



Article

Satellitome Analysis of *Rhodnius prolixus*, One of the Main Chagas Disease Vector Species

Eugenia E. Montiel ¹, Francisco Panzera ² , Teresa Palomeque ¹ , Pedro Lorite ^{1,*} and Sebastián Pita ^{2,*}

¹ Department of Experimental Biology, Genetics, University of Jaén. Paraje las Lagunillas sn., 23071 Jaén, Spain; emontiel@ujaen.es (E.E.M.); tpalome@ujaen.es (T.P.)

² Evolutionary Genetic Section, Faculty of Science, University of the Republic, Iguá 4225, Montevideo 11400, Uruguay; fcopanzera@gmail.com

* Correspondence: plorite@ujaen.es (P.L.); spita@fcien.edu.uy (S.P.)

Abstract: The triatomine *Rhodnius prolixus* is the main vector of Chagas disease in countries such as Colombia and Venezuela, and the first kissing bug whose genome has been sequenced and assembled. In the repetitive genome fraction (repeatome) of this species, the transposable elements represented 19% of *R. prolixus* genome, being mostly DNA transposon (Class II elements). However, scarce information has been published regarding another important repeated DNA fraction, the satellite DNA (satDNA), or satellitome. Here, we offer, for the first time, extended data about satellite DNA families in the *R. prolixus* genome using bioinformatics pipeline based on low-coverage sequencing data. The satellitome of *R. prolixus* represents 8% of the total genome and it is composed by 39 satDNA families, including four satDNA families that are shared with *Triatoma infestans*, as well as telomeric (TTAGG)_n and (GATA)_n repeats, also present in the *T. infestans* genome. Only three of them exceed 1% of the genome. Chromosomal hybridization with these satDNA probes showed dispersed signals over the euchromatin of all chromosomes, both in autosomes and sex chromosomes. Moreover, clustering analysis revealed that most abundant satDNA families configured several superclusters, indicating that *R. prolixus* satellitome is complex and that the four most abundant satDNA families are composed by different subfamilies. Additionally, transcription of satDNA families was analyzed in different tissues, showing that 33 out of 39 satDNA families are transcribed in four different patterns of expression across samples.

Keywords: Chagas disease vector; *Rhodnius prolixus*; satellite DNA; satellitome; fluorescent in situ hybridization; satellite DNA expression; genome evolution



Citation: Montiel, E.E.; Panzera, F.; Palomeque, T.; Lorite, P.; Pita, S. Satellitome Analysis of *Rhodnius prolixus*, One of the Main Chagas Disease Vector Species. *Int. J. Mol. Sci.* **2021**, *22*, 6052. <https://doi.org/10.3390/ijms22116052>

Academic Editors: Miroslav Plohl, Eva Šatović and Raquel Chaves

Received: 4 May 2021

Accepted: 1 June 2021

Published: 3 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rhodnius prolixus, due to its medical importance as one of the main Chagas disease vector species, was the first Triatominae species to be sequenced [1]. Assembled sequences covered about 702 Mb, approximately 95% of the genome taking into consideration that the haploid genome size of *R. prolixus* was estimated at 733 Mb [2]. According to the annotation of this genome assembly, the repeatome of *R. prolixus*—repeated DNA sequences composing a genome [3]—make up to 5.6% of the genome, with Class II transposable elements being the main components [1]. Recently, Castro et al. [4] applied dnaPipeTE software [5] to re-evaluate the transposable element (TEs) quantification in the *R. prolixus* genome, with astonishing results. Using the same raw data obtained in the genome assembly project, Castro et al. [4] estimated that the amount of TEs in the *R. prolixus* genome ranged between 19% and 23%, that is, three to four times higher than the original quantification of Mesquita et al. [1]. In addition, they evaluated other sibling species, *R. montenegrensis* and *R. marabaensis* (formerly *R. robustus* II and III, respectively [6]), with similar results. These findings also confirmed that Class II elements were the most abundant TEs in *Rhodnius* genomes [1,7]. The main issue of genomes with large repeatome is that repeated DNA sequences hinder the genome assembly process, resulting in collapsed

and fragmented genomes and an underestimation of their amount in the genome [8,9]. This underestimation may be caused by several reasons. First, not all repeated DNAs might be present in the genome assembly, or those present could be collapsed. Second, the methodology used was based solely on homology to already known sequences, which makes it likely that new or highly divergent families will not be detected.

Another important component of the repeatome is tandem repeat DNA, in particular, satellite DNA (satDNA). SatDNA is defined as a non-genic repeat sequence organized in arrays of variable length and it can be classified by its repeat unit length as microsatellites, minisatellites and satellites [10]. Currently, data on satDNA repeats are almost completely missing for the *R. prolixus* genome. The only available information on *R. prolixus* satDNA is the existence of four satDNA families that are shared with other triatomine species, *Triatoma infestans* [11], another important Chagas disease vector. However, there is a wealth of information about satDNAs of *T. infestans*. Firstly, in this species, three AT-rich satDNA families strongly related to transposable elements have been characterized from a C0t library [12]. A few years later, Pita et al. [13] described the *T. infestans* repeatome finding that satDNAs are the main component of the repeated fraction. The *T. infestans* satellitome, the collection of satDNA families in a genome [14], includes 42 different satDNA families [13]. Moreover, satellitome analysis could determine that genome size differences between *T. infestans* major lineages (Andean and non-Andean) were due to variations in satDNA abundance [13]. Recently, the satellitome of another heteropteran, *Holhymenia histrio*, has been described, with 34 satDNA families with great variability in their chromosomal location [15].

Currently, satDNA characterization has experienced a huge increase due to the emergence of bioinformatics pipelines using low-coverage sequencing data, i.e., RepeatExplorer or TAREAN [16,17]. In the last five years, more than 40 studies have been published describing satDNA families using those pipelines, among them, the satellitome analysis of *T. infestans* and *H. histrio*. Other successful examples of the employment of this methodology were the satellitome analyses of several grasshoppers [14,18], one cricket species [19], *Drosophila* [20] or beetles [21]. Outside the arthropods, satDNA characterization from low-coverage sequencing data have been published in fish [22–24] and mostly in plants [25–28], among several other species. Those studies have paved the way for understanding satDNA organization and analyzing new roles in the genomes.

Besides their well-known structural function composing heterochromatin, centromeres or telomeres, a relevant role in chromosomal organization, pairing and segregation has been attributed to satDNA [8,29]. For instance, Cabral-de-Mello et al. [30] have recently reported its involvement in the differentiation of the Z sex chromosome in Crambidae moths. Furthermore, non-coding transcripts of satDNA are involved not only in heterochromatin maintenance and kinetochore assembly [31,32], but also in mosquito embryonic development, promoting gene silencing [33]. In *Drosophila melanogaster*, transcripts derived of the (AAGAG)_n satellite are important for viability and male fertility [34], whereas in *Tribolium castaneum*, TCAST satDNA expression affects the epigenetic state of constitutive heterochromatin in heat-shock conditions [35]. In addition, satDNA is transcribed in many insect species during development and exhibits differential expression between tissues or sexes [19,36,37]. Finally, Shatskikh et al. [38] reviewed the transcription of satDNA in *Drosophila*. According to the authors, the generated small RNAs have an important role, among others, in the viability and fertility of the fly and in the regulation of gene expression, also contributing to facilitate dosage compensation.

Herein, we present the description of *R. prolixus* satellitome covering the characterization of their sequences, abundance, divergence and transcriptional activity of each satDNA family. This study is the first to address a genome-wide analysis of the satDNA in this species, contributing to the knowledge of sequences that compose the genome of this important vector of Chagas disease.

2. Results and Discussion

The kissing bug *R. prolixus* is the main vector of Chagas diseases in countries such as Colombia and Venezuela in Latin America [39], and the first Triatomine species whose genome has been sequenced and assembled [1]. Here, we described its satellitome, which is composed by 39 satDNA families, 34 of which were detected by RepeatExplorer2 analysis, and 5 were detected by RepeatMasker mapping. Those results will be discussed below.

Low-coverage sequencing was performed obtaining 6,380,542 paired-end reads (932 Mb) after quality trimming processing. RepeatExplorer2 was used for de novo discovery of satDNAs, together with RepeatMasker to estimate their abundance and divergence. A subset of six million reads were used as input for the RepeatExplorer2 run, and six hundred thousand reads (≈ 90 Mb) were randomly taken as a sample by the software to perform the clusterization (0.12X coverage). Overall, annotation determined that 24% of the genome were multi-copy DNA sequences (Supplementary Figure S1). The analysis of 437 clusters above 0.001% of the genome rendered a total of 34 satDNA families, pointing out that these kinds of tandem repeats bear a great variability.

As commented in the Introduction, only four low copy-number satDNA families have been characterized up to now in the *R. prolixus* genome, all of them shared with *T. infestans* [11]. Although three of these four satDNA families were not detected by the RepeatExplorer2 approach, RepeatMasker masking allowed to confirm their presence. One possible explanation of non-detection of these three satDNA families could be their abundance. Their amount in the *R. prolixus* genome is lower than in the *T. infestans* genome (RproSat13-293, 0.067% vs. TinfSat04-1000, 2.45–4.26%; RproSat25-84, 0.009% vs. TinfSat12-84, 0.02–0.03%; RproSat37-98, 0.0005% vs. TinfSat15-99, 0.01–0.02%) (Table 1). Only the most abundant of the shared satDNAs was present in one cluster obtained by RepeatExplorer2. This satDNA family, RproSat06-136, is one-fold more abundant in *R. prolixus* (0.32%) than in *T. infestans* (TinfSat33-372, 0.01–0.02%). Nevertheless, the amount of two of these families (RproSat13-293 and RproSat25-84) exceeds 0.001% of the genome, which was the limit used for the RepeatExplorer2 analysis, and they should have been detected on the analyzed clusters. It is possible that the high nucleotide diversity observed for these two satDNA families (28.28% and 25.75%) may interfere with the reads' clustering. No new shared satDNA families have been detected between the two species other than the four previously described. The existence of only four low copy-number satDNA families of *R. prolixus* shared with *T. infestans* is in accordance with genomic in situ hybridization analyses in *R. prolixus* using *Triatoma* genomic probes, which revealed that repetitive DNA between both genera were not shared at a great scale [40,41]. These results reinforce the idea of the great genomic difference between both Triatominae genera. Moreover, this scenario is expected for the fast-evolving satDNA sequences. The library evolution hypothesis states that close species tend to share a group of satDNA families, the so-called "library", which in turn, during the divergence of species, may change due to the loss or amplification of each member of the library. Hence, the more distant two species are, the less probable it would be that they share satDNA families between them [42,43]. Since Triatomini and Rhodniini divergence was dated around 18–22 million years ago (Mya) [44], it represents a rather high divergence time under the library evolution hypothesis. Other insects showing similar satDNA families presented shorter divergence periods below 8 Mya, i.e., three species from the *Drosophila obscura* subgroup [45], *Drosophila virilis* and *D. americana* [20], or in several grasshopper species from the *Schistocerca* genus [46]. Nevertheless, there are also some exceptional cases of shared satDNAs among ant species, with great divergence time (74–80 Mya) [47]. Considering other biological groups besides insects, there are extreme cases of conserved satDNAs, such as the dodeca satellite present in *Drosophila melanogaster*, *Arabidopsis thaliana* and humans [48], or the BIV160 and DTHS₃ satDNA families conserved in bivalve mollusks for over 500 My [49].

