

ARTICLE



Out of patterns, the euchromatic B chromosome of the grasshopper *Abracris flavolineata* is not enriched in high-copy repeats

Diogo Milani¹, Francisco J. Ruiz-Ruano^{2,3}, Juan Pedro M. Camacho⁴ and Diogo C. Cabral-de-Mello¹✉

© The Author(s), under exclusive licence to The Genetics Society 2021

In addition to the normal set of standard (A) chromosomes, some eukaryote species harbor supernumerary (B) chromosomes. In most cases, B chromosomes show differential condensation with respect to A chromosomes and display dark C-bands of heterochromatin, and some of them are highly enriched in repetitive DNA. Here we perform a comprehensive NGS (next-generation sequencing) analysis of the repeatome in the grasshopper *Abracris flavolineata* aimed at uncovering the molecular composition and origin of its B chromosome. Our results have revealed that this B chromosome shows a DNA repeat content highly similar to the DNA repeat content observed for euchromatic (non-C-banded) regions of A chromosomes. Moreover, this B chromosome shows little enrichment for high-copy repeats, with only a few elements showing overabundance in B-carrying individuals compared to the 0B individuals. Consequently, the few satellite DNAs (satDNAs) mapping on the B chromosome were mostly restricted to its centromeric and telomeric regions, and they displayed much smaller bands than those observed on the A chromosomes. Our data support the intraspecific origin of the B chromosome from the longest autosome by misdivision, isochromosome formation, and additional restructuring, with accumulation of specific repeats in one or both B chromosome arms, yielding a submetacentric B. Finally, the absence of B-specific satDNAs, which are frequent in other species, along with its euchromatic nature, suggest that this B chromosome arose recently and might still be starting a heterochromatinization process. On this basis, it could be a good model to investigate the initial steps of B chromosome evolution.

Heredity (2021) 127:475–483; <https://doi.org/10.1038/s41437-021-00470-5>

INTRODUCTION

B chromosomes are dispensable genomic elements reported in many plant, animal, and fungal species (Jones and Rees 1982; Camacho 2005; Houben et al. 2014; Jones 2017). B chromosomes were discovered more than a century ago (Wilson 1907) and, for many years, only repetitive DNA had been found on them (for review, see Camacho 2005). However, it is now known that B chromosomes also contain protein-coding genes (Martis et al. 2012; Valente et al. 2014; Navarro-Dominguez et al. 2017). A common characteristic of most B chromosomes is the accumulation of repetitive DNA, which accounts for its evolution and differentiation from A chromosomes (Camacho 2005; Houben et al. 2014). These accumulated repeats include microsatellites and satellite DNAs (satDNAs), multiple classes of transposable elements (TEs), and multigene families (see for example Nur et al. 1988; Ziegler et al. 2003; Coleman et al. 2009; Poletto et al. 2010; Peng and Cheng 2011; Bueno et al. 2013; Klemme et al. 2013; Milani and Cabral-de-Mello 2014; Silva et al. 2014; Coan and Martins 2018; Hanlon et al. 2018; Malimpensa et al. 2018; Marques et al. 2018; Ruiz-Ruano et al. 2018; Felicetti et al. 2021; Stornoli et al. 2021).

In grasshoppers, cytological and molecular analysis in multiple species revealed that most B chromosomes show C-banded heterochromatin and plenty of DNA repeats (for instance, see Ruiz-Ruano et al. 2016a, 2018; Milani et al. 2017a, 2018). For example, most B chromosome variants found in *Eyprepocnemis plorans* are mostly made of rDNA and a satDNA family (Cabrero et al. 1999; 2014; López-León et al. 2008) and they are enriched in R2 retrotransposons (Montiel et al. 2014). The most complete quantification of the repeatome in a grasshopper was recently performed in *Locusta migratoria* and showed that the B chromosome contains 94.9% of repetitive DNA, with a single satDNA comprising 55% of the B chromosome (Ruiz-Ruano et al. 2018). In addition, this B chromosome showed a 17 kb region, including 29 different TEs, which was apparent as a FISH band on the B chromosome. Similarly, heterochromatic B chromosomes rich in a variety of repetitive DNAs have been reported in other grasshopper species, such as *Eumigus monticola*, *Rhammatocerus brasiliensis*, *Xyleus (discoideus) angulatus*, *Schistocerca rubiginosa*, *Podisma sapporensis* and *Dichroplus pratensis* (Bidau et al. 2004; Loreto et al. 2008; Oliveira et al. 2011; Ruiz-Ruano et al. 2016a; Jetybayev et al. 2018; Milani et al. 2018).

¹Departamento de Biologia Geral e Aplicada, UNESP - Univ Estadual Paulista, Instituto de Biociências/IB, Rio Claro, São Paulo, Brazil. ²Evolutionary Biology Centre, Department of Organismal Biology – Systematic Biology, Uppsala University, Uppsala, Sweden. ³Norwich Research Park, School of Biological Sciences, University of East Anglia, Norwich, UK. ⁴Departamento de Genética, Facultad de Ciencias, UGR - Univ de Granada, Granada, Spain. ✉email: cabral.mello@unesp.br

An exception to this general pattern is the South American grasshopper *Abracris flavolineata* ($2n = 22 + X0\delta/XX\phi$) where a submetacentric B chromosome failed to show heterochromatin defined by the C-banding technique, i.e., C-positive blocks (Cella and Ferreira 1991; Bueno et al. 2013). This B chromosome is mitotically stable thus showing the same number in all cells from the same individual and occurring in one or two copies in a natural population sampled at Rio Claro, São Paulo, Brazil (Milani et al. 2017b). Current evidence supports the origin of this B chromosome from the longest chromosome pair (L1), based on the U2 snDNA being only visualized by FISH on these two chromosomes (Bueno et al. 2013). In addition, we detected other repeats on this B chromosome, a satDNA family (Milani et al. 2017a), some microsatellite repeats (Milani and Cabral-de-Mello 2014), and two TEs (Palacios-Gimenez et al. 2014), which were shared with many A chromosomes. Intrigued by the absence of C-positive heterochromatin on the B chromosome of *A. flavolineata*, we decided to perform high-throughput complementary bioinformatic and cytogenetic analyses to characterize its repetitive DNA content. Repeatome analysis including 1744 TEs, 53 satDNAs, and 9 multigene families revealed that, consistent with its C-heterochromatin scarcity, this B chromosome is not enriched in high-copy repetitive DNAs, which makes it unusual among B chromosomes in general. Exceptionally, we found a few repetitive DNAs present on the euchromatic (C-negative) regions of the A chromosomes, which also decorate the interstitial regions of the B chromosome. In contrast, other repetitive DNAs that were enriched in heterochromatic regions (C-bands) of the A complement were mostly restricted to centromeric and distal regions of the B chromosome. In addition, satellitome analysis revealed that the B chromosome shared one satDNA family in exclusivity with the L1 autosome thus supporting B ancestry from this A chromosome. We finally suggest that this B chromosome could be a young element currently being in an initial step of heterochromatinization.

MATERIALS AND METHODS

Biological materials, genomic DNA extraction, and chromosome preparations

For molecular and bioinformatic analysis, we used the same seven male individuals of *Abracris flavolineata* (three 0B, two 1B, and two 2B) previously studied by Bueno et al. (2013), Milani et al. (2017b), and Ahmad et al. (2020). The hind legs of these animals, previously stored in 100% ethanol at -20°C , were used for genomic DNA (gDNA) extraction following the phenol/chloroform-based protocol (Sambrook and Russell 2001), which was used for genomic sequencing and PCR assays (see next topics).

For chromosomal mapping we collected five gravid females, which were maintained alive in cages at the laboratory until oviposition, allowing embryos to be obtained. Mitotic embryo chromosome spreads were prepared according to the protocol proposed by Webb et al. (1978). Chromosome spreads were performed by maceration and spreading of portions of embryos on a slide within a drop of 50% acetic acid, under a hot plate at 45°C .

