DOI: 10.1007/s10535-015-0536-8 BIOLOGIA PLANTARUM 49 (4): 661-670, 2015

Unravelling genome dynamics in *Arabidopsis* synthetic auto and allopolyploid species

M. BENTO, D. TOMÁS, W. VIEGAS, and M. SILVA*

Linking Landscape, Environment, Agriculture and Food (LEAF), Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda 1349-017 Lisboa, Portugal

Abstract

Polyploidization is a major genome modification that results in plant species with multiple chromosome sets. Parental genome adjustment to co-habit a new nuclear environment results in additional innovation outcomes. We intended to assess genomic changes in polyploid model species with small genomes using inter retrotransposons amplified polymorphism (IRAP) and retrotransposon microsatellite amplified polymorphism (REMAP). Comparative analysis among diploid and autotetraploid *A. thaliana* and *A. suecica* lines with their parental lines revealed a marginal fraction of novel bands in both polyploids, and a vast loss of parental bands in allopolyploids. Sequence analysis of some remodelled bands shows that *A. suecica* parental band losses resulted mainly from sequence changes restricted to primer domains. Moreover, in *A. suecica*, both parental genomes presented rearrangement frequencies proportional to their sizes. Overall rates of genomic remodelling events detected in *A. suecica* were similar to those observed in species with a large genome supporting the role of retrotransposons and microsatellite sequences in the evolution of most allopolyploids.

Additional key words: microsatellites, polyploidization, retrotransposons, sequence rearrangement.

Introduction

Recent plant genomic studies have evidenced the crucial relevance of labile repetitive sequences like retrotransposons in plant evolution through polyploidization. Due to the huge amount of genetic data available from species with small genomes like *Arabidopsis* sp., the study of repetitive genome fraction involvement in polyploid induced remodelling events may be particularly revealing in this genus. In fact polyploidy, which result from the fusion of two or more complete genomes into the same nucleus, is present in the evolution route of most flowering plants with both large and small genome (Jiao *et al.* 2011). Distinct types of polyploids can arise depending on the origins of genomes involved: autopolyploids originated from genome doubling within the same species and allopolyploids formed by the combination of two or more genomes from distinct but usually close species (Ma and Gustafson 2008).

The impact of polyploidy on genome organization and function (reviewed in Ainouche 2010 and Stöck and Lamatsch 2013) was extensively pursued using different approaches, and it is presently well established that polyploidy induces innumerable genetic and epigenetic modifications on parental genomes (Feldman and Levy 2009, Jones and Hegarty 2009). Genomic studies using amplified fragment length polymorphism (AFLP), although having the advantage of full genome coverage, usually reveal lower levels of rearrangements associated with polyploidization when compared with techniques targeting major types of repetitive sequences widely present in plant genomes like retrotransposons and microsatellites (Ma *et al.* 2004, Madlung *et al.* 2005, Bento *et al.* 2008, 2013). From these techniques, inter retrotransposons amplified polymorphism (IRAP) and retrotransposon microsatellite amplified polymorphism (REMAP) have recently demonstrated the marked importance of retrotransposons as a particularly dynamic

Submitted 24 November 2014, *last revision* 31 March 2015, *accepted* 1 April 2015.
Abbreviations: AFLP - amplified fragment length polymorphism; IRAP - inter retrotransposons amplified polymorphism; LTR - long terminal repeats; REMAP - retrotransposon microsatellite amplified polymorphism; SSR - simple sequence repeats.
Acknowledgements: M. Bento was funded by a FCT (Fundação para a Ciência e a Tecnologia, Portugal) postdoctoral grant (SFRH/BPD/80550/2011), Diana Tomás was funded by a FCT doctoral scholarship (SFRH/BD/93156/2013), Manuela Silva by the FCT Investigator Programme (IF/00834/2014), and the research work was financed by FCT LEAF Unit (UID/AGR/04129/2013).
* Corresponding author, fax: (+351) 213653238, e-mail address: manuelasilva@isa.ulisboa.pt

fraction in processes underlying polyploid genome adjustments (Bento *et al.* 2008, 2013). Furthermore, the evaluation of microsatellite dynamics in *Brassica*, *Arabidopsis*, and other angiosperms indicate that polyploidization may induce a slight increase of microsatellite frequency in coding regions although associated with a significant decrease in whole genome/non-coding sequences (Shi *et al.* 2013). These evidences reinforce the current well-established notion that genome size changes in allopolyploids are correlated with changes on repetitive genome fractions (Renny-Byfield *et al.* 2013).

Most studies targeting genomic rearrangements in plant allopolyploids were pursued in species with medium and large genomes, such as octoploid triticale (1C=25404 Mb; monoploid genome size 1Cx=6351 Mb), hexaploid *Triticum aestivum* (1C=16944 Mb; 1Cx=5648 Mb), tetraploid *Nicotiana tabacum* (1C=5061 Mb; 1Cx=2530 Mb), dodecaploid *Spartina anglica* (1C=5330 Mb; 1Cx=888 Mb), and tetraploid *Cucumis hytivus* (diploid parental line *C. sativus* 1C=880 Mb) (Bento *et al.* 2008, Feldman and Levy 2009, Jiang *et al.* 2009, Parisod *et al.* 2009, Petit *et al.* 2010). Deoxyribonucleic acid C-values were obtained from the Plant DNA C-values database (Bennett and Leitch 2010; http://www.kew.org/cvalues/) with an estimated repeat content of > 80 %, ~80 %, ~70 %, ~30 %, and ~24 %, respectively (Dvořák 2009, Huang *et al.* 2009, Brenchley *et al.* 2012, Ferreira-de-Carvalho *et al.* 2013, Sierro *et al.* 2013). In species with small genomes, such as *Arabidopsis* with a lower repetitive sequences fraction (~14 %) (AGI 2000), the assessment of alterations induced by polyploidy in such a genome portion was meagerly studied. Due to vast genetic and genomic resources available (http://www.arabidopsis.org/), *Arabidopsis* polyploid species constitute a particularly attractive model to such studies (reviewed in Bomblies and Madlung 2014). Previous studies of newly formed autotetraploid lines of *A. thaliana* revealed minor genomic changes induced by polyploidy (Santos *et al.* 2003, Madlung *et al.* 2005, Ozkan *et al.* 2006), though epigenetic modifications were reported (Yu *et al*. 2010). *A. suecica* (2n=4x=26), the natural allopolyploid between *A. thaliana* and *A. arenosa*, can be synthetically produced by crossing *A. thaliana* (2n=4x=20) and *A. arenosa* (2n=4x=32) autotetraploid lines (Madlung *et al.* 2005). The analysis of *Arabidopsis* allopolyploids revealed both genetic (Madlung *et al.* 2005, Pecinka *et al.* 2011, Henry *et al.* 2014) and epigenetic modifications (Hazzouri *et al.* 2008, Beaulieu *et al.* 2009, Yu *et al.* 2010). Our previous results obtained using IRAP methodology also demonstrated the occurrence of genomic rearrangements induced by allopolyploidization in *A. suecica* (Bento *et al.* 2013). In this work, we intended to understand the implication of retrotransposons and also microsatellite related sequences in genetic modifications induced by auto and allopolyploidization in the *Arabidopsis* species.

