the plant journal

SEB Society for Experimental Biology

The Plant Journal (2019) **99,** 1066–1079 doi: 10.1111/tpj.14380

Genome invasion by a hypomethylated satellite repeat in Australian crucifer *Ballantinia antipoda*

Andreas Finke[1], Terezie Mandáková[2], Kashif Nawaz[1,3], Giang T. H. Vu[1,†], Petr Novák[4], Jiri Macas[4], Martin A. Lysak[2] and Ales Pecinka[1,3,*]

[1]*Max Planck Institute for Plant Breeding Research (MPIPZ), Cologne 50829, Germany,*

[2]*Plant Cytogenomics Research Group, CEITEC – Central-European Institute of Technology, Masaryk University, Brno 62500, Czech Republic,*

[3]*The Czech Academy of Sciences, Institute of Experimental Botany (IEB), Centre of the Region Haná for Agricultural and Biotechnological Research (CRH), Olomouc 77900, Czech Republic, and*

[4]*Biology Centre, The Czech Academy of Sciences, České Budejovice 37005, Czech Republic*

Received 7 July 2017; revised 2 April 2019; accepted 24 April 2019; published online 28 June 2019.

*For correspondence (email pecinka@ueb.cas.cz).

[†]Present address: Leibniz Institute of Plant Genetics and Crop Plant Research, Stadt Seeland, OT Gatersleben, 06466, Germany

SUMMARY

Repetitive sequences are ubiquitous components of all eukaryotic genomes. They contribute to genome evolution and the regulation of gene transcription. However, the uncontrolled activity of repetitive sequences can negatively affect genome functions and stability. Therefore, repetitive DNAs are embedded in a highly repressive heterochromatic environment in plant cell nuclei. Here, we analyzed the sequence, composition and the epigenetic makeup of peculiar non-pericentromeric heterochromatic segments in the genome of the Australian crucifer *Ballantinia antipoda*. By the combination of high throughput sequencing, graph-based clustering and cytogenetics, we found that the heterochromatic segments consist of a mixture of unique sequences and an A–T-rich 174 bp satellite repeat (*BaSAT1*). *BaSAT1* occupies about 10% of the *B. antipoda* nuclear genome in >250 000 copies. Unlike many other highly repetitive sequences, *BaSAT1* repeats are hypomethylated; this contrasts with the normal patterns of DNA methylation in the *B. antipoda* genome. Detailed analysis of several copies revealed that these non-methylated *BaSAT1* repeats were also devoid of heterochromatic histone H3K9me2 methylation. However, the factors decisive for the methylation status of *BaSAT1* repeats remain currently unknown. In summary, we show that even highly repetitive sequences can exist as hypomethylated in the plant nuclear genome.

Keywords: satellite repeats, heterochromatin, DNA methylation, comparative genomics, *Brassicaceae*.

INTRODUCTION

Repetitive sequences, including transposable elements (TEs) and satellite repeats, are ubiquitous components of eukaryotic genomes and have major effects on genome organization, evolution and gene regulation (Lisch, 2013; Mehrotra and Goyal, 2014). In flowering plants, repetitive DNA content varies from less than 10% in miniature genomes of highly specialized carnivorous plants *Utricularia gibba* and *Genlisea nigrocaulis* to 85% in maize (Schnable *et al.*, 2009; Ibarra-Laclette *et al.*, 2013; Vu *et al.*, 2015). The full spectrum and interplay of factors determining repetitive DNA content per genome remain unknown and represent part of the C-value enigma (Gregory, 2005). Many TEs produce their own proteins necessary for amplification, and particularly autonomous RNA transposons (retrotransposons), multiplying via a copy–paste mechanism, have been very successful in invading plant genomes ove short evolutionary times (Piegu *et al.*, 2006; Willing *et al.*, 2015). Recent studies have suggested that retrotransposons contain *cis*-regulatory sequences that are recognized by specific transcription factors and therefore link TE expression with the canonical gene regulatory pathways (Ito *et al.*, 2011; Cavrak *et al.*, 2014; Pietzenuk *et al.*, 2016). In contrast with TEs, which are often several kilobases long and dispersed in the genome, satellite DNAs form homogenous, up to mega base pair long, arrays consisting of a high copy number of typically shorter (150–400 bp) sequence motifs (Heslop-Harrison and Schwarzacher, 2011; Melters *et al.*, 2013; Garrido-Ramos,

2015). The distribution of satellite repeats varies along chromosomes. While the chromosome starts and ends with telomeric repeats, the position of other regions with high density of satellite repeats, including centromeres, nucleolar organizers (NORs) and heterochromatic knobs in some species, for example, maize (Gent *et al.*, 2014), is variable (Mehrotra and Goyal, 2014; Garrido-Ramos, 2015). With exception of ribosomal and telomeric repeats, satellite DNAs are less conserved and mostly specific for a single or few closely related species. The origin of satellites is not yet fully understood, but it has been shown that they can arise *de novo* or by amplification of short tandem repeat arrays already present in the genome as parts of retrotransposons or rDNA ITS sequences (Macas *et al.*, 2003, 2009). Satellite repeats are most likely to amplify via the combinatorial action of unequal crossing over, gene conversion, rolling circle amplification and/or replication slippage (Plohl *et al.*, 2008; Garrido-Ramos, 2015). Some satellite DNAs have essential functions including protection of chromosome ends by telomeres, acting as a platform for kinetochore binding by centromeres or producing high amounts of ribosomal RNA by NORs (Mehrotra and Goyal, 2014). Specialized satellite functions include the regulation of gene expression or effects on chromosome segregation via meiotic drive (Belele *et al.*, 2013; Dawe *et al.*, 2018). Another important function of satellites and other repeats is by creating sequence diversity, which accelerates formation of reproductive barriers (Garrido-Ramos, 2015).

Amplification of TEs, is opposed both epigenetically and genetically. In epigenetic silencing, repeat-derived transcripts are processed into small RNAs, this devalues them as templates for reverse transcription (Mari-Ordonez *et al.*, 2013) and guides the RNA-directed DNA methylation (RdDM) machinery to homologous sequences (reviewed in e.g. Matzke and Mosher, 2014). These regions will be DNA methylated in CG, CHG and CHH contexts (H = A, T or C), histone H3 lysine 9 dimethylated (H3K9me2) and transcriptionally repressed. The given epigenetic state is faithfully transmitted to the next generations and remains robust under various growth situations due to the meristematic silencing centers (Yadav *et al.*, 2009; Du *et al.*, 2012; Baubec *et al.*, 2014). At the same time, TEs are subject to fast mutagenesis via the deamination of methyl-cytosines, microdeletions and deletion-prone homologous recombination events (Devos *et al.*, 2002; Hawkins *et al.*, 2009; Hu *et al.*, 2011; Willing *et al.*, 2016). Although we assume that similar mechanisms control satellite repeats, it is yet to be elucidated if, and how, their proliferation is regulated and eventually suppressed. Data from maize suggest that satellite repeat arrays are less targeted by RdDM at least during vegetative development under ambient conditions (Gent *et al.*, 2014; Fu *et al.*, 2018).

While the distribution of repeats along chromosomes is variable, several common patterns can be observed among plant genomes. In species with small genomes (<500 Mbp/1C) and low repeat content, repetitive DNA forms typically a single major chromosomal cluster containing the centromeric satellite array flanked by pericentromeric regions rich in various TEs and satellite DNA (Ali *et al.*, 2005; Mandáková *et al.*, 2010; Seymour *et al.*, 2014; Vu *et al.*, 2015; Willing *et al.*, 2015). In addition, some species also form repeat-rich domains at the chromosome termini. (Peri)centromeric repeats and inactive rDNA repeats form compact microscopically visible nuclear membrane- or nucleolus-associated heterochromatic chromocenters (CCs), respectively (Fransz *et al.*, 2003, 2006). Increasing genome size, is usually associated with the presence of repeats in chromosome arms. In plants with small genomes, chromosomes often adopt a rosette-like organization during interphase (Fransz *et al.*, 2002), while in plants with large genomes they appear heterochromatic and are often organized with centromeres and telomeres clustered at opposite nuclear poles (Cowan *et al.*, 2001; Tiang *et al.*, 2012). Genomes of crucifers (*Brassicaceae*) with small genomes show the first type of heterochromatin distribution with minor differences caused by the presence of, for example, heterochromatic knobs (Lysak *et al.*, 2005; Mandáková and Lysak, 2008; Hay *et al.*, 2014; Fransz *et al.*, 2016). A remarkable exception in this pattern was found in the endemic Australian species *Ballantinia antipoda* (*B. antipoda*; Southern Shepherd's Purse) with a small genome ($2n = 12$; 1C ~472 Mbp), but with six heterochromatic segments (HSs) occupying up to the entire chromosome arm (Mandáková *et al.*, 2010; Majerová *et al.*, 2014) (Figure 1a).

RESULTS

A 174-bp satellite repeat is a principal component of HSs on *B. antipoda* chromosomes

We hypothesized that the HSs on *B. antipoda* chromosomes are formed by a specific highly amplified repeat. Therefore, we investigated the most abundant repetitive DNA by analyzing *B. antipoda* genomic shotgun reads using a RepeatExplorer pipeline (Novák *et al.*, 2013). The pipeline performs all-to-all pairwise similarity comparisons of sequence reads and identifies genomic repeats as clusters of frequently overlapping read sequences (Novák *et al.*, 2010). Clustering of 865 000 randomly sampled reads (≈0.2× the nuclear genome) resulted in 1000s of clusters ranging from two up to 71 000 reads, and therefore reflecting varying abundance of corresponding genomic sequences. We found 89 clusters, each containing at least 0.05% of the analyzed reads, that were considered to represent abundant repeats. The clusters corresponded to 49.2% of *B. antipoda* nuclear genome and were mostly

