Genome relationships and LTR-retrotransposon diversity in three cultivated *Capsicum* L. (Solanaceae) species

Rafael de Assis

Universidade Estadual de Londrina

Viviane Yumi Baba

Universidade Estadual de Londrina

Leonardo Adabo Cintra

Universidade Estadual de Londrina

Leandro Simões Azeredo Gonçalves

Universidade Estadual de Londrina

Rosana Rodrigues

Universidade Estadual do Norte Fluminense Darcy Ribeiro

Andre Vanzela (✉ andrevanzela@uel.br)

<https://orcid.org/0000-0002-2442-2211>

Research article

Keywords: chili peppers, FISH, LTR retrotransposons, plant genome, transposable elements

Posted Date: October 23rd, 2019

DOI: <https://doi.org/10.21203/rs.2.16380/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Plant genomes are rich in repetitive sequences, and Transposable Elements (TE, mobile fraction) are the most accumulated of them. This mobile fraction can be divided into Class I (retrotransposons) and Class II (transposons). Retrotransposons that are transposed using an intermediated RNA and accumulate in a “copy-and-paste” way have been screened in three completely sequenced genomes of peppers (Solanaceae family). The goal of this study was to understand the genome relationships among *Capsicum annuum*, *C. chinense* and *C. baccatum*, based on a comparative analysis of function, diversity and the chromosome distribution of TE lineages in the *Capsicum* karyotypes. Due to the great commercial importance of pepper in natura, as a spice or as an ornamental plant, these genomes have been widely sequenced, and all the assemblies are available in the SolGenomics group. These sequences have been used to compare all repetitive fractions from a cytogenomic point of view. **Results:** The qualification and quantification of LTR-RT families are contrasted with molecular cytogenetics data, and the results show a strong genome similarity between *C. annuum* and *C. chinense*, compared to *C. baccatum*. The Gypsy superfamily is more abundant than Copia, especially for Tekay/Del family members, including a high representation in *C. annuum* and *C. chinense*. On the other hand, *C. baccatum* accumulates more Athila/Tat sequences. The FISH results show retrotransposons differentially scattered along chromosomes, with the exception of CRM family sequences, which mainly have a proximal accumulation associated with heterochromatin bands. **Conclusions:** The results confirm a close genomic relationship between *C. annuum* and *C. chinense* in comparison to *C. baccatum*, displaying only centromeric GC-rich bands associated with the accumulation of CRM elements, whereas terminal and subterminal AT- and GC-rich bands do not correspond to the accumulation of the retrotransposons in the three *Capsicum* species tested here.

Background

Plant genomes are composed of repetitive and non-repetitive portions, organized in families according to the nature of sequences, amplification/reduction modes, mobility throughout genomes and localization in the chromosomes [1, 2]. Transposable elements are virus-like sequences, and they are grouped in two classes, namely Class I or retrotransposon and Class II or transposon-like. The retrotransposons are the most common elements in plant genomes [1], and they use polygenic chain enzymes, such as reverse transcriptase, for retrotransposition via an intermediate RNA molecule [3-5]. Transposons, on the other hand, use different enzymatic machinery, such as transposase (transposons), helicase/replicase (helitrons), polymerase B (polintons) and tyrosine replicase for cryptons [4, 2] to reposition themselves along the genomes.

The LTR-retrotransposons are classified into *Copia*, *Gypsy*, Bel-Pao, retrovirus and endogenous retrovirus superfamilies, and they can be differentiated by the protein domain organization on the polygenic sequence [6, 2, 5]. The superfamilies *Copia* and *Gypsy* differ from each other in terms of the location of integrase in the polygenic chain, with the integrase in the *Copia* members being upstream from reverse transcriptase, while in *Gypsy* it is downstream [5, 7]. Because LTR-RTs present independent activity in the

chromosomes and different fates in the genomes, closely related species may exhibit variability in their occurrence, amount and chromosome distribution, which influence “fluctuations” in DNA C-values [8-10].

The presence of TEs within genomes can have an influence on gene expression and function; depending on their insertion region, they can change the transcript splicing/processing and closed coding regions (see [11]). In large genomes, such as in some *Fritillaria* (Liliaceae) species, with 31.1 up to 87.3 pg, the genomes are composed mostly of repetitive sequences (~90%), and about ~70% correspond to transposable elements [12]. In wheat (*Triticum aestivum*), for example, about 85% of the 17.3 pg genomes are composed of transposable elements [13]. This can also be seen in plants with smaller genomes, such as in *Cucumis melo* (Cucurbitaceae) with 1 pg of C-value, because ~43% correspond to transposable elements [14]. In some Solanaceae, such as *Solanum tuberosum* and *S. lycopersicum*, also with ~1 pg, for instance, the repetitive elements can be more than 60% [15, 16]. Species of *Capsicum* exhibit a range, 3.16 to 5.77 pg, in this regard, and the repetitive DNA families may represent 76.4% and 79.6% of *C. annuum* and *C. chinense* genomes respectively [17]. In *Nicotiana*, the DNA C-value varies from 1.4 to 6.4 pg [18], and in species such as *N. obtusifolia* with 1.55 pg and *N. attenuata* with 2.48 pg [19], the retrotransposons are 64.8% and 81% of the repetitive fractions respectively [20].

Capsicum species are an excellent model to investigate LTR-RTs’ dynamics and distribution, because they have relatively large genomes, with $2n = 24$ and 26 chromosomes. Representatives of *Capsicum* present a large diversity of repetitive DNA families, especially in relation to number and distribution of ribosomal sequences and heterochromatin sites [21-27]. In this context, we aim to know and compare the occurrence and distribution of LTR-RTs, and their association with other repetitive sequences’ diversity and location on the chromosomes. Peppers are among the most important vegetables in the world due to their high versatility and wide range of applications in cooking, industries and decoration [28]. Because of this, large investments have been made to obtain high throughput sequencing of *C. annuum*, *C. chinense* and *C. baccatum* [17, 29]. This large data volume has increased the possibilities for studying and comparing the genomic organization of these species from the cyto-molecular point of view.

Given the gaps in knowledge regarding *Capsicum* repetitive fractions, some questions about LTR-RTs’ dynamics and distribution remain unanswered, such as the following: Are the heterochromatin rich regions collocated with TE rich regions? Do close genomes share the same LTR-RT families in relation to quantity and chromosome localization? Our discussion focused on the retrotransposons’ characterization, based on a broad cytogenomic comparison using this repetitive fraction available in large *C. annuum*, *C. chinense* and *C. baccatum* datasets.

Methods

Plant materials

Seeds of *Capsicum annuum* cv. Criollo de Morelos (access GBUEL145), *C. chinense* (access GBUEL27) and *C. baccatum* (access GBUEL118), which have been identified by Dr. Leandro S. A. Gonçalves, were obtained from the gene bank of Londrina State University. Samples were sowed in 128-cell polystyrene

trays containing the substrate Vivatto®. Ten seedlings of each species were grown in the Cytogenetics and Plant Diversity Laboratory greenhouse.

