New Phytologist

# Evidence of developmental escape from transcriptional gene silencing in *MESSI* retrotransposons

**Diego H. Sanchez, Hervé Gaubert and Weibing Yang**

The Sainsbury Laboratory, University of Cambridge, 47 Bateman Street, Cambridge, CB2 1LR, UK ; Present address: IFEVA (CONICET-UBA), Facultad de Agronomía, Universidad de Buenos Aires, Av San Martín 4453, Buenos Aires C1417DSE, Argentina

Author for correspondence:
*Diego H. Sanchez*
*Tel: +54 11 528 70112*
*Email: diegosanchez@agro.uba.ar*

Received: *25 November 2018*
Accepted: *12 March 2019*

*New Phytologist* (2019) **223:** 950–964
**doi**: 10.1111/nph.15896

**Key words:** epigenetics, plant development, retrotransposons, *Solanum*, tomato, transcriptional gene silencing, transposable elements, transposons.

## Summary

- Transposable elements (TEs) are ubiquitous genomic features. 'Copy-and-paste' long-terminal-repeat (LTR) retrotransposons have been particularly successful during evolution of the plant kingdom, representing a substantial proportion of genomes. For survival in copious numbers, these TEs may have evolved replicative mobilization strategies that circumvented hosts' epigenetic silencing. Stressful circumstances are known to trigger the majority of known mobilizing plant retrotransposons, leading to the idea that most are activated by environmental signals. However, previous research revealed that plant developmental programs include steps of silencing relaxation, suggesting that developmental signals may also be of importance for thriving parasitic elements.
- Here, we uncover an unusual family of giant LTR retrotransposons from the *Solanum* clade, named *MESSI*, with transcriptional competence in shoot apical meristems of tomato. Despite being recognized and targeted by the host epigenetic surveillance, this family is activated in specific meristematic areas fundamental for plant shoot development, which are involved in meristem formation and maintenance.
- Our work provides initial evidence that some retrotransposons may evolve developmentally associated escape strategies to overcome transcriptional gene silencing in vegetative tissues contributing to the host's next generation.
- This implies that not only environmental but also developmental signals could be exploited by selfish elements for survival within the plant kingdom.

## Introduction

Transposable elements (TEs) are selfish genomic parasites, capable of increasing copy number and modifying their position within host genomes. TEs represent a threat to genome stability due to mutational potential, latent deleterious effects on neighboring gene expression and facilitation of chromosomal rearrangements (Weil & Martienssen, 2008; Hollister & Gaut, 2009; Tenaillon *et al.*, 2010; Blumenstiel, 2011; Gaubert *et al.*, 2017). To counteract TE activity, host genomes evolved the capacity to suppress them through transcriptional gene silencing (TGS) mechanisms (Lisch, 2009; Fultz *et al.*, 2015; Matzke *et al.*, 2015). In plants, the RNA-dependent-DNA-methylation (RdDM) pathway orchestrates sequence-specific targeting of TEs, mainly through the generation of noncoding 24-nucleotide small-RNAs (smRNAs), with subsequent chromatin silencing mediated by epigenetic marks such as DNA methylation (Lisch, 2009; Fultz *et al.*, 2015; Matzke *et al.*, 2015). TGS pathways are particularly vigorous and efficient in plant meristematic tissues, safeguarding from harmful TE activity those cell lineages potentially contributing to the next generation (Baubec *et al.*, 2014). Additionally, post-transcriptional gene silencing (PTGS), effected by 21–22-nucleotide smRNAs, is capable of targeting TE transcripts for degradation if TGS is alleviated (Bucher *et al.*, 2012). Still, occasional transpositional bursts of TEs may occur, most remarkably during plant hybridization, polyploidization or under environmental challenges (Lisch, 2009; Grandbastien, 2015; Galindo-Gonzalez *et al.*, 2017). Previous research has suggested that plant developmental programs include some steps of TGS relaxation, allowing global transcriptional reactivation of TEs. Specifically, this was described in vegetative companion cells of the male gametophyte secondary to global DNA demethylation in this tissue (Slotkin *et al.*, 2009), and also in meristematic tissue of grasses such as maize and rice (Ohtsu *et al.*, 2007; Tamaki *et al.*, 2015). It was speculated that this programmed, seemly reproducible, and cell/tissue-specific reactivation of TEs may provide an advantage for plant genomes in retargeting repressive chromatin, in a fashion reminiscent of animal germline-specific silencing (Martinez & Slotkin, 2012). However, the functional role of this phenomenon, if any, remains poorly understood.

TEs are classically categorized according to their transposition intermediates, where class I elements comprise 'copy-and-paste' retrotransposons while class II elements comprise 'cut-and-paste'

TEs. Class I long-terminal-repeat (LTR) retrotransposons have been particularly successful throughout the evolution of the plant kingdom, resulting in significant increases of genome sizes due to the accumulated inactive relics derived from historical transposition events (Lisch, 2009; Tenaillon *et al.*, 2010). Potentially active LTR retrotransposons typically present two LTRs which flank coding sequences for functional proteins, such as a structural GAG and a polyprotein POL. The latter comprises enzymatic activities required for completion of the elements' life cycle, and includes a protease, a reverse transcriptase/ribonuclease H and an integrase; classical *copia-like* or *gypsy-like* superfamilies are defined according to the arrangement of coding sequences within POL (Kumar & Bennetzen, 1999; Sabot & Schulman, 2006; Wicker *et al.*, 2007). Active elements lacking some of these required open reading frames (ORFs) have also been described, demonstrating that they can be mobilized nonautonomously (Sabot, 2014). Competent LTR retrotransposons also possess a primer binding site that mediates the tRNA priming essential for the reverse transcription of their genomic RNA, which originates life-cycle intermediates in the form of extrachromosomal linear DNA (Kumar & Bennetzen, 1999; Sabot & Schulman, 2006; Wicker *et al.*, 2007). Transcriptional stimulation represents the first key activation step in the life of LTR retrotransposons (Grandbastien, 1998, 2015; Casacuberta & Santiago, 2003; Galindo-Gonzalez *et al.*, 2017). The relatively few plant LTR retrotransposons known to mobilize in experimentally validated controlled bursts do so mainly by *in vitro* tissue culture, mutational disruption of TGS and environmental cues (Casacuberta & Santiago, 2003; Grandbastien, 2015; Galindo-Gonzalez *et al.*, 2017). However, the initial trigger signals for the vast majority of potentially active LTR retrotransposons remains largely unknown, and it is often believed that they can reactivate upon some type of stressful situation (Grandbastien, 1998, 2015; Galindo-Gonzalez *et al.*, 2017). To the best of our knowledge, only a few characterized plant LTR retrotransposons display certain developmental regulation, often resulting from nonphysiological disturbance of epigenetic surveillance and in some cases only after added inductive stress (Grandbastien, 1998; Fukai *et al.*, 2010; Jaaskelainen *et al.*, 2013; Mari-Ordonez *et al.*, 2013; Gaubert *et al.*, 2017). The contribution of vegetative development, as a modulator of plant TE activity, has not yet been thoroughly pursued.

To shed light on the importance of plant development and tissue specificity in the life cycle of LTR retrotransposons, we investigated their activation in shoot apical meristems of tomato plants (*Solanum lycopersicum*). Species of the *Solanum* clade are particularly well suited for harvesting meristems at specific developmental stages (Park *et al.*, 2012). Moreover, they have long been used as models for studying plant development and morphology (Park *et al.*, 2014; Wang *et al.*, 2016), and bear the advantage of having various fully sequenced phylogenetic relatives spanning *c.* 14 million years of divergence time, thus facilitating comparative evolutionary studies (Sarkinen *et al.*, 2013; Tomato Genome Sequencing *et al.*, 2014). Here, we uncover an unusual family of giant *Solanum gypsy-like* elements with transcriptional competence in tomato shoot apical meristems, providing initial clues regarding developmentally associated regulation of TEs in vegetative plant tissues harboring undifferentiated germline cells.

## Materials and Methods

### Growing conditions and plant material

Tomato (*S. lycopersicum*) M82 and *S. pennellii* LA0716 seeds were directly sown in soil (Levington F2, ICL) and stratified in the dark for 1 wk at 4°C. Plants were grown in a Conviron MTPS chamber (Conviron, Winnipeg, Canada) at 25 °C : 18 °C, 16 : 8 h, day/night cycle, 300 μmol $m^{-2}$ $s^{-1}$ light intensity and *c.* 60% humidity. Long-term heat involved a treatment at 35°C over a 16 h daylight period before collection. In all cases, samples were harvested at the end of the day. For next-generation sequencing (NGS), samples consisted of pools of at least 100 primary whole shoot apical meristems, at least four open flowers, or at least six first leaves from 2- or 3-wk-old plants of tomato or *S. pennellii*, respectively. Routinely, shoot apical meristems were carefully dissected under a stereoscope and acetone-fixed in dry ice, removing primordial leaves before harvest. Primary whole shoot apical meristems were at the same developmental window, compatible with the flower meristem (FM) stage (Park *et al.*, 2012), each harboring a flower and sympodial meristems and a sympodial inflorescence. For quantitative reverse transcriptase (qRT)-PCR, samples consisted of pools of at least 100 whole apical meristems collected at different developmental stages according to Park *et al.* (2012): vegetative (VE), transition-to-flowering (TM) and flower (FM) stages, plus sympodial inflorescence (INF). Cotyledons were pooled from at least six plants, while independent calluses were obtained from explants growing on MS media with 0.5 mg $l^{-1}$ 6-benzylaminopurine and 0.1 mg $l^{-1}$ 1-naphthaleneacetic acid.

### Bioinformatics analyses

We performed *de novo* annotation of very young tomato LTR retrotransposons to complement the Xu & Du (2014) curated dataset. Briefly, at least 1000-bp-long preselected LTR-Class derived sequences, as recognized by Jouffroy *et al.* (2016), were first expanded to 5000 bp in both 5′ and 3′ directions along chromosomes. In these sequences, automatic prediction of retrotransposons of at least 98% 5′/3′ LTR similarity was executed using LTRHARVEST (Ellinghaus *et al.*, 2008); with paramenters sets -v -mintsd 3 -maxtsd 6 -seed 30 -xdrop 5 -mat 2 -mis -2 -ins -3 -del -3 -minlenltr 100 -maxlenltr 7000 -mindistltr 1000 -maxdistltr 30000 -similar 98 -overlaps best -vic 60 -longoutput. We recognized 676 nonoverlapping very young retroelements, from which 261 were not documented previously by Xu & Du (2014) (Supporting Information Table S1); these were merged for whole transcriptome analysis. Workflows were developed in-house for these manipulations using BEDTOOLS (Quinlan & Hall, 2010) and custom-made Python scripts (www.python.org), available at https://github.com/diegohernansanchez/.

