

Quantitative sequence characterization for repetitive DNA content in the supernumerary chromosome of the migratory locust

Francisco J. Ruiz-Ruano¹ · Josefa Cabrero¹ · María Dolores López-León¹ · Antonio Sánchez² · Juan Pedro M. Camacho¹

Received: 18 July 2017 / Revised: 23 August 2017 / Accepted: 24 August 2017 / Published online: 4 September 2017
© Springer-Verlag GmbH Germany 2017

Abstract Repetitive DNA is a major component in most eukaryotic genomes but is ignored in most genome sequencing projects. Here, we report the quantitative composition in repetitive DNA for a supernumerary (B) chromosome, in the migratory locust (*Locusta migratoria*), by Illumina sequencing of genomic DNA from B-carrying and B-lacking individuals and DNA obtained from a microdissected B chromosome, as well as the physical mapping of some elements. B chromosome DNA of 94.9% was repetitive, in high contrast with the 64.1% of standard (A) chromosomes. B chromosomes are enriched in satellite DNA (satDNA) (65.2% of B-DNA), with a single satellite (LmiSat02-176) comprising 55% of the B. Six satDNAs were visualized by FISH on the B chromosome, and the only A chromosome carrying all these satellites was autosome 9, pointing to this chromosome, along with autosome 8 (which shares histone genes with the B) as putative ancestors of the B chromosome. We found several transposable elements (TEs) showing nucleotidic variation specific to B-carrying individuals, which was also present in B-carrying transcriptomes. Remarkably, an interstitial region of the B chromosome included a 17 kb chimera composed of 29 different TEs, suggesting reiterative TE insertion in this B chromosome region.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00412-017-0644-7>) contains supplementary material, which is available to authorized users.

✉ Francisco J. Ruiz-Ruano
fjruizruano@ugr.es

¹ Departamento de Genética, Facultad de Ciencias, Universidad de Granada, Avda. Fuentenueva s/n, 18071 Granada, Spain

² Departamento de Biología Experimental, Universidad de Jaén, Jaén, Spain

Keywords B chromosome · Illumina · *Locusta migratoria* · Satellite DNA · Transposable elements

Introduction

Repetitive DNA constitutes a large fraction of most eukaryotic genomes, in the form of transposable elements (TEs), satellite DNA, segmental duplications, ribosomal DNA, multi-copy gene families, pseudogenes, etc. which, collectively, have been called the “repeatome” (Kim et al. 2014). Some of these elements are repeated in tandem, such as satellite DNA and multi-copy gene families, whereas others are scattered across the genome, the most abundant being a multitude of transposable elements. Accurate quantification of repetitive DNA composition is a common caveat of current genome sequencing projects, due to the difficulty in proper assembling mainly posed by satellite DNA (Miga 2015). In fact, a precise account of the amount of different repetitive elements is not currently known for a single chromosome even for the most advanced genome sequencing projects. Our present research tries to fill this gap by developing appropriate NGS and bioinformatic approaches which might be useful in the case of a B chromosome, i.e., a kind of parasitic genomic element being present in only some individuals thus being dispensable (Camacho 2005), but also of possible application to other cases such as sex-specific chromosomes (e.g., W or Y) or aneuploids. Until recently, it was thought that B chromosomes are almost exclusively composed of repetitive DNA, mainly TEs, satellite DNA, and some repetitive gene families, especially 45S ribosomal DNA (rDNA) (for review, see Camacho et al. 2000 and Camacho 2005). However, other gene families have recently been found on B chromosomes, such as histone genes (Teruel et al. 2010; Oliveira et al. 2011; Silva et al. 2014; Utsunomia et al. 2016), 5S rDNA (Oliveira et al. 2011; Kour et al. 2013;

Xie et al. 2014; Jang et al. 2016), and the U2 small nuclear DNA (Bueno et al. 2013; Menezes-de-Carvalho et al., 2015). Likewise, recent literature has confirmed that satellite DNA (Poletto et al. 2010; Peng and Cheng 2011; Klemme et al. 2013; Bauerly et al. 2014; Jang et al. 2016; Utsunomia et al. 2016) and TEs (Martis et al. 2012; Klemme et al. 2013; Kour et al. 2013; Houben et al. 2014; Valente et al. 2014; Huang et al. 2016) are abundant in B chromosomes of several species. This abundance in repetitive elements has hindered the finding of protein-coding genes in B chromosomes, but NGS is showing that these genes appear to be an important component of B chromosomes, at both structural and functional levels (Martis et al. 2012; Valente et al. 2014; Ma et al. 2017; Navarro-Dominguez et al. 2017).

One of the most recent approaches to search for repetitive elements in B-carrying genomes has been facilitated by NGS and the RepeatExplorer software (Novák et al. 2013), and a specific protocol allowing the high-throughput analysis of satellite DNA has recently been put forward by Ruiz-Ruano et al. (2016). These new technologies have been applied in species such as the fungus *Alternaria arborescens* (Hu et al. 2012), rye (Martis et al. 2012; Klemme et al. 2013), the fruitfly *Drosophila albomicans* (Zhou et al. 2012), the fish *Astatotilapia latifasciata* (Valente et al. 2014) and *Moenkhausia sanctaefilomenae* (Utsunomia et al. 2016), and the grasshopper *Eumigus monticola* (Ruiz-Ruano et al. 2017). This has provided fragmented pieces of new and valuable information, but the quantitative composition of B chromosomes for repetitive DNA is largely ignored.

The migratory locust (*Locusta migratoria*) harbors a very widespread B chromosome system, as its presence has been reported in populations from Asia (Itoh 1934; Hsiang 1958; Nur 1969; Kayano 1971), Africa (Rees and Jamieson 1954; Lespinasse 1973; Dearn 1974; Lespinasse 1977), Australia (King and John 1980), and Europe (Cabrero et al. 1984; Viseras et al. 1990). This B chromosome includes a proximal euchromatic region and a distal heterochromatic one (Cabrero et al. 1984). B chromosomes in this species are mitotically unstable, so that they vary in number between different cells of a same individual (Nur 1969; Kayano 1971; Cabrero et al. 1984; Viseras et al. 1990). This mitotic instability is already manifested in 5-day-old embryos (Pardo et al. 1995), and the preferential destiny of cells with higher number of Bs toward the germ line has been suggested to be a mechanism of premeiotic drive in males (Nur 1969). In addition, B chromosomes in this species show drive through females (Pardo et al. 1994), most likely through meiotic drive. The presence of drive in both sexes might explain why these B chromosomes have been so successful in nature.

Here, we develop an approach to estimate the abundance of repetitive DNA in B chromosomes, by means of Illumina sequencing of genomic DNA (gDNA) of B-carrying and B-lacking individuals, followed by subtractive bioinformatic

analysis, Illumina sequencing of DNA extracted from a microdissected B chromosome, and the physical mapping of some elements by means of FISH. We applied this protocol to the B chromosomes in *L. migratoria*, and the results have revealed their extreme enrichment in a single-satellite DNA family and have suggested the possibility that two A chromosomes (8 and 9) might have been involved in B chromosome origin.

Materials and methods

Materials and nucleic acid extraction

In 2014, we collected 20 males and 10 females of *Locusta migratoria* in two natural populations close to Los Barrios (Cádiz, Spain), namely 5 males and 3 females at Finca El Patrón (36.20685N, 5.46481W) and 15 males and 7 females at Puente de Hierro (36.19251N, 5.55131W). The females were cultured to obtain embryos for FISH experiments. We performed an NGS experiment including the two B-lacking males found and four of the B-carrying males. Males were anesthetized before dissection, and several testis tubules were immersed in 3:1 ethanol-acetic acid for cytological analysis of B chromosome frequency. Testis and body remains were immediately frozen in liquid nitrogen and stored at -80°C . We extracted genomic DNA (gDNA) from one hind leg, and RNA from the testes and the other hind leg.

RNA from each individual and body part was sequenced in the Illumina HiSeq2000 platform, which yielded about 8 Gb of 100 nt paired end reads (SRA project SRP066094) per sample. gDNA from each individual was sequenced in the Illumina HiSeq2500 platform yielding about 20 Gb of 125 nt paired end reads per individual, i.e., $\sim 3\times$ genome coverage (SRA project SRP079877). In addition, we used 454 reads of gDNA previously obtained from a B-carrying individual collected at Padul (Granada, Spain) consisting of 292,352 reads summing up ~ 200 Mb, which is available in the SRA database, with SRR1200889 accession number (Ruiz-Ruano et al. 2015).

