the plant journal

SEB Society for Experimental Biology

The Plant Journal (2016) **85,** 229–244 doi: 10.1111/tpj.13103

Diversification, evolution and methylation of short interspersed nuclear element families in sugar beet and related Amaranthaceae species

Katrin Schwichtenberg[1], Torsten Wenke[1], Falk Zakrzewski[1], Kathrin M. Seibt[1], André Minoche[2,3], Juliane C. Dohm[2,4], Bernd Weisshaar[5], Heinz Himmelbauer[3,4] and Thomas Schmidt[1,*]

[1]*Institute of Botany, Technische Universität Dresden, 01069, Dresden, Germany,*
[2]*Max Planck Institute for Molecular Genetics, 14195, Berlin, Germany,*
[3]*Garvan Institute of Medical Research, 2010, Sydney, NSW, Australia,*
[4]*Department of Biotechnology, University of Natural Resources and Life Sciences (BOKU), 1190, Vienna, Austria, and*
[5]*CeBiTec & Department of Biology, University of Bielefeld, 33615, Bielefeld, Germany*

Received 2 September 2015; revised 23 November 2015; accepted 26 November 2015; published online 16 December 2015.
*For correspondence (e-mail Thomas.Schmidt@tu-dresden.de).

SUMMARY

Short interspersed nuclear elements (SINEs) are non-autonomous non-long terminal repeat retrotransposons which are widely distributed in eukaryotic organisms. While SINEs have been intensively studied in animals, only limited information is available about plant SINEs. We analysed 22 SINE families from seven genomes of the Amaranthaceae family and identified 34 806 SINEs, including 19 549 full-length copies. With the focus on sugar beet (*Beta vulgaris*), we performed a comparative analysis of the diversity, genomic and chromosomal organization and the methylation of SINEs to provide a detailed insight into the evolution and age of Amaranthaceae SINEs. The lengths of consensus sequences of SINEs range from 113 nucleotides (nt) up to 224 nt. The SINEs show dispersed distribution on all chromosomes but were found with higher incidence in subterminal euchromatic chromosome regions. The methylation of SINEs is increased compared with their flanking regions, and the strongest effect is visible for cytosines in the CHH context, indicating an involvement of asymmetric methylation in the silencing of SINEs.

Keywords: non-long terminal repeat retrotransposon, short interspersed nuclear elements (SINEs), Amaranthaceae, sugar beet, DNA methylation.

INTRODUCTION

Plants show large variations in their genome sizes, even among closely related species. An explanation for that is the C-value enigma, namely that the size of the genome does not correlate with the complexity of an organism. Transposable elements (TEs), which can exist in high copy numbers in plant genomes, are one of the reasons for that paradox (Baucom *et al.*, 2009; Lisch, 2013). TEs have been identified in all eukaryotic species investigated (Bennetzen and Wang, 2014) and often represent more than half of the nuclear genome (Baucom *et al.*, 2009; Khan *et al.*, 2011). They influence genome organization and gene expression by causing chromosome breaks, illegitimate recombination as well as genome rearrangement, and they are involved in epigenetic control of genes and genomes. (Slotkin and Martienssen, 2007; Hollister and Gaut, 2009; Rebollo *et al.*, 2010; Oliver *et al.*, 2013; Bennetzen and Wang, 2014).

Depending on the mode of transposition, TEs are grouped into two classes: class II elements that transpose from DNA to DNA and class I elements, or retrotransposons, which use reverse transcription of an RNA intermediate to reintegrate into the genome (Finnegan, 1989). Depending on the presence/absence of long terminal repeats (LTRs), retrotransposons are divided into LTR retrotransposons and non-LTR retrotransposons, which differ in their integration mechanism (Kolano *et al.*, 2013; Lisch, 2013). Long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) belong to the group of non-LTR retrotransposons (Wicker *et al.*, 2007). SINEs are well studied and highly abundant in animal and human genomes (Lander *et al.*, 2001) but are poorly investigated in plant genomes. Although SINEs make only a small contribution to the size of plant gen-

omes they can nevertheless reach high copy numbers due to their short length. SINEs have been identified in plant families like Fabaceae, Poaceae, Solanaceae and Brassicaceae, indicating that they are widespread in plants (Deragon and Zhang, 2006; Wenke *et al.*, 2011). Typically, SINE families have a limited taxonomic distribution; the Au element is the sole exception known so far, existing in many plant families like Gramineae, Solanaceae and Fabaceae (Fawcett *et al.*, 2006). SINEs are relatively short [80–500 nucelotides (nt)], non-autonomous retroelements which are mostly derived from tRNA gene sequences. They contain a region that is unrelated to tRNA and are usually terminated by a poly(A) or poly(T) stretch. The TS SINE detected in tobacco (*Nicotiana tabacum*) is the single exception thus far, since it was found to contain a poly (GTT) tail (Deragon and Zhang, 2006; Baucom *et al.*, 2009; Wenke *et al.*, 2011). SINEs are flanked by direct repeats of host DNA (target site duplications, TSDs) as they are inserted into the genome by transposition via target-primed reverse transcription by LINE-encoded enzymes (Zingler *et al.*, 2005). Plant SINEs are primarily dispersed with a preference to gene- and AT-rich regions, but are rarely present in heterochromatic, pericentromeric regions (Zhang and Wessler, 2005; Deragon and Zhang, 2006).

The Amaranthaceae is a family within the order Caryophyllales, which contains the subfamilies Betoideae and Chenopodioideae, including economically important crops like sugar beet (*Beta vulgaris* ssp. *vulgaris*), spinach (*Spinacia oleracea*) and quinoa (*Chenopodium quinoa*). The genus *Beta* includes three sections: the first section *Beta* consists of sugar beet with all cultivars and the wild species *Beta patula*, occurring only on a small island near Madeira. *Beta lomatogona* belongs to the second section, *Corollinae*, and the third section, *Nanae*, contains only *Beta nana* which grows at high altitudes in Greece. *Patellifolia procumbens*, formerly known as *Beta procumbens*, is a wild species, belonging to the genus *Patellifolia* and occurs on the northwest coast of Africa and on the Canary Islands. The subfamily Chenopodioideae includes the distantly related species *C. quinoa* and *S. oleracea*.

Here, we identified and analysed 22 SINE families within the Amaranthaceae with regard to their family structure, sequence diversity and insertion specificity and investigated the methylation of sugar beet SINEs by bisulphite sequencing. We show that SINE families were differently amplified during the evolutionary radiation of Amaranthaceae species. Fluorescent *in situ* hybridization was used to visualize the chromosomal distribution of SINEs.

RESULTS

Molecular structure of SINEs in sugar beet

We analysed the reference genome sequence RefBeet-1.1 (Dohm *et al.*, 2014) of sugar beet (genotype 'KWS2320', 596 Mb assembly size) to extract SINE families populating the genome using the SINE-Finder algorithm (Wenke *et al.*, 2011). SINE-Finder output sequences were grouped into separate SINE clusters. Consensus sequences of these SINE clusters were used as queries for BLAST searches to retrieve all remaining and possibly diversified or truncated copies from the genome sequence. The sugar beet reference sequence contained 6326 SINEs, including 4386 full-length and 1940 truncated copies. The truncated copies were subdivided into 5′-truncated and 3′-truncated SINEs (Table 1, Data S1 in Supporting Information). All families share typical structural SINE features such as RNA polymerase III promoter boxes A and B, poly(A/T) tails and TSDs. Based on a sequence similarity of at least 60%, SINEs were grouped into 19 AmaS (for Amaranthaceae SINE) families designated as AmaS-I to AmaS-XIX. We structured the detected SINEs into subfamilies to reflect their continuous evolution and diversification using the following criteria. Subfamilies were formed if the majority of copies share at least 70% similarity over their entire length. For example, SINEs in the AmaS-II family were grouped into five subfamilies, AmaS-IIa to AmaS-IIe. In contrast, SINEs in the families AmaS-X and AmaS-XI showed similarity of 77%, and the 3′ regions of both families were nearly identical while the 5′ regions showed no similarity, indicating a mosaic-like composition leading to the classification into two separate families instead of subfamilies. Besides the AmaS families we found also a single copy of the Au element.

The average sequence similarity within a family was determined based on the similarity of any full-length SINE copy to the derived consensus. Most families have similarities of between 80 and 90%; the highest similarity was determined for AmaS-XIX with 93%, while the families AmaS-VIa and AmaS-XVIII showed an averaged similarity of only 77%. The lengths of the consensus sequences range from 113 nt (AmaS-I) to 223 nt (AmaS-IX). AmaS families show high differences in the abundance, for example we found 1114 AmaS-I copies whereas AmaS-XIX consisted of only seven copies. Variability has also been found in the ratio of incomplete sequences which are truncated from either the 3′ or 5′ end or both ends. In AmaS-I only 10% of SINE copies were incomplete. In contrast, in AmaS-IVa 53% were 5′- and/or 3′-truncated.

We investigated the length of poly(A) tails and TSDs of all 4385 full-length AmaS SINEs (Figure 1a, b, Data S2). Comparison of poly(A) tails revealed average sizes of between 5 and 13 adenines (Table 1). The length of individual poly(A) tails is highly variable, ranging from 0 up to 48 adenines (Figure 1a). While the vast number of tails (1985 out of 4385) are between 7 and 10 nt long, the number of tails longer than 20 nt is very small. TSDs result from integration of SINEs, and TSDs shorter than 4 nt were not analysed because they are not reliable. The length of TSDs (4–36 nt) is variable, with the majority of copies ranging

Table 1 Short interspersed nuclear element (SINE) families in sugar beet

Family	Full-length SINEs	5′-truncated copies	3′-truncated copies	Similarity (%)[a]	Consensus (nt)[b]	poly(A/T) (nt)[c]
AmaS-I	1004	83	27	85	113	9 (0–32)
AmaS-IIa	220	62	74	81	199	8 (0–22)
AmaS-IIb	43	8	20	78	199	8 (0–21)
AmaS-IIc	163	34	58	87	174	11 (0–28)
AmaS-IId	187	15	101	78	210	8 (0–30)
AmaS-IIe	125	54	77	80	196	8 (0–19)
AmaS-III	210	37	26	92	196	13 (0–34)
AmaS-IVa	19	3	18	81	213	10 (1–31)
AmaS-V	399	7	291	80	192	9 (0–32)
AmaS-VIa	348	25	227	77	193	8 (0–27)
AmaS-VII	292	37	60	78	193	7 (0–21)
AmaS-VIII	16	1	6	85	183	9 (0–17)
AmaS-IX	191	70	56	84	223	10 (0–32)
AmaS-X	174	24	19	79	127	8 (0–48)
AmaS-XI	207	60	30	84	131	9 (0–30)
AmaS-XII	193	29	54	84	184	9 (0–25)
AmaS-XIII	194	33	21	92	160	12 (0–37)
AmaS-XIV	103	18	41	83	180	8 (0–18)
AmaS-XV	127	18	19	87	177	11 (0–31)
AmaS-XVI	59	5	31	91	185	11 (0–33)
AmaS-XVII	25	1	10	79	156	5 (0–16)
AmaS-XVIII	81	16	32	77	184	8 (0–20)
AmaS-XIX	5	0	2	93	127	11 (6–13)
Au	1	0	0	–	181	1
Total	4386	640	1300			

nt, nucleotides.
[a]Averaged identity of full-length SINEs to consensus.
[b]Consensus sequence without poly(A/T).
[c]Averaged length. Numbers in parentheses are minimum and maximum length.

between 4 and 20 nt (Figure 1b). We further investigated the target site preference of families that have at least 100 full-length copies with detectable TSDs. Thus, 2495 of the 4385 full-length SINEs corresponding to 11 families were analysed (Figures 2 and S1, Data S2, sequences marked with an asterisk). We examined four nucleotides of the flanking region upstream of the 5′ TSD (position −4 to −1) and the first four nucleotides of the 5′ TSD (position 1–4) of each SINE copy. Generally, the flanking region upstream of the TSD (position −4 to −2) is AT rich; furthermore, the four nucleotides that are part of the TSD (position 1–4) show a strong preference for an adenine or short adenine stretches. In contrast, the nucleotide directly upstream of the TSD (position −1) is variable and predominantly consists of T, G or C, but rarely of A.

