Theor Appl Genet (2007) 115:373–382
DOI 10.1007/s00122-007-0571-8

ORIGINAL PAPER

The M flax rust resistance pre-mRNA is alternatively spliced and contains a complex upstream untranslated region

Simon Schmidt · Maria Lombardi · Donald M. Gardiner · Michael Ayliffe · Peter A. Anderson

Received: 28 April 2006 / Accepted: 27 April 2007 / Published online: 30 May 2007

Abstract Alternative splicing is an important step in controlling gene expression and has been shown to occur for a number of plant disease resistance (R) genes. The specific biological role of alternatively spliced transcripts from most R genes is unknown, yet in two cases it is clear that functional disease resistance cannot be activated without them. We report 12 splice isoforms of the *M* flax rust resistance gene, a TIR–NBS–LRR class of R gene. Collectively, these isoforms are predicted to encode at least nine different polypeptide products, only one of which is a full length peptide believed to confer functional *M* gene-specific disease resistance. An additional intron to that previously described was found in the 5′ untranslated region. Splicing of this leader intron removes an upstream ORF (μORF) sequence. In some transcripts the leader intron is retained and in this case we predict negligible translation initiation of the full length *M* gene-encoding ORF. The majority of the alternatively spliced isoforms of *M* would encode truncated TIR and TIR–NBS containing proteins. Although the role of alternative splicing and the existence and function of the products they encode is still unclear, the complexities of the splicing profile, and the 5′ UTR of the *M* gene, are likely to serve in mechanisms to regulate R protein levels.

Communicated by D. A. Lightfoot.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-007-0571-8) contains supplementary material, which is available to authorized users.

S. Schmidt · M. Lombardi · D. M. Gardiner · P. A. Anderson (✉)
School of Biological Sciences, Flinders University,
GPO Box 2100, Adelaide, SA 5001, Australia
e-mail: Peter.Anderson@flinders.edu.au

M. Ayliffe
CSIRO, Plant Industry, GPO Box 1600,
Canberra, ACT 2601, Australia

Present Address:
S. Schmidt
Australian Wine Research Institute,
Adelaide, SA, Australia

Introduction

In order to recognise and respond to an invading pathogen, plants employ resistance (R) genes, the products of which are capable of recognising determinants from the invading pathogen. In many cases specific, constitutively expressed R gene products, recognise the presence of specific pathogens via effector molecules expressed from Avirulence (Avr) genes. Such a system implies that the products of R gene expression await the arrival of a detectable (avirulent) pathogen, at which point signalling pathways are activated that result in disease resistance (Staskawicz et al. 1995; Dangl and Jones 2001). Despite an emerging understanding of the signalling machinery (Holt et al. 2005; Glazebrook 2005), little is known about the regulation of R gene expression.

R genes from many different plant species have been cloned, and despite the diversity of plant-pathogen interactions in which they are involved, R proteins contain common conserved motifs. By far the largest class of R genes are those encoding a protein containing a nucleotide binding site (NBS) and a leucine rich repeat region LRR. The NBS–LRR proteins can be further subdivided by their N-terminal region that either contains a putative coiled-coil (CC) region, or a domain sharing homology to the *Drosophila* Toll and mammalian Interleukin-1 receptor proteins (TIR) (Meyers et al. 2003, 2005). The *M* gene of flax

is a TIR–NBS–LRR R gene and is a member of a complex locus containing a tandem array of related sequences (Anderson et al. 1997; Ellis et al. 1995). In contrast, the *L6* gene is a member of a multi-allelic *L* locus (Lawrence et al. 1995; Ellis et al. 1999). The *M* and *L6* genes encode proteins that share 78% amino acid identity.

Alternative splicing is becoming recognised as an important regulatory step in gene expression. Of the one third of human genes that are spliced, 45% display alternatively spliced forms (Gupta et al. 2004), and 21.8 and 21.2% of *Arabidopsis* and rice ESTs respectively, display alternative spliced forms (Wang and Brendel 2006). It is clear, therefore, that alternative splicing is not an infrequent event. Of the TIR–NBS–LRR-like plant disease resistance genes that have been studied, a number are known to produce multiple transcripts by alternative splicing. These genes include *N* from tobacco (Whitham et al. 1994), *L6*, *M* and *N* from flax (Lawrence et al. 1995; Anderson et al. 1997; Dodds et al. 2001a), and *RPP5* and *RPS4* from *Arabidopsis* (Parker et al. 1997; Gassmann et al. 1999). An exception is the *P2* gene of flax, which although containing an additional non-LRR C-terminal domain, does not appear to generate alternatively spliced products (Dodds et al. 2001b). There are less frequent reports of alternative splicing in the CC-NBS-LRR class of R genes, an exception being the *JA1tr* gene from *Phaseolus* in which seven alternative spliced isoforms are reported (Ferrier-Cana et al. 2005), four of which are predicted to encode truncated CC-NBS proteins.

In most cases, alternative R gene transcripts are predicted to encode peptides containing little or none of the LRR region and in some cases composed just of the TIR domain. The existence of these truncated peptides in plant tissue has not been demonstrated; however RT-PCR time course and transgenic experiments indicate that alternate spliced isoforms play a role in controlling the resistance response (Dinesh-Kumar and Baker 2000; Zhang and Gassmann 2003).

Alternative splicing of the tobacco *N* resistance gene occurs via the use of a 70 bp alternative exon (AE), located in the third intron of the gene (Dinesh-Kumar and Baker 2000). This splicing event produces a longer transcript (NL), which due to a frame shift caused by the AE, is predicted to encode a truncated protein lacking the majority of the LRR (Dinesh-Kumar and Baker 2000). The NS transcript, in which the AE is excluded, encodes the full length N protein. During the infection process, the relative abundance of NS and NL transcripts changes dramatically. Both cDNAs, when expressed in separate transgenic plants, are insufficient for complete resistance to TMV, demonstrating the importance of the *N* gene splicing in controlling the resistance response (Dinesh-Kumar and Baker 2000).

The *Arabidopsis RPS4* gene generates alternative transcripts through the retention of either the second, or second and third, introns of the gene. Both isoforms are predicted to encode truncated peptides with only the TIR and NBS domains. Zhang and Gassmann (2003) assessed the biological role of these introns by generating transgenic plants with either intronless versions of the gene, or with a gene capable of encoding only the truncated protein. Both transgenic plants could not confer functional disease resistance indicating that both the processed and alternative processed, transcripts are required.