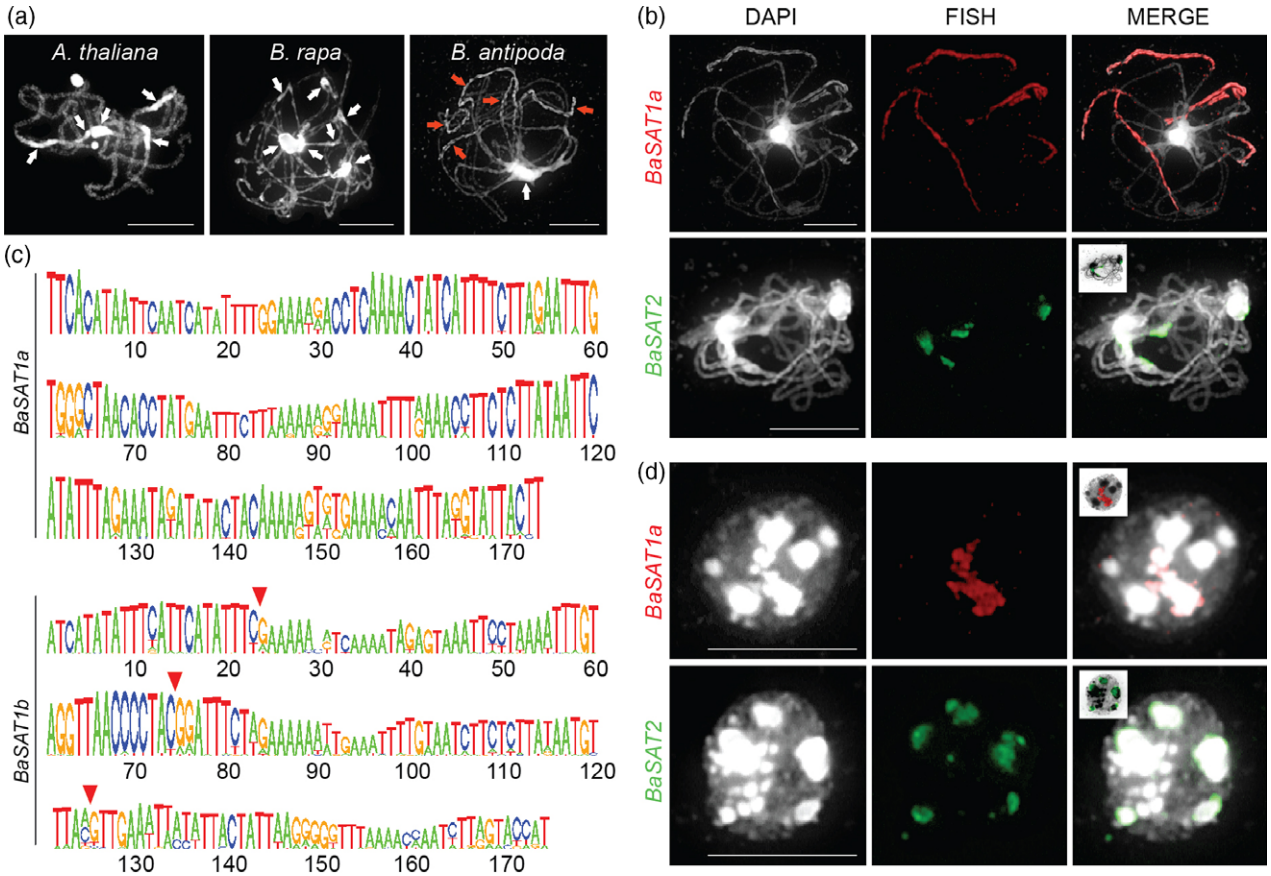

Figure 1. Localization and composition of heterochromatic segments in *B. antipoda*.
(a) Pachytene chromosomes of *Arabidopsis thaliana*, *Brassica rapa* and *B. antipoda* with indicated pericentromeric heterochromatin (white arrows) and *B. antipoda* heterochromatic segments (red arrows).
(b) FISH on *B. antipoda* pachytene chromosomes using probes against *BaSAT1a* (red) and *BaSAT2* (green). The probes had 95% and 94% sequence identity with the *BaSAT1a* and *BaSAT2* consensus sequences, respectively.
(c) Logo plot of the 174-bp monomer consensus sequences of the *BaSAT1a* and *BaSAT1b* repeats. Three CG dinucleotides in *BaSAT1b* are indicated by red triangles.
(d) FISH on *B. antipoda* interphase nuclei using probes against *BaSAT1a* (red) and *BaSAT2* (green). For more images see also Figure S4.All preparations in (a), (b) and (d) were counterstained with DAPI. Scale bars = 10 μm.

composed of LTR-retrotransposons and satellite DNA (16.5% and 15.9% of the genome, respectively, Table 1).

The two major satellite DNA families representing the primary candidates for the HS repeats were named *BaSAT1* and *BaSAT2* (Figures S1 and S2). These subfamilies were split into separate clusters in the RepeatExplorer analysis due to their sequence divergence (66% identity between the consensus sequences). The *BaSAT2* (5.65% of the genome) was composed of about 600-bp long monomers, which contained short arrays of partially degenerated telomeric motifs (TTTAGGG) (Figure S2). Fluorescence *in situ* hybridization (FISH) on the extended meiotic chromosomes using a *BaSAT2* specific probe labeled the middle regions of all six *B. antipoda* chromosomes (Figure 1b), suggesting that it is a (peri)centromeric repeat. The *BaSAT1*, with 174-bp monomers, comprised two distinct subfamilies designated as *BaSAT1a* and *BaSAT1b* (Figures 1c and S1a). The *BaSAT1a/b* probe

Table 1 Composition of the highly repetitive fraction of the *B. antipoda* genome. 'All' indicates the sum of a given repeat type within *B. antipoda* genome according to graph-based clustering

Repeat	Genome proportion (%)
Satellites (all)	(15.85)
BaSAT1a	8.19
BaSAT1b	1.94
BaSAT2	5.65
LTR-retrotransposons (all)	(16.47)
LTR/gypsy	
Athila	9.70
Chromovirus	1.94
LTR/copia	0.99
LTR/unclassified	3.84
DNA transposons (all)	(2.12)
Mutator	1.13
CACTA	0.99
rDNA	3.55
Unclassified repeats	11.23
Total	49.23

unambiguously labeled all six HSs (Figure 1b) and confirmed these repeats as the principal components of HSs. During interphase, *BaSAT1* formed a high number of mini-chromocenters (CCs) without any obvious peripheral localization (Figure 1d; see also Figure S4 interphase nuclei). Considering estimated genome proportions of *BaSAT1* repeats, their prevailing monomer length and haploid genome size (~472 Mbp), we estimated genomic copy numbers of *BaSAT1a* and *BaSAT1b* repeats to be approximately 212 000 and 50 000, respectively. The subfamilies made up 8.19% and 1.94% of the nuclear genome, respectively, making *BaSAT1* the most abundant repeat in *B. antipoda*. Detailed analysis of the *BaSAT1a* and *BaSAT1b* consensus sequences revealed that they are very A–T rich (76.4 and 75.3%; Figures 1b and S1). All cytosines in the *BaSAT1a* consensus sequence were in the CHH (H = C, A or T) context, while the *BaSAT1b* consensus sequence also contained three CG dinucleotides (Figures 1c and S1a).

BaSAT1 repeats are distributed in gene-rich chromosome regions

Comparative chromosome painting using *A. thaliana* gene-rich BAC probes revealed their hybridization to *B. antipoda* HSs (Mandáková *et al.*, 2010), indicating that HSs also contain single copy sequences. To get further insight into the organization of HSs, we combined the *BaSAT1* FISH probe with distinctly labeled *A. thaliana* BAC FISH probes from the bottom arm of chromosome 2 (evolutionary conserved block J), which mark homeologous regions on *B. antipoda* chromosomes 3 and 6 (Mandáková *et al.*, 2010), and hybridized them to *B. antipoda* pachytene chromosomes. Indeed, the BAC and *BaSAT1* signals alternated in a mosaic, proving that HSs contain a mixture of repetitive and evolutionary conserved single copy sequences (Figure 2a). We identified part of these sequences and their organization relative to *BaSAT1* repeats by high-throughput sequencing and *de novo* contig assembly of 115 million *B. antipoda* 100-bp single-end Illumina (San Diego, CA, USA) reads (24-fold genome coverage). This yielded 249 069 contigs consisting from at least two aligned reads. BLAST analysis, using the *BaSAT1* monomeric consensus sequence as the query sequence, revealed 179 contigs with a stretch of *BaSAT1* matching sequence. We excluded 31 contigs that consisted (almost) exclusively of *BaSAT1* repeats (Table S1), and additional 75 contigs, which contained also unique sequences but did not share a significant homology with *A. thaliana* genome (Table S2). The remaining 73 contigs (Table S3) mapped mainly to the euchromatic chromosome arm regions in *A. thaliana* genome (Figure 2b). There was a high concentration of the hits on the bottom arm of *A. thaliana* chromosome 2 and both arms of chromosome 5 (Figure 2b), which is consistent with the positions of HSs on chromosomes 2, 3 and 6 in *B. antipoda* (Mandáková *et al.*, 2010). PCR-based validation of 16 *in silico*-assembled contigs confirmed 13 cases (Figure S3a). Two contigs (c134934 and c217668) mapped with their unique sequence regions to the adjacent genomic positions in the *A. thaliana* genome, suggesting that they may be separated by a single *BaSAT1* repeat array in *B. antipoda*. Indeed, PCR using primers positioned in the unique *BaSAT1* flanking sequences consistently resulted in ~7 kb a product, validating that these contigs are indeed in the same genomic location (Figure S3b). In total, nine of these contigs could be roughly placed to *B. antipoda* chromosomes based on the homology with *A. thaliana* chromosomes (Figure 2b, red arrows). To estimate the position of *BaSAT1* repeats with respect to protein coding genes, we explored the 73 *B. antipoda BaSAT1* contigs showing homology to the *A. thaliana* genome. While, in 33 cases, sequence homology was limited to intergenic regions of *A. thaliana* genome, 40 contained sequences homologous to protein-coding genes. More detailed analysis of the latter cases revealed that 22, 14 and 4 *BaSAT1* sequence contigs were located upstream of the 5′ or downstream of the 3′ ends or directly in the coding region of a putative gene, respectively. The 22 *BaSAT1* copies found upstream of a gene were frequently located close to the translation start site (0–0.5 kb, $n = 12$; 0.5–1 kb, $n = 5$; 1–2.5 kb, $n = 5$; Table S1). Hence, *BaSAT1* satellite repeats were intermingled with single copy sequences and some occur close to, or even disrupt, protein-coding genes.

Most *BaSAT1* repeats are DNA hypomethylated

Repeats are silenced by repressive chromatin marks in plants (Matzke and Mosher, 2014). To explore epigenetic control of *BaSAT1* repeats, we analyzed their DNA methylation and histone modifications profiles. First, we assessed the global distribution of DNA methylation by 5-methyl-2-deoxy-cytosine (5mdC)-specific immunostaining. Contrary to our expectation, there was only a weak signal over *BaSAT1* HSs on pachytene chromosomes and also in CCs of *B. antipoda* nuclei (Figure 3a; Figure S4). We excluded that the lack of signals was due to technical issues because the pericentromeric heterochromatin within the same chromosomes showed strong and continuous staining, indicating ample DNA methylation at other genome regions (Figures 3a and S4).

This prompted us to analyze *BaSAT1* DNA methylation at a single nucleotide resolution level by dideoxy-sequencing of sodium bisulfite-treated DNA. We focused on seven contigs, consisting of *BaSAT1* repeats flanked by unique sequences (Tables S2 and S3), which we were able to amplify by PCR using a combination of unique and *BaSAT1a*-specific primers. The contigs c137937 and c217668 had all cytosines in CHH context, as predicted for the *BaSAT1a* consensus sequence (Figure 1c), but other contigs also contained cytosines in symmetrical context.