Materials and methods

The following plants were used: diploid *Arabidopsis thaliana* L. (cv. Landsberg erecta, Ler-0; 2n=2x=10, NASC stock number NW20) and the equivalent autotetraploid line (LC612; 2n=4x=20, *NASC* stock No. N3900); *A. arenosa* L. (CARE-1; 2n=4x=32, *NASC* stock No. N3901) and the newly synthesized allopolyploid *A. suecica* Fr. line (Ler-0 x CARE-1; 2n=4x=26; F4 generation, *NASC* stock No. N3899). The *A. thaliana* autotetraploid and *A. arenosa* lines used are the exact parents of the referred newly synthesized allopolyploid. All seeds were obtained from the *NASC*-European *Arabidopsis* Stock Centre and germinated and grown in controlled conditions at a 16-h photoperiod, an irradiance of 150 µmol m^{-2} s^{-1}, day/night temperatures of 22/20 ºC, and a relative humidity of 70 %. At least four distinct plants from each genotype were analyzed. Genomic DNA was isolated from fresh young leaves of eight-week-old plants using a modified cetyltrimethylammonium bromide (CTAB) method (Saghai-Maroof *et al.* 1984).

The used PCR-based methodologies included IRAP and REMAP. The IRAP amplify sequences between retrotransposons using primers pointing outward from the long terminal repeats (LTR) end. The REMAP uses the same LTR primers combined with primers matching simple sequence repeats (SSR) or microsatellite motifs with one or more non-SSR anchor nucleotides thus amplifying segments between LTRs, between LTR and SSR, or between nearby microsatellites (Kalendar and Schulman 2006). The IRAP was prepared using primers for LTR regions of distinct *A. thaliana* retrotransposons including *Metaviridae* (*Tat*1, *Athila*4-6, and *Romania*T5) as well as *Pseudoviridae* (*Endovir*1-1). The REMAP was done using an *Endovir*1-1 LTR primer and an anchored microsatellite primer (AAG)6C. All primers used are listed in Table 1. The PCR products were run on 1.7 % (m/v) agarose gels, detected with ethidium bromide staining, and photographed using *BioRad* (Hercules, USA) *GEL DOC 2000* (Tables 2 and 3). Reproducible results presented were accurately observed in at least three replicates of each IRAP and REMAP experiment. Results obtained in *A. suecica* with the *Tat*1 and *Athila* 4 - 6 primers were previously presented in Bento *et al.* (2013). Differences observed between diploid and tetraploid *A. thaliana* profiles unraveled autopoly-ploidization induced genomic changes. Constancy observed in IRAP/REMAP banding profiles on all

A. thaliana and *A. arenosa* plants analyzed allowed an accurate comparison of the neosynthetized allopolyploid with both parental lines. In *A. suecica*, genomic rearrangements were disclosed by differences between its banding profile and the addition of the parental lines banding profiles. Moreover, to confirm that polyploid lost bands did not result from competition between priming sites of parental genomes, a mix of the parental DNAs was used (results not shown) as previously described in detail (Bento *et al.* 2008).

Rearranged sequences were isolated, purified, cloned, and finally sequenced following procedures described in Rocheta *et al.* (2007). The sequences obtained were used for *BLASTN* (algorithm parameters set for default values) on the *NCBI* (the National Center for Biotechnology Information), *TAIR* (the *Arabidopsis* Information Resource), and *PlantSat* (repetitive plant sequences) databases. Additionally, primers were designed to amplify internal fragments of rearranged sequences (Table 1). The PCR amplifications and analysis of internal sequences were performed as described in Bento *et al.* (2010).

Table 1. Primers used for PCR analyses.

Retroelement	GeneBank acc. No.	Primer	Superfamily
*Tat*1	AF056631	5'-TCCTCGTAAAACAACCACAAG	*Gypsy*-like
*Athila*4-6	AF296831	5'-AAATGGATGCTCAAAACATGC	*Gypsy*-like
*Romania*T5	AF073829	5'-GGTTTCTGAGGATGGGTTAG	*Gypsy*-like
*Endovir*1-1	AY016208	5'-CCAAATGAACGAATGACGTG	*Copia*-like
Microsatellite $(AAG)_6C$		5'-AAGAAGAAGAAGAAGAAGC	

Internal segments	Forward	Reverse
Romania_aa951[1]	5'-TGCCTGCTAGGTTATGTTTC	5'-CATGCATAAAGGGAAATAGC
Romania_aa951[2]	5'-ATTCTCTGTACTTGCTACTTGC	5'-GCGCCAAATTGATACTAGG
Romania_aa1357[2]	5'-GTGTGGTCGAGTATCAGGTC	5'-GATTCGTAATGCCTGTTAGG
Romania_aa1357[1]	5'-TCAAGATGCAATCAGATGTC	5'-TAACAAACTGTGCAGTCGAG
End-AAG_as552[2]	5'-ATTCTAGCCATTGTTTCGTC	5'-GTGGGTGTTGTGATTACCC
End-AAG_as552[1]	5'-GATTGGTGGCAGGGATAC	5'-CTTGTTTGTCCTCCTTTTTG

Results

To assess genomic rearrangements induced by polyploidization on *Arabidopsis* autopolyploids and allopolyploids, IRAP and REMAP experiments with five primer combinations were performed including four primers designed for LTR regions of *Arabidopsis* retrotransposons (*Tat*1, *Athila*4-6, *Romania*T5, and *Endovir*1-1) and one anchored microsatellite primer [(AAG)6C] (Table 1). Unique and reproducible banding profiles from diploid and tetraploid *A. thaliana*, *A. arenosa*, and *A. suecica* were consistently obtained in all PCR replicates from at least four distinct plants of each species and primer combination. Thus, a parental comparative evaluation of auto and allopolyploid lines was performed to assess the impact of polyploidization in retrotransposon and microsatellite flanking sequences (Fig. 1).