Alternative splicing of *L6* involves the third intron of the gene and has been shown to generate four transcripts, three of which encode truncated proteins lacking most or all of the LRR (Ayliffe et al. 1999). In contrast to the tobacco *N* and Arabidopsis *RPS4* genes, transgenic flax plants carrying a full-length *L6* cDNA show wild-type resistance. The absence of any truncated L6 products could potentially be compensated by products derived from related R genes in the flax genome (Ayliffe et al. 1999). Analysis of alternative splicing through the rust infection process showed there was no detectable change in the relative abundance of transcripts, although it should be noted that the relative proportion of RNA from rust infected cells may be below a detectable level compared to that of uninfected cells (Ayliffe et al. 1999). It is of interest that *L6* transcripts were found in root tissue, an organ not subject to infection from a rust pathogen.

Alternative splicing is also found in introns located in the 5′ UTR of alleles of the *Mla* locus in barley (Halterman et al. 2003). In this case however, no alternative splicing of introns in protein coding regions of *Mla* genes was detected. In *Mla13*, two alternatively spliced introns were found in the 5′ UTR. The splicing of these introns altered the predicted number and length of upstream micro-open reading frames (μORFs) found in the 5′ UTR, and the profile of intron splicing changed upon pathogen infection. Halterman and Wise (2006) report that the μORFs may function co-operatively to down-regulate translation of the *Mla13* cistron by up to 13-fold. A similar 5′ UTR region containing a μORF is found in the *M* gene, although the existence of introns in the 5′ UTR has not, until now, been investigated.

Collectively, these data indicate that alternative splicing is a common occurrence in R gene transcripts, and that in some instances, alternative transcripts, and the products they may encode, play an essential role in the resistance response. In this report we present a survey of splicing of the *M* flax rust resistance gene pre-mRNA and demonstrate the existence of 12 transcript isoforms derived from this gene. In addition, a μORF located in the 5′UTR of the *M* gene is shown to be removed by the splicing of a leader intron sequence in some transcripts. In other transcripts however, the leader intron is retained. This complex transcriptional profile suggests that multiple levels of post-transcriptional control are imposed upon this gene.

Methods

Plant material

Seed of all flax lines used in this study were kindly provided by G. Lawrence (CSIRO, Canberra). The flax line Forge contains *L6*, *M*, *N*, and *P2* rust resistance genes (Ellis et al. 1992). The flax line Dakota and Williston Brown contain *M* and *M1* rust resistance genes respectively (Islam and Mayo 1990). The line Hoshangabad contains the *LH* allele at the *L* rust resistance locus and is susceptible to all known flax rust strains (Mayo and Shepherd 1980). The Hoshangabad + MM line is derived from the self fertilisation of a transgenic Hoshangabad plant that contains an *M* transgene (Anderson et al. 1997). All plants were grown under glasshouse conditions at 20°C.

RNA extraction

Total RNA was isolated from plant material using modifications of the methods described by Chomczynski and Sacchi (1987), Logemann et al. (1987) and Noonberg et al. (1995). Tissue (0.1–1 g) was frozen and ground in liquid nitrogen. Extraction buffer was added to the frozen powder (1 ml per 100 mg tissue) and the mixture rapidly thawed. The mixture was homogenised by passing three times through a 22 gauge needle using a 3 ml syringe. An equal volume of phenol/chloroform (5:1 pH 4.7) was added and briefly mixed by inverting. Phases were separated by centrifugation and the aqueous phase subjected to two further phenol/chloroform extractions, followed by a single chloroform/isoamylalcohol (24:1) extraction. The aqueous phase was precipitated using 0.1 volumes of sodium acetate pH 5.5 and 0.7 volumes of ethanol. RNA was collected by centrifugation and the pellets washed in 70% ethanol.

5′ and 3′ RACE

Thermoscript™ (Invitrogen) reverse transcriptase was used in both the 5′ and 3′ RACE experiments. 5′ RACE reactions were performed according to the manufacturer's recommendations using 500 ng of total RNA with 100 ng of R4 primer (Table 1) and 1 mM dNTPs. 5′ RACE cDNA products were G-tailed using terminal transferase (New England Biolabs). Two rounds of PCR were used to amplify the 5′ RACE products. The first round of amplification used a reverse primer (R5), nested 15bp 5′ of the R4 primer and the second PCR used a primer (R6) a further 19 bp 5′ (Table 1). An oligo-C primer was used in both first and second round PCR.

3′RACE experiments were conducted using components supplied by Invitrogen (3′ RACE kit). RT reactions were performed as described above, except in this case reverse transcription was primed from an oligo-dT primer with an additional 5′ adapter sequence (Table 1). *M* gene specific transcripts were amplified by PCR using the F4 primer and the adapter sequence (UAP Table 1) used for reverse priming.

5′ and 3′ RACE products were cloned into the pGEM-T Easy vector (Promega), transformed into *E. coli* strain DH10β and individual colonies were selected for plasmid DNA purification and sequence analysis.

Table 1 Primer designation and sequence of all primers used in the 5′, 3′ and RT-PCR analysis of the *M* gene pre-mRNA

Primer designation	Primer sequence
F1	GAAAAACCAGCAAAGCACAA
F2	GCTTTCAATTTCTATCATTTCTGCT
F3	CATACGAAGCTGGAAGAATGC
F4	GATGATACATCTTCTTCC
R1	CACCACACGCTACCTTCATC
R2	CAGCCTCGGAGCCATCTTAATC
R3	CCCCCACTTATCTTTTGTATT
R4	TTCGACTGGTCGATTGCTCTG
R5	TGCTCTGAGGAGGTTGACCTT
R6	CCTTTATCTCTTCTCCCTTAT
Oligo-dT adapter	GGCCACGCGTCGACTAGTAC $(T)_{17}$
UAP	$(CUA)_4$ GGCCACGCGTCGACTAGTAC
Oligo-dC	CCCCCCCCCCCCCCC

RT-PCR

Initial RT-PCR experiments were conducted using primers that annealed either at positions flanking intron 3, or at 5′ and 3′ positions immediately adjacent to the coding region. In the former of these experiments, all transcripts displayed invariant use of intron 3 splice sites (Ayliffe data not shown). In the later experiment, all transcripts displayed invariant use of intron 3 splice sites and a subset of those intron 1 and 2 splice variants listed in Fig. 1 (Schmidt, data not shown).

