Martin Schmidt, Sarah Hense, André E. Minoche, Juliane C. Dohm, Heinz Himmelbauer, Thomas Schmidt, Falk Zakrzewski

Cytosine Methylation of an Ancient Satellite Family in the Wild Beet Beta procumbens

Erstveröffentlichung in / First published in:

Cytogenetic and genome research. 2014, 143 (1-3), S. 157 – 167 [Zugriff am: 19.05.2020]. Karger. ISSN 1424-859X.

DOI: https://doi.org/10.1159/000363485

Cytogenetic and Genome Research

Cytogenet Genome Res 2014;143:157–167
DOI: 10.1159/000363485

Published online: June 27, 2014

Cytosine Methylation of an Ancient Satellite Family in the Wild Beet *Beta procumbens*

Martin Schmidt[a] Sarah Hense[a] André E. Minoche[b–d] Juliane C. Dohm[b–d]
Heinz Himmelbauer[b–d] Thomas Schmidt[a] Falk Zakrzewski[a]

[a]Department of Plant Cell and Molecular Biology, TU Dresden, Dresden, and [b]Max Planck Institute for Molecular Genetics, Berlin, Germany; [c]Centre for Genomic Regulation (CRG), and [d]Universitat Pompeu Fabra (UPF), Barcelona, Spain

Key Words
Beta procumbens · Bisulfite sequencing · DNA methylation · Immunostaining · Satellite DNA · Sugar beet

Abstract
DNA methylation is an essential epigenetic feature for the regulation and maintenance of heterochromatin. Satellite DNA is a repetitive sequence component that often occurs in large arrays in heterochromatin of subtelomeric, intercalary and centromeric regions. Knowledge about the methylation status of satellite DNA is important for understanding the role of repetitive DNA in heterochromatization. In this study, we investigated the cytosine methylation of the ancient satellite family pEV in the wild beet *Beta procumbens*. The pEV satellite is widespread in species-specific pEV subfamilies in the genus *Beta* and most likely originated before the radiation of the Betoideae and Chenopodioideae. In *B. procumbens*, the pEV subfamily occurs abundantly and spans intercalary and centromeric regions. To uncover its cytosine methylation, we performed chromosome-wide immunostaining and bisulfite sequencing of pEV satellite repeats. We found that CG and CHG sites are highly methylated while CHH sites show only low levels of methylation. As a consequence of the low frequency of CG and CHG sites and the preferential occurrence of most cytosines in the CHH motif in pEV monomers, this satellite family displays only low levels of total cytosine methylation.

Plant genomes contain large proportions of repetitive DNA with a high amount of satellite DNA [Schmidt and Heslop-Harrison, 1998]. Satellites consist of homologous sequence motifs (monomers) which vary in size from 150–180 bp and multiples thereof and are organized tandemly. These tandem repeats predominantly occur in large, megabase-sized homogenous arrays in centromeric heterochromatin, intercalary or subtelomeric regions, while smaller arrays may be dispersed along chromosomes [Hemleben et al., 2007; Palomeque and Lorite, 2008; Plohl et al., 2008]. Satellite DNAs are important sequence elements in heterochromatin formation and maintenance and hence are involved in the stability of chromatin structures [Martienssen, 2003; Ugarkovic, 2005; Lisch, 2009; Teixeira and Colot, 2010].

The genus *Beta* is a member of the Amaranthaceae family within the order Caryophyllales. The Amaranthaceae subfamilies Betoideae and the closely related Chenopodioideae [Müller and Borsch, 2005; Stevens, 2013] in-

KARGER

E-Mail karger@karger.com
www.karger.com/cgr

1424–8581/14/1433–0157$39.50/0

Thomas Schmidt
Department of Biology, TU Dresden
Zellescher Weg 20b
DE–01062 Dresden (Germany)
E-Mail Thomas.Schmidt@tu-dresden.de

clude economically important crops, namely sugar beet *(Beta vulgaris)* and quinoa *(Chenopodium quinoa)*. The widely used taxonomic system of Ford-Lloyd [2005] of the genus *Beta* considers the 4 sections *Beta*, *Corollinae*, *Procumbentes*, and *Nanae*. The section *Beta* consists of sugar beet with all cultivars and the wild species *B. patula*, which occurs only on a small island near Madeira. *B. lomatogona* belongs to the section *Corollinae*. The section *Nanae* contains only *B. nana* occurring at high altitudes in Greece. The wild species *B. procumbens*, section *Procumbentes*, grows at the northwest coasts of Africa and on the Canary Islands. *B. procumbens* has mostly submetacentric or acrocentric chromosomes (2n = 18). Similar to sugar beet, a large proportion of repetitive DNA including satellite DNAs has been characterized in *B. procumbens* [Schmidt and Heslop-Harrison, 1996; Dechyeva et al., 2003; Weber and Schmidt, 2009]. A prominent satellite is pEV which was first described in sugar beet [Schmidt and Metzlaff, 1991]. pEV is widespread in the genus and highly abundant in sugar beet and *B. procumbens* [Schmidt and Metzlaff, 1991; Zakrzewski et al., 2010] making it a suitable object for comparative investigation of satellite DNA diversity across species.

Repetitive sequences in plants are epigenetically inactivated by cytosine methylation [Lisch, 2009; Teixeira and Colot, 2010]. First genome-wide analyses of cytosine methylation using bisulfite sequencing were performed in *Arabidopsis* and maize [Cokus et al., 2008; Lister et al., 2008; Gent et al., 2013]. High levels of DNA methylation were detected in pericentromeric and intercalary regions with repetitive DNAs as main targets. However, knowledge of the cytosine methylation of satellite DNA is still limited because of the collapsing of tandem arrays in assemblies of next-generation sequencing (NGS) data. Cytosine methylation is established de novo by domains rearranged methyltransferases DRM1 and DRM2. Cytosine methylation at CG sites is maintained by methyltransferase 1 (MET1) after DNA replication. Additionally, cytosines are methylated in CHG sites (H = A, C, T) and asymmetric CHH sites by chromomethylase 3 (CMT3) or DRM1 and DRM2 [Henderson and Jacobsen, 2007].

In this study, we investigated the diversity of the satellite pEV in the genus *Beta* as well as in *C. quinoa* in silico by analyzing NGS datasets. The characterization of this satellite family included comparative Southern blot and fluorescence in situ hybridization (FISH) on sugar beet and *B. procumbens* chromosomes. For the investigation of the cytosine methylation of pEV, we performed immunostaining of *B. procumbens* chromosomes and bisulfite sequencing of the *B. procumbens*-specific pEV subfamily. We found that pEV arrays display only low levels of cytosine methylation. This is due to the occurrence of many cytosines at CHH sites which are mostly not methylated. CG and CHG sites are highly methylated, but their frequency in pEV monomers is low.

Materials and Methods

Plant Materials and Isolation of Genomic DNA

B. vulgaris ssp. *vulgaris* (genotype KWS2320), *B. patula*, *B. procumbens*, *B. lomatogona*, *B. nana*, and *C. quinoa* were grown under greenhouse conditions. Isolation of genomic DNA was performed from young leaves using the CTAB (cetyltrimethyl/ammonium bromide) standard protocol [Murray and Thompson, 1980].

