the plant journal

SEB Society for Experimental Biology

*The Plant Journal* (2015) **82,** 621–631 doi: 10.1111/tpj.12837

# Evolutionary dynamics of retrotransposons following autopolyploidy in the Buckler Mustard species complex

**Amélie Bardil, Alexandre Tayalé and Christian Parisod***

*Laboratory of Evolutionary Botany, University of Neuchâtel, Neuchâtel 2000, Switzerland*

Received 7 November 2014; revised 18 March 2015; accepted 23 March 2015; published online 30 March 2015.
*For correspondence (e-mail christian.parisod@unine.ch).

## SUMMARY

**Long terminal repeat retrotransposons (LTR-RTs) represent a major fraction of plant genomes, but processes leading to transposition bursts remain elusive. Polyploidy expectedly leads to LTR-RT proliferation, as the merging of divergent diploids provokes a genome shock activating LTR-RTs and/or genetic redundancy supports the accumulation of active LTR-RTs through relaxation of selective constraints. Available evidence supports interspecific hybridization as the main trigger of genome dynamics, but few studies have addressed the consequences of intraspecific polyploidy (i.e. autopolyploidy), where the genome shock is expectedly minimized. The dynamics of LTR-RTs was thus here evaluated through low coverage 454 sequencing of three closely related diploid progenitors and three independent autotetraploids from the young *Biscutella laevigata* species complex. Genomes from this early diverging Brassicaceae lineage presented a minimum of 40% repeats and a large diversity of transposable elements. Differential abundances and patterns of sequence divergence among genomes for 37 LTR-RT families revealed contrasted dynamics during species diversification. Quiescent LTR-RT families with limited genetic variation among genomes were distinguished from active families (37.8%) having proliferated in specific taxa. Specific families proliferated in autopolyploids only, but most transpositionally active families in polyploids were also differentiated among diploids. Low expression levels of transpositionally active LTR-RT families in autopolyploids further supported that genome shock and redundancy are non-mutually exclusive triggers of LTR-RT proliferation. Although reputed stable, autopolyploid genomes show LTR-RT fractions presenting analogies with polyploids between widely divergent genomes.**

**Keywords: autopolyploidy, *Biscutella laevigata*, Brassicaceae, high-throughput sequencing, molecular population genetics, retrotransposon expression levels, transposable elements, whole genome snapshot.**

## INTRODUCTION

Plant genomes are remarkably dynamic, probably in relation to their distinctive propensity for retrotransposition, recombination, genome silencing and polyploidy (Kejnovsky *et al.*, 2009). Transposable elements (TEs) play a pivotal role fuelling such genome restructuring and functional reorganization (Bennetzen, 2005; Biemont and Vieira, 2006). In particular, long terminal repeat retrotransposons (LTR-RTs), amplifying through a copy and paste mechanism involving RNA intermediates, represent the major fraction of most plant genomes and shape their architecture (Fedoroff, 2012; Lisch, 2013). Insertion of TEs indeed changes local genome structure and may further interfere with the expression of nearby genes (Lockton and Gaut, 2009). Accordingly, active TEs have mostly deleterious effects on their host, but specific insertions can rapidly go to fixation when beneficial (Gaut and Ross Ibarra, 2008; Hua-Van *et al.*, 2011). Although TE dynamics produces abundant raw material for evolution in natural populations, its overall impact remains elusive and deserves attention (Doolittle, 2013).

Barbara McClintock (1984) postulated that TE activation plays a significant role in genome changes under evolutionary challenges. Many plant LTR-RTs are accordingly activated in relation to stress (Grandbastien *et al.*, 2005), including hybridization and polyploidy (Parisod *et al.*, 2010a). Polyploid genomes can undergo structural and epigenetic changes quickly after their formation, and empirical evidence suggests that TEs play an influential role in their reorganization (reviewed in Doyle *et al.*, 2008; Freeling *et al.*, 2012; Parisod and Senerchia, 2012). Polyploidy is expected to support TE amplification, because: (i) the merging of diploid progenitors having differentially accu-

mulated TEs provokes their mobilization due to loose epigenetic silencing in the naive hybrid cytoplasm (genome shock hypothesis; Bourc'his and Voinnet, 2010; Josefsson *et al.*, 2006); and (ii) redundancy may relax selective constraints on duplicated gene copies, offering new neutral sites for transcriptionally active TEs to insert and accumulate until the genome is fully diploidized (redundancy hypothesis; Matzke and Matzke, 1998).

To date, TE dynamics was mostly addressed in allopolyploids resulting from hybridization between distantly related genomes and data largely supported the genome shock hypothesis (Comai *et al.*, 2003; Parisod *et al.*, 2010a; Senerchia *et al.*, 2014). Available data indicate considerable genome reorganization, matching expectations of conflict resolution in hybrids derived from genetically divergent parents (reviewed in Parisod and Senerchia, 2012; Tayale and Parisod, 2013). However, few autopolyploids (i.e. with closely related homologous genomes) were examined (reviewed in Parisod *et al.*, 2010b) and almost no study focused on the response of TEs to autopolyploidy (but see Madlung *et al.*, 2005). As the genome shock is expectedly minimized in young autopolyploids originating from slightly divergent progenitors (i.e. typically within species), TE dynamics is predicted to yield limited genome reorganization in such species (Parisod and Senerchia, 2012).

The young Buckler Mustard autopolyploid complex (*Biscutella laevigata;* Brassicaceae) thus represents a good model system to address the underpinnings of TE dynamics in such polyploids having originated within a single species (Manton, 1937). This textbook example includes diploid relict lineages in never glaciated areas of central Europe that hybridized and generated independent autotetraploid lineages having recolonized the Alps after the last glacial maximum (Stebbins, 1971; Raffaelli and Baldoin, 1997; Tremetsberger *et al.*, 2002; Parisod and Besnard, 2007). Accordingly, the diploids (2n = 2x = 18) *B. laevigata* subsp. *austriaca* (A2), *B. brevicaulis* (B2) and *B. laevigata* subsp. *varia* (V2) are representative taxa having contributed to the evolution of different autotetraploids (2n = 4x = 36): *B. laevigata* subsp. *laevigata* (La4), *B. laevigata* subsp. *ossolana* (O4) and *B. laevigata* subsp. *lucida* (Lu4). Here, we examine the genome content and dynamics of LTR-RT families in those six key taxa (thereafter, collectively referred to as *Biscutella*) using high-throughput sequencing. In particular, we address to what extent LTR-RT families have been transpositionally active in diploids, in autopolyploids or in both. Following the rationale that proliferation of LTR-RTs generates closely related sequences with transposition bursts resulting in populations of highly similar TE sequences (Casacuberta *et al.*, 1997), active and quiescent TEs can be distinguished by patterns of sequence variation among individual copies (i.e. insertions at particular chromosomal locations; Le Rouzic *et al.*, 2013; Senerchia *et al.*, 2013). Our detailed survey of 37 LTR-RT families shows that about a third proliferated in autopolyploids and indicated drastic reorganization of TE genome fractions analogous to that otherwise reported in allopolyploids, prompting further evaluation of the underpinnings of TE accumulation in polyploids.

## RESULTS

### Genome content

Low coverage sequencing of the six divergent *Biscutella* genomes using Roche 454 produced a total of 1 385 927 reads (mean length of 367.9 bp). After removal of low complexity reads, a mean of 6.2 and 37.2% of all reads were respectively classified as organelle and coding sequences through BLASTN and TBLASTX against Brassicaceae references (Table 1). Then a mean of 33.4% of reads corresponding to TEs were similarly identified using a custom Brassicaceae TE reference. Reads matched coding regions of TE references well, but consistently showed low coverage in fast evolving regions such as terminal repeats. Such heterogeneous coverage is indicative of significant divergence from other Brassicaceae TEs and called for *de novo* inference of abundant TEs from *Biscutella.*

The assisted automated assembler of repeat families algorithm (AAARF; DeBarry *et al.*, 2008) was used to overlap and walk out reads, producing contigs representative of abundant repeated elements. Such *de novo* assembly resulted in a variable number of consensus repeats in the different genomes (921 for A2, 3799 for B2, 2115 for V2, 1856 for Lu4, 3361 for La4 and 4588 for O4). After removal of low complexity contigs (i.e. <3%), reads matching either assembled sequences of repeated elements in *Biscutella* or our custom Brassicaceae TE reference (BLASTN and TBLASTX) highlighted up to 44% of repeats within a given genome (Table 1). Retrieved reads evenly mapped along *de novo* TE references and, as compared with BLAST of Brassicaceae TE reference, this approach relying on species-specific TEs thus considerably improved the sampling of reads representative of genomic TE insertions.