Table 1. Data of the satDNA families found in *Rhodnius prolixus*: genome abundance (%), length of the repeat unit, A + T content and divergence (%). The number of clusters in the RepeatExplorer2 analysis is also shown. GenBank accession numbers: MW827131 to MW827167.

Name	No. of RE Clusters	Genome Proportion	Repeat Unit Length (bp)	A + T Percentage	Kimura Divergence (%)
RproSat01-165	38	2.13%	165	72.1%	13.42%
RproSat02-169	5	1.94%	169	69.2%	9.81%
RproSat03-124	2	1.18%	124	62.1%	8.81%
RproSat04-133	23	0.861%	133	66.2%	17.38%
RproSat05-208	1	0.460%	208	53.4%	15.48%
RproSat06-136 ¹	1	0.320%	136	62.5%	12.42%
RproSat07-375	1	0.200%	375	68.5%	7.77%
RproSat08-67	1	0.194%	67	61.2%	20.58%
RproSat09-499	1	0.107%	499	71.1%	12.20%
RproSat10-104	1	0.085%	104	62.5%	10.10%
RproSat11-198	1	0.073%	198	83.3%	2.23%
RproSat12-41	1	0.069%	41	58.5%	10.03%
RproSat13-293 ¹	-	0.067%	293	61.1%	28.28%
RproSat14-461	1	0.062%	461	64.2%	6.19%
RproSat15-161	1	0.052%	161	60.9%	8.38%
RproSat16-821	1	0.044%	821	65.3%	6.83%
RproSat17-584	1	0.027%	584	64.7%	11.71%
RproSat18-122	1	0.019%	122	63.1%	25.27%
RproSat19-201	1	0.019%	201	76.1%	19.52%
RproSat20-134	1	0.017%	134	68.7%	16.23%
RproSat21-167	1	0.016%	167	71.3%	16.25%
RproSat22-980	2	0.013%	980	69.4%	4.75%
RproSat23-412	1	0.010%	412	75.2%	5.59%
RproSat24-673	1	0.009%	673	70.0%	2.89%
RproSat25-84 ¹	-	0.009%	84	65.5%	25.75%
RproSat26-146	1	0.009%	146	63.7%	3.55%
RproSat27-187	1	0.008%	187	27.8%	13.06%
RproSat28-199	1	0.008%	199	53.8%	12.72%
RproSat29-31	1	0.006%	31	77.4%	11.21%
RproSat30-201	1	0.006%	201	61.2%	3.02%
RproSat31-75	1	0.005%	75	68.0%	2.64%
RproSat32-59	1	0.004%	59	59.3%	0.88%
RproSat33-123	1	0.003%	123	78.0%	0.96%
RproSat34-415	1	0.003%	415	72.8%	5.15%
RproSat35-279	1	0.003%	279	60.6%	5.71%
RproSat36-40	1	0.003%	40	70.0%	4.17%
RproSat37-98 ¹	-	0.0005%	98	69.4%	6.96%
Telomeric repeat ¹	-	0.003%	5	60.0%	13.9%
(GATA) _n repeat ¹	-	0.001%	4	75.0%	10.1%
Total		8.05%			
Mean			235.23	65.72%	10.56%
SD			223.13	9.10%	6.91%
Median			165	65.50%	10.03%

¹ SatDNA families also present in the *Triatoma infestans* genome.

RepeatExplorer2 analysis has limitations to detect low-complexity sequences, such as the telomeric repeats [16], but RepeatMasker analysis confirmed the presence of the insect canonical telomeric repeat (TTAGG)_n in *R. prolixus*—previously reported by FISH [50]—though at low amounts (0.003% of the genome). In consequence, other repetitive DNAs with short repeat units may have been omitted in this analysis. Therefore, (GATA)_n repeats amount was calculated since this repeat is extremely abundant in the *T. infestans* genome and it seems to be the only repeat DNA shared in the Y chromosomes in *Triatoma*

species [41]. In *R. prolixus*, (GATA)_n repeats are barely 0.001% of the genome, far away from the 4.5% in *T. infestans* [13].

Taking into consideration the telomeric and the (GATA)_n repeats, at least 39 satDNA families are present in the *R. prolixus* satellitome, representing 8.05% of its genome (Table 1). For the nomenclature of the different satDNA families, the proposal of Ruíz-Ruano et al. [14] has been followed, with the satDNA family name bearing the species name abbreviation (Rpro), a number in decreasing abundance and the length of the repeat sequence (Table 1, Supplementary Table S1). The relationships between satDNA families were analyzed by comparison of the consensus sequences. Most of the satDNA families did not show similarity with the sequences of other families. However, four satDNA families presented regions with high similarity. RproSat07-375 and RproSat09-499 families share 79 bp with an identity of 82%, while RproSat22-980 and RproSat24-675 share 69 bp with an 88% identity (Supplementary Figure S2). The high similarity found between the sequences of those satDNA families may suggest that they are evolutionarily related in spite of their different size.

In *R. prolixus*, the C-banding technique revealed the existence of heterochromatin only on the Y chromosome, that is completely heterochromatic. Whereas in *T. infestans*, in addition to an entirely heterochromatic Y chromosome, there are prominent autosomal C-heterochromatic regions whose number and size varies between Andean and non-Andean lineages [2,11,51,52]. This corroborates the higher proportion of satDNA in *T. infestans* (33% and 25% in Andean and non-Andean genomes, respectively) [13] in relation to *R. prolixus* (8.05%). Regarding *R. prolixus*, the amount of the satDNA families descends gradually, with the top family being 2.13%, and just nine families above 0.1%. On the other hand, *T. infestans* genomes present few extremely amplified families, which altogether represent the great majority of the entire satDNA content [13]. Interestingly, a similar situation is observed in the Heteroptera species *Holhymenia histrio*, where the most abundant satDNA family represents 14% of the genome, while all satDNA are 17% of the genome [15].

The size of the repeat units showed great variation, from 31 bp (RproSat29-31) up to near 1 kb (RproSat22-980), regardless of the telomeric and (GATA)_n repeats (Figure 1). Most of the satDNA families have repeat units smaller than 300 bp, although the most frequent sizes were between 100 and 200 bp (median = 163 bp). This is different to the size pattern found in *T. infestans*, in which most of the satDNA families have repeat units smaller than 100 bp (median = 72 bp) (Figure 1). The A + T content of satDNA family sequences ranges between 27.8% (RproSat27-187) and 83.3% (RproSat11-198) (Table 1, Supplementary Table S1). According to our sequencing data, the A + T content of paired-end reads is 65.37%, indicating that satDNA sequences, with a mean of 65.7%, are not especially enriched in A + T. Furthermore, the A + T richness of *R. prolixus* satellitome is just slightly higher than the *T. infestans* one (range of 44.3–81.6%, mean of 64.3%) [13]. Nucleotide divergence of satDNA families in *R. prolixus* is also similar to other satellitomes. Ranging between 0.88% (RproSat32-59) and 28.28% (RproSat13-293), the satellitome divergence of *R. prolixus* shows a median value of 10.03% (Table 1), similar to that described for the grasshopper *Eumigus monticola* (9.21%), the cricket *Gryllus assimilis* (9.3%) or the fish *Megaleporinus macrocephalus* (10.89%) [18,19,23]. Notwithstanding, this divergence value is double that of the satellitome divergence of the beetle *Hippodamia variegata* (5.75%) [21].

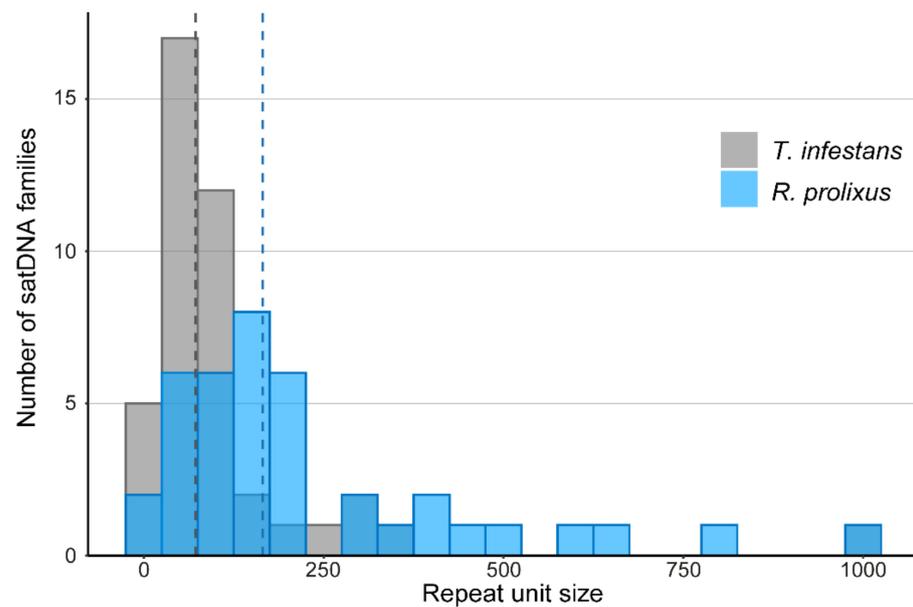


Figure 1. Repeat unit size distribution on *Rhodnius prolixus* and *Triatoma infestans*. Dashed lines represent median value of repeat unit size for *R. prolixus* (163 bp) and *T. infestans* (72 bp).

