex-biased expression [14].

105 It has been demonstrated in *Drosophila* that the initial gender determination and sex-
106 differences result from signals of three key genes: *sex-lethal* (*Sxl*), *transformer* (*tra*) and
107 *doublesex* (*dsx*) [15-17]. Early transcription factors encoded on the X chromosome controls if *Sxl*
108 is transcribed. *Sxl* is active in females but not in males resulting in the production of Sxl protein
109 in females over males. Other steps include differential splicing of sex-determining genes
110 resulting non-functional proteins in males (*Sxl* and *tra*), or even in proteins differing in their C-
111 terminal domains in the two sexes (*dsx*) causing sexual differentiation [15-17]. The *tra* gene also
112 controls the *fruitless* gene (*fru*), an encoding putative transcription factor. The *fru* gene is
113 responsible for the development of anatomical structures needed for reproduction (including
114 motor neurons) as well as sexual behavior, and it has been detected in several insects [15-17].
115 Besides Diptera, some of these genes or their orthologs were documented in other insect orders,

116 like Hymenoptera, Coleoptera, Lepidoptera, and Hemiptera [15]. Genes encoding proteins in
117 insect hormone biosynthesis pathway were also found in evolutionary distant insects [18-20].
118 These genes influence metamorphosis from juvenile to adult and can harmonize SBG expression,
119 orchestrating drastic biological and reproductive changes. Other important genes encoding
120 proteins that influence reproductive fitness are those involving in circadian rhythm. In fruit fly,
121 for instance, loss of circadian clock function in males declines significantly the sperm quantity
122 released from the testis to seminal vesicles decreasing the reproductive fitness in males [21].
123 Testis and seminal vesicles also showed rhythmic and autonomous expression of clock genes
124 [21].

125 Orthoptera species display pronounced sexual dimorphism in body pigmentation, body
126 size, as well as shape and size of reproductive structures. Differences are also evident at the
127 chromosomal level, with males being the heterogametic sex (X0). This feature is observed in
128 most species and considered ancient [22,23]. Despite sex chromosome differences and sexual
129 dimorphism in *Gryllus*, there is no information concerning genes involved in sex determination
130 cascade, sexual differentiation, and reproductive fitness differences between males and females
131 and SBG expression. To address these issues, we gathered RNA-seq data from the heads and
132 gonadal tissues to generate a *de novo* transcriptome assembly for *Gryllus assimilis*, a species
133 with an XX/X0 sex determination system. Furthermore, the transcriptomes of two other XX/X0
134 cricket species (*G. bimaculatus* and *G. firmus*) were also analyzed. Our work enabled
135 comparison of sex-determining genes, and other genes related to sexual fitness, like insect
136 hormone biosynthesis, circadian rhythm, circadian entrainment and circadian rhythm-fly
137 pathway among species of *Gryllus*. Our transcriptome approach was to identify candidate
138 genes/proteins orthologs that might account for phenotypic differences between males and

139 females. The data is also compared against the desert locust model *Locusta migratoria* genome
140 assembly. Consistent with the idea of specialized gene content between male and female, we
141 identified a set of highly expressed genes that differ between males and females. We also
142 identify the known effectors as well as downstream targets of insect sex-determination pathways
143 in crickets. These pathways have been previously identified in other insects but never in
144 Orthoptera. Furthermore, we estimated that protein-encoding reproductive genes evolve faster
145 than non-reproductive genes as result of strong positive selection at those loci. Interestingly, we
146 document that the species studied harbors exceptionally high levels of gene duplications. Our
147 findings suggest that gene duplications may play a role in SBG expression in the field cricket *G.*
148 *assimilis*, a species likely to yield insights into the functional genomics and epigenetics of sex
149 determination.

150

151 **Results**

152 ***Trinity assembly and functional annotation***

153 Following adapter trimming, we used Trinity package to assemble 207,553,576 bases for
154 *G. assimilis*, 25,578,816 bases for *G. bimaculatus* and 41,338,187 bases for *G. firmus*. This
155 resulted in 227,254 contigs with an N50 of 1,818 bases in *G. assimilis*, 37,069 contigs with an
156 N50 of 1,123 bases in *G. bimacullatus* and 63,940 contigs with an N50 of 972 bases in *G. firmus*.
157 The assembly summary statistics for *Gryllus* transcriptomes are given in S1 Table. We then used
158 Trinotate [24] pipeline to functional annotation of the transcripts (including each isoform), which
159 resulted in 48,202 contigs (21%) hits for *G. assimilis*, 13,933 contigs (37.6%) hits for *G.*
160 *bimaculatus* and 27,818 (43.5%) hits for *G. firmus* (S2 Table). One remarkable finding was the
161 failure of a significant number of *Gryllus* contigs to have hits to the non-redundant (NR)

162 database (79% of the 227,254 contigs from *G. assimilis*, 62.4% of the 37,069 contigs for *G.*
163 *bimaculatus* and 56.5% of the 63,940 contigs for *G. firmus*). The observation is in agreement
164 with the recent *G. pennsylvanicus* transcriptome assembly which yielded only ~44% of
165 annotated transcripts [25]. Unmatched contigs may results from mistakes in assembly (contigs
166 from reads with sequences errors), large number of non-coding RNAs, lack of related protein
167 sequences in the database, or even cricket-specific diverged sequences due to the whole genome
168 duplication, for example.

169 We assessed the quality of the *Gryllus* transcriptomes to quantify the putative duplication
170 rate of highly conserved orthologs that are present in only a single copy in related Arthropod
171 species. This is important because genetic diversity, gene duplications and/or the presence or
172 multiples isoforms can lead to misassemblies, where one gene is represented by multiples
173 contigs. BUSCO [26] analysis can be used to assess the frequency of such misassemblies.
174 Applying BUSCO to the *Gryllus* transcriptomes recovered 2,344 (87.6%), 1,433 (53.6%) and
175 1,304 (48.8%) complete gene copies in *G. assimilis*, *G. bimaculatus* and *G. firmus*, respectively.
176 These were identified out of 2,675 genes annotated in Arthropod's BUSCO. Furthermore, we
177 observed 1,415 (52.9%) complete and duplicate copies in *G. assimilis* and 409 (15.3%) ones in
178 both *G. bimaculatus* and *G. firmus* transcriptomes (Fig 1). It means putative episodes of gene
179 duplications recurring over the evolution in *Gryllus*. Applying BUSCO to the *L. migratoria*
180 genome allowed the recovery of 443 (41.6%) complete gene copies and 334 (31.3%) partially
181 recovered or fragmented gene copies. BUSCO groups for which there are no matches that pass
182 the tests of orthology are classified as missing. Missing values are higher in *G. bimaculatus*
183 (30.3%), *G. firmus* (34.7) and *L. migratoria* than *G. assimilis* (4.5). It means cricket-specific
184 diverged sequences due possibly to whole genome duplication events in some of these species

185 Even though *L. migratoria* yielded only 1.5% of duplicated copy in relation to the highly
186 conserved orthologs in Arthropod BUSCO, comparative genomic analysis in this species
187 revealed expansion of many gene families, particularly those that are involved in energy
188 consumption and detoxification [27]. These observations support the view of putative episodes
189 of gene duplications and whole genome duplication during Orthoptera evolution.

190 Using BLASTx tool (e-value 10^{-5}) we recovered putative gene-encoding orthologs
191 involved in the sex determination cascade: *Sxl*, *tra*, *dsx* and *fru* (S3 Table and S1 Appendix).
192 These genes showed high similarity against biological models as *Ceratitis capitata*, *Megaselia*
193 *scalaris*, *Drosophila melanogaster*, *D. virilis* and *Rattus* sp., suggesting conservation. Only *dsx*
194 was not recovered in *G. bimaculatus* transcriptome. Looking at the *Gryllus* transcriptomes, we
195 found out that *G. assimilis* and *G. bimaculatus* displaying two copies of *tra* and *fru* genes. In
196 contrast, both genes are present as single copy in *G. firmus* (S1 Appendix).

197

198 **KEGG pathways analysis**

199 Using the KEGG (Kyoto Encyclopedua of Genes and Genomes) database [28-30] we
200 annotate 16,842 transcripts in *G. assimilis*, 6,439 transcripts in *G. bimaculatus* and 9,033
201 transcripts in *G. firmus*, corresponding to putatively orthologs genes. For *L. migratoria* genome
202 assembly we annotated 20,290 putative orthologs genes. Considering that the species studied
203 belong to different suborders (i.e., *Gryllus* belongs to Ensifera and *L. migratoria* to Caelifera)
204 and that Ensifera is basal in Orthoptera phylogeny [31], we assumed that the differences found
205 among species in relation to putative number of orthologs genes are caused by episodes of gain
206 or loss of orthologs as well as gene duplications (see below). We then used orthology and
207 annotation information to identify genes involved in insect hormone biosynthesis (IHB),

208 circadian rhythm (CR), circadian entrainment (CE) and circadian rhythm-fly (CRF). A summary
 209 of these analyses is given in the Table 1 including number of different genes and isoforms
 210 annotated for *L. migratoria* genome and *Gryllus* transcriptomes. The S1 Fig shows examples of
 211 these analyses denoting the genes identified by our study involved in IHB, CR, CE and CRF in
 212 *G. assimilis* transcriptome. Many protein products of these transcripts are required for
 213 metamorphosis, behavior, biological and reproductive process.

214 **Table 1.** KEGG pathway and number of genes analyzed corresponding to the *L. migratoria* genome and
 215 *Gryllus* transcriptomes.

KEEG Pathway	No. of different annotated genes				No. of annotated isoforms		
	Lmi	Gas	Gbi	Gfi	Gas	Gbi	Gfi
1 Metabolism							
1.2 Metabolism of terpenoids and polyketides							
Insect hormone biosynthesis	13	14	5	7	64	12	20
2 Organismal Systems							
2.1 Environmental adaptation							
Circadian rhythm	14	12	9	13	53	12	25
Circadian entrainment	41	35	17	24	115	21	54
Circadian rhythm-fly	8	8	3	7	25	6	11

216 KEGG= Kyoto Encyclopedia of Genes and Genomes database; Gas= *G. assimilis*; Gbi= *G. bimaculatus*;
 217 Gfi= *G. firmus*; Lmi= *Locusta migratoria* genome.

218
 219 We also compared the number of annotated orthologs genes in these pathways within
 220 each species and those orthologs that are shared between *L. migratoria* genome and *Gryllus*
 221 transcriptomes (Fig 2 and S2 Appendix). To look at putative gene duplication events, we counted
 222 the number of single copy genes and the putative number of multicopy genes in the *L. migratoria*
 223 genome and *Gryllus* transcriptomes (S3 Appendix). We show that Orhoptera genomes have an
 224 exceptionally high number of duplicate genes in these pathways that can be traced to each
 225 species (Fig 3).

226

227 **Differential gene expression analysis between sexes**

228 The data regarding DEG and SBG were compared between gonads and heads from male
 229 and female of *G. assimilis*. This species was used as a reference for this analysis because it was
 230 the only species for which we have biological replicates for different sex and tissues. Because
 231 different isoforms of the transcripts could be differentially expressed within metabolic pathway,
 232 all annotated isoforms were taken to analyze differential expression. When comparing male head
 233 vs female head, we found 29,504 genes with nonzero total read count using adjusted p -value <
 234 0.05. Of these, 2,037 (6.9%) were male-biased genes (MBG) and 1,688 (5.7%) were female-
 235 biased genes (FBG) (Table 2 and Fig 4A). Remarkably, the proportion of DEG in gonads is
 236 comparative higher than somatic tissues due probably to the specificity of tissues studied. In
 237 gonads, we found 15,689 (39%) that were MBG and 12,513 (31.1%) that were FBG, using an
 238 adjusted p -value < 0.05 (Table 2 and Fig 4B). This was out of 39,865 genes with nonzero total
 239 read counts. Heatmaps showing the expression data of the 50 most highly expressed genes for
 240 every comparison is given in the Fig 5, and the annotation for those genes is shown in S4
 241 Appendix.

242 **Table 2.** *G. assimilis* differentially expressed genes between males and females in head and gonad.

Sample	No. of expressed genes	Sex-biased genes (%)	Male-biased genes (%)	Female-biased genes (%)
Head	29,504	3,725 (12.6)	2,037 (6.9)	1,688 (5.7)
Gonad	39,865	28,202 (70)	15,689 (39)	12,513 (31)

243 We report the total number of genes expressed, and those genes that were sex-biased with an adjusted p -
 244 value < 0.05.

245
 246 In our *G. assimilis* dataset, we examined whether the putative genes (and their isoforms)
 247 involved in sex determination cascade showed differential expression. We found that sex-
 248 determining genes were differentially expressed only in gonads (S3 Table). For example, *Slx* and
 249 *fru* are MBG while *tra* is FBG. *dsx* was not differentially expressed between sexes. To
 250 comparatively analyze the biases we visually examined SBG expression in these candidates (Fig
 251 6A).

252 We also checked within *G. assimilis* if the IHB, CR, CE and CRF orthologs genes
 253 showed differential expression between the sexes and compared the intensity and direction of the
 254 expression bias. The statistical analysis is showing in the Table 3. Mostly genes and isoforms
 255 were male-biased (Table 3, S5 Appendix). To comparatively analyze the biases we visually
 256 examined SBG expression in these candidates (Fig 6 B-C for IHB, and S2 Fig for CR, S3 Fig for
 257 CE and S4 Fig for CRF).

258 **Table 3.** *G. assimilis* differentially expressed genes between males and females in head and gonad for the
 259 KEGG pathway studied in this paper.

KEGG pathway	Sample	No. of expressed isoforms	Sex-biased genes (%)	Male-biased genes (%)	Female-biased genes (%)
Insect hormone biosynthesis	Head	48	10 (20.8)	8 (16.7)	2 (4.1)
	Gonad	59	34 (57.6)	21 (35.6)	13 (22)
Circadian rhythm	Head	47	5 (10.6)	0	5 (10.6)
	Gonad	52	48 (92.3)	24 (46.15)	24 (46.15)
Circadian entrainment	Head	97	20 (20.6)	15 (15.5)	5 (5.1)
	Gonad	100	65 (65)	15 (15)	50 (50)
Circadian rhythm-fly	Head	19	3 (15.8)	0 (0)	3 (15.8)
	Gonad	24	14 (58.3)	10 (41.7)	4 (16.6)

260 We report the total number of genes/isoforms expressed, and those genes that were sex-biased with an
 261 adjusted *p*-value < 0.05.

263 *Evolutionary rates analysis*

264 Using sequence data from putative orthologs in three species of field crickets, we
 265 estimated the distribution of number of nonsynonymous (*K_a*) and synonymous (*K_s*) substitutions
 266 for 27 protein-encoding genes considered to represent proteins involved in sex-determination and
 267 reproductive fitness, and contrasted them with a group of five housekeeping genes, randomly
 268 selected to represent different structural proteins and basic metabolic functions. In the first group
 269 of genes, the *K_a* distributions differs significantly from that observed for *K_s* (*p*-value < 2.2e-16,
 270 Fisher's exact test). The *K_a* distribution ranges from 0 to 2.1 (mean = 0.42) while *K_s*
 271 distributions ranges from 0.01 to 1.74 (mean = 0.54) (S4 Table). The estimates of selection
 272 parameter ω in these set of genes range from 0 to 2.25 (mean = 0.8, S4 Table), and although we

273 observed only seven ortholog genes with a clear signature of adaptive evolution ($\omega > 1$), there
274 are another 12 genes with inferred $\omega > 0.5$ (Fig 7A-B and S4 Table). Interestingly, the seven
275 protein-encoding genes with $\omega > 1$ display SBG expression, e.g., FOHSDR (MBG in head and
276 SBG in gonads), CUL1 (FBG in head and SBG in gonads), FBXW1_11 (FBG in head and SBG
277 in gonads), PLCB (MBG in testis), RBX1 (MBG in testis), PRKG1 (FBG in head) and TIM
278 (MBG in testes).

279 The differences of evolutionary rates for protein-encoding genes considered to represent
280 proteins involved in sex-determination and reproductive fitness are more obvious when
281 compared to the five-basic metabolic and structural housekeeping genes. These latter subsets of
282 genes display $\omega < 1$ indicating negative selection driven by a decreased fixation rate of Ka (mean
283 = 0.352, S4 Table).

284

285 **Discussion**

286 With a genome size larger than 2 Gb distributed over 14 autosomal pairs and two X
287 chromosomes in females [32,33], genomic resources for field cricket's species are still lacking.
288 We present here a compressive transcriptome for crickets using species of *Gryllus* as models.
289 The effort should be help to accelerate functional genomics in Orthoptera and knowledge
290 regarding sex-determining genes, IHB, CR, CE, and CRF pathway, in the group. This represents
291 a valuable contribution to the available desert locust model *L. migratoria* genomic resources for
292 at least two reasons. First, field cricket's species have become a model organisms in a variety of
293 fields, from studies in behavioral ecology, to physiological and genetics [34-50]. However, to
294 assess which features of these genomes are lineage-specific and which are shared with other
295 species of orthopterans, it is important to extend genomic studies to other crickets. Furthermore,

296 additional genomes and transcriptomes make comparative genomic studies possible. Second, all
297 three field cricket's species along with the desert locust genome/transcriptome sequences could
298 be plagues of agriculture that cause devastating natural disasters in Asia, Europe, North and
299 South America. Our study thus extends the ecological breadth that is covered and allows for
300 more comparative studies within Orthoptera, while providing new resources for ecological
301 studies around the world.

302 De novo transcriptome assembly without an available genome is largely challenging,
303 including the different transcript isoforms and closely related paralogs. One remarkable findings
304 of the project was the failure of a significant number of contigs to have hits to the NR database
305 i.e., 79% of the 227,254 contigs for *G. assimilis*, 62.4% of the 37,069 contigs for *G. bimaculatus*
306 and 56.5% of the 63,940 contigs for *G. firmus*. Likewise, the recent *G. pennsylvanicus*
307 transcriptome assembly yield ~44% of annotated transcripts [25]. It means that unmatched
308 contigs is challenging for assessing of field cricket's transcriptomes. Unmatched contigs may be
309 the outcome of mistakes during assembly (e.g., contigs from reads with sequences errors), large
310 number of non-coding RNAs, lack of related protein sequences in the database, or even cricket-
311 specific diverged sequences due to the whole genome duplication. The aforesaid is indeed more
312 noticeable for *G. assimilis* than the other congeneric species. It could result from the specificity
313 of tissue studied, i.e., testis and ovary. All these issues should be solved once a full cricket's
314 genome is available.

315

316 **Sex-biased genes and sex-determining genes**

317 Assessing the *G. assimilis* transcriptome, we found out similar proportion SBG in head,
318 i.e., 6.9% MBG and 5.7% FBG. When reproductive-tissues are compared (i.e., testis and ovary),

319 a relatively high proportion of MBG (39%) relative to FBG (31%) in gonads is observed. The
320 proportion of SBG is much higher than that reported in *Drosophila* species where as many as
321 two-thirds of all genes can show sex-biased expression when whole animals are evaluated
322 [51,52]. However, as also noticed in *Drosophila* [53], the proportion of SBG in *G. assimilis*
323 varies greatly when comparing different body parts. These observed differences in SBG can be
324 explained by at least two alternative hypotheses. First, the high proportion of SBG in *G. assimilis*
325 could result from the differences phenotypically visible –strong sexual dimorphism– in body
326 size as well as the specificity of reproductive organs. Second, if *Gryllus* genome experiences
327 selection in male background and only rarely in female background, it would be expected the
328 loss of strongly FBG, especially if their expression is harmful to males. Thus, the high number of
329 SBG, the high degree of sex bias, and the excess of MBG relative to FBG could be a
330 consequence of sexual conflict to achieve optimum levels of gene expression in males,
331 potentially leading many X-linked genes to acquire male-specific functions and to relaxation of
332 negative selection on FBG. However, considering the high variation in SBG expression observed
333 in *G. assimilis*, further experimental evidences are needed to test these hypotheses.

334 We find out key protein-encoding genes considered to be involved in sex-determination
335 pathway: *Sxl*, *tra*, *dsx* and *fru*. To our knowledge, these protein-encoding genes including their
336 isoforms resulting from alternative splicing are reported for first time in Orthoptera. Out these,
337 only *dsx* was not recover in *G. bimaculatus* transcriptome. Searching the *Gryllus* transcriptomes,
338 we found out that *G. assimilis* and *G. bimaculatus* displaying two copies of *tra* and *fru* genes,
339 while in *G. firmus* both genes are present as single copy. However, looking at *G. assimilis*
340 transcriptomes we noted only one copy of *tra* and *fru* along with their corresponding isoforms
341 displaying differential expression, suggesting recent duplication/loss episodes of these genes in

342 *Gryllus*. Looking at the *G. assimilis* transcriptome we identified that *Sxl* and *fru* are MBG in
343 testis while *tra* is FBG in ovary. In many insects, *tra* triggers the expression of female-specific
344 protein isoforms from *dsx* and *fru*, which in turn direct female differentiation [15-17]. When *tra*
345 is absent, a default mechanism give arises to male-specific proteins from *dsx* and *fru* and male
346 development ensues. Our finding seems to agree with this view. While *dsx* seems to be necessary
347 for male trait development in several insects [15-17] it is not SBG in *G. assimilis*, and its
348 function is uncertain in Orthoptera. In insects, *fru* gene is responsible for the development of
349 anatomical structures needed for reproduction including motor neurons, and to regulate sex-
350 specific behavior [15-17]. Considering the bias in *G. assimilis* testis, *fru* could play a crucial role
351 in influencing investment in male sexual reproduction, although this latter requires further
352 experimental evidences. Despite of the cricket's transcriptomes presents a large part of
353 unmatched contigs, the fact that we were able to annotate all the crucial genes involved in sex-
354 determining pathway speaks for the completeness of the *de novo* assembly. Even though gene
355 duplications were detected for *tra* and *fru*, the genes involved in sex-determining are highly
356 conserved in the three-cricket's species. It means that the sex-determination could be conserved
357 across the genus. What remains to investigate is if other gene expression results fit well with our
358 analysis.

359 We find out key genes from IHB, CR, CE and CRF in the cricket's transcriptomes.
360 Benchmarking these finding with the *L. migratoria* genome, we noted that these protein-
361 encoding genes are shared in the two suborders of Orhoptera, Ensifera and Caelifera.
362 Furthermore, our outcomes show that Orhoptera genomes indeed harbor exceptionally high
363 levels of gene duplications for the analyzed genes that can be detected in each species.
364 Considering that the species studied belong to different suborders (i.e., *Gryllus* belongs to

365 Ensifera and *L. migratoria* to Caelifera) and that Ensifera is basal in Orthoptera phylogeny [31],
366 we assumed that the differences observed for copy numbers among species are caused by
367 episodes of gain or loss of orthologs as well as gene duplications. It also means that *L.*
368 *migratoria* underwent recent events of duplications compared to crickets. Likewise, gene
369 duplications are putatively more recent in *G. assimilis* than *G. bimaculatus* and *G. firmus*. The
370 desert locust *L. migratoria* has 6.5 Gb of genome size [27]. It is three times more than the
371 estimated genome size of 2 Gb in crickets [32,33]. We hypothesized that the increasing in copy
372 number in *L. migratoria* compared to crickets might account from putative whole genome
373 duplication. Moreover, comparative genomic analysis in *L. migratoria* revealed expansion of
374 many gene families, particularly those that are involved in energy consumption and
375 detoxification [31]. These observations support the view of putative episodes of gene
376 duplications and whole genome duplication during Orthoptera evolution.

377 Furthermore, these gene duplications and their isoforms are indeed largely responsible for
378 the SBG expression in *G. assimilis*. Many of them showed to be MBG indicating specialized
379 gene expression in males. Genes encoding IHB pathway influencing metamorphosis from
380 juvenile to adult and that can harmonize SBG expression orchestrating drastic biological and
381 reproductive changes, were also found in evolutionary distant insects [18-20]. Some of these
382 protein-encoding genes and their isoforms displaying strong male-biased expression are those
383 resulting for juvenile hormone (JH), juvenile hormone esterase (JHE), aldehyde dehydrogenase
384 (ALDH), NADP⁺-dependent farnesol dehydrogenase (FOHSDR), for example. Other important
385 protein-encoding genes influencing reproductive fitness and adaptation are those involving in
386 circadian rhythm pathway, like for example *clock* (CLOCK), *period* (PER) and *timeless* (TIM).
387 In *G. assimilis* CLOCK is FBG, while PER and TIM are MBG. Recently Des Marteaux et al.

388 [25] reported PER and CLOCK as up-regulated and TIM as down-regulated in different body
389 parts in *G. pennsylvanicus*. However, no information regarding the sex-biased was pointed. In
390 fruitfly, for instances, loss of circadian clock function in males declines significantly the sperm
391 quantity released from the testis to seminal vesicles decreasing the reproductive fitness in males
392 [21]. Testis and seminal vesicles also showed rhythmic and autonomous expression of clock
393 genes [21]. In our findings it is possible that CLOCK plays a larger role in reproductive fitness in
394 females of *G. assimilis*, an assumption that is consistent with its female-biased expression.

395 Comparing the expression data, we noticed some highly differentially expressed genes
396 between gonads and between heads from *G. assimilis*. For example, two major constituents of
397 microtubules, the genes encoding for tubulin beta chain (TBB) and tubulin alpha-1 chain
398 (TUBA1A) (see [54]), display LFC 10.9 and 12.6 in male's gonads, respectively. The Pro-
399 resilin, plays a crucial role in insect flight by providing low stiffness, high strain and efficient
400 energy storage (see [55]), displays a LFC 7.3 in female's heads; the elongation of very long
401 chain fatty acids protein AAEL008004 (AAEL008004) implicated in synthesis of very long
402 chain fatty acids [56], shows a LFC 4.2 in female's heads. Gene encoding for outer dense fibers
403 (ODF3) –filamentous structures located on the outside of the axoneme in the midpiece and
404 principal piece of the mammalian sperm tail– which is thought to help keep the elastic structures
405 passive and elastic recoil of the tail of the sperm [57,58], displays a LFC 10.2 in testis. Except
406 for the encoding gene ODF3 the highly-expressed genes above-mentioned are not evidently
407 related to the sexual phenotypes or reproduction. These DEGs might be a consequence of sexual
408 conflict to achieve optimum levels of gene expression among sexes. Another plausible
409 explanation is the nonrandom distribution of SBG over the genome (see below).

410 It is well-known that SBG are not randomly distributed over the genome, with many of
411 them showing preferential localization on the X-chromosome [1,2]. Also, the content and
412 expression levels between the X and autosomes often differ [2]. In fruitfly, e.g., was found a
413 significant paucity of expression levels with MBG but slightly expressed levels of FBG on the X
414 in relation to the autosomes [13,59,60]. These phenomena are called “desmasculinization” and
415 “feminization” of the X chromosome, respectively. However, recent findings showed that the
416 desmaculinization of the X chromosome is not a general phenomenon. For instances, males and
417 females from some *Drosophila* species did not display MBG when comparing transcripts from
418 different body parts [52]. Conversely, a slight excess of MBG was observed in fruitfly and *D.*
419 *mojavensis* [52]. Other cases reporting significant enrichment of X-linked MBG in head and
420 brain of fruitfly was also documented (see [53,61,62]). Under this background, it is plausible that
421 many of highly SBG could be located on the X-chromosome in crickets, although further
422 experimental evidences are needed to test the possibility of X-linked SBG and the masculinizing
423 or feminizing effect of the X-chromosome in Orthoptera. Our findings prospects raise that the
424 sex chromosomes might have relevance in the biology and SBG evolution, which will set the
425 stage to explore and advance our understanding about sex-related genes and dosage
426 compensation, issues largely unknown in Orthoptera.

427

428 **Evolution of sex-determining and reproductive fitness genes**

429 Using sequence data from putative orthologs in three species of field crickets, we
430 estimated the distribution of Ka and Ks substitutions for 27 protein-encoding genes considered to
431 represent proteins involved in sex-determination pathway and reproductive fitness (including
432 IHB, CR, CE, and CRF). We contrasted them with a group of five housekeeping genes, selected

433 to represent different structural proteins and basic metabolic functions (S4 Table). We estimated
434 that the rate of divergence of 27 protein-encoding genes involved in sex determination and
435 reproductive fitness was significantly higher than that of the five housekeeping genes compared.
436 This supports the idea that protein-encoding genes involved in sex determination and
437 reproductive fitness tends to evolve more rapidly than genes not involved in reproduction
438 [63,64]. Similar results were found for genes encoding of seminal proteins from proteinaceous
439 components of ejaculates in the field crickets *G. firmus* and *G. pennsylvanicus*, supporting this
440 view [38].

441 Considering that the K_s values were similar in protein-encoding genes involved in
442 reproduction and in housekeeping genes, significant differences in mutation rate or codon bias
443 among reproductive genes and housekeeping genes cannot be pointed as explanation. However,
444 the rate of fixation of K_a is 1.5 times higher in genes encoding reproductive proteins than in
445 housekeeping genes. According to Andrés et al. [38], in field crickets, positive selection is more
446 important in the evolution of genes encoding reproductive proteins than in housekeeping genes.
447 Conversely, negative selection resulting from functional constraints is more important in
448 housekeeping genes than in genes encoding reproductive proteins. We found out many genes
449 encoding reproductive proteins with $\omega < 1$, suggesting that many of these proteins are subject to
450 negative selection possibly as result of functional restrictions. However, we also observed a wide
451 range of ω values within the class of genes encoding reproductive proteins. For example, we
452 estimate 12 genes with $\omega > 0.5$ which has been suggested as a threshold to identify candidate
453 genes that may have experienced events of adaptive evolution [63,65,66]. Furthermore, we find
454 out seven genes with $\omega > 1$ suggesting strong positive selection at these loci. On overage, our
455 results indicate that positive selection seems to be more important in the evolution of genes

456 encoding reproductive proteins than in housekeeping genes. Similar results for genes encoding
457 putative seminal proteins were reported in *G. firmus*, *G. pennsylvanicus*, *Drosophila* and
458 primates [38,67,68]. In addition, this selection pressure could be the result of the underlying
459 sexual conflict among different reproductive interest of males and females, thus explaining the
460 observed excess of MBG relative to FBG in gonads.

461

462 **Materials and Methods**

463 *Tissue sampling and RNA extraction*

464 Males and females of *Gryllus assimilis* were obtained from a pool of individuals that had
465 been bred at the Univ. Estadual Paulista - UNESP (Rio Claro, SP, Brazil). RNA from males
466 (head and testes) and females (head and ovary) tissues from were extracted with a phenol-based
467 phase separation using TRIzol Reagent (ThermoFisher Scientific, Invitrogen) following the
468 standard protocol recommended by the supplier. For every tissue it was pooled an equivalent
469 amount of material from three specimens per sex with the same maturity age (i.e., sexually
470 mature males in calling stages and virgin adult females) to produce representative male and
471 female pools. These male and female pools for each tissue were performed in biological
472 triplicates. RNA-seq libraries were prepared on the TruSeq rapid SBS kit or Truseq SBS Kit v4
473 (Illumina Inc., San Diego, CA, USA). Libraries were sequenced (2x126 bp) on a HiSeq 2500
474 platform by Macrogen service facility (Macrogen Inc., South Korea). The *G. assimilis* male and
475 females RNA-seq reads were deposited in the NCBI database as follows: (XXX-XXX).

476 For comparative analysis of transcriptomes, we downloaded from NCBI database RNA-
477 seq data from other two *Gryllus* species, as follows: mixed ovaries and embryos RNA-seq reads
478 (accession SRX023832), mixed-stage embryos RNA-seq reads (accession SRX0238310) and

479 ovaries RNA-seq reads (accession SRX023831) from *G. bimaculatus*; flight muscle from long
480 winged female with histolyzed flight muscle (LWFHFM) samples RNA-seq reads (accession
481 SRX272161, SRX272160, SRX272159), flight muscle from long winged female with functional
482 flight muscle (LWFFM) samples RNA-seq reads (accession SRX272158, SRX272157,
483 SRX272156, SRX272155), fat body from short winged female incapable of flight (FBSWFIF)
484 samples RNA-seq reads (accession SRX272155, SRX272154, SRX272153, SRX272152,
485 SRX272151, SRX272150, SRX272127, SRX272125), and fat body from long winged female
486 with functional flight muscles (FBLWFFM) samples RNA-seq reads (accession SRX272124,
487 SRX272122, SRX272120, SRX272119, SRX272117, SRX272111, SRX272106, SRX272104)
488 from *G. firmus*.

489

490 ***Transcriptome assembly and annotation***

491 The RNA-seq reads per sample were trimmed with Trimmomatic [69] and then used for
492 *de novo* transcriptome assembly using Trinity [70], set to the default parameters. The transcripts
493 (including each isoform) were annotated with Trinotate [24], an extension designed for
494 automatic functional annotation of transcriptomes. It uses several different well-referenced tools
495 for functional annotation, including homology search against sequence databases
496 (BLAST+/SwissProt), protein domain identification (HMMER/PFAM), and comparison to
497 currently curated annotation databases (eggNOG [71] and Gene Ontology terms). Furthermore,
498 Transdecoder [72] was used to identify open reading frame (ORF) with complete coding
499 sequences. Moreover, we annotated orthologs of genes that are presumably present in single
500 copy in arthropod genomes. For this, we used the Benchmarking set of Universal Single-Copy
501 Orthologs (BUSCO) [26] with the Arthropod BUSCO set from OrthoDB version 7 [73], which

502 contains all genes found in single copy in more than 90% of the included 38 species. The
503 Arthropod BUSCO set contains 2,675 genes that were used to search the *Gryllus de novo*
504 transcriptomes, as well as the *Locusta migratoria* genome downloaded from the NCBI (GenBank
505 assembly accession GCA_000516895.1).

506

507 **KEGG pathways analysis**

508 To map KEGG [28,29,30] pathways of assigned crickets orthologs we employed the
509 Automatic Annotation Server (KAAS, [74]). KEGG orthology (KO) assignments were
510 performed based on the bi-directional best hit (BHH) in BLAST [75]. It was used to search for
511 genes involved in insect hormone biosynthesis (IHB), circadian rhythm (CR), circadian
512 entrainment (CE) and circadian rhythm-fly (CRF) pathways. In addition, the data were compared
513 against the genome of the desert locust model *L. migratoria*. Using KO assignments, we counted
514 and compared the number of orthologs genes founding for these pathways between *L. migratoria*
515 genome and *Gryllus* transcriptomes. To look at gene duplication in the abovementioned
516 pathways, we counted the number of single copy genes and the putative number of multicopy
517 genes in the *L. migratoria* genome and *Gryllus* transcriptomes.

518

519 ***Differential gene expression***

520 To look at gene expression, raw reads from *G. assimilis* were first trimmed using
521 Trimmomatic [69], followed by expression quantification for each transcript (including each
522 isoform) using RSEM v1.2.16 [76] with Bowtie2 [77] as the read mapper. The data regarding
523 differentially expressed genes (DEG) and sex-biased genes (SBG) were compared between
524 gonads from male and female and between heads from male and female of *G. assimilis*. This

525 species was taken as reference because it was the only one we have biological replicates (RNA-
526 seq reads) for different sexes and tissues generated in our lab. The statistical analysis of DEG
527 was performed using DESeq2 (version 2.1.14.1; [78]) as implemented in the Bioconductor
528 package (version 3.0; [79]) in R (version 3.3.3; [80]). All p -values were adjusted using Wald test
529 as implemented in DESeq2. To reduce multiples testing, DESeq2 only considers genes that have
530 a set minimum number of reads. This is because adjusted p -values cannot be calculated for genes
531 with high variation among replicates or low median expression across libraries [78]. In our study
532 design, the minimum was three mapped reads across all libraries. A gene was considered biased
533 if the comparison for the factor condition (samples) yield and adjusted p -values < 0.05 (FDR
534 5%). The degree bias was determined by the \log_2 fold-change (\log_2FC) difference between
535 conditions as calculated in DESeq2, i.e., those genes with $LFC > 0$ and $LFC < 0$ with an adjusted
536 p -value < 0.05 were considered as up and down-regulated, respectively.

537

538 ***Evolutionary rates***

539 We assess the selective constrains for protein-encoding genes involved in sex-
540 determining, IHB, CR, CE and CRF among field crickets. To achieve this aim, only orthologs
541 shared among the three species were taken (27 protein-encoding genes). CDS regions of these
542 protein-encoding genes were aligned with ClustalW algorithm [81]. For estimating the number
543 of nonsynonymous (Ka) and synonymous (Ks) substitutions we used the Nei-Gojobori method
544 (Jukes-Cantor) implemented in MEGA7 [82]. Selection estimates (ω) was calculated using the
545 Ka/Ks ratio and considering an p -value < 0.05 (Fisher's exact test). If $\omega = 1$, a neutral model of
546 evolution cannot be rejected, whereas $\omega < 1$ indicates negative selection, and $\omega > 1$ indicates
547 positive selection.

548 To compare the selective constraints between genes encoding nonproductive proteins and
549 those genes encoding influencing reproductive fitness and sexes, we contrasted these latter
550 subsets of genes with a group of five housekeeping genes, selected to represent different
551 structural proteins and basic metabolic functions, i.e., transcription factor E2F4 (E2F4), X-ray
552 repair cross-complementing protein 5 (XRCC5), interleukin enhancer-binding factor 2 homolog
553 (ILF2), RNA-binding protein EIF1AD (EIF1A) and ribonuclease H2 subunit B (RNH2B).

554

555 **Acknowledgments**

556 We thank the anonymous reviewers for the constructive criticism and helpful
557 suggestions. The computations in this paper were run on the Odyssey cluster supported by the
558 FAS Division of Science, Research Computing Group at Harvard University.

559

560 **Founding Statements**

561 OMPG was supported by Research Internships Abroad (BEPE) fellowships from
562 Fundação de Amparo à Pesquisa do Estado de São Paulo-FAPESP (process number 2016/01506-
563 3). This study was supported by FAPESP (process number 2014/11763-8, 2015/16661-1) and
564 Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior-CAPES. DCCM was awarded
565 a research productivity fellowship from the Conselho Nacional de Desenvolvimento Científico e
566 Tecnológico-CNPq (process number 304758/2014-0).

567

568 **Author contribution**

569 OMPG designed the experiments and analyzed the data. OMPG, LB and DCCM wrote
570 the paper. All authors read and approval the final manuscript.

571

572 **Competing interests**

573 Not declared.

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594 **References**

- 595 1. Ellegren H, Parsch J. The evolution of sex-biased genes and sex-biased gene expression. Nat
596 Rev Genet. 2007;8: 689-698.
- 597 2. Parsch J, Ellegren H. The evolutionary causes and consequences of sex-biased gene
598 expression. Nat Rev Genet. 2013;14: 83-87.
- 599 3. van Doorn GS. Intralocus sexual conflict. Ann NY Acad Sci. 2009;1168: 52-71.
- 600 4. Mills SC, Koskela E, Mappes T. Intralocus sexual conflict for fitness: sexual antagonistic
601 alleles for testosterone. Proc R Soc B. 2012;279: 1889-1895
- 602 5. Pannell TM, Morrow EH. Two sexes, one genome: the evolutionary dynamics of intralocus
603 sexual conflict. Ecol Evol. 2013;3: 1819-1834.
- 604 6. Chapman T, Arnqvist G, Bangham J, Rowe L. Sexual conflict. Trends in Ecol Evol. 2003;18:
605 41-47.
- 606 8. Gallach M, Chandrasekaran C, Betrán E. Analyses of nuclearly encoded mitochondrial genes
607 suggest gene duplication as a mechanism for resolving intralocus sexually antagonistic
608 conflict in *Drosophila*. Genome Biol Evol. 2010;2: 835-850.
- 609 9. Gallach M, Betrán E. Gene duplication might resolve intralocus sexual conflict. Trends in
610 Ecol Evol. 2011;26: 558-559.
- 611 10. Connallon T, Clark AG. The resolution of sexual antagonism by gene duplication. Genetics.
612 2011;187: 919-937.
- 613 11. Gallach M, Arnau V, Aldecoa R, Marín I. A sequence motif enriched in regions bound by the
614 *Drosophila* dosage compensation complex. BMC Genomics. 2010;11: 169.
- 615 12. Kaiser VB, Bachtrog D. Evolutions of sex chromosome in insects. Annu Rev Genet.
616 2010;44: 91-112.

- 617 13. Gallach M, Betrán E. Dosage compensation and the distribution of sex-biased gene
618 expression in *Drosophila*: considerations and genomic constraints. J Mol Evol. 2016;82: 199-
619 206.
- 620 14. Kashimada K, Koopman P. *Sry*: the master switch in mammalian sex determination.
621 Development. 2010;137: 3921-3930.
- 622 15. Sánchez L. Sex-determining mechanisms in insects. Int J Dev Biol. 2008;52: 837-856.
- 623 16. Verhulst EC, van de Zande L, Beukeboom LW. Insect sex determination: it all evolves
624 around transformer. Curr Opin in Genet Develop. 2010;20: 376-383.
- 625 17. Mullon C, Pomiankowski A, Reuter M. Molecular evolution of *Drosophila Sex-lethal* and
626 related sex determining genes. BMC Evol Biol. 2012;12: 5.
- 627 18. Noriega FG, Ribeiro JMC, Koener JF, Valenzuela JG, Hernandez-Martinez S, Pham VM,
628 Feyereisen R. Comparative genomics of insect juvenile hormone biosynthesis. Insect
629 Biochem Mol Biol. 2006;36: 366-374.
- 630 19. Noriega FG. Juvenile hormone biosynthesis in insects: what is new, what do we know, and
631 what questions remain? Int Schol Res Not. 2014;124: 1-16.
- 632 20. Niwa YS, Niwa R. Transcriptional regulation of insect steroid hormone biosynthesis and its
633 role in controlling timing of molting and metamorphosis. Develop Growth Differ. 2016;58:
634 94-105.
- 635 21. Beaver LM, Gvakharia BO, Vollintine TS, Hege DM, Stanewsky R, Giebultowicz JM. Loss
636 of circadian clock function decreases reproductive fitness in males of *Drosophila*
637 *melanogaster*. Proc Natl Acad Sci USA. 2002;99: 2134-2139.
- 638 22. White MJD. Animal Cytology and Evolution. Cambridge: Cambridge University Press;
639 1973.