Genomic analysis

The three genomes used for bioinformatic analysis were obtained from Pepper Genome Platform (<http://peppergenome.snu.ac.kr/>). Files containing scaffolds were used to search for autonomous and non-autonomous LTR-RTs. Information on the genomic dataset is shown in Table 1. For genomic comparisons, a database was compiled with conserved protein sequences of TEs from REXdb (<http://www.repeatexplorer.org>), GypsyDB (<http://www.gydb.org>), RepBase (<http://www.girinst.org/censor/>) and NCBI (<http://www.ncbi.nlm.nih.gov/>), and a second database comprised nucleotides 35S and 5S rDNA sequences from NCBI.

The repetitive fraction evaluation was conducted by the BLAST x and n tools, comparing the genomic dataset against the local databases. The results obtained are plotted in the Table S1, which includes quantitative and qualitative estimates according to the phylogeny proposal by Neumann et al. [5], as well as 35S and 5S rDNA fractions. The lineages were grouped into the superfamilies *Copia* (Ale, Alesia, Angela, Bianca, Bryco, Lyco, Gymco, Ikeros, Ivana, Osser, SIRE, TAR and Tork families) and *Gypsy* (CRM, Chlamyvir, Galadriel, Tcn1, Reina, Tekay/Del, Athila, Tat, Ogre, Retand, Phygy and Selgy). The families Athila, Tat, Ogre and Retand, due to their similarity, are hereafter referred to as Athila/Tat clade.

Autonomous and non-autonomous estimates and primer design

The BLAST output files were organized according to the LTR-RT families, including sequences greater than 3500 bp, and they were used to search for sequences with both LTRs with the Dotter I/b [30]. Subsequently, these sequences were screened according to the presence of conserved domains of reverse transcriptase, RNase H and integrase proteins. The complete sequences with the correct protein order were aligned with the MAUVE program (<http://darlinglab.org/mauve/mauve.html>), and grouped as per the similarity graphic. These data were then validated by the ClustalW alignment. From each group, a representative sequence was chosen for the graphic annotation using the IBS program (<http://ibs.biocuckoo.org/>). In order to understand the relationships among these sequences, the alignment was used to construct a bootstrapped maximum likelihood (ML) phylogenetic tree (1.000 bootstraps) in MegaX [31]. The HeatMapper tool [32] was used for the evaluation of putative autonomous elements' distribution in the three datasets using two approaches, one being for the groups of putative autonomous elements and the other for the reverse transcriptase region of these groups.

The most conserved stretch of reverse transcriptase of each LTR-RT family was used for primer designing with the Custom Primers – OligoPerfect™ Designer tool of Thermo Fisher Scientific

(<http://tools.thermofisher.com>, see Table 1). The primers' viability was assessed with the PCR primer stats tool (<http://www.bioinformatics.org>).

DNA extraction, PCR and LTR-RT probes

DNA was isolated from young leaves of each species using the cetyltrimethylammonium bromide (CTAB) method [33], purified with phenol:chloroform (1:1, v/v), chloroform:isoamyl alcohol (24:1, v/v) and RNase (1 mg mL⁻¹) and precipitated in 100% absolute ethanol. The samples were eluted in 10 mM Tris-HCl pH 8 and the concentrations estimated using a NanoDrop 2000 Spectrophotometer (Thermo Scientific).

The LTR-RT probes were obtained by PCR using specific primers for each family, with *C. annuum* as a DNA template. A standard PCR [5 U μ L⁻¹ *Taq* polymerase (0.5 μ L), 10 $\times$ buffer (2.5 μ L), 50 mM MgCl₂ (1.5 μ L), 10 mM dNTP (1 μ L), 5 mM primers (2 μ L each) and H₂O up to a final volume of 25 μ L] was used under the following conditions: 94°C for 2 min, 30 cycles of 94°C for 40 s, 59°C for 40 s and 72°C for 1 min, and a final extension of 72°C for 10 min. The reactions were tested using electrophoresis in an agarose gel at 3 V cm⁻¹ and stained with ethidium bromide. In the end, the reactions were found to be adequate for each primer set. The probes for each LTR-RT family were obtained using a reamplification of PCR products that involved labeling with biotin-11-dUTP (*Gypsy* families) or Cy3-dUTP (*Copia* families).

Cytogenetic analysis

The root tips were pretreated with 0.5% colchicine (1h 30min) and fixed in ethanol-acetic acid (3:1, v:v). The fixed material was treated in a solution of 2% cellulase and 20% pectinase, and squashed in a drop of 60% acetic acid. After liquid nitrogen freezing, the cover slips were removed, and the slides were air dried.

For fluorescence *in situ* hybridization, the slides received a mix containing a solution (30 μ L) composed of 100% formamide (15 μ L), 50% polyethylene glycol (6 μ L), 20 $\times$ SSC (3 μ L), 100 ng of calf thymus DNA (1 μ L), 10% SDS (1 μ L) and 100 ng of probes (4 μ L). The mix was denatured at 90°C for 10 min, and hybridization was performed at 37°C for 24 h in a humid chamber. Post-hybridization washes were carried out with 70% stringency, using an SSC buffer, with pH 7.0. After probe detection with an avidin-fluorescein isothiocyanate (FITC) conjugate, washes were performed in 4 $\times$ SSC/0.2% Tween-20, at room temperature, and the slides were mounted with 25 μ L of DABCO, a solution composed of glycerol (90%), 1,4-diaza-bicyclo (2.2.2)-octane (2.3%), 20 mM Tris-HCl, pH 8.0 (2%), 2.5 mM MgCl₂ (4%) and distilled water (1.7%), in addition to 1 μ L of 2 μ g mL⁻¹ 4,6'-diamidino-2-phenylindole (DAPI).

For C-CMA/DAPI banding, the samples were incubated in a sequence of 45% acetic acid (8 min), saturated solution of Ba(OH)₂ at room temperature (8 min) and in 2 $\times$ saline sodium citrate at 60°C (1h

and 30min). After this, the slides were stained with a CMA₃ for 90 min and DAPI for 30 min [34] and mounted in a solution of Mcllvaine buffer plus glycerol (1:1, v:v) with 2.5 mM MgCl₂.

Chromosome images were acquired in greyscale with a Leica DM4500 B microscope coupled with a DFC300FX camera, pseudo-colored (blue for DAPI, greenish-yellow for FITC and red for Cy3) and contrasted using GIMP 2.8 Linux.

Results

Comparative analyses based on repetitive fraction

The high coverage sequencing scaffolds of *Capsicum annuum* (3,07Gb), *C. chinense* (3,22Gb) and *C. baccatum* (3,01Gb) from the *Pepper Genome Platform* were used for the analysis. The search for conserved stretches of repetitive sequences showed that this fraction represents 2.75%, 17.14% and 2.47% respectively in each *C. annuum*, *C. chinense* and *C. baccatum* dataset (Table S1). When Class I and II elements were compared, it was observed that Class I elements were more abundant in all the *Capsicum* genomes, predominant in *C. annuum* and *C. chinense* (>90%) and less accumulated in *C. baccatum* (~70%). The conserved sequences similar to Class II elements were less represented, with <10% in these genomes (Table S1).

