


# 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-Gimenez<sup>1</sup>, Guilherme Borges Dias<sup>2</sup>, Leonardo Gomes de Lima<sup>2</sup>, Gustavo Campos e Silva Kuhn<sup>2</sup>, Érica Ramos<sup>3</sup> , Cesar Martins<sup>3</sup> & Diogo Cavalcanti Cabral-de-Mello<sup>1</sup>

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.

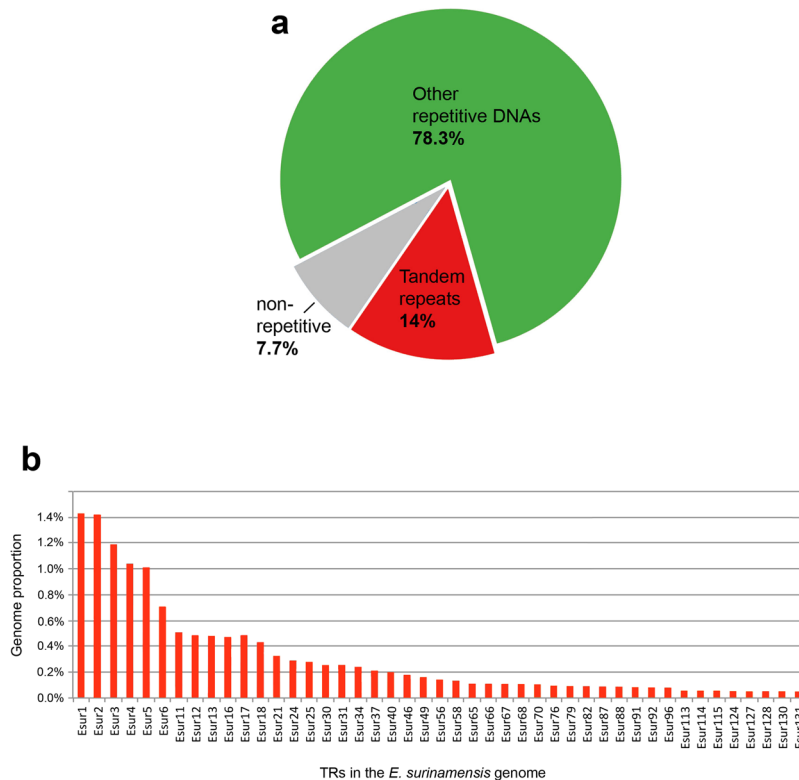
Among the repetitive sequences in the genomes of eukaryotes, tandem repeats (TRs) are very abundant and are mostly represented by satellite DNAs (satDNAs). SatDNA sequences are mainly located in centromeric, telomeric or intercalary heterochromatin<sup>1-3</sup> but, in some cases, also dispersed in euchromatin<sup>4,5</sup>. 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<sup>1,3,6-9</sup>.

SatDNA families, in general, differ in sequence identity, copy number and chromosome distribution<sup>2,10-12</sup>. 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<sup>2-4,7</sup>. Concerted evolution is achieved through multiple mechanisms of non-reciprocal transfer such as unequal cross-over, gene conversion, rolling-circle replication and transposition<sup>13,14</sup>.

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

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

<sup>2</sup>Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil. <sup>3</sup>UNESP - 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](mailto: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<sup>22, 23</sup> but eventually, diverse sex chromosome systems evolved several times, such as neo-XY♂/XX♀<sup>24–26</sup>, X<sub>1</sub>X<sub>2</sub>0♂/X<sub>1</sub>X<sub>1</sub>X<sub>2</sub>X<sub>2</sub>♀<sup>27</sup> and even neo-X<sub>1</sub>X<sub>2</sub>Y♂/X<sub>1</sub>X<sub>1</sub>X<sub>2</sub>X<sub>2</sub>♀<sup>25, 28</sup>. 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<sup>22, 23, 25–28</sup>. 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, C<sub>0</sub>f DNA fraction, telomeric repeats and microsatellite arrays<sup>25–29</sup>.