*MESSI-like* elements within the *Solanum* clade were documented by two rounds of more refined manual bioinformatics

analyses. In the first round, the recognized differentially expressed tomato *MESSI* elements were blasted against tomato, *S. arcanum*, *S. habrochaites*, *S. pennellii*, *S. tuberosum* and *S. melongena* genomes (www.solgenomics.net). BLAST hit areas were extended 5000 bp bidirectionally and were used for *de novo* LTR retrotransposon annotation with LTR_FINDER (Xu & Wang, 2007), with parameters sets -D 30000 -d 1000 -L 5000 -l 200, and *A.thaliana* tRNAs as templates for primer binding site. The resulting list was filtered for unique genomic coordinates and also manually inspected, ruling out potential mis-annotations. For the second round of analysis the same workflow was implemented, but now using all these newly annotated retrotransposons in each *Solanum* genome, for BLAST and subsequent LTR_FINDER analysis. The LTR_FINDER report sometimes included several candidates within the same genomic area; we selected the larger one and otherwise discarded hits that could not be manually resolved. Finally, discontinuous MEGABLAST was implemented to assess similarity with SLY_MESSI_COM_17 (INT_03_147 from Xu & Du, 2014), considered here as an archetypal, young, full-length and differentially expressed family member. Elements defined as incomplete presented strings of ambiguous Ns within their sequences. Retrotransposon age based on 5′/3′ LTR similarity was estimated using a substitution rate of $1.3 \times 10^{-8}$ per site per year (Wicker & Keller, 2007). MEGABLAST, BLASTN and BLASTP were performed using NCBI tools ( https://blast.ncbi.nlm.nih.gov/Blast.cgi) or locally through Python scripts. Multiple sequence alignments were executed with MUSCLE (Edgar, 2004) and CLUSTALW (Thompson *et al.*, 1994), and phylogenetic trees with the Jukes–Cantor model and neighbor-joining method with 10 000 bootstrapping, using the GENEIOUS software ( www.geneious.com). Observed vs expected DNA CpG ratio was calculated as (number of CpGs × sequence length)/(number of Cs × number of Gs). Tandem-repeats were recognized online with the Tandem-Repeats Finder (Benson, 1999).

## Next-generation sequencing and data analysis

Total RNA from shoot apical meristems was extracted with the PicoPure RNA Isolation kit (Life Technologies), whereas the Plant-RNAeasy kit (Invitrogen) was used for flowers and leaves. DNA samples were extracted with the DNeasy Plant Mini Kit (Qiagen). Strand-specific libraries from duplicated biological replicates were prepared with 2 μg of RNA using the TruSeq Stranded mRNA Sample Prep Kit (Illumina, San Diego, CA, USA), while PCR-free libraries for the DNA-seq were prepared with 1 μg of DNA using the TruSeq DNA PCR-Free Library Prep kit (Illumina), following the supplier's instructions. Libraries were analyzed in the 2200 TapeStation (Agilent) and subsequently sequenced in a Next-Seq 500 (Illumina) platform reporting pair-end reads. Raw data were deposited in ArrayExpress ( www.ebi.ac.uk/arrayexpress/) under accession numbers E-MTAB-7823 and E-MTAB-7939.

NGS data were trimmed using TRIMMOMATIC (Bolger *et al.*, 2014) with ILLUMINACLIP parameter set to: 2 : 10 : 5 : 1. Mapping was executed on tomato SL2.50 or *S. pennellii* Spenn_v2.0 genomes ( www.solgenomics.net), and genomic features recovered from the corresponding ITAG2.4 and Spenn_v2.0 annotations. RNA-sequencing (RNA-seq) mapping was performed using STAR (Dobin *et al.*, 2013), with parameter sets –alignEndsType EndToEnd –twopassMode Basic –outReadsUnmapped None –outFilterMultimapNmax 50 –outMultimapperOrder Random –alignMatesGapMax 100000 –alignIntronMax 100000. Beyond the initial analysis, multimapping reads were excluded by remapping with –outFilterMultimapNmax 1. RNA-seq reads from genomic features were counted using HT-seq count (Anders *et al.*, 2015), and further analysed statistically in the R environment with the EDGER package computing the false discovery rate (FDR) through the Benjamini–Hochberg algorithm (Robinson *et al.*, 2010). The present-call threshold was > 2.5 counts-per-million (cpm) in at least two samples, while the statistical threshold was FDR < 0.01. Transcript levels were estimated as mapped read counts normalized by HT-seq total counted library (in cpm). Raw RNA-seq data of sympodial meristems from Park *et al.* (2012) were scrutinized with the same workflow. DNA-seq was mapped with BOWTIE2 (Langmead & Salzberg, 2012), using parameter sets –very-sensitive -X 1000 –non-deterministic. Reads mapping over *MESSI* were prefiltered to diminish ambiguous mapping (first mapped to a masked genome recovering unmapped reads, then remapping these to specific sequences; Sanchez *et al.*, 2017), and were subsequently counted with BEDTOOLS intersect. Open-source software such as SAMTOOLS (Li *et al.*, 2009), PICARD ( http://picard.sourceforge.net.) and DEEPTOOLS (Ramirez *et al.*, 2016) were applied in handling of NGS reads; custom-made workflows for data manipulations and analysis are available at https://github.com/diegohernansanchez/.

M82 tomato meristem smRNAs and DNA methylome raw datasets were shared by Lippman laboratory (Cold Spring Harbor Laboratory, Cold Spring Harbor, MA, USA) and are publicly available (www.solgenomics.net). smRNA data consisted of profiles from vegetative and transition-to-flowering meristems, here merged to increase the depth of available data, and were mapped with BOWTIE2 using the parameter set –score-min L,0,0 –no-mixed –no-discordant –no-unal. Methylomes consisted of profiles from vegetative meristems and were mapped, deduplicated and methylation-called using BISMARK (Krueger & Andrews, 2011). Parameters for bismark mapping were –bowtie2 -N 1 -L 20 -X 1000 -score_min L,0,-0.8 -R 3; whereas the parameter for bismark_methylation_extractor was –comprehensive. Table S2 provides a list of the different NGS datasets produced or used in this study.

## Quantitative real-time RT-PCR

Five micrograms of total RNA was treated with TURBO DNA-free kit (Ambion, Austin, TX, USA), and first-strand cDNA was synthesized by oligo dT priming using the SuperScript III First-Strand Synthesis System (Invitrogen). For real-time PCR we used the LightCycler 480 SYBR Green I Master (Roche), with a 1 : 10 dilution of cDNA in a 10 μl final reaction volume. Cycling and dissociation curves were investigated in a LightCycler 480

Instrument II (Roche). The PCR program was 5 min at 95°C and 45 cycles of 95°C denaturation, 60°C annealing and 72°C extension, each for 10 s. In all samples we used the geometric mean of five housekeeping genes for within-sample normalization: *SlSAND* (Solyc03g115810), *SlPDF2* (Solyc05g009600), *SlEF1A* (Solyc06g005060), *SlUBQ10* (Solyc07g064130) and *SlTIP41* (Solyc10g049850). Primer design, reaction parameters and analysis were performed as described elsewhere (Czechowski *et al.*, 2005). Primer details are given in Table S3.

### *In situ* hybridization

*In situ* hybridization was executed according to Yang *et al.* (2016). Briefly, shoot apical meristems were dissected, fixed in FAA (3.7% formaldehyde, 5% acetic acid, 50% ethanol), and embedded in wax. After being cut into 8 µm sections, the samples were processed with dewaxing, rehydration and dehydration. Sections were then hybridized with a *MESSI* GAG probe at 55°C, and incubated with antidigoxigenin-AP antibody (Roche) for 2 h at room temperature. Hybridization signals were detected via NBT/BCIP (Roche) reaction at 28°C. The probes were obtained by cloning a PCR fragment with primers amplifying a GAG portion of Sly_MESSI_com_17/INT_03_147, with high homology to others. Therefore, the antisense probe reveals transcripts from all *MESSI*. The fragment was cloned into pGEM-T Easy vector (Promega), resequenced, and used as PCR template with primers T7 and SP6. Antisense and sense probes were generated after *in vitro* transcription from these PCR products using the DIG RNA Labelling Kit (Roche). The sense probe assessed nonspecific background binding, showing no evident signals.

## Results

### *MESSI* retrotransposons transcriptionally activate in tomato shoot apical meristems

As a crucial step towards assessing potential LTR retrotransposon activity in meristematic tissues, we performed strand-specific RNA-seq on tomato shoot apical meristems. For comparisons, leaf and flower tissues were also harvested. We investigated these transcriptomes by using a comprehensive annotation of tomato elements tentatively considered active, being full-length (understood here as elements displaying both LTRs) and relatively young (as evaluated by high 5′/3′ LTR sequences similarity, see below). We complemented a previous annotation (Xu & Du, 2014) with our own bioinformatics analysis aimed at finding additional very young LTR retrotransposons, which were automatically *de novo* called from previously recognized LTR retrotransposon-related chromosomal areas (Jouffroy *et al.*, 2016). A target list was constructed including 2335 full-length elements, with 2074 curated from Xu & Du (2014) (designated with the prefix INT) and 261 new additions (designated with the prefix newINT) (Table S1). From these, only 62 passed a stringent present-call threshold aimed at recognizing the reproducible occurrence of retrotransposon-derived transcripts across our whole dataset (Table S4), indicating the apparent absence of global retroelement reactivation in tomato shoot apical meristems. Using stringent statistical tests for robust results (FDR < 0.01 and $\log_2$ fold-change (FC) > 2.5 in two comparisons), we recognized 10 candidates out of the 62 that were differentially up-regulated in meristems compared to leaves and flowers (Table 1). As a first validation approach, we compared the transcript levels of these candidates between our dataset and previously published meristem transcriptomes (Park *et al.*, 2012). The normalized read counts across elements correlated between independent experiments (Fig. S1). As a second validation strategy, we used quantitative real-time RT-PCR with specific primers evaluating independent samples from different tomato tissues; two of the candidates were amenable to unique, specific and efficient PCR amplification in cDNA templates (INT_03_147 and INT_08_206; Fig. S1). High transcript levels from these candidates were detected only in shoot apical meristems regardless of their developmental stage, suggesting persistent tissue-specific activation (Fig. 1a).