In case of body parts with cuticle, materials were frozen into liquid nitrogen before breaking them in a porcelain mortar. For soft tissues, e.g., testis, we used an eppendorf pestil to break them up into 1.5-mL suspension buffer extraction kit. For genomic DNA extractions, we used the “Genomic DNA Gen Elute Mammalian Miniprep” (Sigma-Aldrich) kit following the manufacturer’s protocol. For total RNA extraction from hind legs, we used the “Real Total RNA Spin Plus” (Durviz), whereas for RNA extraction in testis, we used the kit “RNeasy Lipid Tissue” (Qiagen). We measured the concentration of extracted DNA and RNA using the “Infinite 200 NanoQuant” (Tecan) spectrophotometer. We checked DNA integrity in 2- μL sample run on 1% agarose gel and RNA

integrity in 2- μ L sample run on a MOPS gel with DEPC water. After verifying the good concentration and quality of the samples, we sent for sequencing at least 1 μ g of sample, at concentration higher than 50 ng/ μ L.

Acquiring reference sequences for repetitive DNA content analysis

First, we got 107 satellite DNA sequences belonging to 62 families found in *L. migratoria* (Ruiz-Ruano et al. 2016) and 1128 TEs included in RepBase v20.10 (Bao et al. 2015) (last access: 28 October 2015). Then, we assembled mitochondrial DNA and both genes and spacers of the histone cistron, 45S and 5S rDNAs, U1, U2, U4, U5, and U6 snDNAs, using MITObim (Hahn et al. 2013) with known seed sequences of these genes coming from several species: *Locusta migratoria* for H3 histone (GU111931) and mitochondrial DNA (JN858179), *Gomphocerinae* sp. for 45S rDNA (AY859546.1), *Ronderosia bergi* for 5S rDNA (KP213274), *Eyprepocnemis plorans* for U1 (KJ606069), *Abracris flavolineata* for U2 (KP975085), and *Drosophila melanogaster* for U4 (NR_001670), U5 (NR_001933), and U6 (NR_002081) snDNA genes.

In addition, we searched for tRNAs and other repetitive elements in the *L. migratoria* genome assembly (Wang et al. 2014). For tRNA, we used the tRNAscan-SE program (Lowe and Eddy 1997) and removed redundancies with CD-HIT-EST (Li et al. 2012) applying the options -M 0 -aS 0.8 -c 0.8 -G 0 -g 1. This yielded 617 tRNAs sequences. For other repetitive elements, we performed assembly with the RepeatModeler pipeline (Smit and Hubley 2017), and then, we masked the output against the previously mentioned sequences with RepeatMasker (Smit et al. 2017), keeping 1108 sequences with at least 40 consecutive unmasked nucleotides. All together, we used 2969 reference sequences.

Abundance of repetitive sequences in the B chromosome

To estimate the abundance of repetitive elements in B-carrying and B-lacking genomes, we aligned 5 million Illumina read pairs from each gDNA library ($\sim 0.2\times$ of genome coverage) with the reference sequences, by means of RepeatMasker. Abundance for a given repetitive element was calculated as the proportion of nucleotides aligned with the reference sequence, in respect to total library size. In addition to abundance, RepeatMasker provides divergence estimates, and both parameters allow building a repeat landscape which is highly informative about the abundance of repetitive elements showing different values of divergence. By comparing abundance and divergence for the different subclasses of repetitive elements between B-carrying and B-lacking libraries, we built a subtractive repeat landscape which shows the elements with the highest difference in abundance between the

two types of library. This gives a first indication of which DNA sequences are over-represented in the B-carrying libraries in respect to the B-lacking ones, presumably because they are more abundant in B than A chromosomes.

We estimated abundance of repetitive elements in B chromosomes from abundances in B-carrying and B-lacking libraries. For this purpose, we estimated abundance in four B-carrying and two B-lacking males. The use of several individuals within each type, instead of one, has the advantage of diminishing false positives caused by between-individual differences in abundance of repetitive elements in the A chromosomes. Of course, the more individuals in each group, the higher accuracy in B chromosome abundance estimates. For the following calculations, we will use the B-carrying and B-lacking averages. Our estimates are based on the fact that abundance of a given repetitive element in the B-carrying genome is the weighed mean of its abundance in the A chromosomes and that in the B chromosomes. Therefore, we first calculated the weight of A (w_A) and B (w_B) chromosomes in the B-carrying genome on the basis of A ($2C$) and B (C_B) chromosomes DNA content:

$$w_A = \frac{2C}{2C + C_B N_B}, \text{ and } w_B = \frac{C_B N_B}{2C + C_B N_B}$$

where N_B is the average number of B chromosomes in the B-carrying individuals Illumina sequenced.

Therefore, the abundance of a given repetitive element in a B-carrying library (A_{+B}) should be equal to the weighed mean of abundances in the A (A_A) and B (A_B) chromosomes, which are thus multiplied by their weights in the B-carrying genome:

$$A_{+B} = A_A w_A + A_B w_B$$

so that its abundance in the B chromosomes is as follows:

$$A_B = \frac{A_{+B} - A_A w_A}{w_B}$$

Abundances in the B-lacking and B-carrying libraries (A_A and A_{+B} , respectively) were estimated as described above, and the number of B chromosomes (N_B) was scored by cytologically examining 35–111 spermatocytes per male. Genome size ($C = 6.3$ Gb) was given in Wang et al. (2014) and DNA content per B chromosome was calculated from the proportion between C-value and B chromosome size measured in Ruiz-Ruano et al. (2011) ($C_B = 6.3 \times 0.02539 = 0.1599$ Gb). Due to the fact that B chromosomes are mitotically unstable in this species (see Cabrero et al. 1984), N_B in this case was the mean number of Bs per male. For our present calculations, N_B in the equation was the average of N_B values for the four B-carrying males (1.31).

We divided each A_B value between N_B to obtain abundance estimates per B chromosome (A_{1B}) and then calculated the proportion of B sequences corresponding to each repetitive

element, by dividing A_{1B} by the sum of all A_{1B} values being higher than zero. This latter sum provides an estimate of the total proportion of repetitive DNA in the B chromosome. The number of copies per B chromosome for a given repetitive element can be estimated by multiplying A_{1B} by total DNA content in the B chromosome (C_B) and dividing by element length. Likewise, the number of copies in the diploid A chromosome set can be calculated by multiplying its abundance in the B-lacking genome (A_A) by total DNA content in the A chromosomes ($2C$) and dividing by element length.

Validation of B chromosome content by Illumina sequencing of microdissected B chromosomes

A B-carrying male collected in 2012 at Padul (Granada, Spain), sited about 190 km from Los Barrios, was used to perform the microdissection of one B chromosome from a single diplotene cell (as described in Teruel et al. 2014). B chromosome DNA was amplified by the Phi29 DNA Polymerase with the Illustra GenomiPhi V2 amplification kit (GE Healthcare Life Sciences) following manufacturer's instructions, and the DNA obtained was sequenced in the Illumina HiSeq2000 platform.

Search for sequence and structural variation specific to B chromosomes

We first selected the repetitive sequences being clearly more abundant in the B-carrying genomes, compared to B-lacking ones. In these sequences, we searched for two types of variation specific to B chromosomes: sequence variants in the form of single-nucleotide polymorphisms (SNPs) and structural variation as insertions, deletions, and inversions. For this purpose, we selected Illumina reads showing homology with the reference sequences by means of the algorithm BLAT (Kent 2002). On the one hand, we mapped the reads with SSAHA2 (Ning et al. 2001) and selected SNPs showing no variation in B-lacking genomes but showing an additional variant exclusive to B-carrying genomes. We then calculated the frequency of both variants in gDNA and RNA from B-carrying individuals and calculated the transcription intensity (TI) for each variant as the quotient between its frequency in RNA and gDNA from a same individual.

On the other hand, we visualized the former SSAHA2 mappings by the IGV software (Thorvaldsdóttir et al. 2013) and searched for coverage variation, along the contigs, which might indicate structural variation. For a direct comparison of structure in this kind of regions, between B-carrying and B-lacking libraries, we selected 2500 read pairs showing homology with a given element, per library and region, and performed a RepeatExplorer clustering for each region.