### Classification of TE families

Consensus sequences of abundant repeats showing higher than 80% identity were clustered into families using CD-HIT-EST (Li and Godzik, 2006) according to the criteria of Wicker *et al.* (2007). Among the 387 families consistently assembled in at least four genomes (i.e. testifying nonartefactual inferences) and clearly identified by similarity with Brassicaceae TE references, LTR-RTs were the most abundant, with a mean of 86% of TE genome fractions (Figure S1). Among the 99 LTR-RT families satisfying our criteria (67 Gypsy, 17 Copia and 15 unclassified LRT-RTs), several contigs matched with ALYGYPSY36 from *Arabidopsis lyrata,* indicating diversification of this LTR-RT family in the genome of *Biscutella.* Similarly, families of the AT-

**Table 1** Composition of six *Biscutella* genomes based on low coverage 454 sequencing

| | A2[a] | B2[a] | V2[a] | Lu4[a] | O4[a] | La4[a] |
|---|---|---|---|---|---|---|
| Total number of 454 reads | 142212 | 306457 | 167433 | 145818 | 368512 | 255525 |
| Mean size of 454 reads (bp) | 368.9 | 375.6 | 376.1 | 365.8 | 362.8 | 358.0 |
| SSR[b] (%) | 0.06 | 0.11 | 0.06 | 0.05 | 0.09 | 0.07 |
| Organelle[c] (%) | 9.87 | 5.45 | 5.08 | 6.30 | 6.13 | 4.45 |
| Coding sequences (%) | 36.76 | 35.41 | 37.61 | 37.99 | 37.29 | 38.00 |
| Repeats (incl. TEs, %)[d] | 33.26 | 44.50 | 42.01 | 39.37 | 42.11 | 43.28 |
| Unclassified[e] (%) | 20.20 | 14.53 | 15.24 | 16.29 | 14.38 | 14.20 |

[a]Diploid taxa (2n = 2x = 18) were represented by *B. brevicaulis* (B2), *B. laevigata* subsp. *austriaca* (A2) and *B. laevigata* subsp. *varia* (V2), whereas autotetraploid taxa (2n = 4x = 36) were represented by *B. laevigata* subsp. *ossolana* (O4), *B. laevigata* subsp. *laevigata* (La4) and *B. laevigata* subsp. *lucida* (Lu4).
[b]Proportion of total reads corresponding to simple sequence repeats (low complexity).
[c]Proportion of total reads corresponding to plastid and mitochondrial sequences.
[d]Proportion of total reads corresponding to consensus sequences of abundant repeated elements inferred with the assisted automated assembler of repeat families (AAARF) and our custom Brassicaceae TE reference.
[e]Proportion of total reads remaining unclassified following BLAST against the above references.

HILA and the ATGP clades, but also of ALYGYPSY25 or EPAGYPSY9, apparently diversified in *Biscutella.* Some families related to ALYGYPSY36 or ATHILA6 diverged from canonical references just below the 80% similarity threshold and were thus annotated with a '-like' to account for possible relatedness.

In order to infer the evolutionary dynamics of LTR-RT families through patterns of genetic variation among individual copies (Table 2), the end of the LTR sequence was selected as a region offering suitable resolution (Senerchia *et al.*, 2013). Accordingly, LTR regions were manually identified in 37 LTR-RT families through motif recognition and read coverage following DeBarry *et al.* (2008). The remaining LTR-RT families were not further analyzed, as AAARF inferred internal regions only or resulted in ambiguously identified LTR regions.

### Abundance of LTR-RT families

Proportions of reads matching those 37 LTR-RT families were normalized by the total read number within each genome to account for their genomic abundance (Figure 1 and Table S1). These 37 LTR-RT families accounted for more than 20% of *Biscutella* genomes. However, only six families (ATHILA3A, ATHILA3B, ATHILA3C, ATHILA5B ATHILA6B-like and ATGP5A) presented abundance higher than 1% of the genome. Abundances were often non-significantly different among diploid and autotetraploid genomes, but 13 LTR-RT families (35.1%) presented particular genomes with significantly different abundances suggestive of genome-specific dynamics.

### Phylogenetic analysis of LTR-RT families

BLASTN retrieved a sufficient amount of individual reads (i.e. around a minimum of 50) matching the 5′ or the 3′-end of the LTR region for each of the 37 families (Tables 2 and S1). Phylogenetic relationships among those reads representative of individual copies from the six genomes were thus inferred by simultaneous alignment and tree estimation using SATé I (Liu *et al.*, 2009; Figure S2). Families such as ALYGYPSY36B-like or EPAGYPSY9A presented apparent clades of insertions matching with diploid and polyploid taxa to a large extent, indicating proliferation of corresponding LTR-RTs after genome differentiation (Figure 2a, b). Some families presented a similar pattern, but with clades of insertions restricted to either diploid or polyploid taxa, indicating ploidy-specific proliferation (e.g. ALYGYPSY36C in autotetraploids; Figure 2c). Finally, most families presented phylogenetic trees with related insertions randomly shared among taxa, indicating proliferation before genome differentiation (e.g. ATLANTYS2; Figure 2d).

Following Le Rouzic *et al.* (2013), $\beta$ index estimated imbalance of phylogenetic trees and varied among LTR-RT families, indicating differential transposition rates (Tables 2 and S1). Imbalanced trees ($-2 < \beta < -1.5$) were evidenced for 12 LTR-RT families (35%), although only two (ATGP5A and ALYGYPSY36B-like) presented $\beta$ significantly lower than $-1.5$ indicative of one copy giving birth to most daughter copies (i.e. 'master copy model'; Brookfield and Johnson, 2006; Figure 2a). The remaining 25 LTR-RT families revealed random balanced phylogenetic trees ($-1.5 \leq \beta \leq 0$), including eight with a $\beta$ significantly higher than $-1.5$ indicating that all copies duplicated at the same rate (i.e. following the 'transposon model'; Figure 2b).

### Population genetics of LTR-RT families

As expected for intraspecific taxa, hierarchical analysis of molecular variance (Excoffier *et al.*, 1992) revealed low genetic differentiation of LTR-RT copies among genomes ($F_{ST} < 0.077$). Genetic structure among taxa was significantly higher than zero for 14 LTR-RT families (Tables 2 and S1), indicating that closely related copies tend to occur within taxa. Those families also showed taxon-specific

**Table 2** Dynamics of 37 long terminal repeat retrotransposon (LTR-RT) families inferred from phylogenetics and population genetics in six *Biscutella* genomes

