



Mobilization of retrotransposons in synthetic allotetraploid tobacco

M. Petit, C. Guidat, J. Daniel, E. Denis, Elsa Montoriol, Quynh Trang Bui, Ky Lim, A. Kovarik, Ar Leitch, Marie Angele Grandbastien, et al.

► To cite this version:

M. Petit, C. Guidat, J. Daniel, E. Denis, Elsa Montoriol, et al.. Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytologist*, 2010, 186 (1), pp.135-147. 10.1111/j.1469-8137.2009.03140.x . hal-01203912

HAL Id: hal-01203912

<https://hal.science/hal-01203912>

Submitted on 31 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mobilization of retrotransposons in synthetic allotetraploid tobacco

M. Petit¹, C. Guidat¹, J. Daniel¹, E. Denis¹, E. Montoriol¹, Q. T. Bui¹, K. Y. Lim², A. Kovarik³, A. R. Leitch², M.-A. Grandbastien¹ and C. Mhiri¹

¹Institut Jean-Pierre Bourgin, INRA-Centre de Versailles, F-78026, Versailles cedex, France; ²School of Biological and Chemical Sciences, Queen Mary, University of London, London E1 4NS, UK; ³Institute of Biophysics, Academy of Sciences, v.v.i., 612 65 Brno, Czech Republic

Summary

Author for correspondence:
Marie-Angèle Grandbastien
Tel: +33 1 30 83 30 24
Email: gbastien@versailles.inra.fr

Received: 27 August 2009
Accepted: 11 November 2009

New Phytologist (2010) **186**: 135–147
doi: 10.1111/j.1469-8137.2009.03140.x

Key words: allopolyploidy, evolution, retrotransposon, stress, *Nicotiana*, sequence-specific amplification polymorphism (SSAP), Tnt1.

- Allopolyploidy is a major driving force in plant evolution and can induce rapid structural changes in the hybrid genome. As major components of plant genomes, transposable elements are involved in these changes. In a previous work, we observed turnover of retrotransposon insertions in natural allotetraploid tobacco (*Nicotiana tabacum*). Here, we studied the early stages of allopolyploid formation by monitoring changes at retrotransposon insertion sites in the Th37 synthetic tobacco.
- We used sequence-specific amplification polymorphism (SSAP) to study insertion patterns of two populations of the Tnt1 retrotransposon in Th37 S4 generation plants, and characterized the nature of polymorphic insertion sites.
- We observed significant amplification of young Tnt1 populations. Newly transposed copies were amplified from maternal elements and were highly similar to Tnt1A tobacco copies amplified in response to microbial factors. A high proportion of paternal SSAP bands were not transmitted to the hybrid, corresponding to various rearrangements at paternal insertion sites, including indels or the complete loss of the Tnt1/flanking junction.
- These data indicate that major changes, such as retrotransposon amplification and molecular restructuring in or around insertion sites, occur rapidly in response to allopolyploidy.

Introduction

The formation of allopolyploid species is generally associated with major structural, cytogenetic, epigenetic and functional changes to parental genomes, leading to new phenotypes and reproductive isolation (reviewed in Wendel, 2000; Ramsey & Schemske, 2002; Comai *et al.*, 2003; Adams & Wendel, 2005; Chen & Ni, 2006; Feldman & Levy, 2009; Michalak, 2009). Rapid and dynamic changes including DNA loss and gain have been detected at early stages of the allopolyploidy process in many plant models (Song *et al.*, 1995; Chen & Ni, 2006; Petit *et al.*, 2007; Feldman & Levy, 2009). McClintock (1984) proposed that the activation of transposable elements could play a role in genomic changes and may occur as a response to the genome shock of species hybridization. In support of this hypothesis, there are several reports of transcriptional activation of transposable elements following interspecific

hybridization (Liu & Wendel, 2000; Kashkush *et al.*, 2002, 2003; Liu *et al.*, 2004; Madlung *et al.*, 2005; Shan *et al.*, 2005; Josefsson *et al.*, 2006).

There are fewer examples of correlated transposable element mobilization. They include mobility of the Sunfish transposon in *Arabidopsis thaliana* × *Arabidopsis arenosa* synthetic allotetraploids (Madlung *et al.*, 2005) and mobility of retrotransposons and miniature inverted repeat transposable elements (MITEs) in rice (*Oryza sativa*) in response to introgressions from the wild species *Zizania latifolia* (Liu & Wendel, 2000; Shan *et al.*, 2005). Other indirect evidence includes the large-scale spreading of repetitive DNA to homeologous subgenomes in *Gossypium* natural allopolyploids (Hanson *et al.*, 1998; Zhao *et al.*, 1998), and retrotransposon proliferation in several hybrid sunflower (*Helianthus* species), which may have resulted from inter-specific hybridization events, together with environmental changes (Ungerer *et al.*, 2006). By contrast, the strong

transcriptional activation of the *Wis2* retrotransposon in synthetic allopolyploid wheat (*Aegilops sharonensis* × *Triticum monococcum*) does not correlate with its amplification (Kashkush *et al.*, 2003), and no evidence of mobility of several transposons was detected in *A. thaliana* × *Arabidopsis lyrata* synthetic allotetraploids (Beaulieu *et al.*, 2009). Similarly, allotetraploidization appears not to have enhanced proliferation of retrotransposons in natural allopolyploid wheats (Charles *et al.*, 2008) and no evidence of transposable element proliferation in response to allopolyploidy was detected in natural *Brassica* species (Alix & Heslop-Harrison, 2004; Alix *et al.*, 2008). Allopolyploidy-induced mobilization of transposable elements may thus be restricted to some specific elements and/or model systems (for a review, see Parisod *et al.*, 2010).

The *Tnt1A copia*-type retrotransposon of tobacco (*Nicotiana tabacum*) is one of the rare active plant retrotransposons whose amplification has been directly demonstrated, notably in response to microbial stress (Melayah *et al.*, 2001). *Tnt1A* was originally isolated from tobacco (*N. tabacum*) (Grandbastien *et al.*, 1989) and *Tnt1* elements are widely distributed in the genus *Nicotiana* (Melayah *et al.*, 2004). Tobacco is a recent allotetraploid that probably formed by interspecific hybridization of two distantly related diploid species < 200 000 yrs ago (Clarkson *et al.*, 2005; Leitch *et al.*, 2008) or even within the last 20 000 yrs (M. Chase, pers. comm.). The closest living diploid descendants of the progenitors of tobacco are *Nicotiana sylvestris* and *Nicotiana tomentosiformis*, representing the maternally derived S genome and the paternally derived T genome, respectively (Murad *et al.*, 2002). Several papers have described patterns of genetic restructuring subsequent to polyploidy, including intergenomic translocations on both S and T chromosomes (Lim *et al.*, 2004a), reductions in *N. tomentosiformis*-derived endogenous pararetroviruses (Gregor *et al.*, 2004; Matzke *et al.*, 2004) and replacement of rDNA units with novel derivatives (Kovarík *et al.*, 2004; Lim *et al.*, 2004b). Thus the formation of the tobacco genome has been associated with changes to both parental genomes, but these changes have not significantly influenced the expected genome size, that is, the sum of the genome sizes found in the two diploid species (Leitch *et al.*, 2008), as has also been observed in tetraploid cotton (*Gossypium hirsutum*) (Wendel, 2000). Thus the interaction between *Tnt1* elements and their *Nicotiana* hosts provides an excellent model for the study of the evolutionary dynamics of retrotransposons in response to allopolyploidy.

Previously, we compared the distribution of *copia*-type retrotransposons between natural *N. tabacum* and its progenitors, *N. sylvestris* and *N. tomentosiformis*, and showed that the tobacco genome has undergone substantial turnover of retrotransposon sequences, with sequence removals counterbalanced by new insertions (Petit *et al.*, 2007). We detected behavioral patterns that are specific to each retro-

transposon population, both in terms of transmission to the allopolyploid genome and in terms of activation levels, with younger retrotransposon populations showing higher rates of new insertions in tobacco. Most changes were found to be similar across a range of tobacco landraces, suggesting that they occurred at early stages of the formation of the natural tobacco species, perhaps in response to the genome shock of allopolyploidy.

In order to evaluate the timing of these changes, further analysis using newly resynthesized tobacco is necessary. Because of the high levels of divergence between the parental *N. sylvestris* and *N. tomentosiformis* genomes, which split over 10 million yrs ago (Clarkson *et al.*, 2005), attempts to resynthesize the tobacco genome have been largely unsuccessful, with doubled F₁ hybrids showing a lack of fertility in most cases (Lim *et al.*, 2006). So far, the only successful fertile synthetic tobacco line is Th37, made from a cross between *N. sylvestris* (2n = 24) as the maternal parent and *N. tomentosiformis* (2n = 24) as the paternal parent, and converted to a fertile amphidiploid by genome doubling (Burk, 1973). Here we determined if changes previously observed in natural tobacco (Petit *et al.*, 2007) were also detectable in early stages of the allopolyploidy process. This was achieved by evaluating the behavior and fate of retrotransposon populations in the S4 generation of Th37. We monitored two populations of *Tnt1* retrotransposons endogenous to *N. tabacum* via sequence-specific amplification polymorphism (SSAP), a high-resolution retrotransposon-anchored PCR strategy which allows the simultaneous detection of multiple insertions (Waugh *et al.*, 1997). The two populations of *Tnt1* retrotransposons selected were previously described as displaying contrasting behaviors, with the *Tnt1-ol13* population comprising younger *Tnt1* elements with higher rates of new insertions in natural tobacco, while the *Tnt1-ol16* population is more ancient and showed reduced rates of new insertions (Petit *et al.*, 2007). We quantified losses and amplifications of these populations in Th37 by comparing SSAP polymorphisms of Th37 S4 generation plants with progenitor diploid accessions. We also investigated the nature of the molecular changes leading to SSAPs.