After organizing the sequences of Class I elements as LTR, non-LTR and endogenous retrovirus, it was noted that the LTR elements' abundance showed more similarity between *C. annuum* and *C. chinense* (89.32 and 98.78% respectively) than in *C. baccatum* with 70.55%. Both non-LTR and ERVs sequences were less accumulated, and they did not exhibit a regular distribution pattern in these species; for example, 2.21% and 0.15% in *C. annuum*, 0.06% and 0.01% in *C. chinense*, and 1.51% and 0.23% in *C. baccatum* respectively (Table S1).

As our focus was to obtain the physical mapping of LTR-RTs, the sequences of these elements were grouped in families. Members of the *Copia* superfamily were unrepresented in these three datasets (<10%), while *Gypsy* members were the most accumulated. *Capsicum annuum* and *C. chinense* exhibited 85.77% and 98.43% of *Gypsy* related sequences respectively, while in *C. baccatum* they were 63.27% (Table S1). Among the *Gypsy* superfamily, the Tekay/Del elements were predominant in *C. annuum* and *C. chinense*, with 67.68 and 95.43% respectively, while in *C. baccatum* the Tekay/Del elements presented only 15.7%. The Athila/Tat clade was more representative in *C. baccatum* (43.25%) than in *C. annuum* and *C. chinense*, at 16.36 and 2.72% respectively (Table S1). Other *Gypsy* families, such as CRM and Galadriel, had lower representation (see the Table S1 and the Figure 1C).

The other elements belonging to the non-LTR groups, transposons for example, have been less accumulated in these three datasets, such as in LINE and SINE (<5%), ERVs (<1%), CACTA (<0.1%), hAT (<2%), MuDR (<0.1%) and Helitron (<0.1%). Nevertheless, Sola elements presented an interesting contrast because they represented 5.75% and 7.31% in *C. annuum* and *C. baccatum* datasets respectively, and only 0.94% in *C. chinense* dataset. Ribosomal DNAs were also estimated because the literature shows a

large variation in the number of sites between these species. Although 5S rDNA showed no great variation among the three species, the 35S rDNA sequences exhibited a contrasting accumulation, with 18.68% in *C. baccatum*, 1.29% in *C. annuum* and 0.12% in *C. chinense* (Table S1).

Estimation and frequency of putative autonomous elements

The search for autonomous retrotransposons was directed to *Gypsy* superfamily members (Tekay/Del, CRM and Athila/Tat families) because their sequences were the most accumulated in the three datasets. The characterization of retrotransposons was based on LTRs' presence (Dotter I/b tool), annotation of GAG and polygenic chain (Artemis tool) and a minimal size of elements (>3.500 base pairs length). A total of 267 pseudomolecules were selected from these three datasets, and of these, only 7 sequences from CRM, 10 from Tekay/Del and two from the Athila/Tat families were recognized as putative autonomous elements. The other pseudomolecules were then classified as non-autonomous retrotransposons. After two rounds of ClustalW and Mauve alignments, the elected sequences were organized in groups within each family. The Tekay/Del sequences were clustered in five groups, with the sequences varying from 8.105 to 8902 bp length. The sequences of group 2 were shared between *C. annuum* and *C. chinense*, while those of groups 1 and 5 were exclusive for *C. annuum* and those of groups 3 and 4 were exclusive for *C. baccatum* (Figure S1). The CRM sequences were clustered in three groups, varying from 5.259 to 7.328 bp length. Except for group 1 and 2 that appeared in *C. annuum* and *C. baccatum*, the others were shared among the three species (Figure S2). The two sequences of the Athila/Tat family were distinct from each other, making the two groups exclusive of *C. annuum* (Figure S3).

The bootstrapped maximum likelihood phylogenetic tree organized the sequences in three clades: CRM, Athila and Del (Figure 2). The CRM sequences were clustered into 4 groups: A) CRM_annuum_2, CRM_annuum_4 and CRM_chinense_3; B) CRM_baccatum_2; C) CRM_annuum_3 and CRM_chinense_1; and D) CRM_annuum_1. The Athila sequences presented two groups of *C. annuum*. The Del elements were more diverse, being organized into three well supported groups: A) Del_annuum_3; B) Del_annuum_1 and Del_chinense_1; and C) Del_baccatum_3 and Del_baccatum_5. The remaining Del sequences did not form well-supported groups (Figure 2). One sequence was selected from each cluster for graphical representation (Figure 3B).

The clusters of putative autonomous elements were evaluated by the distribution pattern in the genomes; in general, the pattern of these sequences followed the phylogenetic proposal for the *Capsicum* genus, where the elements annotated from *C. annuum* and *C. chinense* were indeed more accumulated in these genomes, with the exception of the group Del I (Del_annuum_1), which on the heatmap (Figure 3A) exhibits greater accumulation in *C. baccatum* genome. The same was observed for the group Del V (Del_baccatum_4 and Del_baccatum_6), which despite being composed of sequences from *C. baccatum* was more accumulated in *C. annuum*. When the reverse transcriptase regions of these putative elements were analyzed, this tendency, although not exactly the same as in the complete elements, was retained,

with some exceptions such as the accumulation of the group Del II, which was composed only of sequences from *C. baccatum* (Del_baccatum_1, Del_baccatum_2, Del_baccatum_3 and Del_baccatum_5), exhibiting more accumulation in *C. chinense*.

Comparative cytogenetics

The sequences of reverse transcriptase of the most representative LTR-RT families were aligned, and four consensus sequences were used for the primer design: three for the *Gypsy* superfamily (CRM, Tekay/Del and Athila) and one for the *Copia* superfamily (Ivana/Oryco). The data can be seen in Table 1.

Fluorescence *in situ* hybridization assays revealed either scattered or clustered signals, depending on the LTR-RT family analyzed. The probe for Ivana/Oryco family (*Copia*) showed a hybridization profile with a few signals scattered along chromosomes, with clear differences between the pairs and, sometimes, with small interstitial and proximal dots (see arrowheads in Figure 4A-B and Figure S6D-I).

The Tekay/Del probe showed scattered signals along the chromosomes in three species (Supplementary figure 4). Although the signals observed in *C. annuum* (Figure 4C) and *C. chinense* (Supplementary figure 4G) were more evident in the interstitial and proximal regions of all chromosomes, in *C. baccatum* the signals accumulated in half of the chromosomes, and in the others with very weak signals (Figure 4D). These results confirm the observations from the bioinformatic analysis, which shows greater accumulation of Tekay/Del sequences in *C. annuum* and *C. chinense* than in *C. baccatum*.