| LTR–RT family | L | N | β[a] | I/R[a] | $F_{ST}$[b] | A2-B2[b] | A2-V2[b] | B2-V2[b] | O4-La4[b] | O4-Lu4[b] | Lu4-La4[b] | D/T[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ATGP1A | 278 | 188 | −1.51 | I | 0.0767* | * | – | – | * | * | – | D/T |
| EPAGYPSY9A | 280 | 245 | −1.31* | R | 0.0724* | * | * | – | – | * | – | D/T |
| ALYGYPSY25A | 299 | 189 | −1.61 | I | 0.0687* | * | * | * | – | – | – | D |
| ALYCOPIA64 | 240 | 84 | −1.67 | I | 0.0674* | * | – | * | – | * | * | D/T |
| ALYGYPSY25B | 166 | 46 | −1.29 | R | 0.0625 | – | – | – | – | – | – | – |
| ATHILA5A | 299 | 113 | −1.59 | I | 0.0552* | – | – | – | * | * | – | T |
| ALYGYPSY2 | 230 | 68 | −1.56 | I | 0.0531* | – | * | * | * | – | – | D/T |
| ALYGYPSY36A-like | 289 | 273 | −1.32* | R | 0.0510* | * | * | * | – | * | – | D/T |
| ALYGYPSY36B-like | 225 | 475 | −1.62* | I | 0.0501* | * | * | * | * | * | – | D/T |
| ATGP5B | 249 | 128 | −1.41 | R | 0.0498* | – | – | * | – | – | – | D |
| ALYGYPSY36C | 299 | 145 | −1.53 | I | 0.0413* | – | – | – | – | * | * | T |
| ATHILA6B-like | 227 | 73 | −1.21* | R | 0.0413* | – | * | * | – | – | – | D |
| ATGP5 | 287 | 106 | −1.26* | R | 0.0353* | * | * | – | – | – | – | D |
| ATCOPIA95 | 198 | 57 | −1.53 | I | 0.0289 | – | – | – | – | – | – | – |
| ATGP5A | 299 | 805 | −1.68* | I | 0.0248* | * | – | * | – | * | – | D/T |
| ALYGYPSY36D | 299 | 100 | −1.27 | R | 0.0207 | – | – | – | – | – | – | – |
| ATHILA5B | 288 | 146 | −1.53 | I | 0.0180 | – | – | – | – | – | – | – |
| ATHILA3C | 287 | 155 | −1.48 | R | 0.0137 | - | – | * | – | – | – | – |
| ATHILA3B | 299 | 116 | −1.20* | R | 0.0126 | – | – | – | – | – | * | – |
| ATLANTYS2 | 290 | 96 | −1.44 | R | 0.0115 | – | – | – | – | – | – | – |
| ATHILA3A | 290 | 168 | −1.19* | R | 0.0104 | – | * | * | – | * | – | – |
| ATGP5C | 290 | 128 | −1.37 | R | 0.0084 | – | – | – | – | – | – | – |
| ALYGYPSY36E | 299 | 169 | −1.40 | R | 0.0080 | – | – | – | – | – | – | – |
| ATHILA4-like | 284 | 703 | −1.29* | R | 0.0068* | * | * | * | * | – | – | D/T |
| ALYGYPSY36F | 161 | 43 | −1.66 | I | 0.0054 | – | – | – | – | – | – | – |
| ATHILA5C | 246 | 163 | −1.50 | R | 0.0029 | – | – | – | – | – | – | – |
| ALYGYPSY36H | 200 | 92 | −1.25 | R | 0.0023 | – | * | – | – | – | – | – |
| ALYGYPSY36G | 299 | 126 | −1.34 | R | 0.0023 | – | – | – | – | – | – | – |
| ATHILA6A-like | 231 | 732 | −1.45 | R | 0.0022 | – | * | – | – | – | – | – |
| ALYGYPSY36I | 288 | 188 | −1.43 | R | 0.0013 | – | – | – | – | – | – | – |
| ATHILA5D | 289 | 373 | −1.43 | R | 0.0009 | – | – | – | – | – | – | – |
| GMCOPIA99 | 180 | 47 | −1.31 | R | 0 | – | – | – | – | – | – | – |
| ALYGYPSY36J | 171 | 50 | −1.49 | R | 0 | – | – | – | – | – | – | – |
| ATGP1B | 189 | 47 | −1.29 | R | 0 | – | – | – | – | – | – | – |
| ATHILA5E | 290 | 119 | −1.30 | R | 0 | – | – | – | – | – | – | – |
| EPAGYPSY9B | 160 | 75 | −0.99* | R | 0 | – | – | – | – | – | – | – |
| ATHILA6C-like | 150 | 42 | −1.60 | I | 0 | – | – | – | – | – | – | – |

Genomes are coded as in Table 1. Inferences are based on the alignment of N reads corresponding to individual LTR-RT copies over L base pairs.

* = significant; – = non-significant.

[a]Imbalance index (β) distinguishes families showing imbalanced trees (I; $-2 < \beta < -1.5$) or balanced random trees (R; $-1.5 \leq \beta \leq 0$).

[b]Genetic differentiation ($F_{ST}$) of LTR-RT copies among genomes (i.e. populations). Pair-wise $F_{ST}$ between diploid and between tetraploid populations are presented, and can be significant among diploids only (D), among tetraploids only (T) or among both diploid and tetraploid taxa (D/T).

clades of TE copies (Figure S2), and thus likely proliferated and differentiated after the split of taxa. In contrast, $F_{ST}$ was non-significantly different from zero for the remaining 23 LTR-RT families (Table 2), showing that closely related copies are randomly shared among taxa and that corresponding TEs were quiescent during diversification of *Biscutella* genomes. The genetic structure of LTR-RT populations within ploidy levels ($F_{SC}$) showed a similar pattern, whereas copies of diploids and independent autotetraploids ($F_{CT}$) were never significantly differentiated (Table S1), pointing to genome-specific rather than consistent ploidy-specific dynamics of LTR-RT families.

Among the 14 LTR-RT families showing a significant $F_{ST}$, eight revealed significant pair-wise $F_{ST}$ among both diploid and tetraploid populations. These eight families included the ALYGYPSY36 group (ALYGYPSY36A-like and ALYGYPSY36B-like), members of the ATGP clade (ATGP1A and ATGP5A) as well as ATHILA4-like, ALYGYPSY2, EPAGYPSY9A and ALYCOPIA64 (Table 2). Four LTR-RT families showed significant pair-wise $F_{ST}$ among diploids only (ATGP5 and ATGP5B, ALYGYPSY25A and ATHILA6B-like), indicating dynamics during the differentiation of diploid taxa, but quiescence in polyploids. In contrast, two LTR-RT families showed significant pair-wise $F_{ST}$ among

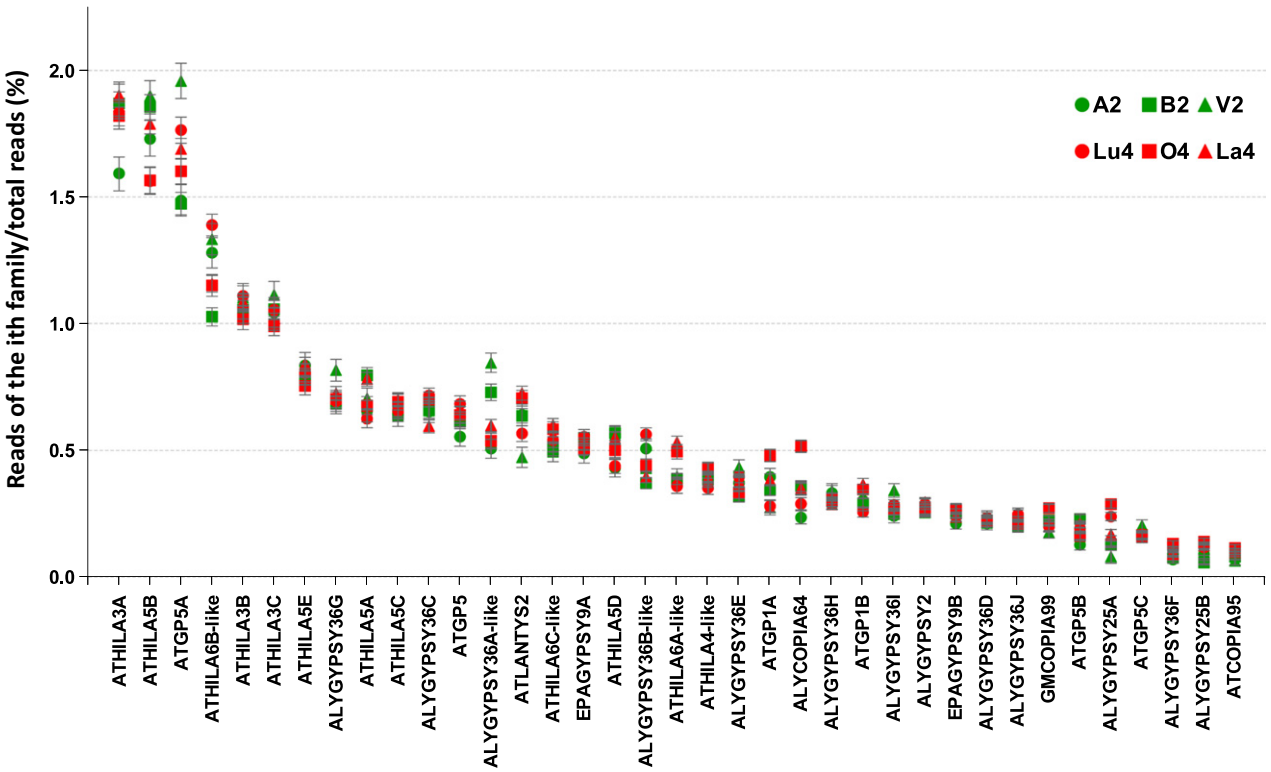


**Figure 1.** Abundance of 37 long terminal repeat retrotransposon (LTR-RT) families estimated by 454 low coverage sequencing of six *Biscutella* genomes. Genomes are coded as in Table 1. Abundance estimated as proportions of matching reads out of the total read number; 95% confidence intervals estimated by resampling. Diploid and autotetraploid genomes are in green and red, respectively.

autotetraploid populations only (ATHILA5A and ALYGYPSY36C), highlighting polyploid-specific dynamics (Table 2 and Figure S1).

