DOI: 10.1007/s11033-019-05047-4

Molecular Biology Reports (2019) 46:6117–6133
https://doi.org/10.1007/s11033-019-05047-4

ORIGINAL ARTICLE

Transposable element discovery and characterization of LTR-retrotransposon evolutionary lineages in the tropical fruit species *Passiflora edulis*

Zirlane Portugal da Costa[1] · Luiz Augusto Cauz-Santos[1] · Geovani Tolfo Ragagnin[2] · Marie-Anne Van Sluys[2] · Marcelo Carnier Dornelas[3] · Hélène Berges[4] · Alessandro de Mello Varani[5] · Maria Lucia Carneiro Vieira[1]

Received: 14 March 2019 / Accepted: 28 August 2019 / Published online: 24 September 2019

Abstract

A significant proportion of plant genomes is consists of transposable elements (TEs), especially LTR retrotransposons (LTR-RTs) which are known to drive genome evolution. However, not much information is available on the structure and evolutionary role of TEs in the Passifloraceae family (Malpighiales order). Against this backdrop, we identified, characterized, and inferred the potential genomic impact of the TE repertoire found in the available genomic resources for *Passiflora edulis*, a tropical fruit species. A total of 250 different TE sequences were identified (96% Class I, and 4% Class II), corresponding to ~ 19% of the *P. edulis* draft genome. TEs were found preferentially in intergenic spaces (70.4%), but also overlapping genes (30.6%). LTR-RTs accounted for 181 single elements corresponding to ~ 13% of the draft genome. A phylogenetic inference of the reverse transcriptase domain of the LTR-RT revealed association of 37 elements with the *Copia* superfamily (*Angela*, *Ale*, *Tork*, and *Sire*) and 128 with the *Gypsy* (*Del*, *Athila*, *Reina*, *CRM*, and *Galadriel*) superfamily, and *Del* elements were the most frequent. Interestingly, according to insertion time analysis, the majority (95.9%) of the LTR-RTs were recently inserted into the *P. edulis* genome (< 2.0 Mya), and with the exception of the *Athila* lineage, all LTR-RTs are transcriptionally active. Moreover, functional analyses disclosed that the *Angela*, *Del*, *CRM* and *Tork* lineages are conserved in wild *Passiflora* species, supporting the idea of a common expansion of *Copia* and *Gypsy* superfamilies. Overall, this is the first study describing the *P. edulis* TE repertoire, and it also lends weight to the suggestion that LTR-RTs had a recent expansion into the analyzed gene-rich region of the *P. edulis* genome, possibly along WGD (Whole genome duplication) events, but are under negative selection due to their potential deleterious impact on gene regions.

Keywords *Passiflora* · Passion fruit · Genome evolution · Mobile genetic elements · Reverse transcriptase

Introduction

First described in maize by Barbara McClintock in the middle of the twentieth century [1], transposable elements (TE) are DNA segments that have the ability to move within the genome [2]. TEs are often found in eukaryotic genomes, with little-known exceptions [3, 4] and, for many years, were considered 'junk DNA' or 'selfish DNA parasites', without any benefit, until they were discovered to cause substantial and deleterious mutations. Because of their mutagenic potential, genomes have evolved and created diverse mechanism to suppress their activity (see [5]). On the other hand, genome-scale studies have revealed that TEs play a key role in genome function, chromosome evolution, speciation and diversity [4, 6].

Long terminal repeat retrotransposons (LTR-RT) are the predominant order of mobile genetic elements found in plant genomes [7], accounting for 97.7% of all TEs and 61.8% of the of *Solanum lycopersicum* genome [8] and consisting of 62.9% of all TEs and 41.7% of the *Pinus taeda*, loblolly pine genome [9]. LTR-RTs are also responsible for genome expansion in some species, such as *Capsicum annuum*, with the accumulation of LTR-RTs and their derivatives [7, 10].

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11033-019-05047-4) contains supplementary material, which is available to authorized users.

✉ Maria Lucia Carneiro Vieira
mlcvieir@usp.br

Extended author information available on the last page of the article

Structurally, these elements have long terminal repeats at both 5′ and 3′ ends in direct orientation that are responsible for controlling the transposition mechanism [2] and downstream located genes [11]. Autonomous LTR-RTs also have internal domains that encode for proteins related to their transposition: GAG (Group-specific Antigen) and POL (synthesized as a polyprotein containing PROT, Aspartic Proteinase; INT, Integrase, RT, Reverse Transcriptase and RH, RNAse H). In some cases, they even include functional ENV (Envelope) proteins [12]. Some LTR-RTs have a chromodomain (Chromatin Organization Modifier Domain, CHD) of 40–50 amino acids involved in chromatin remodeling and gene expression regulation in eukaryotes [13, 14].

LTR-RTs are divided into five superfamilies based on alignments and phylogenetic inference according to the domains of the GAG and POL proteins [15, 16]. Three superfamilies are found only in Metazoans (*Bel-Pao*, *Retrovirus* and *ERV*), and the other two (*Copia* and *Gypsy*) tend to be the most common in plants. *Copia* and *Gypsy* elements differ mostly in the organization of their internal domains [3, 17]. The *Copia* superfamily is further subdivided into evolutionary lineages (*Angela*, *Ale*, *Bianca, Ivana, Oryco, Retrofit, Sire, Tork* and *Maximus*), as are those belonging to the *Gypsy* superfamily (*Athila*, *CRM*, *Del*, *Galadriel* and *Reina*). These lineages share relationships in both sequence similarity and molecular organization, and their identification and characterization are central to understanding the transposition mechanism and evolutionary history of the LTR-RT system [7, 18, 19].

These elements are frequently found in heterochromatic regions, such as regions harboring centromeres, that show very low recombination rates, but others can be found spreading over the chromosome. Some retrotransposon families are found in gene-rich regions where there is strong control, given the deleterious effects that TEs can cause on genes. Thus, large LTR-RTs and TE families with high copy numbers are less tolerated in these regions, and are generally removed due to their large impacts on genome structure. TEs can disrupt genes and influence gene expression and genome evolution, as well as diversity. A number of other roles and impacts are related to the activity of LTR-RTs, such as alternative splicing, epigenetic control, transduction, duplication, recombination and many other nuclear processes [20, 21]. Elucidating the contribution of LTR-RTs to genome structure and function is therefore essential for understanding the evolutionary plasticity of genomes.

Passiflora (Passifloraceae, Malpighiales) is a Neotropical genus accounting for hundreds of species with widespread distribution in the American continent, including the Amazonian and Andean regions. Population pressure in all these regions is high, raising considerable concern about conserving *Passiflora* diversity. The species are highly variable in morphological terms, with substantial variation in genome size (0.212 pg in *P. organensis* up to 2.68 pg in *P. quadrangularis*, [22]). Taxonomically speaking, they are grouped into four subgenera: the ancestral *Astrophea* (57 species); *Decaloba* (220); *Deidamioides* (13); and the recent *Passiflora* subgenus (240), which contains several self-incompatible species [23]. Popularly known as passionflowers or passion fruits, the cultivated species of *Passiflora* are of economic importance worldwide for both juice production and *in natura* consumption. In Brazil in particular, *P. edulis* is the main cultivated species. Despite the interesting attributes of the genus *Passiflora* and the economic relevance of *P. edulis*, during the last decade only a few studies have been conducted to shed light on its transcriptome [24] and genome structure [25].

Our research group has conducted several genetic and genomic studies to enrich our knowledge of *P. edulis* [26–30], including a study by Santos and co-workers [31] that sequenced some 10,000 BES (BAC-end sequences of up to 1000 bp), generating approximately 6.2 Mb of data. Among those sequences, 9.6% were similar to plant genes. Interestingly, it was found that 19.6% of the sequences consisted of repetitive elements, most of which (94.4%) are transposable elements. Additionally, we further sequenced over 100 large-inserts from a *P. edulis* genomic library using a long-read sequencing platform, constituting a gene-rich fraction of the *P. edulis* genome, and representing the most comprehensive genomic resource for the species so far [32]. Furthermore, a draft genome assembled from short-reads is also publicly available at GenBank [25, 33].

Therefore, this scenario prompted us to perform a further and detailed analysis of the TE repertoire to advance our knowledge of *P. edulis* genome structure and evolution. We were able to generate a comprehensive atlas of 250 TE sequences using the standard nomenclature and classification [3] in order to facilitate information exchange. Importantly, the LTR-RT order was analyzed at lineage level. In addition, the transcriptional profile of the *P. edulis* TE repertoire and the presence of LTR-RT lineages in wild *Passiflora* species were also characterized. Compiling our results with the available genomic data, we found that TEs are preferentially found in intergenic spaces, although some overlapped gene sequences. Together with the predicted recent activity and abundance of LTR-RTs characterized herein, our data suggests a recent expansion of these elements into the analyzed gene-rich region of the *P. edulis* genome, which may have occurred along WGD events. However, we hypothesize that a strong negative selection, as a rule, is acting on these elements to prevent deleterious impacts on host fitness. Remarkably, this is the first detailed characterization of mobile genetic elements in Passifloraceae, the passionflower family.

Methods

Identification and characterization of *P. edulis* transposable elements

The *P. edulis* TE repertoire was determined in silico using the gene-rich genome fraction as a template [32] and the draft genome assembly [25].

Initially, the REPET pipeline was used to classify and annotate the TEs [34, 35] recognized in the gene-rich fraction of the *P. edulis* genome, and further classify them into evolutionary lineages. The *P. edulis* draft genome was then screened for these elements using RepeatMasker [36], with the aim of evaluating the order and expansion of evolutionary lineages.

Previously identified domains were annotated using the Conserved Domain Database (CDD) [37] in NCBI's interface. The coordinates of all features of each element were recorded in an Excel sheet and the information used to create a schematic representation of each element in IBS (Illustrator for Biological Sequences) [38].

Terminal inverted repeats (from DIRS and TIR elements) were recognized using the Einvert tool in the EMBOSS open software suite [39]. TSDs (Target site duplications—from LINE, SINE, LARD, TRIM and TIR elements) were identified by comparing the 5′ and 3′ flanked sequences of each element using GenomeView [40] in order to visualize the element location and its flanking sequences.

To define genomic location, TE sequences were mapped against the previously annotated *P. edulis* BAC sequences [32]. To do this, we used the BWA-MEM alignment algorithm [41]. TE locations (within intergenic or genic regions) were examined using GenomeView, based on the *P. edulis* gene models available at the CoGe website (https://genomevolution.org/coge/GenomeInfo.pl?gid=52053). All TE sequences identified in this study are available in GenBank database.

Phylogenetic analysis of LTR-RT elements, assignment to evolutionary lineages, naming of sequences and structural features

The Repeat Explorer web server [42, 43] was used to search for LTR-RT coding domains (GAG, PROT, INT, RT, RH, CHDII, and CHDCR), based on minimum similarity 60% and minimum identity 40%, and the proportion of hit length to database sequence length was set to 0.8.