To unravel potential genomic rearrangements associated with autopolyploidization banding profiles of original diploid *A. thaliana* and those of the autotetraploid line were compared (Table 2). Additionally, a new band of ~1 700 bp, absent in the diploid parental line, was unravelled in the *Romania*T5 IRAP profile of the autotetraploid (Fig. 1). Thus, a reduced overall frequency of 1.7 % novel bands (1/58) was

Table 2. The IRAP and REMAP analyses of autotetraploid *A. thaliana*.

Ploidy	Bands	IRAP				REMAP	Total
		*Tat*1	*Athila*4-6	*Romania*T5	*Endovir*1-1	*Endovir*1-1/ (AAG)6C	
Diploid	observed bands	8	11	7	17	15	58
Tetraploid	conserved bands	8	11	7	17	15	58
	novel bands	0	0	1	0	0	1

Table 3. The IRAP and REMAP analyses of synthetic allopolyploid *A. suecica* and parental lines (* - results previously published in Bento *et al.* (2013); [i] - monomorphic bands: common to both *A. suecica* progenitors; [ii] - polymorphic bands: observed in only one *A. suecica* progenitor; [iii] - *A. suecica* expected bands: the sum of parental bands observed; [iv] - *A. suecica* observed bands: bands observed in the allopolyploid; [v] - *A. suecica* rearranged bands: parental bands absent and novel bands observed only in the polyploidy).

	IRAP				REMAP	Total
	*Tat*1*	*Athila*4-6*	*Romania*T5	*Endovir*1-1	*Endovir*1-1/(AAG)6C	
A. thaliana tetraploid	8	11	8	17	15	59
A. arenosa	12	8	10	12	16	58
A. thaliana vs. *A. arenosa*						
monomorphic bands [i]	2	0	2	8	7	19
polymorphic bands [ii]	16	19	14	13	17	79
A. suecica						
expected [iii]	18	19	16	21	24	98
observed bands [iv]	10	16	7	13	19	65
Rearranged bands [v]	10	3	9	8	9	39
lost from *A. thaliana*	3	1	2	6	0	12
lost from *A. arenosa*	6	2	7	2	7	24
novel bands	1	0	0	0	2	3

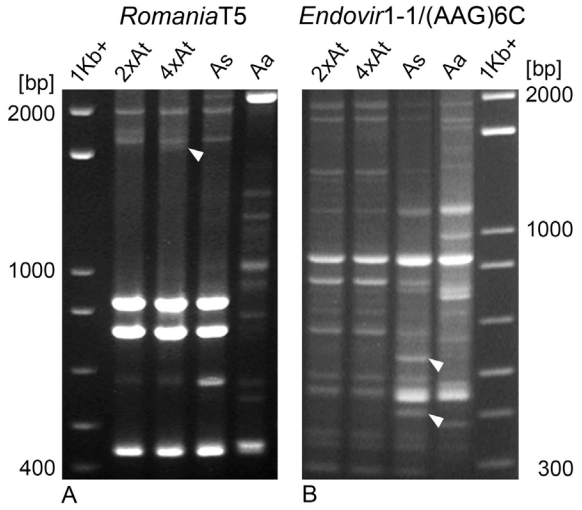

Fig. 1. The IRAP banding profiles obtained with primers for *Romania*T5 and REMAP banding profiles obtained with primers for *Endovir*1-1/(AAG)6C of *A. thaliana* Ler-0 2n=2x=10 (2xAt), *A. thaliana* LC612 2n=4x=20 (4×At), *A. suecica* synthetic allopolyploid 2n=4x=26 (As), and *A. arenosa* CARE-1 2n=4x=32 (Aa). *Arrows* indicate novel bands observed in the autopolyploid (*A*) and in the allopolyploid (*B*) lines. Molecular mass marker is marked as 1 Kb+.

detected in the autopolyploid line. The same comparative analysis of allopolyploid *A. suecica* in relation to the tetraploid *A. thaliana* and *A. arenosa* lines (Table 3) revealed one novel band with *Tat*1 IRAP and two with *Endovir*1-1/(AAG)6C REMAP corresponding to a 4.6 % (3/65) frequency.

Two novel bands uncovered in *A. suecica* with REMAP *Endovir*1-1/(AAG)6C were isolated, purified, and cloned for sequencing, and both sequences revealed to be allocated between microsatellite loci. Moreover, their *BLASTN* analysis on the *NCBI*, *TAIR*, and *PlantSat* databases (Fig. 2 and Table 1 Suppl.) unravelled homology with *A. thaliana* coding sequences mapped on chromosomes 3 and 5, respectively. End-AAG_as415 (415 bp, acc. No. KJ371929) presented a 78 % similarity and a 96 % coverage with an *A. thaliana* coding region corresponding to a gyp1p superfamily protein. Again, End-AAG_as552 sequence (552 bp, acc. No. KJ371930) showed a 77 % homology with the coding region of a RING-H2 finger protein isolated from the same species but only with a 36 % coverage since a 1 - 355 bp region did not produce any significant alignment with the *A. thaliana* published sequence. However, a 1 - 181 bp domain revealed a 92 % homology with an unknown sequence of one *Arabidopsis halleri BAC* library (acc. No. FO203486.1).

To better understand the origin of the *A. suecica* novel sequence, we designed primers to amplify partial internal

fragments of the End-AAG_as552 sequence (primer details in Table 1 and a schematic representation in Fig. 2). The obtained results (Fig. 3A) confirmed the amplification of three targeted internal fragments of *A. suecica* with the expected sizes (Fig. 2 and Table 2 Suppl.): End-AAG_as552_[1] (247 bp); End-AAG_as552_[2] (268 bp); and End-AAG_as552_[1-2] (408 bp) which resulted from the amplification of the forward primer [1] with the reverse primer [2]. In both the parental species, no amplification products were obtained with internal primers confirming the novelty of the End-AAG_as552 sequence identified in *A. suecica*.