When *R. prolixus* satellitome distribution of abundance vs. divergence with respect to each consensus sequence was analyzed, a right-skewed distribution was obtained with a peak below 5% divergence and a long tail (Figure 2a). Amplification and homogenization processes, inversely related to divergence, and point mutation, directly related to divergence [25,53], are the main forces of the satDNA evolution. Bearing that in mind, the satellitome landscapes are very informative about the satDNA structure on the genome, where more homogeneous satDNA families will present narrow and high distribution while dispersed satDNA families will show wide and flattened distributions.

Clustering analysis revealed that the most abundant satDNA families formed superclusters. A supercluster is a set of clusters of the same repetitive DNA [16]. The highest numbers of clusters forming a supercluster were found for RproSat01-165 and RproSat04-133, with 38 and 23 clusters, respectively (Table 1). Each cluster generated a different consensus sequence, which was named with the original cluster name given in the RepeatExplorer2 output, for instance, CL1, CL2, etc. Figure 3 shows the alignment of the consensus sequences from each cluster of these satDNA families. If the different variants of the satDNA family were clustered on different arrays in the genome, they would represent true subfamilies. However, if the different variants of a satDNA family were mixed, they would not belong to different subfamilies, although they have been separated and assigned to different clusters by RepeatExplorer2. In order to test these two alternative hypotheses, we have analyzed the presence of each cluster consensus sequence in the *R. prolixus* genome assembly [1]. Searches in the assembled genome showed that the great majority of the contigs or scaffolds of the assembled genome contain only one of the variants (Figure 4), suggesting the existence of true satDNA subfamilies for the four most abundant satDNA families. The only exception was found in the RproSat04-133, where two of the sequence variants appeared together in more than 80% of the scaffolds (Figure 4d, clusters 80 and 148). This may be an artefact generated by the high similarity between the consensus sequences of these two subfamilies (over 97%), which makes it difficult to discriminate them from each other, and hence hampers the analysis. Additionally, minimum spanning networks were generated to analyze satDNA subfamilies' relationships within each family. The most complex networks correspond to satDNA families with higher amounts of members, RproSat01-165 and RproSat04-133. In the RproSat01-165 family, the most abundant subfamilies are close in the network, with the CL11 subfamily acting as a network node (Figure 5a). In the RproSat04-133 family, the network is more complex and reticulate than

in the RproSat01-165 one, coinciding with a higher Kimura divergence and broader and more flattened landscape (Figures 2e and 5d).

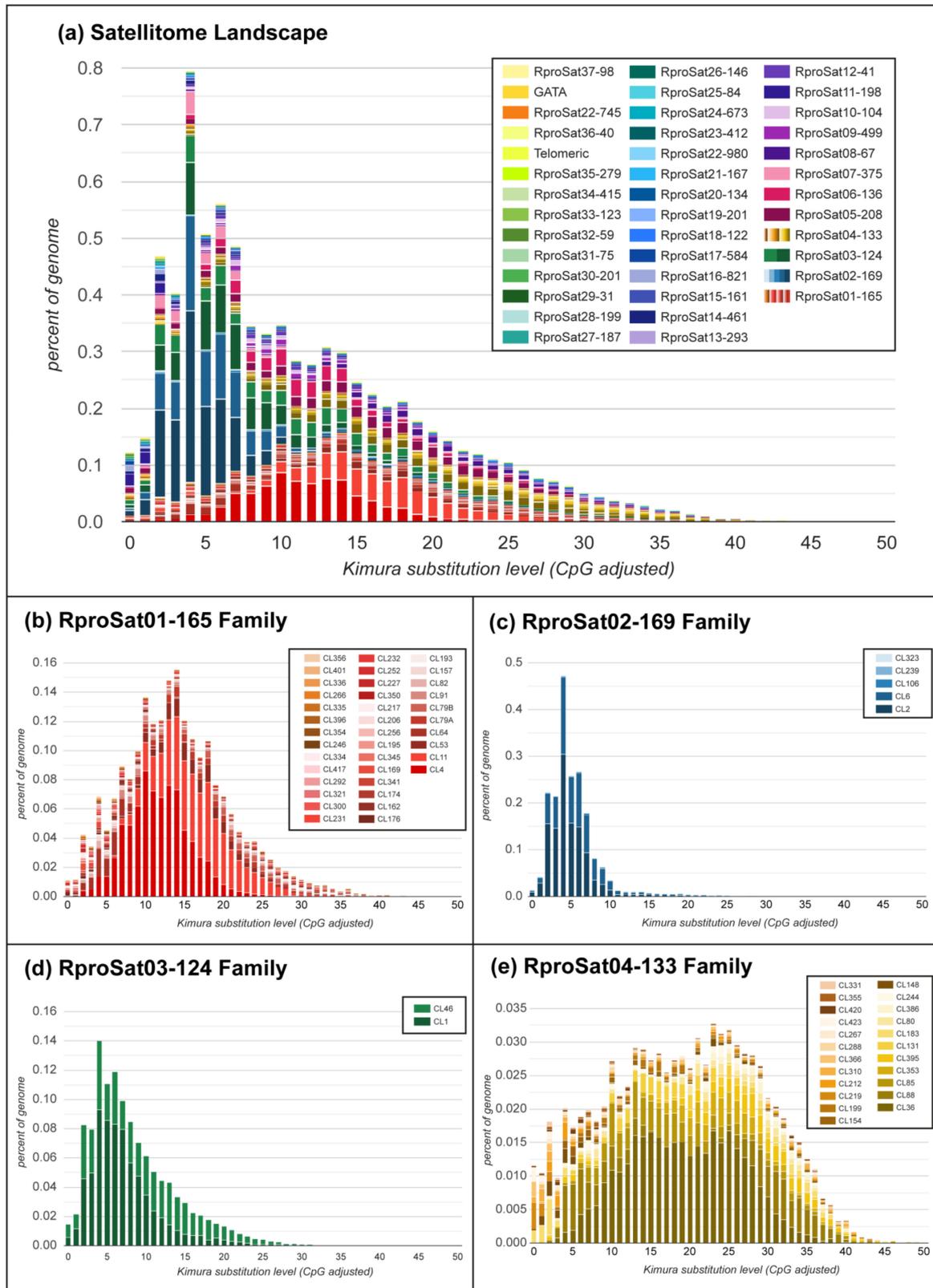
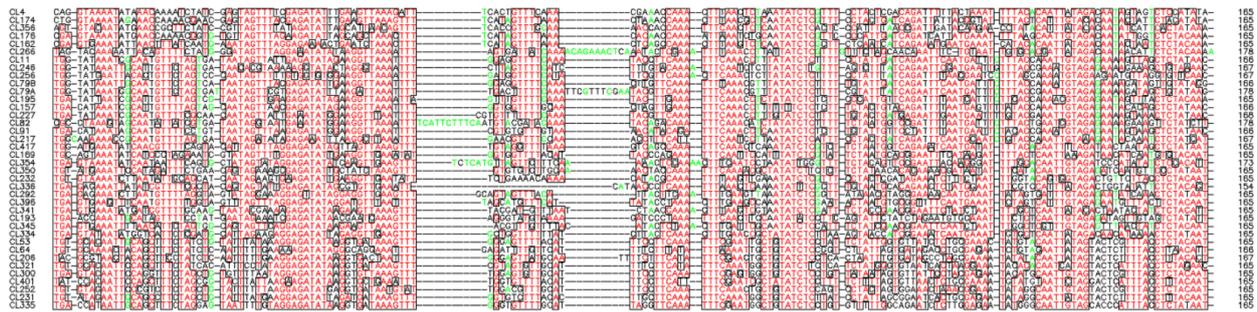
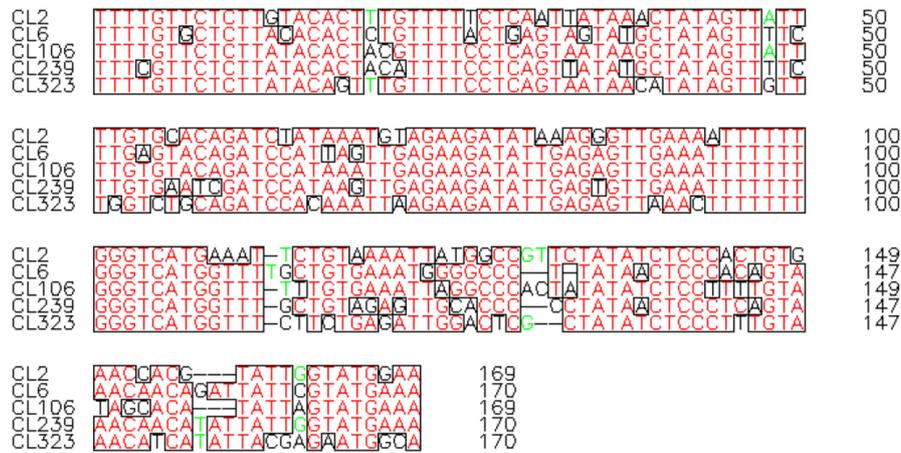


Figure 2. (a) Satellitome landscape of the satDNA families in *Rhodnius prolixus*. In the landscape, abundance vs. Kimura divergence from satDNA consensus sequences is plotted. (b–e) Landscape of most abundant satDNA families divided by their subfamilies. The order of satDNA families and subfamilies are in order of their position on a stacked histogram.

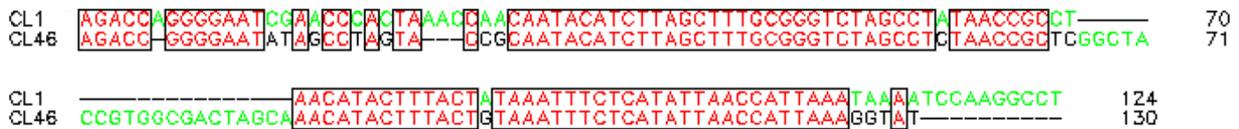
RproSat01-165



RproSat02-169



RproSat03-124



RproSat04-133

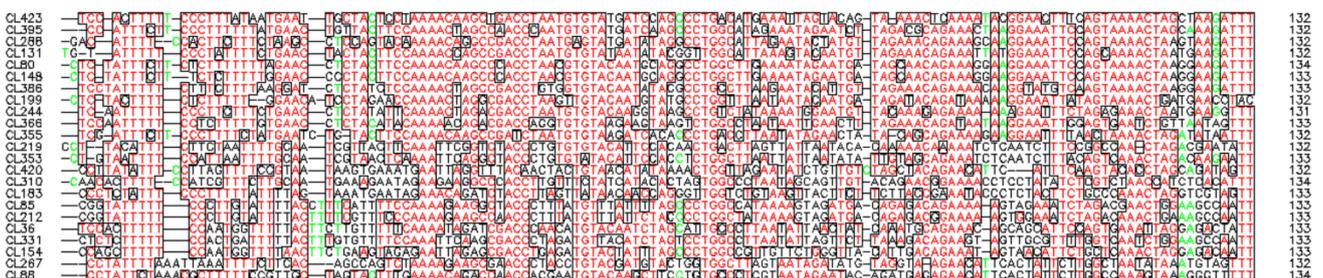


Figure 3. Consensus sequences alignments of the different subfamilies found for the most abundant satDNA families in *Rhodnius prolixus*. Boxed red letters correspond with conserved position among sequences. Less conserved positions are indicated with green and black letters.