Our work shows that the *Tnt1-ol13* retrotransposon population has amplified significantly in the newly resynthesized tobacco and that a high proportion of paternal *N. tomentosiformis* SSAP bands are absent in synthetic tobacco. Insertion sites associated with *N. tomentosiformis* SSAP bands that disappeared in Th37 S4 generation plants were variously restructured, for example through the acquisition of indels, or the complete loss of *Tnt1*/flanking junctions. Together these data indicate that significant levels of amplification of young retrotransposon populations, as well as sequence divergence and molecular restructuring in or around parental insertion sites, occur rapidly in response to allopolyploidy in *Nicotiana*.

Materials and Methods

Plant material and DNA extraction

The S4 progeny of the synthetic hybrid Th37 tobacco (*Nicotiana*) line was obtained from the US National Plant Germplasm System (NPGS; USDA, North Carolina State University, Raleigh, NC, USA) as a kind gift from Professor Verne A. Sissons (USDA, Agricultural Research Station, North Carolina State University). The original Th37 F₁ hybrid ($n + n'$) was converted to a fertile amphidiploid ($2n + 2n'$) by regeneration from *in vitro* callus culture (Burk, 1973). Selfed S1 seeds were obtained from a single regenerated S0 plant. The F₁, S0, S1, S2 and S3 generations of Th37 are no longer available. The S4 generation was obtained by pooling the seeds of several S3 plants, and the 18 S4-progeny plants used in this work (Th37-1 to Th37-19, excluding Th37-18) are the same as those analyzed in previous works (Skalicka *et al.*, 2003, 2005; Kovarik *et al.*, 2004; Lim *et al.*, 2004a). S5 progenies were obtained by selfing of Th37-1, Th37-3 and Th37-9 S4 plants. No published information is available on the *Nicotiana sylvestris* Speg. & Comes and *Nicotiana tomentosiformis* Goodsp. accessions used as parents for the creation of the Th37 hybrid (Burk, 1973). However, it is likely that accessions available in the USDA germplasm were used (information kindly provided by Dr Earl Wernsman, Crop Science Department, North Carolina State University, and Dr Jennifer Levin, Tobacco Genetics and Breeding, North Carolina State University). We therefore used as parental controls all three *N. sylvestris* accessions (ac. TW136, TW137 and TW138) and the single *N. tomentosiformis* accession (ac. TW142) available from the US NPGS (USDA, North Carolina State University), obtained as a kind gift from Dr Jennifer S. Levin. We previously observed very low levels of amplified fragment length polymorphism (AFLP) and SSAP diversity across several *N. sylvestris* accessions obtained from diverse germplasm sources (including TW137) (Petit *et al.*, 2007). Similarly, very few SSAPs were detected among the TW136, TW137 and TW138 *N. sylvestris* accessions (see Fig. 1b). As none of these polymorphic bands was detected in Th37 S4 plants, we did not take them into account in our analyses. As a *Nicotiana tabacum* L. control we used the XHFD8 cv. Xanthi line (Bourgin *et al.*, 1973). Nuclear DNA was extracted from fresh young leaves as in Kovarik *et al.* (2000).

SSAP analysis and characterization of SSAP fragments

The two retrotransposon primers Tnt1-ol13 and Tnt1-ol16, used to target two different Tnt1 populations, are located in the Tnt1 LTR U5 region, and have been described in Petit *et al.* (2007). The SSAP strategy is schematized in Fig. 1(a), and the protocol has been described in

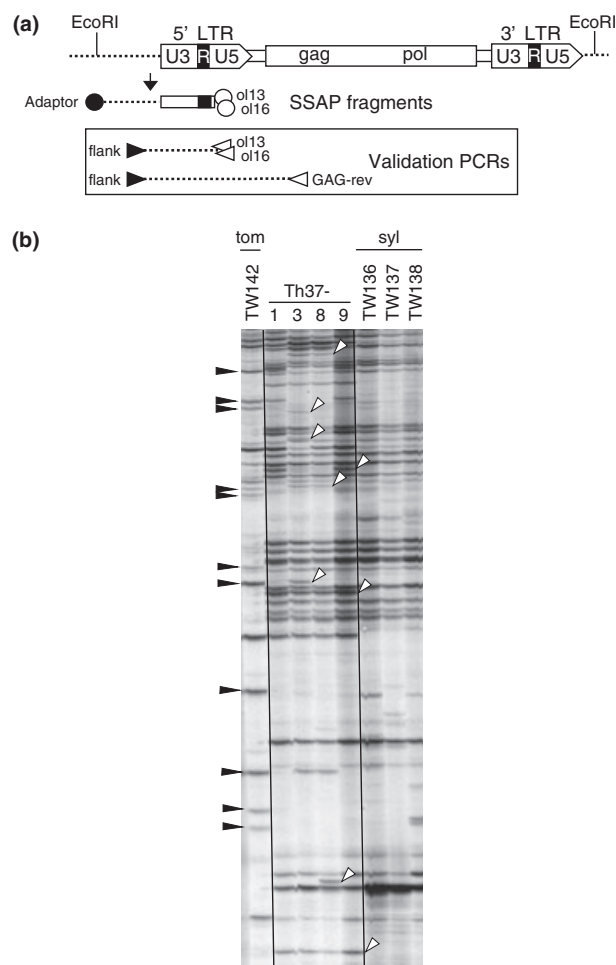


Fig. 1 Diagram showing the sequence-specific amplification polymorphism (SSAP) strategy and an example of SSAP profiles. (a) After DNA digestion by the *EcoRI* restriction enzyme and adaptor ligation, Tnt1/flanking junctions are amplified between a primer designed against the Tnt1 long terminal repeat (LTR) and a primer against the *EcoRI* adaptor. SSAP primers are represented as circles, and primers used for validation PCRs are represented as triangles. (b) Example of SSAP polymorphisms among the four Th37 S4 plants and parental accessions for the Tnt1-ol13 population. Only partial profiles are shown. tom, paternal *Nicotiana tomentosiformis* TW142; syl, the three *N. sylvestris* TW136, TW136 and TW137 accessions, of which one was used as the maternal parent for the production of the Th37 hybrid. White arrowheads, new SSAP bands present in all or some Th37 S4 plants; black arrowheads, *N. tomentosiformis* bands missing in some or all Th37 S4 plants.

Petit *et al.* (2007). Tnt1-ol13 and Tnt1-ol16 SSAP bands of interest were excised from the dried gels using a razor blade and re-suspended in 60 μ l of TE buffer (10 mM Tris-HCl and 1 mM EDTA). DNA was eluted from the gel bands after an overnight incubation at 37°C, and was collected after a quick centrifugation to remove the gel and associated paper. A 5- μ l sample was then used as a DNA template in PCR amplification using the same primers and protocol as those used for SSAP. Amplification products were visualized and quantified on 1% agarose gels. PCR

products were directly cloned in the pGEMT-Easy Vector kit (Promega) according to the manufacturer's recommendations. A volume of 1 µl of ligation product was electroporated with 40 µl of laboratory-made electrocompetent *Escherichia coli* cells (DH10B strain). Clones containing inserts were selected by PCR on bacteria using the T7 and SP6 universal primers located on each side of the cloning site. Sequencing was performed by Genoscreen (Lille, France). Nucleotide sequences of the SSAP clones are presented in Supporting Information Note S1. SSAP bands isolated from Tnt1-ol13 profiles were numbered NB13-xy or DTB13-xy, and SSAP bands isolated from Tnt1-ol16 profiles were numbered NB16-xy or DTB16-xy (see Results).

Primers were designed against the genomic region flanking each Tnt1 insertion ('flank', Fig. 1a; primer sequences in Note S1). Direct PCRs, using these primers in combination with the Tnt1-ol13 or the Tnt1-ol16 primer, were performed on the Th37 S4 progenies, and on the control TW142 (*N. tomentosiformis*), TW137 (*N. sylvestris*) and XHFD8 (*N. tabacum*) accessions. The PCR products were directly sequenced by Genoscreen. When no amplification product was detected with these primer combinations, suggesting a deletion of the insertion site, we tested whether the deletion extended into the Tnt1 internal region by performing additional PCRs using several primers designed against the flanking region in combination with the GAG-Rev primer (5'-CCATTGAATTTTGCTACCTCGT-3', position 708–729) designed in the Tnt1 (X13777) Gag coding region.

