Chromosome Res (2015) 23:571–582
DOI 10.1007/s10577-015-9492-6

ORIGINAL ARTICLE

Chromosomal distribution and evolution of abundant retrotransposons in plants: *gypsy* elements in diploid and polyploid *Brachiaria* forage grasses

Fabíola Carvalho Santos · Romain Guyot · Cacilda Borges do Valle · Lucimara Chiari · Vânia Helena Techio · Pat Heslop-Harrison · André Luís Laforga Vanzela

Published online: 19 September 2015

Abstract Like other eukaryotes, the nuclear genome of plants consists of DNA with a small proportion of low-copy DNA (genes and regulatory sequences) and very abundant DNA sequence motifs that are repeated thousands up to millions of times in the genomes including transposable elements (TEs) and satellite DNA. Retrotransposons, one class of TEs, are sequences that amplify via an RNA intermediate and reinsert into the genome, are often the major fraction of a genome. Here, we put research on retrotransposons into the larger context of plant repetitive DNA and genome behaviour, showing features of genome evolution in a grass genus, *Brachiaria*, in relation to other plant species. We show the contrasting amplification of different retroelement fractions across the genome with characteristics for various families and domains. The genus *Brachiaria* includes both diploid and polyploid species, with similar chromosome types and chromosome basic numbers x= 6, 7, 8 and 9. The polyploids reproduce asexually and are apomictic, but there are also sexual species. Cytogenetic studies and flow cytometry indicate a large variation in DNA content (C-value), chromosome sizes and genome organization. In order to evaluate the role of transposable elements in the genome and karyotype organization of species of *Brachiaria*, we searched for sequences similar to conserved regions of TEs in RNAseq reads library produced in *Brachiaria decumbens*. Of the 9649 TE-like contigs, 4454 corresponded to LTR-retrotransposons, and of these, 79.5 % were similar to members of the gypsy superfamily. Sequences of conserved protein domains of gypsy were used to design primers for producing the probes. The probes were used in FISH against chromosomes of accesses of *B. decumbens*, *Brachiaria brizantha*, *Brachiaria ruziziensis* and *Brachiaria humidicola.* Probes showed hybridization signals predominantly in proximal regions, especially those for retrotransposons of the clades *CRM* and *Athila*, while elements of *Del* and *Tat* exhibited dispersed signals, in addition to those proximal signals. These results show that the proximal region of *Brachiaria* chromosomes is a hotspot for retrotransposon insertion,

Responsible Editors: Maria Assunta Biscotti and Ettore Olmo

F. C. Santos · A. L. L. Vanzela (✉)
Department of General Biology, Center of Biological Sciences, State University of Londrina, Londrina 86057-970 Paraná State, Brazil
e-mail: andrevanzela@uel.br

R. Guyot
Institut de Recherche pour le Développement (IRD), UMR IPME, BP 64501, 34394 Montpellier Cedex, France

C. B. do Valle · L. Chiari
Embrapa Gado de Corte, 79106-550 Campo Grande, Mato Grosso do Sul State, Brazil

V. H. Techio
Department of Biology, Federal University of Lavras, 37200-000, Lavras, Minas Gerais State, Brazil

P. Heslop-Harrison (✉)
Department of Genetics, University of Leicester, Leicester LE1 7RH, UK
e-mail: phh@molcyt.com

particularly for the gypsy family. The combination of high-throughput sequencing and a chromosome-centric cytogenetic approach allows the abundance, organization and nature of transposable elements to be characterized in unprecedented detail. By their amplification and dispersal, retrotransposons can affect gene expression; they can lead to rapid diversification of chromosomes between species and, hence, are useful for studies of genome evolution and speciation in the *Brachiaria* genus. Centromeric regions can be identified and mapped, and retrotransposon markers can also assisting breeders in the developing and exploiting interspecific hybrids.

Keywords centromeres · retrotransposons · FISH · in situ hybridization · metaviridae · grasses · genomics · genome organization · transposons · transposable elements · genetics · repetitive DNA · chromosomes

Abbreviations

PBS	Primer binding site
PR	Protease
RT	Reverse transcriptase
RT-*Athila*	Reverse transcriptase of *Athila* lineage
RT-*CRM*	Reverse transcriptase of *CRM* lineage
RT-*Tat*	Reverse transcriptase of *Tat* lineage
RNAse H	Ribonuclease H
INT	Integrase
IRAP	Inter-retroelement amplified polymorphism
PPT	Polypurine tract
LTRs	Long terminal repeats
LTR-RTs	Retrotransposons with LTR
TEs	Transposable elements
POL	Polygenic string
FISH	Fluorescent *in situ* hybridization
CRM	Centromere-specific retrotransposons of Maize

Introduction

The nuclear genome size of plants varies more than 2000-fold, from 63.40 Mbp (1C=0.0648 pg) in *Genlisea margaretae* (Greilhuber et al. 2006) to 259,000 Mbp (1C=with 264.9 pg) in the allohexaploid *Trillium hagae* (Zonneveld 2010; Pellicer et al. 2010). Sequencing and study of composition of genomes have shown that transposable elements (TEs) and responsible for much of the genome size variation in plants, in addition to those variations arising from differences in the levels of ploidy (Adams and Wendel 2005; Heslop-Harrison and Schmidt 2012). Transposable elements represent up to 85 % of the DNA (genome size) in cereals (Daron et al. 2014) or loblolly pine (Kamm et al. 1996; Neale et al. 2014), and TEs may increase their copy number, leading to a rapid expansion in genome size between closely related species (Gregory 2005; Pearce et al. 1996). In *Arabis alpina*, Willing et al. (2015) show the genome expansion through amplification of a retrotransposon family compared to *Arabidopsis thaliana*. TEs can also involve genome contraction through a wide range of illegitimate recombination and deletion processes, which depend on a balance between insertion and removal events in distinct chromosome regions (Bennetzen and Wang 2014).

Transposable elements are classified according to their mode of mobility into class 1, retrotransposons that transpose via an RNA intermediate using a copy-and-past mechanisms, and class 2, transposons that move via a DNA molecule using cut-and-paste events and may amplify before or during the transposition cycle (Hansen and Heslop-Harrison 2004). The LTR-retrotransposons (LTR-RTs) represent the majority of TEs identified in plant genomes (e.g. Bertioli et al. 2013). They are subdivided into two big super families: copia (Pseudovirideae) and gypsy (Metavirideae). The most striking difference between members of these two super families is the order of the integrase (INT) amino acid domain along the polyprotein gene encoded by the LTR retrotransposons (Hansen and Heslop-Harrison 2004; Wicker et al. 2007). *Copia* are the best studied and understood retrotransposons in plant species, probably because they seem to be more conserved in plant evolution.

