Scientific Correspondence

A TILLING Reverse Genetics Tool and a Web-Accessible Collection of Mutants of the Legume *Lotus japonicus*[1]

Jillian A. Perry, Trevor L. Wang, Tracey J. Welham, Sarah Gardner, Jodie M. Pike, Satoko Yoshida, and Martin Parniske*

The Sainsbury Laboratory (J.A.P., S.G., J.M.P., S.Y., M.P.), John Innes Centre (Tre.L.W., Tra.W.), Colney Lane, Norwich NR4 7UH, United Kingdom

Reverse genetics aims to identify the function of a gene with known sequence by phenotypic analysis of cells or organisms in which the function of this gene is impaired. Commonly used strategies for reverse genetics encompass transposon mutagenesis (Tissier et al., 1999) and RNA-mediated gene silencing or RNA interference (Voinnet, 2002). We adopted a complementary strategy to set up a reverse genetics tool for the legume *Lotus japonicus* that identifies individuals carrying point mutations in any gene of interest within a large population of ethyl methanesulfonate (EMS)-mutagenized M_2 plants. This strategy was first described by McCallum et al. (2000a,b) using the acronym TILLING (Targeted Induced Local Lesions in Genomes). The target sequence is PCR amplified from pooled M_2 individuals. DNA with point mutations are detected by melting and reannealing of the PCR products. This results in the formation of heteroduplex DNA in which one strand originates from the mutant and the other from the wild-type PCR product. A mismatch occurs at the site of the point mutation, which can be detected using mismatch-specific endonucleases such as CEL I from celery (*Apium graveolens*; Yang et al., 2000). This enzyme recognizes mismatches in heteroduplex DNA and cleaves DNA specifically at the mismatched site. The cleavage products can be separated by gel electrophoresis, typically sequencing-type denaturing PAGE. This method of mismatch detection is amenable to pooling strategies. In the Arabidopsis TILLING facility, DNA of eight M_2 plants is mixed to form a pool (Colbert et al., 2001). At this pool size, a population of 768 individuals can be screened by PCR in a 96-well microtiter plate, and run on one 96-well gel, each well representing eight individuals. Individuals from pools yielding cleavage products are then PCR amplified individually to identify the mutation bearing plant, progeny of which will segregate the mutation of interest.

Although a high-throughput TILLING facility has been successfully established for Arabidopsis (Colbert et al., 2001), several aspects of plant biology cannot be studied in this model plant; for example, root symbiosis with rhizobia and arbuscular mycorrhiza fungi, compound leaf development, aspects of flower development, and perenniality. We chose the model legume *L. japonicus* (Handberg and Stougaard, 1992) to establish a legume reverse genetics tool. The genome of this plant is subject of a sequencing project (VandenBosch and Stacey, 2003).

One problem associated with EMS mutagenesis is the high frequency of infertile plants not only in the M_1 but also in subsequent generations. Our aim was to assemble a general TILLING population that would ideally be composed of fertile individuals only, so that progeny of a plant carrying a mutant allele can be directly recovered. A general TILLING population of 3,697 independent M_2 plants was established biased against the occurrence of severe developmental phenotypes to maximize M_3 seed availability (Fig. 1). DNA was prepared in 96-well microtiter plate format and seed from each individual was collected.