- 640 23. Hewitt GM. Grasshoppers and crickets, Animal Cytogenetics, vol 3, Insecta 1, Orthoptera,
641 Gebrüder Borntraeger, Berlin; 1979.
- 642 24. Trinotate: Transcriptome Functional Annotation and Analysis. Available from:
643 <https://trinotate.github.io>
- 644 25. Des Marteaux LE, McKinnon AH, Udaka H, Toxopeus J, Sinclair BJ. Effects of cold-
645 acclimation on gene expression in fall field cricket (*Gryllus pennsylvanicus*) ionoregulatory
646 tissues. BMC Genomics. 2017;18: 357.
- 647 26. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing
648 genome assembly and annotation completeness with single-copy orthologs. Bioinformatics.
649 2015;31: 3210-3212.
- 650 27. Wang X, Fang X, Yang P, Jiang X, Jiang F, Zhao D, Li B, et al. The locust genome provides
651 insight into swarm formation and long-distance flight. Nat Commun. 2014;5: 2957.
- 652 28. Kanehisa, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on
653 genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45(D1): D353-D361.
- 654 29. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference
655 resource for gene and protein annotation. Nucleic Acids Res. 2016;44(D1): D457-D462.
- 656 30. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucl Acids Res.
657 2000;28: 27-30.
- 658 31. Song H, Amédégno C, Cigliano MM, Desutter-Grandcolas L, Heads SW, Huang Y, et al.
659 300 million years of diversification: elucidating the patterns of orthopteran evolution based on
660 comprehensive taxon and gene sampling. Cladistics. 2015;31: 621-651.
- 661 32. Palacios-Gimenez OM, Carvalho CR, Ferrari Soares FA, Cabral-de-Mello DC. Contrasting
662 the chromosomal organization of repetitive DNAs in two Gryllidae crickets with highly

- 663 divergent karyotypes. Plos ONE. 2015;10: e0143540.
- 664 33. Gregory TR. Animal genome size database. Available from: <http://www.genomesize.com>
- 665 34. Howard DJ, Marshall JL, Hampton DD, Britch SC, Draney ML, Chu J, Cantrell RG. The
666 genetics of reproductive isolation: a retrospective and prospective look with comments on
667 ground crickets. American Nat. 2002;3: S8-S21.
- 668 35. Fedorka KM, Mousseau TA. Female mating bias results in conflicting sex-specific offspring
669 fitness. Nature. 2004;429: 65-67.
- 670 36. Bussiere LF, Hunt J, Jennions MD, Brooks R. Sexual conflict and cryptic female choice in
671 the black field cricket, *Teleogryllus commodus*. Evolution. 2006;60: 792-800.
- 672 37. Zera AJ. Evolutionary genetics of juvenile hormone and ecdysteroid regulation in *Gryllus*: a
673 case study in the microevolution of endocrine regulation. Comp Biochem Physiol A Mol
674 Integr Physiol. 2006;144: 365-379.
- 675 38. Andrés JA, Maroja LS, Bogdanowicz SM, Swanson WJ, Harrison RG. 2006. Molecular
676 evolution of seminal proteins in field crickets. Mol Biol Evol. 23: 1574-1584.
- 677 39. Andrés JA, Maroja LS, Harrison RG. Searching for candidate speciation genes using a
678 proteomic approach: seminal proteins in field crickets. Proc R Soc B. 2008;275: 1975-1983.
- 679 40. Andrés JA, Larson ER, Bogdanowicz SM, Harrison RG. Patterns of transcriptome
680 divergence in the male accessory gland of two closely related species of field crickets.
681 Genetics. 2013;193: 501-513.
- 682 41. Shaw KL, Parsons YM, Lesnick SC. QTL analysis of a rapidly evolving speciation
683 phenotype in the Hawaiian cricket *Laupala*. Mol Ecol. 2007;16: 2879-2892.
- 684 42. Shaw KL, Lesnick SC. Genomic linkage of male song and female acoustic preference QTL
685 underlying a rapid species radiation. Proc Natl Acad Sci USA. 2009;106: 9737-9742.

- 686 43. Maroja LS, Clark ME, Harrison RG. *Wolbachia* plays no role in the one-way reproductive
687 incompatibility between the hybridizing field crickets *Gryllus firmus* and *G. pennsylvanicus*.
688 *Heredity*. 2008;101: 435-444.
- 689 44. Hartfelder K, Emlen DJ. Endocrine control of insect polyphenism. In: Gilbert LI, editor.
690 *Insect Endocrinology*: Elsevier; 2011. pp. 464-522.
- 691 45. Danbara Y, Sakamoto T, Uryu O, Tomioka K. RNA interference of timeless gene does not
692 disrupt circadian locomotor rhythms in the cricket *Gryllus bimaculatus*. *J Insect Physiol*.
693 2010;56: 1738-1745.
- 694 46. Lynch JA, Peel AD, Drechsler A, Averof M, Roth S. EGF signaling and the origin of axial
695 polarity among the insects. *Curr Biol*. 2010;20: 1042-1047.
- 696 47. Ellison CK, Wiley C, Shaw KL. The genetics of speciation: genes of small effect underlie
697 sexual isolation in the Hawaiian cricket *Laupala*. *J Evol Biol*. 2011;24: 1110-1119.
- 698 48. Mito T, Shinmyo Y, Kurita K, Nakamura T, Ohuchi H, Noji S. Ancestral functions of
699 Delta/Notch signaling in the formation of body and leg segments in the cricket *Gryllus*
700 *bimaculatus*. *Development*. 2011;138: 3823-3833.
- 701 49. Tomioka K, Uryu O, Kamae Y, Umezaki Y, Yoshii T. Peripheral circadian rhythms and their
702 regulatory mechanism in insects and some other arthropods: a review. *J Comp Physiol B*
703 *Bioch Syst Envir Physiol*. 2012;182: 729-740.
- 704 50. Zeng V, Ewen-Campen B, Horch HW, Roth S, Mito T, Extravour CS. Developmental gene
705 discovery in a hemimetabolous insect: de novo assembly and annotation of a transcriptome for
706 the cricket *Gryllus bimaculatus*. *PLoS ONE*. 2013;8: e61479.
- 707 51. Gnad F, Parsch J. Sebida: a database for the functional and evolutionary analysis of genes
708 with sex-biased expression. *Bioinformatics*. 2006;22: 2577-2579.

- 709 52. Meisel RP, Malone JH, Clark AG. Disentangling the relationship between sex-biased gene
710 expression and X-linkage. *Genome Res.* 2012;22: 1255-1265.
- 711 53. Huylmans AK, Parsch J. Variation in the X:autosome distribution of male-biased genes
712 among *Drosophila melanogaster* tissues and its relationship with dosage compensation.
713 *Genome Biol Evol.* 2015;7: 1960-1971.
- 714 54. Hall JL, Dudley L, Dobner PR, Lewis SA, Cowan NJ. Identification of two human β -tubulin
715 isoforms. *Mol Cell Biol.* 1983;3: 854-862.
- 716 55. Ardell DH, Andersen SO. Tentative identification of a resilin gene in *Drosophila*
717 *melanogaster*. *Insect Biochem Mol Biol.* 2001;31: 965e970.
- 718 56. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, et al. Genome sequence of
719 *Aedes aegypti*, a major arbovirus vector. *Science.* 2007;316: 1718-1723.
- 720 57. Nakagawa Y, Yamane Y, Okanoue T, Tsukita S, Tsukita S. Outer dense fiber 2 is a
721 widespread centrosome scaffold component preferentially associated with mother centrioles:
722 its identification from isolated centrosomes. *Mol Biol Cell.* 2001;2: 1687-1697.
- 723 58. Ishikawa H, Kubo A, Tsukita S, Tsukita S. Odf2-deficient mother centrioles lack
724 distal/subdistal appendages and the ability to generate primary cilia. *Nat Cell Biol.* 2005;7:
725 517-524.
- 726 59. Parisi M, Nuttall R, Naiman D, Bouffard G, Malley J, Andrews J, et al. Paucity of genes on
727 the *Drosophila* X chromosome showing male-biased expression. *Science.* 2003;299: 697-700.
- 728 60. Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. Sex-dependent gene expression and
729 evolution of the *Drosophila* transcriptome. *Science.* 2003;300: 1742-1745.
- 730 61. Chang PL, Dunham JP, Nuzhdin SV, Arbeitman MN. Somatic sex-specific transcriptome
731 differences in *Drosophila* revealed by whole transcriptome sequencing. *BMC Genomics.*

732 2011;12: 364.

733 62. Catalán A, Hutter S, Parsch J. Population and sex differences in *Drosophila melanogaster*
734 brain gene expression. BMC Genomics. 2012;13: 654.

735 63. Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. Evolutionary EST
736 analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc Natl Acad
737 Sci USA. 2001;95: 4051-4.

738 64. Kern AD, Jones CD, Begun DJ. Molecular genetics of male accessory gland in the
739 *Drosophila simulans* complex. Genetics. 2004;167: 725-35.

740 65. Swanson WJ, Wong A, Wolfner MF, Aquadro CF. Evolutionary expressed sequence tag
741 analysis of *Drosophila* identifies genes subjected to positive selection. Genetics. 2004;168:
742 1457-65.

743 66. Clark NJ, Swanson WJ. Pervasive adaptive evolution in primate seminal proteins. PLoS
744 Genet. 2005;1: 335-42.

745 67. Aguadé M. Different forces drive the evolution of the Acp26Aa and Acp26Ab accessory
746 gland genes in the *Drosophila melanogaster* species. Genetics. 1998;150: 1079-89.

747 68. Aguadé M. Positive selection drives the evolution of the Acp29AB accessory gland protein
748 in *Drosophila*. Genetics. 1999;152: 543-51.

749 69. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence
750 data. Bioinformatics. 2014;30: 2114-2120.

751 7. Innocenti P, Morrow EH. The sexually antagonistic genes of *Drosophila melanogaster*. PLoS
752 Biol. 2010;8: e1000335.

753 70. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. *De novo*
754 transcript sequence reconstruction from RNA-seq: reference generation and analysis with

755 Trinity. Nat Prot. 2014; 8:1-8.

756 71. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. eggNOG
757 4.5: a hierarchical orthology framework with improve functional annotations for eukaryotic,
758 prokaryotic and viral sequences. Nucl Acids Res. 2016;44(D1): D286-D293.

759 72. Transdecoder: Find Coding Regions within Transcripts. Available from:
760 <https://github.com/TransDecoder/TransDecoder.wiki.git>

761 73. Waterhouse RM, Zdobnov EM, Tegenfeldt F, Li J, Kriventseva EV. OrthoDB: a hierarchical
762 catalog of animal, fungal and bacterial orthologs. Nucl Acids Res. 2013;41: D358-D365.

763 74. Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation
764 and pathway reconstruction server. Nucl Acids Res. 2007;35: W182-W185.

765 75. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucl Acids Res.
766 2000;28: 27-30.

767 76. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-seq data with or
768 without a reference genome. BMC Bioinformatics. 2011;12: 323.

769 77. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie2. Nature Meth. 2012;9:
770 357-359.

771 78. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-
772 seq data with DESeq2. Genome Biol. 2014;15: 550.

773 79. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B.
774 Bioconductor: open software development for computational biology and bioinformatics.
775 Genome Biol. 2004;5: R80.

776 80. R Core Team. R: A language and environment for statistical computing. R Foundation for
777 Statistical Computing, Vienna, Austria. 2017.

778 81. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of
779 progressive multiple sequence alignment through sequence weighting, position-specific gap
780 penalties and weight matrix choice. Nucl Acids Res. 1994;22: 4673-80.

781 82. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version
782 7.0 for bigger datasets. Mol Biol Evol. 2016;33: 1870-4.

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801 **Figures**

802 **Figure 1.** BUSCO analysis for *G. assimilis* (Gas), *G. bimaculatus* (Gbi) and *G. firmus* (Gfi)
803 transcriptomes and *Locusta migratoria* (Lmi) genome. Assessing highly conserved orthologs
804 was made considering the related Arthropod species (n = 2,675 for transcriptomes and n = 1,066
805 for genome). The recovered matches are classified as complete if their lengths are within the
806 expectation of the BUSCO profile match lengths. If these are found more than once they are
807 classified as duplicated. The matches that are only partially recovered are classified as
808 fragmented, and BUSCO groups for which there are no matches that pass the tests of orthology
809 are classified as missing. Note the high values of duplicated and missing copy due putatively to
810 events of gene duplications in cricket's transcriptomes. Also note the high missing values in Lmi
811 genome.

812

813 **Figure 2.** The Venn diagram illustrates the results given by KEGG candidate orthologues
814 sequences corresponding to insect hormone biosynthesis (A), circadian rhythm (B), circadian
815 entrainment (C) and circadian rhythm-fly (D) pathways for *L. migratoria* genome (Lmi), *G.*
816 *assimilis* (Gas), *G. bimaculatus* (Gbi) and *G. firmus* (Gfi) transcriptomes. The sequences sets
817 show the number of orthologs for each species and the number of orthologs genes shared
818 between species are shown in the intersection areas of the diagram. Note that the number of
819 orthologs recovered varies among species, suggesting gain or loss of orthologs genes throughout
820 Orthoptera evolution.

821

822 **Figure 3.** Barplot comparing the number of single copy genes and the putative number of genes
823 with two or more than two copies in *G. assimilis* (Gas), *G. bimaculatus* (Gbi) and *G. firmus* (Gfi)

824 transcriptomes and *L. migratoria* genome (Lmi) for insect hormone biosynthesis (A), circadian
825 rhythm (B), circadian entrainment (C) and circadian rhythm-fly (D) KEGG pathways. Note the
826 extensive number of duplicated genes in many pathways. Considering that the species studied
827 belong to different suborders (i.e., *Gryllus* belongs to Ensifera and *L. migratoria* to Caelifera)
828 and that Ensifera is basal in Orthoptera phylogeny (Song et al. 2015), we assumed that the
829 differences observed for copy numbers among species are caused by episodes of gain or loss of
830 orthologs as well as gene duplications over Orthoptera evolution. It also means that Lmi
831 underwent recent events of duplications compared to crickets. Furthermore, gene duplications
832 are putatively more recent in Gas than Gbi and Gfi.

833

834 **Figure 4.** MA-plot of RNA-seq data comparing male head vs female head (A) and testis vs
835 ovary (B) in *G. assimilis*. The x-axis shows mean expression for each genes and the y-axis shows
836 log₂ fold change, as estimated in DESeq2. Points in red are differentially regulated between
837 treatments at an adjusted *p*-value < 0.05.

838

839 **Figure 5.** Heatmap showing the expression data of the top 50 up-regulated and down-regulated
840 genes by adjusted *p*-value (< 0.05) with different fold changes in *G. assimilis*. Male head vs
841 female head (A) and testis vs ovary (B). The log₂ fold change expression differences are
842 indicated by colors as denoted in the bar scale. MH= male head biological replicates; FH=
843 female head biological replicates; MT= male testis biological replicates; FO= female ovary
844 biological replicates.

845

846 **Figure 6.** Dotchart showing the expression profiles of genes corresponding to sex determination
847 in testis vs ovary (A), insect hormone biosynthesis in male head vs female head (B) and testis vs
848 ovary (C) for *G. assimilis*. We report only those genes considered as up- and down-regulated
849 within each condition using an adjusted p -value < 0.05 . In the plot are indicated by colors the
850 gene name followed by contig number. Note the male-biased genes expression in A-C.

851

852 **Figure 7.** Estimating of Ka vs Ks values (A) and plot for estimation of ω values (B) for 27
853 putative protein-encoding orthologs influencing the sex and reproductive fitness in field crickets
854 *G. assimilis*, *G. bimaculatus* and *G. firmus* (see Table 7). The red lines in (A) and (B) represent
855 the neutral expectation of $Ka/Ks = 1$. The black line in (A) represents the threshold value of
856 Ka/Ks considered to identify genes that may have undergone episodes of positive selection.

857

858

859

860

861

862

863

864

865

866

867

868 **Supporting Information**

869 **S1 Table.** Summary statistics of *Gryllus* species based on all transcripts contigs.

870

871 **S2 Table.** Trinotate summary annotation of *Gryllus* species based on all the isoforms.

872

873 **S3 Table.** *G. assimilis* orthologs of genes involved in the sex determination including their
874 isoforms and differential expression information.

875

876 **S4 Table.** Detection of selection by Nei-Gojobori (Jukes-Cantor) method considering an p -value
877 < 0.05 (Fisher's exact test). We contrasted protein-encoding genes influencing reproductive
878 fitness and genes encoding nonproductive proteins (housekeeping genes).

879

880 **S1 Figure.** KEGG pathway analysis [30] shows orthologs in the cricket *G. assimilis* involved in
881 insect hormone biosynthesis (A), circadian rhythm (B), circadian entrainment (C) and circadian
882 rhythm-fly (D), as indicated by green highlighting.

883

884 **S2 Figure.** Dotchart showing the expression profiles of genes corresponding to circadian rhythm
885 KEGG orthologs genes in male head vs female head (A) and testis vs ovary (B) for *G. assimilis*.
886 We report only those genes considered as up- and down-regulated within each condition using an
887 adjusted p -value < 0.05 . In the plot are indicated by colors the gene name followed by contig
888 number. Note the female-biased genes expression in A.

889

890 **S3 Figure.** Dotchart showing the expression profiles of genes corresponding to circadian
891 entrainment KEGG orthologs genes in male head vs female head (A) and testis vs ovary (B) for
892 *G. assimilis*. We report only those genes considered as up- and down-regulated within each
893 condition using an adjusted p -value < 0.05 . In the plot are indicated by colors the gene name
894 followed by contig number. Note the male-biased genes expression in A and the female-biased
895 genes expression in B.

896

897 **S4 Figure.** Dotchart showing the expression profiles of genes corresponding to circadian rhythm
898 fly KEGG orthologs genes in male head vs female head (A) and testis vs ovary (B) for *G.*
899 *assimilis*. We report only those genes considered as up- and down-regulated within each
900 condition using an adjusted p -value < 0.05 . In the plot are indicated by colors the gene name
901 followed by contig number. Note the female-biased genes expression in A and the male-biased
902 genes expression in B.

903

904 **S1 Appendix:** *G. assimilis*, *G. bimaculatus* and *G. firmus* orthologs of genes involved in the sex
905 determination including their isoforms.

906

907 **S2 Appendix:** KEGG unique orthologs involved in Insect hormone biosynthesis, Circadian
908 rhythm, Circadian entrainment and Circadian rhythm-fly sharing between *L. migratoria* genome
909 (Lmi) and transcriptomes for *G. assimilis* (Gas), *G. bimaculatus* (Gbi) and *G. firmus* (Gfi).
910 Positive (+) and negative (-) sign indicate presence or absence of genes, respectively.

911

912 **S3 Appendix:** *L. migratoria*, *G. assimilis*, *G. bimaculatus* and *G. firmus* orthologs of genes

913 involved in insect hormone biosynthesis, circadian rhythm, circadian entrainment and circadian
914 rhythm-fly KEGG pathway including their isoforms. KEGG= Kyoto Encyclopedia of Genes and
915 Genomes; KO= KEGG orthology number. *Insect hormone biosynthesis is known to regulate
916 development, reproductive maturation and behavior in insects.

917

918 **S4 Appendix:** *G. assimilis* expression data of the most highly expressed genes isoforms/genes
919 corresponding to the heatmap. Adjusted *p*-value < 0.05. MBG= male biased gene; FBG= female
920 biased gene.

921

922 **S5 Appendix:** *G. assimilis* orthologs of genes involved in insect hormone biosynthesis, circadian
923 rhythm, circadian entrainment and circadian rhythm-fly KEGG pathway including their isoforms
924 and differential expression information. KEGG= Kyoto Encyclopedia of Genes and Genomes;
925 KO= KEGG orthology number; log₂FC= log₂ fold-change; MBG= male-biased gene; FBG=
926 female-biased gene. We report only those isoforms considered as sex-biased genes within each
927 tissue, i.e., those isoforms with log₂FC > 0 and adjusted *p*-value < 0.1. *Insect hormone
928 biosynthesis are known to regulate development, reproductive maturation and behavior in
929 insects.

930

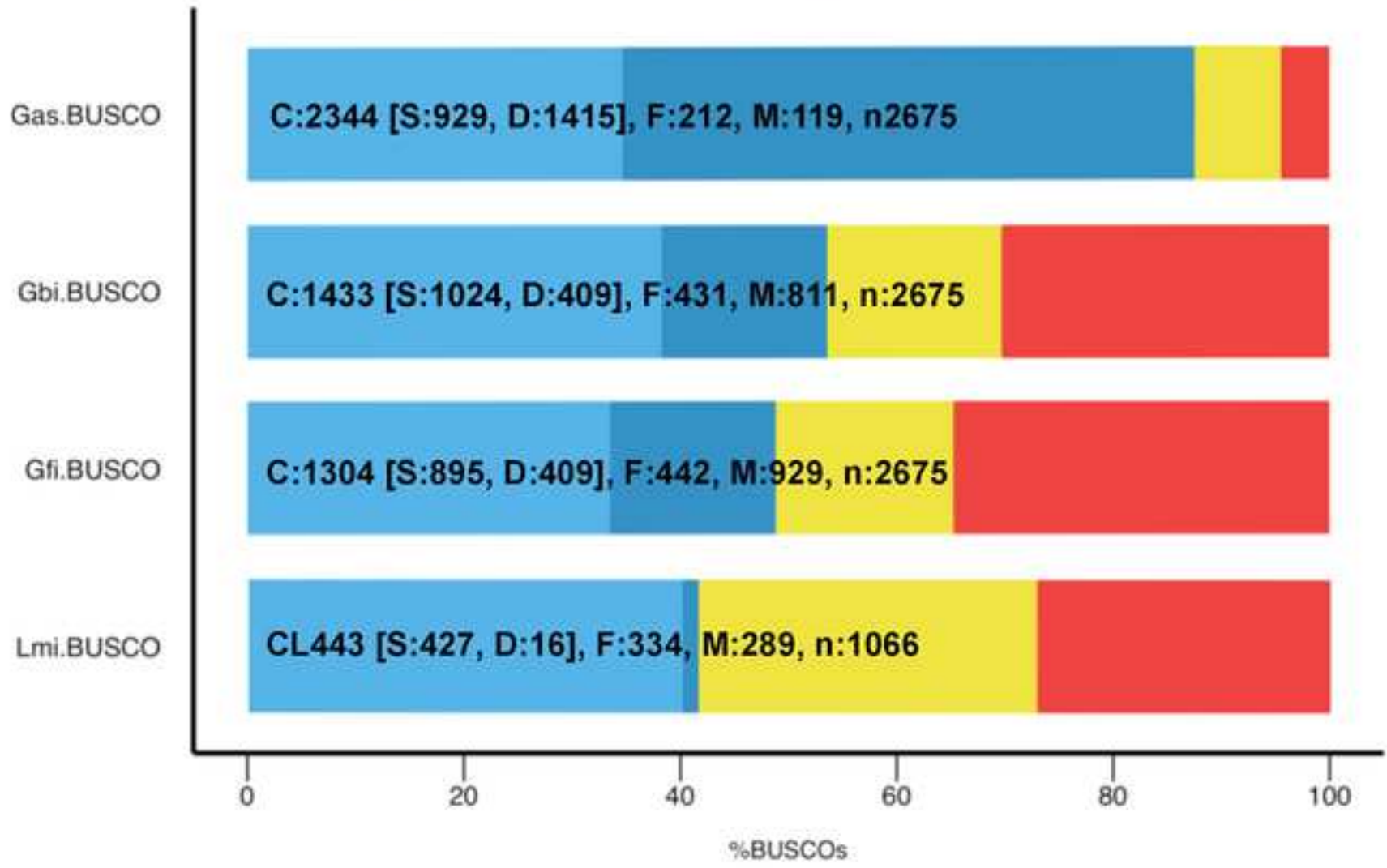
931

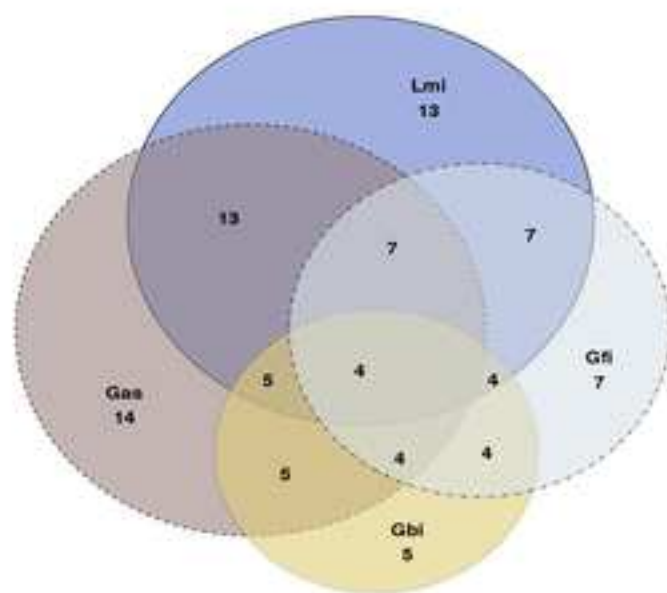
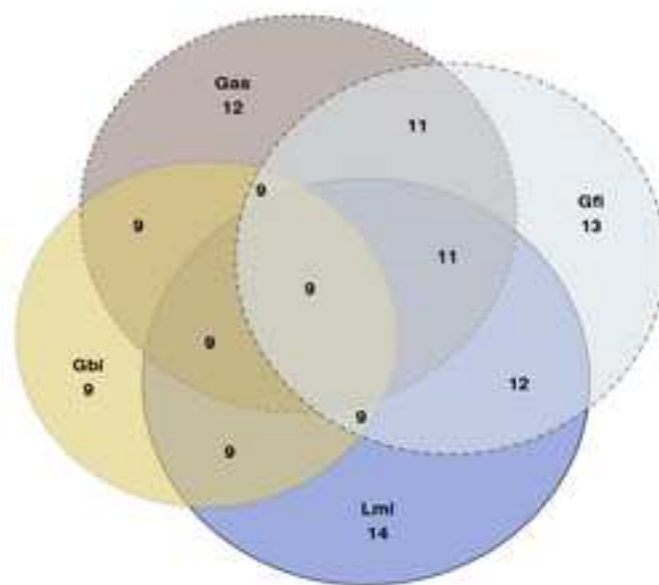
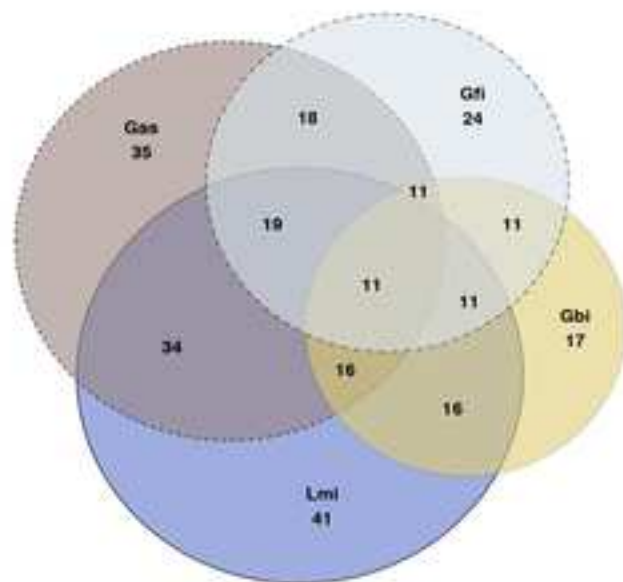
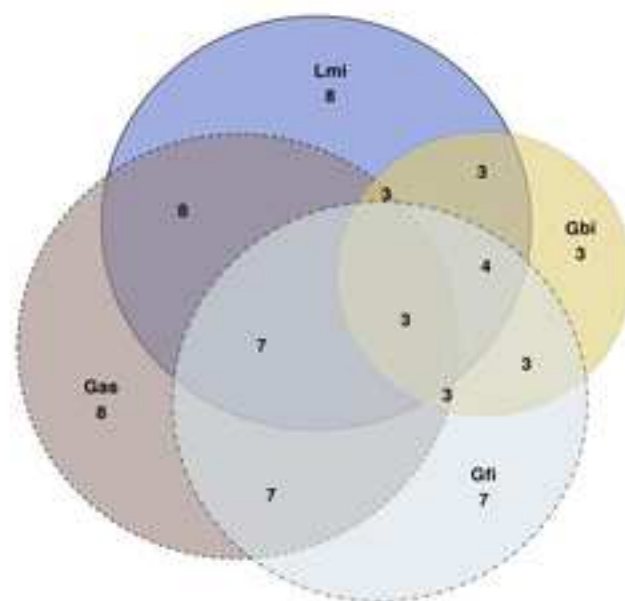
932

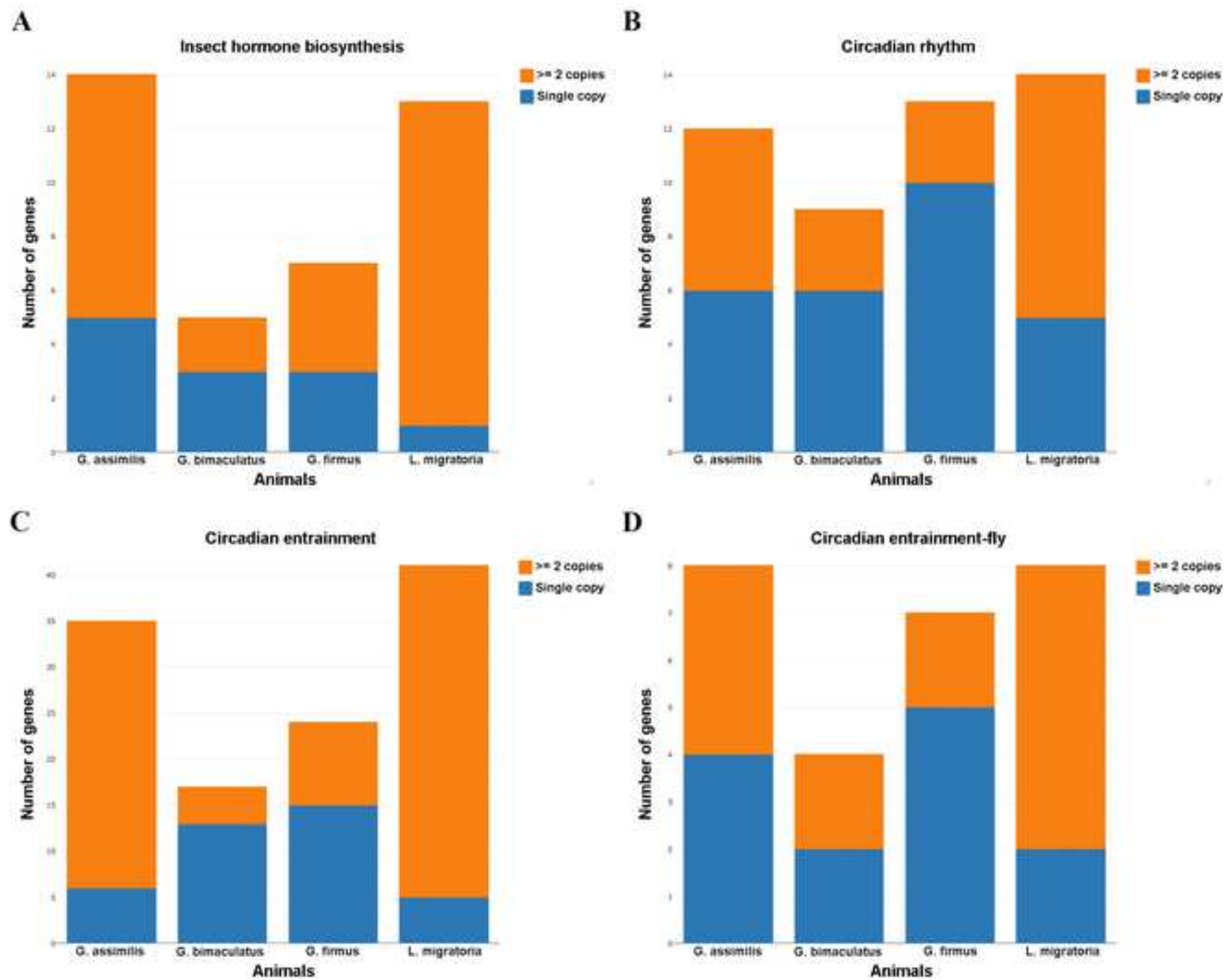
933

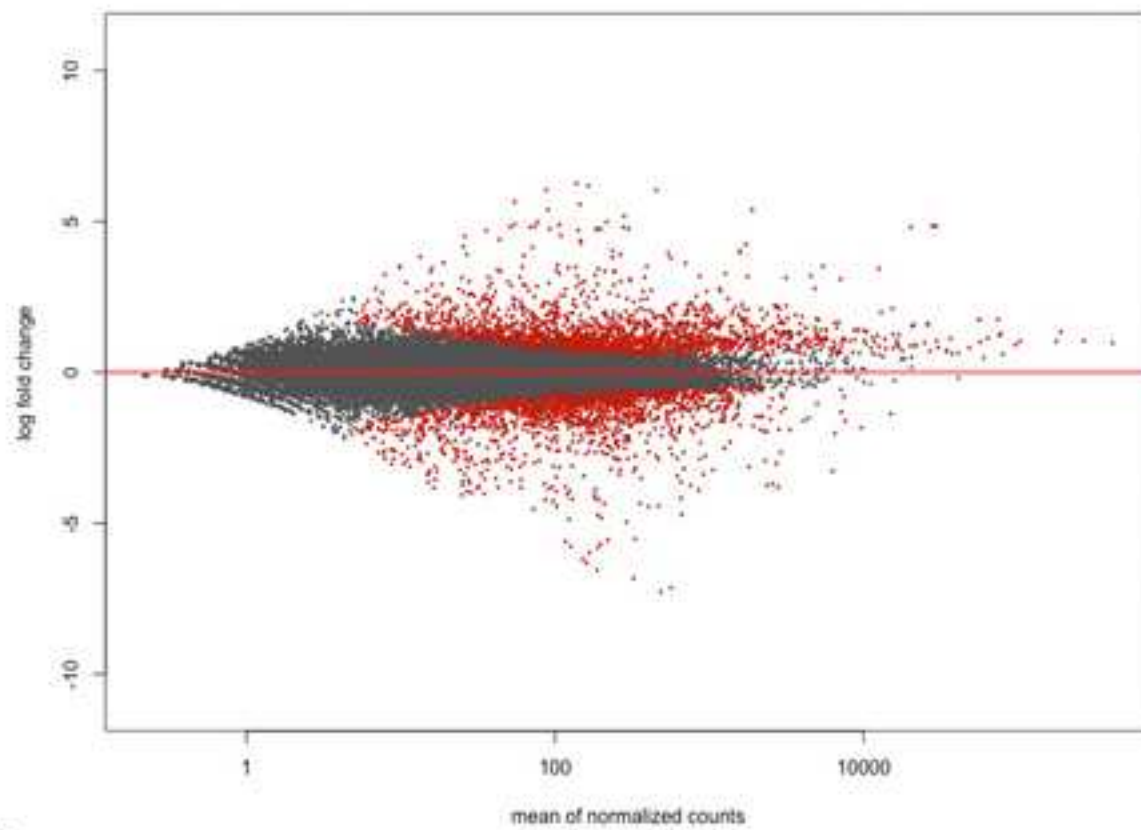
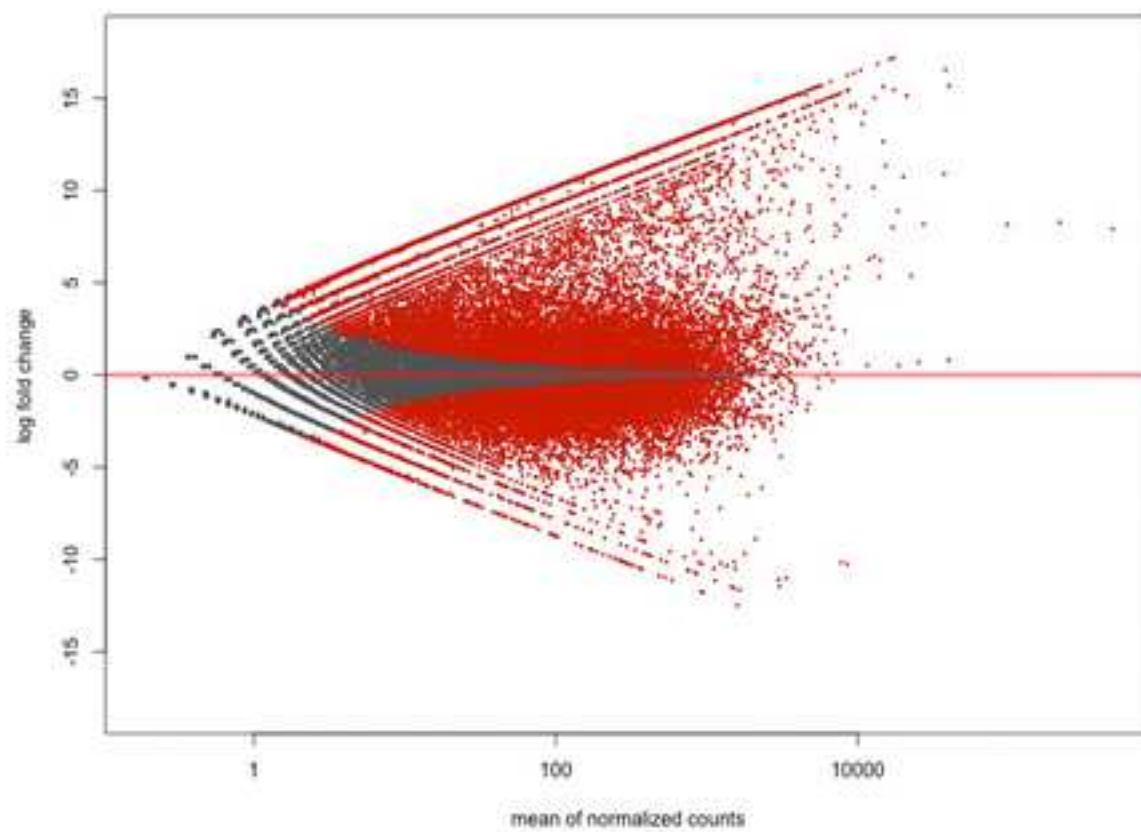
BUSCO Assessment Results