All subsequent RT reactions were performed using the R1 primer located in exon 4 of the *M* gene. Reverse transcription used either Thermoscript, or Superscript II RNaseH-Reverse Transcriptase, in which 250 ng of total RNA was mixed with 150 ng RT primer R1 (Fig. 1). The RNA/primer mix was incubated at 65°C for 1 min and placed immediately on ice. This mixture was incubated for 1 min at 40°C prior to the addition of reverse transcriptase. All RT reactions were carried out according to the manufacturer's specifications, and compared with control reactions in which the reverse transcriptase was omitted. Reactions were either carried out at 42 or 65°C depending on the thermal stability of the reverse transcriptase.

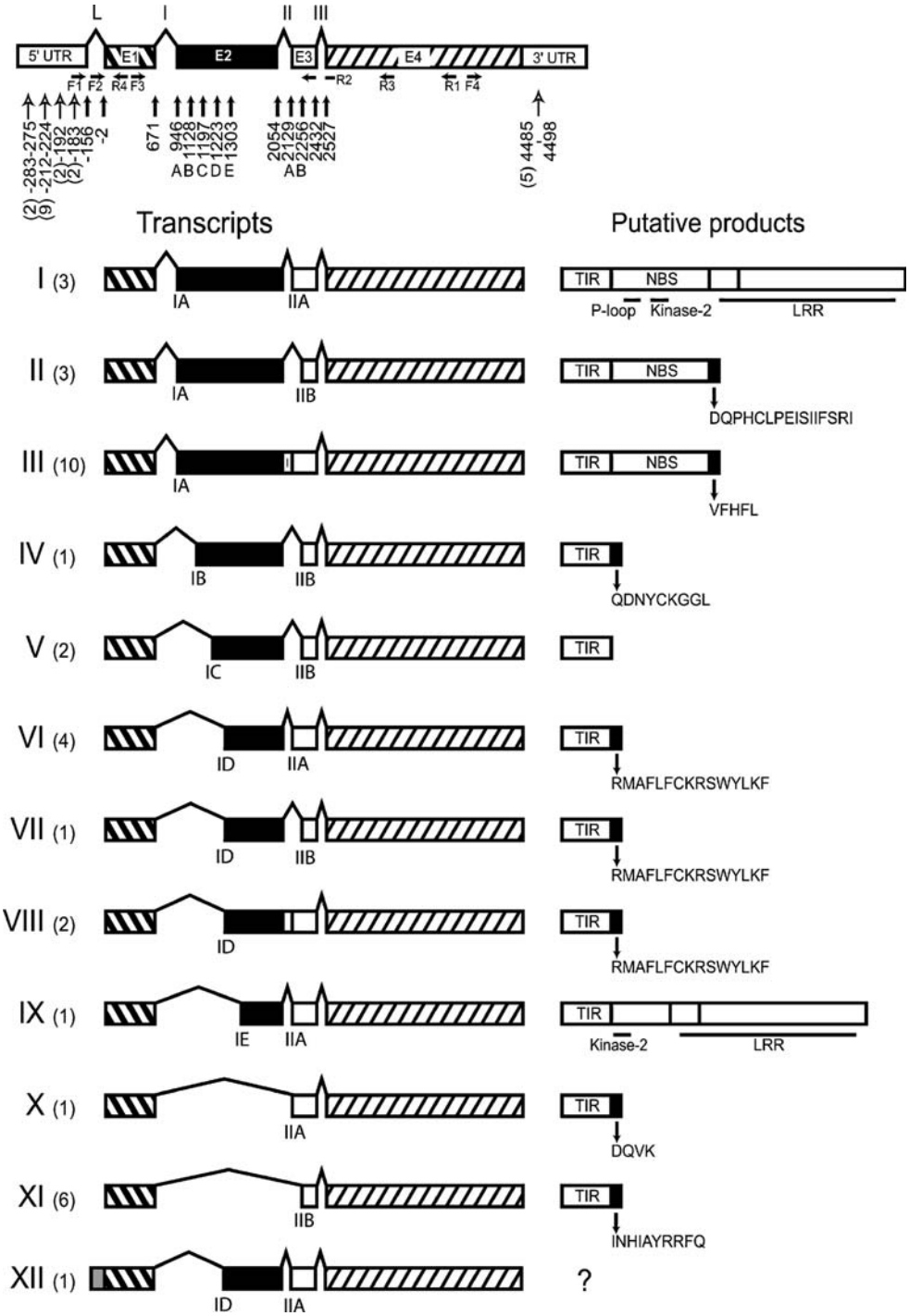

Fig. 1 The transcript and predicted products derived from the *M* flax rust resistance gene. RT-PCR products derived from Forge cDNA were obtained from two experiments using either F3 and R2, or F1 and R3 primers in the PCR step. Analysis of leader intron retained transcripts was conducted using the F2 and R3 primers. The sequence of each transcript type is submitted as a supplementary file. Splice sites are indicated by *vertical closed arrows*. The five splice acceptor sites at intron 1, and the two splice acceptor sites at intron 2, are labelled IA, IB, IC, ID and IE, and IIA and IIB respectively. The splice sites are labelled below each transcript isoform. All intron 1 and intron 2 splice acceptor sites are used in conjunction with a single intron 1 or intron 2 splice donor site. Variation of splice acceptor and donor sites at intron 3 has not been detected. Three transcripts (indicated by *grey boxes*) were identified in which either the leader or intron 2 were retained. 5′ and 3′ RACE analysis can predict the likely start of transcription and polyadenylation, respectively. Potential transcription start and stop sites are indicated by vertical open arrows in the 5′ and 3′ UTRs, respectively. The number of independent clones displaying the locations is shown in parentheses. Approximate locations of primers used in the RT-PCR and RACE experiments are shown by *closed arrows*. The predicted products of each splice variant are shown on the right. *Closed boxes* represent unique amino acid sequences not found in the predicted full length M peptide. The existence of these products is yet to be demonstrated, and in the case of leader intron retention in transcript *XII*, we can only predict that translation may re-commence downstream at other start codons