Selection of pEV Monomers from Plant Sequence Databases

Reference sequences of *B. vulgaris* ssp. *vulgaris* (genotype KWS2320) [Dohm et al., 2014], *B. patula*, *B. procumbens*, *B. lomatogona*, *B. nana*, and *C. quinoa* [H. Himmelbauer, pers. commun.] were screened for pEV monomers using localFASTA, blastn with default parameters and assembly algorithms (gaps: 10%, size: 16, overlap: 100, identity: 40%, word length: 12, index: 12, maximum mismatches: 60%, ambiguity: 4) integrated in the Geneious software version 6.1.6 (http://www.geneious.com). The majority consensus sequence of previously published pEV monomers from *B. vulgaris* was adopted from Zakrzewski et al. [2011]. For the detection of pEV sequences in all NGS datasets, NGS reads were mapped against an artificial pEV dimer, and positive hits were selected. pEV sequences and full-length monomers were extracted, and sequence alignments were generated using MUSCLE [Edgar, 2004] and MAFFT [Katoh et al., 2005] (online suppl. files 1–7; see www.karger.com/doi/10.1159/000363485 for all online suppl. material). The start and end positions of the pEV monomer were adopted from Zakrzewski et al. [2011]. Majority consensus sequences, monomer length, monomer number and pairwise identities of species-specific pEV monomers were deducted using Geneious software version 6.1.6.

Southern Blot Hybridization

Three micrograms of genomic DNA of *B. vulgaris* ssp. *vulgaris* (genotype KWS2320), *B. patula*, *B. procumbens*, *B. lomatogona*, *B. nana*, and *C. quinoa* were digested with *Eco*RI (Thermo-Fischer Scientific) and *Alw*NI (New England Biolabs), separated on a 1.2% agarose gel (online suppl. fig. 2) and transferred onto Hybond-XL nylon membrane (GE Healthcare) using alkaline transfer. Southern hybridization using ^{32}P-labeled probes was performed according to standard protocols [Sambrook et al., 1989]. A representative pEV probe originating from *B. procumbens* was used. Filters were hybridized at 60°C and washed in 2× SSC/0.1% SDS for 80 min at 60°C. Signals were detected by autoradiography.

FISH and Immunofluorescence

Meristem tissues of young leaves were used for the preparation of mitotic chromosomes. The maceration of plant material was performed in an enzyme mixture consisting of 0.3% (w/v) cytohelicase (Sigma), 1.8% (w/v) cellulose from *Aspergillus niger* (Sigma), 0.2% (w/v) cellulase Onozuka-R10 (Serva), and 20% (v/v) pectinase from *A. niger*, followed by spreading of the nuclei on slides.

Species-specific probes of pEV were labeled with biotin-16-dUTP (Roche) by PCR. Hybridization and detection were performed according to Schmidt et al. [1994]. Chromosome preparations were counterstained with DAPI (4′,6-diamidino-2-phenylindole) and mounted in antifade solution (CitiFluor). Examination of slides was carried out with a Zeiss Axioplan2 Imaging fluorescent microscope. Images were acquired with the Applied Spectral Imaging v.3.3 software coupled with the CCD camera ASI BV300-20A. The images were optimized by Adobe Photoshop software using only functions affecting the whole image equally.

For detection of 5-methylcytosine (5mC) and pEV satellite arrays along pachytene chromosomes, preparations were made according to Zakrzewski et al. [2011]. Slides were fixed with 4% paraformaldehyde in 1× PBS. After washing 3 times with 1× PBS, slides were incubated in 70% formamide at 80°C for 3 min followed by a dehydration series in 70% and 100% ethanol. After incubation in 200 µl blocking solution, slides were incubated with 50 µl of primary antibody against 5mC (anti-5mC, Calbiochem, No. 162 33 D3; diluted 1:250) at 4°C overnight. Detection of the 5mC antibody was carried out with 50 µl secondary antibody in antibody solution (antimouse-Alexa Fluor 488, Invitrogen, No. A-11001; diluted 1:250) at 37°C for 1 h. For the subsequent FISH, slides were washed 3 times with 2× SSC for 10 min before fixing in 4% paraformaldehyde and 3.5% sucrose. The FISH procedure followed after washing the slides 3 times in 2× SSC for 10 min and dehydrating in 70 and then 100% ethanol. The examination of slides was carried out as described above.

Bisulfite Sequencing of Satellite Monomers

Genomic DNA (1 µg) from young sugar beet leaves was bisulfite-converted using the Epimark Bisulfite Conversion Kit (NEB) according to the manufacturer's instructions.

As control for conversion efficiency, previously amplified PCR products of pEV were bisulfite-converted in parallel. The bisulfite-converted genomic DNA and the bisulfite-converted PCR products of pEV were used as template for the amplification of pEV satellite sequences. Two primer pairs for forward and reverse DNA strand, respectively, were used to amplify fragments of the entire length of 1 pEV monomer and to cover the diversity of pEV variants. The primers contain wobble bases at positions where a cytosine occurs allowing equally binding to methylated cytosine or uracil (converted unmethylated cytosine). The amplification products were ligated into pGEM-T vector and transformed into XL1-Blue *E. coli* cells. The plasmid library was grown in 96-well plates and Sanger sequenced. Sequence alignments of the bisulfite-sequenced satellite monomers were generated using MAFFT with default parameters (online suppl. files 8 and 9). Conserved cytosine positions were analyzed for conversion events (online suppl. tables 2 and 3). Used primers are listed in online supplementary table 4.

Results

The pEV Satellite Is Differentially Amplified in the Genus Beta

To investigate the abundance and sequence diversity of pEV, we performed a large-scale identification and characterization of pEV sequences in sugar beet, *B. patula*, *B. procumbens*, *B. lomatogona*, and *B. nana* as well as in *C. quinoa*. The full-length monomers of pEV were determined in the reference sequence of each genome. Furthermore, because satellite DNA is often collapsed in genome assemblies, we used Illumina reads of each species for an estimation of the total amount of pEV in each genome and to draw conclusions about the homogenization of pEV monomers in the reference sequences. The genome proportion and the number of analyzed full-length monomers, the diversity and pairwise identity in the investigated genomes are summarized in table 1.

In sugar beet, 2,309 full-length monomers were isolated from the reference sequence. Furthermore, 409,000 Illumina reads containing pEV sequences were detected allowing estimation of a total genome proportion of 41 Mb. In the draft of *B. patula*, which is closely related to sugar beet and a species of the section *Beta*, 2,733 pEV full-length monomers were detected showing high identity to sugar beet pEV monomers. In *B. patula*, pEV makes up 31 Mb and is represented by 306,000 Illumina reads. These results confirm the high abundance of the satellite pEV in the section *Beta*. In the *B. procumbens* reference sequence, 5,019 full-length monomers were detected making up the highest amount of full-length pEV monomers identifiable in *Beta* genome sequences. In the Illumina dataset, 61,000 pEV sequences were detected representing 6 Mb of the genome. Furthermore, 38 and 28 full-length pEV monomers occur in the reference sequences of *B. lomatogona* and *B. nana*, respectively. In *C. quinoa*, 7 pEV full-length monomers were detected. In each of these 3 genomes, approximately 1,000 Illumina sequences contain pEV sequences.