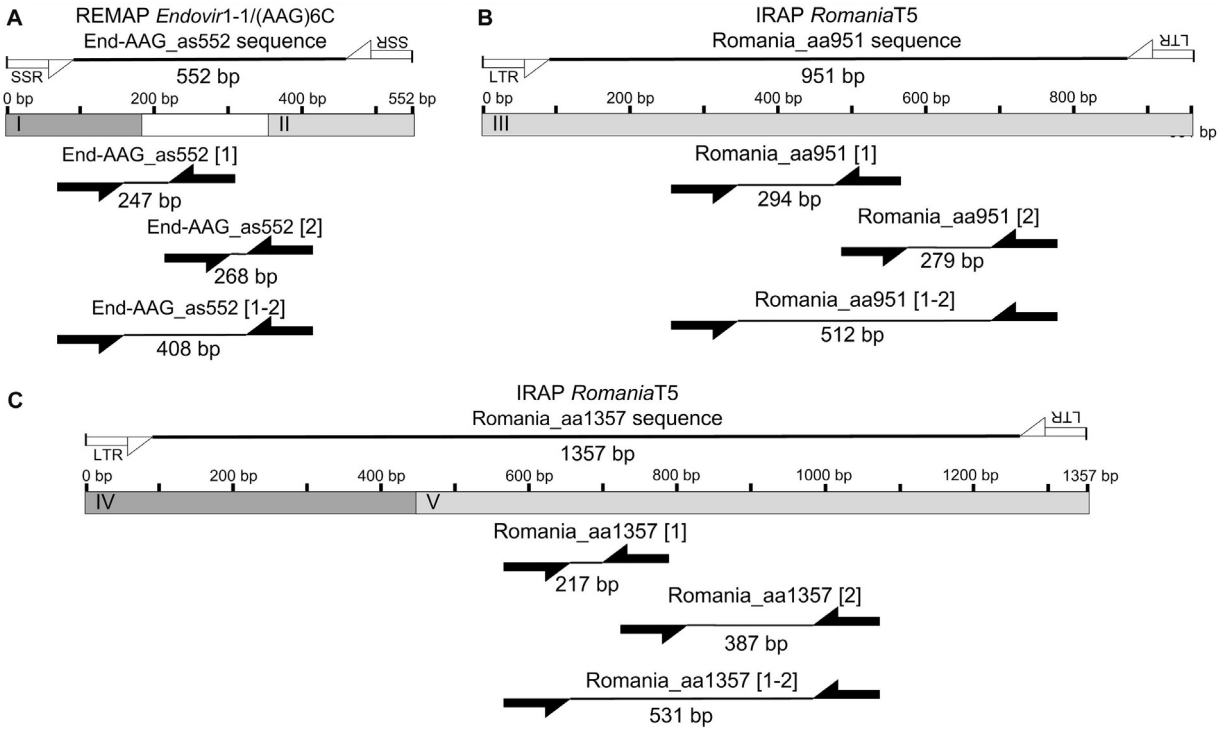

Fig. 2. *A. suecica* rearranged sequence characterization (alignments detailed in Table 1 Suppl.) and primers designed to amplify internal segments (primer details in Tables 1 and 2 Suppl.). Primers used in IRAP and REMAP initial experiments are represented in *white* and regions revealing similarity with published sequences (the *NCBI*, *TAIR*, and *PlantSat* databases) are indicated (I, II, III, IV, and V). Primers designed to amplify internal segments are represented in *black* and the dimensions of expected internal fragments are also illustrated. *A* - an REMAP End-AAG_as552 *A. suecica* novel band: I - a 1 to 181 bp region similar to *A. halleri* BAC; II - a 356 to 552 bp region similar to *A. thaliana* RING-H2 finger protein ATL74 coding region; expected internal segments: End-AAG_as552 [1], End-AAG_as552 [2], and End-AAG_as552 [1-2]. *B* - an IRAP Romania_aa951 sequence from *A. arenosa* absent in *A. suecica*: III - the total span of the sequence similar to the partial region of a gypsy-like retrotransposon family (*Athila*); LTR - expected internal segments: Romania_aa951 [1], Romania_aa951 [2], and Romania_aa951 [1-2]. *C* - an IRAP Romania_aa1357 sequence from *A. arenosa* absent in *A. suecica*: IV - a 1-446pb region composed by ~2.5 copies of the 180 bp centromeric repeats of *A. arenosa*; V - a 447-1357 bp region similar to *A. thaliana Athila* family retrotransposons; expected internal segments: Romania_aa1357 [1], Romania_aa1357 [2], and Romania_aa1357 [1-2].

A. suecica progenitor lines (tetraploid *A. thaliana* and *A. arenosa*) had markedly distinct profiles with a total of 59 and 58 bands, respectively, comprehending mainly species-specific bands (polymorphic bands) and a minor fraction (19/98 ~19 %) of common bands (monomorphic bands) (Table 3). Thus, absolutely additive banding profiles were expected to render 98 distinctive bands in the allopolyploid, markedly contrasting with the total number of 65 bands detected in the neosynthetized *A. suecica*. The high frequency (36/98 ~37 %) of bands absent in *A. suecica* comprehended 67 % (24/36) of *A. arenosa*-specific and 33 % (12/36) of *A. thaliana*-specific bands.

Sequence analysis of parental bands undetected in *A. suecica* through IRAP and REMAP (Table 1 Suppl.) included one *A. thaliana*-origin band and five *A. arenosa*-origin bands, namely: *1*) Tat_at944 (*A. thaliana Tat*1 IRAP, 944 bp, acc. No. KJ371934); *2*) Tat_aa740 (*A. arenosa Tat*1 IRAP, 740 bp, acc. No. KJ371931); *3*) Endovir_aa929 (*A. arenosa Endovir*1-1 IRAP, 929 bp, acc. No. KJ371933); *4*) End-AAG_aa387 (*A. arenosa Endovir*1-1/(AAG)6C REMAP, 387 bp, acc. No. KJ371928); *5*) Romania_aa951 (*A. arenosa Romania*T5 IRAP, 951 bp, acc. No. KJ371932); and *6*) Romania_aa1357 (*A. arenosa Romania*T5 IRAP, 1 357 bp, acc. No. KJ371935). The *BLAST* alignments of

A. suecica lost sequences revealed a significant homology with both repetitive and coding sequences. Four restructured sequences presented very significant alignments with protein coding regions of the *A. thaliana* genome mapped on distinct chromosomes: Tat_at944 - chromosome 1; Tat_aa740 - chromosome 3; Endovir_aa929 - chromosome 3; and End-AAG_aa387 - chromosome 4. From them, the REMAP End-AAG_aa387 sequence was proved to be amplified through the (AAG)6C anchored primer as described for the *A. suecica* novel bands obtained through the same REMAP experiment. On the other hand, Romania_aa951 and Romania_aa1357 sequences produced significant alignments with all *A. thaliana* centromeric regions. Romania_aa951 corresponded mainly to a partial region of the LTR of a gypsy-like retrotransposon family (*Athila*), and the Romania_aa1357 initial part (1 - 446 pb) was composed by ~2.5 units of 180 bp centromeric repeats whereas the remaining part corresponded also to *Athila*-like retrotransposon sequences.