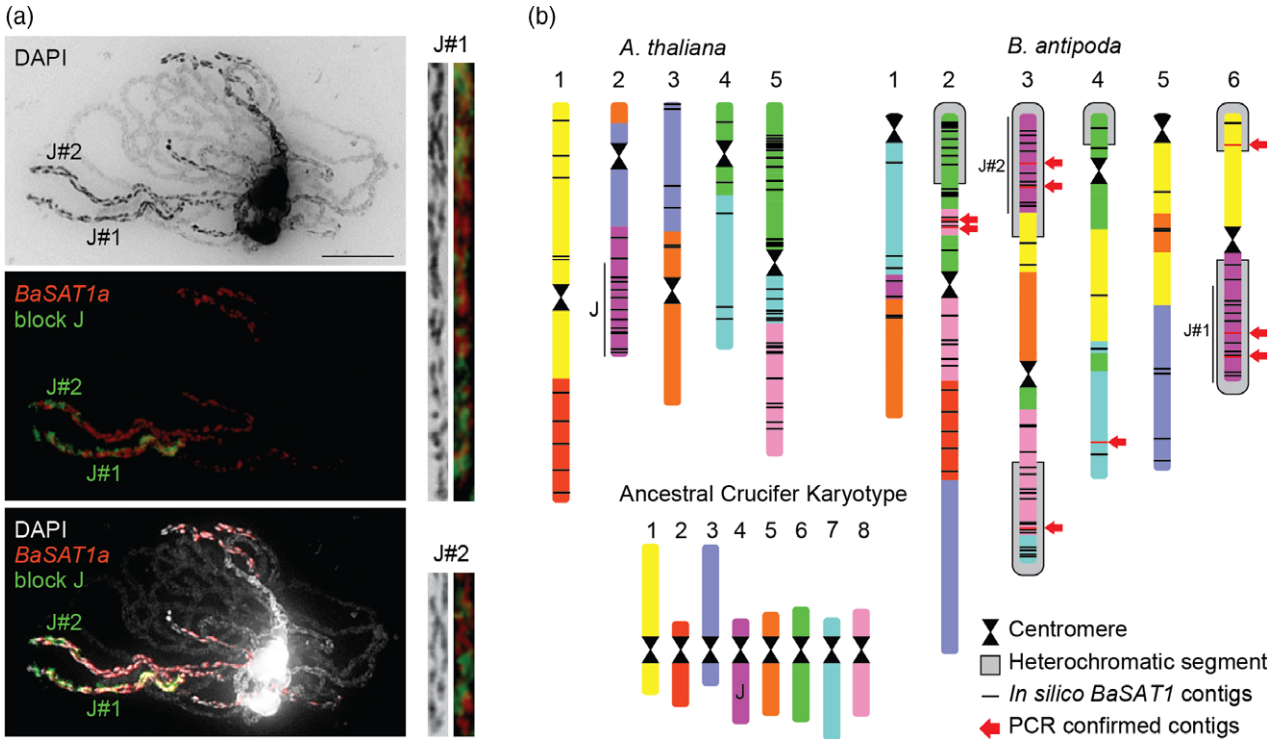

Figure 2. Genomic organization of heterochromatic segments (HSs) in *B. antipoda*.
(a) FISH on *B. antipoda* pachytene chromosomes using the *BaSAT1a* repeat (red) probe combined with comparative chromosome painting signal for ancient karyotype genomic block J (green). The block J appears twice due to the past polyploidization event. Chromosomes were counterstained with DAPI. The chromosomes in J#1 and J#2 were straightened using the 'straighten-curved-objects' plugin in the Image J software. Scale bar = 10 μm.
(b) Comparison of the extant karyotypes of *A. thaliana* and *B. antipoda* and the reconstructed ancestral karyotype (modified from Mandáková *et al.*, 2010). Homologous regions are indicated in the same color. Centromeres are depicted as black double-triangle structures and HSs as the gray expanded sectors below corresponding to parts of *B. antipoda* chromosomes. The genome locations of *in silico* reconstructed contigs containing *BaSAT1* are shown on *B. antipoda* and are homologous to *A. thaliana* chromosomes as black bars. Red bars indicate the position of contigs confirmed by PCR that were used for analysis of DNA methylation and histone modifications by chromatin immunoprecipitations.

The contigs c240383 and c213788 had additionally two CG sites, the contigs c118277 and c214317 had at least one cytosine in CHG context and the contig c97472 contained cytosines in all sequence contexts. Analysis of DNA methylation revealed that *BaSAT1* repeats were hypermethylated over their entire length in all contigs, except for c13721 and c217668, which were DNA hypomethylated (Figure 3b, c), suggesting that some *BaSAT1* copies may be indeed hypomethylated. We excluded this pattern to be tissue specific, as the same DNA methylation patterns were found in DNA extracted from leaves and flowers (Figure S5).

To get a representative picture of DNA methylation for more *BaSAT1* repeats, we performed DNA methylation analysis based on high-throughput bisulfite sequencing (BS-seq). The BS-seq reads were mapped to *de novo* assembled scaffolds on the *B. antipoda* genome, on which genes were predicted using Augustus software with support of *A. thaliana* TAIR10 genome annotation. This confirmed that *BaSAT1* repeats are indeed interspersed in genomic regions containing putative genes and may form complex arrays of monomers divided by spacers of variable length (Figures 4a–c and S6a). Analyzing DNA methylation over multiple genomic regions revealed that some of the putative genes contained only CG methylation, which resembled gene body methylation (Figures S6b–c and S7), while other regions contained DNA methylation in all sequence contexts (Figure 4a–c; Figure S6a). Next we looked for DNA methylation specifically in *BaSAT1* repeats. We found 39 778 *BaSAT1* repeats on the assembled genome, out of which 7742 repeats had four or more cumulative BS-seq reads mapping to given positions; Figure 4d). Because of the absence of high quality reference genome and high repetitiveness of *BaSAT1* repeats, we estimated DNA methylation to be the percentage of methylated versus non-methylated sequenced molecules at each cytosine position covered by at least four BS-seq reads. In *BaSAT1*, there were 7.5% (out of a total 16 757) CG positions methylated, for CHG context it was 5.4% (out of a total 13 082) and for CHH context 9.8% (out of a total 39 779). For comparison, we quantified DNA methylation at (peri)centromeric regions localized *BaSAT2* repeats, which appeared DNA methylated on meiotic spreads (Figure 3a). In whole assembled genome, we found 23 594 *BaSAT2* copies, out of which 12 465 had four or more

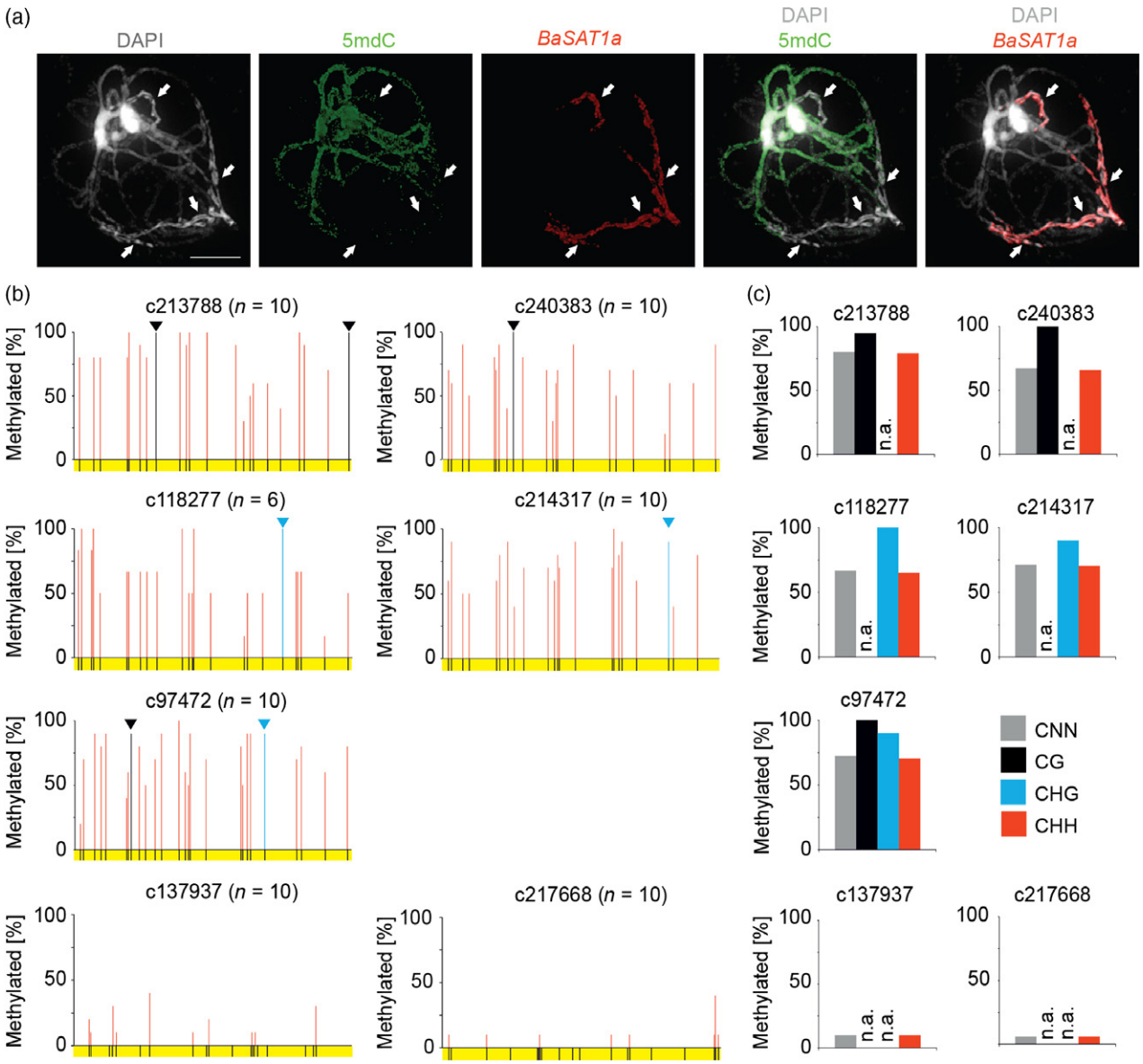

Figure 3. Analysis of DNA Methylation at *BaSAT1* repeats.
(a) Combination of immunostaining with 5mdC antibody (green) and FISH using a *BaSAT1a*-specific probe (red) on *B. antipoda* pachytene chromosomes. Heterochromatic segments are indicated by white arrows. Pachytenes were counterstained with DAPI. Scale bar = 10 μm.
(b) DNA methylation analysis of *BaSAT1* repeats (yellow, 174 bp) by bisulfite treatment followed by dideoxy sequencing. Each repeat was flanked by a unique sequence, into which one PCR primer was placed (see Experimental procedures). The positions of all cytosines in the reference sequence (irrespective of their methylation status) are indicated by the black bars in the yellow field. Quantitative data on the average, CG, CHG and CHH methylation are represented in gray, black, blue and red, respectively. CG and CHG methylation is further highlighted by black and blue triangles, respectively. The number of analyzed DNA molecules for each repeat is indicated as n behind the contig name.
(c) Quantitative data for (b). CNN shows the % of methylated cytosines irrespective of their sequence context; n.a., not applicable, cytosines at these contexts were absent.