The species-specific alignments (online suppl. files 1–7) of pEV monomers extracted from the reference sequences were used for the generation of consensus monomer sequences from each individual species (fig. 1A). Although these consensus sequences contain some wobble positions, they can be used to analyze the diversity and to estimate the level of homogenization of pEV in each genome. The higher the differences in nucleotide diversity, the more variable are the positions in the consensus sequence. The sugar beet and *B. patula* pEV consensus sequences contain only very few wobble nucleotides and are highly similar to each other (98.1%) indicating a high degree of homogenization of pEV in both species (fig. 1B). In *B. procumbens*, the consensus sequence shows only a minor number of wobble nucleotides which also indicates a homogenization of pEV in this species. The *B. procumbens* pEV consensus is diverged by point mutations

Table 1. Distribution, abundance and diversity of the pEV satellite in the genus *Beta* and in *C. quinoa*

Section	Species	Assembled genome					Illumina reads			
		database (size)	full-length monomers	monomer size, nt	GC content, %	pairwise identity, %	database reads (size)	coverage	hits	hits/1.0× coverage[a]
Beta	*B. vulgaris* (sugar beet)	RefBeet 1.0 (596 Mb)	2,309	158	43.6	77.4	46,132,887 (4,659 Mb)	7.8×	3,204,040	409,000 (41 Mb)
	B. patula	Ref 0.1 (607 Mb)	2,733	158, 159	43.5	76.8	98,621,695 (9,861 Mb)	16.2×	4,968,609	306,000 (31 Mb)
Corollinae	*B. lomatogona*	Ref 0.1 (775 Mb)	38	157, 158, 159	39.1	65.6	159,198,401 (15,920 Mb)	20.5×	18,724	<1,000 (91 kb)
Procumbentes	*B. procumbens*	Ref 0.1 (695 Mb)	5,019	152, 157	44.4	62.8	344,501,043 (34,019 Mb)	49.0×	2,989,585	61,000 (6 Mb)
Nanae	*B. nana*	Ref 0.1 (513 Mb)	28	154, 158	36.8	59.2	170,224,864 (17,022 Mb)	33.1×	44,081	>1,000 (133 kb)
Chenopodium	*C. quinoa*	Ref 0.1 (1.4 Gb)	7	156	37.6	54.3	106,333,666 (10,633 Mb)	7.4×	8,976	>1,000 (121 kb)

[a] Rounded to the nearest thousand.

and indels from pEV of sugar beet and *B. patula*, suggesting the evolution of a species-specific pEV subfamily in *B. procumbens*. The *B. lomatogona* and *B. nana* consensus sequences are more similar (88%) to each other than to the consensus sequence of the remaining species (61–66%) and are characterized by many wobbles indicating a low degree of homogenization. The consensus sequence of pEV in *C. quinoa* shows a high number of point mutations and hence is highly diverged from members of the genus *Beta* (fig. 1A).

Potential scenarios of relationships can be estimated by comparison of the pEV consensus sequences of all investigated species using a neighbor-joining dendrogram (fig. 1C). The pEV consensus of sugar beet and *B. patula* groups both species close to each other suggesting the generation of a pEV subfamily that is restricted to the section *Beta*. The *B. procumbens*-specific pEV consensus is separated from the section *Beta* demonstrating that a subfamily of pEV specific for *B. procumbens* evolved. Similarly, subfamilies of pEV occur in the genomes of *B. nana* and *B. lomatogona*. They might be more related to each other because of their close grouping. The pEV consensus of *C. quinoa* is diverged from members in the genus *Beta* and is placed in a separate position in the dendrogram.

The Genomic Organization of the pEV Satellite in the Genus Beta and Its Localization on Sugar Beet and B. procumbens Chromosomes

The highest numbers of pEV reads in NGS data are detectable for sugar beet, *B. patula* and *B. procumbens*. To analyze the large-scale genomic organization of pEV in these species, comparative Southern blot hybridization was performed. Moreover, *B. lomatogona*, *B. nana* and *C. quinoa* were analyzed in parallel to compare the estimated abundance and separation into species-specific pEV subfamilies.

Conserved recognition sites for the restriction endonucleases *Eco*RI and *Alw*NI in pEV monomers were identified in silico in sugar beet, *B. patula* and *B. procumbens*, and these endonucleases were used for restriction of genomic DNA. Strong hybridization signals using a representative pEV probe from *B. procumbens* were detected in the closely related species sugar beet and *B. patula* (fig. 2), confirming high abundance in these species. In *B. procumbens*, similarly strong signals were detectable which demonstrates the high copy number of pEV in the section *Procumbentes*. The typical ladder-like banding pattern of satellite DNA is clearly observable for sugar beet, *B. patula* and *B. procumbens* demonstrating the organization of pEV in large tandem arrays. Moreover, the recognition site of *Eco*RI is largely conserved in sugar beet and *B. patula* indicating homogenization of pEV in the section *Beta* (fig. 2). The weaker hybridization signals to mono- and multimers and strong hybridization of larger pEV fragments show low conservation of *Alw*NI recognition sites in the section *Beta*. Recognition sites of *Eco*RI and *Alw*NI are not largely conserved in *B. procumbens* which leads to the lower number of mono- and multimers after digestion with both enzymes. In *B. lomatogona*, only large DNA fragments showing weak signals are observ-