A specific advantage of EMS mutagenesis is that a series of allelic mutations can be obtained, displaying a range of phenotypes that can serve as the basis of detailed structure function studies. This also has the potential to recover weak alleles with subtle changes in functionality of genes that would be lethal when more strongly affected. However, the recovery of strongly affected or knockout alleles will often be the primary interest in species for which no insertional knockout mutagenesis tool is established. To enrich for plants bearing functionally impaired mutant alleles of genes involved in a variety of developmental and metabolic processes, we scored approximately 45,600 M_2 progeny of 4,190 EMS-mutagenized M_1 plants, and isolated mutants affected in metabolism, morphology, and the root nodule symbiosis (Figs. 1 and 2; Table I). A database comprising information on individual mutant plants including photographs where appropriate is accessible at http://www. lotusjaponicus.org/finder.htm. This allows for the
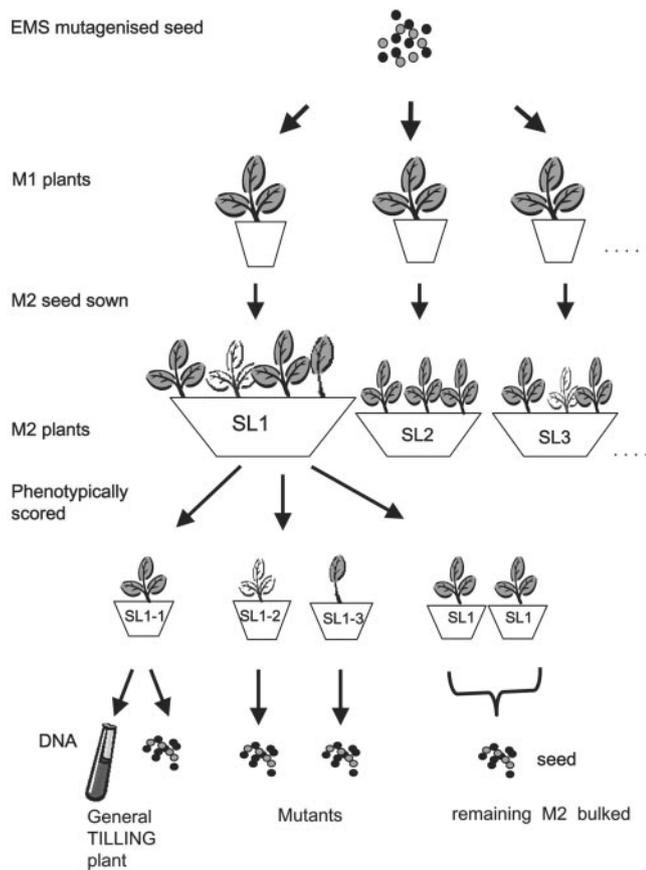
Figure 1. Structure of the TILLING population. EMS-mutagenized seed of *L. japonicus* B-129 "Gifu" resulted in approximately 4,190 fertile M$_1$ plants. M$_2$ seed was harvested from individual M$_1$ plants, and M$_2$ families were grown individually. The families were kept separate to maximize diversity of point mutations represented in the TILLING population. Many M$_1$ plants suffered from poor seed set. Dependent on availability, up to 30 seeds were sown per M$_2$ family, and the resulting total of 45,600 plants were inoculated after 2 weeks with *M. loti* and scored after a total of 6 weeks for the development of symbiotic nitrogen-fixing root nodules, and screened for other developmental defects (Table I). A single healthy-looking plant was chosen for the general TILLING population and designated with the number SL n-1, where n is a sequential number representing the family, whereas the number behind the dash identifies the sibling. All interesting mutants identified in a family were recorded and designated with consecutive sibling numbers (SL n-2, n-3, etc.) and seed harvested individually. Seed from all remaining plants was collected in bulk so that each seed pack contained seeds from only one family. This seed was collected for the purpose of future forward screens and as a backup for the isolation of mutant alleles known to segregate in a particular family, in case the phenotypically interesting M$_2$ mutants were infertile.

assembly of trait-specific or theme-based TILLING populations enriched for mutants affected in a particular developmental process.