To deeper assess the involvement of retrotransposon related sequences in parental genome adjustment in the allopolyploid line, internal primers were designed for *A. arenosa* remodelled sequences Romania_aa951 and Romania_aa1357 (primers in Table 1 and a schematic representation in Fig. 2). The results obtained in *A. arenosa* confirmed the amplification of three Romania_aa951 internal fragments with expected sizes (Table 2 Suppl.): Romania_aa951_[1] (294 bp); Romania_aa951_[2] (279 bp); and Romania_aa951_[1-2] (512 bp), which resulted from the amplification of the forward primer [1] with the reverse primer [2]. Contrary to the original Romania_aa951 band, the three predictable internal segments were also detected in the *A. suecica* allopolyploid. Romania_aa951 segments [1] and [2] were absent in *A. thaliana* di- and tetra-ploid lines (Fig. 3*B*) although the primers for segment [1-2] amplified two bands with ~400 and ~550 bp in both *A. thaliana* lines.

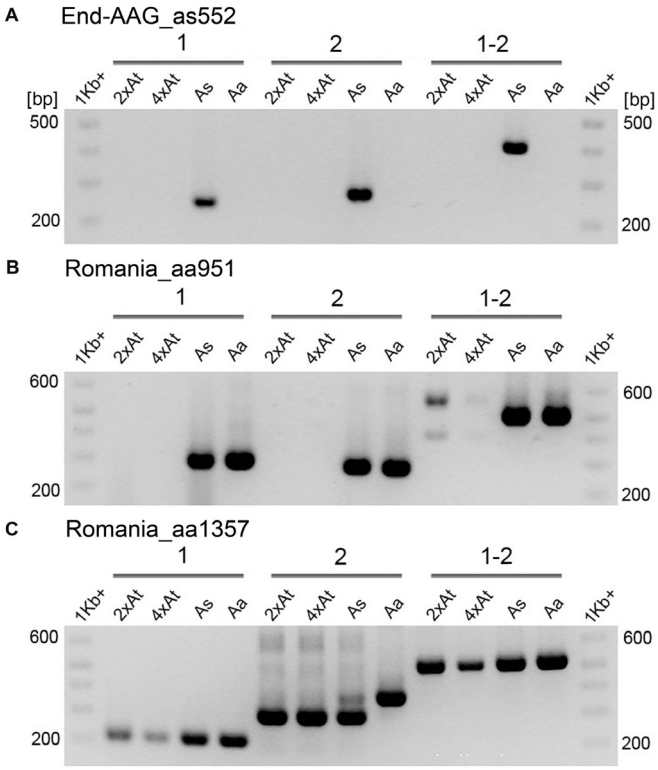

Fig. 3. The PCR banding profiles obtained with primers designed to amplify internal segments of rearranged sequences identified in *A. suecica*: *A* - End-AAG_as552 [1], [2], and [1-2], *B* - Romania_aa951 [1], [2], and [1-2], *C* - Romania_aa1357[1], [2], and [1-2]. *A. thaliana* (2×At), *A. thaliana* (4×At), *A. suecica* synthetic allopolyploid (As), and *A. arenosa* (Aa). The details of all primers used are presented in Table 1 and Table 2 Suppl. Molecular mass marker is marked as 1 Kb+.

Concerning Romania_aa1357, the designed primers targeted only a domain similar to *Athila* retrotransposons (Fig. 2) and the results obtained in *A. arenosa* confirmed the amplification of three fragments with expected sizes (Table 2 Suppl.): Romania_aa1357_[1] (217 bp); Romania_aa1357_[2] (387 bp); and Romania_aa1357_[1-2] (531 bp), which resulted from the amplification of the forward primer [1] with the reverse primer [2]. As observed for Romania_aa951, all Romania_aa1357 internal segments were found in the *A. suecica* allopolyploid (Fig. 3*C*). Moreover, the internal segments Romania_aa1357_[1] and Romania_aa1357_[1-2] were

also amplified in diploid and tetraploid *A. thaliana* yielding bands with the same sizes in all the genotypes. Amplification with primers for the segment Romania_aa1357_[2] revealed one band with ~300 bp in both diploid and tetraploid *A. thaliana* and *A. suecica* but absent in *A. arenosa*.

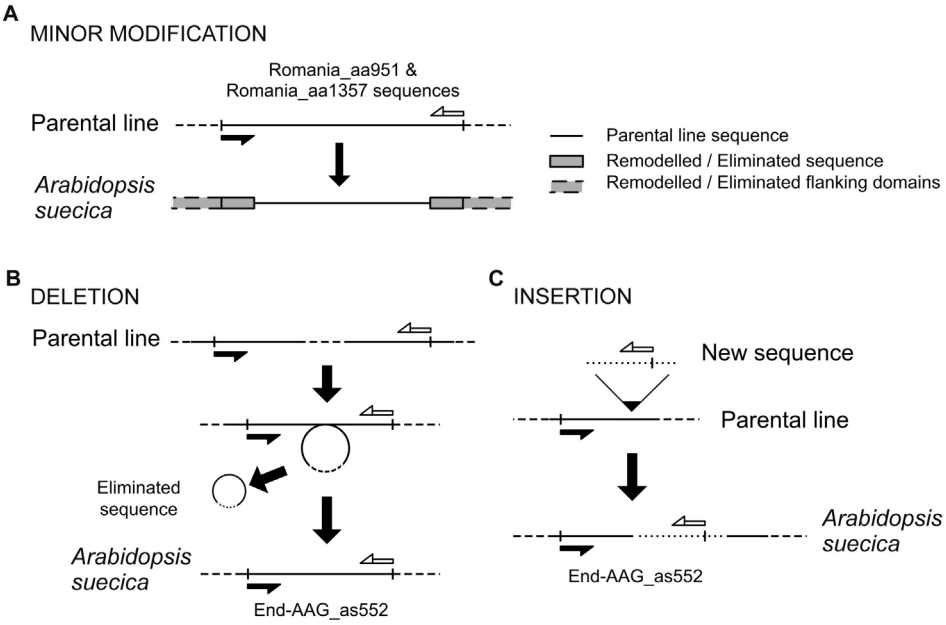

Fig. 4. *A* - Schematic representation of IRAP RomaniaT5 *A. arenosa* origin sequences (Romania_aa951 and Romania_aa1357). Although original IRAP bands were absent in *A. suecica*, their internal fragments were detected in the allopolyploid genotype revealing that polyploidization probably induced changes restricted to flanking domains of primer annealing sites. *B* and *C* - Possible rearrangements induced by polyploidization in the origin of a novel REMAP Endovir1-1/(AAG)6C sequence (End-AAG_as552) detected exclusively in *A. suecica* (*B*). It illustrates a deletion in one parental genome that allows the amplification of a novel PCR product; in *C*, it is illustrated the insertion of a sequence containing a primer annealing site into a new location near a second primer annealing site originating a new PCR product.