Genome sequencing and identification of repetitive DNA sequences being overabundant in B-carrying individuals

Genomic DNA sequencing was performed by the Illumina HiSeq 4000 platform using the Macrogen Inc. service (Seoul, Republic of Korea). Sequencing yielded 27–41 Gb DNA (per sample) of 151 bp paired reads. The genomes from seven individuals are deposited in the Sequence Read Archive (SRA) under the accession numbers SRX7784770–SRX7784772. Repetitive sequences making up the repeatome of *A. flavolineata* were recovered and characterized using different approaches, including a thorough search for the satDNA families making up the satellitome, multigene families, and TEs (see details below).

To find and characterize the maximum number of different satDNA families, we applied the satMiner protocol (Ruiz-Ruano et al. 2016b). For this purpose, we randomly selected $2 \times 5,000,000$ reads from each

individual using SeqTK (<https://github.com/lh3/seqtk>) and pooled those belonging to the same type of genome (0B, 1B, or 2B) by concatenating them. We then performed sequence preprocessing for each group of reads using the “`rexp_prepare_normaltag.py`” script (<https://github.com/fjuizruano/ngs-protocols>), which uses Trimmomatic (Bolger et al. 2014) to remove adapters and low-quality nucleotides ($Q < 20$), and finally selected only completely paired reads after trimming, i.e., those read pairs with 151 bp in both members. The script then interleaves forward and reverse reads and converts them to fastq format. We obtained 100,000 read pairs for each of the three libraries (0B, 1B, and 2B) and concatenated them into a single file. We then applied the satMiner protocol (Ruiz-Ruano et al. 2016b) consisting of several rounds of clustering with RepeatExplorer (RE) software (Novák et al. 2013) alternated with the DeconSeq filtering tool (Schmieder and Edwards 2011) to remove those satDNA sequences identified in previous RE rounds and added 100,000 of these cleaned read pairs from each pool sample (0B, 1B, and 2B) prior to each new RE round (again summing up 300,000 read pairs).

RE clusters putatively containing satDNAs were selected by visual graph inspection to identify those showing spherical or ring shapes, which are characteristic of this type of DNA sequence. Then, we performed manual curation of the selected contigs by Geneious v4.8 software (Drummond et al. 2009), checked their tandem structure by dotplot graphic inspection, and recovered the consensus sequence for repeat units of each satDNA family or subfamily. To search for homology between different satDNA families we first compared their consensus sequences using multiple sequence alignments with Muscle (Edgar 2004) implemented in Geneious v4.8 software (Drummond et al. 2009), and second, we ran a homology test based on RepeatMasker (Smit et al. 2017) with “`rm_homology.py`” (<https://github.com/fjuizruano/ngs-protocols>). The results of these analyses were used to classify the satDNA collection into superfamilies, families or subfamilies according to the identity criterion proposed in Ruiz-Ruano et al. (2016b).

For TE identification, we randomly selected 100,000 read pairs from each pool of genomes (0B, 1B, and 2B), for a total of 600,000 reads, which were used as input for a single RE round followed by a reclustering-specific tool available in the Galaxy platform (<https://repeatexplorer-elixir.cerit-sc.cz/galaxy/>). This tool was used for merging clusters showing homology into larger contigs, which are prone to improve TE assembly. Then, we analyzed all the cluster contigs for sequence extraction with Geneious v4.8 software (Drummond et al. 2009). Since this method allowed the recovery of fewer than 200 different TE families, we also used the dnaPipeTE pipeline (Goubert et al. 2015), which uses Trinity (Grabherr et al. 2011) as an assembler, followed by recurrent TE annotation and quantification in the raw reads compared with a custom database previously built by Ruiz-Ruano et al. (2018) from B-carrying genomes of *Locusta migratoria*. This analysis was performed using only forward reads and default parameters recommended for dnaPipeTE. Next, by means of a custom script (https://github.com/fjuizruano/ngs-protocols/blob/master/dnapipe_createdb.py) we used dnaPipeTE assembly and annotation to generate a fastq file with annotated contigs in the RepeatMasker format (Smit et al. 2017) for further analysis.

Finally, the multigene families (H3 histone gene, 18S, 28S, 5.8S, and 5S rDNAs, U1, U2, and U6 snDNAs) and full mitochondrial DNA (mtDNA) were recovered using MITObim (Hahn et al. 2013) with the seed sequences used for *Locusta migratoria* in Ruiz-Ruano et al. (2018).

All the repeats obtained by these different methods were later concatenated, and redundancy was removed by CD-HIT-EST clustering (Li and Godzik 2006) using an 80% sequence identity level, implying that those repeats showing at least 80% identity were considered the same family.

Estimation of repetitive DNA sequence abundances and divergences in the *A. flavolineata* genome

Sequence abundance and divergence of each repetitive DNA family were determined in each of the seven genomes analyzed by means of RepeatMasker (Smit et al. 2017) using the Cross_match search engine on 5,000,000 read pairs from each library. SatDNA families were named in decreasing order of abundance in 0B genomes, following Ruiz-Ruano et al. (2016b). Sequence divergence was estimated by the Kimura 2-parameter (K2P) model using the `calcDivergenceFromAlign.pl` script within RepeatMasker software (Smit et al. 2017). Abundance for a given repetitive DNA family was calculated as a genome proportion, represented by the sum of all mapped nucleotides belonging to it (including all subfamilies) with respect to the total number of nucleotides in the selected reads from each

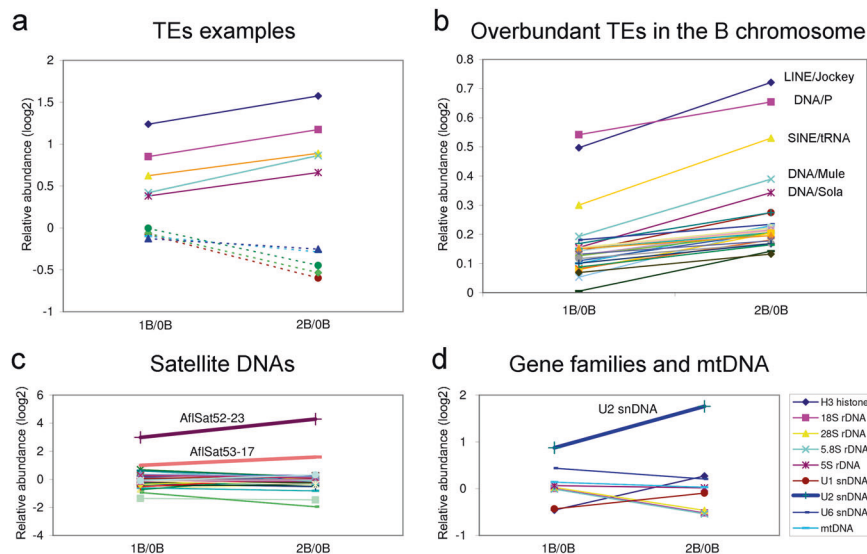


Fig. 1 Relative abundance of several types of repetitive DNA in *A. flavolineata* genomes carrying 1B and 2B, in comparison with the B-lacking genome, measured by the log₂ transformed 1B/0B and 2B/0B quotients. **a** Five examples of TEs showing overabundance in the B chromosome (solid lines), and five others showing the dilution effect (dotted lines) with negative values for both quotients. **b** Overabundant TEs in the B chromosome, indicating TE type for the five showing the highest quotients. **c** Only two satDNA families were overabundant in the B chromosome. Note the dilution effect for many other satDNAs. **d** Only the U2 snDNA showed clear overabundance in the B-carrying genomes, whereas the other families showed the dilution effect or quotients close to zero, suggesting their scarcity in the B chromosome.

Illumina library. Abundance and divergence for each family were separately estimated for each individual and later averaged for 0B (three individuals), 1B (two individuals), and 2B (two individuals) genomes. We then calculated two sequence abundance quotients, 1B/0B and 2B/0B, to search for repeats being overabundant in the B-carrying genomes so that those repeats showing both quotients clearly higher than 1 and that 2B/0B was higher than 1B/0B were considered overabundant in B-carrying individuals and thus enriched in the B chromosome. However, those repeats showing quotients lower than 1 are considered less abundant (or absent) in the B chromosome rather than in the average A chromosome. All satDNA families and some TEs showing overabundance in B-carrying genomes were selected for subsequent chromosomal mapping (see below).

