Parental transposable element loads influence their dynamics in young Nicotiana hybrids and allotetraploids

Corinne Mhiri, Christian Parisod, Julien Daniel, Maud Petit, K. Yoong Lim, François Dorlhac de Borne, Ales Kovarik, Andrew Leitch, Marie-Angele Grandbastien

DR CORINNE MHIRI (Orcid ID : 0000-0002-1377-6604)

DR CHRISTIAN PARISOD (Orcid ID : 0000-0001-8798-0897)

Article type : MS - Regular Manuscript

Parental transposable element loads influence their dynamics in young *Nicotiana* hybrids and allotetraploids

Corinne Mhiri[1], Christian Parisod[2], Julien Daniel[1], Maud Petit[1], K. Yoong Lim[3], François Dorlhac de Borne[4], Ales Kovarik[5], Andrew R. Leitch[3] and Marie-Angèle Grandbastien[1]

[1]Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, F-78000 Versailles, France; [2]Ecological Genomics, Institute of Plant Sciences, University of Bern, CH-3013 Bern, Switzerland; [3]School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK; [4]Leaf Research, Imperial Brands, 24100 Bergerac, France; [5]Institute of Biophysics, Academy of Sciences of the Czech Republic, CZ-61265 Brno, Czech Republic

ORCID ID for authors :

Corinne MHIRI : 0000-0002-1377-6604

Christian PARISOD: 0000-0001-8798-0897

Ales KOVARIK: 0000-0003-2896-0698

Andrew R LEITCH: 0000-0001-8574-302X

Marie-Angèle GRANDBASTIEN : 0000-0003-2379-3448

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/nph.15482

Author for correspondence:

Corinne MHIRI

Tel: +33 1 30 83 30 24

Email: Corinne.Mhiri@inra.fr

Received: *10 July 2018*

Accepted: *6 September 2018*

Summary

- The genomic shock hypothesis suggests that allopolyploidy is associated with genome changes driven by transposable elements, as a response to imbalances between parental insertion loads. To explore this hypothesis, we compared three allotetraploids *Nicotiana arentsii*, *N. rustica* and *N. tabacum,* which arose over comparable time frames from hybridization between increasingly divergent diploid species.
- We used sequence-specific amplification polymorphism (SSAP) to compare the dynamics of six transposable elements in these allopolyploids, their diploid progenitors, and in corresponding synthetic hybrids.
- We show that element-specific dynamics in young *Nicotiana* allopolyploids reflect their dynamics in diploid progenitors. Transposable element mobilization is not concomitant with immediate genome merger, but occurs within the first generations of allopolyploid formation. In natural allopolyploids, such mobilizations correlate with imbalances in the repeat profile of the parental species, which increases with their genetic divergence. Other restructuring leading to locus loss is immediate, non-random, and targeted at specific subgenomes, independently of cross orientation.
- The correlation between transposable element mobilization in allopolyploids and quantitative imbalances in parental transposable element loads supports the genome shock hypothesis proposed by McClintock.

Keywords: allopolyploidy, evolution, genome shock, *Nicotiana arentsii, Nicotiana rustica*, *Nicotiana tabacum*, sequence-specific amplification polymorphism (SSAP), transposable element.

Introduction

Allopolyploidy, i.e. interspecific hybridization associated with whole-genome duplication, is a recurrent plant speciation process that has been extensively described in crops (cotton, wheat, rapeseed), model and wild species (for example in genera *Arabidopsis*, *Spartina*, *Tragopogon*) (Soltis & Soltis, 2009; Jiao *et al.*, 2011; Soltis *et al.*, 2014; Soltis *et al.*, 2015; Alix *et al.*, 2017). Allopolyploidy is thought to stimulate genetic and epigenetic changes (Wendel, 2000; Doyle *et al.*, 2008), potentially resulting in short-term "revolutionary changes" such as DNA elimination, gene loss, chromosome exchanges, amplification of high-copy DNA sequences and chromatin or methylation modifications (Feldman & Levy, 2009). Over longer time frames genome diploidisation processes can lead to genome downsizing, loss of gene duplicates, genome fraction and chromosome number reductions (Dodsworth *et al.*, 2015).

In synthetic hybrids and allopolyploids that mimic natural species (i.e. newly created by crossing diploid species), the extent of genome restructuring generated by allopolyploidy seems to be species-dependent or even line-dependent. For example, no significant structural rearrangements were found in synthetic cotton allopolyploids (Liu *et al.*, 2001), whereas translocations, chromosome loss and abnormal segregation patterns were observed in the first generations *Tragopogon* and *Brassica napus* allopolyploids (Lim *et al.*, 2008; Sarilar *et al.*, 2013) and up to 14 % of parental genomic loci loss was reported in *Aegilops-Triticum* synthetic hybrids and allopolyploids (Shaked *et al.*, 2001). The timing and level of genetic changes associated with allopolyploidy appear to be dependent on species and genotypes, and it remains unclear what is driving these different responses.

The plant genus *Nicotiana* L. (Solanaceae) represents an excellent model to investigate allopolyploidy, as recurrent and independent allopolyploidy involving different parental diploids occured in nature. Three young allotetraploids, *N. arentsii*, *N. tabacum* and *N. rustica,* that formed over similar time frames (Clarkson *et al.*, 2017), are studied here: (1) *N. arentsii* formed *ca*. 0.36 million years ago (Mya) from diploid progenitors of *N. undulata* (maternal) and *N. wigandioides* (paternal), which diverged *ca.* 5.4 Mya. (2) *N. rustica* formed *ca.* 0.57 Mya from progenitors of *N. paniculata* (maternal) and *N. undulata* (paternal), which diverged *ca*. 7.7 Mya. (3) *N. tabacum* (tobacco) formed *ca*. 0.40 Mya from progenitors of *N. sylvestris* (maternal) and *N. tomentosiformis* (paternal), which diverged *ca*. 12.9 Mya. The three allopolyploids thus arose within recent and comparable time frames from interspecific hybridizations between diploid progenitor pairs that

diverged at highly variable times, and represent a rare experimental system to compare the impact of parental divergence on genomic restructuring associated with allopolyploid formation.

Transposable elements (TEs) represent a substantial and variable part of plant genomes, accounting for genome size variation in angiosperms (Tenaillon *et al.*, 2010). They are classified according to the nature of their transposition intermediate (Wicker *et al.*, 2007), with class I elements, or retrotransposons, transposing via a RNA-based replicative mechanism that can lead to rapid increases in copy numbers, whereas most class II elements excise and re-insert via a conservative mechanism. Class I elements are generally predominant in large plant genomes (Oliver *et al.*, 2013). They are central players in genome reorganization after interspecific hybridization, thought to generate a genome shock associated with the mobilization of transposable elements (McClintock, 1984) in response to imbalance between parental insertion loads (Michalak, 2009; Parisod *et al.*, 2012). While transposition in response to allopolyploidy can be restricted to specific elements, transposable elements also appear associated with major structural changes such as recombination-driven sequence loss (Parisod *et al.*, 2010). *Nicotiana* represents a material of choice for the study of dynamics of transposable element associated with allopolyploidy because it contains some of the best-characterized, active plant retrotransposons (Grandbastien *et al.*, 1989; Hirochika *et al.*, 1996). Previous studies of retrotransposons in *N. tabacum* have revealed the rapid turnovers of insertion sites and element-specific dynamics (Petit *et al.*, 2007), including in the S4 generation of synthetic allotetraploid mimics of tobacco (Petit *et al.*, 2010).

Here, we compared genome changes associated with several transposable element families in the allopolyploids *N. arentsii*, *N. rustica* and *N. tabacum.* We also compared the natural species with a large panel of F1/ S0 synthetic hybrids, including reciprocal hybrids. We selected six *Nicotiana* transposable elements for study, these being one class II element and five class I retrotransposons (Table 1). These elements were selected because they displayed contrasted dynamics, with some of them young and active, while others are older, including some ancient elements conserved among eudicots and monocots.

Using genome-wide SSAP (Sequence-Specific Amplification Polymorphism) (Waugh *et al.*, 1997), and a range of natural and synthetic hybrids, we compared the insertion polymorphisms of these transposable element families and examined the impact of variable genome shock intensities on structural changes associated with allopolyploidy. Our results show that: (1) the same families show similar dynamics in diploids and derived hybrids and allopolyploids; (2) transposable element are not mobilized at very early stages but other non stochastic restructuring occurs, targeted at particular insertions and specific subgenomes; (3) much genome changes occur within the first generations of allopolyploid formation. Finally, we show that transposable element mobilization transposition in

allopolyploids is greatest when parental repeat profiles are most imbalanced, congruent with the genome shock hypothesis.

Material and methods

Plant material

Natural accessions: We used 49 *Nicotiana* accessions (Table 2) originating from various sources and reflecting the available diversity in seed collections. *N. wigandioides* and *N. arentsii* had only one accession available. For *N. tabacum* we used landraces that were not derived from breeding programs, and are considered "ancestral types" (Petit *et al.*, 2007), and two feral tobaccos.