Phylogenetic analysis

Partial sequences of genomic Tnt1 copies were previously characterized in *N. sylvestris* and *N. tomentosiformis* (Vernhettes *et al.*, 1998). The U3-R-U5 region of these sequences

was manually extracted up to the Tnt1-ol13 primer annealing site. Similarly, partial long terminal repeat (LTR) sequences obtained during this work were curated up to the Tnt1-ol13 primer annealing site in cases where they were amplified using the Tnt1-ol16 primer, located further downstream (Fig. 1a). Sequences were aligned using the CLUSTAL W multiple-alignment program at EMBL-EBI (European Molecular Biology Laboratory – European Bioinformatics Institute; <http://www.ebi.ac.uk/Tools/clustalw2/index.html>; Larkin *et al.*, 2007). Alignments were manually curated using JALVIEW version 2.4 (Clamp *et al.*, 2004), to take into account variations in the number of BII repeated boxes in the U3 region of Tnt1A copies (Vernhettes *et al.*, 1998). Phylogenetic analyses were performed using Bayesian inference (BI) with MRBAYES v3.1.2 (<http://mrbayes.csit.fsu.edu/index.php>; Ronquist & Huelsenbeck, 2003).

Results

SSAP analysis of Th37 S4 plants

Previous work showed that Th37 S4 generation plants are polymorphic for several genomic loci, allowing their classification into three groups according to rDNA polymorphisms and cytogenetic studies (Skalicka *et al.*, 2003, 2005; Kovarik *et al.*, 2004; Lim *et al.*, 2004a). Group 3 plants are characterized by higher genomic changes than Group 1 plants, with Group 2 plants showing intermediate patterns. For SSAP analysis, we selected four plants of the S4 generation, Th37-1 as a representative of Group 2, Th37-3 and Th37-8 as representatives of Group 3 and Th37-9 as a representative of Group 1.

SSAP insertion polymorphisms of the two Tnt1 populations were compared across the four synthetic tobacco S4 plants (Th37-1, Th37-3, Th37-8 and Th37-9) and parental

			Th37 S4 plants				
			G1 ^b	G2 ^b	G3 ^b		
	tom ^a	syl ^a	9	1	3	8	All ^c
Tnt1-ol13							
Number of scored bands	23	40	54	54	59	56	65
Number of new bands	–	–	6	6	12	10	15
% new bands/AP ^d	–	–	9.5	9.5	19.0	15.9	23.8
Tnt1-ol16							
Number of scored bands	34	43	55	55	56	55	58
Number of new bands	–	–	3	3	4	4	5
% new bands/AP ^d	–	–	3.9	3.9	5.2	5.2	6.5

^atom, all SSAP bands scored in *Nicotiana tomentosiformis*; syl, all SSAP bands scored in *N. sylvestris*.

^bGx, groups of Th37 S4 plants as described in Skalicka *et al.* (2003). Numbers within each Gx group refer to each S4 plant number, as described in Skalicka *et al.* (2003).

^cBands scored across the four Th37 S4 plants.

^dAP = additivity of parental *N. sylvestris* and *N. tomentosiformis* SSAP bands.

Table 1 New sequence-specific amplification polymorphism (SSAP) bands detected in Th37 S4 generation plants

diploid accessions. A typical SSAP profile obtained for Tnt1-ol13 is shown in Fig. 1(b). We scored new bands not observed in either parental species (Table 1, Fig. 1b). The proportion of new bands was evaluated relative to the additivity of *N. sylvestris* and *N. tomentosiformis* parental bands. Summing all four Th37 S4 plants together revealed the percentage of new bands to be 23.8% for Tnt1-ol13 and 6.5% for Tnt1-ol16. Although some new bands were common to all four Th37 plants, inter-plant variability was observed for several new bands (as illustrated in Fig. 1b). The same new bands were often observed together in the Th37-3 and Th37-8 plants, both occurring in Group 3 (as illustrated in Fig. 1b).

We scored profiles for parental bands inherited in Th37 S4 generation plants. *N. sylvestris* SSAP bands of the two Tnt1 populations are nearly fully transmitted and only one Tnt1-ol16 band is absent in Th37 S4 plants (data not shown). By contrast, a large number of *N. tomentosiformis*-derived SSAP bands are absent in some or all Th37 S4 generation plants analyzed (Table 2; Fig. 1b). Percentages of missing bands were calculated relative to the fraction of bands found in *N. tomentosiformis*, excluding bands shared with *N. sylvestris*. SSAP bands shared between *N. sylvestris* and *N. tomentosiformis* are thought to correspond to ancient homeologous insertions predating the divergence of the two species (Petit *et al.*, 2007). As SSAP generates dominant markers, the fate of these common bands in the synthetic hybrid cannot be evaluated, and we thus excluded them from our calculations. Scoring all four Th37 S4 generation plants together revealed the proportion of *N. tomentosiformis*-derived parental bands missing in some or all of the

four plants to vary between 76.9% for Tnt1-ol16 and 79% for Tnt1-ol13. Inter-plant variability was also observed, with some *N. tomentosiformis*-derived bands still present in one, two or three Th37 S4 plants (as illustrated in Fig. 1b), especially for the Tnt1-ol13 population (Table 2).

Characterization of new SSAP bands in Th37

We successfully extracted and cloned eight new Th37 bands from Tnt1 SSAP profiles; six were from Tnt1-ol13 and two from Tnt1-ol16 populations (NB bands; Tables 3,4). PCR primers were designed against flanking sequences and used in combination with Tnt1-ol13 or Tnt1-ol16 primers to validate the presence/absence of the insertion (validation PCRs; see Fig. 1a) in parental species and 18 plants of the Th37 S4 generation used previously for rDNA and cytogenetic studies (Skalicka *et al.*, 2003, 2005; Kovarik *et al.*, 2004; Lim *et al.*, 2004a), which includes the four plants used for SSAP analysis (Table 4).

For four new Tnt1-ol13 bands in Th37 (NB13-22, NB13-29b, NB13-33 and NB13-41), PCR amplification patterns from genomic DNA was consistent with the patterns of occurrence of new bands on Th37 SSAP profiles. No PCR products were derived from genomic DNA of the parental species or natural tobacco, confirming that the new SSAP bands are newly inserted Tnt1 copies in Th37. We observed different segregation patterns for each of the new insertions across the 18 Th37 S4 plants and all new insertions were restricted to Group 3 plants (Table 4). Confirmation of their transpositional nature was provided by sequence analysis of the four SSAP fragments. Tnt1 primers

Table 2 *Nicotiana tomentosiformis*-derived parental sequence-specific amplification polymorphism (SSAP) bands missing in Th37 S4 generation plants

	tom-sp ^a	Th37 S4 plants				
		G1 ^b	G2 ^b	G3 ^b		All ^c
		9	1	3	8	
		<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Tnt1-ol13						
Scored bands	19					
Bands missing in all four S4		9 (47.4)	9 (47.4)	9 (47.4)	9 (47.4)	9 (47.4)
Bands missing in some S4		1 (5.3)	2 (10.5)	3 (15.8)	4 (21.0)	6 (31.6)
Total missing bands		10 (52.7)	11 (57.9)	12 (63.2)	13 (68.4)	15 (79.0)
Tnt1-ol16						
Scored bands	26					
Bands missing in all four S4		16 (61.5)	16 (61.5)	16 (61.5)	16 (61.5)	16 (61.5)
Bands missing in some S4		2 (7.7)	2 (7.7)	3 (11.5)	3 (11.5)	4 (15.4)
Total missing bands		18 (69.2)	18 (69.2)	19 (73.0)	19 (73.0)	20 (76.9)

^atom-sp, insertions specific to *N. tomentosiformis* (i.e. not shared with *Nicotiana sylvestris*).

^bGx, groups of Th37 S4 plants as described in Skalicka *et al.* (2003). Numbers within each Gx group refer to each S4 plant number, as described in Skalicka *et al.* (2003).

^cBands scored across the four Th37 S4 plants.

Table 3 Summary of polymorphic sequence-specific amplification polymorphism (SSAP) bands characterized in Th37 S4 plants

Band name ^a	Tnt1 type ^b	Nature of polymorphism	Nature of insertion site
New SSAP bands in Th37			
NB13-22	Tnt1A1 (4 BII)	New insertion	In intron cytochrome P450-like gene
NB13-29b	Tnt1A1 (4 BII)	New insertion	In intron of alpha-N acetyl-glucosaminidase gene
NB13-33	Tnt1A1 (4 BII)	New insertion	In exon of lipase-like gene
NB13-41	Tnt1A1 (4 BII)	New insertion	In intron of amino acid decarboxylase-like gene
NB16-59	Tnt1A1 (4 BII)	New insertion (putative)	Unknown (2 bp of flanking sequence)
NB13-54/NB16-55 ^c	Tnt1A6 (2BII)	Indel 1 (= DTB-ol13-42)	No significant hit (149 bp)
NB16-20	Tnt1B2	Indel 2 (= DTB-ol16-23)	In intron of protein kinase (232 bp)
<i>Nicotiana tomentosiformis</i>-derived parental SSAP bands missing in Th37			
DTB13-42	Tnt1A6 (2 BII)	Indel 1 (= NB13-54/NB16-55)	Flanking sequence not identified
DTB13-52	Tnt1A6 (3 BII)	Loss of Tnt1/flanking junction	Flanking sequence not identified
DTB13-59	? (U3 truncated in 5')	Loss of Tnt1/flanking junction	Flanking sequence not identified, repetitive
DTB16-23	Tnt1B2	Indel 2 (= NB16-20)	In intron of protein kinase-like gene
DTB16-47	Tnt1C1	Loss of Tnt1/flanking junction	In gypsy-type retrotransposon polyprotein
DTB16-66	Tnt1C4	Loss of Tnt1/flanking junction	Flanking sequence not identified, repetitive

^aNB, new band; DTB, *Nicotiana tomentosiformis* bands missing in Th37 SSAP profiles; 13 or 16 refers to the two different Tnt1 populations assayed and the last number refers to the band identification on SSAP gels.