As most plant SINEs are derived from tRNA genes, we compared the 5′ regions of all SINE families with 702 Viridiplantae tRNA genes (Jühling *et al.*, 2009) and detected no notable similarity to specific tRNA genes. However, we found significant similarities of boxes A and B of sugar beet SINEs to tRNA polymerase III promoters. The number of corresponding tRNA genes that have at least eight to eleven nucleotides identical to the consensus of box A and box B, respectively, is shown in Figure 3 for each SINE family. For example, in AmaS-V we found high sequence similarity (91–100%) of the box A motif to 200 tRNA genes. A total of 59 out of the 200 sequences share 100% similarity to box A. For box B, 359 tRNA genes share high similarity (91–100%) to the box B motif of AmaS-V, and 156 of them contain identical sequences. Generally, SINE B boxes show significantly more conservation to the corresponding motif of tRNA genes than SINE A boxes. The high similarities of SINE boxes A and B to tRNA polymerase III promoters suggest that AmaS-SINEs are ancestrally derived from tRNAs.

DNA methylation of sugar beet SINEs

The transposition of SINEs is usually suppressed in plants. A typical mechanism for the inactivation of transposable elements is DNA methylation of cytosine nucleotides. We investigated the cytosine methylation of the SINE body without poly(A) and TSDs of all sugar beet AmaS copies using high-throughput bisulphite sequencing. In addition, 100 flanking nucleotides up- and downstream of each SINE body sequence were analysed. A total of 6325 AmaS SINEs were investigated for methylation at symmetric CG and CHG (H = A, T, C) and asymmetric CHH sites separately on both DNA strands.

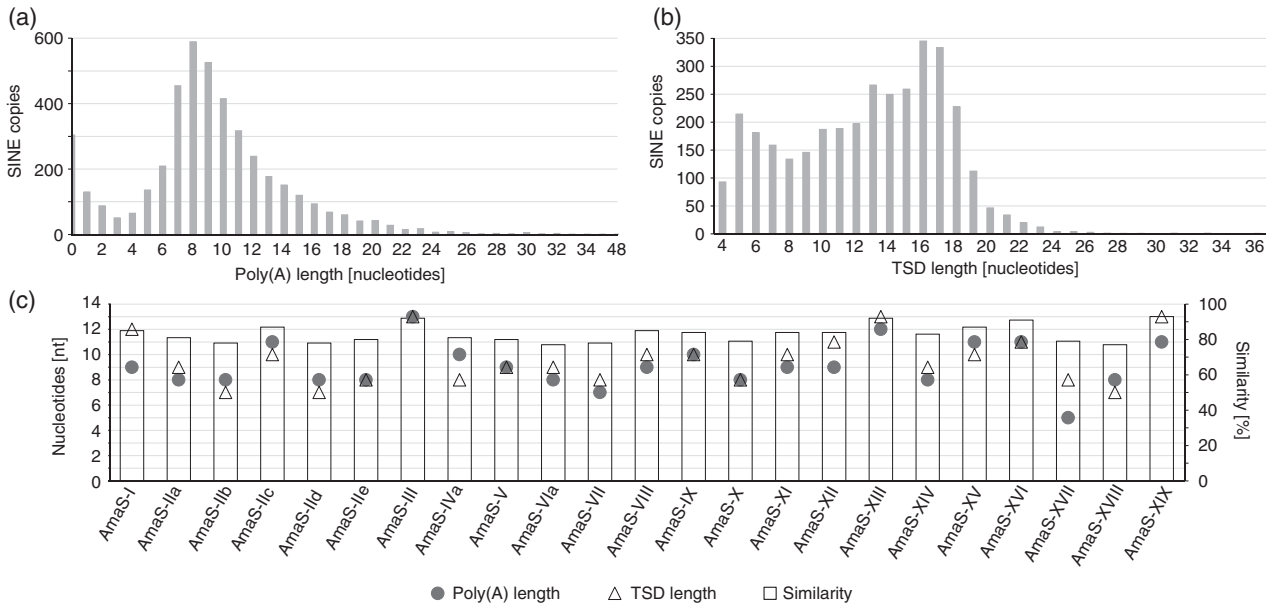

Figure 1. Length of poly(A) (a), target site duplication (TSD) (b) and comparison with the similarity to the consensus sequences (c) of full-length sugar beet AmaS short interspersed nuclear elements (SINEs).

Bisulphite-converted DNA from sugar beet leaves was subjected to next-generation sequencing representing about 11-fold coverage of the genome. The bisulphite reads were mapped against each genomic AmaS SINE copy. A total of 6242 reads were mapped without ambiguity, revealing 255 853 analysed cytosines. Of these, 208 793 cytosines occurred in the asymmetric CHH context, 26 688 were found in CHG sites and 20 372 in the CG context. The frequency of cytosine methylation was the highest for cytosines in the CG context (91%), followed by cytosines in the CHG context (83%). We found the lowest frequency of methylation (43%) for cytosines at CHH sites.

Expansion of the analysis to the flanking regions showed a lower overall methylation compared with SINEs. Mapping against SINEs and flanking regions yielded 80 556 uniquely mapping reads and a total of 3 015 374 analysed cytosines. Most cytosines occurred in the asymmetric CHH context (2 441 428), followed by CHG sites (291 142) and CG sites (282 804). The highest frequency for cytosine methylation was the CG context, accounting for 86%. Cytosines in the CHG context showed a methylation frequency of 76.8%. The lowest ratio of cytosine methylation was found at CHH sites, accounting for 39.1%.

A quantification trend blot displaying the methylation frequency of all cytosines along all SINE copies including flanking regions was generated (Figure 4). The frequency of cytosine methylation increases from approximately 50 nt upstream and downstream flanking regions, respectively, towards the SINE sequences for cytosines occurring in the CG motif (Figure 4a), the CHG motif (Figure 4b) and the CHH motif (Figure 4c). The strongest increase is detectable for CHH sites (Figure 4c). Individual quantification trend blots were generated for each SINE family (Figure S2). The methylation frequency is increased along the SINE sequences compared with the flanking regions. Again, the strongest increase in methylation is observable for the cytosines in the CHH motif. This suggests a crucial role of asymmetric methylation in silencing of SINE copies in the sugar beet genome.

Diversification and evolution of SINE families within the Amaranthaceae

To investigate the distribution of SINE families within species related to sugar beet we analysed species of the subfamilies Betoideae and Chenopodioideae. We used genome sequence data of *B. patula* with a draft genome assembly size of 607 Mb, *B. lomatogona* (assembly size 775 Mb) and *B. nana* (assembly size 513 Mb) as representative species of the genus *Beta* sections *Beta*, *Corollinae* and *Nanae*, respectively; *P. procumbens* (assembly size 695 Mb) represents the genus *Patellifolia*. As distantly related species we chose *C. quinoa* (assembly size 1.4 Gb) and *S. oleracea* (assembly size 661 Mb). These genomes were assembled *de novo* using Illumina paired end and mate pair sequences and will be published elsewhere. The spinach genome is already publicly available (Dohm *et al.*, 2014).

Identification and grouping of SINEs was carried out as described for sugar beet. The absence of SINE families in single species was confirmed by PCRs with family-specific primers.

We identified 34 810 copies containing 19 549 full-length SINEs in the Amaranthaceae species including sugar beet

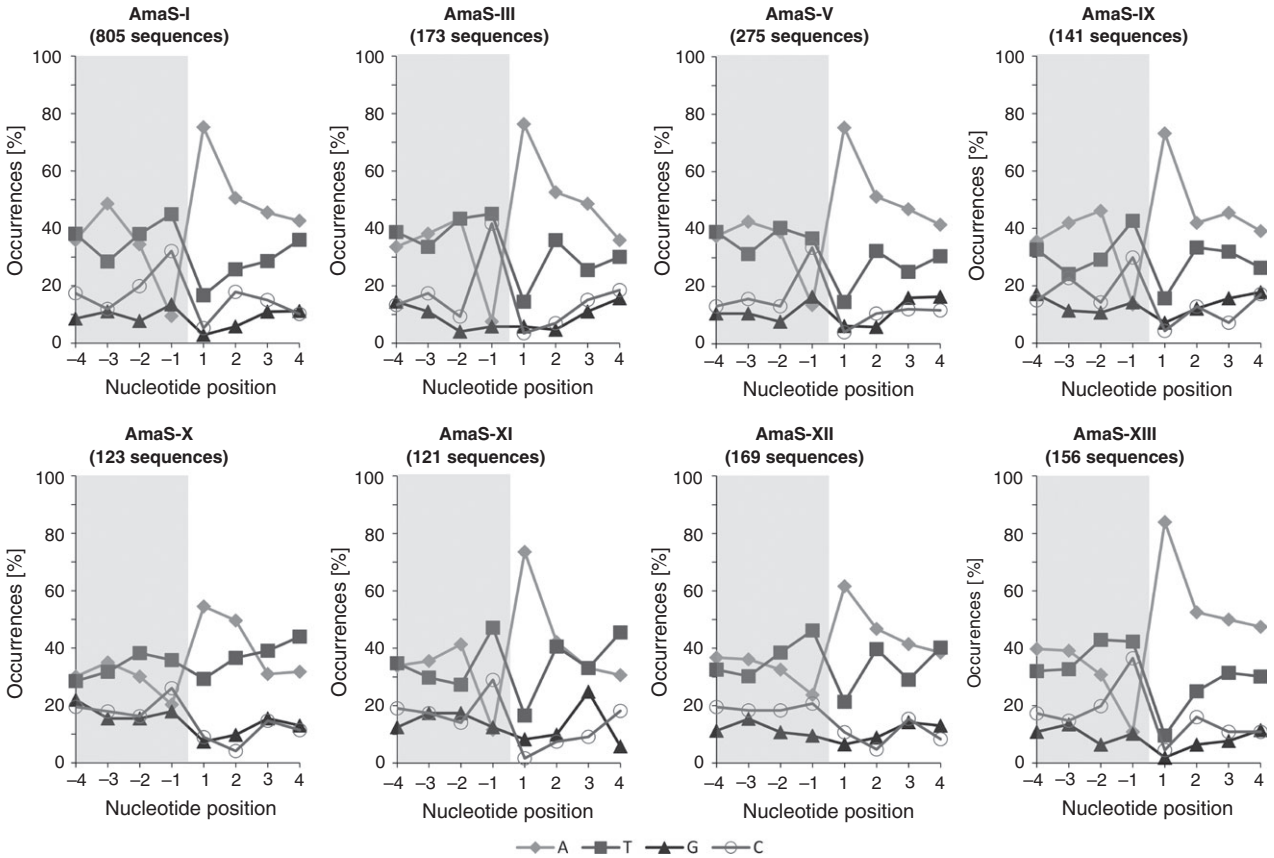

Figure 2. 5′ insertion site preference of sugar beet AmaS short interspersed nuclear elements (SINEs).
The figure shows nucleotide frequencies at the 5′ nicking sites of AmaS-I, AmaS-III, AmaS-V, AmaS-IX, AmaS-X, AmaS-XI, AmaS-XII and AmaS-XIII. The grey shaded positions are nucleotides of the flanking DNA upstream of the 5′ target site duplication (TSD). Positions 1 to 4 mark the first four nucleotides of the 5′ TSD.