A pilot experiment was performed in which the *SYMRK* gene was subjected to the TILLING procedure (Fig. 3). The *SYMRK* gene is required for the formation of root symbioses by *L. japonicus*, and pre-

Table I. *Morphological, symbiotic and metabolic mutants of L. japonicus Gifu*

Mutants in categories architecture, flower, leaf, root, and stem originate from scoring an initial 2415 M2 families (28,500 M2 plants). A total of 3843 M2 families (45,600 plants) were screened for nodulation mutants[a], and mutants with either no nodules, or small white nodules as well as plants with a reduced or increased nodule number (supernodulating mutants) were isolated. 1428 M2 families (17,100 plants) were screened for starch biosynthesis and breakdown (see supplementary materials)[b]. The numbers recorded represent characters scored on individual plants and thus may be sibling aggregates. Furthermore, a single plant may be altered in more than one character and will be recorded in more than one section. A database comprising all the mutant information including photographs can be found at http://www.lotusjaponicus.org/finder.htm

Morphological Character	No. of M$_2$ Plants
Architecture	
Shape	
Bushy	54
Single stem	35
Size	
Compact	7
Giant	1
Miniature	49
Stature	
Erect	25
Horizontal (forced)	18
Flower	
Color	
Pale yellow	1
Morphology	
Abnormal all	5
Abnormal standard	2
Abnormal wing	3
Veins pale/absent	10
Reproductive organs	
Abnormal anthers	1
Abnormal both	1
No flowers	4
Size	
Absent	411
Small	2
Fruit	
Fertility	
Infertile	540
Low fertility	134
Pod Shape	
Curved	6
Distorted	1
Short	87
Seeds	
Abnormal	1
Few	6
Leaf	
Appearance	
Flecked	28
Mottled	89
Spotted	55
Variegated	31
Color	
Dark green	63
Lime green	149
Pale green	140
Veins/lamina contrast	13
Yellow	59
Yellow at nod screen	130
Yellow when young	26

(*Table continues on next page*)

Table I. Continued

Morphological Character	No. of M$_2$ Plants
Leaf shape	
Broad	9
Narrow	275
Pinched	12
Rounded	24
Thick	23
Very narrow	42
Leaflets	
Fewer (than five)	8
Variable no.	8
Shape	
Crinkled	185
Crinkled and downcurled	34
Distorted	2
Downcurled	110
Margins raised	118
Twisted	3
Upcurling	17
Size	
Large	11
Small	485
Tiny	127
Nodule	
Color	
Brown	2
Green	1
White[a]	178
Yellow	4
No.	
Few[a]	57
Many (super. hyper)	8
None	207
Size	
Bump	2
Giant	9
Small[a]	159
Product[b]	
Starch	
Biosynthesis	110
Starch	
Breakdown	73
Root	
Color	
Brown	1
Transparent	3
Yellow	2
Length	
Long	1
Short	21
Stubby	1
Stem	
Diameter	
Thick	36
Thin	16
Length	
Dwarf ($^1/_4$–$^2/_3$ normal)	333
Extreme dwarf	146
Shape	
Angled	3
Stiff	4
Twisted	1