Passiflora edulis LTR-RTs were assigned to evolutionary lineages based on phylogenetic analysis. For complete LTR-RT elements, phylogenetic analysis was based on the translated RT domain. For incomplete LTR-RT elements, lacking the RT domain, the GAG domain was used, as this domain was the most frequent and conserved. Domains from previously described lineages were retrieved from the *Gypsy* Database (GyDB) [15] and domains from previously characterized LTR-RT lineages from sugarcane [44] were used in phylogenetic analysis to assign *P. edulis* elements to lineages. A total of nine *Copia* and six *Gypsy* lineages were used to initially classify the LTR-RTs from *P. edulis*. Phylogenetic inferences were then made for each superfamily individually.

Translated domains were aligned using MUSCLE [45] implemented in MEGA 7.0 [46], with default parameters, and then manually verified and edited in order to correct amino acids incorrectly aligned and remove alignment blocks of low quality to obtain an optimal curated alignment suitable for use in phylogenetic analysis. Model Generator was used to find the best amino acid substitution model for molecular evolution and gamma rate heterogeneity using the Akaike Information Criterion (AIC) [47]. All phylogenetic analyses were drawn using the highest-ranked substitution model available. Phylogenetic trees were constructed using RAxML [48], applying the Maximum Likelihood method with 1000 bootstrap replicates. Only groups supported with a high bootstrap value (> 50) were considered. Trees were visualized using FigTree v1.4 [49] and edited using the graphical editors Dendroscope [50] and TreeGraph 2 [51] to generate an Image Format. The line colors indicate each group representing a lineage.

Passiflora edulis LTR-RTs were assigned to evolutionary lineages based on phylogenetic inferences. Only groups supported with a high bootstrap value (> 50) were considered. On the basis of a proposed universal classification of TEs [3], we were able to assign names to LTR sequences. We standardized the name of *P. edulis* LTR-RT sequences, following the example of Domingues and co-workers [44] for sugarcane sequences. Sequences were named 'RLC' (*Copia*) or 'RLG' (*Gypsy*), 'pe' for '*Passiflora edulis*' and the lineage name, e.g. '*Angela*'. The BAC sequence within the element was located (e.g. 'Pe1K19'); then each sequence of the same lineage and within the same BAC clone was numbered sequentially. For instance, 'RLC_*peAngela*_Pe1K19-1' is the first *P. edulis Angela* element found in BAC Pe1K19, from the superfamily *Copia*.

Entire sequences of LTR-RTs were submitted for pairwise comparisons with the blast2seq tool [52]. In addition, full lengths elements were self-aligned using MUSCLE [45], implemented in MEGA 7.0 [46], to confirm the presence of the boundaries formed by long terminal repeats.

TSDs were identified by comparing the 4 to 6 bp of the 5′ and the 3′ flanked sequences of each element using GenomeView [40], and to visualize each LTR-RT location and its flanking sequences. An LTR-RT was considered a complete

element only if it exhibited all the coding domains necessary for its transposition (GAG, PROT, INT, RT and RH) and the two long terminal repeats. Coordinates of all features of each of the complete LTR-RTs were recorded in an Excel sheet and the information used to create a schematic representation in IBS [38].

In addition, to briefly investigate the evolutionary role of LTR-RTs in *P. edulis* genome expansion, a WGD analysis was performed based on transcriptome data previously obtained from three RNA-seq libraries of shoot apexes of juvenile, vegetative and reproductive adult *P. edulis* plants (details provided in [32]). For that, we employed the transdecoder pipeline (https://github.com/TransDecoder/TransDecoder/wiki) coupled with the WGDdetector pipeline [53] to predict the WGD events.

Estimation of LTR-RT insertion time

At each transposition cycle, an LTR-RT creates a new copy of itself. It is assumed that both LTRs are often identical at the time of transposition. Therefore, in this study, the insertion time of intact LTR-RT copies was obtained based on the supposition that LTRs are identical at the time of integration [54, 55]. This entailed aligning the 5′ and 3′ LTR sequences of each element using MUSCLE [45]. The divergence among LTRs (K) was obtained in MEGA 7.0 [46] using the Kimura-2-parameter distance [56]. The insertion time (t) for each full-length element was calculated according to the formula: $t = K/2r$ [57], where t is the insertion time in Mya (million years ago), K is the number of nucleotide substitutions per site and r is the nucleotide substitution rate. For r we used the value 1.5×10^{-8}, as reported for the *Arabidopsis* chalcone synthase and alcohol dehydrogenase genes [58], and used for dating LTR-RTs in *Linum usitatissimun* [59], *Vitis vinifera* and *Solanum* spp. [60] and *Eucalyptus grandis* [61].

In silico analysis of LTR-RT transcription

Herein, the assembled transcripts of the transcriptome data previously obtained [32] were mapped against all full-length LTR-RTs. Mapping was performed using the BWA-MEM package [41], and transcripts similar to LTR-RTs were assigned to a lineage based on Wicker criteria [3], i.e., 80% coverage and 80% nucleotide identity. In addition, the number of transcripts associated with LTR-RTs was normalized using the Cufflinks package [62] available in the Galaxy suite [63].

RNA extraction and reverse transcriptase PCR analysis

Plants in vegetative growth cultivated under greenhouse conditions were used to collect leaf tissues. Total RNA was extracted with TRIzol (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. Primers were manually designed based on the sequences inside the RT (reverse transcriptase) domain of the full-length LTR-RT elements (Supplementary Table 1). Gene Runner [64] was used to verify the quality of the primer sequences as follows: primers from 18 to 22 bases in length; GC content of 50–60%; absence of secondary structures, like dimers or hairpins; T_m (melting temperature) from 58 to 62 °C.

To remove residual genomic DNA, total RNA preparation was then treated with DNAse (Promega, Madison, WI, USA). One RNA µg was used to generate the first cDNA strand using the SMARTer™ PCR cDNA Synthesis Kit (Clontech Laboratories, Mountain View, CA, USA) following the manufacturer's instructions. The reaction contained 1 × M-MLV RT buffer, 0.6 mM of each dNTP, 25 U of RNasin® (Promega), 200 U of M-MLV RT (Promega) and 0.25 µM of primer CDS. Diethylpyrocarbonate-treated water (0.01%) was added to make up the final volume of 25 µL, and the reaction was incubated at 42° C for 1 h.

cDNA dilutions were used in PCR reactions as follows: 3.0 µL of cDNA, 1 × PCR buffer, 0.25 mM of dNTP, 0.6 µM of primer, 2 mM of $MgCl_2$, and 1.2 U of Taq DNA polymerase (Promega). Ultrapure water was added to make up the final volume of 20 µL. Reactions were incubated for 5 min at 95 °C (denaturing step), then 30 cycles of 1 min at 95 °C, 1 min at 55 °C, 1 min at 72 °C and then 8 min at 72 °C. PCR products were analyzed in electrophoresis with 1.2% agarose gels stained with SYBR SAFE® (Invitrogen), using 1 × TBE as the running buffer, and with the 100 bp DNA Ladder (Invitrogen) as a marker for comparison.

LTR-RTs from wild *Passiflora* species

We also investigated the presence of the LTR-RTs in other species of the *Passiflora* genera. The following wild species were screened: *P. edmundoi*, *P. setacea* and *P. alata* (subgenus *Passiflora*); *P. organensis* (subgenus *Decaloba*); *P. deidamioides* and *P. contracta* (subgenus *Deidamioides*) and *P. rhamnifolia* (subgenus *Astrophea*). These species belong to the *Passiflora* collection kept in our laboratory (Genetics Department, 'Luiz de Queiroz' College of Agriculture, University of São Paulo, Brazil).

The primers described above (Supplementary Table 1) were used to amplify the Reverse Transcriptase domain using the total DNA of the wild species as a template; this DNA was extracted from young leaf tissues using the acetyltrimethyl ammonium bromide method adapted from Murray and Thompson [65]. PCR and gel electrophoresis procedures were as detailed above.

Results

Identification and characterization of Transposable Elements in the *Passiflora edulis* genome

We first exploited the *P. edulis* TE repertoire recognized in a gene-rich fraction of the *P. edulis* genome [32], and performed in-depth characterization of these elements up to evolutionary lineages. Subsequently, we examined the draft genome to evaluate the most common orders and evolutionary lineages' expansion trends. It is important to clarify that the *P. edulis* draft genome [25] shows a high level of fragmentation and low-quality assembly (> 234,000 scaffolds, N50: 1311 bp; 194 scaffolds > 10 Kb, spanning 165 Mb), so it cannot be used for primary TE identification.

The search for TEs resulted in the identification of 250 unique elements corresponding to 19% of the *P. edulis* draft genome (Table 1, Supplementary file 1, Supplementary Table 2). Most of Class I and II elements correspond to incomplete elements (177 out of 250). Furthermore, TEs were predominantly located in intergenic spaces of the gene-rich fraction of *P. edulis* (176/250), but 74 were found to overlap genes (70 exonic and only four intronic sequences).

Retrotransposons were the most common. LTR-RTs were the most frequent elements, accounting for ~ 13% of the data set. Eleven incomplete elements (7933 to 13,640 bp in length) were recognized as belonging to the DIRS (*Dictyostelium* Intermediate Repeat Sequence) order, and were found to have some typical domains in different arrangements (Supplementary Fig. 1), suggesting they are relics of ancient amplification events. LINEs (Long Interspersed Nuclear Elements) and SINEs (Short Interspersed Nuclear Elements) were represented by only 7 and 2 copies, respectively, corroborating their low frequency in plant genomes [66]. Thirty-six elements were recognized as belonging to the LARDs (Large Retrotransposon Derivatives) order, including 3 LARDs that harbor genes inside their sequences and another 4 that were found within introns (Supplementary Table 3).

Only 3.6% (9/250) of the TEs were assigned to Class II (DNA transposons), most of them (6) to the TIR order. We were also able to identify two Helitrons which contained an additional domain bearing fragments of the toll-Interleukin-1 Receptor (IPR000157) that is involved in signaling processes [67], and Leucine-rich Repeats (IPR001611). Both domains might be products of the Helitron gene capturing process. DNA transposons accounted only for 0.26% of the *P. edulis* genome, the majority in the TIR order.