The Athila/Tat probe also exhibited scattered signals along the chromosomes, being much more accumulated in interstitial regions. In *C. annuum*, except for a pair with less intense signals, the remainder exhibited brighter signals at the proximal to interstitial (close to proximal) regions, without signals in terminal ones (Figure 4E), similar to those found in *C. chinense* (Supplementary figure 6 J-L). *Capsicum baccatum* exhibited four chromosomes with less intense signals, and brighter FISH signals in the remaining chromosomes. However, four chromosomes showed stronger signals than those observed in *C. annuum* and *C. chinense* (Figure 4H). The CRM probes showed FISH signals accumulated in the proximal regions of all chromosomes in the three species (see Figures 4F-G and Supplementary figure 5). Nevertheless, there was a clear difference in signal intensity among chromosome pairs, with a minor signal in a pair of *C. baccatum* and another in *C. chinense* (see arrowheads).

The C-CMA/DAPI banding was performed to verify that LTR-RTs' accumulation areas corresponded to the AT- and GC-rich band regions as well as to check if the diversity in the distribution profiles of repetitive sequences is equivalent in these species. *Capsicum annuum* showed two pairs without bands, three pairs with terminal dots C-DAPI⁺ (dots) co-located with more intense C-CMA⁺ bands and terminal C-CMA⁺ bands in 11 chromosomes, including three pairs with stronger terminal C-CMA⁺ bands (Figures 5A-B). *Capsicum chinense* exhibits a larger number of intense C-CMA⁺ bands (seven pairs), and of these, two pairs exhibited intense adjacent C-DAPI⁺ bands. Smaller terminal C-CMA⁺ bands were observed in 11 chromosomes, and proximal bands were observed as centromeric dots in nine pairs. Some of these

bands were evidenced with DAPI and CMA₃ staining. Thinner interstitial bands were stained with C-DAPI but not with C-CMA (Figures 5C-D). *Capsicum baccatum* showed six pairs with terminal dots and six with centromeric and/or interstitial C-DAPI bands. In three pairs, these bands appeared as CMA⁺/DAPI⁺. The strongest terminal C-CMA bands were detected in four pairs in addition to thinner terminal bands in at least one arm of all chromosomes. Proximal centromeric dot bands were observed in 11 pairs, of which only three also appeared as DAPI⁺ (Figure 5E-F).

Discussion

Differential accumulation of repetitive DNA families on Capsicum genomes

TEs can move through genomes, representing an evolutionary force that modifies genome structure using different mechanisms, such as illegitimate recombination, gene capture, shuffling of regulatory motifs and generation of new functionality or silencing (see [35]). In the last instance, TEs may cause polymorphism on genomes' global structure and fluctuations in the DNA C-values [10]. TEs have been useful to compare genomes and karyotypes in evolutionary studies as well as applicable approaches, such as in grape and blood orange, where the insertion of TEs near the genes leads to a change in the gene expression [36].

The set of TEs called "Mobilome" occupy the most portion in plant genomes and play an important role in the physical and functional aspects of chromosomal structures, such as those of the CR family (centromeric retrotransposons) that are associated with chromosomal kinetics [37, 38]. In some monocotyledons, for instance, the Mobilome represents around to 75% of the genomes, such as the 80% in maize [6], and the LTR-RTs are the most dynamic elements in them [39, 7]. In *Nicotiana attenuata* and *N. obtusifolia*, for example, LTR retrotransposons reach up to 81% and 64% of genomes respectively [40]. The knowledge of LTR-RTs in a chromosomal landscape can help us understand some of the regulatory potential of TEs along chromosomes, and holds the prospect of its possible application in crop breeding programs, such as for peppers. A good example is maize, in which there are several TE families near the genes that can act as an enhancer/repressor under stressful conditions [41, 42].

Previous studies addressing TEs in the Solanaceae family have shown that *Gypsy* elements of *Solanum lycopersicum* have been more frequent than the *Copia* superfamily members [43], although an approach using only autonomous elements showed a predominance of the Tekay/Del family (*Gypsy*) and Tork (*Copia*) in this species [44]. This is interesting because it shows that estimates can vary when we consider only autonomous elements or also sequences of non-autonomous elements. Our results indicate that only 0.002% of the *C. annuum* dataset corresponds to the autonomous elements, followed by 0.001% in *C. chinense* and 0.004% in *C. baccatum*. The remaining sequences correspond to non-autonomous elements. It is important to point out that the three genomic datasets were obtained by high-covered sequencing [17, 29], which may support the assembly of the pseudochromosomes as complete elements. According to Lisch [36], the major part of coding repetitive fractions relates to fragments of

non-autonomous elements, which may be amplified by the activity of the autonomous. This could explain the high percentage of LTR-RT fragments in these three datasets.

According to Qin et al. [29], the *Gypsy* members were the most abundant LTR retrotransposons in *Capsicum*, with the highest insertion activity among Solanaceae species. When we furthered this comparison to the lowest clades of *Gypsy* and *Copia* between *Capsicum annuum* with 89% of LTR-RTs, *C. chinense* with 98% and *C. baccatum* with 70%, a predominance of *Gypsy* (>70%) over *Copia* superfamily (<10%) was evident as well as a contrasting accumulation of Tekay/Del, Athila/Tat and CRM families of *Gypsy* in these three genomes. These data point toward the importance of LTR-RTs' fate in the process of genome organization and differentiation between related species, even considering that *C. annuum* and *C. chinense* (Annum clade) are closest in relation to *C. baccatum* of Baccatum clade [45]. These three species also differ in the accumulation of 35S rDNA, with about 20% more sequences in *C. baccatum* than in *C. annuum* and *C. chinense*, besides the number of rDNA sites [25]. These genomic differences may be responsible for certain difficulties in performing interspecific crosses between these species of distinct clades because the pre- and post-zygotic barriers, as reported by Manzur et al. [46] and Cremona et al. [47].

The differential activity of retrotransposons among close genomes was also reported in *Helianthus* [48] and *Solanum* [49, 50], and the results agree with those obtained here in terms of *Capsicum*. Our data has also shown differences among *C. annuum* and *C. chinense*, especially in Tekay/Del, ERVs and Line-RTE accumulation, suggesting that other elements, besides the LTR-RT ones, evolve independently. The differential accumulation of Tnt1 retrotransposon in *Nicotiana* may be a good example for comparison as well as to support the idea of an independent fate of TEs on genome differentiation [51, 52].

Recovered Del, CRM and Athila/Tat autonomous elements support the Gypsy LTR-RTs' predominance

The ability of retrotransposons to activate and invade plant genomes may be associated with some internal and external factors, such as biotic and abiotic stresses, breeding processes, injuries, climatic changes, polyploidization, hybridization and other events (see [35, 7]). However, the activation and proliferation of TEs also depend on the ability to cheat cellular silencing controls [53]. However, only autonomous elements containing the complete and functional enzymatic machinery can do that. According to Kumar and Bennetzen's [54] criteria, autonomous elements are those that require a complete polygenic chain, regulators and both LTRs. The absence of any region qualifies it as a non-autonomous element. Following these criteria, the difference observed between potentially autonomous sequences in the three *Capsicum* datasets and those considered non-autonomous was 87.27%. It suggests that these repetitive element classes may have undergone different events of degeneration along genome differentiation.