Plant LTR-RTs often accumulate in heterochromatin regions of chromosomes (Gao et al. 2008; Kejnovsky et al. 2012; Pearce et al. 1996) and are also found in proximal and non-recombining regions of sex chromosomes (VanBuren et al. 2015). An example of expansion of retrotransposons in centromeric heterochromatin is seen in a comparison of related species of *Alstroemeria* (Kuipers et al. 1998). In *Arabidopsis*, *copia* elements are clustered at centromeres (Heslop-Harrison et al. 2003) and also occur along the chromosomes, while *gypsy* elements are preferentially inserted in pericentromeric regions (Pereira 2004, who also considers evolutionary mechanisms and consequences for copy number and

distribution of elements). In most monocots and dicots examined so far, centromere regions contain a specific lineage of *gypsy* LTR-RTs called Centromeric Retrotransposons (CRs, called CRM in Maize), suggesting that these sequences play an important role in the function and evolution of centromeres (Zhong et al. 2002; Gao et al. 2015).The gypsy family of retroelements is also responsible for the genome expansion seen in *A. alpina* (Willing et al. 2015), while insertion and excision of DNA transposable element families such as hAT elements in *Musa* (Menzel et al. 2015) or MITEs in *Brassica* (Nouroz et al. 2015) leads to differences both in genome size and organization.

The activity of retrotransposons with LTRs can be modulated by DNA methylation and silencing processes and is induced by a wide range of internal and external factors, including biotic and abiotic stresses (Takeda et al. 1998; Casacuberta and González 2013). As consequence, LTR-RTs may proliferate and induce mutations with biological effects, depending of their insertion sites. Hybridization and allopolyploidization are also considered as potential genomic shocks involved in the accumulation of transposable elements and genome reorganization in plants (Parisod et al. 2010; Renny-Byfield and Wendel 2014; Zou et al. 2011). As well as amplification of genome sizes through transposable elements, most plant species have one or more rounds of whole-genome doubling events or polyploidy in their ancestry (see Heslop-Harrison 2012). These events may be followed by chromosomal loss and rearrangements, and there may be rapid genomic changes at the time of the polyploidization or hybridization events (Ma and Gustafson 2008; Gaeta et al. 2007).

The grass genus *Brachiaria* (Poaceae) is of African origin and comprises about 100 species (Renvoize et al. 1996), with many polyploids and hybrid species (Boldrini et al. 2009; Akiyama et al. 2010; Nielen et al. 2010) propagated sexually through seeds and vegetatively. The basic chromosome number is x=6, 7 and 9, with species including 2n=2x=18 diploids and various ploidies to 2n=10x=90. Four of the species, *Brachiara ruziziensis*, *Brachiara decumbens*, *Brachiara brizantha* and *Brachiara humidicola*, have great economic importance in production of forage and seeds in weak and acids soils (Nakamura et al. 2005). In general, hybrids between these species exhibit variable frequencies of univalent and multivalent pairing, asynchrony in cell divisions and abnormal development of the microsporogenesis (Mendes-Bonato et al. 2002), suggesting that polyploidy and hybridization may cause a loss of 'genomic homeostasis'. Together, these observations suggest the *Brachiaria* genus is a good model to use to compare the distribution and study the impact of TEs between diploid and polyploid species.

To understand the diversity of LTR-RTs in the organization of proximal chromosome regions in *Brachiaria*, we aimed to identify and characterize transcriptionally active *gypsy* LTR-RTs using RNAseq data generated from the diploid species *B. decumbens*. Probes related to four transcribed *gypsy* lineages were generated and used for in situ hybridization to chromosomes of diploids (*B. decumbens*, *B. ruziziensis* and *Brachiara brizantha*) and polyploids (*B. decumbens*, *B. humidicola* and *B. brizantha*). Probes of 5S rDNA and telomeres were used to identify chromosomes and their morphology. By showing different retroelement distributions, our data exemplify the genomic distribution of this important group of abundant sequences and contribute to the knowledge of genome structure, composition and evolution of diploid and polyploid *Brachiaria* species.

Materials and methods

Biological samples

Samples of ten accessions belonging to four species of *Brachiaria* were used: (i) B72 ($2n$=4x=36) and B183 ($2n$=5x=45) of *B. brizantha*, (ii) R102 ($2n$=2x=18) of *B. ruziziensis*, (iii) H16 ($2n$=4x=36) and H36, H38 and H112 ($2n$=9x=54) of *Brachiara humidicola*, (iv) D4 ($2n$=2x=18) of *B. decumbens* and (v) two interspecific hybrids with $2n$=4x=36. Accessions were vegetatively propagated and are maintained in the Embrapa Gado de Corte (Brazilian Research Institute), Campo Grande, Mato Grosso do Sul State, Brazil.

RNAseq analysis and design of FISH probes

RNAseq data were produced from *B. decumbens* roots using Illumina HiSeq 2000 sequencing technology. RNA sequencing reads were assembled with Trinity (Grabherr et al. 2011). All contigs were first compared to the RepBase (Jurka et al. 2005) amino acids database (version 19.06; http://www.girinst.org/repbase/) using BLAST (BLASTx, *E-value* 1e-4) to classify them according to their similarities. The nucleotide contigs

showing significant similarities to LTR-RTs coding regions were further compared to the reverse transcriptase (RT), integrase (INT) and capsid (GAG) amino acid domains database downloaded from the Gypsy Database 2.0 (http://gydb.org/) (Llorens et al. 2011). The RT amino acid domains from *B. decumbens* nucleotide contigs were extracted using GeneWise (https://www.ebi.ac.uk/Tools/psa/genewise/) with at least 150 amino acid residues. Sequences were aligned using muscle tools (Edgar 2004) and a neighbor-joining tree was constructed using ClustalW and edited with FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Primers were designed on selected contigs with Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/) and conserved regions were amplified by PCR. Sequences, including primers, regions and sizes are available in the Table 1. Inserts of the p*Ta*794 clone containing the 5S rDNA sequence of *T. aestivum* (Gerlach and Dyer 1980) and the pLT11 clone containing the telomeric insert $(TTTAGGG)_n$ of p*At*T4 from *A. thaliana* (Richards and Ausubel 1988) were also used in FISH. Probes were labeled with digoxigenin-11-dUTP or biotin-11-dUTP using the Invitrogen Bioprime CGH labelling kit.

Fluorescent in situ hybridization

Root tips were pretreated in 2 mM 8-hydroxyquinoline, 5 % dimethylsulfoxide for 24 h at 9 °C and fixed in ethanol/acetic acid (3:1, *v/v*). Chromosome spreads were made as described by Schwarzacher and Heslop-Harrison (2000). Samples were digested in 2 % cellulase Onozuka R-10 (Serva) and 20 % pectinase (Sigma) for 5 h at 37 °C, squashed in a drop of 45 % acetic acid and coverslips removed in liquid nitrogen. Slides were sequentially treated with RNaseA (100 μg mL^{-1}) and pepsin (10 μg mL^{-1}), washed in 2× SSC and dehydrated in an ethanol series. A hybridization mix (40 μL/slide), which consisted of 50 % formamide, 2× SSC, 10 % dextran sulphate, 25 ng salmon sperm DNA, 0.15 % SDS, 400 ng of labeled probe and 0.25 mM EDTA, was denatured at 80 °C for 10 min applied onto slide. Both were denatured at 70° for 7 min using a thermal cycler, followed by an overnight incubation at 37 °C in a humid chamber. Slides were washed in SSC buffer at 80 % stringency, and detection was done with Alexa 594-antibiotin (red) and FITC-antidigoxigenin (green) in 5 % BSA in SSC buffer. Post-detection washes were done in the SSC buffer followed by staining with DAPI (4 μg mL^{-1}/slide). Samples were mounted with antifade (Citifluor).