Subsequently, 91.2% (165/181) of the LTR-RTs were classified into evolutionary lineages within each superfamily. Thirty-seven of them were assigned to *Copia* lineages and 128 to *Gypsy* lineages. *Copia* and *Gypsy* elements had similar proportions of internal recognizable domains. The majority of *Gypsy* elements (~ 60%) had up to four internal domains. Six *Gypsy* elements had none of the main internal domains, but contained putative sequences of chromodomains, used to classify them as *Gypsy*. All *Copia* elements had at least one domain; the majority exhibited five domains (Fig. 1a). Taking each domain separately, their frequencies

Table 1 Classification and abundance of transposable elements (TEs) identified in two data sets of *Passiflora edulis*

	Order	Gene-rich fraction[a]			Draft genome[b]		
		Number of elements	Total length of elements (bp)	Percentage of nucleotides	Fractional element occurrences in the genome	Total length of elements (bp)	Percentage of nucleotides
Class I	SINE	2	614	0.01	1428	120,742	0.07
	LINE	7	54,468	0.52	10,959	1,923,670	1.16
	LTR/Copia	44	324,682	3.12	25,508	4,204,432	2.54
	LTR/Gypsy	137	1,093,707	10.52	102,612	18,945,287	11.44
	DIRS	11	115,640	1.11	7642	1,125,943	0.68
	LARD	36	189,161	1.82	31,159	4,512,069	2.72
	TRIM	4	4981	0.05	1026	140,829	0.09
Class I total		241	1,783,253	17.15	180,334	30,972,972	18.70
Class II	TIR	6	32,748	0.31	2643	358,602	0.22
	Helitron	2	13,827	0.13	507	71,418	0.04
	MITE	1	792	0.01	3	134	0.00
Class II total		9	47,367	0.45	3153	430,154	0.26
Total		250	1,830,620	17.60	183,487	31,403,126	18.96

[a]TEs identified in a gene-rich fraction of the *Passiflora edulis* genome (Munhoz et al. [32])

[b]TEs identified in Munhoz et al. [32] and mapped against the draft *P. edulis* genome released (Araya et al. [25])

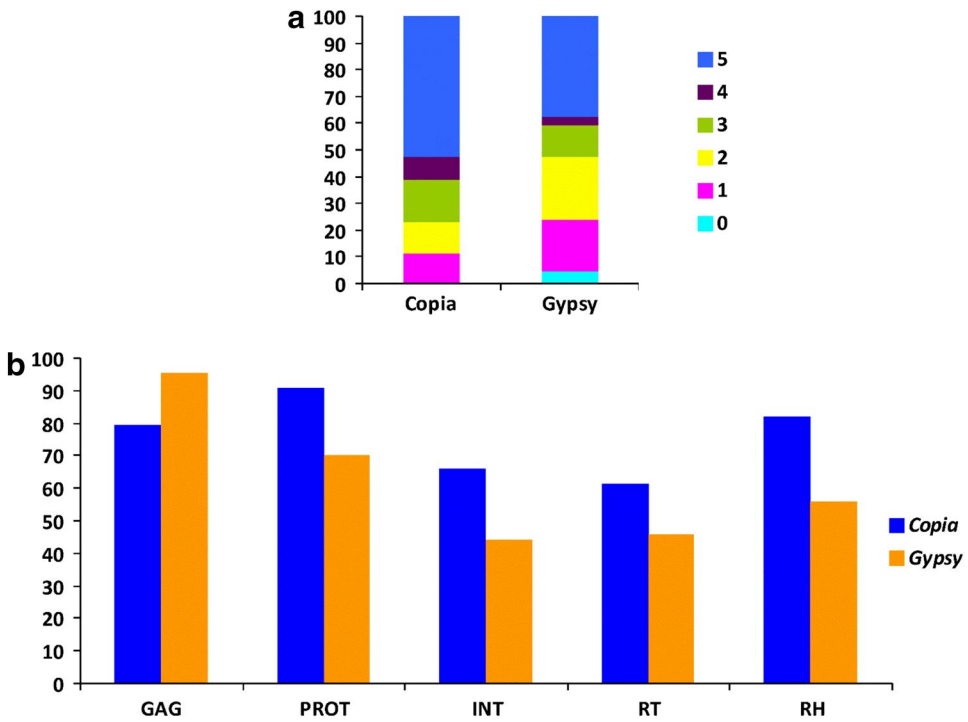


Fig. 1 LTR-RT elements of *Passiflora edulis*: **a** percentage of elements from *Copia* and *Gypsy* superfamilies that contain up to five domains, displayed in different colors; **b** percentage of elements from *Copia* and *Gypsy* superfamilies containing the following internal domains: GAG, PROT, INT, RT and RH

were higher in *Copia* elements than in *Gypsy*, except for the GAG domain, which is prevalent in *Gypsy* (Fig. 1b). This is evidence of the higher level of conservation of *Copia* compared to *Gypsy* elements, although in absolute terms, *Gypsy* contained over three times more elements. In fact, *Gypsy* elements were the most common, accounting for 10.52% and 11.4% of the gene-rich fraction and the draft genome, respectively.

TSDs ranged from 4 to 6 bp (average 4 bp), and were identified in 134 LTR-RTs., while LTRs were recognized in 162 elements. In total, 73 LTR-RTs were complete elements (22 *Copia* and 51 *Gypsy*).

Four plant evolutionary *Copia* lineages (*Angela*, *Ale*, *Tork* and *Sire*) and five *Gypsy* lineages (*Del*, *Athila*, *Reina*, *CRM* and *Galadriel*) were identified (Figs. 2 and 3; Supplementary Figs. 2 and 3). Element lengths varied within the superfamilies. For *Copia*, RLC_*peAngela* ranged from 5.5 to 14.3 kb, with an average of ~ 9.7 kb. RLC_*peTork* exhibited elements with median lengths from 2.2 to 6.2 kb, with an average of ~ 4.2 kb. For *Gypsy*, the larger lineage (RLG_*peDel*) exhibited elements ranging in length from 0.6 to 13.3 kb, with an average of ~ 8.3 kb. RLG_*peAthila* length ranged from 1.0 to 13.7 kb, with an average of ~ 7.7 kb (Table 2, Supplementary Table 4).

The *Angela* lineage was the most abundant in the *Copia* superfamily and the *Del* lineage in *Gypsy*. *Angela* corresponded to 1.9% of the total data and *Del* to 8.5%. RLC_*peAngela* was the only lineage with a higher proportion of complete elements than incomplete ones (Supplementary Fig. 5). In the *P. edulis* draft genome, the same lineages were the most representative, *Angela* corresponding to 0.9% (~ 1.5 Mb) and *Del* to 9.3%, comprising ~ 15.4 Mb.

Estimation of LTR-RT insertion time

We estimated the insertion time of all 73 LTR-RTs with intact and complete copies (Fig. 4). The majority (74.3%, 29/39) of the RLG_*peDel* copies were recently inserted into the *P. edulis* genome (< 1.0 Mya), but insertion of three other copies was dated at > 2.0 Mya. RLG_*peReina,* RLG_*peGaladriel* and RLG_*peCRM*, with 3, 1, and 2 copies respectively, had a similar insertion time pattern (< 1.0 Mya); interestingly, one copy of RLG_*peReina* had an estimated age of 0.0 Mya, while RLG_*peAthila* copies (6) were inserted less than 1.5 Mya.

All RLC_*peAngela* copies were inserted into the genome < 2.0 Mya, and one of them was found to be dated at 0.0 Mya. One of the two copies of the RLC_*peTork* lineage was inserted into the genome up to 0.5 Mya, and the other between 1.5 and 2.0 Mya. The RLC_*peSire* copy had an estimated age of between 1.5 and 2.0 million years.

Overall, the results presented in Fig. 4 indicate that 40% of *Copia* LTR-RTs were inserted up to 1.0 Mya, and 60% between 1 and 2.0 Mya; for *Gypsy* 70.6% were inserted up to 1.0 Mya and 23.5% between 1 and 2.0 Mya.

LTR-RT in silico transcriptional activity and Reverse Transcriptase PCR analysis

We were able to associate 2821 transcripts derived from the RNA-seq libraries with full-length LTR-RTs in all *Copia*

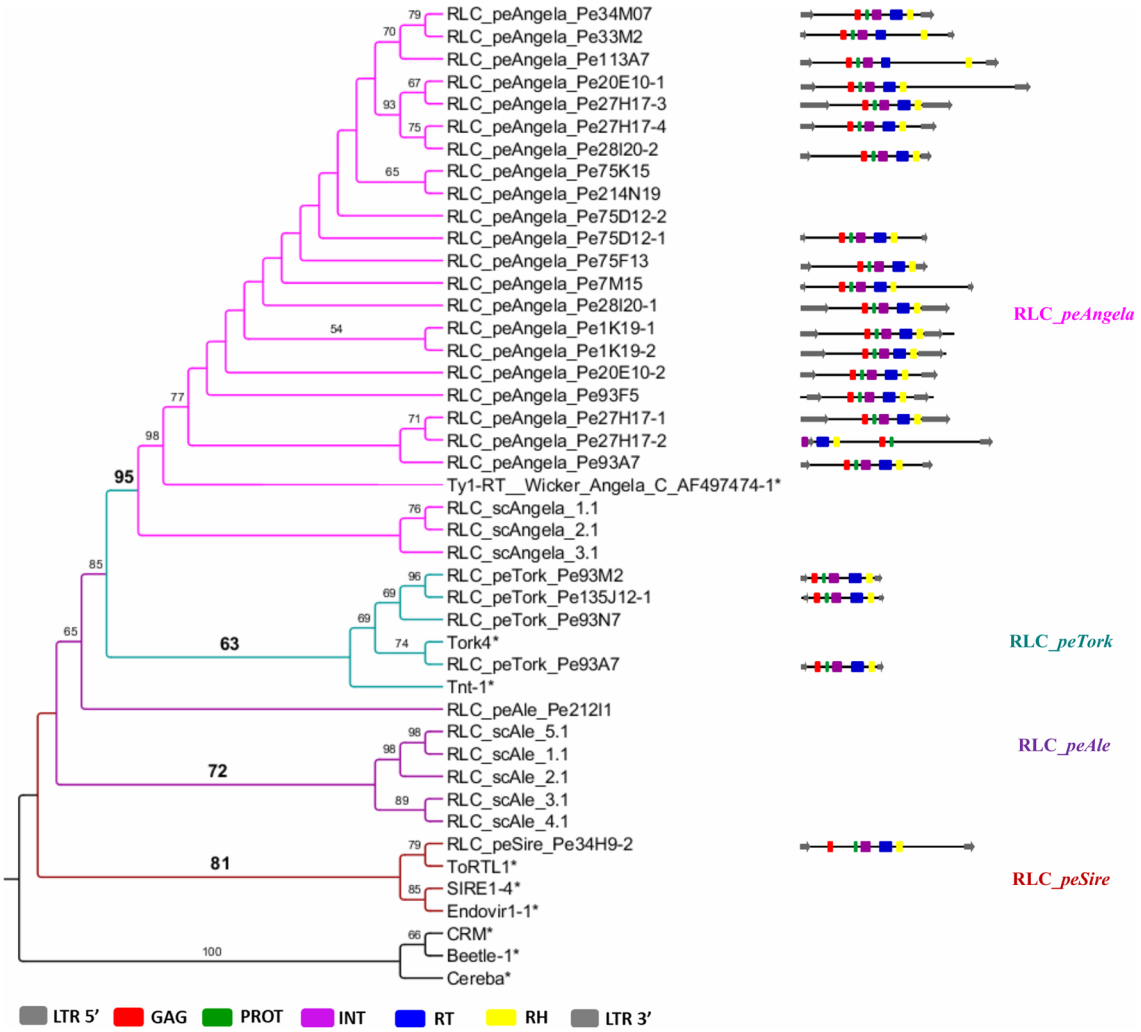


Fig. 2 Phylogenetic tree of *Copia* lineages inferred from the complete amino acid sequence of the Reverse Transcriptase domain. Maximum Likelihood analysis was based on the JTT + F model. The bootstrap values for each node are indicated in bold type. Sequences from the Gypsy database are indicated by an asterisk. *Angela* and *Ale* are sugarcane-characterized lineages. Names and colored lines indicate lineages. The *CRM* lineage from the *Gypsy* superfamily was used as an outgroup to produce a rooted tree. A schematic representation of full-length elements is shown on the right. Abbreviations and color-coding of domains: *LTR* long terminal repeat (gray), *GAG* gag (red), *PROT* protease (green), *INT* integrase (purple), *RT* reverse transcriptase (blue), *RH* ribonuclease H (yellow). (Color figure online)

and *Gypsy* lineages. The number of associated transcripts was correlated with the number of representative elements from each lineage, i.e. the higher the number of elements representative of the lineage, the higher the number of transcripts associated. Hence, there is no evidence to show that lineages have different expression levels. The largest number of transcripts was associated with the RLG_*peDel* lineage, probably because this lineage has more elements (Fig. 5a). In addition, the highest number of transcripts was identified in the RNA-seq library prepared from vegetative adult plants (Fig. 5b).