For subsequent PCR amplification, the reverse primer R2 that flanks intron 3 was used in conjunction with the forward primer F3. Following the discovery of the leader intron, RT-PCR analysis involved the use of the F1 and F2 primers in combination with the R3 primer. The data presented in Fig. 1 was derived from the analysis of multiple clones obtained from these two RT-PCR experiments. For the second of these RT-PCR experiments, Forge total RNA was treated with TURBO RNase-*free*™ DNase (Ambion) according to the manufacturer's recommendations. All cloned and sequenced RT-PCR products were derived from Forge cDNA. Sequence files lodged as supplementary material represent all transcript isoforms derived from RT-PCR using F3 and R2, except for transcript *XII* that was derived from RT-PCR using primers F2 and R3. PCR was performed with either Platinum Pfx DNA polymerase (Gibco BRL) or Phusion Taq (Finnzymes) and reactions were A-tailed by the addition of dATP by Taq polymerase (New England Biolabs) prior to ligation of products into the pGEM-T Easy vector (Promega) and transformation of DH10B *E. coli* cells. Plasmid DNA extracted from white colonies was digested with *Eco*RV, a unique site for which exists in intron 2. This discriminated between different splice isoforms and the identity of the cloned RT-PCR products to the *M* gene was then confirmed by sequence analysis.

Southern analysis

RT-PCR products were separated by acrylamide gel electrophoresis and transferred to Hybond N^+ nylon membrane (Amersham) by standard methods. A deoxygenin labelled *M* probe was generated by PCR, and DNA blot hybridisation and washing was done as recommended by the manufacturer (Amersham).

Results

We have conducted a survey of the *M* flax rust resistance gene mRNAs using both 5′ and 3′ RACE and RT-PCR. 5′ RACE cDNA was generated using the R4 primer located in exon 1 (Fig. 1) and then G-tailed prior to PCR with a forward oligo-dC primer and 5′ nested reverse primers, R5 and R6, located in exon 1. 3′ RACE cDNA was generated using an oligo-dT primer with a unique 5′ sequence. Subsequent PCR amplification used a forward primer (F4) located in exon 4 of the *M* gene, and a reverse primer containing the unique 5′ sequence of the oligo-dT primer. RT-PCR amplification of *M* cDNA used reverse primers located in exon 4 and forward primers in exon 1 and the 5′ UTR. Independent RT-PCR clones were separated into 12 classes based on their *Eco*RV digestion profile. The positions of the primers, the RT-PCR products that were amplified, the number of independent clones that were sequenced, and the putative protein products derived from these transcripts are summarised in Fig. 1.

From these experiments, 12 splice isoforms derived from the *M* flax rust resistance gene were identified (designated *I-XII* Fig. 1). The 5′ RACE analysis identified an additional intron to that previously reported (Anderson et al. 1997).

In total, five splicing events were identified at intron 1, which made invariant use of the splice donor site and five different splice acceptor sites. Three splicing events were observed at intron 2. In addition, two exon-skipping events were observed which used the intron 1 splice donor site, and either of the two intron 2 splice acceptor sites. No variation of splicing was detected at intron 3 (this study and previous work of Schmidt and Ayliffe, data not shown). This is in stark contrast to the related flax rust resistance gene, *L6*, which displayed four splice isoforms at intron 3 (Ayliffe et al. 1999). In the study of the *L6* gene, alternative splicing at introns 1 and 2 was not investigated.

Given that the *M* gene is a member of a highly conserved gene family, and also highly homologous to the *L* locus, it was essential to demonstrate that all these transcript isoforms are *M* gene products. All RT-PCR clones that were sequenced showed no deviation from the *M* gene sequence, indicating all 12-splice variants were derived from *M*. To further confirm this origin, RT-PCR analyses were undertaken on transgenic Hoshangabad containing an *M* transgene (Anderson et al. 1997). The resultant amplification products were compared with those derived from RNA extracted from a non-transgenic Hoshangabad plant. RT-PCR analysis was also conducted on total RNA extracted from other flax lines, including Williston Brown (*M1*), Dakota (*M*), and Forge (*L6*, *M*, *N*, *P2*). These RT-PCR amplification products were analysed by DNA blot using an *M* gene-specific probe (Fig. 2). Amplification products homologous to the *M* gene probe were obtained only when total RNA extracted from flax tissue containing the *M* gene was used as a template for RT-PCR.

M gene splice isoforms

Transcript *I* is predicted to encode a protein containing the complete TIR, NBS and LRR regions of M (Fig. 1). Two transcripts, *II* and *III*, are predicted to encode TIR-NBS proteins lacking the LRR, but include a unique sixteen and five amino acid C-termini, respectively. Alternative splicing of intron 1 of *M* using splice acceptor sites 1B, 1C, 1D and both exon 2 skipping events, results in five transcripts encoding proteins with only the TIR domain. These are homologous with *M* to the point of intron removal, and differ only in their 3′ ends which, as the result of a frameshift, are

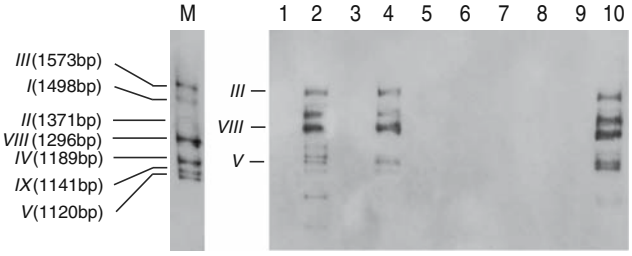

Fig. 2 RT-PCR products separated by 5% acrylamide gel electrophoresis and blotted and probed with a deoxygenin labelled *M*-probe. The probe was generated using a forward primer (GATCAACCACATTG CTTAC) and a reverse primer (CCACTCTGTTTTGCTATTG) both located in exon 3. Following the RT reaction, PCR was conducted with the F3 and R2 primers (Fig. 1). *Lanes M* A mixture of PCR products amplified from cloned splice isoforms of the *M* gene. *1* Forge total RNA − RT, *2* Forge total RNA + RT, *3* Dakota total RNA − RT, *4* Dakota total RNA + RT, *5* Williston Brown total RNA − RT, *6* Williston Brown total RNA + RT, *7* Hoshangabad total RNA − RT, *8* Hoshangabad total RNA + RT, *9*: Hoshangabad + MM − RT, *10* Hoshangabad + MM + RT

predicted to encode between zero and 15 unique amino acid extensions that do not appear in the mature full-length M protein.