cumulative BS-seq reads mapping at specific sites. For *BaSAT2*, there were 23.9% of CG, 20.0% of CHG and 27.4% of CHH methylated cytosines (out of a total 20 004, 19 345 and 23 595 sites, respectively), which is about three-fold more than for *BaSAT1* (all pairwise comparisons were *P*-value = 2.2E-16 in chi-squared tests; Figure 4d). For both repeats, there were no differences in frequency of DNA methylation at different strands (Figure 4d). Hence, also BS-seq data supported DNA hypomethylation of *BaSAT1* repeats. Next, we used these data also to look into the composition of *BaSAT1* arrays with respect to both subtypes. We performed BLAST analysis using *BaSAT1a* and *BaSAT1b* consensus sequences (Figure S1) and looked whether both types occurred separately or were

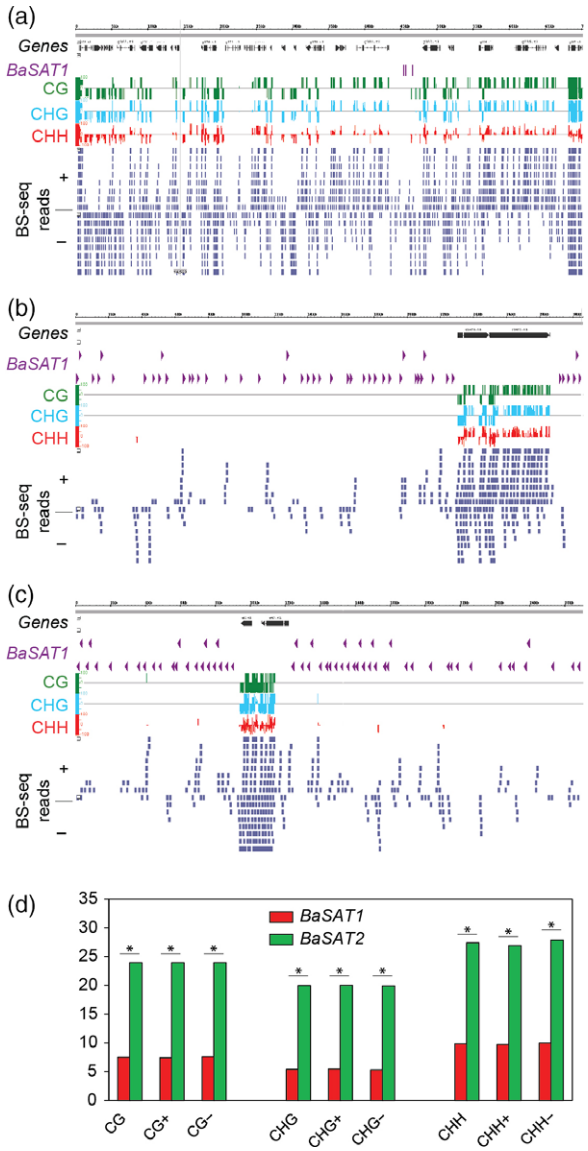

Figure 4. DNA methylation analysis by bisulfite sequencing (BS-seq).
(a–c) Examples of *B. antipoda* scaffolds with predicted putative genes (black), *BaSAT1* repeats (violet arrowheads) and DNA methylation information in sequence and strand-specific contexts. BS-seq reads show the coverage of the individual positions with sequencing reads on the Watson (+) and Crick (–) strands. Note that only some *BaSAT1* copies could be analyzed for DNA methylation due to limited unique mapping. (a) Shows the heavily methylated genomic region.
(b, c) Represent arrays of *BaSAT1* repeats with a single DNA methylated gene in each snapshot.
(d) Analysis of DNA methylation in *BaSAT1* and *BaSAT2* repeats based on BS-seq. We quantified the percentage of cytosine methylation in CG, CHG and CHH contexts on both and single (+ and –) DNA strands. The percentages correspond to methylated cytosine positions versus non-methylated ones. Each cytosine position had to be covered by at least four reads to be considered for analysis. All indicated comparisons were statistically significantly different (*) with a *P*-value = 2.2E-16 (chi-squared test).

intermingled. Visual inspection of multiple scaffolds revealed that, most of the time, *BaSAT1* subtypes do not intermingle (Figure S8) and only in one case we found a BaSAT1 array that also contained two *BaSAT1b* copies (Figure S8e).

Next, we scored for global distribution of heterochromatin- and euchromatin-specific modifications H3K9me2 and H3K4me3, respectively, in *B. antipoda* nuclei by immunostaining (Figure 5a). Intense H3K9me2 and H3K4me3 signals alternated and but a weaker H3K9me2 signal was dispersed also over the middle part of the flattened nuclei (see the overlapping images in Figure 5a). At this low resolution level, the *BaSAT1* FISH signals overlapped with both H3K4me3 and H3K9me2 signals (Figure 5a). To test this at finer scale, we determined the abundance of the H3K4me3 and H3K9me2 in specific *BaSAT1* repeats by chromatin immunoprecipitation (ChIP) along the selected contigs with known DNA methylation status (Figure 3b,c), plus contig c126293 containing a presumably euchromatic control locus *BaACTIN7* (Data S1). We found that the highly DNA methylated contigs c97472, c118277, c213788, c240383 were enriched for the H3K9me2 and depleted for the H3K4me3 modification, whereas the sparsely DNA methylated contigs c13721 and c217668 showed lower levels of H3K9me2 but high levels of H3K4me3 (Figure 5b). This suggests that individual copies of *BaSAT1* displayed either heterochromatic or euchromatic features.

BaSAT1-like sequences are found in several other Australian *Microlepidieae taxa*

Unusual features of *BaSAT1* raised our curiosity about its origin and via this possibly also its dynamics in the *B. antipoda* genome. Detailed investigations into the phylogeny of the Australian *Brassicaceae* recently led to the assignment of the monotypic genus *Ballantinia* to the tribe Microlepidieae, endemic to Australasia (Heenan *et al.*, 2012). To determine whether the *BaSAT1* repeats might have originated before divergence of the Microlepidieae genera, we performed PCRs using *BaSAT1* consensus sequence-specific primers on the DNA of eight additional species (representing eight different genera) of this tribe: *Arabidella eremigena*, *Blennodia canescens*, *Cuphonotus andraeanus*, *Drabastrum alpestre*, *Harmsiodoxa puberula*, *Menkea villosula*, *Phlegmatospermum richardsii* and *Stenopetalum nutans*, as well as of *A. thaliana*. Genomic BLASTs excluded the presence of *BaSAT1*-like sequences in *A. thaliana* and therefore we used this species as control. None of the PCRs yielded a regular ladder indicative of tandem repeats, but we obtained specifically 1–1.5 kb PCR products for *P. richardsii* and an approximately 5 kb product for *M. villosula* (Figure 6a). To analyze underlying sequences, we extracted, cloned and sequenced the 1.5 kb PCR amplicon of *P. richardsii*. This revealed that (among other sequences) the band contained satellite sequences. The monomer of *PrSAT1* resembled the *BaSAT1* repeat in terms of the average length (168 bp) and A–T content

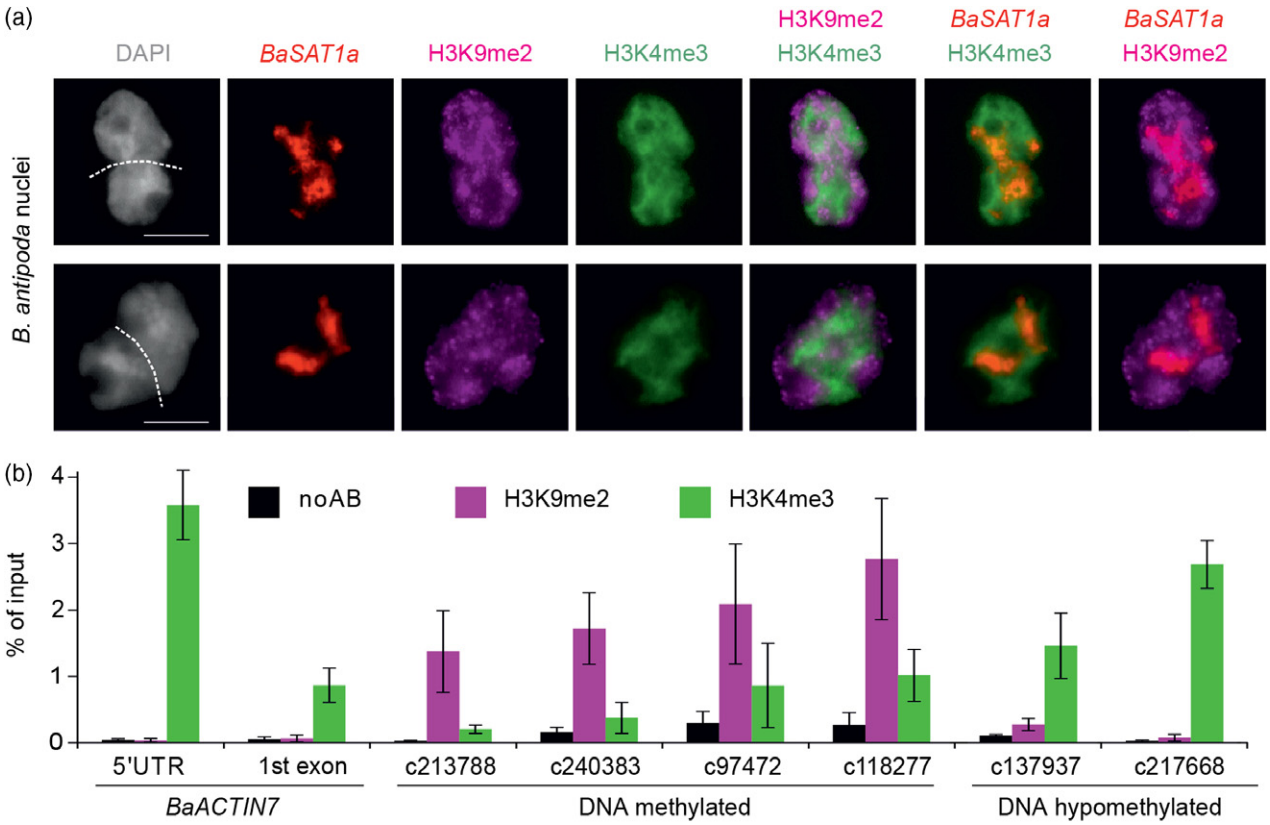

Figure 5. Histone modifications at *B. antipoda* heterochromatic segments.
(a) Immunostaining of *B. antipoda* nuclei with H3K9me2 (pink) and H3K4me3 antibodies (green) followed by FISH with *BaSAT1a* probe (red). Please note that both lanes showed two attached nuclei (attachment zone is indicated by the dashed line). Nuclei were counterstained with DAPI. Scale bars = 10 μm.
(b) Chromatin immunoprecipitation assaying abundance of H3K9me2 and H3K4me3 along the indicated *BaSAT1* contigs. Error bars indicate the standard deviation of two independent biological replicates. A putative *B. antipoda ACTIN 7* (*BaACTIN7*) was identified based on sequence homology to the *A. thaliana ACTIN 7* locus and used as a euchromatic control.

