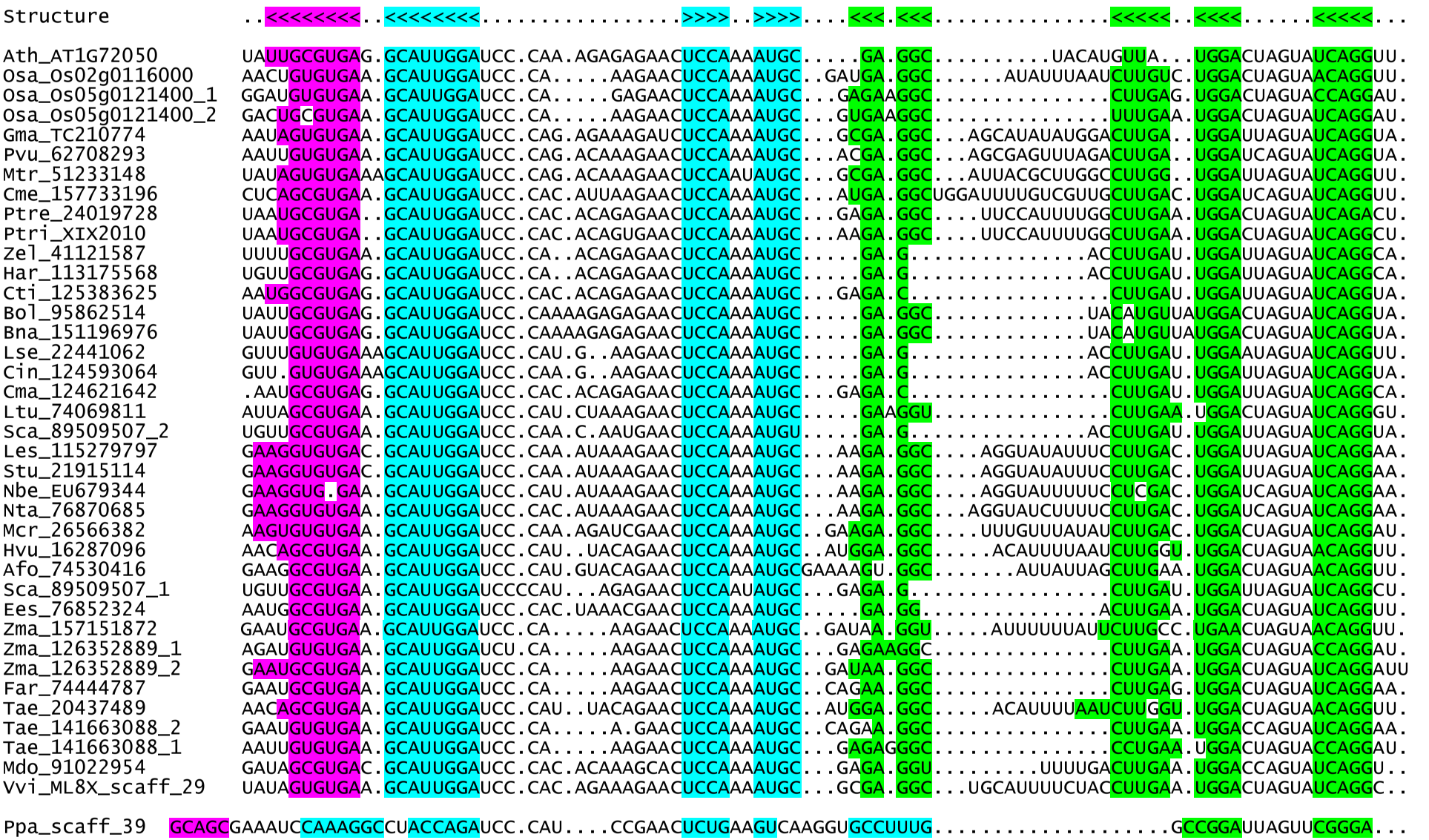
SUPPLEMENTARY INFORMATION

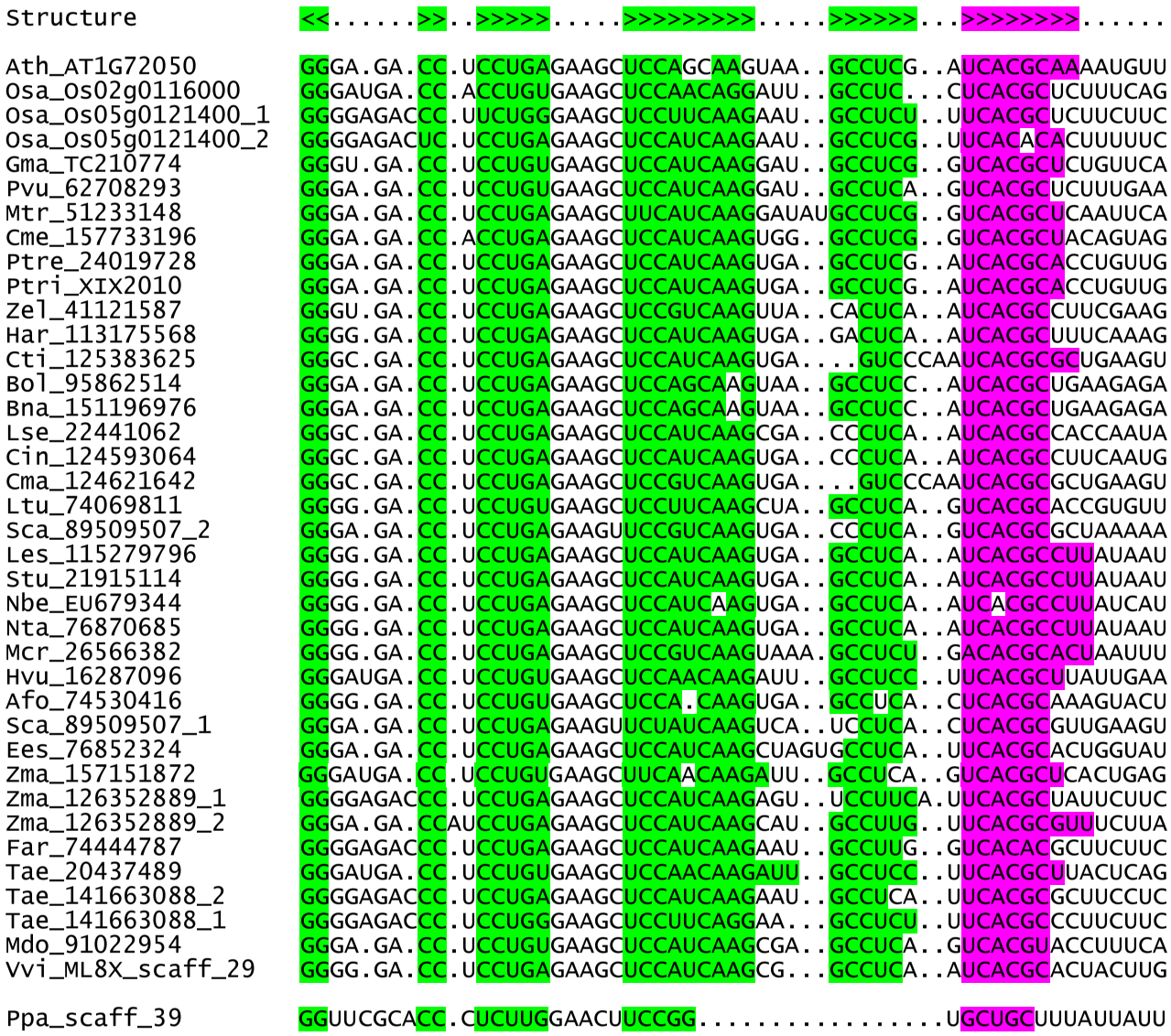
A Plant 5S Ribosomal RNA Mimic Regulates Alternative Splicing of Transcription Factor IIIA Pre-mRNAs