A comparison between the eukaryotic consensus splice donor and acceptor sites and that of each *M* transcript variant is shown in Table 2. All the splice sites conform to the dinucleotide acceptor and donor splice site consensus, apart from the donor site of the leader intron and the acceptor site of the 1C transcript. The leader intron uses a rare GC splice donor site although the remainder of the surrounding sequence adheres well to the splice donor site consensus sequence. In 99% of *Arabidopsis* splice donors sites, GU is used where the most common alternative to this is GC (Brown et al. 1996). A rare UG splice acceptor site is used in the 1C transcript, although the adjacent sequence also adheres well to the consensus sequence.

Several transcripts (*III*, *VIII*, and *XII*) were identified in which an intron was retained (Fig. 1). Intron 2 was retained in transcripts *III* and *VIII*, and the leader intron was retained in transcript *XII*. In addition, 2 of the 15 5′ RACE clones showed leader intron retention.

Transcriptional leader region

From 5′ RACE analysis, a previously unidentified intron within the 5′ untranslated region of the *M* gene was identified. Fifteen independent clones were sequenced, all of which contained a G-tail indicating that the clones were derived from cDNA. Two of the clones contained the leader intron sequence and are thus examples of transcripts in which the leader intron is retained. The remainder showed invariant acceptor and donor splice site use in the splicing of the leader intron and varying lengths of 5′ sequence

Table 2 Comparison of *M* splice sites to dicotyledonous splice site consensus

Donor sites	Exon				Intron					
	−3	−2	−1	↓	+1	+2	+3	+4	+5	+6
Consensus		**A_{62}**	**G_{79}**	↓	**G_{100}**	**U_{99}**	**A_{70}**	**A_{58}**	**G_{49}**	**U_{53}**
L Intron	C	A	G	↓	G	C	A	A	G	A
Intron 1	C	G	A	↓	G	U	A	C	G	U
Intron 2	A	A	G	↓	G	U	A	U	U	U
Intron 3	A	A	G	↓	G	U	G	U	G	U
Acceptor sites	Intron							Exon		
	−6	−5	−4	−3	−2	−1	↓	+1	+2	+3
Consensus		**U_{64}**	**G_{42}**	**Y_{95}**	**A_{100}**	**G_{100}**	↓	**G_{57}**	**U_{44}**	
L Intron	U	U	G	A	A	G	↓	G	A	A
Intron 1										
A	U	C	G	U	A	G	↓	G	C	A
B	G	A	A	U	A	G	↓	G	C	A
C	U	C	G	U	U	G	↓	A	U	A
D	G	A	G	C	A	G	↓	A	A	G
E	U	G	A	U	A	G	↓	U	G	G
Intron 2										
A	U	U	A	C	A	G	↓	G	C	A
B	U	U	G	U	A	G	↓	G	A	U
Intron 3	A	C	A	U	A	G	↓	A	T	A

Subscripts following consensus residues denote the percentage that a particular residue is observed. Arrows indicate the intron/exon borders. L = leader intron, Y = T or C. The consensus sequence and percentages are sourced from Lorkovic et al. (2000) and Simpson and Filipowicz (1996)

homologous to the *M* gene (Fig. 1). Of greatest significance is that upon splicing, an upstream μORF in the UTR of the *M* gene is removed (Fig. 3). This μORF, which is in a different reading frame to the *M* ORF, is defined by an initiation codon located 69 bp 5′ to the initiation codon of the *M* protein ORF. Furthermore, the μORF terminates within the predicted initiation codon of the M protein ORF. This leader sequence intron therefore has the capacity to regulate *M* gene expression by alternative splicing. 5′ RACE analysis also revealed several potential transcription start points for the *M* gene, with the relative abundance of each 5′ RACE product indicated in parentheses in Fig. 1.

Given the potential regulatory role of the leader intron, examples of leader intron retention were of interest to us. In order to assess the complexity of transcripts in which the leader intron is retained, we conducted RT-PCR experiments using a forward primer nested within the leader sequence intron (F2 primer, Fig. 1). Cloning and sequence analysis of the RT-PCR products obtained from this reaction identified clones identical to *M*. One transcript was found in which introns 1–3 were spliced and yet the leader intron was retained (transcript *XII*, Fig. 1).

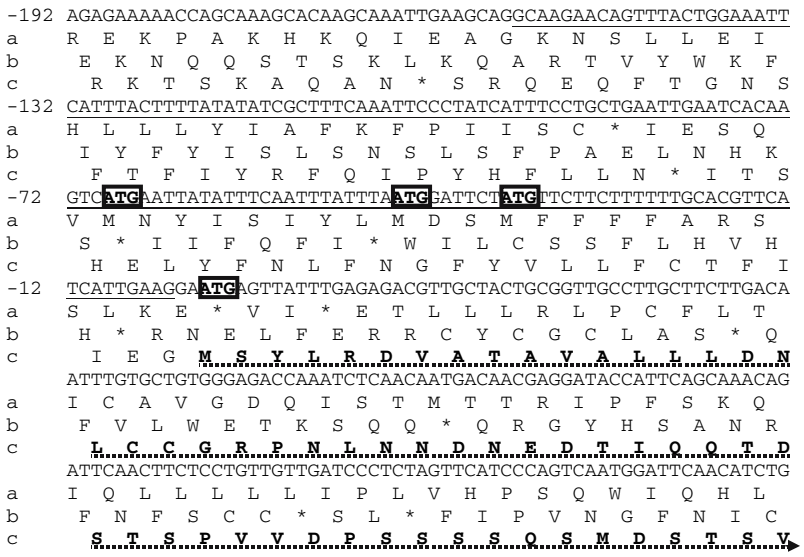

Fig. 3 The *M* gene 5′ UTR and first 170 bases of the main open reading frame of *M* (Anderson et al. 1997). The main ORF of *M* is shown by a *dotted arrow*. The leader intron is *underlined* and potential start codons are *boxed*. In the 5′ UTR is an ORF that terminates at the TGA stop codon partially imbedding in the initiation codon of the main ORF

3′ RACE analysis

Five 3′ RACE products were cloned and sequenced. All the products shared 100% sequence homology to the *M* gene which terminated between 123 and 137 bp 3′ of the predicted translation termination codon of the *M* ORF. 5′ of the polyadenylated tail is a poly U stretch that is consistent with the polyadenylation signal sequence in eukaryotes (Li and Hunt 1997). A consensus polyadenylation signal (AUAAA) however, could not be identified, and thus polyadenylation of the *M* transcripts reported here must occur via a non-consensus signal. Because of the use of the F4 primer located in exon 4, we cannot rule out the existence of additional polyadenylation sites further 5′ in the *M* gene. Prematurely polyadenylated transcripts have been reported in other genes (Cui et al. 2003; Lou et al. 1998), and in other R genes (Ayliffe et al. 2004).