Synthetic hybrids: We analyzed 26 *Nicotiana* synthetic hybrids (Table 3). Synthetics hybrids were coded according to their progenitor species, arranged in the maternal paternal order. For *N. sylvestris* x *N. tomentosiformis* ST hybrids, we used several independent lines formed from different progenitor accession pairs. We distinguished these lines with a supplementary letter: ST-H refers to ST hybrids between diploid parents, ST-D refers to ST hybrids between tetraploidized parents, and ST-Th refers to the Th37 allopolyploid (see the section "S4 allopolyploids", below).

F1 hybrids: Three independent *N. tabacum* ST-H-F1 hybrids were obtained from a cross between *N. sylvestris* (TW137, S-subgenome donor) and *N. tomentosiformis* (ITG645, T-subgenome donor). The *N. arentsii* F1 hybrids (UW), involving diploid parents *N. undulata* (U-subgenome donor) and *N. wigandioides* (W-subgenome donor) and the *N. rustica* F1 hybrids (PU1, UP1, UP2), involving diploid parents *N. undulata* (U-subgenome donor) and *N. paniculata* (P-subgenome donor) were described previously (Dadejova *et al.*, 2007). The PU and UP F1 hybrids correspond to reciprocal crosses between the same *N. undulata* and *N. paniculata* parental accessions.

S0 allopolyploids: Doubled S0 versions of F1 hybrids were obtained by plantlet regeneration from *in vitro* calli, which efficiently induces genome doubling in tobacco (Kasperbauer & Collins, 1972). F1 plants were germinated and maintained *in vitro* on S-medium (Duchefa Biochemie, NL) in a growth chamber under a 21°C – 70% humidity - 16 h light / 8 h dark photoperiod. F1 leaf explants were placed on B-medium supplemented with IAA (Indole-3-acetic acid) 0.5 $mg.ml^{-1}$ and BA (benzyl adenine) $1mg.ml^{-1}$ (Bourgin *et al.*, 1979). Plantlets regenerated from proliferating calli were rooted and subcultured on S-medium. Successful genome doubling was identified by flow cytometry on a CyFlow SL3 (Partec) cytometer. We obtained S0 plants from the ST-H15-F1 and the PU-F1 hybrids. No calli were obtained from UW F1 leaf explants, and no S0 plants were identified in 33 regenerants

of UP-F1 hybrids. Ten independent synthetic ST-D-S0 hybrids, obtained through crossing of tetraploidized *N. sylvestris* B39 with tetraploidized *N. tomentosiformis* TW142, were included in our study.

S4 allopolyploids: We also included in this work ST-Th-S4 plants corresponding to S4 segregating progenies of the synthetic Th37 tobacco allotetraploid (Burk, 1973). The four S4 lines used in this work (Th1, Th3, Th8, Th9) were analyzed previously (Skalicka *et al.*, 2003; Kovarik *et al.*, 2004; Lim *et al.*, 2004; Skalicka *et al.*, 2005; Petit *et al.*, 2010) and are representative of the Th37 segregant groups defined in (Skalicka *et al.*, 2003).

Plant terminology: For simplification, we assembled taxa into three groups named Arentsii, Rustica and Tabacum, each containing accessions, as appropriate, of the natural allotetraploid, of the diploid progenitor species, and of their derived synthetic hybrids and allopolyploids.

SSAP genotyping

SSAP was performed as in (Petit *et al.*, 2007) and (Parisod *et al.*, 2012). Briefly, leaf DNA was digested with EcoRI and ligated with EcoRI-specific adapters. Regions encompassing transposable element termini and flanking sequences were amplified by PCR using an EcoRI adaptor primer and a primer designed in the extremity of each element: Ns1-1F, Au-1F, TS-a, Tnt1-OL16, Tnt2d and TRIM-b, as in (Parisod *et al.*, 2012). For each element, all Tabacum lines were run side-by-side on the same polyacrylamide gel, while the Rustica and Arentsii lines were run side-by-side on the same polyacrylamide gel, because of their common progenitor *N. undulata*. At least two replicates of each analysis were performed and only reproducible and clearly resolved SSAP loci were manually scored as present (1) or absent (0).

For each element, independent matrices were assembled to assess genetic diversity in natural species, synthetics and parental accessions (Supporting Information Table S1). Proportions of polymorphic SSAP loci were estimated, along with the Shannon Index H that suitably estimates diversity for dominant markers in autogamous species as considered here (Bussell, 1999).

Phylogenetic Reconstructions

Relationships among *Nicotiana* taxa assessed with the SSAP datasets were evaluated by the neighbor-net method, using SplitsTree 4.16 (Huson & Bryant, 2006). Each neighbor-net diagram was produced from distances computed as the number of pairwise differences among individual genotypes

(UncorrectedP option). This procedure was applied to each of the different element datasets independently, as well as to the compiled matrix obtained with all six elements.

Distribution of SSAP variation within and among groups

For each locus *j*, the frequencies of band presence (1) and absence (0) were estimated among accessions of the paternal ($P(1)_j$, $P(0)_j$) or maternal ($M(1)_j$, $M(0)_j$) diploid progenitors as well as among accessions of the hybrid/allopolyploid ($H(1)_j$, $H(0)_j$). These (H) frequencies were compared to the (E) profile expected under additivity of diploid progenitors profiles, which was estimated as: $E(1)_j = [P(1)_j \times M(0)_j] + [P(0)j \times M(1)_j] + [P(1)_j \times M(1)_j]$ for locus *j* presence frequency and $E(0)_j = P(0)_j \times M(0)_j$ for locus *j* absence frequency. Frequencies of additive (A), new (N) or missing parental loci (L) were calculated for each hybrid/allopolyploid and each locus *j* as follows: (i) additive locus $A_j = [E(1)_j \times H(1)_j] + [E(0)_j \times H(0)_j]$ (i.e. sum of the products of locus *j* presence or absence frequency in the hybrid/allopolyploid (H) and in the expected (E) profiles); (ii) new locus $N_j = E(0)_j \times H(1)_j$ (i.e product of locus *j* absence frequency in expected (E) profile and locus *j* presence frequency in the hybrid/allopolyploid (H)); (iii) missing parental locus $L_j = E(1)_j \times H(0)_j$ (i.e. product of locus *j* presence frequency in the expected profile (E) and absence frequency in the hybrid/allopolyploid). Frequencies of additive, new or missing parental loci were summed across all *j* loci, and converted into proportions following Senerchia *et al.* (2014). We also calculated and summed across loci the frequencies of missing loci of paternal or maternal origin (LM and LP) as follows: $LP_j = P(1)_j \times M(0)_j \times H(0)_j$ for loci of paternal origin and $LM_j = M(1)_j \times P(0)_j \times H(0)_j$ for loci of maternal origin, as well as the frequencies of common loci shared by progenitors (C) : $C_j = P(1)_j \times M(1)_j$, and the frequencies of paternal-specific loci (PB) and maternal-specific loci (MB): $PB_j = P(1)j \times M(0)_j$ and $MB_j = M(1)_j \times P(0)_j$. The R script for these calculations is provided in Supporting Information Notes S1.

For each natural allotetraploid species and synthetic hybrids with at least three accessions, one-way ANOVA followed by multiple comparisons with Tukey's honest significant differences (HSD) were carried out to assess transposable elements displaying significantly different proportions of non additive bands (new or lost loci) using the package 'agricolae' on R cran (https://cran.r-project.org/). We used as internal baseline TRIM (Witte *et al.*, 2001; Yasui *et al.*, 2001), an ancient element showing a high sequence heterogeneity indicative of low transpositional activity in *N. tabacum* (unpublished data), and presenting a basal restructuring activity in older *Nicotiana* allopolyploids of the *Repandae* section (Parisod *et al.*, 2012). Hence, only non-additive locus percentages significatively different from those of TRIM after Tukey HSD Tests were considered as relevant.

The correlation between the proportion of non-shared loci between progenitors and the proportions of new or missing loci in allopolyploids was assessed with a General Linear Mixed Model (GLMM) including variation within species and types of transposable elements as random effects in our data, according to Senerchia *et al.* (2014).

Results

Phylogenetic relationships of natural and synthetic *Nicotiana* hybrids

SSAP characterised six transposable elements, giving a total of 426 reproducible loci for the Arentsii group (i.e. *N. arentsii*, synthetic mimics and progenitor species), 368 for the Rustica group (i.e. *N. rustica*, synthetic mimics and progenitor species), and 542 loci for the Tabacum group (i.e. *N. tabacum*, synthetic mimics and progenitor species) (Supporting Information Tables S1, S2). We computed the neighbor-net diagram for each group using a combination of the six elements, as well as for each element separately. The topologies of each combined tree (Fig. 1) presented the parental taxa on opposite sides of the neighbor-net, with the natural allotetraploids and synthetic hybrids branching at intermediate positions, supporting their hybrid origin. F1 and S0 lines from the same cross (see Table 3) or hybrids produced from reciprocal crosses (such as *N. paniculata* x *N. undulata* PU and *N. undulata* x *N. paniculata* UP hybrids) clustered together. In contrast, the synthetic ST-Th-S4 segregants of the Tabacum group were differentiated from the other Tabacum synthetics (Fig. 1c). The topologies placed *N. arentsii*, *N. rustica* and their corresponding synthetics closer to the paternal species, whereas *N. tabacum* and its corresponding synthetics were closer to the maternal species. As expected, branch lengths separating the parents from natural and synthetic hybrids were shorter for the latter, indicating that they present more additive profiles than the natural allopolyploids.