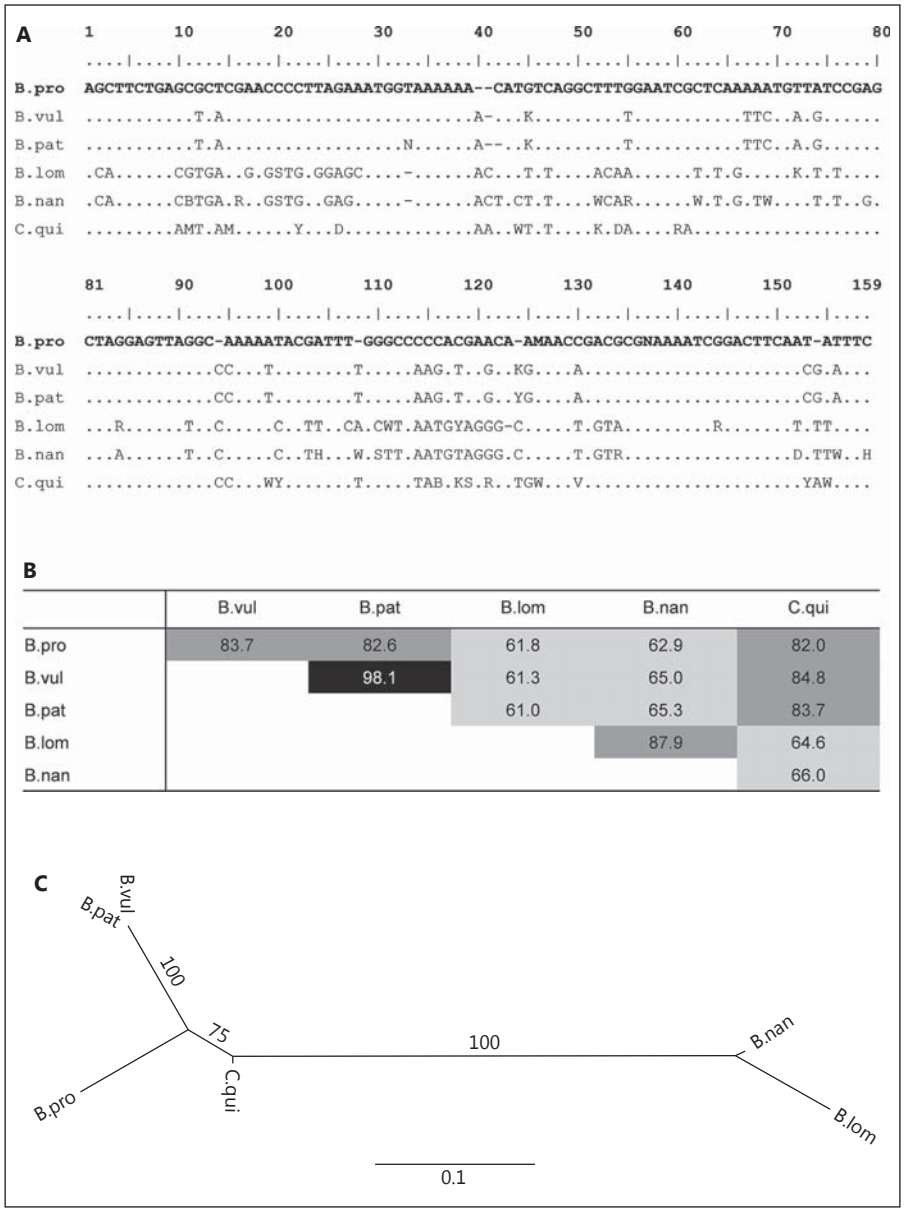

Fig. 1. Diversity of the pEV satellite in the genus *Beta* and in *C. quinoa*. **A** Consensus sequences of species-specific pEV monomers of the genus *Beta* and of *C. quinoa*. The consensus sequences were generated using pEV alignment of all monomers isolated from species-specific reference genome sequences. The following wobble nucleotides are shown: B = C, G or T; D = A, G or T; H = A, C or T; K = G or T; M = A or C; N = any nucleotide; R = A or G; S = G or C; Y = C or T; V = A, C or G; W = A or T. **B** Identity matrix of species-specific pEV consensus sequences. Light grey shaded boxes indicate identity values <80%, grey boxes show identity values of 80–90% and black boxes display identity values >90%. **C** Cladogram of species-specific pEV monomer consensus sequences using neighbor-joining algorithm.

able. The low number of pEV copies detected is most likely caused by the divergence of the *B. lomatogona* pEV subfamily. In *B. nana*, a signal smear representing large DNA fragments hybridizing with the pEV probe confirms the moderately abundant character of pEV. The ladder-like banding pattern is observable after digestion with *Eco*RI but not with *Alw*NI, revealing conservation of the *Eco*RI recognition site in a subset of pEV monomers and organization in tandem arrays. Faint but clear signals of large DNA fragments hybridizing with pEV are detectable in *C. quinoa*. The organization of pEV monomers in arrays is confirmed by PCR using species-specific primers in all analyzed species (online suppl. fig. 1).

In order to detect the chromosomal localization of the pEV satellite in sugar beet and *B. procumbens*, FISH experiments were performed on sugar beet and *B. procumbens* mitotic metaphase chromosomes. In sugar beet, pEV is localized in large arrays in intercalary heterochromatic regions on both arms of all metacentric chromosomes (fig. 3A). In contrast, in *B. procumbens* pEV arrays are amplified in the brightly DAPI-stained centromeric and/or intercalary regions of many chromo-

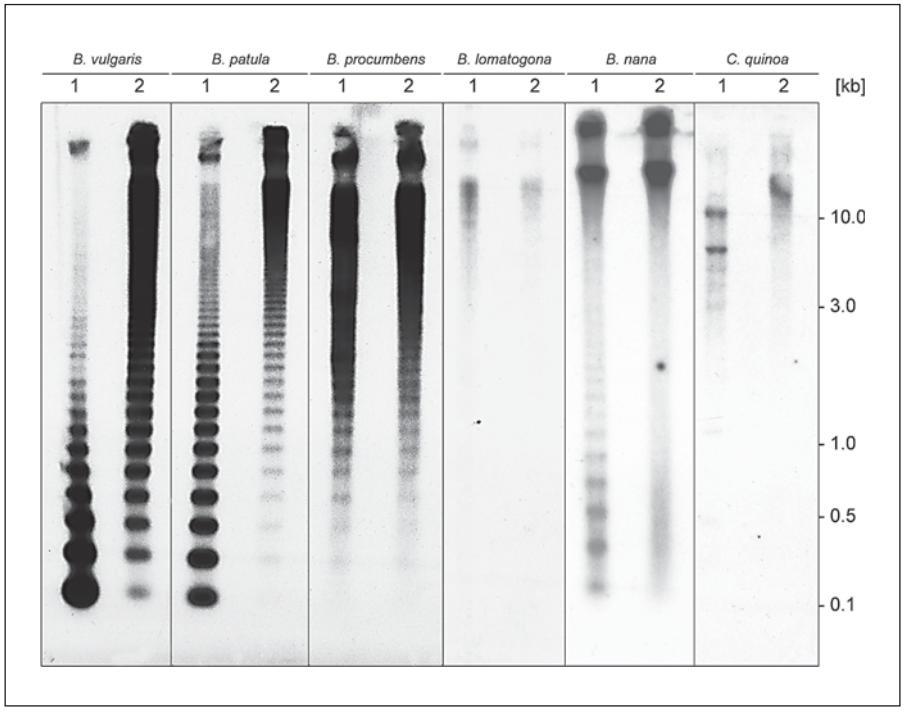

Fig. 2. Southern blot hybridization showing the distribution of the pEV satellite in the genus *Beta* and in *C. quinoa*. Equal amounts of genomic DNA of *B. vulgaris*, *B. patula*, *B. procumbens*, *B. lomatogona*, *B. nana*, and *C. quinoa* were digested with *Eco*RI (1) and *Alw*NI (2), separated on a 1.2% agarose gel, blotted on a nylon membrane, and hybridized with a representative probe of pEV from *B. procumbens*.

somes (fig. 3C). Ten chromosomes are characterized by the localization of pEV arrays in centromeric heterochromatin, and 4 of these chromosomes show additional signals in intercalary regions. In the remaining 8 chromosomes, pEV is localized only in the intercalary heterochromatin.