Ming C. Hammond, Andreas Wachter, and Ronald R. Breaker

Table of Contents

Supplementary Figure 1. Representative sequences for plant 5S rRNA mimic





ORGANISM CODE LEGEND:

Afo	*Aquilegia formosa x Aquilegia pubescens*
Ath	***Arabidopsis thaliana***
Bol	*Brassica oleracea*
Bna	*Brassica napus*
Cin	*Cichorium intybus*
Cma	*Centaurea maculosa*
Cme	*Cucumis melo*
Cti	*Carthamus tinctorius*
Ees	*Euphorbia esula*
Far	*Festuca arundinacea*
Gma	*Glycine max*
Har	*Helianthus argophyllus*
Hvu	*Hordeum vulgare*
Les	*Lycopersicon esculentum*
Lse	*Lactuca serriola*
Ltu	*Liriodendron tulipifera*
Mcr	*Mesembryanthemum crystallinum*
Mdo	*Malus domestica*
Mtr	*Medicago truncatula*
Nbe	*Nicotiana benthamiana*
Nta	*Nicotiana tabacum*
Osa	***Oryza sativa***
Ppa	***Physcomitrella patens***
Ptre	*Populus tremula x Populus tremuloides*
Ptri	***Populus trichocarpa***
Pvu	*Phaseolus vulgaris*
Sca	*Senecio cambrensis*
Stu	*Solanum tuberosum*
Tae	*Triticum aestivum*
Zel	*Zinnia elegans*
Zma	*Zea mays*
Vvi	***Vitis vinifera***