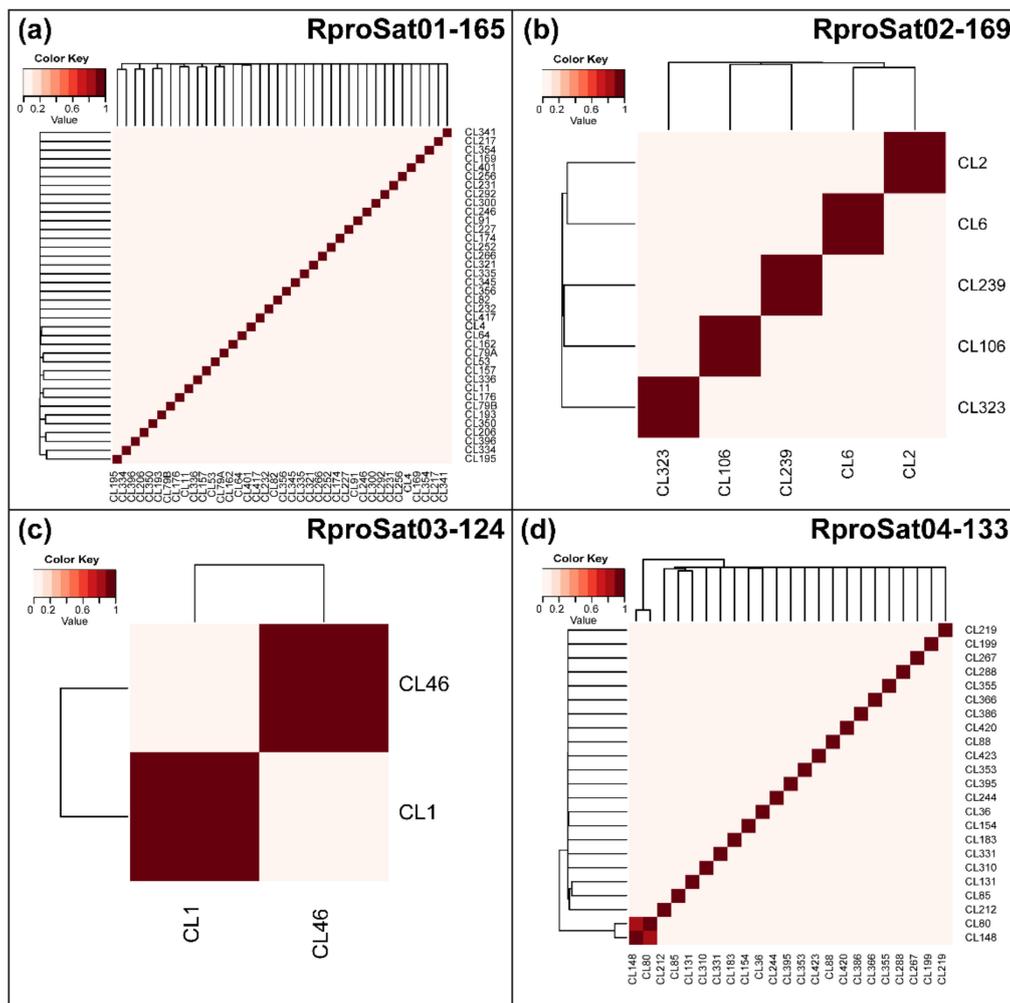


Figure 4. Heatmap representing the fraction of *Rhodnius prolixus* genome scaffolds shared by two subfamilies: (a) RproSat01-165, (b) RproSat02-169, (c) RproSat03-124 and (d) RproSat04-133. The figure shows that almost no subfamilies coincide in the same scaffold.

As discussed above, repetitive DNAs may not be well-represented in assembled genomes. In order to know how well-represented the satellitome is on *R. prolixus* genome assembly, pseudo reads were generated from it to estimate their abundance and divergence with RepeatMasker (Supplementary Table S1). Thereupon, just 5.6% of the assembled genome corresponded to satDNA sequences, and four families detected by RepeatExplorer2 were missing (Supplementary Table S1). This result is significantly lower than that obtained by us (8.05%) from the unassembled reads. Nevertheless, it is not possible to be sure if these repeated DNA sequences consist in a great part of those which were left out from the genome assembly, or if they may be collapsed. In any case, this shows the importance of the knowledge of satDNA abundance prior to the sequencing assembly of a Triatominae genome. Discrepancies of satDNA abundance estimations between our analysis and assembled genome are smaller in *R. prolixus* than in other insects [54]. This might be due to the fact that the *R. prolixus* satDNAs are scattered on the genome organized into small arrays. In species with large heterochromatic blocks, genome assembly will cover until those block edges, so intern repeats could not be assembled and the amounts of discarded sequences will be higher, underestimating the amount of satDNA.

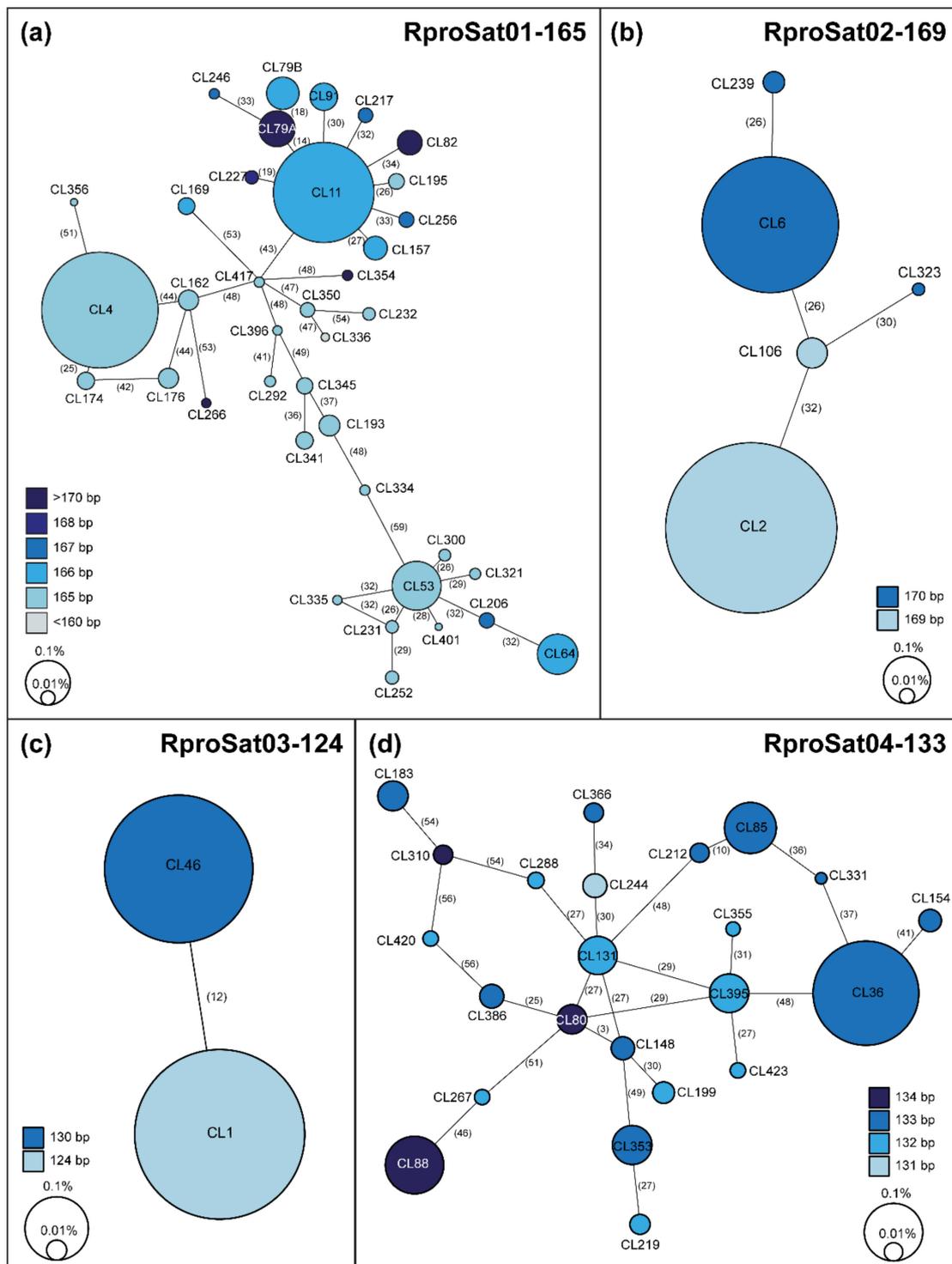


Figure 5. Minimum spanning networks for the four satDNA families forming superclusters in *Rhodnius prolixus*: (a) RproSat01-165, (b) RproSat02-169, (c) RproSat03-124 and (d) RproSat04-133. Numbers between brackets are the mutational steps. Each circle corresponds to a RepeatExplorer2 cluster or subfamily, where the size is proportional to its abundance in the genome. Colors denote the consensus monomer length.

Available NCBI *R. prolixus* raw reads from genomic DNA were also included in the analysis (SRR6749969, SRR6749971, SRR6749972 and SRR6749978). It is important to note that *R. prolixus* belong to a complex of cryptic species with the ability to hybridize. This issue could lead to erroneous interpretation of the data, as it has been recently revised [55]. Hence, the correct species classification was checked using the ribosomal internal transcribed

spacer 2 (ITS-2) region of each dataset (Supplementary Figure S3). The amount of satDNA in these four archives of sequencing data ranges between 8.81% and 9.62%, closer to our estimate (8.05%) than that obtained from the assembled genome (5.6%) (Supplementary Table S1, Figure S1). Notwithstanding, as already shown on other insect species [13,14,20], variations in the amount—and even the absence—of some satDNA families were found between individuals. In spite of this, the general aspect of the repeat landscape for each genome was conserved (Supplementary Figure S1b–e).

