

TECHNICAL ADVANCE

Large-scale insertional mutagenesis using the *Tnt1* retrotransposon in the model legume *Medicago truncatula*

Million Tadege¹, Jiangqi Wen¹, Ji He¹, Haidi Tu¹, Younsig Kwak^{1,†}, Alexis Eschstruth², Anne Cayrel², Gabriella Endre³, Patrick X. Zhao¹, Mireille Chabaud⁴, Pascal Ratet² and Kirankumar S. Mysore^{1,*}

¹Plant Biology Division, The Samuel Roberts Noble Foundation, 2510 Sam Noble Parkway, Ardmore, OK 73401, USA,

²Institut des Sciences du Végétal, CNRS, Avenue de la Terrasse, 91198 Gif sur Yvette, France,

³Institute of Genetics, Biological Research Center of the Hungarian Academy of Sciences, Temesvári krt. 62, H-6726 Szeged, Hungary, and

⁴Laboratoire de Biologie Moléculaire des Relations Plantes-Microorganismes, INRA-CNRS, BP27, F-31326 Castanet-Tolosan Cedex, France

Received 18 December 2007; accepted 21 December 2007.

*For correspondence (fax +1 580 224 6692; e-mail ksmysore@noble.org).

†Present address: Department of Plant Pathology, 367A Johnson Hall, PO Box 646430, Washington State University, Pullman, WA 99164 6430, USA.

Summary

Medicago truncatula is a fast-emerging model for the study of legume functional biology. We used the tobacco retrotransposon *Tnt1* to tag the *Medicago* genome and generated over 7600 independent lines representing an estimated 190 000 insertion events. *Tnt1* inserted on average at 25 different locations per genome during tissue culture, and insertions were stable during subsequent generations in soil. Analysis of 2461 *Tnt1* flanking sequence tags (FSTs) revealed that *Tnt1* appears to prefer gene-rich regions. The proportion of *Tnt1* insertion in coding sequences was 34.1%, compared to the expected 15.9% if random insertions were to occur. However, *Tnt1* showed neither unique target site specificity nor strong insertion hot spots, although some genes were more frequently tagged than others. Forward-genetic screening of 3237 R₁ lines resulted in identification of visible mutant phenotypes in approximately 30% of the regenerated lines. Tagging efficiency appears to be high, as all of the 20 mutants examined so far were found to be tagged. Taking the properties of *Tnt1* into account and assuming 1.7 kb for the average *M. truncatula* gene size, we estimate that approximately 14 000–16 000 lines would be sufficient for 90% gene tagging coverage in *M. truncatula*. This is in contrast to more than 500 000 lines required to achieve the same saturation level using T-DNA tagging. Our data demonstrate that *Tnt1* is an efficient insertional mutagen in *M. truncatula*, and could be a primary choice for other plant species with large genomes.

Keywords: *Tnt1* tagging, *Medicago*, legume, retrotransposon, insertion mutagenesis, functional genomics.

Introduction

Legumes constitute one of the largest crop families and are second only to grasses in economic importance to mankind (Graham and Vance, 2003). A key contribution of legumes to sustainable agriculture and the nitrogen cycle is their ability to fix atmospheric nitrogen, which saves an estimated US\$10 billion annually in commercial fertilizer (Graham and Vance, 2003). Over the past decade, *Medicago truncatula* has been selected as one of two model legume species because of its close relationship to many crop and forage legumes

and its amenability to genetic and genomic studies (Cook, 1999; Frugoli and Harris, 2001; May and Dixon, 2004; Tadege *et al.*, 2005; Udvardi *et al.*, 2005; Young and Shoemaker, 2006). The *M. truncatula* consortium has made significant progress in expressed sequence tag (EST) and genome sequencing, with most of the gene space anticipated to be sequenced by 2008. Various genomic tools such as gene expression, protein and metabolite profiling, as well as mutagenesis tools such as RNAi, fast neutron bombard-

ment, ethyl methane sulfonate (EMS) and target-induced local lesions in genomes (TILLING), have been developed to aid identification of gene function (Lei *et al.*, 2005; Limpens *et al.*, 2004; Modolo *et al.*, 2007; Penmesta and Cook, 2000; Wang *et al.*, 2006), but a significant number of tagged *M. truncatula* lines have not been generated. Although T-DNA tagging has been attempted at a small scale (Scholte *et al.*, 2002), a high level of somaclonal mutations was seen without gene disruption by the T-DNA (Brocard *et al.*, 2006). *Medicago truncatula* functional genomic studies will greatly benefit from an indexed set of insertion mutants analogous to the T-DNA-tagged lines in Arabidopsis. The relative importance of insertional mutagenesis for *M. truncatula* functional genomics and its complementarity to other genomic tools has been discussed previously (Tadege *et al.*, 2005).

Insertional mutagenesis is a powerful tool to mutagenize and tag genomes, facilitating subsequent cloning of mutated genes. *Agrobacterium* T-DNA and maize transposable elements have been extensively used as insertional mutagens in plants (An *et al.*, 2005; Bortiri *et al.*, 2006; Fladung *et al.*, 2004; Kumar and Hirochika, 2001; Tadege and Mysore, 2006; Walbot, 2000). In *Arabidopsis thaliana*, T-DNA tagging has proved to be one of the most powerful tools to determine gene function. Due to the small genome size and the ability to perform *in planta* transformation, near-saturation mutagenesis of the Arabidopsis genome by T-DNA tagging has been achieved (Alonso *et al.*, 2003).

T-DNA tagging is not feasible in *M. truncatula* because of the large genome size (approximately 500 Mb; four times that of Arabidopsis) and the absence of an efficient *in planta* transformation system (Somers *et al.*, 2003). Using the equation derived by Krysan *et al.* (1999) and considering 1.7 kb to be the average gene size (based on the International *Medicago* Genome Annotation Group, and current genome annotation data), it is estimated that over 1.3 million insertions or at least 903 000 independent transformation events would be required to achieve 99% saturation mutagenesis in *M. truncatula* using T-DNA as a mutagen. Transposable elements represent an attractive alternative for generating a large number of insertions in a relatively fewer number of mutant lines. The class II type of maize DNA transposons such as *Ac/Ds* and *En/Spm* move into new locations in a genome by a 'cut and paste' mechanism causing unstable mutations (Wessler, 2006). The class I type of transposable elements, known as retrotransposons, on the other hand, move into new locations in a genome by a 'copy and paste' mechanism (Kumar and Bennetzen, 1999). Retrotransposons copy themselves via an mRNA intermediate and insert into new chromosomal locations after reverse transcription (Wessler, 2006). As there is no excision during replicative transposition, mutations generated by retrotransposon insertions are stable.

We have chosen the transposable element of *Nicotiana tabacum* cell type 1 (*Tnt1*) to initiate saturation mutagenesis in the model legume *M. truncatula*. *Tnt1* is one of the very few well-characterized plant long terminal repeat (LTR) retrotransposons (Grandbastien *et al.*, 1989). It is very active in *M. truncatula* and transposes in multiple copies during tissue culture but is stable during seed-to-seed propagation in soil (d'Erfurth *et al.*, 2003). We chose an *M. truncatula* transgenic line Tnk88-7-7 containing five *Tnt1* insertions as the starting material, and activated *Tnt1* transposition via somatic embryogenesis from leaf explants. Over the last four years, we have regenerated 7600 independent lines that contain an estimated 190 000 *Tnt1* insertions. We have sequenced and analyzed more than 2460 non-redundant *Tnt1* flanking sequence tags (FSTs) and show that *Tnt1* frequently targets genes with a preference for exons.