The putative autonomous elements recovered in our analysis, i.e., ten sequences of Tekay/Del, two sequences of Athila/Tat of *C. annuum* and seven of CRM, have been quantitatively different in each

dataset, supporting the idea of the independent fate of TEs among genomes [55]. The retrotransposons' independent fate can be exemplified by the occurrence of some exclusive Tekay/Del elements in *C. baccatum* against the three in *C. annuum* and *C. chinense* as well as by the thirty-fold difference in CRM amount in *C. baccatum* in relation to *C. annuum* and *C. chinense*. This is in accordance with the report of Hawkins et al. [56], which suggests that in *Gossypium* species, different lineages of LTR-RTs evolved at different moments along genome evolutionary history, generating a threefold difference in DNA content among diploid species. In other example, De Castro Nunes et al. [38] observed a greater accumulation of CRM copies in the diploid *Coffea* species, in comparison to the hybrid tetraploid *C. arabica*.

Not all LTR-RT rich regions in Capsicum chromosomes are heterochromatin hotspots

There is a widespread idea in related literature that TEs, especially LTR-RT superfamilies, occupy "specific" chromosomal regions, with the agreement that *Copia* elements are distributed preferentially along the chromosomes associated with euchromatin, while *Gypsy* elements are resident in heterochromatin-rich regions (see [7]). In *Coffea*, *Brachiaria* and *Secale*, for example, *Gypsy* probes were located in proximal heterochromatin-rich chromosome regions [57–59], but in *Gossypium* species, *Gypsy* probes are hybridized along chromosomes [60]. However, when the elements are considered according to their phylogenetic positions, i.e., families of *Copia* and *Gypsy* [61, 5, 7], it becomes evident that there are many differences in the TE distribution profiles, in both plants (see [10, 62]) and animals [63, 64]. Thus, it seems wiser to believe that each element has its own characteristics, including chromosomal positioning, genome impact, epigenetic influence, diversification rate and other features.

Previous studies using FISH in *Capsicum* spp. have been restricted to rDNA probes [25], which demonstrated that *C. baccatum* accumulates more in terminal 35S rDNA sites compared to *C. annuum* and *C. chinense*, which exhibited just two to four pairs. This is interesting because the reports of Moscone et al. [21, 65], Scaldaferrro et al. [24] and Martins et al. [27] have shown a wide variability in the presence of terminal, interstitial and proximal heterochromatic bands in *Capsicum* species, such as the large and minor heterochromatic terminal bands in *C. annuum*, *C. chinense* and *C. baccatum* observed here. The FISH results using different LTR-RT probes showed hybridization signals accumulated from proximal (CRM) to interstitial region (Athila/Tat and Tekay/Del), scattered, or as minor dots along chromosomes (Tekay/Del, Oryco and Tork), such as in *Brachiaria* [59]. However, in no case were preferential accumulation or strong signals found in terminal chromosome regions, suggesting there is no accumulation of LTR-RT sequences at regions containing rDNA or terminal heterochromatin in *Capsicum* chromosomes. This is different from what Balint-Kurti et al. [66] found in *Musa*, where the *Copia*-Monkey retrotransposons appeared accumulated in the NOR.

The FISH using the Athila/Tat probe was strongly hybridized at proximal to interstitial regions in almost all the chromosomes, and in this case, there was also no evident co-location with heterochromatic regions, although there were small AT- and GC-rich minor bands in very few chromosomes, i.e., without evident correlation with heterochromatin hotspots. In this case as well, the scattered signals (or dots)

observed after FISH with Tekay/Del and Oryco probes are in agreement with the concept of dispersed localization of retroelements within plant genomes but without a dependence on co-localization with heterochromatin blocks. This Athila/Tat dispersion pattern, such as interstitial dots, has been described by Park et al. [67] in *C. annuum*. Using the *Passiflora edulis* for comparison, members of Ty3/Gypsy superfamily were the most accumulated, and their sequences appeared scattered along chromosomes, including at the pericentromeric regions [68]. In some Solanaceae species, such as tomato and peppers, elements of the Tekay/Del Gypsy superfamily had a scattered accumulation profile, as reported by Park et al. [68], with hybridization in chromosomes of *Solanum lycopersicum* (tomato) and *Capsicum annuum* (pepper), where pepper had a higher number of and intense signals than those observed in tomato.

Different from the other LTR-RT probes, the CRM probe exhibited intense signals in the proximal regions, associated with centromeres, and in *Capsicum* chromosomes, these regions were rich in CMA⁺ and DAPI⁺ signals. One notable exception is the Centromeric Retrotransposon lineage of Chromovirus (CRM or centromeric retrotransposon of maize), which occurs preferentially in proximal chromosome regions. CRMs carry particular domains called chromodomain (CHRoMatin Organization MOdifer DOMAIN) and CR motifs that have a potential to interact with the CENH3 centromeric protein and to participate in the centromere function [69, 37]. FISH centromere signals using CRM family probes have been described in several plant species, for example in some monocotyledon groups [70, 71, 59], suggesting that besides association with specific centromeric proteins, this accumulation may also be associated with recombination-poor regions.

Conclusions

This comparative cytogenomic analysis using the three most economically important *Capsicum* species showed great diversity in genome composition, although there is a closer approximation between *C. annuum* and *C. chinense* as compared to *C. baccatum* as is suggested for the Annuum and Baccatum complex phylogeny. The dataset screening of these three species showed that there was a differential accumulation of transposition elements, especially those from the families Tekay/Del, CRM and Athila/Tat from Gypsy, while those from the Copia superfamily were underrepresented. From the chromosomal point of view, these transposition elements were dispersed along the chromosomes (*Copia*) as well as in blocks (*Gypsy*), highlighting those of the CRM family that predominated the centromeric region. Another aspect is that LTR-RTs are not always associated with heterochromatin-rich regions. These data support the idea of the independent fate of LTR-RTs. Such genomic and chromosomal differences between closely related species should be taken into account in breeding programs, as they may interfere with the success of interspecific crosses and the introgression of agronomic traits of interest. *Capsicum* spp. proved to be a good model for most studies on the repetitive fraction, from the genomic and chromosomal point of view, considering the diversity in the accumulation and genomic distribution of LTR-RTs.

Abbreviations

Centromeric Retrotransposon of Maize: CRM

Endogenous Retrovirus: ERV

Fluorescence in situ hybridization: FISH

LTR-retrotransposons: LTR-RT

Long Terminal Repeats: LTRs

Long Interspersed Elements: LINEs

Picograms: pg

Ribosomal DNA: rDNA

Saline-Sodium Citrate: SSC

Short interspersed nuclear elements: SINEs

Sodium Dodecyl Sulfate: SDS

Transposable elements: TEs

Declarations

Ethics approval and consent to participate: not applicable

Consent for publication: not applicable

Availability of data and material: Original sequences of these three genomes can be accessed at Pepper Genome Platform (<http://peppergenome.snu.ac.kr/>)

Competing interests: The authors declare that they have no competing interests

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The authors thank the Brazilian agencies FINEP, Fundação Araucária, CNPq, CAPES, and ProPPG-Uel for their financial support.