Discussion

In general, alternative spliced isoforms can be generated by a variety of mechanisms including multiple splice donor and acceptor sites, intron retention, alternate exon, and exon skipping events. The *M* gene transcript isoforms (Fig. 1) are generated by multiple splice acceptor sites, intron retention and exon skipping. Neither alternate exon, nor alternate splice donor site use is observed.

As the *M* gene is a member of a gene family in flax, of which many members display sequence similarity (Ellis et al. 1995), it was essential to demonstrate that the transcripts identified in this study were derived from the *M* gene, and not from a related R gene. This was confirmed in two ways. First, RT-PCR products were sequenced and shown to be identical to the sequence of the *M* gene. Second, DNA blot analysis revealed that no equivalent RT-PCR products were generated from RNA extracted from flax lines that did not contain the *M* gene. We can be confident therefore that these amplification products are derived from *M* gene transcripts.

Non-canonical splice sites (eukaryotic consensus)

Each of the donor and acceptor splice sites used for processing of *M* transcripts were compared to plant consensus splice sites. While the plant splice site consensus sequences are loose, in all but one case the observed *M* gene splice sites were consistent with these consensus sequences. Only the U present in position −2 at splice acceptor site 1C of the *M* gene does not strictly conform to the splice consensus sequence. The existence of *Arabidopsis* genes with non-consensus splice sites, including several with a UG acceptor site, is reported (http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis_nonconsensus_splice_sites.shtml). The use of a GC donor site in the leader intron, although rare, represents the most frequent alternative to the GU donor splice site, being found in 1% of mammalian splice sites (Chong et al. 2004). The remainder of the sequence adjacent to the acceptor and donor dinucleotide splice sites of *M* adhere well to the consensus and therefore the use of such an array of splice acceptor sites may represent some flexibility in the binding of the splicesome complex. These findings suggest a lack of co-operative splicing as was observed by Simpson et al. (1999).

The role, if any, of alternative spliced transcripts

In general, the position of introns is conserved between R genes of the TIR–NBS–LRR class, and in cases where alternative transcripts are reported they are predicted to

encode truncated products of either TIR–NBS or TIR domains. In the case of the tobacco *N* gene and the *Arabidopsis RPS4* gene, the alternative transcripts, and most probably the products they encode, are required for full disease resistance. Furthermore, outside the plant kingdom, similarities are reported between the splicing of Toll-like receptor genes involved in animal innate immunity and that of TIR-containing R genes in plants (Jordan et al. 2002). It is likely, therefore, that a system controlling the fine balance in transcripts encoding full-length and truncated products governs the provision of *M* gene-mediated rust resistance in flax.

We acknowledge, however, that a transcript surveillance system exists in eukaryotic cells that enable incomplete or incorrectly spliced transcripts to be identified and degraded. This process is called nonsense mediated decay (NMD) (reviewed in Conti and Izaurralde 2005; Lejeune and Maquat 2005). In short, this process involves a pioneer round of translation whereby if a premature termination codon (PTC) is identified this transcript is targeted for degradation. NMD is conserved among eukaryotes; however, different mechanisms exist in different organisms. In mammals, a PTC located >50–55 nt from an exon–exon boundary predestines this transcript for destruction, although a recent report indicates that the role of NMD in regulating transcript levels may not be as widespread as once thought (Pan et al. 2006). Although a NMD pathway is not clearly defined in plants, several lines of evidence point to its existence. PTC-containing mRNAs of the *waxy* gene in rice show reduced stability (Isshiki et al. 2001). In addition, an *Arabidopsis* mutant of the UPF3 homologue, a factor required for NMD in mammals and yeast, is less able to degrade mRNAs with PTCs than wild-type plants (Hori and Watanabe 2005). Wang and Brendel (2006) found that >98% of *Arabidopsis* genes contain stop codons within 50 bp or downstream of the last exon–exon boundary, and that of the alternative spliced isoforms that are generated, 43% produces candidate targets for NMD.

If a NMD pathway were to operate on the profile of *M* gene transcripts, we would predict that all except transcripts *I* and *IX* would be subject to NMD. The products of transcripts *I* and *IX* both contain the TIR and LRR regions; transcript *IX*, however, would encode a protein without all the functional domains of the nucleotide binding site. Transcript *I* is the only one capable of encoding a full-length M protein.

It is still unclear what role NMD may play in plants. Regardless of this, the complex array of splicing isoforms generated by the M gene still presents a curiosity. Why would such an array of transcripts be produced, only for the majority of them to be targeted for destruction? One possibility is that the signals directing splicing of the *M* gene are weak and the aberrant transcripts are filtered out by NMD. Alternatively, it is possible that the shuffling of alternative splice sites within the *M* pre-mRNA provides only a means to down-regulate levels of transcript *I*. If NMD is not operating on these transcripts, then what role do the numerous truncated TIR and TIR–NBS proteins play? Both the TIR and TIR–NBS truncations, and/or, the transcripts that encode them, may act as negative regulators of the disease resistance signalling pathway. This is consistent with the aberrant or even lethal phenotype associated with the overexpression of some R genes (Mindrinos et al. 1994; Oldroyd and Staskawicz 1998; Tang et al. 1999; Tao et al. 2000; Grant et al. 2003), and also with the down regulation of *Mla*13 mediated through leader intron alternative splicing (Halterman and Wise 2006). A mechanism of down regulation does, however, not explain why intronless versions of *N* and *RPS4* were not capable of conferring complete disease resistance or why, in the case of *L6*, an intronless gene was capable of providing full resistance (Dinesh-Kumar and Baker 2000; Zhang and Gassmann 2003; Ayliffe et al. 1999). It is clear that further study of the stability and function of alternative R gene transcripts is required, as well as the identification of the products they encode.