With the exception of INT_11_119 belonging to the *copia-like* superfamily (Xu & Du, 2014), the remaining recognized meristem-induced LTR retrotransposons shared sequence homologies in-between LTRs and belonged to the *gypsy-like* superfamily (see below). We focused on these, which we initially named *Meristem-Expressed-Solanum-Specific-Identity*, in short '*MESSI*'. Through additional refined bioinformatics analyses and manual curation (see Materials and methods), we finally annotated 80 complete and 16 incomplete *MESSI* or *MESSI-like* elements in the tomato genome ('complete' referring to those elements presenting known and available whole sequences, as opposed to 'incomplete' where elements feature a variable number of ambiguous Ns from reported suboptimal chromosome sequencing). These were designated with the prefix Sly_MESSI, with indication of their status as complete or incomplete elements as _com or _inc, respectively (Table S5). After present-call filtering and reduced threshold stringency ($\log_2$ FC > 1.5 in at least one comparison), reassessment of our transcriptome data showed that 14 listed *MESSI* were differentially induced in meristems (12 complete and two incomplete), presenting

**Table 1** Tomato LTR retrotransposons differentially expressed in shoot apical meristems as compared with leaves and flowers.

| | Meristems vs leaves | | Meristems vs flowers | |
|---|---|---|---|---|
| LTR retrotransposon | $\log_2$ FC | FDR | $\log_2$ FC | FDR |
| INT_03_147 | 10.88 | 1.55014E-25 | 5.41 | 1.57941E-15 |
| INT_07_162 | 11.60 | 2.90642E-30 | 11.17 | 6.72539E-25 |
| INT_08_159 | 11.32 | 3.77169E-28 | 10.89 | 6.75822E-23 |
| INT_08_206 | 9.97 | 2.05788E-56 | 6.61 | 8.24816E-37 |
| INT_09_24 | 5.03 | 1.76354E-17 | 3.33 | 7.55596E-10 |
| INT_10_167 | 11.68 | 3.7499E-31 | 8.34 | 4.85729E-24 |
| INT_10_173 | 8.40 | 3.79114E-41 | 7.33 | 5.24731E-34 |
| INT_10_180 | 10.04 | 9.06196E-36 | 11.96 | 1.73729E-30 |
| INT_11_119 | 2.63 | 1.47062E-12 | 5.91 | 3.19634E-39 |
| newINT_01_81 | 7.26 | 3.22988E-30 | 2.89 | 1.32323E-09 |

These were from the tomato annotation of young and potentially active LTR retrotransposons. FC, fold change; FDR, false discovery rate.

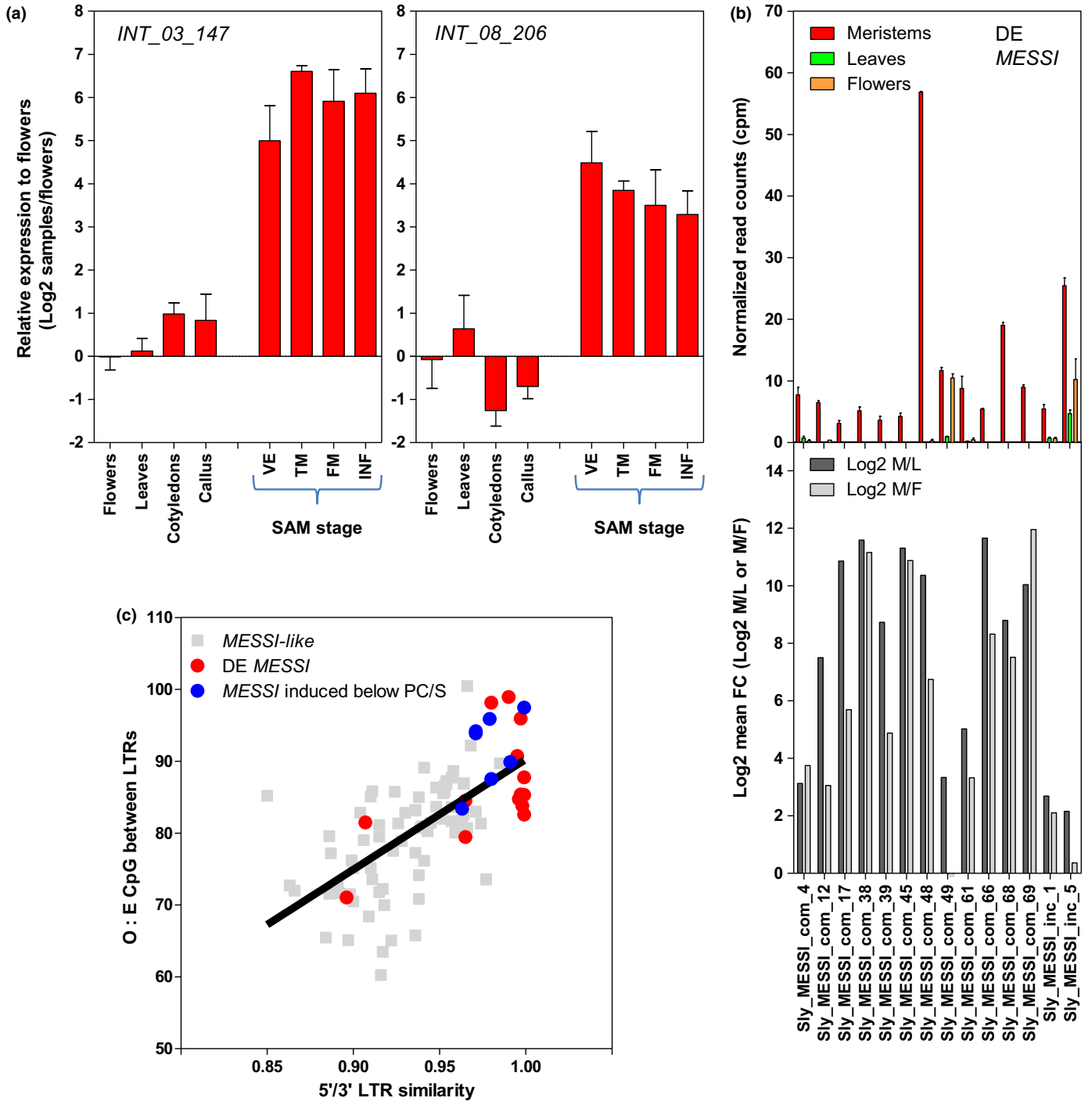


**Fig. 1** *MESSI* LTR retrotransposons are active in tomato shoot apical meristems. (a) Quantitative real-time RT-PCR analysis of transcript levels across different tissues for INT_03_147 and INT_08_206 LTR retrotransposons from Xu & Du (2014). Bars represent mean + SD of duplicated independent biological replicates, depicted on $\log_2$ scale relative to flower samples. SAM, shoot apical meristem; VE, late vegetative stage; TM, transition-to-flower stage; FM, flower stage; INF, sympodial inflorescence. (b) Transcript levels of the 12 complete and two incomplete differentially expressed *MESSI* from RNA-seq data. Upper panel: normalized read count as counts per million (cpm) in tomato meristems (red), leaves (green) or flowers (orange). Bars represent mean + SD of duplicated independent biological replicates. Lower panel: relative expression levels as mean fold changes (FC) on $\log_2$ scale, between meristems and leaves (M/L, dark gray) or between meristems and flowers (M/F, light gray). (c) Correlation across all tomato *MESSI* and *MESSI-like* elements between 5′/3′ LTR similarity (1 = 100%, identical) and the ratio of observed vs expected DNA CpG sites (O : E CpG as %) for sequences in-between LTRs. However, Sly_MESSI_inc_6 and Sly_MESSI_inc_15 were severe outliers for the O : E CpG ratio and were discarded for this analysis. The linear regression model is depicted as a black line ($r^2 = 0.43$). DE, differentially expressed; PC/S, present-call/significance.

negligible or no expression in leaves or flowers (Fig. 1b; Table S6). Seven extra *MESSI* appeared induced in meristems, but their levels were below our present-call filter or significance thresholds; in addition, three old *MESSI* (as inferred by lower 5′/3′ LTR sequence similarity, see below) were not considered as being differentially expressed because they presented activity also in leaves and flowers (Table S6).

The age of a full-length LTR retrotransposon can be estimated by scoring 5′/3′ LTR sequence similarity. Identical LTRs are diagnostic of very recent insertions, while increased dissimilarities reflect degeneration due to accumulated random mutations, diverging LTRs at a rate proportional to the element's age (Pereira, 2004; Wicker & Keller, 2007). 5′/3′ LTR similarity for differentially activated *MESSI* was between 89.6% and 99.9% (Table S6), implying they transposed within the last *c.* 4 million years (Wicker & Keller, 2007). As an independent estimator of relative age, we investigated across all annotated *MESSI* and *MESSI-like* sequences the ratio of observed vs expected CpG sites (O : E CpG). As evolutionary time proceeds, this ratio is expected to decrease in methylated DNA sequences such as TEs, due to C-to-T mutational bias through deamination of 5-methylcytosine (Shen *et al.*, 1994). Remarkably, *MESSI* and *MESSI-like* O : E CpG ratios correlated with 5′/3′ LTR similarity (Fig. 1c). The highest O : E CpG ratio was observed in most meristem-induced *MESSI*, an indication of their very young age (Fig. 1c). Together, these observations confirm the idea that very young and full-length *MESSI* preserve the meristem-activation property. However, note that some recent insertions seem to have already lost meristematic activity, whereas few older elements still display it (Fig. 1c).

## *MESSI* display characteristic features of the *OGRE* group

To analyze the sequence features of tomato *MESSI*, we focused on those that could be considered potentially competent for transposition: differentially expressed, complete full-length and very young insertions (5′/3′ LTR similarity > 99.5%, suggesting an age of no more than *c.* 0.2 million years; Fig. 2). Their median total and LTR sizes were *c.* 20.8 kb and 3.1 kb, respectively, suggesting that giant size is a founding character. This places them among the largest TEs alongside *gypsy-like* retrotransposons of the *OGRE* and *SNARE* families (Neumann *et al.*, 2003; Macas &