Authors' contributions: RA performed all the experiments, wrote and corrected the manuscript. VYB and LAC performed bioinformatic analysis the experiments. LSA G and RR were responsible for varieties supply, plant maintenance, analyses and manuscript corrections. ALLV designed the study, checked the data analyses, and organized, wrote and corrected the manuscript, as the head of the group. All authors have read and approved the manuscript.

Acknowledgments: ALLV, LSAG and RR thanks CNPq for the productivity scholarship (numbers 309902/2018-5, 307911/2018-7 and 307569/2017-9, respectively) and RR for the scholarship FAPERJ E-26/202.985/2017.

References

1. Heslop-Harrison JS, Schwarzacher T. Organisation of the plant genome in chromosomes. *The Plant Journal*. 2011, 66.1: 18-33.
2. Bennetzen JL, Wang H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annual review of plant biology*. 2014, 65:505-530.
3. Kazazian HH. Mobile elements: drivers of genome evolution. *Science*. 2004, 303(5664):1626-1632.
4. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nature reviews genetics*. 2007, 8(12): 973.
5. Neumann P, Novák P, Hošťáková N, Macas J. Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mobile DNA*. 2019, 10(1): 1.
6. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: complexity, diversity, and dynamics. *Science*. 2009, 326(5956):1112-1115.
7. Orozco-Arias S, Isaza G, Guyot R. Retrotransposons in plant genomes: Structure, identification, and classification through bioinformatics and machine learning. *International journal of molecular sciences*. 2019, 20(15):3837.
8. Bennetzen JL. Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica*. 2002, 115(1):29-36.
9. Dodsworth S, Jang TS, Struebig M, Chase MW, Weiss-Schneeweiss H, Leitch AR. Genome-wide repeat dynamics reflect phylogenetic distance in closely related allotetraploid *Nicotiana* (Solanaceae). *Plant systematics and evolution*. 2017, 303(8):1013-1020.
10. De Souza TB, Chaluvadi SR, Johnen L, Marques A, González-Elizondo MS, Bennetzen JL, Vanzela ALL. Analysis of retrotransposon abundance, diversity and distribution in holocentric *Eleocharis* (Cyperaceae) genomes. *Annals of botany*. 2018, 122(2): 279-290.
11. Hirsch CD, Springer NM. Transposable element influences on gene expression in plants. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 2017. 1860(1), 157-165.
12. Ambrožová K, Mandáková T, Bureš P, Neumann P, Leitch IJ, Koblízková A, et al. Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria lilies*. *Annals of Botany*. 2011, 107(2): 255-268.
13. Wicker T, Gundlach H, Spannagl M, Uauy C, Borrill P, Ramirez-Gonzalez RH, et al. Impact of transposable elements on genome structure and evolution in bread wheat. *Genome biology*. 2018, 19(1):103.

14. Morata J, Tormo M, Alexiou KG, Vives C, Ramos-Onsins SE, Garcia-Mas J, Casacuberta JM. The evolutionary consequences of transposon-related pericentromer expansion in melon. *Genome biology and evolution*. 2018, 10(6):1584-1595.
15. Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*. 2012, 485(7400):635.
16. Mehra M, Gangwar I, Shankar R. A deluge of complex repeats: the *Solanum* PloS one. 2015, 10(8):e0133962.
17. Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, et al. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* *Nature genetics*. 2014, 46(3):270.
18. Narayan RKJ. Nuclear DNA changes, genome differentiation and evolution in *Nicotiana* (Solanaceae). *Plant systematics and evolution*. 1987, 157(3-4):161-180.
19. Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR. The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Annals of botany*. 2008, 101(6):805-814.
20. Brockmøller T, Ling Z, Li D, Gaquerel E, Baldwin IT, Xu S. *Nicotiana attenuata* Data Hub (Na DH): an integrative platform for exploring genomic, transcriptomic and metabolomic data in wild tobacco. *BMC genomics*. 2017, 18(1):79.
21. Moscone EA, Lambrou M, Hunziker AT, Ehrendorfer F. Giemsa C-banded karyotypes in *Capsicum* (Solanaceae). *Plant systematics and evolution*. 1993, 186(3-4):213-229.
22. Park YK, Kim BD, Kim BS, Armstrong KC, Kim NS. Karyotyping of the chromosomes and physical mapping of the 5S rRNA and 18S-26S rRNA gene families in five different species in *Capsicum*. *Genes & genetic systems*. 1999, 74(4):149-157.
23. Park YK, Park KC, Park CH, Kim NS. Chromosomal localization and sequence variation of 5S rRNA gene in five *Capsicum* *Molecules and cells*. 2000, 10(1):18-24.
24. Scaldaferrro MA, Grabiele M, Moscone EA. Heterochromatin type, amount and distribution in wild species of chili peppers (*Capsicum*, Solanaceae). *Genetic resources and crop evolution*. 2013, 60(2):693-709.
25. Scaldaferrro MA, da Cruz MVR, Cecchini NM, Moscone EA. FISH and AgNor mapping of the 45S and 5S rRNA genes in wild and cultivated species of *Capsicum* (Solanaceae). *Genome*. 2015, 59(2):95-113.
26. Aguilera PM, Debat HJ, Grabiele M. An integrated physical map of the cultivated hot chili pepper, *Capsicum baccatum* Pendulum. *International Journal of Agriculture & Biology*. 2017, doi: 10.17957/IJAB/15.0303
27. Martins LDV, Peron AP, Lopes ÂCDA, Gomes RLF., Carvalho RD, Feitoza LDL. Heterochromatin distribution and histone modification patterns of H4K5 acetylation and H3S10 phosphorylation in *Capsicum* *Crop Breeding and applied biotechnology*. 2018, 18(2):161-168.
28. Moreira AFP, Ruas PM., De Fátima Ruas, C, Baba VY, Giordani W, Arruda IM, et al. Genetic diversity, population structure and genetic parameters of fruit traits in *Capsicum chinense*. *Scientia*