DNA amplification and chromosomal mapping of repetitive DNAs

We designed primers for PCR amplification either manually or else using Primer3 software (Untergasser et al. 2012) (Supplementary Table 1), and PCR conditions followed the same protocol described in Milani et al. (2018). For satDNA sequences, the monomeric bands were isolated and purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corp., The Epigenetics Company, CA, USA) according to the manufacturer's recommendations. The same method was applied for TE isolation, taking care of isolating fragments showing the size expected from computational annealing of primers. These products were used for reamplification using the same PCR conditions. All amplified sequences were sequenced by the Sanger method using Macrogen Inc. (Seoul, Republic of Korea) service to confirm the actual amplification of the target sequence.

We performed fluorescence in situ hybridization (FISH) on mitotic chromosome spreads from embryos using one or two probes simultaneously, according to Cabral-de-Mello and Marec (2021). Probes were labeled by digoxigenin-11-dUTP (Roche, Mannheim, Germany) or biotin-14-dATP (Invitrogen) and detected by antidigoxigenin-rhodamine (Roche) and streptavidin, Alexa Fluor 488-conjugated (Invitrogen), respectively. The chromosomes were counterstained using 4',6-diamidino-20-phenylindole dihydrochloride (DAPI) and slides were mounted in VECTASHIELD (Vector, Burlingame, CA, USA). The preparations were observed and images were captured using a BX61 Olympus microscope equipped with a fluorescence lamp and appropriate filters and a DP70 cooled digital camera. All images were processed and optimized using Adobe Photoshop CS6. According to the results observed, we classified the satDNA families into three types: (i) visible FISH bands covering the whole chromosome width (B-pattern), (ii) occurrence of dot-like scattered signals across the chromosome (D-pattern), and (iii) no FISH signal at all (NS-pattern).

Statistical methods

We compared repeat abundance between the 0B, 1B, and 2B genomic libraries by means of nonparametric Friedman ANOVA and the Wilcoxon matched pairs test.

RESULTS

Comparative genomic abundance reveals little enrichment for high-copy repeats in the B chromosome

The overall mean repetitive DNA abundance in *A. flavolineata* genomes from the Rio Claro, São Paulo, Brazil population was 52.94% in 0B individuals, 52.59% in 1B individuals, and 52.00% in 2B individuals; this figure thus decreased with an increasing number of B chromosomes (Friedman ANOVA: $\chi^2 = 8.08$, $N = 1806$, $df = 2$, $P < 0.018$). This result suggests that this B chromosome shows lower repetitive DNA content than the A chromosomes, on average, so that when a given repetitive element is scarce in the B chromosome, its genomic proportion will decrease as the number of Bs grows. This "dilution effect" was significant for TEs ($\chi^2 = 10.12$, $N = 1744$, $df = 2$, $P < 0.0064$), marginally significant for satDNA ($\chi^2 = 4.57$, $N = 53$, $df = 2$, $P > 0.10$), and not significant for multigene families ($\chi^2 = 1.56$, $N = 9$, $df = 2$, $P > 0.45$) (Fig. 1). However, a few repeats showed the reverse pattern, i.e., their abundance increased with increasing numbers of B chromosomes. This pattern suggested the presence of these repeats in the B chromosome. For quantitative application of this criterion, we calculated the 1B/0B and 2B/0B quotients and selected those elements showing 1B/0B > 1 and 2B/0B > 1B/0B, as the two conditions, as a whole, allowed selection for repeats showing increasing abundance with B number.

We found 53 satDNA families in *A. flavolineata*, all of which were present in the three genome libraries analyzed (Supplementary Table 2), thus revealing the absence of B-specific satDNAs. The dilution effect was also apparent for satDNA, as its genomic content decreased in the presence of B chromosomes (4.52% in 0B, 4.03% in 1B, and 3.99% in 2B) (see Supplementary Table 2 and Fig. 1c) (Wilcoxon matched pairs test: 0B vs. 1B: $z = 3.27$, $P = 0.001$; 0B vs. 2B: $z = 2.19$, $P = 0.028$). However, we found no significant difference in satDNA content between the 1B and 2B libraries ($z = 0.27$, $P = 0.79$), perhaps due to some degree of B chromosome heterogeneity. Consistent with the general dilution effect for satDNA, abundance

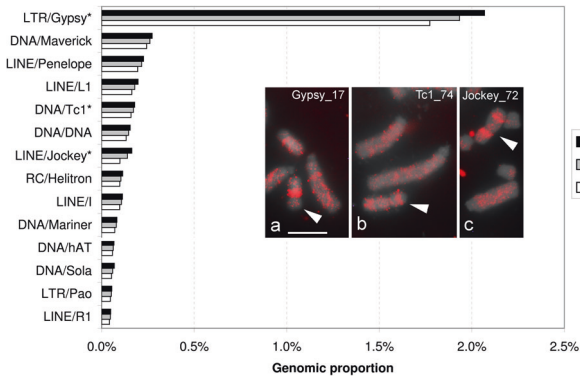


Fig. 2 Comparative genomic proportion between the 0B, 1B, and 2B genomes for the most abundant TEs in *Abracris flavolineata*. The asterisks indicate the superfamilies in which one representative was selected for FISH mapping on chromosomes (a–c). Note the spread distribution on long arms and absence of signals on pericentromeric C-heterochromatic region of A chromosomes. In the B chromosome (arrowheads) observe the differential distribution of TEs, i.e., first interstitial half of long arm (a), spread signal along the entire extension of the B chromosome, except distal regions (b) and enrichment on interstitial areas of both arms, and faint signals in proximal half of long arm (c). This last repeat was also absent in the terminal regions. Bar = 10 µm.

comparisons between libraries revealed that only two satDNA families (AflSat52-23 and AflSat53-17) were overabundant in the B-carrying genomes (see Table S2 and Fig. 1c).

The analysis of coding tandem repeats (including rRNA, U snRNA, and H3 histone multigene families) revealed that only the U2 snRNA family showed overabundance in the B-carrying genomes (Fig. 1d). The presence of U2 snDNA on the *A. flavolineata* B chromosome was previously shown by FISH analyses (Bueno et al. 2013; Menezes-de-Carvalho et al. 2015; Milani et al. 2017b). The remaining gene families and mtDNA failed to show differences in relative abundance between B-carrying and B-lacking genomes, but some of them displayed the dilution effect (Fig. 1d).

In the case of TEs, we found 212 elements (out of the 1744 analyzed) meeting the $1B/0B > 1$ and $2B/0B > 1B/0B$ criteria. These elements belonged to 28 families (Supplementary Table 3), the most abundant being LTR/Gypsy elements (Fig. 2). To test whether these results actually reflect overabundance in the B chromosome, we performed FISH for one element belonging to three distinct superfamilies, LTR/Gypsy (Gypsy_17), DNA/Tc1 (Tc1_74), and LINE/Jockey (Jockey_72). This analysis revealed their concentration on certain B chromosome regions with the appearance of chromosome bands (Fig. 2). As these three families were among the seven most abundant, additional FISH work would reveal whether the observed pattern critically depends on abundance, a highly feasible possibility (see also Supplementary Figure 1).

High-throughput analysis of the satellitome reveals that satDNA is scarce on the B chromosome

One of the 53 satDNA families found (named here as AflSat02-391) had previously been described as AflaSAT-1 (Milani et al. 2017a). The repeat unit length (RUL) of the 53 families ranged from 7 to 832 bp (mean = 224, SD = 167.6), and the total A + T content ranged from 30.43% to 76.50% (mean = 57.1%, SD = 8%). Homology tests between all satDNA families revealed the occurrence of only two superfamilies (SFs), with AflSat15-299, AflSat16-298, and AflSat26-296 comprising SF1, and AflSat20-233 and AflSat28-247 constituting SF2. As expected, the families belonging to each SF showed highly similar sequence properties (RUL and A + T content) (Supplementary Table 2).