(79%). As observed for *BaSAT1*, a BLAST search for sequence homologs as well as the search against the PlantSAT database failed to identify repeats with a *PrSAT1* sequence similarity. Intraspecific comparison of the identified *PrSAT1* sequences revealed an average sequence similarity of 68% (ranging from 58 to 100%), which was close to the sequence variation found between *BaSAT1a* monomers (71% on average; Figures 6b and S9). FISH using the cloned *PrSAT1* sequence to *P. richardsii* mitotic chromosomes revealed one large and one small locus, suggesting that the *PrSAT1* sequences occupied specific chromosome regions in high densities, but did not span the entire chromosome arms as did *BaSAT1* repeats.

DISCUSSION

Using a combination of low coverage genome sequencing, graph-based clustering and FISH, we identified the *BaSAT1* satellite repeat (monomer 174 bp; ca. 10% of the nuclear genome; >200 000 copies) as the principal component of the peculiar HS in the *B. antipoda* genome. Based on the survey of tandem satellite repeats in 282 species from various kingdoms (Melters *et al.*, 2013), *BaSAT1* would be an ideal candidate for centromeric repeat sequences. However, the centromeric function of *BaSAT1* is not supported by its absence at (peri)centromeres of *B. antipoda* monocentric chromosomes, presence on both arms of chromosomes 3 and 6 (would cause dicentric chromosomes) and microscopically estimated absence on chromosomes 1 and 5 (would cause acentric chromosomes). Instead, the primary candidate for the centromeric sequence in *B. antipoda* is the second most abundant repeat *BaSAT2* with a 600-bp monomer length, which localizes to a (peri)centromeric region of all chromosomes. *BaSAT2* contains several Arabidopsis-type telomeric repeat motifs; this siituation most likely explains a strong localization of the Arabidopsis-derived telomeric repeat FISH signals within the (peri)centromeric regions of all *B. antipoda* chromosomes (Mandáková *et al.*, 2010; Majerová *et al.*, 2014).

The origin of *BaSAT1* is unclear and it is very likely to be species specific, a characteristic common to many satellite repeats (e.g. Kamm *et al.*, 1995; Ohmido *et al.*, 2000; Nouzová *et al.*, 2001). However, we found *BaSAT1*-like

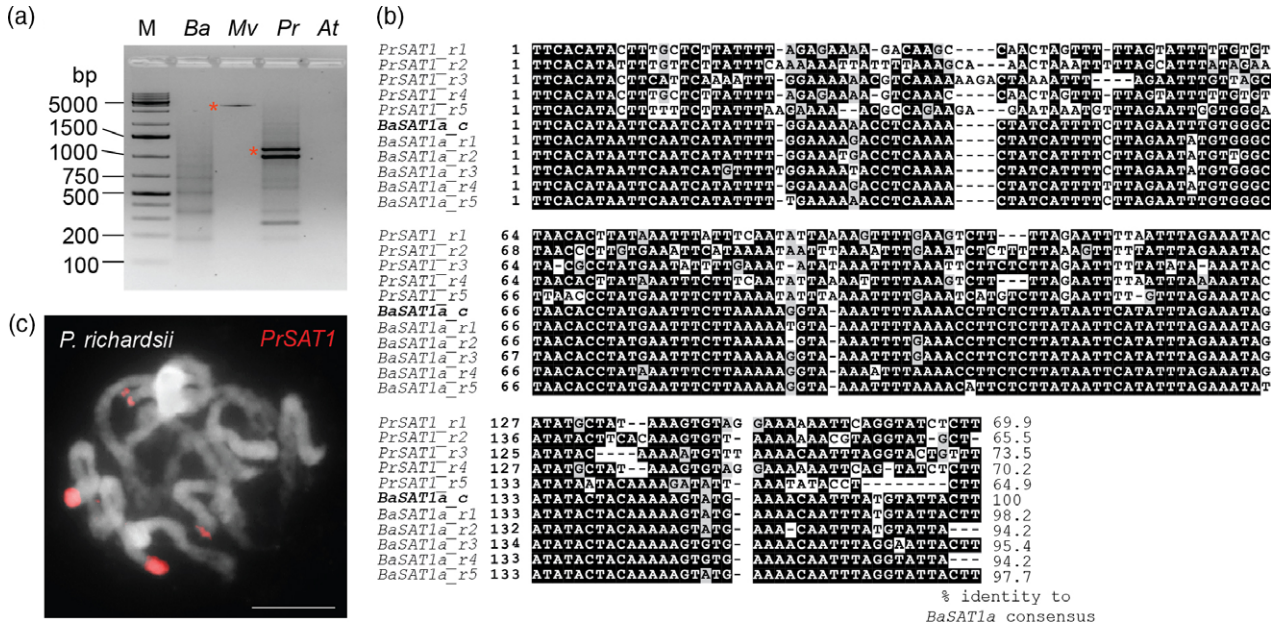

Figure 6. *BaSAT1*-like sequences in other Australian *Microlepidieae*.
(a) PCR using *BaSAT1a*-consensus sequence-derived primers and genomic DNA of *B. antipoda* (*Ba*), *Menkea villosula* (*Mv*), *Phlegmatospermum richardsii* (*Pr*) and *Arabidopsis thaliana* (*At*) control. Fragment sizes of the DNA marker (M) used are indicated. Asterisks indicate fragments, which were excised, cloned and sequenced.
(b) Shading of ClustalW2 alignments of five *PriSAT1* monomers of *P. richardsii* (*Pr_rep1* to *Pr_rep5*) and five *BaSAT1* monomers of *B. antipoda* (*Ba_rep1* to *Ba_rep5*) to the *BaSAT1*-consensus sequence. Shaded nucleotides were conserved in at least 50% of the aligned sequences. Identity of aligned sequences with the *BaSAT1*-consensus sequence is given after the alignment.
(c) FISH with *PrSAT1*-specific probe (red) to *P. richardsii* mitotic chromosomes counterstained with DAPI. Scale bar = 10 μm.

sequences (*PrSAT1*) occurring at two cytologically detectable genomic regions in *P. richardsii* among the eight tested species of tribe Microlepidieae. This finding suggests that *BaSAT1*-like repeats were present already in the common ancestor of at least some Australian Microlepidieae. However, at this point, we cannot exclude that other *BaSAT1*-like sequences, which are not amplified with our *BaSAT1* primers, exist in the nuclear genomes of other closely related genera.

Genomes of higher plants show a bias towards a higher A–T content, which ranges from approximately 53–67% (Barow and Meister, 2002; Lysak *et al.*, 2007; Šmarda *et al.*, 2012, 2014). The *BaSAT1* and *PrSAT1* repeats are very A–T rich (77 and 78.2%, respectively) and resemble the satellite *FriSAT1* (87% A–T) identified in the North American *Fritillaria* species (Ambrožová *et al.*, 2011). The *FriSAT1* is also present in very high copy numbers (>200 000) and can occupy large portions of the *Fritillaria* genomes, for example, up to 36% in *F. falcata*. Both *BaSAT1* and *FriSAT1* occur at many genomic positions rather than in a single or few arrays. The pattern of *BaSAT1* is even more intriguing, as it is scattered over evolutionary well conserved chromosome blocks and suggests that *BaSAT1* is capable of spreading by a currently unidentified mechanism. Speculatively, this could occur via reintegration of previously excised repeat arrays into new genomic positions and/or many microinversions, which would intermingle *BaSAT1* with gene-rich sequences. Whether and to what extent is the amplification of *BaSAT1* repeats and related sequences allowed by the duplicated nature of *Microlepidieae* genomes (Mandáková *et al.*, 2010; Mandáková *et al.*, 2017) remains currently unknown.

Although repeats are generally fully and stably DNA methylated in plants (Mathieu *et al.*, 2003; Stroud *et al.*, 2013), recent studies from *Brassicaceae* have suggested some species-specific variability, including a reduced degree of simultaneous methylation at both cytosines in symmetrical sites in *Arabis alpina* and lacking gene body methylation in *Eutrema salsugineum* and *Conringia planisiliqua* (Willing *et al.*, 2015; Bewick *et al.*, 2016). Here, *B. antipoda* may represent yet another example. Based on the intensity of immunostaining signals, HSs appeared to be only poorly DNA methylated when compared with euchromatic chromosome arms and pericentromeric regions. A similar phenotype was described, based on cytogenetic studies, for centromeric repeats of *A. thaliana*, *Beta vulgaris*, *Zea mays* and *Oryza sativa* (Zhang *et al.*, 2008; Yan *et al.*, 2010; Zakrzewski *et al.*, 2011, 2014); however, molecular analysis by bisulfite sequencing revealed that these repeats carried a good proportion of methylated cytosines (Zakrzewski *et al.*, 2011, 2014; Schmidt *et al.*, 2014). In contrast, we found by both immunostaining and bisulfite sequencing that *BaSAT1* repeats are hypomethylated. Only about 5–10% of cytosines (depending on the

context) were methylated in *BaSAT1*, while it was about three times more (20–27%) for (peri)centromeric *BaSAT2* repeats, which appeared DNA methylated in immunostaining. The answer to which factors determine the DNA methylation status of individual *BaSAT1* repeats remains unknown. The lack of DNA methylation at most *BaSAT1* repeats is not caused by defective DNA methylation pathways, but rather by their modulation. This is suggested by the presence of dense DNA methylation in all sequence contexts at multiple genomic regions flanking *BaSAT1* repeats and also gene body methylation at many putative genes. Speculatively, DNA methylation of *BaSAT1* repeats could be influenced by the genomic neighborhood of other (DNA methylated) repeats and/or transcription over the *BaSAT1* repeat arrays, leading to the production of small interfering RNAs. However, even if existing, such small RNAs are apparently not sufficient or not abundant enough to induce genome-wide *BaSAT1* DNA methylation. Analysis of the histone modifications on six individual *BaSAT1* repeats with contrasting DNA methylation patterns revealed that DNA methylated *BaSAT1* copies are marked by repressive histone modification H3K9me2 methylation, while the low methylated ones are enriched in the permissive modification H3K4me3. Surprisingly, we also observed the H3K9me2 mark at the two repeats, which lack cytosines in CHG context. H3K9me2 is directed to specific positions by the interaction between CMT3 CHG DNA methyltransferase and SuvH4/KYP histone methyltransferase (Du *et al.*, 2012). At present it is not clear whether *B. antipoda* uses an H3K9me2 establishment mechanism independent of CHG methylation or the presence of this methylation is simply an effect of spreading from the neighboring CHG containing *BaSAT1* copies. Furthermore, our data demonstrated that individual *BaSAT1* repeats carry either heterochromatic or euchromatic features. Hence, our observations challenge the paradigm of repetitive DNA hypermethylation, and show that even highly repetitive non-coding DNA sequences can adopt euchromatic-like features in plant nuclear genomes. In conclusion, the data suggest a differential use of epigenetic pathways to control tandem repeats versus transposons.