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-X<sub>1</sub>X<sub>2</sub>Y♂/X<sub>1</sub>X<sub>1</sub>X<sub>2</sub>X<sub>2</sub>♀ sex-determining system<sup>28</sup>. During male meiosis the multiple sex chromosomes remain unpaired and do not form chiasmata, suggesting that they do not proceed recombination<sup>22, 23</sup>. The neo-Y chromosome is the largest element and it exhibits multiple heterochromatic bands, while the neo-X<sub>1</sub> and neo-X<sub>2</sub> are poor in C-positive heterochromatin. Of the repetitive DNAs, two multigene families (5S rDNA and histone genes), C<sub>0</sub>f DNA fraction, and diverse microsatellites mapped to the neo-Y<sup>28</sup>.

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)<sup>30</sup>, 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-X<sub>1</sub>X<sub>2</sub>Y 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*<sup>28</sup>, this represents about 0.16× genome coverage. The clustering analysis through the RepeatExplorer<sup>31, 32</sup> 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  $\alpha$  and a composite repeat called  $\alpha/\beta$  (see Supplementary Table S2 and Supplementary Fig. S1). There was no significant sequence similarity between  $\alpha$  and  $\beta$  sequences. The size of the Esur17  $\alpha$ -repeat was 163 bp while the Esur17  $\alpha/\beta$  composite repeat was 317 bp, each consisting of an  $\alpha$ -unit and additional 154 bp  $\beta$ -unit. The nucleotide divergence between both Esur17- $\alpha$  and Esur17- $\alpha/\beta$  repeats was 9.4% (see Supplementary Table S2). Regarding Esur18, our sequences analysis showed that this is an array composed of  $\alpha/\beta$  composite units with 517 bp, consisting of a 329 bp  $\alpha$ -unit and a 188 bp  $\beta$ -unit (see Supplementary Fig. S1). The Esur18  $\alpha/\beta$  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.*<sup>28</sup>. 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-  $\alpha/\beta$ , 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-  $\alpha/\beta$  repeats (see Supplementary Fig. S4). For Esur17, the  $\alpha$  and  $\alpha/\beta$  repeats were located in the secondary constriction of pair 1 and in the X<sub>2</sub> chromosome; for Esur18, the  $\alpha$  and  $\beta$  repeats were located in secondary constriction of pair 1, but additional multiple loci of the  $\alpha$  repeats were detected in the Y chromosome (see Supplementary Table S2 and Supplementary Fig. S5).