^bU3 type classified according to Vernhettes *et al.* (1998); in parentheses: number of regulatory BII repeats found in copies of the Tnt1A sub-family.

^cSame variant insertion site obtained from Tnt1-ol13 (NB13-54) and Tnt1-ol16 (NB16-55) SSAP profiles.

used in this study were designed against the LTR U5 region (Fig. 1a), allowing characterization of the upstream U3 region that contains promoter and regulatory sequences involved in the expression of active copies (Casacuberta &

Grandbastien, 1993; Grandbastien *et al.*, 2005). The U3 region of *Nicotiana* Tnt1 elements is highly variable (Vernhettes *et al.*, 1998) and Tnt1 elements that are transcriptionally and transpositionally active in response to

Table 4 Segregation of polymorphic Tnt1 insertion sites in the Th37 S4 progeny: analysis by presence/absence PCR^a

				Th37 S4 plants																	
				G1 ^d			G2 ^d	G3 ^d													
Band name ^b	tom ^c	syl ^c	tab ^c	7	9 ^e	16	1 ^e	2	3 ^e	4	5	6	8 ^e	10	11	12	13	14	15	17	19
New insertions in Th37																					
NB13-22	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	1	1	1	1	0
NB13-29b	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	1	1	1	1	1	1
NB13-33	0	0	0	0	0	0	0	1	1	1	1	1	0	0	1	1	0	1	1	0	0
NB13-41	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1	1	0	1	1
Indel(s) at insertion sites in Th37																					
Indel 1	1	0	1	1–	1–	1–	1–	1–	1–	1–	1–	1–	1–	1–	1–	1–	1–	1–	1–	1–	1–
Indel 2	1	0	1	1	1	1	1	1+	1+	1+	1+	1+	1+	1+	1+	1+	1+	1+	1+	1+	1+
<i>Nicotiana tomentosiformis</i> -derived parental insertion sites missing in Th37																					
DTB13-52	1	0	1	0	0	0	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1
DTB13-59	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DTB16-47	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DTB16-66	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

^a1, presence of PCR product at expected size; 0, absence of PCR product; 1–, PCR product of smaller size than expected from the control *Nicotiana tomentosiformis*; 1+, PCR product of larger size than expected from the control *N. tomentosiformis*.

^bNB, new band; DTB, *N. tomentosiformis* bands missing in Th37 sequence-specific amplification polymorphism (SSAP) profiles; 13 or 16 refers to the two different Tnt1 populations assayed and the last number refers to the band identification on SSAP gels.

^ctom, *N. tomentosiformis*; syl, *Nicotiana sylvestris*; tab, *Nicotiana tabacum*.

^dGx, groups of Th37 S4 segregants as described in Skalicka *et al.* (2003). Numbers within each Gx group refer to each S4 plant number, as described in Skalicka *et al.* (2003).

^eThe four Th37 S4 plants used for the original SSAP screening.

microbial factors belong to a conserved subclass of elements of the Tnt1A subfamily that are characterized by the conservation of three or four tandemly repeated regulatory BII motifs (Casacuberta *et al.*, 1995; Melayah *et al.*, 2001). The four newly transposed copies contained identical LTR regions with four BII repeats (Note S1), and were very similar (99% identity) to several Tnt1A1 copies newly transposed in response to fungal extracts (Melayah *et al.*, 2001; Fig. 2).

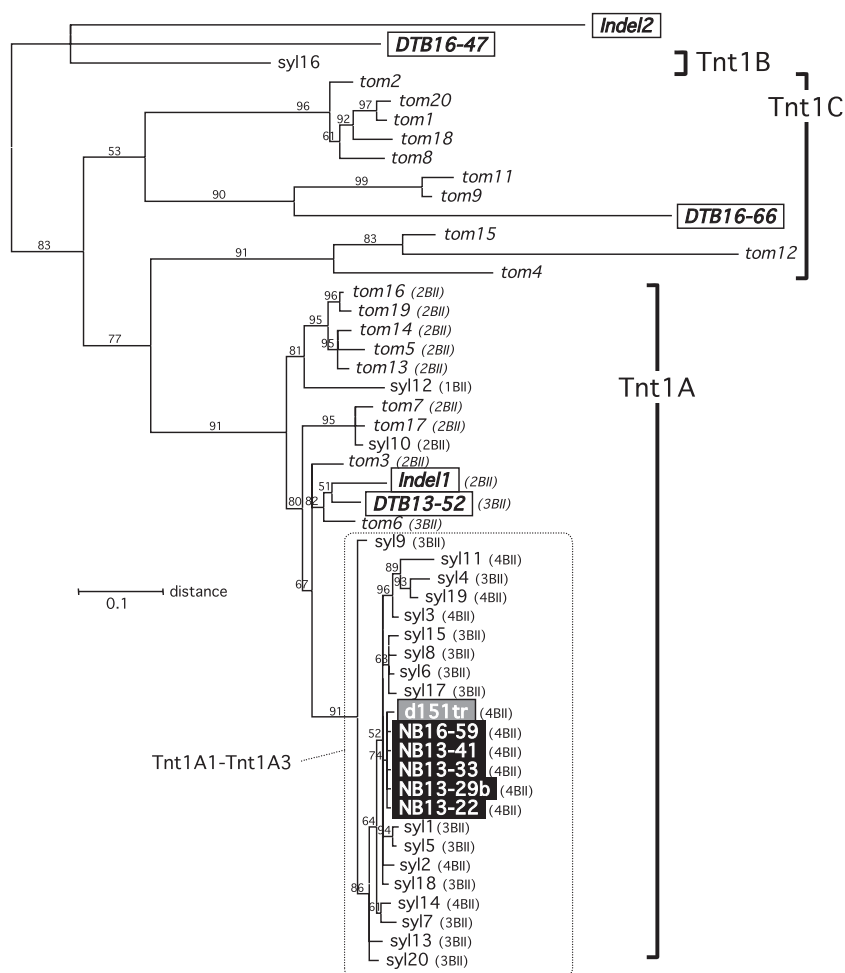
A fifth new SSAP band (NB16-59) had a Tnt1A1 U3 LTR sequence that was identical to those of the four other copies (Fig. 2; Note S1), indicating a newly transposed copy. Unfortunately, however, only 2 bp of flanking sequence was recovered, because of the close proximity of the *EcoRI* site to the LTR extremity, and it was not possible to perform PCR tests to study its segregation in the S4 generation (Table 3; Note S1).

For the three remaining new SSAP bands (NB13-54, NB16-20 and NB16-55), PCRs revealed that the insertion was present in *N. tomentosiformis* ac. TW142, indicating that they were not newly transposed copies. The NB16-55 and NB13-54 insertion sites were identical, indicative of

recognition of the same copy by both Tnt1-ol13 and Tnt1-ol16 primers. The presence of the new SSAP bands in Th37 S4 plants was correlated with changes in the size of the PCR products, decreasing for NB13-54/NB16-55 and increasing for NB16-20 (Tables 3 and 4). Sequence analysis of SSAP clones and PCR products confirmed that the new SSAP bands were generated by sequence deletions or insertions, respectively, within the Tnt1/flanking junction of pre-existing *N. tomentosiformis*-derived copies. These size modifications led to their artefactual appearance as 'new' SSAP bands.

For the NB16-55/NB13-54 insertion site, the variant SSAP band was present in all four Th37 S4 plants used for SSAP profiles, and was generated by deletions of 31- and 60-bp sequences from the LTR of a Tnt1A6 copy (Fig. 3a; variant Indel 1). The first deletion removed one of the two BII-related repeats present in the *N. tomentosiformis* ac. TW142 parental sequence, and the second larger deletion removed a large sequence around the TATA box. Similarities were found between nucleotide motifs immediately upstream of each deleted sequence and at their terminal 3' end, suggestive of illegitimate recombination events leading

Fig. 2 Phylogenetic tree obtained from the alignments of partial long terminal repeat (LTR) sequences of *Nicotiana sylvestris* (sylx) and *Nicotiana tomentosiformis* (tomx) genomic copies, and of Tnt1 insertions that are polymorphic in the Th37 synthetic hybrid. *Nicotiana sylvestris* and *N. tomentosiformis* genomic copies, as well as subfamilies and groups, were described in Vernhettes *et al.* (1998). Newly transposed copies in Th37 are indicated in white on a black background. Clone d151tr, against a gray background, is a Tnt1A1 copy representative of 30% of the newly transposed copies in response to microbial stress (Melayah *et al.*, 2001). *Nicotiana tomentosiformis* paternal insertions that were modified in Th37 are boxed. The number of regulatory BII boxes in Tnt1A elements is indicated in parentheses. Bootstrap values are shown above branches. Notes: the DTB13-59 insertion was not included because of the 5' truncation of its U3 region, and the Indel 1 and Indel 2 sequences are the parental versions of the insertion sites, which are devoid of the indels.