When the structural variations which were characteristic of B-carrying individuals, and hence the B chromosome, were complex, we tried to increase the known contig sequence, toward both sides, by searching 454 reads of gDNA showing homology with the 300–400 nt at both ends, by means of the BLASTN software (Altschul et al. 1990). We then assembled the reads with the Geneious v4.8 software (Drummond et al. 2009) and repeated the process several times until no more lengthening was possible.

PCR amplification and fluorescent in situ hybridization

We designed primer pairs by using the Primer3 software (Untergasser et al. 2012) (Table S1). For elements others than satellite DNA, the PCR program employed started with an initial denaturation step at 95 °C for 5 min, followed by 30 cycles including heating at 94 °C for 30 s, 55–60–65 °C as annealing temperatures for 30 s, and 72 °C for 30 s, and a final extension step at 72 °C for 7 min. For satellites, we used the primers and PCR conditions previously described in Ruiz-Ruano et al. (2016). All the PCR products used for FISH were labeled by nick translation with 2.5 units of DNA polymerase/DNAse I (Invitrogen), following the standard protocol. The 18S rDNA probe was obtained by amplification of a 1113 bp fragment using the 18S-E and 1100R primers described in Littlewood and Olson (2001), according to PCR conditions described in Ruiz-Estévez et al. (2014). The physical mapping of these probes was performed by FISH following the protocol described in Cabrero et al. (2003). The probes were labeled with tetramethylrhodamine-5-dUTP or fluorescein-12-dUTP (Roche).

MinION validation of the TE chimera

To test the physical reliability of the TE chimera, we performed a sequencing experiment with the Oxford Nanopore MinION system using flow cell version R9. The 1D library preparation was carried out using the Nanopore Genomic Kit (SQK-LSK108) with the CleanNGS magnetic beads (CleanNA), with a starting DNA quantity of 5 µg without previous fragmentation. The sequencing was performed using the Nanopore's local base-calling software and yielded a total of 130 Mb in 63,346 reads. We first searched for reads showing homology with the TE chimera, using BLAST (Altschul et al. 1990), and then selected those reads aligning along its entire length with chimera sequence, using Geneious (Drummond et al. 2009), mapped them with SSAHA2 (Cabrero et al. 2003), and visualized the mapping with IGV (Thorvaldsdóttir et al. 2013).

Results

Subtractive analysis of Illumina sequenced B-carrying and B-lacking genomic DNA

Global abundance analysis showed that about 64.1% of the B-lacking *L. migratoria* genome consists of repetitive DNA (Fig. 1 and Table S2), a figure being slightly higher than the 60% estimated by Wang et al. (2014) in individuals from China. Remarkably, B-carrying libraries showed a slightly higher figure (64.6%) due to B chromosome enrichment in repetitive DNA (see below).

Out of 2841 repetitive elements showing positive results in RepeatMasker, our subtraction procedure indicated that 1533 of them showed higher abundance in the B-carrying genomes, suggesting that they are present in the B chromosome (Table S2). Illumina sequencing of the microdissected B chromosome showed the presence of 275 out of the 2841 elements detected by RepeatMasker. Considering both approaches,

1199 elements were not present in the B chromosome by both methods, 1367 were found in the B by the subtractive procedure only, 109 were found by microdissection only, and 166 were found by both methods. Contingency analysis showed significant positive association between results with both methods ($\chi^2 = 5.03$, $df = 1$, $P = 0.025$, odds ratio = 1.34, 95% CI: 1.04–1.72). Likewise, Spearman rank correlation analysis showed significant positive correlation between A_{1B} values and coverage observed in the Illumina sequences obtained from the microdissected B chromosome ($r_s = 0.33$, $N = 1533$, $t = 13.7$, $P < 0.000001$), suggesting that the higher the abundance of a repetitive element in the B chromosome, the higher the coverage in the DNA sequences obtained by B microdissection. In fact, 86% of the elements obtained by microdissection were predicted to be in the B chromosome by the subtraction approach.

Summarizing the information obtained by both methods, at subclass level, showed the presence of 48 subclasses of repetitive elements in the B chromosome (Table S3), with very

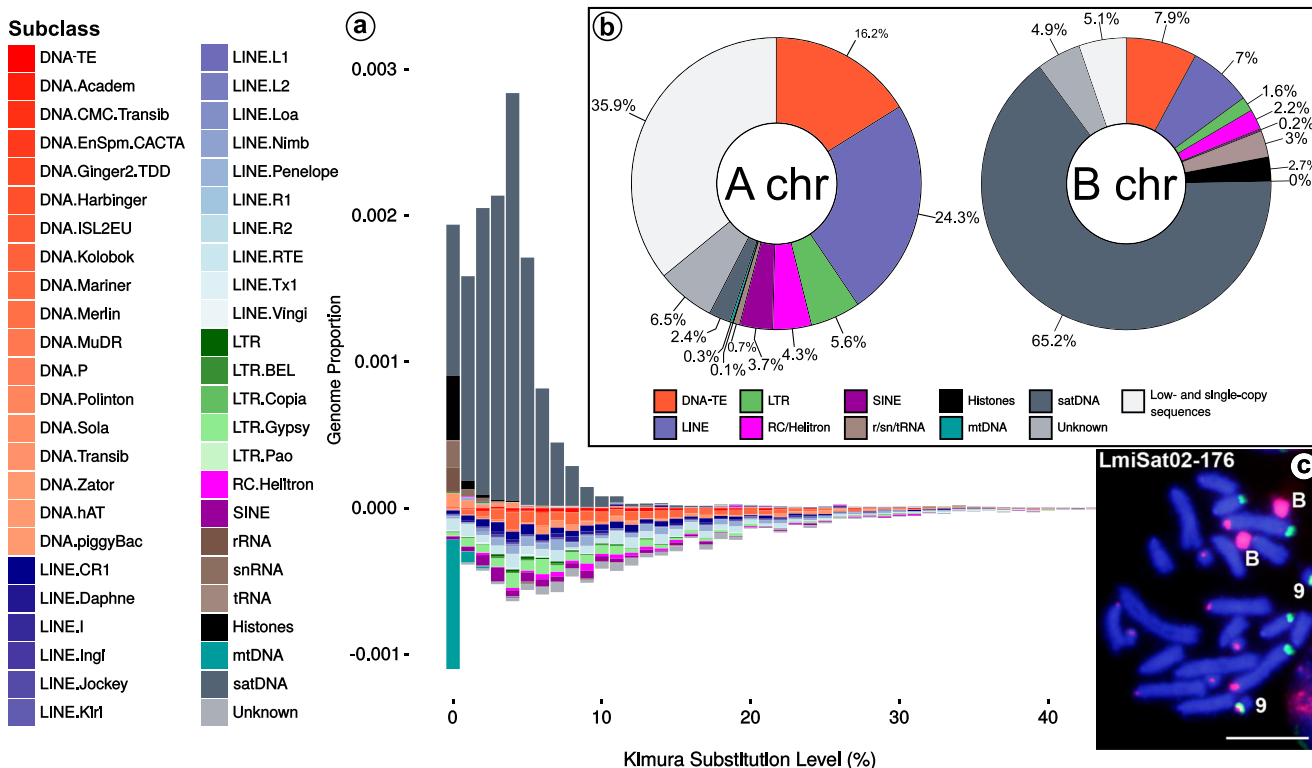


Fig. 1 **a** Subtractive repetitive landscape obtained from average counts for repetitive DNA elements observed in gDNA in four B-carrying and two B-lacking males of *L. migratoria*. It represents the difference in abundance for each repeat subclass showing different degrees of divergence in respect to the consensus sequence. It shows positive values for subclasses showing higher abundance in the B-carrying genomes than in the B-lacking ones, and negative ones for subclasses being underrepresented in the B-carrying genomes. Note that B-carrying genomes are enriched in satellite DNA and histone genes (positive values) but impoverished in TEs (negative values). **b** Relative frequency of repetitive elements in A and B chromosomes, the former being

inferred from the average counts in B-lacking males, and the latter as explained in materials and methods. The 0% in the B chromosome subfigure refers to the absence of mtDNA in it. **c** Fluorescent in situ hybridization (FISH) for the LmiSat02-176 satellite DNA (in red color) on a mitotic metaphase cell from a diploid B-carrying *L. migratoria* embryo, merged with DAPI (blue). This satellite DNA is present across the whole B chromosome length, and bioinformatic analysis indicates that it constitutes more than half of DNA content in the B chromosome. Note the presence of green signals, corresponding to 18S rDNA, on three pairs of A chromosomes (2, 6, and 9) and their absence on the B chromosome. Bar in **c** represents 5 μ m

highly significant positive correlation between the proportions resulting from both methods ($r_S = 0.79$, $t = 8.74$, $df = 46$, $P < 0.000001$), showing that 34 of these subclasses were detected by B microdissection. Bearing in mind that both methods were applied to individuals from two populations placed 190 km apart, we consider that the high consistency between the two methods support the high reliability of the B chromosome composition in repetitive DNA inferred here. In fact, all 11 repetitive classes in which the 48 subclasses can be summarized were present in the Illumina sequencing of the microdissected B chromosome, with very high positive correlation between the proportions yielded by both methods ($r_S = 0.81$, $t = 4.13$, $df = 9$, $P = 0.0026$).