Discussion

The role of retrotransposon and microsatellite associated sequences in polyploid genome adjustment is hereby expressly corroborated through REMAP/IRAP evaluation of *Arabidopsis* auto- and allo-polyploid synthetic lines. The banding profiles of the *A. thaliana* autopolyploid line and its diploid progenitor revealed an overall genomic stability since all bands from the diploid line were observed in the autotetraploid one. Lack of parental sequence alterations was previously described in the *A. thaliana* autopolyploid through AFLP analysis (Ozkan *et al.* 2006). The undetection of band loss in the tetraploid *A. thaliana* line was expected considering the presence of multiple identical sequences due to its autopolyploid nature probably responsible for the maintenance of the described undistinguishable banding profiles. In fact, in the autotetraploid *A. thaliana* line, the only genomic alteration observed was a novel band detected with IRAP *Romania*T5. Conversely, in the *A. suecica* synthetic allopolyploid, an extensive rate of genome remodelling was observed corresponding to a major parental band elimination (~37 %) and a minor appearance of novel bands (4.6 %). Sequence elimination/modification revealed by band loss was commonly detected in almost all allopolyploid species previously studied, such as triticale, *Triticum*, *Brassica*, *Nicotiana*, and *Spartina* (Song *et al.* 1995, Shaked *et al.* 2001, Kashkush *et al.* 2002, Ma *et al.* 2004, Dong *et al.* 2005, Ma and Gustafson 2006, Bento *et al.* 2008, Eilam *et al.* 2008, Petit *et al.* 2010). However, the present results unravelled in *A. suecica* through the IRAP/REMAP methodologies do not corroborate the preferential loss of bands from the larger genome parental species as observed using similar methodologies in allopolyploid systems with large genomes like wheat, triticale (Bento *et al.* 2010), and *Cucumis* × *hytivus* (Jiang *et al.* 2009). In triticale, the proportion of rye/wheat origin band loss (20/5=4) (Bento *et al.* 2008) is considerably higher than the respective parental genome size ratio (8 600/5 648 Mbp=1.52) (Bennett and Leitch 2010). Conversely, differences detected in the frequencies of *A. arenosa* origin and *A. thaliana* origin band losses (24/36≅67 % and 12/36 ≅33 %, respectively) in *A. suecica* (24/12=2) were

equivalent to the parental genome size proportion (196/125 Mbp=1.57) [196 bp - *A. arenosa* (Bennett and Leitch 2010); 125 bp - *A. thaliana* (AGI 2000).

The high parental band loss here detected (37 %) markedly contrasts with the extremely low frequency of genomic modifications (2.3 %) detected in *A. suecica* when compared with the parental lines using AFLP (Madlung *et al.* 2005). The AFLP detects changes in restriction fragments occurring widely in all genome fractions and also detects epigenetic genome restructuring events if cytosine methylation sensitive restriction enzymes are used like in works previously referred (Madlung *et al.* 2005, Ozkan *et al.* 2006). Since the REMAP/IRAP methodologies are focused particularly on genome repetitive regions enriched in retrotransposons and/or microsatellites (Kalendar *et al.* 1999), the modifications now unravelled reinforces the idea that those repetitive sequences are preferentially altered by polyploidization as recently suggested (Bento *et al.* 2013). It would be tempting to address the involvement of those particularly labile sequences in natural *A. suecica* genome adjustment. In fact, major restructuring events involving rDNA *loci* were similarly detected in both natural and synthetic *A. suecica* lines (Pontes *et al.* 2004). However, since the exact parental lines of the natural *A. suecica* lines are virtually unknown, it is impossible to accurately evaluate that interesting issue using molecular tools like IRAP and REMAP.

The *BLASTN* analysis through search in the *NCBI*, *TAIR*, and *PlantSat* databases of *A. suecica* remodelled bands revealed a high degree of similarity with coding sequences (six of eight bands sequenced, Table 1 Suppl.). From them, the analysis of all three remodelled bands detected through REMAP *Endovir*1-1/(AAG)6C unravelled sequences flanking microsatellites. This finding corroborates the idea that although microsatellites are present in both non-coding and coding regions, their frequency is higher in transcribed regions (Rakoczy-Trojanowska and Bolibok 2004). Moreover, this result clearly emphasises the role of microsatellite sequences in allopolyploid genome dynamics as recently proposed by Shi *et al.* (2013). The remaining restructured sequences detected through IRAP analysis, presenting a high similarity with *A. thaliana* coding sequences, suggest that although retrotransposons in *Arabidopsis* tend to be preferentially clustered within pericentromeric heterochromatin regions (Peterson-Burch *et al.* 2004), polyploidy induced genome adjustment seems to affect mainly LTR flanking sequences intermingled with gene-rich domains. Despite the contrasting dimensions of repetitive sequence fractions in a plant with a large genomes like wheat (70 - 80 %) (Dvořák 2009) and with a small genome like *A. thaliana* (~14 %) (AGI 2000) the present work ultimately shows that the levels of restructuring events targeting repetitive sequences of neighbouring regions are similar in allopolyploid species with both large (Bento *et al.* 2008) and small genome.

The amplification of internal segments of *A. suecica* remodelled sequences allowed a deeper understanding of major rearrangement occurrences. The internal amplify-cation of the Romania_aa951 and Romania_aa1357 retrotransposon-related sequences in *A. suecica* suggests that the absence of parental bands in the allopolyploid might result from rearrangements confined to primer annealing site regions (Fig. 4*A*). Contrastingly, the amplification of the End-AAG_as552 allopolyploid novel band using the primers targeting a domain with no significant homology with any known sequence corroborates the novelty of this restructured band since no amplification was detected in both the parental lines (Fig. 3*A*). The origin of such a new band in *A. suecica* could result either from the deletion of an intercalary sequence or from an insertion between the primer regions allowing the emergence of a new PCR amplicon (Fig. 4*B*,*C*). Likewise, allopolyploid genome remodelling events now unravelled may be explained by a high rate of homoeologous pairing and intergenomic exchange recently reported to occur during meiosis of equivalent *A. suecica* neoallopolyploid lines (Henry *et al.* 2014).