Chromosomal location of the most abundant satDNA families (over 1% of the genome) was performed by fluorescence in situ hybridization (FISH). Hybridization with satDNA probes showed disperse signals over euchromatin of all chromosomes, autosomes and both sex chromosomes (Figure 6). These cytogenetic results are in agreement with the molecular results. The analysis of the satDNA subfamilies in the assembled *R. prolixus* genome showed that these satDNA were present in a high number of scaffolds (Supplementary Figure S4), but the number of monomers in each scaffold was low, most of them with less than 50 repeats (Supplementary Figure S5). This hybridization pattern is similar to that found for other less abundant families, such as the four satDNA families shared between *R. prolixus* and *T. infestans* [11]. All data support the different composition of the heterochromatic Y chromosome between Triatomini and Rhodniini tribes [40,41]. The satDNA family TinfSat01-33 and (GATA)_n repeats are the main components of the *T. infestans* Y chromosome heterochromatin, and no other satDNA families were present in this chromosome [41]. On the contrary, the *R. prolixus* Y chromosome contains several satDNA families, in the same way as autosomes and the X chromosome. In *Triatoma* species, (GATA)_n repeats are especially accumulated on the Y chromosome, and these repeats seem to be the only repetitive DNA shared by the Y chromosomes of this genus [40,41]. However, (GATA)_n repeats are not abundant in the *R. prolixus* genome, and FISH with (GATA)_n repeats showed no signals on the Y chromosome (data not shown).

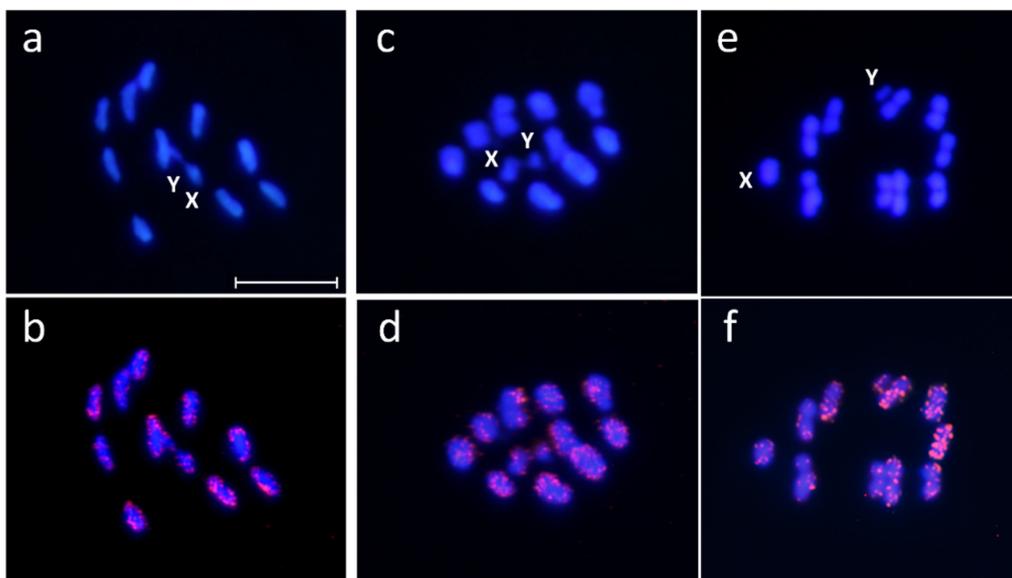


Figure 6. Chromosomal location of most abundant satDNA families of *Rhodnius prolixus*. Male meiotic metaphases stained with DAPI (a,c,e). Merged images of FISH with RproSat01-165 probe (b), RproSat02-169 probe (d) and RproSat03-124 (f). Scale bar = 10 μ m.

The possible role of transcripts of satDNA has been questioned in the past, but accumulation of evidence has changed that view. Currently, it is widely accepted that satellite non-coding RNAs might have functions in different cellular contexts, such as cancer, stress response, development or cell proliferation [32,33,56]. Hence, we have analyzed whether satDNA families of *R. prolixus* are transcribed in different tissues. Samples from two available RNA-seq experiments from tissues were selected. The first one was an RNA-seq

analysis of antenna, the main chemosensory structures in insects, performed in nymph, male and female (SRX1011796, SRX1011778 and SRX1011769, respectively), and the second one, an RNA-seq analysis performed in female and male gonads (SRX6380683 and SRX6380682, respectively). Due to the taxonomy identification issue commented on above, the ITS2 region sequence was obtained from each dataset in order to check the correct species classification (Supplementary Figure S3). After mapping dataset reads to our satDNA consensus sequences, satDNA families poorly represented were discarded and read count was normalized. Library preparation for RNA-sequencing is highly decisive for results. Therefore, library selection from antenna was random, while enrichment for messenger RNA sequences was applied at gonads library. Bearing that in mind, comparison between experiments should be made with caution since non-coding RNA could probably be under-represented on gonads samples, and satDNA transcription could be underestimated. We found that 33 satDNA families were transcribed at least in two samples (Figure 7 and Supplementary Figure S6, Table S2), showing four different patterns. The first pattern would correspond to satDNA families transcribed in all studied tissues (antenna, ovaries and testis), although with different proportions, such as RproSat02-169 and RproSat04-133. The second pattern would represent satDNA families highly transcribed in antenna, with less or no transcription in gonads, such as RproSat03-124 and RproSat26-146. The third pattern would correspond to satDNA families highly transcribed in gonads, such as RproSat19-201 and RproSat33-123. Finally, the fourth pattern would belong to satDNA families highly transcribed in one gonadal tissue only, such as RproSat06-136 (testis) and RproSat29-31 (ovaries). Additionally, the transcription level of satDNA families appears to be significantly correlated with their abundance, whether combined transcription (Spearman's correlation $r_s = 0.52$, $p = 0.002$) or tissue transcription (Supplementary Table S3) is considered. However, one exception should be pointed out, the Rpro34-415 family. This family, which represents only 0.003% of the genome, showed transcription levels similar to the most abundant families, being higher at male antennae and gonads. Together, those results indicated the satellitome is generally expressed in *R. prolixus*, although each family is transcribed at a different level and at a different pattern, suggesting that satDNA transcription could have a specific role in those tissue environments. Satellite DNA transcription is an accepted feature, as we commented on above, and it has been seen before in other insects, such as Coleoptera [57–59], Hymenoptera [60–62], Orthoptera [19], Lepidoptera [63] or Diptera [33,34,64]. In *D. melanogaster*, Mills et al. [34] found that satDNA derived from (AAGA)_n tandem repeat is highly transcribed at neuron and testis, being necessary for male fertility. Another *D. melanogaster* satDNA, the 1.688 satDNA family, contains a member with a dense X-linked distribution (1.688^X), which plays an important role in marking the X chromosome during dosage compensation [65]. In males, the small interfering RNAs generated from 1.688^X sequences promote X localization of the male-specific lethal complex, which increases X-linked gene expression by modification of chromatin [38,63]. In *D. buzzatti* and *D. mojavensis*, satDNA families pBuM and CDSTR198 were transcribed, particularly in pupae and male tissues, even when both satDNAs have different genomic environment (heterochromatin and euchromatin, respectively) [64]. Outside of insects, satDNA has proven their importance. For instance, in humans, SATIII is associated with cell response to stress, recruiting RNA-processing factor and downregulating cellular transcription [32]. Our findings suggest that satDNA transcription might have functionality on the *R. prolixus* genome and open the door to future studies to address whether those satDNAs contribute to gene regulation or chromatin modulation.

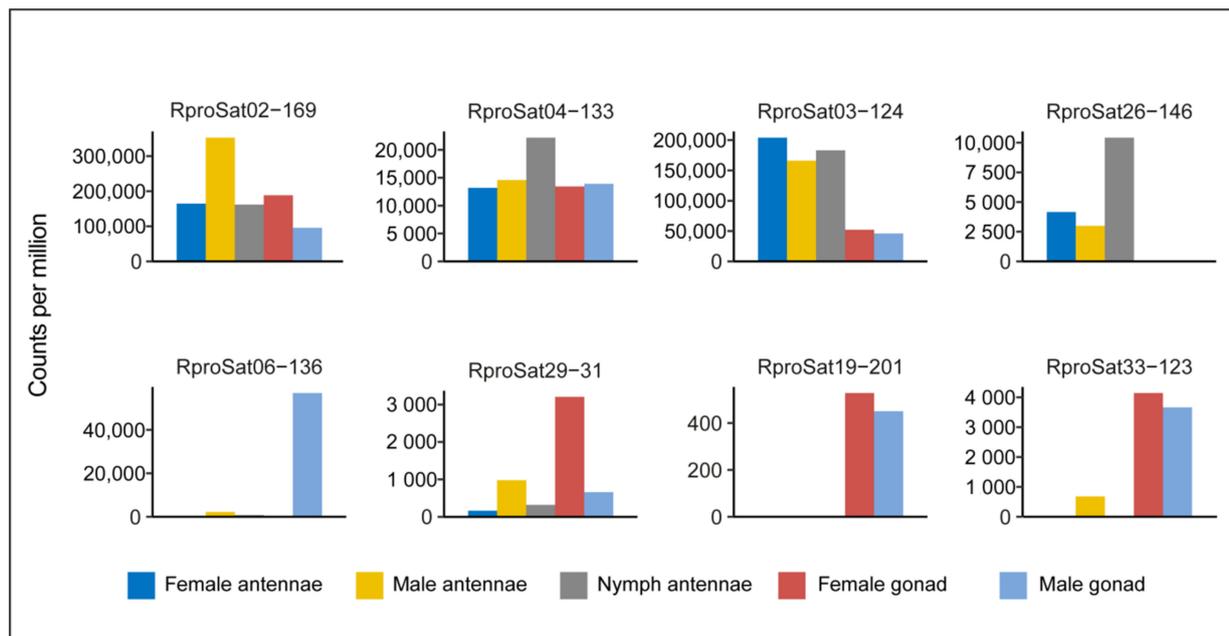


Figure 7. Examples of the four transcription patterns shown by satDNA families of *R. prolixus* in different tissues.

3. Materials and Methods

3.1. Samples, DNA Extraction and Chromosome Preparation

Domestic *R. prolixus* individuals were collected from Colombia (Department Casanare, Municipality Yopal). DNA extraction for sequencing was performed from the head of an adult male using the NucleoSpin Tissue kit (Macherey-Nagel Co., Düren, Germany). For cytogenetic analysis, adult males were dissected, and the testes were fixed in an ethanol–glacial acetic acid mixture (3:1) and stored at -20°C . Squashes were made in a 50% acetic acid drop, coverslips were removed after freezing in liquid nitrogen and the slides were air-dried and stored at 4°C [13].