An RT-PCR analysis was performed to confirm the transcriptional activity of selected LTR-RTs. The youngest LTR-RT from each lineage was selected based on estimated insertion time, which led to the following selection: RLC_*peAngela*_Pe93F5 (0.0 Mya), RLC_*peTork*_Pe93M2 (0.07 Mya), RLG_*peAthila*_Pe93M4-1 (0.5 Mya), RLG_*peCRM*_Pe1M17 (0.53 Mya), RLG_*peDel*_Pe99P16-2 (0.03 Mya), RLG_*peGaladriel*_Pe164A12 (0.4 Mya), RLG_*peReina*_Pe212I1 (0.0 Mya). RT-PCR analysis confirmed the transcriptional activity of all elements, except RLG_*peAthila*_Pe93M4-1 (Fig. 5c).

LTR-RTs from wild *Passiflora* species

The presence of the selected *P. edulis* TEs was tested in the wild *Passiflora* species (Fig. 6).

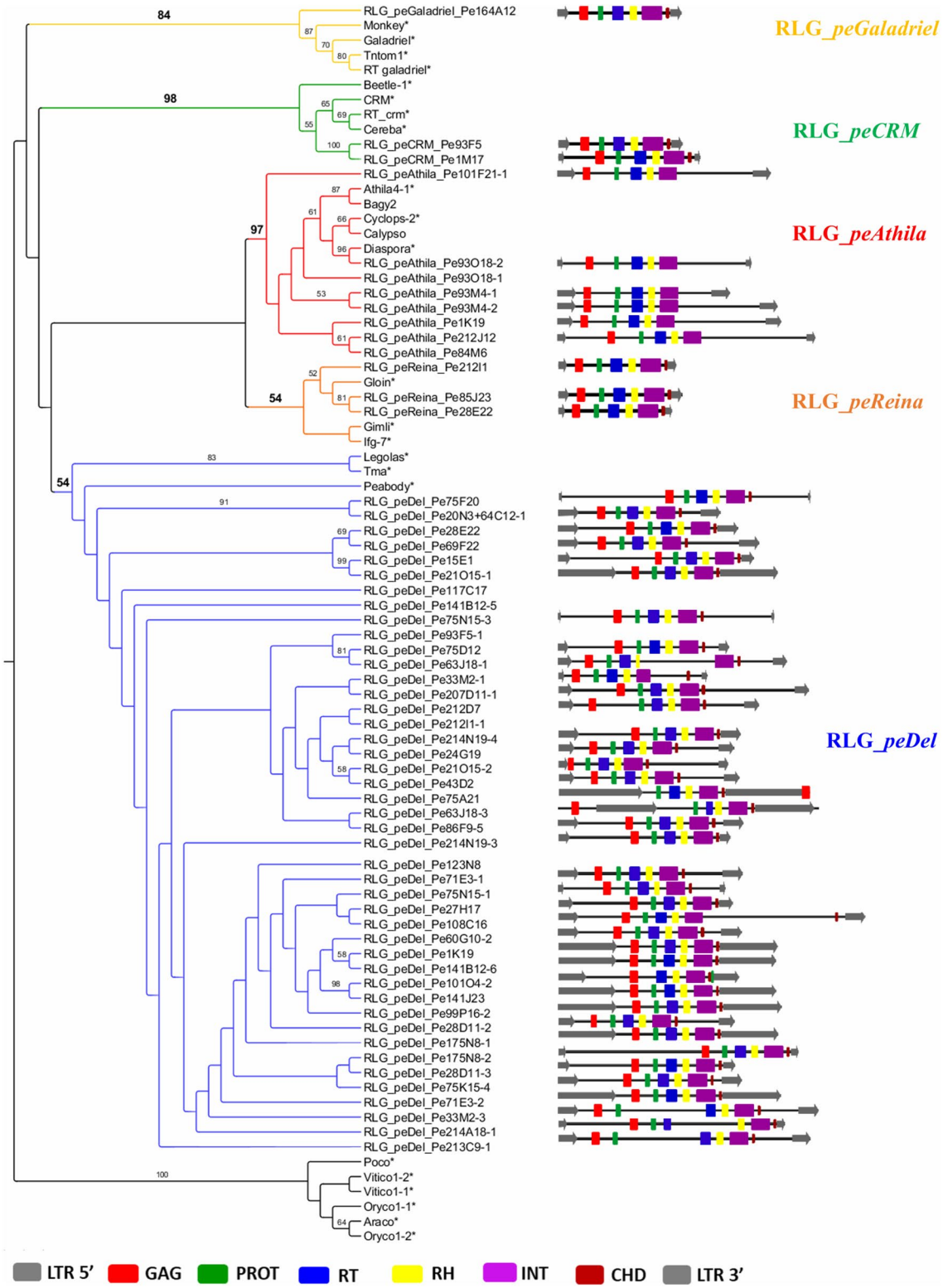


Fig. 3 Phylogenetic tree of *Gypsy* lineages inferred from the complete amino acid sequence of the Reverse Transcriptase domain. Maximum Likelihood analysis was based on the WAG + F model. The bootstrap values for each lineage node are indicated in bold type. Sequences from the Gypsy database are indicated by an asterisk. Names and colored lines indicate lineages. The *Oryco* lineage from the *Copia* superfamily was used as the outgroup to produce a rooted tree. A schematic representation of full-length elements is shown on the right. Abbreviations and color-coding of domains: *LTR* long terminal repeat (gray), *GAG* gag (red), *PROT* protease (green), *RT* reverse transcriptase (blue), *RH* ribonuclease H (yellow), *INT* integrase (purple), *CHD* chromodomain (carmine). (Color figure online)

Table 2 General features of *Passiflora edulis* LTR-RT lineages in *Copia* and *Gypsy* superfamilies identified in two data sets of *Passiflora edulis*

Superfamily/lineage	Gene-rich fraction[a]					Draft genome[b]		
	Element length range (bp)	LTR length range (bp)	Number of elements	Total length of lineage elements (bp)	Percentage related to total data	Fractional element occurrences in the genome	Total length of lineage elements (bp)	Percentage related to total data
Copia								
RLC_*peAngela*	5688–14,300	316–1831	21	204,129	1.96	11,029	1,543,735	0.93
RLC_*peAle*	4640	285–288	1	4640	0.04	563	91,652	0.06
RLC_*peTork*	2267–6289	147–987	10	55,198	0.53	6480	1,372,577	0.83
RLC_*peSire*	2108–13,073	102–1270	5	60,715	0.58	7436	1,196,468	0.72
Total *Copia*			37	324,682	3.11	25,508	4,204,432	2.54
Gypsy								
RLG_*peDel*	669–15,362	103–4096	104	886,522	8.52	84,754	15,465,373	9.34
RLG_*peAthila*	1033–13,706	290–1401	17	139,694	1.34	11,443	2,204,103	1.33
RLG_*peReina*	5224–5684	293–475	3	16,334	0.16	2063	687,815	0.42
RLG_*peCRM*	6,529–6888	255–460	2	13,417	0.13	276	32,989	0.02
RLG_*peGaladriel*	1624–5708	513	2	7332	0.07	682	132,002	0.08
Total *Gypsy*			128	1,063,299	10.22	99,218	18,522,282	11.19
Total			165	1,387,981	13.33	124,726	22,726,714	13.73

[a]TEs identified in a gene-rich fraction of the *Passiflora edulis* genome (Munhoz et al. [32])

[b]TEs identified in Munhoz et al. [32] and mapped against the draft *P. edulis* genome released (Araya et al. [25])

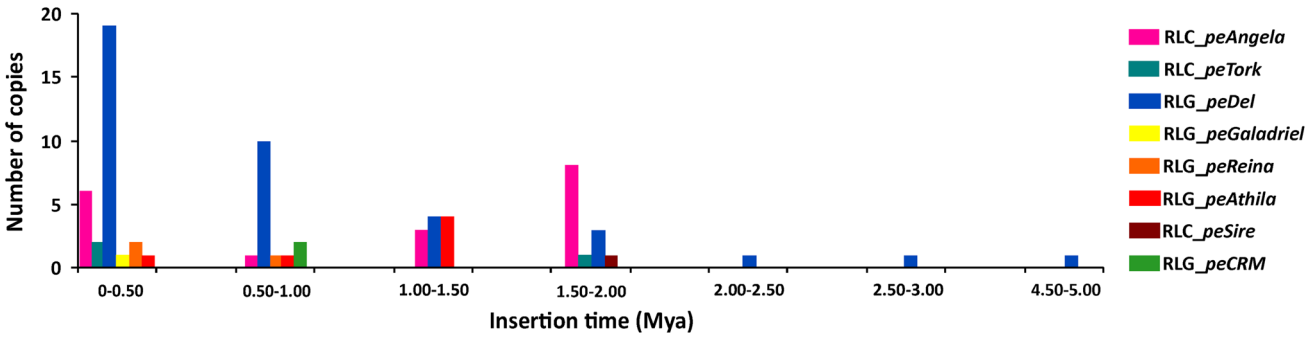


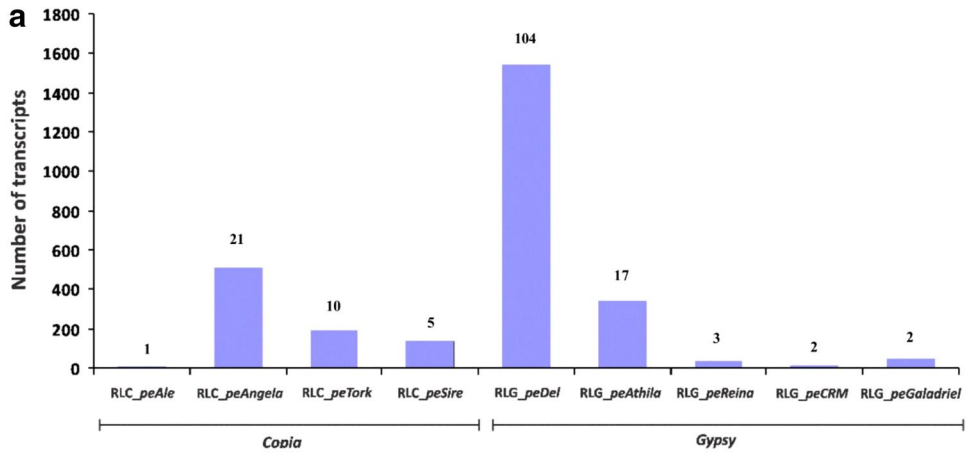
Fig. 4 Estimated insertion times of 73 full-length LTR-RT lineages of *Passiflora edulis*