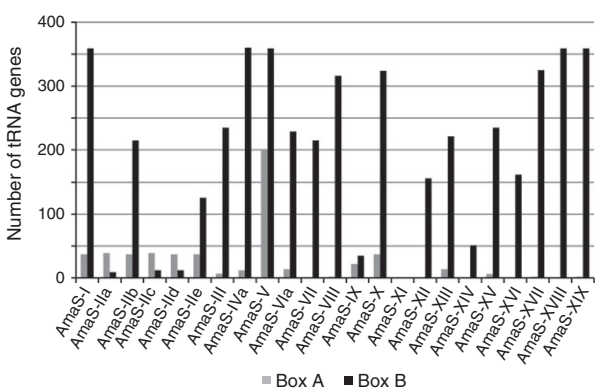

Figure 3. Similarity of sugar beet short interspersed nuclear elements to Viridiplantae tRNA genes.
Based on similarity (a minimum of eight out of eleven nucleotides), the RNA polymerase III promoter boxes A and B of 702 tRNA genes of Viridiplantae were assigned to the corresponding boxes A and B of the AmaS consensus sequences of sugar beet.

(Table 2, Data S1). The lengths of consensus sequences range from 113 nt (AmaS-I in sugar beet and *C. quinoa*) to 224 nt (AmaS-IX in *B. lomatogona* and *B. nana*).

In order to verify the computationally identified SINE copy numbers (see Table 2) we performed comparative Southern blot hybridization using as a representative example a probe of the widespread sugar beet AmaS-I family. Blot hybridization revealed different intensities in all species. AmaS-I is highly abundant in the section *Beta*, moderately amplified in *B. lomatogona* and spinach and exists in low copy numbers in *Patellifolia*, *C. quinoa* and *S. oleracea* (Figure S3). The Southern hybridization results are consistent with the abundance of SINE families retrieved from genome sequences (Table 2).

To visualize the classification and divergence of the SINE families, we constructed an unrooted dendrogram of the 10 most representative copies of each family with the lowest divergence from the consensus sequence (see Figure 5, Data S3). This dendrogram serves only as visualization of the diversity of families and represents no phylogenetic relationships. All SINE families form separate branches. However, subfamilies group together, verifying their relationships. AmaS-X and AmaS-XI do not group together, but have a mosaic-like composition and are evolutionary related. A similar composite structure has also been

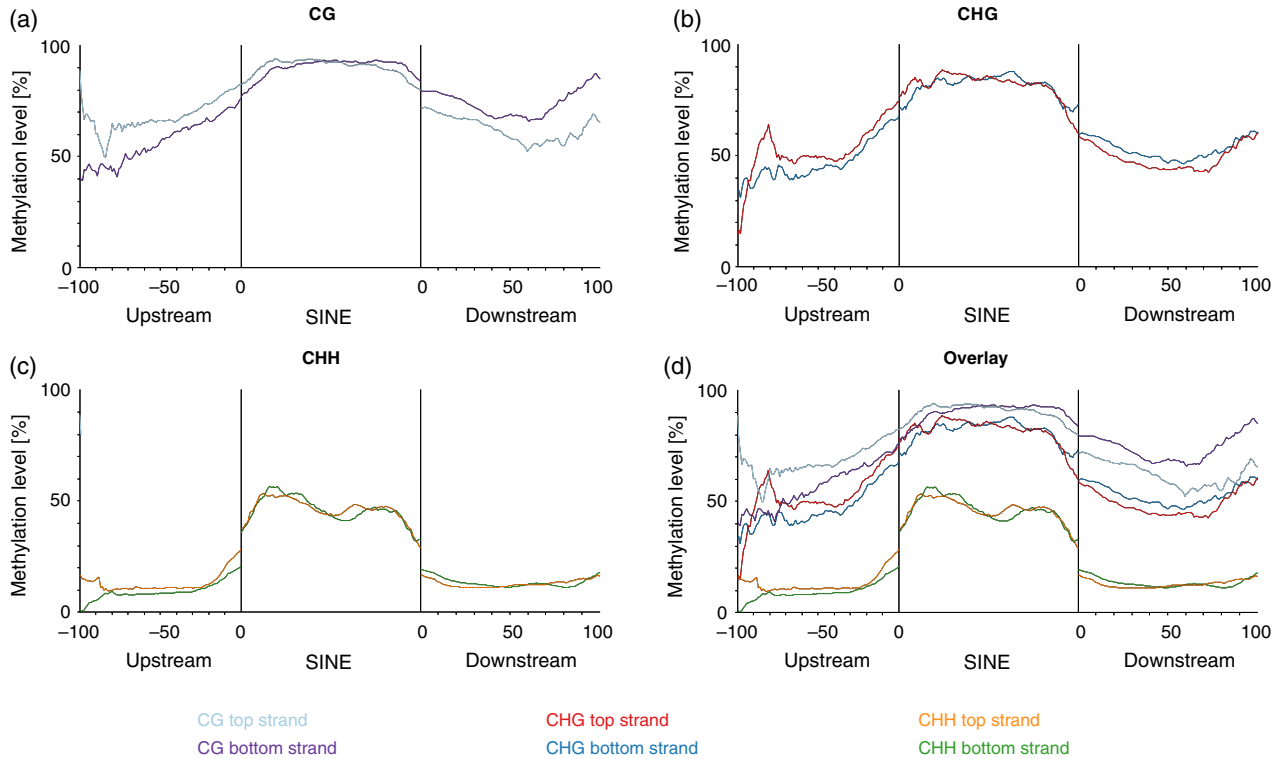

Figure 4. Methylation of 6325 sugar beet AmaS short interspersed nuclear elements (SINEs) and 100 nucleotides up- and downstream of their flanking region. (a) CG, (b) CHG, (c) CHH and (d) overlay.

detected in the families AmaS-V, AmaS-IX and AmaS-XX. AmaS-XX shares the 5′ region with AmaS-IX, but the 3′ region shows a high similarity to AmaS-V. Alignments for these families are shown in Figure S4. No evolutionary relationship could be detected in any of the other families.

Generally, species of the genus *Beta* contain a higher number of SINE families (18–21 families) than *P. procumbens* (12 families), *C. quinoa* (10 families) and *S. oleracea* (8 families). Species-specific SINE families or subfamilies (shaded in Figure 5) were found in all analysed species except for *B. patula* and *B. nana*.

By inspection of the data (Table 2) we found six SINE families existing in all analysed species of the Amaranthaceae, indicating conservation of families across distant taxonomic borders and therefore most likely to be of ancient origin. The species-specific percentage abundance of these six SINE families is shown in Figure 6(a). Considering only those six families, AmaS-I is highly abundant in the genus *Beta*, representing more than 50% of these SINEs in sugar beet, *B. patula* and *B. nana*, whereas only a smaller portion of AmaS-I SINEs are present in distantly related species. In contrast, AmaS-XIII has a higher abundance in *P. procumbens*, *C. quinoa* and *S. oleracea* than in the genus *Beta*. This analysis showed that the amplification of SINEs only roughly follows taxonomic grouping but shows a discordant distribution.

Comparative physical mapping and chromosomal assignment of SINEs

In order to investigate the chromosomal distribution of Amaranthaceae SINEs, comparative fluorescent *in situ* hybridization (FISH) was performed. We used the widespread AmaS-I as probe to compare the chromosomal organization of SINEs in Amaranthaceae species. We observed dispersed signals on mitotic metaphase chromosomes in all analysed species. Increased signals in distal to subterminal euchromatic chromosome regions indicate a local accumulation of SINEs (Figure 7a–g), detected in particular for sugar beet (Figure 7a) and spinach (Figure 7g). In other chromosomal regions such as the centromeres the SINE copy number is reduced and SINEs are mostly excluded from sites of 18S-5.8S-25S rRNA genes (Figure 7a, b, d), but no region is completely depleted of SINEs. Furthermore, we localized AmaS-II comparatively to provide an insight into different genera of Amaranthaceae. The probe used for hybridization with AmaS-II is not specific for subfamilies, therefore signals detected include AmaS-IIa to AmaS-IIe. In both sugar beet and *P. procumbens*, AmaS-II is dispersed on all chromosomes and mostly excluded from heterochromatic regions (Figure 7h, i) and rDNA sites of sugar beet (Figure 7h). A dispersed pattern on chromosomes was also observed for AmaS-XIII. Here, hybridization was carried out on sugar beet and the