Cytosine Methylation of pEV in B. procumbens

In order to detect the cytosine methylation of pEV arrays in all *B. procumbens* chromosomes simultaneously, we performed immunostaining using an antibody against 5mC combined with FISH. Pachytene chromosomes were used for immunostaining to achieve higher resolution enabling the detection of the long-range distribution of DNA methylation with respect to pEV arrays. The centromeric heterochromatin can be identified as brightly DAPI-stained region in each chromosome (fig. 3B, D). The centromeres of selected chromosomes are exemplarily indicated by arrows in figure 3B–D. Notably, the *B. procumbens* centromeres show little staining with 5mC antibody compared to adjacent chromosomal regions, demonstrating that the centromeres are characterized by only low levels of cytosine methylation (blue fluorescence in fig. 3F). The intercalary heterochromatin is identifiable by bright DAPI staining (of smaller size than centromeric heterochromatin) and slightly increased levels of cytosine methylation compared to the centromeric heterochromatin (fig. 3D, F).

However, after digital enlargement, it becomes obvious that the pEV arrays in centromeric and intercalary heterochromatin show similar levels of methylation (fig. 3G–M). In the centromeric pEV clusters (arrows in fig. 3I), the red pEV signals can be clearly distinguished from green 5mC signals, hence demonstrating little cytosine methylation. Overlapping signals would be visible as yellow areas showing strong cytosine methylation. Likewise, in intercalary arrays (arrowheads in fig. 3I), the sparsely overlapping signals of pEV (red) and 5mC (green) indicate low cytosine methylation of the corresponding satellite arrays.

Despite the advantages of immunostaining in the chromosome-wide investigation of the cytosine methylation of pEV arrays, this technique does not allow the identification and characterization of methylation levels at a resolution of single nucleotides. Therefore, we performed bisulfite sequencing of pEV sequences for strand-specific analysis of the cytosine methylation level at CG, CHG (H = A, C, T) and CHH sites. The efficiency of bisulfite conversion was controlled and accounts for 99.5%, hence validating experimental conditions (online suppl. table 2). The pEV satellite displays high levels of methylation at symmetric CG sites on both DNA strands (fig. 4).

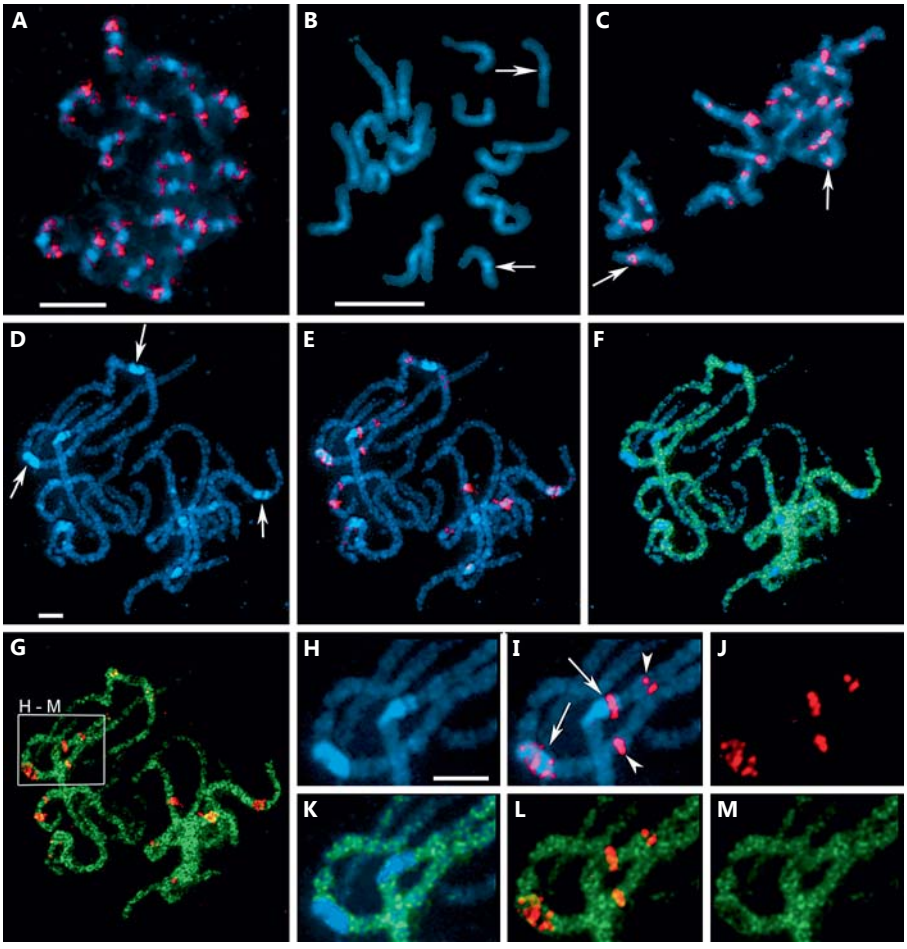

Fig. 3. FISH and immunostaining on mitotic prophase (**A–C**) and meiotic pachytene chromosomes (**D–M**) of sugar beet (**A**) and *B. procumbens* (**B–M**). Chromosomes are stained with DAPI (blue). Heterochromatic regions are visible as strong DAPI signals (e.g. **B**, **D**, **H**). Hybridization with representative probes of the pEV satellite (red) originating from sugar beet (**A**) and *B. procumbens* (**C–M**) visualizes pEV arrays in heterochromatin of intercalary regions of sugar beet (**A**), and in centromeric and intercalary regions of *B. procumbens* (**C**). **C** Two centromeric pEV arrays are indicated by arrows. **D–G** The same pachytene chromosomes of *B. procumbens* after DAPI staining (**D**), FISH with pEV satellite probe (**E**), immunolocalization of 5mC (**F**) and superposition of FISH and immunostaining pictures (**G**). **D** Exemplarily, 3 centromeres are indicated by arrows. **E** pEV arrays (red) are localized in centromeric and intercalary regions. **F** The large heterochromatic blocks show less methylation signals and hence appear blue. **G** pEV arrays show low levels of cytosine methylation. **H–M** Digital enlargement of the region framed in **G**. **H** The 2 centromeres are visible as brightly DAPI-stained regions. **I** pEV arrays are localized in the 2 centromeres (arrows) and in 2 intercalary regions (arrowheads). **J** Only pEV arrays are shown. **K** The 2 centromeres show low levels of 5mC. **L** pEV signals are distributed in low-methylated chromosomal regions visible as reduced overlapping of green and red signals. **M** The reduced number of green signals demonstrates low methylation of centromeric regions. Bars = 5 μm.

The level of CG methylation varies between 46 and 100%. Similarly strong methylation is detectable for the CHG site, which varies between 43 and 79%. Remarkably low is the asymmetric CHH methylation varying between 0 and 35% (fig. 4). On average, 75% of cytosines are methylated at CG sites and 61% at CHG sites. In contrast, only 9% of cytosines are methylated at the CHH sites. Taking into account that a cytosine mostly occurs in the CHH motif, followed by CG and CHG, the overall methylation of pEV is low. In total, 36% of all cytosines are methylated in a pEV monomer of *B. procumbens* (fig. 4). We observed no differences in the methylation at CG, CHG and CHH sites between both DNA strands.