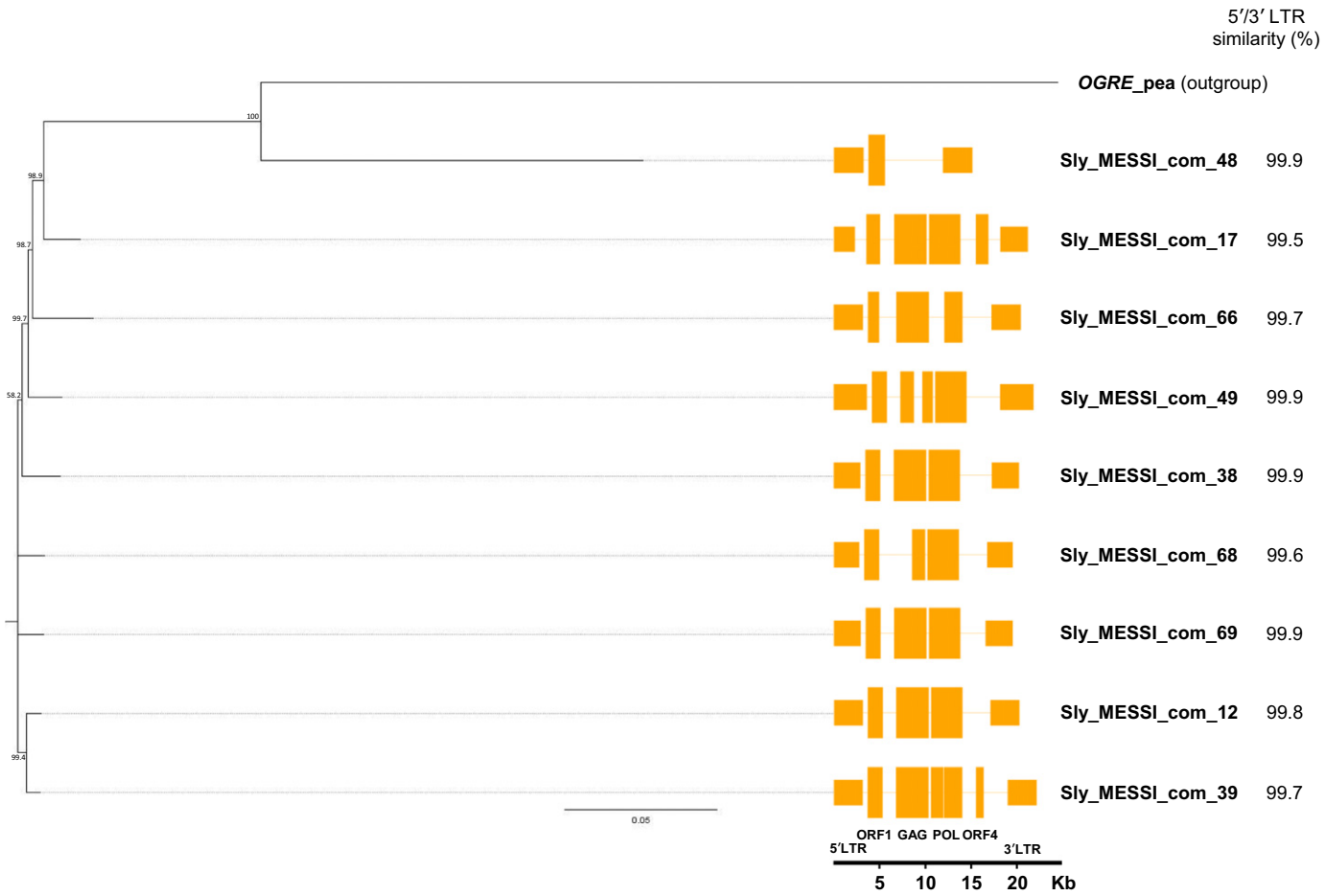


**Fig. 2** *MESSI* elements represent a novel LTR retrotransposon multimember family. Phylogenetic analysis of complete full-length, very young and differentially expressed *MESSI*. The tree was built aligning in-between LTR sequences, and using the Jukes–Cantor model and neighbor-joining method with 10 000 bootstraps. Pea's *OGRE* was used as outgroup (Neumann *et al.*, 2003). LTRs and ORFs are denoted in orange; only those ORFs with 280 or more amino acids were considered.

Neumann, 2007; Du *et al.*, 2010). They shared the primer binding site compatible with priming through tRNA-Arg, similar to the *OGRE* family (Fig. S2a; Macas & Neumann, 2007). As they presented high homology in-between LTRs (> 96% identity at the nucleotide level for any given pairwise comparison), *MESSI* can be defined as a multimember family (Wicker *et al.*, 2007). They displayed at least three nonoverlapping coding areas, the second and third being enlarged GAG/protease and POL (reverse transcriptase/ribonuclease H/integrase), each of *c.* 3400–3600 bp (Fig. 2). According to the coding arrangement within POL, *MESSI* belong to the *gypsy-like* superfamily. Despite their young age, some of them showed mutated GAG and/or POL, and could be theoretically considered nonautonomous elements. Before GAG, these *MESSI* presented an ORF1 of *c.* 1600 bp, also reported for other *gypsy-like* elements such as *OGRE* from pea (Neumann *et al.*, 2003; Steinbauerova *et al.*, 2011). However, we found no protein homology between ORF1 from *MESSI* and *OGRE* (Fig. S2b), implying that these may have different functions despite the similar placement within the retrotransposons. *MESSI* ORF1 showed no known protein domains in public protein databases. Furthermore, all but one presented an area with tandem repeats upstream of the 3′ LTR (Fig. S2c), a feature reported also for *OGRE* and *SNARE* families (Macas & Neumann, 2007; Du *et al.*, 2010). In two cases, an ORF4 resulted from these repeats (Fig. 2), but we did not find any homology database hits. Although it is not uncommon for plant *gypsy-like* elements to display extra ORFs and tandem repeats, their functional roles remain enigmatic (Steinbauerova *et al.*, 2011). Of note, Sly_MESSI_com_48 (INT_08_206 in Xu & Du, 2014) presented a smaller size and derived heavily mutated GAG while lacking POL sequences, but we recognized a complete ORF1 with low homology to the others (*c.* 35% identity at the protein sequence level); overall this suggests a nonautonomous element which lineage diverged a long time ago. The rest of the differentially expressed *MESSI* not considered before comprised older elements (lower 5′/3′ LTR sequence similarity, Table S6), with mutated and reduced ORFs but still in-between LTRs homology to the very young elements (> *c.* 84% nucleotide sequence identity in > *c.* 83% cover for pairwise comparisons); therefore, they probably represent aged *MESSI* remnants still conserving transcriptional capabilities.

Although a typical character of LTR retrotransposon families is the high homology of coding areas across members, with > 80-% sequence identity, many families sometimes present an identity below 30% in LTR sequences (Casacuberta & Santiago, 2003; Wicker *et al.*, 2007; Grandbastien, 2015). Compared to coding regions, LTR homologies across very young *MESSI* tended to be low (overall *c.* 48% pairwise identity). Nonetheless, interspersed by indels in some elements, high homology blocks comprising *c.* 250 bp were shared at the 5′ and 3′ edges of LTRs (Fig. S2d). Remarkable similarity was manifest in the first 30–32 bp block of the 5′ LTR edge, which probably includes conserved an integrase-binding site, as expected from elements of the same family (Wicker *et al.*, 2007). This was also true for the putative nonautonomous Sly_MESSI_com_48/INT_08_206 (Fig. S2).

We further analyzed in more detail the RNA-seq mapped data. These showed a nonrandom distribution of reads resulting from the coding strand, mostly in GAG and POL areas, and in some cases the LTRs (Figs 3a,b, S3). This may imply selective splicing of spacer sequences between ORFs, like pea's *OGRE* (Neumann *et al.*, 2003). Interestingly, Sly_MESSI_com_48/INT_08_206 presented high expression levels of LTRs and ORF1, with apparent long-distance spillover transcription towards downstream sequences (Fig. S3). Older meristem-activated *MESSI* showed reads mapping only to the LTRs, with no visible expression of

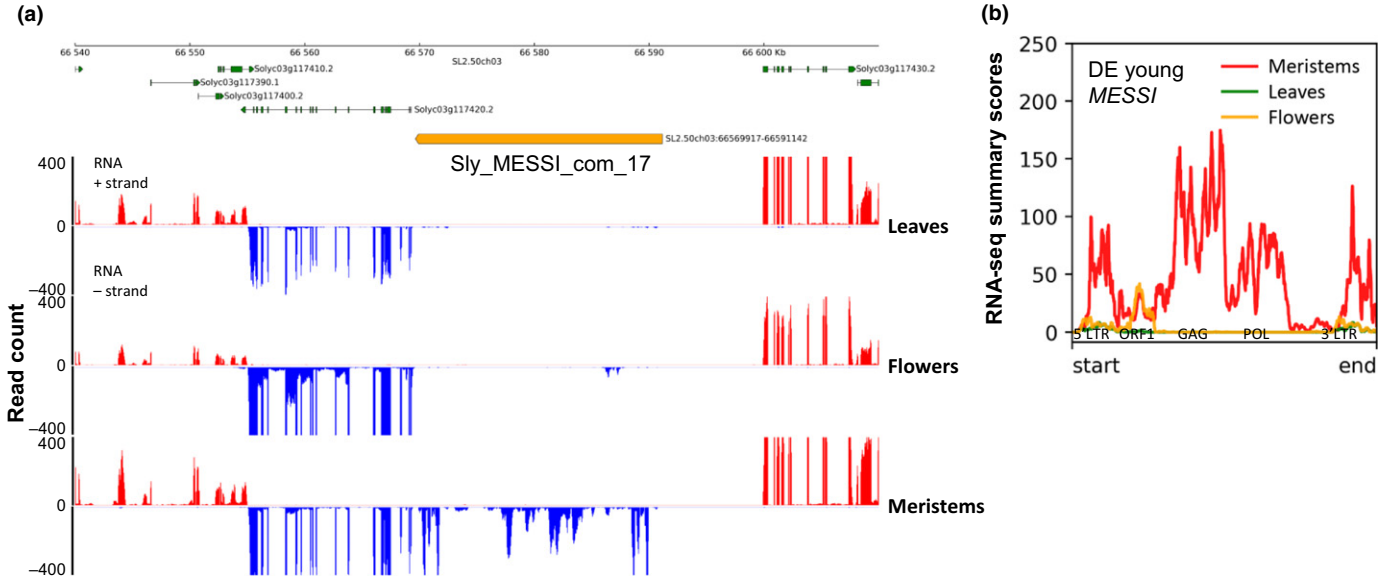


**Fig. 3** Transcriptional profile of *MESSI* LTR retrotransposons. (a) Example genome browser plot of tomato leaf, flower and shoot apical meristem RNA-seq data mapped to genomic coordinates on chromosome 3 around Sly_MESSI_com_17/INT_03_147. Plus strand reads are depicted in red and minus strand reads in blue. Genes and *MESSI* models on top are colour coded green and orange, respectively. (b) Scaled summary plot for RNA-seq data over the entire sequence of very young differentially expressed *MESSI*.

other areas (Fig. S3). Taking these together, we conclude that *MESSI* represents an unusual *gypsy-like* multimember plant LTR retrotransposon family, with transcriptional competence mainly in tomato shoot apical meristems and unorthodox features related to the *OGRE* group (Macas & Neumann, 2007).