EXPERIMENTAL PROCEDURES

Plant materials and growth conditions

For the origin of the analyzed species accessions, see Table S4. Seed and plant material was provided by TM and MAL. For surface sterilization, *B. antipoda* seeds were incubated in 8% sodium hypochloride solution for 10 min and subsequently washed four times in distilled water and plated on ½MS medium supplied with 15 μM gibberellic acid (GA4 + 7). Plated seeds were kept at 4°C for 48–72 h and subsequently grown under long day conditions (16 h light, 8 h dark) at 21°C. Next, 3-week-old seedlings were transferred to soil and cultivated under long-day greenhouse conditions.

Nucleic acids isolation

DNA was prepared using the DNeasy plant mini kit (Qiagen, Hilden, Germany) or Phytopure Nucleon DNA isolation kit (GE Healthcare, Chicago, IL, USA). RNA was prepared using the RNeasy plant mini kit (Qiagen).

Next-generation sequencing

The sequencing library of *B. antipoda* was prepared from 1 μg genomic DNA with the TruSeq DNA kit (Illumina). Library quality was assessed on a Bioanalyzer (Agilent, Santa Clara, CA, USA). The library was sequenced in a single-end 101 nt read mode using a HiSeq 2500 instrument (Illumina). The reads were quality filtered and those containing parts of the adapter sequences were discarded using FAST-X tools (http://hannonlab.cshl.edu/fastx_toolkit/).

Identification and characterization of genomic repeats

Repeat identification by similarity-based clustering of Illumina reads was performed using local installation of the *RepeatExplorer* pipeline (Novák *et al.*, 2013), which was run on a Debian Linux server with 32 CPU cores and 64 GB RAM. In total, 1 115 000 reads were analyzed using default clustering parameter settings. The pipeline employs graphical representation of read similarities to identify clusters of frequently overlapping reads representing various repetitive elements or their parts (Novák *et al.*, 2010). In addition, it provides information about repeat quantities (estimated from the number of reads in a cluster) and outputs from BLASTn and BLASTx (Altschul *et al.*, 1990) similarity searches of our custom databases of repetitive elements and repeat-encoded conserved protein domains that aid in repeat annotation (Novák *et al.*, 2013). This information was combined and used for final manual annotation and quantification of repeats from all clusters, making up at least 0.05% of investigated genomes. Clusters containing plastid and mitochondrial sequences representing a contamination of nuclear DNA preparations by organellar DNA were excluded from the analysis, leaving 864 771 reads. Potential satellite repeats were identified based on the circular shapes of their cluster graphs (Novák *et al.*, 2010) and further analyzed using TAREAN tool of the RepeatExplorer that uses k-mer analysis of unassembled reads to reconstruct consensus sequences of tandem repeats (Novák *et al.*, 2017).

De novo assembly of *B. antipoda* scaffolds and contigs

For analysis of *BaSAT1* repeat distribution on *B. antipoda* chromosomes and for local bisulfite sequencing, we performed *de novo* contig assembly using trimmed single-end reads with CLC Genomics Workbench Software (Version 5.5), using the following parameters: word size: automatic, bubble size: automatic, minimum contig length: 200. The reads were mapped back to the contigs and mismatch, insertion and deletions were penalized with 2, 3 and 3, respectively. The length fraction was set to 0.5 and similarity fraction to 0.8.

For the whole genome BS-seq analysis, we performed an additional DNA-seq experiment. Here, 500 ng of *B. antipoda* genomic DNA were dissolved in 130 μL of EB buffer and fragmented to an average size of ca. 600 bp with the S2 focused ultrasonicator (Covaris Ltd, Brighton, UK) set to the following parameters: Intensity: 3, Duty Cycle: 5%, Cycles per Burst: 200, Treatment time: 70 sec. Subsequently, fragmented DNA was concentrated using Ampure XP magnetic beads and DNA-seq libraries were constructed using the NEBNext Ultra 2 DNA library prep kit (NEB, Cat. No. E7645S)

according to the manufacturer's instructions. From these libraries, 47 345 811 PE read of 250-bp length were obtained. We assembled scaffolds and contigs using the SOAPdenovo2 program, Version 2.04 (Luo *et al.*, 2012). We used filtered paired-end and single-end DNA-seq reads with k-mer size 101 and default parameters. In total, 2 293 915 scaffolds and contigs were assembled from all the reads. The scaffolds and contigs containing *BaSAT1a* and *BaSAT1b* (Figure S1a) repeats were identified using global alignment. First, the aligned sequence files were used to generate a motif matrix file of 174 bp through the MEME application of MEME Suite (Bailey *et al.*, 2009). Then the matrix file was used to scan for repeat locations throughout the 2 293 915 assembled *B. antipoda* scaffolds and contigs using another MEME Suite application FIMO (Grant *et al.*, 2011). This yielded 35 791 scaffolds and contigs with one or more (in total 84 587) *BaSAT1* repeat regions with $q_{val} \leq 10^{-4}$.

Chromosome preparation

Inflorescences of the analyzed accessions were fixed in ethanol: acetic acid (3:1) overnight and stored in 70% ethanol at −20°C. Selected inflorescences were rinsed in distilled water and in citrate buffer (10 mM sodium citrate, pH 4.8; 2 × 5 min) and incubated in an enzyme mix (0.3% cellulase, cytohelicase, and pectolyase; all Sigma-Aldrich) in citrate buffer at 37°C for 3–6 h. Individual flower buds were disintegrated on a microscope slide in a drop of citrate buffer and 15–30 μL of 60% acetic acid. The suspension was spread on a hot plate at 50°C for 0.5–2 min. Chromosomes were fixed by adding 100 μL of ethanol:acetic acid (3:1). The slide was dried with a hair dryer, post-fixed in 4% formaldehyde dissolved in distilled water for 10 min, and air dried. Chromosome preparations were treated with 100 μg/mL RNase in 2× sodium saline citrate (SSC; 20× SSC: 3 M sodium chloride, 300 mM trisodium citrate, pH 7.0) for 60 min and with 0.1 mg/mL pepsin in 0.01 M HCl at 37°C for 5 min; then post-fixed in 4% formaldehyde in 2× SSC, and dehydrated in an ethanol series (70, 90, and 100%, 2 min each).

Fluorescence *in situ* hybridization

Satellite repeats of *Ballantinia* (*BaSAT1* and *BaSAT2*) and *Phlegmatospermum* (*PrSAT1*), and *Arabidopsis thaliana* BAC clones corresponding to genomic block J of the Ancestral Crucifer Karyotype (ACK; Schranz *et al.*, 2006; Lysak *et al.*, 2016) were used as FISH probes. All DNA probes were labeled with biotin-dUTP or digoxigenin-dUTP by nick translation as described (Mandáková and Lysak, 2016). Selected labeled DNA probes were pooled together, ethanol precipitated, dissolved in 20 μL of 50% formamide, 10% dextran sulfate in 2× SSC and pipetted onto microscopic slides. The slides were heated at 80°C for 2 min and incubated at 37°C overnight. Post-hybridization washing was performed in 20% formamide in 2× SSC at 42°C. Hybridized probes were visualized through fluorescently labeled antibodies against biotin-dUTP and digoxigenin-dUTP (Mandáková and Lysak, 2016). Chromosomes were counterstained with 4,6-diamidino-2-phenylindole (DAPI, 2 μg/mL) in Vectashield antifade. Fluorescence signals were analyzed and photographed using a Zeiss Axioimager epifluorescence microscope and a CoolCube camera (MetaSystems, Heidelberg, Germany). Individual images were merged and processed using Photoshop CS software (Adobe Systems, San Jose, CA, USA). Pachytene chromosomes in Figure 2 were straightened using the 'straighten-curved-objects' plugin in the Image J software (Kocsis *et al.*, 1991).

5-Methyl-2′-deoxy-cytosine (5mdC) immunodetection

For immunostaining of 5mdC, standard chromosome preparations (see above) were used. A denaturation mixture containing 20 μL of 50% formamide, 10% dextran sulfate in 2× SSC was pipetted onto microscopic slides. The slides were heated at 80°C for 2 min and washed in 2× SSC (2 × 5 min). Slides were blocked for 30 min with 5% BSA solution (5% bovine serum albumin, 0.2% Tween-20 in 4× SSC) at 37°C for 30 min and then incubated with 100 μL of primarily antibody against 5mC (diluted 1:100, Diagenote) at 37°C for 30 min. After washing two times in 2× SSC the primary antibody was detected with the secondary antibody coupled with AlexaFluor488 (diluted 1:200, Invitrogen) at 37°C for 30 min followed by washing two times in 2× SSC and a dehydration in an ethanol series (70, 90, and 100%, 2 min each). Chromosomes were counterstained with DAPI, fluorescence signals analyzed and photographed, and slides rehybridized by satellite probes as described above.

Histone immunolabeling

Leaf tissue (0.5–1 g) with 0.5 mL of NIB buffer (10 mM Tris−HCl, 10 mM EDTA, 100 mM KCl, 0.5 M sucrose, 4 mM spermidine, 1 mM spermine, 0.1% 2-mercaptoethanol) was placed into a Petri dish on ice and chopped to a fine suspension with the razor blade. The suspension was pipetted into an Eppendorf tube, fixed in an equal volume of 4% paraformaldehyde on ice for 20 min, filtered through 50 and 20 μm mesh filters and centrifuged at 595 ***g*** at 4°C for 3 min. The supernatant was removed and the pellet with nuclei resuspended in 40 μL of NIB. Then, 2 μL of the suspension were pipetted onto a slide, dried at 4°C for 1 h and post-fixed in 4% formaldehyde in 2× SSC for 30 min. Slides were blocked for 30 min with 5% BSA solution at 37°C and then incubated with 100 μL of primarily antibodies against H3K4met3 and H3K9met2 (diluted 1:100, Abcam, Cambridge, UK) at 37°C for 2 h. After washing two times in 2× SSC the primary antibodies were detected with the secondary antibodies coupled with AlexaFluor488 (diluted 1:200, Invitrogen) and Cy5 (diluted 1:100, Jackson ImmunoResearch) at 37°C for 30 min followed by washing two times in 2× SSC. Chromosomes were counterstained with DAPI, fluorescence signals analyzed and photographed, and slides re-hybridized by satellite probes as described above.

DNA methylation analysis

For local DNA methylation analysis, 150–200 ng of *B. antipoda* genomic DNA was treated with sodium bisulfite using the Epitect Bisulfite Kit (Qiagen). PCR fragments were amplified for 32–35 cycles using MethylTaq DNA polymerase (Diagenode, Seraing, Belgium) according to manufacturers' recommendations, purified with QIAQuick PCR purification kit (Qiagen) and cloned into the pJet1.2 vector using the ClonJet PCR cloning kit (Thermo Scientific, Waltham, MA, USA). Colony PCR was performed to identify positive clones and the positive plasmids were isolated using the NucleoSpin Plasmid Mini Prep Kit (Macherey Nagel, Düren, Germany) and sequenced on an 3730XL Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). The sequences were trimmed, aligned with the ClustalW algorithm and analyzed using CyMATE (Hetzl *et al.*, 2007).