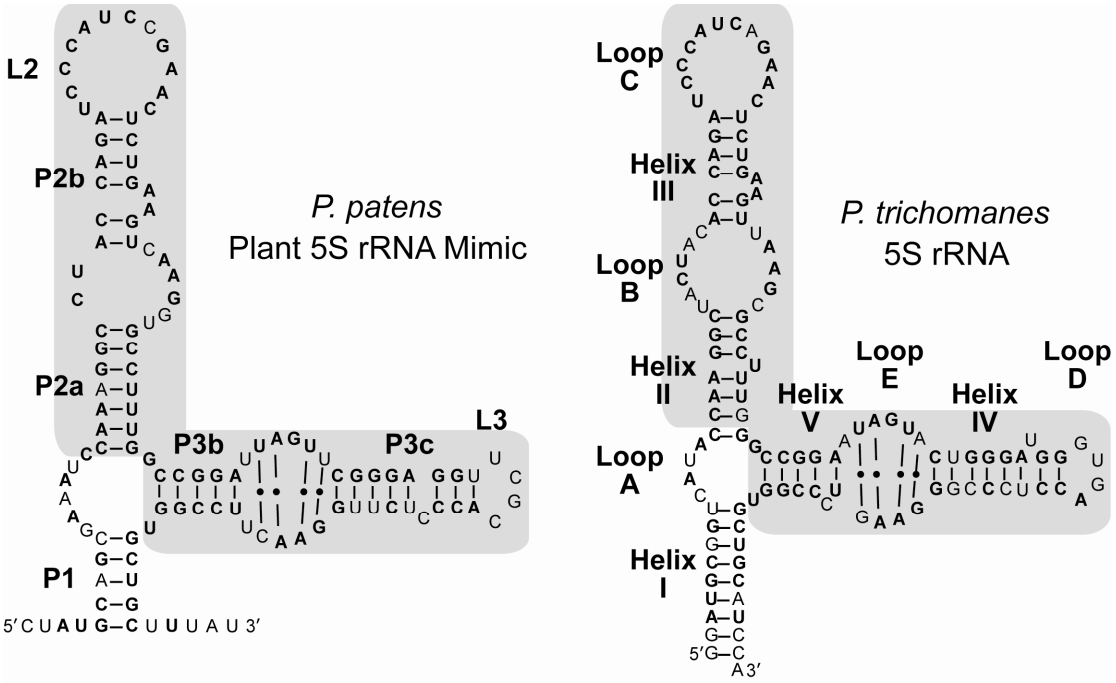
Bold = Genome sequenced

Sequence alignment of plant 5S rRNA mimic representatives from various plant species shows the conservation of sequence and structure. Nucleotides forming pairing interactions P1 through P3 are highlighted in color. Each sequence is annotated with the organism code (see Legend) and the corresponding NCBI gi accession number (for ESTs / cDNAs) or gene locus id (for sequenced genomes). The *Physcomitrella patens* (moss) sequence is displayed on a separate line because it is significantly diverged from the other sequence examples, which are from angiosperms. All sequences except for the example from *P. patens* were used to calculate conservation of nucleotide identity and presence, and covariation or compatible mutations in base-pairing interactions (**Fig. 1a**), as described[1].

We acknowledge the following sources for genomic data: TAIR for *Arabidopsis thaliana*[2], RAP-DB for *Oryza sativa*[3], DoE Joint Genome Institute (JGI) and Poplar Genomic Consortium for *Populus trichocarpa*, Genoscope for *Vitis vinifera*[4], and Cosmoss for *Physcomitrella patens*[5].