### RNA-Seq and expression of LTR-RT families

More than 75 millions SOLiD reads corresponding to polyadenylated transcripts from leaf and root tissues of two autopolyploid individuals from the *B. laevigata* subsp. *laevigata* (La4) showed that all LTR-RT families, except four (ALYGYPSY2 and ALYGYPSY36J, EPAGYSPSY9A and ATGP1B), were expressed (Figure 3 and Table S2). Housekeeping genes (Tong *et al.*, 2009) showed here constant expression levels in both tissues, ranging from 1.78 to 2.10 for polyubiquitin (UBQ10) and 0.79 to 1.24 for glyceraldehyde 3-phosphate dehydrogenase (GADPH), whereas actin 7 (ACT7) showed lower expression (0.05 in leaves and 0.11 in roots). Despite exceptions (e.g. ATLANTYS2, ATGP1A and ALYGYPSY25A), LTR-RT families presented relatively low expression, ranging from 0 to 0.014 and thus showing 10-fold to 100-fold lower levels than housekeeping genes. Although 18 families (i.e. 48.6%) revealed differential expression in leaf and root tissues, those LTR-RTs contributed little to the La4 polyploid transcriptome.

Families with several genomic copies generally showed relatively low expression levels in the La4 polyploid. Only ATLANTYS2 revealed single nucleotide polymorphism (SNP) variants indicative of transcripts from different genomic copies. Most LTR-RT families with evidence of proliferation in polyploids based on $F_{ST}$ revealed low expression levels, whereas several quiescent families showed relatively high expression in at least one tissue (e.g. ALYGYPSY36F and ALYGYPSY36G, ATHILA5E, ATLANTYS2, ALYGYPSY25A and ATCOPIA95). Only one family (ATGP1A) showed evidence of both recent transpositional and transcriptional activities, highlighting a loose coupling between past proliferation and current expression in polyploids.

## DISCUSSION

### Common TE pool in Brassicaceae

Long terminal repeat retrotransposons (LTR-RTs) represent the prevalent component of most plant genomes, but processes shaping their genome-wide landscape remain elusive (reviewed in Brookfield, 2005; Bonchev and Parisod, 2013). The abundance and dynamics of LTR-RTs lead to ambiguous assemblies with available sequencing approaches and thus challenges interpretations from comparative genomics. Accordingly, low coverage sequencing (0.1×) of random genome-wide fragments, coupled with appropriate statistical approaches, currently represents a valuable strategy to address variation in TE genome fractions (Senerchia *et al.*, 2013). In particular, for non-model

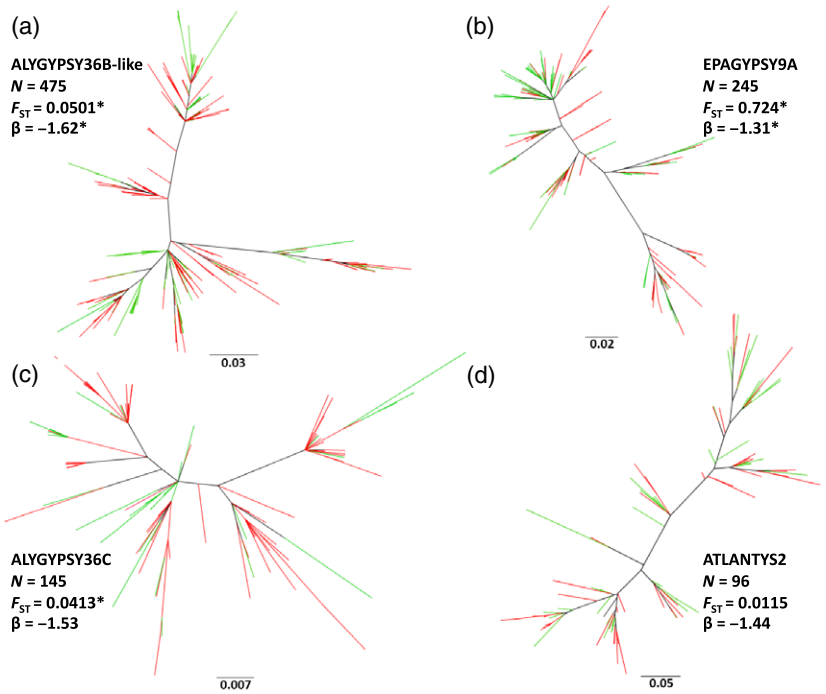


**Figure 2.** Examples of phylogenetic relationships among copies of long terminal repeat retrotransposon (LTR-RT) families in diploids and autopolyploids of *Biscutella*.
Unrooted phylogenies based on sequences at the end of the LTR region distinguish contrasted topologies: trees with clades of insertions specific to both diploid (green) and autotetraploid (red) taxa that are either: (a) imbalanced ($-2 < \beta < -1.5$) as shown here by the ALYGYPSY36B-like family; or (b) balanced ($\beta > -1.5$) as shown here by the EPAGYPSY9A family; (c) trees with autotetraploid-specific clades of insertions, as shown by the ALYGYPSY36C family; and (d) trees clades of insertions shared among taxa, as shown here by the ATLANTYS2 family. Bars are scaled to branch lengths. Number of LTR-RT copies (N), genetic differentiation ($F_{ST}$) and imbalance (β) according to Table 2. Trees for all 37 LTR-RT families are presented in Figure S2.

species lacking high quality genome assembly and that are divergent from model species, such as the focal Buckler Mustards (Couvreur *et al.*, 2010), *de novo* inference of TE references proved necessary to accurately examine and compare TE genome fractions. Compilation of comprehensive databases from related Brassicaceae (mostly Arabidopsis and *Brassica* species) provided divergent TE references that resulted in biased sampling of insertions in *Biscutella,* whereas species-specific TE references assembled from AAARF (DeBarry *et al.*, 2008) supported here more complete assessments of the TE diversity (also see Estep *et al.*, 2013).

Such sequence sampling in Buckler Mustards revealed a highly repetitive genome (Table 1), which is coherent with its relatively large genome (i.e. 920 Mb) as compared with other Brassicaceae (Lysak *et al.*, 2009). This first TE survey in an early diverging clade of core Brassicaceae thus adequately complements insights from other lineages comprising Arabidopsis (Hu *et al.*, 2011) and *Brassica* (Zhao *et al.*, 2013; Chalhoub *et al.*, 2014) and thus offers new knowledge for further examination of mechanisms ruling genome size variation across this model family (El Baidouri and Panaud, 2013). Repeats highlighted as TEs in *Biscutella* revealed a quite diverse set of abundant LTR-RTs showing clear relationships with canonical Brassicaceae TE references. This study further shows that the overall TE repertoire is consistent across Brassicaceae (Zhang and Wessler, 2004). The ATHILA clade, which includes a variety of pericentromeric families in Arabidopsis (Marco and Marin, 2008), presented here a diversity of closely related and abundant families (Figure 1). Congruently, several other LTR-RT groups such as ATGP5, EPAGYPSY9, ALYGYPSY25 and ALYGYPSY36 revealed considerable variation consistent with their diversification during the evolution of this early diverging genome. Accordingly, the diversity of LTR-RTs inferred in *Biscutella* supports the hypothesis of an ancestral pool of TE sequences (Jurka *et al.*, 2011) that diversified and fuelled genome evolution during radiation of the Brassicaceae clades, as otherwise shown in Triticeae (Middleton *et al.*, 2012; Senerchia *et al.*, 2013).

### Proliferation of LTR-RTs and species diversification

Most LTR-RT families analyzed here presented similar abundances in the six closely related genomes of the Buckler Mustard species complex. Specific LTR-RT families (e.g. ATGP5A, ALYGYPSY36A-like and ALYGYPSY25A) nevertheless showed significantly different abundances among genomes, suggesting contrasted dynamics after the split of corresponding taxa. Significantly lower or higher abundances varied among genomes according to the LTR-RT family considered, pointing to family-specific dynamics during species diversification. Noticeably, the genome of the diploid *B. laevigata* subsp. *austriaca,* which is considered as having diverged early from other taxa of the *B. laevigata* complex (Tremetsberger *et al.*, 2002), presented low abundances for several LTR-RTs, indicating that they may have amplified during the radiation of Buckler Mustards.