## Evidence of developmental escape from TGS in *MESSI* family

We speculated that developmental relaxation of TGS may be responsible for *MESSI* activation in shoot apical meristems (Martinez & Slotkin, 2012). To test this, we first analyzed available tomato meristem smRNA data to assess whether the *MESSI* family is a recognized target of the RdDM machinery (Fultz *et al.*, 2015). Perfect-matching uniquely mapped smRNAs showed enrichment for 24-nucleotide smRNAs in all documented *MESSI* and *MESSI-like* elements together, and also for just the differentially expressed *MESSI* (Fig. 4a). The ratio 24 + 23-nucleotide : 21 + 22-nucleotide smRNAs was 3.69 for all *MESSI* and *MESSI-like* and 4.63 for only the differentially expressed *MESSI*, compared to 1.55 for the total smRNA population. Very young differentially expressed *MESSI* displayed smRNAs targeting both DNA strands, with an apparent bias towards the 5′ end sequences (Figs 4b, S4). These

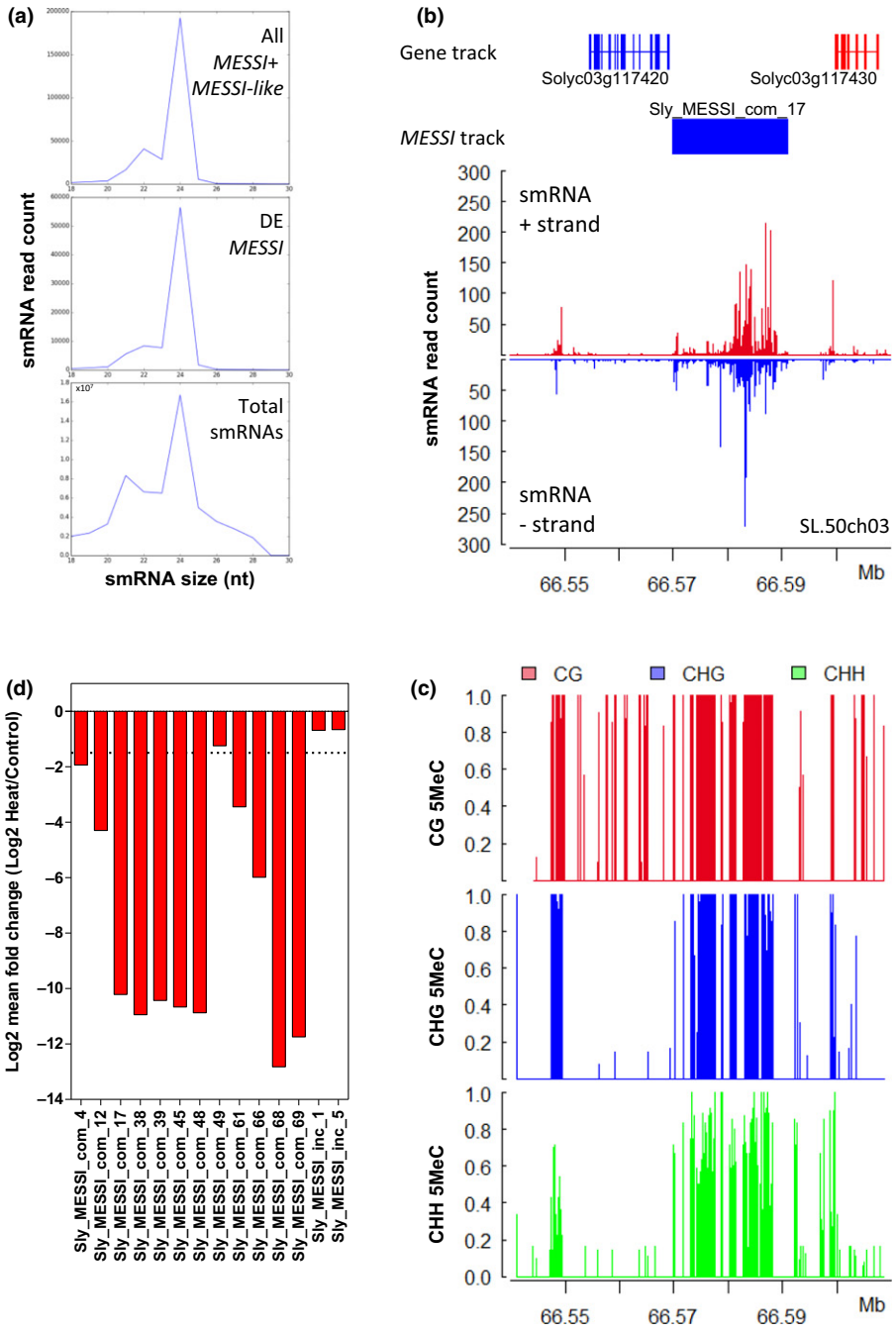


**Fig. 4** *MESSI* LTR retrotransposons are targeted by RdDM in shoot apical meristems, but are refractory to heat-induced release from silencing. (a) smRNA count vs size in profiles of tomato shoot apical meristems for: top, all *MESSI* and *MESSI-like* elements; middle, meristem differentially expressed *MESSI*; bottom, whole smRNA universe. DE, differentially expressed; nucleotide, nucleotides. (b) and (c) Example genome browser of tomato shoot apical meristem smRNA and DNA methylation profiles, mapped to genomic coordinates on chromosome 3 around Sly_MESSI_com_17/INT_03_147. (b) smRNA counts. Reads are depicted in red for 5′ to 3′ plus strand and in blue for 3′ to 5′ minus strand. Genes and *MESSI* models on top are color coded according to their coding strand. (c) Cytosine methylation profile (1 = 100% methylation) for the three DNA contexts CG (red), CHG (blue) and CHH (green); H = C, A or T. Genes and *MESSI* models correspond to (b). (d) Relative transcript levels of the 14 differentially expressed *MESSI* in tomato shoot apical meristems under heat stress. Bars represent the relative expression as mean fold changes (FC) in $\log_2$ scale between heat and control samples.

observations suggest an effective recognition of *MESSI* sequences by the tomato canonical RdDM pathway. Because this pathway mediates epigenetic silencing of target sequences through DNA methylation (Lisch, 2009; Fultz *et al.*, 2015; Matzke *et al.*, 2015), we then investigated available methylome profiles in tomato meristems. We observed cytosine methylation signals in the three DNA contexts (CG, CHG and CHH) along the whole sequences of very young differentially expressed *MESSI*, typical of RdDM-dependent DNA methylation (Figs 4c, S4). Visual inspection of freely available data in the form of genome browser profiles also showed smRNA and DNA methylation over *MESSI* coordinates in tomato fruit samples (accessible through the Tomato Epigenomics Database: http://ted.bti.cornell.edu/epigenome/index.html; Zhong *et al.*, 2013). Taking this together, we conclude that the *MESSI* family is recognized and targeted by TGS in tomato shoot apical meristems and other tissues, implying that epigenetic surveillance is not completely relaxed for their transcriptional competence. Therefore, these TEs may have evolved an innate capacity to developmentally evade some layer of the epigenetic silencing. Because completely unrestrained and fully activated LTR retrotransposons typically generate copious life-cycle intermediates in the form of extrachromosomal DNA (Mirouze *et al.*, 2009; Ito *et al.*, 2011), we wondered if this was also the case for *MESSI*. We performed tissue-specific PCR-free DNA-seq and counted reads mapping to the target TEs, potentially revealing not only extrachromosomal DNA but also putative new chromosomal copies (Sanchez *et al.*, 2017). However, tomato shoot apical meristems did not show an increased number of DNA-seq reads mapping to the very young transcriptionally activated *MESSI* (Fig. S5). This suggests there is no substantial meristematic proliferation of *MESSI* DNA copies, providing initial evidence for the absence of significant extrachromosomal DNA synthesis.

It has been recognized that transcriptional stimulation of silenced TEs and genes may be initiated in plant vegetative tissues under long-term heat treatment, which disturbs TGS (Sanchez & Paszkowski, 2014). As target of TGS, we wondered whether heat reinforces *MESSI* activity in meristematic tissues. We thus performed whole transcriptome analysis of shoot apical meristems from tomato plants subjected to long-term heat. Examination of heat-responsive gene markers corroborated that the temperature treatment was effective; moreover, we were able to recognize the triggering of several heat-induced LTR retrotransposons (Fig. S6a). However, differentially expressed *MESSI* elements did not show an induction upon this temperature shift. Most of them were actually repressed, with a few exceptions that did not change significantly (Fig. 4d). The majority of the other recognized *MESSI-like* elements showed no differential behavior after the heat treatment, with the exception of Sly_MESSI_com_1 and the previously documented older *MESSI* with activity across tissues, which also seemed to be repressed (Fig. S6b). Although it cannot be ruled out that other environmental conditions influence their induction, our results support the idea that *MESSI* escape from TGS may be mediated by developmental rather than stress signals.

We then investigated whether *MESSI* escape from TGS in a cell-type-specific fashion. For this, we analyzed the spatial distribution of *MESSI* transcripts using *in situ* hybridization in tomato shoot apical meristems at different developmental stages. Remarkably, we observed area-specific signals. *MESSI* appeared transcriptionally competent in the outer layers and central zones of meristems in the transition-to-flowering stage, and also at the boundaries between the meristems and initiating organs (either leaf or inflorescence primordia) (Figs 5, S7). This configuration evokes the expression patterns of some plant developmental transcription factors involved in the specification of stem cells and organ boundaries, such as *Arabidopsis thaliana SHOOT MERISTEMLESS* (Long & Barton, 2000; Landrein *et al.*, 2015). Furthermore, strong signals were observed at the primordia and flank of initiating leaflets (Figs 5, S7), reminiscent of organ boundary transcription factors such as tomato *GOBLET* (Berger *et al.*, 2009). Organ boundary areas are known to have key

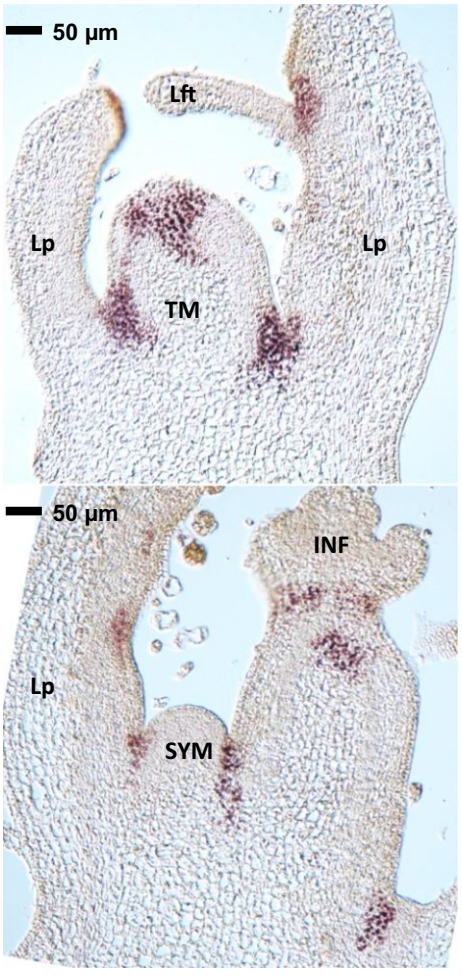


**Fig. 5** *In situ* hybridization for *MESSI* LTR retrotransposons. *MESSI*'s antisense GAG hybridization signals were detected by the NBT/BCIP reaction in 8 µm sections in representative tomato shoot apical meristems. Upper panel: transition meristem stage. Lower penal: inflorescence and sympodial vegetative meristem stage. Lp, leaf primordia; Lft, leaflets; TM, transition meristem; INF, sympodial inflorescence; SIM, sympodial meristem.

developmental functions such as the formation and maintenance of all meristems and compound-leaf patterning (Wang *et al.*, 2016). We conclude that these observed developmentally associated patterns are compatible with *MESSI* transcriptional activation in strategic meristematic cell lineages possibly contributing to the germline.