Although similar to combined neighbor-nets, topologies from each element differed slightly (Supporting Information Fig. S1a,b,c), as expected from their specific dynamics. This is particularly clear for the Tabacum group. Some of the element-specific patterns are explained by differential parental element loads, which also correlate with the degree of genetic distance between diploid progenitors, as described below.

Transposable element loads and related genetic diversity in diploid progenitors

To evaluate relative transposable element loads and related genetic diversity in diploid progenitors, we determined the total number of SSAP loci and the proportion of polymorphic loci as well as the Shannon Diversity Index (H) (Fig. 2, Supporting Information Table S2). We also

evaluated the percentage of loci that were species-specific or shared by progenitors, taking intraspecific polymorphism into account (Table 4).

For *N. arentsii* progenitors (*N. undulata* and *N. wigandioides*), the global number of SSAP loci is higher in paternal *N. wigandioides* than in *N. undulata*, driven by significantly higher numbers of TS and Ns1, and to a lesser extent Tnt2 and Tnt1. The ancient Au and TRIM elements showed insignificant differences in parental loads. It was not possible to compare the diversity between progenitors, as only one *N. wigandioides* accession was available.

For *N. rustica* progenitors (*N. paniculata* and *N. undulata*), the global number of SSAP loci is higher in paternal *N. undulata* than in *N. paniculata*. Both diploid species showed similar diversity overall (H for *N. paniculata* = 0.643 and H for *N. undulata* = 0.71), but a particularly high diversity for TS in *N. paniculata* (H = 0.237) compared with *N. undulata* (H = 0.074). TS elements are less abundant in *N. paniculata* and are mainly responsible for *N. paniculata* having fewer SSAP loci overall. In contrast, Tnt2 SSAP loci occur in *N. undulata* and *N. paniculata* in similar numbers, but with higher polymorphisms in *N. undulata* (H = 0.210), potentially revealing its recent activity in this species. The ancient Au and TRIM elements displayed insignificant difference in parental loads.

For *N. tabacum* progenitors (*N. sylvestris* and *N. tomentosiformis*), the global number of SSAP loci is greater in maternal *N. sylvestris*, largely caused by larger numbers of Tnt2, TS and Tnt1 loci. Although displaying significantly more loci, *N. sylvestris* showed a low diversity (global H = 0.271, with H < 0.09 for each type of element analysed) compared with *N. tomentosiformis* (global H = 0.895, with H > 0.08 for each element), suggesting that insertions are mainly fixed in the *N. sylvestris* population. The ancient elements Au and TRIM displayed insignificant difference in parental loads.

We observed significant variations in the proportion of shared SSAP loci between progenitors (Table 4), varying from 4.1% (for Tnt2 between *N. sylvestris* and *N. tomentosiformis*) to more than 36% (for TRIM between *N. undulata* and *N. wigandioides*). However, the global proportion of shared SSAP loci (23.2% for *N. undulata* and *N. wigandioides,* 16.3% for *N. paniculata* and *N. undulata* and 12.1% for *N. sylvestris* and *N. tomentosiformis*) reflects the increasing genetic distances between progenitor pairs (Clarkson *et al.*, 2017). Out of the six elements considered, the putatively inactive TRIM displayed one of the highest proportion of shared loci in the three *Nicotiana* parental pairs, whereas the recently active elements TS and Tnt2 were amongst the elements with the lowest proportion of shared loci.

Transposable element-associated genome restructuring in natural allotetraploids

We compared the additivity of progenitor profiles to the profiles observed for natural allopolyploids, assessing the potential that the locus is additive (i.e. locus pre-existing in one or both parents), new (i.e. locus absent from parental accessions) or missing (parental locus missing in the hybrid).

As expected for recent allopolyploids, the global mean proportion of additive SSAP loci (Fig. 3 and Supporting Information Table S3) was high, varying from 62.69% for *N. rustica* to 70.22% and 70.32% for *N. tabacum* and *N. arentsii* respectively. New loci represented 7.55%, 13.54% and 10.78% of total loci in *N. arentsii*, *N. rustica* and *N. tabacum* respectively, whereas missing loci accounted for higher values of 21.42%, 23.77% and 18.99% respectively.

To analyze the relative contribution of each transposable element to these genome changes (Fig. 4 and Supporting Information Table S3), we used TRIM, an ancient element with low transpositional activity in *Nicotiana* (Material and Methods) as a baseline highlighting elements with distinctive evolutionary trajectories. TRIM indeed displayed a high level of additive loci (Fig. 4a) in *N. arentsii*, *N. rustica* and *N. tabacum* (74.55%, 72.47%, 83.43% respectively) and only Ns1 presented a similar pattern (77.10%, 68.90%, 72.55% respectively). Other elements showed variable proportions of additive loci according to the species being considered, with high levels of additivity observed for TS/Tnt1 in *N. arentsii* (73.04%/72.56%), Tnt2 in *N. rustica* (67.56%), and Tnt1 in *N. tabacum* (70.34%).

Different patterns of parental loci missing in hybrids (Fig. 4b) were found according to species. TRIM showed low levels of missing loci (21.43% in *N. arentsii*, 18.57% in *N. rustica*, 12.60% in *N. tabacum*). TS showed a similar pattern (18.70% in *N. arentsii*, 21.65% in *N. rustica*, 14.96% in *N. tabacum*), whereas Au displayed a high level of missing parental loci in the three allopolyploids (28.62% in *N. arentsii*, 30.85% in *N. rustica*, 21.69% in *N. tabacum*), similar to Tnt2 (27.42% in *N. arentsii*, 24.38% in *N. rustica*, 24.60% in *N. tabacum*).

Each allopolyploid displayed an element-specific profile of new loci (Fig. 4c), but globally TS and Au generated the highest number of new loci. TS produced 19.79% and 18.06% of new loci in *N. rustica* and *N. tabacum* respectively, congruent with its recent activity in diverse *Solanaceae* (Wenke *et al.*, 2011). Tnt1 generated a high percentage of new loci (20.45%) in *N. rustica* only, with two-fold lower levels in the two other allopolyploids (9.76% in *N. arentsii*, 8.50% in *N. tabacum*), whereas Tnt2 reached a high proportion of new loci (14.36%) in *N. tabacum* only, in accordance with the potential of this element to be active in this species (Petit *et al.*, 2007).

Transposable element-associated genome restructuring in synthetic hybrids

We estimated the extent and timing of restructuring after hybridization and polyploidization at different steps of the formation of synthetic hybrids of the three groups Arentsii (UW hybrids), Rustica (PU and UP hybrids) and Tabacum (ST-H, ST-D and ST-Th hybrids) (Fig. 3, Fig. 4 and Supporting Information Table S3).

As expected, the global percentage of additive SSAP loci was high in F1/S0 synthetic hybrids, ranging from 87.77% to 96.69%. Only the ST-Th-S4 segregants displayed a lower value of 77.41%. In contrast to natural allopolyploids, very few (none or one) new loci were detected in any F1/S0, with only ST-Th-S4 segregants showing a significant global level of new loci (7.63%) approaching the level of new loci in natural *N. tabacum* (10.78%). Early parental locus loss appeared as the major restructuring event in the F1/S0 synthetic hybrids, with proportions ranging from 3.29% to 7.55%, except for the UW-F1 hybrid that displays a substantially higher level (12.08%). Here too, the ST-Th-S4 segregants showed a level of missing loci (14.96%) that was closer to natural *N. tabacum* (18.99%) than to early F1/S0 synthetics (3.29 to 4.35%).

Levels of additive loci were high for all transposable elements in F1/S0, excepting Tnt2 and Au in the UW-F1 hybrid displaying high levels of parental locus loss (Tnt2: 17.86% and Au: 25.00%). Interestingly, these two elements were also chiefly associated with parental locus loss in its natural counterpart *N. arentsii*. The ST-Th-S4 segregants displayed a lower level of additive loci for the TS, Ns1 and Au elements (respectively 56.33%, 74.40% and 75.69%) corresponding to a high percentage of new loci for TS and Ns1 (respectively 17.17% and 11.31%) and a high level of missing loci for TS and Au (respectively 26.51% and 18.40%). In these ST-Th-S4 segregants, TS and Ns1 revealed the highest level of new loci (17.17% and 11.31%), suggesting that these elements may have been active within the four generations after hybridization. Accordingly, TS was also the element generating the highest level of new loci in natural *N. tabacum* (18.06%).

A low level of restructuring associated to genome doubling step (F1 to S0 transition) was observed (Supporting Information Table S1), with only sporadic changes of parental missing loci between F1 and corresponding S0 (nine in ST-H-F1 to -S0, two in PU-F1 to PU-S0).

Interestingly, we observed within synthetics of the Tabacum or Rustica groups that similar parental loci were missing in independent synthetic hybrids or in synthetic hybrids derived from different parental accessions (Supporting Information Table S1). In the Tabacum group, low polymorphism levels were found accross the three ST-H-F1 hybrids and across the ten ST-D-S0 hybrids. Furthermore, a high proportion (8 out of 18 - 44.4%) of parental loci missing in ST-D-S0 hybrids were also missing in ST-H-F1 hybrids, and a proportion (13.1%) of parental missing loci were also shared by at least two of the three Tabacum group synthetic types (ST-H, ST-D and ST-Th).