Results

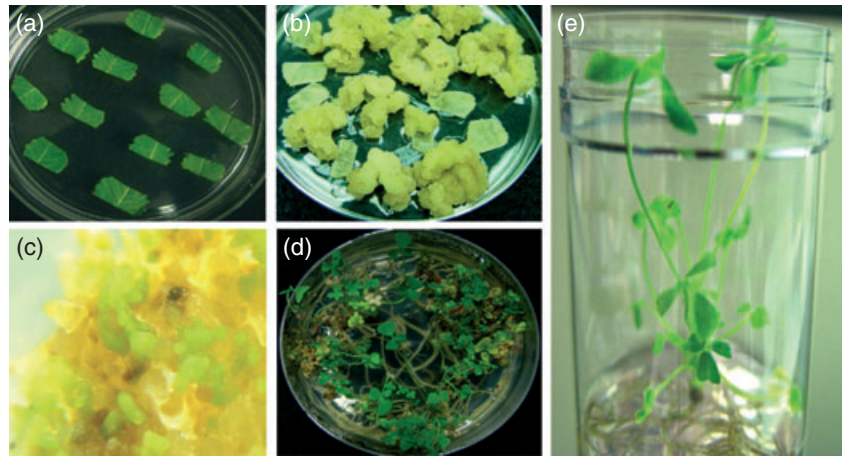
Large-scale generation of insertion mutants in M. truncatula

The complete *Tnt1* retro-element was cloned into a binary vector and transformed into *M. truncatula* via *Agrobacterium*-mediated transformation (d'Erfurth *et al.*, 2003). A low-copy *Tnt1*-containing single transgenic line (with five *Tnt1* copies) was used as the starting material to generate mutants. Although *Tnt1* is reported to be transcriptionally induced by various stresses and chemical treatments (Mhiri *et al.*, 1997; Vernhettes *et al.*, 1997), we were not able to detect transposition by vernalization at 4°C for 4–6 weeks, or by treatment with chemicals and phytohormones such as 5-azacytidine, methyl jasmonate, salicylic acid, 2,4-dichlorophenoxyacetic acid (2,4-D) or 6-benzylaminopurine (BAP), either individually or in various combinations (data not shown). Regeneration of new plants via the cotyledon transformation method (Zhou *et al.*, 2004) with or without *Agrobacterium* infection also did not induce transposition (data not shown). Complete callus formation in the presence of auxin, on the other hand, was necessary and sufficient for effective *Tnt1* transposition in *M. truncatula*.

Leaf explants were cultured on Schenk and Hildebrandt medium (SHM; Schenk and Hildebrandt, 1972) supplemented with auxin and cytokinin (see Appendix S1) to form calli (4–6 weeks). Calli were subsequently transferred to embryogenesis medium and then to hormone-free medium to ultimately regenerate new plants (Figure 1). Transposition probably occurs very early during the first 2–3 weeks of callus formation, but the precise times of commencement and termination of transposition are not known. To avoid possible redundancy, we selected only one plant from each callus derived from a single leaf explant (Figure 1). The total number of new inserts per line showed wide variation and ranged between 6 and 59, with an average of 25, estimated

Figure 1. Generation of mutants via tissue culture.

(a) Leaf explants from 4–7-week-old *Tnt1* transgenic plants were cultured on SHM medium supplemented with auxin and cytokinin.
 (b) Four-week-old calli developing on leaf explants.
 (c) A single callus developing somatic embryos on SHM medium supplemented with cytokinin only.
 (d) Plants regenerating from a single embryogenic callus on hormone-free medium.
 (e) A single plant, representing an independent line, selected from (d) for transfer to soil. Only one plant was regenerated from each leaf segment. Transposition occurs between the stages shown in (a) and (b).



based on a combination of Southern blot analysis (Figure S1) and sequencing of FSTs. The reason for this variability is not understood. Shorter (2 weeks) or longer (8–10 weeks) subculture in the callus induction medium reduced the efficiency of regeneration but had no consistently significant effect on the total number of transpositions. Over the last four years, approximately 7600 independent lines have been regenerated with an estimated 190 000 insertions, providing approximately 70% coverage of the euchromatin gene space.

Tnt1 preferentially targets exons

To evaluate the distribution of *Tnt1* in the *M. truncatula* genome and determine its site specificity, we sequenced 2461 *Tnt1* FSTs and analyzed their distribution. The results presented in Table 1 show that at least 78.6% of the FSTs matched *M. truncatula* coding sequence (CDS) in International *Medicago* Genome Annotation Group (IMGAG), *Medicago truncatula* Gene Index (MTGI) ESTs, bacterial artificial chromosomes (BACs) or genes from other species in the National Center for Biotechnology Information (NCBI) non-redundant (NR) database. NCBI matches were considered with a cut-off *e*-value $\leq 1.00\text{E-}5$, whereas all other

Medicago matches were considered with a cut-off *e*-value $\leq 1.00\text{E-}10$. The remaining 21.4% of the FSTs had no significant matches in any of the databases, suggesting that they could be part of unsequenced repetitive regions of the genome or unique *M. truncatula* genes that have not yet been sequenced. As the *M. truncatula* genome sequencing effort is preferentially targeting BACs from gene-rich euchromatic arms, it may be conservatively concluded that a minimum of 78.6% of the inserts are within or near the gene-rich regions. Of the 1935 FSTs that constituted the 78.6% match, 1164 (60.2%) had high homology (*e*-values $< 1.00\text{E-}25$) to a very wide range of known genes, including transcription factors, protein kinases, cell-cycle regulators, transporters, photosynthetic and metabolic enzymes (see Table S1).

To determine the proportion of inserts that hit genes or intergenic regions, we removed the EST and NCBI protein matches and determined the exact insertion site for a subset of FSTs (shown in Table 1) that we were able to map on fully annotated *M. truncatula* BACs. Analysis of 964 such FSTs revealed that 329 (34.1%) of the insertions were in coding exons, 224 (23.2%) were in introns plus UTRs, and 411 (42.6%) were in intergenic regions (Table 2). If random insertion of *Tnt1* in the euchromatin gene space was assumed (i.e. normally distributed), based on the size of the BACs and current annotation, 153, 210 and 601 insertions (corresponding to 15.9%, 21.8% and 62.3%) for exons, introns plus UTRs, and intergenic regions, respectively,

Table 1 BLAST match distribution of *Tnt1* FSTs

BLAST hit ^a	Number of FSTs	Percentage
<i>Medicago</i> pseudogenome/BACs	1666	67.7
ESTs in MTGI (not included above)	161	6.5
Any protein in NCBI NR (not included above)	108	4.4
Subtotal	1935	78.6
No hit	526	21.4
Total FSTs	2461	100

^aAn NCBI NR hit is based on a BLASTX search with a 1e-5 threshold. All other hits are based on BLASTN searches with a 1e-10 threshold.

Table 2 *Tnt1* preferentially targets exons

Insertion site	Number of FSTs	Percentage of FSTs	Length of BAC	
			sequence analyzed (bp)	Percentage coverage
IMGAG CDS	329	34.1	29 524 300	15.9
Introns + UTRs	224	23.2	40 593 722	21.8
Intergenic	411	42.6	116 103 826	62.3
Total	964	100	186 221 848	100

would have been expected (Table 2). Based on a χ^2 goodness-of-fit test (Snedecor and Cochran, 1989), the overall observed insertion numbers are significantly different from expectations ($P < 0.001$). Based on a z-test of individual insertion frequencies (Madden *et al.*, 1997), the differences between the actual *Tnt1* insertions in CDS (329) and intergenic regions (411) and the expected frequencies (153 and 601, respectively) in the analyzed genomic regions are significant ($P < 0.001$). Interestingly, the number of insertions in introns plus UTRs (224) was comparable ($P = 0.64$) to the expected value (210) for unbiased insertion (Table 2). This suggests that the preference of *Tnt1* for gene-rich regions is due to preference for coding exons. The IMGAG annotation is considered to be 95% accurate for exons (Town, 2006), but, irrespective of this accuracy, the exon preference of *Tnt1* appears to be quite clear.

Tnt1 has no strong target site specificity

To determine whether *Tnt1* has insertion hot spots, we mapped the FSTs onto *M. truncatula* pseudochromosomes. As the pseudochromosomes are linear distributions of ordered BACs, this configuration roughly estimates the distribution of inserts on BACs across the chromosomes. We

found that FSTs were fairly uniformly distributed across all the pseudochromosomes (Figure 2). Closer inspection of a portion of the pseudochromosomes that had accumulated higher than average FSTs revealed that some genes or BACs are more frequently hit than others (see inset in Figure 2). However, overall, there are no specific chromosomal regions or BACs that are heavily targeted or avoided altogether in the sequenced gene space.