Preparations were analyzed with a Nikon E800 imaging epifluorescence microscope. Images were overlaid using Adobe Photoshop 6.0; image brightness and contrast in each colour were optimized using only operations affecting the whole image equally.

Results

The RNA sequences of *B. decumbens* were obtained using Illumina sequencing, and reads assembled into 126,601 unigenes, with an average size of 1 kb (total 126,383,241 bp). Fifty-eight percent of unigenes (73,553; N50 1823) have a length larger than 500 bp. The functional annotation of transposable elements using BLASTx against the RepBase amino acid database showed that 9649 unigenes (7.6 %) have at least one match on these database of known proteins of TEs. Based on the BLASTx results, 4454 TE unigenes (47.1 %, Fig. 1) were classified into the class I LTR-RTs group of retrotransposons. Among LTR-RT unigenes, *gypsy* super family members represented the majority (79.5 %). The reverse transcriptase domains of identified LTR-RTs were extracted and used to draw a neighbor-joining tree with the reference RT domains from the Gypsy Database. Results showed that most of the lineages of *gypsy* and *copia* super families typically present in plants were also identified in the transcriptome of *B. decumbens* with the exception of the *Galadriel* lineage. Among the clades (Fig. 2), seven contigs were selected from four different *gypsy* lineages to design primers for PCR and FISH: *Del*, *CRM*, *Athila* and *Tat* (Table 1 and Fig. 2; Figs. 3, 4 and 5 for in situ hybridization results).

The universal probes of 5S rDNA (p*Ta*794) of *T. aestivum* and $TTTAGGG_{(n)}$ (telomeric sequence, pLT11 clone of *A. thaliana*) were used to help identify chromosomes and detect possible variations in relation to hybridization site numbers. The 5S rDNA probe showed hybridization sites predominantly in proximal regions, varying from four sites in *B. decumbens* 2× (Fig. 3c) to six in *B. brizantha* 5× (Fig. 4d). The telomeric probe hybridized always at the chromosome ends without evidence of ectopic telomeric sites (polyploid *B. humidicola* Fig. 4a, c; *B. brizantha* Fig. 5a).

FISH using *gypsy* probes showed differences in the chromosome distribution between lineages and between

Table 1 List and characteristics of *B. decumbens gypsy* members that were used as probes in FISH

Clades	Primer F	Primer R	Name of contigs	Length of contigs (bp)	Protein domains
Del	TGTCCCCCTGACTGTCTAGG	ACATCGCTGCAAGTGGAGAA	CL1419.Contig1	1721	GAG
Del	CCAGACGTGTTCCCGAAAGA	GCATAGCTTCTGCGCTTGTC	CL4156.Contig1	7044	RT
Del	AGGGTCCTGTGGCTGGATAT	GCCCCTTCTGGAACTTCCTC	CL16780.Contig1	4607	GAG
CRM	CGCACTTCGAGCATTACGTG	ATCCACCTCAATGCCCTGTG	Unigene32063	2959	RT
Athila	TGCGATGAGGGAGGTAGTCA	TGAGGGTGTAGAAGGCGAGA	Unigene48044	1039	RT
Tat	TCTGCAGGATGACGACAAGG	AGCCAAACTCTGTCACCACC	Unigene431	3781	RNase H
Tat	ACCAGGTTATTGACTCCACGG	GCTTCCACGTTGCGATGGAG	Unigene431	3781	RT
Tat	TTCACACATGTGCTGGTTGC	TCTCGACTTGTCCATTGGCC	CL2826.Contig5	5007	INT

ploidy levels and species. The *Athila* probe, for instance, was located predominantly in the centromeric-pericentromeric regions of chromosomes of diploids *B. decumbens* (Fig. 3a, b), *B. ruziziensis* (Fig. 3e) and *B. brizantha* (Fig. 3i), but with few dispersed signals (Fig. 3a). In the polyploid *B. brizantha* (Fig. 5c), the *Athila* probe showed also centromeric-pericentromeric hybridization signals, but differences in location and abundance (strength of signal) were evident between chromosomes, with some minor sites (Fig. 5c).

The *Tat* probes showed relatively strong hybridization signals in the centromeric-pericentromeric regions, with dispersed signals in interstitial regions of diploid *B. decumbens* (Fig. 3c) and *B. ruziziensis* (Figs. 3d, e). Differences in the intensity and size of signals were seen between chromosomes within a genome (Fig. 3e). However, the polyploids showed differences in the signal location. *B. humidicola*, for instance, showed a more dispersed profile of hybridization, except for 14 chromosomes that exhibited accumulation of the *Tat* probe in centromeric-pericentromeric regions (Fig. 4a). In the polyploid *B. brizantha* (Fig. 5c, d), the signals produced using the reverse transcriptase as probe were weak (low intensity, most likely due to a weak probe), but they showed differential labeling of chromosomes (Fig. 5c). FISH using an INT probe of the same element showed stronger signals that were scattered and/or concentrated in about half of the chromosomes, while the other half of the signals were weak and dispersed (Fig. 5d).

FISH with *CRM* showed centromeric signals, with a few inconspicuous signals at pericentromeric-interstitial regions in diploids of *B. ruziziensis* (Fig. 3f, g) and *B. brizantha* (Fig. 3j). In the polyploids, the *CRM* probe hybridized in the proximal regions, but the signals were less intense. No dispersed signals were noted along chromosome arms of *B. humidicola* (Fig. 4c). In contrast, chromosomes of polyploid *B. brizantha* exhibited

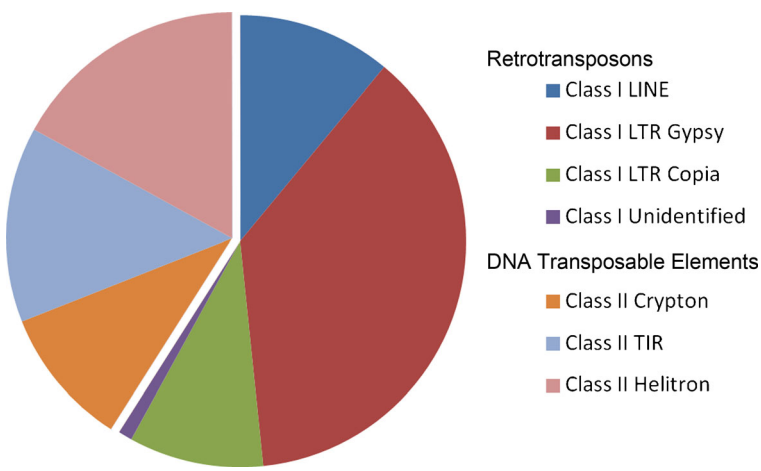

Fig. 1 Transposable element families in the transcriptome of *Brachiaria decumbens*, obtained from RNAseq. The proportions of major families of class I (retrotransposons, 59 %) and class II (DNA, 41 %) transposable elements are shown