viously identified mutations lead to a non-nodulating phenotype (Stracke et al., 2002). The genomic sequence of *SYMRK* from start to stop codon extends over 5,472 kb, but the maximum length of PCR products that can be conveniently resolved by sequencing-type PAGE is just over 1 kb. Therefore, we used the program CODDLE (Codons Optimized to Discover Deleterious Lesions; http://www.proweb.org/input/) to identify a region within the *SYMRK* gene that would have the highest likelihood to be functionally affected by EMS mutagenesis. EMS induces mostly G/C to A/T transitions (Anderson, 1995), and CODDLE uses this information to predict the EMS-induced changes in a coding sequence, and plots the probability of missense and nonsense mutations along the coding sequence (Fig. 3). Second, the program searches the predicted protein sequence under study for the occurrence of conserved amino acid sequence blocks, with the idea that changes in such conserved regions are less likely to be functionally neutral. Figure 3 shows part of a CODDLE output for the *SYMRK* gene. We confined the screen for mutant alleles to the region encoding the protein kinase domain. Primers for three overlapping PCR amplicons spanning the entire kinase domain were designed (Figs. 3 and 4) using CODDLE in combination with the PRIMER3 program (Rozen and Skaletsky, 2000) as set up on the CODDLE Web page (http://www.proweb.org/input/).
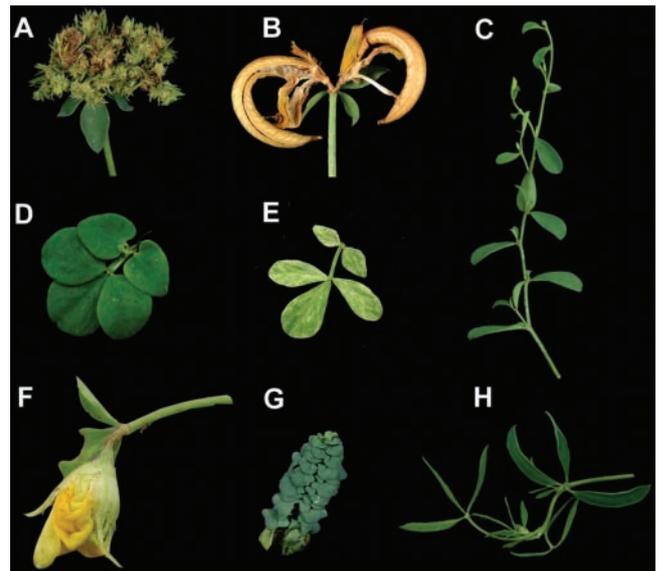