To assess target site specificity, we looked at the occurrence of bases in 40 bp sequences upstream and downstream of the target site duplication (TSD). *Tnt1* creates a 5 bp duplication at the insertion site. Analysis of FSTs with exact insertion sites on *M. truncatula* BACs showed that there was no strongly preferred sequence on either side of the TSD, and the 5 bp TSD itself was not highly conserved (Figure 3). To evaluate the base preferences of *Tnt1* around the insertion site, we calculated the base preferences of the *M. truncatula* pseudogenome (assembled BACs), and compared this with the IMGAG CDS, introns plus UTRs, and intergenic regions, and the FSTs (Table 3). The total base compositions for G, C, A and T were found to be 17.1, 17.1, 32.4 and 33.4%, respectively, for the FSTs, 16.6, 16.6, 33.4 and 33.4%, respectively, for the assembled BACs, and 22.4, 18.6, 29.9

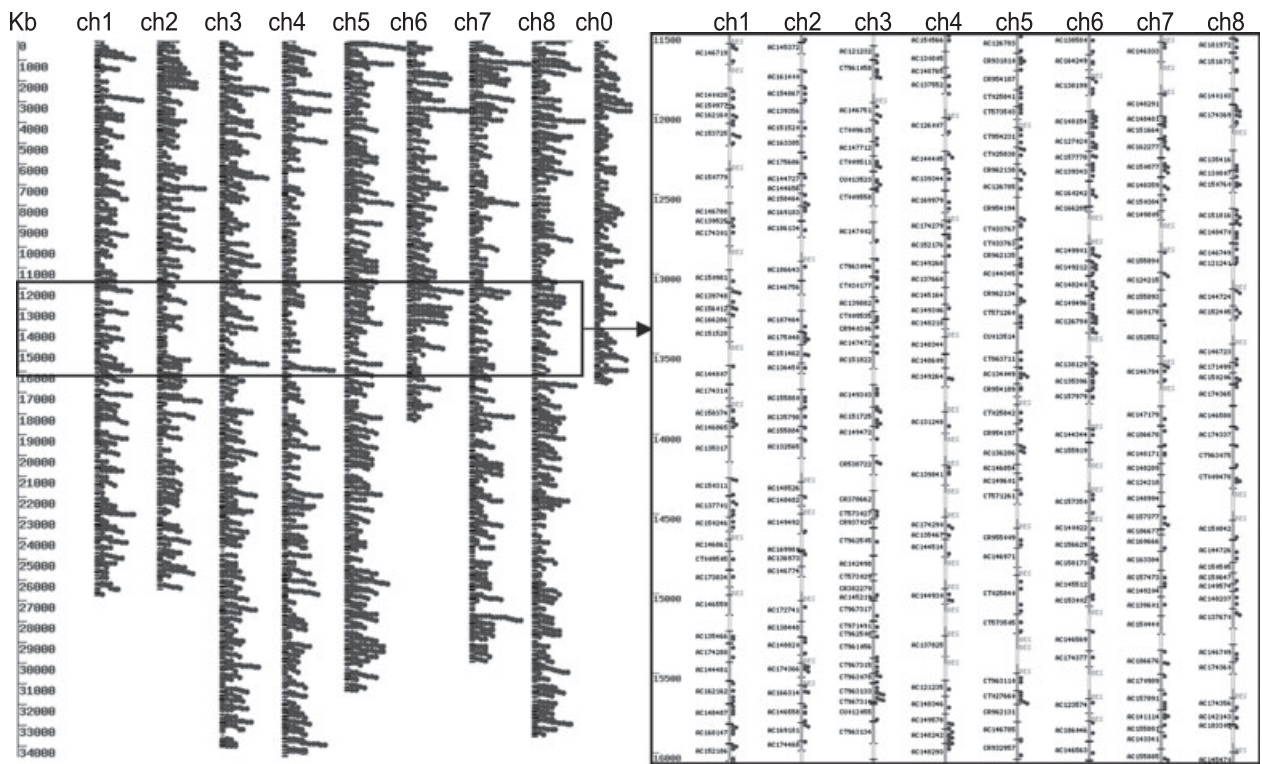


Figure 2. Distribution of FSTs on pseudochromosomes. All FSTs with a BLAST e-value cut-off $\leq 1.00E-50$ are shown. A portion of the FST distribution (inset) on pseudochromosomes with high-density FSTs is shown for a clearer view. The BACs or chromosomal regions are not targeted with absolutely equal frequency, but there are no regions that are heavily targeted or completely avoided in the sequenced euchromatic region.

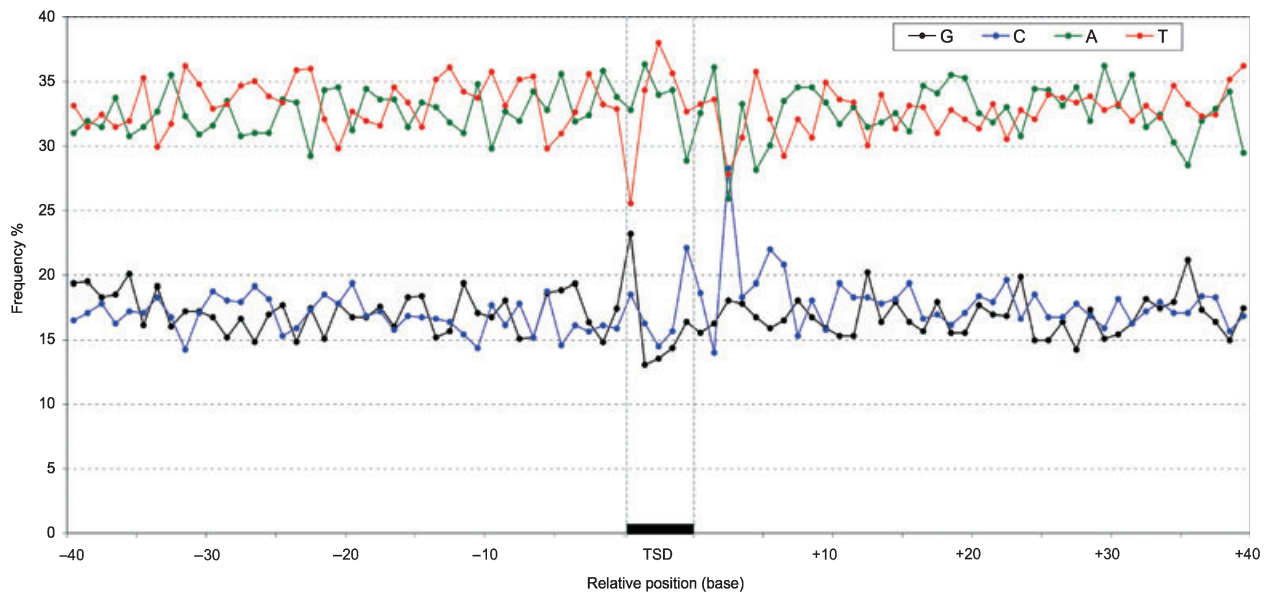


Figure 3. *Tnt1* has no unique target site specificity.

Forty base pairs on either side of the 5 bp target site duplication (TSD) and within the TSD itself showed no strong preference for a particular set of sequences, but weak statistically significant preferences for GC and against AT were detected within the TSD and 7 bp immediately downstream of the TSD. The strongest of these was at position TSD+3, where the preference for C (28.3%) was greater by 11.7% than the expected 16.6% in the pseudogenome. The corresponding values are 18.6% in IMGAG CDS, 15.4% in introns plus UTRs, and 16.5% in the intergenic region. The other observed preferences were weaker but follow a similar trend in being closer to the IMGAG CDS than to the rest of the genome. Under this scenario, coding exons have a better chance of being targeted by *Tnt1* than untranslated regions, introns or intergenic regions.

Table 3 Base composition (%) of *M. truncatula* genomic sequences and FSTs

Sequence	G	C	A	T	Total bp ^a
Pseudogenome	16.6	16.6	33.4	33.4	185 926 858
IMGAG CDS	22.4	18.6	29.9	29.1	29 503 461
Introns + UTRs	16.4	15.4	30.3	37.8	40 550 322
Intergenic	15.2	16.5	35.4	32.9	115 873 075
FSTs	17.1	17.1	32.4	33.4	1 073 570
85 bp around TSD	16.9	17.5	32.6	32.9	71 570
TSD1	23.2**	18.5	32.8	25.5**	71 570
TSD2	13.1*	16.3	36.3	34.3	71 570
TSD3	13.5	14.5	34.0	38.0*	71 570
TSD4	14.4	15.7	34.3	35.6	71 570
TSD5	16.4	22.1*	28.9*	32.7	71 570
TSD+1	15.6	18.6	32.5	33.3	71 570
TSD+2	16.3	14.0	36.1	33.6	71 570
TSD+3	18.1	28.3**	25.9**	27.8*	71 570
TSD+4	17.8	18.3	33.3	30.6	71 570
TSD+5	16.7	19.4	28.1*	35.7	71 570
TSD+6	15.9	22.0*	30.0	32.1	71 570
TSD+7	16.5	20.5*	35.5	29.2	71 570

^aThe total bp reported here is less than the corresponding length of BAC sequence analyzed in Table 2 because of a small proportion of unknown bases (Ns) in the sequences.