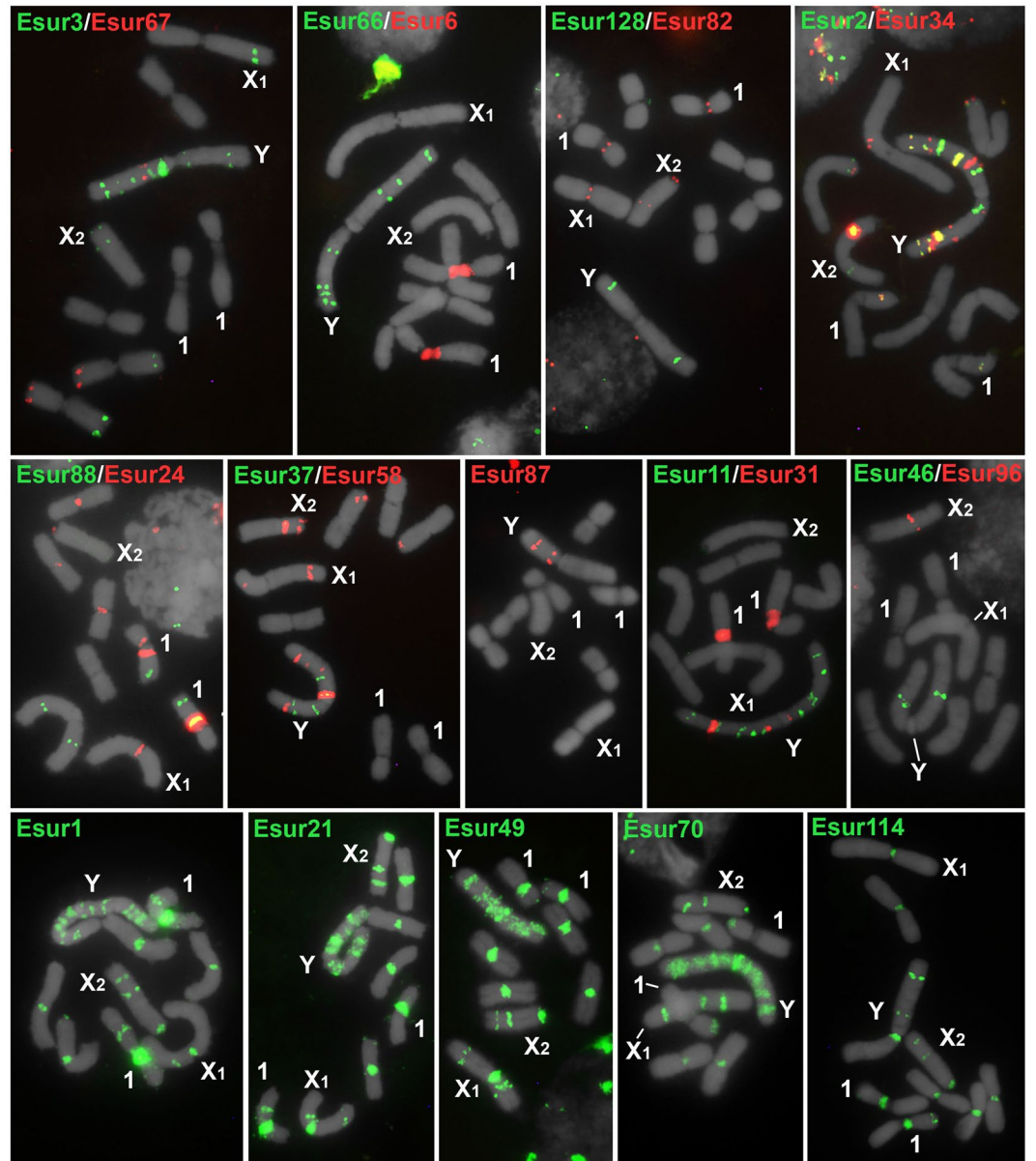
Some of satDNAs families were placed in the neo-X<sub>1</sub> and neo-X<sub>2</sub> but the abundance (number of loci) of satDNAs on the neo-Y chromosomes was remarkable, showing either multiple discrete loci or a scattered pattern (Fig. 3). The abundance in relation to the number of loci on the neo-Y chromosome was also noteworthy compared to autosomes and the neo-X<sub>1</sub> and neo-X<sub>2</sub> 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 families<sup>33</sup>, *Tribolium castaneum* with nine<sup>34</sup>, and *Camellia japonica* with four<sup>35</sup>, reaching up to 16 satDNAs in *Drosophila melanogaster* (reviewed by ref. 36) and 62 in the grasshopper *L. migratoria*<sup>30</sup>. 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<sup>33–36</sup>.