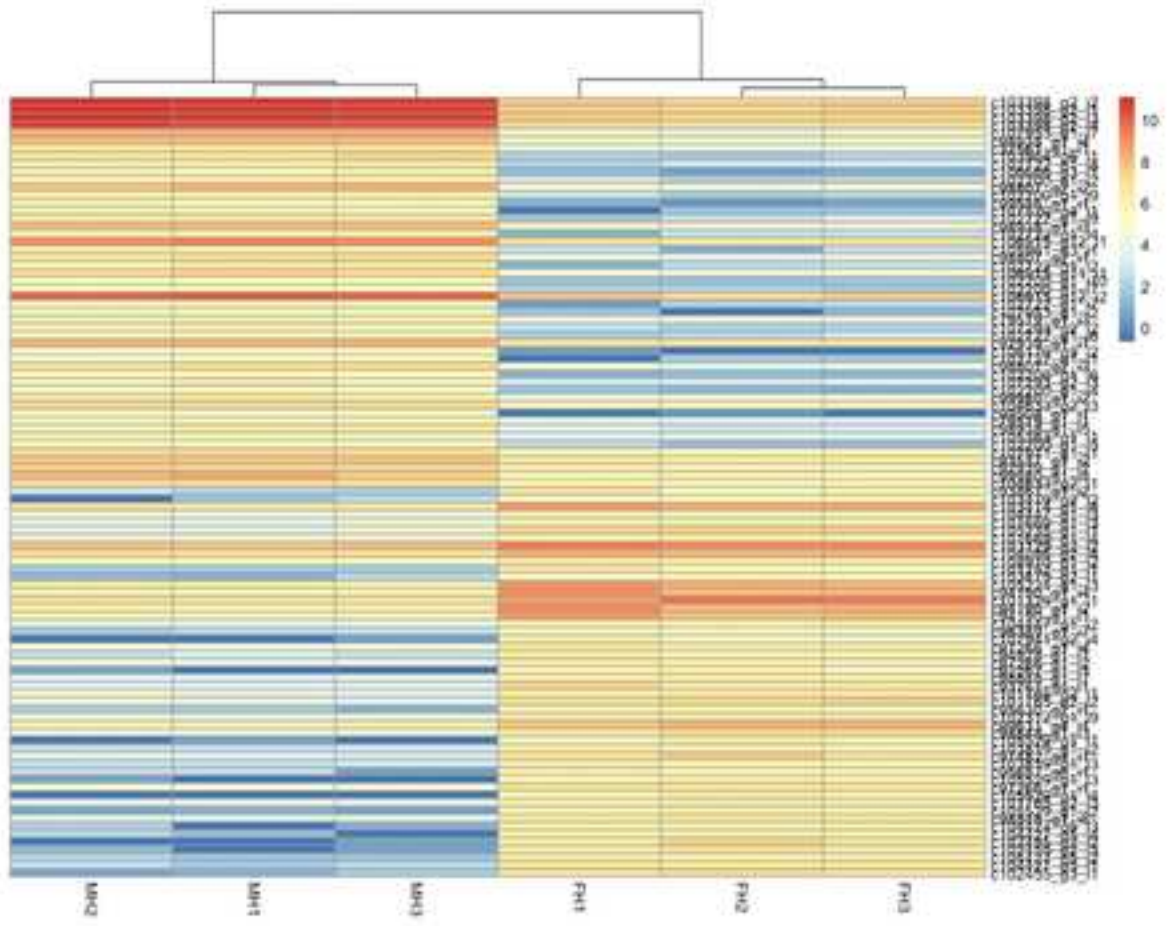
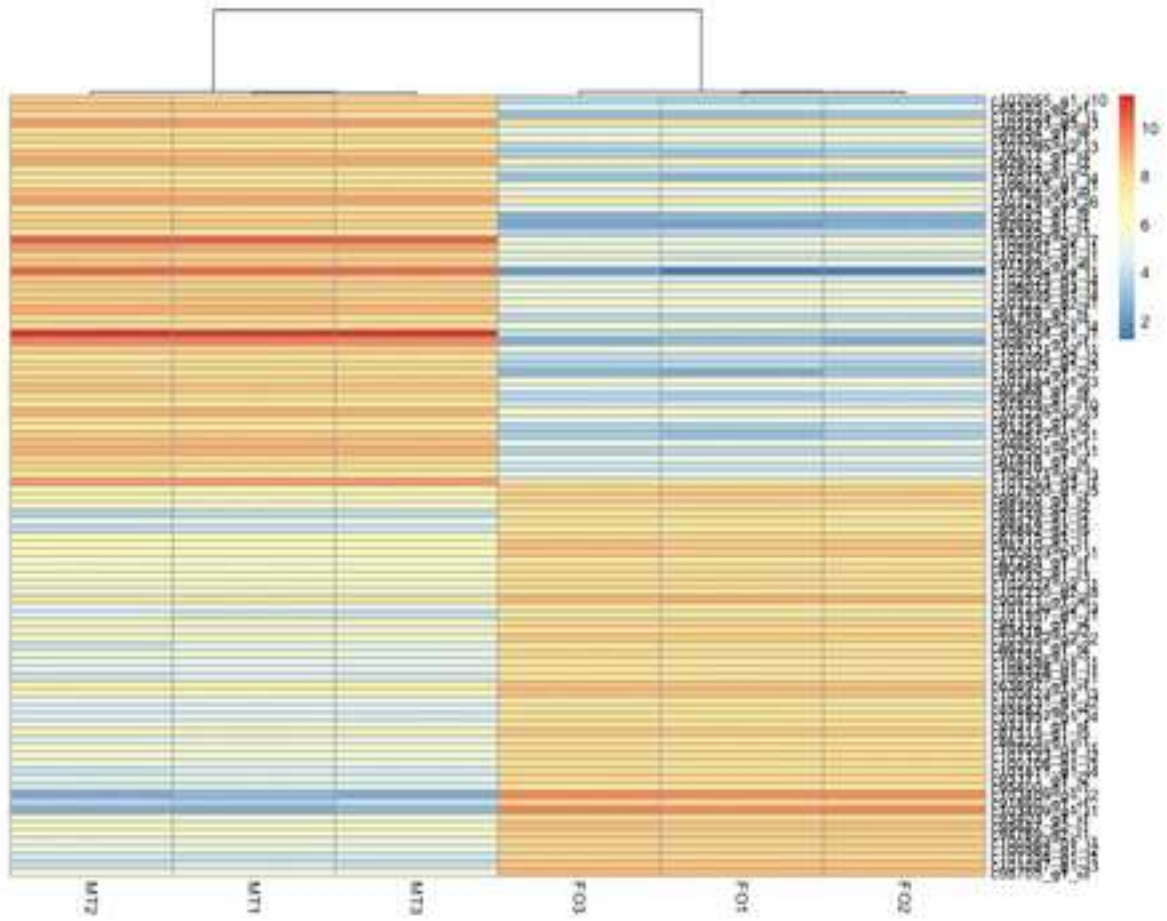
Complete (C) and single-copy (S) Complete (C) and duplicated (D)
Fragmented (F) Missing (M)

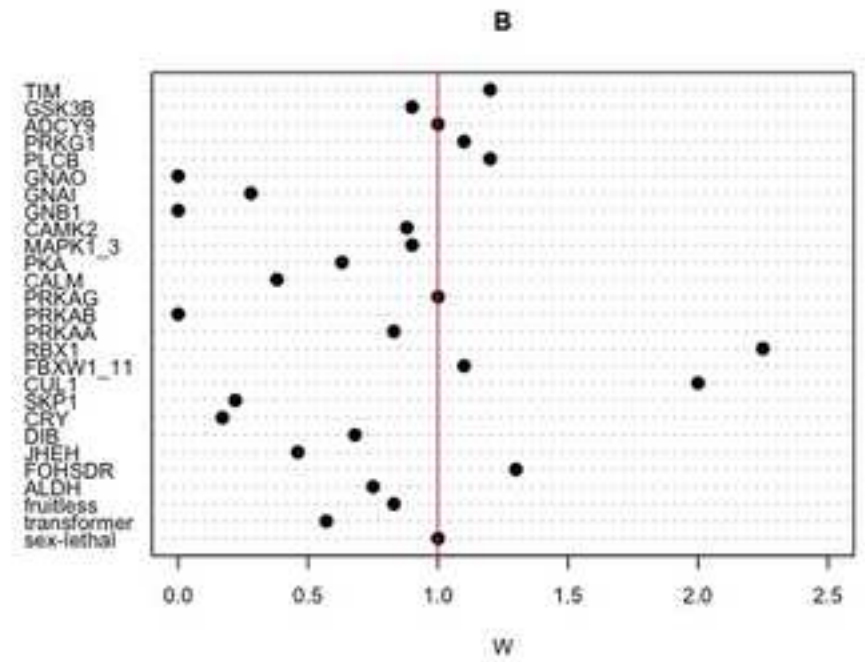
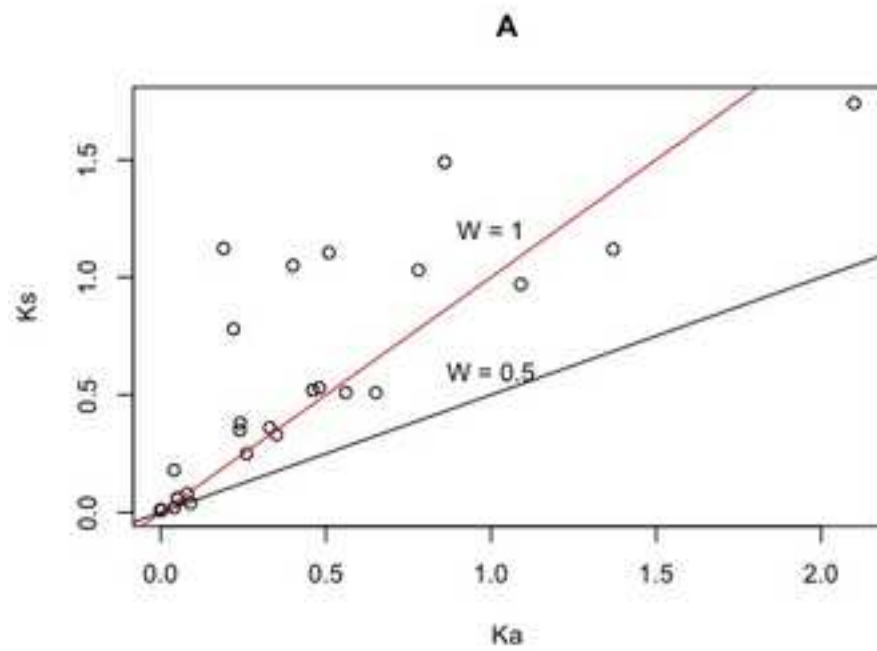


A Insect hormone biosynthesis**B** Circadian rhythm**C** Circadian entrainment**D** Circadian rhythm fly



A**DESeq2: *G. assimilis* Male head vs. Female head****B****DESeq2: *G. assimilis* Male testis vs. Female ovary**

A**B**





[Click here to access/download](#)

Supporting Information - Compressed/ZIP File Archive
supporting_information.zip



Supplementary Table 1: Summary statistics of *Gryllus* species based on all transcripts contigs.

Trinity outputs	<i>G. assimilis</i>	<i>G. bimaculatus</i>	<i>G. firmus</i>
Total trinity transcripts	227,254	37,069	63,940
Total trinity “genes”	150,981	30,847	46,790
Percent GC	39.15	37.57	39.91
Longest contig (bp)	28,164	8,208	57,165
Shortest contig (bp)	201	201	201
Transcript contig N10	5,978	2,768	3,048
Transcript contig N20	4,256	2,157	2,143
Transcript contig N30	3,199	1,763	1,644
Transcript contig N40	2,433	1,427	1,270
Transcript contig N50	1,818	1,123	972
Median contig length	422	383	377
Average contig	913.31	690.03	646.52
Total assembled bases	207,553,576	25,578,816	41,338,187

Supplementary Table 2. Trinotate summary annotation of *Gryllus* species based on all the isoforms.

Annotations	<i>G. assimilis</i>	<i>G. bimaculatus</i>	<i>G. firmus</i>
Total BLASTx hit	48,202	13,933	27,818
<i>Homo sapiens</i> ¹	10,538	3,070	4,904
<i>D. melanogaster</i> ¹	9,307	2,464	3,058
<i>Mus</i> ¹	7,837	2,545	8,089
<i>Danio rerio</i> ¹	1,767	655	691
<i>Gallus</i> ¹	1,169	353	488
<i>C. elegans</i> ¹	936	141	250
<i>Arabidopsis</i> ¹	237	67	80
<i>Bombix mori</i> ¹	204	57	81
<i>Locusta migratoria</i> ¹	107	21	22
<i>Yeast</i> ¹	68	20	49
<i>Nasonia vitripennis</i> ¹	58	0	16
Total no. of ORFs	60,889	14,369	22,270
No. of complete ORFs	31,964	5,043	4,661
Longest ORF*	19,464	5,529	57,162
GO BLAST			
<i>Cellular component</i>	37,852	11,968	24,301
<i>Biological process</i>	500	165	232
<i>Molecular function</i>	4,499	1,073	2,056
eggNOG	21,749	6,933	14,405

The number of annotations obtained for the assembly is given from best BLASTx hits against other protein sequence databases (E-value 1×10^{-5});

¹Some biological models in which *Gryllus* species have putative orthologs genes;

*Longest ORF in base pairs;

GO= gene ontology BLAST;

eggNOG= evolutionary genealogy of genes: Non-supervised Orthologous Groups.

Supplementary Table 3. *G. assimilis* orthologs of genes involved in the sex determination including their isoforms and differential expression information.

Gene name	<i>G. assimilis</i> ID	Differential expression (log2FC)		BLASTx ID - species
		Head	Gonad	
<i>sex-lethal</i>	c105826_g1_i2	-	MBG (0.7)	SXL_CERCA - <i>Ceratitis capitata</i>
	c105826_g1_i4	-	MBG (0.9)	SXL_CERCA - <i>C. capitata</i>
	c105826_g1_i5	-	MBG (0.5)	SXL_CERCA - <i>C. capitata</i>
<i>transformer</i>	c155351_g1_i1	-	-	TRA2_DROVI - <i>Drosophila virilis</i>
	c100340_g1_i1	-	FBG (0.6)	TRA2B_RAT - <i>Rattus</i> sp.
	c100340_g1_i3	-	FBG (0.5)	TRA2B_RAT - <i>Rattus</i> sp.
	c100340_g1_i4	-	FBG (0.8)	TRA2B_RAT - <i>Rattus</i> sp.
	c100340_g1_i5	-	FBG (1)	TRA2B_RAT - <i>Rattus</i> sp.
<i>doublesex</i>	c97760_g1_i2	-	-	DSX_DROME - <i>Drosophila melanogaster</i>
	c97760_g1_i5	-	-	DSX_DROME - <i>D. melanogaster</i>
<i>fruitless</i>	c104957_g1_i1	-	-	FRU_DROME - <i>D. melanogaster</i>
	c104957_g2_i1	-	MBG (0.6)	FRU_DROME - <i>D. melanogaster</i>
	c104957_g2_i2	-	MBG (2.5)	FRU_DROME - <i>D. melanogaster</i>
	c104957_g2_i3	-	-	FRU_DROME - <i>D. melanogaster</i>
	c104957_g2_i4	-	MBG (1.6)	FRU_DROME - <i>D. melanogaster</i>
	c104957_g2_i6	-	-	FRU_DROME - <i>D. melanogaster</i>
	c104957_g2_i7	-	-	FRU_DROME - <i>D. melanogaster</i>
	c104957_g2_i8	-	-	FRU_DROME - <i>D. melanogaster</i>
	c104957_g2_i9	-	MBG (1.4)	FRU_DROME - <i>D. melanogaster</i>
	c104957_g2_i10	-	MBG (0.6)	FRU_DROME - <i>D. melanogaster</i>
	c104957_g2_i12	-	-	FRU_DROME - <i>D. melanogaster</i>
	c104957_g2_i13	-	-	FRU_DROME - <i>D. melanogaster</i>
	c21731_g1_i1	-	-	FRU_DROME - <i>D. melanogaster</i>
	c21731_g1_i2	-	-	FRU_DROME - <i>D. melanogaster</i>

log2FC= log2 fold-change; MBG= male-biased gene; FBG= female-biased gene. Those isoforms with log2FC > 0 and log2FC < 0 and adjusted *p*-value < 0.05 were considered as sex-biased genes within each tissue.

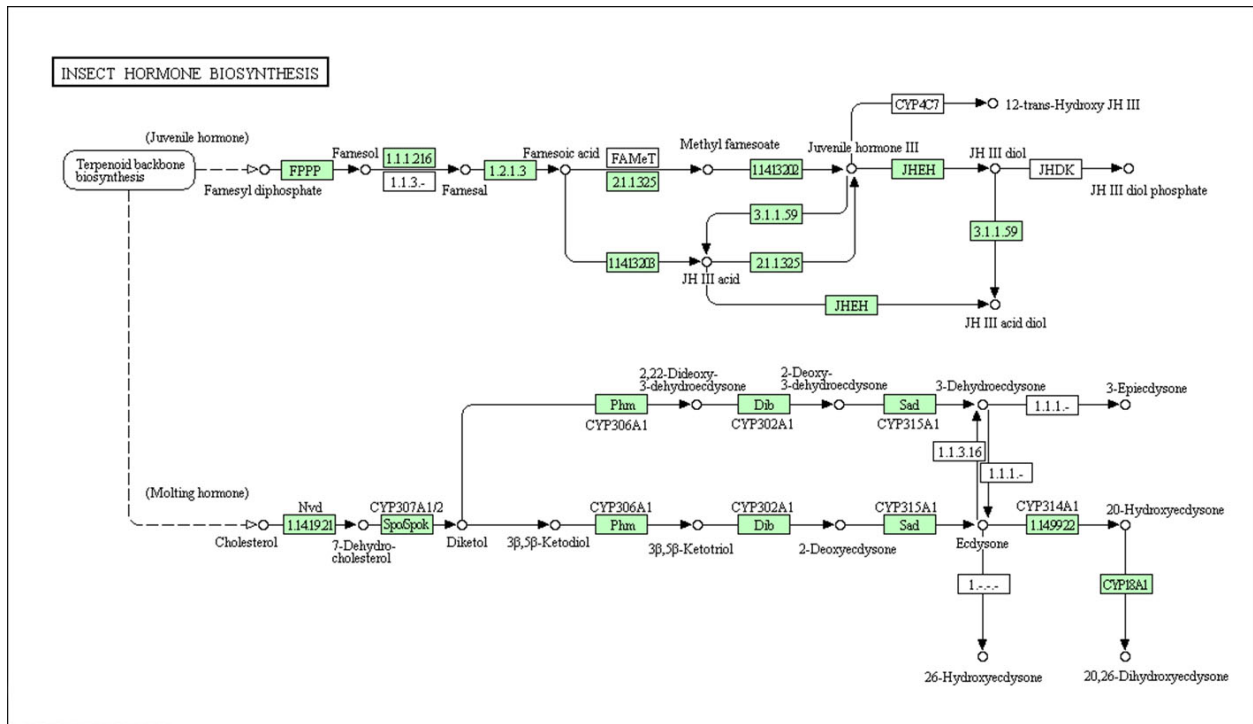
Supplementary Table 4. Detection of selection by Nei-Gojobori (Jukes-Cantor) method considering an p -value < 0.05 (Fisher's exact test). We contrasted protein-encoding genes influencing reproductive fitness and genes encoding nonproductive proteins (housekeeping genes).

Gene name	Ka	Ks	$\omega = Ka/Ks$
<i>sex-lethal</i>	0.26	0.25	1
<i>transformer</i>	0.86	1.49	0.57
<i>fruitless</i>	0.05	0.06	0.83
ALDH	0.78	1.03	0.75
FOHSDR	0.651	0.509	1.3
JHEH	0.509	1.105	0.46
DIB	0.239	0.35	0.68
CRY	0.19	1.123	0.17
SKP1	0.04	0.18	0.22
CUL1	0.04	0.02	2
FBXW1_11	0.56	0.51	1.1
RBX1	0.09	0.04	2.25
PRKAA	0.05	0.06	0.83
PRKAB	0	0.01	0
PRKAG	0.08	0.08	1
CALM	0.4	1.05	0.38
PKA	0.24	0.38	0.63
MAPK1_3	0.48	0.53	0.9
CAMK2	0.46	0.52	0.88
GNB1	0	0.01	0
GNAI	0.22	0.78	0.28
GNAO	0	0.01	0
PLCB	1.37	1.12	1.2
PRKG1	1.09	0.97	1.1
ADCY9	0.35	0.33	1
GSK3B	0.33	0.36	0.9
TIM	2.1	1.74	1.2
E2F4*	0.5	0.7	0.7
XRCC5*	0.75	0.97	0.77
ILF2*	0	0.02	0
EIF1A*	0	0.01	0
RNH2B*	0.51	0.86	0.6

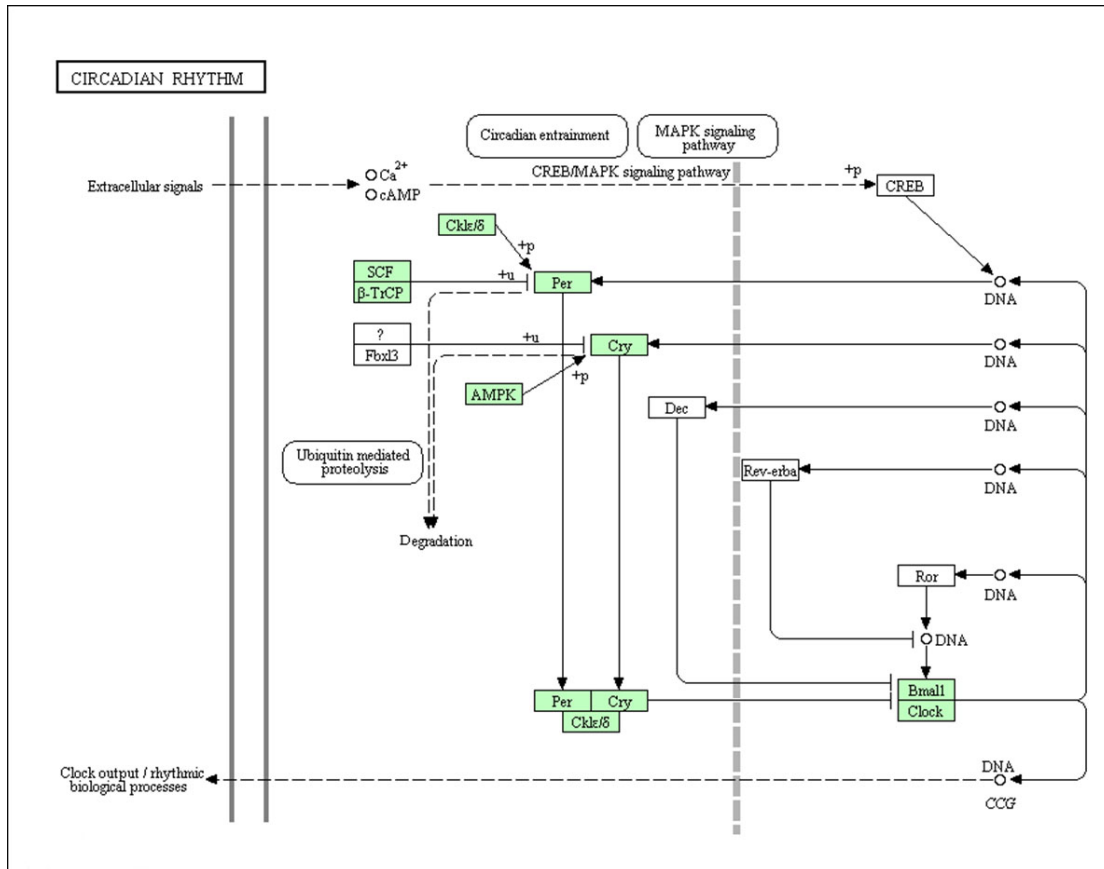
Ka = nonsynonymous substitutions;

Ks = synonymous substitutions;

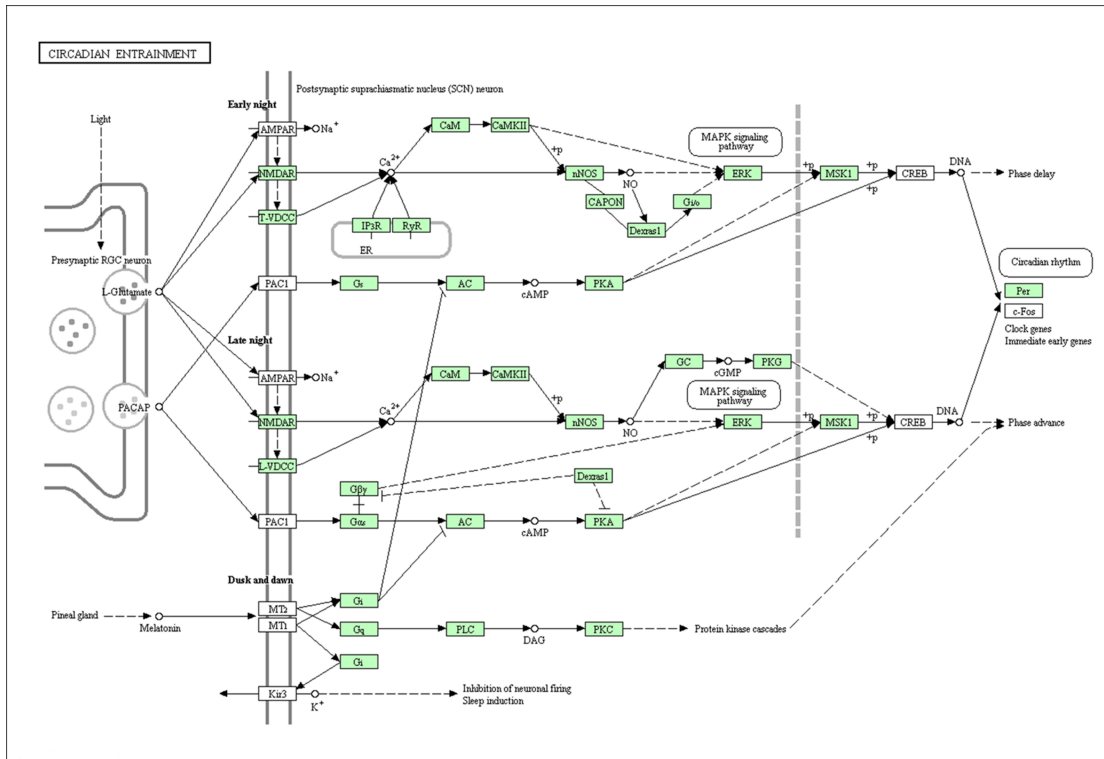
*= Housekeeping genes.



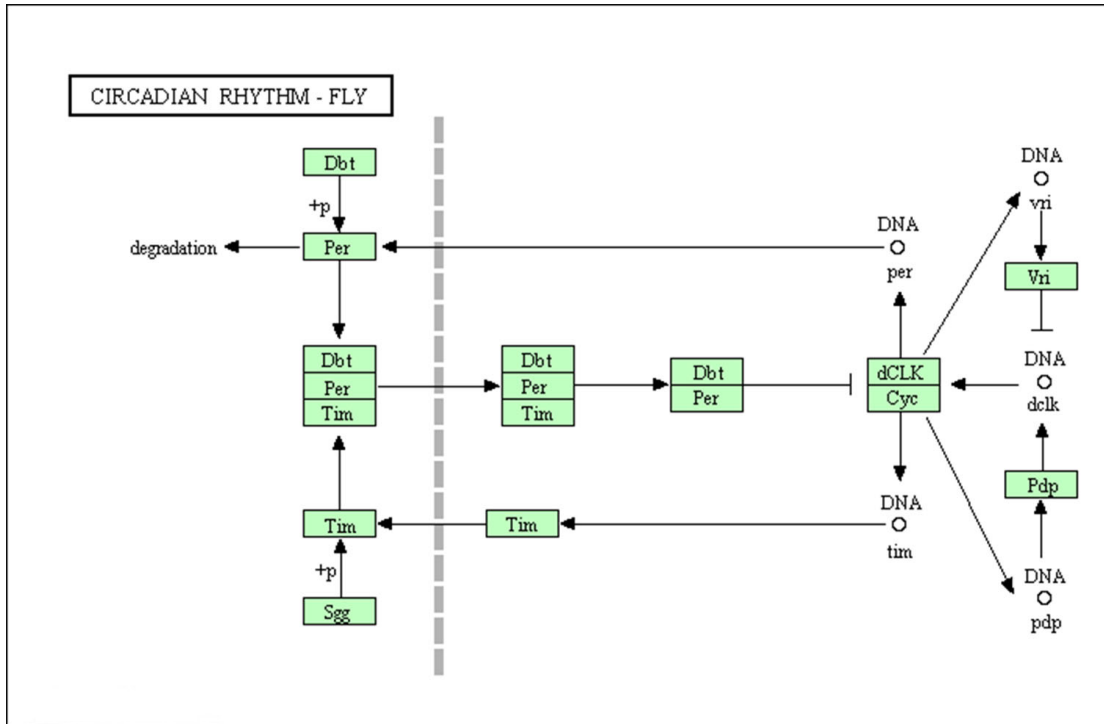
Supplementary Figure S1 A. Insect hormone biosynthesis pathway in *G. assimilis*. KEGG pathway analysis [30] shows orthologs in the cricket *G. assimilis* involved in insect hormone biosynthesis, as indicated by green highlighting.



Supplementary Figure S1 B. Circadian rhythm pathway in *G. assimilis*. KEGG pathway analysis [30] shows orthologs in the cricket *G. assimilis* involved in circadian rhythm, as indicated by green highlighting.



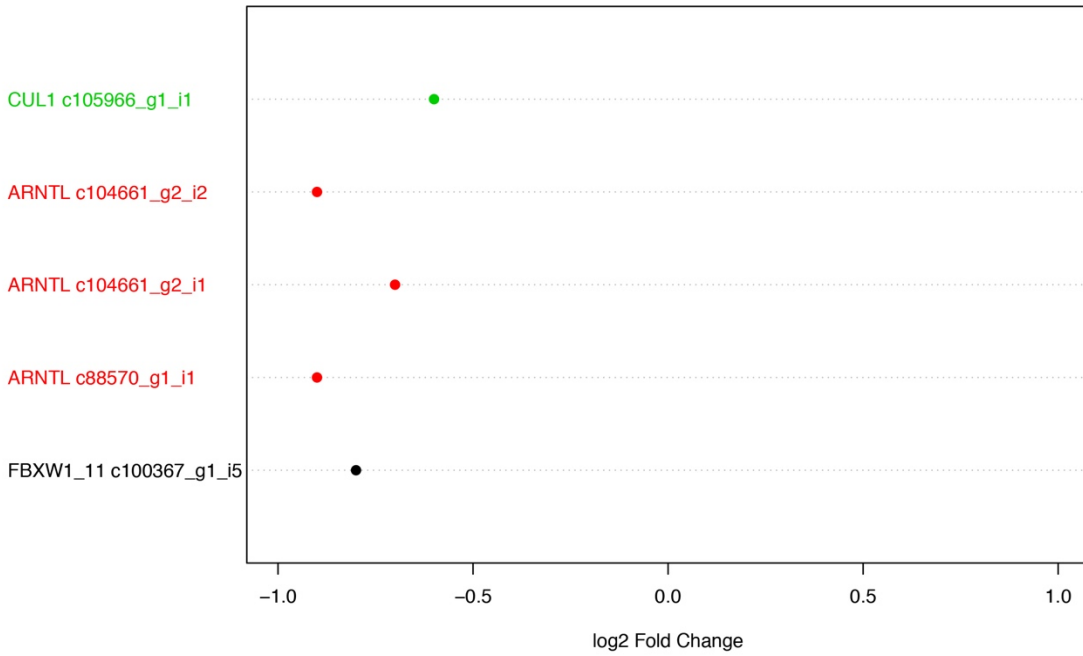
Supplementary Figure S1 C. Circadian entrainment pathway in *G. assimilis*. KEGG pathway analysis [30] shows orthologs in the cricket *G. assimilis* involved in circadian entrainment, as indicated by green highlighting.



Supplementary Figure S1 D. Circadian rhythm-fly pathway in *G. assimilis*. KEGG pathway analysis [30] shows orthologs in the cricket *G. assimilis* involved in circadian rhythm-fly, as indicated by green highlighting.

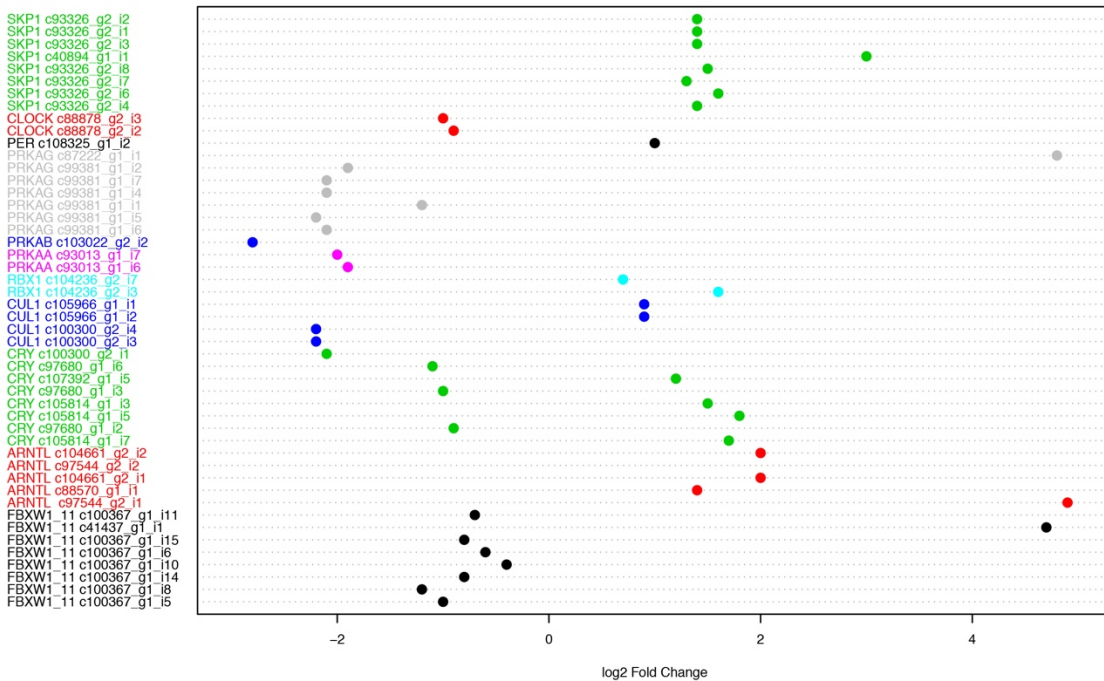
A

DEG: Circadian rhythm genes in *G. assimilis* male vs. female head



B

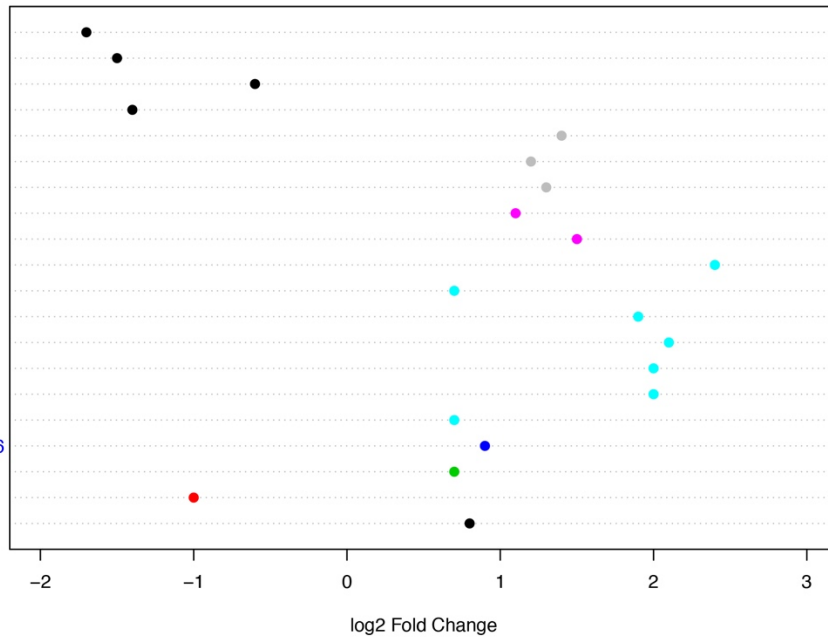
DEG: Circadian rhythm genes in *G. assimilis* testis vs. ovary



A

DEG: Circadian entrainment genes in *G. assimilis* male vs. female head

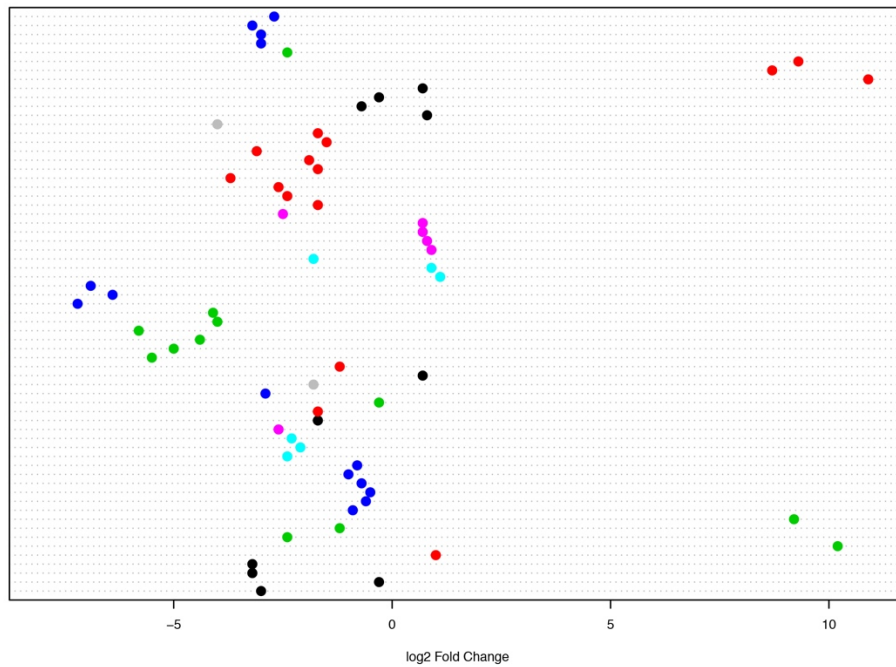
PRKG1 c103658_g1_i3
 PRKG1 c103658_g1_i2
 PRKG1 c107317_g1_i3
 PRKG1 c103658_g1_i1
 RYR2 c103458_g3_i2
 RYR2 c107689_g1_i1
 RYR2 c109001_g1_i2
 ADCY2 c74935_g1_i3
 ADCY2 c71860_g1_i2
 CALM c92733_g1_i6
 CALM c108404_g1_i1
 CALM c92733_g1_i1
 CALM c92733_g1_i4
 CALM c92733_g1_i3
 CALM c92733_g1_i2
 CALM c108404_g1_i2
 CACNA1D c102695_g2_i6
 PKA c94301_g2_i1
 GNG13 c98518_g1_i1
 GNAS c102442_g1_i3



B

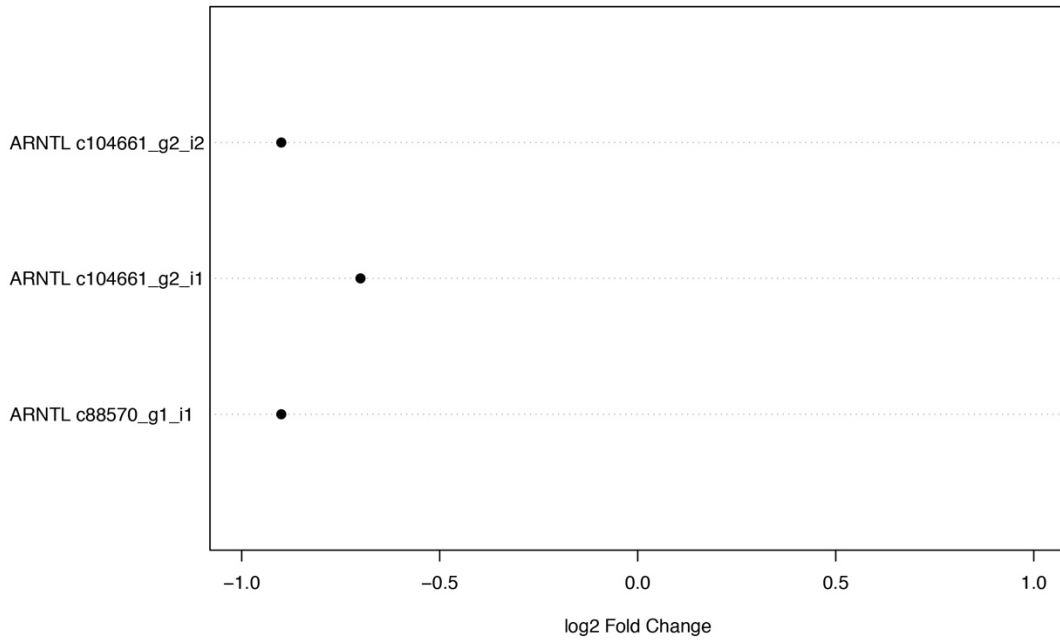
DEG: Circadian entrainment genes in *G. assimilis* testis vs. ovary

MAPK1 c103658_g1_i3
 MAPK1 c103658_g1_i2
 MAPK1 c107317_g1_i3
 MAPK1 c103658_g1_i1
 RYR2 c103458_g3_i2
 RYR2 c107689_g1_i1
 RYR2 c109001_g1_i2
 ADCY2 c74935_g1_i3
 ADCY2 c71860_g1_i2
 CALM c92733_g1_i6
 CALM c108404_g1_i1
 CALM c92733_g1_i1
 CALM c92733_g1_i4
 CALM c92733_g1_i3
 CALM c92733_g1_i2
 CALM c108404_g1_i2
 CACNA1D c102695_g2_i6
 PKA c94301_g2_i1
 GNG13 c98518_g1_i1
 GNAS c102442_g1_i3
 GNB1 c68653_g1_i2
 GNAO1 c98545_g1_i1
 MAPK1 c103658_g1_i3
 MAPK1 c103658_g1_i2
 MAPK1 c107317_g1_i3
 MAPK1 c103658_g1_i1
 RYR2 c103458_g3_i2
 RYR2 c107689_g1_i1
 RYR2 c109001_g1_i2
 ADCY2 c74935_g1_i3
 ADCY2 c71860_g1_i2
 CALM c92733_g1_i6
 CALM c108404_g1_i1
 CALM c92733_g1_i1
 CALM c92733_g1_i4
 CALM c92733_g1_i3
 CALM c92733_g1_i2
 CALM c108404_g1_i2
 CACNA1D c102695_g2_i6
 PKA c94301_g2_i1
 GNG13 c98518_g1_i1
 GNAS c102442_g1_i3
 GNB1 c68653_g1_i2
 GNAO1 c98545_g1_i1