Our subtractive approach indicated that 94.9% of B chromosome DNA is repetitive (Fig. 1, Tables S2 and S3), leaving about 5.1% of low and single-copy DNA. The most abundant types of repetitive DNA in the B chromosome are satellite DNA (65.2% of B chromosome content) and TEs, especially DNA transposons (7.9%) and LINEs (7%). In addition, the B chromosome contains several gene families, such as histone genes (2.7%), 45S (0.25%) and 5S (0.78%) rRNA genes, snRNA genes (1.3%), especially U2 (1.1%), and tRNA genes (0.7%) (Fig. 1 and Table S3). These results reveal remarkable differences in repetitive DNA content between A and B chromosomes since, in A chromosomes, TEs comprise 54% of their total DNA (19.1% in the B chromosomes) whereas all remaining classes of repetitive elements sum only 10%, with satellites representing a mere 2.4% (see also Ruiz-Ruano et al. 2016) (Fig. 1a, b). In brief, A chromosomes contain 64% of repetitive DNA and 36% of low and single-copy DNA. As a reflect of the difference between A and B chromosomes in DNA content, the abundance of the 11 classes of repetitive DNA showed no significant correlation between A and B chromosomes ($r_S = 0.49$, $t = 1.69$, $df = 9$, $P = 0.125$).

The presence of histone genes in the B chromosomes of this species had previously been shown by Teruel et al. (2010), by means of FISH analysis. However, previous FISH analyses had failed to detect the presence of 45S and 5S rDNA (Teruel et al. 2010), or else U1 (Anjos et al. 2015) or U2 snRNA genes (DC Cabral del Mello, personal communication) on the B chromosome, but all these genes show much lower abundance in the B chromosome than histone genes.

A single satellite DNA family comprises about half of repetitive DNA in the B chromosome

It was highly remarkable that a single satDNA (LmiSat02-176) comprised more than half of B chromosome DNA (54.8%), representing 84% of all satellite DNA contained in the B chromosome. This satellite is thus the most abundant repetitive element in the B chromosome, the remaining elements individually representing 3.1% or less of B chromosome content, and only three satellites (LmiSat02-176,

LmiSat01-193, and LmiSat04-18) surpassed 1% of B content (Table S4). The physical mapping by FISH for all 59 satDNAs reported in Ruiz-Ruano et al. (2016) confirmed that LmiSat02-176 is extremely abundant in the B chromosome as it shows intense FISH signal covering its whole length (Fig. 1c). Five other satDNAs were observed by FISH on the B chromosome (Table 1 and Fig. S1b-f). It was noteworthy that the bioinformatic analysis suggested the presence of four satDNAs in the B chromosome showing higher abundance than the telomeric repeat (i.e., > 0.2%) but failing to show FISH signals on the B chromosome even though they did on A chromosomes (Table 1).

TE abundance in the B chromosome

The most abundant TEs in the B chromosome were RTE-2_LMi (0.57% of B-DNA), followed by hAT-4_LMi (0.47%), Sola2-3N1_LMi (0.41%), Helitron-N11B_LMi (0.39%), and Penelope-42_LMi (0.38%) (Table S2). To get additional insights on the structure of the TEs found in the B chromosome, we selected seven TEs showing different levels of abundance in the B chromosome. Coverage analysis along their full length, performed by SSAHA2, revealed that the three most abundant TEs (hAT-4_LMi, Sola2-3N1_LMi, and Penelope-42_LMi) showed uniformly high coverage along most of their length in the B-carrying genomes (Fig. S2), whereas the TEs with intermediate (Penelope-59_LMi and Tx1-1_LMi) (Fig. 2) or low (Helitron-N17B_LMi and DNA2-4_LMi) (Fig. S3) abundance in the B chromosome showed an irregular coverage pattern by displaying high coverage only for a short region. Interestingly, the high coverage region in DNA2-4_LMi yielded a satellite DNA which was present only in B-carrying males, but it was not apparent by FISH analysis (see below).

To investigate the chromosome distribution of these seven TEs, we selected one or two regions in TEs with uniform or irregular coverage, respectively, for PCR amplification (Figs. 2, S2 and S3) and probe generation for FISH analysis. Remarkably, six of these TEs showed a large FISH signal on a same interstitial region of the B chromosome, but they were not apparent on A chromosomes, whereas the remaining TE (DNA2-4_LMi) showed a conspicuous cluster on autosome 9 but not on the B chromosome (Table 1, Figs. 2 and S4). Remarkably, the two TEs showing intermediate abundance and irregular coverage (Penelope-59_LMi and Tx1-1_LMi) displayed the FISH signal in the interstitial region of the B chromosome only for the probe generated on the high coverage region, suggesting that the low coverage regions of these TEs are absent or poorly represented in the B chromosome (Fig. 2b, d). Taken together, these results revealed a very high consistency between the bioinformatic estimates of TE abundance in the B chromosome and those that can be inferred by FISH.

Table 1 FISH analysis of some repetitive elements found in the *L. migratoria* genome

| Annotation | length | Proportion in | | Presence in micro-B | FISH signals on | | Figure |
|-------------------------------|--------|---------------|--------------|---------------------|-----------------|--------------|---------------------------|
| | | A chromosomes | B chromosome | | A chromosomes | B chromosome | |
| Satellite/LmiSat02-176 | 176 | 0.48% | 54.76% | Yes | X,5,7,9-11 | Yes | 1 |
| Satellite/LmiSat01-193 | 193 | 0.88% | 3.08% | Yes | all | Yes | S1 |
| Satellite/LmiSat04-18 | 18 | 0.05% | 2.24% | No | 9,11 | Yes | S1 |
| Satellite/LmiSat16-278 | 278 | 0.01% | 0.80% | No | 2 | No | Ruiz-Ruano et al. 2016 |
| Satellite/LmiSat18-210 | 210 | 0.01% | 0.59% | No | 7 | No | Ruiz-Ruano et al. 2016 |
| Satellite/LmiSat09-181 | 181 | 0.03% | 0.48% | No | 8 | No | Ruiz-Ruano et al. 2016 |
| Satellite/LmiSat06-185 | 185 | 0.07% | 0.37% | Yes | 2,3,8,9,11 | Yes | S1 |
| Satellite/LmiSat10-9 | 9 | 0.04% | 0.36% | No | 9,11 | No | Ruiz-Ruano et al. 2016 |
| Satellite/LmiSat53-47 | 47 | 0.01% | 0.21% | No | 9 | Yes | S1 |
| Satellite/LmiSat07-5-tel | 5 | 0.10% | 0.20% | No | all | Yes | S1 |
| hAT-4_LMi#DNA/hAT | 2392 | 0.01% | 0.47% | Yes | No | Yes | S4 |
| Sola2-3N1_LMi#DNA/Sola2 | 3259 | 0.05% | 0.41% | Yes | No | Yes | S4 |
| Penelope-42_LMi#LINE/Penelope | 2844 | 0.02% | 0.38% | Yes | No | Yes | S4 |
| Penelope-59_LMi#LINE/Penelope | 2206 | 0.02% | 0.11% | Yes | No | Yes | 2,S4 |
| Tx1-1_LMi#LINE/Tx1 | 5679 | 0.02% | 0.11% | No | No | Yes | 2,S4 |
| Helitron-N17B_LMi#RC/Helitron | 2926 | 0.01% | 0.05% | No | No | Yes | S4 |
| DNA2-4_LMi#DNA | 1127 | 0.01% | 0.03% | No | Yes | No | S4 |