* $P < 0.05$; ** $P < 0.001$, significance of difference compared to the pseudogenome base composition.

and 29.1% for IMGAG CDS. The total base compositions in the introns plus UTRs and intergenic regions for G, C, A and T were found to be 16.4, 15.4, 30.3 and 37.8%, and

15.2, 16.5, 35.4 and 32.9%, respectively (Table 3), suggesting that coding exons are more GC-rich than other sequences in the *Medicago* genome. For individual positions, some weak but statistically significant preferences for GC sequences were detected within and immediately downstream of the TSD (TSD1 to TSD+7) as shown in Figure 3 and Table 3. For example, within the TSD, at position TSD1, the value for G was 23.2% versus the expected 16.6%, and that for T was 25.5% versus the expected 33.4%, and downstream of the TSD, at position TSD+3, the value for C was 28.3% versus the expected 16.6%, and that for A was 25.9% versus the expected 33.4%. Only in two cases was this trend reversed; at TSD+2, G (13.1%) was selected against, and at TSD+3, T (38.0%) was preferred. However, in general, these differences were not particularly strong ($P < 0.05$ for most of them, Table 3), and in most cases the observed changes were rather closer to the total base composition of the IMGAG CDS than the total assembled BACs (Table 3). This modest preference of *Tnt1* for GC versus AT at eight positions within and immediately downstream of the TSD is consistent with the exon preference of *Tnt1*, but fails to constitute a unique target site as might be expected for many transposons and retrotransposons. From these observations, we conclude that *Tnt1* has neither strong insertion hot spots nor unique target site specificity, but targets most of the gene-rich euchromatin with preference for transcribed regions.

Forward-genetic screens identified 30% visible mutant phenotypes in the Tnt1 population

To evaluate the effectiveness of *Tnt1*-mediated mutagenesis in *M. truncatula* and catalogue altered visible phenotypes, we organized yearly phenotypic screening events with interested colleagues from Europe and the USA. To date, 3237 lines (R_1 progeny of regenerated R_0 lines) have been screened for visible phenotypes after inoculation with *Sinorhizobium meliloti* and arbuscular mycorrhizal fungi under low-nitrogen and low-phosphorus conditions. Over 30% of the lines displayed visually recognizable phenotypes (Table 4). Germination, fertility, trichome development and other mutant phenotypes that were not visible to the naked eye are not included in Table 4. When two or more phenotypes were observed in the same line, phenotypes were recorded as separate only if they clearly segregated in the population. When the segregation pattern was not obvious, only the major phenotype was taken into account even if more than two phenotypes could be scored. Thus, the approximately 30% visible phenotype frequency is a conservative estimate and represents only the phenotypes that we analyzed. It should be noted that this is a preliminary screen under one set of conditions, which are obviously not the optimal conditions for some of the phenotypes such as flowering time and plant size as the poor nutrition could normally affect these processes. In addition, most of the symbiotic mutants have not yet been reconfirmed by a secondary screen or additional testing, and it is not unusual to find some false positives.

Figure 4 shows some of the representative mutant phenotypes summarized in Table 4. Some of the non-nodulating mutants had short and compact roots (Figure 4a), some had reduced red pigment on stems and leaves (Figure 4h), and some flowered late (Figure 4e). Many non-nodulating mutants were also not colonized by arbuscular mycorrhizal fungi (Figure 4k). Another group of mutants worth noting is the non-flowering mutants (Figure 4f). The mutants in this category were bushy, had lost apical dominance, and did not form internodes. They grew slowly at the seedling stage and did not flower after two years of long-day conditions even after vernalization. These results show that we were able to obtain a wide variety of mutants ranging from symbiotic to developmental. Interestingly, several identical phenotypes were identified from the developmentally altered mutants. For example, *unifoliolate* (*uni*), *stemless*, *stamina pistilloida* (*stp*) and *narrow lamina* phenotypes were identified in more than one independent line.

One concern with forward-genetic screening of the *Tnt1* population is that the phenotype-to-tag ratio is not one-to-one, as there are multiple copies of the *Tnt1* tag in each line. It is necessary to perform a few back-crosses, depending on the number of inserts, to obtain a mutant with a single insert. The *Tnt1* lines can be back-crossed to R108

Table 4 Summary of visible phenotypic classes observed in the preliminary screening of 3237 *Tnt1* lines in R_1 progeny

Major phenotypes ^a	Number of lines	Frequency (%) ^b
Symbiotic		
Nodulation minus (no nodules)	128	3.95
Super-nodulator (more nodules than usual)	18	0.56
Fixation minus (white nodules)	135	4.17
Reduced mycorrhiza association (no or reduced external hyphae)	105	3.23
Aberrant nodules (club-shaped, brown/black, big, etc)	35	1.08
Morphology		
Extreme dwarf with dark green leaves	12	0.37
Extreme dwarf, bright green, deformed shoots	5	0.15
Extreme dwarf	14	0.43
Moderate dwarf	168	5.19
Stunted growth	11	0.34
Unusual branching and delayed flowering	6	0.18
Epinastic leaves on a normal plant	4	0.12
Variegated leaves	5	0.15
Chlorotic leaves	20	0.62
Spotted leaves (lesion-mimic type)	13	0.4
Short and compact roots	7	0.21
Deformed roots (growth-arrested)	3	0.09
Long and thin roots, reduced/normal lateral branches	12	0.37
Development and flowering		
Narrow lamina	6	0.18
Unifoliolate and cauliflower-like inflorescence	5	0.15
Mixed foliate (3, 4 and 5 leaflets) with more flowers	4	0.12
No shoot apical meristem (only cotyledons)	7	0.22
Late-flowering	87	2.69
Bushy, late-flowering	4	0.12
Bushy, non-flowering (stemless)	7	0.22
Early flowering	50	1.54
Inflorescence and floral organ defect	25	0.77
Pigmentation		
Purple leaves (more anthocyanin)	16	0.49
Bluish light green (not pale) plants (less anthocyanin)	21	0.65
Pale green leaves (light or yellow green)	41	1.27
Dark green leaves	8	0.25
Albino	12	0.37
Total visible phenotypes	994	30.70

^aAll of the phenotypes were visually scored, with use of a stereo dissecting microscope when necessary.

^bNumbers are before confirmation by secondary screening or additional testing.

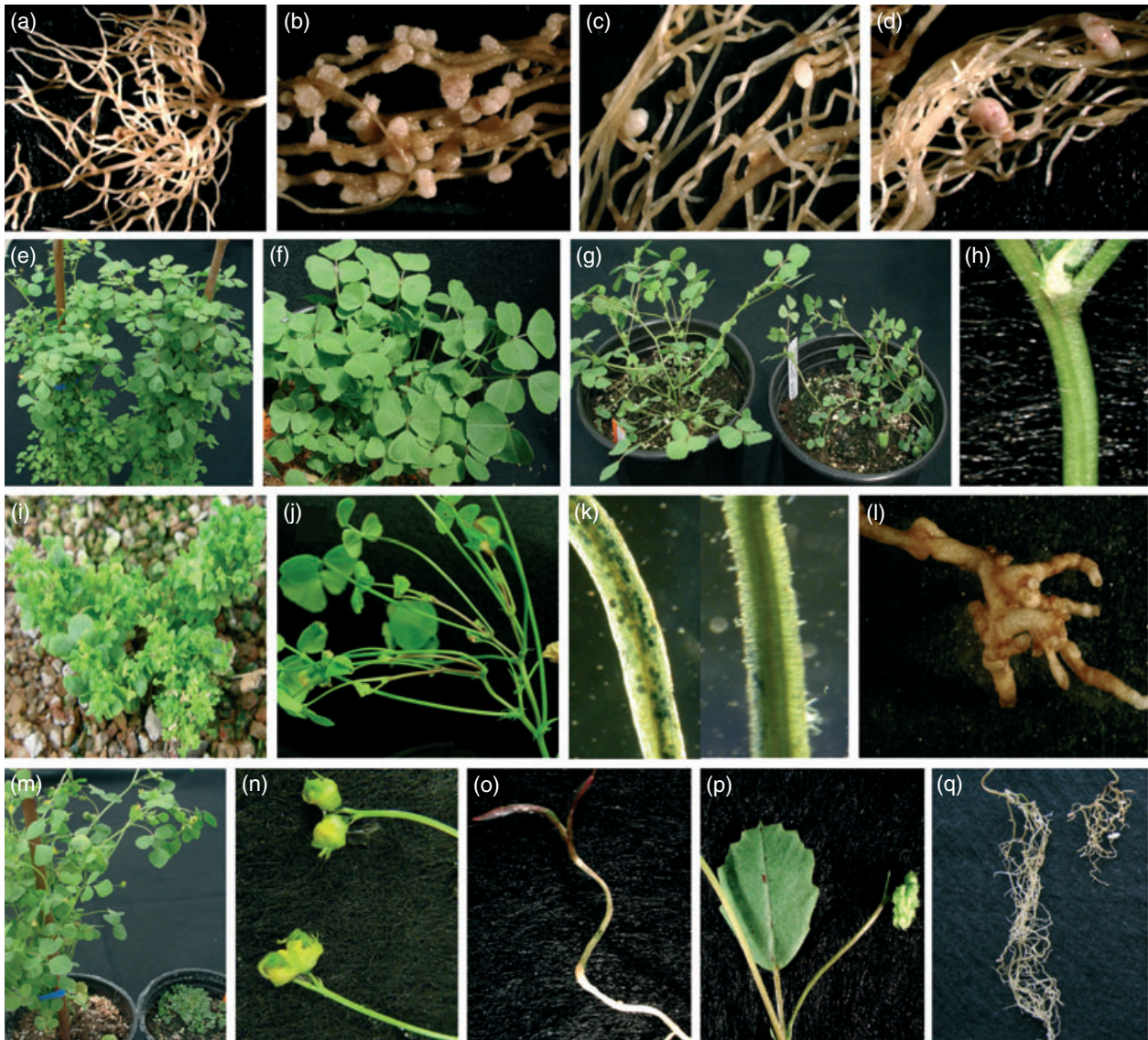


Figure 4. Visible phenotypes of representative *Tnt1* mutants.