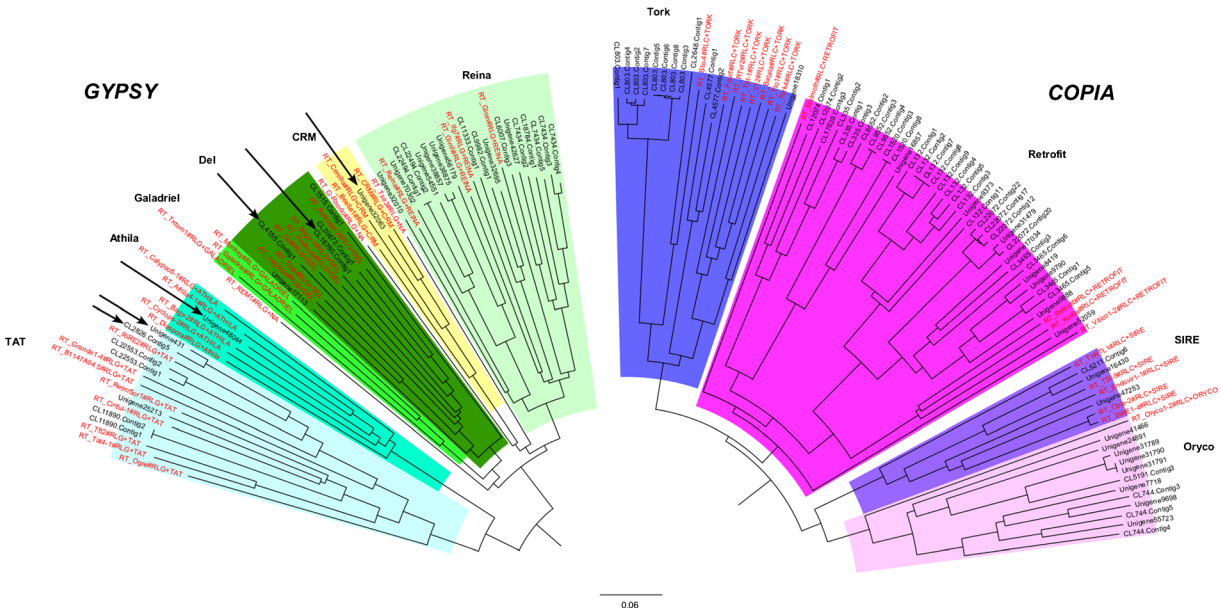

Fig. 2 Cladogram showing the relationships between *Ty*1-*copia* and *Ty*3-*gypsy* retrotransposable element superfamilies identified in the transcriptome of *B. decumbens* within contigs/unigenes (*black names*) or as characteristic domains (*red*). *Arrows* indicate the *gypsy* LTR-RTs sequences used to design primers for in situ hybridization (Table 1)

proximal signals, besides some interstitial dots in interstitial regions (Fig. 5a, and in the box below and to the right of the image). The *Del* probe showed dots scattered in all along chromosome regions of diploid *B. brizantha* (Fig. 3h, i), but with some few chromosomes with proximal signals. Likewise, in the polyploid *B. brizantha*, the *Del* probes showed signals as dots scattered along chromosomes, with marking in some proximal regions (Fig. 5b and inset box v and iv).

Discussion

Transposable element DNA families are relatively easily identified in genome sequences due to their repetitive nature and sequence signatures that facilitate recognition using bioinformatic tools (Lerat 2010; Janicki et al. 2011; Heitkam et al. 2014; Menzel et al. 2015). In plants, class I transposable elements, particularly those of *gypsy* and *copia* LTR-RTs super families, are the most frequent in relation to the other elements that occupy the plant genomes (Wicker et al. 2007). Because these elements are commonly expressed in plant genomes, we can identify them from a collection of RNAseq reads using a model plant with no available sequenced genome. According to data obtained from RNAseq in *Prunus persica* (Rosaceae), for instance, the LTR-RTs represent 18.6 %, (10 % *gypsy* and 8.6 % *copia*-like; Verde et al. 2013), while in DNA analysis of other monocots, retrotransposons (LTR-RTs) sequences can vary from 21.4 % in *Brachypodium* (a small genome with most TEs located in the centromeric regions; Vogel et al. 2010) to 26 % in rice, about 50 % in sorghum and banana (including 25.7 % *copia*-like and 11.6 % *gypsy*-like; d'Hont et al. 2012), and more than 80 % in wheat. Transposable elements and retrotransposon-related sequences are normally found in transcriptomes. Some of these sequences are 'active', and their insertion into the genome leads to variation that is detectable using Inter-Retroelement Amplified Polymorphism (IRAP) markers (Vicient et al. 2001; Alsayied et al. 2015). However, few, if any of the transcripts, are likely to be translated (only from the small proportion of TE sequences with functional open reading frames) or reverse-transcribed, nor reinserted into the genome, in a single generation. Activation of transposable elements by stress conditions (including environmental, tissue culture or sexual hybridization) is well known, and new insertions can sometimes be detected following such events (Takeda et al. 1998; Parisod et al. 2010).

The assembly obtained from an RNAseq assembly of *B. decumbens* showed a predominance of expressed *gypsy* elements (79.5 %) when compared with *copia* super family members, as well as members of other