59 satDNA families were analyzed, but only the ten most abundant in the B chromosome are shown, as the remaining 49 satDNAs showed no FISH signals on the B chromosome. In addition, seven TEs were analyzed by FISH. In all cases, the proportion represented by each element in the content of A and B chromosomes is shown, as well as whether they were present in the Illumina library obtained from microdissected B chromosomes, and the A chromosomes showing FISH signals for each contig

A TE chimera in the B chromosome

For a better understanding of the structure of the TE-rich interstitial region found in the B chromosome, we performed an additional bioinformatic analysis starting with Penelope-59_LMi and Tx1-1_LMi elements. First, we searched for reads showing homology with the high coverage regions of these elements, in B-carrying and B-lacking libraries, and then, we performed RepeatExplorer clustering in each case. For both elements, the graphs obtained in the B-lacking library showed a lineal pattern, typical for TEs, and the contig only contained the target TE flanking the selected region (Fig. 2a, c). In the B-carrying library, however, the graph showed two lineal structures joined by the central (selected) region (insets in Fig. 2a, c). One of these lineal structures corresponded with that found in the B-lacking genome, i.e., including target TE sequences, whereas the other contained the high coverage region of the target TE flanked by other different TEs. This difference between B-carrying and B-

lacking genomes suggested the existence of a TE chimera in the B chromosome.

In order to lengthen the known DNA sequence for these B chromosome contigs, we reiteratively aligned 454 reads with the known DNA sequences and manually extended the known contig as much as possible (Fig. S5). The contig finally obtained was 17,327 nt long and contained fragments of 28 TEs and a full SINE, in different orientations (Tables S5 and Fig. S6). We submitted this B chromosome chimeric sequence to GenBank with accession number KX611679. Remarkably, bioinformatic search failed to find evidence for the presence of the TE chimera in the two B-lacking males analyzed here, reinforcing the conclusion that it is specific to B chromosomes. Finally, in order to validate the inferred TE chimera structure, we performed a MinION sequencing experiment on gDNA from one of the four B-carrying males analyzed here. The finding of 24 reads showing full length homology with different regions of the chimera sequence, ranging from 496 to 4430 nt (Fig. S5), supported the reliability of the structure inferred from the 454 reads assembling. However, we did

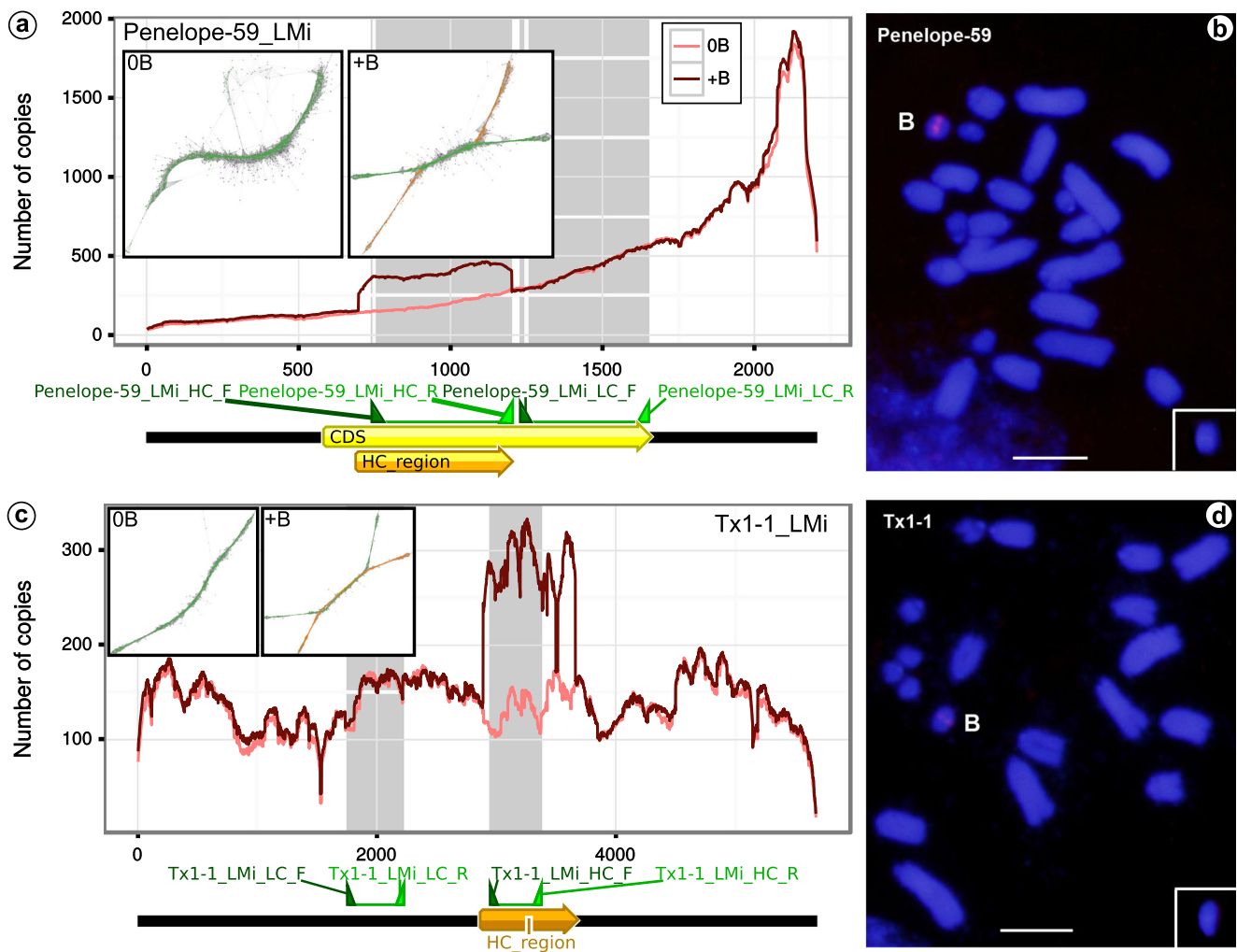


Fig. 2 Coverage analysis for two selected TEs (Penelope-59_LMi (a) and Tx1-1_LMi (c) each representing about 0.1% of B chromosome content. Note that only a discrete region of both TEs was highly overrepresented in the B-carrying libraries. In both cases, we selected the Illumina reads showing homology with the high-coverage region for each of these TEs, separately in B-carrying and B-lacking libraries, and then performed a RepeatExplorer run, which yielded different graphs for B-lacking and B-carrying libraries (see inset graphs). In the B-carrying graph, the cluster components being common with B-lacking ones are

noted in green color, whereas those being specific to the B-carrying libraries (i.e., the TE chimera) are noted in orange color. **c, d** The physical mapping by FISH for Penelope-59_LMi and Tx1-1_LMi, respectively. Note that, for both TEs, only the high coverage region showed conspicuous signals on an interstitial region of the B chromosome, located in the heterochromatic distal half being adjacent to the euchromatic proximal region. However, no signal was found using probes for a region showing no coverage differences between 0B and +B individuals (insets in **b** and **d**). Bar in **b** and **d** represents 5 μm