In the Rustica group hybrids, 63 % (17 out of 27) of missing parental loci were common to the reciprocal PU-F1 and UP-F1 synthetic hybrids.

Parental origin of transposable element loci not transmitted to natural and synthetic hybrids

To estimate whether restructuring of insertion sites differentially affects each parental genome, we calculated the proportions of loci of paternal or maternal origin missing in natural and synthetic hybrids (Fig. 5 and Supporting Information Table S3).

In natural *N. tabacum* and *N. rustica*, global restructuring levels appeared to be equivalent for both parental subgenomes (Fig. 5a). In contrast, higher levels of maternal loci are missing in *N. arentsii*. However, the dynamics of each element greatly differed (Fig. 5b). The proportion of missing paternal loci was similar for all elements in *N. tabacum* and *N. arentsii*, with a higher variation for missing maternal loci. In *N. rustica*, the proportion of both paternal and maternal missing loci displayed a large variation. For example, Au showed a very high level of missing maternal loci (20.72%) in *N. arentsii*, and a very low level of missing paternal loci (4.28%), although the two parental species are phylogenetically closely related. TS (in *N. rustica*) but also TRIM (in *N. tabacum*) also showed contrasted levels of missing parental loci.

In contrast to natural *N. tabacum*, a predominance of missing paternal loci was observed in all Tabacum group synthetics, with higher levels in ST-Th S4 segregants (11.71%) compared to 2.82% of missing maternal loci. In the Arentsii group UW-F1 hybrids, losses of maternal loci were higher (8.19%) than those of paternal loci (3.89%), similar to natural *N. arentsii*. Interestingly, the comparison of reciprocal UP-F1 and PU-F1 hybrids of the Rustica group indicated preferential losses of loci from the *N. paniculata* genome, independent of its paternal or maternal origin.

No element-specific trend could be evidenced in most synthetics, due to low levels of restructuring. The high level of missing maternal loci observed in the UW-F1 hybrid is restricted to a few elements (Au, Tnt2 and Ns1) and is not strictly correlated with *N. undulata*-specific element locus abundance (see Table 4). In ST-Th S4 segregants, highest levels of missing paternal loci were detected for TS, Au and Ns1.

Correlation between transposable element-related restructuring and parental transposable element loads in natural *Nicotiana* allopolyploids

We investigated to what extent restructuring in natural allotetraploids is correlated to imbalances in parental transposable element loads, using the percentage of non-shared insertions between diploid progenitors as a proxy for parental transposable element load divergence. We modeled the log-transformed percentage of new (N) or missing (M) SSAP loci as a function of the percentage of non-shared insertions for each element in each allotetraploid (Fig. 6). In both cases, simple linear regression slopes progressively increased from *N. arentsii* (N=0.04248, M=0.01104) to *N. rustica* (N=0.05589, M=0.02518) and *N. tabacum* (N=0.09486, M=0.04216). We refined our analysis following Senerchia *et al.* (2014) by generating a linear mixed model to include variation within species and types of transposable elements as random effects in our data. The maximum likelihood linear mixed model indicated that the proportion of new loci is significantly associated with the proportion of non-shared loci between progenitors (t-value = 3.9952, p-value = 0.0013, R^2 = 0.963), confirming that increasing global parental divergence is associated with increasing levels of putative new insertions in the derived allotetraploid. In contrast, the proportion of missing parental loci is not significantly associated with the proportion of non-shared progenitor loci progenitors (t-value = 0.8423, p-value = 0.4778, R^2 = 0.878).

Discussion

Transposable element evolutionary dynamics in allopolyploids reflects their dynamics in diploids

Phylogenetic networks reconstructed with SSAP data for six transposable elements, whether analysed independently or together, were congruent with prior phylogenetic studies (Clarkson *et al.*, 2017). SSAP markers show a clear split between the parental and hybrid taxa, and between synthetic hybrids and natural allopolyploids. Natural and synthetic hybrids are positioned closer to the paternal species for the Arentsii and Rustica groups, but closer to the maternal species for the Tabacum group. This directly reflects divergences in global relative transposable element loads between diploid progenitor pairs, with greater transposable element loads in the paternal diploid progenitors of *N. arentsii* and *N. rustica* (*N. wigandioides* and *N. undulata*, respectively) and in the maternal progenitor of *N. tabacum* (*N. sylvestris*).

Parental divergences also reflect element-specific trajectories during parental species evolution. Ancient transposable elements, such as Au and TRIM, show little divergence of insertion loads between parental pairs involved in allopolyploid formation, while elements known to be active

(notably TS) show much more divergent profiles. Similar results for TS was previously observed in the formation of the *Nicotiana* polyploid section *Repandae* that formed ca. 4.5 Mya (Parisod *et al.*, 2012), indicating TS activity across the genus. For other elements, activity appeared restricted to particular *Nicotiana* lineages, with e.g. Tnt2 or Tnt1 showing many differences between diploid progenitors of *N. tabacum* and *N. arentsii*, but not those of *N. rustica*. Similarly, Ns1 showed a significant divergence of insertion loads restricted to *N. arentsii* diploid progenitors.

The three natural allopolyploid species showed a significant fraction (30%-40%) of non-additive loci having likely arisen through genetic changes in the allopolyploid. However, although we have used a wide panel of parental accessions that are known in germplasms, we do not know whether more closely related parental lineages exist anywhere, or have gone extinct.

The occurrence of non-additive loci reflects element-specific trajectories similar to those observed in diploid progenitors. We detected a high level of new SSAP loci for the presumably active TS in the three allopolyploids, and in addition for Tnt2 elements in *N. tabacum*. New Tnt1 loci were also observed at high frequency in natural *N. rustica* and *N. arentsii,* but not in *N. tabacum*. The appearence of new loci in allopolyploids is thus congruent with indications of activity in at least one parental lineage. Quite surprinsingly, high numbers of new Au loci were observed in all three allopolyploids, contrasting with a lack of apparent activity in diploid lineages. Interestingly, the proportion of new Au loci in ST-Th S4 synthetics hybrids of the Tabacum group is also significantly higher than the TRIM baseline. Potentially, this dynamic could represent allopolyploid induced activation of Au, and perhaps a direct consequence of the genome shock induced by allopolyploidy. Such contrasting evolutionary dynamics of different transposable elements following hybridization were previously observed for the same elements in *Nicotiana* section *Repandae* (Parisod *et al.*, 2012) or for unrelated transposable elements in genus *Aegilops* (Senerchia *et al.*, 2014). The predominance of apparently missing parental loci for all six transposable elements in the allopolyploids is coherent with sequence loss and fits the pattern of global genome downsizing measured after allopolyploidy (Leitch *et al.*, 2008).

Early genome changes are not stochastic and occur within the first generations of allopolyploid formation

Genome changes in F1/S0 *Nicotiana* synthetics were almost exclusively parental locus losses. The very rare occurence of new loci suggests no, or minimal transposition of the six tested elements immediately following genome merger. Similar results have been reported in Triticeae (Senerchia *et al.*, 2015) and Brassicaceae (Sarilar *et al.*, 2013). Similarly, no transposition of Wis2 was detected in early synthetic wheat polyploids (Kashkush *et al.*, 2003), supporting the hypothesis of active

repressive mechanisms of transposable elements mobilization in early hybrids (Michalak, 2009). These repressive mechanisms could possibly be epigenetic and/or even post-transcriptional, since Wis2 transcriptional upregulation has been observed (Kashkush *et al.*, 2003).

In contrast, losses of parental loci were detected in F1/S0 *Nicotiana* synthetics. There were striking similarities in profiles of parental losses across the three independent ST-H-F1/S0 hybrids and across the ten independent ST-D-S0 hybrids of the Tabacum group, indicating that the same changes occured in independent hybrid zygotes. Moreover, nearly half of loci missing from ST-D-S0 hybrids were also missing in ST-H-F1/S0 hybrids, and a significant proportion was also missing in ST-Th-S4 hybrids, although the three hybrid sets were produced by crossing different combinations of diploid progenitor accessions. These data indicate that the early restructuring of parental insertion loci is not random. Similarly, identical genomic indels occurred in nearly 30% of independent isogenic *Brassica napus* synthetics (Lukens *et al.*, 2006), and non-random restructuring at insertion loci has also been detected in *Aegilops* F1 hybrids (Senerchia *et al.*, 2015). A large proportion of missing parental loci are also identical in PU-F1 and UP-F1 reciprocal hybrids of the Rustica group, revealling a limited effect of the parental direction of cross.

Very few further changes were observed after genome doubling (S0 versus F1), indicating that interspecific hybridization impacts transposable element dynamics more than polyploidy. In our study, genome doubling also involved *in vitro* tissue culture, which is thought to mobilize transposable elements (Grandbastien, 2015). Thus most transposable elements studied here are not responding significantly to tissue culture.