(a) Non-nodulating compact root; (b) super-nodulator; (c) white nodules (fixation minus); (d) wild-type R108 control for nodulation and fixation; (e) late-flowering (the plant on the right flowered at 120 days compared to 35 days for the R108 control shown on the left); (f) non-flowering, bushy and stemless; (g) early-flowering (the plant on the right flowered 2 weeks earlier than the control shown on the left); (h) pigmentation mutant (accumulates less red pigment in the stem and leaves than the control); (i) dwarf and deformed shoots (bright green, tiny cotyledon-like bushes but no true leaves or branches); (j) unusual branching, bushy and late-flowering; (k) mycorrhizal mutant (the mutant root on the right lacks arbuscules and vesicles compared to the control root on the left); (l) root meristem mutant; (m) extreme dwarf (control plant of the same age on the left); (n) reproductive structures mutant (petals aborted half way and do not open, stamen not developed); (o) no shoot apical meristem; (p) single leaflet with cauliflower head-like inflorescence; (q) dwarf root (right) with wild-type control (left).

wild-type or to known A17 mutant lines for allelism testing (P.R., unpublished data). However, as most of the phenotypes segregate in the R_1 progeny, it may not be necessary to perform any back-crossing just for the sake of identifying the disrupted gene. By recovering and sequencing all the FSTs in a particular mutant R_1 plant and PCR genotyping each FST in the segregating R_1 population, we were able to identify a single mutation among a population of 41 *Tnt1* FSTs (Figure S2; M.T. and K.S.M., unpublished data). In a sufficiently large segregating R_1 population, this simple PCR

technique leads to identification of a single FST. In some cases, there may be two or rarely three FSTs co-segregating with the phenotype. These additional candidates can be readily eliminated if there is another allele for the mutation in the population. In cases where alleles are not found, the candidate FSTs may be evaluated by complementation of the mutant phenotype. Complementation may be performed in one of two ways; somatic embryogenesis using leaf explants or organogenesis using cotyledons. Complementation via somatic embryogenesis has the disadvantage of

inducing new transpositions. However, as there is no excision during transposition, the mutant phenotype will not disappear unless it is complemented. We have complemented two leaf mutants by this method with unambiguous results (M.T. and K.S.M., unpublished data; Rujin Chen, personal communication). Complementation via organogenesis using cotyledons (Zhou *et al.*, 2004) is the preferred method for transformation as it is faster and does not induce transposition. All of the mutants that have been selected for in-depth characterization by our group or our collaborators have so far been found to be tagged. We have confirmed tagging of 20 independent mutant lines that fall into nine phenotypic classes. These tagged genes include some that have already been described in legumes, such as *UNI* (Hofer *et al.*, 1997), *NIN* (Marsh *et al.*, 2007), *SUNN* (Elise *et al.*, 2005) and *AP1* (Benlloch *et al.*, 2006), and some phenotypes that have not yet been described, including leaf shape, nitrogen fixation and pigment production. Gene tagging in these lines was confirmed by either a combination of segregation analysis and complementation, or by PCR-based segregation analysis alone in cases where alleles were identified (Figure S2). This, however, does not necessarily mean that the tagging efficiency is 100% as analyses were not exhaustive. Nevertheless, the fact that we found no single untagged mutation in 20 independent mutants selected for analysis suggests that the efficiency of gene tagging for *Tnt1* is much higher than for other insertional mutagens described to date, including *Tos17* and T-DNA.

Development of publicly available reverse genetics resources

To increase the utility of the *Tnt1*-tagged *M. truncatula* lines to the legume research community, we have created a user-friendly *Tnt1* FST database (http://bioinfo.noble.org/mt_insertion/) that is freely accessible. This *Tnt1* FST database currently contains the non-redundant sequences described above and can be searched by using a gene name or a BLAST search. The sequence and BLAST information from each line are linked to a window where the major visible phenotype(s) of that line are described and displayed to give gene-to-phenotype associations for gene cloning *in silico*. Efforts are in progress to significantly expand the database by sequencing more FSTs in collaboration with our European partners (<http://www.eugrainlegumes.org>). In addition to the FST database, genomic DNA has been extracted from 6000 *Tnt1* lines for customized reverse-genetic screening of DNA pools using a combination of *Tnt1* and gene-specific primers. DNA from 5000 lines has been organized into 50 'super pools' of 100 lines each, and preliminary screening suggests that identification of *Tnt1* inserts in a desired gene is very efficient (data not shown). Efforts are underway to increase the size of these DNA pools and establish a cost-effective delivery system to the legume research community. Details of this

strategy and the screening service will be announced to the community in future when this service is fully operational. Currently, we welcome enquiries for screening of DNA pools on collaborative basis. Participation in the forward-genetic screening (non-destructive assay) of the *Tnt1* lines has been freely available to the research community once a year over a period of 2 weeks for the last 3 years, and will continue until all the lines are screened. Interested parties are encouraged to contact the corresponding author.

Discussion

Legumes account for approximately one third of the world's primary crop production, human dietary protein, and processed vegetable oil. We have generated a substantial number of publicly available insertion mutants in the model legume *M. truncatula*. To date, 7600 independent *Tnt1*-tagged *M. truncatula* lines have been generated, with an average of 25 inserts per line, and most of the insertions are in gene-rich regions of the genome. We have shown that *Tnt1* has no unique target site specificity (Figure 3), but prefers coding exons (Table 2). Interestingly, the only other plant retrotransposon used for large-scale mutagenesis, rice *Tos17*, also shows preference for exons (Miyao *et al.*, 2003), but, unlike *Tnt1*, *Tos17* has moderate target site specificity. Although this is the first time that *Tnt1* has been used for large-scale mutagenesis, previous studies in *Arabidopsis*, *Medicago* and, more recently, in its natural host tobacco have led to similar conclusions that *Tnt1* copies insert within or close to host gene coding sequences (Courtial *et al.*, 2001; d'Erfurth *et al.*, 2003; Le *et al.*, 2007; Lucas *et al.*, 1995). This is in sharp contrast to many other retro-elements which are found nested in intergenic regions, predominantly concentrated in pericentromeric regions, or in specific heterochromatin domains (Kumar and Bennetzen, 1999; SanMiguel *et al.*, 1996; Zhu *et al.*, 2003).

The mechanism by which *Tnt1* prefers coding regions is not known. However, as shown in Table 3 and Figure 3, the *Medicago truncatula* genome is AT-rich, with a calculated GC content of 33.2% overall, 41.0% in coding exons, 31.8% in introns plus UTRs, and 31.7% in intergenic regions. In the 85 bp region around the target site duplication, we found a statistically significant preference for GC in five positions (TSD1, TSD5, TSD+3, TSD+6, TSD+7) within and immediately downstream of the TSD (Table 3). Within this 12 bp putative target site (positions TSD1 to TSD+7), four positions (TSD4, TSD+1, TSD+2, TSD+4) appeared to be neutral. In TSD2, G was selected against (13.1%), in TSD3, T was preferred (38%), and in TSD+5, A was selected against (28.1%). The most significant statistical differences were seen at TSD1 (23.2, 18.5, 32.8 and 25.5%, for G, C, A and T, respectively) and at TSD+3 (18.1, 28.3, 25.9 and 27.8%) ($P < 0.001$, Table 3). These preferences, however, are too

weak to constitute a recognizable consensus sequence because the least-favored base T at position TSD1 still has a 25.5% chance of being targeted, and the most-favored base C at position TSD+3 is preferred by only 11.7% over the average pseudogenome base composition. By comparison, a recognizable consensus sequence for *Tos17* retrotransposon in rice has preferences for T (66%) at position -2, A (63%) at position +2, T (9%) and G (53%) at position -3, A (9%) and C (52%) at position +3, and a weaker preference for A (38%) at position -5 and T (40%) at position +5 (Miyao *et al.*, 2003). In this case, the 5 bp TSD itself was not conserved and the consensus has been described as relatively weak. In the stronger 9 bp consensus for the rice *mPing* or maize *mPIF* transposable elements, the 3 bp TSD is defined by TAA or TTA and flanked by moderately conserved 3 bp sequences on either side (Naito *et al.*, 2006; Zhang *et al.*, 2004). Nevertheless, the very weak target site preferences of *Tnt1* could account for the observed preferential targeting of coding exons, because statistical significance was observed with reference to the overall pseudogenome, and the observed base preferences are closer to the base composition of IMGAG CDS than to the rest of the genome. This shows that *Tnt1* has the ability to target a very wide area of the genome, and most coding regions could serve as potentially preferred target sites.