Table 2 Short interspersed nuclear element (SINE) families in Amaranthaceae

Family	*Beta vulgaris*	*Beta patula*	*Beta lomatogona*	*Beta nana*	*Patellifolia procumbens*	*Chenopodium quinoa*	*Spinacia oleracea*
AmaS-I	1114 (85%)	1545 (88%)	633 (91%)	265 (90%)	44 (78%)	47 (77%)	510 (79%)
AmaS-IIa	356 (80%)	0	0	0	0	0	0
AmaS-IIb	71 (78%)	25 (81%)	239 (83%)	287 (83%)	0	265 (81%)	240 (79%)
AmaS-IIc	255 (87%)	245 (87%)	242 (93%)	0	1044 (88%)	0	0
AmaS-IId	303 (78%)	427 (78%)	487 (85%)	141 (88.2%)	0	0	0
AmaS-IIe	256 (80%)	395 (78%)	276 (87%)	92 (86%)	330 (84%)	3 (–)	77 (77%)
AmaS-III	273 (92%)	748 (89%)	657 (95%)	0	0	0	0
AmaS-IVa	40 (81%)	433 (78%)	472 (87%)	254 (92%)	1 (–)	158 (79%)	0
AmaS-IVb	0	0	0	0	0	0	139 (80%)
AmaS-V	697 (80%)	736 (81%)	485 (87%)	238 (92%)	0	0	0
AmaS-VIa	600 (77%)	369 (81%)	727 (90%)	553 (91%)	0	0	0
AmaS-VIb	0	0	0	0	263 (86%)	0	0
AmaS-VII	389 (78%)	493 (79%)	530 (77%)	658 (78%)	0	0	0
AmaS-VIII	23 (85%)	24 (84%)	8 (98%)	5 (98%)	246 (90%)	0	0
AmaS-IX	317 (84%)	241 (83%)	401 (88%)	516 (90%)	0	0	0
AmaS-X	217 (79)	206 (80%)	368 (81%)	728 (81%)	16 (83%)	129 (78%)	0
AmaS-XI	297 (84%)	597 (80%)	652 (83%)	148 (89%)	0	0	0
AmaS-XII	276 (84%)	290 (84%)	193 (94%)	29 (95%)	245 (93%)	12 (90%)	381 (83%)
AmaS-XIII	248 (92%)	264 (93%)	425 (97%)	3 (–)	1109 (92%)	308 (84%)	511 (85%)
AmaS-XIV	162 (83%)	203 (85%)	176 (87%)	70 (85%)	80 (83%)	186 (78%)	102 (84%)
AmaS-XV	164 (87%)	167 (86%)	52 (82%)	50 (87%)	296 (79%)	0	0
AmaS-XVI	95 (91%)	123 (88%)	92 (94%)	0	0	0	0
AmaS-XVII	36 (79%)	44 (77%)	707 (92%)	182 (85%)	0	0	0
AmaS-XVIII	129 (77%)	134 (75%)	545 (86%)	689 (87%)	0	0	0
AmaS-XIX	7 (93%)	5 (93%)	4 (90%)	7 (90%)	16 (88%)	5 (89%)	0
AmaS-XX	0	0	478 (96%)	0	0	0	0
AmaS-XXI	0	0	0	0	0	183 (71%)	0
Au	1	1	5	2	6	5 (94%)	41 (81%)
Total	6326	7715	8854	4917	3696	1301	2001

The table shows all copies of SINE families in Amaranthaceae.
Values in parentheses represent the average similarity of full-length SINEs to the consensus sequence.

distantly related species *C. quinoa* and spinach (Figure 7j–l). The strongest signals were detected in spinach, and again SINEs were excluded from 18S-5.8S-25S rDNA sites (Figure 7l).

Next, we analysed in detail the chromosomal distribution of the SINE families investigated by FISH and of all sugar beet AmaS SINEs with computational methods using the most recent sugar beet assembly RefBeet-1.2 (which has improved long-range continuity despite the same sequence content as RefBeet-1.1). Consistent with the FISH results AmaS-I, -II and -XIII show a dispersed pattern (Figure 8a–c). All SINE families show a dispersed distribution with a preference for distal to subterminal chromosome regions on all chromosomes with arm-specific differences in SINE density. On chromosome 1, 2 and 9 the abundance of SINEs is higher on one arm, while the remaining chromosomes show a similar abundance of SINEs on both chromosome arms (Figure 8d).

Estimation of the age and transpositional activity of SINEs

We estimated the age of SINEs by comparing the length of TSDs and 3′ tails. Both are indicators for the relative insertion time of individual SINE copies (Roy-Engel *et al.*, 2002; Gentles *et al.*, 2005). Typically, old SINEs have shorter poly (A) tails and diverged TSDs caused by substitutions over time compared with younger copies. We correlated the lengths of poly(A) tails and TSDs with the sequence similarity between the members of a SINE family and found a clear correlation. The higher the average similarity of a family, the longer are the poly(A) tracts and the TSDs. For example, AmaS-III has a sequence similarity of 92% and an average poly(A)- and TSD-length of 13 nt each, indicating that this is a family which emerged or expanded relatively recently. In contrast, AmaS-XVIII (77% similarity) (see Figure 1c) has an average poly(A) length of 8 nt and a TSD length of 7 nt, suggesting that most SINEs of this family are relatively ancient.

A possible way to estimate the activity of a SINE family relates the copy numbers to sequence similarity intervals. Usually, ancient families have high copy numbers and low similarities due to accumulated mutations. However, transpositional bursts of a SINE family can result in high similarities in a large number of copies even if the family is old. Therefore, the average similarity could not make a

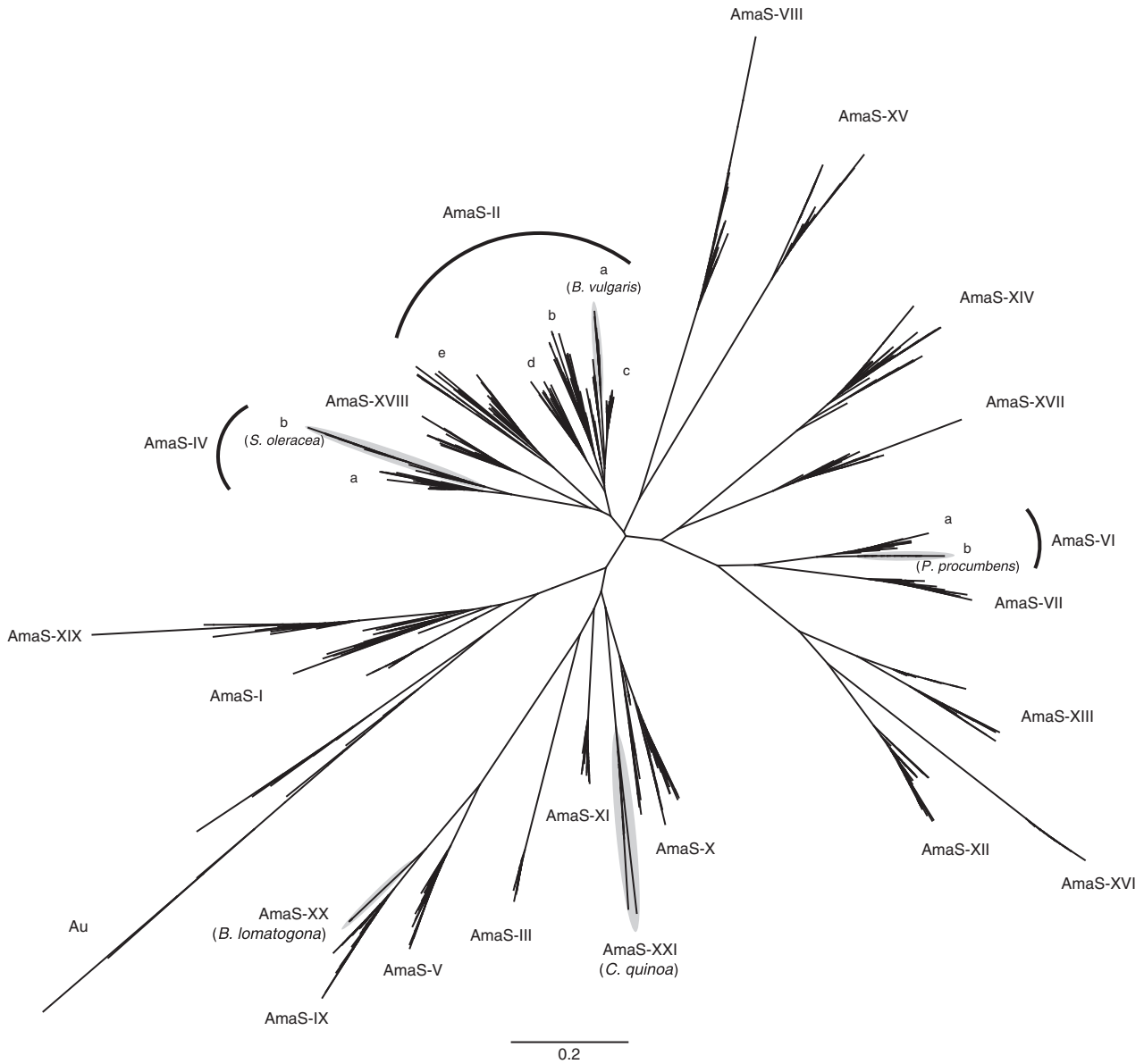

Figure 5. Dendrogram of representative copies showing the highest similarity to the consensus sequences of the short interspersed nuclear element families in Amaranthaceae.
Species-specific families and subfamilies are shaded grey. Letters a to e correspond to subfamilies. Subfamilies are embraced by curves.

clear prediction of the age of a family, but nevertheless it can be a good indication in combination with other features like species phylogeny or TSDs and 3′ tails. Figures 9 and S5 combine the copy numbers and similarity to the consensus sequence of all AmaS SINE families in all investigated species. Figure 9(a) shows typical examples for different transpositional behaviours. In some families, for example AmaS-XXI in *C. quinoa*, SINEs were active a long time ago, as they show high copy numbers and most copies have similarities lower than 80%. Other families indicated consistent activity over a long period, as suggested by similarity levels detected for AmaS-XV in *P. procumbens*. The similarities range between 60 and 100% and the number of copies in each similarity interval is relatively consistent. An additional mode of amplification is a transpositional burst, as detected for AmaS-I in *B. patula*. While 593 copies have low similarities of 60–80%, indicating a greater age of this fraction, 304 of the remaining 952 copies show high similarities of about 90% which indicates that these consist of younger SINEs. In Figure 9(b) and (c) the similarity and copy number of a probably recent family (AmaS-III) and an ancient family (AmaS-I) are shown. Most AmaS-III copies (95% in sugar beet, 82% in *B. patula* and 71% in *B. lomatogona*) show similarities