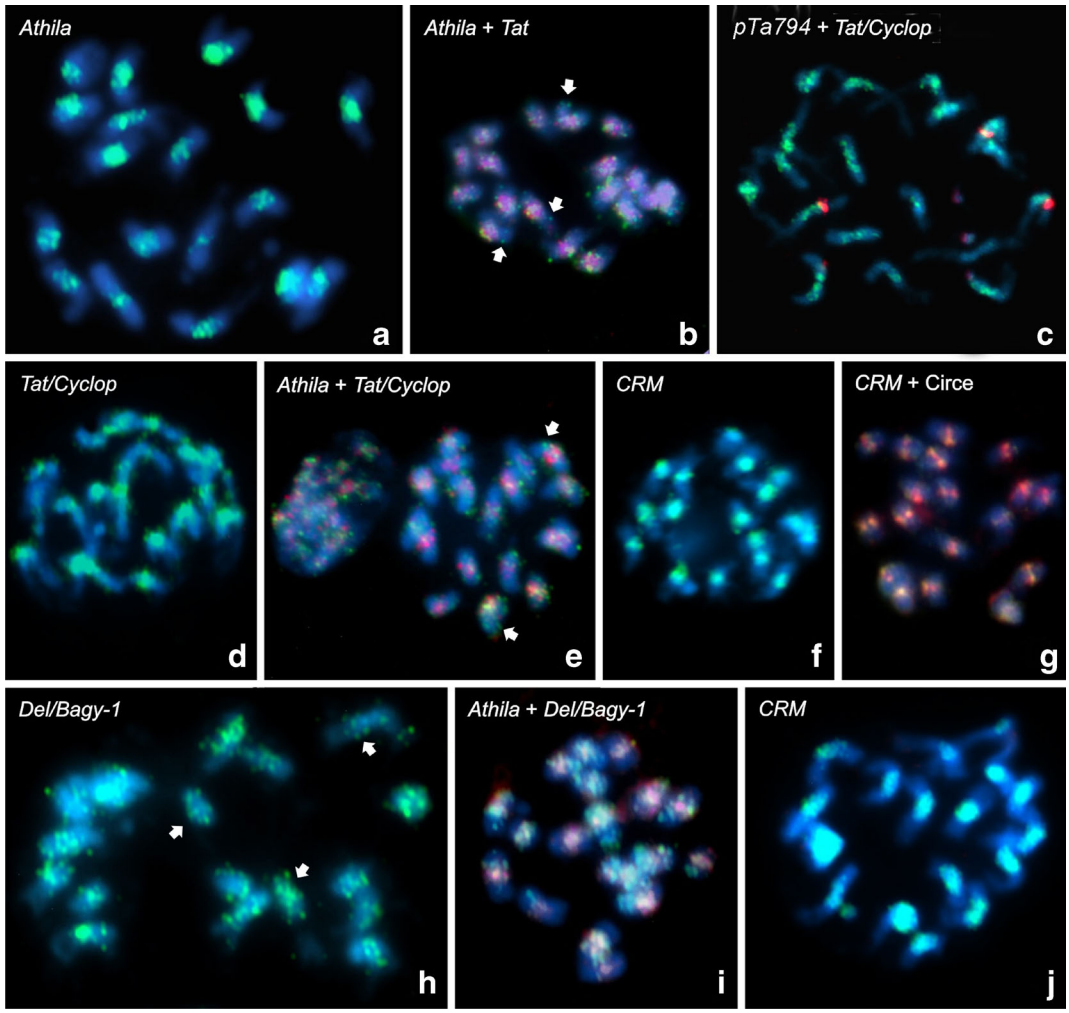

Fig. 3 In situ hybridization of *gypsy* retrotransposon probes to chromosomes (fluorescing blue with DAPI) of diploid species of *Brachiaria*. **a** Prometaphase of *B. decumbens* hybridized with RT-*Athila* probe (*green*). Signals were accumulated in centromeric-pericentromeric regions of all chromosomes, and no signals were detected in distal regions. **b** Metaphase of *B. decumbens* hybridized with RT-*Athila* probe (observe proximal signals in *red*) and with RT-*Tat* probe (*green*) hybridizing as dots along the chromosome arms, with some pericentromeric signals. **c** Prometaphase of *B. decumbens* hybridized with p*Ta*794 probe (*red*) and INT-*Tat*/*Cyclops* probe (*green*). The 5S rDNA shows four interstitial signals, while *Tat/Cyclops* probe exhibited signals spread from the centromeric region well into the interstitial regions. **d** Prometaphase of *B. ruziziensis* hybridized with INT-*Tat/Cyclops* probe (*green*), which showed also signals spread from the centromeric region, reaching interstitial regions. **e** Metaphase of *B. ruziziensis* hybridized with RT-*Athila* probe (*red*), with signals accumulated in centromeric-pericentromeric regions of all chromosomes, and INT-*Tat*/*Cyclops* probe (*green*) with dot-like signals in pericentromeric regions and in interstitial-distal signals of few chromosomes. Note an interphase nucleus showing collocation of both probes. **f** Metaphase of *B. ruziziensis* hybridized with RT-*CRM* probe (*green*). Observe only proximal signals, without interstitial-distal ones. **g** Metaphase of *B. ruziziensis* hybridized with RT-*CRM* (*red*) and GAG-*Circe* (*green*) probes, showing collocation at centromeric region in all chromosomes. **h** Metaphase of *B. brizantha* hybridized with RT-*Del/bagy*-1 (*green*), showing dot-like signals in proximal, interstitial and distal regions. However, note that there are differences of signals amount among chromosomes. **i** Metaphase of *B. brizantha* hybridized with RT-*Del/bagy*-1 (*green*), showing dot-like signals in proximal, interstitial and distal regions, and RT-*Athila* probe (*red*) with centromeric-pericentromeric signals. **j** Metaphase of *B. brizantha* hybridized with RT-*CRM* probe (*green*), showing only centromeric-pericentromeric signals, without distal ones. Bar=5 μm

repetitive DNA families (Fig. 1). The main *gypsy* lineages in *B. decumbens* were of the *Reina*, *Athila*, *Tat*, *Del* and *CRM*, typical of many plant species with a range of families. Primers were designed for these four last lineages to amplify PCR products. *Athila* and *Tat* elements comprise two families of large *gypsy* LTR-RTs, with

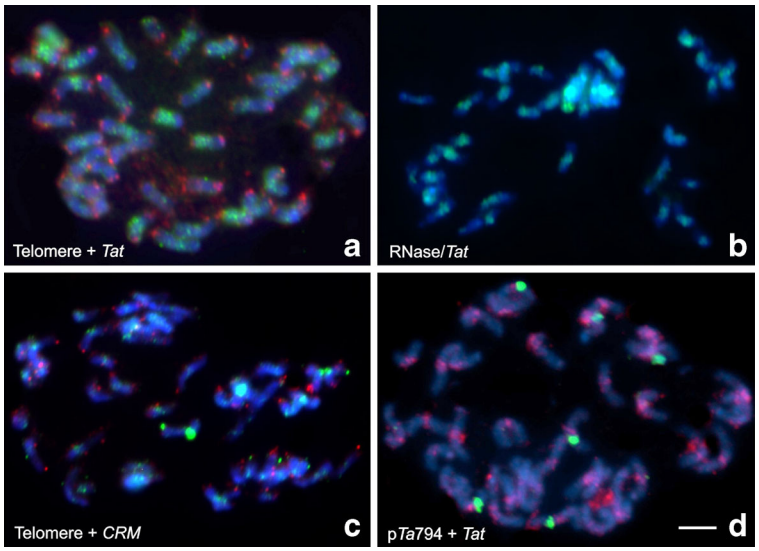

Fig. 4 In situ hybridization of *gypsy* retrotransposon probes to chromosomes in polyploids of *B. humidicola* (**a**, **c**) and *B. brizantha* (**b**, **d**): **a** partial metaphase hybridized with telomeric probe (*red*), which shows dot-like signals distally positioned, and INT-*Tat/Cyclops* probe (*green*) along chromosomes with concentration of signals in centromeric-pericentromeric regions in almost half chromosomes. **b** Partial metaphase hybridized with RNAse-*Tat* probe (*green*). Signals appeared predominantly in centromeric-pericentromeric regions, with some signal extending into the interstitial regions of few chromosomes. **c** Chromosomes of *B. humidicola* hybridized with telomeric probe (*red*), which showed terminal signals, and with RT-*CRM* probe (*green*), that showed proximal weak signals, and few interstitial dot-like signals. *Strong green signals* are nonspecific. **d** Chromosomes of *B. brizantha* hybridized with p*Ta*794 probe (*green*), showing six interstitial-proximal signals and INT-*Tat-cyclop* probe (*red*), with signals predominantly centromeric-pericentromeric and few interstitial. Bar=5 μm