The exon preference of *Tnt1* has important implications for efficient genome tagging in crop species, which are in most cases dominated by transposable elements. Approximately, 250–300 Mb out of the 500 Mb genome of *M. truncatula* is estimated to be in the gene-dense euchromatin region (<http://www.medicago.org>), and, according to current IMGAG predictions, approximately 62% of this euchromatin gene space is intergenic or at least does not encode full-length proteins (Table 2). This suggests that less than a quarter of the *M. truncatula* genome contains protein coding regions. Thus, *Tnt1* offers a substantial advantage over any random insertion mutagen for saturation mutagenesis.

Assuming random insertion and 1.7 kb as the average gene size (obtained from current IMGAG gene prediction data), it can be estimated that 90% coverage of the entire *M. truncatula* genome would require over 670 000 inserts. As the average copy number during T-DNA tagging is 1.5 (Alonso *et al.*, 2003), over 450 000 individual transformation and regeneration events would be required to achieve 90% coverage using T-DNA. This probably under-estimates the number of inserts required because multiple T-DNA inserts usually occur in tandem and essentially have the effect of mutating a single locus even though three or more copies may be found per genome. Nevertheless, 450 000 transformation and regeneration events is still huge compared to only one transformation and approximately 27 000 regeneration events required to reach the same level of saturation using *Tnt1*. In addition, there is no need to assume random insertion, as we have demonstrated that *Tnt1*

prefers exons. As shown in Table 1, *Tnt1* has at least a 78.6% preference for gene-rich regions, and, within the euchromatic gene space, transcription units (exon plus intron) are targeted at a frequency of 57.3% (Table 2). Taking these facts into account and considering 300 Mb for the euchromatin gene space of *M. truncatula*, a more appropriate estimate could be derived from the equation $P = 1 - (1 - [x/y])^{R_{tr} \cdot R_{gs} \cdot N}$ where P is the probability of finding one *Tnt1* insert within a given gene, x is the average length of a gene transcript (1.7 kb), y is the total length of transcribed region in the gene space [$300 \times 37.7\%$ (15.9 ± 21.8), i.e. 113.1 Mb; Table 2], R_{tr} is the probability of insertion into a transcription region within the gene space (0.573), R_{gs} is the probability of insertion into gene space in the whole genome (0.786), and N is the total number of *Tnt1* inserts required. Based on this equation, and given that there are 25 inserts per line, we estimate that approximately 14 000–16 000 lines will be sufficient for a 90% probability of tagging any given gene in *M. truncatula* using *Tnt1* as an insertional mutagen. This estimate is highly dependent on gene size; while genes with a size of more than 2 kb can be easily saturated, genes smaller than 1 kb will be difficult to hit with more than 80% probability. The probability estimates for various gene size classes are shown in Figure 5.

Irrespective of the type of mutagen used, all procedures that require somatic embryogenesis have the inherent problem of generating tissue culture-induced mutations, and *M. truncatula* regeneration is no exception. In a *Tos17*-mutagenized rice population (R_1 progeny), less than 10% of the mutants were tagged (Miyao *et al.*, 2003). As this is a concern for *Tnt1* tagging, it is important to clarify the differences between the two systems. First, the rice mutants were generated from cell-suspension cultures, but the *M. truncatula* mutants were generated from individual leaf explant callus. As there is no means of tracking individual cells in liquid medium, cells that are mutated by other retrotransposons or DNA transposons may occur in quite a large number of the regenerated lines. However, in the *Tnt1* case, there is no possibility that one event could occur in two or more lines as only one plant was regenerated from one callus of a single leaf explant origin. Second, and most important, the rice cell culture was maintained for at least five months in tissue culture medium, and it was shown that the rate of *Tos17* transposition and other background mutations was directly proportional to the duration of tissue culture (Hirochika, 2001; Miyao *et al.*, 2003). In our system, the calli were maintained on auxin-containing tissue culture medium for an average of five weeks.

Despite the above concerns, we have direct evidence that the *M. truncatula* *Tnt1* mutants that we have characterized so far are *Tnt1*-tagged. Of the 20 mutant lines that we have analyzed so far, we confirmed *Tnt1* tagging in all 20 cases, which include new genes or genes that have been already described. Based on these observations, we believe that

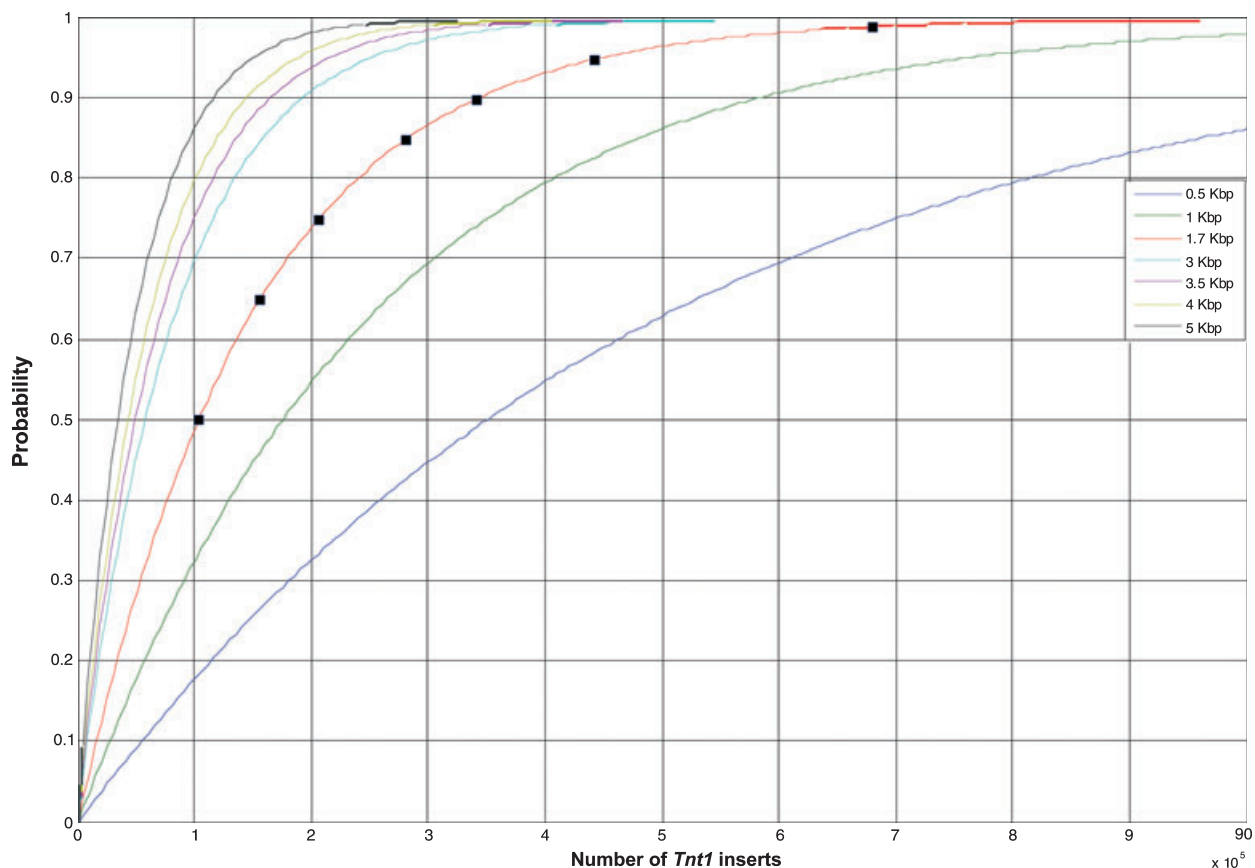


Figure 5. Genome coverage estimate for *Tnt1* inserts.