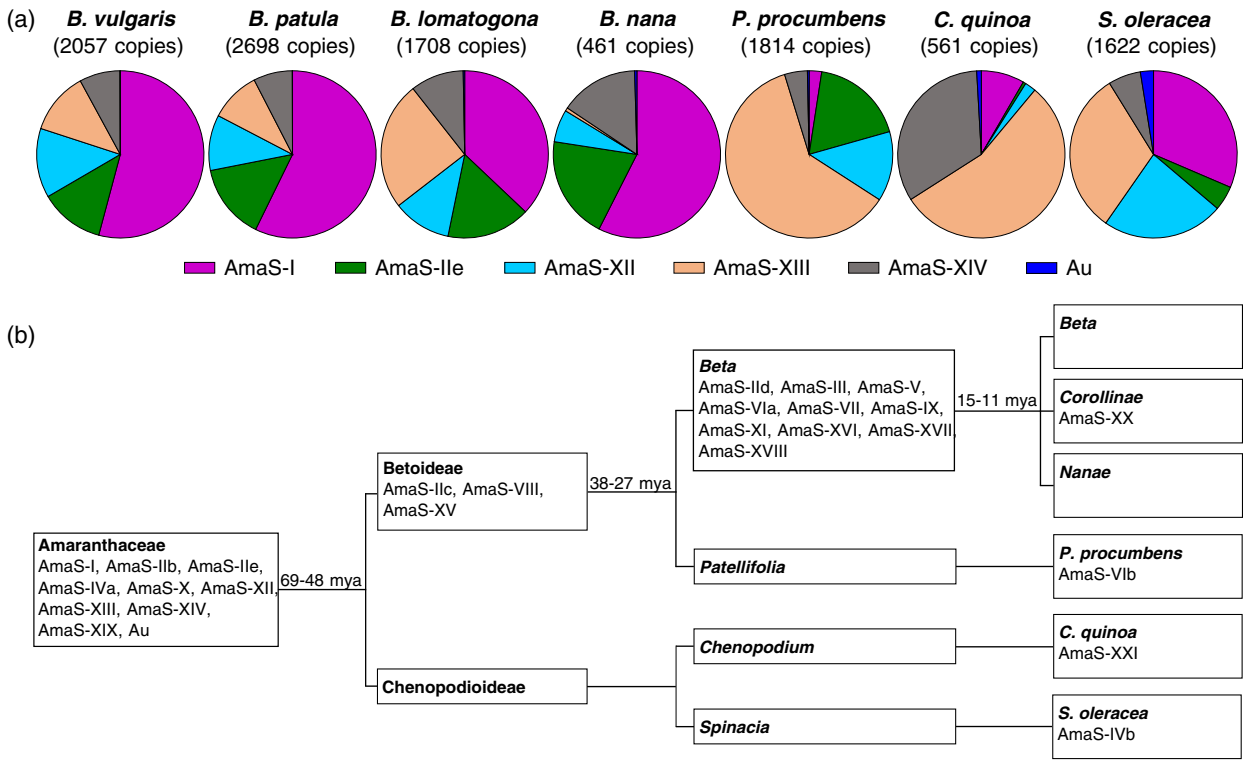

Figure 6. Distribution and phylogeny of short interspersed nuclear element (SINE) families within the Amaranthaceae.
(a) Distribution of conserved SINE families. The figure shows copy numbers of AmaS-I, AmaS-IIe, AmaS-XII, AmaS-XIII, AmaS-XIV and Au for *Beta vulgaris, Beta patula, Beta lomatogona, Beta nana, Patellifolia procumbens, Chenopodium quinoa* and *Spinacia oleracea*.
(b) A scenario of the evolution of SINE families within the Amaranthaceae. The age of the different groups has been estimated in millions of years ago (mya) by Hohmann *et al.* (2006).

to the consensus sequence of at least 80% in all three species (Figure 9b), indicating a recent age or activity of this family. AmaS-I was predicted to be an old family due to its presence in all investigated species of Amaranthaceae. 82% of the copies have a similarity of at least 80% in sugar beet (Figure 9c), but only 30% of the copies have such high similarities in *C. quinoa*, consistent with the high copy numbers in sugar beet and low copy numbers in *C. quinoa*. The more recent activity of AmaS-I in sugar beet, recognizable in higher copy numbers and higher similarity, suggests that AmaS-I proliferated in sugar beet but not in *C. quinoa* after the radiation of Amaranthaceae.

Based on the evolutionary radiation of the Amaranthaceae into Betoideae and Chenopodioideae (Hohmann *et al.*, 2006), we estimated the minimum age of some families. Those SINE families not detectable in some species were assigned to the last common ancestor. The families AmaS-I, AmaS-IIb, AmaS-IIe, AmaS-IVa, AmaS-X, AmaS-XII-XIV, AmaS-XIX and Au exist in all species of the family Amaranthaceae and emerged in the last common ancestor at least 69–48 million years ago (mya) (Figure 6b). In contrast, AmaS-XX was found only in the section *Corollinae* and hence is probably younger than 15–11 my.

The Au SINE family shows a broad distribution in plants. It exists in all Amaranthaceae species and is the only SINE family in the Amaranthaceae with a poly(T) tail. Copy numbers range from one copy in sugar beet, *B. patula* and *B. lomatogona* to 41 copies in spinach. For similarity analysis we used consensus sequences of *P. procumbens*, *C. quinoa* and spinach Au elements. No consensus sequences were available for the other species due to the low copy numbers or incompleteness of Au copies, and therefore we used the conserved region of the longest detected copy. The Au elements of all analysed species of the Amaranthaceae have relatively low averaged similarities with the Au consensus sequence of other plants (Fawcett *et al.*, 2006), ranging from 68% in *P. procumbens* and spinach to 76% in *B. patula*. Interestingly, we found identical or nearly identical TSDs and flanking regions of individual Au copies in the genus *Beta* (Figure S6), therefore this copy is most likely located at the same locus. This indicates that this Au copy existed before the division of the genus *Beta* into different sections, and according to Hohmann *et al.* (2006) integrated at least 15–11 mya. BLAST searches of the SINE with up- and downstream flanking regions on the sugar beet genome (http://bvseq.molgen.mpg.de) revealed that the Au is located in an intron of

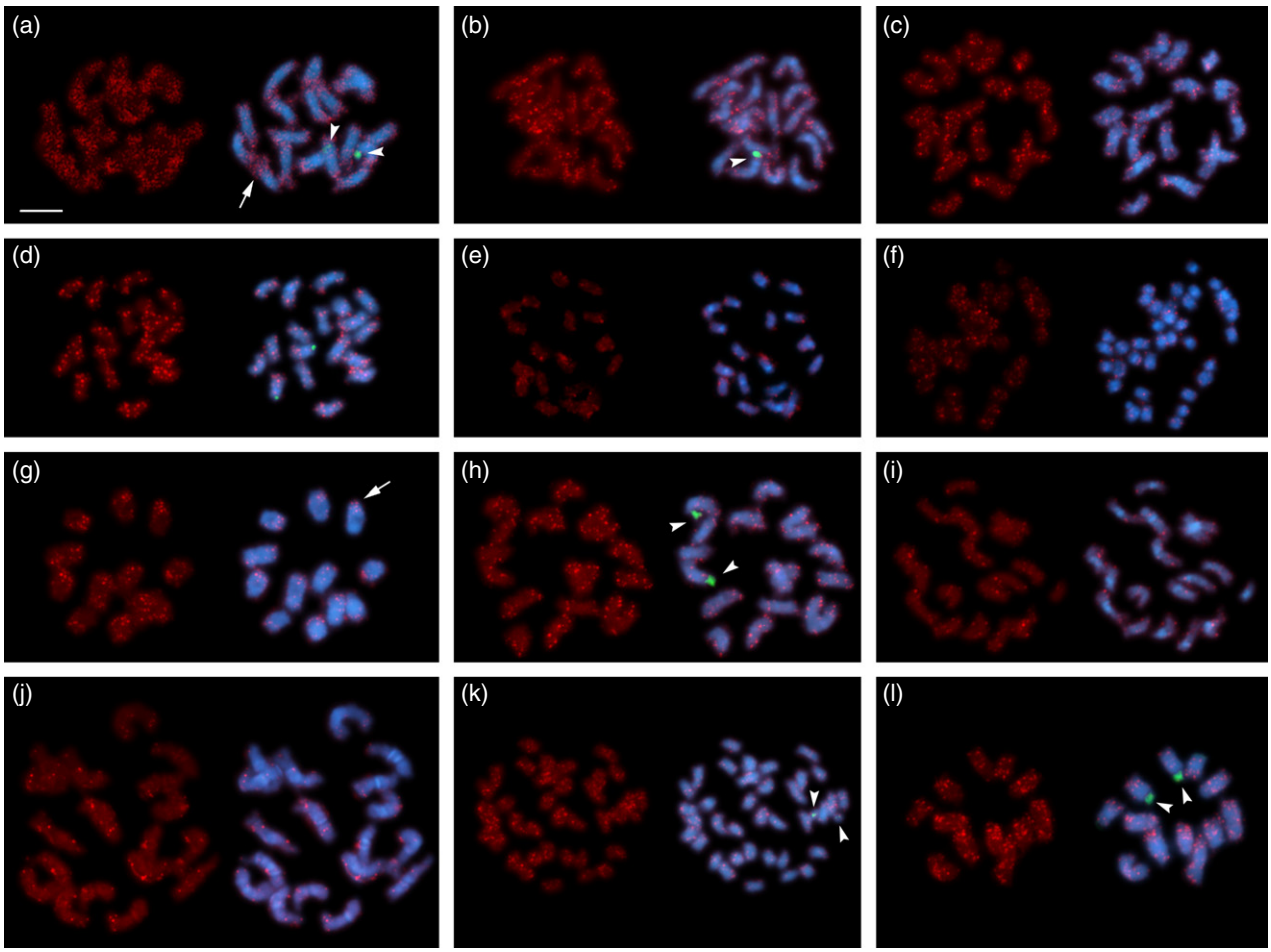

Figure 7. Fluorescent *in situ* hybridization of short interspersed nuclear elements (SINEs) in Amaranthaceae.
4′,6-Diamidino-2-phenylindole (DAPI) staining (blue fluorescence) shows the chromosomal morphology. Heterochromatic regions are visible as strong DAPI signals. Red signals are sites of SINE hybridization and green signals (in a, h, k, i) label the 18S-5.8S-25S rRNA genes (arrowhead). AmaS-I is localized on sugar beet (a), *Beta patula* (b), *Beta lomatogona* (c), *Beta nana* (d), *Patellifolia procumbens* (e), *Chenopodium quinoa* (f) and *Spinacia oleracea* (g) metaphase chromosomes. SINEs have a dispersed distribution over all mitotic metaphase chromosomes with a preference for distal to terminal euchromatic regions (examples arrowed). AmaS-II is shown on metaphase chromosomes from sugar beet (h) and *P. procumbens* (i). The probe used for hybridization is unspecific for subfamilies. AmaS-XIII was mapped on sugar beet (j), *C. quinoa* (k) and *S. oleracea* (l) metaphase chromosomes. The scale bar in panel (a) corresponds to 5 μm.

a gene, which is predicted as putative ubiquitin-protein ligase gene Bv_47830_mkjc (Dohm *et al.*, 2014). The location of the SINE in a gene may be the reason for the survival of Au in the genus *Beta*. The coding region of the gene and the SINE region is highly conserved in the genus *Beta*, whereas the 5′ and 3′ untranslated regions of the gene are diverged.

DISCUSSION

We identified 34 810 SINE copies of 22 families in the genomes of Amaranthaceae species. Application of the SINE-Finder algorithm turned out to be crucial for the identification of SINE families. Analysis with AmaS SINEs as a query in the NCBI database revealed their absence in unrelated plant genomes, and BLAST searches with SINEs from other plant species, with exception of the Au element, retrieved no copies in *Beta* species. Hence, AmaS SINEs are independent and restricted to the Amaranthaceae, and existence in related species is typical for plant SINEs. The number of families is in the same range as in other plant species: Twelve families were identified in rice and Brassicaceae contain 15 different families (Deragon and Zhang, 2006; Khan *et al.*, 2011). Three families, AmaS-II, AmaS-IV and AmaS-VI, were subdivided into subfamilies. This is similar to SINEs in the Solanaceae, where two out of nine families contain subfamilies (Wenke *et al.*, 2011), indicating their ongoing evolutionary diversification.