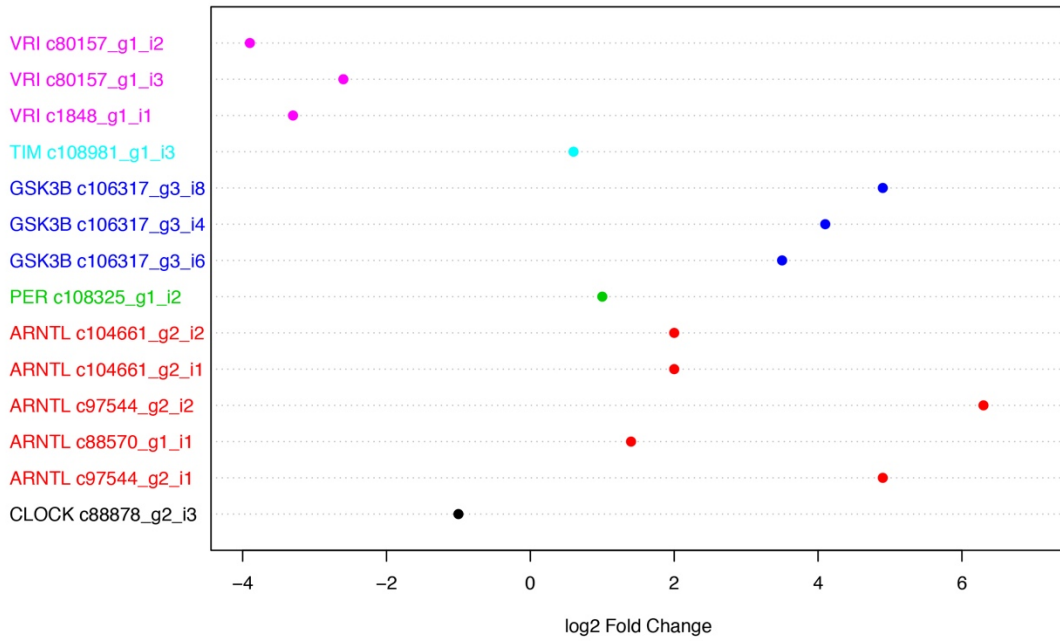
A

DEG: Circadian rhythm fly genes in *G. assimilis* male vs. female head



B

DEG: Circadian rhythm fly genes in *G. assimilis* testis vs. ovary



S1 Appendix

Gene name	Transcript ID			BLASTx ID - species		
	<i>G. assimilis</i>	<i>G. bimaculatus</i>	<i>G. firmus</i>	<i>G. assimilis</i>	<i>G. bimaculatus</i>	<i>G. firmus</i>
<i>sex-lethal</i>	c105826_g1_i2	c7093_g1_i1	c21247_g1_i1	SXL_CERCA - <i>Ceratitis capitata</i>	SXL_DROME - <i>Drosophila melanogaster</i>	SXL_MEGSC - <i>Megaselia scalaris</i>
	c105826_g1_i4	c7093_g1_i2	c21247_g1_i3	SXL_CERCA - <i>C. capitata</i>	SXL_DROME - <i>D. melanogaster</i>	SXL_DROME - <i>D. melanogaster</i>
	c105826_g1_i5	c7093_g1_i3		SXL_CERCA - <i>C. capitata</i>	SXL_DROME - <i>D. melanogaster</i>	
		c7093_g1_i4			SXL_DROME - <i>D. melanogaster</i>	
<i>transformer</i>	c155351_g1_i1	c9564_g1_i1	c33796_g1_i2	TRA2_DROVI - <i>Drosophila virilis</i>	TRA2B_RAT - <i>Rattus</i> sp.	TRA2B_RAT - <i>Rattus</i> sp.
	c100340_g1_i1	c9564_g1_i2	c33796_g1_i5	TRA2B_RAT - <i>Rattus</i> sp.	TRA2B_RAT - <i>Rattus</i> sp.	TRA2B_RAT - <i>Rattus</i> sp.
	c100340_g1_i3	c9564_g1_i3	c33796_g1_i6	TRA2B_RAT - <i>Rattus</i> sp.	TRA2B_RAT - <i>Rattus</i> sp.	TRA2B_RAT - <i>Rattus</i> sp.
	c100340_g1_i4	c9564_g1_i4	c33796_g1_i7	TRA2B_RAT - <i>Rattus</i> sp.	TRA2B_RAT - <i>Rattus</i> sp.	TRA2B_RAT - <i>Rattus</i> sp.
	c100340_g1_i5	c9564_g1_i5		TRA2B_RAT - <i>Rattus</i> sp.	TRA2B_RAT - <i>Rattus</i> sp.	
		c9564_g1_i6			TRA2B_RAT - <i>Rattus</i> sp.	
<i>doublesex</i>	c97760_g1_i2		c41697_g1_i1	DSX_DROME - <i>D. melanogaster</i>		DSX_DROME - <i>D. melanogaster</i>
	c97760_g1_i5			DSX_DROME - <i>D. melanogaster</i>		
<i>fruitless</i>	c104957_g1_i1	c28219_g1_i1	c46699_g1_i1	FRU_DROME - <i>D. melanogaster</i>	FRU_DROME - <i>D. melanogaster</i>	FRU_DROME - <i>D. melanogaster</i>
	c104957_g2_i1	c305_g1_i1	c49032_g1_i1	FRU_DROME - <i>D. melanogaster</i>	FRU_DROME - <i>D. melanogaster</i>	FRU_DROME - <i>D. melanogaster</i>
	c104957_g2_i2	c305_g1_i2		FRU_DROME - <i>D. melanogaster</i>	FRU_DROME - <i>D. melanogaster</i>	
	c104957_g2_i3			FRU_DROME - <i>D. melanogaster</i>		
	c104957_g2_i4			FRU_DROME - <i>D. melanogaster</i>		
	c104957_g2_i6			FRU_DROME - <i>D. melanogaster</i>		
	c104957_g2_i7			FRU_DROME - <i>D. melanogaster</i>		
	c104957_g2_i8			FRU_DROME - <i>D. melanogaster</i>		
	c104957_g2_i9			FRU_DROME - <i>D. melanogaster</i>		
	c104957_g2_i10			FRU_DROME - <i>D. melanogaster</i>		

c104957_g2_i12

c104957_g2_i13

c21731_g1_i1

c21731_g1_i2

FRU_DROME - *D. melanogaster*

FRU_DROME - *D. melanogaster*

FRU_DROME - *D. melanogaster*

FRU_DROME - *D. melanogaster*

S2 Appendix

KEGG pathway	KO number	KEGG gene name	<i>L. migratoria</i>	<i>G. assimilis</i>	<i>G. bimaculatus</i>	<i>G. firmus</i>	
Insect hormone biosynthesis	K00128	ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	+	+	+	+	
	K01063	JHE; juvenile-hormone esterase [EC:3.1.1.59]	+	+	-	+	
	K10718	JHAMT; juvenile hormone-III synthase [EC:2.1.1.325]	+	+	+	-	
	K10719	JHEH; juvenile hormone epoxide hydrolase	+	+	+	+	
	K10720	PHM; CYP306A1; ecdysteroid 25-hydroxylase	+	+	-	-	
	K10721	DIB; CYP302A1; ecdysteroid 22-hydroxylase	+	+	+	+	
	K10722	SAD, CYP315A1; ecdysteroid 2-hydroxylase	+	+	-	+	
	K10723	SHD, CYP314A1; ecdysone 20-monooxygenase [EC:1.14.99.22]	+	+	-	-	
	K14937	CYP15A1_C1; methyl farnesoate epoxidase / farnesoate epoxidase [EC:1.14.13.202 1.14.13.203]	+	+	-	+	
	K14938	NVD, DAF36; cholesterol 7-desaturase [EC:1.14.19.21]	+	+	-	-	
	K14939	spo, spok, CYP307A; cytochrome P450 family 307 subfamily A	+	+	-	-	
	K14985	CYP18A1; 26-hydroxylase [EC:1.14.-.-]	+	+	-	-	
	K15890	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]	+	+	+	+	
	K21013	FPPP; farnesyl diphosphate phosphatase [EC:3.1.3.-]		+	-	-	
	Circadian rhythm	K02223	CLOCK, KAT13D; circadian locomoter output cycles kaput protein [EC:2.3.1.48]	+	+	-	-
		K02295	CRY; cryptochrome	+	+	+	+
		K02296	ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1	+	+	-	+
		K02633	PER; period circadian protein	+	+	-	+
		K03094	SKP1, CBF3D; S-phase kinase-associated protein 1	+	+	+	+
K03347		CUL1, CDC53; cullin 1	+	+	+	+	
K03362		FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/11	+	+	+	+	
K03729		BHLHB2, DEC1; class B basic helix-loop-helix protein 2		-	-	+	
K03868		RBX1, ROC1; RING-box protein 1	+	+	+	+	
K07198		PRKAA, AMPK; 5'-AMP-activated protein kinase, catalytic alpha subunit [EC:2.7.11.11]	+	+	+	+	
K07199	PRKAB; 5'-AMP-activated protein kinase, regulatory beta subunit	+	+	+	+		
K07200	PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit	+	+	+	+		

Circadian entrainment	K08532	NR1F1, RORA; RAR-related orphan receptor alpha	+	-	-	-
	K08960	CSNK1E; casein kinase 1, epsilon [EC:2.7.11.1]	+	+	+	+
	K10269	FBXL3; F-box and leucine-rich repeat protein 3	+	-	-	+
	K02183	CALM; calmodulin	+	+	+	+
	K02633	PER; period circadian protein	+	+	-	+
	K02677	PRKCA; classical protein kinase C alpha type [EC:2.7.11.13]	+	+	+	+
	K04345	PKA; protein kinase A [EC:2.7.11.11]	+	+	+	-
	K04371	MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]	+	+	+	+
	K04379	FOS; proto-oncogene protein c-fos	-	-	-	+
	K04445	MSK1, RPS6KA5; ribosomal protein S6 kinase alpha-5 [EC:2.7.11.1]	+	+	-	+
	K04515	CAMK2; calcium/calmodulin-dependent protein kinase (CaM kinase) II [EC:2.7.11.17]	+	+	+	+
	K04534	GNAO, G-ALPHA-O; guanine nucleotide-binding protein G(o) subunit alpha	+	+	+	-
	K04536	GNB1; guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	+	+	+	+
	K04539	GNB5; guanine nucleotide-binding protein subunit beta-5	+	+	+	-
	K04542	GNG5; guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-5	-	-	-	+
	K04543	GNG7; guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-7	-	+	+	-
	K04546	GNG11; guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-11	-	-	-	+
	K04547	GNG13; guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-13	+	+	-	-
	K04630	GNAI; guanine nucleotide-binding protein G(i) subunit alpha	+	+	+	+
	K04632	GNAS; guanine nucleotide-binding protein G(s) subunit alpha	+	+	-	+
	K04634	GNAQ; guanine nucleotide-binding protein G(q) subunit alpha	+	+	+	+
	K04851	CACNA1D; voltage-dependent calcium channel L type alpha-1D	+	+	-	+
	K04854	CACNA1G; voltage-dependent calcium channel T type alpha-1G	+	+	-	-
	K04958	ITPR1; inositol 1,4,5-triphosphate receptor type 1	+	+	+	+
	K04961	RYR1; ryanodine receptor 1	+	-	-	+
	K04962	RYR2; ryanodine receptor 2	+	+	-	+
	K04963	RYR3; ryanodine receptor 3	+	-	-	+
	K05197	GRIA1; glutamate receptor 1	+	-	-	-
	K05199	GRIA3; glutamate receptor 3	+	-	-	-

K05200	GRIA4; glutamate receptor 4	+	-	-	-
K05208	GRIN1; glutamate receptor ionotropic, NMDA 1	+	+	+	-
K05210	GRIN2B; glutamate receptor ionotropic, NMDA 2B	+	+	-	-
K05858	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	+	+	+	+
K07376	PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	+	+	+	+
K07825	GNB3; guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-3	-	-	-	+
K07843	RASD1; RAS, dexamethasone-induced Ras-related protein 1	+	+	-	-
K08041	ADCY1; adenylyl cyclase 1 [EC:4.6.1.1]	+	+	-	-
K08042	ADCY2; adenylyl cyclase 2 [EC:4.6.1.1]	+	+	-	+
K08043	ADCY3; adenylyl cyclase 3 [EC:4.6.1.1]	+	+	-	-
K08044	ADCY4; adenylyl cyclase 4 [EC:4.6.1.1]	+	-	-	-
K08045	ADCY5; adenylyl cyclase 5 [EC:4.6.1.1]	+	+	-	+
K08047	ADCY7; adenylyl cyclase 7 [EC:4.6.1.1]	+	-	-	-
K08048	ADCY8; adenylyl cyclase 8 [EC:4.6.1.1]	+	+	-	-
K08049	ADCY9; adenylyl cyclase 9 [EC:4.6.1.1]	+	+	+	+
K11265	ADCY10; adenylyl cyclase 10 [EC:4.6.1.1]	+	+	-	-
K12318	GUCY1A; guanylate cyclase soluble subunit alpha [EC:4.6.1.2]	+	+	-	-
K12319	GUCY1B; guanylate cyclase soluble subunit beta [EC:4.6.1.2]	+	+	-	-
K13240	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	+	+	+	-
K16513	NOS1AP, CAPON; carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase	+	+	-	-
Circadian rhythm fly					
K02223	CLOCK, KAT13D; circadian locomoter output cycles kaput protein [EC:2.3.1.48]	+	+	-	-
K02296	ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1	+	+	-	+
K02633	PER; period circadian protein	+	+	-	+
K03083	GSK3B; glycogen synthase kinase 3 beta [EC:2.7.11.26]	+	+	+	+
K08960	CSNK1E; casein kinase 1, epsilon [EC:2.7.11.1]	+	+	+	+
K09057	HLF; hepatic leukemia factor	+	+	-	+
K12074	TIM; timeless	+	+	+	+
K12114	VRI; vrilie	+	+	-	+

S3 Appendix

KEGG pathway	KO number	KEGG gene name	<i>L. migratoria</i> ID	<i>G. assimilis</i> ID	<i>G. bimaculatus</i> ID	<i>G. firmus</i> ID
Insect hormone biosynthesis*	K00128	ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP010008985.1	c103318_g1_i1	c10247_g1_i1	c29739_g1_i1
		ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP010008987.1	c104195_g4_i1	c12272_g1_i1	c31891_g1_i1
		ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP010023066.1	c104386_g1_i10	c12277_g1_i1	c32667_g1_i1
		ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP010165191.1	c104386_g1_i2	c12277_g1_i2	c32667_g1_i4
		ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP010165193.1	c104386_g1_i5	c12277_g1_i3	c35265_g1_i1
		ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP010314925.1	c107419_g1_i3	c11236_g1_i2	c35518_g1_i3
		ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP010314930.1	c107419_g1_i5		
		ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP010546925.1	c107419_g1_i6		
		ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP010574230.1	c90017_g1_i3		
		ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP010574231.1	c90017_g1_i4		
		ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP010640791.1			
		ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP010820493.1			
		ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP010820499.1			
		ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP011139379.1			
	ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP011382677.1				
	ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	AVCP011394878.1				
	K01063	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010020421.1	c102917_g1_i1		c34612_g1_i1
		JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010020483.1			
		JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010088941.1			
		JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010133559.1			
		JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010153472.1			
		JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010157482.1			
		JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010202440.1			
		JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010218183.1			
JHE; juvenile-hormone esterase [EC:3.1.1.59]		AVCP010270924.1				
JHE; juvenile-hormone esterase [EC:3.1.1.59]		AVCP010314599.1				
JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010409093.1					

	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010444439.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010473895.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010473937.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010526077.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010526085.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010526099.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010696607.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010725541.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010762774.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010762792.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010762798.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010762803.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010800742.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010904551.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP010978539.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP011040284.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP011040290.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP011059383.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP011070276.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP011070285.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP011126241.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP011127689.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP011250171.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP011314829.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP011352197.1		
	JHE; juvenile-hormone esterase [EC:3.1.1.59]	AVCP011390336.1		
K10718	JHMT; juvenile hormone-III synthase [EC:2.1.1.325]	AVCP011124322.1	c100034_g1_i1	c11236_g1_i4
	JHMT; juvenile hormone-III synthase [EC:2.1.1.325]		c100034_g1_i3	
	JHMT; juvenile hormone-III synthase [EC:2.1.1.325]		c100034_g1_i4	

	JHAMT; juvenile hormone-III synthase [EC:2.1.1.325]		c105887_g1_i1		
K10719	JHEH; juvenile hormone epoxide hydrolase	AVCP010626649.1	c109061_g1_i3	c2307_g1_i1	c30656_g1_i2
	JHEH; juvenile hormone epoxide hydrolase	AVCP010895675.1	c63479_g1_i1		c32216_g1_i1
	JHEH; juvenile hormone epoxide hydrolase	AVCP010895684.1	c63479_g1_i2		c32216_g1_i2
	JHEH; juvenile hormone epoxide hydrolase		c91728_g1_i3		c32216_g1_i3
	JHEH; juvenile hormone epoxide hydrolase		c96010_g1_i11		c33813_g2_i1
	JHEH; juvenile hormone epoxide hydrolase		c96010_g1_i12		
	JHEH; juvenile hormone epoxide hydrolase		c96010_g1_i6		
	JHEH; juvenile hormone epoxide hydrolase		c96010_g1_i9		
	JHEH; juvenile hormone epoxide hydrolase		c96730_g1_i2		
	JHEH; juvenile hormone epoxide hydrolase		c96730_g1_i3		
	JHEH; juvenile hormone epoxide hydrolase		c96730_g1_i4		
K10720	PHM; CYP306A1; ecdysteroid 25-hydroxylase	AVCP010205869.1	c100334_g1_i2		
	PHM; CYP306A1; ecdysteroid 25-hydroxylase	AVCP010480474.1	c100334_g1_i3		
	PHM; CYP306A1; ecdysteroid 25-hydroxylase	AVCP010938175.1	c84397_g1_i5		
	PHM; CYP306A1; ecdysteroid 25-hydroxylase	AVCP010938180.1			
	PHM; CYP306A1; ecdysteroid 25-hydroxylase	AVCP011168642.1			
K10721	DIB; CYP302A1; ecdysteroid 22-hydroxylase	AVCP010426404.1	c107624_g1_i1	c7130_g1_i1	c52890_g1_i1
	DIB; CYP302A1; ecdysteroid 22-hydroxylase	AVCP010426408.1	c107624_g1_i5	c7130_g1_i2	c12565_g1_i2
	DIB; CYP302A1; ecdysteroid 22-hydroxylase		c107624_g1_i7	c7130_g1_i3	
	DIB; CYP302A1; ecdysteroid 22-hydroxylase		c107624_g1_i8		
K10722	SAD, CYP315A1; ecdysteroid 2-hydroxylase	AVCP010507933.1	c86091_g1_i1		c17196_g1_i1
	SAD, CYP315A1; ecdysteroid 2-hydroxylase		c86091_g1_i2		
K10723	SHD, CYP314A1; ecdysone 20-monooxygenase [EC:1.14.99.22]	AVCP010995239.1	c105588_g1_i3		
	SHD, CYP314A1; ecdysone 20-monooxygenase [EC:1.14.99.22]	AVCP010995247.1	c105588_g1_i6		
	SHD, CYP314A1; ecdysone 20-monooxygenase [EC:1.14.99.22]		c85061_g1_i2		
	SHD, CYP314A1; ecdysone 20-monooxygenase [EC:1.14.99.22]		c85061_g1_i3		
K14937	CYP15A1_C1; methyl farnesoate epoxidase / farnesoate epoxidase [EC:1.14.13.202 1.14.13.203]	AVCP010011325.1	c107918_g2_i3		c32572_g1_i4
	CYP15A1_C1; methyl farnesoate epoxidase / farnesoate epoxidase [EC:1.14.13.202 1.14.13.203]	AVCP010063589.1	c94971_g1_i4		

	CYP15A1_C1; methyl farnesoate epoxidase / farnesoate epoxidase [EC:1.14.13.202 1.14.13.203]	AVCP010190331.1			
	CYP15A1_C1; methyl farnesoate epoxidase / farnesoate epoxidase [EC:1.14.13.202 1.14.13.203]	AVCP010265835.1			
	CYP15A1_C1; methyl farnesoate epoxidase / farnesoate epoxidase [EC:1.14.13.202 1.14.13.203]	AVCP010307808.1			
	CYP15A1_C1; methyl farnesoate epoxidase / farnesoate epoxidase [EC:1.14.13.202 1.14.13.203]	AVCP010477990.1			
K14938	NVD, DAF36; cholesterol 7-desaturase [EC:1.14.19.21]	AVCP010451115.1	c51442_g1_i1		
	NVD, DAF36; cholesterol 7-desaturase [EC:1.14.19.21]	AVCP010500987.1	c87477_g1_i1		
	NVD, DAF36; cholesterol 7-desaturase [EC:1.14.19.21]	AVCP010522058.1	c87477_g1_i3		
	NVD, DAF36; cholesterol 7-desaturase [EC:1.14.19.21]	AVCP010821147.1	c87477_g1_i4		
K14939	spo, spok, CYP307A; cytochrome P450 family 307 subfamily A	AVCP010105460.1	c148959_g1_i1		
	spo, spok, CYP307A; cytochrome P450 family 307 subfamily A	AVCP010105462.1	c151578_g1_i1		
	spo, spok, CYP307A; cytochrome P450 family 307 subfamily A	AVCP010416115.1	c16438_g1_i1		
	spo, spok, CYP307A; cytochrome P450 family 307 subfamily A	AVCP010416118.1			
	spo, spok, CYP307A; cytochrome P450 family 307 subfamily A	AVCP010416120.1			
K14985	CYP18A1; 26-hydroxylase [EC:1.14.-.-]	AVCP010265822.1	c135509_g1_i1		
	CYP18A1; 26-hydroxylase [EC:1.14.-.-]	AVCP010938189.1	c64866_g1_i1		
K15890	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]	AVCP010270499.1	c103730_g1_i2	c7362_g1_i1	c14342_g1_i1
	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]	AVCP010270524.1	c103730_g2_i1		c33941_g1_i2
	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]	AVCP010465314.1	c104977_g2_i4		c33941_g2_i1
	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]	AVCP010541381.1	c104977_g2_i6		c35191_g1_i1
	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]	AVCP010651088.1	c104977_g3_i1		
	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]	AVCP010804682.1	c104977_g3_i2		
	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]	AVCP011059295.1	c104977_g3_i4		
	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]		c62560_g1_i1		
	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]		c90176_g1_i2		
	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]		c97494_g2_i1		
K21013	FPPP; farnesyl diphosphate phosphatase [EC:3.1.3.-]		c107901_g2_i1		
	FPPP; farnesyl diphosphate phosphatase [EC:3.1.3.-]		c107901_g2_i2		
	FPPP; farnesyl diphosphate phosphatase [EC:3.1.3.-]		c107901_g2_i3		
	FPPP; farnesyl diphosphate phosphatase [EC:3.1.3.-]		c107901_g2_i4		
Circadian rhythm	K02223	CLOCK, KAT13D; circadian locomoter output cycles kaput protein [EC:2.3.1.48]	AVCP010552063.1	c88878_g2_i2	

	CLOCK, KAT13D; circadian locomoter output cycles kaput protein [EC:2.3.1.48]		c88878_g2_i3		
K02295	CRY; cryptochrome	AVCP010106028.1	c105814_g1_i3	c12848_g1_i1	c28906_g1_i1
	CRY; cryptochrome	AVCP010303865.1	c105814_g1_i5	c1294_g1_i1	
	CRY; cryptochrome	AVCP010535649.1	c105814_g1_i7		
	CRY; cryptochrome	AVCP010808824.1	c107392_g1_i5		
	CRY; cryptochrome	AVCP010944781.1	c97680_g1_i1		
	CRY; cryptochrome		c97680_g1_i2		
	CRY; cryptochrome		c97680_g1_i3		
	CRY; cryptochrome		c97680_g1_i6		
K02296	ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1	AVCP010331992.1	c104661_g2_i1		c24663_g1_i2
	ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 2	AVCP010786666.1	c104661_g2_i2		
	ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 3		c88570_g1_i1		
	ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 4		c97544_g2_i1		
	ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 5		c97544_g2_i2		
K02633	PER; period circadian protein	AVCP010544742.1	c108325_g1_i2		c11358_g1_i1
	PER; period circadian protein	AVCP010544746.1			c11358_g1_i2
K03094	SKP1, CBF3D; S-phase kinase-associated protein 1	AVCP011119043.1	c146290_g1_i1	c888_g1_i1	c2009_g1_i1
	SKP1, CBF3D; S-phase kinase-associated protein 2		c40894_g1_i1		
	SKP1, CBF3D; S-phase kinase-associated protein 3		c93326_g2_i1		
	SKP1, CBF3D; S-phase kinase-associated protein 4		c93326_g2_i2		
	SKP1, CBF3D; S-phase kinase-associated protein 5		c93326_g2_i3		
	SKP1, CBF3D; S-phase kinase-associated protein 6		c93326_g2_i4		
	SKP1, CBF3D; S-phase kinase-associated protein 7		c93326_g2_i6		
	SKP1, CBF3D; S-phase kinase-associated protein 8		c93326_g2_i7		
	SKP1, CBF3D; S-phase kinase-associated protein 9		c93326_g2_i8		
K03347	CUL1, CDC53; cullin 1	AVCP010000002.1	c100300_g2_i1	c12042_g1_i1	c30805_g1_i1
	CUL1, CDC53; cullin 1	AVCP010348421.1	c100300_g2_i2	c12042_g1_i2	c30805_g1_i3
	CUL1, CDC53; cullin 1	AVCP010428821.1	c100300_g2_i3	c12593_g1_i2	
	CUL1, CDC53; cullin 2	AVCP010517399.1	c100300_g2_i4		
	CUL1, CDC53; cullin 3	AVCP011051027.1	c105966_g1_i1		

	CUL1, CDC53; cullin 4		c105966_g1_i2		
K03362	FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/11	AVCP010002621.1	c100367_g1_i10	c10291_g1_i1	c32567_g1_i1
	FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/11		c100367_g1_i11	c10291_g1_i2	c32567_g1_i2
	FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/11		c100367_g1_i14	c10291_g1_i3	c43416_g1_i1
	FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/12		c100367_g1_i15		
	FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/13		c100367_g1_i5		
	FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/14		c100367_g1_i6		
	FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/15		c100367_g1_i8		
	FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/16		c41437_g1_i1		
K03729	BHLHB2, DEC1; class B basic helix-loop-helix protein 2				c4768_g1_i1
K03868	RBX1, ROC1; RING-box protein 1	AVCP010094236.1	c104236_g2_i3	c10592_g1_i1	c18550_g1_i1
	RBX1, ROC1; RING-box protein 1	AVCP010658826.1	c104236_g2_i5	c10592_g1_i2	c25821_g1_i2
	RBX1, ROC1; RING-box protein 2	AVCP010658829.1	c104236_g2_i7		
K07198	PRKAA, AMPK; 5'-AMP-activated protein kinase, catalytic alpha subunit [EC:2.7.11.11]	AVCP011071697.1	c93013_g1_i6	c9412_g1_i1	c34103_g1_i1
	PRKAA, AMPK; 5'-AMP-activated protein kinase, catalytic alpha subunit [EC:2.7.11.11]	AVCP011297814.1	c93013_g1_i7		c34103_g1_i3
	PRKAA, AMPK; 5'-AMP-activated protein kinase, catalytic alpha subunit [EC:2.7.11.11]				c34103_g1_i4
K07199	PRKAB; 5'-AMP-activated protein kinase, regulatory beta subunit	AVCP010021664.1	c103022_g2_i2	c11937_g1_i2	c19146_g1_i2
	PRKAB; 5'-AMP-activated protein kinase, regulatory beta subunit	AVCP010729344.1			
K07200	PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit	AVCP010378400.1	c87222_g1_i1	c15397_g1_i1	c35125_g1_i2
	PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit	AVCP011034505.1	c99381_g1_i1	c3207_g1_i1	c35125_g1_i3
	PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit	AVCP011034507.1	c99381_g1_i2		
	PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit	AVCP011034510.1	c99381_g1_i4		
	PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit	AVCP011034511.1	c99381_g1_i5		
	PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit		c99381_g1_i6		
	PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit		c99381_g1_i7		
K08532	NR1F1, RORA; RAR-related orphan receptor alpha	AVCP010057892.1			
K08960	CSNK1E; casein kinase 1, epsilon [EC:2.7.11.1]		c105680_g1_i2	c10344_g1_i1	c27216_g1_i1
	CSNK1E; casein kinase 1, epsilon [EC:2.7.11.1]	AVCP010057707.1		c10344_g1_i2	c27216_g1_i4
	CSNK1E; casein kinase 1, epsilon [EC:2.7.11.1]	AVCP011068991.1			c27216_g1_i5

		CSNK1E; casein kinase 1, epsilon [EC:2.7.11.1]	AVCP011068996.1			
		CSNK1E; casein kinase 1, epsilon [EC:2.7.11.1]	AVCP011069001.1			
		CSNK1E; casein kinase 1, epsilon [EC:2.7.11.1]	AVCP011245865.1			
	K10269	FBXL3; F-box and leucine-rich repeat protein 3	AVCP010558384.1			c16691_g1_i1
		FBXL3; F-box and leucine-rich repeat protein 3				c16691_g1_i2
		FBXL3; F-box and leucine-rich repeat protein 3				c4279_g1_i1
Circadian entrainment	K02183	CALM; calmodulin	AVCP010364210.1	c103247_g2_i1	c1319_g1_i1	c12382_g1_i1
		CALM; calmodulin	AVCP010507976.1	c103247_g2_i3	c3317_g1_i1	c30265_g1_i2
		CALM; calmodulin	AVCP010507991.1	c108404_g1_i1		c31502_g1_i1
		CALM; calmodulin	AVCP010633788.1	c108404_g1_i2		c31502_g1_i2
		CALM; calmodulin	AVCP010837610.1	c108404_g1_i3		c31746_g1_i1
		CALM; calmodulin	AVCP010972300.1	c156622_g1_i1		c31746_g1_i2
		CALM; calmodulin		c92733_g1_i1		
		CALM; calmodulin		c92733_g1_i2		
		CALM; calmodulin		c92733_g1_i3		
		CALM; calmodulin		c92733_g1_i4		
		CALM; calmodulin		c92733_g1_i6		
	K02633	PER; period circadian protein	AVCP010544742.1	c108325_g1_i2		c11358_g1_i1
		PER; period circadian protein	AVCP010544746.1			c11358_g1_i2
	K02677	PRKCA; classical protein kinase C alpha type [EC:2.7.11.13]	AVCP010074224.1	c106522_g1_i7	c4252_g1_i1	
		PRKCA; classical protein kinase C alpha type [EC:2.7.11.13]	AVCP010074231.1			
		PRKCA; classical protein kinase C alpha type [EC:2.7.11.13]	AVCP010373753.1			
	K04345	PKA; protein kinase A [EC:2.7.11.11]	AVCP010161759.1	c105127_g3_i3	c10847_g1_i1	c30380_g1_i2
		PKA; protein kinase A [EC:2.7.11.11]	AVCP010288143.1	c94301_g2_i1		c33559_g1_i1
		PKA; protein kinase A [EC:2.7.11.11]	AVCP010608875.1	c97605_g1_i1		c33559_g1_i2
		PKA; protein kinase A [EC:2.7.11.11]	AVCP010774621.1	c97605_g1_i2		c33559_g1_i3
	K04371	MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]	AVCP010005810.1	c101078_g1_i1	c2133_g1_i1	c15700_g1_i1
		MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]	AVCP010233258.1	c101078_g1_i2	c2133_g1_i2	
		MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]		c101078_g1_i3		

	MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]				c101078_g1_i4
	MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]				c101078_g1_i5
	MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]				c101078_g1_i6
	MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]				c101078_g1_i7
K04379	FOS; proto-oncogene protein c-fos				c15858_g1_i1
K04445	MSK1, RPS6KA5; ribosomal protein S6 kinase alpha-5 [EC:2.7.11.1]	AVCP011012268.1	c105901_g2_i1		c25231_g1_i1
	MSK1, RPS6KA5; ribosomal protein S6 kinase alpha-5 [EC:2.7.11.1]	AVCP011012295.1	c105901_g2_i2		
	MSK1, RPS6KA5; ribosomal protein S6 kinase alpha-5 [EC:2.7.11.1]	AVCP011121210.1	c105901_g2_i3		
K04515	CAMK2; calcium/calmodulin-dependent protein kinase (CaM kinase) II [EC:2.7.11.17]	AVCP010974924.1	c105422_g2_i2	c19483_g1_i1	c29945_g1_i1
	CAMK2; calcium/calmodulin-dependent protein kinase (CaM kinase) II [EC:2.7.11.17]	AVCP010974925.1	c105422_g2_i6		c29945_g1_i2
	CAMK2; calcium/calmodulin-dependent protein kinase (CaM kinase) II [EC:2.7.11.17]				c29945_g1_i3
	CAMK2; calcium/calmodulin-dependent protein kinase (CaM kinase) II [EC:2.7.11.17]				c29945_g1_i5
K04534	GNAO, G-ALPHA-O; guanine nucleotide-binding protein G(o) subunit alpha	AVCP010160264.1	c96645_g1_i1	c1962_g1_i1	
	GNAO, G-ALPHA-O; guanine nucleotide-binding protein G(o) subunit alpha	AVCP010160278.1			
K04536	GNB1; guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	AVCP010126887.1	c68653_g1_i2	c12706_g1_i2	c10967_g1_i1
	GNB1; guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2	AVCP010126891.1			
	GNB1; guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-3	AVCP011266109.1			
	GNB1; guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-4	AVCP011321127.1			
K04539	GNB5; guanine nucleotide-binding protein subunit beta-5	AVCP010152993.1	c106262_g1_i2	c7378_g1_i1	
	GNB5; guanine nucleotide-binding protein subunit beta-6	AVCP010152999.1			
	GNB5; guanine nucleotide-binding protein subunit beta-7	AVCP010387620.1			
K04542	GNG5; guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-5				c32379_g1_i2
K04543	GNG7; guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-7		c89856_g1_i1	c4749_g1_i1	
K04546	GNG11; guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-11				c25246_g1_i1
K04547	GNG13; guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-13	AVCP010180265.1	c98518_g1_i1		
K04630	GNAI; guanine nucleotide-binding protein G(i) subunit alpha	AVCP010003311.1	c90417_g2_i1	c9381_g1_i1	c32989_g1_i1
	GNAI; guanine nucleotide-binding protein G(i) subunit alpha	AVCP010003314.1			
	GNAI; guanine nucleotide-binding protein G(i) subunit alpha	AVCP010003316.1			
	GNAI; guanine nucleotide-binding protein G(i) subunit alpha	AVCP010160277.1			

	GNAI; guanine nucleotide-binding protein G(i) subunit alpha	AVCP011289419.1			
	GNAI; guanine nucleotide-binding protein G(i) subunit alpha	AVCP011319480.1			
K04632	GNAS; guanine nucleotide-binding protein G(s) subunit alpha	AVCP010581788.1	c102442_g1_i3		c35117_g1_i1
	GNAS; guanine nucleotide-binding protein G(s) subunit alpha	AVCP010581791.1			c35117_g1_i4
	GNAS; guanine nucleotide-binding protein G(s) subunit alpha	AVCP010581796.1			c35117_g1_i5
K04634	GNAO, G-ALPHA-O; guanine nucleotide-binding protein G(o) subunit alpha	AVCP010350490.1	c108928_g1_i2	c12047_g1_i2	c34975_g1_i2
	GNAO, G-ALPHA-O; guanine nucleotide-binding protein G(o) subunit alpha	AVCP010350495.1		c12047_g1_i3	
	GNAO, G-ALPHA-O; guanine nucleotide-binding protein G(o) subunit alpha	AVCP011093443.1			
K04851	CACNA1D; voltage-dependent calcium channel L type alpha-1D	AVCP010457135.1	c102695_g2_i16		c27592_g1_i1
	CACNA1D; voltage-dependent calcium channel L type alpha-1D	AVCP010457164.1	c102695_g2_i17		c42376_g1_i1
	CACNA1D; voltage-dependent calcium channel L type alpha-1D	AVCP010457168.1	c102695_g2_i18		
	CACNA1D; voltage-dependent calcium channel L type alpha-1D	AVCP010806846.1	c102695_g2_i1		
	CACNA1D; voltage-dependent calcium channel L type alpha-1D		c102695_g2_i2		
	CACNA1D; voltage-dependent calcium channel L type alpha-1D		c102695_g2_i6		
K04854	CACNA1G; voltage-dependent calcium channel T type alpha-1G	AVCP010144274.1	c111276_g1_i1		
	CACNA1G; voltage-dependent calcium channel T type alpha-1G	AVCP010473551.1	c131265_g1_i1		
	CACNA1G; voltage-dependent calcium channel T type alpha-1G	AVCP010643498.1	c61724_g1_i1		
	CACNA1G; voltage-dependent calcium channel T type alpha-1G	AVCP010643524.1	c89689_g1_i2		
	CACNA1G; voltage-dependent calcium channel T type alpha-1G	AVCP010709941.1	c89689_g1_i3		
	CACNA1G; voltage-dependent calcium channel T type alpha-1G	AVCP010831543.1			
	CACNA1G; voltage-dependent calcium channel T type alpha-1G	AVCP011139997.1			
	CACNA1G; voltage-dependent calcium channel T type alpha-1G	AVCP011140001.1			
	CACNA1G; voltage-dependent calcium channel T type alpha-1G	AVCP011185454.1			
	CACNA1G; voltage-dependent calcium channel T type alpha-1G	AVCP011278195.1			
K04958	ITPR1; inositol 1,4,5-triphosphate receptor type 1	AVCP010294788.1	c107526_g6_i2	c6547_g1_i1	c40300_g1_i1
	ITPR1; inositol 1,4,5-triphosphate receptor type 1	AVCP010875595.1	c107526_g6_i5		c46837_g1_i1
	ITPR1; inositol 1,4,5-triphosphate receptor type 2	AVCP010875603.1	c107526_g6_i6		
K04961	RYR1; ryanodine receptor 1	AVCP010459525.1			c35732_g2_i12

	RYR1; ryanodine receptor 1	AVCP010733099.1	c35732_g2_i14
	RYR1; ryanodine receptor 1	AVCP010935611.1	c35732_g2_i3
	RYR1; ryanodine receptor 1		c35732_g2_i7
	RYR1; ryanodine receptor 1		c35732_g2_i8
K04962	RYR2; ryanodine receptor 2	AVCP010037553.1	c103458_g3_i2
	RYR2; ryanodine receptor 2	AVCP010037557.1	c107689_g1_i1
	RYR2; ryanodine receptor 2	AVCP010144978.1	c109001_g1_i2
	RYR2; ryanodine receptor 3	AVCP010225348.1	
	RYR2; ryanodine receptor 4	AVCP010248248.1	
	RYR2; ryanodine receptor 5	AVCP010260819.1	
	RYR2; ryanodine receptor 6	AVCP010304485.1	
	RYR2; ryanodine receptor 7	AVCP010354263.1	
	RYR2; ryanodine receptor 8	AVCP010368708.1	
	RYR2; ryanodine receptor 9	AVCP010459530.1	
	RYR2; ryanodine receptor 10	AVCP010547494.1	
	RYR2; ryanodine receptor 11	AVCP010630966.1	
	RYR2; ryanodine receptor 12	AVCP010630967.1	
	RYR2; ryanodine receptor 13	AVCP010630970.1	
	RYR2; ryanodine receptor 14	AVCP010684369.1	
	RYR2; ryanodine receptor 15	AVCP010733102.1	
	RYR2; ryanodine receptor 16	AVCP010733105.1	
	RYR2; ryanodine receptor 17	AVCP010935609.1	
	RYR2; ryanodine receptor 18	AVCP010945010.1	
	RYR2; ryanodine receptor 19	AVCP011013694.1	
	RYR2; ryanodine receptor 20	AVCP011334508.1	
	RYR2; ryanodine receptor 21	AVCP011335685.1	
	RYR2; ryanodine receptor 22	AVCP011377094.1	
	RYR2; ryanodine receptor 23	AVCP011382369.1	
RYR2; ryanodine receptor 24	AVCP011384383.1		

K04963	RYR3; ryanodine receptor 3	AVCP011361824.1		c35732_g2_i10
	RYR3; ryanodine receptor 3			c35732_g2_i18
	RYR3; ryanodine receptor 3			c35732_g2_i5
	RYR3; ryanodine receptor 3			c35732_g2_i9
K05197	GRIA1; glutamate receptor 1	AVCP010022368.1		
	GRIA1; glutamate receptor 2	AVCP010598232.1		
K05199	GRIA3; glutamate receptor 3	AVCP010028921.1		
K05200	GRIA4; glutamate receptor 4	AVCP010027284.1		
K05208	GRIN1; glutamate receptor ionotropic, NMDA 1	AVCP011146984.1	c98719_g1_i2	c25924_g1_i1
	GRIN1; glutamate receptor ionotropic, NMDA 2	AVCP011146993.1	c98719_g1_i3	
	GRIN1; glutamate receptor ionotropic, NMDA 3		c98719_g1_i5	
K05210	GRIN2B; glutamate receptor ionotropic, NMDA 2B	AVCP010261932.1	c135807_g1_i1	
	GRIN2B; glutamate receptor ionotropic, NMDA 2B	AVCP010261953.1	c61974_g2_i1	
	GRIN2B; glutamate receptor ionotropic, NMDA 2B	AVCP010261963.1	c75896_g1_i1	
K05858	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP010041366.1	c103398_g1_i3	c16277_g1_i1
	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP010068771.1	c106270_g2_i1	c21068_g1_i1
	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP010186040.1	c106270_g2_i2	
	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP010332785.1	c106270_g2_i3	
	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP010339678.1	c106270_g2_i4	
	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP010379466.1	c106270_g2_i5	
	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP010379471.1	c108574_g1_i1	
	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP010528347.1	c108574_g1_i3	
	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP010588974.1		
	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP010609446.1		
	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP010701310.1		
	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP010973861.1		
	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP011000111.1		
	PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP011109216.1		
PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	AVCP011109218.1			