3.2. DNA Sequencing and Graph-Based Clustering of Sequencing Reads

Low-coverage sequencing was performed using the DNBSseq™ sequencing platform at BGI, Hong Kong, which yielded 1.2 Gbp of PE150 reads. Raw reads were first quality trimmed with Trimmomatic [66]. Fastq files were modified, i.e., discarded reads containing Ns, fastq to fasta, with the FastX toolkit (http://hannonlab.cshl.edu/fastx_toolkit, accessed on 3 May 2021). Sequences corresponding to mitochondrial DNA were eliminated from the repeat analysis. NCBI-deposited genomes' raw reads were downloaded using prefetch and fastq-dump tools (SRR6749969, SRR6749971, SRR6749972 and SRR6749978).

As *R. prolixus* species determination can be tricky, the ITS2 ribosomal region sequence was extracted from raw data by mapping with bbmap (sourceforge.net/projects/bbmap/, accessed on 3 May 2021) against the *R. prolixus* ITS2 sequence (DQ118978). A phylogenetic maximum likelihood tree was performed using all the available sequences from other *Rhodnius* species from GenBank. Alignment was performed with MAFFT [67] and a phylogenetic tree was constructed using the ML method in RaxML [68]. The tree was edited with iTOL [69].

Graph-based clustering was performed using the RepeatExplorer2 pipeline, which includes the TAREAN analysis, on the Galaxy portal environment (<https://repeatexplorer-elixir.cerit-sc.cz>, accessed on 3 May 2021). A set of six million paired-end reads were randomly selected for clustering analysis. Clusters containing satDNAs were identified based on the graph topology with sphere or ring-like shape. For each candidate cluster, we chose the longest and the highest coverage contig assembled by RepeatExplorer2 to generate a dot plot with the Dotmatcher tool (<http://emboss.bioinformatics.nl/cgi-bin/>

[emboss/dotmatcher/](#), accessed on 3 May 2021). Afterwards, contigs were separated in monomers to align them using MAFFT and generate a consensus sequence. All satDNA consensus sequences were submitted to NCBI (Acc. Numbers MW827131-MW827167). When all satDNA clusters were annotated and a monomer consensus was obtained, similarity between them was tested using the Basic Local Alignment Search Tool (BLAST), with `blastn` and `-e 0.001` options. Additionally, divergence and abundance for each satDNA were calculated using RepeatMasker (<http://www.repeatmasker.org>, accessed on 3 May 2021) with “-a” option and the RMBlast search engine. For this, we randomly selected a million reads and aligned against the total collection of satDNA dimers or monomer concatenations of approximately 200 bp length. We estimated the average divergence and generated a satellite landscape considering distances from the sequences applying the Kimura 2-parameter model with the perl script `calcDivergenceFromAlign.pl` and `createRepeatLandscape.pl` from the RepeatMasker suite. Subfamilies’ consensus alignments were plotted with Prettyplot (<https://www.bioinformatics.nl/cgi-bin/emboss/prettyplot>, accessed on 3 May 2021) and minimum spanning networks were built according to the pairwise distance of subfamilies’ consensus sequences and considering the relative abundances using PopART v1.7 [70,71].

3.3. Transcription of Satellite DNA

We downloaded *R. prolixus* RNA-seq data from two bio-projects from the NCBI database: an antenna transcriptome project (SRX1011796, SRX1011778, SRX1011769) and a sex differentiation of gonad transcription project (SRX6380683, SRX6380682). To check that insects used were *R. prolixus*, ribosomal ITS-2 spacer reads were extracted by mapping reads from all sets with `bbmap` (sourceforge.net/projects/bbmap/, accessed on 3 May 2021) against the *R. prolixus* ITS-2 spacer (DQ118978).

Raw RNA-seq data from all tissues were mapped to each satDNA consensus using `bbmap` to obtain the output as a sam file. We used the same satDNA dimers or monomer concatenations as used for abundance analysis as references. The aligned reads were counted using `samtools` [72]. Read counts were analyzed in R base version 4.0.1 [73] using the `edgeR` package [74]. In brief, read counts from all tissues were normalized to counts per million (CPM) and filtered satDNAs with more than 50 CPM in at least 2 samples. Correlation between satDNA transcription and abundance was analyzed by means of Spearman correlation and graphs were obtained with the `ggplot2` package [75] in R.

3.4. Cytogenetic Mapping

The consensus sequences of the most abundant satDNA families, over 1% of the genome, were used to design a set of oligonucleotides (Supplementary Table S4). These labeled oligonucleotides were used as probes (final concentration of 5 ng/mL in 50% formamide) to perform fluorescence in situ hybridizations (FISH) according to the procedure described by Palomeque et al. [76] and Pita et al. [13]. The fluorescent immunological detection was carried out using the avidin-FITC/anti-avidin-biotin system with three amplification rounds. Slides were mounted in Vectashield-DAPI (Vector Laboratories, Burlingame, CA, USA). DAPI, in the antifade solution, was used to counterstain the chromosomes. Images were taken with a BX51 Olympus® fluorescence microscope (Olympus, Hamburg, Germany) equipped with a CCD camera (Olympus® DP70) and processed using Adobe® Photoshop® software.

3.5. *Rhodnius Prolixus* Genome Assembly satDNA Families Searches

Rhodnius prolixus genome assembly was downloaded from VectorBase (<https://vectorbase.org/vectorbase/app/>, accessed on 3 May 2021), which is the same as that available in GenBank: GCA_000181055.3.

To include this data in the RepeatExplorer analysis, a simulated Illumina paired-end 150 bp reads run was performed using ART [77].

The search for the described satDNA families was carried on with a Basic Local Alignment Search Tool (BLAST) analysis. Only those hits with 90% of query coverage per HSP and a 90% identity were taken into account. To render the heatmaps, BASH text-processing tools and R base version 4.0.1 [73] with gplots [78] and RColorBrewer [79] packages were employed. The same BLAST results were used to evaluate the amount of satDNA families by contig or scaffold, and figures were obtained with the ggplot2 package [74] in R.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms22116052/s1>. Figure S1: Repeat composition of *Rhodnius prolixus* genome obtained from RepeatExplorer2 annotation (a). Satellitome landscape of the satDNA families in NCBI samples (b–d). Figure S2: Alignment of consensus sequences of satDNA families with similarity: RproSat07-375 and RproSat09-499 (a), and RproSat22-980 and Rpro24-675 (b). Similar positions are marked with red font and red boxes indicate regions with high similarity. Sequences are numbered according to the original position of consensus sequences. Figure S3: ML phylogenetic tree depicting the position of samples used in our study within *Rhodnius prolixus*. Figure S4: Number of scaffolds where subfamilies are distributed on *Rhodnius prolixus* genome assembly. Figure S5: Monomer distribution of most abundant satDNA families in *Rhodnius prolixus* assembled genome. Histograms relate number of scaffolds and monomers. For all families, long arrays are present on few scaffolds. Figure S6: SatDNA transcription in different tissues. Table S1: Estimations of genome abundance and nucleotide divergence of satDNA families and subfamilies found in *Rhodnius prolixus* genome sequenced in this work (Rpro) and in pseudo-reads generated from *Rhodnius prolixus* genome assembly (GCA_000181055.3) (RproSim). Subfamilies' names are followed by the size of their consensus sequence. Table also shows the same data for available NCBI *Rhodnius prolixus* raw reads from genomic DNA: Rpro1 (SRR6749969), Rpro2 (SRR6749971), Rpro3 (SRR6749972) and Rpro4 (SRR6749978). Table S2: SatDNA transcription on different tissues. Combined column refers to sum of all samples' transcription for each satDNA family. Table S3: Results of Spearman correlation between satDNA transcription and abundance in different samples. Table S4: Designed oligonucleotides for three main satellite DNA families of *Rhodnius prolixus*.

Author Contributions: E.E.M., F.P., T.P., P.L. and S.P. conceived and designed the experiment; E.E.M., F.P., T.P., P.L. and S.P. performed the experiment; E.E.M., P.L. and S.P. analyzed the data; E.E.M., F.P., T.P., P.L. and S.P. acquired the funds; E.E.M., P.L. and S.P. wrote the draft. All authors have read and agreed to the published version of the manuscript.

Funding: This article was written within the framework of a project funded by “Comisión Sectorial de Investigación Científica” (No. 160, CSIC-Udelar, Uruguay) assigned to F.P., a project funded by “Programa Operativo FEDER Andalucía 2014–2020” assigned to P.L., the program “Plan de Apoyo a la Investigación 2019–2020, Acción 1” (Univ. Jaén, Spain), Group RNM924, assigned to P.L., and the program “Plan de Apoyo a la Investigación 2017–2018, Acción 6” (Univ. Jaén, Spain) assigned to E.E.M. and P.L. F.P. and S.P. are members of the “Sistema Nacional de Investigadores (ANII)” and researchers from the “Programa de Desarrollo de las Ciencias Básicas (PEDECIBA)”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All satDNA consensus sequences were submitted to NCBI (Acc. Numbers MW827131–MW827167).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mesquita, R.D.; Vionette-Amaral, R.J.; Lowenberger, C.; Rivera-Pomar, R.; Monteiro, F.A.; Minx, P.; Spieth, J.; Carvalho, A.B.; Panzera, F.; Lawson, D.; et al. Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 14936–14941. [[CrossRef](#)] [[PubMed](#)]
2. Panzera, F.; Pérez, R.; Panzera, Y.; Ferrandis, I.; Ferreiro, M.J.; Calleros, L. Cytogenetics and genome evolution in the subfamily Triatominae (Hemiptera, Reduviidae). *Cytogenet. Genome Res.* **2010**, *128*, 77–87. [[CrossRef](#)]
3. Maumus, F.; Quesneville, H. Ancestral repeats have shaped epigenome and genome composition for millions of years in *Arabidopsis thaliana*. *Nat. Commun.* **2014**, *5*, 4104. [[CrossRef](#)] [[PubMed](#)]