sizes ranging between 10 and 12 kb. These are recognized by a primer binding site (PBS) complementary to tRNA-Glu. *Athila/Tat* form a large clade in the phylogeny proposed by Llorens et al. (2009) (see also *Gypsy* Database 2.0 http://gydb.org/index.php/Phylogeny:POL_LTR_retroelements), but both can be separated by differences in the homology and number of 3'-ORFs (see Chavanne et al. 1998). Probes of both groups of LTR retrotransposons were obtained from PCR from various conserved regions including RT from *Athila*-like, and RT, INT or RNaseH from *Tat*-like and in situ hybridization showed these were located preferentially in proximal chromosome regions; although in polyploids of *B. decumbens* and *B. humidicola*, the *Tat* probe was more scattered in interstitial regions than the *Athila* probe. Although both probes have been located in the proximal chromosome region, they are not typical centromeric retrotransposons, perhaps because they do not carry any chromodomain (Weber and Schmidt 2009). The *Bagy-1* LTR-RT, an element from the *Del gypsy* lineage that was initially reported from the barley genome, showed hybridization signals such as dots distributed in the terminal and interstitial chromosome regions, but with a greater concentration in the pericentromeric region. This dispersed distribution of *Bagy*-1 may be associated with the fact that this LTR-RT belongs to a group considered to be quite active in the grass genomes (Vicient et al. 2001). Notably, in the tetraploid accession of *B. humidicola* (Fig. 4a), the INT-*Tat/Cyclops* probe labels about half the chromosomes more strongly than the others, suggesting that the sequence has amplified in one of the ancestral genomes before they came together in the tetraploid.

Of all the LTR-RTs tested here, only the *CRM*-like elements belong to chromoviruses, generally the most widespread clade of *gypsy*-like elements (Gorinšek et al. 2004). Centromere-specific chromoviruses are very common in both angiosperm and gymnosperm genomes and are close to *Reina*, *Tekay* and *Galadriel* clades (Llorens et al. 2009). All the members of these clades present an integrase chromodomain at C-terminus (Gorinšek et al. 2004), so they can recognize and interact with modified centromeric histones, suggesting that these elements present an important role for the structure and the function of the centromere (Gao et al. 2008; Houben et al. 2007 for another centromeric retroelement in barley; Gao et al. 2015 in rice species).

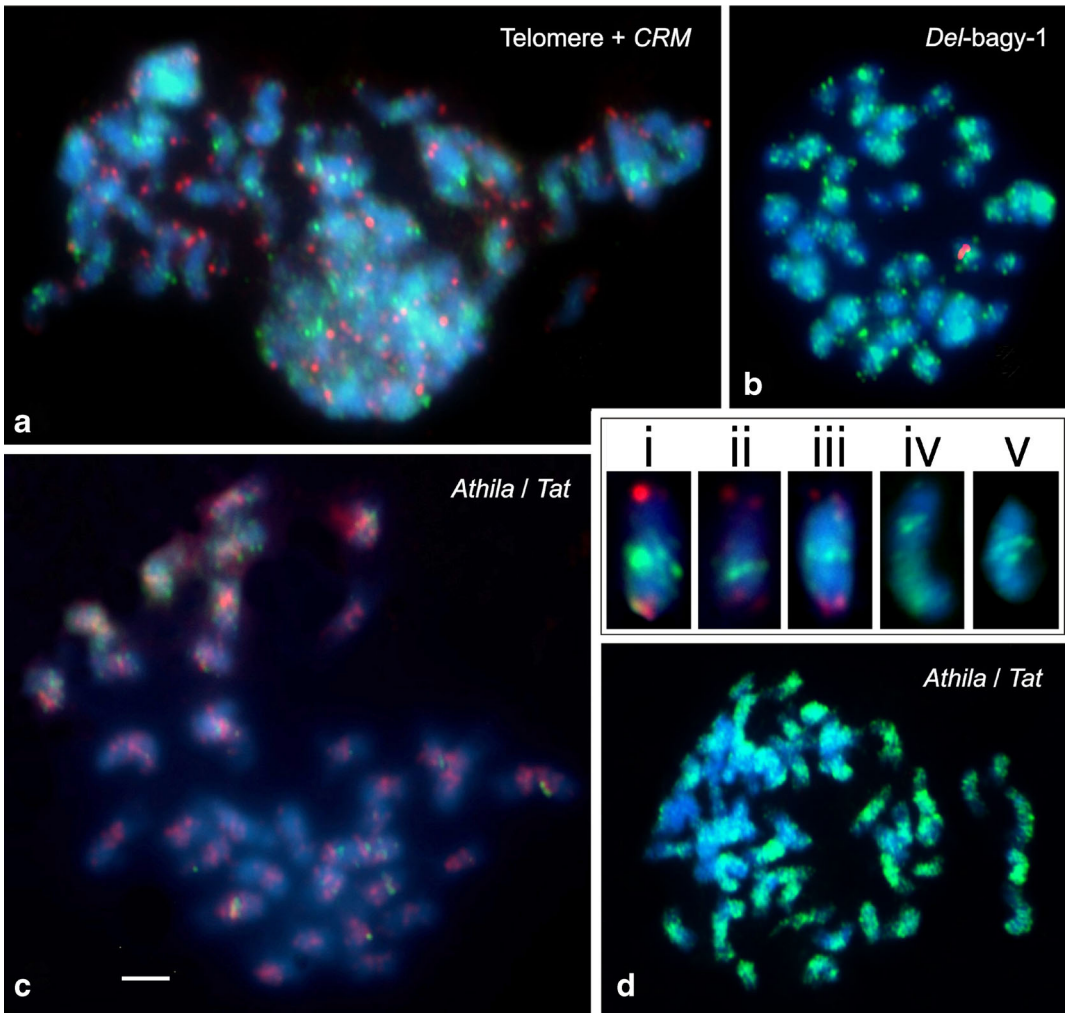

Fig. 5 In situ hybridization of *gypsy* retrotransposon probes to chromosomes in polyploids of *B. brizantha*: **a** metaphase and nucleus hybridized with telomeric probe (*red*), which shows dot-like signals terminally positioned, and RT-*CRM* probe (*green*) with signals concentrated in centromeric-pericentromeric regions. In the *box inset* below (*i*, *ii* and *iii*) telomeric dots and proximal *CRM* signals are presented. **b** Metaphase hybridized with RT-*Del/baggy*-1 probe (*green*), showing dispersed signals along chromosomes. Box below shows two chromosomes with more interstitial signals (*iv*) and more interstitial to pericentromeric ones (*v*). **c** Metaphase hybridized with RT-*Athila* (*red*) and RT-*Tat* (*green*) probes, showing RT-*Athila* accumulated in centromeric-pericentromeric regions and RT-*Tat* as dot-like in interstitial positions. **d** Metaphase hybridized with INT-*Tat/Cyclop* probe (*green*), showing signals predominantly scattered along chromosomes, with some of them exhibiting accumulation of signals in proximal regions. Bar=5 μm

The conserved regions of four representatives of *gypsy* lineages (*CRM*, *Athila*, *Del* and *Tat*) of *B. decumbens* showed a preference for integration into proximal chromosome regions, which is targeted to heterochromatin locations and suppression of recombination. This shows the non-random distribution of *CR* (centromeric retrotransposons) of *Brachiaria*. These retrotransposon lineages appeared also more concentrated in proximal chromosome regions of *Brachypodium distachyon*, with transposons-like distributed closer to the gene-rich regions (Vogel et al. 2010). In some cereals, such as barley, wheat and rice, centromeric and pericentromeric regions contain large occurrence of *gypsy* LTR-RTs (Cheng and Murata 2003; Nagaki et al. 2005). The accumulation of these elements into the proximal chromosome regions in these four species of *Brachiaria* is in agreement with the proposal of Bennetzen and Wang (2014), suggesting that TEs predominate in recombination-poor regions, which are genomic locations of low rate of unequal homologous

recombination and gene conversion. These are also gene-depleted regions where retrotransposons insertion may have minimal effect.

