

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*

*mansoni*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 $2n = 29\text{♂}/30\text{♀}$ and $X0\text{♂}/XX$,^{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 $2n = 29\text{♂}/30\text{♀}$ karyotype, an $X0\text{♂}/XX\text{♀}$ 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 18 S 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

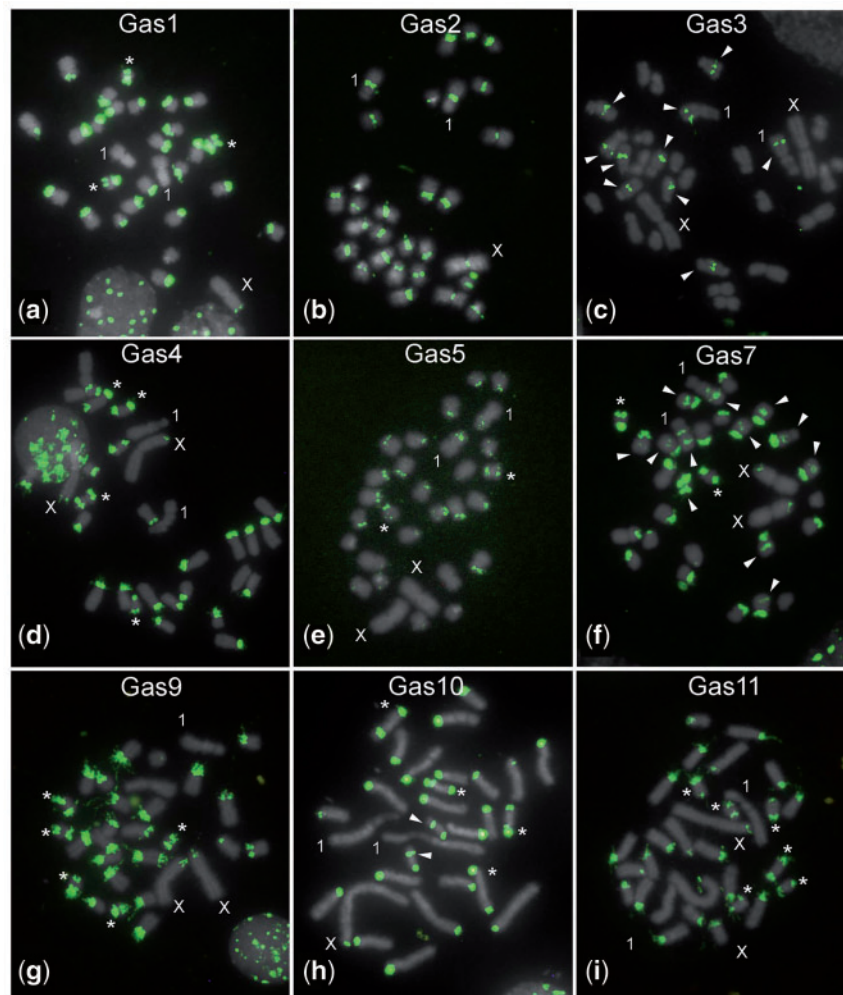


Figure 1. Chromosomal location of nine satDNAs in mitotic chromosomes of embryos of *G. assimilis* by FISH. The satDNA names are shown in the images. Asterisks indicate chromosomes with signals in both termini.

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), LWFWM (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* LWFWM (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

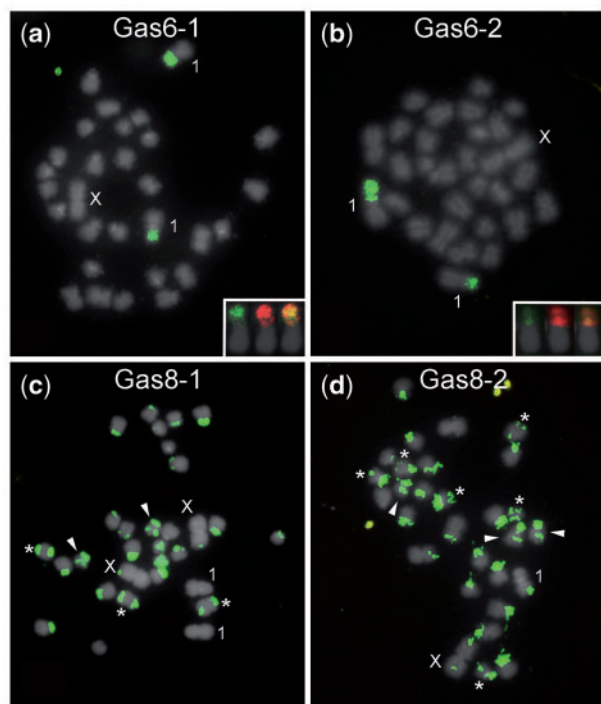


Figure 2. Chromosomal location of the Gas6 (a and b) and Gas8 (c and d) satDNAs subfamilies in mitotic chromosomes of embryos of *G. assimilis* by FISH. The satDNA names are shown in the images. Note the specific chromosomal localization on the pair 1 of the Gas6 subfamilies (a and b) contrasting with 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