Conclusively, the IRAP/REMAP comparative analysis performed in *Arabidopsis* auto- and allo-polyploids clearly revealed distinct outcomes in newly formed polyploids. The overall genomic stability detected in the autopolyploid *A. thaliana* line, confirming previously described results using different techniques (Ozkan *et al.* 2006), contrasts with high parental genome remodelling in synthetic *A. suecica* involving repetitive and coding sequences. Furthermore, a deeper examination of parental lost sequences revealed that parental band loss mainly involved punctual changes on primers annealing sites in accordance with previous studies on allopolyploids with large genomes (Bento *et al.* 2010). However, contrary to allopolyploids with large genomes like triticale (Bento *et al.* 2008), a preferential modification of a larger parental genome was not observed revealing distinct parental genome behaviours in allopolyploid species with small genomes. The high parental band loss here reported suggest a major role of microsatellites and retrotransposons-related sequences in remodelling routes necessary to the establishment of new fertile polyploids. Moreover, repetitive sequence rearrangements may be highly involved in stable genetic changes needed to resolve parental genome conflicts that Comai *et al.* (2000) suggested to occur following the phenotypic instabilities and gene silencing events observed in the first generations of the same *A. suecica* synthetic line.

References

AGI (The Arabidopsis Genome Initiative): Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. - Nature **408**: 796-815, 2000.

Ainouche, M.L.: Plant polyploidy. - New Phytol. **186** (Special Issue): 1-261, 2010.

Beaulieu, J., Jean, M., Belzile, F.: The allotetraploid *Arabidopsis thaliana-Arabidopsis lyrata* subsp. *petraea* as an alternative model system for the study of polyploidy in plants. - Mol. Genet. Genomics. **281**: 421-435, 2009.

Bennett, M.D., Leitch, I.J.: Plant DNA C-values Database Release 6.0, Dec. 2012. http://data.kew.org/cvalues.

Bento, M., Gustafson, P., Viegas, W., Silva, M.: Genome merger: from sequence rearrangements in triticale to their elimination in wheat-rye addition lines. - Theor. appl. Genet. **121**: 489-497, 2010.

Bento, M., Pereira, S., Rocheta, M., Gustafson, P., Viegas, W., Silva, M.: Polyploidization as a retraction force in plant genome evolution: sequence rearrangements in triticale. - PLoS ONE **3**: e1402, 2008.

Bento, M., Tomas, D., Viegas, W., Silva, M.: Retrotransposons represent the most labile fraction for genomic rearrangements in polyploid plant species. - Cytogenet. Genome Res. **140**: 286-294, 2013.

Bomblies, K., Madlung, A.: Polyploidy in the *Arabidopsis* genus. - Chromosome Res. **22**: 117-134, 2014.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G.L., D'Amore, R., Allen, A.M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M.C., Sehgal, S., Gill, B., Kianian, S., Anderson, O., Kersey, P., Dvorak, J., McCombie, W.R., Hall, A., Mayer, K.F., Edwards, K.J., Bevan, M.W., Hall, N.: Analysis of the bread wheat genome using whole-genome shotgun sequencing. - Nature **491**: 705-710, 2012.

Comai, L., Tyagi, A.P., Winter, K., Holmes-Davis, R, Reynolds, S.H., Stevens, Y., Byers, B.: Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. - Plant Cell **12**: 1551-1567, 2000.

Dong, Y.Z., Liu, Z.L., Shan, X.H., Qiu, T., He, M.Y., Liu, B.: Allopolyploidy in wheat induces rapid and heritable alterations in DNA methylation patterns of cellular genes and mobile elements. - Russ. J. Genet. **41**: 890-896, 2005.

Dvořák, J.: Triticeae genome structure and evolution. - In: Muehlbauer, G.J., Feuillet, C. (ed.): Genetics and Genomics of the Triticeae. Pp. 685-711. Springer, New York 2009.

Eilam, T., Anikster, Y., Millet, E., Manisterski, J., Feldman, M.: Nuclear DNA amount and genome downsizing in natural and synthetic allopolyploids of the genera *Aegilops* and *Triticum*. - Genome **51**: 616-627, 2008.

Feldman, M., Levy, A.A.: Genome evolution in allopolyploid wheat-a revolutionary reprogramming followed by gradual changes. - J. Genet. Genomics **36**: 511-518, 2009.

Ferreira-de-Carvalho, J., Chelaifa, H., Boutte, J., Poulain, J., Couloux, A., Wincker, P., Bellec, A., Fourment, J., Berge`s, H., Salmon, A., Ainouche, M.: Exploring the genome of the salt-marsh *Spartina maritima* (*Poaceae*, *Chloridoideae*) through BAC end sequence analysis. - Plant mol. Biol. **83**: 591-606, 2013.

Hazzouri, R.M., Mohajer, A., Dejak, S.I., Otto, S.P., Wright, S.I.: Contrasting patterns of transposable-element insertion polymorphism and nucleotide diversity in autotetraploid and allotetraploid *Arabidopsis* species. - Genetics **179**: 581-592, 2008.

Henry, I.M., Dilkes, B.P., Tyagi, A., Gao, J., Christensen, B., Comai, L.: The BOY NAMED SUE quantitative trait locus confers increased meiotic stability to an adapted natural allopolyploid of *Arabidopsis*. - Plant Cell. **26**: 181-194, 2014.

Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., Lucas, W.J., Wang, X., Xie, B., Ni, P., Ren, Y., Zhu, H., Li, J., Lin, K., Jin, W., Fei, Z., Li, G., Staub, J., Kilian, A., Van der Vossen, E.A., Wu, Y., Guo, J., He, J,, Jia, Z., Ren, Y., Tian, G., Lu, Y., Ruan, J., Qian, W., Wang, M., Huang, Q., Li, B., Xuan, Z., Cao, J., Asan, Wu, Z., Zhang, J., Cai, Q., Bai, Y., Zhao, B., Han, Y., Li, Y., Li, X., Wang, S., Shi, Q., Liu, S., Cho, W.K., Kim, J.Y., Xu, Y., Heller-Uszynska, K., Miao, H., Cheng, Z., Zhang, S., Wu, J., Yang, Y., Kang, H, Li, M., Liang, H., Ren, X., Shi, Z., Wen, M., Jian, M., Yang, H., Zhang, G., Yang, Z., Chen, R., Liu, S., Li, J., Ma, L., Liu, H., Zhou, Y., Zhao, J., Fang, X., Li, G., Fang, L., Li, Y., Liu, D., Zheng, H., Zhang, Y., Qin, N., Li, Z., Yang, G., Yang, S., Bolund, L., Kristiansen, K., Zheng, H., Li, S., Zhang, X., Yang, H., Wang, J., Sun, R., Zhang, B., Jiang, S., Wang, J., Du, Y., Li, S.: The genome of the cucumber, *Cucumis sativus* L. - Nat. Genet. **41**: 1275-1281, 2009.