Phylogenetic relationships among individual LTR-RT copies from the different genomes (i.e. populations) as

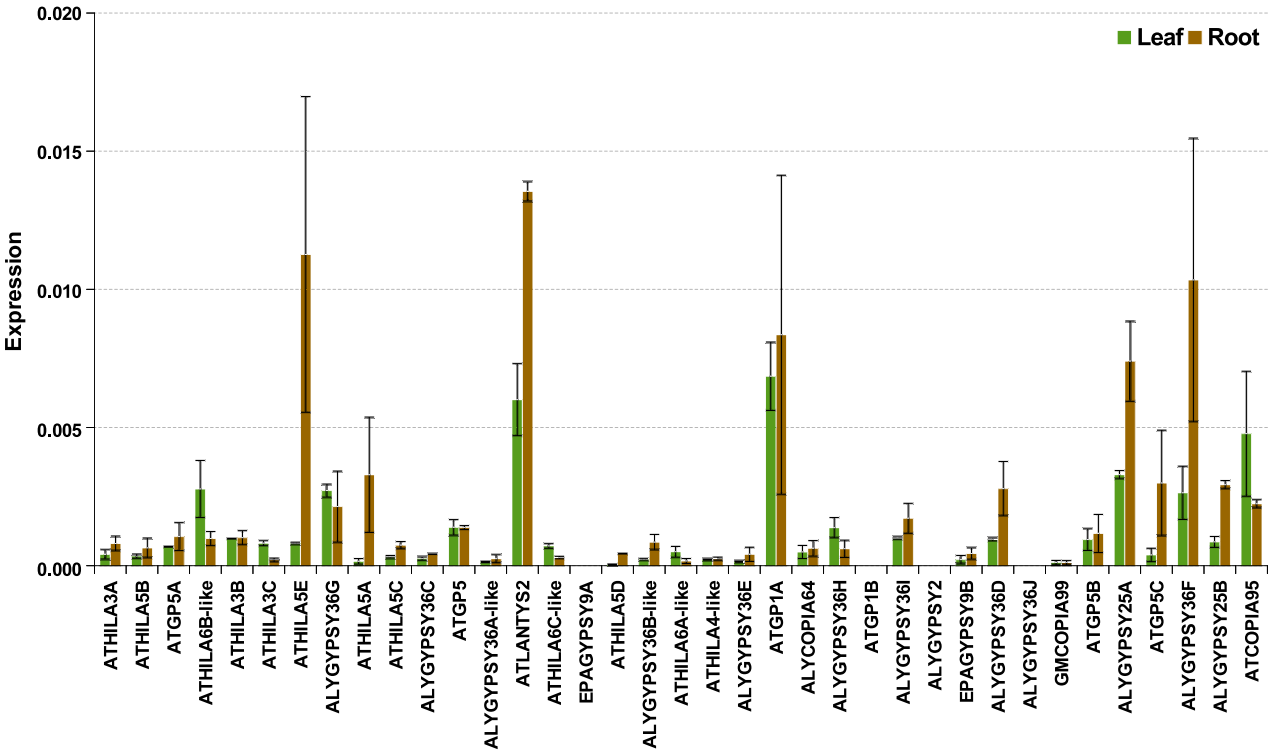


**Figure 3.** Expression of 37 long terminal repeat retrotransposons (LTR-RT) families in autotetraploid individuals of *Biscutella laevigata* subsp. *laevigata* (La4) through RNA-Seq.

Mean expression in leaf (green) and in root (brown) tissues estimated as the proportion of mapped cDNA reads on the consensus sequences of LTR-RT families out of the total mapped reads, normalized by the reference length. Error bars represent the standard error among biological replicates. LTR-RT families are ranked according to genomic abundance (Figure 1). Housekeeping genes such as ubiquitin 10 (UBQ10) and glyceraldehyde-3-phosphate dehydrogenase (GAD-PH) showed expression levels ranging for 0.78 to 2.10, whereas actin7 (ACT7) showed expression between 0.05 and 0.11 (not drawn).

well as further demographic inferences from molecular population genetics confirmed insights based on abundance. In contrast to similar surveys in widely divergent species (e.g. Senerchia *et al.*, 2013), sequence variation revealed low genetic structure among LTR-RT populations ($F_{ST} < 0.077$), which is coherent with the diversification of Buckler Mustards during the late Pleistocene (Tremetsberger *et al.*, 2002). In particular, the majority of the 37 studied LTR-RTs revealed non-significant $F_{ST}$ among genomes, indicating that neutral mutations *per se* were insufficiently high to assign insertions to genomes and to delineate taxa. Such evidence thus points to closely related copies being randomly shared among taxa and highlights LTR-RT families having remained quiescent during differentiation of *Biscutella* genomes. In contrast, more than one-third of the studied LTR-RTs showed significant $F_{ST}$ among genomes, highlighting closely related copies within taxa and indicating transpositional activity after the differentiation of corresponding genomes (Table 2).

Non-significant genetic structure among ploidy levels ($F_{CT}$) refuted consistent amplification of LTR-RTs in response to autopolyploidy *per se,* but significant structure among pairs of LTR-RT populations pointed to families with transpositional activity during the divergence of specific genomes. Congruent with quantitative assessment through molecular population genetics, phylogenies of active LTR-RT families revealed taxon-specific clades of insertions indicative of proliferation. Active families however presented balanced and imbalanced trees in equal proportions, indicating that proliferation through either the 'transposon model' (i.e. all copies duplicate at the same rate) or the 'master copy model' (i.e. one copy gives birth to most daughter copies) seemingly has a limited impact on their long-term dynamics (Brookfield and Johnson, 2006). Such a high proportion of LTR-RTs with evidence of recent transpositional activity emphasizes the significant influence of genome dynamics within species (Kalendar *et al.*, 2000; Bonchev and Parisod, 2013; Gustafsson *et al.*, 2014).

### Dynamics of LTR-RTs following autopolyploidy

Most of the 14 LTR-RT families inferred as active in *Biscutella* presented evidence of proliferation at the polyploid level. In contrast to available evidence of reduced genome reorganization in autopolyploids as compared to allopolyploids (reviewed in Parisod *et al.*, 2010b), the large proportion of transpositionally active TEs detected here in Buckler Mustards indicates substantial genome changes following

such duplication. In particular, two LTR-RTs (ALYGYPSY36C and ATHILA5A) showed genetic differentiation indicative of proliferation restricted to autopolyploids. Such a pattern accounts for a minority of LTR-RTs proliferating in autopolyploids, but favors the hypothesis that genetic redundancy *per se* sustained transpositional activity (Matzke and Matzke, 1998). In contrast, eight active LTR-RT families presented significant genetic differentiation among diploids and autopolyploids, together with specific clades of insertions, indicating transpositional activity before and after polyploidy (Figure 2c). The majority of LTR-RTs thus presented differentiation among progenitors and evidence of further proliferation when merged in polyploids. Such a pattern is coherent with effective transposition due to increased availability of neutral sites for expressed LTR-RTs to insert (i.e. redundancy hypothesis) and/or transient transcriptional activation of LTR-RTs (i.e. genome shock hypothesis) in polyploid genomes (Parisod and Senerchia, 2012).

The redundancy hypothesis strictly predicts the coupling of transcriptional and transpositional activity for proliferating LTR-RTs in young *B. laevigata* autopolyploids (Tremetsberger *et al.*, 2002; Parisod and Besnard, 2007). RNA-Seq however revealed low expression levels for all LTR-RT families in the polyploid *B. laevigata* subsp. *laevigata* (La4), including those having proliferated in polyploids exclusively (i.e. ALYGYPSY36C and ATHILA5A). Such evidence does not contradict with transcription and transposition at the origin of polyploids, but suggests transient LTR-RT activity, as reported in experimental autopolyploids (Madlung *et al.*, 2005) and allopolyploids (Ha *et al.*, 2009). Noticeably, higher expression levels were detected for LTR-RTs, such as ATLANTYS2, that revealed quiescent in Buckler Mustards and certainly represent co-transcripts from adjacent genes (Kashkush *et al.*, 2003). The pervasive LTR-RT expression detected here probably represents noise to a large extent (Palazzo and Gregory, 2014).