A subtractive landscape (2B/0B) revealed a clear dilution effect for satDNA abundance, as the 2B genome showed a high deficit

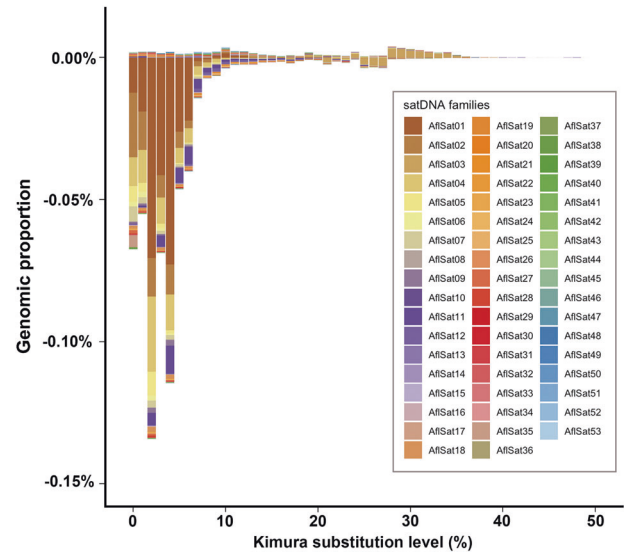


Fig. 3 Subtractive repetitive landscape (genome proportion versus sequence divergence based on Kimura substitution level) obtained from average counts for satDNAs in two males with 2B chromosomes and three with no B chromosome of *Abracris flavolineata*. Abundance values show the difference between the 2B minus the 0B genomes. Thus, positive values indicate overabundance in the 2B genomes, and negative values indicate overabundance in the 0B genomes. Note the occurrence of mainly negative values indicating the low enrichment of satDNAs in B-carrying genomes.

for most satDNA families, especially for the most abundant ones (Fig. 3). To analyze whether these genomic results are reflected at the cytogenetic level, we performed the physical mapping by FISH on A and B chromosomes of *A. flavolineata* for all 53 satDNA families identified by bioinformatic analysis. Similar to other grasshopper species (for instance, see Ruiz-Ruano et al. 2016a, 2018), we observed three different patterns, with 44 families showing bands on chromosomes (B-pattern), three families showing many small dots scattered on chromosomes (D-pattern), and the six remainder showing no FISH signals (NS-pattern) (Table 1 and Supplementary Figure 2).

A summary of chromosome locations for the 53 satDNA families (Table 1) indicated that 47% of the 205 FISH bands found on A chromosomes were located on pericentromeric regions involving the centromere and the short chromosomal arm. The location of these satDNAs thus coincided with the heterochromatin location in this species, as revealed by C-banding (Bueno et al. 2013). However, the other half of the satDNA bands were found on euchromatic regions at proximal (5%), interstitial (30%), or distal (18%) locations of the long A chromosome arms (Table 1 and Supplementary Figure 2a–x). Notwithstanding, it is clear that the pericentric heterochromatic regions were enriched in satDNA as they contained the five most abundant families representing 81% of all satDNA content in the 0B genome (Supplementary Table 2) (i.e., 3.67% out of the total 4.52%) (Fig. 4, Supplementary Figure 2a, b, Table 1). Remarkably, of these five satDNAs, only the satDNA showing the highest abundance (AflSat01-179) was present on all A chromosomes (Fig. 4a, Table 1), thus most likely playing a centromeric function (Melters et al. 2013). However, the least abundant satDNA families tended to show FISH bands on a single chromosome pair (Fig. 4b, Supplementary Figure 2), as 15 of the 20 families with this condition showed abundance under the median value of all 53 families, and only 5 showed abundance above the median (Table 1). X was the A chromosome showing more satDNA FISH bands in exclusivity (three interstitially and three distally located), followed by S10 (4), M6 (3), L2 and M8 (2),

Table 1. Chromosome location of the 53 satDNA families found in *Abracris flavolineata*.

| SatDNA family | Pattern | Chromosome pair and location | | | | | | | | | | | | |
|------------------|---------|------------------------------|-----|-----|-----|-----|-----|-----|-------|----|-----|-----|-----|-----|
| | | L1 | L2 | L3 | M4 | M5 | M6 | M7 | M8 | S9 | S10 | S11 | X | B |
| AflSat01-179 | B | p | p | p | p | p | p | p | p | p | p | p | p | p |
| AflSat02-391 | B | p | p | p | p | p | p | p | p | p | p | | p | p |
| AflSat03-17 | B | p | p | p | p | p | p | p | p | p | | p | p | p |
| AflSat04-161 | B | | p* | p | | p | p | p | | p | | p | | |
| AflSat05-437 | B | p | p | p | p | p | p | p | p | p | | p | p | |
| AflSat06-265 | B | d | d | d | d | d | d | d | d | d | | | d | |
| AflSat07-36 | B | i | i | i | i | i | i+d | i+d | i | p | | | i | p+d |
| AflSat08-184 | D | | | | | | | | | | | | | |
| AflSat09-187 | B | | | d | d | d | d | | | | d | | d | |
| AflSat10-297 | B | p | p | p | p | p | p | p | p | p | p | | p | |
| AflSat11-407 | NS | | | | | | | | | | | | | |
| AflSat12-237 | B | | | | | pr | pr | | | | p | p | | |
| AflSat13-177 | D | | | | | | | | | | | | | |
| AflSat14-110 | B | i | | i | i | | | d* | | | | | i | |
| AflSat15-299/SF1 | B | pr | | pr | | | | | | | | | | |
| AflSat16-298/SF1 | B | | | | | | pr | | | | | | | |
| AflSat17-231 | B | | i | | | | | | | | | | | |
| AflSat18-7 | B | d | d | d | d | d | d | d | | | | | i | |
| AflSat19-17 | NS | | | | | | | | | | | | | |
| AflSat20-233/SF2 | B | | | | | | | pr | pr | | | | | |
| AflSat21-260 | NS | | | | | | | | | | | | | |
| AflSat22-135 | B | i | i | i | i | i | | i | | d | | d | i | |
| AflSat23-286 | B | | | | | | | p | | | | | | |
| AflSat24-308 | B | | | | | | | | p | | | | | |
| AflSat25-40 | B | p+i* | p+i | p+i | p+i | p+i | p+i | p+i | p+i+d | p | p | | p+i | p+d |
| AflSat26-296/SF1 | B | | | | | | pr | | | | | | | |
| AflSat27-609 | B | | | i | | | | | | | | | | |
| AflSat28-247/SF2 | B | | i | | | | | i | i | | | | | |
| AflSat29-616 | B | | | | | | | | | | p | | | |
| AflSat30-197 | B | | | | | d | | | | | | | d | |
| AflSat31-226 | B | | | | | | d | | | | | | | |
| AflSat32-429 | B | | | | | | | | | d | p | | | |
| AflSat33-295 | B | | | | | | pr | | | | | | i | |
| AflSat34-30 | B | | i* | i | | | | | | | | | | |
| AflSat35-362 | B | | | | | | | | | | i | | | |
| AflSat36-41 | NS | | | | | | | | | | | | | |
| AflSat37-134 | B | i | i | i | i | | i | i | | | | | pr | |
| AflSat38-377 | B | | | | | | | | | | p | | | |
| AflSat39-832 | B | | i | | | | | | | | | | | |
| AflSat40-218 | B | | | | | | | | | | d | | i | p+i |
| AflSat41-186 | B | | | | | | | | p | | | | | |
| AflSat42-75 | D | | | | | | | | | | | | | |
| AflSat43-131 | B | | | | | | | | | | | | i | |
| AflSat44-254 | B | | | | | | | | | | d | | | |
| AflSat45-414 | NS | | | | | | | | | | | | | |
| AflSat46-153 | B | p | | | | | | | | | | | | p+d |
| AflSat47-75 | B | | | | | | | | | | | | d | |
| AflSat48-167 | B | | | | | | | | | | | | d | |
| AflSat49-138 | B | | | | | | | | | | | | d | |
| AflSat50-102 | B | | | | | | | | | | | | i | |