Lisch and Bennetzen (2011) review the different rates of amplification, insertion and removal of transposable elements and show the association with epigenetic control. Together, it is clear that transposable elements have a major consequence on the complexity and organization of genomes in different plant species. Different transposable elements, as exemplifed by *gypsy* retrotransposons in the *Brachiaria* species examined here, can have characteristic distributions which have different impacts on behavior of genomes and amplification of elements and hence on the diversification of species.

The elements are not only important to understand for fundamental biological and evolutionary reasons, but also because of their abundance and consequences for gene expression. For breeding of crops, including the *Brachiaria* forage grasses, it is important to exploit the biodiversity within the whole genus by making crosses, synthetic hybrids and polyploids and transposable elements can be used to provide markers, to define relationships between genomes in hybrids, and examine pairing and recombination at meiosis.

Acknowledgments The authors thank the Brazilian agencies Fundação Araucária, CNPq and CAPES for financial support. JSHH thanks IAEA-FAO for support via Cooperative Research Programme D23029 Climate Proofing of Food Crops: Genetic Improvement for Adaptation to High Temperatures in Drought Prone Areas and Beyond.

References

Adams KL, Wendel JF (2005) Polyploidy and genome evolution in plants. Curr Opin Plant Biol 8:135–141. doi:10.1016/j.pbi.2005.01.001

Akiyama Y, Yamada-Akiyama H, Ebina M (2010) Morphological diversity of chromosomes bearing ribosomal DNA loci in *Brachiaria* species. Grassl Sci 56(4):217–223

Alsayied NF, Fernández JA, Schwarzacher T, Heslop-Harrison JS (2015) Diversity and relationships of *Crocus sativus* and its relatives analysed by inter-retroelement amplified polymorphism (IRAP). Ann Bot 116:359–368

Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol 65:19.1–19.26. doi:10.1146/annurev-arplant-050213-035811

Bertioli DJ, Vidigal B, Nielen S, Ratnaparkhe MB, Lee T-JH, Leal-Bertioli SCM, Kim C, Guimaraes PM, Seijo G, Schwarzacher T, Paterson AH, Heslop-Harrison P, Araujo ACG (2013) The repetitive component of the a genome of peanut (*Arachis hypogaea*) and its role in remodelling intergenic sequence space since its evolutionary divergence from the B genome. Ann Bot 112(3):545–559. doi:10.1093/aob/mct128

Boldrini KR, Micheletti PL, Gallo PH, Mendes-Bonato AB, Pagliarini MS, Valle CB (2009) Origin of a polyploid accession of *Brachiaria humidicola* (Poaceae: Panicoideae: Paniceae). Genet Mol Res 8:888–895

Casacuberta E, González J (2013) The impact of transposable elements in environmental adaptation. Mol Ecol 22:1503–1517

Chavanne F, Zhang D, Liaud M, Cerff R (1998) Structure and evolution of cyclops: a novel giant retrotransposon of the *Ty3*/*Gypsy* family highly amplified in pea and other legume species. Plant Mol Biol 37:363–375. doi:10.1023/A:1005969626142

Cheng Z, Murata M (2003) A centromeric tandem repeat family originating from a part of *Ty3*/*gypsy*-retroelement in wheat and its relatives. Genetics 164:665–672

D'Hont A, Denoeud F, Aury J-M, Baurens, FC, Carreel F, Garsmeur O, … & Jaillon O (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature, 488(7410), 213–217

Daron J, Glover N, Pingault L, Theil S, Jamilloux V, Paux E, Barbe V, Mangenot S, Alberti A, Wincker P, Quesneville H, Feuillet C, Choulet F (2014) Organization and evolution of transposable elements along the bread wheat chromosome 3B. Genome Biol 15(12):546. doi:10.1186/s13059-014-0546-4

Edgar (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792–1797

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC (2007) Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell 19:3403–3417

Gao X, Hou Y, Ebina H, Levin HL, Voytas DF (2008) Chromodomains direct integration of retrotransposons to heterochromatin. Genome Res 18:359–369

Gao D, Jiang N, Wing RA, Jiang J, Jackson SA (2015) Transposons play an important role in the evolution and diversification of centromeres among closely related species. Frontiers in plant science 6:216

Gerlach WL, Dyer TA (1980) Sequence organization of the repeating units in the nucleus of wheat which contain 5S rRNA genes. Nucleic Acids Res 8(21):4851–4865

Gorinšek B, Gubenšek F, Kordiš D (2004) Evolutionary genomics of chromoviruses in eukaryotes. Mol Biol Evol 21(5):781–798

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29(7):644–652

Gregory RT (2005) The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. Ann Bot 95:133–146. doi:10.1093/aob/mci009

Greilhuber J, Borsch T, Müller K, Worberg A, Porembski S, Barthlott W (2006) Smallest angiosperm genomes found in

Lentibulariaceae with chromosomes of bacterial size. Plant Biol 8:770–777. doi:10.1055/s-2006-924101

Hansen CN, Heslop-Harrison JS (2004) Sequences and phylogenies of plant pararetroviruses, viruses and transposable elements. Adv Bot Res 41:165–193

Heitkam T, Holtgräwe D, Dohm JC, Minoche AE, Himmelbauer H, Weisshaar B, Schmidt T (2014) Profiling of extensively diversified plant LINEs reveals distinct plant-specific subclades. Plant J 79:385–397

Heslop-Harrison JS (2012) Genome evolution: extinction, continuation or explosion? Curr Opin Plant Biol 15:115–121

Heslop-Harrison JS, Schmidt T (2012) Plant nuclear genome composition. In: Encyclopedia of life sciences. eLS 2012. Wiley, Chichester. doi:10.1002/9780470015902.a0002014.pub2, http://www.els.net/

Heslop-Harrison JS, Brandes A, Schwarzacher T (2003) Tandemly repeated DNA sequences and centromeric chromosomal regions of Arabidopsis species. Chromosom Res 11:241–253

Houben A, Schroeder-Reiter E, Nagaki K, Nasuda S, Wanner G, Murata M, Endo TR (2007) CENH3 interacts with the centromeric retrotransposon *cereba* and GC-rich satellites and locates to centromeric substructures in barley. Chromosoma 116(3):275–283

Janicki M, Rooke R, Yang G (2011) Bioinformatics and genomic analysis of transposable elements in eukaryotic genomes. Chromosom Res 19:787–808. doi:10.1007/s10577-011-9230-7

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110(1–4): 462–467

Kamm A, Doudrick RL, Heslop-Harrison JS, Schmidt T (1996) The genomic and physical organization of Ty1-copia-like sequences as a component of large genomes in *Pinus elliottii* var. *elliottii* and other gymnosperms. Proc Natl Acad Sci 93: 2708–2713