Patterns of genetic variation among genome-wide copies as well as LTR-RT expression in the *B. laevigata* autopolyploid complex support overall TE dynamics due to LTR-RT activation following the merging of closely related genomes (i.e. genome shock) and/or proliferation independent of any differentiation among progenitors (i.e. redundancy *per se*). Accordingly, TE activation through genome shock and facilitation of TE accumulation though genetic redundancy appear as mutually non-exclusive factors sustaining effective transposition in plants. In contrast with expectations of limited changes after whole genome duplication *per se* (Hegarty *et al.*, 2006; Leitch and Leitch, 2008; Parisod *et al.*, 2010b), long-term genome dynamics in natural autopolyploids is indicative of drastic TE reorganization analogous to that otherwise reported in allopolyploids. The relative impact of genome merging and genome duplication should thus be further assessed through the analysis of resynthesized autopolyploids mimicking established ones. Accordingly, future works addressing the evolutionary consequences of autopolyploidy may thus fully consider hybridization between closely related individuals to reach insightful conclusions (Tayale and Parisod, 2013).

## EXPERIMENTAL PROCEDURES

### Plant material and low coverage sequencing

Three diploids (2n = 2x = 18) and three autotetraploids (2n = 4x = 36) taxa involved in the evolution of the *Biscutella laevigata* autopolyploid complex were selected and collected in natural populations (Manton, 1937; Raffaelli and Baldoin, 1997; Tremetsberger *et al.*, 2002; Parisod and Besnard, 2007). Diploids were represented by *B. brevicaulis* (B2, Montagne d'Aurouze, Maritime Alps, France; 44°37′07.33″N–5°54′31.32″E), *B. laevigata* subsp. *austriaca* (A2, Gainfarn, Austria; 47°58′18.89″N–16°10′39.73″E) and *B. laevigata* subsp. *varia* (V2, Beuron, Germany; 48°03′05.81″N–8°59′00.81″E). Autotetraploids were represented by *B. laevigata* subsp. *ossolana* (O4, Simplon, Southern Alps, Switzerland; 46°11′05.99″N–8°04′31.59″E), *B. laevigata* subsp. *laevigata* (La4, Rochers de Naye, Northern Alps, Switzerland; 46°26′07.37″N–6°58′49.37″E) *and B. laevigata* subsp. *lucida* (Lu4, Recoaro, Italy; 45°42′07.98″N–11°09′38.21″E).

Leaves of one individual from each taxon were stored in silica gel and genomic DNA was extracted using the Qiagen DNeasy Plant Mini Kit (www.qiagen.com/DNeasyPlantMini). Approximately 40 ng of genomic DNA was mechanically shotguned and individually tagged random fragments were then sequenced on two half plates of the Roche 454 GS FLX titanium platform (service provided by Microsynth, Balgach, Switzerland). Sequences have been deposited at the NCBI Sequence Read Archive (SRA) under BioProject PRJNA280682.

### Classification of 454 reads using Brassicaceae references

Following Senerchia *et al.* (2013), low complexity 454 reads showing biased base distribution were removed (i.e. <0.2% of simple sequence repeats). Then, mitochondrial and chloroplastic sequences were identified by a BLASTN against reference organelle genomes from Brassicaceae species available in NCBI (NC_016118, NC_008285, NC_016123, NC_016125, NC_016120, NC_001284, NC_009265, NC_009266, NC_000932, NC_016734, NC_015139, NC_015139, NC_009274 and NC_009272; www.ncbi.nlm.nih.gov/; 2014). Coding sequences were identified by BLASTN followed by TBLASTX against the updated version of the TAIR10 model sequences (Lamesch *et al.*, 2012). Unassigned reads were further inspected to identify TE sequences.

A custom Brassicaceae TE reference was produced by merging and curating resources from *Arabidopsis arenosa, Arabidopsis lyrata, Arabidopsis thaliana, Brassica campestris, Brassica carinata, Brassica napus, Brassica nigra, Brassica oleracea* and *Brassica rapa* available in Repbase (Jurka *et al.*, 2005), in TIGR Plant Repeat Databases (Ouyang and Buell, 2004), in a work comparing *A. lyrata* and *A. thaliana* (de la Chaux *et al.*, 2012), and the mining of the *B. rapa* genome (available from V. Sarilar, P. Martinez Palacios, J. Joets, K. Alix, UMR Génétique Végétale Le Moulon). This custom reference contains 3214 non-redundant TEs classified according to Wicker *et al.* (2007). This reference was used to classify previously unassigned 454 reads using BLASTN followed by a TBLASTX. Given divergence between *Biscutella* TE sequences and available references (see Results), *de novo* assembly inferred repeated elements

(see below) and the procedure was repeated after the merging of both databases.

### Identification and abundance of repeats and TEs from Biscutella

Abundant repeated elements were assembled in the six genomes independently, using AAARF (DeBarry *et al.*, 2008). After exploration of the parameter space as advised by Estep *et al.* (2013), the following specifications optimally resulted in the lowest number of longest contigs: Minimum length for BLAST hits of 100 bp, minimum identity for BLAST hits of 85%, minimum coverage depth of 5, minimum number of five BLAST hits for contig extension and maximum $10^{-5}$ e-value for BLAST and BL2SEQ hits. Resulting contigs were trimmed for low complexity sequences using PRINSEQ v0.20.1 (Schmieder and Edwards, 2011), and their quality revealed normally distributed GC content and <1.5% of ambiguous bases occurring over <3% of sequences.

Contigs from each genome were merged and those presenting sequence identity higher that 80% over at least 80% of their alignment were clustered into families using CD-HIT-EST (Li and Godzik, 2006). Only contigs inferred independently in at least four genomes and presenting length higher than 3000 bp were further considered by keeping the longest contig as representative.

Contigs were annotated by BLASTN and/or TBLASTX (best hit) against our custom Brassicaceae TE reference. Repeats were annotated as belonging to a known TE family when sharing at least 80% sequence identity over 80% of the alignment of coding/internal domains or terminal regions of TE references (Wicker *et al.*, 2007). Stretches of homology shorter than 80 bp were not considered. Non-clustered contigs matching the same TE reference following those criteria were distinguished with letters following the family name. Contigs unambiguously related to a TE reference, but showing lower sequence identity (i.e. 70–80%), were annotated with a '-like' to account for surmised relatedness.

Following Senerchia *et al.* (2013), the abundance of LTR-RT families within each genome was estimated as the proportion of reads matching the corresponding family consensus sequences out of the total read number. Possible sampling effects were assessed by resampling reads assignment 999 times with replacement using R (http://www.R-project.org). Ninety-five percent confidence intervals around proportions were estimated according to binomial distributions using Wald's method for large numbers in the R package binGroup. Non-overlapping confidence intervals identified significantly different abundances.

### Sampling of individual LTR-RT copies

The 3′ or the 5′ end of the LTR region of each LTR-RT family was manually delimited through identification of primer binding sites and polypurine tracts (Du *et al.*, 2011) as well as coverage according to DeBarry *et al.* (2008). After broad delimitation of LTR sequences through identification of internal coding sequences by BLASTX (i.e. using conserved domains from plant TEs; Llorens *et al.*, 2011), LTR borders were characterized by consistent increased in read mapping and confirmed by either conserved CA at the 3′ end or TG at the 5′ end of the LTR region.

Individual reads corresponding to a minimum of 150 bp at either the 5′ or the 3′ end of the LTR region were retrieved for each LTR-RT family from sequenced genomes by BLASTN. Identical 454 reads starting within two nucleotides of one another (i.e. <10% of the total read number) were removed using 454 Replicate Filter (Gomez-Alvarez *et al.*, 2009) to eliminate technical artifacts arising during sequencing. Given the shallow sequencing depth (0.1×), those individual 454 reads were considered representative of independent copies (Senerchia *et al.*, 2013).

### Phylogenetic inferences of LTR-RT families

SATé I simultaneously inferred alignment and phylogenetic tree from retrieved reads corresponding to individual copies of each LTR-RT family (Liu *et al.*, 2009). Using a maximum likelihood (ML) approach and treating gaps as missing data, this algorithm iteratively aligned sequences and progressively merged local alignments into an alignment over the full data set until the tree/alignment pair with the best ML score is reached. Iteration limit was set to 10, using 'decomposition = longest' and 'return = best'. Final ML trees were inferred using RaxML in SATé I and drawn in Figtree (http://tree.bio.ed.ac.uk/software/figtree/).