Table 1 continued

| SatDNA family | Pattern | Chromosome pair and location | | | | | | | | | | | | |
|---------------|---------|------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | L1 | L2 | L3 | M4 | M5 | M6 | M7 | M8 | S9 | S10 | S11 | X | B |
| AflSat51-30 | B | | | | | | | | | | | | | i |
| AflSat52-23 | B | p+i | p+i | p+i | p+i | p+i | p+i | p+i | p+i | p+i | p+i | p+i | p+i | p+d |
| AflSat53-17 | NS | | | | | | | | | | | | | |
| Total (p) | | 8 | 8 | 8 | 7 | 8 | 8 | 9 | 9 | 9 | 9 | 6 | 7 | 8 |
| Total (pr) | | 1 | 0 | 1 | 0 | 1 | 4 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| Total (i) | | 6 | 9 | 8 | 6 | 4 | 4 | 6 | 4 | 1 | 2 | 1 | 11 | 1 |
| Total (d) | | 2 | 2 | 3 | 3 | 4 | 4 | 4 | 2 | 3 | 3 | 1 | 6 | 4 |
| Total | | 17 | 19 | 20 | 16 | 17 | 20 | 20 | 16 | 13 | 14 | 8 | 25 | 13 |
| Shared with B | | 7 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 5 | 3 | 7 | – |

SF = superfamily, L1–L3 = three long autosomes, M4–M8 = five medium autosomes, S9–S11 = three short autosomes, X = X chromosome, B = B chromosome, p = pericentromeric (centromere plus short chromosomal arms), pr = proximal (near to centromere), i = interstitial, d = distal. B = banded pattern, D = dotted pattern, and NS = no signal. Repeats with dotted patterns occurred on all chromosomes, including the B chromosome (see text for details). Asterisks (*) indicate the occurrence of heteromorphisms for the presence/absence of bands in homologous chromosomes.

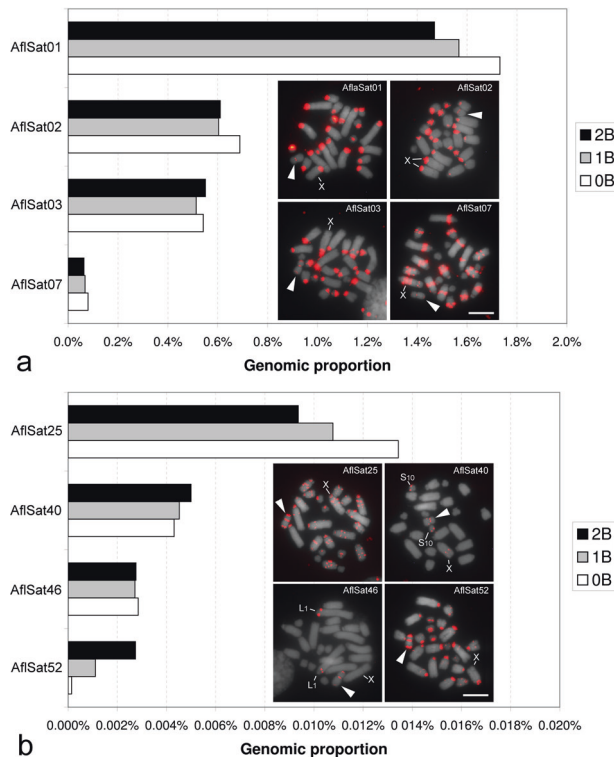


Fig. 4 Comparative genomic proportion and FISH mapping for eight satDNAs occurring in the B chromosomes with a banded pattern. SatDNAs showing high (a) or low (b) abundance (expressed as genome proportion). Repeat names are indicated on the left. Some A chromosomes and the B chromosome is indicated by arrowheads. The differential satDNA distribution on the B chromosome was observed, with pericentromeric signals for AflSat01 and AflSat02, pericentromeric plus distal signals for AflSat03, AflSat07, AflSat25, AflSat46, and AflSat52, and pericentromeric plus interstitial (on the long arm) signals for AflSat40. In addition, note the exclusive presence of AflSat46 bands on the B chromosome and the L1 pair. Bar = 10 μ m.

and L3 and M7 (1). The X chromosome harbored the highest number of satDNA families (25) and it was the A chromosome showing the highest number of interstitial and distal satDNA bands (Table 1).

We noticed a clear-cut difference in chromosome location between the two superfamilies existing in the genome of *A. flavolineata*, as the three families belonging to SF1 always showed proximal locations on one (AflSat16–298 and AflSat26–296) or two (AflSat15–299) A chromosome pairs (Table 1, Supplementary Figure 2g,i,n), whereas the two SF2 family members showed either proximal (AflSat20–233) or interstitial (AflSat28–247) locations (Table 1, Supplementary Figure 2m,w).

Finally, there were nine other satDNA families where the location on A chromosomes was not in the form of FISH bands, three of which showed the dotted pattern (D) (Table 1, Supplementary Figure 3), and the six remaining showed no FISH signal at all (NS) (Table 1, Supplementary Figures 2y, 4).

Regarding the B chromosome, we observed that eight of the 44 satDNA families showing the B-pattern on A chromosomes (AflSat01–179, AflSat02–391, AflSat03–17, AflSat07–36, AflSat025–40, AflSat40–218, AflSat46–153, and AflSat52–23), were also present on the B chromosome, whereas the three families showing the D-pattern also showed multiple small dots on the B chromosome (Table 1 and Fig. 4, Supplementary Figure 3). Among the 13 satDNA bands observed on the B chromosome, eight were pericentromeric, one was interstitial and four were distal. Most of the eight satDNA families showing FISH bands on the B chromosome showed multichromosomal locations on A chromosomes, except two showing locations on only one (AflSat46–153 on L1) or two (AflSat40–218 on S10 and X) A chromosomes (Table 1 and Fig. 4b). Among all A chromosomes, L1 and X were the A chromosomes sharing the highest number of satDNA families with the B chromosome (seven each; see Table 1). Bearing in mind that L1 also shares the U2 snDNA in exclusivity with the B chromosome (Milani et al. 2017b), we consider that, with the available data (repetitive DNA only), L1 is the best candidate to be the ancestor of this B chromosome.

The satDNA families with dotted patterns occupied virtually the entire extension of the B chromosome, but AflSat08–184 and AflSat42–75 were less abundant on pericentromeric and terminal regions (Supplementary Figure 3a,c) whereas AflSat13–177 was less evident on the proximal region of the short arm (Supplementary Figure 3b). They also showed FISH signals on the euchromatic (non-C-banded) regions of the long arm of all A chromosomes, but they were absent in their C-banded regions located on the pericentromeric region and the short arm (Supplementary Figure 3).

A comparative analysis of abundance for the eight satDNA families displaying the B FISH pattern on the B chromosome revealed why the global abundance of satDNA in the 0B, 1B, and 2B genomes showed a dilution effect. For this purpose, we separately represented the most and the least abundant families



Fig. 5 Comparative C-banding and FISH for repeat location between the L1 and B chromosomes of *A. flavolineata*. The L1 autosome showed, like the remaining A chromosomes, a large pericentromeric C-band including the pericentromeric region and the short arm. The FISH analysis for seven satDNA families and the U2 snDNA repeat showed pericentromeric and telomeric locations on the B chromosome whereas they were located on the pericentromeric region and the short arm of L1 (AflSat01, AflSat02, AflSat03, the pericentromeric region (AflSat46), interstitial (AflSat07), interstitial region and the short arm (AflSat25 and AflSat52). satDNAs thus might suggest that B originates from the proximal third of L1, including the interstitial region containing several satDNAs. However, the U2 snDNA is located on an L1 region outside the former proximal region, so the presence of U2 on B is not explained by a single rearrangement event.