Kejnovsky E, Hawkins JS, Feschotte C (2012) Plant Transposable Elements: Biology and Evolution, Wendel JF et al. (eds.), Plant Genome Diversity Volume 1

Kuipers AGJ, Heslop-Harrison JS, Jacobsen E (1998) Characterisation and physical localisation of Ty1-copia-like retrotransposons in four *Alstroemeria* species. Genome 41: 357–367

Lerat E (2010) Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. Heredity 104:520–533. doi:10.1038/hdy.2009.165

Lisch D, Bennetzen JL (2011) Transposable element origins of epigenetic gene regulation. Curr Opin Plant Biol 14:156–161. doi:10.1016/j.pbi.2011.01.003

Llorens C, Muñoz-Pomer A, Bernad L, Botella H, Moya A (2009) Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. Biol Direct 4:41. doi:10.1186/1745-6150-4-41

Llorens C, Futami R, Covelli L, Dominguez-Escriba L, Viu JM, Tamarit D, Aguilar-Rodriguez J, Vicente-Ripolles M, Fuster G, Bernet GP, Maumus F, Munoz-Pomer A, Sempere JM, LaTorre A, Moya A (2011) The gypsy database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res (NARESE) 39(suppl 1):D70–D74. doi:10.1093/nar/gkq1061

Ma X-F, Gustafson JP (2008) Allopolyploidization-accommodated genomic sequence changes in triticale. Ann Bot 101:825–832

Mendes-Bonato AB, Pagliarini MS, Forli F, Valle CB, de Oliveira Penteado MI (2002) Chromosome numbers and microsporogenesis in *Brachiaria brizantha* (Gramineae). Euphytica 125(3):419–425

Menzel G, Heitkam T, Seibt KM, Nouroz F, Müller-Stoerme M, Heslop-Harrison JS, Schmidt T (2015) The diversification and activity of hAT transposons in Musa genomes. Chromosom Res 22:559–571. doi:10.1007/s10577-014-9445-5

Nagaki K, Neumann P, Zhang D, Ouyang S, Buell CR, Cheng Z, Jiang J (2005) Structure, divergence, and distribution of the CRR centromeric retrotransposon family in rice. Mol Biol Evol 22:845–855. doi:10.1093/molbev/msi069

Nakamura T, Miranda CH, Ohwaki Y, Valéio JR, Kim Y, Macedo M (2005) Characterization of nitrogen utilization by *Brachiaria* grasses in Brazilian Savannas (Cerrados). Soil Sci Plant Nutr 51(7):973–979

Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, Cardeno C, Koriabine M, Holtz-Morris AE, Liechty JD, Martínez-García PJ, Vasquez-Gross HA, Lin BY, Zieve JJ, Dougherty WM, Fuentes-Soriano S, Wu LS, Gilbert D, Marçais G, Roberts M, Holt C, Yandell M, Davis JM, Smith KE, Dean JF, Lorenz WW, Whetten W, Sederoff R, Wheeler N, McGuire PE, Main D, Loopstra CA, Mockaitis K, de Jong PJ, Yorke JA, Salzberg SL, Langley CH (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biol 4;15(3):R59

Nielen S, Almeida LM, Carneiro VT, Araujo ACG (2010) Physical mapping of rDNA genes corroborates allopolyploid origin in apomictic *Brachiaria brizantha*. Sex Plant Reprod 23(1):45–51

Nouroz F, Shumaila N, Heslop-Harrison JS (2015) Evolutionary genomics of miniature inverted-repeat transposable elements (MITEs) in *Brassica*. Mol Gen Genomics. doi:10.1007/s00438-015-1076-9

Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien MA (2010) Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol 186:37–45

Pearce SR, Li D, Flavell AJ, Harrison G, Heslop-Harrison JS, Kumar A (1996) The Ty1-copia group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. Mol Gen Genet MGG 250:305–315

Pellicer J, Fay MF, Leitch IJ (2010) The largest eukaryotic genome of them all? Bot J Linn Soc 164:10–15. doi:10.1111/j.1095-8339.2010.01072.x

Pereira V (2004) Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. Genome Biol 5(10):R79

Renny-Byfield S, Wendel JF (2014) Doubling down on genomes: polyploidy and crop plants. Am J Botf 101(10):1711–1725

Renvoize SA, Clayton WD, Kabuye CH (1996) Morphology, taxonomy, and natural distribution of *Brachiaria* (Trin.) Griseb. In: *Brachiaria*: biology, agronomy, and improvement CIAT publication no. 259. Miles JW, Maass BL, Valle CB, Kumble V (eds.) CIAT, Campo Grande, Brazil

Richards EJ, Ausubel FM (1988) Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. Cell 53(1):127–136

Schwarzacher T, Heslop-Harrison JS (2000) Practical in situ hybridization. Oxford: Bios. 203+xii pp

Takeda S, Sugimoto K, Otsuki H, Hirochika H (1998) Transcriptional activation of the tobacco retrotransposon Tto1 by wounding and methyl jasmonate. Plant Mol Biol 36:365–376

VanBuren R, Zeng F, Chen C, Zhang J, Wai CM, Han J, Aryal R, Gschwend AR, Wang J, Na JK, Huang L, Zhang L, Miao W, Gou J, Arro J, Guyot R, Moore RC, Wang ML, Zee F, Charlesworth D, Moore PH, Yu Q, Ming R (2015) Origin and domestication of papaya Yh chromosome. Genome Res 25:524–533

Verde I et al (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet 45: 487–494. doi:10.1038/ng.2586

Vicient CM, Jääskeläinen MJ, Kalendar R, Schulman AH (2001) Active retrotransposons are a common feature of grass genomes. Plant Physiol 125:1283–1292

Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, Barry K et al (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463: 763–768. doi:10.1038/nature08747

Weber B, Schmidt T (2009) Nested Ty3-gypsy retrotransposons of a single *Beta procumbens* centromere contain a putative chromodomain. Chromosom Res 17(3):379–396

Wicker T, Sabot F, Hua-Van A, Bennetzen J, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, San-Miguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982. doi:10.1038/nrg2165

Willing EM, Rawat V, Mandáková T, Maumus F, James GV, Nordström KJ, … & Schneeberger K (2015) Genome expansion of Arabis alpina linked with retrotransposition and reduced symmetric DNA methylation. Nature Plants, 1(2) Article number: 14023

Zhong CX, Marshalla JB, Topp C, Mroczek R, Kato A, Nagaki K, Birchler JA, Jiang J, Dawe RK (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 14:2825–2836. doi:10.1105/tpc.006106

Zonneveld BJM (2010) New record holders for maximum genome size in eudicots and monocots. J Bot 2010, 527357. doi:10.1155/2010/527357, 4 pages

Zou J, Fu D, Gong H, Qian W, Xia W, Pires JC, Li RY, Long Y, Mason AS, Yang T, Lim YP, Park BS, Meng J (2011) De novo genetic variation associated with retrotransposon activation, genomic rearrangements and trait variation in a recombinant inbred line population of *Brassica napus* derived from interspecific hybridization with *Brassica rapa*. Plant J 68:212–224