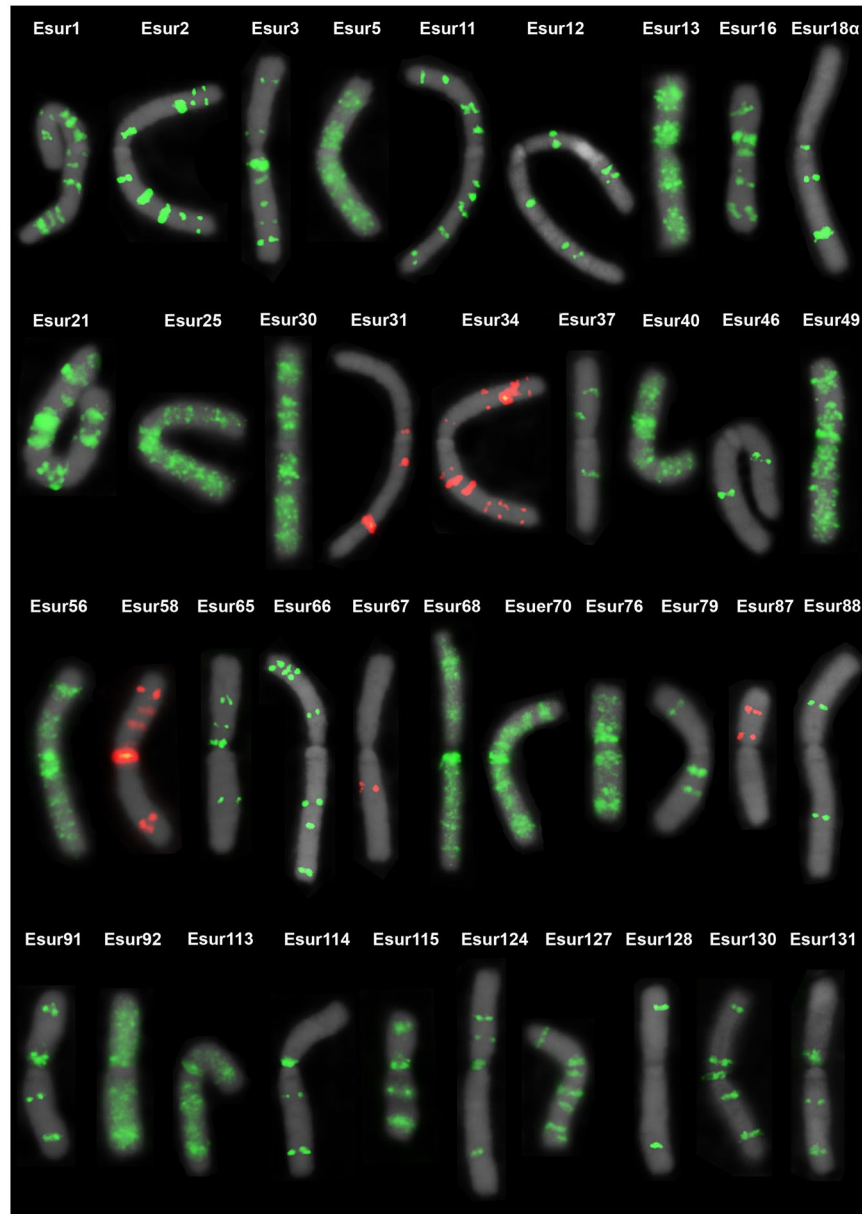
We could speculate that the accumulation of satDNAs in *E. surinamensis* could have been favored by highly rearranged karyotype ( $2n = 9\delta$ ) in comparison with the modal diploid chromosome number in Gryllidae crickets<sup>23</sup>. Alternatively, the high diversity of repetitive DNA itself might 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 X<sub>1</sub> and X<sub>2</sub> 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<sup>37–39</sup>. 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<sup>1,8</sup>. 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*.