AmaS SINEs are characterized by typical features of plant SINEs: a tRNA related 5′ region containing the promoter boxes A and B, a tail represented by poly(A/T) stretches or simple sequence repeats putative and flanking TSDs. Comparison of the family consensus sequences shows that 11 out of 21 AmaS SINE families begin with 5′-ACCAA-3′ or have at least this conserved motif in their 5′

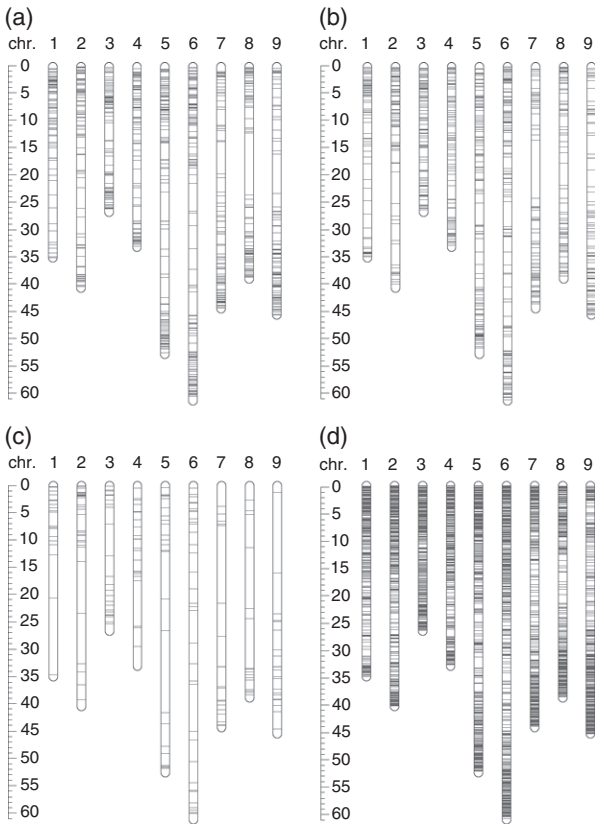

Figure 8. Chromosomal mapping of sugar beet AmaS short interspersed nuclear elements (SINEs).
Chromosomal mapping of AmaS-I (a), AmaS-II (b), AmaS-XIII (c) and all AmaS SINEs (d) in sugar beet. Scale provided in Mbp.

region (Data S4), suggesting conservation during evolution. A similar observation was made for the *BoS* elements of *B. oleracea*, where nearly all families contain the motif 5′-AACCRRG-3′ (R = A or G) (Zhang and Wessler, 2005). The vast majority of AmaS SINE families have a length below 200 nt which is in line with other plant SINEs (Gadzalski and Sakowicz, 2011; Wenke *et al.*, 2011; Ben-David *et al.*, 2013).

SINE families can achieve high copy numbers in small genomes; for example 13 487 copies, representing approximately 1.8% of the genome (430 Mbp) were found in rice (Khan *et al.*, 2011). We identified 6326 SINEs in the sugar beet genome sequence (596 Mbp genome assembly) and estimate a copy number of about 7800 SINEs in the whole genome (estimated genome size of 731 Mbp; Dohm *et al.*, 2014). Based on this information, sugar beet SINEs cover about 0.18% of the genome. This is in a similar range to that in *B. oleracea* (0.16%) and potato (0.15%) (Deragon and Zhang, 2006; Wenke *et al.*, 2011). In contrast, *Arabidopsis thaliana* SINEs represent only 0.05% of the genome (Deragon and Zhang, 2006) and the maize genome (2300 Mbp) consists of 0.02% SINEs (Baucom *et al.*, 2009). However, the maize genome is three times larger than that of sugar beet, so there is no stringent correlation between the number of SINEs and genome size. SINEs seem to occur typically in high copy numbers, but due to their small size they have only a minor influence on the variation in genome size in plants, which is in contrast to human and animal genomes. The genus *Beta* showed higher overall copy numbers compared with other Amaranthaceae genera, indicating a higher activity of SINEs in the genus *Beta*. This may result from accumulated mutations, leading to a more efficient mobilization by LINEs and therefore a higher copy number in *Beta*. Other reasons for the higher copy number could be the quality or quantity of the genome sequences, the length of assembly or the diversity of SINEs.

Generally, we observed a high number of families in the genus *Beta* and a decreasing number in distantly related species. One hypothesis for this is the impact of the Ice Age on the species distribution and radiation of the Amaranthaceae into Betoideae and Chenopodioideae (Figure 6b). Glaciation resulted in lower global temperatures. Thus, due to the higher environmental stress level Betoideae may have tolerated more SINE families. Another aspect supporting this theory is the activation of LINE families in the genus *Beta* but not in *Patellifolia* and the subfamily Chenopodioideae (Heitkam *et al.*, 2014). The climatic stress may have caused the reactivation of repetitive elements.

Usually, ancient SINE copies have a high sequence diversity caused by errors of the LINE reverse transcriptase or by nucleotide substitutions and insertions/deletions after retrotransposition. The presence of six AmaS families and Au in all analysed species of the family Amaranthaceae indicates an ancient age of at least 69–48 million years for these SINE families, according to the division of Amaranthaceae into Betoideae and Chenopodioideae (Figure 6b). After the division, a transpositional burst may have occurred in some species, resulting in a higher sequence similarity and high copy numbers of SINEs in these species. It is tempting to speculate that some AmaS SINEs are very ancient sequences because they have been found in *Beta* and *Chenopodium*. Quinoa occurs in South America and the SINEs may have existed in the progenitor of both genera. Other SINE families like AmaS-V or AmaS-VII probably did not evolve before the division of Betoideae into *Beta* and *Patellifolia*. This should be verified by the analysis of additional Amaranthaceae species.

Some SINE families exist in all investigated species of the genus *Beta* but not in other genera such as *Patellifolia*. This indicates that these families arose in the genus *Beta*, after the radiation of Betoideae into *Beta* and *Patellifolia*. We postulate that the absence of SINE families in the species distribution is the result of the evolutionary extinction and decay of SINE families in some species. For example,

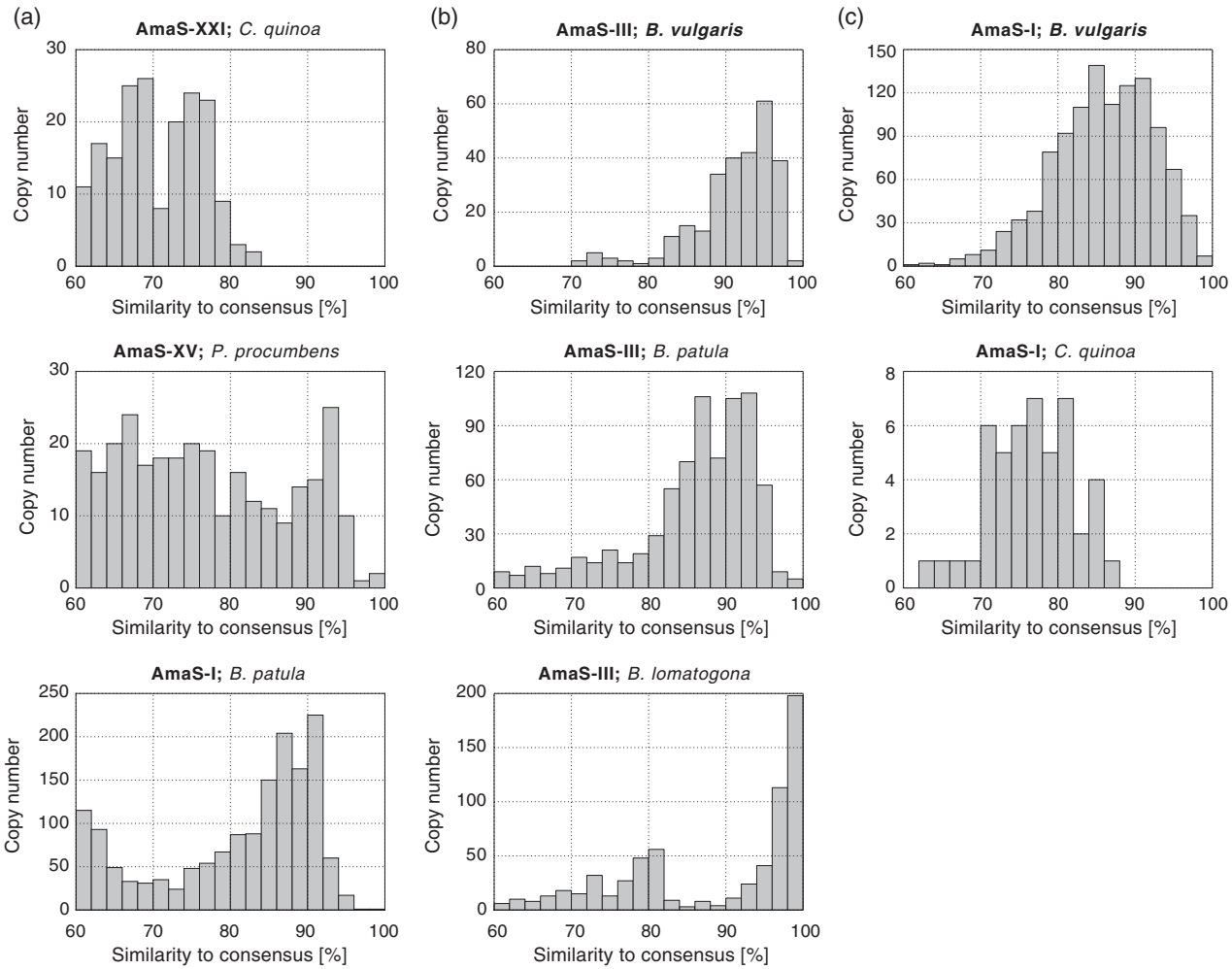

Figure 9. Comparison of copy numbers and similarity of AmaS short interspersed nuclear element (SINE) families.
(a) Examples of transposition patterns: transpositional activity a long time ago, AmaS-XXI in *Chenopodium quinoa*; consistent transpositional activity over a long period, AmaS-XV in *Patellifolia procumbens* and a transpositional burst, AmaS-I in *Beta patula*.
(b) Copy numbers and similarity of AmaS-III in sugar beet, *Beta patula* and *Beta lomatogona*.
(c) Copy numbers of AmaS-I in sugar beet and *C. quinoa*.