- Horticulturae. 2018, 236, 1-9.
29. Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, et al. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. Proceedings of the national academy of sciences. 2014, 111(14):5135-5140.
 30. Sonnhammer EL, Durbin R. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene. 1995, 167(1-2):GC1-GC10.
 31. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution. 2018, 35(6):1547-1549.
 32. Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, Wishart DS. Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res. 2016. (epub ahead of print). doi:10.1093/nar/gkw419
 33. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical bulletin. 1987, 19:11-15.
 34. Schwarzbacher T, Ambros P, Schweizer D. Application of Giemsa banding to orchid karyotype analysis. Plant systematics and evolution. 1980, 134(3-4):293-297.
 35. Galindo-González L, Mhiri C, Deyholos MK, Grandbastien LTR-retrotransposons in plants: Engines of evolution. Gene. 2017, 626:14-25.
 36. Lisch D. How important are transposons for plant evolution?. Nature reviews genetics. 2013, 14(1):49.
 37. Neumann P, Navrátilová A, Koblížková A, Kejnovský E, Hřibová E, Hobza R, Macas J. Plant centromeric retrotransposons: a structural and cytogenetic perspective. Mobile DNA. 2011, 2(1):4.
 38. De Castro Nunes R, Orozco-Arias S, Crouzillat D, Mueller LA, Strickler SR, Descombes P, Vanzela AL. Structure and distribution of centromeric retrotransposons at diploid and allotetraploid *Coffea* centromeric and pericentromeric regions. Frontiers in plant science. 2018, 9:175.
 39. Negi P, Rai AN, Suprasanna P. Moving through the stressed genome: emerging regulatory roles for transposons in plant stress response. Frontiers in plant science. 2016, 7:1448
 40. Xu S, Brockmüller T, Navarro-Quezada A, Kuhl H, Gase K, Ling Z, et al. Wild tobacco genomes reveal the evolution of nicotine biosynthesis. Proceedings of the National Academy of Sciences. 2017, 114(23), 6133-6138.
 41. Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM. Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS genetics. 2015, 11(1), e1004915.
 42. Mao H, Wang H, Liu S, Li Z, Yang X, Yan J et al. A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. Nature Communications. 2015, 6, 8326.
 43. Jouffroy O, Saha S, Mueller L, Quesneville H, Maumus F. Comprehensive repeatome annotation reveals strong potential impact of repetitive elements on tomato ripening. BMC genomics. 2016, 17(1):624.

44. Paz RC, Kozaczek ME, Rosli HG, Andino NP, Sanchez-Puerta MV. Diversity, distribution and dynamics of full-length Copia and Gypsy LTR retroelements in *Solanum lycopersicum*. *Genetica*. 2017, 145(4-5):417-430.
45. Carrizo García C, Barfuss MH, Sehr EM, Barboza GE, Samuel R, Moscone EA, Ehrendorfer F. Phylogenetic relationships, diversification and expansion of chili peppers (*Capsicum*, Solanaceae). *Annals of botany*. 2016, 118(1):35-51.
46. Manzur JP, Fita A, Prohens J, Rodríguez-Burruezo A. Successful wide hybridization and introgression breeding in a diverse set of common peppers (*Capsicum annuum*) using different cultivated Aji (*baccatum*) accessions as donor parents. *PLoS One*. 2015, 10(12), e0144142.
47. Cremona G, Iovene M, Festa G, Conicella C, Parisi M. Production of embryo rescued hybrids between the landrace "Friariello" (*Capsicum annuum annuum*) and *C. baccatum* var. *pendulum*: phenotypic and cytological characterization. *Euphytica*. 2018, 214(8), 129.
48. Staton SE, Bakken BH, Blackman BK, Chapman MA, Kane NC, Tang S, Burke JM. The sunflower (*Helianthus annuus*) genome reflects a recent history of biased accumulation of transposable elements. *The plant journal*. 2012, 72(1):142-153.
49. Di Filippo M, Traini A, D'Agostino N, Frusciante L, Chiusano ML. Euchromatic and heterochromatic compositional properties emerging from the analysis of *Solanum lycopersicum* BAC sequences. *Gene*. 2012, 499(1):176-181.
50. Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, Fich EA. The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nature genetics*. 2014, 46(9):1034.
51. Vernhettes S, Grandbastien MA, Casacuberta JM. The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. *Molecular biology and evolution*. 1998, 15(7):827-836.
52. Melayah D, Lim KY, Bonnivard E, Chalhoub B, Borne FDD, Mhiri C, Grandbastien, MA. Distribution of the Tnt1 retrotransposon family in the amphidiploid tobacco (*Nicotiana tabacum*) and its wild *Nicotiana* *Biological journal of the linnean society*. 2004, 82(4):639-649.
53. Casacuberta E, González J. The impact of transposable elements in environmental adaptation. *Molecular ecology*. 2013, 22(6):1503-1517.
54. Kumar A, Bennetzen, JL. Plant retrotransposons. *Annual review of genetics*. 1999, 33(1):479-532.
55. Jurka J, Bao W, Kojima KK. Families of transposable elements, population structure and the origin of species. *Biology direct*. 2011, 6(1):44.
56. Hawkins JS, Proulx SR, Rapp RA, Wendel JF. Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proceedings of the national academy of sciences*. 2009, 106(42):17811-17816.
57. Yuyama PM, Pereira LFP, Dos Santos TB, Sera T, Vilas-Boas LA, Lopes FR, et al. FISH using a gag-like fragment probe reveals a common Ty 3-gypsy-like retrotransposon in genome of *Coffea* *Genome*. 2012, 55(12), 825-833.

58. Santos FC, Guyot R, Do Valle CB, Chiari L, Techio VH, Heslop-Harrison P, Vanzela ALL. Chromosomal distribution and evolution of abundant retrotransposons in plants: gypsy elements in diploid and polyploid *Brachiaria* forage grasses. *Chromosome research*. 2015, 23(3):571-582.
59. Zhang H, Koblířková A, Wang K, Gong Z, Oliveira L, Torres GA, et al. Boom-bust turnovers of megabase-sized centromeric DNA in *Solanum* species: rapid evolution of DNA sequences associated with centromeres. *The Plant Cell*. 2014, 26(4), 1436-1447.
60. Lu H, Cui X, Liu Z, Liu Y, Wang X, Zhou Z. Discovery and annotation of a novel transposable element family in *Gossypium*. *BMC plant biology*. 2018, 18(1):307.
61. Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viu JM, Tamarit D, et al. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic acids research*. 2010, 39(suppl_1), D70-D74.
62. Ma B, Xin Y, Kuang L, He N. Distribution and characteristics of transposable elements in the mulberry genome. *The plant genome*. 2019; doi:10.3835/plantgenome2018.12.0094
63. Schemberger MO, Nogaroto V, Almeida MC, Artoni RF, Valente GT, Martins C, et al. Sequence analyses and chromosomal distribution of the Tc1/Mariner element in Parodontidae fish (Teleostei: Characiformes). 2016, 593(2), 308-314.
64. Schemberger MO, Nascimento VD, Coan R, Ramos É, Nogaroto V, Ziemniczak K, et al. DNA transposon invasion and microsatellite accumulation guide W chromosome differentiation in a Neotropical fish genome. *Chromosoma*. 2019, 1-14.
65. Moscone EA, Lambrou M, Ehrendorfer F. Fluorescent chromosome banding in the cultivated species of *Capsicum* (Solanaceae). *Plant systematics and evolution*. 1996, 202(1-2):37-63.
66. Balint-Kurti PJ, Clendennen SK, Doležalová M, Valárik M, Doležal J, Beetham PR, May GD. Identification and chromosomal localization of the monkey retrotransposon in *Musa* *Molecular and general genetics* MGG. 2000, 263(6):908-915.
67. Park M, Jo S, Kwon JK, Park J, Ahn JH, Kim S, Kim BD. Comparative analysis of pepper and tomato reveals euchromatin expansion of pepper genome caused by differential accumulation of Ty3/Gypsy-like elements. *BMC genomics*. 2011, 12(1):85.
68. Pamponét VCC, Souza MM, Silva GS, Micheli F, de Melo CAF, de Oliveira S G, Corrêa RX. Low coverage sequencing for repetitive DNA analysis in *Passiflora edulis* Sims: cytogenomic characterization of transposable elements and satellite DNA. *BMC genomics*. 2019, 20(1):262.
69. Houben A, Schroeder-Reiter E, Nagaki K, Nasuda S, Wanner G, Murata M, et al. CENH3 interacts with the centromeric retrotransposon cereba and GC-rich satellites and locates to centromeric substructures in barley. *Chromosoma*. 2007, 116(3), 275-283.
70. Nagaki K, Neumann P, Zhang D, Ouyang S, Buell CR, Cheng Z, Jiang J. Structure, divergence, and distribution of the CRR centromeric retrotransposon family in rice. *Molecular biology evolution*. 2005, 22:845-855.
71. Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, Barry K et al. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. 2010, 463:763–768. doi: 10.1038/nature08747