For genome-wide DNA methylation analysis, 1000 ng of *B. antipoda* genomic DNA was isolated and provided to the Max Planck Genome Centre in Cologne, Germany (https://mpgc.mpipz.mpg.de/home/) for library construction. The genomic DNA was sheared to fragments of ca. 400 bp using a S2 focused

ultrasonicator (Covaris Inc., Brigthon). A BS-seq library was constructed using the Bioo Scientific NEXTFLEX® Bisulfite Library Prep Kit (Perkin Elmer, Waltham, MA, USA) according to manufacturer's recommendations. The library was sequenced as 150-bp long paired-end reads. The reads were mapped to 35 791 *BaSAT1* repeats containing scaffolds and contigs using Bismark aligner software (Krueger and Andrews, 2011). Then, we used the Bismark methylation extractor (Krueger and Andrews, 2011) for strand-specific identification of methylated cytosines. This software yielded a bedgraph file of 5mdC, in which each methylation was reported in terms of location, context, and frequency. The scaffolds contained in total 84 587 *BaSAT1* repeats (i.e. often multiple repeats per one scaffold). For DNA methylation analysis, we identified cytosines in BaSAT1 and BaSAT2 repeats covered by at least four BS-seq reads, which resulted in 7742 and 12 463 analyzable *BaSAT1* and BaSAT2 copies, respectively. The percentage of methylated and non-methylated positions was calculated for each cytosine and then summed up for the whole repeat.

ChIP

ChIP was done as described (Gendrel *et al.*, 2005) with modifications: 3 g of leaves of 12-week-old soil grown plants were harvested. Crosslinking was performed in 37 mL of 1% (w/v) formaldehyde solution under vacuum for 20 min and subsequently quenched with 2.5 mL of 2 M glycine solution (final concentration 0.125 M) under vacuum for 7 min. Crosslinked material was snap frozen in liquid nitrogen, homogenized under liquid nitrogen, suspended in 30 mL, filtered through four layers of Miracloth and subsequently centrifuged for at 2000 ***g*** for 20 min at 4°C. After resuspension of the pellet in 1 mL of extraction buffer 2 the solution was centrifuged at 17 000 ***g*** for 15 min at 4°C. The resulting pellet was resuspended in 300 μL of extraction buffer 3, overlayed on 300 μL of extraction buffer 3 and centrifuged at 17 000 ***g*** for 1 h at 4°C. This step was repeated once. The nuclei pellet was suspended in 300 μL of ice-cold nuclei lysis buffer and chromatin was sheared at 4°C using a Diagenode Disruptor for six cycles with 30 sec of high energy sonication and a 30 sec break. Subsequently a centrifugation at 17 000 ***g*** for 10 min at 4°C was performed to remove nuclear debris. An aliquot of the chromatin extract was set aside to serve as the input control. The remaining extract was diluted 1:10 in ChIP dilution buffer. The chromatin solution was divided into four aliquots, 40 μL of Protein A Magnetic Sepharose beads (GE Healthcare) per mL were added and incubated for 45 min at 4°C with slight agitation. Subsequently the solution was centrifuged at 12 000 ***g*** for 30 sec at 4°C and the supernatant was transferred to a fresh tube. Three microlitres of the following antibodies were added were added to the respective tubes. H3K4me3: #39159 Histone H3 trimethyl Lys4 Rabbit pAB, Activemotif, H3K9me2: #720092 dimethyl-histone H3 Lys9 pAB, (Invitrogen). One aliquot served as the no Ab control. The immunoprecipitation reaction was incubated overnight at 4°C under slight agitation. After incubation, 40 μL/mL of Protein A Magnetic Agarose beads were added the solution, incubated for 1 h and centrifuged for 10 min at 12 000 ***g*** to pellet the beads. Beads were washed twice 5 min each with low or high salt buffer (150 and 500 mM NaCL, respectively; plus 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris–HCl (pH 8.1)); LiCl wash buffer: 0.25 LiCl, 1% NP40, 1% Na deoxycholate, 1 mM EDTA, 10 mM Tris–HCl (pH 8.1) and TE buffer: 10 mM Tris–HCl (pH 8.0), 1 mM EDTA. The DNA was eluted twice by incubation with 250 μL elution buffer (Qiagen) at 65°C for 15 min.

Quantitative (real-time) PCR was performed using the QPCR Green Master Mix Fluorescein Kit (BiotechRabbit, BR0501203, Berlin, Germany) in 12 μL QPCR reaction according to manufacturer's protocols. The samples were amplified using a CFX384 Touch real-time PCR Detection System (Bio-Rad, Hercules, CA, USA), and quantified with a calibration line made with DNA isolated from crosslinked, sonicated chromatin. With all experiments, no-template controls, No Ab controls and input samples were taken along for every primer set used. As the control, abundance of the respective histone modifications at the 5′UTR and the first exon of a putative *B. antipoda Actin7* (*BaActin7*) gene was assayed. The *BaActin7* sequence was identified by BLAST analysis of the *B. antipoda* contig library using the *A. thaliana* Actin7 gene as query.

Primers

All primers and oligonucleotides used in this study are defined in Table S5.

Sequence deposition

Adapter-trimmed raw sequencing reads generated in this study are deposited at the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under the accession number PRJEB21350 (for the content see Table S6).

ACKNOWLEDGEMENTS

We thank O. Mittelsten Scheid for numerous discussions, B. Eilts, P. Pecinkova and R. Gentges for technical assistance. We acknowledge the Millenium Seed Bank Project (Royal Botanic Gardens, Kew, UK) for providing seeds of selected Microlepidieae species. This work was supported by the grants from German Research Foundation in the frame of SPP Adaptomics to A.P. and K.N. (grant no. PE1853/2), Purkynĕ fellowship to A.P (grant no. KAV-2861/OPV/2017), research grants from the Czech Science Foundation awarded to JM, MAL and TM (grant nos. P501/12/G090 and 17-13029S), and by the CEITEC 2020 (grant no. LQ1601) project.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTIONS

GTHV prepared samples for genome sequencing. AF performed all molecular and TM all cytogenetic experiments. PN and JM performed graph-based clustering analysis and generated repeat consensus sequences. KN performed genome assembly, *in silico* repeat analysis and analyzed bisulfite sequencing data. AP, AF, MAL and JM designed the experiments and wrote the manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Consensus sequences of the *BaSAT1a* and *BaSAT1b* satellite repeats.

Figure S2. Consensus sequence of the *BaSAT2* repeat.

Figure S3. PCR confirmation of *BaSAT1* containing contigs.

Figure S4. Immunostaining reveals low DNA methylation of *BaSAT1* repeats.

Figure S5. *BaSAT1* bisulfite sequencing analysis at different developmental stages.

Figure S6. Examples of DNA methylation in the *B. antipoda* genome.

Figure S7. Additional examples of DNA methylation in the *B. antipoda* genome.

Figure S8. *BaSAT1a* and *BaSAT1b* repeats are organized into separate arrays.

Figure S9. *PrSAT1* sequence analysis.

Table S1. *BaSAT1* repeat only contigs.

Table S2. Unchanchored *BaSAT1* contigs.

Table S3. Homeology of *BaSAT1* contigs to the *A. thaliana* genome.

Table S4. Origin of materials from Australian Brassicaceae species used in this study.

Table S5. Oligonucleotides used in this study.

Table S6. Content of the dataset PRJEB21350 deposited at the European Nucleotide Archive (http://www.ebi.ac.uk/ena).

Data S1. Sequences of assembled contigs (fasta-format).

REFERENCES

Ali, H.B.M., Lysak, M.A. and Schubert, I. (2005) Chromosomal localization of rDNA in the Brassicaceae. *Genome*, **48**, 341–346.

Altschul, S., Gish, W., Miller, W., Myers, E. and Lipman, D. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Ambrožová, K., Mandáková, T., Bureš, P., Neumann, P., Leitch, I.J., Koblížková, A., Macas, J. and Lysak, M.A. (2011) Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of Fritillaria lilies. *Ann. Bot.* **107**, 255–268.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. (2009) MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208.

Barow, M. and Meister, A. (2002) Lack of correlation between AT frequency and genome size in higher plants and the effect of nonrandomness of base sequences on dye binding. *Cytometry*, **47**, 1–7.

Baubec, T., Finke, A., Mittelsten Scheid, O. and Pecinka, A. (2014) Meristem-specific expression of epigenetic regulators safeguards transposon silencing in Arabidopsis. *EMBO Rep.* **15**, 446–452.

Belele, C.L., Sidorenko, L., Stam, M., Bader, R., Arteaga-Vazquez, M.A. and Chandler, V.L. (2013) Specific tandem repeats are sufficient for paramutation-induced trans-generational silencing. *PLoS Genet.* **9**, e1003773.

Bewick, A.J., Ji, L., Niederhuth, C.E. *et al.* (2016) On the origin and evolutionary consequences of gene body DNA methylation. *Proc. Natl Acad. Sci. USA*, **113**, 9111–9116.

Cavrak, V.V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L.M. and Mittelsten Scheid, O. (2014) How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genet.* **10**, e1004115.

Cowan, C.R., Carlton, P.M. and Cande, W.Z. (2001) The polar arrangement of telomeres in interphase and meiosis. Rabl organization and the bouquet. *Plant Physiol.* **125**, 532–538.

Dawe, R.K., Lowry, E.G., Gent, J.I. *et al.* (2018) A Kinesin-14 motor activates neocentromeres to promote meiotic drive in maize. *Cell*, **173**, 839–850.e18.

Devos, K.M., Brown, J.K.M. and Bennetzen, J.L. (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. *Genome Res.* **12**, 1075–1079.

Du, J., Zhong, X., Bernatavichute, Y.V. *et al.* (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell*, **151**, 167–180.

Fransz, P., de Jong, J.H., Lysak, M., Castiglione, M.R. and Schubert, I. (2002) Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl Acad. Sci. USA*, **99**, 14584–14589.

Fransz, P., Soppe, W. and Schubert, I. (2003) Heterochromatin in interphase nuclei of Arabidopsis thaliana. *Chromosome Res.* **11**, 227–240.

Fransz, P., ten Hoopen, R. and Tessadori, F. (2006) Composition and formation of heterochromatin in Arabidopsis thaliana. *Chromosome Res.* **14**, 71–82.

Fransz, P., Linc, G., Lee, C.R. *et al.* (2016) Molecular, genetic and evolutionary analysis of a paracentric inversion in Arabidopsis thaliana. *Plant J.* **88**, 159–178.

Fu, F.-F., Dawe, R.K. and Gent, J.I. (2018) Loss of RNA-directed DNA methylation in maize chromomethylase and DDM1-type nucleosome remodeler mutants. *Plant Cell*, **30**, 1617–1627.