4. Castro, M.R.J.; Goubert, C.; Monteiro, F.A.; Vieira, C.; Carareto, C.M.A. Homology-free detection of transposable elements unveils their dynamics in three ecologically distinct *Rhodnius* species. *Genes* **2020**, *11*, 170. [[CrossRef](#)]
5. Goubert, C.; Modolo, L.; Vieira, C.; ValienteMoro, C.; Mavingui, P.; Boulesteix, M. De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). *Genome Biol. Evol.* **2015**, *7*, 1192–1205. [[CrossRef](#)] [[PubMed](#)]
6. Monteiro, F.A.; Weirauch, C.; Felix, M.; Lazoski, C.; Abad-Franch, F. Chapter five—Evolution, systematics, and biogeography of the Triatominae, vectors of Chagas disease. In *Advances in Parasitology*; Rollinson, D., Stothard, J.R., Eds.; Academic Press: Cambridge, MA, USA, 2018; Volume 99, pp. 265–344. [[CrossRef](#)]
7. Fernández-Medina, R.D.; Granzotto, A.; Ribeiro, J.M.; Carareto, C.M.A. Transposition burst of mariner-like elements in the sequenced genome of *Rhodnius prolixus*. *Insect Biochem. Mol. Biol.* **2016**, *69*, 14–24. [[CrossRef](#)] [[PubMed](#)]
8. Garrido-Ramos, M.A. Satellite DNA: An evolving topic. *Genes* **2017**, *8*, 230. [[CrossRef](#)]
9. Lower, S.S.; McGurk, M.P.; Clark, A.G.; Barbash, D.A. Satellite DNA evolution: Old ideas, new approaches. *Curr. Opin. Genet. Dev.* **2018**, *48*, 70–78. [[CrossRef](#)]
10. Ruíz-Ruano, F.J.; Castillo-Martínez, J.; Cabrero, J.; Gómez, R.; Camacho, J.P.M.; López-León, M.D. High-throughput analysis of satellite DNA in the grasshopper *Pyrgomorpha conica* reveals abundance of homologous and heterologous higher-order repeats. *Chromosoma* **2018**, *127*, 323–340. [[CrossRef](#)] [[PubMed](#)]
11. Pita, S.; Mora, P.; Vela, J.; Palomeque, T.; Sánchez, A.; Panzera, F.; Lorite, P. Comparative Analysis of repetitive DNA between the main vectors of Chagas disease: *Triatoma infestans* and *Rhodnius prolixus*. *Int. J. Mol. Sci.* **2018**, *19*, 1277. [[CrossRef](#)]
12. Bardella, V.B.; da Rosa, J.A.; Vanzela, A.L.L. Origin and distribution of AT-rich repetitive DNA families in *Triatoma infestans* (Heteroptera). *Infect. Genet. Evol.* **2014**, *23*, 106–114. [[CrossRef](#)] [[PubMed](#)]
13. Pita, S.; Panzera, F.; Mora, P.; Vela, J.; Cuadrado, A.; Sánchez, A.; Palomeque, T.; Lorite, P. Comparative repeatome analysis on *Triatoma infestans* Andean and Non-Andean lineages, main vector of Chagas disease. *PLoS ONE* **2017**, *12*, e0181635. [[CrossRef](#)]
14. Ruíz-Ruano, F.J.; López-León, M.D.; Cabrero, J.; Camacho, J.P.M. High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci. Rep.* **2016**, *6*, 28333. [[CrossRef](#)] [[PubMed](#)]
15. Bardella, V.B.; Milani, D.; Cabral-de-Mello, D.C. Analysis of *Holhymenia histrio* genome provides insight into the satDNA evolution in an insect with holocentric chromosomes. *Chromosome Res.* **2020**, *28*, 369–380. [[CrossRef](#)] [[PubMed](#)]
16. Novák, P.; Neumann, P.; Macas, J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat. Protoc.* **2020**, *15*, 3745–3776. [[CrossRef](#)] [[PubMed](#)]
17. Novák, P.; Ávila Robledillo, L.; Koblížková, A.; Vrbová, I.; Neumann, P.; Macas, J. TAREAN: A computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res.* **2017**, *45*, e111. [[CrossRef](#)]
18. Ruíz-Ruano, F.J.; Cabrero, J.; López-León, M.D.; Camacho, J.P.M. Satellite DNA content illuminates the ancestry of a supernumerary (B) chromosome. *Chromosoma* **2017**, *126*, 487–500. [[CrossRef](#)] [[PubMed](#)]
19. Palacios-Gimenez, O.M.; Bardella, V.B.; Lemos, B.; Cabral-de-Mello, D.C. Satellite DNAs are conserved and differentially transcribed among *Gryllus* cricket species. *DNA Res.* **2018**, *25*, 137–147. [[CrossRef](#)]
20. Silva, B.S.M.L.; Heringer, P.; Dias, G.B.; Svartman, M.; Kuhn, G.C.S. De novo identification of satellite DNAs in the sequenced genomes of *Drosophila virilis* and *D. americana* using the RepeatExplorer and TAREAN pipelines. *PLoS ONE* **2019**, *14*, e0223466. [[CrossRef](#)]
21. Mora, P.; Vela, J.; Ruíz-Ruano, F.J.; Ruiz-Mena, A.; Montiel, E.E.; Palomeque, T.; Lorite, P. Satellitome analysis in the ladybird beetle *Hippodamia variegata* (Coleoptera, Coccinellidae). *Genes* **2020**, *11*, 783. [[CrossRef](#)]
22. Utsunomia, R.; Ruíz-Ruano, F.J.; Silva, D.M.Z.A.; Serrano, É.A.; Rosa, I.F.; Scudeler, P.E.S.; Hashimoto, D.T.; Oliveira, C.; Camacho, J.P.M.; Foresti, F. A Glimpse into the Satellite DNA Library in Characidae Fish (Teleostei, Characiformes). *Front. Genet.* **2017**, *8*, 103. [[CrossRef](#)]
23. Utsunomia, R.; Silva, D.; Ruíz-Ruano, F.J.; Goes, C.; Melo, S.; Ramos, L.P.; Oliveira, C.; Porto-Foresti, F.; Foresti, F.; Hashimoto, D.T. Satellitome landscape analysis of *Megaleporinus macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. *Sci. Rep.* **2019**, *9*, 5856. [[CrossRef](#)]
24. Dos Santos, R.G.; Calegari, R.M.; Silva, D.M.Z.A.; Ruiz-Ruano, F.J.; Melo, S.; Oliveira, C.; Foresti, F.; Uliano-Silva, M.; Porto-Foresti, F.; Utsunomia, R. A Long-Term Conserved Satellite DNA That Remains Unexpanded in Several Genomes of Characiformes Fish Is Actively Transcribed. *Genome Biol. Evol.* **2021**, *13*, evab002. [[CrossRef](#)]
25. Ruíz-Ruano, F.J.; Navarro-Domínguez, B.; Camacho, J.P.M.; Garrido-Ramos, M.A. Characterization of the satellitome in lower vascular plants: The case of the endangered fern *Vandenboschia speciosa*. *Ann. Bot.* **2019**, *123*, 587–599. [[CrossRef](#)]
26. Belyayev, A.; Jandová, M.; Josefiová, J.; Kalendar, R.; Mahelka, V.; Mandák, B.; Krak, K. The major satellite DNA families of the diploid *Chenopodium album* aggregate species: Arguments for and against the “library hypothesis”. *PLoS ONE* **2020**, *15*, e0241206. [[CrossRef](#)]
27. Sader, M.; Vaio, M.; Cauz-Santos, L.A.; Dornelas, M.C.; Vieira, M.L.C.; Melo, N.; Pedrosa-Harand, A. Large vs small genomes in Passiflora: The influence of the mobilome and the satellitome. *Planta* **2021**, *253*, 86. [[CrossRef](#)]
28. Heitkam, T.; Schulte, L.; Weber, B.; Liedtke, S.; Breitenbach, S.; Kögler, A.; Morgenstern, K.; Brückner, M.; Tröber, U.; Wolf, H.; et al. Comparative repeat profiling of two closely related conifers (*Larix decidua* and *Larix kaempferi*) reveals high genome similarity with only few fast-evolving satellite DNAs. *bioRxiv* **2021**. [[CrossRef](#)]
29. Thakur, J.; Packiaraj, J.; Henikoff, S. Sequence, chromatin and evolution of satellite DNA. *Int. J. Mol. Sci.* **2021**, *22*, 4309. [[CrossRef](#)]