### *MESSI* elements occur across *Solanum*

It can be hypothesized that *MESSI*-specific meristematic escape from TGS in the wild-type background under normal growing conditions may have evolved recently in tomato. Alternatively, it may be an older attribute shared by many *MESSI-like* retrotransposons within the *Solanum* clade. To answer this, we used comparative genomics: in a first step we attempted to annotate *MESSI-like* elements in several fully sequenced *Solanum* species, while in a second step we aimed to analyze shoot apical meristem activity in a set of predicted elements. Although low sequence quality in some of the available genomes may hamper the finding of such large TEs, our bioinformatics analyses using homologies to tomato *MESSI* did recognize full-length *MESSI-like* elements in all explored *Solanum* (Table 2). They shared the characteristic giant size, between 18 and 21 kb on average, with some extreme examples reaching *c.* 33 kb (Fig. S8). In all cases, the documented *MESSI-like* elements peaked in numbers between 96% and 92% 5′/3′ LTR similarity (Fig. S8), diagnostic of maximal historical transposition bursts occurring between *c.* 1.5 and *c.* 3 million years ago (Wicker & Keller, 2007). This implies an overall low recent mobilization of this family. However, from this it can be inferred that they were more intensely active in *S. tuberosum* and *S. melongena* long after their divergence from tomato (*c.* 8 and 14 million years ago, respectively; Sarkinen *et al.*, 2013). Also, because *S. arcanum*, *S. habrochaites* and *S. pennellii* split from tomato within the last 2–3 million years (Sarkinen *et al.*, 2013), *MESSI* must have been active during and after these speciation events. We therefore focused only on a subset of relatively modern *MESSI-like* elements across *Solanum* genomes with 5′/3′ LTR similarity ≥ 97.5%, which denotes insertions within the last *c.* 1 million years, thus ensuring their transposition events took place after the speciation of their hosts. Most of them belonged to tomato and *S. pennellii*, and phylogenetic analyses of their aligned sequences suggested shared ancestry (Fig. S9). Taking this evidence together, we conjecture that *Solanum* species probably present a low number of extant young *MESSI* insertions. We then aimed at exploring putative *MESSI* competence in *S. pennellii* shoot apical meristems (designated with the prefix Spenn_MESSI, Table S7), performing RNA-seq analysis in comparison with leaves and flowers. After present-call filtering and statistical assessment for differential expression ($\log_2$ FC > 1.5 in at least one comparison), we observed differential transcriptional activation in *S. pennellii* shoot apical meristems of some of the predicted *MESSI* elements (Fig. 6a; TableS8). However, most were deemed old, with 5′/3′ LTR similarity ranging between 97.7% and 88.8% (Table S8), implying an estimated historical mobilization between *c.* 0.88 and *c.* 4.6 million years ago. RNA-seq reads mapped to these elements showed transcript signals mostly in one or both LTRs, and in some cases not only in meristems but also in leaves and flowers, resembling some of the old tomato *MESSI* (Figs 6b,c, S10). We interpret these data as evidence that *MESSI* elements in *S. pennellii* are now probably transpositionally inactive, but with some old remnants still conserving the meristem transcriptional competence distinctive of their tomato counterparts. Together, this could imply that the escape from TGS during development evolved in *MESSI* ancestors before the split between tomato and *S. pennellii* lineages. However, the meristem-activated elements may be posed to disappear in the latter, probably due to aging.

**Table 2** Number of *MESSI* and *MESSI-like* LTR retrotransposons annotated in different *Solanum* genomes.

| Species | Complete | Incomplete |
|---|---|---|
| *S. lycopersicum* | 80 | 16 |
| *S. arcanum* | 75 | – |
| *S. habrochaites* | 82 | – |
| *S. pennellii* | 78 | 35 |
| *S. tuberosum* | 21 | 41 |
| *S. melongena* | 7 | 15 |

Incomplete elements refers to those in which the sequences in-between the LTRs are at present not entirely known, displaying a variable number of ambiguous bases represented by Ns.

## Discussion

### Evidence of the *MESSI* family evading layers of epigenetic silencing

Under natural conditions, nonsilent TEs must be those engaging in bursts of transposition (Lisch & Bennetzen, 2011; Bousios & Gaut, 2016). It has been speculated that such a dearth of host silencing pressure may be mediated by natural variation at epigenetic regulators, by environmental/genomic challenges leading to a transient failure of the epigenetic machinery, by a programmed developmental relaxation of TGS, or due to the invasion of TEs not yet recognized by the target genome as in horizontal transfer (Weil & Martienssen, 2008; Blumenstiel, 2011; Lisch & Bennetzen, 2011; Martinez & Slotkin, 2012; Bousios & Gaut, 2016; Lanciano & Mirouze, 2018). However, a plausible and nonmutually exclusive alternative may be the existence of inherent TE attributes that could have been selected as adaptations to counteract or evade host suppression, allowing sporadic activation and transposition of recognized and otherwise silenced elements. In plants, scenarios of this kind have been documented in class II 'cut-and-paste' type TEs. For example, the *VANDAL21* family of *A. thaliana* was reported to sport an antisilencing factor that mediates DNA methylation loss of TEs (Hosaka *et al.*, 2017). By contrast, the rice autonomous *Ping* is actively transcribed despite DNA methylation of regulatory terminal sequences, transposing during the propagation of rice landraces with shared ancestry together with its nonautonomous *mPing* partner; this supports the idea that some TEs may efficiently elude silencing (Lu *et al.*, 2017). To the best of our knowledge, no operative antisilencing

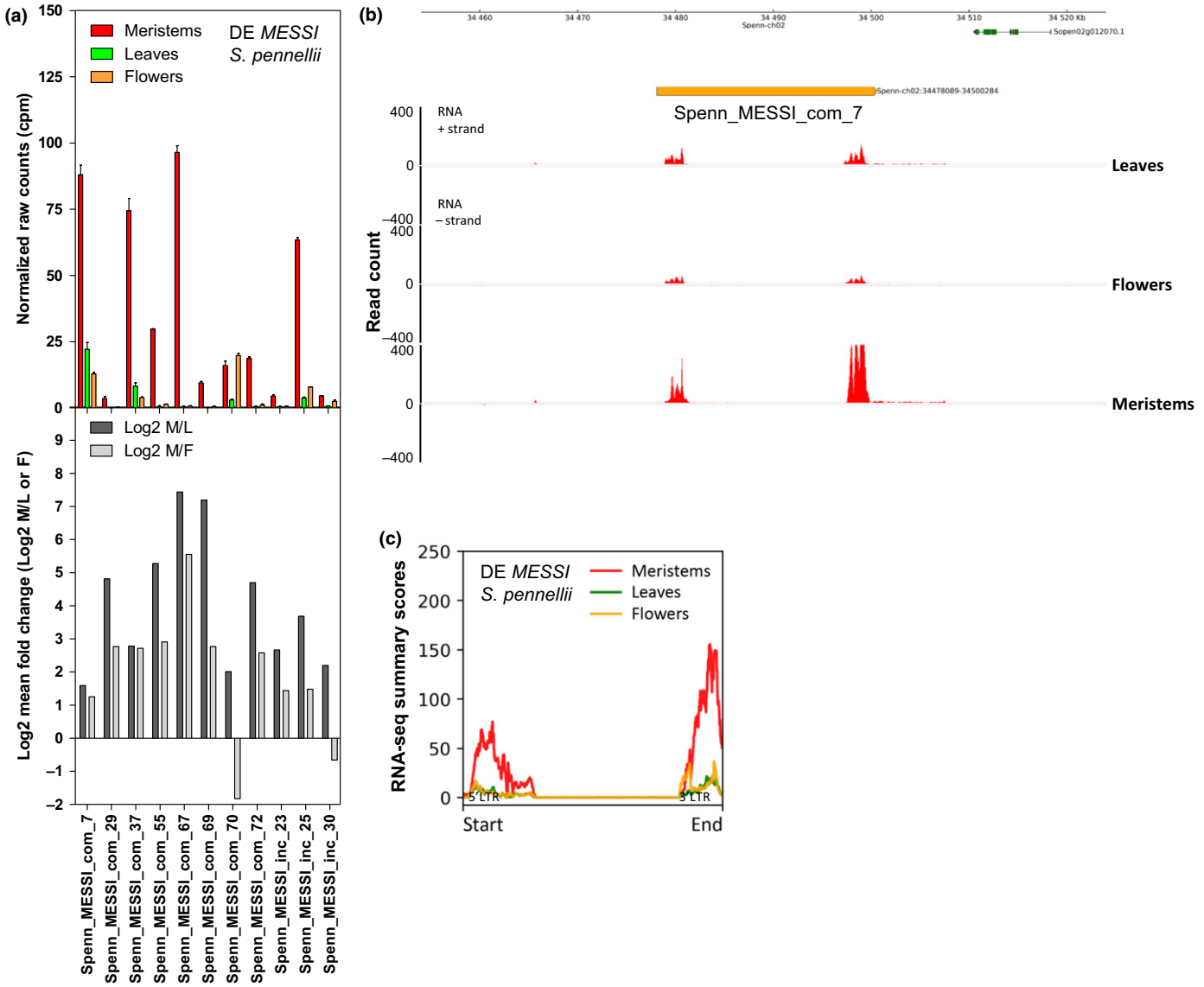


**Fig. 6** Transcriptional profile of *MESSI* LTR retrotransposons in *Solanum pennellii*. (a) Transcript levels of complete differentially expressed *S. pennellii MESSI* from RNA-seq data. Upper panel: normalized read count as counts-per-million (cpm) in meristems (red), leaves (green) or flowers (orange). Bars represent mean + SD of duplicated independent biological replicates. Lower panel: relative expression levels as mean fold changes (FC) on $\log_2$ scale, between meristems and leaves (M/L, dark gray) or between meristems and flowers (M/F, light gray). (b) Example genome browser plot of *S. pennell* leaf, flower and shoot apical meristem RNA-seq data mapped to genomic coordinates on chromosome 2 around Spenn_MESSI_com_7. Plus strand reads are depicted in red and minus strand reads in blue. Genes and *MESSI* models on top are colour coded green and orange, respectively. (c) Scaled summary plot for RNA-seq data over the entire sequence of complete differentially expressed *S. pennellii MESSI*.