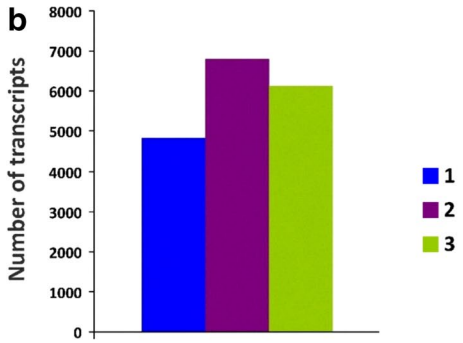
RLC_*peAngela*_Pe93F5, RLC_*peTork*_Pe93M2, RLG_*peCRM*_Pe1M17 and RLG_*peDel*_Pe99P16-2 were identified in all species tested. RLG_*peAthila*_Pe93M4-1 was identified in *P. alata* and *P. setacea*, and RLG_*peGaladriel*_Pe164A12 was identified only in *P. setacea*. RLG_*peReina*_Pe212I1 was not present in any of the wild species.

The element from the *CRM* lineage showed a weak PCR band, probably because this lineage is less represented in these species compared to the others.

Discussion

This study reveals a conserved *Passiflora* transposable element collection comprising 250 TEs fully characterized in *P. edulis*, with most of Class I and Class II lineages represented; this is the first set of complete sequences of TEs for the species. Comparing the two sets of genomic data, a gene-rich fraction and a fragmented draft genome (respectively 17.6% and 18.9% comprising TEs), both were found





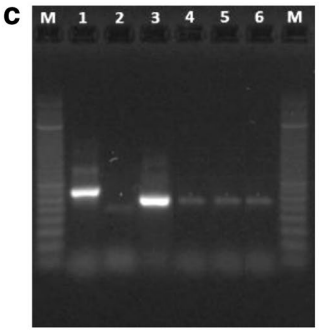


Fig. 5 **a** Number of transcripts assigned to LTR-RT lineages of the *Copia* and *Gypsy* superfamilies. The number of representative elements is indicated at the top of each bar; **b** number of normalized transcripts assigned to LTR-RT lineages. The transcripts were obtained from three *Passiflora edulis* RNA-seq libraries of shoot apexes of juvenile (1, blue), vegetative (2, purple) and reproductive adult plants (3, green) (Dornelas M.C., unpublished data); **c** Agarose (2%) gel electrophoresis of RT-PCR products from the selected transposable elements: RLC_*peAngela*_Pe93F5 (1), RLC_*peTork*_Pe93M2 (2), RLG_*peDel*_Pe99P16-2 (3), RLG_*peCRM*_Pe1M17 (4), RLG_*peGaladriel*_Pe164A12 (5) and RLG_*peReina*_Pe212I1 (6), using cDNA templates of *P. edulis*. M, 100 bp ladder (Invitrogen). (Color figure online)

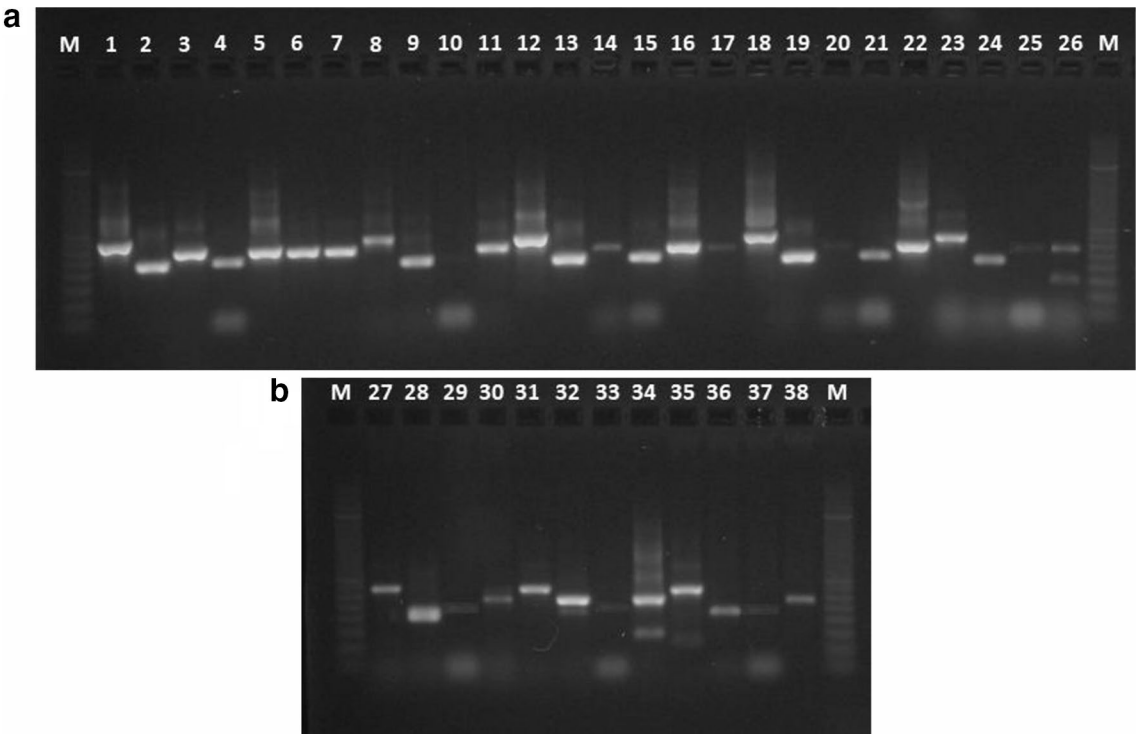


Fig. 6 Agarose (2%) gel electrophoresis of PCR products from the following elements: RLC_*peAngela*_Pe93F5 (1), RLC_*peTork*_Pe93M2 (2), RLG_*peAthila*_Pe93M4-1 (3), RLG_*peCRM*_Pe1M17 (4), RLG_*peDel*_Pe99P16-2 (5), RLG_*peGaladriel*_Pe164A12 (6) and RLG_*peReina*_Pe212I1 (7). M: 100 bp ladder (Invitrogen). **a** Lines 1–7, *P. edulis* (TEs 1–7); 8–11, *P. edmundoi* (TEs 1, 2, 4 and 5); 12–17, *P. setacea* (TEs 1, 2, 3, 4, 5 and 6); 18–22, *P. alata* (TEs 1, 2, 3, 4 and 5); 23–26, *P. organensis* (TEs 1, 2, 4 and 5); **b** Lines 27–30, *P. deidamioides* (TEs 1, 2, 4 and 5); 31–34, *P. contracta* (TEs 1, 2, 4 and 5); 35–38, *P. rhamnifolia* (TEs 1, 2, 4 and 5)

to have very similar proportions of mobile elements. This result was expected, since the search for TEs in the draft genome was based on the TEs recognized in the gene-rich fraction.

Passiflora edulis is an understudied species, with few genomic resources available. Therefore, it should be borne in mind that the TE content found herein may be underestimated as it reflects the approach adopted to pick up the elements, and consequently the data source we analyzed, though of high quality. In addition, the fact that we studied a gene-rich region may have introduced a bias towards the type and age of TEs identified. Nonetheless, we do show a set of manually curated TEs that could be used in forthcoming comparative studies.

Genome size and structure is largely determined by TEs and their abundance in plant genomes is highly variable, sometimes even in related species [66]. This variation is a consequence of species' genome evolution, which affects genome size [68, 69]; species-specific factors that influence TE amplification and repression [70]; and polyploidization events [7, 21].

In completely sequenced genomes of Malpighiales, genome size and TE content are extremely variable. For instance, the genome of *Hevea brasiliensis* is estimated to be 1470 Mb in size and 71% of it is made up of mobile genetic elements [71]. These figures are 742 Mb and 24.4% in *Manihot esculenta* [72]; 485 Mb and 42.0% in *Populus trichocarpa* [73]; 416 Mb and 45.9% in *Jatropha curcas* [74]; 373 Mb and ~ 23% in *Linum usitatissimun* [59] and 320 Mb and 50.0% in *Ricinus communis* [75].

The genome size of *P. edulis* is estimated at ~ 1230 Mb [22] and we believe that the TE content is high in this genome, which might possibly explain its reasonably large size. Very recently, Pamponét et al. [76], using very low-coverage (2.2×) sequencing data for repeat identification, found that the *P. edulis* genome possibly consists of at least 59% repetitive DNA, 53% of this proportion consisting of LTR-RTs. This is a preliminary analysis and corroborates our data with regard the proportion of LTR-RTs. However only when the whole genome is available will we be able to confirm the content and types of TEs in *P. edulis.*

Retrotransposons are predominant in flowering plants, the majority belonging to the LTR order [66]. In *P. edulis*, retrotransposons are also prominent, most consisting of LTR-RTs, which are also predominant in other Malpighiales (74.6% in *Linum*, 74.2% in *Jatropha*, 71.7% in *Hevea*, 60% in *Populus* and 45.6% in *Manihot*).

Retrotransposons play key roles in gene expansion, function, diversification and evolution. In *P. edulis*, we identified LARDs that captured genes and others that have been inserted into intronic sequences. There are a few reports of the capture of plant genes by retrotransposons [2, 54]. In rice, 672 captured genes have been identified, ~ 33% of them being expressed, implying that a percentage of captured genes could have evolved into pseudogenes. In sorghum, however, a total of 1343 captured genes have been identified, ~ 70% being expressed, indicating that most of the captured genes might be still functional [77].

Only 2.6% of *P. edulis* elements were classified as DNA transposons, which are in fact less abundant in plant genomes, like in the sequenced Malpighiales, in which the proportion of DNA transposons is 2.7% in *M. esculenta* [72], 1.8% in *R. communis* [75] and 15.7% in *L. usitatissimun* [59]. However, there are little-known exceptions, such as *Arabidopsis thaliana*, whose genome is composed of 59.4% of DNA transposons (see [66]). This is probably a consequence of its compact genome size, with little space for insertions of long elements and possibly strong selections against them [20]. In sequenced Malpighiales, the proportion of DNA transposons is 2.7% in *M. esculenta* [72], 1.8% in *R. communis* [75] and 15.7% in *L. usitatissimun* [59].