K07376	PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	AVCP010020864.1	c103658_g1_i1	c13089_g1_i1	c32781_g1_i2
	PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	AVCP010020865.1	c103658_g1_i2		c33833_g1_i1
	PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	AVCP010025019.1	c103658_g1_i3		c34442_g1_i3
	PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	AVCP010025020.1	c105127_g3_i1		c43039_g1_i1
	PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	AVCP010025028.1	c107317_g1_i1		c8413_g1_i1
	PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	AVCP010100178.1	c107317_g1_i2		
	PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	AVCP010100179.1	c107317_g1_i3		
	PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	AVCP010100184.1	c107317_g1_i6		
	PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	AVCP010420287.1			
	PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	AVCP010885563.1			
	PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	AVCP010891327.1			
K07825	GNB3; guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-3				c35184_g2_i2
K07843	RASD1; RAS, dexamethasone-induced Ras-related protein 1	AVCP010155003.1	c131519_g1_i1		
	RASD1; RAS, dexamethasone-induced Ras-related protein 1	AVCP010442741.1			
K08041	ADCY1; adenylyate cyclase 1 [EC:4.6.1.1]	AVCP010338299.1	c87575_g1_i1		
	ADCY1; adenylyate cyclase 1 [EC:4.6.1.1]	AVCP010387153.1	c87575_g1_i3		
	ADCY1; adenylyate cyclase 1 [EC:4.6.1.1]	AVCP010387154.1	c87575_g1_i4		
	ADCY1; adenylyate cyclase 1 [EC:4.6.1.1]	AVCP010387168.1	c87575_g1_i5		
	ADCY1; adenylyate cyclase 1 [EC:4.6.1.1]	AVCP010485448.1	c87575_g1_i6		
	ADCY1; adenylyate cyclase 1 [EC:4.6.1.1]	AVCP010596783.1	c91549_g2_i1		
	ADCY1; adenylyate cyclase 1 [EC:4.6.1.1]	AVCP010711854.1	c91549_g2_i2		
	ADCY1; adenylyate cyclase 1 [EC:4.6.1.1]	AVCP010859012.1	c91549_g2_i3		
	ADCY1; adenylyate cyclase 1 [EC:4.6.1.1]	AVCP010930530.1	c91549_g2_i4		
	ADCY1; adenylyate cyclase 1 [EC:4.6.1.1]	AVCP010930535.1			
K08042	ADCY2; adenylyate cyclase 2 [EC:4.6.1.1]	AVCP010088636.1	c71860_g1_i1		c15919_g1_i2
	ADCY2; adenylyate cyclase 2 [EC:4.6.1.1]	AVCP010278700.1	c71860_g1_i2		
	ADCY2; adenylyate cyclase 2 [EC:4.6.1.1]	AVCP010649783.1	c71860_g1_i3		
	ADCY2; adenylyate cyclase 2 [EC:4.6.1.1]	AVCP010701253.1	c74935_g1_i3		
	ADCY2; adenylyate cyclase 2 [EC:4.6.1.1]	AVCP010796195.1			

	ADCY2; adenylate cyclase 2 [EC:4.6.1.1]	AVCP010902049.1		
	ADCY2; adenylate cyclase 2 [EC:4.6.1.1]	AVCP010902063.1		
	ADCY2; adenylate cyclase 2 [EC:4.6.1.1]	AVCP010959910.1		
	ADCY2; adenylate cyclase 2 [EC:4.6.1.1]	AVCP011141695.1		
K08043	ADCY3; adenylate cyclase 3 [EC:4.6.1.1]	AVCP010338290.1	c130825_g1_i1	
	ADCY3; adenylate cyclase 3 [EC:4.6.1.1]	AVCP010351156.1	c56321_g1_i2	
	ADCY3; adenylate cyclase 3 [EC:4.6.1.1]	AVCP010387161.1		
	ADCY3; adenylate cyclase 3 [EC:4.6.1.1]	AVCP010457753.1		
	ADCY3; adenylate cyclase 3 [EC:4.6.1.1]	AVCP010792940.1		
	ADCY3; adenylate cyclase 3 [EC:4.6.1.1]	AVCP010792956.1		
K08044	ADCY4; adenylate cyclase 4 [EC:4.6.1.1]	AVCP010859015.1		
K08045	ADCY5; adenylate cyclase 5 [EC:4.6.1.1]	AVCP010112970.1	c112889_g1_i1	c23703_g1_i1
	ADCY5; adenylate cyclase 5 [EC:4.6.1.1]	AVCP010207049.1	c59442_g1_i1	
	ADCY5; adenylate cyclase 5 [EC:4.6.1.1]	AVCP010312489.1		
	ADCY5; adenylate cyclase 5 [EC:4.6.1.1]	AVCP010406942.1		
	ADCY5; adenylate cyclase 5 [EC:4.6.1.1]	AVCP010607373.1		
	ADCY5; adenylate cyclase 5 [EC:4.6.1.1]	AVCP010665562.1		
	ADCY5; adenylate cyclase 5 [EC:4.6.1.1]	AVCP010795138.1		
	ADCY5; adenylate cyclase 5 [EC:4.6.1.1]	AVCP010859011.1		
K08047	ADCY7; adenylate cyclase 7 [EC:4.6.1.1]	AVCP010457749.1		
K08048	ADCY8; adenylate cyclase 8 [EC:4.6.1.1]	AVCP010686997.1	c102939_g1_i1	
	ADCY8; adenylate cyclase 8 [EC:4.6.1.1]	AVCP010859007.1	c83296_g1_i2	
	ADCY8; adenylate cyclase 8 [EC:4.6.1.1]	AVCP010859014.1		
	ADCY8; adenylate cyclase 8 [EC:4.6.1.1]	AVCP011339888.1		
K08049	ADCY9; adenylate cyclase 9 [EC:4.6.1.1]	AVCP010026864.1	c94429_g1_i1	c15074_g1_i1 c38331_g1_i1
	ADCY9; adenylate cyclase 9 [EC:4.6.1.1]	AVCP010026876.1	c94429_g1_i2	
	ADCY9; adenylate cyclase 9 [EC:4.6.1.1]		c94429_g1_i3	
	ADCY9; adenylate cyclase 9 [EC:4.6.1.1]		c94429_g1_i6	
K11265	ADCY10; adenylate cyclase 10 [EC:4.6.1.1]	AVCP010729349.1	c107118_g1_i6	

	ADCY10; adenylate cyclase 10 [EC:4.6.1.1]	AVCP011139129.1	c99025_g1_i1	
	ADCY10; adenylate cyclase 10 [EC:4.6.1.1]	AVCP011139136.1	c99025_g1_i2	
	ADCY10; adenylate cyclase 10 [EC:4.6.1.1]	AVCP011387609.1		
K12318	GUCY1A; guanylate cyclase soluble subunit alpha [EC:4.6.1.2]	AVCP010209819.1	c99019_g1_i2	
	GUCY1A; guanylate cyclase soluble subunit alpha [EC:4.6.1.2]	AVCP010546574.1		
	GUCY1A; guanylate cyclase soluble subunit alpha [EC:4.6.1.2]	AVCP010546592.1		
	GUCY1A; guanylate cyclase soluble subunit alpha [EC:4.6.1.2]	AVCP010546594.1		
	GUCY1A; guanylate cyclase soluble subunit alpha [EC:4.6.1.2]	AVCP010697511.1		
	GUCY1A; guanylate cyclase soluble subunit alpha [EC:4.6.1.2]	AVCP010697513.1		
	GUCY1A; guanylate cyclase soluble subunit alpha [EC:4.6.1.2]	AVCP010796523.1		
K12319	GUCY1B; guanylate cyclase soluble subunit beta [EC:4.6.1.2]	AVCP010065847.1	c97977_g1_i2	
	GUCY1B; guanylate cyclase soluble subunit beta [EC:4.6.1.2]	AVCP010297104.1		
	GUCY1B; guanylate cyclase soluble subunit beta [EC:4.6.1.2]	AVCP010571532.1		
	GUCY1B; guanylate cyclase soluble subunit beta [EC:4.6.1.2]	AVCP010698031.1		
	GUCY1B; guanylate cyclase soluble subunit beta [EC:4.6.1.2]	AVCP011128259.1		
	GUCY1B; guanylate cyclase soluble subunit beta [EC:4.6.1.2]	AVCP011156050.1		
	GUCY1B; guanylate cyclase soluble subunit beta [EC:4.6.1.2]	AVCP011375505.1		
K13240	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP010273556.1	c96059_g3_i1	c20543_g1_i1
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP010273557.1	c96059_g3_i2	c26909_g1_i1
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP010484058.1	c96059_g3_i3	
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP010530439.1	c96059_g3_i4	
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP010575208.1		
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP010653731.1		
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP010715751.1		
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP010756308.1		
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP010831357.1		
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP010831360.1		
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP010845927.1		
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP010903877.1		

	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP011013093.1				
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP011151039.1				
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP011336368.1				
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP011338136.1				
	NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	AVCP011393087.1				
Circadian rhythm-fly	K16513	NOS1AP, CAPON; carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase protein	AVCP010174918.1	c100237_g1_i2		
		NOS1AP, CAPON; carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase protein	AVCP010174951.1	c100237_g1_i4		
		NOS1AP, CAPON; carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase protein	AVCP010175016.1	c100237_g1_i6		
		NOS1AP, CAPON; carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase protein		c104796_g1_i3		
		NOS1AP, CAPON; carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase protein		c98844_g1_i1		
		NOS1AP, CAPON; carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase protein		c98844_g1_i2		
	K02223	CLOCK, KAT13D; circadian locomoter output cycles kaput protein [EC:2.3.1.48]	AVCP010552063.1	c88878_g2_i2		
		CLOCK, KAT13D; circadian locomoter output cycles kaput protein [EC:2.3.1.48]	AVCP010786666.1	c88878_g2_i3		
	K02296	ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1		c104661_g2_i1		
		ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 2		c104661_g2_i2		
		ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 3		c88570_g1_i1		
		ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 4		c97544_g2_i1		
		ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 5		c97544_g2_i2		
	K02633	PER; period circadian protein	AVCP010544742.1	c108325_g1_i2		
		PER; period circadian protein	AVCP010544746.1			
	K03083	GSK3B; glycogen synthase kinase 3 beta [EC:2.7.11.26]	AVCP010022305.1	c106317_g3_i4	c28740_g1_i1	c31952_g1_i2
		GSK3B; glycogen synthase kinase 3 beta [EC:2.7.11.26]		c106317_g3_i6		c31952_g2_i3
		GSK3B; glycogen synthase kinase 3 beta [EC:2.7.11.26]		c106317_g3_i8		
	K08960	CSNK1E; casein kinase 1, epsilon [EC:2.7.11.1]	AVCP010057707.1	c105680_g1_i2	c10344_g1_i1	c27216_g1_i1
		CSNK1E; casein kinase 1, epsilon [EC:2.7.11.1]	AVCP011068991.1		c10344_g1_i2	c27216_g1_i4
	CSNK1E; casein kinase 1, epsilon [EC:2.7.11.1]	AVCP011068996.1			c27216_g1_i5	
	CSNK1E; casein kinase 1, epsilon [EC:2.7.11.1]	AVCP011069001.1				
	CSNK1E; casein kinase 1, epsilon [EC:2.7.11.1]	AVCP011245865.1				
K12074	TIM; timeless	AVCP011153060.1	c108981_g1_i1	c20038_g1_i1	c35592_g3_i1	

	TIM; timeless	AVCP011153061.1	c108981_g1_i2	c2668_g1_i1	c5483_g1_i1
	TIM; timeless	AVCP010421397.1	c108981_g1_i3	c4195_g1_i1	
	TIM; timeless	AVCP011174069.1	c108981_g1_i4		
	TIM; timeless		c84037_g1_i2		
	TIM; timeless		c1848_g1_i1		
	TIM; timeless		c80157_g1_i2		
	TIM; timeless		c80157_g1_i3		
K02296	ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1				c24663_g1_i2
K02633	PER; period circadian protein				c11358_g1_i1
	PER; period circadian protein				c11358_g1_i2
K09057	HLF; hepatic leukemia factor	AVCP010669702.1	c105631_g3_i3		c39383_g1_i1
	HLF; hepatic leukemia factor	AVCP010678706.1	c92137_g1_i3		
	HLF; hepatic leukemia factor	AVCP010899555.1	c92137_g1_i4		
	HLF; hepatic leukemia factor	AVCP010899556.1	c92137_g1_i5		
	HLF; hepatic leukemia factor		c92137_g1_i7		

S4 appendix

Sample	<i>G. assimilis</i> ID	Gene name - Full name	Differential expression (log2FC)
Gonad	c107055_g1_i10	ORB2_DROME^ORB2_DROME^Q:7332-6550,H:444-704^94.64%ID^E:4e-165^RecName: Full=Probable RNA-binding protein orb2;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 6.9
	c88385_g2_i1	RIBC2_HUMAN^RIBC2_HUMAN^Q:1293-400,H:3-309^34.42%ID^E:1e-42^RecName: Full=RIB43A-like with coiled-coils protein 2;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	MBG 5.3
	c105223_g2_i1	LCA5_MOUSE^LCA5_MOUSE^Q:938-336,H:97-297^30.85%ID^E:1e-15^RecName: Full=Lebercilin;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 7.2
	c101293_g3_i3	PABP1_MOUSE^PABP1_MOUSE^Q:4583-2703,H:1-626^70.62%ID^E:0^RecName: Full=Polyadenylate-binding protein 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 2.5
	c99584_g1_i8	RN123_MOUSE^RN123_MOUSE^Q:4505-963,H:86-1298^40.35%ID^E:0^RecName: Full=E3 ubiquitin-protein ligase RNF123;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 4.1
	c92530_g1_i1	TM20B_DANRE^TM20B_DANRE^Q:1394-1140,H:67-146^61.18%ID^E:3e-13^RecName: Full=Mitochondrial import receptor subunit TOM20 homolog B;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio	MBG 3.7
	c107095_g2_i3	LZTL1_HUMAN^LZTL1_HUMAN^Q:1497-586,H:2-299^43.93%ID^E:5e-59^RecName: Full=Leucine zipper transcription factor-like protein 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	MBG 5.3
	c76517_g1_i3	HOME1_RAT^HOME1_RAT^Q:463-128,H:1-112^75.89%ID^E:8e-53^RecName: Full=Homer protein homolog 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus	MBG 7.5
	c92902_g1_i2	IPO7_HUMAN^IPO7_HUMAN^Q:5122-1973,H:1-1036^58.29%ID^E:0^RecName: Full=Importin-7;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	MBG 3.1
	c92844_g1_i2	ATP5J_DROME^ATP5J_DROME^Q:819-511,H:1-101^58.25%ID^E:2e-30^RecName: Full=ATP synthase-coupling factor 6, mitochondrial;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 5.3
	c100170_g1_i4	TTL3A_DROME^TTL3A_DROME^Q:1474-797,H:461-673^56.19%ID^E:2e-139^RecName: Full=Tubulin glycyclase 3A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 6.7
	c106025_g1_i1	CLU_AEDAE^CLU_AEDAE^Q:4992-943,H:35-1374^69.13%ID^E:0^RecName: Full=Clustered mitochondria protein homolog {ECO:0000255 HAMAP-Rule:MF_03013};^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae; Culicinae; Aedini; Aedes; Stegomyia	MBG 3.1
	c97368_g1_i9	TRI37_MOUSE^TRI37_MOUSE^Q:3434-2106,H:6-446^76.75%ID^E:0^RecName: Full=E3 ubiquitin-protein ligase TRIM37;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 5.5
	c101293_g3_i6	PABP1_MOUSE^PABP1_MOUSE^Q:4418-2703,H:56-626^67.86%ID^E:0^RecName: Full=Polyadenylate-binding protein 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 2.6
	c99584_g1_i6	RN123_MOUSE^RN123_MOUSE^Q:3977-435,H:86-1298^40.35%ID^E:0^RecName: Full=E3 ubiquitin-protein ligase RNF123;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 4.3
	c90442_g1_i4	K1328_HUMAN^K1328_HUMAN^Q:964-668,H:59-155^35.35%ID^E:3e-06^RecName: Full=Uncharacterized protein KIAA1328;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	MBG 7.9
	c92692_g1_i1	SPT17_HUMAN^SPT17_HUMAN^Q:2053-1523,H:1-177^28.25%ID^E:4e-14^RecName: Full=Spermatogenesis-associated protein 17;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	MBG 8.5
	c88385_g2_i3	RIBC2_HUMAN^RIBC2_HUMAN^Q:1248-400,H:18-309^34.13%ID^E:1e-42^RecName: Full=RIB43A-like with coiled-coils protein 2;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	MBG 5.5
	c100597_g2_i7	ETHE1_ARATH^ETHE1_ARATH^Q:2081-1392,H:52-289^61.76%ID^E:2e-94^RecName: Full=Persulfide dioxygenase ETHE1 homolog, mitochondrial;^Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae; Pentapetalae; rosids; malvids; Brassicales; Brassicaceae; Camelineae; Arabidopsis	MBG 7.9
	c108947_g1_i1	MDN1_HUMAN^MDN1_HUMAN^Q:13623-10354,H:3221-4312^29.79%ID^E:3e-122^RecName: Full=Midasin;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	MBG 4.3
		MDN1_HUMAN^MDN1_HUMAN^Q:7950-6457,H:5098-5594^45.01%ID^E:5e-108^RecName: Full=Midasin;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	MBG 4.3
		Homo^MDN1_HUMAN^MDN1_HUMAN^Q:9513-8620,H:4549-4868^34.66%ID^E:2e-15^RecName: Full=Midasin;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	

c102525_g1_i1	DRC1_MOUSE^DRC1_MOUSE^Q:2473-359,H:94-747^31.98%ID^E:7e-92^RecName: Full=Dynein regulatory complex protein 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 5.7
c97195_g1_i6	GLD2A_DROME^GLD2A_DROME^Q:1668-613,H:905-1259^55.49%ID^E:2e-122^RecName: Full=Poly(A) RNA polymerase gld-2 homolog A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 4.5
c105604_g4_i1	TBA1_DROME^TBA1_DROME^Q:840-1,H:123-402^96.79%ID^E:0^RecName: Full=Tubulin alpha-1 chain;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 12.6
c102525_g1_i4	DRC1_MOUSE^DRC1_MOUSE^Q:3040-722,H:25-747^32.19%ID^E:9e-99^RecName: Full=Dynein regulatory complex protein 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 5.9
c106012_g3_i1	GLMN_MOUSE^GLMN_MOUSE^Q:2202-508,H:44-584^28.85%ID^E:2e-44^RecName: Full=Glomulin;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 3.4
c105695_g1_i4	PAIP1_XENLA^PAIP1_XENLA^Q:1697-981,H:136-385^27.27%ID^E:2e-20^RecName: Full=Polyadenylate-binding protein-interacting protein 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Pipoidae; Xenopodinae; Xenopus; Xenopus	MBG 4.7
c103225_g2_i1	ALF_DROME^ALF_DROME^Q:2163-1072,H:1-361^76.65%ID^E:0^RecName: Full=Fructose-bisphosphate aldolase;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 3.5
c97368_g1_i7	TR137_MOUSE^TR137_MOUSE^Q:3234-1912,H:8-446^76.64%ID^E:0^RecName: Full=E3 ubiquitin-protein ligase TRIM37;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 5.5
c98758_g2_i4	MAOX_ANAPL^MAOX_ANAPL^Q:1829-183,H:3-550^67.09%ID^E:0^RecName: Full=NADP-dependent malic enzyme;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Testudines + Archosauria group; Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria; Aves; Neognathae; Galloanserae; Anseriformes; Anatidae; Anas	MBG 6.3
c106025_g1_i4	CLU_AEDAE^CLU_AEDAE^Q:4992-943,H:35-1374^69.13%ID^E:0^RecName: Full=Clustered mitochondria protein homolog {ECO:0000255 HAMAP-Rule:MF_03013};^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae; Culicinae; Aedini; Aedes; Stegomyia	MBG 3.3
c108454_g1_i1	TBB_STRPU^TBB_STRPU^Q:2425-1613,H:1-271^86.72%ID^E:4e-175^RecName: Full=Tubulin beta chain;^Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus	MBG 10.9
c90601_g1_i1	ODF3B_DANRE^ODF3B_DANRE^Q:719-264,H:100-257^38.75%ID^E:1e-23^RecName: Full=Outer dense fiber protein 3-B;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio	MBG 10.2
c105121_g2_i1	PO210_RAT^PO210_RAT^Q:5094-46,H:2-1645^29.19%ID^E:0^RecName: Full=Nuclear pore membrane glycoprotein 210;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus	MBG 4
c107095_g2_i2	LZTL1_HUMAN^LZTL1_HUMAN^Q:1500-586,H:1-299^44.12%ID^E:1e-59^RecName: Full=Leucine zipper transcription factor-like protein 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae; Homo	MBG 5.4
c103002_g1_i3	MPV17_XENLA^MPV17_XENLA^Q:769-425,H:62-176^47.83%ID^E:4e-29^RecName: Full=Protein Mpv17;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Pipoidae; Pipidae; Xenopodinae; Xenopus; Xenopus	MBG 6.6
c76517_g1_i2	MPV17_XENLA^MPV17_XENLA^Q:1527-1030,H:10-175^36.31%ID^E:4e-22^RecName: Full=Protein Mpv17;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Pipoidae; Pipidae; Xenopodinae; Xenopus; Xenopus	MBG 7.1
c107884_g1_i3	HOME3_RAT^HOME3_RAT^Q:460-128,H:5-115^73.87%ID^E:3e-52^RecName: Full=Homer protein homolog 3;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus	MBG 3.5
c97368_g1_i5	SORL_MOUSE^SORL_MOUSE^Q:7504-1151,H:78-2215^32.4%ID^E:0^RecName: Full=Sortilin-related receptor;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 5.5
c99809_g1_i4	TRI37_MOUSE^TRI37_MOUSE^Q:3671-2343,H:6-446^76.75%ID^E:0^RecName: Full=E3 ubiquitin-protein ligase TRIM37;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 6.1
c97848_g1_i10	UBX11_RAT^UBX11_RAT^Q:1159-413,H:61-313^30%ID^E:2e-26^RecName: Full=UBX domain-containing protein 11;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus	MBG 4.6
c103225_g2_i3	THIO2_DROYA^THIO2_DROYA^Q:2560-2279,H:1-95^65.26%ID^E:5e-35^RecName: Full=Thioredoxin-2;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 4.1
c97195_g1_i2	ALF_DROME^ALF_DROME^Q:2925-1834,H:1-361^76.1%ID^E:0^RecName: Full=Fructose-bisphosphate aldolase;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 4.7
c93253_g1_i2	GLD2A_DROME^GLD2A_DROME^Q:1668-613,H:905-1259^55.49%ID^E:3e-126^RecName: Full=Poly(A) RNA polymerase gld-2 homolog A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 5.6
c104617_g1_i1	K1841_MOUSE^K1841_MOUSE^Q:2382-1336,H:118-471^40.67%ID^E:2e-79^RecName: Full=Uncharacterized protein KIAA1841;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 7.3
c94650_g2_i1	TOM7_HUMAN^TOM7_HUMAN^Q:1644-1483,H:1-54^40.74%ID^E:5e-06^RecName: Full=Mitochondrial import receptor subunit TOM7 homolog;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae; Homo	MBG 4.8
	TYPH_RAT^TYPH_RAT^Q:1219-308,H:156-463^37.82%ID^E:3e-71^RecName: Full=Thymidine phosphorylase;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus	
	TYPH_RAT^TYPH_RAT^Q:1574-1335,H:79-159^70.37%ID^E:3e-71^RecName: Full=Thymidine phosphorylase;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus	

c100501_g1_i1	TMM35_DANRE^TMM35_DANRE^Q:3096-2701,H:6-134^37.12%ID^E:3e-21^RecName: Full=Transmembrane protein 35;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio	MBG 7.3
c97848_g1_i7	THIO2_DROYA^THIO2_DROYA^Q:538-257,H:1-95^65.26%ID^E:5e-38^RecName: Full=Thioredoxin-2;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 4.2
c94010_g1_i3	PEPD_RAT^PEPD_RAT^Q:794-168,H:272-479^63.16%ID^E:3e-70^RecName: Full=Xaa-Pro dipeptidase;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus	MBG 5.6
c108571_g3_i3	MFRN2_DANRE^MFRN2_DANRE^Q:2101-1625,H:199-366^66.67%ID^E:3e-60^RecName: Full=Mitoferrin-2;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio	MBG 4.3
c101293_g3_i1	PABP1_MOUSE^PABP1_MOUSE^Q:4583-2703,H:1-626^70.62%ID^E:0^RecName: Full=Polyadenylate-binding protein 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 2.5
c107900_g1_i5	VATL_AEDAE^VATL_AEDAE^Q:2477-2007,H:1-157^87.9%ID^E:7e-80^RecName: Full=V-type proton ATPase 16 kDa proteolipid subunit;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae; Culicinae; Aedini; Aedes; Stegomyia	FBG 2.5
c99520_g1_i1	POLO_DROME^POLO_DROME^Q:3083-1410,H:1-563^55.46%ID^E:0^RecName: Full=Serine/threonine-protein kinase polo;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 2.3
c98305_g2_i2	RL4_DROME^RL4_DROME^Q:1773-694,H:3-364^71.82%ID^E:1e-156^RecName: Full=60S ribosomal protein L4;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 2.4
c98149_g1_i2	UBR7_HUMAN^UBR7_HUMAN^Q:1992-1462,H:36-212^57.78%ID^E:2e-64^RecName: Full=Putative E3 ubiquitin-protein ligase UBR7;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 5.6
c98578_g1_i1	424^42.55%ID^E:4e-26^RecName: Full=Putative E3 ubiquitin-protein ligase UBR7;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 3.8
c85882_g2_i5	AKT1_DROME^AKT1_DROME^Q:3947-2484,H:109-588^75%ID^E:0^RecName: Full=RAC serine/threonine-protein kinase;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 4.7
c87515_g1_i1	UCHL3_BOVIN^UCHL3_BOVIN^Q:1701-991,H:19-230^45.38%ID^E:2e-68^RecName: Full=Ubiquitin carboxyl-terminal hydrolase isozyme L3;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos	FBG 2.4
c98740_g2_i1	CCNB_PATVU^CCNB_PATVU^Q:1246-344,H:127-400^41.86%ID^E:1e-68^RecName: Full=G2/mitotic-specific cyclin-B;^Eukaryota; Metazoa; Lophotrochozoa; Mollusca; Gastropoda; Patellogastropoda; Patelloidea; Patellidae; Patella	FBG 3.1
c100423_g1_i1	L2EFL_DROME^L2EFL_DROME^Q:704-207,H:1-167^47.34%ID^E:2e-37^RecName: Full=Protein lethal(2)essential for life;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 2.7
c87293_g1_i1	CCNB3_DROME^CCNB3_DROME^Q:1117-260,H:273-559^56.51%ID^E:1e-95^RecName: Full=G2/mitotic-specific cyclin-B3;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 2.7
c90460_g1_i1	PK3C3_XENLA^PK3C3_XENLA^Q:3942-1216,H:4-886^64.81%ID^E:0^RecName: Full=Phosphatidylinositol 3-kinase catalytic subunit type 3;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Pipoidae; Pipidae; Xenopodinae; Xenopus; Xenopus	FBG 2.6
c93743_g1_i3	S35F6_HUMAN^S35F6_HUMAN^Q:2940-1759,H:1-369^50.38%ID^E:1e-106^RecName: Full=Solute carrier family 35 member F6;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 2.2
c102027_g2_i1	PKRA_DROME^PKRA_DROME^Q:2845-1625,H:1-406^70.27%ID^E:0^RecName: Full=Protein krasavietz;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 3.6
c107230_g2_i8	STIP1_MACFA^STIP1_MACFA^Q:4752-3130,H:4-542^56.43%ID^E:0^RecName: Full=Stress-induced-phosphoprotein 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Cercopithecidae; Cercopithecinae; Macaca	FBG 1.9
c90871_g1_i6	CDR2L_HUMAN^CDR2L_HUMAN^Q:4584-3970,H:27-234^38.94%ID^E:3e-25^RecName: Full=Cerebellar degeneration-related protein 2-like;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 3.6
c107230_g2_i3	ARF1_LOCF1^ARF1_LOCF1^Q:2995-2450,H:1-182^99.45%ID^E:6e-120^RecName: Full=ADP-ribosylation factor 1;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera; Acridomorpha; Acridoidea; Acrididae; Oedipodinae; Locusta	FBG 3.6
c101957_g1_i1	CDR2L_HUMAN^CDR2L_HUMAN^Q:3551-2925,H:27-238^38.68%ID^E:2e-25^RecName: Full=Cerebellar degeneration-related protein 2-like;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 4.7
c94332_g1_i4	GOR_HUMAN^GOR_HUMAN^Q:3186-2428,H:408-656^52.34%ID^E:2e-75^RecName: Full=Putative exonuclease GOR;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 2.2
c85419_g1_i2	AP2M1_CHICK^AP2M1_CHICK^Q:1783-476,H:1-433^88.53%ID^E:0^RecName: Full=AP-2 complex subunit mu;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Testudines + Archosauria group; Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria; Aves; Neognathae; Galloanserae; Galliformes; Phasianidae; Phasianinae; Gallus	FBG 3.7
c103652_g2_i2	GBLP_ORENI^GBLP_ORENI^Q:1743-802,H:1-312^76.11%ID^E:8e-179^RecName: Full=Guanine nucleotide-binding protein subunit beta-2-like 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Neoteleostei; Acanthomorpha; Ovalentaria; Cichlomorphae; Cichliformes; Cichlidae; African cichlids; Pseudocrenilabrinae; Oreochromini; Oreochromis	FBG 2.8
	IAP_GVCPM^IAP_GVCPM^Q:1683-742,H:3-275^37.14%ID^E:3e-69^RecName: Full=Apoptosis inhibitor IAP;^Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Betabaculovirus	

c86223_g1_i2	ROST_DROME^ROST_DROME^Q:1745-1062,H:12-236^29.11%ID^E:3e-29^RecName: Full=Protein rolling stone;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 4.4
c98740_g1_i3	L2EFL_DROME^L2EFL_DROME^Q:678-184,H:1-170^45.88%ID^E:2e-45^RecName: Full=Protein lethal(2)essential for life;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 3.8
c108385_g1_i1	CCD50_HUMAN^CCD50_HUMAN^Q:6820-6059,H:12-208^31.1%ID^E:5e-08^RecName: Full=Coiled-coil domain-containing protein 50;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 3.7
c108528_g1_i1	HMMR_MOUSE^HMMR_MOUSE^Q:1655-1398,H:658-743^39.53%ID^E:5e-10^RecName: Full=Hyaluronan mediated motility receptor;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Mus; Mus	FBG 3.8
c100349_g1_i1	PELI_DROME^PELI_DROME^Q:1626-508,H:45-424^56.15%ID^E:5e-143^RecName: Full=Protein pellino;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 4.6
c63697_g1_i1	GDIA_PANTR^GDIA_PANTR^Q:2514-1198,H:1-440^67.95%ID^E:0^RecName: Full=Rab GDP dissociation inhibitor alpha;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Pan	FBG 2.2
c100423_g1_i3	CCNB3_DROME^CCNB3_DROME^Q:1117-260,H:273-559^56.51%ID^E:3e-96^RecName: Full=G2/mitotic-specific cyclin-B3;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 2.8
c107531_g1_i2	RT31_BOVIN^RT31_BOVIN^Q:3538-2729,H:133-386^38.13%ID^E:6e-33^RecName: Full=28S ribosomal protein S31, mitochondrial;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos	FBG 3.9
c85882_g2_i4	UCHL3_BOVIN^UCHL3_BOVIN^Q:1662-991,H:6-230^52.44%ID^E:2e-81^RecName: Full=Ubiquitin carboxyl-terminal hydrolase isozyme L3;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos	FBG 4.5
c101957_g1_i4	GOR_HUMAN^GOR_HUMAN^Q:3186-2428,H:408-656^52.34%ID^E:2e-75^RecName: Full=Putative exonuclease GOR;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 4.3
c93371_g1_i1	VAT1L_HUMAN^VAT1L_HUMAN^Q:2046-1027,H:42-379^57.65%ID^E:5e-126^RecName: Full=Synaptic vesicle membrane protein VAT-1 homolog-like;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 3.9
c87515_g1_i5	CCNB_PATVU^CCNB_PATVU^Q:1180-344,H:127-400^45.16%ID^E:2e-73^RecName: Full=G2/mitotic-specific cyclin-B;^Eukaryota; Metazoa; Lophotrochozoa; Mollusca; Gastropoda; Patellogastropoda; Patelloidea; Patellicidae; Patella	FBG 2.3
c86223_g1_i1	ROST_DROME^ROST_DROME^Q:1745-1062,H:12-236^29.11%ID^E:3e-29^RecName: Full=Protein rolling stone;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 4.2
c107001_g1_i1	NOL9_DROME^NOL9_DROME^Q:1890-364,H:485-964^28.41%ID^E:3e-52^RecName: Full=Polynucleotide 5'-hydroxyl-kinase NOL9;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 2.8
c107123_g1_i2	RAB6_DROME^RAB6_DROME^Q:1565-972,H:11-208^76.88%ID^E:1e-95^RecName: Full=Ras-related protein Rab6 {ECO:0000312 EMBL:AAF53168.1};^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 3.3
c100706_g1_i1	TRF_BLADI^TRF_BLADI^Q:2316-433,H:51-715^29.18%ID^E:2e-82^RecName: Full=Transferrin;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattodea; Blaberoidea; Blaberidae; Blaberinae; Blaberus	FBG 3.9
c101811_g1_i4	TRF_BLADI^TRF_BLADI^Q:2421-1606,H:367-640^23.08%ID^E:8e-11^RecName: Full=Transferrin;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattodea; Blaberoidea; Blaberidae; Blaberinae; Blaberus	FBG 5.1
c93371_g1_i2	RAB32_HUMAN^RAB32_HUMAN^Q:2134-1502,H:20-225^68.25%ID^E:5e-92^RecName: Full=Ras-related protein Rab-32;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 5.1
c95400_g1_i1	VAT1L_HUMAN^VAT1L_HUMAN^Q:2046-1027,H:42-379^57.65%ID^E:5e-126^RecName: Full=Synaptic vesicle membrane protein VAT-1 homolog-like;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 3.9
c103409_g1_i2	CDC6_HUMAN^CDC6_HUMAN^Q:1512-280,H:148-556^44.71%ID^E:3e-97^RecName: Full=Cell division control protein 6 homolog;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 4.7
c91860_g1_i1	VGR_SOLIN^VGR_SOLIN^Q:5429-345,H:33-1690^34.38%ID^E:0^RecName: Full=Vitellin receptor {ECO:0000312 EMBL:AAP92450.1};^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Vespoidea; Formicidae; Myrmicinae;	FBG 10.2
c103409_g1_i1	Solenopsis VGR_SOLIN^VGR_SOLIN^Q:2456-1971,H:32-200^34.5%ID^E:4e-18^RecName: Full=Vitellin receptor {ECO:0000312 EMBL:AAP92450.1};^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Vespoidea; Formicidae; Myrmicinae; Solenopsis	FBG 6.5
c91860_g1_i1	AUB_DROME^AUB_DROME^Q:2543-195,H:75-861^46.03%ID^E:0^RecName: Full=Protein aubergine {ECO:0000312 EMBL:AGA18944.1};^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 6.5
c103409_g1_i1	VGR_SOLIN^VGR_SOLIN^Q:5357-330,H:33-1697^34.68%ID^E:0^RecName: Full=Vitellin receptor {ECO:0000312 EMBL:AAP92450.1};^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Vespoidea; Formicidae; Myrmicinae;	FBG 10.2
c93403_g1_i1	Solenopsis VGR_SOLIN^VGR_SOLIN^Q:2384-1899,H:32-200^34.5%ID^E:4e-18^RecName: Full=Vitellin receptor {ECO:0000312 EMBL:AAP92450.1};^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Vespoidea; Formicidae; Myrmicinae; Solenopsis	FBG 10.2
c93403_g1_i1	NASP_HUMAN^NASP_HUMAN^Q:1144-464,H:511-757^42.19%ID^E:2e-34^RecName: Full=Nuclear autoantigenic sperm protein;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 3.8
c98687_g2_i1	NASP_HUMAN^NASP_HUMAN^Q:1600-1394,H:42-110^56.52%ID^E:1e-16^RecName: Full=Nuclear autoantigenic sperm protein;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 3.8
c98687_g2_i1	TBA1_DROME^TBA1_DROME^Q:902-3,H:95-394^99.67%ID^E:0^RecName: Full=Tubulin alpha-1 chain;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 3.6

c94769_g2_i1	SUI1_ANOGA^SUI1_ANOGA^Q:2857-2528,H:1-110^99.09%ID^E:7e-68^RecName: Full=Protein translation factor SUI1 homolog;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae; Anophelinae; Anopheles	FBG 3.3
c100562_g1_i1	L2EFL_DROME^L2EFL_DROME^Q:789-289,H:1-170^50%ID^E:3e-49^RecName: Full=Protein lethal(2)essential for life;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 4.1
c100562_g1_i2	L2EFL_DROME^L2EFL_DROME^Q:789-289,H:1-170^50%ID^E:2e-48^RecName: Full=Protein lethal(2)essential for life;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 4
c107354_g1_i1	TOLL_DROME^TOLL_DROME^Q:4188-1252,H:44-1044^35.1%ID^E:5e-165^RecName: Full=Protein toll;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 5.5
c101087_g1_i3	UN13D_HUMAN^UN13D_HUMAN^Q:3246-151,H:54-1078^25.51%ID^E:6e-72^RecName: Full=Protein unc-13 homolog D;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 6.1
c98705_g1_i3	LSD2_DROME^LSD2_DROME^Q:899-249,H:34-285^29.3%ID^E:2e-29^RecName: Full=Lipid storage droplets surface-binding protein 2;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 4.4
Head		
c103398_g2_i2	KARG_SCHAM^KARG_SCHAM^Q:1855-788,H:1-356^89.33%ID^E:0^RecName: Full=Arginine kinase;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera; Acridomorpha; Acridoidea; Acrididae; Cyrtacanthacridinae; Schistocerca	MBG 4.8
c103398_g2_i1	KARG_SCHAM^KARG_SCHAM^Q:1855-788,H:1-356^89.33%ID^E:0^RecName: Full=Arginine kinase;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera; Acridomorpha; Acridoidea; Acrididae; Cyrtacanthacridinae; Schistocerca	MBG 4.9
c103398_g2_i3	KARG_SCHAM^KARG_SCHAM^Q:1855-788,H:1-356^89.33%ID^E:0^RecName: Full=Arginine kinase;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera; Acridomorpha; Acridoidea; Acrididae; Cyrtacanthacridinae; Schistocerca	MBG 4.8
c103398_g2_i4	KARG_DROME^KARG_DROME^Q:1290-1090,H:242-308^88.06%ID^E:4e-55^RecName: Full=Arginine kinase;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 4.8
c107953_g1_i7	KARG_DROME^KARG_DROME^Q:1543-1397,H:196-244^89.8%ID^E:4e-55^RecName: Full=Arginine kinase;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 4.8
c107953_g1_i7	KARG_DROME^KARG_DROME^Q:940-788,H:306-356^84.31%ID^E:3e-19^RecName: Full=Arginine kinase;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 4.8
c98938_g1_i4	HEXA_BLADI^HEXA_BLADI^Q:1256-324,H:382-700^44.06%ID^E:4e-80^RecName: Full=Hexamerin;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattodea; Blaberoidea; Blaberidae; Blaberinae; Blaberus	MBG 5.3
c32561_g1_i1	KAD1_CAEEL^KAD1_CAEEL^Q:1456-848,H:3-208^56.31%ID^E:2e-71^RecName: Full=Adenylate kinase isoenzyme 1 {ECO:0000255 HAMAP-Rule:MF_03171};^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	MBG 4
c103964_g6_i1	BMP3_MOUSE^BMP3_MOUSE^Q:1627-1319,H:366-468^48.54%ID^E:2e-28^RecName: Full=Bone morphogenetic protein 3;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 3.5
c102722_g1_i8	LPSBP_PERAM^LPSBP_PERAM^Q:1216-602,H:47-256^32.39%ID^E:2e-19^RecName: Full=Hemolymph lipopolysaccharide-binding protein;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattodea; Blattodea; Blattidae; Blattinae; Periplaneta	MBG 6
c102722_g1_i8	HUTH_BOVIN^HUTH_BOVIN^Q:2058-163,H:6-638^64.74%ID^E:0^RecName: Full=Histidine ammonia-lyase;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos	MBG 4.8
c106666_g3_i1	LRC15_MOUSE^LRC15_MOUSE^Q:2471-1650,H:29-309^28.87%ID^E:1e-20^RecName: Full=Leucine-rich repeat-containing protein 15;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 6.2
c102200_g1_i2	LRC15_MOUSE^LRC15_MOUSE^Q:2438-1650,H:133-405^29.3%ID^E:5e-17^RecName: Full=Leucine-rich repeat-containing protein 15;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 6.2
c98807_g2_i2	VIT6_CAEEL^VIT6_CAEEL^Q:4739-2994,H:34-655^21.79%ID^E:7e-13^RecName: Full=Vitellogenin-6;^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	MBG 4.7
c102200_g1_i9	VIT6_CAEEL^VIT6_CAEEL^Q:959-513,H:1340-1489^28.76%ID^E:7e-07^RecName: Full=Vitellogenin-6;^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	MBG 4.7
c99546_g1_i1	MYPH_ECHGR^MYPH_ECHGR^Q:546-79,H:22-179^57.59%ID^E:3e-57^RecName: Full=Myophilin;^Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda; Cyclophyllidea; Taeniidae; Echinococcus	MBG 4.2
c107520_g4_i1	VIT6_CAEEL^VIT6_CAEEL^Q:4818-3073,H:34-655^21.79%ID^E:7e-13^RecName: Full=Vitellogenin-6;^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	MBG 4.7
c107520_g4_i1	VIT6_CAEEL^VIT6_CAEEL^Q:959-513,H:1340-1489^28.76%ID^E:6e-07^RecName: Full=Vitellogenin-6;^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	MBG 4.7
c107520_g4_i1	MRJP2_APIME^MRJP2_APIME^Q:933-436,H:223-394^25.56%ID^E:3e-10^RecName: Full=Major royal jelly protein 2;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis	MBG 6.2
c107520_g4_i1	ISFD2_CHRS^ISFD2_CHRS^Q:568-368,H:175-244^32.86%ID^E:2e-07^RecName: Full=Sulfoacetaldehyde reductase 2;^Bacteria; Proteobacteria; Gammaproteobacteria; Oceanospirillales; Halomonadaceae; Chromohalobacter	MBG 5.6
c102722_g1_i9	HUTH_BOVIN^HUTH_BOVIN^Q:1761-163,H:106-638^71.29%ID^E:0^RecName: Full=Histidine ammonia-lyase;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos	MBG 4.8