Jiang, B., Lou, Q., Wu, Z., Zhang, W., Wang, D., Mbira, K.G., Weng, Y., Chen, J.: Retrotransposon- and microsatellite sequence-associated genomic changes in early generations of a newly synthesized allotetraploid *Cucumis* × *hytivus* Chen & Kirkbride. - Plant mol. Biol. **77**: 225-233, 2009.

Jiao, Y.N., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H.Y., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E., Schuster, S.C., Ma, H., Leebens-Mack, J., De Pamphilis, C.W.: Ancestral polyploidy in seed plants and angiosperms. - Nature **473**: 97-100, 2011.

Jones, R.N., Hegarty, M.: Order out of chaos in the hybrid plant nucleus. - Cytogenet. Genome Res. **126**: 376-389, 2009.

Kalendar. R., Grob, T., Regina, M., Suoniemi, A., Schulman, A.: IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. - Theor. appl. Genet. **98**: 704-711, 1999.

Kalendar, R., Schulman, A.H.: IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. - Nat. Protocols **1**: 2478-2484, 2006.

Kashkush, K., Feldman, M., Levy, A.A.: Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. - Genetics **160**: 1651-1659, 2002.

Ma, X.F., Fang, P., Gustafson, J.P.: Polyploidization-induced genome variation in triticale. - Genome **47**: 839-848, 2004.

Ma, X.F., Gustafson, J.P.: Timing and rate of genome variation in triticale following allopolyploidization. - Genome **49**: 950-958, 2006.

Ma, X.F., Gustafson, J.P.: Allopolyploidization-accommodated genomic sequence changes in triticale. - Ann. Bot. **101**: 825-832, 2008.

Madlung, A., Tyagi, A.P., Watson, B., Jiang, H.M., Kagochi, T., Doerge, R.W., Martienssen, R., Comai, L.: Genomic changes in synthetic *Arabidopsis* polyploids. - Plant J. **41**: 221-230, 2005.

Ozkan, H., Tuna, M., Galbrainth, D.W.: No DNA loss in

autotetraploid of *Arabidopsis thaliana*. - Plant Breed. **125**: 288-291, 2006.

Parisod, C., Salmon, A., Zerjal, T., Tenaillon, M., Grandbastien, M.A., Ainouche, M.: Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. - New Phytol. **184**: 1003-1015, 2009.

Pecinka, A., Fang, W., Rehmsmeier, M., Levy, A.A., Scheid, O.M.: Polyploidization increases meiotic recombination frequency in *Arabidopsis*. - BMC Biol. **9**: 24, 2011.

Peterson-Burch, B.D., Nettleton, D., Voytas, D.F.: Genomic neighborhoods for *Arabidopsis* retrotransposons: a role for targeted integration in the distribution of the *Metaviridae*. - Genome Biol. **5**: R78, 2004.

Petit, M., Guidat, C., Daniel, J., Denis, E., Montoriol, E., Bui, Q.T., Lim, K.Y., Kovarik, A., Leitch, A.R., Grandbastien, M.A., Mhiri, C.: Mobilization of retrotransposons in synthetic allotetraploid tobacco. - New Phytol. **186**: 135-147, 2010.

Pontes, O., Neves, N., Silva, M., Lewis, M.S., Madlung, A., Comai, L., Viegas, W., Pikaard, C.S.: Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid *Arabidopsis suecica* genome. - PNAS **101**:18240-18245, 2004.

Rakoczy-Trojanowska, M., Bolibok, H.: Characteristics and a comparison of three classes of microsatellite-based markers and their application in plants. - Cell. mol. Biol. Lett. **9**: 221-238, 2004.

Renny-Byfield, S., Kovarik, A., Kelly, L.J., Macas, J., Novak, P., Chase, M.W., Nichols, R.A., Pancholi, M.R., Grandbastien, M.A., Leitch, A.R.: Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. - Plant J. **74**: 829-839, 2013.

Rocheta, M., Cordeiro, J., Oliveira, M., Miguel, C.: PpRT1: the first complete gypsy-like retrotransposon isolated in *Pinus pinaster*. - Planta **225**: 551-562, 2007.

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A., Allard, R.W.: Ribosomal DNA spacer-length polymorphisms in barley - Mendelian inheritance, chromosomal location, and population-dynamics. - PNAS **81**: 8014-8018, 1984.

Santos, J.L., Alfaro, D., Sanchez-Moran, E., Armstrong, S.J., Franklin, F.C.H.: Partial diploidization of meiosis in autotetraploid *Arabidopsis thaliana*. - Genetics **165**: 1533-1540, 2003.

Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., Levy, A.A.: Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. - Plant Cell **13**: 1749-1759, 2001.

Shi, J.Q., Huang, S.M., Fu, D.H., Yu, J.Y., Wang, X.F., Hua, W., Liu, S.Y., Liu, G.H., Wang, H.Z.: Evolutionary dynamics of microsatellite distribution in plants: insight from the comparison of sequenced *Brassica*, *Arabidopsis* and other angiosperm species. - Plos One. **8**: e59988, 2013.

Sierro, N., Battey, J.N., Ouadi, S., Bovet, L., Goepfert, S., Bakaher, N., Peitsch, M.C., Ivanov, N.V.: Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. - Genome Biol. **14**: R60, 2013.

Song, K.M., Lu, P., Tang, K.L., Osborn, T.C.: Rapid genome change in synthetic polyploids of *Brassica* and Its implications for polyploid evolution. - PNAS **92**: 7719-7723, 1995.

Stöck, M., Lamatsch, D.K.: Trends in polyploidy research in animals and plants. - Cytogenet. Genome Res. **140**: 2-4, 2013.

Yu, Z., Haberer, G., Matthes, M., Rattei, T., Mayer, K.F.X., Gierl, A., Torres-Ruiz, R.A.: Impact of natural genetic variation on the transcriptome of autotetraploid *Arabidopsis thaliana*. - PNAS **107**: 17809-17814, 2010.