There are some reports on the acquisition of genes or gene fragments by DNA transposons that tend to interact with host genes [78–80]. We report the occurrence of gene fragments in Helitrons, which could be non-functional pseudogenes or pseudogenes with some degree of function, as already documented [81]. The capture of genes or gene fragments by Helitrons was reported in rice [82] and *Arabidopsis* [83]. In maize, Helitrons have had significant impact on genome structure and function, including the synthesis of new genes and modification of expression patterns [84, 85].

In general, TEs are frequently recognized as genomic fossils that were once autonomous, but at some point in time experienced a mutation that rendered them inactive [86]. The majority of *P. edulis* TEs (70.8%) were incomplete, corroborating previous findings showing that most TE copies are either defective or fossilized [87]. However, a non-autonomous TE can remain active using the enzymatic machinery required for transposition provided by an autonomous partner [88, 89].

TEs are preferentially located in intergenic spaces [7, 86], or in gene neighborhoods or even in regulatory regions [90], which is consistent with our findings. For instance, in *Linum* the location of TEs was not completely random and some regions had equal coverage of genes and TEs, with a predominance of certain superfamilies [59]. TEs close to genes can become positive regulators of gene expression, but they can also be responsible for up- and downregulation of gene expression, causing beneficial or deleterious effects depending on the process involved. TEs affect adjacent gene regions by becoming targets for epigenetic silencing [21, 91].

This led us to consider the activity of LTR-RTs in this gene-rich region of the *P. edulis* genome and we therefore investigated them in detail.

At superfamily level, 81.8% of the LTR-RTs belonged to *Gypsy* and 18.2% to *Copia*. The *Gypsy* superfamily is

the most representative in the majority of plant genomes adequately characterized (reviewed in [66]). Based only or our data, one cannot assume that the presence of *Gypsy* in the examined gene-rich fraction implies prevalence in the whole *P. edulis* genome. In related species of *P. edulis*, *Gypsy* accounted for 85% in *Hevea*, 70.6% in *Ricinus*, 54.5% in *Jatropha* and 52% in *Populus*. On the other hand, the majority of LTR-RTs are *Copia* in *Linum* (54.1%).

Some reports have shown that *Gypsy* elements are associated with heterochromatic regions and *Copia* with euchromatic regions [44, 59, 92]. In fact, pericentromeric regions are, in general, *Gypsy*-enriched in plants, although this could vary among families (see [93]). Evidence to date suggests that there is a negative association between TEs and high recombination rates along the chromosomes. Thus, most TEs tend to accumulate in heterochromatic, non-recombining regions of the genome, such as centromeres and, more commonly, in pericentromeric regions [20].

Four *Copia* and five *Gypsy* evolutionary lineages were identified. The RLC_*peAngela* lineage was the most representative within *Copia*. The *Angela* lineage has been described in other plant species. In fact, *Angela* was significantly predominant in the genome of *Setaria italica* (Poaceae), corresponding to 28.2% of all complete LTR-RTs [19, 44]. The RLC_*peAngela* lineage is well-conserved in the genome fraction analyzed herein, since most of the *P. edulis* elements were complete (78.26%, 18/23) and had similar structure in terms of total length and internal domain length and organization.

The RLG_*peDel* lineage dominated the population of TEs in the gene-rich fraction, representing 48.4% of all TEs. Interestingly, in the fragmented draft genome, the *Del* lineage was also the most representative, accounting for 49.3% of all TEs. We observed that the RLG_*peDel* lineage is not well-conserved in terms of the number of complete elements (37.5%, 39/104), total length and internal domain organization.

Although less frequent, *Copia* elements seems to be more tolerated in this gene-rich region, because they have a more conserved structure. Fragmented elements, most of them *Gypsy*, could have resulted from old insertions that underwent strong purifying selection. Insertions of TEs can have drastic effects on genic regions by disrupting coding sequence and the regulatory environment. Given the strong mutagenic potential, molecular mechanisms evolved to suppress TE activity [5]. In addition, natural selection is a powerful force that affects the distribution and accumulation of TEs. Hence, insertions that are strongly deleterious are rapidly removed while insertions that have little or no negative effects on host fitness are free to drift in frequency, occasionally to fixation [94].

The success and diversity of TEs in a genome depend on the properties of the elements and the evolutionary forces acting at the level of the host species [95]. TEs we have seen in this gene region of *P. edulis* resulted from the combined processes of new transposition and removal by selection. TEs that are conserved and still active are being controlled by silencing mechanisms or have not yet been removed by selection. TEs with no conserved structure are possibly older inactive elements, with silencing maintained to repress transposition and ectopic recombination events, which have deleterious effects on host fitness [20].

Regarding the representation of the families, an ascertainment bias was undoubtedly introduced to the effect that identified elements were those that tend to inhabit gene-rich regions. Moreover, we were unable to find elements in the gene-poor regions of the draft genome and those well-known as pericentromeric specialists that were not sampled. The general pattern very recently found by Pamponét et al. [76] seems consistent with our findings, since the families *Ty3/Gypsy/Athila*, *Ty1/Copia/Angela* and *Ty1/Copia/Maximus-Sire* have shown similar genome proportions and were uniformly distributed along the FISH labeled *P. edulis* chromosomes. On the other hand, the *Ty3/Gypsy/Chromovirus* was represented by two families, which were the most frequent, one uniformly distributed, but the other pericentromeric-specific.

The prevalence of particular families is highly variable among plant species [7, 19]. It is well known that the expansion of some plant genomes is due to the accumulation of few types of elements, showing a clear correlation between genome size and the activity of these highly repeated families [7, 18]. One example is the expansion of the *Capsicum annuum* genome through a massive accumulation of a single-type *Gypsy* element that belongs to the *Del* subgroup [10].

Passiflora edulis has one of the largest genomes compared to other *Passiflora* species [22]. We therefore believe that *P. edulis* has undergone a whole-genome duplication event [32], that may be accompanied by TE proliferation, specifically of *Gypsy* elements, along evolutionary time. Corroborating this, our WGD analysis, based on the distribution of synonymous substitutions per synonymous site for pairwise paralogs (dS), showed that *P. edulis* has undergone at least two WGD events, since the first peak is possibly a noise signal caused by unmerged allelic haplotypes in the genome data or alternative splice transcripts within the transcriptome data [53] (Supplementary Fig. 5). These WGD events may have contributed to increased TE proliferation, especially in the pericentromeric regions that were not analyzed, explaining the increase in genome size, which is composed of at least 53% of LTR-RTs, as very recently reported [76]. In addition, this may have contributed in part to the TE patterns in the genic regions.

In fact, it is hypothesized that WGD is followed by a reduction in purifying selection, caused by genome

redundancy, which may have led to an increase in TE insertions [96]. Furthermore, due to genome stress after polyploidization, silencing mechanisms that control TE activity may have been temporarily disrupted. This combination of relaxed selection and a breakdown of silencing mechanisms could potentially drive the dramatic spread of TEs through the genome [97, 98].

Another important aspect is the possible association with mating system, as some population genetic models predict that highly outcrossing species, such as the passionflowers, including *P. edulis*, are expected to experience higher rates of transposable element activity and copy number (see [98]). *Gypsy* was predominant in the gene-rich fraction, and represented a large proportion of the draft genome. The data recently released by [76, 99] corroborate our hypothesis, implying that this element is highly frequent and could have significantly influenced genome size.

We found that full-length LTR-RT elements became active quite recently. Most elements (95.9%, 70/73) were inserted into the genome < 2.0 Mya and this seems to be the result of a relaxed selection. These recently expanded TEs are kept silenced and will likely be removed by ectopic recombination, in order to suppress deleterious effects on genes.

An analysis of in silico transcriptional activity showed that all *Copia* and *Gypsy* lineages were associated with transcripts, which is an indication of LTR-RT activity. Young insertions are expected to be active, as selection has not yet removed them. Interestingly, some older elements (4 Mya) still seem to be active, but with fewer transcripts associated (data not shown). This could represent a degree of tolerance to insertions in this region. However, more data is needed to make considerable assumptions.

Some LTR-RT lineages in *L. usitatissimun* and *P. trichocarpa* were also active [59, 92]. In addition, some reports have highlighted TE' expression in different tissues over different time periods, including sugarcane [44, 100, 101], *Eucalyptus* [61] and oil palm (*Elaeis guineensis*) [102].

Reverse Transcriptase PCR analysis confirmed the transcriptional activity of the youngest element of each LTR-RT lineage, except for RLG_*peAthila*_Pe93M4-1. RLC_*peAngela*_Pe93F5 and RLG_*peDel*_Pe99P16-2 exhibited stronger PCR bands. These elements seem to be more transcribed than the others, which is plausible since these elements are respectively among the most representative elements of the *Copia* and *Gypsy* superfamilies in both data sets analyzed. Amplifications of RLC_*peTork*_Pe93M2, RLG_*peCRM*_Pe1M17, RLG_*peGaladriel*_Pe164A12 and RLG_*peReina*_Pe212I1 exhibited weak PCR bands. These lineages were poorly represented, possibly because they are less represented in the whole genome.

Regarding all lineages examined, RLC_*peAngela*, RLC_*peTork*, RLG_*peCRM* and RLG_*peDel* seem to have a more conserved structure, since they were found in all four subgenera. This could be the result of their shared evolutionary history from the common ancestor, indicating that these lineages may have undergone amplifications before the *Passiflora* evolutionary diversification into subgenera.

RLG_*peAthila* and RLG_*peGaladriel* lineages seem to be conserved throughout the subgenus *Passiflora* since they were only found in *P. alata* and *P. setacea*. This is plausible since these species together with *P. edulis* belong to the same subgenus [103]. These lineages could have undergone relatively recent proliferation in this subgenus. RLG_*peReina* was the only lineage not found in any of the wild species examined, and is possibly a specific lineage of *P. edulis*, due to its recent proliferation in *P. edulis* or decay in the other. However, this pattern could be a result of the sample examined. We used the RT domain sequence from young elements to test for the presence of these lineages in wild *Passiflora* species. Hence, we could have missed old elements that may not have the RT domain or that are highly mutagenized in the RT sequence.

In addition, our data suggest that *P. edulis* genome underwent at least two WGD events. These events may have occurred in a common ancestor of *Passiflora*, which emerged ~ 40 Mya [104]. Hence, TEs could have expanded in a shared WGD event, explaining the conservation of some LTR-RT families in wild species, and differential retention of other families due to genomic fractionation following WGD events.

Conclusions

This is the first report on the annotation and detailed characterization of transposable elements (TEs) in the tropical fruit species *Passiflora edulis* (Passifloraceae, Malpighiales), including structural features, protein domain organization and proliferation in wild species. Most of the TEs are located in intergenic spaces, although some TEs overlap genes. LTR retroelements predominate, consisting mainly of *Gypsy* elements, with over-representation of the RLG_*peDel* lineage. *P. edulis* LTR-RTs are reported as active over the last 2 million years. We were also able to confirm the transcriptional activity of full-length LTR-RTs by means of transcript association and reverse transcriptase PCR analysis. The recent activity and abundance of LTR-RTs in a gene-rich portion of *P. edulis* indicate that WGD events could have led to a recent expansion of these lineages, which will probably be selectively removed due to its deleterious potential. Interestingly, some lineages seem to be conserved in wild species of *Passiflora* and further detailed characterization of the repeat portion will contribute to understanding their influence on

the evolution of this genus. Our study provides references of single elements in the *P. edulis* genome, and will be helpful in understanding the transposon landscape in further evolutionary studies on *Passiflora*.