(a) INDEL 1

```

tobacco      ataggttatataaaaatagttattcttctatcctatcacacTGATGGTTTCTATCTCATTAATAGCCAAAGTAATAGGCATGTGCTTAAGGAAATTTCTTCTGTTGATAGTCAACT
tomento      ataggttatataaaaatagttattcttctatcctatcacacTGATGGTTTCTATCTCATTAATAGCCAAAGTAATAGGCATGTGCTTAAGGAAATTTCTTCTGTTGATAGTCAACT
Th37-1       ataggttatataaaaatagttattcttctatcctatcacacTGATAGTTTCTATCTCATTAATAGCCAAAGTAATAGGCATGTGCTTAAGGAAATTTCTTCTGTTGATAGTCAACT
                                     <-----BI-box----->      <-----BII box----->
                                     <----->      <----->
tobacco      TTGTTGAATTTCTTTGGATTGATAGCCAGCTTTGTTGAATTTCTCAACATTGAAGAAAAAAGGAAAAAGTGTGTGCAAAATTGTCAAATATAGTAGGCTTTAGAGAGTGAGACTTCGGC
tomento      TTGTTGAATTTCTTTGGATTGATAGCTAAGCTTTGTTGAATTTCTCAACATTGAAGAAAAAAGGAAAAAGTGTGTGCAAAATTGTCAAATATAGTAGGCTTTAGAGAGTGAGACTTCGGC
Th37-1       TTGTTGAATTTCTTTGGATTGATAGCTAAGCTTTGTTGAATTTCTCAACATTGAAGAAAAAAGGAAAAAGTGTGTGCAAAATTGTCAAATATAGTAGGCTTTAGAGAGTGAGACTTCGGC
                                     *****
tobacco      TATAAAGGAGAGCTTCAACTCTCATTTCCACATACCAACAAAGAGAGAAAGAAAGAGTGT
tomento      TATAAAGGAGAGCTTCAACTCTCATTTCCACATACCAACAAAGAGAGAAAGAAAGAGTGT
Th37-1       TATAAAGGAGAGCTTCAACTCTCATTTCCACATACCAACAAAGAGAGAAAGAAAGAGTGT
                                     *****

```

(b) INDEL 2

```

tobacco      aagatagtttaatttccactcaagttttctgttaaagaattatgcaatcctgttgattctactcaacattcttgacatagagggattaggtcagaactacgatactcatattttt
tomento      aagatagtttaatttccactcaagttttctgttaaagaattatgcaatcctgttgattctactcaacattcttgacatagagggattaggtcagaactacgatactcatattttt
Th37-1       aagatagtttaatttccactcaagttttctgttaaagaattatgcaatcctgttgattctactcaacattcttgacatagagggattaggtcagaactacgatactcatattttt
Th37-3       aagatagtttaatttccactcaagttttctgttaaagaattatgcaatcctgttgattctactcaacattcttgacatagagggattaggtcagaactacgatactcatattttt

tobacco      tgcctcttaagttgtatgcactaaagttttaataatagttatacaatgagatcacttagttttatgataatcacta-----actaacaTGTTGGGTC
tomento      tgcctcttaagttgtatgcactaaagttttaataatagttatacaatgagatcacttagttttatgataatcacta-----actaacaTGTTGGGTC
Th37-1       tgcctcttaagttgtatgcactaaagttttaataatagttatacaatgagatcacttagttttatgataatcacta-----actaacaTGTTGGGTC
Th37-3       tgcctcttaagttgtatgcactaaagttttaataatagttatacaatgagatcacttagttttatgataatcacta-----actaacaTGTTGGGTC
                                     *****

tobacco      CATCCCATAAATTAAGGAAGAAATGATTGCGCAGCCATCTTTCTTTCTTATTGTCAAATAAGGTAGGCTCGAGAAAGAAATGTCAAACCTCTTAGGCGGAAGAAAGAAAAATTTAC
tomento      CATCCCATAAATTAAGGAAGAAATGATTAGCAGCCATCTTTCTTTCTTATTGTCAAATAAGGTAGGCTCGAGAAAGAAATGTCAAACCTCTTAGGCGGAAGAAAGAAAAATTTGC
Th37-1       CATCCCATAAATTAAGGAAGAAATGATTAGCAGCCATCTTTCTTTCTTATTGTCAAATAAGGTAGGCTCGAGAAAGAAATGTCAAACCTCTTAGGCGGAAGAAAGAAAAATTTGC
Th37-3       CATCCCATAAATTAAGGAAGAAATGATTGCGCAGCCATCTTTCTTTCTTATTGTCAAATAAGGTAGGCTCGAGAAAGAAATGTCAAACCTCTTAGGCGGAAGAAAGAAAAATTTGC

tobacco      ATTGCAAAATTTCAACCAATCCTTCATTCTGATTGGTTGAAGAGGTAGGCCATTTCTCCATATAAAGGAGAGCTTCGGCTTCTTCATTTTATCACACAGAAAGAGAAGCTACACAAATAG
tomento      ATTGCAAAATTTCAACCAATCCTTCATTCTGATTGGTTGAAGAGGTAGGCCATTTCTCCATATAAAGGAGAGCTTCGGCTTCTTCATTTTATCACACAGAAAGAGAAGCTACACAAATAG
Th37-1       ATTGCAAAATTTCAACCAATCCTTCATTCTGATTGGTTGAAGAGGTAGGCCATTTCTCCATATAAAGGAGAGCTTCGGCTTCTTCATTTTATCACACAGAAAGAGAAGCTACACAAATAG
Th37-3       ATTGCAAAATTTCAACCAATCCTTCATTCTGATTGGTTGAAGAGGTAGGCCATTTCTCCATATAAAGGAGAGCTTCGGCTTCTTCATTTTATCACACAGAAAGAGAAGCTACACAAATAG

tobacco      TTGCAGAGAGTATTCTCAGGCATTGTGAGAAATAGAC
tomento      TTGCAGAGAGTATTCTCAGGCATTGTGAGAAATAGAC
Th37-1       TTGCAGAGAGTATTCTCAGGCATTGTGAGAAATAGAC
Th37-3       TTGCAGAGAGTATTCTCAGGCATTGTGAGAAATAGAC

```

Fig. 3 Alignment of sequences of the two Tnt1/flanking junctions modified by indels in Th37. Sequences are from PCR products amplified using validation primers designed against flanking sequences combined with Tnt1-ol16 (Indel 1) or Tnt1-ol13 (Indel 2) primers. The flanking sequence is in lowercase and the 5' long terminal repeat (LTR) sequence is in uppercase. Tobacco = *Nicotiana tabacum* cv. Xanthi XHFD8; tomento = paternal *Nicotiana tomentosiformis* ac. TW142. Identical nucleotides are indicated by dark gray backgrounds, and partial similarities by light gray backgrounds. Dashes are missing nucleotides. Stars indicate duplicated motifs adjacent to deleted sequences in Indel 1 or direct repeats flanking the insertion in Indel 2. The TATA box is underlined, and the position of regulatory BI and BII boxes found in Tnt1A elements (Vernhettes *et al.*, 1998) is indicated for Indel 1. Note: for Indel 1, Th37-3, Th37-8 and Th37-9 sequences (not shown) are identical to the Th37-1 sequence, and for Indel 2, the Th37-8 sequence (not shown) is identical to the Th37-3 sequence, and the Th37-9 sequence (not shown) is identical to the Th37-1 sequence.

to both deletions. A number of point mutations were observed in the Th37 variant (96.6% identity with *N. tomentosiformis* ac. TW142, not taking indels into account). PCR confirmed that the Indel 1 variant was present in all Th37 S4 plants tested (Table 4).

For the NB16-20 insertion, a new SSAP band specific to Th37-3 and Th37-8, but not detected in Th37-1 and Th37-9 SSAP profiles, was generated by a short 24-bp insertion in the flanking DNA in close proximity to the LTR of a Tnt1B2 copy (Fig. 3b, variant Indel 2). Interestingly, the insertion was bound by a 4-bp direct duplication, suggesting a possible insertion by staggered nick mechanisms, such as those used for insertion of transposable elements. A number of point mutations were also observed in the Th37 variant (96.4% identity with the *N. tomentosiformis* ac. TW142, not taking the indel into account). In contrast to the Indel 1 variant, the Indel 2 variant was present only in Group 3 plants (Table 4). Sequence analysis of PCR products obtained from Th37-1 and Th37-9 plants, in which the parental SSAP band was detected, confirmed that the sequences were similar to that from *N. tomentosiformis* ac. TW142 (Fig. 3b).

These data show that four, and very likely five, out of seven Tnt1 insertions originally identified as new SSAP bands in Th37 plants are *bona fide* newly inserted copies, resulting from transpositional activation of Tnt1A subsequence to *de novo* allopolyploidy.

Characterization of *N. tomentosiformis*-derived SSAP bands missing in Th37

We extracted and sequenced from *N. tomentosiformis* ac. TW142 six SSAP bands that were missing in at least some of the Th37 profiles (DTB bands; Tables 3,4). Two of the six SSAP clones (DTB13-42 and DTB16-23) were the paternal Tnt1 insertions that led to Indel 1 and Indel 2 variants in Th37, for which modification of their size led to their artefactual appearance as 'new' SSAP bands, and in turn to the disappearance of the parental *N. tomentosiformis*-derived SSAP band. For the four other *N. tomentosiformis* bands missing in Th37 (DTB13-52, DTB13-59, DTB16-47 and DTB16-66), PCR primers designed against flanking sequences were used in combination with Tnt1-

ol13 or Tnt1-ol16 primers to characterize the presence/absence of the insertion site across 18 Th37 S4 plants (Table 4).