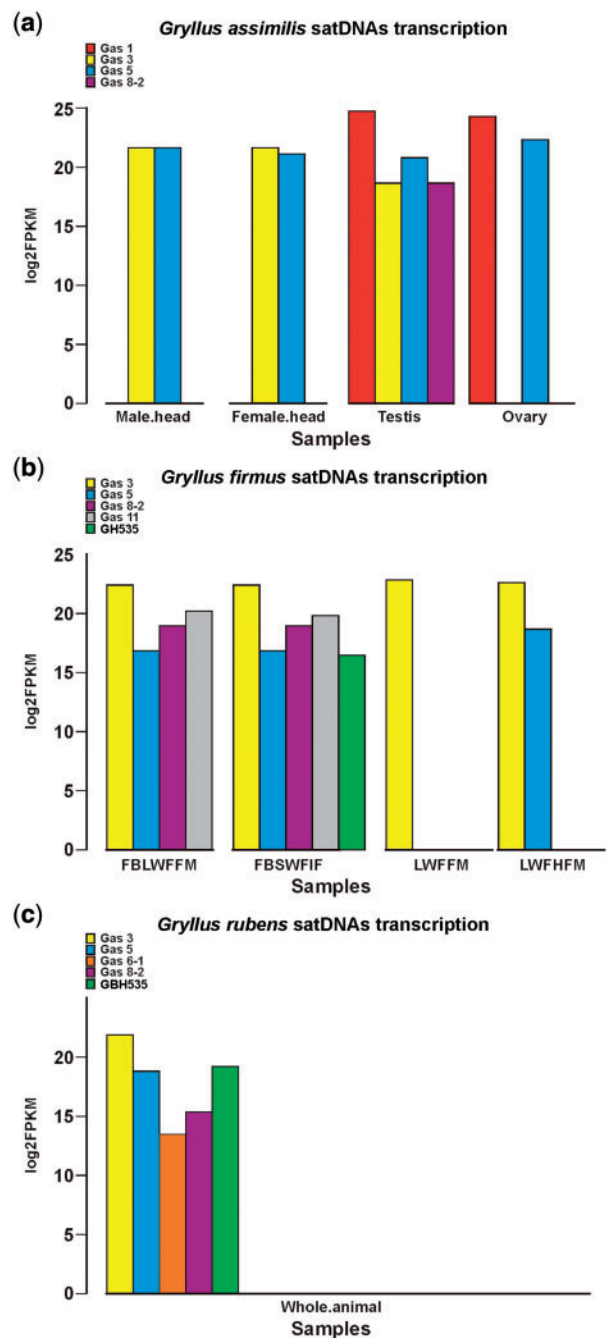


Figure 3. Differential expression of satDNAs between male and female in distinct body parts in the *G. assimilis* genome (a), between different tissues in *G. firmus* female (b) and whole animal samples in *G. rubens* (c). The quantification method applied is FPKM (fragments per kilo-base of transcript 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 Reads	Proportion	FPKM	Female head Reads	Proportion	FPKM	Testis Reads	Proportion	FPKM	Ovary Reads	Proportion
Gas1	0.0135	0	0	0	0	0	714	5.3E-06	2.7E+07	28	2.1E-07	1.9E+07
Gas3	0.00246	62	4.3E-07	3.2E+06	134	8.6E-07	106	7.9E-07	392,449	0	0	0
Gas5	0.00212	26	1.8E-07	3.1E+06	70	4.5E-07	318	2.4E-06	11.8E+06	32	2.4E-07	4.7E+06
Gas8-2	0.00127	0	0	0	0	0	66	4.9E-07	385,392	0	60	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).

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 histolyzed 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).

in 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 C_{0t} -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- X_1X_2Y .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 Plohl, 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 Larracunte, 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. Larracuent, 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, Đ. 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, Đ. 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, Đ. 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. Palomque, 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.