Figure 2. A selection of morphological mutants from M$_2$ families of *L. japonicus*. A, SL1203-3 is a flower mutant bearing abnormal reiterated flower structures. B, SL5428 exhibits a fruit phenotype with curled pods and delayed senescence of the petals. This represents a dominant mutation that was manifest in the M$_1$ plant. C, In SL2143-2, a single leaf replaces the normal five leaflets of the compound leaf. D, SL2112-4 is a leaf shape mutant that has rounded leaflets. E, SL189-2 shows chlorophyll variegation in the leaflets. F, SL1077-2 has abnormal flowers in which the petals do not expand fully. G, SL3534-2 is an extreme dwarf. H, SL2137-2 exhibits narrow leaflets.
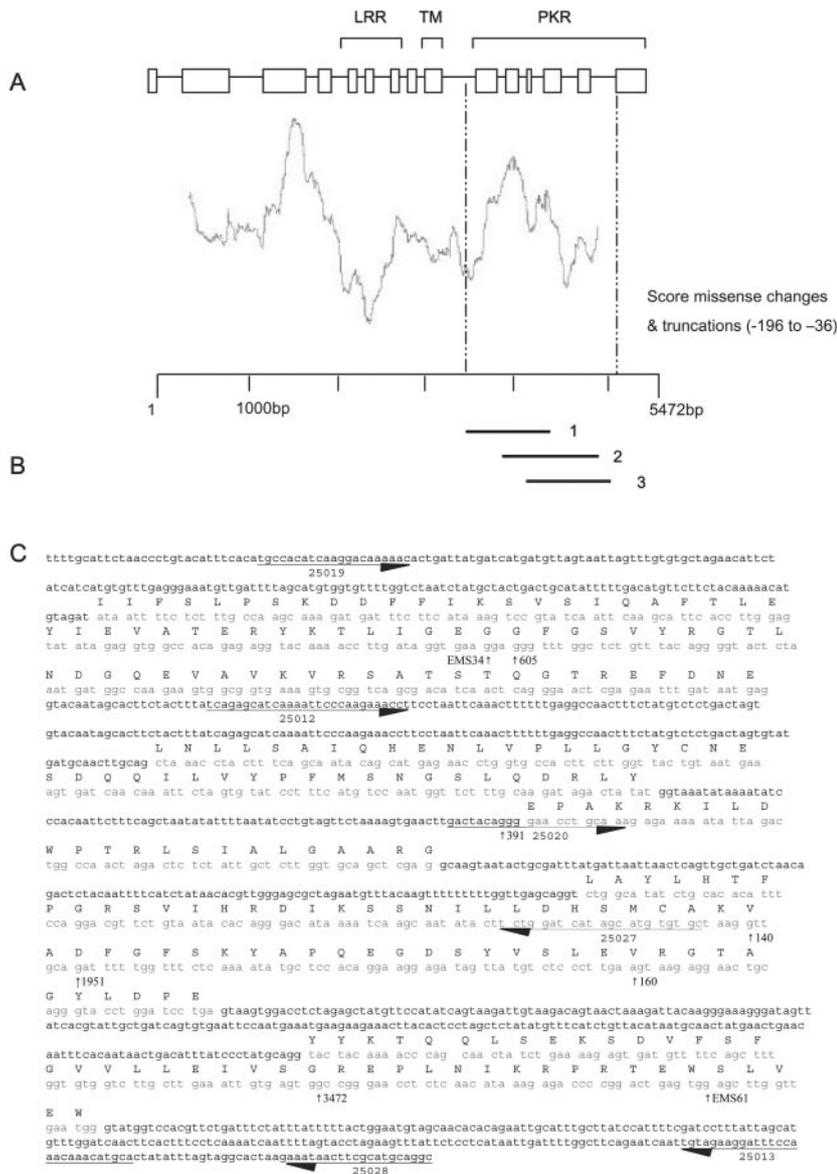
Figure 3. Effect of EMS mutagenesis on the SYMRK gene. A, Output of the CODDLE program for 5,472 bp of genomic sequence of *SYMRK*. Exons are represented by white boxes and introns by black lines. Regions encoding Leu-rich repeat (LRR), transmembrane domain (TM), and protein kinase region (PKR) are indicated by brackets. The CODDLE program was used to identify regions of the *SYMRK* gene in which G/C to A/T transitions are most likely to result in deleterious effects. Each point on the plot is the sum of scores calculated for a 500-bp window centered at that residue. A residue susceptible to a nonsense change scored +6, to a missense change scored 0, to a silent change scored −1, and to a splice junction scored +4 (McCallum et al., 2000b; http://www.proweb.org/glossary.html). B, Three amplicons (primer combination [forward and reverse[: 1, 25019 and 25027; 2, 25012 and 25013; and 3, 25020 and 25028) chosen to cover the protein kinase region. C, Sequence of the