For DTB13-59 and DTB16-47, no PCR product was detected in any of the 18 Th37 S4 plants, while a PCR product of the expected size was detected in *N. tomentosiformis* ac. TW142 (Table 4). For DTB13-52 and DTB16-66, PCR products of the expected size were detected in some, but not all Th37 S4 plants, in agreement with the presence/absence on SSAP profiles of the paternally derived SSAP band for Th37-1, Th37-3, Th37-8 and Th37-9 (Table 4), while a PCR product of the expected size was detected in *N. tomentosiformis* ac. TW142. Different segregation patterns were observed for the missing insertion sites in the three groups of Th37 S4 plants (Table 4). These data indicate that the disappearance in Th37 of the four paternal SSAP bands, DTB13-52, DTB13-59, DTB16-47 and DTB16-66, is a result of the disappearance of the Tnt1/flanking junctions. In order to evaluate the extent of the deletion, several flanking primers were further tested in combination with a primer designed against the internal coding region of Tnt1 (GAG-Rev; Fig. 1b). PCR results were similar to those obtained with Tnt1-ol13 or Tnt1-ol16 primers (data not shown).

In summary, our data show that the six *N. tomentosiformis*-derived SSAP bands that are missing in some or all Th37 S4 plants correspond either to variants containing indels in or near the Tnt1 insertion (two cases) or to complete losses of the Tnt1/flanking junction, which include all of the LTR and some Tnt1 internal coding regions (four cases).

Nature and distribution of Tnt1 copies polymorphic in Th37

In order to determine the parental origin of Tnt1 copies newly inserted in Th37, we compared their partial LTR sequences with the corresponding sequences of randomly selected Tnt1 genomic copies previously characterized from both parental diploid species (Vernhettes *et al.*, 1998). Tnt1 U3 regions are known to be variable and to define different subfamilies (Tnt1A, Tnt1B and Tnt1C), whose distribution differs among *Nicotiana* host species (Vernhettes *et al.*, 1998). *Nicotiana sylvestris* contains mostly Tnt1A elements with a majority of closely related members of the Tnt1A1-Tnt1A3 groups, as defined in Vernhettes *et al.* (1998). *Nicotiana tomentosiformis* contains less conserved Tnt1A elements and more highly diverged Tnt1C elements compared with *N. sylvestris* (illustrated in Fig. 2). Our results show that the five newly transposed copies in Th37 are most similar to *N. sylvestris* copies (Fig. 2), suggesting that they have been amplified from one or a few closely related mother copies located on the Th37 maternal subgenome. We also BLASTed LTR sequences of newly inserted elements against

low coverage (*c.* 0.5% genome) 454 titanium high-throughput sequencing reads of *N. sylvestris* and *N. tomentosiformis* (hits with > 95% homology). We obtained hits only with the *N. sylvestris* database, while none came up with the *N. tomentosiformis* database (A. Kovarik, A.R. Leitch, M.-A. Grandbastien, unpublished data), confirming the maternal origin of the Tnt1A copies activated in Th37.

However, *N. tomentosiformis*-derived Tnt1 copies that have been modified in Th37, either by indel formation or deletion of the Tnt1/flanking junction, have much diversity within the Tnt1A and Tnt1C subfamilies. The most diverged *N. tomentosiformis* copies (DTB16-47, DTB16-66 and variant Indel 2) were all of the Tnt1C subfamily and were amplified using the Tnt1-ol16 primer, while those obtained from Tnt1-ol13 profiles (variant Indel 1 and DTB13-52) belong to the less diverged Tnt1A subfamily. Although only a few sequences were examined, preventing statistical comparisons, this observation is consistent with our previous observation that Tnt1-ol16 populations are older than Tnt1-ol13 populations (Petit *et al.*, 2007).

With the exception of the putative new insertion NB16-59, for which the genomic target sequence is too short to allow identification, newly transposed copies in Th37 had all inserted into genic regions, within or close to coding sequences (Table 3, Table S1). In addition, the fact that the sequences flanking new copies were not found among low coverage (*c.* 0.5% genome) 454 titanium high-throughput sequencing reads of *N. sylvestris* and *N. tomentosiformis* (A. Kovarik, A.R. Leitch, M.-A. Grandbastien, unpublished data) confirms that Tnt1A did not insert into the highly repetitive, presumably heterochromatic, fraction of the genome. By contrast, none of the *N. tomentosiformis*-derived bands missing in Th37 was associated with recognizable genic regions, with the exception of the insertion that led to the Indel 2 variant. The *N. tomentosiformis* counterpart of DTB16-47 was found inserted into a *gypsy*-type retrotransposon, and the nature of the flanking sequences could not be determined in other cases, but was often repetitive (Table 3, Table S1).

Discussion

The Tnt1A retrotransposon is mobilized in response to the genome shock of allopolyploidy

Tobacco Tnt1A retrotransposons are activated by biotic and abiotic stresses, and their transcriptional activation by microbial factors is correlated with transposition (reviewed in Grandbastien *et al.*, 2005). Stress activation of retrotransposons may induce genomic changes and arguably be an adaptive response to stressful conditions (Grandbastien, 1998; Kalendar *et al.*, 2000). We show here that Tnt1A can amplify early in response to allopolyploidy, a situation that has been compared to a genome shock (McClintock,

1984). New SSAP bands are detected in the S4 generation of the Th37 synthetic allopolyploid, and the majority of them are newly transposed copies. The SSAP profiles performed here used only one restriction enzyme and reveal only a fraction of all existing insertion sites. Consequently, only a fraction of potential changes were revealed. We detected no new insertions in the S5 generation (data not shown), suggesting that Tnt1A was no longer active in the S4 germline.

Newly transposed Tnt1A copies in Th37 are all identical to each other across the partial LTR sequence available, indicating that they were amplified from one or a few closely related mother copies. Newly transposed copies inserted in low copy genic regions, and the mother copies activated in Th37 almost certainly originate from the maternal *N. sylvestris* subgenome. Most Tnt1 *N. tomentosiformis* may no longer be active, as genomic copies are highly divergent (Vernhettes *et al.*, 1998; see Fig. 2). A differential behavior of parental retrotransposons was also observed in Arabidopsis allopolyploids, in which only the more conserved paternal set of elements is expressed (Josefsson *et al.*, 2006).

Interspecific hybridization has been proposed to induce bursts of transposition (McClintock, 1984); however, careful bibliographic surveys reveal that activation of transposable elements appears restricted to some types of element (for a review see Parisod *et al.*, 2010). Tnt1A is one of the rare plant retrotransposons whose amplification has been directly demonstrated, and its mobilization in Th37 is therefore not surprising. Activation levels vary among different Tnt1 populations, with proportions of new SSAP bands 3.6-fold higher for Tnt1-ol13 elements compared with Tnt1-ol16 elements. A similar trend was observed in natural tobacco, with a nearly twofold difference in new SSAP band rates (Petit *et al.*, 2007). The consistent increase in new SSAP bands observed for Tnt1-ol13 elements in natural and synthetic tobacco argues that mobilization may be specific to these younger retrotransposon populations. Newly transposed Tnt1A copies in Th37 belong to the Tnt1A1-A3 subclasses, and their regulatory U3 regions show high sequence identity to copies activated in response to microbial factors. The activation of Tnt1A1 elements by microbial stress is mediated by four repeated 31-bp BII sequences containing motifs similar to regulatory motifs of plant defense response genes, and involved in the binding of stress-specific proteic factors (Casacuberta & Grandbastien, 1993; Casacuberta *et al.*, 1995; Vernhettes *et al.*, 1997). The high conservation of BII sequences and numbers in Tnt1A copies activated by both microbial factors and allopolyploidy suggests that similar stress pathways and defense factors may be involved.

Chromosome doubling of the Th37 F₁ hybrid was performed by *in vitro* tissue culture (Burk, 1973), a process that can activate plant transposons, notably via demethylation (Kubis *et al.*, 2003). However, Tnt1A expression is

poorly activated in cell cultures (Béguiristain *et al.*, 2001). Tnt1A amplification was detected in 25% of plants regenerated from protoplasts submitted to cell-wall degrading fungal extracts, but in only 2.7% of control plants regenerated from tissue culture (Melayah *et al.*, 2001). Similarly, very few new SSAP bands were detected in F₁ plants produced from other *N. sylvestris* × *N. tomentosiformis* crosses, or in corresponding S0 plants obtained through *in vitro* tissue culture doubling (M. Petit, C. Mhiri & M-A. Grandbastien, unpublished). Tnt1A transpositions detected in the Th37 S4 generation are therefore unlikely to have been generated by tissue culture, and are a direct response to the genome shock of allopolyploidy, possibly after the meiosis step in the Th37 S0 germline.