Table

Table 1: General description of scaffolds files from *Capsicum* species, and primers list of LTR-RTs conserved proteins.

Features	<i>C. annuum</i>	<i>C. baccatum</i>	<i>C. chinense</i>
% GC	35	35.4	34.9
% AT	65	64.6	65.1
Number of sequences	37.989	25.349	51.917
Max length	18.549.843	16.429.935	20.331.050
Min length	264	267	274
Primers			
Athila/Tat - RT	F 5' GGGTGGTATTGCTTCTTGGA 3' R 5' GAATCACCTACCACAGAG 3'		
CRM - RT	F 5' CCACCAACAAGATAACGG 3' R 5' CCATCCATTCATAGAGACC 3'		
Tekay/De - RTI	F 5' GTTCAGGGTGCCAAGTGT 3' R 5'GGGCGTTAGTCAACCTGAAG 3'		
Ivana/Oryco - RT	F 5' GGTTCAAGGATCGGTTGATAG 3' R 5' GTTGAGCTTGCACACCATGT3'		

RT = reverse transcriptase

Figures

Figure 1

Comparison of the relative distribution (%) of repetitive DNA families among three Capsicum datasets. (A) Note that the Class I elements are more representative than Class II and rDNA sequences in the three cases. (B) LTR-RT elements predominated over non-LTR and ERVs, but note that *C. baccatum* genome exhibited 30% less sequences than other two datasets. Observe also that Copia superfamily elements (green) were less accumulated (<10%) than Gypsy ones. (C) Observed that the most accumulated elements were Tekay/Del, Athila/Tat and CRM, but, except to CRM family, there were a big difference in the elements quantity in each dataset.

Figure 2

Phylogenetic tree based on putative autonomous sequences of Capsicum using the maximum likelihood method with bootstrap 1000. CRM sequences organize four groups (dark grey) and Athila sequences two groups (blue). The Del sequences were organized in two well supported groups (orange), corroborating with the MAUVE alignment (Figure S1), and a third group of sequences without well supported values (green), except for the sequences Del_baccatum_3 and Del_baccatum_5 (Figure S1) corroborating with the MAUVE alignment.

Figure 3

Comparative distribution of the autonomous LTR-RTs and the reverse transcriptase sequences along the *C. annuum*, *C. chinense* and *C. baccatum* datasets. (A) The Heatmap image shows that Athila/Tat elements accumulated more in *C. annuum* and *C. chinense* (marked in yellow and light-yellow), the CRM groups were differentially accumulated, highlighting the absence of CRM III and the predominance of Del I and II in *C. baccatum*, and bigger accumulation of Del III, IV and V in *C. annuum* and *C. chinense*. The reverse transcriptase of these elements exhibits a similar pattern of distribution than the one observed for the complete elements, Athila/Tat I and II were more accumulate in *C. annuum* and *C. chinense*, respectively. CRM I and II were more accumulated in *C. chinense*, while the group CRM III was more accumulated in *C. baccatum*. Capsicum annuum and *C. chinense* had a bigger accumulation of Del I, II and II than *C. baccatum*, while the groups Del III and IV exhibited more accumulation in *C. baccatum*. (B) Graphical representation of LTR-RTs groups. LTR – long terminal repeat, GAG – nucleocapsid, RT – reverse transcriptase, RH – RNase H, INT – integrase, ASP – aspartase. Note a difference in extension (bp length) among elements, including GAG and POL positioning, and LTR sizes. Note also that only CRM II exhibits a chromodomain sequence and that all the Del elements present an additional aspartase locus.

Figure 4

FISH using LTR-RTs probes against metaphases and prometaphases of Capsicum species. Chromosomes were counter-stained with DAPI (blue), Copia probes with Cy3-11-dUTP (red) and Gypsy probes labelled with biotin-11-dUTP / avidin-FITC conjugate (green). The Copia Ivana/Oryco probe showed few hybridization signals scattered along chromosomes, with a low accumulated profile in both

C. chinense (A) and *C. baccatum* (B). The Gypsy Tekay/Del probe exhibited hybridization signals dispersed along the chromosomes in the three species, but with a larger accumulation in *C. annuum* chromosomes (C) than the other two species, such as in *C. baccatum* (D). The Gypsy Athila/Tat probe showed brighter hybridization signals than Tekay/Del, accumulating in the pericentromeric to interstitial regions of all *C. annuum* chromosomes (E), differently of *C. baccatum* because some chromosomes accumulated many signals and others very few (H). The boxes i, ii, iii and iv are highlighting differences in the pericentromeric and interstitial Athila/Tat signals in two *C. baccatum* chromosomes. The Gypsy CRM probe showed FISH signals accumulated in the centromeric regions, but with two pairs in each species with much less intense signals. Note the arrows in *C. baccatum* (F) and *C. chinense* (G). Bar represents 10 μm .

Figure 5

Capsicum species shows a considerable diversity in the C-CMA/DAPI banding profiles. Observe that *C. annuum* presents four chromosomes without fluorescent signals. C-DAPI interstitial dots were detected in three chromosome pairs and terminal bands two pairs (A), all of them were co-located with C-CMA bands. Ten pairs exhibit C-CMA signals, being three pairs with stronger terminal and the other as terminal to subterminal small signals (B). *Capsicum chinense* showed four pairs with strongest DAPI signals,

besides minor centromeric, interstitial and terminal signals on a few chromosomes (C), while C-CMA bands were observed in all the chromosomes, varying as strongest terminal bands in seven pairs, minor centromeric bands in six pairs and as terminal and interstitial dots in ten pairs (D). Some of these bands have been evidenced with DAPI and CMA3 (C-D). Observe that *C. baccatum* exhibits six pairs with minor terminal and six with centromeric and/or interstitial C-DAPI (E). C-CMA signals were detected in four pairs as strongest terminal bands, but minor terminal bands were observed in all chromosomes, as well as minor centromeric in almost all the chromosomes (F). Bar represents 10 μm

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterials.pdf](#)