Higher levels of new and missing loci were detected in the S4 generation of the Th37 hybrid of the Tabacum group. Both types of changes reached levels fairly close to those observed in natural *N. tabacum*, indicating that most transposable element-associated genome changes, if not immediate, accumulate within the first generations after formation of allopolyploids. Similarly, *Veju* transpositions were only detected after the S1 generation of wheat synthetic hybrids (Kraitshtein *et al.*, 2010). Potentially, repressive epigenetic mechanisms can be overcome over generations, as demonstrated in *Arabidopsis* interspecific hybrids (Ha *et al.*, 2009). Significant genome structural changes were reported in S4, but not in S0 generations of *Arabidopsis* synthetic polyploids (Madlung *et al.*, 2005) and sequence deletions in *Veju* LTRs were observed in the S1-S3 generation of synthetic allopolyploids of wheat (Kraitshtein *et al.*, 2010).

Perhaps meiotic processes, i.e. chromosomal interactions and meiotic recombination trigger increased frequencies of transposable element mobilisation and sequence restructuring, as demonstrated for genomic rearrangements in *B. napus* synthetic hybrids (Szadkowski *et al.*, 2010). Early restructuring in F1 hybrids prior to meiosis was however observed in our work, as well as in *Aegilops* and *Brassica* F1 hybrids (Sarilar *et al.*, 2013; Senerchia *et al.*, 2015). This could result from

somatic recombinations at early embryonic stages, possibly resulting from rapid methylation changes shown to be correlated to deletions of *Veju* sequences in wheat allopolyploids (Kraitshtein *et al.*, 2010). Efforts are ongoing to obtain subsequent S1, S2 and S3 generations of synthetics hybrids to enable the detailed comparison of the accumulation of structural changes over generations.

Early genome changes affect specific subgenomes

Global restructuring of parental insertion loci occurs at a similar frequency for both parental subgenomes in *N. tabacum* and *N. rustica*, but not in *N. arentsii*, the latter showing greater restructuring of the maternal subgenome. In synthetic hybrids, a subgenome imbalance in the degree of restructuring observed at parental loci is prominent.

In all synthetics of the Tabacum group, restructuring was predominantly observed for paternal *N. tomentosiformis* loci. This targeting of the *N. tomentosiformis* subgenome is consistent with other studies showing copy number reduction of several *N. tomentosiformis* repetitive sequences in these hybrids (Skalicka *et al.*, 2005; Renny-Byfield *et al.*, 2012). Other studies in *Tragopogon mirus* (Koh *et al.*, 2010), synthetic *B. napus* (Lukens *et al.*, 2006), and synthetic wheat (Kraitshtein *et al.*, 2010) have also reported a global restructuring biaised towards the paternal subgenomes. The mechanisms underlying this genomic imbalance are not established (Hegarty *et al.*, 2013) although Song *et al.* (1995) suggested a nuclear-cytoplasmic effect, i.e. that the maternal subgenome, occuring with cytoplasmic organelles with which it has co-evolved, may be less perturbed by the genome shock of allopolyploidy than the paternal subgenome. Such nuclear-cytoplasmic interactions may be linked to epigenetic reprogramming of parental genomes, including methylation changes, chromatin remodeling and deregulation of maternally-inherited small RNA (Michalak, 2009; Parisod *et al.*, 2010).

In contrast to these data, there were greater numbers of insertion locus changes to the maternal genome than to the paternal genome in *N. arentsii* and the *N. undulata* x *N. wigandioides* F1 hybrid. These results argue against the nuclear-cytoplasmic interaction hypothesis. Several similar examples have been reported previously, for example, for insertion loci in *Spartina* hybrids (Parisod *et al.*, 2009) and chromosome loss in *A. thaliana x A. lyrata* synthetics (Beaulieu *et al.*, 2009). Furthermore, our comparison of reciprocal *N. undulata* x *N. paniculata* and *N. paniculata* x *N. undulata* F1 synthetics showed that genome restructuring targeted the *N. paniculata* subgenome more than the *N. undulata* subgenome, regardless of its parental origin. The same reciprocal hybrids also displayed a unidirectional silencing of *N. paniculata* origin rDNA units (Dadejova *et al.*, 2007).

Patterns and rates of divergence of a specific subgenome could be influenced by the specificities of the transposable elements it carries. Potentially transposable element type, distribution (clustered or not), chromosomal localization (euchromatic or heterochromatic sequences) and abundance may impact the element's subsequent fate in hybrids or polyploids. For instance, Tnt1 and Tnt2 populations are more degenerated and clustered in *N. tomentosiformis* and in the *N. tomentosiformis*-derived subgenome of *N. tabacum* (Vernhettes *et al.*, 1998; Melayah *et al.*, 2004; Petit *et al.*, 2007), and a greater level of erosion of various LTR retrotransposons was found in this subgenome than in the *N. sylvestris*-derived subgenome, which itself also revealed evidence for recent bursts of repeat amplification (Renny-Byfield *et al.*, 2011). Differences in transposable element populations between *N. tomentosiformis* and *N. sylvestris* could thus potentially explain preferential targeting the T-subgenome in synthetic hybrids of the Tabacum group. Restructuring at insertion loci may also reflect the subgenome's propensity for recombination, which may be influenced by, for example, differences in subgenome asymmetrical DNA methylation (Kraitshtein *et al.*, 2010; Senerchia *et al.*, 2015). Interactions with the second merged subgenome may thus also play a role, explaining that the *N. undulata*-derived subgenome shows higher destabilization when merged with the *N. wigandioides*-derived subgenome (as in the Arentsii group) than when it is merged with the *N. paniculata*-derived subgenome (as in the Rustica group).

Discrepancies in patterns of subgenome divergence between natural allopolyploids and their corresponding synthetic allopolyploids were also observed for other repeats, indicating that genomic changes over extended generations during evolution may cause new patterns to emerge that are not apparent in early generations. For example, *N. sylvestris* 35S rDNA units are maintained in Th37 and F1 synthetic hybrids, and mostly lost in *N. tabacum* (Skalicka *et al.*, 2003; Kovarik *et al.*, 2004). Similarly, *Nicotiana arentsii* and *N. rustica* show 35S rDNA homogenization towards *N. undulata*-derived units, but additive rDNA profiles in their synthetic hybrids. In addition, *Nicotiana* allopolyploid species of section *Polydicliae* show DNA changes that are not detectable in resynthesized allopolyploids at S5 generation (Anssour *et al.*, 2009). The extent of major genomic changes observed in *Nicotiana* allopolyploids, perhaps associated with long-term diploidization processes, is thought to correlate with the age of *Nicotiana* allopolyploid species (Lim *et al.*, 2007).

The genome shock intensity influences transposable element mobilization

Interspecific hybridization is thought to generate a genome shock that has been associated with the mobilization of transposable elements (McClintock, 1984). As suggested previously (Michalak, 2009; Parisod *et al.*, 2012), the qualitative and/or quantitative imbalance between parental transposable element loads could result in conflicting interactions between subgenomic element populations and in inefficient epigenetic repression of transposable elements. It is thus expected that

greater imbalance leads to stronger genome shock intensity, as previously proposed by Lim *et al.* (2004). The comparative analysis of the same transposable elements in *N. arentsii*, *N. rustica* and *N. tabacum* offers an excellent test for this hypothesis, especially as we observed here that the global transposable element load shared between diploids declined with an increasing genetic distances between the diploids.

To refine this analysis, and take into account element-specific trajectories, we evaluated the correlation between the proportion of new or missing loci in the three allopolyploids, and the proportion of loci that are not shared between parental diploids for each element, representing a proxy for quantitative transposable element load imbalance. The proportion of new loci, i.e. putative mobilization of transposable elements in allopolyploids, correlates significantly with the extent of each transposable element load imbalance, confirming the hypothesis of an influence of the genome shock intensity on transposable element activation. In turn, this is also correlated with the phylogenetic distance between the parents. In contrast, the extent of parental locus loss does not correlate with transposable element imbalance in the parents, despite a weak correlation with the degree of phylogenetic divergence between the parents. This suggests that mechanisms involved in genome restructuring influencing locus losses (e.g. recombination/deletion) are substantially different processes to those associated by repeat amplification.

Due to their ubiquity, repetitive nature, transposition ability and epigenetic sensivitity, transposable elements are suspected to be key players in genome evolution, especially associated with hybridization and polyploidization. Although qualitative differences between parental transposable element populations have not been assessed here, our results highlight transposable element quantitative imbalance among progenitors as an important factor for transposable element evolution in allopolyploids, and largely support the genome shock hypothesis of McClintock (1984). Further investigations including post-S0 generations of reciprocal *Nicotiana* synthetic hybrids are needed to decipher the biological processes involved to the initial genome shock (such as the meiosis step), to estimate the kinetics of genomic changes over generations, and to further evaluate the importance of transposable elements in genome evolution in polyploid speciation.

Acknowledgements

We wish to dedicate this work to our friend KY Lim who passed away, we miss him deeply. We thank N Senerchia for statistical advices and glasshouse holders P Grillot and M Lebrusq for their plant care assistance. The work was funded by the BTH01453 ARN (Association pour la Recherche sur les Nicotianées) research contract, by the ANR (Agence Nationale de la Recherche) Biodiversity 2005 program (ANR-05-BDIV-015, "Effect of polyploidy on plant genome biodiversity and

evolution"), by a NREC funding and by the F-UK Alliance (08782YM) and F-CZ Barrande (09150SH and GACR 501/12/G090) bi-national programs. The IJPB benefits from the support of the Labex Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS).