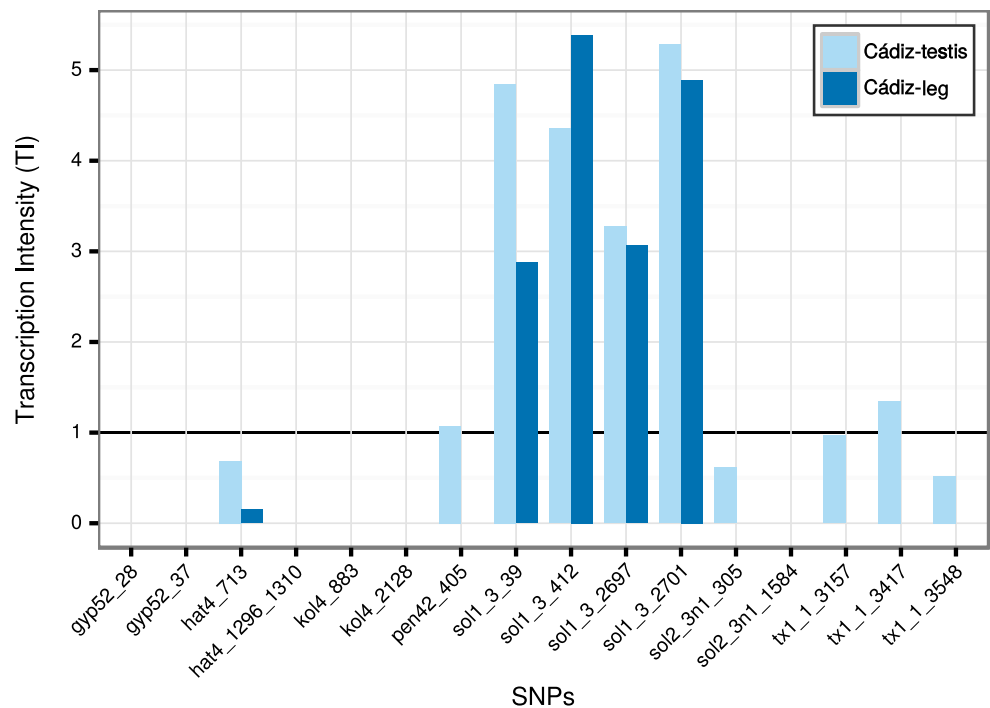
not obtain longer enough reads providing further evidence on the complete structure of the TE chimera.

Some TEs residing in the B chromosome are transcriptionally active

We searched for SNPs showing variants being specific to B-carrying individuals thus being absent from B-lacking genomes and transcriptomes. We named “Ref” the reference variant present in all individuals, thus being on A chromosomes, and “B-Alt” the B-specific alternative variant. We then calculated the frequency of Ref and B-Alt variants in genome and transcriptome libraries, and then got an estimate of

transcription intensity (TI) of the B-Alt variant as the quotient between its transcriptome and genome frequencies in each of the four B-carrying males. This analysis indicated that some of the TEs located in the B chromosome (Gypsy-52, Kolobok-4, and Sola2-3N1) showed no transcription, whereas others appear to be active (Fig. 3). For instance, hAT-4 showed transcriptional activity only for copies carrying SNPs but not for other copies carrying deletions (see also Table S6). However, the TE showing the highest TI, with high consistency between different SNPs, was Sola1-3_LMi, with slight differences between body parts and individuals (Fig. 3). In addition, although with lower intensity, we observed transcriptional activity for the Tx1-1 element, but only in testis.

Fig. 3 Transcription intensity (TI) for B-specific SNPs in B-carrying males, calculated as the quotient between the frequency of the Alt variant in RNA and gDNA from a same individual



Discussion

B chromosomes in *L. migratoria* are highly enriched in repetitive DNA, especially satellite DNA

Our present results suggest that, as a whole, repetitive DNA represents 94.9% of the B chromosome, confirming previous predictions that B chromosomes are enriched in repetitive DNA (for review, see Camacho 2005). It is noteworthy that the quantitative composition of a B chromosome in repetitive DNA had never been reported with the level of detail performed here. Remarkably, it highly differs from that of A chromosomes in *L. migratoria*, as the latter are mainly composed of TEs (54.1%) whereas TEs only constitute 19.1% of B chromosome content. In contrast, B chromosome is mainly made of satellite DNA (65.2%), with a single satellite (LmiSat02-176) constituting more than 54.8% of it, whereas satellites comprise only 2.4% of A chromosomes. In rye, A and B chromosomes contain about similar proportions of the different classes of repeats, but B chromosomes show accumulation of B-specific satellite repeats and organellar DNA (Martis et al. 2012), and several transposons are either amplified or depleted on the B chromosome (Klemme et al. 2013).

A remarkable fact of *L. migratoria* B chromosomes is that they do not show B-specific satellite DNA families, with the only exception of the satellite DNA emerged within DNA2-4 TEs (see Fig. S3b), but it was not apparent by FISH (results not shown). However, in rye and the grasshopper *Eumigus monticola*, two satellite DNA families have been observed

by FISH on B but not A chromosomes (Klemme et al. 2013; Ruiz-Ruano et al. 2017).

Bioinformatic analysis indicated that only three satellite DNAs (LmiSat02-176, LmiSat01-193, and LmiSat04-18) showed abundance higher than 1% in the B chromosome, representing 54.8, 3.1, and 2.2%, respectively, of all DNA sequences in it. However, only the two former were visualized in the B chromosome by FISH (Fig. S1). Three other satDNAs (LmiSat06-185, LmiSat53-47, and LmiSat07-5-tel) were also visualized by FISH on the B chromosome even though they showed lower abundances (0.2–0.37%). Only three of these six satDNAs (i.e., LmiSat01-193, LmiSat02-176, and LmiSat06-185) were found in the Illumina reads from the microdissected B chromosome (Table S4). All five satDNAs visualized on the B chromosome show a clustered pattern on A chromosomes (Ruiz-Ruano et al. 2016). Remarkably, the bioinformatic approach suggested that the B chromosome contains four other satDNAs (LmiSat09-181, LmiSat10-9, LmiSat16-278, and LmiSat18-210) showing abundances (0.36–0.8%) being higher than that of the least abundant satDNA which was visualized on the B chromosome by FISH (i.e., LmiSat07-5-tel). These four satellites are clustered on A chromosomes (Ruiz-Ruano et al. 2016) but were not apparent on the B chromosome, most likely because they show arrays shorter than 1.5 kb, which is the minimum threshold for FISH visualization (Schwarzacher and Heslop-Harrison 2000). The bioinformatic approach also suggested the presence of 27 other satDNAs on the B chromosome in abundances ranging from 0.002 to 0.19%, but none of them was visualized by FISH or appeared in the Illumina reads of

the microdissected B chromosome (Table S4). Likewise, our bioinformatic results suggested the presence of 45S and 5S rDNA, as well as U1 and U2 snRNA genes on the B chromosome, but they had never been visualized by FISH (Teruel et al. 2010; Anjos et al. 2015), perhaps because of their low abundance (1% or less) and their organization in small arrays. The actual presence of these satDNAs on the B chromosome cannot be elucidated with the present approach and future research might perhaps solve this conundrum. Anyway, the 27 satDNAs represent, as a whole, only 3.5% of all satDNA content in the B chromosome, and their effect on global B chromosome content is actually negligible.

Whereas the 62 satDNA families found in the A chromosomes of *L. migratoria* represent only 2.4% of the genome (Ruiz-Ruano et al. 2016), those found in the B chromosome represent 65.2% of all B sequences. This is mainly due to the massive amplification of a single satellite in the B chromosome (LmiSat02-176) as it constitutes almost 84% of all satDNA found in the B chromosome. In fact, the four B-carrying males analyzed carried 672,537 repeats of this satDNA, on average, whereas the two B-lacking individuals carried only 174,798 repeats, on which basis we estimate that about 497,739 repeats were present on B chromosomes. We can thus infer that a B chromosome in this species carries about double amount of LmiSat02-176 than all A chromosomes together, a fact that is consistent with FISH results (see Fig. 1c). The amplification of this satellite in the B chromosome might have been facilitated by the high frequency of B-bivalent formation during meiotic first prophase, frequently persisting till metaphase I thus being chiasmate (Cabrero et al. 1984), as this facilitates their amplification through unequal crossover (Smit 1976). In *L. migratoria* B chromosomes, chiasma formation is most frequent in the proximal euchromatic region (Cabrero et al. 1984), whereas LmiSat02-176 is present along whole B chromosome length (Fig. 1c). It is thus conceivable that satDNA amplification in the B chromosome could be more often achieved by sister chromatid unequal crossover than by chiasma-dependent homologous meiotic unequal crossover, as suggested for ribosomal DNA (Eickbush and Eickbush 2007).