Major restructuring occurs at Tnt1 insertion sites in the *N. tomentosiformis* paternal subgenome

A high proportion of *N. tomentosiformis*-derived SSAP bands were not observed in the synthetic Th37 tobacco. As retrotransposons do not excise, the loss of SSAP bands is caused by other processes. Changes that occurred at *N. tomentosiformis*-derived Tnt1 insertion sites included small indels as well as larger deletions of the Tnt1/flanking junctions, extending up to internal Tnt1 coding sequences. These data indicate rapid molecular restructuring in, or around, Tnt1 insertion sites in response to allopolyploidy, processes that have specifically destabilized the paternal *N. tomentosiformis*-derived subgenome. This is consistent with other data that showed that the *N. tomentosiformis*-derived subgenome was most unstable in Th37, with reductions in copy numbers of geminivirus-related DNA, pararetrovirus sequences, and *Nicotiana tomentosiformis* repetitive sequences (NTRS) (Skalicka *et al.*, 2003, 2005; Kovarik *et al.*, 2004; Lim *et al.*, 2004a). In addition, substantial modifications were observed in the *N. tomentosiformis*-type rDNA sequences (Skalicka *et al.*, 2003). Rearrangements in repetitive DNA sequences in response to interspecific hybridization have been frequently observed (reviewed in Parisod *et al.*, 2010; Koukalova *et al.*, 2010). The reasons for such instabilities remain unknown. Rearrangements at Tnt1 insertion sites were not observed in natural tobacco plants regenerated from tissue culture (Melayah *et al.*, 2001). This indicates that, like Tnt1A mobilization, rearrangements observed in Th37 S4 plants were not stimulated by the *in vitro* genome doubling step and were likely caused by the allopolyploidy process. They may reflect nontargeted DNA lesions and genome rearrangements resulting from failed chromosome pairing between homeologous chromosomes (Le Comber *et al.*, 2010), which may influence repetitive DNA, the major genome component. However, no restructuring of parental Tnt1 insertions (M. Petit *et al.*, unpublished) and rDNA loci (Lim *et al.*, 2006) were detected in the F₁ and S0 generations of other *N. sylves-*

tris × *N. tomentosiformis* crosses, suggesting that, as for Tnt1A activation, meiosis may be required to trigger genomic restructuring.

Genetic changes predominantly influencing one parental genome have also been observed in other allopolyploids (Song *et al.*, 1995; Shaked *et al.*, 2001). It has been proposed that nuclear–cytoplasmic interactions may cause greater instability of the paternal genome. Unfortunately, no reciprocal synthetic *N. tabacum* is available to test this hypothesis (Leitch *et al.*, 2005). Alternatively, the preferential targeting of paternally derived insertions may be attributable to structural specificities of the *N. tomentosiformis* Tnt1 populations. Tnt1 elements are more clustered on *N. tomentosiformis* chromosomes, compared with *N. sylvestris* (Melayah *et al.*, 2001). In addition, nearly all targeted Tnt1 copies are associated with repetitive sequences, which may be preferential substrates for recombination-based deletion processes. Other sequences eliminated in Th37, for example NTRS satellites or geminivirus-related DNA (GRD5) repeats, are also clusters composed of long tandem arrays (Skalicka *et al.*, 2005).

The near-complete transmission of *N. sylvestris*-derived SSAP bands to Th37 is in contrast to observations of SSAP band losses observed in the S genome of natural tobacco (Petit *et al.*, 2007). Similar discrepancies between synthetic and natural tobacco have been observed for other repetitive genomic components. For example, *N. sylvestris*-derived 18–26S rDNA intergenic sequences are maintained in Th37 and are mostly lost in natural tobacco (Skalicka *et al.*, 2003; Kovarik *et al.*, 2004). There are no intergenomic gene conversions of rDNA units and fewer intergenomic translocations in Th37 compared with natural tobacco (Skalicka *et al.*, 2003; Lim *et al.*, 2004a). Ongoing divergence of the T genome in natural tobacco has also resulted in substantial reductions in the copy numbers of *N. tomentosiformis*-derived endogenous pararetrovirus sequences (*NtoEPRV*), NTRS repeats and A1/A2 satellite repeats, all of which are much less reduced in Th37 plants (Gregor *et al.*, 2004; Matzke *et al.*, 2004; Skalicka *et al.*, 2005). More puzzling, the *N. tomentosiformis* geminivirus-related GRD5 repeats, which are deleted in 60% of the Th37 plants, are maintained in tobacco (Skalicka *et al.*, 2005). Discrepancies between the evolutionary trajectories of natural tobacco and Th37 might suggest a stochastic factor influencing genomic changes. It is also likely that further genomic changes to both parental genomes have occurred during the evolution of tobacco after its formation, and the low insertion polymorphism levels observed among tobacco landraces (Petit *et al.*, 2007) probably reflect a narrow genetic base. The extent of major genomic changes leading to long-term diploidization is proportional to the age of *Nicotiana* allopolyploids (Lim *et al.*, 2007), and 1-million-yr-old allopolyploid species of section *Polydicliae* show DNA changes that are not detectable in resynthesized allopolyploids at only the S5

generation (Anssour *et al.*, 2009). Although no changes in Tnt1 populations in the Th37 S5 generation were observed compared with the S4 generation (data not shown), it is possible that this merely reflects a reduced rate of genetic changes after initial bursts of genetic changes in the first generations of allopolyploidy.

Acknowledgements

We wish to dedicate this work to our dear friend Yoong K. Lim (1956–2008). We are greatly indebted to him for his participation in this work, and his continuous support, help and advice, and we truly miss him. This work was funded by a BTH01453 Association pour la Recherche sur les Nicotianées (ANR) research contract, the Agence Nationale de la Recherche (ANR) Biodiversity 2005 program (ANR-05-BDIV-015, ‘Effect of polyploidy on plant genome biodiversity and evolution’), NERC and the F-UK (Alliance, 08782YM) and F-CZ (Barrande, 09150SH and MEB-020823) bi-national Programmes d’Actions Intégrées (PAI).

References

- Adams KL, Wendel JF. 2005. Novel patterns of gene expression in polyploid plants. *Trends in Genetics* 21: 539–543.
- Alix K, Heslop-Harrison JS. 2004. The diversity of retroelements in diploid and allotetraploid Brassica species. *Plant Molecular Biology* 54: 895–909.
- Alix K, Joets J, Ryder CR, Moore J, Barker GC, Bailey JP, King GJ, Heslop-Harrison JS. 2008. The CACTA transposon *Bot1* played a major role in *Brassica* genome divergence and gene proliferation. *Plant Journal* 56: 1030–1044.
- Anssour S, Krügel T, Sharbel TF, Saluz HP, Bonaventure G, Baldwin IT. 2009. Phenotypic, genetic and genomic consequences of natural and synthetic polyploidization of *Nicotiana attenuata* and *Nicotiana obtusifolia*. *Annals of Botany (London)* 103: 1207–1217.
- Beaulieu J, Jean M, Belzile F. 2009. The allotetraploid *Arabidopsis thaliana*–*Arabidopsis lyrata* subsp. *petraea* as an alternative model system for the study of polyploidy in plants. *Molecular Genetics and Genomics* 281: 421–435.
- Béguiristain T, Grandbastien M-A, Puigdomenech P, Casacuberta P. 2001. Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. *Plant Physiology* 127: 212–221.
- Bourgin J-P, Chupeau Y, Missonier C. 1973. Plant regeneration from mesophyll protoplasts of several *Nicotiana* species. *Physiologia Plantarum* 45: 288–292.
- Burk LG. 1973. Partial self-fertility in a theoretical amphiploid progenitor of *N. tabacum*. *Journal of Heredity* 64: 348–350.
- Casacuberta JM, Grandbastien M-A. 1993. Characterisation of LTR sequences involved in the protoplast specific expression of the tobacco Tnt1 retrotransposon. *Nucleic Acids Research* 21: 2087–2093.
- Casacuberta JM, Vernhettes S, Grandbastien M-A. 1995. Sequence variability within the tobacco retrotransposon Tnt1 population. *EMBO Journal* 14: 2670–2678.
- Charles M, Belcram H, Just J, Huneau C, Viollet A, Couloux A, Segurens B, Carter M, Huteau V, Coriton O *et al.* 2008. Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. *Genetics* 180: 1071–1086.