region covered by the amplicons 1, 2, and 3, delimited by dashed lines in A. The positions and family identifier of mutations are indicated by vertical arrows. In all identified mutants, a G residue was mutated to an A, which is consistent with the predominant G/C to A/T transitions induced by EMS (Anderson, 1995). Primer positions and orientations are indicated by horizontal arrows. Forward primers 25019 GCACACAT-GCTATGATCCAGA, 25012 TCAGAGCAT-CAAAATTCCCAAGAAACC, and 25020 GAC-TACAGGGGAACCTGCAA were labeled with 6-carboxyfluorescein (6-FAM). Reverse primers 25027 GCACACATGCTATGATCCAGA, 25013 TGCATGTTTGTTTGGAAATCCTTCTACA, and 25028 GCCTGCATGCGAAGTTATTT were labeled with 4,7,2′,7′-tetrachloro-6-carboxy-fluorescein (TET).
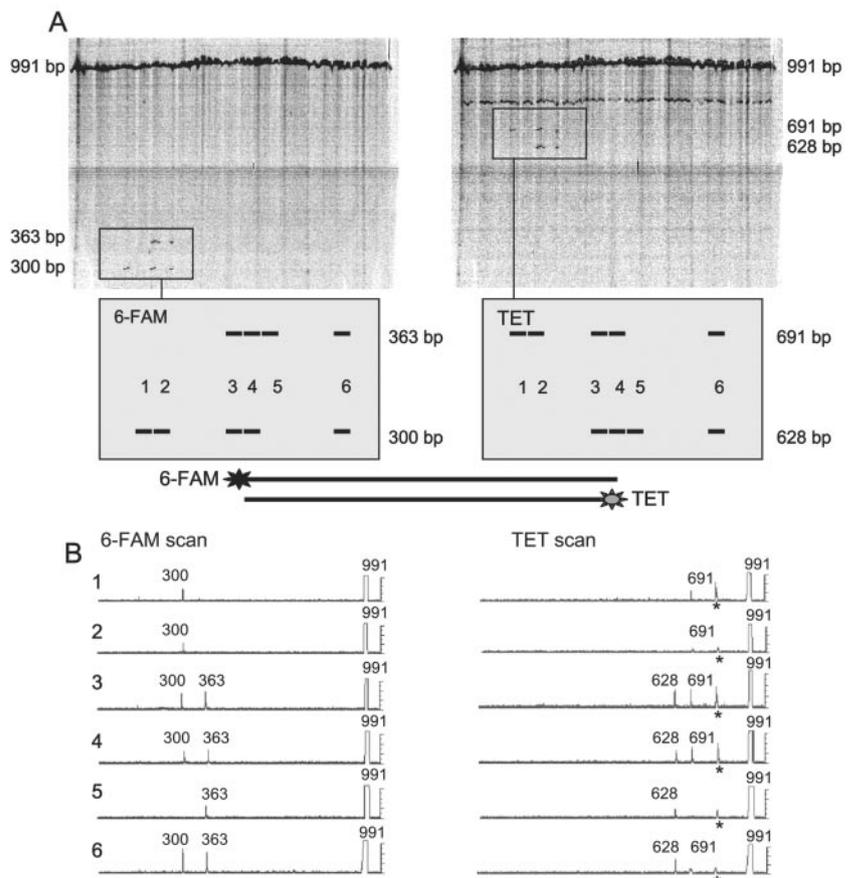
A population of 117 non-nodulating plants from 75 families, 135 plants from 112 families carrying few or small or white nodules, and 23 plants from 21 families with abnormalities in root development were inspected for the occurrence of point mutations in the SYMRK kinase domain. Thirteen symbiosis-defective mutants that were isolated in an independent mutagenesis program in other laboratories were included as reference. Two of these (EMS34 and EMS61) had been assigned previously to the *SYMRK* (*Ljsym2*) complementation group (Szczyglowski et al., 1998; Stracke et al., 2002). These 288 mutants were screened at a pool size of three individuals per pool. In this preselected population, 15 homozygous mutants representing six different alleles were identified that carried missense mutations. Furthermore, a mutation in plant SL391-2 in the splice acceptor site of the 11th intron was found that is likely to affect