Frequent TE insertion in an interstitial B chromosome region

At first sight, the low TE content of *L. migratoria* B chromosomes is in high contrast with the general view that these dispensable chromosomes are havens for TE proliferation because, by landing on them, TEs do not decrease host genome fitness (for review, see Camacho et al. 2000 and Camacho 2005). In fact, B chromosome scarcity for TEs, shown here, is consistent with previous observations by FISH on B chromosomes of the grasshopper *Eyprepocnemis plorans* (Montiel et al. 2012) and is also consistent with our present FISH

analysis (see Fig. S4). Bearing in mind that B chromosomes in *L. migratoria* most likely derived from A chromosomes (Teruel et al. 2010), it is worth inferring that massive amplification of satellite DNA in the B chromosome (especially LmiSat02-176) explains the sharp difference in repetitive DNA content between A and B chromosomes and is responsible for TE loss in the B chromosome after its origin, replaced by satellite DNA. This would be feasible if the B chromosome would have some constraint for maximum size, as suggested by the absence of B variants being larger than the typical one observed in all natural populations (Pardo et al. 1995).

In this context, the presence in the B chromosome of a highly complex chimeric region including many different TEs is highly remarkable, most of which are incomplete and thus inactive. The chimera comprised 17,327 bp and included DNA sequences showing homology with 29 TEs belonging to 18 different subclasses (see Table S5 and Fig. S6). The only full element was SINE/Lm2 (0.37% of TE DNA in the B), which shows homology with SINE/Lm1, a non-clustered sequence described by Bradfield et al. (1985). The reliability of this chimera was confirmed by the MinION sequencing.

This maremagnum of incomplete TEs suggests a scenario of frequent TE insertion in the B chromosome, with some elements inserted over others, successively inactivating one another. This suggests that this region is dispensable for B chromosome maintenance, for which reason it tolerates such a concentration of TE insertions. The TE chimera is located in an interstitial region of the B chromosome, within its heterochromatic half, next to the euchromatic proximal region. FISH analysis showed high abundance for six TEs located in this region (hAT-4_LMi, Helitron-N17B_LMi, Sola2-3N1_LMi, Penelope-42_LMi, Penelope-59_LMi, and Tx1-1_LMi), four of them forming part of the chimera (Fig. S6) in very high correspondence with our bioinformatic analysis of abundance in the B chromosome. The apparent clustering of these TEs on this region indicates that they are tandemly repeated, and the fact that different TEs yielded variable intensity for the FISH signal, in high correspondence with bioinformatic estimations of TE abundance, indicates that the methods developed here provide accurate estimations of TE abundance in B chromosomes. Anyway, the presence of the TE chimera means that *L. migratoria* B chromosomes can act as evolutionary sinks for transposons, as previously suggested for the ribosomal DNA of B chromosomes in the grasshopper *E. plorans* and the R2 retrotransposon (Montiel et al. 2014).

Whereas we failed to find evidence for the presence of the TE chimera in the two B-lacking individuals analyzed here, it was highly surprising to find it in the *L. migratoria* genome published by Wang et al. (2014) (Fig. S6). Given the complex structure of the chimera, it would be highly unlikely that it might have arisen twice independently, and we believe that the *L. migratoria* genome sequencing project was performed on B-carrying individuals, as is also suggested by the abundance

of LmiSat02-176 in their Illumina reads, which was twice that found in Spanish B-lacking individuals (Ruiz-Ruano et al. 2016). Likewise, Valente et al. (2014) noted a similar problem in cichlids, in the species *Metriacлина zebra*, and called attention on the need of karyotyping individuals prior to start a genome sequencing project, to avoid contamination by B chromosome sequences.

B chromosome origin

The present results on repetitive DNA content of B chromosomes in *L. migratoria* have important implications for B chromosome origin. The only previous hypothesis on this subject claimed that the B chromosome had derived from autosome 8, as it was the only A chromosome carrying histone genes, likewise B chromosomes (Teruel et al. 2010). The observed sequence divergence for H3 and H4 genes, between the A and B chromosomes, suggested that the B chromosome arose more than 750,000 years ago (Teruel et al. 2010). This long age makes it difficult identifying the A chromosome from which the B chromosome derived, as during such a long evolutionary time, the B chromosome can have experienced many changes in sequence and structure, giving it features being unrecognizable on A chromosomes, such as the TE chimera. However, the absence of B-specific satellite DNAs is puzzling for old B chromosomes like this.

The present results on satellite DNA abundance and location in the B chromosome are not fully consistent with the origin of this B chromosome from autosome 8, as autosome 9 is the only A chromosome carrying all 5 clustered satDNAs found on the B chromosome, a fact pointing to autosome 9 as possible B ancestor. This is reinforced by the exclusive presence of LmiSat53–47 on these two chromosomes. However, evidences based on satellite DNA location are weakened by the intragenomic dissemination of this kind of DNA sequences (Ruiz-Ruano et al. 2016), as the coincidence in satDNA location between autosome 9 and the B chromosome could also be due to the post-B-origin amplification of some of these 5 satDNAs, even if the B had derived from autosome 8. Therefore, with the available information, it is most parsimonious to consider autosomes 8 and 9 as putative ancestors of the B chromosome. Likewise, in rye, NGS has shown that B chromosomes arose from two different A chromosomes (Martis et al. 2012).

Although we have greatly delimited B chromosome origin in *L. migratoria*, this is a difficult problem to solve as these B chromosomes are quite old (see above). Therefore, other kinds of genetic markers would help to complete this task, especially protein-coding genes, as well as the completion of this species genome, as this would allow ascertaining which A chromosome carries the genes found in the B chromosomes, which is more reliable information than that coming from repetitive elements.

Acknowledgements We thank Francisco Jiménez-Cazalla and Jorge Doña for their collaboration in obtaining the material from Los Barrios (Cádiz). This study was supported by grants from the Spanish Plan Andaluz de Investigación (CVI-6649) and Secretaría de Estado de Investigación, Desarrollo e Innovación (CGL2015-70750-P), and was partially performed by FEDER funds. FJ Ruiz-Ruano was supported by a Junta de Andalucía fellowship.

Integrity of research All experiments comply with the current Spanish laws.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Human and animal rights and informed consent All applicable international, national, and/or institutional guidelines for the care and use of animals were followed

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Anjos A, Ruiz-Ruano FJ, Camacho JPM, Loreto V, Cabrero J, de Souza MJ (2015) U1 snDNA clusters in grasshoppers: chromosomal dynamics and genomic organization. *Heredity* 114:207–219
- Bao W, Kojima KK, Kohany O (2015) Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 6:11
- Bauerly E, Hughes SE, Vietti DR, Miller DE, McDowell W, Hawley RS (2014) Discovery of supernumerary B chromosomes in *Drosophila melanogaster*. *Genetics* 196:1007–1016
- Bradfield JY, Locke J, Wyatt GR (1985) An ubiquitous interspersed DNA sequence family in an insect. *DNA* 4:357–363
- 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
- Cabrero J, Viseras E, Camacho JPM (1984) The B-chromosomes of *Locusta migratoria* I. Detection of negative correlation between mean chiasma frequency and the rate of accumulation of the B's; a reanalysis of the available data about the transmission of these B-chromosomes. *Genetica* 64:155–164
- Cabrero J, Bakkali M, Bugrov A, Warchalowska-Sliwa E, López-León MD, Perfectti F, Camacho JPM (2003) Multiregional origin of B chromosomes in the grasshopper *Eyprepocnemis plorans*. *Chromosoma* 112:207–211
- Camacho JPM, Sharbel TF, Beukeboom LW (2000) B-chromosome evolution. *Philos Trans R Soc Lond Ser B Biol Sci* 355:163–178
- Camacho JPM (2005) B chromosomes. In: Gregory TR (ed) *The evolution of the genome*. Elsevier, San Diego, pp 223–286
- Dearn JM (1974) Phase transformation and chiasma frequency variation in locusts. *Chromosoma* 45:321–338
- Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M (2009) Geneious v.4.8.5, biomatters ltd. Auckland, New Zealand
- Eickbush TH, Eickbush DG (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* 175:477–485
- Hahn C, Bachmann L, Chevreaux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res* 41:e129