Garrido-Ramos, M.A. (2015) Satellite DNA in plants: more than just rubbish. *Cytogenet Genome Res.* **146**, 153–170.

Gendrel, A.-V., Lippman, Z., Martienssen, R. and Colot, V. (2005) Profiling histone modification patterns in plants using genomic tiling microarrays. *Nat. Methods*, **2**, 213–218.

Gent, J.I., Madzima, T.F., Bader, R., Kent, M.R., Zhang, X., Stam, M., McGinnis, K.M. and Dawe, R.K. (2014) Accessible DNA and relative depletion of H3K9me2 at maize loci undergoing RNA-directed DNA methylation. *Plant Cell*, **26: 4903**, LP-4917.

Grant, C.E., Bailey, T.L. and Noble, W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.

Gregory, T.R. (2005) The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. *Ann. Bot.* **95**, 133–146.

Hawkins, J.S., Proulx, S.R., Rapp, R.A. and Wendel, J.F. (2009) Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proc. Natl Acad. Sci. USA*, **106**, 17811–17816.

Hay, A.S., Pieper, B., Cooke, E. *et al.* (2014) Cardamine hirsuta: a versatile genetic system for comparative studies. *Plant J.* **78**, 1–15.

Heenan, P.B., Goeke, D.F., Houliston, G.J. and Lysak, M.A. (2012) Phylogenetic analyses of ITS and rbcL DNA sequences for sixteen genera of Australian and New Zealand Brassicaceae result in the expansion of the tribe Microlepidieae. *Taxon*, **61**, 970–979.

Heslop-Harrison, J.S. and Schwarzacher, T. (2011) Organisation of the plant genome in chromosomes. *Plant J.* **66**, 18–33.

Hetzl, J., Foerster, A.M., Raidl, G. and Mittelsten Scheid, O. (2007) CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. *Plant J.* **51**.

Hu, T.T., Pattyn, P., Bakker, E.G. *et al.* (2011) The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nat. Genet.* **43**, 476–481.

Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G. *et al.* (2013) Architecture and evolution of a minute plant genome. *Nature*, **498**, 94–98.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I. and Paszkowski, J. (2011) An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature*, **472**, 115–119.

Kamm, A., Galasso, I., Schmidt, T. and Heslop-Harrison, J.S. (1995) Analysis of a repetitive DNA family from Arabidopsis arenosa and relationships between Arabidopsis species. *Plant Mol. Biol.* **27**, 853–862.

Kocsis, E., Trus, B.L., Steer, C.J., Bisher, M.E. and Steven, A.C. (1991) Image averaging of flexible fibrous macromolecules: the clathrin triskelion has an elastic proximal segment. *J. Struct. Biol.* **107**, 6–14.

Krueger, F. and Andrews, S.R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, **27**, 1571–1572.

Lisch, D. (2013) How important are transposons for plant evolution? *Nat. Rev. Genet.* **14**, 49–61.

Luo, R., Liu, B., Xie, Y. *et al.* (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*, **1**, 18.

Lysak, M.A., Koch, M.A., Pecinka, A. and Schubert, I. (2005) Chromosome triplication found across the tribe Brassiceae. *Genome Res.* **15**, 516–525.

Lysak, M.A., Cheung, K., Kitschke, M. and Bureš, P. (2007) Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. *Plant Physiol.* **145**, 402–410.

Lysak, M.A., Mandáková, T. and Schranz, M.E. (2016) Comparative paleogenomics of crucifers: ancestral genomic blocks revisited. *Curr. Opin. Plant Biol.* **30**, 108–115.

Macas, J., Navrátilová, A. and Mészáros, T. (2003) Sequence subfamilies of satellite repeats related to rDNA intergenic spacer are differentially amplified on *Vicia sativa* chromosomes. *Chromosoma*, **112**, 152–158.

Macas, J., Koblížková, A., Navrátilová, A. and Neumann, P. (2009) Hypervariable 3′ UTR region of plant LTR-retrotransposons as a source of novel satellite repeats. *Gene*, **448**, 198–206.

Majerová, E., Mandáková, T., Vu, G.T.H., Fajkus, J., Lysak, M.A. and Fojtová, M. (2014) Chromatin features of plant telomeric sequences at terminal vs. internal positions. *Front. Plant Sci.* **5**, 593.

Mandáková, T. and Lysak, M.A. (2008) Chromosomal phylogeny and karyotype evolution in x = 7 crucifer species (Brassicaceae). *Plant Cell*, **20**, 2559–2570.

Mandáková, T. and Lysak, M.A. (2016). Chromosome preparation for cytogenetic analyses in Arabidopsis. In *Curr Protocols Plant Biol*. John Wiley & Sons, Inc., pp. 43–51.

Mandáková, T., Joly, S., Krzywinski, M., Mummenhoff, K. and Lysak, M.A. (2010) Fast diploidization in close mesopolyploid relatives of Arabidopsis. *Plant Cell*, **22**, 2277–2290.

Mandáková, T., Pouch, M., Harmanová, K., Zhan, S.H., Mayrose, I. and Lysak, M.A. (2017) Multispeed genome diploidization and diversification after an ancient allopolyploidization. *Mol. Ecol.* **26**, 6445–6462.

Mari-Ordonez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V. and Voinnet, O. (2013) Reconstructing de novo silencing of an active plant retrotransposon. *Nat. Genet.* **45**, 1029–1039.

Mathieu, O., Jasencakova, Z., Vaillant, I., Gendrel, A.-V., Colot, V., Schubert, I. and Tourmente, S. (2003) Changes in 5S rDNA chromatin organization and transcription during heterochromatin establishment in Arabidopsis. *Plant Cell*, **15**, 2929–2939.

Matzke, M.A. and Mosher, R.A. (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**, 394–408.

Mehrotra, S. and Goyal, V. (2014) Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. *Genomics Proteomics Bioinformatics*, **12**, 164–171.

Melters, D.P., Bradnam, K.R., Young, H.A. *et al*. (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* **14**, R10–R10.

Nouzová, M., Neumann, P., Navrátilová, A., Galbraith, D.W. and Macas, J. (2001) Microarray-based survey of repetitive genomic sequences in *Vicia* spp. *Plant Mol. Biol.* **45**, 229–244.

Novák, P., Neumann, P. and Macas, J. (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics*, **11**, 378.

Novák, P., Neumann, P., Pech, J., Steinhaisl, J. and Macas, J. (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*, **29**, 792–793.

Novák, P., Robledillo, L.A., Koblížková, A., Vrbová, I., Neumann, P. and Macas, J. (2017) TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res.* **45**, e111.

Ohmido, N., Kijima, K., Akiyama, Y., de Jong, J.H. and Fukui, K. (2000) Quantification of total genomic DNA and selected repetitive sequences reveals concurrent changes in different DNA families in indica and japonica rice. *Mol. Gen. Genet.* **263**, 388–394.

Piegu, B., Guyot, R., Picault, N. *et al*. (2006) Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. *Genome Res.* **16**, 1262–1269.

Pietzenuk, B., Markus, C., Gaubert, H., Bagwan, N., Merotto, A., Bucher, E. and Pecinka, A. (2016) Recurrent evolution of heat-responsiveness in Brassicaceae COPIA elements. *Genome Biol.* **17**, 209.

Plohl, M., Luchetti, A., Meštrović, N. and Mantovani, B. (2008) Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene*, **409**, 72–82.

Schmidt, M., Hense, S., Minoche, A.E., Dohm, J.C., Himmelbauer, H., Schmidt, T. and Zakrzewski, F. (2014) Cytosine methylation of an ancient satellite family in the wild beet Beta procumbens. *Cytogenet Genome Res.* **143**, 157–167.

Schnable, P.S., Ware, D. and Fulton, R.S. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.

Schranz, M.E., Lysak, M.A. and Mitchell-Olds, T. (2006) The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends Plant Sci.* **11**, 535–542.

Seymour, D.K., Koenig, D., Hagmann, J., Becker, C. and Weigel, D. (2014) Evolution of DNA methylation patterns in the Brassicaceae is driven by differences in genome organization. *PLoS Genet.* **10**, e1004785.

Šmarda, P., Bureš, P., Šmerda, J. and Horová, L. (2012) Measurements of genomic GC content in plant genomes with flow cytometry: a test for reliability. *New Phytol.* **193**, 513–521.

Šmarda, P., Bureš, P., Horová, L., Leitch, I.J., Mucina, L., Pacini, E., Tichý, L., Grulich, V. and Rotreklová, O. (2014) Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc. Natl Acad. Sci. USA*, **111**, E4096–E4102.

Stroud, H., Greenberg, M.V.C., Feng, S., Bernatavichute, Y.V. and Jacobsen, S.E. (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell*, **152**, 352–364.

Tiang, C.-L., He, Y. and Pawlowski, W.P. (2012) Chromosome organization and dynamics during interphase, mitosis, and meiosis in plants. *Plant Physiol.* **158**, 26–34.

Vu, G.T.H., Schmutzer, T., Bull, F. (2015) Comparative genome analysis reveals divergent genome size evolution in a carnivorous plant genus. *Plant Genome* **8**. https://doi.org/10.3835/plantgenome2015.04.0021.

Willing, E.M., Rawat, V., Mandáková, T. *et al*. (2015) Genome expansion of Arabis alpina linked with retrotransposition and reduced symmetric DNA methylation. *Nat. Plants*, **1**, 1–5.

Willing, E.-M., Piofczyk, T., Albert, A., Winkler, B.J., Schneeberger, K. and Pecinka, A. (2016) UVR2 ensures trans-generational genome stability under simulated natural UV-B in Arabidopsis thaliana. *Nat. Commun.* **7**, 13522.

Yadav, R.K., Girke, T., Pasala, S., Xie, M. and Reddy, G.V. (2009) Gene expression map of the Arabidopsis shoot apical meristem stem cell niche. *Proc. Natl Acad. Sci. USA*, **106**, 4941–4946.

Yan, H., Kikuchi, S., Neumann, P., Zhang, W., Wu, Y., Chen, F. and Jiang, J. (2010) Genome-wide mapping of cytosine methylation revealed dynamic DNA methylation patterns associated with genes and centromeres in rice. *Plant J.* **63**, 353–365.

Zakrzewski, F., Weisshaar, B., Fuchs, J., Bannack, E., Minoche, A.E., Dohm, J.C., Himmelbauer, H. and Schmidt, T. (2011) Epigenetic profiling of heterochromatic satellite DNA. *Chromosoma*, **120**, 409–422.

Zakrzewski, F., Schubert, V., Viehoever, P., Minoche, A.E., Dohm, J.C., Himmelbauer, H., Weisshaar, B. and Schmidt, T. (2014) The CHH motif in sugar beet satellite DNA: a modulator for cytosine methylation. *Plant J.* **78**, 937–950.

Zhang, W., Lee, H.-R., Koo, D.-H. and Jiang, J. (2008) Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in *Arabidopsis thaliana* and maize. *Plant Cell*, **20**, 25–34.