Role of the leader intron

The intron in the 5′ UTR of the *M* gene removes a μORF that terminates coincident with the predicted translation start codon of the *M* gene ORF. *M* gene transcripts do exist, however, where the leader intron is retained. The rules governing translation initiation in eukaryotes (Kozak 1999) leads us to predict that a transcript containing the leader intron would commence translation at the defining μORF start codon at position −69 (Fig. 3). The presence of two additional start codons within the μORF would almost certainly preclude translation initiation at the main ORF via leaky scanning. Termination of the μORF would occur within the start codon (AUGA) of the M protein ORF. According to the known constraints on translation re-initiation (Kozak 2001), termination in such a position would severely limit re-initiation at the first AUG of the M protein ORF. However, translation could re-commence at either an out of frame start codon (79 bp 3′), or further 3′ (93 bp) at an in frame start codon, with an efficiency approaching that of the first start codon (position −69) (Kozak 2001).

Although it is plausible that transcripts in which the leader intron is retained may represent partially spliced isoforms, it is noted that transcription and splicing are almost coincident in eukaryotic cells (Bently 1999). In addition, a number of examples exist where intron-retained transcripts are expressed and encode functionally important products (Black 2003), or are selectively edited and retained in the nucleus (Prasanth et al. 2005). Regardless, the leader intron

has the potential to impart a significant regulatory effect on *M* gene expression. In transcripts where the leader intron is retained, we predict that the ribosomal complex after translating a small peptide from the μORF, would bypass translation initiation of the full-length M ORF, and could re-initiate further downstream. This would have significant consequences on the spectrum of *M* gene products that are generated from these transcripts. It is worth noting that all flax rust resistance alleles at the *L* locus contain a μORF, 5′ of the full-length ORF. It is not known however, whether these μORFs are removed by splicing of the 5′ UTR.

The potential regulatory role of the *M* gene μORF is revealed when analysing the N-terminal region of the M protein. The first 22 amino acids translated from the M protein ORF, are predicted to encode a signal peptide (http://www.cbs.dtu.dk/services/SignalP/). Despite this, Catanzariti et al. (2006) demonstrate that AvrM recognition occurs in the cytoplasm, and recently we have demonstrated that the first 52 amino acids of the M protein direct cytosolic plasma membrane localisation of an M:GUS fusion protein (Anderson, unpublished data). Therefore, protein translated from a downstream in frame initiation codon would not contain this N-terminal trafficking information and therefore could encode a TIR–NBS–LRR protein with a different cellular localisation to that of full-length M protein.

Although we present here the existence of an array of splicing isoforms of the *M* gene, we have not yet thoroughly investigated the relative abundance of each isoform and how this may change in different plant tissues and under different conditions. Such a quantitative RT-PCR study of the *M* gene transcripts is clearly warranted given the change in *N* gene splicing associated with TMV infection. One obvious question would be to ask if the profile of *M* gene transcripts changes upon activation of the resistance response. The cloning of the *AvrM* gene from the flax rust pathogen (Catanzariti et al. 2006) now enables the analysis of *M* gene splicing in flax tissue transiently infected with *Agrobacterium tumefaciens* containing *AvrM*. Given the existence of *L6* transcripts in all flax tissues analysed, including those not subject to rust infection, it would also be of interest to survey *M* gene splicing profiles in a range of flax tissue.

Acknowledgments We wish to acknowledge the input of the referees in improvements made to the manuscript prior to publication.

Reference

Anderson PA, Lawrence GJ, Morrish BC, Ayliffe MA, Finnegan EJ, Ellis JG (1997) Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the Leucine-Rich repeat coding region. Plant Cell 9:641–651

Ayliffe MA, Frost DV, Finnegan EJ, Lawrence GJ, Anderson PA, Ellis JG (1999) Analysis of alternative transcripts of the flax L6 rust resistance gene. Plant J 17:287–292

Ayliffe MA Steinau M, Park RF, Rooke L, Pacheco MG, Hulbert SH, Trick HN, Pryor AJ (2004) Aberrant mRNA processing of the maize Rp1-D rust resistance gene in wheat and barley. Mol Plant Microbe Interact 17(8):853–864

Bently D (1999) Coupling RNA polymerase II transcription with pre-mRNA processing. Curr Opin Cell Biol 11:347–351

Black DL (2003) Mechanisms of alternative pre-messanger RNA splicing. Annu Rev Biochem 72:291–336

Brown JWS, Simpson CG (1998) Splice site selection in plant pre-mRNA splicing. Annu Rev Plant Physiol Plant Mol Biol 49:77–95

Brown JW, Smith SP, Simpson CG (1996) *Arabidopsis* consensus intron sequences. Plant Mol Biol 32(3):531–535

Catanzariti AM, Dodds PN, Lawrence GJ, Ayliffe MA, Ellis JG (2006) Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence. Plant Cell 18:243–256

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyantate-phenol-chloroform extraction. Ann Appl Biol 162:156–159

Chong A, Zhang G, Bajic VB (2004) Information for the coordinates of exons (ICE): a human splice site database. Genomics 84:762–766

Conti E, Izaurralde E (2005) Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. Curr Opin Cell Biol 17:316–325

Cui XQ, Hsia AP, Liu F, Ashlock DA, Wise RP, Schnable PS (2003) Alternative transcription initiation sites and polyadenylation sites are recruited during Mu suppression at the rf2a locus of maize. Genetics 163(2):685–698

Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. Nature 411:826–833

Dodds PN, Lawrence GJ, Ellis JG (2001a) Contrasting modes of evolution acting on the complex N locus for rust resistance in flax. Plant J 27:439–453

Dodds PN, Lawrence GJ, Ellis JG (2001b) Six amino acid changes confined to the leucine-rich repeat beta strand/beta-turn motif determine the difference between the P and P2 rust resistance specificities in flax. Plant Cell 13:163–178

Dinesh-Kumar SP, Baker BJ (2000) Alternatively spliced *N* resistance gene transcripts: their possible role in tobacco mosaic virus resistance. Proc Natl Acad Sci USA 97:1908–1913

Ellis JG, Lawrence GJ, Finnegan EJ (1992) Developing a transposon tagging system to isolate rust resistance genes from flax. Theor Appl Genet 85:46–54

Ellis JG, Lawrence GJ, Finnegan EJ, Anderson PA (1995) Contrasting the complexity of two rust resistance loci in flax. Proc Natl Acad Sci USA 92:4185–4188

Ellis JG, Lawrence GJ, Luck JE, Dodds PN (1999) Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. Plant Cell 11:495–506

Ferrier-Cana E, Macadré C, Sévignac M, David P, Langin T, Geffroy V (2005) Distinct post-transcriptional modification result into seven alternative transcripts of the CC-NBS-LRR gene *JA1tr* of *Phaseolus vulgaris*. Theor Appl Genet 110:895–905