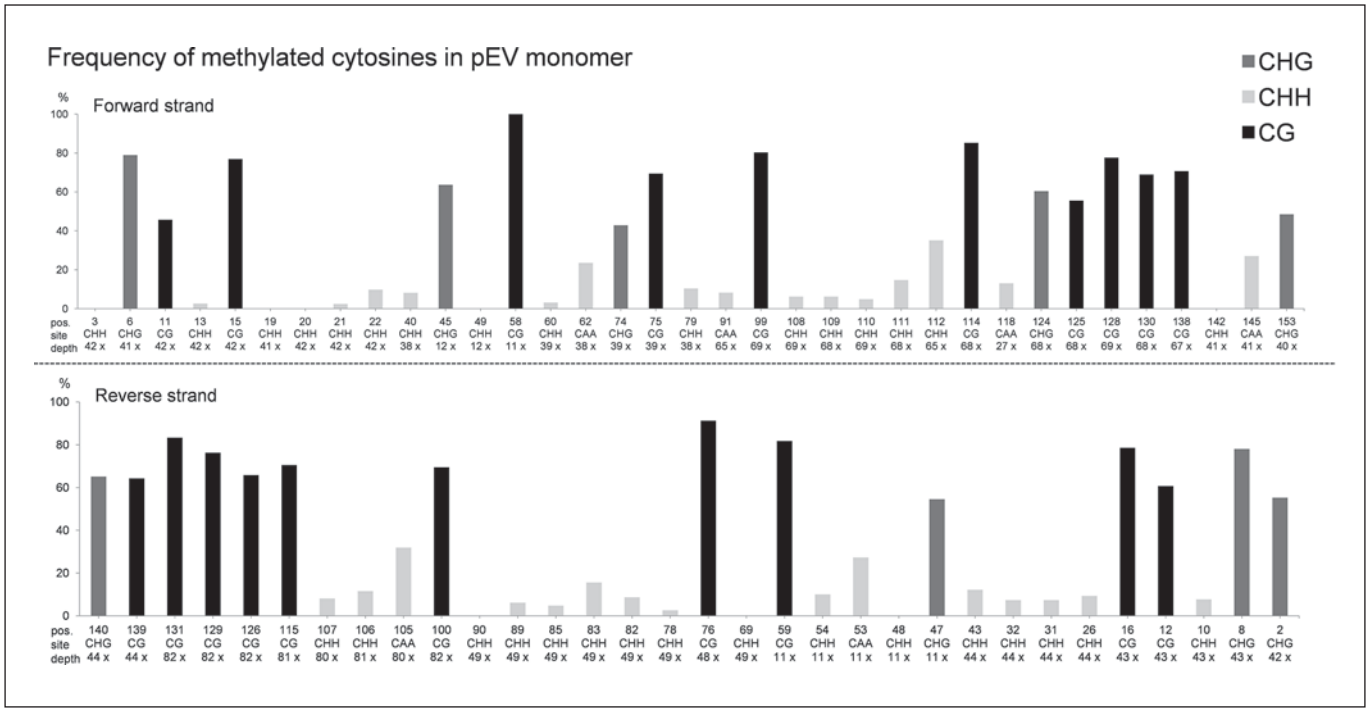

Fig. 4. Illustration of cytosine sites and their average methylation in the pEV monomer in *B. procumbens*. The x-axis indicates conserved cytosine positions, the corresponding cytosine sites (CG, CHG, CHH) and the bisulfite sequence depth. The y-axis shows the frequency of methylated cytosines.

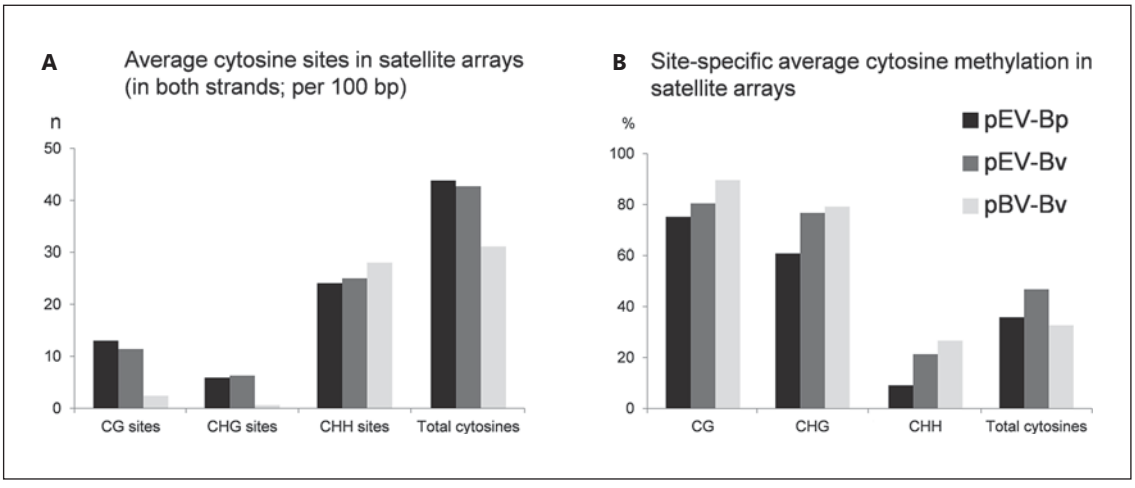

Fig. 5. Dependence of global cytosine methylation of satellite DNA on frequency of CG, CHG and CHH sites. **A** The frequency of CG, CHG and CHH and all cytosine sites scaled on a 100-bp window is displayed for the *B. procumbens* pEV subfamily (pEV-Bp), the sugar beet pEV subfamily (pEV-Bv) and the sugar beet pBV satellite (pBV-Bv). **B** The level of methylation at CG, CHG, CHH and all cytosines sites is indicated for each of the 3 satellites.

Cytogenet Genome Res 2014;143:157–167
DOI: 10.1159/000363485

Discussion

In this study, we investigated the diversity and cytosine methylation of the satellite family pEV. By analyzing genome assemblies and large-scale NGS datasets of representative species of the genus *Beta*, we found that pEV is present in all investigated species, which suggests its origin before the radiation of species in the genus *Beta*. This is also demonstrated by the occurrence of pEV as a moderately repetitive and highly diverged repeat in *C. quinoa*. Because the split of the Betoideae and Chenopodioideae was estimated 52.7 MYA [Kadereit et al., 2012], we speculate that the pEV satellite might be older than 50–60 million years, characterizing pEV as an ancient satellite family that occurs in species of Betoideae and Chenopodioideae. In *Anemone blanda*, the age of a satellite DNA family was estimated to be 27 million years [Hagemann et al., 1993] suggesting that satellite families in plant genomes are already present for millions of years.

The large abundance of pEV in the sections *Beta* and *Procumbentes* indicates a massive amplification of pEV. Most likely, pEV has occurred as low abundant or even unique precursor sequence in the ancestors of the Betoideae and Chenopodioideae, and amplification took place to a large extent during the evolution of the genus *Beta*. Subsequently, pEV copies might have been vastly removed by recombination events in species of the sections *Nanae* and *Corollinae*, while highly abundant pEV subfamilies evolved in the sections *Procumbentes* and *Beta*. However, it is more likely that in the sections *Beta* and *Procumbentes* a massive amplification of pEV occurred independently from each other after the radiation of each section.