AmaS-X is present in all species except *S. oleracea* and has relatively low similarities, ranging from 78% in *C. quinoa* to 83% in *P. procumbens*, indicating that transposition occurred long ago. It is possible that AmaS-X existed in spinach or a progenitor of spinach in low copy numbers, but no transpositional burst or continuous low level of amplification occurred in this species, resulting possibly in only few copies. Due to nucleotide substitutions over time these few copies are no longer recognizable as SINEs (Huang *et al.*, 2012), whereas in other species the family is still present. Similarly, AmaS-III and AmaS-XVI are present in sugar beet, *B. patula* and *B. lomatogona*, but are absent in *B. nana*. The families are amplified in the sections *Beta* and *Corollinae*, but not in *Nanae*, which is perhaps the result of divergence and the impossibility of detecting these families in *B. nana*. Therefore, knowledge on the distribution of a SINE family alone does not enable any inference about the relationships of species. In theory, during radiation all existing SINE families are inherited from ancestral species. Following speciation, a family can diversify completely and the occurrence of mutations or rearrangements can result in subfamilies. The development of subfamilies could occur on two levels: mutations in integrated genomic SINEs or alternatively during reverse transcription by LINE reverse transcriptase which is prone to errors (Grandbastien, 1998; Wenke *et al.*, 2009). AmaS-II is a good example of the evolution of subfamilies. An alignment of AmaS-II in sugar beet is shown in Figure S7. Due to the existence of AmaS-IIe in all species we believe that this family is the oldest subfamily, and therefore the founder family. AmaS-IIb exists in all species except *P. procumbens* and we assume that AmaS-IIb was present in *P. procumbens* too, but due to mutations over time SINEs of this family are no longer detectable. AmaS-

IIb has few substitutions in the 5′ and 3′ regions compared with AmaS-IIe, while AmaS-IIc has a core region similar to AmaS-IIb but substitutions in the 5′ region and deletions in the 3′ region. AmaS-IIc only appears in the Betoideae, indicating an age younger than AmaS-IIb. AmaS-IIa is similar to AmaS-IIb, but has substitutions in the 3′ region. This family only exists in sugar beet and therefore is the youngest subfamily.

Comparison of SINE families existing across species (Figure 6a) revealed extreme differences among species. On the one hand SINE families can behave similarly in closely related species, but on the other hand some families are amplified differently in closely related species, as also shown for Solanaceae (Wenke *et al.*, 2011).

The Au element is a ubiquitous SINE family in plants and was detected with low copy numbers in all analysed species of the Amaranthaceae. This widespread distribution in plants indicates that Au is an evolutionary ancestral SINE. The low copy numbers detected in Amaranthaceae support the old age of this family. The Au element of spinach differs from the other Amaranthaceae Au, harbouring a deletion in the 3′ region. This deletion could be the reason for the marginally higher copy number in spinach. It is possible that the transcription machinery of LINEs is able to recognize this Au sequence more efficiently, resulting in a higher amplification of Au in spinach. According to the gene annotation in *Beta*, the element was found in an intron of a ubiquitin-protein ligase gene, resulting in high conservation in a safe haven. It is unlikely that this copy will be deleted because removal could affect the function of the gene. However, it is also possible that the presence of Au has no effect and is non-functional and evolutionarily neutral.

Plant SINEs prefer gene-rich regions and are rarely present in heterochromatic pericentromeric regions (Deragon and Zhang, 2006; Baucom *et al.*, 2009). All AmaS SINEs are dispersed distributed with a preference for distal chromosome regions. This is in accordance with SINEs in other species (Deragon and Zhang, 2006; Wenke *et al.*, 2011). Because of their small size, SINEs are less deleterious than larger TEs and therefore probably more tolerated, also resulting in a mostly genome-wide dispersed integration. A general feature of SINEs is their tendency for location in euchromatic chromosome regions, often closely related to genes (Lenoir *et al.*, 2001; Deragon and Zhang, 2006; Baucom *et al.*, 2009). The tolerance of SINEs in euchromatin can be due to several reasons, for example the LINE endonuclease can nick euchromatic DNA better than tightly packed heterochromatic DNA and therefore, euchromatin is more accessible for integration than heterochromatin (Cost *et al.*, 2001). AmaS SINEs have a strong target site preference, as shown for other species (Zhang and Wessler, 2005; Khan *et al.*, 2011; Wenke *et al.*, 2011). All AmaS SINEs have an A/T-rich upstream flanking region and integrate upstream of an adenine or stretch of adenines, indicating the need for weak hydrogen bonds for the insertion of SINEs. Another typical feature of SINE integration seems to be an arbitrary nucleotide directly upstream of the 5′ TSD except adenine, as also found by Zhang and Wessler (2005) and Wenke *et al.* (2011) for SINEs with poly (A) tails.

The transposition of SINEs depends on the reverse transcriptional machinery of LINEs. The 3′ end of the SINE has to be recognized by the reverse transcriptase of an active LINE. A partnership between SINEs and LINEs has been observed in several species (Ohshima and Okada, 2005; Baucom *et al.*, 2009); however, among the 1600 LINEs known in sugar beet (Wenke *et al.*, 2009; Heitkam *et al.*, 2014) we did not find a LINE family corresponding to one of the SINE families detected in this study. Therefore, we concluded that the poly(A), and hence most likely a relaxed SINE/LINE correlation, is sufficient for the binding of the LINE reverse transcriptase as shown for other species (Lenoir *et al.*, 2001; Myouga *et al.*, 2001).

Due to their ability to cause rearrangements, TEs are silenced by the host, resulting in inactive copies. Environmental stresses, including wounding and pathogen attack, can reactivate TEs resulting in transposition (Grandbastien, 1998). In rice, an activation of retrotransposons after introgression of DNA from a wild species was observed, resulting in an increase in copy number associated with *de novo* methylation of these retrotransposons and possibly also SINEs (Liu and Wendel, 2000). The epigenetic silencing of repetitive elements occurs by cytosine methylation over their entire length (Lisch, 2009; Teixeira and Colot, 2010). This methylation not only modifies CG sites, but also CHG and CHH sites. In the Arabidopsis genome, 24% of CG, 6.7% of CHG and 1.7% of CHH are methylated to maintain genome stability (Cokus *et al.*, 2008). Sugar beet SINEs are mostly localized in euchromatic regions, which are characterized by a high density of genes (Dohm *et al.*, 2014). AmaS SINEs are more methylated on both DNA strands than the flanking regions. Higher levels of symmetric CG and CHG methylation of SINEs compared with flanking regions indicates that SINEs are also transcriptionally silenced by epigenetic modifications such as histone modifications. CHG methylation induces dimethylation of histone H3 at lysine 9 (H3K9me2) and a similar mechanism is proposed for CG methylation (Johnson *et al.*, 2007; Law and Jacobsen, 2010). However, the strongest effect is detectable at asymmetric CHH sites, and most cytosines occur in this motif. SINEs may be involved in the regulation of genes in the close vicinity of their integration site via CHH methylation as has been demonstrated for repetitive DNA in Arabidopsis and maize (Henderson and Jacobsen, 2008; Gent *et al.*, 2013). A similar mechanism is postulated for small satellite arrays being dispersed along sugar beet euchromatin (Zakrzewski *et al.*, 2014).

SINEs have a strong effect on the organization and diversity of genomes. We give a comprehensive insight into the evolution and age of Amaranthaceae SINEs and provide detailed information on the structure, integration and methylation of sugar beet SINEs. The evolutionary dynamics of SINEs, including emergence, diversification or decay, are important factors affecting plant genome organization and evolution.

EXPERIMENTAL PROCEDURES

DNA preparation and hybridization experiments

Plant material. For hybridization experiments, *B. vulgaris* ssp. *vulgaris* 'KWS 2320', *B. patula* (BETA 548), *B. lomatogona* (BETA 674), *B. nana* (BETA 543), *P. procumbens* (BETA 2316), *C. quinoa* (CHEN 125) and *S. oleracea* ('Matador') were used. Plants were grown under greenhouse conditions. Sugar beet was obtained from KWS SAAT SE (http://www.kws.com/). The wild beet species were obtained from the GenBank of the Leibniz Institute of Plant Genetics and Crop Plant Research Gatersleben (http://www.ipk-gatersleben.de).

DNA gel blot hybridization. For DNA gel blot hybridization, isolation of genomic DNA was performed using the cetyltrimethyl ammonium bromide (CTAB) protocol (Saghai-Maroof *et al.*, 1984). Six micrograms of genomic DNA was digested with *Hae*III (Thermo Scientific, http://www.thermoscientific.de.) and separated on 1.2% agarose gels. For alkaline transfer of the DNA we used positively charged nylon membranes (GE Healthcare, http://www.gehealthcare.com/). The hybridization was performed according to standard protocols using probes labelled with ^{32}P by random priming (Sambrook *et al.*, 1989). SINE family-specific probes were isolated from sugar beet DNA using PCR and the primers listed in Table S1. Filters were hybridized at 60°C and washed in 2 × SSC/01% SDS for 40 min at 60°C to reach a stringency of at least 75%. Signals were detected by autoradiography.

Fluorescent in situ hybridization. For the preparation of mitotic chromosomes the meristems of young leaves were used. Leaves were incubated in 2 mM 8-hydroxychinoline and fixed in methanol:acetic acid (3:1). Maceration of plant material was performed in an enzyme mixture consisting of 0.3% (w/v) cytohelicase (Sigma, http://www.sigmaaldrich.com), 1.8% (w/v) cellulase from *Aspergillus niger* (Sigma), 0.2% (w/v) cellulase Onozuka-R10 (Serva, http://www.serva.de) and 20% (v/v) pectinase from *A. niger*, followed by spreading of the nuclei on acid-cleaned slides. Species-specific probes were labelled by PCR (for primers see Table S1) in the presence of biotin-11-dUTP. Then FISH was performed according to Schmidt *et al.* (1994). Slides were examined with a Zeiss Axioplan2 Imaging fluorescent microscope. Images were obtained with Applied Spectral Imaging v.3.3 software coupled with the high-resolution CCD camera ASI BV300-20A. The images were optimized by Adobe Photoshop software (http://www.adobe.com) using only functions affecting the whole image equally.

Bioinformatic analysis

Data and resources. Reference sequences of *B. vulgaris* ssp. *vulgaris* (genotype 'KWS2320', 596 Mb), and *S. oleracea* (661 Mb) (Dohm *et al.*, 2014), available at http://bvseq.molgen.mpg.de, *B. patula* (607 Mb), *B. lomatogona* (775 Mb), *B. nana* (513 Mb), *P. procumbens* (695 Mb) and *C. quinoa* (1.4 Gb) were used for bioinformatics analysis. The assembled genomes of *B. patula*, *B. lomatogona*, *B. nana*, *P. procumbens* and *C. quinoa* will be published elsewhere. All reference sequences were assembled from Illumina paired ends and mate pairs, with a targeted genome coverage between 50 × and 100 ×. Information about sequenced plants is listed in Table S2.

SINE-Finder. To identify SINE families in genome sequences we used the SINE-Finder program (Wenke *et al.*, 2011). The search motif was as follows: a 5′ TSD region of 40 nt, followed by RVTGG as box A motif, 25–50 nt as a spacer, the box B motif GTTCRA, a spacer of 20–500 nt, six adenines or thymines as a poly(A/T) stretch or single sequence repeats, and a 3′ TSD region of 40 nt.