Imbalance of inferred phylogenetic trees was estimated using the $\beta$ index (Mooers and Heard, 1997; Aldous, 2001). Interpreting branch duplication as transposition events (Le Rouzic *et al.*, 2013), TE families with all copies having transposed at the same rate (i.e. 'transposon model') show balanced random trees ($-1.5 \leq \beta \leq 0$), whereas imbalanced trees ($-2 < \beta < -1.5$) highlight TE insertions having transposed from only one copy (i.e. 'master copy model'; Brookfield and Johnson, 2006). The $\beta$ index was here estimated through ML using the R package apTreeshape (Bortolussi *et al.*, 2006). As the tip number was large, 95% confidence intervals were computed from 1000 bootstraps using the 'profile' method.

### Population genetics of LTR-RT copies

Partitioning of molecular variance among copies of LTR-RT families within and among genomes (i.e. populations), nested within ploidy levels, was assessed through $F_{ST}$-based hierarchical AMOVA (Excoffier *et al.*, 2005). Using Tamura and Nei distances between sequences, three differentiation indexes among TE copies were estimated in ARLEQUIN version 3.5: among all populations ($F_{ST}$), among populations within ploidy levels ($F_{SC}$) and among ploidy levels ($F_{CT}$). Genetic differentiation ($F_{ST}$) was similarly tested between pairs of populations. Significance at the 0.05 level was tested by permuting haplotypes 100 times between partitions.

### RNA-Seq and transcript expression of LTR-RTs

Tissues from young leaves and roots were simultaneously collected in liquid nitrogen from two autotetraploid individuals of the La4 taxon, grown for 1 year in a common garden (Botanical Garden of Neuchatel). Total RNA was extracted from each tissue separately using PureLink RNA Mini Kit (Life Technologies, Zug, Switzerland) and its quality was checked on an Agilent bioanalyzer. Quantitative cDNA libraries were prepared following poly(A) enrichment and barcoded for SOLiD, mixed and single-end sequenced on two lanes of SOLiD 5500×l (service provided by Microsynth, Balgach, Switzerland). The exact call chemistry was further performed to improve the quality of SOLiD reads.

Short single-end reads (70 bp) corresponding to cDNA fragments were mapped to consensus sequences of the 37 LTR-RT families and to three housekeeping genes using the BWA-backtrack algorithm for short reads (Li and Durbin, 2009). UBQ10, GADPH and ACT7 show limited expression fluctuation across tissues and experimental conditions (Tong *et al.*, 2009), and were selected here as housekeeping genes. Descriptive statistics on the alignment were provided by Bam files produced with SAMtools (Li *et al.*, 2009) and further used for calling variants with GATK (DePristo *et al.*, 2011).

Following Slotte *et al.* (2013), expression levels were estimated for each LTR-RT family and housekeeping gene in each tissue as

the proportion of matching cDNA reads out of the total mapped reads and normalized by the reference length. Expression levels were averaged among individual replicates. Expression from multiple LTR-RT copies was checked from SNP variants based on a minimal coverage depth of 10 and a minimum minor allele frequency of 5, using SNiPloid (Peralta *et al.*, 2013).

## ACKNOWLEDGEMENTS

We thank K. Alix and N. Senerchia for sharing data as well as anonymous reviewers for valuable comments on the manuscript. Data were analyzed using facilities from the Genetic Diversity Centre (GDC), ETH Zurich. We are not aware of any conflict of interest. This work was supported by grants from the Velux Stiftung (705), the Novartis Foundation, and the Swiss National Science Foundation (PZ00P3-131950).

## SUPPORTING INFORMATION

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

**Figure S1.** Relative abundance of transposable element fractions in *Biscutella* genomes.

**Figure S2.** Phylogenetic relationships among copies of 37 long terminal repeat retrotransposon families in diploids and polyploids of the autopolyploid complex *Biscutella laevigata*.

**Table S1.** Abundance and dynamics (phylogenetics and population genetics) of 37 long terminal repeat retrotransposons (LTR-RT) families in six *Biscutella* genomes (coded as in Table 1).

**Table S2.** Expression of 37 long terminal repeat retrotransposons (LTR-RT) families in tetraploid individuals of the *Biscutella laevigata* subsp. *laevigata* (La4) population through RNA-Seq.

## REFERENCES

**Aldous, D.J.** (2001) Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Stat. Sci.* **16**, 23–24.

**Bennetzen, J.L.** (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. *Curr. Opin. Genet. Dev.* **15**, 621–627.

**Biemont, C. and Vieira, C.** (2006) Junk DNA as an evolutionary force. *Nature*, **443**, 521–524.

**Bonchev, G. and Parisod, C.** (2013) Transposable elements and microevolutionary changes in natural populations. *Mol. Ecol. Resour.* **13**, 765–775.

**Bortolussi, N., Durand, E., Blum, M. and François, O.** (2006) apTreeshape: statistical analysis of phylogenetic tree shape. *Bioinformatics*, **22**, 363–364.

**Bourc'his, D. and Voinnet, O.** (2010) A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science*, **330**, 617–622.

**Brookfield, J.F.Y.** (2005) The ecology of the genome - Mobile DNA elements and their host. *Nat. Rev. Genet.* **6**, 128–136.

**Brookfield, J.F.Y. and Johnson, L.J.** (2006) The evolution of mobile DNAs: when will transposons create Phylogenies that look as if there is a master gene? *Genetics*, **173**, 1115–1123.

**Casacuberta, J.M., Vernhettes, S., Audeon, C. and Grandbastien, M.A.** (1997) Quasispecies in retrotransposons: a role for sequence variability in Tnt1 evolution. *Genetica*, **100**, 109–117.

**Chalhoub, B., Denoeud, F., Liu, S. *et al.*** (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, **345**, 950–953.

**de la Chaux, N., Tsuchimatsu, T., Shimizu, K.K. and Wagner, A.** (2012) The predominantly selfing plant *Arabidopsis thaliana* experienced a recent reduction in transposable element abundance compared to its outcrossing relative *Arabidopsis lyrata*. *Mob. DNA*, **3**, 2.

**Comai, L., Madlung, A., Josefsson, C. and Tyagi, A.** (2003) Do the different parental 'heteromes' cause genomic shock in newly formed allopolyploids? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**, 1149–1155.

**Couvreur, T.L.P., Franzke, A., Al-Shehbaz, I.A., Bakker, F.T., Koch, M.A. and Mummenhoff, K.** (2010) Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Mol. Biol. Evol.* **27**, 55–71.

**DeBarry, J.D., Liu, R. and Bennetzen, J.L.** (2008) Discovery and assembly of repeat family pseudomolecules from sparse genomic sequence data using the Assisted Automated Assembler of Repeat Families (AAARF) algorithm. *BMC Bioinformatics*, **9**, 235.

**DePristo, M.A., Banks, E., Poplin, R. *et al.*** (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498.

**Doolittle, W.F.** (2013) Is junk DNA bunk? A critique of ENCODE. *Proc. Natl Acad. Sci. USA*, **110**, 5294–5300.

**Doyle, J.J., Flagel, L.E., Paterson, A.H., Rapp, R.A., Soltis, D., Soltis, P. and Wendel, J.F.** (2008) Evolutionary genetics of genome merger and doubling in plants. *Annu. Rev. Genet.* **42**, 443–461.

**Du, J.C., Tian, Z.X., Hans, C.S., Laten, H.M., Cannon, S.B., Jackson, S.A., Shoemaker, R.C. and Ma, J.X.** (2011) Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *Plant J.* **63**, 584–598.

**El Baidouri, M. and Panaud, O.** (2013) Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biol. Evol.* **5**, 954–965.

**Estep, M.C., DeBarry, J.D. and Bennetzen, J.L.** (2013) The dynamics of LTR retrotransposon accumulation across 25 million years of panicoid grass evolution. *Heredity*, **110**, 194–204.

**Excoffier, L., Smouse, P.E. and Quattro, J.M.** (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

**Excoffier, L., Laval, G. and Schneider, S.** (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinf. Online*, **1**, 47–50.