(Fig. 4), thus revealing that three families (AflSat01-179, AflSat07-36, and AflSat25-40) showed a clear decrease in abundance with an increasing number of B chromosomes, whereas only two (AflSat40-218 and AflSat52-23) showed the reverse pattern, due to B-enrichment, but these two satDNA families were among the least abundant in the genome.

DISCUSSION

Genome low-pass sequencing combined with computational and chromosomal analysis provides a comprehensive understanding of the organization and evolution of DNA repeats on B chromosomes (Kumke et al. 2016; Ruiz-Ruano et al. 2018; Milani et al. 2018; Ebrahimzadegan et al. 2019; Serrano-Freitas et al. 2019). Through this approach, we found that the B chromosome of the grasshopper *A. flavolineata* is poorly enriched in repetitive DNA. Only three of the 53 satDNA families found in this species (AflSat40-218, AflSat52-23, and AflSat53-17), which are among the less abundant in the OB genome, were overabundant in B-carrying genomes. Likewise, only 28 TE families, containing 212 elements, representing only 12% of the 1744 TEs found, showed overabundance in B-carrying genomes. This scenario contrasts with the general idea that B chromosomes are enriched in repetitive DNA (Camacho 2005; Houben et al. 2014; Marques et al. 2018). Consistently, repeat-enriched B chromosomes have been reported in fish (Ziegler et al. 2003; Coan and Martins 2018; Stornioli et al. 2021), reptiles (Kichigin et al. 2019), plants (Martis et al. 2012; Kumke et al. 2016; Ebrahimzadegan et al. 2019), and insects (Hanlon et al. 2018; Ruiz-Ruano et al. 2018). Among the repeats found in B chromosomes, satDNA is the most frequent component (McAllister 1995; Klemme et al. 2013; Hanlon et al. 2018; Ruiz-Ruano et al. 2018; Ebrahimzadegan et al. 2019; Langdon et al. 2000; Stornioli et al. 2021).

This accumulation of repetitive DNAs on B chromosomes is commonly assumed to be due to their genetic isolation from A chromosomes, with which they do not recombine (Camacho 2005; Houben et al. 2014). In this way, the nonenrichment in repetitive DNA, the absence of C-heterochromatin blocks, and the absence of B-specific satDNA families would be consistent with the hypothesis that this B is a young element, resembling the composition of the A chromosome from which it derived (most likely L1, see below). The high similarity between B and A chromosomes is also supported also by B chromosome microdissection of *A. flavolineata* followed by chromosome painting, as all C-negative A chromosome regions and the B chromosome were similarly labeled (Menezes-de-Carvalho et al. 2015).

Based on FISH mapping of the U2 snDNA, the B chromosome in *A. flavolineata* was suggested to have derived from the L1 autosome, as only these two chromosomes harbor this sequence (Bueno et al.

2013). Here, chromosomal mapping of the full satellitome of this species has provided additional clues about B chromosome ancestry and evolution. We observed that the L1 autosome and the X chromosome both share the highest number of satDNA families with the B chromosome, i.e., seven families. However, the absence of U2 on the X chromosome and the fact that the L1 autosome is the only A chromosome sharing AflSat46-153 with the B chromosome, reinforce the conclusion that L1 is the most likely B ancestor. Although our data support the possible derivation of the B chromosome from the L1 autosome, with possible subsequent restructuring of the B chromosome, additional research is necessary to obtain accurate information on the possible synteny of the repeats shared by these chromosomes, as it would help to unveil the precise origin of the B chromosome. In addition, some repeats present on the L1 autosome were not found on the B chromosome, indicating some additional degree of B differentiation attributed to the intense dynamism of repetitive DNAs. Among grasshoppers, the origin of B chromosomes from large A chromosomes, as the current results suggest in *A. flavolineata* appears to be uncommon, as the few cases where B chromosome ancestry was claimed involved medium (M) or small (S) A chromosomes, such as S11 in *E. plorans* (Teruel et al. 2014), S8 in *E. monticola* (Ruiz-Ruano et al. 2016a), M8 and S9 in *L. migratoria* (Ruiz-Ruano et al. 2018), S9 in *S. rubiginosa*, S11 in *R. brasiliensis*, and S10 in *X. d. angulatus* (Milani et al. 2018). These medium- or small-sized A chromosomes are enriched in repetitive DNAs because their pericentromeric C-banded regions are the same size as the pericentromeric C-banded regions in long A chromosomes, but their non-C-banded regions are much smaller. Therefore, M and S chromosomes are more prone to be involved in chromosome rearrangements, which might be an initial step for B chromosome origin (Hewitt 1974; Perfectti and Werren 2001; Camacho 2005; Raskina et al. 2008; Houben et al. 2014; Ruiz-Ruano et al. 2016a; Milani et al. 2018). In *A. flavolineata*, the low amount of repeats on the B chromosome would be consistent with the low proportion of the C-banded region in L1 (see Fig. 5) and the loss of most of the C-banded chromatin in the B. In contrast, B derivation from medium or small A chromosomes with a lower proportion of non-C-banded chromatin should most likely render heterochromatic Bs, likewise in cases with B chromosome ancestry related to highly heterochromatic chromosomes, such as sex chromosomes (Sharbel et al. 1998; Pansonato-Alves et al. 2014; Ventura et al. 2015; Serrano-Freitas et al. 2019).

Remarkably, satDNAs displaying FISH bands on the *A. flavolineata* B chromosome frequently showed a symmetrical pattern for the FISH bands located on pericentromeric and distal regions, such as the U2 snDNA in both B chromosome arms (Fig. 5, Table 1), suggesting the isochromosome nature of this B chromosome and the involvement of centromeric misdivision in its origin. The small size of the FISH bands observed on the B chromosome for most

satDNA families (e.g., AflSat01-179, AflSat02-391, and AflSat03-17), would be consistent with the loss of the L1 short arm (which contains the largest amount of C-heterochromatin and satDNA families) during the B-forming misdivision. Isochromosomes arising from misdivision have been reported in grasshoppers such as *Eyprepocnemis plorans* (López-León et al. 1993), *Omocestus burri* (Del Cerro et al. 1994) and *Metaleptea brevicornis adpersa* (Grieco and Bidau 2000), plants such as *Zea mays* (Carlson and Phillips 1986), *Crepis capillaris* (Leach et al. 2005) and *S. cereale* (Marques et al. 2012), the fish *Astyanax scabripinnis* (Mestriner et al. 2000) and *Drosophila melanogaster* (Hanlon et al. 2018). Against the isochromosome hypothesis in *A. flavolineata* would be the B chromosome being not perfectly metacentric, as this would require additional events of inversion or differential duplications or deletions between B arms. More intense amplification of DNA repeats on one of the B chromosome arms has been noticed for TEs such as Gypsy_17, Tc1_74, and *Afmar2* (Palacios-Gimenez et al. 2014), and two satDNAs analyzed here (AflaSat07-36 and AflaSat40-218). This kind of event could have contributed to the emergence of the submetacentric B chromosome, which is currently prevalent in *A. flavolineata*. Notwithstanding, the evidence for L1 derivation of the B chromosome is still preliminary, as we all are still in the initial steps to disentangle the conundrum of B chromosome origin.

Altogether our results indicate that the B chromosome in *A. flavolineata* is unusually little enriched in repetitive DNAs, presumably because this B chromosome arose from the longest A chromosome, with a low proportion of C-heterochromatin, the most part of which was lost during the misdivision that yielded the B chromosome from the L1 autosome. The B chromosome is enriched in only a few repetitive elements, to a low extent, and the absence of B-specific satDNAs suggests that this B chromosome is young. This fact might be helpful in testing the L1-derivation hypothesis of the B chromosome, as a putatively young element could still conserve high similarity in gene content with its ancestor chromosome.

Data archiving

Genomes have been deposited at the Sequence Read Archive (SRA) under accession numbers SRX7784770–SRX7784772.