The estimate assumes a size of 300 Mb for the euchromatin region of *M. truncatula* and takes into account the property of *Tnt1* in preferentially targeting coding exons. The probability of tagging any given gene with *Tnt1* is a function of gene size and is derived from the equation $P = 1 - (1 - [x/y])^{R_{Tnt1} \cdot P_{gs} \cdot N}$ (see text for definition of variables). The most likely estimate, which assumes 1.7 kb for the average gene length, is shown in red. The data points on the red curve show the corresponding number of *Tnt1* inserts required to achieve 50–99% genome coverage.

most of the mutant phenotypes in our population are caused by *Tnt1* insertions. Although such analyses cannot be exhaustive and we have only tested a tiny fraction of the mutants so far, the fact that all the 20 mutants were tagged out of the 20 tested, compared to only four tagged out of 40 mutants tested in the rice *Tos17* system (Miyao *et al.*, 2003), suggests that *Tnt1* is a very efficient mutagen. Taken together, our results show that *Tnt1* tagging in *M. truncatula* is a very valuable system for legume functional genomics, and demonstrate that saturation insertional mutagenesis in model legumes is within reach. *Tnt1* appears to be a superior insertional mutagen and could be the primary choice for gene tagging in other plant species with large genomes. Consistent with this assertion, *Tnt1* has been recently shown to efficiently transpose and tag genes in lettuce (Mazier *et al.*, 2007), a plant whose genome size is approximately five times that of *M. truncatula*. Although its transposition efficiency has yet to be demonstrated in cereals, *Tnt1* has potential as a universal gene knockout tool for plant functional genomics as we move forward from the model *Arabidopsis* to more challenging crop plants.

Experimental procedures

Mutagenesis and regeneration

Transformation of *M. truncatula* genotype R108 with a construct containing the complete *Tnt1* element was performed as previously described (d'Erfurth *et al.*, 2003). One transgenic line (Tnk88-7-7) that contains five *Tnt1* copies was used as the starting material for generating all the lines via somatic embryogenesis as described in Appendix S1.

FST isolation and sequencing

Genomic DNA was isolated according to the method described by Doyle and Doyle (1990). The *Tnt1* flanking sequence tags (FSTs) were recovered by thermal asymmetric interlaced PCR (Liu *et al.*, 1995, 2005) or inverse PCR (Lucas *et al.*, 1995) as described in Appendix S1.

Homology searching and categorization

Tnt1 flanking sequences were aligned against the *Mt* pseudo-genome release version 1.0 (2007/04/30) (<http://www.medicago.org/genome/downloads/Mt1/>), the latest phase III BAC sequences

(2007/05/21) (<http://www.medicago.org/genome/downloads.php>) and MTGI version 8 (2005/01/19) (ftp://occams.dfci.harvard.edu/pub/bio/tgi/data/Medicago_truncatula) using BLASTN searches with an e-value cut-off $\leq 1.00E-10$, and against the NCBI protein database (NR, 2007/05/20) (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA>) using BLASTX with an e-value cut-off $\leq 1.00E-5$ to obtain the data in Table 1. The personal BLAST navigator (He *et al.*, 2007) and some in-house programs were used to analyze the BLAST alignments. The sizes of the IMGAG CDS, transcribed (TRANS) and pseudogenome version 1.0 databases were used to determine the sizes of the CDS, introns plus UTRs, and intergenic regions in Table 2 for comparison.

Target site specificity detection

For 843 FST sequences whose exact insertion sites were identified on the *M. truncatula* pseudogenome and whose up- and downstream sequences were both available in the pseudogenome, the frequency of individual base pairs in the target site duplication (TSD) and 40 bp on either side of the TSD was scored using an in-house program. For hot spot detection, all FSTs that matched to the *M. truncatula* pseudogenome were used for BLAST searches and mapping onto *M. truncatula* pseudochromosomes with an e-value cut-off $\leq 1.00E-50$ using the CVIT BLAST utility (http://www.medicago.org/genome/cvit_blast.php).

Forward-genetic screening

R₁ seeds were scarified with sulfuric acid, cold-treated for 1 week at 4°C on filter paper, and germinated for 24 h at room temperature. A maximum of 12 seedlings for each line were transferred into a 72-cell (6 × 12) tray (one seedling per cell) containing a mixture of commercial sand and perlite (in 1:3 ratio) and 15–20% arbuscular mycorrhizal fungi (Endorize-TA, Biorize; <http://www.biorize.com>). One-week-old seedlings were inoculated with *Sinorhizobium meliloti*, and trays were watered with low-nitrogen and low-phosphorus medium until phenotypic screening in the 5th week. Visible phenotypes were recorded on the 5th and 6th week with the aid of dissecting stereoscopes. After this initial screening, plants were transferred to soil with fertilizer and grown to maturity, and additional phenotypes were recorded as they appeared.

Acknowledgements

We thank Miki Hartwell, Janie Gallaway, Hee-Kyung Lee, Shuirong Zhang, and Trina Cottrell for their excellent technical assistance in tissue culture and plant maintenance in the greenhouse, and Andrea Harper (University of York, UK), Silvina Gonzalez-Rizzo, Viviane Cosson, Patricia Durand, Sumanjari Das (University of North Texas, USA), Guangming Li, Drs Rebecca Dickstein (University of North Texas, USA), Michael Schultze (University of York, UK), Michelle McConn and Paul Nakata (USDA, Bayers College of Medicine, USA), Alison Blancaflor, Aline Valster, Hongliang Wang, Willem Van-de-Velde, Kay Jones (Massachusetts Institute of Technology, USA), and all members of the Mysore laboratory, Udvardi laboratory and Dixon laboratory for their help in mutant screening. We thank Dr Michael Udvardi and Dr Richard A. Dixon for critical reading of the manuscript, Dr Yuhong Tang and Ann Harris for sequencing, and Drs Michael Schultze, Rujin Chen and Jeremy Murray for sharing unpublished data. This work was supported by the Samuel Roberts Noble Foundation, in part by the National Science Foundation plant genome grant number DBI 0703285, and by the European Union (EU FP6-GLIP project).

Supplementary Material

The following supplementary material is available for this article online:

Appendix S1. Methods used for mutagenesis and regeneration, FST isolation and sequencing, Southern blotting, confirmation of tagging by PCR analysis of alleles, and genetic complementation.

Table S1. Summary of FSTs that show high homology to known or predicted proteins and nucleotide sequences.

Table S2. Primers used for FST recovery and genotyping of *leaf lamina* mutants.

Figure S1. Estimation of copy number by Southern blot analysis in regenerated *Tnt1* lines.

Figure S2. Allelic confirmation of tagging of a *leaf lamina* mutant by *Tnt1*.

This material is available as part of the online article from <http://www.blackwell-synergy.com>

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

- Alonso, J.M., Stepanova, A.N., Leisse, T.J. *et al.* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, **301**, 653–657.
- An, G., Jeong, D.H., Jung, K.H. and Lee, S. (2005) Reverse genetic approaches for functional genomics of rice. *Plant Mol. Biol.* **59**, 111–123.
- Benlloch, R., d'Erfurth, I., Ferrandiz, C., Cosson, V., Beltran, J.P., Canas, L.A., Kondorosi, A., Madueno, F. and Ratet, P. (2006) Isolation of *mtpim* proves *Tnt1* a useful reverse genetics tool in *Medicago truncatula* and uncovers new aspects of AP1-like functions in legumes. *Plant Physiol.* **142**, 972–983.
- Bortiri, E., Jackson, D. and Hake, S. (2006) Advances in maize genomics: the emergence of positional cloning. *Curr. Opin. Plant Biol.* **9**, 164–171.
- Brocard, L., Schultze, M., Kondorosi, A. and Ratet, P. (2006) T-DNA mutagenesis in the model plant *Medicago truncatula*: is it efficient enough for legume molecular genetics? *CAB Rev.* **1**, 023. doi: 10.1079/PAVSNNR2006023.
- Cook, D.R. (1999) *Medicago truncatula* – a model in the making! *Curr. Opin. Plant Biol.* **2**, 301–304.
- Courtial, B., Feuerbach, F., Eberhard, S., Rohmer, L., Chiapello, H., Camilleri, C. and Lucas, H. (2001) *Tnt1* transposition events are induced by in vitro transformation of *Arabidopsis thaliana*, and transposed copies integrate into genes. *Mol. Genet. Genomics*, **265**, 32–42.
- Doyle, J.J. and Doyle, J.L. (1990) Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13–15.
- Elise, S., Etienne-Pascal, J., Fernanda, C.-N., Gerard, D. and Julia, F. (2005) The *Medicago truncatula* *SUNN* gene encodes a CLV1-like leucine-rich repeat receptor kinase that regulates nodule number and root length. *Plant Mol. Biol.* **58**, 809–822.
- d'Erfurth, I., Cosson, V., Eschstruth, A., Lucas, H., Kondorosi, A. and Ratet, P. (2003) Efficient transposition of the *Tnt1* tobacco retrotransposon in the model legume *Medicago truncatula*. *Plant J.* **34**, 95–106.
- Fladung, M., Deutsch, F., Honicka, H. and Kumar, S. (2004) T-DNA and transposon tagging in aspen. *Plant Biol.* **6**, 5–11.
- Frugoli, J. and Harris, J. (2001) *Medicago truncatula* on the move. *Plant Cell*, **13**, 458–463.