- Chen ZJ, Ni Z. 2006. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *Bioessays* 28: 240–252.
- Clamp M, Cuff J, Searle SM, Barton GJ. 2004. The Jalview Java alignment editor. *Bioinformatics* 20: 426–427.
- Clarkson JJ, Lim KY, Kovarik A, Chase MW, Knapp S, Leitch AR. 2005. Long-term genome diploidization in allopolyploid *Nicotiana* section *Repandae* (Solanaceae). *New Phytologist* 168: 241–252.
- Comai L, Madlung A, Josefson C, Tyagi A. 2003. Do the different parental “heteromes” cause genomic shock in newly formed allopolyploids? *Philosophical Transactions of the Royal Society London B: Biological Sciences* 358: 1149–1155.
- Feldman M, Levy AA. 2009. Genome evolution in allopolyploid wheat - a revolutionary reprogramming followed by gradual changes. *Journal of Genetics and Genomics* 36: 511–518.
- Grandbastien M-A. 1998. Activation of plant retrotransposons under stress conditions. *Trends in Plant Sciences* 3: 181–187.
- Grandbastien M-A, Spielmann A, Caboche M. 1989. Tnt1, a mobile retroviral-like transposable element of tobacco isolated via plant cell genetics. *Nature* 337: 376–380.
- Grandbastien M-A, Audeon C, Bonnivard E, Casacuberta JM, Chalhoub B, Costa APP, Le QH, Melayah D, Petit M, Poncet C *et al.* 2005. Stress activation and genomic impact of Tnt1 retrotransposons in *Solanaceae*. *Cytogenetic and Genome Research, Special Issue on “Retrotransposable Elements and Genome Evolution”* 110: 229–241.
- Gregor W, Mette MF, Staginnus C, Matzke MA, Matzke AJM. 2004. A distinct endogenous pararetrovirus family in *Nicotiana tomentosiformis*, a diploid progenitor of polyploid tobacco. *Plant Physiology* 134: 1191–1199.
- Hanson RE, Zhao XP, Islam-Faridi MN, Paterson AH, Zwick MS, Crane CF, McKnight TD, Stelly DM, Price HJ. 1998. Evolution of interspersed repetitive elements in *Gossypium* (Malvaceae). *American Journal of Botany* 85: 1364–1368.
- Josefsson C, Dilkes B, Comai L. 2006. Parent-dependent loss of gene silencing during interspecies hybridization. *Current Biology* 16: 1322–1328.
- Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman A. 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proceedings of the National Academy of Sciences, USA* 97: 6603–6607.
- Kashkush K, Feldman M, Levy AA. 2002. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* 160: 1651–1659.
- Kashkush K, Feldman M, Levy AA. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics* 33: 102–106.
- Koukalova B, Moraes AP, Byfield SR, Matyasek R, Leitch AR, Kovarik A. 2010. Fall and rise of satellite repeats in allopolyploids of *Nicotiana* over c. 5 million years. *New Phytologist* 186: 148–160.
- Kovarik A, Koukalova B, Lim KY, Matyasek R, Lichtenstein CP, Leitch AR, Bezdek M. 2000. Comparative analysis of DNA methylation in tobacco heterochromatic sequences. *Chromosome Research* 8: 527–541.
- Kovarik A, Matyasek R, Lim KY, Skalicka K, Koukalova B, Knapp S, Chase M, Leitch AR. 2004. Concerted evolution of 18–5.8–26S rDNA repeats in *Nicotiana* allotetraploids. *Biological Journal of the Linnean Society* 82: 615–663.
- Kubis SE, Castilho AM, Vershinin AV, Heslop-Harrison JS. 2003. Retroelements, transposons and methylation status in the genome of oil palm (*Elaeis guineensis*) and the relationship to somaclonal variation. *Plant Molecular Biology* 52: 69–79.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al.* 2007. ClustalW and ClustalX version 2. *Bioinformatics* 23: 2947–2948.
- Le Comber SC, Ainouche ML, Kovarik A, Leitch AR. 2010. Making a functional diploid: from polysomic to disomic inheritance. *New Phytologist* 186: 113–122.
- Leitch AR, Lim KY, Skalicka K, Kovarik A. 2005. Nuclear cytoplasmic interaction hypothesis and the role of translocations in *Nicotiana* allopolyploids. In: Cigna AADM, ed. *NATO Advanced Research Workshop on Impact of radiation risk estimates in normal and emergency situations*. Yerevan, Armenia: Springer, 319–326.
- Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR. 2008. The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Annals of Botany* 101: 805–814.
- Lim KY, Matyasek M, Kovarik A, Leitch AR. 2004a. Genome evolution in allotetraploid *Nicotiana*. *Biological Journal of the Linnean Society* 82: 599–606.
- Lim KY, Skalicka K, Koukalova B, Volkov R, Matyasek R, Hemleben V, Leitch AR, Kovarik A. 2004b. Dynamic changes in the distribution of a satellite homologous to intergenic 26-18S rDNA spacer in *Nicotiana* genomes. *Genetics* 166: 1935–1946.
- Lim KY, Souckova-Skalicka K, Sarasan V, Clarkson JJ, Chase MW, Kovarik A, Leitch AR. 2006. A genetic appraisal of a new synthetic *Nicotiana tabacum* (Solanaceae) and the Kostoff synthetic tobacco. *American Journal of Botany* 93: 875–883.
- Lim KY, Kovarik A, Matyasek R, Chase MW, Clarkson JJ, Grandbastien MA, Leitch AR. 2007. Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. *New Phytologist* 175: 756–763.
- Liu B, Wendel JF. 2000. Retrotransposon activation followed by rapid repression in introgressed rice plants. *Genome* 43: 874–880.
- Liu ZL, Wang YM, Shen Y, Guo W, Hao S, Liu B. 2004. Extensive alterations in DNA methylation and transcription in rice caused by introgression from *Zizania latifolia*. *Plant Molecular Biology* 54: 571–582.
- Madlung A, Tyagi AP, Watson B, Jiang H, Kageuchi T, Doerge RW, Martienssen R, Comai L. 2005. Genomic changes in synthetic Arabidopsis polyploids. *Plant Journal* 41: 221–230.
- Matzke MA, Gregor W, Mette MF, Aufsatz W, Kanno T, Jakowitsch J, Matzke AJ. 2004. Endogenous pararetroviruses of polyploid tobacco (*Nicotiana tabacum*) and its diploid progenitors, *N. sylvestris* and *N. tomentosiformis*. *Biological Journal of the Linnean Society* 82: 627–638.
- McClintock B. 1984. The significance of responses of the genome to challenge. *Science* 226: 792–801.
- Melayah D, Bonnivard E, Chalhoub B, Audéon C, Grandbastien M-A. 2001. The mobility of the tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors. *The Plant Journal* 28: 159–168.
- Melayah D, Lim KY, Bonnivard E, Chalhoub B, Dorlhac de Borne F, Mhiri C, Leitch AR, Grandbastien M-A. 2004. Distribution of the Tnt1 retrotransposon family in the amphidiploid tobacco (*Nicotiana tabacum*) and its wild *Nicotiana* relatives. *Biological Journal of the Linnean Society* 82: 639–649.
- Michalak P. 2009. Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. *Heredity* 102: 45–50.
- Murad L, Lim KY, Christopodoulou V, Matyasek R, Lichtenstein CP, Kovarik A, Leitch AR. 2002. The origin of tobacco's T genome is traced to a particular lineage within *Nicotiana tomentosiformis* (Solanaceae). *American Journal of Botany* 89: 921–928.
- Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien M-A. 2010. Impact of transposable elements in organization and functioning of allopolyploid genomes. *New Phytologist* 186: 37–45.
- Petit M, Lim KY, Julio E, Poncet C, Dorlhac De Borne F, Kovarik A, Leitch AR, Grandbastien M-A, Mhiri C. 2007. Differential impact of retrotransposon populations on the genome of allotetraploid tobacco (*Nicotiana tabacum*). *Molecular Genetics and Genomics* 278: 1–15.
- Ramsey J, Schemske DW. 2002. Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics* 33: 589–639.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

- Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. 2001. 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.
- Shan XH, Liu ZL, Dong ZY, Wang YM, Chen Y, Lin XY, Long LK, Han FP, Dong YS, Liu B. 2005. Mobilization of the active MITE transposons *mPing* and *Pong* in rice by introgression from wild rice (*Zizania latifolia* Griseb.). *Molecular Biology and Evolution* 22: 976–990.
- Skalicka K, Lim KY, Matyasek R, Koukalova B, Leitch AR, Kovarik A. 2003. Rapid evolution of parental rDNA in a synthetic tobacco allotetraploid line. *American Journal of Botany* 90: 988–996.
- Skalicka K, Lim KY, Matyasek R, Matzke M, Leitch AR, Kovarik A. 2005. Preferential elimination of repeated DNA sequences from the paternal, *Nicotiana tomentosiformis* genome donor of a synthetic, allotetraploid tobacco. *New Phytologist* 166: 291–303.
- Song K, Lu P, Tang K, Osborn TC. 1995. Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences, USA* 92: 7719–7723.
- Ungerer MC, Strakosh SC, Zhen Y. 2006. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Current Biology* 16: R872–R873.
- Vernhettes S, Grandbastien M-A, Casacuberta JM. 1997. In vivo characterization of transcriptional regulatory sequences involved in the defence-associated expression of the tobacco retrotransposon Tnt1. *Plant Molecular Biology* 35: 673–679.
- Vernhettes S, Grandbastien M-A, Casacuberta JM. 1998. The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high plasticity of its regulatory sequences. *Molecular Biology and Evolution* 15: 827–836.
- Waugh R, Mclean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT, Powell W. 1997. Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence specific amplification polymorphisms (SSAP). *Molecular and General Genetics* 253: 687–694.
- Wendel JF. 2000. Genome evolution in polyploids. *Plant Molecular Biology* 42: 225–249.
- Zhao XP, Si Y, Hanson RE, Crane CF, Price HJ, Stelly DM, Wendel JF, Paterson AH. 1998. Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. *Genome Research* 8: 479–492.

Supporting Information

Additional supporting information may be found in the online version of this article.

Note S1 Sequences of Tnt1/flanking junctions characterized as sequence-specific amplification polymorphism (SSAP) bands polymorphic in Th37 S4 plants.

Table S1 Nature of genomic target sequences flanking Tnt1 insertions

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.



About New Phytologist

- *New Phytologist* is owned by a non-profit-making **charitable trust** dedicated to the promotion of plant science, facilitating projects from symposia to open access for our Tansley reviews. Complete information is available at **www.newphytologist.org**.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as-ready' via *Early View* – our average submission to decision time is just 29 days. Online-only colour is **free**, and essential print colour costs will be met if necessary. We also provide 25 offprints as well as a PDF for each article.
- For online summaries and ToC alerts, go to the website and click on 'Journal online'. You can take out a **personal subscription** to the journal for a fraction of the institutional price. Rates start at £151 in Europe/\$279 in the USA & Canada for the online edition (click on 'Subscribe' at the website).
- If you have any questions, do get in touch with Central Office (**newphytol@lancaster.ac.uk**; tel +44 1524 594691) or, for a local contact in North America, the US Office (**newphytol@ornl.gov**; tel +1 865 576 5261).