factors have been described for LTR retrotransposons, but mounting evidence suggests that several plant retrotransposon families could override certain layers of epigenetic suppression. For instance, inductive environmental conditions activated *ONSEN/COPIA78* of *A. thaliana* and *Tnt1* of tobacco despite TGS recognition and targeting (Ito *et al.*, 2011; Hernandez-Pinzon *et al.*, 2012; Sanchez *et al.*, 2017). Other plant retrotransposons such as *RIDER* of tomato, *OGRE* of pea, *SARE*$^{A}$ of soybean and *BARE1* of barley appeared transcriptionally unrestricted – and certainly translationally competent at least in the last case – across most wild-type tissues where the epigenetic machinery is intact (Neumann *et al.*, 2003; Jiang *et al.*, 2009; Du *et al.*, 2010; Jaaskelainen *et al.*, 2013). However, no particular developmental pattern of evasion has been reported in these cases. Our results showed that *MESSI*, a novel giant retrotransposon family, appears transcriptionally competent in tomato shoot apical meristems in spite of being recognized and targeted by the TGS machinery. At present, it cannot be ruled out that silencing is relaxed specifically in those cell lines presenting *MESSI in situ* hybridization signals. Nevertheless, the pattern of transcript spatial distribution supports the idea that *MESSI* hijack tomato's developmental signaling to elude those layers of epigenetic surveillance that evolved to restrict the initial stimulation of TEs. This behavior is remarkable, considering that the activation in

cell lineages supplying genetic information to the host's next generation must be a prerequisite for the survival and evolution of genomic parasites (Blumenstiel, 2011). It is conceivable that, during plant ontogeny, *MESSI* may take advantage of their eluding activity to secure new copies in meristematic tissues carrying the plant germline, which differentiates late in vegetative development (Lanfear, 2018). Currently, the underlying molecular bases of such developmental escape remain unknown. It is well described that *cis*-elements within LTR sequences mediate the initial transcriptional activation, effected by transcription factors from the host (Grandbastien, 2015; Galindo-Gonzalez *et al.*, 2017). Compelling evidence demonstrates the involvement of stress-associated factors in the trigger of environmentally responsive plant retrotransposons (Vernhettes *et al.*, 1997; Cavrak *et al.*, 2014; Pietzenuk *et al.*, 2016; Zervudacki *et al.*, 2018). Thus, for the case concerned here, we venture to predict that evasion is probably mediated by specific plant developmental transcription factors recognizing regulatory sequence motifs presumably occurring in *MESSI* LTRs, subsequently recruiting the Pol II-dependent machinery and perhaps simply overpowering TGS. As an alternative to overpowering TGS, triggering competence might depend critically upon silencing targeting *MESSI* sequences, for example, if transcriptional activity would require or may be enhanced by DNA methylation (Williams *et al.*, 2015). The validation of such hypothesis, along with the exploration of the identity and nature of the corresponding *cis*-elements and *trans*-acting factors, remains a green venue for future research. In addition, the finding of *MESSI* transcript accumulation despite being targeted by 21 + 22-nucleotide smRNAs might also point to an escape from PTGS; however, additional research would be necessary to confirm that these smRNAs truly mediate cleavage of *MESSI* transcripts or instead inhibit their translation.

### Escape from silencing evolved in *Solanum* ancestors

It is interesting to consider when *MESSI* developmentally associated escape evolved. Our observation that old *MESSI* remnants retain only LTR activity, in tomato as well as in *S. pennellii*, may be interpreted as aging acting on elements originally presenting the tissue-specific evasion property. Alternatively, transcriptional competence in old LTRs across different tissues in either tomato or *S. pennellii* may point to a common origin as constitutively active TEs, on which tissue-specific triggering was later adaptively selected. Further research efforts are underway to ascertain these possibilities; but under both scenarios, the timing for evolution of evasion is taken back to a common ancestor pre-dating the tomato and *S. pennellii* speciation event. Regardless, the proficient escape from TGS might be a prerequisite for the subsistence of oversized TEs such as *MESSI*, given that longer eukaryotic elements are more efficiently selected against due to an increased likelihood of deleterious ectopic recombination (Petrov *et al.*, 2011). Supporting this idea, giant dicot retrotransposons such as *OGRE* and *SARE*[4] appeared constitutively active across tissues, although this attribute is also shared by other more standard-sized monocot elements (Neumann *et al.*, 2003; Jiang *et al.*, 2009; Du *et al.*, 2010; Jaaskelainen *et al.*, 2013). Also, the escape property seems to be shared by tomato INT_11_119, which belongs to a different superfamily. Together, this suggest that LTR retrotransposon evasion strategies, whatever the underlying mechanisms, must have evolved many times independently during plant kingdom evolution.

### Defenses other than TGS probably restrain *MESSI*

In view of their age, homologies, variable numbers and nonsyntenic chromosomal locations, we infer that *MESSI* undoubtedly mobilized after the divergence of their *Solanum* hosts. However, we have not yet empirically recorded a contemporary burst of tomato young *MESSI*. As with other TE families, transposition may rarely occur under normal circumstances, indicating that additional layers of the host epigenetic surveillance may effectively restrain *MESSI* downstream of their initial activation step (Bucher *et al.*, 2012). Indeed, our initial results hint at the absence of substantial amounts of extrachromosomal DNA intermediaries in tomato meristems despite *MESSI* characteristic expression, suggesting that their life cycle is most likely being blocked post-transcriptionally by other defenses before the initiation of genomic RNA reverse-transcription. This interpretation is compatible with the existence of host multilayer silencing mechanisms individualized for particular TEs, as described for the LTR retrotransposon *EVADE/COPIA93* in *A. thaliana* (Mirouze *et al.*, 2009); however, the evolution of self-restraint in *MESSI* cannot at present be ruled out (Tucker *et al.*, 2015; Gaubert *et al.*, 2017). Yet unexplored physiological conditions may be required to increase the occurrence likelihood of *MESSI* operative mobilization. Nonetheless, our results provide preliminary evidence that not only sensed external environmental signals, but also internal developmental signals, may be exploited by particular retrotransposons for survival within plant genomes.

## Acknowledgements

We thank Thomas Jones, Carlos Bayon and Jayne Griffiths for assistance with RNA-seq library preparation/sequencing and technical help, and Jerzy Paszkowski for general support and for providing reagents and access to the equipment required. We are most grateful to Zachary Lippman and Zachary Lemmon for generous sharing of unpublished smRNA and DNA methylation data. We also greatly appreciate the generous and opportune contribution of *S. pennellii* seeds by Dani Zamir, Itay Zemach, Itai Ofner, Michal Lieberman-Lazarovich and Ilan Levin. DHS would like to acknowledge Hajk-Georg Drost, Varodom Charoensawan, Hugo Tavares, Jeremy Gruel, Anna Gogleva and Yassin Refahi for continuous and kind support with bioinformatics, and Adrian Valli for comments on the manuscript.

## Author contributions

DHS performed the experiments with exception of callus tissue-culture, DNA-seq and *in situ* hybridizations, the first two executed by HG and the third by WY. DHS examined the data, performed the bioinformatics analyses and wrote the paper.

## ORCID

Diego H. Sanchez https://orcid.org/0000-0002-4234-1124
Weibing Yang https://orcid.org/0000-0002-2379-5729

## References

**Anders S, Pyl PT, Huber W. 2015.** HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

**Baubec T, Finke A, Mittelsten Scheid O, Pecinka A. 2014.** Meristem-specific expression of epigenetic regulators safeguards transposon silencing in Arabidopsis. *EMBO Reports* **15**: 446–452.

**Benson G. 1999.** Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**: 573–580.

**Berger Y, Harpaz-Saad S, Brand A, Melnik H, Sirding N, Alvarez JP, Zinder M, Samach A, Eshed Y, Ori N. 2009.** The NAC-domain transcription factor GOBLET specifies leaflet boundaries in compound tomato leaves. *Development* **136**: 823–832.

**Blumenstiel JP. 2011.** Evolutionary dynamics of transposable elements in a small RNA world. *Trends in Genetics* **27**: 23–31.

**Bolger AM, Lohse M, Usadel B. 2014.** Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

**Bousios A, Gaut BS. 2016.** Mechanistic and evolutionary questions about epigenetic conflicts between transposable elements and their plant hosts. *Current Opinion in Plant Biology* **30**: 123–133.

**Bucher E, Reinders J, Mirouze M. 2012.** Epigenetic control of transposon transcription and mobility in Arabidopsis. *Current Opinion in Plant Biology* **15**: 503–510.

**Casacuberta JM, Santiago N. 2003.** Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. *Gene* **311**: 1–11.

**Cavrak VV, Lettner N, Jamge S, Kosarewicz A, Bayer LM, Mittelsten Scheid O. 2014.** How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genetics* **10**: e1004115.

**Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR. 2005.** Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiology* **139**: 5–17.

**Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013.** STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

**Du J, Tian Z, Bowen NJ, Schmutz J, Shoemaker RC, Ma J. 2010.** Bifurcation and enhancement of autonomous-nonautonomous retrotransposon partnership through LTR swapping in soybean. *Plant Cell* **22**: 48–61.

**Edgar RC. 2004.** MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.

**Ellinghaus D, Kurtz S, Willhoeft U. 2008.** LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinformatics* **9**: 18.

**Fukai E, Umehara Y, Sato S, Endo M, Kouchi H, Hayashi M, Stougaard J, Hirochika H. 2010.** Derepression of the plant Chromovirus LORE1 induces germline transposition in regenerated plants. *PLoS Genetics* **6**: e1000868.

**Fultz D, Choudury SG, Slotkin RK. 2015.** Silencing of active transposable elements in plants. *Current Opinion in Plant Biology* **27**: 67–76.

**Galindo-Gonzalez L, Mhiri C, Deyholos MK, Grandbastien MA. 2017.** LTR-retrotransposons in plants: engines of evolution. *Gene* **626**: 14–25.

**Gaubert H, Sanchez DH, Drost HG, Paszkowski J. 2017.** Developmental restriction of retrotransposition activated in Arabidopsis by environmental stress. *Genetics* **207**: 813–821.

**Grandbastien M-A. 1998.** Activation of plant retrotransposons under stress conditions. *Trends in Plant Science* **3**: 181–187.

**Grandbastien MA. 2015.** LTR retrotransposons, handy hitchhikers of plant regulation and stress response. *Biochimica et Biophysica Acta* **1849**: 403–416.

**Hernandez-Pinzon I, Cifuentes M, Henaff E, Santiago N, Espinas ML, Casacuberta JM. 2012.** The Tnt1 retrotransposon escapes silencing in tobacco, its natural host. *PLoS ONE* **7**: e33816.