Gassmann W, Hinsch ME, Staskawicz BJ (1999) The *Arabidopsis RPS4* bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. Plant J 20:265–277

Glazebrook J (2005) Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Ann Rev Phytopath 43:205–227

Grant JJ, Chini A, Basu D, Loake GJ (2003) Targeted activation tagging of the *Arabidopsis* NBS-LRR gene, ADR1, Conveys resistance to virulent pathogens. Mol Plant Microbe Interact 16:669–680

Gupta S, Zink D, Korn B, Vingron M, Haas SA (2004) Genome wide identification and classification of alternative splicing based on EST data. Bioinformatics 20:2579–2585

Halterman DA, Wise RP (2006) Upstream open reading frames of the barley Mla13 powdery mildew resistance gene function co-operatively to down-regulate translation. Mol Plant Pathol 7(3):167–176

Halterman D, Wei F, Wise R (2003) Powdery mildew induced *Mla* mRNAs are alternatively spliced and contain multiple upstream open reading frames. Plant Physiol 131:558–567

Holt BF, Belkhadir Y, Dangl JL (2005) Antagonistic control of disease resistance protein stability in the plant immune system. Science 309(5736):929–932

Hori K, Watanabe Y (2005) UPF3 suppresses aberrant spliced mRNA in *Arabidopsis*. Plant J 43:530–540

Islam MR, Mayo GME (1990) A compendium on host genes in flax conferring resistance to flax rust. Plant Breed 104:89–100

Isshiki M, Yamamoto Y, Satoh H, Shimamoto K (2001) Nonsense-mediated decay of mutant waxy mRNA in rice. Plant Physiol 125(3):1388–1395

Jordan T, Schornack S, Lahaye T (2002) Alternative splicing of transcripts encoding Toll-like plant resistance proteins—what's the functional relevance to innate immunity. Trends Plant Sci 7:392–398

Kozak M (1999) Initiation of translation in prokaryotes and eukaryotes. Gene 234(2):187–208

Kozak M (2001) Constraints on reinitiation of translation in mammals. Nucleic Acids Res 29(24):5226–5232

Lawrence GJ, Finnegan EJ, Ayliffe MA, Ellis JG (1995) The L6 gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene RPS2 and the Tobacco viral resistance gene N. Plant Cell 7:1195–1206

Lejeune F, Maquat LE (2005) Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. Curr Opin Cell Biol 17(3):309–15

Logemann J, Schell J, Willmitzer L (1987) Improved methods for the isolation of RNA from plant tissue. Anal Biochem 163:16–20

Li QS, Hunt AG (1997) The polyadenylation of RNA in plants. Plant Physiol 115(2):321–325

Lorkovic ZJ, Wieczorek Kirk DA, Lambermon MHL, Filipowicz W (2000) Pre-mRNA splicing in higher plants. Trends Plant Sci 5:160–167

Lou H, Neugebauer KM, Gagel RF, Berget SM (1998) Regulation of alternative polyadenylation by U1 snRNPs and SRp20. Mol Cell Biol 18(9):4977–4985

Mayo GME, Shepherd KW (1980) Studies of genes controlling specific host-pathogen interactions in flax and its rust. I. Fine structure analysis of the M group in the host. Heredity 44:212–227

Meyers BC, Kozik A, Griego A, Kuang HH, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15(4):809–834

Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes. Curr Opin Plant Biol 8(2):129–134

Mindrinos M, Katagiri F, Yu G, Ausbel FM (1994) The *A. thaliana* disease resistance gene RPS2 encodes a protein containing a nucleotide-binding site and leucine-rich repeats. Cell 78:1089–1099

Noonberg SB, Scott GK, Benz CC (1995) Effect of pH on RNA degradation during guanidinium extraction. Biotechniques 19:731–733

Oldroyd GED, Staskawicz BJ (1998) Genetically engineered broad-spectrum disease resistance in tomato. Proc Natl Acad Sci USA 95(17):10300–10305

Pan Q, Saltzman AL, Kim YK, Misquitta C, Shai O, Maquat LE, Frey BJ, Blencowe BJ (2006) Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. Genes Dev 20:153–158

Parker JE, Coleman MJ, Frost LN, Schmidt R, Van Der Biezen EA, Moores T, Dean C, Daniels MJ, Jones JDG (1997) The *Arabidopsis* downy mildew resistance gene *RPP5* shares similarity to the toll and interleukin-1 receptors with *N* and *L6*. Plant Cell 9:879–894

Prasanth KV, Prasanth SG, Xuan ZY, Hearn S, Freier SM, Bennett CF, Zhang MQ, Spector DL (2005) Regulating gene expression through RNA nuclear retention. Cell 123(2):249–263

Simpson GC, Filipowicz W (1996) Splicing of precursors to messenger RNA in higher plants: mechanism, regulation and sub-nuclear organisation of the spiceosomal machinery. Plant Mol Biol 32:1–41

Simpson CG, Clark GP, Lyon JM, Watters J, McQuade C, Brown JWS (1999) Interactions between introns via exon definition in plant pre- mRNA splicing. Plant J. 18(3):293–302

Staskawicz BJ, Ausubel FM, Baker BJ, Ellis JG, Jones JDG (1995) Molecular genetics of plant disease resistance. Science 268:661–667

Tao Y, Yuan F, Leister RT, Ausubel FM, Katagiri F (2000) Mutational analysis of the *Arabidopsis* nucleotide binding site-leucine-rich repeat resistance gene RPS2. Plant Cell 12:2541–2554

Tang XY, Xie MT, Kim YJ, Zhou JM, Klessig DF, Martin GB (1999) Overexpression of Pto activates defense responses and confers broad resistance. Plant Cell 11(1):15–29

Wang B-B, Brendel V (2006) Genome-wide comparative analysis of alternative splicing in plants. Proc Natl Acad Sci USA 103:7175–7180

Whitham S, Dinesh-Kumar SP, Choi D, Hehl R, Corr C, Baker B (1994) The product of the tobacco mosaic virus resistance gene N—similarity to toll and the interleukin-1 receptor. Cell 78:1101–1115

Zhang XC, Gassmann W (2003) RPS4-Mediated disease reistance requires the combined presence of RPS4 transcripts with full length and truncated open reading frames. Plant Cell 15:2333–2342