c98938_g1_i3	KAD1_CAEEL^KAD1_CAEEL^Q:1456-848,H:3-208^56.31%ID^E:2e-71^RecName: Full=Adenylate kinase isoenzyme 1 {ECO:0000255 HAMAP-Rule:MF_03171};^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	MBG 3.9
c102722_g1_i4	HUTH_BOVIN^HUTH_BOVIN^Q:1221-163,H:286-638^73.09%ID^E:0^RecName: Full=Histidine ammonia-lyase;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos	MBG 4.7
c106915_g12_i1	TNNC1_DROME^TNNC1_DROME^Q:1993-1565,H:1-143^74.83%ID^E:1e-62^RecName: Full=Troponin C, isoform 1;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 3.5
c105961_g3_i1	CRPI_PERAM^CRPI_PERAM^Q:244-134,H:640-676^75.68%ID^E:4e-10^RecName: Full=Allergen Cr-PI;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattodea; Blattoidea; Blattidae; Blattinae; Periplaneta	MBG 4.9
c98807_g2_i1	MP20_DROME^MP20_DROME^Q:1493-1179,H:15-120^52.83%ID^E:2e-31^RecName: Full=Muscle-specific protein 20;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 3.9
c102722_g1_i2	HUTH_BOVIN^HUTH_BOVIN^Q:1830-85,H:6-588^66.33%ID^E:0^RecName: Full=Histidine ammonia-lyase;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos	MBG 5.2
c106915_g11_i1	TNNC1_DROME^TNNC1_DROME^Q:299-3,H:1-99^66.67%ID^E:1e-34^RecName: Full=Troponin C, isoform 1;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 3.6
c102200_g1_i10	VIT6_CAEEL^VIT6_CAEEL^Q:979-533,H:1340-1489^28.76%ID^E:2e-07^RecName: Full=Vitellinogenin-6;^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	MBG 4.9
c102200_g1_i7	VIT6_CAEEL^VIT6_CAEEL^Q:3602-1857,H:34-655^21.79%ID^E:5e-13^RecName: Full=Vitellinogenin-6;^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	MBG 4.5
c106915_g12_i2	TNNC1_DROME^TNNC1_DROME^Q:1566-1114,H:1-151^76.16%ID^E:2e-69^RecName: Full=Troponin C, isoform 1;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 3.4
c102722_g1_i7	HUTH_BOVIN^HUTH_BOVIN^Q:1446-319,H:6-382^57.66%ID^E:6e-129^RecName: Full=Histidine ammonia-lyase;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos	MBG 4.7
c107953_g1_i2	PPO3_DROME^PPO3_DROME^Q:509-231,H:576-673^41.58%ID^E:2e-13^RecName: Full=Phenoloxidase 3;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 5.4
c79579_g1_i3	FBLN2_MOUSE^FBLN2_MOUSE^Q:3071-858,H:553-1219^33.47%ID^E:1e-94^RecName: Full=Fibulin-2;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 2.9
c107293_g2_i2	CBPC1_MOUSE^CBPC1_MOUSE^Q:2511-1393,H:781-1130^46.51%ID^E:2e-106^RecName: Full=Cytosolic carboxypeptidase 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 4.4
c102722_g1_i6	HUTH_HUMAN^HUTH_HUMAN^Q:645-163,H:478-638^71.43%ID^E:4e-65^RecName: Full=Histidine ammonia-lyase;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae; Homo	MBG 4.3
c92939_g1_i1	PYG_DROME^PYG_DROME^Q:3924-1417,H:5-840^83.73%ID^E:0^RecName: Full=Glycogen phosphorylase;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 2.1
c106170_g3_i2	LPSBP_PERAM^LPSBP_PERAM^Q:1445-816,H:47-256^34.82%ID^E:6e-31^RecName: Full=Hemolymph lipopolysaccharide-binding protein;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattodea; Blattoidea; Blattidae; Blattinae; Periplaneta	MBG 6
c102722_g1_i1	HUTH_BOVIN^HUTH_BOVIN^Q:1240-401,H:6-286^56.18%ID^E:1e-78^RecName: Full=Histidine ammonia-lyase;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos	MBG 4.7
c98807_g2_i3	MP20_DROME^MP20_DROME^Q:1507-1193,H:15-120^52.83%ID^E:2e-31^RecName: Full=Muscle-specific protein 20;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 3.8
c102200_g1_i6	VIT6_CAEEL^VIT6_CAEEL^Q:959-513,H:1340-1489^28.76%ID^E:2e-07^RecName: Full=Vitellinogenin-6;^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	MBG 4.8
c107293_g2_i3	CBPC1_MOUSE^CBPC1_MOUSE^Q:2955-1393,H:622-1130^48.79%ID^E:8e-159^RecName: Full=Cytosolic carboxypeptidase 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 4
c102200_g1_i4	VIT6_CAEEL^VIT6_CAEEL^Q:3000-1255,H:34-655^21.79%ID^E:4e-13^RecName: Full=Vitellinogenin-6;^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	MBG 4.8

c99980_g1_i2	BHMT1_XENLA^BHMT1_XENLA^Q:506-231,H:304-395^57.61%ID^E:2e-28^RecName: Full=Betaine--homocysteine S-methyltransferase 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Pipoidea; Pipidae; Xenopodinae; Xenopus; Xenopus^BHMT1_XENLA^BHMT1_XENLA^Q:1093-968,H:264-305^73.81%ID^E:3e-13^RecName: Full=Betaine--homocysteine S-methyltransferase 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Pipoidea; Pipidae; Xenopodinae; Xenopus; Xenopus	MBG 3.1
c104633_g2_i3	PGAM2_HUMAN^PGAM2_HUMAN^Q:1119-370,H:2-250^64.4%ID^E:8e-94^RecName: Full=Phosphoglycerate mutase 2;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	MBG 2.1
c99508_g1_i1	LPSBP_PERAM^LPSBP_PERAM^Q:1738-1166,H:49-253^33.5%ID^E:1e-27^RecName: Full=Hemolymph lipopolysaccharide-binding protein;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattodea; Blattellidae; Blattellinae; Periplaneta	MBG 5.6
c79579_g1_i1	FBLN1_DANRE^FBLN1_DANRE^Q:2420-852,H:176-681^39.27%ID^E:1e-93^RecName: Full=Fibulin-1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio^FBLN1_DANRE^FBLN1_DANRE^Q:2651-1524,H:193-542^34.44%ID^E:2e-38^RecName: Full=Fibulin-1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio	MBG 3
c98938_g1_i5	KAD1_CAEEL^KAD1_CAEEL^Q:721-176,H:3-185^58.47%ID^E:8e-72^RecName: Full=Adenylate kinase isoenzyme 1 {ECO:0000255 HAMAP-Rule:MF_03171};^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	MBG 3.9
c105363_g1_i1	SV2B_MOUSE^SV2B_MOUSE^Q:2535-1807,H:81-324^32.93%ID^E:2e-34^RecName: Full=Synaptic vesicle glycoprotein 2B;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	MBG 2.7
c102200_g1_i5	VIT6_CAEEL^VIT6_CAEEL^Q:959-513,H:1340-1489^28.76%ID^E:2e-07^RecName: Full=Vitellin-6;^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Peloderinae; Caenorhabditis	MBG 4.1
c107971_g1_i1	FAU_DROME^FAU_DROME^Q:877-509,H:153-279^29.32%ID^E:1e-06^RecName: Full=Protein anoxia up-regulated;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 2.1
	FLNA_DROME^FLNA_DROME^Q:2276-1254,H:1869-2210^73.98%ID^E:3e-165^RecName: Full=Filamin-A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora^FLNA_DROME^FLNA_DROME^Q:2276-1254,H:1493-1796^30.32%ID^E:2e-33^RecName: Full=Filamin-A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora^FLNA_DROME^FLNA_DROME^Q:2276-1254,H:1676-1986^27.67%ID^E:2e-27^RecName: Full=Filamin-A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora^FLNA_DROME^FLNA_DROME^Q:2279-1254,H:808-1121^28.61%ID^E:5e-22^RecName: Full=Filamin-A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora^FLNA_DROME^FLNA_DROME^Q:2276-1254,H:1290-1606^25.61%ID^E:1e-21^RecName: Full=Filamin-A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora^FLNA_DROME^FLNA_DROME^Q:2288-1254,H:421-734^25.71%ID^E:3e-19^RecName: Full=Filamin-A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora^FLNA_DROME^FLNA_DROME^Q:2279-1266,H:324-631^26.72%ID^E:3e-18^RecName: Full=Filamin-A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora^FLNA_DROME^FLNA_DROME^Q:2276-1257,H:1774-2078^26.86%ID^E:9e-18^RecName: Full=Filamin-A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora^FLNA_DROME^FLNA_DROME^Q:2279-1254,H:1400-1698^27.01%ID^E:4e-14^RecName: Full=Filamin-A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora^FLNA_DROME^FLNA_DROME^Q:2135-1254,H:274-544^25.82%ID^E:1e-12^RecName: Full=Filamin-A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora^FLNA_DROME^FLNA_DROME^Q:2258-1254,H:1105-1423^23.71%ID^E:9e-09^RecName: Full=Filamin-A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora^FLNA_DROME^FLNA_DROME^Q:1946-1254,H:245-447^25%ID^E:4e-06^RecName: Full=Filamin-A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora^FLNA_DROME^FLNA_DROME^Q:2276-1266,H:1195-1511^23.68%ID^E:5e-06^RecName: Full=Filamin-A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 2
c92939_g1_i2	PYG_DROME^PYG_DROME^Q:1563-1417,H:792-840^83.67%ID^E:4e-18^RecName: Full=Glycogen phosphorylase;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	MBG 2.2
c99980_g1_i4	BHMT1_BOVIN^BHMT1_BOVIN^Q:1397-234,H:9-397^71.98%ID^E:0^RecName: Full=Betaine--homocysteine S-methyltransferase 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos	MBG 3.2
c104633_g2_i1	PGAM2_HUMAN^PGAM2_HUMAN^Q:924-175,H:2-250^64.4%ID^E:1e-94^RecName: Full=Phosphoglycerate mutase 2;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	MBG 1.9
c93867_g1_i2	PCE_TACTR^PCE_TACTR^Q:1514-777,H:118-374^36.4%ID^E:2e-36^RecName: Full=Proclotting enzyme;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Chelicerata; Merostomata; Xiphosura; Limulidae; Tachypleus	FBG 4.3
c103479_g2_i2	LPSBP_PERAM^LPSBP_PERAM^Q:902-513,H:120-254^35.77%ID^E:4e-18^RecName: Full=Hemolymph lipopolysaccharide-binding protein;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattodea; Blattellidae; Blattellinae; Periplaneta	FBG 4.3
c103474_g1_i4	UAP1L_XENTR^UAP1L_XENTR^Q:5946-5740,H:421-504^41.67%ID^E:7e-09^RecName: Full=UDP-N-acetylhexosamine pyrophosphorylase-like protein 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana	FBG 2.6

c104447_g1_i3	FACR1_DROME^FACR1_DROME^Q:2336-1335,H:262-594^34.13%ID^E:1e-65^RecName: Full=Putative fatty acyl-CoA reductase CG5065 {ECO:0000250 UniProtKB:Q922J9, ECO:0000312 EMBL:AAF57974.1};^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 2.4
c101669_g1_i3	B3GT1_PONPY^B3GT1_PONPY^Q:6546-5764,H:53-315^33.96%ID^E:1e-36^RecName: Full=Beta-1,3-galactosyltransferase 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Pongo	FBG 2.1
c105235_g1_i1	ELVL1_AEDAE^ELVL1_AEDAE^Q:4132-3362,H:2-259^53.1%ID^E:7e-83^RecName: Full=Elongation of very long chain fatty acids protein AAEL008004;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae; Culicinae; Aedini; Aedes; Stegomyia	FBG 3.9
c101669_g1_i4	B3GT1_PONPY^B3GT1_PONPY^Q:6545-5763,H:53-315^33.96%ID^E:1e-36^RecName: Full=Beta-1,3-galactosyltransferase 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Pongo	FBG 2.2
c103129_g2_i2	PGBD3_HUMAN^PGBD3_HUMAN^Q:1265-186,H:237-573^26.94%ID^E:1e-17^RecName: Full=PiggyBac transposable element-derived protein 3;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 2
c104086_g2_i4	GFPT2_MOUSE^GFPT2_MOUSE^Q:3264-1246,H:4-682^68.38%ID^E:0^RecName: Full=Glutamine--fructose-6-phosphate aminotransferase [isomerizing] 2;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus; Mus	FBG 1.9
c108910_g1_i2	LHPL3_HUMAN^LHPL3_HUMAN^Q:5620-5099,H:20-190^39.08%ID^E:8e-28^RecName: Full=Lipoma HMGIC fusion partner-like 3 protein;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 1.9
c101252_g1_i1	LIPP_MYOCO^LIPP_MYOCO^Q:1963-1256,H:10-246^43.93%ID^E:9e-59^RecName: Full=Pancreatic triacylglycerol lipase;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Hystricognathi; Myocastoridae; Myocastor	FBG 4.7
c103479_g2_i1	LPSBP_PERAM^LPSBP_PERAM^Q:902-513,H:120-254^35.77%ID^E:4e-18^RecName: Full=Hemolymph lipopolysaccharide-binding protein;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattodea; Blattodea; Blattidae; Blattinae; Periplaneta	FBG 4.4
c105224_g1_i3	TAKT_DROME^TAKT_DROME^Q:2038-1370,H:25-245^27.19%ID^E:2e-18^RecName: Full=Protein takeout;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 3
c94190_g1_i4	RDH13_HUMAN^RDH13_HUMAN^Q:2772-1822,H:2-319^49.69%ID^E:4e-87^RecName: Full=Retinol dehydrogenase 13;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 3.7
c101329_g1_i1	RESIL_DROME^RESIL_DROME^Q:1991-1806,H:345-405^43.55%ID^E:8e-07^RecName: Full=Pro-resilin;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 3.3
c94190_g1_i2	RDH13_HUMAN^RDH13_HUMAN^Q:2772-1822,H:2-319^49.69%ID^E:3e-87^RecName: Full=Retinol dehydrogenase 13;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 3.8
c94190_g1_i1	RDH13_HUMAN^RDH13_HUMAN^Q:2772-1822,H:2-319^49.69%ID^E:5e-86^RecName: Full=Retinol dehydrogenase 13;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 3.7
c104447_g1_i2	FACR1_DROME^FACR1_DROME^Q:2768-1347,H:123-594^36.29%ID^E:1e-104^RecName: Full=Putative fatty acyl-CoA reductase CG5065 {ECO:0000250 UniProtKB:Q922J9, ECO:0000312 EMBL:AAF57974.1};^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 2.9
c96389_g1_i7	CUC1B_TENMO^CUC1B_TENMO^Q:267-130,H:34-79^80.43%ID^E:3e-17^RecName: Full=Pupal cuticle protein C1B;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Tenebrionidae; Tenebrio	FBG 3.8
c107941_g2_i4	CHIT3_DROME^CHIT3_DROME^Q:2571-550,H:692-1347^48.76%ID^E:0^RecName: Full=Probable chitinase 3;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora`CHIT3_DROME^CHIT3_DROME^Q:1695-562,H:1404-1780^64.55%ID^E:2e-149^RecName: Full=Probable chitinase 3;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora`CHIT3_DROME^CHIT3_DROME^Q:2058-565,H:1805-2279^49.9%ID^E:1e-143^RecName: Full=Probable chitinase 3;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora`CHIT3_DROME^CHIT3_DROME^Q:1677-568,H:220-594^44.33%ID^E:3e-89^RecName: Full=Probable chitinase 3;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora`CHIT3_DROME^CHIT3_DROME^Q:369-1,H:1408-1530^77.24%ID^E:2e-54^RecName: Full=Probable chitinase 3;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora`CHIT3_DROME^CHIT3_DROME^Q:447-1,H:937-1086^57.62%ID^E:4e-47^RecName: Full=Probable chitinase 3;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora`CHIT3_DROME^CHIT3_DROME^Q:441-1,H:1884-2029^55.78%ID^E:2e-44^RecName: Full=Probable chitinase 3;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora`CHIT3_DROME^CHIT3_DROME^Q:363-1,H:220-337^48.76%ID^E:2e-29^RecName: Full=Probable chitinase 3;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora`CHIT3_DROME^CHIT3_DROME^Q:2292-1846,H:638-848^24.88%ID^E:6e-06^RecName: Full=Probable chitinase 3;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 5.6
c97266_g1_i4	DPS1_HUMAN^DPS1_HUMAN^Q:2444-2022,H:275-415^51.77%ID^E:3e-34^RecName: Full=Decaprenyl-diphosphate synthase subunit 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 2.8
c94928_g1_i1	NRF6_CAEEL^NRF6_CAEEL^Q:2535-916,H:250-822^19.83%ID^E:1e-06^RecName: Full=Nose resistant to fluoxetine protein 6;^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	FBG 3.9

c97266_g1_i2	DPS1_HUMAN^DPS1_HUMAN^Q:2243-2022,H:342-415^47.3%ID^E:5e-07^RecName: Full=Decaprenyl-diphosphate synthase subunit 1;^Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo	FBG 2.9
c92267_g1_i1	RESIL_DROME^RESIL_DROME^Q:1029-844,H:341-402^58.06%ID^E:8e-14^RecName: Full=Pro-resilin;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 5.8
c98848_g1_i7	FACR1_DROME^FACR1_DROME^Q:1417-1127,H:492-588^32.99%ID^E:2e-09^RecName: Full=Putative fatty acyl-CoA reductase CG5065 {ECO:0000250 UniProtKB:Q922J9, ECO:0000312 EMBL:AAF57974.1};^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 3.2
c93761_g4_i1	NPL11_CAEL^NPL11_CAEL^Q:1821-385,H:379-848^39.05%ID^E:1e-116^RecName: Full=Nepriylsin-11;^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	FBG 4.4
c101765_g2_i1	CUO6_BLACR^CUO6_BLACR^Q:302-111,H:24-86^46.88%ID^E:2e-08^RecName: Full=Cuticle protein 6;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattodea; Blaberoidea; Blaberidae; Blaberinae; Blaberus	FBG 3.2
c101765_g2_i2	CUO6_BLACR^CUO6_BLACR^Q:302-111,H:24-86^46.88%ID^E:8e-10^RecName: Full=Cuticle protein 6;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattodea; Blaberoidea; Blaberidae; Blaberinae; Blaberus	FBG 3.7
c95630_g2_i1	TAKT_DROME^TAKT_DROME^Q:818-153,H:19-239^24.78%ID^E:2e-18^RecName: Full=Protein takeout;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 4.9
c102312_g1_i9	SPZ_DROME^SPZ_DROME^Q:3743-3444,H:229-326^28.16%ID^E:2e-06^RecName: Full=Protein spaetzle;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 3.1
c99631_g4_i1	CCCP_DROYA^CCCP_DROYA^Q:2884-2288,H:56-260^26.57%ID^E:1e-12^RecName: Full=Circadian clock-controlled protein;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 3.1
c98848_g1_i3	FACR1_DROME^FACR1_DROME^Q:2498-1131,H:135-588^39.47%ID^E:9e-100^RecName: Full=Putative fatty acyl-CoA reductase CG5065 {ECO:0000250 UniProtKB:Q922J9, ECO:0000312 EMBL:AAF57974.1};^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 3.6
c105529_g1_i1	RESIL_DROME^RESIL_DROME^Q:752-567,H:341-402^62.9%ID^E:6e-16^RecName: Full=Pro-resilin;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 6.2
c103819_g1_i2	ELVL1_AEDAE^ELVL1_AEDAE^Q:2945-2175,H:5-262^45.35%ID^E:6e-66^RecName: Full=Elongation of very long chain fatty acids protein AAEL008004;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae; Culicinae; Aedini; Aedes; Stegomyia	FBG 4.2
c97482_g4_i1	TCB1_CAEBR^TCB1_CAEBR^Q:133-2,H:173-215^50%ID^E:2e-07^RecName: Full=Transposable element Tcb1 transposase;^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	FBG 4.7
c103819_g1_i3	TCB1_CAEBR^TCB1_CAEBR^Q:287-165,H:120-160^29.27%ID^E:2e-07^RecName: Full=Transposable element Tcb1 transposase;^Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis	FBG 4
c95657_g6_i1	ELVL1_AEDAE^ELVL1_AEDAE^Q:2945-2175,H:5-262^45.35%ID^E:2e-66^RecName: Full=Elongation of very long chain fatty acids protein AAEL008004;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae; Culicinae; Aedini; Aedes; Stegomyia	FBG 4.8
c105529_g1_i3	RTBS_DROME^RTBS_DROME^Q:619-960,H:564-684^34.43%ID^E:3e-31^RecName: Full=Probable RNA-directed DNA polymerase from transposon BS;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 6.3
c97266_g1_i1	Sophophora^RTBS_DROME^RTBS_DROME^Q:314-616,H:460-560^38.61%ID^E:3e-31^RecName: Full=Probable RNA-directed DNA polymerase from transposon BS;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 2.9
c105529_g1_i4	Sophophora^RTBS_DROME^RTBS_DROME^Q:951-1499,H:686-865^25.81%ID^E:3e-31^RecName: Full=Probable RNA-directed DNA polymerase from transposon BS;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 6.5
c101765_g2_i3	RESIL_DROME^RESIL_DROME^Q:752-567,H:341-402^62.9%ID^E:5e-16^RecName: Full=Pro-resilin;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 3.4
c105529_g1_i2	RESIL_DROME^RESIL_DROME^Q:752-567,H:341-402^62.9%ID^E:1e-15^RecName: Full=Pro-resilin;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 5.9
c98848_g1_i8	FACR1_DROME^FACR1_DROME^Q:1927-1127,H:348-588^31.09%ID^E:8e-42^RecName: Full=Putative fatty acyl-CoA reductase CG5065 {ECO:0000250 UniProtKB:Q922J9, ECO:0000312 EMBL:AAF57974.1};^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 3.5
c104427_g5_i3	CUA3A_TENMO^CUA3A_TENMO^Q:415-230,H:35-96^83.87%ID^E:1e-29^RecName: Full=Larval cuticle protein A3A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Tenebrionidae; Tenebrio	FBG 5.8
c104427_g5_i4	CUA3A_TENMO^CUA3A_TENMO^Q:415-230,H:35-96^83.87%ID^E:6e-29^RecName: Full=Larval cuticle protein A3A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Tenebrionidae; Tenebrio	FBG 5.7
c102455_g2_i2	RESIL_DROME^RESIL_DROME^Q:864-307,H:214-403^44.5%ID^E:4e-21^RecName: Full=Pro-resilin;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 7.3

c102455_g2_i3	RESIL_DROME^RESIL_DROME^Q:495-307,H:341-403^66.67%ID^E:1e-18^RecName: Full=Pro-resilin;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 6.8
c104427_g5_i2	CUA3A_TENMO^CUA3A_TENMO^Q:415-230,H:35-96^83.87%ID^E:1e-29^RecName: Full=Larval cuticle protein A3A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Tenebrionidae; Tenebrio	FBG 5.5
c104427_g5_i1	CUA3A_TENMO^CUA3A_TENMO^Q:394-191,H:35-102^85.29%ID^E:2e-34^RecName: Full=Larval cuticle protein A3A;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Tenebrionidae; Tenebrio	FBG 5.5
c102455_g3_i1	RESIL_DROME^RESIL_DROME^Q:755-567,H:341-403^65.08%ID^E:6e-17^RecName: Full=Pro-resilin;^Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora	FBG 7.1

S5 Appendix

KEGG pathway	KEGG gene name	KO number	<i>G. assimilis</i> ID	Differential expression (log2FC)	
				Head	Gonad
Insect hormone biosynthesis*	JHEH; juvenile hormone epoxide hydrolase	K10719	c96010_g1_i11	MBG (0.9)	
	JHEH; juvenile hormone epoxide hydrolase	K10719	c63479_g1_i2	MBG (2.5)	MBG (3.2)
	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]	K15890	c97494_g2_i1	FBG (0.9)	
	JHEH; juvenile hormone epoxide hydrolase	K10719	c91728_g1_i3	MBG (1)	FBG (2.9)
	CYP18A1; 26-hydroxylase [EC:1.14.-.-]	K14985	c64866_g1_i1	FBG (2.3)	
	ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	K00128	c90017_g1_i3	MBG (1.3)	
	ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	K00128	c103318_g1_i1	MBG (1.3)	FBG (1.1)
	JHEH; juvenile hormone epoxide hydrolase	K10719	c96010_g1_i12	MBG (1)	MBG (2)
	ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	K00128	c90017_g1_i3	MBG (1.3)	
	JHEH; juvenile hormone epoxide hydrolase	K10719	c109061_g1_i3	MBG (2)	
	JHEH; juvenile hormone epoxide hydrolase	K10719	c96010_g1_i9		MBG (1.7)
	SAD, CYP315A1; ecdysteroid 2-hydroxylase	K10722	c86091_g1_i2		FBG (4.7)
	JHAMT; juvenile hormone-III synthase [EC:2.1.1.325]	K10718	c105887_g1_i1		MBG (4.2)
	PHM; CYP306A1; ecdysteroid 25-hydroxylase	K10720	c100334_g1_i2		FBG (2.5)
	NVD, DAF36; cholesterol 7-desaturase [EC:1.14.19.21]	K14938	c87477_g1_i1		MBG (0.5)
	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]	K15890	c103730_g2_i1		MBG (2.7)
	ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	K00128	c107419_g1_i6		MBG (0.8)
	SHD, CYP314A1; ecdysone 20-monoxygenase [EC:1.14.99.22]	K10723	c105588_g1_i3		MBG (9.8)
	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]	K15890	c103730_g1_i2		MBG (2)
	CYP15A1_C1; methyl farnesoate epoxidase / farnesoate epoxidase [EC:1.14.13.202 1.14.13.203]	K14937	c107918_g2_i3		MBG (2.8)
	NVD, DAF36; cholesterol 7-desaturase [EC:1.14.19.21]	K14938	c87477_g1_i4		MBG (0.5)
	FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]	K15890	c104977_g2_i6		FBG (3)
	CYP15A1_C1; methyl farnesoate epoxidase / farnesoate epoxidase [EC:1.14.13.202 1.14.13.203]	K14937	c94971_g1_i4		MBG (2.9)
CYP18A1; 26-hydroxylase [EC:1.14.-.-]	K14985	c135509_g1_i1		MBG (3.4)	
FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]	K15890	c97494_g2_i1		MBG (2.6)	

JHEH; juvenile hormone epoxide hydrolase	K10719	c96730_g1_i2	FBG (2.4)
SHD, CYP314A1; ecdysone 20-monoxygenase [EC:1.14.99.22]	K10723	c105588_g1_i6	MBG (8.7)
JHEH; juvenile hormone epoxide hydrolase	K10719	c96010_g1_i6	MBG (1.5)
FPPP; farnesyl diphosphate phosphatase [EC:3.1.3.-]	K21013	c107901_g2_i3	MBG (2.7)
JHE; juvenile-hormone esterase [EC:3.1.1.59]	K01063	c102917_g1_i1	MBG (8.5)
spo, spok, CYP307A; cytochrome P450 family 307 subfamily A	K14939	c148959_g1_i1	FBG (4)
JHEH; juvenile hormone epoxide hydrolase	K10719	c96730_g1_i4	FBG (2.5)
ALDH; aldehyde dehydrogenase (NAD+) [EC:1.2.1.3]	K00128	c107419_g1_i3	MBG (0.9)
SAD, CYP315A1; ecdysteroid 2-hydroxylase	K10722	c86091_g1_i1	FBG (4.9)
FPPP; farnesyl diphosphate phosphatase [EC:3.1.3.-]	K21013	c107901_g2_i4	MBG (1.4)
JHEH; juvenile hormone epoxide hydrolase	K10719	c63479_g1_i1	MBG (2.3)
JHEH; juvenile hormone epoxide hydrolase	K10719	c96730_g1_i3	FBG (2.6)
PHM; CYP306A1; ecdysteroid 25-hydroxylase	K10720	c100334_g1_i3	FBG (2.8)
FOHSDR; NADP+-dependent farnesol dehydrogenase [EC:1.1.1.216]	K15890	c104977_g2_i4	FBG (1)
JHEH; juvenile hormone epoxide hydrolase	K10719	c109061_g1_i3	MBG (4.9)
FPPP; farnesyl diphosphate phosphatase [EC:3.1.3.-]	K21013	c107901_g2_i2	MBG (3.6)
SHD, CYP314A1; ecdysone 20-monoxygenase [EC:1.14.99.22]	K10723	c85061_g1_i2	FBG (3.1)

Circadian rhythm

FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/11	K03362	c100367_g1_i5	FBG (0.8)	FBG (1)
ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1	K02296	c88570_g1_i1	FBG (0.9)	MBG (1.4)
CUL1, CDC53; cullin 1	K03347	c105966_g1_i1	FBG (0.6)	MBG (0.9)
ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1	K02296	c104661_g2_i1	FBG (0.7)	MBG (2)
CRY; cryptochrome	K02295	c105814_g1_i7		MBG (1.7)
CUL1, CDC53; cullin 1	K03347	c100300_g2_i4		FBG (2.2)
CUL1, CDC53; cullin 1	K03347	c105966_g1_i2		MBG (0.9)
RBX1, ROC1; RING-box protein 1	K03868	c104236_g2_i3		MBG (1.6)
FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/11	K03362	c100367_g1_i14		FBG (0.8)
FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/11	K03362	c100367_g1_i10		FBG (0.4)
PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit	K07200	c99381_g1_i6		FBG (2.1)

ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1	K02296	c97544_g2_i1	MBG (4.9)
CRY; cryptochrome	K02295	c97680_g1_i2	FBG (0.9)
RBX1, ROC1; RING-box protein 1	K03868	c104236_g2_i7	MBG (0.7)
FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/11	K03362	c100367_g1_i6	FBG (0.6)
CRY; cryptochrome	K02295	c105814_g1_i5	MBG (1.8)
PER; period circadian protein	K02633	c108325_g1_i2	MBG (1)
PRKAA, AMPK; 5'-AMP-activated protein kinase, catalytic alpha subunit [EC:2.7.11.11]	K07198	c93013_g1_i6	FBG (1.9)
CRY; cryptochrome	K02295	c105814_g1_i3	MBG (1.5)
FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/11	K03362	c100367_g1_i15	FBG (0.8)
SKP1, CBF3D; S-phase kinase-associated protein 1	K03094	c93326_g2_i4	MBG (1.4)
SKP1, CBF3D; S-phase kinase-associated protein 1	K03094	c93326_g2_i7	MBG (1.3)
CUL1, CDC53; cullin 1	K03347	c100300_g2_i2	FBG (1.9)
CRY; cryptochrome	K02295	c97680_g1_i3	FBG (1)
PRKAB; 5'-AMP-activated protein kinase, regulatory beta subunit	K07199	c103022_g2_i2	FBG (2.8)
FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/11	K03362	c41437_g1_i1	MBG (4.7)
PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit	K07200	c99381_g1_i5	FBG (2.2)
CRY; cryptochrome	K02295	c107392_g1_i5	MBG (1.2)
CRY; cryptochrome	K02295	c97680_g1_i6	FBG (1.1)
SKP1, CBF3D; S-phase kinase-associated protein 1	K03094	c40894_g1_i1	MBG (3)
SKP1, CBF3D; S-phase kinase-associated protein 1	K03094	c93326_g2_i3	MBG (1.4)
PRKAA, AMPK; 5'-AMP-activated protein kinase, catalytic alpha subunit [EC:2.7.11.11]	K07198	c93013_g1_i7	FBG (2)
PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit	K07200	c99381_g1_i1	FBG (1.2)
FBXW1_11, BTRC, beta-TRCP; F-box and WD-40 domain protein 1/11	K03362	c100367_g1_i11	FBG (0.7)
CLOCK, KAT13D; circadian locomotor output cycles kaput protein [EC:2.3.1.48]	K02223	c88878_g2_i3	FBG (1)
PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit	K07200	c99381_g1_i7	FBG (2.1)
CUL1, CDC53; cullin 1	K03347	c100300_g2_i3	FBG (2.2)
PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit	K07200	c99381_g1_i4	FBG (2.1)
ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1	K02296	c97544_g2_i2	MBG (6.3)
SKP1, CBF3D; S-phase kinase-associated protein 1	K03094	c93326_g2_i1	MBG (1.4)

	SKP1, CBF3D; S-phase kinase-associated protein 1	K03094	c93326_g2_i2		MBG (1.4)
	PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit	K07200	c99381_g1_i2		FBG (1.9)
	PRKAG; 5'-AMP-activated protein kinase, regulatory gamma subunit	K07200	c87222_g1_i1		MBG (4.8)
	SKP1, CBF3D; S-phase kinase-associated protein 1	K03094	c93326_g2_i6		MBG (1.6)
Circadian entrainment	GNAS; guanine nucleotide-binding protein G(s) subunit alpha	K04632	c102442_g1_i3	MBG (0.76)	MBG (0.7)
	GNG13; guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-13	K04547	c98518_g1_i1	FBG (1.04)	FBG (2.9)
	CALM; calmodulin	K02183	c108404_g1_i2	MBG (0.7)	
	PKA; protein kinase A [EC:2.7.11.11]	K04345	c94301_g2_i1	MBG (0.75)	FBG (1.2)
	CACNA1D; voltage-dependent calcium channel L type alpha-1D	K04851	c102695_g2_i16	MBG (1.1)	FBG (4.4)
	CALM; calmodulin	K02183	c92733_g1_i3	MBG (2.1)	MBG (3)
	PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	K07376	c107317_g1_i2	FBG (0.6)	MBG (0.8)
	ADCY2; adenylate cyclase 2 [EC:4.6.1.1]	K08042	c71860_g1_i2	MBG (1.53)	
	PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	K07376	c103658_g1_i3	FBG (1.7)	
	RYR2; ryanodine receptor 2	K04962	c109001_g1_i2	MBG (1.3)	
	RYR2; ryanodine receptor 2	K04962	c107689_g1_i1	MBG (1.2)	
	CALM; calmodulin	K02183	c108404_g1_i1	MBG (0.7)	FBG (3.2)
	CACNA1D; voltage-dependent calcium channel L type alpha-1D	K04851	c102695_g2_i6	MBG (0.9)	FBG (4)
	KG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	K07376	c103658_g1_i1	FBG (1.4)	
	CALM; calmodulin	K02183	c108404_g1_i3	MBG (0.7)	FBG (3.2)
	CACNA1D; voltage-dependent calcium channel L type alpha-1D	K04851	c102695_g2_i2	MBG (0.8)	FBG (4)
	KG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	K07376	c107317_g1_i3	FBG (0.6)	MBG (0.7)
	KG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	K07376	c103658_g1_i2	FBG (1.5)	
	CALM; calmodulin	K02183	c92733_g1_i6	MBG (2.4)	
	ADCY2; adenylate cyclase 2 [EC:4.6.1.1]	K08042	c74935_g1_i3	MBG (1)	
	CALM; calmodulin	K02183	c92733_g1_i4	MBG (2.1)	
	CALM; calmodulin	K02183	c92733_g1_i1	MBG (1.9)	MBG (3.1)
	RYR2; ryanodine receptor 2	K04962	c103458_g3_i2	MBG (1.4)	
MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]	K04371	c101078_g1_i6		FBG (0.4)	

MTNR1B; melatonin receptor type 1B	K04851	c102695_g2_i17	FBG (5.5)
NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	K13240	c96059_g3_i3	FBG (3)
ADCY1; adenylate cyclase 1 [EC:4.6.1.1]	K08041	c91549_g2_i2	FBG (1.7)
MSK1, RPS6KA5; ribosomal protein S6 kinase alpha-5 [EC:2.7.11.1]	K04445	c105901_g2_i2	FBG (2.4)
ADCY9; adenylate cyclase 9 [EC:4.6.1.1]	K08049	c94429_g1_i1	MBG (0.8)
PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	K07376	c107317_g1_i6	MBG (0.9)
PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	K05858	c108574_g1_i3	MBG (1.1)
PKA; protein kinase A [EC:2.7.11.11]	K04345	c97605_g1_i1	MBG (10.2)
ADCY10; adenylate cyclase 10 [EC:4.6.1.1]	K11265	c107118_g1_i6	MBG (10.9)
CALM; calmodulin	K02183	c108404_g1_i2	FBG (3)
PKA; protein kinase A [EC:2.7.11.11]	K04345	c105127_g3_i3	FBG (2.4)
ADCY1; adenylate cyclase 1 [EC:4.6.1.1]	K08041	c87575_g1_i4	FBG (2.4)
GUCY1B; guanylate cyclase soluble subunit beta [EC:4.6.1.2]	K12319	c97977_g1_i2	FBG (2.3)
ADCY1; adenylate cyclase 1 [EC:4.6.1.1]	K08041	c87575_g1_i6	FBG (2.6)
PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	K05858	c108574_g1_i1	MBG (0.9)
GNB1; guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	K04536	c68653_g1_i2	FBG (1.7)
MSK1, RPS6KA5; ribosomal protein S6 kinase alpha-5 [EC:2.7.11.1]	K04445	c105901_g2_i3	FBG (2.1)
MTNR1B; melatonin receptor type 1B	K04851	c102695_g2_i1	FBG (5)
ADCY1; adenylate cyclase 1 [EC:4.6.1.1]	K08041	c87575_g1_i1	FBG (3.7)
CALM; calmodulin	K02183	c103247_g2_i3	FBG (0.3)
GUCY1A; guanylate cyclase soluble subunit alpha [EC:4.6.1.2]	K12318	c99019_g1_i2	FBG (1.4)
MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]	K04371	c101078_g1_i5	FBG (0.9)
PER; period circadian protein	K02633	c108325_g1_i2	MBG (1)
NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	K13240	c96059_g3_i4	FBG (3)
PKA; protein kinase A [EC:2.7.11.11]	K04345	c97605_g1_i2	MBG (9.2)
NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	K13240	c96059_g3_i2	FBG (3.2)
ADCY1; adenylate cyclase 1 [EC:4.6.1.1]	K08041	c91549_g2_i3	FBG (1.7)
ADCY8; adenylate cyclase 8 [EC:4.6.1.1]	K08048	c102939_g1_i1	FBG (4)
MSK1, RPS6KA5; ribosomal protein S6 kinase alpha-5 [EC:2.7.11.1]	K04445	c105901_g2_i1	FBG (2.3)

GRIN2B; glutamate receptor ionotropic, NMDA 2B	K05210	c61974_g2_i1	FBG (1.8)
GNAO; G-ALPHA-O; guanine nucleotide-binding protein G(o) subunit alpha	K04534	c96645_g1_i1	FBG (2.6)
PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	K07376	c107317_g1_i1	MBG (0.7)
ADCY10; adenylate cyclase 10 [EC:4.6.1.1]	K11265	c99025_g1_i2	MBG (8.7)
MTNR1B; melatonin receptor type 1B	K04851	c102695_g2_i18	FBG (5.8)
ADCY1; adenylate cyclase 1 [EC:4.6.1.1]	K08041	c91549_g2_i4	FBG (1.9)
MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]	K04371	c101078_g1_i3	FBG (0.6)
ADCY9; adenylate cyclase 9 [EC:4.6.1.1]	K08049	c94429_g1_i3	FBG (0.7)
ADCY1; adenylate cyclase 1 [EC:4.6.1.1]	K08041	c87575_g1_i5	FBG (3)
NOS1; nitric-oxide synthase, brain [EC:1.14.13.39]	K13240	c96059_g3_i1	FBG (2.7)
GRIN2B; glutamate receptor ionotropic, NMDA 2B	K05210	c75896_g1_i1	FBG (1.8)
GNAI; guanine nucleotide-binding protein G(i) subunit alpha	K04630	c90417_g2_i1	FBG (1.7)
GNB5; guanine nucleotide-binding protein subunit beta-5	K04539	c106262_g1_i2	FBG (1.7)
GRIN1; glutamate receptor ionotropic, NMDA 1	K05208	c98719_g1_i2	FBG (7.2)
MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]	K04371	c101078_g1_i1	FBG (0.5)
MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]	K04371	c101078_g1_i7	FBG (0.7)
MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]	K04371	c101078_g1_i2	FBG (1)
ADCY9; adenylate cyclase 9 [EC:4.6.1.1]	K08049	c94429_g1_i6	MBG (0.7)
PRKG1; cGMP-dependent protein kinase 1 [EC:2.7.11.12]	K07376	c105127_g3_i1	FBG (2.5)
CAMK2; calcium/calmodulin-dependent protein kinase (CaM kinase) II [EC:2.7.11.17]	K04515	c105422_g2_i6	MBG (0.4)
ADCY1; adenylate cyclase 1 [EC:4.6.1.1]	K08041	c87575_g1_i3	FBG (1.5)
ADCY1; adenylate cyclase 1 [EC:4.6.1.1]	K08041	c91549_g2_i1	FBG (1.7)
MAPK1_3; mitogen-activated protein kinase 1/3 [EC:2.7.11.24]	K04371	c101078_g1_i4	FBG (0.8)
ADCY10; adenylate cyclase 10 [EC:4.6.1.1]	K11265	c99025_g1_i1	MBG (9.3)
PLCB; phosphatidylinositol phospholipase C, beta [EC:3.1.4.11]	K05858	c103398_g1_i3	FBG (1.8)
GRIN1; glutamate receptor ionotropic, NMDA 1	K05208	c98719_g1_i5	FBG (6.4)
GNAQ; guanine nucleotide-binding protein G(q) subunit alpha	K04634	c108928_g1_i2	FBG (1.2)
GRIN1; glutamate receptor ionotropic, NMDA 1	K05208	c98719_g1_i3	FBG (6.9)

Circadian rhythm-fly

TIM; timeless	K12074	c108981_g1_i3	MBG (0.6)	MBG (0.6)
TIM; timeless	K12074	c108981_g1_i2	MBG (0.6)	
ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1	K02296	c88570_g1_i1	FBG (0.87)	MBG (1.4)
ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1	K02296	c104661_g2_i1	FBG (0.7)	MBG (1.9)
ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1	K02296	c104661_g2_i2	FBG (0.86)	MBG (2)
ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1	K02296	c97544_g2_i1		MBG (4.9)
VRI; vrille	K12114	c1848_g1_i1		FBG (3.3)
GSK3B; glycogen synthase kinase 3 beta [EC:2.7.11.26]	K03083	c106317_g3_i6		MBG (3.5)
PER; period circadian protein	K02633	c108325_g1_i2		MBG (1)
HLF; hepatic leukemia factor	K09057	c105631_g3_i3		MBG (1)
VRI; vrille	K12114	c80157_g1_i3		FBG (2.6)
GSK3B; glycogen synthase kinase 3 beta [EC:2.7.11.26]	K03083	c106317_g3_i4		MBG (4.1)
CLOCK, KAT13D; circadian locomoter output cycles kaput protein [EC:2.3.1.48]	K02223	c88878_g2_i3		FBG (1)
ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1	K02296	c97544_g2_i2		MBG (6.3)
VRI; vrille	K12114	c80157_g1_i2		FBG (3.9)
GSK3B; glycogen synthase kinase 3 beta [EC:2.7.11.26]	K03083	c106317_g3_i8		MBG (4.9)
TIM; timeless	K12074	c84037_g1_i2		FBG (3.1)

5. CONCLUSÕES GERAIS

Um dos objetivos dessa tese de doutorado foi investigar a estrutura, composição molecular e processos evolutivos envolvidos na diferenciação de cromossomos sexuais em Orthoptera. Genomas de Orthoptera apresentam características peculiares como tamanho gigante e abundância em DNA repetitivo, que certamente representam um desafio para montagens de genomas. Portanto, nos Capítulos 1, 2 e 3, esses tópicos foram abordados mediante uma associação comparativa de dados citogenômicos e genômicos das famílias de DNA repetitivo mais abundantes dentro deste grupo de insetos, isolados principalmente mediante técnicas clássicas de biologia celular e molecular e mediante *reads* de genomas sequenciados. As espécies de grilos analisadas nesses capítulos apresentaram distintos sistemas sexuais, *Grillus assimilis* (X0), *Ciclotyloides americanus* (X₁X₂0) e *Eneoptera surinamensis* (neo-X₁X₂Y).