Characterization of SINEs. For alignments and BLAST searches we used Geneious software v.6.1.8 (http://www.geneious.com, Kearse *et al.*, 2012) and stand-alone versions of muscle (Edgar, 2004) and fasta (ftp://ftp.ebi.ac.uk/pub/software/unix/fasta/fasta36/). Majority consensus sequences and pairwise identities were constructed using Geneious. Dendrograms were created using mega 6 software (Tamura *et al.*, 2007), applying the neighbour-joining distance method and the maximum composite likelihood nucleotide model with 1000 replicates.

The age distributions of the detected SINE copies for the individual families and species were visualized based on the sequence similarities to the consensus sequence. Histograms were created using a Python-script including the Matpotlib module (Hunter, 2007).

For chromosomal mapping of SINEs, the SINE copies (including flanking regions) were used as query for a BLAT search against all anchored scaffolds of the most recent *B. vulgaris* assembly (v.1.2). Due to the use of different assembly versions and highly similar SINE queries, duplicate hits were retrieved by BLAT. Moreover, the inclusion of flanking sequences in the queries led to overlapping hits resulting from closely neighbouring SINEs. Thus, duplicates were excluded according to the following criteria: sequences were removed if they (i) had a chromosomal and positional overlap and either the start or stop positions of two hits differed by less than 30 nt, (ii) the overlapping part was longer than the non-overlapping part, or (iii) the distance between stop positions was smaller than 100 nt. Chromosomal coordinates were obtained from the GenDBE project (https://gendbe.cebitec.uni-bielefeld.de) including a 50-nt spacer between the scaffolds.

Data filtering and processing for visualization in MapChart 2.2 (Voorrips, 2002) were realized with custom Python routines.

Bisulphite sequencing of SINEs and data analysis. Genomic DNA from sugar beet leaves was bisulphite converted using the EpiTect Bisulfite Kit from Qiagen (cat. no. 59104; http://www.qiagen.com/). We bisulphite sequenced 81 650 740 raw read pairs with a length of 101 nt on an Illumina HiSeq2000 platform, resulting in approximately 11-fold coverage of the genome. The reads were quality trimmed using TrimGalore with default parameters (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Following quality trimming, 80 751 920 reads were mapped against all 6325 AmaS SINEs detected in the reference sequence of sugar beet genotype 'KWS2320' (Dohm *et al.*, 2014). In addition, 100 nt up- and downstream flanking the integration sites of each SINE were included. Mapping was per-

formed using BISMARK v.0.12.2 (Krueger and Andrews, 2011) with Bowtie 1 and the following parameters: -q -n 1 -l 50 -k 2 —best —maxins 500 —chunkmbs 512 Option —directional. Quantification trend blots were generated using SEQMONK software (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/). Only cytosine positions where a cytosine has been sequenced at least three times were included.

ACKNOWLEDGEMENTS

We thank Nadin Fliegner for technical assistance. Genomic sequencing other than bisulphite sequencing was performed at the Genomics Unit of the CRG in Barcelona. This work was supported by the BMBF grant 'AnnoBeet: Annotation des Genoms der Zuckerrübe unter Berücksichtigung von Genfunktionen und struktureller Variabilität für Nutzung von Genomdaten in der Pflanzenbiotechnologie' FKZ 0315962 A, 0315962 B and 0315962 C (to BW, HH and TS).

SUPPORTING INFORMATION

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

Figure S1. 5′ insertion site preference of sugar beet short interspersed nuclear elements.

Figure S2. Methylation of all sugar beet short interspersed nuclear element families and 100 nucleotides up- and downstream of their flanking region.

Figure S3. Genomic organization of short interspersed nuclear element families in Amaranthaceae.

Figure S4. Evolutional relationship of AmaS short interspersed nuclear element families.

Figure S5. Comparison of copy number and similarity to consensus sequence of all species of Amaranthaceae.

Figure S6. Alignment of Au in the genus *Beta*.

Figure S7. Alignment of 10 representative copies of AmaS-II subfamilies in sugar beet.

Table S1. Primers used for the generation of AmaS SINE probes for Southern blot and fluorescent *in situ* hybridization.

Table S2. Amaranthaceae sequence data.

Data S1. Fasta file containing all short interspersed nuclear elements found in the family Amaranthaceae.

Data S2. Fasta file containing all full-length sugar beet short interspersed nuclear elements.

Data S3. Fasta file containing sequences used for the dendrogram.

Data S4. Fasta file containing consensus sequences of all short interspersed nuclear element families found in the family Amaranthaceae.

REFERENCES

Baucom, R.S., Estill, J.C., Chaparro, C., Upshaw, N., Jogi, A., Deragon, J.-M., Westerman, R.P., Sanmiguel, P.J. and Bennetzen, J.L. (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. *PLoS Genet.* **5**, e1000732.

Ben-David, S., Yaakov, B. and Kashkush, K. (2013) Genome-wide analysis of short interspersed nuclear elements SINES revealed high sequence conservation, gene association and retrotranspositional activity in wheat. *Plant J.* **76**, 201–210.

Bennetzen, J.L. and Wang, H. (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* **65**, 1–26.

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

Cost, G.J., Golding, A, Schlissel, M.S. and Boeke, J.D. (2001) Target DNA chromatinization modulates nicking by L1 endonuclease. *Nucleic Acids Res.* **29**, 573–577.

Deragon, J.-M. and Zhang, X. (2006) Short interspersed elements (SINEs) in plants: origin, classification, and use as phylogenetic markers. *Syst. Biol.* **55**, 949–956.

Dohm, J.C., Minoche, A.E., Holtgräwe, D. *et al.* (2014) The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature*, **505**, 546–549.

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

Fawcett, J.A., Kawahara, T., Watanabe, H. and Yasui, Y. (2006) A SINE family widely distributed in the plant kingdom and its evolutionary history. *Plant Mol. Biol.* **61**, 505–514.

Finnegan, D.J. (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet.* **5**, 103–107.

Gadzalski, M. and Sakowicz, T. (2011) Novel SINEs families in *Medicago truncatula* and *Lotus japonicus*: bioinformatic analysis. *Gene*, **480**, 21–27.

Gent, J.I., Ellis, N.A., Guo, L., Harkess, A.E., Yao, Y., Zhang, X. and Dawe, R.K. (2013) CHH islands: *de novo* DNA methylation in near-gene chromatin regulation in maize. *Genome Res.* **23**, 628–637.

Gentles, A.J., Kohany, O. and Jurka, J. (2005) Evolutionary diversity and potential recombinogenic role of integration targets of non-LTR retrotransposons. *Mol. Biol. Evol.* **22**, 1983–1991.

Grandbastien, M.A. (1998) Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.* **3**, 181–187.

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

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

Hohmann, S., Kadereit, J.W. and Kadereit, G. (2006) Understanding Mediterranean-Californian disjunctions: molecular evidence from Chenopodiaceae-Betoideae. *Taxon*, **55**, 67–78.

Hollister, J.D. and Gaut, B.S. (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* **19**, 1419–1428.

Huang, C., Burns, K. and Boeke, J. (2012) Active transposition in genomes. *Annu. Rev. Genet.* **46**, 651–675.

Hunter, J.D. (2007) Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95.

Johnson, L.M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J. and Jacobsen, S.E. (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr. Biol.* **17**, 379–384.

Jühling, F., Mörl, M., Hartmann, R.K., Sprinzl, M., Stadler, P.F. and Pütz, J. (2009) tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.* **37**, D159–D162.

Kearse, M., Moir, R., Wilson, A. *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Khan, M.F., Yadav, B.S., Ahmad, K. and Kumar, A. (2011) Mapping and analysis of the LINE and SINE type of repetitive elements in rice. *Bioinformation*, **7**, 276–279.

Kolano, B., Bednara, E. and Weiss-Schneeweiss, H. (2013) Isolation and characterization of reverse transcriptase fragments of LTR retrotransposons from the genome of Chenopodium quinoa (Amaranthaceae). *Plant Cell Rep.* **32**, 1575–1588.

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

Lander, E.S., Linton, L.M., Birren, B. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.

Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220.

Lenoir, A., Lavie, L., Prieto, J.L., Goubely, C., Coté, J.C., Pélissier, T. and Deragon, J.M. (2001) The evolutionary origin and genomic organization of SINEs in Arabidopsis thaliana. *Mol. Biol. Evol.* **18**, 2315–2322.

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

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

Liu, B. and Wendel, J.F. (2000) Retrotransposon activation followed by rapid repression in introgressed rice plants. *Genome*, **43**, 874–880.

Myouga, F., Tsuchimoto, S., Noma, K., Ohtsubo, H. and Ohtsubo, E. (2001) Identification and structural analysis of SINE elements in the *Arabidopsis thaliana* genome. *Genes Genet. Syst.* **76**, 169–179.

Ohshima, K. and Okada, N. (2005) SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. *Cytogenet. Genome Res.* **110**, 475–490.

Oliver, K.R., McComb, J.A. and Greene, W.K. (2013) Transposable elements: powerful contributors to angiosperm evolution and diversity. *Genome Biol. Evol.* **5**, 1886–1901.

Rebollo, R., Horard, B., Hubert, B. and Vieira, C. (2010) Jumping genes and epigenetics: towards new species. *Gene*, **454**, 1–7.

Roy-Engel, A.M., Salem, A.H., Oyeniran, O.O., Deininger, L., Hedges, D.J., Kilroy, G.E., Batzer, M.A. and Deininger, P.L. (2002) Active Alu element "A-tails": size does matter. *Genome Res.* **12**, 1333–1344.

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A. and Allard, R.W. (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl Acad. Sci. USA*, **81**, 8014–8018.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. 3 ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Schmidt, T., Schwarzacher, T. and Heslop-Harrison, J.S. (1994) 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.

Slotkin, R.K. and Martienssen, R. (2007) Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8**, 272–285.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599.

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

Voorrips, R.E. (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **93**, 77–78.

Wenke, T., Holtgräwe, D., Horn, A.V., Weisshaar, B. and Schmidt, T. (2009) An abundant and heavily truncated non-LTR retrotransposon (LINE) family in *Beta vulgaris*. *Plant Mol. Biol.* **71**, 585–597.

Wenke, T., Döbel, T., Sörensen, T.R., Junghans, H., Weisshaar, B. and Schmidt, T. (2011) Targeted identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in plant genomes. *Plant Cell*, **23**, 3117–3128.

Wicker, T., Sabot, F., Hua-Van, A. *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982.

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, X. and Wessler, S.R. (2005) BoS: a large and diverse family of short interspersed elements (SINEs) in *Brassica oleracea*. *J. Mol. Evol.* **60**, 677–687.

Zingler, N., Willhoeft, U., Brose, H., Schoder, V., Jahns, T., Hanschmann, K.O., Morrish, T.A. and Schumann, G.G. (2005) Analysis of 5J junctions of human LINE-1 and *Alu* retrotransposons suggests an alternative model for 5′-end attachment requiring microhomology-mediated end-joining. *Genome Res.* **15**, 780–789.