Author's contribution

CM designed and conducted the experiments, performed the analysis and wrote the manuscript. CP contributed to the statistical analysis and edited the final version of the manuscript. JD maintained *in vitro* the *Nicotiana* hybrid collection. AK, ARL, KYL and FDB created F1 *Nicotiana* hybrids, provided *Nicotiana* accessions and edited the final version of the manuscript. JD maintained *in vitro* the *Nicotiana* hybrid collection. MP and KYL produced some S0 doubled material. MAG obtained the fundings, supervised the research and edited the final version of the manuscript.

References

Alix K, Gerard PR, Schwarzacher T, Heslop-Harrison JSP. 2017. Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Annals of botany* **120**: 183-194.

Anssour S, Krugel 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* **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.

Bourgin JP, Chupeau Y, Missonier C. 1979. Plant Regeneration from mesophyll protoplasts of several *Nicotiana* species. *Physiol Plant* **45**: 288-292.

Bureau TE, Wessler SR. 1994. Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *The Plant cell* **6**: 907-916.

Burk LG. 1973. Partial self-fertility in a theoretical amphidiploid progenitor of *N. tabacum*. *Journal of Heredity* **64**: 348-350.

Bussell JD. 1999. The distribution of random amplified polymorphic DNA (RAPD) diversity amongst populations of *Isotoma petraea* (Lobeliaceae). *Molecular Ecology* **8**: 775-789.

Clarkson JJ, Dodsworth S, Chase MW. 2017. Time-calibrated phylogenetic trees establish a lag between polyploidisation and diversification in *Nicotiana* (Solanaceae). *Plant Systematics and Evolution* **303**: 1001-1012.

Dadejova M, Lim KY, Souckova-Skalicka K, Matyasek R, Grandbastien MA, Leitch A, Kovarik A. 2007. Transcription activity of rRNA genes correlates with a tendency towards intergenomic homogenization in *Nicotiana* allotetraploids. *New Phytologist* **174**: 658-668.

Dodsworth S, Leitch AR, Leitch IJ. 2015. Genome size diversity in angiosperms and its influence on gene space. *Current opinion in genetics & development* **35** (Supplement C): 73-78.

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. 2008. Evolutionary genetics of genome merger and doubling in plants. *Annual review of genetics* **42**: 443-461.

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. 2015. LTR retrotransposons, handy hitchhikers of plant regulation and stress response. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1849**: 403-416.

Grandbastien MA, Spielmann A, Caboche M. 1989. Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* **337**: 376-380.

Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang XJ, Chen ZJ. 2009. Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proceedings of the National Academy of Sciences, USA* **106**: 17835-17840.

Hegarty M, Coate J, Sherman-Broyles S, Abbott R, Hiscock S, Doyle J. 2013. Lessons from natural and artificial polyploids in higher plants. *Cytogenetic and genome research* **140**: 204-225.

Hirochika H, Otsuki H, Yoshikawa M, Otsuki Y, Sugimoto K, Takeda S. 1996. Autonomous transposition of the tobacco retrotransposon Tto1 in rice. *The Plant cell* **8**(4): 725-734.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution* **23**: 254-267.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**: 97-100.

Kashkush K, Feldman M, Levy AA. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics* **33**: 102-106.

Kasperbauer MJ, Collins GB. 1972. Reconstitution of diploids from leaf tissue of anther-derived haploids in tobacco. *Crop Science* **12**: 98-101.

Koh J, Soltis PS, Soltis DE. 2010. Homeolog loss and expression changes in natural populations of the recently and repeatedly formed allotetraploid *Tragopogon mirus* (Asteraceae). *BMC genomics* **11**: 97.

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**: 615-625.

Kraitshtein Z, Yaakov B, Khasdan V, Kashkush K. 2010. Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. *Genetics* **186**: 801-812.

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, 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.

Lim KY, Matyasek R, Kovarik A, Leitch A. 2004. Genome evolution in allotetraploid *Nicotiana*. *Biological Journal of the Linnean Society* **82**: 599-606.

Lim KY, Soltis DE, Soltis PS, Tate J, Matyasek R, Srubarova H, Kovarik A, Pires JC, Xiong Z, Leitch AR. 2008. Rapid chromosome evolution in recently formed polyploids in *Tragopogon* (Asteraceae). *PloS one* **3**: e3353.

Liu B, Brubaker CL, Mergeai G, Cronn RC, Wendel JF. 2001. Polyploid formation in cotton is not accompanied by rapid genomic changes. *Genome* **44**: 321-330.

Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T. 2006. Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. *Plant physiology* **140**: 336-348.

Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, Martienssen R, Comai L. 2005. Genomic changes in synthetic *Arabidopsis* polyploids. *The Plant Journal* **41**: 221-230.

McClintock B. 1984. The significance of responses of the genome to challenge. *Science* **226**: 792-801.

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.

Oliver KR, McComb JA, Greene WK. 2013. Transposable elements: powerful contributors to angiosperm evolution and diversity. *Genome biology and evolution* **5**: 1886-1901.

Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien MA. 2010. Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytologist* **186**: 37-45.

Parisod C, Mhiri C, Lim KY, Clarkson JJ, Chase MW, Leitch AR, Grandbastien MA. 2012. Differential dynamics of transposable elements during long-term diploidization of *Nicotiana* section Repandae (Solanaceae) allopolyploid genomes. *PloS one* **7**: e50352.

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

Petit M, Guidat C, Daniel J, Denis E, Montoriol E, Bui QT, Lim KY, Kovarik A, Leitch AR, Grandbastien MA, et al. 2010. Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytologist* **186**: 135-147.

Petit M, Lim KY, Julio E, Poncet C, Dorlhac de Borne F, Kovarik A, Leitch AR, Grandbastien MA, Mhiri C. 2007. Differential impact of retrotransposon populations on the genome of allotetraploid tobacco (*Nicotiana tabacum*). *Molecular Genetics and Genomics* **278**: 1-15.

Renny-Byfield S, Chester M, Kovarik A, Le Comber SC, Grandbastien MA, Deloger M, Nichols RA, Macas J, Novak P, Chase MW, et al. 2011. Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs. *Molecular biology and evolution* **28**: 2843-2854.

Renny-Byfield S, Kovarik A, Chester M, Nichols RA, Macas J, Novak P, Leitch AR. 2012. Independent, rapid and targeted loss of highly repetitive DNA in natural and synthetic allopolyploids of *Nicotiana tabacum*. *PloS one* **7**: e36963.

Sarilar V, Palacios PM, Rousselet A, Ridel C, Falque M, Eber F, Chèvre A-M, Joets J, Brabant P, Alix K. 2013. Allopolyploidy has a moderate impact on restructuring at three contrasting transposable element insertion sites in resynthesized *Brassica napus* allotetraploids. *New Phytologist* **198**: 593-604.

Senerchia N, Felber F, Parisod C. 2014. Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. *New Phytologist* **202**: 975-985.

Senerchia N, Felber F, Parisod C. 2015. Genome reorganization in F1 hybrids uncovers the role of retrotransposons in reproductive isolation. *Proceedings of the Royal Society B - Biological Sciences* **282**: 20142874.

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. *The Plant cell* **13**: 1749-1759.

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.

Soltis DE, Visger CJ, Soltis PS. 2014. The polyploidy revolution then...and now: Stebbins revisited. *American journal of botany* **101**: 1057-1078.

Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. 2015. Polyploidy and genome evolution in plants. *Current opinion in genetics & development* **35**: 119-125.

Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annual review of plant biology* **60**: 561-588.

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.

Szadkowski E, Eber F, Huteau V, Lode M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux MJ, Delourme R, King GJ, et al. 2010. The first meiosis of resynthesized *Brassica napus*, a genome blender. *New Phytologist* **186**: 102-112.

Tenaillon MI, Hollister JD, Gaut BS. 2010. A triptych of the evolution of plant transposable elements. *Trends in plant science* **15**: 471-478.

Vernhettes S, Grandbastien MA, Casacuberta JM. 1998. The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability 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 retrotransposon elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Molecular and General Genetics* **253**: 687-694.

Wendel JF. 2000. Genome evolution in polyploids. *Plant molecular biology* **42**(1): 225-249.

Wenke T, Dobel T, Sorensen TR, Junghans H, Weisshaar B, Schmidt T. 2011. Targeted identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in plant genomes. *The Plant cell* **23**: 3117-3128.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews and Genetics* **8**: 973-982.

Witte CP, Le QH, Bureau T, Kumar A. 2001. Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. *Proceedings of the National Academy of Sciences, USA* **98**: 13778-13783.

Yasui Y, Nasuda S, Matsuoka Y, Kawahara T. 2001. The Au family, a novel short interspersed element (SINE) from *Aegilops umbellulata*. *Theoretical and Applied Genetics* **102**: 463-470.

Yoshioka Y, Matsumoto S, Kojima S, Ohshima K, Okada N, Machida Y. 1993. Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. *Proceedings of the National Academy of Sciences, USA* **90**: 6562-6566.