The pEV satellite occurs as diverged subfamilies in species of the genus *Beta* and in *C. quinoa*, which is consistent with the observed turnover of satellite families and the formation of species-specific subfamilies in closely related species. This might be explained when taking into account the different copy numbers of pEV in the analyzed genomes. The lower the copy number of a satellite, the less homogenization takes place which inhibits the formation of a species-specific satellite family [Hemleben et al., 2007] as can be seen in *B. lomatogona*, *B. nana* and *C. quinoa*. However, when the copy number strongly increases, fast homogenization might occur, and a species-specific and homogenized satellite family may evolve [Koukalova et al., 2010]. Because pEV is highly abundant in species of the sections *Beta* and *Procumbentes*, the species-specific pEV subfamilies contain highly similar and homogenized pEV copies which might have been generated by rolling circle amplification [Cohen et al., 2008, 2010; Navrátilová et al., 2008; Cohen and Segal, 2009].

The *B. procumbens*-specific pEV subfamily is distributed in the centromeric regions of 10 chromosomes. For centromeres, it was shown that satellite DNA is a sequence platform for the binding of CENH3, the centromeric variant of histone H3 which is essential for the establishment of the kinetochore. Moreover, modification of satellite DNAs on the level of DNA sequence and chromatin contributes to the epigenetic identity of the centromere [Nagaki et al., 2003, 2004; Lamb et al., 2007; Plohl et al., 2008]. This might point to the importance of the cytosine methylation of the *B. procumbens* centromeric pEV arrays in the epigenetic characterization of the centromere. Together with the centromeric satellites pTS4 and pTS5 [Schmidt and Heslop-Harrison, 1996], and the centromeric *gypsy*-like retrotransposons *Beetle1* and *Beetle2* [Weber and Schmidt, 2009], the pEV satellite might be involved in the recruitment of CENH3.

In addition to the centromeric pEV clusters, the intercalary pEV arrays might participate in the regulation of heterochromatin by cytosine methylation [Pikaard et al., 2013]. In sugar beet, the cytosine methylation of pEV and of the centromeric pBV satellite has been investigated using immunostaining experiments [Zakrzewski et al., 2011], and for both satellites low levels of cytosine methylation compared to adjacent regions were detected. However, pBV arrays show lower levels than pEV demonstrating hypomethylation of sugar beet centromeres. Similarly, the *B. procumbens* centromere is characterized by reduced levels of cytosine methylation as it has also been observed in *Arabidopsis* and jute [Lister et al., 2008; Zhang et al., 2008; Begum et al., 2013], which might indicate that low levels of cytosine methylation are typical for plant centromeres. The intercalary pEV arrays are similarly low-methylated as centromeric pEV arrays in *B. procumbens*. A low methylation of satellites has been also detected for individual satellite families in *Arabidopsis*, rice, maize, sugar beet, *Pennisetum glaucum* and jute by immunostaining [Kamm et al., 1994; Zhang et al., 2008; Yan et al., 2010; Zakrzewski et al., 2011; Begum et al., 2013], demonstrating that satellite DNAs are not necessarily heavily methylated in plant genomes.

The low cytosine methylation might be the result of deamination of 5mC which is the most common nucleotide mutation and might lead to higher AT contents [Hendrich et al., 1999]. Another explanation might be a potential site-preference for cytosine methylation at satellite repeats. Using bisulfite sequencing, high levels of methylation were detected for the rarely occurring sym-

metric CG and CHG sites in the *B. procumbens* pEV subfamily in contrast to low levels of methylation of cytosines at frequent CHH sites. Therefore, methylation of cytosines in satellites is established in dependence on the sequence context and hence strongly influences the total methylation level of the corresponding arrays [Zakrzewski et al., 2014].

We compared the frequency of CG, CHG and CHH sites (online suppl. table 1) and the level of the site-specific methylation between the sugar beet pEV subfamily, the pBV satellite in sugar beet and the *B. procumbens* pEV subfamily. Interestingly, we found that the site-specific methylation levels increase the lower the total frequency of cytosines is (fig. 5; see pBV-Bv), and the strongest effect is observable for CHH sites. For example, the frequency of CG and CHG sites in pBV is much lower than in the pEV subfamilies while the CHH frequency is similar (fig. 5A). A consequence might be that the CHH methylation is elevated in pBV and also CG and CHG methylation is slightly increased hence leading to similar levels of methylation compared to the pEV arrays in *B. procumbens* (fig. 5B; see total cytosines of pEV-Bp and pBV-Bv). Therefore, we assume that higher CHH methylation might compensate reduced frequency of CG and CHG sites as has been also discussed for dispersed satellite families in sugar beet [Zakrzewski et al., 2014]. However, the intercalary pEV arrays in sugar beet show slightly elevated levels of methylation compared to pEV arrays in *B. procumbens* while both subfamilies share similar frequencies of CG, CHG and CHH sites. These differences might depend on the chromosomal localization because the sugar beet pEV subfamily is localized only in intercalary regions while the *B. procumbens* pEV also occurs in centromeres of many chromosomes. In accordance with the major satellites, smaller satellite families which are dispersed on chromosomes in sugar beet are characterized by low CHH methylation and high levels of methylation at CG and CHG sites [Zakrzewski et al., 2014].

Asymmetric CHH sites are more frequent than CG and CHG sites in the pEV subfamily of *B. procumbens*. Nevertheless, the CG and CHG methylation might be sufficient for the maintenance of heterochromatin because CHG methylation induces KRYPTONITE, a histone H3 methyltransferase, which methylates histone H3 at position lysine 9 (H3K9me2). H3K9me2 is a strong marker for heterochromatin and, in turn, H3K9me2 stimulates the CMT3 DNA methyltransferase to methylate cytosines in the CHG motif, hence inducing the maintenance of heterochromatin via a feedback loop [Law and Jacobsen, 2010]. Similarly, the CG methylation was shown to be involved in the induction of H3K9m2 [Johnson et al., 2007]. Therefore, we propose that the pEV satellite is actively involved in the maintenance of the intercalary and the centromeric heterochromatin of *B. procumbens* by its CG and CHG methylation. Furthermore, this most likely demonstrates that even low frequencies of CG and CHG sites in satellites are sufficient for heterochromatization. It has been shown that methylation at CHH sites in repetitive loci, such as tandem repeats, participates in the regulation of genes in *Arabidopsis* and maize [Henderson and Jacobsen, 2008; Gent et al., 2013]. However, because pEV forms large arrays in intercalary and centromeric heterochromatin showing most likely a reduced density of genes, it is unlikely that pEV is involved in the regulation of genes by its low level of CHH methylation.

Taken together, our results indicate that satellite DNA might fulfill an essential function in the regulation and maintenance of heterochromatin via its site-specific cytosine methylation. The DNA methylation of satellite-rich heterochromatin depends strongly on the level and frequency of CG and CHG methylation of the underlying highly abundant and homogenized satellite repeats.