- Houben A, Banaei-Moghaddam AM, Klemme S, Timmis JN (2014) Evolution and biology of supernumerary B chromosomes, *Cell Mol. Life Sci* 71:467–478
- Hsiang W (1958) Cytological studies on migratory locust hybrid, *Locusta migratoria migratoria* L. x *Locusta migratoria manilensis* Meyen. *Acta Zool Sin* 10:53–59
- Hu J, Chen C, Peever T, Dang H, Lawrence C, Mitchell T (2012) Genomic characterization of the conditionally dispensable chromosome in *Alternaria arborescens* provides evidence for horizontal gene transfer. *BMC Genomics* 13:171
- Huang W, Du Y, Zhao X, Jin W (2016) B chromosome contains active genes and impacts the transcription of a chromosomes in maize (*Zea mays* L.) *BMC Plant Biol* 16:1
- Itoh H (1934) Chromosomal variation in the spermatogenesis of a grasshopper, *Locusta danica*. *Jap J Genet* 10:115–134
- Jang TS, Parker JS, Weiss-Schneeweiss H (2016) Structural polymorphisms and distinct genomic composition suggest recurrent origin and ongoing evolution of B chromosomes in the *Prospero autumnale* complex (Hyacinthaceae). *New Phytol* 210:669–679
- Kayano H (1971) Accumulation of B chromosomes in the germ line of *Locusta migratoria*. *Heredity* 27:119–123
- Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12:656–664
- Kim YB, Oh JH, McIver LJ, Rashkovetsky E, Michalak K, Garner HR, Kang L, Nevo E, Korol AB, Michalak P (2014) Divergence of *Drosophila melanogaster* repeatomes in response to a sharp microclimate contrast in evolution canyon, Israel. *Proc Natl Acad Sci U S A* 111:10630–10635
- King M, John B (1980) Regularities and restrictions governing C-band variation in acridoid grasshoppers. *Chromosoma* 76:123–150
- 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. *New Phytol* 199:550–558
- Kour G, Kaul S, Dhar MK (2013) Molecular characterization of repetitive DNA sequences from B chromosome in *Plantago lagopus* L. *Cytogenet Genome Res* 142:121–128
- Lepinasse R (1973) Comportement des chromosomes surnuméraires et relation avec le taux de mortalité dans une population africaine de *Locusta migratoria migratorioides*. *Chromosoma* 44:107–122
- Lepinasse R (1977) Analyse de la transmission des chromosomes surnuméraires chez *Locusta migratoria migratorioides* R. Et F. (Orthoptera, Acrididae). *Chromosoma* 59:307–322
- Li W, Fu L, Niu B, Wu S, Wooley J (2012) Ultrafast clustering algorithms for metagenomic sequence analysis. *Brief Bioinform* 13:656–668
- Littlewood DTJ, Olson PD (2001) Small subunit rDNA and the Platyhelminthes: signal, noise, conflict and compromise. In: Littlewood DTJ, Bray RA (eds) *Interrelationships of the Platyhelminthes*. Taylor & Francis, London, pp 262–278
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964
- Ma W, Gabriel TS, Martis MM, Gursinsky T, Schubert V, Vrána J, Doležel J, Grundlach H, Alschmied L, Scholz U, Himmelbach A, Behrens S, Banaei-Moghaddam AM, Houben A (2017) Rye B chromosomes encode a functional Argonaute-like protein with in vitro slicer activities similar to its a chromosome paralogue. *New Phytol* 213:916–928
- Martis MM, Klemme S, Banaei-Moghaddam AM, Blattner FR, Macas J, Schmutz T, Scholz U, Gundlach H, Wicker T, Šimková H, Novák P, Neumann P, Kubaláková M, Bauer E, Haseneyer G, Fuchs J, Doležel J, Stein N, Mayer KF, Houben A (2012) Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. *Proc Natl Acad Sci U S A* 109:13343–13346
- 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 Gen Genomics* 290:1787–1792
- Miga KH (2015) Completing the human genome: the progress and challenge of satellite DNA assembly. *Chromosom Res* 23:421–426
- Montiel EE, Cabrero J, Camacho JPM, López-León MD (2012) Gypsy, RTE and mariner transposable elements populate *Eyprepocnemis plorans* genome. *Genetica* 140:365–374
- 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
- Ning Z, Cox AJ, Mullikin JC (2001) SSAHA: a fast search method for large DNA databases. *Genome Res* 11:1725–1729
- 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 (1969) Mitotic instability leading to an accumulation of B-chromosomes in grasshoppers. *Chromosoma* 27:1–19
- 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
- Pardo MC, López-León MD, Cabrero J, Camacho JPM (1994) Transmission analysis of mitotically unstable B chromosomes in *Locusta migratoria*. *Genome* 37:1027–1034
- Pardo MC, Lopez-Leon MD, Viseras E, Cabrero J, Camacho JPM (1995) Mitotic instability of B chromosomes during embryo development in *Locusta migratoria*. *Heredity* 74:164–169
- Peng SF, Cheng YM (2011) Characterization of satellite CentC repeats from heterochromatic regions on the long arm of maize B-chromosome. *Chromosom Res* 19:183–191
- Poletto AB, Ferreira IA, Martins C (2010) The B chromosomes of the African cichlid fish *Haplochromis obliquoides* harbour 18S rRNA gene copies. *BMC Genet* 11:1
- Rees H, Jamieson A (1954) A supernumerary chromosome in *Locusta*. *Nature* 173:43–44
- Ruiz-Estévez M, Cabrero J, Camacho JPM, López-León MD (2014) B chromosomes in the grasshopper *Eyprepocnemis plorans* are present in all body parts analyzed and show extensive variation for rDNA copy number. *Cytogenet Genome Res* 143:268–274
- Ruiz-Ruano FJ, Ruiz-Estévez M, Rodríguez-Pérez J, López-Pino JL, Cabrero J, Camacho JPM (2011) DNA amount of X and B chromosomes in the grasshoppers *Eyprepocnemis plorans* and *Locusta migratoria*. *Cytogenet Genome Res* 134:120–126
- Ruiz-Ruano FJ, Cuadrado Á, Montiel EE, Camacho JPM, López-León MD (2015) Next generation sequencing and FISH reveal uneven and nonrandom microsatellite distribution in two grasshopper genomes. *Chromosoma* 124:221–234
- Ruiz-Ruano FJ, López-León MD, Cabrero J, Camacho JPM (2016) High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci Rep* 6:28333
- Ruiz-Ruano FJ, Cabrero J, López-León MD, Camacho JPM (2017) Satellite DNA content illuminates the ancestry of a supernumerary (B) chromosome. *Chromosoma* 126:487–500
- Schwarzacher T, Heslop-Harrison P (2000) *Practical in situ hybridization*. BIOS Scientific Publishers Ltd.
- Teruel M, Cabrero J, Perfectti F, Camacho JPM (2010) B chromosome ancestry revealed by histone genes in the migratory locust. *Chromosoma* 119:217–225
- 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
- Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14:178–192
- Silva DM, Pansonato-Alves JC, Utsunomia R, Araya-Jaime C, Ruiz-Ruano FJ, Daniel SN, Hashimoto DT, Oliveira C, JPM C, Porto-Foresti F, Foresti F (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 (2017) RepeatModeler Open-1.0. <http://www.repeatmasker.org>
- Smit AFA, Hubley R, Green P (2017) RepeatMasker Open-4.0. <http://www.repeatmasker.org>
- Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* 191:528–535
- 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
- Utsunomia R, Silva DM, Ruiz-Ruano FJ, Araya-Jaime C, Pansonato-Alves JC, Scacchetti PC, Hashimoto DT, Oliveira C, Trifonov VA, Porto-Foresti F, Camacho JPM, Foresti F (2016) Uncovering the ancestry of B chromosomes in *Moenkhausia sanctaefilomenae* (Teleostei, Characidae). *PLoS One* 11:e0150573
- 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
- Viseras E, Camacho JPM, Cano MI, Santos JL (1990) Relationship between mitotic instability and accumulation of B chromosomes in males and females of *Locusta migratoria*. *Genome* 33:23–29
- Wang X, Fang X, Yang P, Jiang X, Jiang F, Zhao D, Li B, Cui F, Wei J, Ma C et al (2014) The locust genome provides insight into swarm formation and long-distance flight. *Nat Commun* 5:2957
- Xie S, Marasek-Ciolakowska A, Ramanna MS, Arens P, Visser RG, van Tuyl JM (2014) Characterization of B chromosomes in *Lilium* hybrids through GISH and FISH. *Plant Syst Evol* 300:1771–1777
- Zhou Q, Zhu HM, Huang QF, Zhao L, Zhang GJ, Roy SW, Vicoso B, Xuan ZL, Ruan J, Zhang Y, Zhao RP, Ye C, Zhang XQ, Wang J, Wang W, Bachtrog D (2012) Deciphering neo-sex and B chromosome evolution by the draft genome of *Drosophila albomicans*. *BMC Genomics* 13:109