Figure Legends

Fig. 1 Neighbor-net diagrams of phylogenetic relationships between natural and synthetic *Nicotiana* allotetraploids (a) *N. arentsii*, (b) *N. rustica,* (c) *N. tabacum*, and their corresponding diploid progenitors, based on combined sequence-specific amplification polymorphism (SSAP) profiles obtained for six transposable elements (TEs). (d) Species tree cladogram showing relationships between natural diploid and allotetraploid species, as described in Leitch *et al.* (2008). The genome size of each species (in pg) has been taken from RBG Kew (data.kew.org/cvalues). Accessions of natural and synthetic hybrids are coded as in Tables 2 and 3.

Fig. 2 Evaluation of transposable element (TE)–associated diversity in *Nicotiana* parental accessions. (a) Species tree cladogram showing relationships between natural diploid and allotetraploid species, as described in Leitch *et al.* (2008). (b) Distribution of TE insertions and (c) Shannon index genetic diversity revealed by TE insertions in diploid progenitors. Graphs present means ± SE obtained from the different accessions of each species (see Supporting Information Table S2 for details). Transposable elements are color-coded according to the key on the right: SYL, *N. sylvestris*; UND, *N. undulata*; WIG, *N. wigandioides*; PAN, *N. paniculata*; TOM, *N. tomentosiformis*. SYL and TOM are the *N. tabacum* progenitors, PAN and UND are the *N. rustica* progenitors and UND and WIG are the *N. arentsii* progenitors, as described in Fig. 1(d). SSAP, sequence-specific amplification polymorphism.

Fig. 3 Additive, new and missing sequence-specific amplification polymorphism (SSAP) loci in natural and synthetic *Nicotiana* hybrids. The graph presents mean locus percentages (± SE) obtained with the combined transposable element (TE) set for the different natural accessions of *N. arentsii* (ARE), *N. rustica* (RUS) or *N. tabacum* (TAB), or for the different synthetic hybrids crosses (for nomenclature, see Table 3). Sample sizes are indicated above the panel.

Fig. 4 Distribution of additive, new and missing sequence-specific amplification polymorphism (SSAP) loci in natural and synthetics *Nicotiana* hybrids. (a) Additive loci, (b) missing parental loci, (c) new loci. Each symbol represents the mean percentage of loci for a given transposable element (TE), calculated from all individuals representing each hybrid set. The overall distribution of loci is illustrated by boxplots, with grey boxes delimited by the upper and the lower quartiles and displaying the median (thick black horizontal line). The whiskers (dashed vertical lines) extend to the most extreme values. Transposable element symbols and color codes are indicated on the panel on the right. Significantly different proportions assessed by Tukey tests are shown by different letters on the left of the symbols. *, No statistical tests possible (number of accessions/lines < 3). ARE, *N. arentsii*; RUS, *N. rustica*; TAB, *N. tabacum*; for nomenclature of synthetic hybrids, see Table 3.

Fig. 5 Parental origin of missing loci in synthetic and natural *Nicotiana* hybrids. (a) Mean global percentage (± SE) of parental missing loci for the six transposable elements (TEs). Hybrids represented by only one accession are labelled *. (b) Distribution of mean percentages of maternal and paternal missing loci for each TE family. The overall distribution of missing parental loci is illustrated by boxplots, with grey boxes delimited by the upper and the lower quartiles and displaying the median (thick black horizontal line). The whiskers (dashed vertical lines) extend to the most extreme values. Transposable element symbols and color codes are indicated in the key on the right. ML-mat, loci of maternal origin; ML-pat, loci of paternal origin. ARE, *N. arentsii*; RUS, *N. rustica*; TAB, *N. tabacum*; for nomenclature of synthetic hybrids, see Table 3.

Fig. 6 Association of the transposable element (TE)-related divergence between progenitor species and TE-associated restructuring in the three recent natural *Nicotiana* allotetraploids. General linear mixed model associating proportions of non-shared loci between progenitors with (a) the proportion of new loci, (b) the proportion of missing parental loci. ARE, *N. arentsii*; RUS, *N. rustica*; TAB, *N. tabacum*. For better clarity, individual TE families within each species were not coded.

Supporting Information Fig. S1: Neighbor-net diagrams obtained from SSAP profiles of each transposable element family.

Supporting Information Table S1: SSAP matrices produced in this work for each transposable element and assessment of missing loci shared across hybrids.

Supporting Information Table S2: Recorded SSAP loci and diversity calculated for each transposable element.

Supporting Information Table S3: SSAP locus frequencies and fate in hybrids.

Supporting Information Notes S1: R Script used to calculate SSAP loci frequencies in *Nicotiana* allopolyploids

Table 1 Transposable elements investigated in the present study.

Name	Type	Characteristics	References
Au	SINE[a]	ancient, shared among eudicots and monocots	(Witte *et al.*, 2001; Yasui *et al.*, 2001)
TS	SINE[a]	recently amplified in *N. tabacum*, putatively active	(Yoshioka *et al.*, 1993; Wenke *et al.*, 2011)
Ns1	*Stowaway* MITE[b]	only sequence data available	(Bureau & Wessler, 1994)
Tnt1	*copia* LTR-RT[c]	limited amplification in response to allopolyploidy	(Petit *et al.*, 2007; Petit *et al.*, 2010)
Tnt2	*copia* LTR-RT[c]	recently amplified in *N. tabacum*, putatively active	(Petit *et al.*, 2007)
TRIM[d]	LTR-RT[c]	ancient, shared among eudicots and monocots, putatively inactive	(Witte *et al.*, 2001; Yasui *et al.*, 2001)

[a]Short interspersed nuclear element; [b]miniature inverted-repeat transposable element; [c]LTR retrotransposon; [d]terminal-repeat retrotransposon in miniature.

Table 2 *Nicotiana* accessions investigated in the present study.

Section	Ploidy	Species (Acronym)	Reference	Code	Source
Tomentosae	2n	*N. tomentosiformis* (TOM)	Nee *et al.* 51771	tom1	New York Botanic Garden, New York (US)[a,b]
	2n	*N. tomentosiformis* (TOM)	ITG647	tom2	Imperial Brands, Bergerac (F)[c]
	2n	*N. tomentosiformis* (TOM)	TW142	tom3	USDA (US)[d]
	2n	*N. tomentosiformis* (TOM)	ITG645	tom4	Imperial Brands, Bergerac (F)[c]
	2n	*N. tomentosiformis* (TOM)	ITG646	tom5	Imperial Brands, Bergerac (F)[c]
	2n	*N. tomentosiformis* (TOM)	Nic479/84	tom6	IPK, Gatersleben (D)[e]
Sylvestres	4n	*N. sylvestris* (SYL)	B39 (4n)	syl1	provided by A. Kovarik, Institute of Biophysics, Brno (CZ)
	2n	*N. sylvestris* (SYL)	TW136	syl2	USDA (US)[d]
	2n	*N. sylvestris* (SYL)	TW137	syl3	USDA (US)[d]
	2n	*N. sylvestris* (SYL)	TW138	syl4	USDA (US)[d]
	2n	*N. sylvestris* (SYL)	Ducrettet	syl5	"Only the lonely", Ducrettet horticultural seeds [g]
	2n	*N. sylvestris* (SYL)	ITG626	syl6	Imperial Brands, Bergerac (F)[c]
	2n	*N. sylvestris* (SYL)	A04750326	syl7	Nijmegen Botanical Garden (NL)[h,i]
	2n	*N. sylvestris* (SYL)	934750319	syl8	Nijmegen Botanical Garden (NL)[h]
Paniculatae	2n	*N. paniculata* (PAN)	TW99	pan1	USDA (US)[d]
	2n	*N. paniculata* (PAN)	ITG496	pan2	Imperial Brands, Bergerac (F)[c]
	2n	*N. paniculata* (PAN)	Nic435/84	pan3	IPK, Gatersleben (D)[e]
	2n	*N. paniculata* (PAN)	A34750324	pan4	Nijmegen Botanical Garden (NL)[h]
Undulatae	2n	*N. undulata* (UND)	TW145	und1	USDA (US)[d]
	2n	*N. undulata* (UND)	ITG628	und2	Imperial Brands, Bergerac (F)[c]
	2n	*N. undulata* (UND)	TW147 (PI306637)	und3	USDA (US)[d]
	2n	*N. undulata* (UND)	Nee *et al.* 51816	und4	New York Botanic Garden, New York (US)[a]
	2n	*N. wigandioides* (WIG)	Nee *et al.* 51764	wig	New York Botanical Gardens, New York (US)[a]