Acknowledgements

Falk Zakrzewski acknowledges a fellowship and financial support of the FAZIT foundation. This work is funded by the BMBF grant 'Annotation des Genoms der Zuckerrübe unter Berücksichtigung von Genfunktionen und struktureller Variabilität für die Nutzung von Genomdaten in der Pflanzenbiotechnologie (ANNOBEET)' – 0315962 to T.S. and H.H. We thank Nadin Fliegner and Ines Walther for excellent technical assistance.

References

Begum R, Zakrzewski F, Menzel G, Weber B, Alam SS, Schmidt T: Comparative molecular cytogenetic analyses of a major tandemly repeated DNA family and retrotransposon sequences in cultivated jute *Corchorus* species (Malvaceae). Ann Bot 112:123–134 (2013).

Cohen S, Segal D: Extrachromosomal circular DNA in eukaryotes: possible involvement in the plasticity of tandem repeats. Cytogenet Genome Res 124:327–338 (2009).

Cohen S, Houben A, Segal D: Extrachromosomal circular DNA derived from tandemly repeated genomic sequences in plants. Plant J 53: 1027–1034 (2008).

Cohen S, Agmon N, Sobol O, Segal D: Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. Mob DNA 1: 11 (2010).

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, et al: Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. Nature 452:215–219 (2008).

Dechyeva D, Gindullis F, Schmidt T: Divergence of satellite DNA and interspersion of dispersed repeats in the genome of the wild beet *Beta procumbens*. Chromosome Res 11:3–21 (2003).

Dohm JC, Minoche AE, Holtgräwe D, Gutierrez SC, Zakrzewski F, et al: The genome sequence of sugar beet *(Beta vulgaris)*. Nature 505:546–549 (2014).

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

Ford-Lloyd BV: Sources of genetic variation, genus *Beta*, in Biancardi E, Campbell LG, Skaracis GN, De Biaggi M (eds): Genetics and Breeding of Sugar Beet, pp 25–33 (Science Publishers, Inc., Enfield 2005).

Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, et al: CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. Genome Res 23:628–637 (2013).

Hagemann S, Scheer B, Schweizer D: Repetitive sequences in the genome of *Anemone blanda*: identification of tandem arrays and of dispersed repeats. Chromosoma 102:312–324 (1993).

Hemleben V, Kovarik A, Torres-Ruiz RA, Volkov RA, Beridze T: Plant highly repeated satellite DNA: molecular evolution, distribution and use for identification of hybrids. Syst Biodivers 5:277–289 (2007).

Henderson IR, Jacobsen SE: Epigenetic inheritance in plants. Nature 447:418–424 (2007).

Henderson IR, Jacobsen SE: Tandem repeats upstream of the *Arabidopsis* endogene *SDC* recruit non-CG DNA methylation and initiate siRNA spreading. Genes Dev 22:1597–1606 (2008).

Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A: The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature 401:301–304 (1999).

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, et al: The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol 17:379–384 (2007).

Kadereit G, Ackerly D, Pirie MD: A broader model for C_4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae s.s.). Proc Biol Sci 279:3304–3311 (2012).

Kamm A, Schmidt T, Heslop-Harrison JS: Molecular and physical organization of highly repetitive, undermethylated DNA from *Pennisetum glaucum*. Mol Gen Genet 244:420–425 (1994).

Katoh K, Kuma K, Toh H, Miyata T: MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33: 511–518 (2005).

Koukalova B, Moraes AP, Renny-Byfield S, Matyasek R, Leitch AR, Kovarik A: Fall and rise of satellite repeats in allopolyploids of *Nicotiana* over c. 5 million years. New Phytol 186: 148–160 (2010).

Lamb JC, Yu W, Han F, Birchler JA: Plant chromosomes from end to end: telomeres, heterochromatin and centromeres. Curr Opin Plant Biol 10:116–122 (2007).

Law JA, Jacobsen SE: Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11:204–220 (2010).

Lisch D: Epigenetic regulation of transposable elements in plants. Annu Rev Plant Biol 60:43–66 (2009).

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al: Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. Cell 133:523–536 (2008).

Martienssen RA: Maintenance of heterochromatin by RNA interference of tandem repeats. Nat Genet 35:213–214 (2003).

Müller K, Borsch T: Phylogenetics of Amaranthaceae using *matK/trnK* sequence data – evidence from parsimony, likelihood and Bayesian approaches. Ann Missouri Bot Gard 92: 66–102 (2005).

Murray MG, Thompson WF: Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4326 (1980).

Nagaki K, Talbert PB, Zhong CX, Dawe RK, Henikoff S, Jiang JM: Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. Genetics 163:1221–1225 (2003).

Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, et al: Sequencing of a rice centromere uncovers active genes. Nat Genet 36:138–145 (2004).

Navrátilová A, Koblízková A, Macas J: Survey of extrachromosomal circular DNA derived from plant satellite repeats. BMC Plant Biol 8: 90 (2008).

Palomeque T, Lorite P: Satellite DNA in insects: a review. Heredity (Edinb) 100:564–573 (2008).

Pikaard CS, Haag JR, Pontes OMF, Blevins T, Cocklin R: A transcription fork model for Pol IV and Pol V-dependent RNA-directed DNA methylation. Cold Spring Harb Symp Quant Biol 77:205–212 (2013).

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

Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, ed 2 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1989).

Schmidt T, Heslop-Harrison JS: High-resolution mapping of repetitive DNA by in situ hybridization: molecular and chromosomal features of prominent dispersed and discretely localized DNA families from the wild beet species *Beta procumbens*. Plant Mol Biol 30:1099–1113 (1996).

Schmidt T, Heslop-Harrison JS: Genomes, genes and junk: the large-scale organization of plant chromosomes. Trends Plant Sci 3:195–199 (1998).

Schmidt T, Metzlaff M: Cloning and characterization of a *Beta vulgaris* satellite DNA family. Gene 101:247–250 (1991).

Schmidt T, Schwarzacher T, Heslop-Harrison JS: Physical mapping of rRNA genes by fluorescent in-situ hybridization and structural analysis of 5S rRNA genes and intergenic spacer sequences in sugar beet *(Beta vulgaris)*. Theor Appl Genet 88:629–636 (1994).

Stevens PF: Angiosperm Phylogeny Website, http://www.mobot.org/MOBOT/research/APweb/ (2013).

Teixeira FK, Colot V: Repeat elements and the *Arabidopsis* DNA methylation landscape. Heredity (Edinb) 105:14–23 (2010).

Ugarkovic D: Functional elements residing within satellite DNAs. EMBO Rep 6:1035–1039 (2005).

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

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

Zakrzewski F, Wenke T, Holtgräwe D, Weisshaar B, Schmidt T: Analysis of a c_0t-1 library enables the targeted identification of minisatellite and satellite families in *Beta vulgaris*. BMC Plant Biol 10:8 (2010).

Zakrzewski F, Weisshaar B, Fuchs J, Bannack E, Minoche AE, et al: Epigenetic profiling of heterochromatic satellite DNA. Chromosoma 120:409–422 (2011).

Zakrzewski F, Schubert V, Viehöfer P, Minoche AE, Dohm JC, et al: The CHH motif in sugar beet satellite DNA: a modulator for cytosine methylation. Plant J 78:937–950 (2014).

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