REFERENCES

Ahmad SF, Jehangir M, Cardoso AL, Wolf IR, Margarido VP, Cabral-de-Mello DC et al. (2020) B chromosomes of multiple species have intense evolutionary dynamics and accumulated genes related to important biological processes. *BMC Genomics* 21:656

Bidau CJ, Rosato M, Marti DA (2004) FISH detection of ribosomal cistrons and assortment-distortion for X and B chromosomes in *Dichroplus pratensis* (Acrididae). *Cytogenet Genome Res* 106:295–301

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120

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

Cabral-de-Mello DC, Marec F (2021) Universal fluorescence in situ hybridization (FISH) protocol for mapping repetitive DNAs in insects and other arthropods. *Mol Genet Genomics* 296:513–526

Cabrero J, López-León MD, Bakkali M, Camacho JPM (1999) Common origin of B chromosome variants in the grasshopper *Eyprepocnemis plorans*. *Heredity* 83:435–439

Cabrero J, López-León MD, Ruiz-Estévez M, Gómez R, Petitpierre E, Rufas JS et al. (2014) B1 was the ancestor B chromosome variant in the western Mediterranean area in the grasshopper *Eyprepocnemis plorans*. *Cytogenet Genome Res* 142:54–58

Camacho JPM (2005) B chromosomes. In: Gregory TR (ed) *The evolution of the genome*. Elsevier, San Diego, p 223–286

Carlson WR, Phillips RL (1986) The B chromosome of maize. *Crit Rev Plant Sci* 3:201–226

Cella MD, Ferreira A (1991) The cytogenetics of *Abracris flavolineata* (Orthoptera, Caelifera, Ommatolampinae, Abracrini). *Rev Brasileira Genética* 14:315–329

Coan RL, Martins C (2018) Landscape of transposable elements focusing on the B chromosome of the cichlid fish *Astatotilapia latifasciata*. *Genes* 9:269

Coleman JJ, Rounsley SD, Rodríguez-Carres M, Kuo A, Wasmann CC, Grimwood J et al. (2009) The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet* 5:e1000618

Del Cerro AD, Fernández A, Santos JL (1994) Spreading synaptonemal complexes of B isochromosomes in the grasshopper *Omocestus burri*. *Genome* 37:1035–1040

Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M (2009) Geneious v.4.8.5. Biomatters Ltd, Auckland, New Zealand

Ebrahimzadegan R, Houben A, Mirzaghaderi G (2019) Repetitive DNA landscape in essential A and supernumerary B chromosomes of *Festuca pratensis* Huds. *Sci Rep* 9:1–11

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797

Felicetti D, Haerter CA, Baumgärtner L, Paiz LM, Takagui FH, Margarido VP et al. (2021) A new variant B chromosome in auchenipteridae: the role of (GATA)_n and (TTAGGG)_n sequences in understanding the evolution of supernumeraries in *Trachelyopterus*. *Cytogenet Genome Res* 161:70–81

Goubert C, Modolo L, Vieira C, ValienteMoro C, Mavingui P, Boulesteix M (2015) 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* 7:1192–1205

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I et al. (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol* 29:644

Grieco ML, Bidau CJ (2000) The dicentric nature of the metacentric B chromosome of *Metaleptea brevicornis adpersa* (Acridinae, acrididae). *Heredity* 84:639–646

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res* 41:e129–e129

Hanlon SL, Miller DE, Eche S, Hawley RS (2018) Origin, composition, and structure of the supernumerary B chromosome of *Drosophila melanogaster*. *Genetics* 210:1197–1212

Hewitt GM (1974) The integration of supernumerary chromosomes into the orthopteran genome. *Cold Spring Harb symposia Quant Biol* 38:183–194. Cold Spring Harbor Laboratory Press

Houben A, Banaei-Moghaddam AM, Klemme S, Timmis JN (2014) Evolution and biology of supernumerary B chromosomes. *Cell Mol Life Sci* 71:467–478

Jetybayev IY, Bugrov AG, Dzyubenko VV, Rubtsov NB (2018) B chromosomes in grasshoppers: different origins and pathways to the modern Bs. *Genes* 9:509

Jones RN (2017) New species with B chromosomes discovered since 1980. *Nucleus* 60:263–281

Jones RN, Rees H (1982) B chromosomes. Academic press

Kichigin IG, Lisachov AP, Giovannotti M, Makunin AI, Kabilov MR, O'Brien PC et al. (2019) First report on B chromosome content in a reptilian species: the case of *Anolis carolinensis*. *Mol Genet Genomics* 294:13–21

Klemme S, Banaei-Moghaddam AM, Macas J, Wicker T, Novák P, Houben A (2013) High-copy sequences reveal distinct evolution of the rye B chromosome. *N. Phytol* 199:550–558

Kumke K, Macas J, Fuchs J, Altschmied L, Kour J, Dhar MK et al. (2016) Plantago lagopus B chromosome is enriched in 5S rDNA-derived satellite DNA. *Cytogenet Genome Res* 148:68–73

Langdon T, Seago C, Jones RN, Ougham H, Thomas H, Forster JW, Jenkins G (2000) De novo evolution of satellite DNA on the rye B chromosome. *Genetics* 154:869–884

Leach CR, Houben A, Field B, Pistrick K, Demidov D, Timmis JN (2005) Molecular evidence for transcription of genes on a B chromosome in *Crepis capillaris*. *Genetics* 171:269–278

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659

López-León MD, Cabrero J, Dzyubenko VV, Bugrov AG, Karamysheva TV, Rubtsov NB, Camacho JPM (2008) Differences in ribosomal DNA distribution on A and B chromosomes between eastern and western populations of the grasshopper *Eyprepocnemis plorans*. *Cytogenet Genome Res* 121:260–265

López-León MD, Cabrero J, Pardo MC, Viseras E, Camacho JPM, Santos JL (1993) Generating high variability of B chromosomes in *Eyprepocnemis plorans* (grasshopper). *Heredity* 71:352–362

Loreto V, Cabrero J, López-León MD, Camacho JPM, Souza MJ (2008) Possible autosomal origin of macro B chromosomes in two grasshopper species. *Chromosome Res* 16:233–241

Malimpensa GC, Traldi JB, Toyama D, Henrique-Silva F, Vicari MR, Moreira-Filho O (2018) Chromosomal mapping of repeat DNA in *Bergiaria westermanni* (Pime-lodidae, Siluriformes): localization of 45S rDNA in B chromosomes. *Cytogenet Genome Res* 154:99–106

Marques A, Klemme S, Guerra M, Houben A (2012) Cytomolecular characterization of de novo formed rye B chromosome variants. *Mol Cytogenet* 5:34

Marques A, Klemme S, Houben A (2018) Evolution of plant B chromosome enriched sequences. *Genes* 9:515

Martis MM, Klemme S, Banaei-Moghaddam AM, Blattner FR, Macas J, Schmutz T et al. (2012) Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. *Proc Natl Acad Sci USA* 109:13343–13346