O aumento de tamanho genômico em grilos está relacionado com a abundância genômica de DNA repetitivo. Em *G. assimilis* o DNA repetitivo representa ~ 40% do genoma, enquanto que o genoma de *E. surinamensis* contém mais de 90% de DNA repetitivo. Por sua vez, satDNAs contribuíram grandemente para o aumento de tamanho genômico em *E. surinamensis* quando comparado com *G. assimilis*. Nas duas espécies, os DNAsat foram encontrados como repetições arranjadas em tandem tanto na heterocromatina considerada pobre em genes quanto na eucromatina considerada rica em genes. Considerando o tamanho genômico (*G. assimilis* 2,13 Gb e *E. surinamensis* 5,42 Gb) e a quantidade de DNAsat em cada espécie (*G. assimilis* 4% e *E. surinamensis* 14%), a quantidade de DNAsat foi incrementada cerca de 9 vezes no genoma de *E. surinamensis*. Interessantemente, *E. sunimanensis* apresenta um não usual alto número de 45 famílias de DNAsat. Enriquecimento de DNAsat em cromossomos sexuais em relação aos autossomos foi também observado para esta espécie. Além disso, um surpreendente acúmulo de número de loci de DNAsat foi observado no neo-Y altamente diferenciado, incluindo 39 satDNAs sobre representados neste cromossomo, seis sendo exclusivos, correspondendo a maior diversidade de satDNAs até agora relatada para cromossomos sexuais. A expansão diferencial de DNAsat entre os sexos foi corroborada por dados de qPCRs. Os resultados claramente demonstram que as doses de DNAsat diferem significativamente entre os sexos, com os machos tendo em média duas a dez vezes mais cópias do que as fêmeas. Os dados sugerem também que o neo-Y acumulou DNAsat após sua origem, causando seu aumento de tamanho em comparação com o neo-X₁ e neo-X₂.

Ressalta-se ainda que foi observado uma origem complexa para os cromossomos neo-sexuais em *E. surinamensis*, envolvendo mais de duas translocações simples como relatada em geral para os cromossomos neo- X_1X_2Y de Orthoptera (WHITE, 1973; HEWITT, 1979; PALACIOS-GIMENEZ et al., 2013). A evolução dos cromossomos sexuais de *E. surinamensis* provavelmente envolveu fusões em tandem e cêntricas assim como inversões, como documentado por exemplo em *Muntiacus muntjac* (Frönicke e Scherthan, 1997; Hartmann e Scherthan, 2004). Os múltiplos loci intersticiais de satDNAs no neo-Y de *E. surinamensis* podem representar remanescentes do material centromérico ancestral em locais de fusão cromossômica, ou simplesmente são consequência de expansão massiva de repetições de DNA que poderiam induzir novos rearranjos cromossômicos.

A ampla diversidade observada para os distintos sistemas sexuais nas espécies estudadas indica que, assim como em outros grupos de organismos, os padrões de evolução de sistemas sexuais em Orthoptera são complexos, ocorrendo distintas vezes ao longo de sua história evolutiva. Os resultados também reforçam a importância de rearranjos cromossômicos (fusões, fissões e inversões) como um mecanismo evolutivo primário para gerar neo-cromossomos sexuais em grilos, seguido de acúmulo de DNAs repetitivos que modificam a estrutura da cromatina. Putativos eventos de fissões, inversões, fusões assim como acúmulo de DNA repetitivo são hipotetizados e evidenciados para origem dos cromossomos sexuais X_1X_20 e neo- X_1X_2Y de *C. americanus* e *E. surinamensis*, respectivamente. Os eventos relatados, *de novo* origem de cromossomos sexuais mediante rearranjos cromossômicos, assim como acúmulo de DNA repetitivo que levaram a diferenciação entre cromossomos sexuais, foram documentados em outros grupos de insetos e vertebrados (BACHTROG et al., 2011, 2014; BEUKEBOON e PERRIN, 2014; PEICHEL, 2017). Por tanto, as evidências em grilos representam outro caso notável de convergência evolutiva devido a que cromossomos sexuais não relacionados compartilham muitas propriedades entre táxons distantes.

No Capítulo 4 da presente tese, os DNAsat encontrados em *G. assimilis* e *G. bimaculatus* (YOSHIMURA et al., 2006) foram testados quanto a transcrição diferencial entre diferentes tecidos, sexos e espécies de grilos (*Gryllus assimilis*, *G. bimaculatus*, *G. firmus* e *G. rubens*). O objetivo foi testar a possível conservação de satDNAs em grilos, e visar para possíveis papéis funcionais destas sequencias na regulação gênica, modulação da cromatina e como componente funcional de importantes estruturas como telômeros, centrômeros e cromossomos sexuais. Em

muitas espécies de insetos, os DNAsat são expressos ao longo do desenvolvimento e exibem diferenças de expressão entre os tecidos e entre os sexos (PEZER et al., 2011; UGARKOVIĆ, 2005; CAMACHO et al., 2015; DALÍKOVÁ et al., 2017). Foram encontradas evidências de que sete DNAsat incluindo aqueles isolados de *G. assimilis* (Gas1 Gas3, Gas5, Gas6-1, Gas8-2, Gas11) e aquele isolado em *G. bimaculatus* (GBH535) (YOSHIMURA et al., 2006) são compartilhados entre grilos, mas são transcritos diferencialmente em diferentes partes do corpo, bem como entre os sexos. Os dados sugerem ampla conservação de DNAsat entre espécies de grilos após a divergência da biblioteca de DNAsat possivelmente como consequência de sua funcionalidade. Três repetições (Gas3, Gas5 e Gas8-2) são comumente transcritas no genoma de *G. assimilis*, *G. firmus* e *G. rubens*. Considerando a localização cromossômica dessas repetições na heterocromatina constitutiva de *G. assimilis*, tais transcritos podem ser importantes para a formação de heterocromatina. Desta forma, os transcritos de DNAsat poderiam ser processados em pequenos siRNAs que participam da formação de heterocromatina e controle de expressão gênica (UGARKOVIĆ, 2005).

Em outros casos, foi observado que os DNAsat Gas mostram transcrição diferencial para diferentes tecidos entre espécies. No entanto, os tecidos estudados foram todos diferentes entre espécies e não posso especular sobre transcrição espécie-específica ou tecido-específico. Em *G. assimilis*, Gas1 está localizado em todos os centrômeros e no cromossomo X. Além disso, vários loci para Gas8-2 ocorreram em ambos os autossomos e cromossomo X. O Gas1 foi transcrito em testículos e ovários e nenhuma transcrição deste elemento foi vista em cabeças de machos ou fêmeas. Para Gas8-2 é destacada a transcrição diferencial nas gônadas provavelmente devido à alta especificidade dos tecidos estudados. Esses achados sugerem possível envolvimento de satDNAs (Gas1 e Gas8-2) durante a meiose masculina e feminina ou alguma relação para a maturação e função das gônadas. No entanto, outras evidências experimentais são necessárias para testar esta possibilidade. Em outros casos, quando os mesmos tecidos foram comparados, observei transcrição de GBH535 no FBSWFIF, mas não no FBLWFFM. Além disso, Gas5 é transcrito no LWFHFM mas não no LWFFM. Esse achado parece indicar alguma relação entre transcrição de satDNAs e habilidade de vôo em *G. firmus*, embora isto mereça validação experimental. Tendo em mente a grande diversidade de DNAsat e seus transcritos, vários sinais regulatórios específicos de sequência podem residir dentro deles, atuando como *barcode*, permitindo que a célula identifique territórios cromossômicos específicos. Por exemplo, esses sinais podem envolver

DNAs, RNAs ou proteínas, bem como estruturas secundárias ou terciárias de catálise mediada por RNA, como foi documentado no grilo de caverna (ROJAS et al., 2000) e besouro (FELICIELLO et al., 2015).

Por meio de análises *in silico* de satDNAs em *G. assimilis*, foi encontrado que sete deles são capazes de adotar estruturas secundárias com extensões de RNA de fita dupla. Tais estruturas secundárias podem determinar interações RNA-proteína sugerindo papéis funcionais. Transcritos de DNAsat capazes de adotar estruturas tipo *hummer-head* foram detectadas por exemplo em salamandras, esquistossomas e espécies do gênero *Dolichopoda* (ROJAS et al., 2000; UGARKOVIĆ, 2005; PEZER et al., 2011). Foi demonstrado também que essas estruturas tipo *hummer-head* podem funcionar como ribozimas com atividades de auto-clivagem, embora o papel fisiológico deles não parece ser claro (UGARKOVIĆ, 2005; PEZER et al., 2011). Considerando tais evidências é possível que o *foldings* ajude a dispersar os DNAsat dentro do genoma por meio do mecanismo de *rolling-circle replication*, em que o monômero circular resulta do processamento de RNA de estrutura secundária em monômeros lineares e, subsequentemente circularização por uma ligase de RNA específica do hospedeiro (Van Tole et al., 1991; Reid et al., 2000).

Estas descobertas sugerem que o grande repertório de DNAsat e seus transcritos podem ter relevância na organização e regulação de genomas, preparando o cenário para uma análise funcional adicional do genoma em *Gryllus*. Tendo em mente a transcrição de satélites, é plausível que os transcritos possam ser responsáveis pela modificação epigenética da cromatina e também possam ter efeito sobre a expressão gênica. Embora seja necessária uma pesquisa adicional nesta área, esse estudo estrutural e funcional fornece um passo importante para entender a biologia de DNAsat em Orthoptera, destacando a importância dos grilos como organismos modelo clássico para estudos evolutivos.

Por várias décadas, grilos de campo se tornaram modelos para estudos em ecologia comportamental, fisiologia e genética. Grilos apresentam características interessantes, como dimorfismo sexual pronunciado no tamanho do corpo, bem como para forma e tamanho das estruturas reprodutivas. Em vários grupos de eucariotas, machos e as fêmeas geralmente mostram diferenças fenotipicamente visíveis enquanto compartilham o mesmo genoma. Tipicamente, essas diferenças sexuais resultam da presença de genes enviesados do sexo, ou seja, genes diferencialmente expressos entre machos e fêmeas. Além disso, a expressão de SBG é considerada

como um resultado do antagonismo sexual que foi resolvido entre os sexos (PARSCH e ELLEGREN, 2013). No último capítulo dessa tese (Capítulo 5) foram explorados os transcriptomas de três espécies de grilos (*Gryllus assimilis*, *G. bimaculatus* e *G. firmus*) e a montagem genômica de uma espécie modelo para Orthoptera, *Locusta migratoria*. O primeiro objetivo foi buscar genes codificadores de proteínas relacionados com a determinação sexual, genes codificadores de proteínas envolvidos com o *fitness* reprodutivo (biossíntese de hormônio de insetos, ritmo circadiano, ritmo circadiano *fly*) e finalmente, genes enviesados do sexo. O segundo objetivo foi tentar elucidar de uma maneira comparativa os fatores evolutivos atuando em aqueles loci.

Usando *G. assimilis* como modelo para estudos de genes enviesados do sexo, foi encontrado um conjunto de genes altamente expressos no macho quando comparados com fêmeas. Uma interpretação plausível para estes achados seria conflito sexual resolvido entre os sexos com o fim de alcançar níveis de expressão ótimo em machos. A presença maior de SBG em machos poderia explicar as diferenças fenotípicas entre os sexos. Esse trabalho também possibilitou a comparação de genes determinantes do sexo e outros genes relacionados à aptidão sexual, como a biossíntese de hormônios de insetos, ritmo circadiano e ritmo circadiano *fly*, entre as espécies de *Gryllus* e *L. migratoria*. Os genes efetores e os alvos das vias de determinação do sexo foram previamente identificados em outros insetos, mas nunca em Orthoptera. Os resultados mostraram que os genes determinantes do sexo são altamente conservados em grilos, indicando que a determinação sexual poderia ser uma característica conservada dentro do gênero.

Outro achado relevante foi que a taxa de divergência de genes codificadores de proteínas envolvidas na determinação do sexo e no *fitness* reprodutivo foi significativamente maior comparado com os genes *housekeeping*. Esta observação apoia a hipótese que genes codificadores de proteínas envolvidas na determinação sexual e na aptidão reprodutiva tendem a evoluir mais rapidamente que genes não envolvidos com a reprodução. Foi demonstrado que a rápida evolução destas sequências resulta de uma forte seleção positiva nestes loci. Finalmente, foi documentado que os genomas das espécies estudadas apresentam níveis altos de duplicações gênicas. As descobertas sugerem que as duplicações gênicas podem desempenhar um papel relevante na evolução e divergência de genomas, e particularmente na expressão de genes enviesados do sexo em *G. assimilis*, uma espécie que nos seguintes anos provavelmente irá fornecer informações sobre genômica funcional e epigenética da determinação do sexo.

6. PERSPECTIVAS FUTURAS

Espécies de eucariotos apresentam diferenças dramáticas em número e estrutura de cromossomos. Os pioneiros biólogos evolutivos pensaram que essas diferenças no cariótipo dentro e entre as espécies eram fatores de adaptação e especiação, uma ideia que está atualmente despertando interesse científico. No entanto, sabemos pouco sobre os mecanismos moleculares subjacentes a evolução das mudanças no cariótipo e temos poucas evidências diretas para apoiar um papel dessas mudanças no cariótipo para adaptação e especiação. Seria interessante usar a diversidade cromossômica presente em grilos para identificar em detalhe os mecanismos moleculares e as consequências evolutivas das alterações nos cromossomos sexuais, no número de cromossomos e nos centrômeros.

Uma das diferenças mais óbvias no cariótipo dentro das espécies de grilos é a presença de cromossomos sexuais. Apesar de muitos trabalhos teóricos e empíricos para explicar a evolução convergente dos cromossomos sexuais nos eucariotas, ainda não temos uma visão molecular detalhada de como os cromossomos sexuais evoluem neste grupo. Também temos poucos testes empíricos claros de hipóteses sobre o motivo pelo qual os cromossomos sexuais evoluem. Abordar essas questões fundamentais para nosso conhecimento requer estudos de cromossomos sexuais em diferentes estágios de diferenciação. A diversidade de cromossomos sexuais de Orthoptera pode ser usada para explorar e determinar os fatores moleculares e evolutivos que promovem transições entre cromossomos sexuais homomórficos e heteromórficos. Um passo importante em Orthoptera seria identificar primeiro o conteúdo gênico dos cromossomos sexuais. Posteriormente, determinar os mecanismos que contribuem para a manutenção da dosagem de genes em cromossomos sexuais degenerados o não degenerado.

Trabalhos teóricos sugerem que as regiões cromossômicas com recombinação reprimida, como por exemplo cromossomos sexuais, inversões e fusões cromossômicas, provavelmente abriguem loci importantes para a adaptação e especiação. Em consonância com estas previsões, trabalhos empíricos recentes demonstraram que *clusters* gênicos importantes para a adaptação e especiação geralmente são mapeados em regiões genômicas de baixa recombinação. No entanto, não entendemos como esses *clusters* genômicos foram construídos através da evolução. Eles são criados por fusões cromossômicas inteiras, ou por transposições de genes únicos? Ou, esses *clusters* pré-existentes em que mecanismos para suprimir a recombinação, como inversões,

evoluíram mais tarde? Construir e comparar montagens de genomas inteiros para identificar a história evolutiva e os mecanismos moleculares subjacentes à construção de *hotspots* genômicos conhecidos de adaptação seria tudo um desafio em grilos, um grupo extraordinariamente diverso com mais de 90 espécies distribuídas em diversos ambientes ao redor do mundo.

REFERÊNCIAS BIBLIOGRÁFICAS

- AMREIN, H.; GORMAN, M.; NOTHIGER, R. The sex-determining gene *tra-2* of *Drosophila* encodes a putative RNA binding protein. **Cell**, v. 55, n. 2, p. 1025-1035, 1988.
- ANDRÉS, J.A.; LARSON, E.R.; BOGDANOWICZ, S.M.; HARRISON, R.G. Patterns of transcriptome divergence in the male accessory gland of two closely related species of field crickets. **Genetics**, v. 193, n. 3, p. 501-513, 2013.
- ANDRÉS, J.A.; MAROJA, L.S.; BOGDANOWICZ, S.M.; SWANSON, W.J.; HARRISON, R. G. Molecular evolution of seminal proteins in field crickets. **Molecular Biology and Evolution**, v. 23, n. 2, p. 1574-1584, 2006.
- ANDRÉS, J.A.; MAROJA, L.S.; HARRISON, R.G. Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. **Proceedings Biological Sciences.**, v. 275, n. 275, p. 1975-1983, 2008.
- ANDREWS, S. FastQC, A quality control tool for high throughput sequence data <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. 2012.
- ANJOS, A.; RUIZ-RUANO, F.J.; CAMACHO, J.P.M.; LORETO, V.; CABRERO, J.; DE SOUZA M.J.; CABRAL-DE-MELLO, D.C. U1 snDNA clusters in grasshoppers: chromosomal dynamics and genomic organization. **Heredity**, v. 114, n. 2, p. 207-219, 2015.
- AUGÉ-GOUILLOU, C.; BIGOT, Y.; POLLER, N.; HAMELIN, M.H.; MEUNIER-ROTIVAL, M.; PERIQUET, G. Human and other mammalian genomes contain transposons of the mariner family. **FEBS Letters**, v. 368, n. 3, p. 541-546, 1995.
- AYLING, L.J.; GRIFFIN, D.K. The evolution of the sex chromosomes. **Cytogenetic and Genome Research**, v. 99, n. 1-4, p.125-140, 2002.
- BACHTROG, D.; KIRKPATRICK, M.; MANK, J.E.; MCDANIEL, S.F.; PIRES, J.C.; RICE, W.; VALENZUELA, N. Are all sex chromosomes created equal? **Trends in Genetics**, v. 27, n. 9, p. 350-7, 2011.
- BACHTROG, D.; MANK, J.E.; PEICHEL, C.L.; KIRKPATRICK, M.; OTTO, S.P; et al. Sex determination: why so many ways of doing it? **PLoS Biology**, v. 12, n. 7, p. e1001899, 2014.
- BARZOTTI, R.; PELLICCIA, F.; ROCCHI, A. Identification and characterization of U1 small nuclear RNA genes from two crustacean isopod species. **Chromosome Research**, v. 11, n. 4, p. 365-373, 2003.
- BENSON, G. Tandem repeats finder: a program to analyze DNA sequences. **Nucleic Acids Research**, v. 27, n. 2, p. 573-580, 1999.

- BERGERO, R.; CHARLESWORTH, D. The evolution of restricted recombination in sex chromosomes. **Trends in Ecology and Evolution**, v. 24, n. 2, p. 94-102, 2008.
- BERGERO, R.; QIU, S.; CHARLESWORTH, D. Gene loss from a plant sex chromosome system. **Current Biology**, v. 25, n. 9, p. 1234-1240, 2015.
- BERGERO, R.; QIU, S.; FORREST, A.; BORTHWICK, H.; CHARLESWORTH, D. Expansion of the pseudo-autosomal region and ongoing recombination suppression in the *Silene latifolia* sex chromosomes. **Genetics**, v. 194, n. 3, p. 673-686, 2013.
- BETRÁN, E.; THORNTON, K.; LONG, M. Retroposed new genes out of the X in *Drosophila*. **Genome Research**, v. 12, n. 12, p. 1854-1859, 2002.
- BEUKEBOON, L.W.; PERRIN, N. The Evolution of Sex Determination. Oxford University Press. 2014.
- BIDAU, C.J.; MARTI, D.A. Meiosis and the Neo-XY system of *Dichroplus vittatus* (Melanoplinae, Acrididae): a comparison between sexes. **Genetica**, v. 110, n. 2, p. 185-194, 2000.
- BIÉMONT, C. A brief history of the status of transposable elements: from junk DNA to major players in evolution. **Genetics**, v. 186, n. 4, p. 1085-1093, 2010.
- BIÉMONT, C.; VIEIRA, C. Genetics: junk DNA as an evolutionary force. **Nature**, v. 443, n. 7111, p. 521-524, 2006.
- BISCOTTI, M.A.; CANAPA, A.; FORCONI, M.; OLMO, E.; BARUCCA, M. Transcription of tandemly repetitive DNA: functional roles. **Chromosome Research**, v. 23, n. 3, p. 463-477, 2015.
- BÖHNE, A.; BRUNET, F.; GALIANA-ARNOUX, D.; SCHULTHEIS, C.; VOLFF, J.N. Transposable elements as drivers of genomic and biological diversity in vertebrates. **Chromosome Research**, v. 16, n. 1, p. 203-215, 2008.
- BOLGER, A.M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114-2120, 2014.
- 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, n. 13, p. 3509-3516, 1986.
- BUENO, D.; PALACIOS-GIMENEZ, O.M.; CABRAL-DE-MELLO, D.C. Chromosomal mapping of repetitive DNAs in the grasshopper *Abracris flavolineata* reveal possible ancestry of the B chromosome and H3 histone spreading. **PLoS ONE**, v. 8, n. 6, p. e66532, 2013.
- BUENO, D.; PALACIOS-GIMENEZ, O.M.; MARTÍ, D.A.; MARIGUELA, T.C.; CABRAL-DE-MELLO, D.C. The 5S rDNA in two *Abracris* grasshoppers (Ommatolampidinae: Acrididae):

molecular and chromosomal organization. **Molecular Genetics and Genomics**, v. 291, n. 4, p. 1607-1613, 2016.

BULL, J.J. Evolution of sex determining mechanisms. The Benjamin/Cummings Publishing Company, Inc., 1983.

BUSSIÈRE, L.F.; HUNT, J.; JENNIONS, M.D.; BROOKS, R. Sexual conflict and cryptic female choice in the black field cricket, *Teleogryllus commodus*. **Evolution**, v. 60, n. 4, p. 792-800, 2006.

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, n. 2, p. 279-292, 2012.

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, n. 4, 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, n. 2, p. 279-292, 2012.

CAMACHO, J.P.M.; RUIZ-RUANO, F.J.; MARTÍN-BLÁZQUEZ, R.; LÓPEZ-LEÓN, M.D.; CABRERO, J.; LORITE, P.; CABRAL-DE-MELLO, D.C.; et al. A step to the gigantic genome of the desert locust: chromosome sizes and repeated DNAs. **Chromosoma**, v. 124, n. 2, p. 263-275, 2015.

CARDOSO, H.; DUTRA, A. The Neo-X Neo-Y sex pair in Acrididae, its structure and association. **Chromosoma**, v. 70, n. 3, p. 323-336, 1979.

CARVALHO, A.B. Origin and evolution of the *Drosophila* Y chromosome. **Current Opinion in Genetics and Development**, v. 12, n. 6, p. 664-668, 2002.

CARVALHO, A.B.; CLARCK, A.W. Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. **Genome Research**, v. 23, n. 11, p. 1894-1907, 2013.

CASTILLO, E.R.; MARTI, D.A.; BIDAU, C.J. Sex and neo-sex chromosomes in Orthoptera: a review. **Journal of Orthoptera Research**, v. 19, n. 2, p. 213-231, 2010b.

CASTILLO, E.R.D.; BIDAU, C.J.; MARTÍ, D.A. Neo-sex chromosome diversity in Neotropical melanopline grasshoppers (Melanoplineae, Acrididae). **Genetica**, v. 138, n. 7, p. 775-786, 2010a.

CASTILLO, E.R.D.; TAFFAREL, A.; MARIOTTINI, Y.; FERNANDEZ-ARHEX, V.; MARTÍ, D.A.; BIDAU, C.J. Neo-sex chromosomes in the *Maculipennis* species group (*Dichroplus*: Acrididae, Melanoplineae): the cases of *D. maculipennis* and *D. vittigerum*. **Zoological Science**, v. 33, n. 3, p. 303-310, 2016.

- CASTILLO, E.R.D.; TAFFAREL, A.; MARTÍ, D.A. The early evolutionary history of neo-sex chromosomes in Neotropical grasshoppers, *Boliviacris noroestensis* (Orthoptera: Acrididae: Melanoplinae). **European Journal of Entomology**, v. 111, n. 3, p. 321-327, 2014.
- CATALÁN, A.; HUTTER, S.; PARSCH, J. Population and sex differences in *Drosophila melanogaster* brain gene expression. **BMC Genomics**, v. 13, p. 654, 2012.
- CHANG, P.L.; DUNHAM, J.P.; NUZHIDIN, S.V.; ARBEITMAN, M.N. Somatic sex-specific transcriptome differences in *Drosophila* revealed by whole transcriptome sequencing. **BMC Genomics**, v. 12, p. 364, 2011.
- CHARLESWORTH, B. The evolution of chromosomal sex determination and dosage compensation. **Current Biology**, v. 6, n. 2, p. 149-162, 1996.
- CHARLESWORTH, B.; COYNE, J.B.; BARTON, N.H. The relative rates of evolution of sex-chromosomes and autosomes. **American Naturalist**, v. 130, p. 113-146, 1987.
- CHARLESWORTH, B.; SNIEGOWSKI, P.; STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature**, v. 371, p. 215-220, 1994.
- CHARLESWORTH, D.; CHARLESWORTH, B.; MARAIS, G. Steps in the evolution of heteromorphic sex chromosomes. **Heredity**, v. 95, n. 2, p. 118-128, 2005.
- CHILDS, G.; NOCENTE-MCGRATH, C.; LIEBER, T.; HOLT, C.; KNOWLES, J.A. Sea urchin (*Lytechinus pictus*) late-stage histone H3 and H4 genes: characterization and mapping of a clustered but nontandemly linked multigene family. **Cell**, v. 31, n. 2 Pt 1, p. 383-93, 1982.
- CHINTAUAN-MARQUIER, I.C.; LEGENDRE, F.; HUGEL, S.; ROBILLARD, T.; GRANDCOLAS, P.; NEL, A.; ZUCCON, D.; DESUTTER-GRANDCOLAS, L. Laying the foundations of evolutionary and systematic studies in crickets (Insecta, Orthoptera): a multilocus phylogenetic analysis. **Cladistics**, v. 32, n. 1, p. 54-81, 2016.
- CIGLIANO, M.M.; LANGE, C.E. Orthoptera. In: Morrone, J.J.; Coscarón, S. (Eds). Biodiversidad de Artrópodos Argentinos. Una perspectiva biotaxonómica. Ediciones Sur, La Plata, Argentina, p. 67-83, 1998.
- CLARK N.J.; SWANSON, W.J. Pervasive adaptive evolution in primate seminal proteins. **PLoS Genetics**, v. 1, n. 3, p. e35, 2005.
- COHN, R.H.; KEDES, L.H. Nonallelic histone gene clusters of individual sea urchins (*Lytechinus pictus*): mapping of homologies in coding and spacer DNA. **Cell**, v. 18, n. 3, p. 855-64, 1979a.
- COHN, R.H.; KEDES, L.H. Nonallelic histone gene clusters of individual sea urchins (*Lytechinus pictus*): polarity and gene organization. **Cell**, v. 18, n. 3, p. 843-53, 1979b.

COLGAN, D.J. MCLAUCHLA, G.D.; WILSON, F.; LIVINGSTON, S.P.; EDGEcombe, G.D.; MACARANAS, J.; CASSIS, G.; GRAY, M.R. Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. **Australian Journal of Zoology**, v. 46, n. 5, p. 419-437, 1998.

DALÍKOVÁ, M.; ZRZAVÁ, M.; KUBÍCKOVÁ, S.; MAREC, F. W-enriched satellite sequence in the Indian meal moth, *Plodia interpunctella* (Lepidoptera, Pyralidae). **Chromosome Research**, 2017, doi 10.1007/s10577-017-9558-8.

DANBARA, Y.; SAKAMOTO, T.; URYU, O.; TOMIOKA, K. RNA interference of timeless gene does not disrupt circadian locomotor rhythms in the cricket *Gryllus bimaculatus*. **Journal of Insect Physiology**, v. 56, n. 12, p. 1738-1745, 2010.

DE BARROS, A.V.; SCZEPANSKI, T.S.; CABRERO, J.; CAMACHO, J.P.M.; VICARI, M.R.; ARTONI, R.F. Fiber FISH reveals different patterns of high-resolution physical mapping for repetitive DNA in fish. **Aquaculture**, v. 322-323, p. 47-50, 2001.

DIAS, G.B.; HERINGER, P.; KUHN, G.C.S. Helitrons in *Drosophila*: chromatin modulation and tandem insertions. **Mobile Genetic Elements**, v. 6, n. 2, p. e1154638, 2016.

DIAS, G.B.; HERINGER, P.; SVARTMAN, M.; KUHN, G.C.S. Helitrons shaping the genomic architecture of *Drosophila*: enrichment of DINE-TR1 in α and β -heterochromatin, satellite DNA emergence and piRNA expression. **Chromosome Research**, v. 23, n. 3, p. 597-613, 2015.

DIAS, G.B.; SVARTMAN, M.; DELPRAT, A.; RUIZ, A.; KUHN, G.C.S. Tetris is a foldback transposon that provided the building blocks for an emerging satellite DNA of *Drosophila virilis*. **Genome Biology and Evolution**, v. 6, n. 6, p. 1302-1313, 2014.

DÍAZ, M.O.; SÁEZ, F.A. DNA synthesis in the neo-X neo-Y sex determination system of *Dichroplus bergi* (Orthoptera: Acrididae). **Chromosoma**, v. 24, n. 1, p. 10-16, 1968.

DOOLITTLE, W.F.; SAPIENZA, C. Selfish genes, the phenotype paradigm and genome evolution. **Nature**, v. 284, n. 5757, p. 601-603, 1980.

EDGAR, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792-1797, 2004.

ELLEGREN, H.; PARSCH, J. The evolution of sex-biased genes and sex-biased gene expression. **Nature Review Genetics**, v. 8, p. 689-698, 2007.

ELLISON, C.K.; WILEY, C.; SHAW, K.L. The genetics of speciation: genes of small effect underlie sexual isolation in the Hawaiian cricket *Laupala*. **Journal of Evolutionary Biology**, v. 24, n. 5, p. 1110-1119, 2011.

- EZAZ, T.; STIGLEC, R.; VEYRUNES, F.; GRAVES, J.A.M. Relationships between vertebrate ZW and XY sex chromosome systems. **Current Biology**, v. 16, n. 17, p. 736-743, 2006.
- FAGEGALTIER, D.; BOUGÉ, A.L.; BERRY, B.; POISOT, E.; SISMEIRO, O.; COPPÉE, J.Y.; et al. The endogenous siRNA pathway is involved in heterochromatin formation in *Drosophila*. **Proceedings of National Academy of Sciences of the United States of America**, v. 106, n. 50, p. 21258-21263, 2009.
- FEDORKA, K.M.; MOUSSEAU, T.A. Female mating bias results in conflicting sex-specific offspring fitness. **Nature**, v. 429, n. 6987, 65-67, 2004.
- FELICIELLO, I.; AKRAP, I.; UGARKOVIĆ, D. Satellite DNA modulates gene expression in the beetle *Tribolium castaneum* after heat stress. **PLoS Genetics**, v. 11, n. 8, p. e1005547, 2015.
- FERREIRA, A.; CELLA, D.M. Chromosome structure of *Eneoptera surinamensis* (Orthoptera, Grylloidea, Eneopterinae) as revealed by C, NOR and N banding techniques. **Chromosome Science**, v. 9, n. 2, p. 47-51, 2006.
- FERREIRA, A.; MESA, A. Cytotaxonomy of the genus *Dichromatos* Cigliano 2007 (Orthoptera, Acridoidea, Melanoplinae). **Journal of Orthoptera Research**, v. 19, n. 2, 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.
- FESCHOTTE, C.; PRITHAM, E.J. DNA transposons and the evolution of eukaryotic genomes. **Annual Review of Genetics**, v. 41, p. 331-368, 2007.
- FESCHOTTE, C.; WESSLER, S.R. Mariner-like transposases are widespread and diverse in flowering plants. **Proceedings of National Academy of Sciences of the United States of America**, v. 99, n. 1, p. 280-285, 2002.
- FLOT, J. F. et al. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. **Nature**, v. 500, p. 453-457, 2013.
- FRÖNICKE, L.; SCHERTHAN, H. Zoo-fluorescence *in situ* hybridization analysis of human and Indian muntjac karyotypes (*Muntiacus muntjak vaginalis*) reveals satellite DNA clusters at the margins of conserved syntenic segments. **Chromosome Research**, v. 5, v. 4, p. 254-261, 1997.
- FRY, J.D. 2010. The genomic location of sexually antagonistic variation: some cautionary comments. **Evolution**, v. 64, n. 5, p. 1510-1516, 2010.
- GALL, J.G.; PARDUE, M. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. **Proceedings of National Academy of Sciences of the United States of America**, v. 63, n. 2, p. 378-383, 1969.
- GALLACH, M.; BETRÁN, E. Dosage compensation and the distribution of sex-biased gene expression in *Drosophila*: considerations and genomic constraints. **Journal of Molecular**

Evolution, v. 82, n. 4-5, p. 199-206, 2016.

GENTLEMAN, R.C.; CAREY, V.J.; BATES, D.M.; BOLSTAD, B.; DETTLING, M.; DUDOIT, S.; ELLIS, B. Bioconductor: open software development for computational biology and bioinformatics. **Genome Biology**, v. 5, n. 10, p. R80, 2004.

GILBERT SF. *Developmental Biology*. 6th edition. Sinauer Associates. Sunderland, Massachusetts. 2000.

GORDON, A.; HANNON, G.J. Fastx-toolkit, FASTQ/A short-reads pre-processing tools, Unpublished http://hannonlab.cshl.edu/fastx_toolkit. 2010.

GRAVES, J.A.M. Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. **Annual Review of Genetics**, v. 42, p. 565-586, 2008.

GREENFIELD, MD. Acoustic communication in Orthoptera. In: Gangwere, S.K., Muralirangan, M.C., Muralirangan, M. (Eds.), *The Bionomics of Grasshoppers, Katydid and Their Kin*. CAB International, Wallingford, p. 197-230. 1997.

GREGORY, T.R. *The Evolution of the Genome*. 1st ed. Elsevier Academic Press. 2004.

GREWAL, S.I.; ELGIN, S.C. Transcription and RNA interference in the formation of heterochromatin. **Nature**, v. 447, n. 339-406, p. 399-406, 2007.

HAAS, B.J.; PAPANICOLAOU, A.; YASSOUR, M.; GRABHERR, M.; BLOOD, P.D.; BOWDEN, J.; COUGER, M.T.; et al. *De novo* transcript sequence reconstruction from RNA-seq: reference generation and analysis with Trinity. **Nature Protocols**, v. 8, n. 8, p. 1-8, 2014.

HAMADA, M.; KIRYU, H.; SATO, K.; MITUYAMA, T.; ASAI, K. Prediction of RNA secondary structure using generalized centroid estimators. **Bioinformatics**, v. 25, n. 4, p. 465-473, 2009.

HARTFELDER, K.; EMLÉN, D.J. Endocrine control of insect polyphenism. In: Gilbert LI, editor, *Insect Endocrinology*, Elsevier, p. 464-522. 2011.

HARTMANN, N.; SCHERTHAN, H. Characterization of ancestral chromosome fusion points in the Indian muntjac deer. **Chromosoma**, v. 112, v. 5, p. 213-220, 2004.

HENSE, W.; BAINES, J.F.; PARSCH, J. X chromosome inactivation during *Drosophila* spermatogenesis. **PLoS Biology**, v. 5, n. 10, p. e273, 2007.

HEWITT, G.H. Orthoptera. Grasshoppers and Crickets. In: *Animal Cytogenetics 3. Insecta 1*, (Eds) by B. John. Gerbrüder Borntraeger: Berlin-Suttgart, pp 1-170. 1979.

HOLMQUIST, G.P.; DANCIS, B. Telomere replication, kinetochore organizers, and satellite DNA evolution. **Proceedings of National Academy of Sciences of the United States of**

America, v. 76, n. 9, p. 4566-4570, 1979.

HOLT, C.A.; CHILDS, G. A new family of tandem repetitive early histone genes in the sea urchin *Lytechinus pictus*: evidence for concerted evolution within tandem arrays. **Nucleic Acids Research**, v. 12, n. 16, p. 6455-71, 1984.

HOWARD, D.J.; MARSHALL, J.L.; HAMPTON, D.D.; BRITCH, S.C.; DRANEY, M.L.; et al. The genetics of reproductive isolation: a retrospective and prospective look with comments on ground crickets. **American Naturalist**, v. 3, n. S3, p. S8-S21, 2002.

HSU, T.C.; PATHAK, S.; BASEN, B.M.; STAHL, G.J. Induced Robertsonian fusions and tandem translocations in mammalian cell cultures. **Cytogenetics and Cell Genetics**, v. 21, n. 1-2, 86-98, 1978.

HUYLMANS, A.K.; PARSCH, J. Variation in the X:autosome distribution of male-biased genes among *Drosophila melanogaster* tissues and its relationship with dosage compensation. **Genome Biology and Evolution**, v. 7, n. 7, p. 1960-1971, 2015.

IJDO, J.W.; WELLS, R.A.; BALDINI, A.; REEDERS, S.T. Improved telomere detection using a telomere repeat probe (TTAGGG)_n generated by PCR. **Nucleic Acids Research**, v. 19, n. 17, p. 4780, 1991.

INGRAM, V.M. Gene evolution and the haemoglobins. **Nature**, v. 189, p. 704-708, 1961.

INNOCENTI, P.; MORROW, E.H. The sexually antagonistic genes of *Drosophila melanogaster*. **PLoS Biology**, v. 8, n. 3, p. e1000335, 2010.

JEONG, H.; MASON, S.P.; BARABÁSI, A.L.; OTTAVAI, Z.N. Lethality and centrality in protein networks. **Nature**, v. 411, n. 6833, p. 41-42, 2001.

JOHN, H.A.; BIRNSTIEL, M.L.; JONES, K. RNA-DNA hybrids at the cytological level. **Nature**, v. 223, n. 5206, p. 582-587, 1969.

JUNIER, T.; PAGNI, M. Dotlet: diagonal plots in a web browser. **Bioinformatics**, v. 6, n. 2, p. 178-179, 2000.

KAISER, V.B.; BACHTROG, D. Evolution of sex chromosomes in insects. **Annual Review of Genetics**, v. 44, p. 91-112, 2010.

KANEHISA, M.; GOTO, S. KEGG: Kyoto encyclopedia of genes and genomes. **Nucleic Acids Research**, v. 28, n. 1, p. 27-30, 2000.

KAPITONOV, V.V.; JURKA, J. A universal classification of eukaryotic transposable elements implemented in Repbase. **Nature Review Genetics**, v. 9, n. 5, p. 411-412, 2008.

KASHIMADA, K.; KOOPMAN, P. *Sry*: the master switch in mammalian sex determination.

Development, v. 137, n. 23, p. 3921-3930, 2010.

KAZAZIAN, H.H.Jr. Mobile elements: drivers of genome evolution. **Science**, v. 303, n. 5664, p. 1625-1632, 2004.

KEJNOVSKY, E. MICHALOVOVA, M.; STEFLOVA, P.; KEJNOVSKA, I.; MANZANO, S.; HOBZA R.; KUBA, Z.; KOVARIK, J.; JAMILENA, M.; VYSKOT, B. Expansion of microsatellites on evolutionary young Y chromosome. **PLoS ONE**, v. 8, n. 1, p. e45519, 2013.

KEJNOVSKY, E.; HOBZA R.; CERMAK, T.; KUBAT, Z.; VYSKOT, B. The role of repetitive DNA in structure and evolution of sex chromosomes in plants. **Heredity**, v. 102, n. 6, p. 533-541, 2009.