**Fedoroff, N.V.** (2012) Transposable elements, epigenetics, and genome evolution. *Science*, **338**, 758–767.

**Freeling, M., Woodhouse, M.R., Subramaniam, S., Turco, G., Lisch, D. and Schnable, J.C.** (2012) Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Curr. Opin. Plant Biol.* **15**, 131–139.

**Gaut, B. and Ross Ibarra, J.** (2008) Selection on major components of angiosperm genomes. *Science*, **320**, 484–486.

**Gomez-Alvarez, V., Teal, T.K. and Schmidt, T.M.** (2009) Systematic artifacts in metagenomes from complex microbial communities. *ISME J.* **3**, 1314–1317.

**Grandbastien, M., Audeon, C., Bonnivard, E. *et al.*** (2005) Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenet. Genome Res.* **110**, 229–241.

**Gustafsson, A.L.S., Skrede, I., Rowe, H.C., Gussarova, G., Borgen, L., Rieseberg, L.H., Brochmann, C. and Parisod, C.** (2014) Genetics of cryptic speciation within an arctic mustard, *Draba nivalis*. *PLoS One*, **9**, e93834.

**Ha, M., Lu, J., Tian, L., Ramachandran, V., Kasschau, K.D., Chapman, E.J., Carrington, J.C., Chen, X.M., Wang, X.J. and Chen, Z.J.** (2009) Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. *Proc. Natl Acad. Sci. USA*, **106**, 17835–17840.

**Hegarty, M.J., Barker, G.L., Wilson, I.D., Abbott, R.J., Edwards, K.J. and Hiscock, S.J.** (2006) Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Curr. Biol.* **16**, 1652–1659.

**Hu, T.T., Pattyn, P., Bakker, E.G. *et al.*** (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* **43**, 476–481.

**Hua-Van, A., Le Rouzic, A., Boutin, T.S., Filee, J. and Capy, P.** (2011) The struggle for life of the genome's selfish architects. *Biol. Direct*, **6**, 19.

**Josefsson, C., Dilkes, B. and Comai, L.** (2006) Parent-dependent loss of gene silencing during interspecies hybridization. *Curr. Biol.* **16**, 1322–1328.

**Jurka, J., Pavlicek, A., Klonowski, P., Kohany, O., Kapitonov, V.V. and Walichiewicz, J.** (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467.

**Jurka, J., Bao, W. and Kojima, K.K.** (2011) Families of transposable elements, population structure and the origin of species. *Biol. Direct*, **6**, 44.

**Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E. and Schulman, A.H.** (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl Acad. Sci. USA*, **97**, 6603–6607.

**Kashkush, K., Feldman, M. and Levy, A.A.** (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat. Genet.* **33**, 102–106.

**Kejnovsky, E., Leitch, I.J. and Leitch, A.R.** (2009) Contrasting evolutionary dynamics between angiosperm and mammalian genomes. *Trends Ecol. Evol.* **24**, 572–582.

**Lamesch, P., Berardini, T.Z., Li, D.H. *et al.*** (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210.

**Le Rouzic, A., Payen, T. and Hua-Van, A.** (2013) Reconstructing the evolutionary history of transposable elements. *Genome Biol. Evol.* **5**, 77–86.

**Leitch, A.R. and Leitch, I.J.** (2008) Genomic plasticity and the diversity of polyploid plants. *Science*, **320**, 481–483.

**Li, H. and Durbin, R.** (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

**Li, W. and Godzik, A.** (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.

**Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R.** (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

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

**Liu, K., Raghavan, S., Nelesen, S., Linder, C.R. and Warnow, T.** (2009) Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. *Science*, **324**, 1561–1564.

**Llorens, C., Futami, R., Covelli, L. *et al.*** (2011) The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res.* **39**, D70–D74.

**Lockton, S. and Gaut, B.S.** (2009) The contribution of transposable elements to expressed coding sequence in *Arabidopsis thaliana*. *J. Mol. Evol.* **68**, 80–89.

**Lysak, M.A., Koch, M.A., Beaulieu, J.M., Meister, A. and Leitch, I.J.** (2009) The dynamic ups and downs of genome size evolution in Brassicaceae. *Mol. Biol. Evol.* **26**, 85–98.

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

**Manton, I.** (1937) The problem of *Biscutella laevigata* L. II. The evidence from meiosis. *Ann. Bot.* **51**, 439–465.

**Marco, A. and Marin, I.** (2008) How Athila retrotransposons survive in the Arabidopsis genome. *BMC Genom.* **9**, 219.

**Matzke, M.A. and Matzke, A.J.M.** (1998) Polyploidy and transposons. *Trends Ecol. Evol.* **13**, 241.

**McClintock, B.** (1984) The significance of responses of the genome to challenge. *Science*, **226**, 792–801.

**Middleton, C., Stein, N., Keller, B., Kilian, B. and Wicker, T.** (2012) Comparative analysis of genome composition in Triticeae reveals strong variation in transposable element dynamics and nucleotide diversity. *Plant J.* **73**, 347–356.

**Mooers, A.O. and Heard, S.B.** (1997) Inferring evolutionary process from the phylogenetic tree shape. *Q. Rev. Biol.* **72**, 31–54.

**Ouyang, S. and Buell, C.R.** (2004) The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. *Nucleic Acids Res.* **32**, D360–D363.

**Palazzo, A.F. and Gregory, T.R.** (2014) The case of junk DNA. *PLoS Genet.* **10**, e1004351.

**Parisod, C. and Besnard, G.** (2007) Glacial in situ survival in the Western Alps and polytopic autopolyploidy in *Biscutella laevigata* L. (Brassicaceae). *Mol. Ecol.* **16**, 2755–2767.

**Parisod, C. and Senerchia, N.** (2012) Responses of transposable elements to polyploidy. In *Plant Transposable Elements* (Grandbastien, M.-A. and Casacuberta, J.M. eds). Berlin, Heidelberg: Springer-Verlag, pp. 147–148.

**Parisod, C., Alix, K., Just, J., Petit, M., Sarilar, V., Mhiri, C., Ainouche, M., Chalhoub, B. and Grandbastien, M.A.** (2010a) Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytol.* **186**, 37–45.

**Parisod, C., Holderegger, R. and Brochmann, C.** (2010b) Evolutionary consequences of autopolyploidy. *New Phytol.* **186**, 5–17.

**Peralta, M., Combes, M.-C., Cenci, A., Lashermes, P. and Dereeper, A.** (2013) SNiPloid: a utility to exploit high-throughput SNP data derived from RNA-seq in allopolyploid species. *Int. J. Plant Genomics*, **2013**, 890123.

**Raffaelli, M. and Baldoin, L.** (1997) Il complesso di *Biscutella laevigata* L. (Crucifera) in Italia. *Webbia*, **52**, 87–128.

**Schmieder, R. and Edwards, R.** (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **15**, 863–864.

**Senerchia, N., Wicker, T., Felber, F. and Parisod, C.** (2013) Evolutionary dynamics of LTR retrotransposons in wild wheats assessed with high throughput sequencing. *Genome Biol. Evol.* **5**, 1010–1020.

**Senerchia, N., Felber, F. and Parisod, C.** (2014) Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. *New Phytol.* **202**, 975–985.

**Slotte, T., Hazzouri, K.M., Agren, J.A. *et al.*** (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* **45**, 831–835.

**Stebbins, G.** (1971) *Chromosomal evolution in higher plants*. London: Addison-Wesley.

**Tayale, A. and Parisod, C.** (2013) Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenet. Genome Res.* **140**, 79–96.

**Tong, Z., Gao, Z., Wang, F., Zhou, J. and Zhang, Z.** (2009) Selection of reliable reference genes for gene expression studies in peach using real-time PCR. *BMC Mol. Biol.* **10**, 71.

**Tremetsberger, K., Konig, C., Samuel, R., Pinsker, W. and Stuessy, T.F.** (2002) Infraspecific genetic variation in *Biscutella laevigata* (Brassicaceae): new focus on Irene Manton's hypothesis. *Plant Syst. Evol.* **233**, 163–181.

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

**Zhang, X.Y. and Wessler, S.R.** (2004) Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. *Proc. Natl Acad. Sci. USA*, **101**, 5589–5594.

**Zhao, M., Du, J., Lin, F., Tong, C., Yu, J., Huang, S., Wang, X., Liu, S. and Ma, J.** (2013) Shifts in evolutionary rate and intensity of purifying selection between two *Brassica* genomes revealed by analyses of orthologous transposons and relicts of a whole genome triplication. *Plant J.* **76**, 211–222.

 