Acknowledgements Mr. Steve Simmons for proofreading the manuscript.

Author contributions Conceptualization, MLCV; Data curation, ZPC and AMV; Formal analysis, ZPC, LACS, GTR and AMV; Funding acquisition, MLCV; Investigation, ZPC; Methodology, ZPC; Resources, MCD; Supervision, MAVS, HB and MLCV; Writing – original draft, ZPC and MLCV; Writing – review & editing, MAVS and AMV.

Funding This research was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant Number 2014/25215-2, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, scholarship awarded to ZPC) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, scholarships awarded to ZPC and GTR).

Compliance with ethical standards

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

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Ravindran S (2012) Barbara McClintock and the discovery of jumping genes. Proc Natl Acad Sci 109:20198–20199. https://doi.org/10.1073/pnas.1219372109
2. Lisch D (2012) How important are transposons for plant evolution? Nat Rev Genet 14:49–61. https://doi.org/10.1038/nrg3374
3. Wicker T, Sabot FF, Hua-Van AA et al (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982. https://doi.org/10.1038/nrg2165
4. Garcia-Perez MM-L, García-Pérez JL (2010) DNA Transposons: nature and applications in genomics. Curr Genom 11:115–128. https://doi.org/10.2174/138920210790886871
5. Schrader L, Schmitz J (2019) The impact of transposable elements in adaptive evolution. Mol Ecol 28:1537–1549. https://doi.org/10.1111/mec.14794
6. Klein SJ, O'Neill RJ (2018) Transposable elements: genome innovation, chromosome diversity, and centromere conflict. Chromosome Res 26(1–2):5–23. https://doi.org/10.1007/s10577-017-9569-5
7. Vicient CM, Casacuberta JM (2017) Impact of transposable elements on polyploid plant genomes. Ann Bot 120:195–207. https://doi.org/10.1093/aob/mcx078
8. The Tomato Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641. https://doi.org/10.1038/nature11119
9. Wegrzyn JL, Liechty JD, Stevens KA et al (2014) Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. Genetics 196:891–909. https://doi.org/10.1534/genetics.113.159996
10. Park M, Park J, Kim S et al (2012) Evolution of the large genome in *Capsicum annuum* occurred through accumulation of single-type long terminal repeat retrotransposons and their derivatives. Plant J 69:1018–1029. https://doi.org/10.1111/j.1365-313X.2011.04851.x
11. Hernandez-Pinzon I, de Jesus E, Santiago N, Casacuberta JM (2009) The frequent transcriptional readthrough of the tobacco Tnt1 retrotransposon and its possible implications for the control of resistance genes. J Mol Evol 68:269–278. https://doi.org/10.1007/s00239-009-9204-y
12. Neumann P, Požárková D, Koblížková A, Macas J (2005) PIGY, a new plant envelope-class LTR retrotransposon. Mol Genet Genom 273:43–53. https://doi.org/10.1007/s00438-004-1092-7
13. Novikova A, Smyshlyaev G, Novikova O (2012) Evolutionary history of LTR retrotransposon chromodomains in plants. Int J Plant Genomics 2012:874743. https://doi.org/10.1155/2012/874743
14. Novikova O (2009) Chromodomains and LTR retrotransposons in plants. Commun Integr Biol 2:158–162. https://doi.org/10.4161/cib.7702
15. Llorens C, Futami R, Covelli L et al (2011) The Gypsy database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res 39:D70–D74. https://doi.org/10.1093/nar/gkq1061
16. Llorens C, Muñoz-Pomer A, Bernad L et al (2009) Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. Biol Direct 4:41. https://doi.org/10.1186/1745-6150-4-41
17. Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol 65:505–530. https://doi.org/10.1146/annurev-arplant-050213-035811
18. El Baidouri M, Panaud O (2013) Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. Genome Biol Evol 5(5):954–965. https://doi.org/10.1093/gbe/evt025
19. Ochoa Cruz EA, Cruz GMQ, Vieira AP, Van Sluys M-A (2016) Virus-like attachment sites as structural landmarks of plants retrotransposons. Mob DNA 7:14. https://doi.org/10.1186/s13100-016-0069-5
20. Kent TV, Uzunovic J, Wright SI (2017) Coevolution between transposable elements and recombination. Philos Trans R Soc B Biol Sci 372:20160. https://doi.org/10.1098/rstb.2016.0458
21. Galindo-González L, Mhiri C, Deyholos MK, Grandbastien M-A (2017) LTR-retrotransposons in plants: engines of evolution. Gene 626:14–25. https://doi.org/10.1016/j.gene.2017.04.051
22. Yotoko KSC, Dornelas MC, Togni PD et al (2011) Does variation in genome sizes reflect adaptive or neutral processes? New clues from *Passiflora*. PLoS ONE 6:e18212. https://doi.org/10.1371/journal.pone.0018212
23. Ulmer T, MacDougal JM (2004) *Passiflora*: passionflowers of the world. Timber Press, Portland, p 430
24. Munhoz CF, Santos AA, Arenhart RA et al (2015) Analysis of plant gene expression during passion fruit-*Xanthomonas axonopodis* interaction implicates lipoxygenase 2 in host defence. Ann Appl Biol 167:135–155. https://doi.org/10.1111/aab.12215
25. Araya S, Martins AM, Junqueira NTV et al (2017) Microsatellite marker development by partial sequencing of the sour passion fruit genome (*Passiflora edulis* Sims). BMC Genomics 18:549. https://doi.org/10.1186/s12864-017-3881-5
26. Carneiro MS, Camargo LEA, Coelho ASG et al (2002) RAPD-based genetic linkage maps of yellow passion fruit (*Passiflora edulis* Sims. f. flavicarpa Deg.). Genome 45:670–678. https://doi.org/10.1139/g02-035
27. Lopes R, Lopes MTG, Carneiro MS et al (2006) Linkage and mapping of resistance genes to *Xanthomonas axonopodis* pv. passiflorae in yellow passion fruit. Genome 49:17–29. https://doi.org/10.1139/G05-081

28. Oliveira EJ, Vieira MLC, Garcia AAF et al (2008) An integrated molecular map of yellow passion fruit based on simultaneous maximum-likelihood estimation of linkage and linkage phases. J Am Soc Hortic Sci 133:35–41
29. Costa ZP, de Munhoz C, Vieira MLC (2017) Report on the development of putative functional SSR and SNP markers in passion fruits. BMC Res Notes 10:445. https://doi.org/10.1186/s13104-017-2771-x
30. Moraes MC, Geraldi IO, De Pina Matta F et al (2005) Genetic and phenotypic parameter estimates for yield and fruit quality traits from a single wide cross in yellow passion fruit. HortScience 40:1978–1981
31. Santos A, Penha H, Bellec A et al (2014) Begin at the beginning: a BAC-end view of the passion fruit (*Passiflora*) genome. BMC Genom 15:816. https://doi.org/10.1186/1471-2164-15-816
32. Munhoz CF, Costa ZP, Cauz-Santos LA et al (2018) A gene-rich fraction analysis of the *Passiflora edulis* genome reveals highly conserved microsyntenic regions with two related Malpighiales species. Sci Rep 8:13024. https://doi.org/10.1038/s41598-018-31330-8
33. Wu Y, Tian Q, Liu J et al (2019) The genomic sequence and comparative genomic analysis of cultivated passion fruit (*Passiflora edulis* L.). bioRxiv. https://doi.org/10.1101/522128
34. Quesneville H, Bergman CM, Andrieu O et al (2005) Combined evidence annotation of transposable elements in genome sequences. PLoS Comput Biol 1:e22. https://doi.org/10.1371/journal.pcbi.0010022
35. Flutre T, Duprat E, Feuillet C, Quesneville H (2011) Considering transposable element diversification in de novo annotation approaches. PLoS ONE 6:e16526. https://doi.org/10.1371/journal.pone.0016526
36. Smit A, Hubley R, Green P (2018) RepeatMasker 4.0.8
37. Marchler-Bauer A, Bo Y, Han L et al (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45:D200–D203. https://doi.org/10.1093/nar/gkw1129
38. Liu W, Xie Y, Ma J et al (2015) IBS: an illustrator for the presentation and visualization of biological sequences. Bioinformatics 31:3359–3361. https://doi.org/10.1093/bioinformatics/btv362
39. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European molecular biology open software suite. Trends Genet 16:276–277. https://doi.org/10.1016/S0168-9525(00)02024-2
40. Abeel T, Van Parys T, Saeys Y et al (2012) GenomeView: a next-generation genome browser. Nucleic Acids Res 40:e12. https://doi.org/10.1093/nar/gkr995
41. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303:3997
42. Novák P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics 11:378. https://doi.org/10.1186/1471-2105-11-378
43. Novák P, Neumann P, Pech J et al (2013) RepeatExplorer: a galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29:792–793. https://doi.org/10.1093/bioinformatics/btt054
44. Domingues DS, Cruz GMQ, Metcalfe CJ et al (2012) Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. BMC Genom 13:137. https://doi.org/10.1186/1471-2164-13-137
45. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340
46. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054
47. Keane TM, Creevey CJ, Pentony MM et al (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol Biol 6:29. https://doi.org/10.1186/1471-2148-6-29
48. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033
49. Rambaut A (2012) FigTree v. 1.4.0. http://tree.bio.ed.ac.uk/software/figtree/
50. Huson DH, Richter DC, Rausch C et al (2007) Dendroscope: an interactive viewer for large phylogenetic trees. BMC Bioinformatics 8:460. https://doi.org/10.1186/1471-2105-8-460
51. Stöver BC, Müller KF (2010) TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics 11:7. https://doi.org/10.1186/1471-2105-11-7
52. Sayers EW, Barrett T, Benson DA et al (2011) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 39:D38–D51. https://doi.org/10.1093/nar/gkq1172
53. Yang Y, Li Y, Chen Q et al (2019) WGDdetector: a pipeline for detecting whole genome duplication events using the genome or transcriptome annotations. BMC Bioinformatics 20:75. https://doi.org/10.1186/s12859-019-2670-3
54. Zhao D, Ferguson AA, Jiang N (2016) What makes up plant genomes: the vanishing line between transposable elements and genes. Biochim Biophys Acta 1859:366–380. https://doi.org/10.1016/j.bbagrm.2015.12.005
55. SanMiguel P, Gaut BS, Tikhonov A et al (1998) The paleontology of intergene retrotransposons of maize. Nat Genet 20:43–45. https://doi.org/10.1038/1695
56. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120
57. Yin H, Du J, Wu J et al (2015) Genome-wide annotation and comparative analysis of long terminal repeat retrotransposons between pear species of *P. bretschneideri* and *P. communis*. Sci Rep 5:1–15. https://doi.org/10.1038/srep17644
58. Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Mol Biol Evol 17:1483–1498. https://doi.org/10.1093/oxfordjournals.molbev.a026248
59. Gonzalez LG, Deyholos MK (2012) Identification, characterization and distribution of transposable elements in the flax (*Linum usitatissimum* L.) genome. BMC Genomics 13:644. https://doi.org/10.1186/1471-2164-13-644
60. Moisy C, Schulman AH, Kalendar R et al (2014) The Tvv1 retrotransposon family is conserved between plant genomes separated by over 100 million years. Theor Appl Genet 127:1223–1235. https://doi.org/10.1007/s00122-014-2293-z
61. Marcon HS, Domingues DS, Silva JC et al (2015) Transcriptionally active LTR retrotransposons in *Eucalyptus* genus are differentially expressed and insertionally polymorphic. BMC Plant Biol 15:1–16. https://doi.org/10.1186/s12870-015-0550-1
62. Trapnell C, Williams BA, Pertea G et al (2010) Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511. https://doi.org/10.1038/nbt.1621
63. Afgan E, Baker D, van den Beek M et al (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res 44:W3–W10. https://doi.org/10.1093/nar/gkw343
64. Spruyt M, Buquicchio F (2004) Gene Runner for Windows. http://www.generunner.net/

65. Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325. https://doi.org/10.1093/nar/8.19.4321
66. Oliver KR, McComb JA, Greene WK (2013) Transposable elements: powerful contributors to angiosperm evolution and diversity. Genome Biol Evol 5:1886–1901. https://doi.org/10.1093/gbe/evt141
67. Narayanan KB, Park HH (2015) Toll/interleukin-1 receptor (TIR) domain-mediated cellular signaling pathways. Apoptosis 20:196–209. https://doi.org/10.1007/s10495-014-1073-1
68. Lee S-I, Kim N-S (2014) Transposable elements and genome size variations in plants. Genomics Inform 12:87. https://doi.org/10.5808/GI.2014.12.3.87
69. Civáň P, Švec M, Hauptvogel P (2011) On the coevolution of transposable elements and plant genomes. J Bot 2011:1–9. https://doi.org/10.1155/2011/893546
70. Hua-van A, Le Rouzic A, Boutin TS et al (2011) The struggle for life of the genome's selfish architects. Biol Direct 6:19. https://doi.org/10.1186/1745-6150-6-19
71. Tang C, Yang M, Fang Y et al (2016) The rubber tree genome reveals new insights into rubber production and species adaptation. Nat Plants 2:1–10. https://doi.org/10.1038/NPLANTS.2016.73
72. Wang W, Feng B, Xiao J et al (2014) Cassava genome from a wild ancestor to cultivated varieties. Nat Commun 5:5110. https://doi.org/10.1038/ncomms6110
73. Tuskan GA, DiFazio S, Jansson S et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604. https://doi.org/10.1126/science.1128691
74. Wu P, Zhou C, Cheng S et al (2015) Integrated genome sequence and linkage map of physic nut (*Jatropha curcas* L.), a biodiesel plant. Plant J 81:810–821. https://doi.org/10.1111/tpj.12761
75. Chan AP, Crabtree J, Zhao Q et al (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol 28:951–956. https://doi.org/10.1038/nbt.1674
76. Pamponét VCC, Souza MM, Silva GS et al (2019) Low coverage sequencing for repetitive DNA analysis in *Passiflora edulis* Sims: cytogenomic characterization of transposable elements and satellite DNA. BMC Genomics 20(262):1–17. https://doi.org/10.1186/s12864-019-5576-6
77. Jiang S-Y, Ramachandran S (2013) Genome-wide survey and comparative analysis of LTR retrotransposons and their captured genes in rice and sorghum. PLoS ONE 8:e71118–e71118. https://doi.org/10.1371/journal.pone.0071118
78. Gupta S, Gallavotti A, Stryker GA et al (2005) A novel class of Helitron-related transposable elements in maize contain portions of multiple pseudogenes. Plant Mol Biol 57:115–127. https://doi.org/10.1007/s11103-004-6636-z
79. Morgante M, Brunner S, Pea G et al (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet 37:997–1002. https://doi.org/10.1038/ng1615
80. Dooner HK, Weil CF (2007) Give-and-take : interactions between DNA transposons and their host plant genomes. 486–492. https://doi.org/10.1016/j.gde.2007.08.010
81. Thomas J, Pritham EJ (2015) Helitrons, the eukaryotic rolling-circle transposable elements. Mob DNA III. https://doi.org/10.1128/microbiolspec.mdna3-0049-2014
82. Fan C, Zhang Y, Yu Y et al (2008) The subtelomere of *Oryza sativa* chromosome 3 short arm as a hot bed of new gene origination in rice. Mol Plant 1:839–850. https://doi.org/10.1093/mp/ssn050
83. Gaut BS, Hollister JD (2007) Population and evolutionary dynamics of Helitron transposable elements in *Arabidopsis thaliana*. Mol Biol Evol 24:2515–2524. https://doi.org/10.1093/molbev/msm197
84. Du C, Fefelova N, Caronna J et al (2009) The polychromatic Helitron landscape of the maize genome. Proc Natl Acad Sci 106:19916–19921. https://doi.org/10.1073/pnas.0904742106
85. Barbaglia AM, Klusman KM, Higgins J et al (2012) Gene capture by Helitron transposons reshuffles the transcriptome of maize. Genetics 190:965–975. https://doi.org/10.1534/genetics.111.136176
86. Zhao M, Ma J (2013) Co-evolution of plant LTR-retrotransposons and their host genomes. Protein Cell 4:493–501. https://doi.org/10.1007/s13238-013-3037-6
87. Cruz GMQ, Metcalfe CJ, De Setta N et al (2014) Virus-like attachment sites and plastic CpG Islands: landmarks of diversity in plant Del retrotransposons. PLoS ONE 9:1–14. https://doi.org/10.1371/journal.pone.0097099
88. Piskurek O, Jackson DJ (2012) Transposable elements: from DNA parasites to architects of metazoan evolution. Genes (Basel) 3:409–422. https://doi.org/10.3390/genes3030409
89. Jiang N, Jordan IK, Wessler SR (2002) Dasheng and RIRE2: a nonautonomous long terminal repeat element and its putative autonomous partner in the rice genome. Plant Physiol 130:1697–1705. https://doi.org/10.1104/pp.015412
90. Contreras B, Vives C, Castells R, Casacuberta J (2015) The impact of transposable elements in the evolution of plant genomes. In: Pontarotti P (ed) Evol Biol biodiversification from genotype to phenotype. Springer, Cham, pp 93–105
91. Uzunovic J, Josephs EB, Stinchcombe JR, Wright SI (2019) Transposable elements are important contributors to standing variation in gene expression in *Capsella grandiflora*. Mol Biol Evol 36(8):1734–1745. https://doi.org/10.1093/molbev/msz098
92. Cossu RM, Buti M, Giordani T et al (2012) A computational study of the dynamics of LTR retrotransposons in the *Populus trichocarpa* genome. Tree Genet Genomes 8:61–75. https://doi.org/10.1007/s11295-011-0421-3
93. Stitzer MC, Anderson SN, Springer NM, Ross-Ibarra J (2019) The genomic ecosystem of transposable elements in maize. bioRxiv. https://doi.org/10.1101/559922
94. Lynch M (2007) The origins of genome architecture, 1st edn. Sinauer Associates, Sunderland
95. Bourque G, Burns KH, Gehring M et al (2018) Ten things you should know about transposable elements. Genome Biol 19:199. https://doi.org/10.1186/s13059-018-1577-z
96. Parisod C, Senerchia N (2012) Responses of transposable elements to polyploidy. In: Parisod C, Senerchia N (eds) Plant transposable elements. Springer, Berlin, pp 147–168
97. Wendel JF, Jackson SA, Meyers BC, Wing RA (2016) Evolution of plant genome architecture. Genome Biol 17:37. https://doi.org/10.1186/s13059-016-0908-1
98. Ågren JA, Huang HR, Wright SI (2016) Transposable element evolution in the allotetraploid *Capsella bursa-pastoris*. Am J Bot 103:1197–1202. https://doi.org/10.3732/ajb.1600103
99. Sader MA, Dias Y, Costa ZP, Munhoz C, Penha H, Bergés H, Vieira MLC, Pedrosa-Harand A (2019) Identification of passion fruit (*Passiflora edulis*) chromosomes using BAC-FISH. Chromosom Res. https://doi.org/10.1007/s10577-019-09614-0
100. de Setta N, Monteiro-Vitorello CB, Metcalfe CJ et al (2014) Building the sugarcane genome for biotechnology and identifying evolutionary trends. BMC Genomics 15:540. https://doi.org/10.1186/1471-2164-15-540
101. de Araujo PG, Rossi M, de Jesus EM et al (2005) Transcriptionally active transposable elements in recent hybrid sugarcane. Plant J 44:707–717. https://doi.org/10.1111/j.1365-313X.2005.02579.x
102. Beulé T, Agbessi MDT, Dussert S et al (2015) Genome-wide analysis of LTR-retrotransposons in oil palm. BMC Genomics 16:1–14. https://doi.org/10.1186/s12864-015-2023-1

103. Muschner VC, Lorenz AP, Cervi AC et al (2003) A first molecular phylogenetic analysis of *Passiflora* (Passifloraceae). Am J Bot 90:1229–1238. https://doi.org/10.3732/ajb.90.8.1229
104. Muschner VC, Zamberlan PM, Bonatto SL, Freitas LB (2012) Phylogeny, biogeography and divergence times in *Passiflora* (Passifloraceae). Genet Mol Biol 35:1036–1043. https://doi.org/10.1590/S1415-47572012000600019

Affiliations

Zirlane Portugal da Costa[1] · Luiz Augusto Cauz-Santos[1] · Geovani Tolfo Ragagnin[2] · Marie-Anne Van Sluys[2] · Marcelo Carnier Dornelas[3] · Hélène Berges[4] · Alessandro de Mello Varani[5] · Maria Lucia Carneiro Vieira[1]

Zirlane Portugal da Costa
zirlane@usp.br

Luiz Augusto Cauz-Santos
luizcauz@usp.br

Geovani Tolfo Ragagnin
geovaniragagnin@usp.br

Marie-Anne Van Sluys
mavsluys@usp.br

Marcelo Carnier Dornelas
dornelas@unicamp.br

Hélène Berges
helene.berges@inra.fr

Alessandro de Mello Varani
amvarani@fcav.unesp.br

[1] Departamento de Genética, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba 13418-900, Brazil

[2] Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, São Paulo 05508-090, Brazil

[3] Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas 13083-862, Brazil

[4] Institut National de la Recherche Agronomique (INRA), Centre National de Ressources Génomique Végétales, Castanet-Tolosan 31326, France

[5] Departamento de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal 14884-900, Brazil