KEMKEMER, C.; CATALÁN, A.; PARSCH, J. “Escaping” the X chromosome leads to increased gene expression in the male germline of *Drosophila melanogaster*. **Heredity**, v. 112, n. 2, p. 149-155, 2014.

KEMKEMER, C.; HENSE, W.; PARSCH, J. Fine-scale analysis of X chromosome inactivation in the male germline of *Drosophila melanogaster*. **Molecular Biology and Evolution**, v. 28, n. 5, p. 1561-1563, 2011.

KERN, A.D; JONES, C.D.; BEGUN, D.J. Molecular genetics of male accessory gland in the *Drosophila simulans* complex. **Genetics**, v. 167, n. 2, p. 725-35, 2004.

KIDWELL, M.G.; LISCH, D. Transposable elements as sources of variation in animals and plants. **Proceedings of National Academy of Sciences of the United States of America**, v. 94, n. 15, p. 77047711, 1997.

KING, K.; JOBST, J; HEMLEBEN, V. Differential homogenization and amplification of two satellite DNAs in the genus *Cucurbita* (Cucurbitaceae). **Journal of Molecular Evolution**, v. 41, n. 6, p. 996-1005, 1995.

KIRKPATRICK, K. How and why chromosome inversions evolve. **PLoS Biology**, v. 8, n. 9, p. e1000501, 2010.

KONG, A.; THORLEIFSSON, G.; GUDBJARTSSON, D.F.; MASSON, G.; SIGURDSSON, A.; JONASDOTTIR, A.; WALTERS, G.B.; et al. Fine-scale recombination rate differences between sexes, populations and individuals. **Nature**, v. 467, 1099-1103, 2010.

KORPELAINEN, H. Sex-ratios and conditions required for environmental sex determination in animals. **Biological Reviews of the Cambridge Philosophical Society**, v. 65, n. 2, p. 147-184, 1990.

KRAMEROV, D.A.; VASSETZKY, N.S. Origin and evolution of SINEs in eukaryotic genomes. **Heredity**, v. 107, n. 6, p. 487-495, 2011.

- KUHN, G.C.S.; KÜTTLER, H.; MOREIRA-FILHO, O.; HESLOP-HARRISON, J.S. The 1.688 repetitive DNA of *Drosophila*: concerted evolution at different genomic scales and association with genes. **Molecular Biology and Evolution**, v. 29, n. 1, p. 7-11, 2012.
- KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. **Molecular Biology and Evolution**, v. 33, n. 7, p. 1870-4, 2016.
- LAHN, B.T.; PAGE, D.C. Four evolutionary strata on the human X chromosome. **Science**, v. 286, n. 5441, p. 964-967, 1999.
- LANDE, R. Models of speciation by sexual selection on polygenic traits. **Proceedings of National Academy of Sciences of the United States of America**, v. 78, n. 6, p. 3721-3725, 1981.
- LANGIN, T.; CAPY, P.; DABOUSSI, M.J. The transposable element impala, a fungal member of the Tc1-mariner superfamily. **Molecular and General Genetics**, v. 246, n. 1, p. 19-28, 1995.
- LANGMEAD, B.; SALZBERG, S.L. Fast gapped-read alignment with Bowtie2. **Nature Methods**, v. 9, n. 4, p. 357-359, 2012.
- LEMOS, B.; ARARIPE, L.; O. HARTL, D.L. Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. **Science**, v. 319, n. 5859, p. 91-93, 2008.
- LEMOS, B.; BRANCO, A.T.; HARTL, D.L. Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. **Proceedings of National Academy of Sciences of the United States of America**, v. 107, n. 36, p. 15826-15831, 2010.
- LENORMAND, T.; DUTHEIL, J. Recombination difference between sexes: A role for haploid selection. **PLoS Biology**, v. 3, n. 3, p. e63, 2005.
- LEVINE, M.T.; JONES, C.D.; KERN, A.D.; LINDFORS, H.A.; BEGUN, D.J. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. **Proceedings of National Academy of Sciences of the United States of America**, v. 103, n. 26, p. 9935-9939, 2006.
- LEWIN, B. Genes VIII: And Molecular Biology of the Gene. Published by Pearson Prentice Hall. Pearson Education, Inc. Upper Saddle River, NT 07458. 2004.
- LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; et al. The sequence alignment/map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078-2079, 2009.
- LIAO, D.; PAVELITZ, T.; WEINER, A.M. Characterization of a novel class of interspersed LTR elements in primate genomes: structure, genomic distribution, and evolution. **Journal of Molecular Evolution**, v. 46, n. 6, p. 649-60, 1998.

- LIFSCHYTZ, E.; LINDSLEY, D.L. The role of X-chromosome inactivation during spermatogenesis. **Proceedings of National Academy of Sciences of the United States of America**, v. 69, n. 1, p. 182-186, 1972.
- LONG, E.O.; DAWID, I.B. Repeated genes in eukaryotes. *Annual Review of Biochemistry*, v. 49, p. 727-764, 1980.
- LOPES, D.M.; CARVALHO, C.R.; CLARINDO, W.R.; PRAÇA, M.M.; TAVARES, M.G. Genome size estimation of three stingless bee species (Hymenoptera, Meliponinae) by flow cytometry. **Apidologie**, v. 40, n. 5, p. 517-523, 2009.
- LÓPEZ-FLORES, I.; GARRIDO-RAMOS, M.A. Repetitive DNA in the Repetitive DNA Content of Eukaryotic Genomes (ed. Garrido- Ramos, M.A.) p. 1-28 (Genome Dyn. Basel, Karger, vol 7. 2012.
- LORETO, V.; CABRERO, J.; LÓPEZ-LEÓN, M.A.; CAMACHO, J.P.M.; SOUZA, M.J. Possible autosomal origin of macro B chromosomes in two grasshopper species. **Chromosome Research**, v. 16, n. 2, p. 233-241, 2008.
- LOUREIRO, J.; RODRIGUEZ, E.; DOLEŽEL, J.; SANTOS, C. Comparison of four nuclear isolation buffers for plant DNA flow cytometry. **Annals of Botany**, v. 98, n. 3, p. 679-689, 2006a.
- LOUREIRO, J.; RODRIGUEZ, E.; DOLEŽEL, J.; SANTOS, C. Flow cytometric and microscopic analysis of the effect of tannic acid on plant nuclei and estimation of DNA content. **Annals of Botany**, v. 98, n. 3, p. 515-527, 2006b.
- LOVE, M.I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 550, 2014.
- LUGER, K.; DECHASSA, M.L.; TREMETHICK, D.J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? **Nature Reviews Molecular Cell Biology**, v. 13, n. 7, p. 436-447, 2012.
- LYNCH, J.A.; PEEL, A.D.; DRECHSLER, A.; AVEROF, M.; ROTH, S. EGF signaling and the origin of axial polarity among the insects. **Current Biology**, v. 20, n. 11, p. 1042-1047, 2010.
- LYNCH, M. *The Origins of the Genome Architecture*. Sinauer Associates, Inc. Publishers. Sunderland, Massachusetts. 2007.
- MANCHADO, M.; ZUASTI, E.; CROSS, I.; MERLO, 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, n. 1, p. 79-86, 2006.
- MANDL, B.; BRANDT, W.F.; SUPERTI-FURGA, G.; GRANINGER, P.G.; BIRNSTIEL,

M.L.; BUSSLINGER, M. The five cleavage-stage (CS) histones of the sea urchin are encoded by a maternally expressed family of replacement histone genes: functional equivalence of the CS H1 and frog H1M(B4) proteins. **Molecular and Cell Biology**, v. 17, n. 3, p. 1189-200, 1997.

MANK, J.E.; ELLEGREN, H. All dosage compensation is local: gene-by- gene regulation of sex-biased expression on the chicken Z chromosome. **Heredity**, v. 102, n. 3, p. 312-320, 2009.

MARIN, I.; BAKER, B.S. The evolutionary dynamics of sex determination. **Science**, v. 281, n. 5385, p. 1990-1994, 1998.

MAROJA, L.S.; CLARK, M.E.; HARRISON, R.G. *Wolbachia* plays no role in the one-way reproductive incompatibility between the hybridizing field crickets *Gryllus firmus* and *G. pennsylvanicus*. **Heredity**, v. 101, n. 5, p. 435-444, 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, n. 18, p. 6255-6270, 1983.

MATSUNAGA, S. Junk DNA promotes sex chromosome evolution. **Heredity**, v. 102, n. 6, p. 525-526, 2009.

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, n. 1, p. 1883-1891, 1983.

MEIKLEJOHN, C.D.; LANDEEN, E.L.; COOK, J.M.; KINGAN, S.B.; PRESGRAVES, D.C. Sex chromosome-specific regulation in the *Drosophila* male germline but little evidence for chromosomal dosage compensation or meiotic inactivation. **PLoS Biology**, v. 9, n. 8, p. e1001126, 2011.

MEIKLEJOHN, C.D.; PRESGRAVES, D.C. Little evidence for demasculinization of the *Drosophila* X chromosome among genes expressed in the male germline. **Genome Biology and Evolution**, v. 4, n. 10, p. 1007-1016, 2012.

MEISEL, R.P.; MALONE, J.H.; CLARK, A.G. Disentangling the relationship between sex-biased gene expression and X-linkage. **Genome Research**, v. 22, n. 7, p. 1255-1265, 2012.

MERLO, M.A.; CROSS, I.; MANCHADO, M.; CÁRDENAS, S.; REBORDINOS, L. The 5S rDNA high dynamism in *Diplodus sargus* is a transposon-mediated mechanism. Comparison with other multigene families and Sparidae species. **Journal of Molecular Evolution**, v. 76, n. 3, p. 83-97, 2013.

MESA, A.; DE MESA, R.S. Complex sex-determining mechanisms in three species of South American grasshoppers (Orthoptera, Acridoidea). **Chromosoma**, v. 21, n. 2, p. 163-180, 1967.

- MESA, A.; FONTANETTI, C.S.; GARCÍA-NOVO, P. Does an X-autosome centric fusion in Acridoidea condemn the species to extinction? **Journal of Orthoptera Research**, v. 10, n. 2, p. 141-146, 2001.
- Mesa, A.; García-Novo, P. *Neometrypus badius* a new species of cricket with an unusual sex determining mechanism. **Journal Orthoptera Research**, v. 10, n. 1, p. 81-87, 2001.
- MESA, A.; GARCÍA-NOVO, P.; DOS SANTOS, D. X1X2O (male)-X1X1X2X2 (female) chromosomal sex determining mechanism in the cricket *Ciclotyloides americanus* (Orthoptera, Grylloidea, Mogoplistidae). **Journal of Orthoptera Research**, v. 11, n. 1, p. 87-90, 2002.
- MEŠTROVIĆ, N.; MRAVINAC, B.; PAVLEK, M.; VOJVODA-ZELJKO, T.; ŠATOVIĆ, E.; PLOHL, M. Structural and functional liaisons between transposable elements and satellite DNAs. **Chromosome Research**, v. 23, n. 3, p. 583-596, 2015.
- MIGA, K.H. Completing the human genome: the progress and challenge of satellite DNA assembly. **Chromosome Research**, v. 23, n. 3, p. 421-426, 2015.
- MIKHAYLOVA, L.M.; NURMINSKY, D.I. Lack of global meiotic sex chromosome inactivation, and paucity of tissue-specific gene expression on the *Drosophila* X chromosome. **BMC Biology**, v. 9, p. 29, 2011.
- MIKHAYLOVA, L.M.; NURMINSKY, D.I. No severe and global X chromosome inactivation in meiotic male germline of *Drosophila*. **BMC Biology**, v. 10, p. 50, 2012.
- MILLS, RE.; BENNETT, E.A.; ISKOW, R.C.; DEVINE, S.E. Which transposable elements are active in the human genome? **Trends in Genetics**, v. 23, n. 4, p. 183-191, 2007.
- MITO, T.; SHINMYO, Y.; KURITA, K.; NAKAMURA, T.; OHUCHI, H.; et al. Ancestral functions of Delta/Notch signaling in the formation of body and leg segments in the cricket *Gryllus bimaculatus*. **Development**, v. 138, n. 17, p. 3823-3833, 2011.
- MOGHADAM, H.K.; POINTER, M.A.; WRIGHT, A.E.; BERLIN, S.; MANK, J.E. W chromosome expression responds to female-specific selection. **Proceedings of National Academy of Sciences of the United States of America**, v. 109, v. 21, p. 8207-8211, 2012.
- MONTIEL, E.E.; CABRERO, J.; CAMACHO, J.P.M.; LÓPEZ-LEÓN, M.D. Gypsy, RTE and Mariner transposable elements populate *Eyprepocnemis plorans* genome. **Genetica**, v. 140, n. 7-9, p. 365-374, 2012.
- MORENO-MENDOZA, N.; HARLEY, V.R.; MERCHANT-LARIOS, H. Differential expression of SOX9 in gonads of the sea turtle *Lepidochelys olivacea* at male- or female-promoting temperatures. **Journal of Experimental Zoology**, v. 284, n. 6, p. 705-710, 1999.
- MORIYA, Y.; ITOH, M.; OKUDA, S.; YOSHIZAWA, A.C.; KANEHISA, M. KAAS: an automatic genome annotation and pathway reconstruction server. **Nucleic Acids Research**, v. 35, p. W182-W185, 2007.

- MULLON, C.; POMIANKOWSKI, A.; REUTER, M. Molecular evolution of *Drosophila Sex-lethal* and related sex determining genes. **BMC Evolutionary Biology**, p. 12:5, 2010.
- NATRI, H.M.; SHIKANO, T.; MERILA, J. Progressive recombination suppression and differentiation in recently evolved neo-sex chromosomes. **Molecular Biology and Evolution**, v. 30, n. 5, p. 1131-1144, 2013.
- NAVAJAS-PÉREZ, R.; DE LA HERRÁN, R.; JAMILENA, M.; LOZANO, R.; RUIZ REJÓN, C.; RUIZ REJÓN, M.; GARRIDO-RAMOS, M.A. Reduced rates of sequence evolution of Y-linked satellite DNA in *Rumex* (Polygonaceae). **Journal of Molecular Evolution**, v. 60, n. 3, p. 391-399, 2005.
- NEI, M.; ROONEY, A.P. Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics*, v. 39, p. 121-152, 2005.
- NGUYEN, P.; SÝKOROVÁ, M.; ŠÍCHOVÁ, J.; KŮTA, V.; DALÍKOVÁ, M.; FRYDRYCHOVÁ, R.C.; NEVEN, L.G.; SAHARA, K.; MAREC, F. Neo-sex chromosomes and adaptive potential in tortricid pests. **Proceedings of the National Academy of Sciences**, v. 110, n. 17, p. 6931-6936, 2013.
- NICOLAS, M.; MARAIS, G.; VLADKA H.; V.; BOHUSLAV, J.; LAPORTE, V.; VYSKOT, B.; MOUCHIROUD, D.; NEGRUTIU, I.; CHARLESWORTH, D.; MONÉGER, F. A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. **PLoS Biology**, v. 3, n. 1, p. e4, 2005.
- NILSEN, T.W. The spliceosome: the most complex macromolecular machine in the cell? **BioEssays**, v. 25, n. 12, p. 1147-1149, 2003.
- NOVAK, P.; NEUMANN, P.; MACAS, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. **BMC Bioinformatics**, v. 11, p. 378, 2010.
- NOVAK, P.; NEUMANN, P.; PECH, P.; STEINHAIŠL, 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, n. 6, p. 792-793, 2013.
- NOVÁK, P.; NEUMANN, P.; PECH, P.; STEINHAIŠL, 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, n. 6, p. 792-793, 2013.
- OHNO, S. Sex Chromosomes and Sex-linked Genes. Springer-Berlin, v. 1, 1967.
- ORGEL, L.E.; CRICK, F.H. Selfish DNA: the ultimate parasite. **Nature**, v. 284, n. 5757, p. 604-607, 1980.
- OTTE, D. Evolution of cricket songs. **Journal of Orthoptera Research**, v. 1, n.1, p. 25-49,

1992.

PAÇO, A.; AGEDA, F.; MEŠTROVIĆ, N.; PLOHL, M.; CHAVES, R. The puzzling character of repetitive DNA in *Phodopus* genomes (Cricetidae, Rodentia). **Chromosome Research**, v. 23, n. 3, p. 427-440, 2015.

PAL-BHADRA, M.; LEIBOVITCH, B.A.; GANDHI, S.G.; CHIKKA, M.R.; BHADRA, U.; BIRCHLER, J.A.; et al. Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. **Science**, v. 303, n. 5658, p. 669-672, 2004.

PALACIOS-GIMENEZ, O.M.; BUENO, D.; CABRAL-DE-MELLO, D.C. Chromosomal mapping of two *Mariner*-like elements in the grasshopper *Abracris avolineata* (Orthoptera: Acrididae) reveals enrichment in euchromatin. **European Journal of Entomology**, v. 111, n. 3, p. 329-334, 2014

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.

PALACIOS-GIMENEZ, O.M.; MARTI, D.A.; CABRAL-DE-MELLO, D.C. Neo-sex chromosomes of *Ronderosia bergii*: insight into the evolution of sex chromosomes in grasshoppers. **Chromosoma**, v. 124, n. 3, p. 353-365, 2015.

PALACIOS-GIMENEZ, O.M.; MILANI, D.; LEMOS, B.; CASTILLO, E.R.; MARTÍ, D.A.; RAMOS, E.; MARTINS, C.; CABRAL-DE-MELLO, D.C. Uncovering the evolutionary history of neo-XY sex chromosomes in the grasshopper *Ronderosia bergii* (Orthoptera, Melanoplinae) through satellite DNA analysis. **BMC Evolutionary Biology**, submitted, 2017.

PALMIERI, N.; KOSIOL, C.; SCHLÖTTERER, C. The life cycle of *Drosophila* orphan genes. **ELife**, v. 3, p. e01311, 2014.

PALOMEQUE, T.; LORITE, P. Satellite DNA in insects: a review. **Heredity**, v. 100, n. 6, p. 564-573, 2008.

PARISI, M.; NUTTALL, R.; NAIMAN, D.; BOUFFARD, G.; MALLEY, J.; ANDREWS, J.; EASTMAN, S.; OLIVER, B. Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. **Science**, v. 299, n. 5607, p. 697-700, 2003.

PARSCH, J.; ELLEGREN, H. The evolutionary causes and consequences of sex-biased gene expression. **Nature Review Genetics**, v. 14, p. 83-87, 2013.

PASCOAL, S.; CEZARD, T.; EIK-NES, A.; GHARBI, K.; MAJEWSKA, J.; PAYNE, E.; RITCHIE, M.G.; ZUK, M.; BAILEY, N.W. Rapid convergent evolution in wild crickets. **Current Biology**, v. 24, n. 12, p. 1369-1374, 2014.

- PAVELITZ, T.; LIAO, D.; WEINER, A.M. Concerted evolution of the tandem array encoding primate U2 snRNA (the RNU2 locus) is accompanied by dramatic remodeling of the junctions with flanking chromosomal sequences. **EMBO Journal**, v. 18, n. 13, p. 3783-92, 1999.
- PAVELITZ, T.; RUSCHE, L.; MATERA, A.G.; SCHARF, J.M.; WEINER, A.M. Concerted evolution of the tandem array encoding primate U2 snRNA occurs *in situ*, without changing the cytological context of the RNU2 locus. **EMBO Journal**, v. 14, n. 1, p. 169-77, 1995.
- PEICHEL, C.L. Convergence and divergence in sex-chromosome evolution. **NATURE GENETICS**, v. 49, n. 3, p. 321-322, 2017.
- 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, n. 3, p. 331-335, 2001.
- PERINA, A.; SEOANE, D.; GONZÁLEZ-TIZÓN, A.M.; RODRÍGUEZ-FARIÑA, F.; MARTÍNEZ-LAGE, A. Molecular organization and phylogenetic analysis of 5S rDNA in crustaceans of the genus *Pollicipes* reveal birth-and-death evolution and strong purifying selection. **BMC Evolutionary Biology**, v. 11, p. 304, 2011.
- PETERSON, B.G; CARL, P.; BOUDT, K.; BENNETT, T.; ULRICH, U.; ZIVOT, E.; LESTEL, M.; et al. PerformanceAnalytics package. Econometric tools for performance and risk analysis <http://r-forge.r-project.org/projects/returnanalytics/>. 2014.
- PEZER, Z.; BRAJKOVIĆ, J.; FELICIELLO, I.; UGARKOVIĆ, D. Transcription of satellite DNAs in insects. **Progress in Molecular and Subcellular Biology**, v. 51, p. 161-178, 2011.
- PIEAU, C.; GIRONDOT, N.; RICHARD-MERCIER, G.; DESVAGES, M.; DORIZZI, P.; ZABORSKI, P. Temperature sensitivity of sexual differentiation of gonads in the European pond turtle. **Journal of Experimental Zoology**, v. 270, n. 1, p. 86-93, 1994.
- PIMPINELLI, S.; BERLOCO, M.; FANTI, L.; DIMITRI, P.; BONACCORSI, S.; MARCHETTI, E.; CAIZZI, R.; et al. Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. **Proceedings of National Academy of Sciences of the United States of America**, v. 92, n. 9, p. 3804-3808, 1995.
- PINHAL D., YOSHIMURA, T.S.; ARAKI, C.S.; MARTINS, C. The 5S rDNA family evolves through concerted and birth-and-death evolution in fish genomes: an example from freshwater stingrays. **BMC Evolutionary Biology**, v. 11, p. 151-164, 2011.
- PINKEL, D.; STRAUME, T.; GRAY, J.W. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. **Proceedings of the National Academy of Sciences**, v. 83, n. 9, p. 2934-2938, 1986.
- PLASTERK, R.H.A.; IZSVÁK, Z.; IVICS, Z. Resident aliens: the Tc1/*mariner* superfamily of transposable elements. **Trends in Genetics**, v. 15, n. 8, p. 326-332, 1999.

PLOHL, M.; LUCHETTI, A.; MEŠTROVIĆ, N.; MANTOVANI, B. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. **Gene**, v. 409, n. 1-2, p. 72-82, 2008.

PODGORNAYA, O.; GAVRILOVA, E.; STEPHANOVA, V.; KOMISSAROV, A. Large tandem repeats make up the chromosome bar code: a hypothesis. **Advances in Protein Chemistry and Structural Biology**, v. 90, p. 1-30, 2013.

R, CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2017.

RANZ, J.M.; CASTILLO-DAVIS, C.I.; MEIKLEJOHN, C.D.; HARTL, D.L. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. **Science**, v. 300, n. 5626, p. 1742-1745, 2003.

REBUZZINI, P.; CASTIGLIA, R.; NERGADZE, G.; et al. Quantitative variation of LINE-1 sequences in five species and three subspecies of the subgenus *Mus* and in five Robertsonian races of *Mus musculus domesticus*. **Chromosome Research**, v. 17, n. 1, p. 65-76, 2009.

REID, C.E.; LAZINSKI, D.W. A host-specific function is required for ligation of a wide variety of ribozyme-processed RNAs. **Proceedings of National Academy of Sciences of the United States of America**, v. 97, n. 1, p. 424-429, 2000.

RICE, W.R. Evolution of the Y sex chromosome in animals. **Bioscience**, v. 46, n. 5, p. 331-343, 1996.

RICE, W.R. Sex chromosomes and the evolution of sexual dimorphism. **Evolution**, v. 38, n. 4, p. 735-742, 1984.

RICHARD, G.F.; KERREST, A.; DUJON, B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. **Microbiology and Molecular Biology Reviews**, v. 72, n. 4, p. 686-727, 2008.

RIDEOUT, E.J.; BILLETER, J.C.; GOODWIN, S.F. The sex-determination gene *fruitless* and *doublesex* specify a neural substrate required for courtship song. **Current Biology**, v. 17, n. 17, p. 1473-1478, 2007.

ROBILLARD, T.; DESUTTER-GRANDCOLAS, L. A revision of the Neotropical Eneopterinae crickets (Orthoptera, Grylloidea, Eneopteridae) with a phylogenetic discussion. **Insect. Syst. Evol**, v. 35, n. 4, p. 411-435, 2005.

ROBILLARD, T.; GRANDCOLAS, P.; DESUTTER-GRANDCOLAS, L. A shift toward harmonics for high-frequency calling shown with phylogenetic study of frequency spectra in Eneopterinae crickets (Orthoptera, Grylloidea, Eneopteridae). **Canadian Journal of Zoology**, v. 85, n. 12, p. 1264-1275, 2007.

- ROESTI, M.; MOSER, D.; BERNER, D. Recombination in the threespine stickleback genome—patterns and consequences. **Molecular Ecology**, v. 22, n. 11, p. 3014-3027, 2013.
- ROJAS, A.A.; VAZQUEZ-TELLO, A.; FERBEYRE, G.; VENANZETTI, F.; BACHMANN, L.; PAQUIN, B.; et al. Hammerhead-mediated processing of satellite pDo500 family transcripts from *Dolichopoda* cave crickets. **Nucleic Acids Research**, v. 28, n. 20, p. 4037-4043, 2000.
- ROONEY, A.P.; WARD, T. Evolution of a large ribosomal RNA multigene family in filamentous fungi: birth and death of a concerted evolution paradigm. **Proceedings of National Academy of Sciences of the United States of America**, v. 102, n. 14, p. 5084-5089, 2005.
- ROSS, M.T.; GRAFHAM, D.V.; COFFEY, A.J.; SCHERER, S.; McLAY, K.; Platzer, M.; Howell, G.R.; et al. The DNA sequence of the human X chromosome. **Nature**, v. 434, n. 7031, p. 325-337, 2005.
- ROZEN, S.; SKALETSKY, H. Primer3 on the WWW for general users and for biologist programmers, p. 365-386. In *Bioinformatics Methods and Protocols*, Humana Press. 1999.
- RYNER, L.C.; BAKER, B.S. Regulation of *doublesex* pre-mRNA processing occurs by 3'-splice site activation. **Genes and Development**, v. 5, n. 11, p. 2071-2085, 1991.
- SAMBROOK, J.; RUSSELL, D.W. *Molecular Cloning: A Laboratory Manual*. The 3rd ed., Cold Spring Harbor laboratory. 2001.
- SÁNCHEZ, L. Sex-determining mechanisms in insects. **The International Journal of Developmental Biology**, v. 52, n. 7, p. 837-856, 2008.
- SANTOS, C.F.; GUYOT, R.; BORGES DOVALLE, C.; CHIARI, L.; TECHIO, V.H.; HESLOP-HARRISON, P.; VANZELA LAFORGA, A.L. Chromosomal distribution and evolution of abundant retrotransposons in plants: gypsy elements in diploid and polyploid *Brachiaria forage* grasses. **Chromosome Research**, v. 23, n. 3, p. 571-582, 2015.
- SCHMIDT, T.; HESLOP-HARRISON, J.S. Genomes, genes and junk: the large-scale organization of plant chromosomes. **Trends in Plant Science**, v. 3, n. 5, p. 195-199, 1998.
- SCHWEIZER, D. MENDELAK, M.; WHITE, M.J.D.; CONTRERAS, N. Cytogenetics of the parthenogenetic grasshopper *Warramaba virgo* and its bisexual relatives. **Chromosoma**, v. 88, n. 3, p. 227-236, 1983.
- SHAPIRO, H.M. *Practical flow cytometry*, 4th ed., John Wiley and Sons, New Jersey. 2003.
- SHAW, K.L.; LESNICK, S.C. Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 24, p. 9737- 9742, 2009.
- SHAW, K.L.; PARSONS, Y.M.; LESNICK, S.C. QTL analysis of a rapidly evolving speciation

phenotype in the Hawaiian cricket *Laupala*. **Molecular Ecology**, v. 16, n. 14, p. 2879-2892, 2007.

SHIRANGI, T.R.; TAYLOR, B.J.; MCKEOWN, M. A double-switch system regulates male courtship behavior in male and female *Drosophila melanogaster*. **Nature Genetics**, v. 38, n. 12, p. 1435-1439, 2006.

SINGER, D.; DOWNHOWER, L. Highly repeated DNA of the baboon: organization of sequences homologous to highly repeated DNA of the African green monkey. **Journal of Molecular Biology**, v. 134, n. 4, p. 835-842, 1979.

SINGER, M.; BERG, P. Genes and Genomes. A Changing Perspective. University Science Books, Mill Valley, California. 1991.

SKALETSKY, H.; KURODA-KAWAGUCHI, T.; MINX, P.J.; CORDUM, H.S.; HILLIER, L.; BROWN, L.G.; REPPING, S.; PYNTIKOVA, T.; ALI, J.; BIERI, T.; et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. **Nature**, v. 423, n. 6942, p. 825-837, 2003.

SMEDS, L.; WARMUTH, V.; BOLIVAR, P.; UEBBING, S.; BURRI, R.; SUH, A.; NATER, A.; et al. Evolutionary analysis of the female-specific avian W chromosome. **Nature Communications**, v. 6, p. 7330-7310, 2015.

SONG, H.; AMÉDÉGNATO, C.; CIGLIANO, M.M.; DESUTTER-GRANDCOLAS, L.; HEADS, S.W.; HUANG, Y.; OTTE, D.; WHITING, M.F. 300 million years of diversification: elucidating the patterns of orthopteran evolution based on comprehensive taxon and gene sampling. **Cladistics**, v. 31, n. 6, p. 621-651, 2015.

SPOTILA, L.D.; SPOTILA, J.R.; HALL, S.E. Sequence and expression analysis of WT1 and SOX9 in the red-eared slider turtle, *Trachemys scripta*. **Journal of Experimental Zoology**, v. 281, n. 5, p. 417-427, 1998.

STEFLOVÁ, P.; TOKAN, V.; VOGEL, I.; LEXA, M.; MACAS, J.; NOVAK, P., HOBZA, R.; et al. Contrasting patterns of transposable element and satellite distribution on sex chromosomes (XY₁Y₂) in the dioecious plant *Rumex acetosa*. **Genome Biology and Evolution**, v. 5, n. 4, p. 769-782, 2013.

STEINEMANN, M.; STEINEMANN, S. Degeneratin in Y chromosome of *Drosophila miranda*: a tap for retrotransposons. **Genetics**, v. 89, p. 7591-7595, 1992.

STEINEMANN, M.; STEINEMANN, S.; LOTTSPREICH, F. How Y chromosomes become genetically inert. **Proceedings of the National Academy of Sciences**, v. 90, n. 12, p. 5737-5741, 1993.

STEINEMANN, S.; STEINEMANN, M. Retroelements: tools for sex chromosome evolution. **Cytogenetic and Genome Research**, v. 110, n. 1-4, p. 134-143, 2005.

STEPHAN, W.; CHO, S. Possible role of natural selection in the formation of tandem-repetitive noncoding DNA. **Genetics**, v. 156, n. 1, p. 3330-341, 1994.

SUMNER, A.T. A simple technique for demonstrating centromeric heterochromatin. **Experimental Cell Research**, v. 75, n. 1, p. 304-306, 1972.

SWANSON WJ, WONG A, WOLFNER MF, AQUADRO CF. Evolutionary expressed sequence tag analysis of *Drosophila* identifies genes subjected to positive selection. **Genetics**, v. 168, n. 3, p. 1457-65, 2004.

SWANSON, W.J; CLARK, A.G; WALDRIP-DAIL, H.M.; WOLFNER, M.F.; AQUADRO, C.F. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. **Proceedings of National Academy of Sciences of the United States of America**, v. 95, n. 13, p. 4051-4, 2001.

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

THOMPSON, J.D.; HIGGIS, D.G.; GIBSON, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, v. 22, n. 22, p. 4673-80, 1994.

TIAN, M.; MANIATIS, T. A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA. **Cell**, v. 74, n. 1, p. 105-114, 1993.

TOMASZKIEWICZ, M.; MEDVEDEV, P.; MAKOVA, K.D. Y and W chromosome assemblies: approaches and discoveries. **Trends in Genetics**, v. 33, n. 4, p. 266-282, 2017.

TOMIOKA, K.; URYU, O.; KAMAE, Y.; UMEZAKI, Y.; YOSHII, T. Peripheral circadian rhythms and their regulatory mechanism in insects and some other arthropods: a review. **Journal of Comparative Physiology. B, Biochemical, Systematic and Environmental Physiology**, v. 182, n. 6, p. 729-740, 2012.

TRAPNELL, C.; WILLIAMS, B.A.; PERTEA, G.; MORTAZAVI, A.; KWAN, G.; VAN BAREN, M.J.; et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nature Biotechnology**, v. 28, n. 5, 511-515, 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, n. 7, p. 787-794, 2010.

- UGARKOVIĆ, D. Functional elements residing within satellite DNAs. **EMBO Reports**, v. 6, n. 11, p. 1035-1039, 2005.
- VALADKHAN, S. snRNAs as the catalysts of pre-mRNA splicing. **Current Opinion in Chemical Biology**, v. 9, n. 6, p. 603-608, 2005.
- VAN ARSDELL, S.W.; WEINER, A.M. Human genes for U2 small nuclear RNA are tandemly repeated. **Molecular and Cell Biology**, v. 4, n. 3, p. 492-499, 1984.
- VAN TOL, H.; BUZAYAN, J.M.; BRUENING, G. Evidence for spontaneous circle formation in the replication of the satellite RNA of tobacco ringspot virus. **Virology**, v. 180, n. 1, p. 23-30, 1991.
- VERHULST, E.C.; VAN DE ZANDE, L.; BEUKEBOOM, L.W. Insect sex determination: it all evolves around transformer. **Current Opinion in Genetics and Development**, v. 20, n. 4, p. 376-383, 2010.
- VIBRANOVSKI, M.D. Meiotic sex chromosome inactivation in *Drosophila*. **Journal of Genomics**, v. 2, p. 104-117, 2014.
- VIBRANOVSKI, M.D.; LOPES, H.F.; KARR, T.L.; LONG, M. Stage-specific expression profiling of *Drosophila* spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. **PLoS Genetics**, v. 5, n. 11, p. e1000731, 2009.
- VIBRANOVSKI, M.D.; ZHANG, Y.E.; KEMKEMER, C.; LOPES, H.F.; KARR, T.L.; LONG M. Re-analysis of the larval testis data on meiotic sex chromosome inactivation revealed evidence for tissue-specific gene expression related to the *Drosophila* X chromosome. **BMC Biology**, v. 10, p. 49, 2012.
- VICOSO, B.; EMERSON, J.J.; ZEKTSER, Y.; MAHAJAN, S.; BACHTROG, D. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. **PLoS Biology**, v. 11, n. 8, p. e1001643, 2013a.
- VICOSO, B.; KAISER, V.B.; BACHTROG, D. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. **Proceedings of the National Academy of Sciences of United States of America**, v. 110, n. 16, p. 6453-6458, 2013b.
- 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, n. 2, p. 127-142, 2011.
- VIERNA, J.; WEHNER, S.; ZU SIEDERDISSEN, C.H.; MARTÍNEZ-LAGE, A.; MARZ, M. Systematic analysis and evolution of 5S ribosomal DNA in metazoans. **Heredity**, v. 111, n. 5, p. 410-421, 2013.

VITTE, C.; BENNETZEN, J.L. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. **Proceedings of National Academy of Sciences of the United States of America**, v. 103, n. 47, p. 17638-43, 2006.

VIZOSO, M.; VIERNA, J.; GONZÁLEZ-TIZÓN, A.M.; MARTÍNEZ-LAGE, A. The 5S rDNA gene family in Mollusks: characterization of transcriptional regulatory regions, prediction of secondary structures, and long-term evolution, with special attention to *Mytilidae* mussels. **Journal of Heredity**, v. 102, n. 4, p. 433-447, 2011.

VOLPE, T.A.; KIDNER, C.; HALL, I.M.; TENG, G.; GREWAL, S.I.; MARTIENSSSEN, R.A. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. **Science**, v. 297, n. 5588, p. 1833-1837, 2002.

WANG, G.X.; HE, Q.Y.; MACAS, J.; NOVÁK, P.; NEUMANN, P.; MENG, D.X.; ZHAO, H.; GUO, N.; et al. Karyotypes and distribution of tandem repeat sequences in *Brassica nigra* determined by fluorescence *in situ* hybridization. **Cytogenetics and Genome Research**, v. 152, n. 3, p. 158-165, 2017.

WANG, Z.X.; KURATA, N.; SAJI, S.; KATAYOSE, Y.; MINOBE, Y. A chromosome 5-specific repetitive DNA-sequence in rice (*Oryza sativa* L.). **Theoretical Applied Genetics**, v. 90, n. 7-8, p. 907-913, 1995.

WATANABE-NAGASU, N.Y.; ITOH, T.; TANI, T.; OKANO, K.; KOGA, N.; OKADA, N.; OHSHIMA, Y. Structural analysis of gene loci for rat U1 small nuclear RNA. **Nucleic Acids Research**, v. 11, n. 6, p. 1791-1801, 1983.

WEBB, G.C.; WHITE, M.J.D.; CONTRERAS, N.; CHENEY, J (1978). Cytogenetics of the parthenogenetic grasshopper *Warramaba* (formerly *Moraba*) *virgo* and its bisexual relatives. IV. Chromosome banding studies. **Chromosoma**, v. 67, n. 4, p. 309-339, 1978.

WEST-EBERHARD, M.J. Sexual selection, social competition, and speciation. **Quarterly Review of Biology**, v. 58, n. 2, p. 155-183, 1983.

WEST, S. The increasing functional repertoire of U1 snRNA. **Biochemical Society Transactions**, v. 40, n. 4, p. 846-849, 2012.

WHITE, M.; KITANO, J.; PEICHEL, C.L. Purifying selection maintains dosage sensitive genes during degeneration of the threespine stickleback Y chromosome. **Molecular Biology and Evolution**, v. 32, v. 8, p. 1981-1995, 2015.

WHITE, M.J.D. Animal Cytology and Evolution. 3rd ed. Cambridge University Press, Cambridge. 1973.

WICKER T.; SABOT F.; HUA-VAN, A.; BENNETZEN, J.L.; CAPY, P.; CHAL-HOUB, B.; FLAVELL, A.; LEROY, P.; et al. A unified classification system for eukaryotic transposable

elements. **Nature Review Genetics**, v. 8, n. 12, p. 973-982, 2007.

WISE, J.A.; WEINER, A.M. *Dictyostelium* small nuclear RNA D2 is homologous to rat nucleolar RNA U3 and is encoded by a dispersed multigene family. **Cell**, v. 22, v. 1 Pt 1, p. 109118, 1980.

WITHERSPOON, D.J.; ROBERTSON, H.M. Neutral evolution of ten types of *mariner* transposons in the genomes of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. **Journal of Molecular Evolution**, v. 56, n. 6, p. 751-769, 2003.

WRIGHT, A.E.; DEAN, R.; ZIMMER, F.; MANK, J.E. How to make a sex chromosome. **Nature Communications**, v. 7, n. 12087, p. 12087, 2016.

WRIGHT, A.E.; HARRISON, P.W.; MONTGOMERY, S.H.; POINTER, M.A.; MANK, J.E. Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. **Evolution**, v. 68, n.11, p. 3281-3295, 2014.

YOSHIDO, A.; ŠÍCHOVÁ, J.; KUBÍČKOVÁ, S.; MAREC, F.; SAHARA, K. Rapid turnover of the W chromosome in geographical populations of wild silkmoths, *Samia cynthia* ssp. **Chromosome Research**, v. 21, n. 2, p. 1-16, 2013.

YOSHIMURA, A.; NAKATA, A.; MITO, T.; NOJI, S. The characteristics of karyotype and telomeric satellite DNA sequences in the cricket, *Gryllus bimaculatus* (Orthoptera, Gryllidae). **Cytogenetic and Genome Research**, v. 112, n. 3-4, p. 329-336, 2006.

ZEFA, E. Autosomal rearrangement in *Gryllus assimilis* Fabricius, 1775 (Orthoptera, Gryllidae). **Genetics and Molecular Biology**, v. 122, n. 3, p. 333-336, 1999.

ZEFA, E.; REDU, D.R.; COSTA, M.K.M.; FONTANETTI, C.S.; GOTTSCHALK, M.S.; PADILHA, G.B.; SILVA, A.F.; MARTINS, L.P. A new species of *Endecous* Saussure, 1878 (Orthoptera, Gryllidae) from northeast Brazil with the first X₁X₂0 chromosomal sex system in Gryllidae. **Zootaxa**, v. 3847, n. 1, p. 125-132, 2014.

ZENG, V.; EWEN-CAMPEN, B.; HORCH, H.W.; ROTH, S.; MITO, T.; et al. Developmental gene discovery in a hemimetabolous insect: de novo assembly and annotation of a transcriptome for the cricket *Gryllus bimaculatus*. **PLoS ONE**, v. 8, n. 5, p. e61479, 2013.

ZERA, A.J. Evolutionary genetics of juvenile hormone and ecdysteroid regulation in *Gryllus*: a case study in the microevolution of endocrine regulation. **Comparative Biochemistry and Physiology - Part A Molecular & Integrative Physiology**, v. 144, n. 3, p. 365-379, 2006.

ZHANG, Y.E.; VIBRANOVSKI, M.D.; KRINSKY, B.H.; LONG, M. Age-dependent chromosomal distribution of male-biased genes in *Drosophila*. **Genome Research**, v. 20, n. 11, p. 1526-1533, 2010.

ZHOU, Z.; ZHAO, L.; LIU, N.; GUO, H.; GUAN, B.; DI, J.; SHI, F. Towards a higher-level

Ensifera phylogeny inferred from mitogenome sequences. **Molecular Phylogenetics and Evolution**, v. 108, p. 22-33, 2017.

ZWICK, M.S.; HANSON, R.E.; MCKNIGHT, T.D.; NURUL-ISLAM-FARIDI, M.; STELLY, D.M. A rapid procedure for the isolation of *C₀t*-1 DNA from plants. **Genome**, v. 40, n. 1, p. 138-142, 1997.