- McAllister BF (1995) Isolation and characterization of a retroelement from B chromosome (PSR) in the parasitic wasp *Nasonia vitripennis*. *Insect Mol Biol* 4:253–262
- Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG et al. (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* 14:R10
- Menezes-de-Carvalho NZ, Palacios-Gimenez OM, Milani D, Cabral-de-Mello DC (2015) High similarity of U2 snDNA sequence between A and B chromosomes in the grasshopper *Abracris flavolineata*. *Mol Genet Genomics* 290:1787–1792
- Mestriner CA, Galetti PM, Valentini SR, Ruiz IR, Abel LD, Moreira-Filho O, Camacho JPM (2000) Structural and functional evidence that a B chromosome in the characid fish *Astyanax scabripinnis* is an isochromosome. *Heredity* 85:1–9
- Milani D, Bardella VB, Ferretti AB, Palacios-Gimenez OM, Melo ADS, Moura RC et al. (2018) Satellite DNAs unveil clues about the ancestry and composition of B chromosomes in three grasshopper species. *Genes* 9:523
- Milani D, Cabral-de-Mello DC (2014) Microsatellite organization in the grasshopper *Abracris flavolineata* (Orthoptera: Acrididae) revealed by FISH mapping: remarkable spreading in the A and B chromosomes. *PLoS ONE* 9:e97956
- Milani D, Palacios-Gimenez OM, Cabral-de-Mello DC (2017b) The U2 snDNA is a useful marker for B chromosome detection and frequency estimation in the grasshopper *Abracris flavolineata*. *Cytogenet Genome Res* 151:36–40
- Milani D, Ramos É, Loreto V, Martí DA, Cardoso AL, de Moraes KCM et al. (2017a) The satellite DNA AfaSAT-1 in the A and B chromosomes of the grasshopper *Abracris flavolineata*. *BMC Genet* 18:81
- Montiel EE, Cabrero J, Ruiz-Estévez M, Burke WD, Eickbush TH, Camacho JPM, López-León MD (2014) Preferential occupancy of R2 retroelements on the B chromosomes of the grasshopper *Eyprepocnemis plorans*. *PLoS ONE* 9:e91820
- Navarro-Dominguez B, Ruiz-Ruano FJ, Cabrero J, Corral JM, López-León MD, Sharbel TF, Camacho JPM (2017) Protein-coding genes in B chromosomes of the grasshopper *Eyprepocnemis plorans*. *Sci Rep* 7:45200
- Novák P, Neumann P, Pech J, Steinhaisl J, Macas J (2013) RepeatExplorer: a galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29:792–793
- Nur U, Werren JH, Eickbush DG, Burke WD, Eickbush TH (1988) A “selfish” B chromosome that enhances its transmission by eliminating the paternal genome. *Science* 240:512–514
- Oliveira NL, Cabral-de-Mello DC, Rocha MF, Loreto V, Martins C, Moura RC (2011) Chromosomal mapping of rDNAs and H3 histone sequences in the grasshopper *Rhammatocerus brasiliensis* (Acrididae, Gomphocerinae): extensive chromosomal dispersion and co-localization of 5S rDNA/H3 histone clusters in the A complement and B chromosome. *Mol Cytogenet* 4:24
- Palacios-Gimenez OM, Bueno D, Cabral-De-Mello DC (2014) Chromosomal mapping of two Mariner-like elements in the grasshopper *Abracris flavolineata* (Orthoptera: Acrididae) reveals enrichment in euchromatin. *Eur J Entomol* 111:329–334
- Pansonato-Alves JC, Serrano ÉA, Utsunomia R, Camacho JPM, da Costa Silva GJ, Vicari MR et al. (2014) Single origin of sex chromosomes and multiple origins of B chromosomes in fish genus *Characidium*. *PLoS ONE* 9:e107169
- Peng SF, Cheng YM (2011) Characterization of satellite CentC repeats from heterochromatic regions on the long arm of maize B-chromosome. *Chromosome Res* 19:183–191
- Perfectti F, Werren JH (2001) The interspecific origin of B chromosomes: experimental evidence. *Evolution* 55:1069–1073
- Poletto AB, Ferreira IA, Martins C (2010) The B chromosomes of the African cichlid fish *Haplochromis obliquidens* harbour 185 rRNA gene copies. *BMC Genet* 11:1
- Raskina O, Barber JC, Nevo E, Belyayev A (2008) Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenet Genome Res* 120:351–357
- Ruiz-Ruano FJ, Cabrero J, López-León MD, Camacho JPM (2016a) Satellite DNA content illuminates the ancestry of a supernumerary (B) chromosome. *Chromosoma* 126:487–500
- Ruiz-Ruano FJ, Cabrero J, López-León MD, Sánchez A, Camacho JPM (2018) Quantitative sequence characterization for repetitive DNA content in the supernumerary chromosome of the migratory locust. *Chromosoma* 127:45–57
- Ruiz-Ruano FJ, López-León MD, Cabrero J, Camacho JPM (2016b) High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci Rep* 6:28333
- Sambrook J, Russell DW (2001) Molecular cloning. Sambrook Russe 1(2):3. Cold Springs Harbour Laboratory Press
- Schmieder R, Edwards R (2011) Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS ONE* 6:e17288
- Serrano-Freitas ÉA, Silva DM, Ruiz-Ruano FJ, Utsunomia R, Araya-Jaime C, Oliveira C et al. (2019) Satellite DNA content of B chromosomes in the characid fish *Characidium gomesi* supports their origin from sex chromosomes. *Mol Genet Genomics* 295:195–207
- Sharbel TF, Green DM, Houben A (1998) B-chromosome origin in the endemic New Zealand frog *Leiopelma hochstetteri* through sex chromosome devolution. *Genome* 41:14–22
- Silva DMDA, Pansonato-Alves JC, Utsunomia R, Araya-Jaime C, Ruiz-Ruano FJ, Daniel SN et al. (2014) Delimiting the origin of a B chromosome by FISH mapping, chromosome painting and DNA sequence analysis in *Astyanax paranae* (Teleostei, Characiformes). *PLoS ONE* 9:e94896
- Smit AFA, Hubley R, Green P (2017) RepeatMasker Open-4.0. <http://www.repeatmasker.org>
- Stornioli JHF, Goes CAG, Calegari RM, dos Santos RZ, Giglio LM, Foresti F et al. (2021) The B chromosomes of *Prochilodus lineatus* (Teleostei, Characiformes) are highly enriched in satellite DNAs. *Cells* 10:1527
- Teruel M, Ruiz-Ruano FJ, Marchal JA, Sánchez A, Cabrero J, Camacho JPM, Perfectti F (2014) Disparate molecular evolution of two types of repetitive DNAs in the genome of the grasshopper *Eyprepocnemis plorans*. *Heredity* 112:531–542
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Res* 40:e115–e115
- Valente GT, Conte MA, Fantinatti BEA, Cabral-de-Mello DC, Carvalho RF, Vicari MR, Kocher TD, Martins C (2014) Origin and evolution of B chromosomes in the cichlid fish *Astatotilapia latifasciata* based on integrated genomic analyses. *Mol Biol Evol* 31:2061–2072
- Ventura K, O'Brien PCM, do Nascimento Moreira C, Yonenaga-Yassuda Y, Ferguson-Smith MA (2015) On the origin and evolution of the extant system of B chromosomes in *Oryzomyini* radiation (Rodentia, Sigmodontinae). *PLoS ONE* 10:e0136663
- Webb GC, White MJD, Contreras N, Cheney J (1978) Cytogenetics of the parthenogenetic grasshopper *Warramaba* (formerly *Moraba*) *virgo* and its bisexual relatives. IV. Chromosome banding studies. *Chromosoma* 67:309–339
- Wilson EB (1907) The supernumerary chromosomes of Hemiptera. *Sci NY* 26:870–871
- Ziegler CG, Lamatsch DK, Steinlein C, Engel W, Scharl M, Schmid M (2003) The giant B chromosome of the cyprinid fish *Alburnus alburnus* harbours a retrotransposon-derived repetitive DNA sequence. *Chromosome Res* 11:23–35

ACKNOWLEDGEMENTS

We acknowledge the three anonymous reviewers for helpful comments that contributed significantly to improve the manuscript. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES), by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (process numbers 2014/11763-8 and 2015/16661-1), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). DCC-d-M is a recipient of a research productivity fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (process number 308290/2020-8). FJRR hold postdoctoral fellowships from Junta de Andalucía fellowship (Spain), Sven och Lilly Lawskis fond (Sweden), and a Marie Skłodowska-Curie Individual Fellowship (grant agreement 875732, European Commission).

AUTHOR CONTRIBUTIONS

DM and DCC-d-M conceived the study; DM and FJRR performed the bioinformatic analysis; DM performed the chromosomal analysis; DM, FJRR, JPMC, and DCC-d-M performed formal data analysis and wrote the paper; DCC-d-M led the project management. All authors read and approved the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41437-021-00470-5>.

Correspondence and requests for materials should be addressed to Diogo C. Cabral-de-Mello.

Reprints and permission information is available at <http://www.nature.com/reprints>

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