- Graham, P.H. and Vance, C.P. (2003) Legumes: importance and constraints to greater use. *Plant Physiol.* **131**, 872–877.
- Grandbastien, M.-A., Spielmann, A. and Caboche, M. (1989) *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature*, **337**, 376–380.
- He, J., Dai, X. and Zhao, X. (2007) PLAN: a web platform for automating high-throughput BLAST searches and for managing and mining results. *BMC Bioinformatics*, **8**, 53, doi: 10.1186/1471-2105-8-53.
- Hirochika, H. (2001) Contribution of the *Tos17* retrotransposon to rice functional genomics. *Curr. Opin. Plant Biol.* **4**, 118–122.
- Hofer, J., Turner, L., Hellens, R., Ambrose, M., Matthews, P., Michael, A. and Ellis, N. (1997) *UNIFOLIATA* regulates leaf and flower morphogenesis in pea. *Curr. Biol.* **7**, 581–587.
- Krysan, P.J., Young, J.C. and Sussman, M.R. (1999) T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell*, **11**, 2283–2290.
- Kumar, A. and Bennetzen, J.L. (1999) Plant retrotransposons. *Annu. Rev. Genet.* **33**, 479–532.
- Kumar, A. and Hirochika, H. (2001) Applications of retrotransposons as genetic tools in plant biology. *Trends Plant Sci.* **6**, 127–134.
- Le, Q.H., Melayah, D., Bonnivard, E., Petit, M. and Grandbastien, M.-A. (2007) Distribution dynamics of the *Tnt1* retrotransposon in tobacco. *Mol. Genet. Genomics*, **278**, 639–651.
- Lei, Z., Elmer, A.M., Watson, B.S., Dixon, R.A., Mendes, P.J. and Sumner, L.W. (2005) A two-dimensional electrophoresis proteomic reference map and systematic identification of 1367 proteins from a cell suspension culture of the model legume *Medicago truncatula*. *Mol. Cell. Proteomics*, **4**, 1812–1825.
- Limpens, E., Ramos, J., Franken, C., Raz, V., Compaan, B., Franssen, H., Bisseling, T. and Geurts, R. (2004) RNA interference in *Agrobacterium rhizogenes*-transformed roots of *Arabidopsis* and *Medicago truncatula*. *J. Exp. Bot.* **55**, 983–992.
- Liu, Y.G., Mitsukawa, N., Oosumi, T. and Whittier, R.F. (1995) Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* **8**, 457–463.
- Liu, Y.G., Chen, Y. and Zhang, Q. (2005) Amplification of genomic sequences flanking T-DNA insertions by thermal asymmetric interlaced polymerase chain reaction. *Methods Mol. Biol.* **286**, 341–348.
- Lucas, H., Feuerbach, F., Kunert, K., Grandbastien, M.A. and Caboche, M. (1995) RNA-mediated transposition of the tobacco retrotransposon *Tnt1* in *Arabidopsis thaliana*. *EMBO J.* **14**, 2364–2373.
- Madden, S.L., Galella, E.A., Zhu, J., Bertelsen, A.H. and Beaudry, G.A. (1997) SAGE: transcript profiles for p53-dependent growth regulation. *Oncogene*, **15**, 1079–1085.
- Marsh, J.F., Rakocevic, A., Mitra, R.M., Brocard, L., Sun, J., Eschstruth, A., Long, S.R., Schultze, M., Ratet, P. and Oldroyd, G.E. (2007) *Medicago truncatula* *NIN* is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. *Plant Physiol.* **144**, 324–335.
- May, G.D. and Dixon, R.A. (2004) *Medicago truncatula*. *Curr. Biol.* **14**, R180–R181.
- Mazier, M., Botton, E., Flamain, F., Bouchet, J.-P., Courtial, B., Chupeau, M.-C., Chupeau, Y., Maisonneuve, B. and Lucas, H. (2007) Successful gene tagging in lettuce using the *Tnt1* retrotransposon from tobacco. *Plant Physiol.* **144**, 18–31.
- Mhiri, C., Morel, J.-B., Vernhettes, S., Casacuberta, J.M., Lucas, H. and Grandbastien, M.-A. (1997) The promoter of the tobacco *Tnt1* retrotransposon is induced by wounding and by abiotic stress. *Plant Mol. Biol.* **33**, 257–266.
- Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K. and Hirochika, H. (2003) Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell*, **15**, 1771–1780.
- Modolo, L., Blount, J., Achnine, L., Naoumkina, M., Wang, X. and Dixon, R. (2007) A functional genomics approach to (iso)flavonoid glycosylation in the model legume *Medicago truncatula*. *Plant Mol. Biol.* **64**, 499–518.
- Naito, K., Cho, E., Yang, G., Campbell, M.A., Yano, K., Okumoto, Y., Tanisaka, T. and Wessler, S.R. (2006) Eukaryotic transposable elements and genome evolution special feature: dramatic amplification of a rice transposable element during recent domestication. *Proc. Natl Acad. Sci. USA*, **103**, 17620–17625.
- Penmesta, R.V. and Cook, D.R. (2000) Production and characterization of diverse developmental mutants of *Medicago truncatula*. *Plant Physiol.* **123**, 1387–1397.
- SanMiguel, P., Tikhonov, A., Jin, Y. et al. (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science*, **274**, 765–768.
- Schenk, R.U. and Hildebrandt, A.C. (1972) Medium and techniques for induction and growth of monocotyledonous and dicotyledonous plant cell cultures. *Can. J. Bot.* **50**, 199–204.
- Scholte, M., d'Erfurth, I., Ripa, S. et al. (2002) T-DNA tagging in the model legume *Medicago truncatula* allows efficient gene discovery. *Mol. Breeding*, **10**, 203–215.
- Snedecor, G.W. and Cochran, W.G. (1989) *Statistical Methods*, 8th edn. Ames, IA: Iowa State University Press.
- Somers, D.A., Samac, D.A. and Olhott, P.M. (2003) Recent advances in legume transformation. *Plant Physiol.* **131**, 892–899.
- Tadege, M. and Mysore, K.S. (2006) Insertional mutagenesis for plant functional genomics. In *Floriculture, Ornamental and Plant Biotechnology: Advances and Topical Issues* (Teixeira da Silva, J.A., ed). Isleworth, UK: Global Science Books, pp. 608–618.
- Tadege, M., Ratet, P. and Mysore, K.S. (2005) Insertional mutagenesis: a Swiss Army knife for functional genomics of *Medicago truncatula*. *Trends Plant Sci.* **10**, 229–235.
- Town, C.D. (2006) Annotating the genome of *Medicago truncatula*. *Curr. Opin. Plant Biol.* **9**, 122–127.
- Udvardi, M.K., Tabata, S., Parniske, M. and Stougaard, J. (2005) *Lotus japonicus*: legume research in the fast lane. *Trends Plant Sci.* **10**, 222–228.
- Vernhettes, S., Grandbastien, M.-A.I. and Casacuberta, J.M. (1997) In vivo characterization of transcriptional regulatory sequences involved in the defence-associated expression of the tobacco retrotransposon *Tnt1*. *Plant Mol. Biol.* **35**, 673–679.
- Walbot, V. (2000) Saturation mutagenesis using maize transposons. *Curr. Opin. Plant Biol.* **3**, 103–107.
- Wang, H., Li, G. and Chen, R. (2006) Fast neutron bombardment (FNB) mutagenesis for forward and reverse genetic studies in plants. In *Floriculture, Ornamental and Plant Biotechnology Advances and Topical Issues* (Teixeira da Silva, J.A., ed). Isleworth, UK: Global Science Books, pp. 629–639.
- Wessler, S.R. (2006) Transposable elements and the evolution of eukaryotic genomes. *Proc. Natl Acad. Sci. USA*, **103**, 17600–17601.
- Young, N.D. and Shoemaker, R.C. (2006) Genome studies and molecular genetics. Part 1: model legumes. Exploring the structure, function and evolution of legume genomes. *Curr. Opin. Plant Biol.* **9**, 95–98.

- Zhang, X., Jiang, N., Feschotte, C. and Wessler, S.R.** (2004) PIF- and Pong-like transposable elements: distribution, evolution and relationship with Tourist-like miniature inverted-repeat transposable elements. *Genetics*, **166**, 971–986.
- Zhou, X., Chandrasekharan, M.B. and Hall, T.C.** (2004) High rooting frequency and functional analysis of GUS and GFP expression in transgenic *Medicago truncatula* A17. *New Phytol.* **162**, 813–822.
- Zhu, Y., Dai, J., Fuerst, P.G. and Voytas, D.F.** (2003) Controlling integration specificity of a yeast retrotransposon. *Proc. Natl Acad. Sci. USA*, **100**, 5891–5895.