splicing (Table II). The nonsense mutation in EMS61 published previously (Stracke et al., 2002) served as an internal reference control. All of these mutations were detected in individuals that did not produce root nodules upon inoculation with *Mesorhizobium loti*.

An M_1 plant EMS treated at the embryo stage represents a mosaic of differentially mutagenized cells. Because more than one cell of the embryo gives rise to the germ line, less than 75% of the M_2 progeny are expected to segregate a particular mutant allele. Only one (SL1951) of the six novel mutant alleles was represented in a heterozygous state in the nodulating sibling of the general TILLING population.

The high number of functionally affected mutant alleles in the preselected mutants shows that, depending on the purpose of the resource, it might be advisable to include a phenotypic screen for the trait

Figure 4. CEL1 mismatch detection. A, Fluorescence emission scans of a 96-well gel showing CEL1-treated PCR products (primer combination 25020 and 25028) that result in an uncut band (991 bp) and cleavage products from mismatched heteroduplexes. Emission scans for forward primer labeled with 6-FAM (left image) and reverse primer labeled with TET (right image) shows cleavage products of 300, 363, 628, and 691 bp, respectively. For clarity, the insets show schematic diagrams of the cleavage products observed in the original gels. The numbered lanes 1 to 6 correspond to six different 3x pools containing the following mutants (Table II): pool 1, SL1951-4; pool 2, SL1951-5; pool 3, SL1951-6 and SL160-4; pool 4, SL1951-7 and SL160-5; pool 5, SL160-6; and pool 6, SL1951-2 and SL160-3. B, Fluorescent traces of six individual lanes representing the pools above from the gel shown in A. The traces on the left show the scan for 6-FAM, which labels the top strand. Cleaved (300 and 363 bp) and non-cleaved (991 bp) products were detected, and fragment length could be estimated due to the inclusion of size standards (not shown). The right traces represent the scan for TET, which labels the bottom strand. Cleaved (628 and 691 bp) and non-cleaved products (991 bp) were detected. The CEL I assay was essentially performed as described by Colbert et al. (2001), but PCR was carried out in the absence of unlabeled primer. Instead of Sephadex filtration, reaction products were precipitated with ethanol and separated by PAGE using an ABI 377 sequencer.

of interest when setting up TILLING in a particular organism.

ACKNOWLEDGMENTS

Table II. *Mutations in the region encoding the kinase domain of SYMRK*

The base and amino acid change are listed in addition to the position of the mutation in genomic DNA sequence relative to the A of the start codon. Non-nodulating M2 plants homozygous for alleles SL605, SL391, SL 1951, and SL160 set seed, and the non-nodulating phenotype was stably inherited to the M3 generation, but the two non-nodulating individuals, SL140−2 and SL3472−2, did not set seed. We therefore tested M3 seed from the bulk harvested families SL140 and SL3472, but we could not identify non-nodulating plants in the sibling progeny.

Plant Identifier	Mutation	Amino Acid Change	Amino Acid Position	Nucleic Acid Position
EMS 34[a]	GGA to AGA	Gly to Arg	603	3726
SL605−2 and 3	GGG to AGG	Gly to Arg	604	3729
SL391−2	GGG to AGG		Splice acceptor	4155
SL140−2	GTT to ATT	Val to Ile	736	4441
SL3472−2	GGC to GAC	Gly to Asp	793	4832
EMS 61[b]	TGG to TGA	Trp to Stop	806	4872
SL1951−2, 3, 4, 5, 6, and 7	GAT to AAT	Asp to Asn	738	4447
SL160−2, 3, 4, 5, and 6	GGA to AGA	Gly to Arg	759	4510

[a]From Szczyglowski et al. (1998). [b]From Szczyglowski et al. (1998); Stracke et al. (2002).

LITERATURE CITED

Anderson P (1995) Methods Cell Biol **48:** 31–58

Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung AT, McCallum CM, Comai L, Henikoff S (2001) Plant Physiol **126:** 480–484

Handberg K, Stougaard J (1992) Plant J **2:** 487–496

McCallum CM, Comai L, Greene EA, Henikoff S (2000a) Nat Biotechnol **18:** 455–457

McCallum CM, Comai L, Greene EA, Henikoff S (2000b) Plant Physiol **123:** 439–442

Rozen S, Skaletsky H (2000) Methods Mol Biol **132:** 365–386

Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J, Szczyglowski K, Parniske M (2002) Nature **417:** 959–962

Szczyglowski K, Shaw RS, Wopereis J, Copeland S, Hamburger D, Kasiborski B, Dazzo FB, de Bruijn FJ (1998) Mol Plant-Microbe Interact **11:** 684–697

Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, Jones JD (1999) Plant Cell **11:** 1841–1852

VandenBosch K, Stacey G (2003) Plant Physiol **131:** 840–865

Voinnet O (2002) Curr Opin Plant Biol **5:** 444

Yang B, Wen X, Kodali NS, Oleykowski CA, Miller CG, Kulinski J, Besack D, Yeung JA, Kowalski D, Yeung AT (2000) Biochemistry **39:** 3533–3541