30. Cabral-de-Mello, D.C.; Zrzavá, M.; Kubičková, S.; Rendón, P.; Marec, F. The role of satellite DNAs in genome architecture and sex chromosome evolution in Crambidae moths. *Front. Genet.* **2021**, *12*, 661417. [[CrossRef](#)]
31. Perea-Resa, C.; Blower, M.D. Satellite transcripts locally promote centromere formation. *Dev. Cell* **2017**, *42*, 201–202. [[CrossRef](#)]
32. Louzada, S.; Lopes, M.; Ferreira, D.; Adegá, F.; Escudeiro, A.; Gama-Carvalho, M.; Chaves, R. Decoding the role of satellite DNA in genome architecture and plasticity—an evolutionary and clinical affair. *Genes* **2020**, *11*, 72. [[CrossRef](#)]
33. Halbach, R.; Miesen, P.; Joosten, J.; Taşkoprü, J.; Rondeel, I.; Pennings, B.; Vogels, C.B.F.; Merklings, S.H.; Koenraadt, C.J.; Lambrechts, L.; et al. A satellite repeat-derived piRNA controls embryonic development of *Aedes*. *Nature* **2020**, *580*, 274–277. [[CrossRef](#)]
34. Mills, W.K.; Lee, Y.C.G.; Kochendoerfer, A.M.; Dunleavy, E.M.; Karpen, G.H. RNA from a simple-tandem repeat is required for sperm maturation and male fertility in *Drosophila melanogaster*. *Elife* **2019**, *8*, 48940. [[CrossRef](#)] [[PubMed](#)]
35. Pezer, Z.; Ugarković, D. Satellite DNA-associated siRNAs as mediators of heat shock response in insects. *RNA Biol.* **2012**, *9*, 587–595. [[CrossRef](#)]
36. Palomeque, T.; Lorite, P. Satellite DNA in insects: A review. *Heredity* **2008**, *100*, 564–573. [[CrossRef](#)]
37. Dalíková, M.; Zrzavá, M.; Kubičková, S.; Marec, F. W-enriched satellite sequence in the Indian meal moth, *Plodia interpunctella* (Lepidoptera, Pyralidae). *Chromosome Res.* **2017**, *25*, 241–252. [[CrossRef](#)]
38. Shatskikh, A.S.; Kotov, A.A.; Adashev, V.E.; Bazylev, S.S.; Olenina, L.V. Functional significance of satellite DNAs: Insights from *Drosophila*. *Front. Cell Dev. Biol.* **2020**, *8*, 312. [[CrossRef](#)] [[PubMed](#)]
39. De Souza, R.C.M.; Gorla, D.E.; Chame, M.; Jaramillo, N.; Monroy, C.; Diotaiuti, L. Chagas disease in the context of the 2030 agenda: Global warming and vectors. *Mem. Inst. Oswaldo Cruz* **2021**, *116*, e200479. [[CrossRef](#)]
40. Pita, S.; Panzera, F.; Sánchez, A.; Panzera, Y.; Palomeque, T.; Lorite, P. Distribution and evolution of repeated sequences in genomes of Triatominae (Hemiptera-Reduviidae) inferred from genomic in situ hybridization. *PLoS ONE* **2014**, *9*, e114298. [[CrossRef](#)] [[PubMed](#)]
41. Pita, S.; Lorite, P.; Vela, J.; Mora, P.; Palomeque, T.; Thi, K.P.; Panzera, F. Holocentric chromosome evolution in kissing bugs (Hemiptera-Reduviidae-Triatominae): Diversification of repeated sequences. *Parasit. Vectors* **2017**, *10*, 410. [[CrossRef](#)]
42. Fry, K.; Salser, W. Nucleotide sequences of HS-alpha satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell* **1977**, *12*, 1069–1084. [[CrossRef](#)]
43. Mestrovic, N.; Plohl, M.; Mravinac, B.; Ugarković, D. Evolution of satellite DNAs from the genus *Palorus*—Experimental evidence for the “library” hypothesis. *Mol. Biol. Evol.* **1998**, *15*, 1062–1068. [[CrossRef](#)]
44. De Paula, A.S.; Barreto, C.; Telmo, M.C.M.; Diotaiuti, L.; Galvão, C. Historical biogeography and the evolution of hematophagy in Rhodniini (Heteroptera: Reduviidae: Triatominae). *Front. Ecol. Evol.* **2021**, *9*, 660151. [[CrossRef](#)]
45. Bachmann, L.; Sperlich, D. Gradual evolution of a specific satellite DNA family in *Drosophila ambigua*, *D. tristis*, and *D. obscura*. *Mol. Biol. Evol.* **1993**, *10*, 647–659. [[CrossRef](#)]
46. Palacios-Gimenez, O.M.; Milani, D.; Song, H.; Marti, D.A.; López-León, M.D.; Ruíz-Ruano, F.J.; Camacho, J.P.M.; Cabral-de-Mello, D.C. Eight million years of satellite DNA evolution in grasshoppers of the genus *Schistocerca* illuminate the ins and outs of the library hypothesis. *Genome Biol. Evol.* **2020**, *12*, 88–102. [[CrossRef](#)] [[PubMed](#)]
47. Lorite, P.; Muñoz-López, M.; Carrillo, J.; Sanlloriente, O.; Vela, J.; Mora, P.; Tinaut, A.; Torres, M.I.; Palomeque, T. Concerted evolution, a slow process for ant satellite DNA: Study of the satellite DNA in the *Aphaenogaster* genus (Hymenoptera, Formicidae). *Org. Divers. Evol.* **2017**, *17*, 595–606. [[CrossRef](#)]
48. Abad, J.P.; Carmena, M.; Baars, S.; Saunders, R.D.; Glover, D.M.; Ludeña, P.; Sentis, C.; Tyler-Smith, C.; Villasante, A. Dodeca satellite: A conserved G+C-rich satellite from the centromeric heterochromatin of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 4663–4667. [[CrossRef](#)] [[PubMed](#)]
49. Šatović, E.; Plohl, M. Distribution of DTHS₃ satellite DNA across 12 bivalve species. *J. Genet.* **2018**, *97*, 575–580. [[CrossRef](#)] [[PubMed](#)]
50. Pita, S.; Panzera, F.; Mora, P.; Vela, J.; Palomeque, T.; Lorite, P. The presence of the ancestral insect telomeric motif in kissing bugs (Triatominae) rules out the hypothesis of its loss in evolutionarily advanced Heteroptera (Cimicomorpha). *Comp. Cytogenet.* **2016**, *10*, 427–437. [[CrossRef](#)]
51. Panzera, F.; Dujardin, J.P.; Nicolini, P.; Caraccio, M.N.; Rose, V.; Tellez, T.; Bermúdez, H.; Barges, M.D.; Mas Coma, S.; O’Connor, J.E.; et al. Genomic changes of Chagas disease vector, South America. *Emerg. Infect. Dis.* **2004**, *10*, 438–446. [[CrossRef](#)]
52. Panzera, F.; Ferreira, M.J.; Pita, S.; Calleros, L.; Pérez, R.; Basmadján, Y.; Guevara, Y.; Brenière, S.F.; Panzera, Y. Evolutionary and dispersal history of *Triatoma infestans*, main vector of Chagas disease, by chromosomal markers. *Infect. Genet. Evol.* **2014**, *27*, 105–113. [[CrossRef](#)] [[PubMed](#)]
53. Lorite, P.; Carrillo, J.A.; Aguilar, J.A.; Palomeque, T. Isolation and characterization of two families of satellite DNA with repetitive units of 135 bp and 2.5 kb in the ant *Monomorium subopacum* (Hymenoptera, Formicidae). *Cytogen. Genome Res.* **2004**, *105*, 83–92. [[CrossRef](#)]
54. Wang, S.; Lorenzen, M.D.; Beaman, R.W.; Brown, S.J. Analysis of repetitive DNA distribution patterns in the *Tribolium castaneum* genome. *Genome Biol.* **2008**, *9*, R61. [[CrossRef](#)]
55. Brito, R.N.; Souza, R.C.M.; Abad-Franch, F. Dehydration-stress resistance in two sister, cryptic *Rhodnius* species—*Rhodnius prolixus* and *Rhodnius robustus* genotype I (Hemiptera: Reduviidae). *J. Med. Entomol.* **2019**, *56*, 1019–1026. [[CrossRef](#)]

56. Ferreira, D.; Escudeiro, A.; Adegas, F.; Chaves, R. DNA Methylation patterns of a satellite non-coding sequence—FA-SAT in cancer cells: Its expression cannot be explained solely by DNA methylation. *Front. Genet.* **2019**, *10*, 101. [[CrossRef](#)] [[PubMed](#)]
57. Pezer, Z.; Ugarković, D. Transcription of pericentromeric heterochromatin in beetles—Satellite DNAs as active regulatory elements. *Cytogenet. Genome Res.* **2009**, *124*, 268–276. [[CrossRef](#)] [[PubMed](#)]
58. Feliciello, I.; Akrap, I.; Ugarković, D. Satellite DNA modulates gene expression in the beetle *Tribolium castaneum* after heat stress. *PLoS Genet.* **2015**, *11*, e1005466. [[CrossRef](#)]
59. Mora, P.; Vela, J.; Ruiz-Mena, A.; Palomeque, T.; Lorite, P. Characterization and transcriptional analysis of a subtelomeric satellite DNA family in the ladybird beetle *Henosepilachna argus* (Coleoptera: Coccinellidae). *Eur. J. Entomol.* **2017**, *114*, 481–487. [[CrossRef](#)]
60. Rouleux-Bonnin, F.; Renault, S.; Bigot, Y.; Periquet, G. Transcription of four satellite DNA subfamilies in *Diprion pini* (Hymenoptera, Symphyta, Diprionidae). *Eur. J. Biochem.* **1996**, *238*, 752–759. [[CrossRef](#)] [[PubMed](#)]
61. Renault, S.; Rouleux-Bonnin, F.; Periquet, G.; Bigot, Y. Satellite DNA transcription in *Diadromus pulchellus* (Hymenoptera). *Insect Biochem. Mol. Biol.* **1999**, *29*, 103–111. [[CrossRef](#)]
62. Lorite, P.; Renault, S.; Rouleux-Bonnin, F.; Bigot, S.; Periquet, G.; Palomeque, T. Genomic organization and transcription of satellite DNA in the ant *Aphaenogaster subterranea* (Hymenoptera, Formicidae). *Genome* **2002**, *45*, 609–616. [[CrossRef](#)] [[PubMed](#)]
63. Věchtová, P.; Dalíková, M.; Sýkorová, M.; Žurovcová, M.; Füssy, Z.; Zrzavá, M. CpSAT-1, a transcribed satellite sequence from the codling moth, *Cydia pomonella*. *Genetica* **2016**, *144*, 385–395. [[CrossRef](#)] [[PubMed](#)]
64. Lima, L.G.; Svartman, M.; Kuhn, G.C.S. Dissecting the Satellite DNA landscape in three cactophilic *Drosophila* sequenced genomes. *G3-Genes Genom. Genet.* **2017**, *7*, 2831–2843. [[CrossRef](#)]
65. Joshi, S.S.; Meller, V.H. Satellite repeats identify X chromatin for dosage compensation in *Drosophila melanogaster* males. *Curr. Biol.* **2017**, *27*, 1393–1402. [[CrossRef](#)] [[PubMed](#)]
66. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
67. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [[CrossRef](#)]
68. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)]
69. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [[CrossRef](#)] [[PubMed](#)]
70. Leigh, J.W.; Bryant, D. PopART: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **2015**, *6*, 1110–1116. [[CrossRef](#)]
71. Bandelt, H.; Forster, P.; Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **1999**, *16*, 37–48. [[CrossRef](#)]
72. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)] [[PubMed](#)]
73. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019. Available online: <https://www.R-project.org/> (accessed on 3 May 2021).
74. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [[CrossRef](#)] [[PubMed](#)]
75. Wickham, H. ggplot2. *WIREs Comp. Stat.* **2011**, *3*, 180–185. [[CrossRef](#)]
76. Palomeque, T.; Muñoz-López, 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.* **2005**, *13*, 795–807. [[CrossRef](#)]
77. Huang, W.; Li, L.; Myers, J.R.; Marth, G.T. ART: A next-generation sequencing read simulator. *Bioinformatics* **2012**, *28*, 593–594. [[CrossRef](#)]
78. Warnes, G.R.; Bolker, B.; Bonebakker, L.; Gentleman, R.; Huber, W.; Liaw, A.; Lumley, T.; Maechler, M.; Magnusson, A.; Moeller, S.; et al. ggplots: Various R Programming Tools for Plotting Data. R Package Version 3.1.1. 2020. Available online: <https://CRAN.R-project.org/package=ggplots> (accessed on 3 May 2021).
79. Neuwirth, E. RColorBrewer: ColorBrewer Palettes. R Package Version 1.1-2. 2014. Available online: <https://CRAN.R-project.org/package=RColorBrewer> (accessed on 3 May 2021).