	4n	*N. arentsii* (ARE)	TW12	are	USDA (US)[d]
Genuinae	4n	*N. tabacum* (TAB)	Alipes (ITG383)	tab1	Imperial Brands, Bergerac (F)[c]
Tabacum	4n	*N. tabacum* (TAB)	Calycina (ITG388)	tab2	Imperial Brands, Bergerac (F)[c]
	4n	*N. tabacum* (TAB)	Chinensis (ITG389)	tab3	Imperial Brands, Bergerac (F)[c]
	4n	*N. tabacum* (TAB)	Fructicosa (ITG981)	tab4	Imperial Brands, Bergerac (F)[c]
	4n	*N. tabacum* (TAB)	Lacerata (ITG982)	tab5	Imperial Brands, Bergerac (F)[c]
	4n	*N. tabacum* (TAB)	Petiolaris (ITG985)	tab6	Imperial Brands, Bergerac (F)[c]
	4n	*N. tabacum* (TAB)	Purpurea B (ITG986)	tab7	Imperial Brands, Bergerac (F)[c]
	4n	*N. tabacum* (TAB)	Ambalema (ITG384)	tab8	Imperial Brands, Bergerac (F)[c]
	4n	*N. tabacum* (TAB)	Bolivie	tab9	Feral tobacco collected in Bolivia[f]
	4n	*N. tabacum* (TAB)	Chulumani	tab10	Feral tobacco collected in Bolivia[f]
Rusticae	4n	*N. rustica* (RUS)	Pavonii (Nic640/75)	rus1	IPK, Gatersleben (D)[e]
	4n	*N. rustica* (RUS)	Texana (Nic613/77)	rus2	IPK, Gatersleben (D)[e]
	4n	*N. rustica* (RUS)	Texana F (ITG507)	rus3	Imperial Brands, Bergerac (F)[c]
	4n	*N. rustica* (RUS)	Asiatica (Nic616/47)	rus4	IPK, Gatersleben (D)[e]
	4n	*N. rustica* (RUS)	Brasilia (Nic59/78)	rus5	IPK, Gatersleben (D)[e]
	4n	*N. rustica* (RUS)	Brasilia g (ITG481)	rus6	Imperial Brands, Bergerac (F)[c]
	4n	*N. rustica* (RUS)	Major Alef (Nic55/75)	rus7	IPK, Gatersleben (D)[e]
	4n	*N. rustica* (RUS)	Cordata (Nic604/76)	rus8	IPK, Gatersleben (D)[e]
	4n	*N. rustica* (RUS)	Cerinthoides (ITG482)	rus9	Imperial Brands, Bergerac (F)[c]
	4n	*N. rustica* (RUS)	Souffii Gabes (ITG504)	rus10	Imperial Brands, Bergerac (F)[c]
	4n	*N. rustica* (RUS)	Zlag 221 (ITG510)	rus11	Imperial Brands, Bergerac (F)[c]

[a]http://www.nybg.org; [b]obtained from J. J. Clarkson, Royal Botanic Garden, Kew (UK); [c]http://www.imperialbrandsscience.com/en/our-science/plant-science.html; [d]http://www.ars-grin.gov/; [e]http://gbis.ipk-gatersleben.de; [f]obtained from S. Knapp (Natural History Museum, London, UK); [g]http://www.ducrettet.com; [h]http://www.bgard.science.ru.nl/zapp/search/index/search/; [i]provided by Dr P Maliga, Rutgers University, NJ, USA.

Table 3 *Nicotiana* synthetic hybrids investigated in the present study.

Cross (maternal x paternal)	Cross Name	Code	Nature	Source
N. sylvestris x N. tomentosiformis	ST-Th-S4	Th1	4n S4 progeny from Th37[a]	USDA (US)[b]; Burk, 1973
		Th3		
		Th8		
		Th9		
	ST-H-F1	H1	2n F1 hybrids between *N. sylvestris* TW137 (syl3) and *N. tomentosiformis* ITG645 (tom4)	This study
		H15		
		H39		
	ST-H-S0	H15-2	4n S0 hybrids obtained from the H15 F1 hybrid	This study
		H15-32		
		H15-70		
	ST-D-S0	D1 to D10	Ten 4n S0 hybrids from a cross between doubled *N. tomentosiformis* issued from TW142 (tom3) and doubled B39 *N. sylvestris* (syl1)	This study
N. undulata x N. wigandioides	UW-F1	UW	2n F1 hybrid between *N. undulata* TW145 (und1) and *N. wigandioides* Nee *et al*. 51764 (wig)	Dadejova *et al*, 2007
N. undulata x N. paniculata	UP-F1	UP1	2n F1 hybrids between *N. undulata* TW145 (und1) and *N. paniculata* TW99 (pan1)	Dadejova *et al*, 2007
		UP2		
N. paniculata x N. undulata	PU-F1	PU1	2n F1 hybrid between *N. paniculata* TW99 (pan1) and *N. undulata* TW145 (und1)	Dadejova *et al*, 2007
	PU-S0	PU1-1	4n S0 hybrids derived from PU1 hybrid	This study
		PU1-2		

[a]S4 progeny described in Skalicka *et al*. (2003, 2005), Kovarik *et al*. (2004), Lim *et al.* (2004) and Petit *et al*. (2010); [b]http://www.ars-grin.gov/

Table 4 Distribution of sequence-specific amplification polymorphism (SSAP) loci from the different transposable elements (TEs) among the progenitor diploid taxa of the Tabacum, Rustica and Arentsii groups (SYL, *Nicotiana sylvestris*; UND, *N. undulata*; WIG, *N. wigandioides*; PAN, *N. paniculata*; TOM, *N. tomentosiformis*).

TE	TOT[a]	MAT-tot[b]	PAT-tot[b]	MAT/PAT[c]	MAT-spe[d]	PAT-spe[d]	MAT-spe vs. PAT-spe (Chi-Square)[e]
ARENTSII group progenitors (MAT = UND ; PAT = WIG)							
Au	59.00	38.25	36.00	15.25 (25.8%)	23.00 (59.7%)	20.75 (57.6%)	NS
TS	103.00	47.25	76.00	20.25 (19.7%)	27.00 (57.1%)	55.75 (73.4%)	MAT < PAT : 18.61***
Ns1	99.00	51.75	71.00	23.75 (24.0%)	28.00 (54.1%)	47.25 (66.5%)	MAT < PAT : 8.85***
Tnt1	36.25	15.25	26.00	5.00 (13.8 %)	10.25 (67.2%)	21.00 (80.8%)	MAT < PAT : 6.08**
Tnt2	27.50	10.75	21.00	4.25 (15.5%)	6.50 (60.5%)	16.75 (79.8%)	MAT <PAT : 7.36**
TRIM	52.75	31.75	40.00	19.00 (36.0%)	12.75 (40.2%)	21.00 (52.5%)	NS
Total	***377.50***	***195.00***	***270.00***	***87.50 (23.2%)***	***107.50 (55.1%)***	***182.50 (67.6%)***	**MAT < PAT : 38.28*** **
RUSTICA group progenitors (MAT = PAN ; PAT = UND)							
Au	58.94	31.00	38.25	10.31 (17.5%)	20.69 (66.7%)	27.94 (73.0%)	NS
TS	53.25	11.00	47.25	5.00 (9.4%)	6.00 (54.5%)	42.25 (89.4%)	MAT < PAT : 51.50***
Ns1	82.00	43.25	51.75	13.00 (15.9%)	30.25 (69.9%)	38.75 (74.9%)	NS
Tnt1	33.50	21.00	15.25	2.75 (8.2%)	18.25 (86.9%)	12.5 (82.0%)	NS
Tnt2	18.69	12.25	10.75	4.31 (23.1%)	7.94 (64.8%)	6.44 (59.9%)	NS
TRIM	47.00	27.75	31.75	12.5 (26.6%)	15.25 (55.0%)	19.25 (60.6%)	NS
Total	***293.38***	***146.25***	***195.00***	***47.88 (16.3%)***	***98.38 (67.3%)***	***147.13 (75.5%)***	***MAT < PAT : 18.57 *** ***
TABACUM group progenitors (MAT = SYL ; PAT = TOM)							

Au	65.08	39.75	33.66	8.33 (12.8%)	31.42 (79.0%)	25.33 (75.3%)	NS
TS	68.42	31.75	45.00	8.33 (12.2%)	23.42 (73.8%)	36.67 (81.5%)	MAT > PAT : 4.99**
Ns1	70.80	45.01	36.67	10.88 (15.4%)	34.13 (75.8%)	25.79 (70.3%)	NS
Tnt1	68.64	48.37	28.33	8.06 (11.7%)	40.31 (83.3%)	20.27 (71.5%)	MAT > PAT : 11.97 **
Tnt2	82.97	63.51	22.84	3.38 (4.1%)	60.13 (94.7%)	19.46 (85.2%)	MAT > PAT : 39.54***
TRIM	61.58	38.75	34.16	11.33 (18.4%)	27.42 (70.8%)	22.83 (66.8%)	NS
Total	***417.49***	***267.14***	***200.66***	***50.31 (12.1%)***	***216.81 (81.2%)***	***150.35 (74.9%)***	***MAT > PAT: 23.34******

[a]Total cumulated probabilities (Shared and parental-specific) for SSAP bands presence in both parental accessions (TOT= « MAT/PAT »+ « MAT-spe » + « PAT-spe »).

[b]Total pondered SSAP bands number for paternal (PAT) or maternal (MAT) parental taxa.

[c]Pondered SSAP bands number shared between parental taxa. Percentages were calculated for each TE relative to the TOT column.

[d]Pondered SSAP bands number restricted to maternal (MAT) or paternal (PAT) taxa. Percentages were calculated for each TE relative to either MAT-tot or PAT-tot columns.

[e]Comparison of parental-specific SSAP bands numbers (Yate's one-sided chi-square tests). NS, non significant; X>Y, significantly higher proportion of SSAP bands specific to X as compared to Y-specific bands; ***, significant at α=0.005; **, significant at α=0.05.