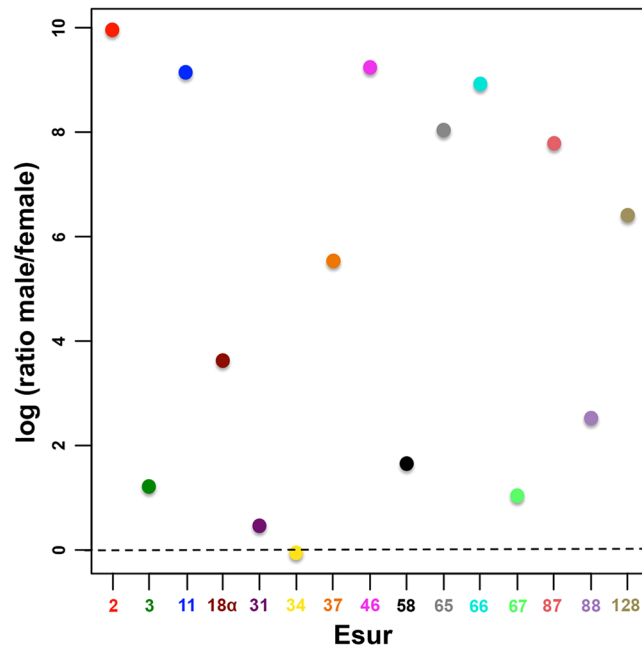




**Figure 3.** Distribution of the 39 satDNAs in the neo-Y chromosome of *E. surinamensis*.

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*<sup>40</sup>. This finding disagrees with predictions from computer simulations, which anticipated a direct correlation between the monomer size and the divergence due to low recombination frequency<sup>41</sup>. However, as noted by Bachmann *et al.*<sup>40</sup>, 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  $\Delta$ 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<sup>43–47</sup>. 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  $\alpha/\beta$  repeats for Esur17 and Esur18. satDNAs with composite units have been reported for example in *Chrysolina carnifex*<sup>48</sup>, *Tribolium brevicornis*<sup>49</sup> and primates<sup>50</sup>. 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*<sup>51</sup> the pBuM-1 and pBuM-2 subfamilies in the *Drosophila buzzatii* species complex<sup>52</sup> and the psr2 and psr18 subfamilies in *Nasonia vitripennis*<sup>53</sup>. 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* species<sup>54</sup> the pSc200 and pSC250 families in the rye chromosomes<sup>55</sup> and the PROsat, PSUchr1sat and PsatDNA satDNAs in two species of hamsters<sup>56</sup>.

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<sup>1, 18, 19, 21, 57, 58</sup>. In some species, it was shown that these repetitive sequences are involved in genetic degeneration, chromatin organization and regulation of expression<sup>36, 57, 59</sup>. 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<sup>22, 23</sup>. Regarding satDNAs, more than one family harboring sex chromosomes were reported, for example, in *D. melanogaster*<sup>60</sup>, *Muntiacus muntjac*<sup>61, 62</sup>, *T. castaneum*<sup>34</sup>, *Rumex acetosa*<sup>63–65</sup> and *Silene latifolia*<sup>66</sup>. 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-X<sub>1</sub> and neo-X<sub>2</sub>. A similar pattern was reported for the large Y chromosomes of plants that also accumulated repetitive DNAs<sup>65,66</sup> in contrast to the small mammalian Y chromosomes<sup>67,68</sup>.

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-X<sub>1</sub>X<sub>2</sub>Y sex chromosomes<sup>22,23,25</sup>. The evolution of *E. surinamensis* sex chromosomes probably included centric and tandem fusions and inversions, like in *M. muntjac*<sup>61,62</sup>. 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 chromosomes<sup>65</sup>. 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 length<sup>25</sup> 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<sup>22,23</sup>. 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<sup>22,23</sup> – a common feature for old sex chromosomes<sup>69</sup> – suggest the occurrence of old neo-sex chromosome system in *E. surinamensis*, which contrasts with the described young sex chromosomes<sup>21,70,71</sup>. Moreover considerable accumulation of various satDNAs and other repetitive DNA classes in the neo-Y chromosome<sup>25</sup> reinforce this view, as reported in other species<sup>67,69</sup>. This could mean that there are no evolutionary strata on the *E. surinamensis* sex chromosomes, similar to the strata found in human X chromosome<sup>72</sup>. 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 elsewhere<sup>73</sup>. 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 Russel<sup>74</sup>.

**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 FASTQC<sup>75</sup> and we did a quality filtering with the FASTX-Toolkit suit<sup>76</sup>. 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 suit<sup>76</sup> 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<sup>31,32</sup>. Afterward, we searched for clusters that showed repeat graph density in summary output, which is a typical characteristic of satDNAs families in this approach<sup>31</sup>, and refined this search using Dot plot charts implemented in Dotlet<sup>77</sup>.

**Isolation and sequence analysis of satDNAs.** Clusters with high graph density were submitted to the Tandem Repeats Finder (TRF) algorithm<sup>78</sup> 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 Dotlet<sup>77</sup> 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 Muscle<sup>79</sup> implemented in MEGA5<sup>80</sup>. 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 software<sup>81</sup> 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 MgCl<sub>2</sub>, 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.*<sup>82</sup> with modifications<sup>83</sup> using mitotic chromosome preparations. Fiber-FISH experiments were conducted as described in de Barros *et al.*<sup>84</sup> 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  $\Delta$ Ct method of relative quantification<sup>85</sup>. 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 3.3.1<sup>86</sup> and edited using Adobe Photoshop CS2.

**Statistic correlation analysis.** We used the PerformanceAnalytics package<sup>87</sup> implemented in the R statistical software version 3.3.1<sup>86</sup> 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 (XY<sub>1</sub>Y<sub>2</sub>) 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.cshl.edu/fastx\\_toolkit](http://hannonlab.cshl.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