**Hollister JD, Gaut BS. 2009.** Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research* **19**: 1419–1428.

**Hosaka A, Saito R, Takashima K, Sasaki T, Fu Y, Kawabe A, Ito T, Toyoda A, Fujiyama A, Tarutani Y *et al.* 2017.** Evolution of sequence-specific anti-silencing systems in Arabidopsis. *Nature Communications* **8**: 2161.

**Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. 2011.** An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **472**: 115–119.

**Jaaskelainen M, Chang W, Moisy C, Schulman AH. 2013.** Retrotransposon BARE displays strong tissue-specific differences in expression. *New Phytologist* **200**: 1000–1008.

**Jiang N, Gao D, Xiao H, van der Knaap E. 2009.** Genome organization of the tomato sun locus and characterization of the unusual retrotransposon Rider. *The Plant Journal* **60**: 181–193.

**Jouffroy O, Saha S, Mueller L, Quesneville H, Maumus F. 2016.** Comprehensive repeatome annotation reveals strong potential impact of repetitive elements on tomato ripening. *BMC Genomics* **17**: 624.

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

**Kumar A, Bennetzen JL. 1999.** Plant retrotransposons. *Annual Review of Genetics* **33**: 479–532.

**Lanciano S, Mirouze M. 2018.** Transposable elements: all mobile, all different, some stress responsive, some adaptive? *Current Opinion in Genetics & Development* **49**: 106–114.

**Landrein B, Kiss A, Sassi M, Chauvet A, Das P, Cortizo M, Laufs P, Takeda S, Aida M, Traas J *et al.* 2015.** Mechanical stress contributes to the expression of the STM homeobox gene in Arabidopsis shoot meristems. *eLife* **4**: e07811.

**Lanfear R. 2018.** Do plants have a segregated germline? *PLoS Biology* **16**: e2005439.

**Langmead B, Salzberg SL. 2012.** Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

**Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009.** The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

**Lisch D. 2009.** Epigenetic regulation of transposable elements in plants. *Annual Review of Plant Biology* **60**: 43–66.

**Lisch D, Bennetzen JL. 2011.** Transposable element origins of epigenetic gene regulation. *Current Opinion in Plant Biology* **14**: 156–161.

**Long J, Barton MK. 2000.** Initiation of axillary and floral meristems in Arabidopsis. *Developmental Biology* **218**: 341–353.

**Lu L, Chen J, Robb SMC, Okumoto Y, Stajich JE, Wessler SR. 2017.** Tracking the genome-wide outcomes of a transposable element burst over decades of amplification. *Proceedings of the National Academy of Sciences, USA* **114**: E10550–E10559.

**Macas J, Neumann P. 2007.** Ogre elements – a distinct group of plant Ty3/gypsy-like retrotransposons. *Gene* **390**: 108–116.

**Mari-Ordonez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O. 2013.** Reconstructing *de novo* silencing of an active plant retrotransposon. *Nature Genetics* **45**: 1029–1039.

**Martinez G, Slotkin RK. 2012.** Developmental relaxation of transposable element silencing in plants: functional or byproduct? *Current Opinion in Plant Biology* **15**: 496–502.

**Matzke MA, Kanno T, Matzke AJ. 2015.** RNA-directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. *Annual Review of Plant Biology* **66**: 243–267.

**Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D, Paszkowski J, Mathieu O. 2009.** Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* **461**: 427–430.

**Neumann P, Pozarkova D, Macas J. 2003.** Highly abundant pea LTR retrotransposon Ogre is constitutively transcribed and partially spliced. *Plant Molecular Biology* **53**: 399–410.

**Ohtsu K, Smith MB, Emrich SJ, Borsuk LA, Zhou R, Chen T, Zhang X, Timmermans MC, Beck J, Buckner B *et al.* 2007.** Global gene expression

analysis of the shoot apical meristem of maize (*Zea mays* L.). *The Plant Journal* **52**: 391–404.

**Park SJ, Eshed Y, Lippman ZB. 2014.** Meristem maturation and inflorescence architecture – lessons from the Solanaceae. *Current Opinion in Plant Biology* **17**: 70–77.

**Park SJ, Jiang K, Schatz MC, Lippman ZB. 2012.** Rate of meristem maturation determines inflorescence architecture in tomato. *Proceedings of the National Academy of Sciences, USA* **109**: 639–644.

**Pereira V. 2004.** Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. *Genome Biology* **5**: R79.

**Petrov DA, Fiston-Lavier AS, Lipatov M, Lenkov K, Gonzalez J. 2011.** Population genomics of transposable elements in *Drosophila melanogaster*. *Molecular Biology and Evolution* **28**: 1633–1644.

**Pietzenuk B, Markus C, Gaubert H, Bagwan N, Merotto A, Bucher E, Pecinka A. 2016.** Recurrent evolution of heat-responsiveness in Brassicaceae COPIA elements. *Genome Biology* **17**: 209.

**Quinlan AR, Hall IM. 2010.** BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

**Ramirez F, Ryan DP, Gruning B. 2016.** deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**(W1): W160–W165.

**Robinson MD, McCarthy DJ, Smyth GK. 2010.** edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

**Sabot F. 2014.** Tos17 rice element: incomplete but effective. *Mobile DNA* **5**: 10.

**Sabot F, Schulman AH. 2006.** Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. *Heredity* **97**: 381–388.

**Sanchez DH, Gaubert H, Drost HG, Zabet NR, Paszkowski J. 2017.** High-frequency recombination between members of an LTR retrotransposon family during transposition bursts. *Nature Communications* **8**: 1283.

**Sanchez DH, Paszkowski J. 2014.** Heat-induced release of epigenetic silencing reveals the concealed role of an imprinted plant gene. *PLoS Genetics* **10**: e1004806.

**Sarkinen T, Bohs L, Olmstead RG, Knapp S. 2013.** A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evolutionary Biology* **13**: 214.

**Shen JC, Rideout WM 3rd, Jones PA. 1994.** The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Research* **22**: 972–976.

**Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, Martienssen RA. 2009.** Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**: 461–472.

**Steinbauerova V, Neumann P, Novak P, Macas J. 2011.** A widespread occurrence of extra open reading frames in plant Ty3/gypsy retrotransposons. *Genetica* **139**: 1543–1555.

**Tamaki S, Tsuji H, Matsumoto A, Fujita A, Shimatani Z, Terada R, Sakamoto T, Kurata T, Shimamoto K. 2015.** FT-like proteins induce transposon silencing in the shoot apex during floral induction in rice. *Proceedings of the National Academy of Sciences, USA* **112**: E901–E910.

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

**Thompson JD, Higgins DG, Gibson TJ. 1994.** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673–4680.

**Tomato Genome Sequencing C, Aflitos S, Schijlen E, de Jong H, de Ridder D, Smit S, Finkers R, Wang J, Zhang G, Li N *et al.* 2014.** Exploring genetic variation in the tomato (*Solanum* section Lycopersicon) clade by whole-genome sequencing. *The Plant Journal* **80**: 136–148.

**Tucker JM, Larango ME, Wachsmuth LP, Kannan N, Garfinkel DJ. 2015.** The Ty1 retrotransposon restriction factor p22 targets Gag. *PLoS Genetics* **11**: e1005571.

**Vernhettes S, Grandbastien MA, Casacuberta JM. 1997.** *In vivo* characterization of transcriptional regulatory sequences involved in the defence-associated expression of the tobacco retrotransposon Tnt1. *Plant Molecular Biology* **35**: 673–679.

**Wang Q, Hasson A, Rossmann S, Theres K. 2016.** *Divide et impera*: boundaries shape the plant body and initiate new meristems. *New Phytologist* **209**: 485–498.

**Weil C, Martienssen R. 2008.** Epigenetic interactions between transposons and genes: lessons from plants. *Current Opinion in Genetics & Development* **18**: 188–192.

**Wicker T, Keller B. 2007.** Genome-wide comparative analysis of copia retrotransposons in Triticeae, rice, and Arabidopsis reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. *Genome Research* **17**: 1072–1081.

**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 Genetics* **8**: 973–982.

**Williams BP, Pignatta D, Henikoff S, Gehring M. 2015.** Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. *PLoS Genetics* **11**: e1005142.

**Xu Y, Du J. 2014.** Young but not relatively old retrotransposons are preferentially located in gene-rich euchromatic regions in tomato (*Solanum lycopersicum*) plants. *The Plant Journal* **80**: 582–591.

**Xu Z, Wang H. 2007.** LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research* **35**(Web Server issue): W265–W268.

**Yang W, Schuster C, Beahan CT, Charoensawan V, Peaucelle A, Bacic A, Doblin MS, Wightman R, Meyerowitz EM. 2016.** Regulation of meristem morphogenesis by cell wall synthases in Arabidopsis. *Current Biology* **26**: 1404–1415.

**Zervudacki J, Yu A, Amesefe D, Wang J, Drouaud J, Navarro L, Deleris A. 2018.** Transcriptional control and exploitation of an immune-responsive family of plant retrotransposons. *EMBO Journal* **37**: pii: e98482.

**Zhong S, Fei Z, Chen YR, Zheng Y, Huang M, Vrebalov J, McQuinn R, Gapper N, Liu B, Xiang J *et al.* 2013.** Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nature Biotechnology* **31**: 154–159.

## Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

**Fig. S1** Validation of transcriptome data in independent experiments.

**Fig. S2** Analyses of alignments and tandem-repeats.

**Fig. S3** Tomato leaf, flower and shoot apical meristem RNA-seq profiles for very young differentially expressed full-length *MESSI*.

**Fig. S4** Tomato shoot apical meristem smRNA and DNA methylation profiles for very young differentially expressed full-length *MESSI*.

**Fig. S5** Tissue-specific PCR-free DNA-seq.

**Fig. S6** Transcript levels in tomato shoot apical meristem samples under long-term heat.

**Fig. S7** Complete set of tomato shoot apical meristem sections for *in situ* hybridization signals of *MESSI* LTR retrotransposons.

**Fig. S8** *MESSI* LTR retrotransposons occur across *Solanum*.

**Fig. S9** Phylogenetic analyses of *MESSI* and *MESSI-like* elements across *Solanum* with estimated insertion within the last *c.* 1 million years.

**Fig. S10** *Solanum pennellii* leaf, flower and shoot apical meristem RNA-seq profiles for differentially expressed *MESSI*.

**Table S1** List of the examined tomato LTR retrotransposons.

**Table S2** List of the different NGS dataset produced or used in this study.

**Table S3** List of primers used.

**Table S4** Transcriptome data of LTR retrotransposons deemed expressed in our RNA-seq experiments.

**Table S5** *MESSI* and *MESSI*-like LTR retrotransposons annotated in tomato.

**Table S6** Transcriptome data of tomato *MESSI* differentially expressed in meristems.

**Table S7** *MESSI* and *MESSI*-like LTR retrotransposons annotated in *Solanum pennellii*.

**Table S8** Transcriptome data of *Solanum pennelli MESSI* differentially expressed in meristems.

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