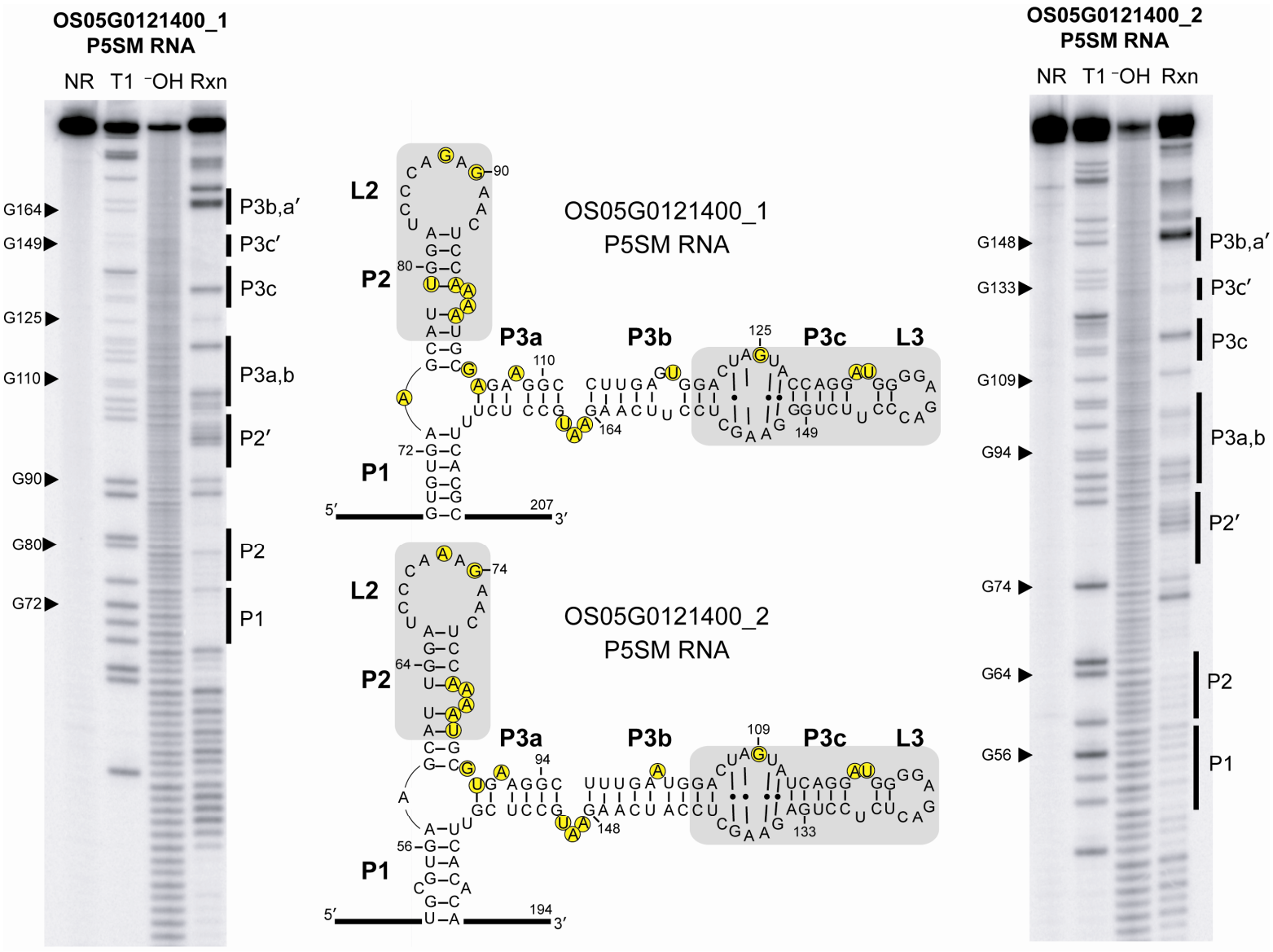
Supplementary Figure 2. Variant P5SM representative in the moss *Physcomitrella patens*

```
CLUSTAL 2.0.3 multiple sequence alignment[6]

Lhe_5S          GGAUGCGGUCAUA-CCAA-GGCUACUACACCAGAUCCCAUCAGAACUCUGCAGUUAAGC-GCCCUUUGGGCCGGAAUAGUACUGGGAUGGGUG--ACCUCCCGGGAAG
Mpo_5S          GGAUGCGGUCAUA-CCAG-GGCUACUACACCAGAUCCCAUCAGAACUCUGCAGUUAAGC-GCC-CUUGGGCCGGAAUAGUACUGGGAUGGGUG--ACCUCCCGGGAAG
Apu_5S          -GGUGCGGUCAUA-CCAG-GGCUACUACACCGGAUCCCAUCAGAACUCCGUAGUUAAGC-GCC-CUUGGGCCGGAUCAGUACUGGGAUGGGUG--ACCUCCCGGGAAG
Plag_5S         GGAUGCGGUCAUA-CCAA-GGCUACUACACCAGAUCCCAUCAGAACUCUGAAGUUAAGC-GCC-UUUGGGCCGGAAUAGUACUGGGAUGGGUG--ACCUCCCGGGAAG
Ppa_P5SM        CUAUGCAGCGAAAUCCAAAGGC--CU--ACCAGAUCCCAUCCGAACUCUGAAGUCAAGGUGCC-UUU-GGCCGGAUUAGUUC-GGGA-GGUUCGCACCCUCUUGGAAC
                   *** *  * * ***  ***  **  *** ********* ****** * *** ***  ***  ** *******  *** * **** ** *   ***  *  ****

Lhe_5S          UCCCGGUGCUGCAUCCA
Mpo_5S          UCCCGGUGCUGCAUCCA
Apu_5S          UCCCGGUGCUGCACCCU
Plag_5S         UCCCGGUGCUGCAUCCA
Ppa_P5SM        UUCCGGUGCUGCUUUAU
                * **********
```



Through tBLASTn searches of the *Physcomitrella patens* genome using the amino acid sequence for zinc fingers 2 and 3 of *At*TFIIIA, we identified a variant of P5SM in a putative TFIIIA gene of the moss *P. patens*. Since the 5S rRNA sequence for this organism is not yet available, we aligned the *P. patens* P5SM sequence to the 5S rRNA sequences for four Bryophyta species: *Marchantia polymorpha* (Mpo, liverwort), *Lophocolea heterophylla* (Lhe, liverwort), *Plagiomnium trichomanes* (Plag, moss), and *Anthoceros punctatus* (Apu, hornwort)[7]. Asterisks indicate conserved nucleotides across the five sequences. The secondary structure model for the *P. patens* P5SM was manually determined based upon comparisons to *P. trichomanes* 5S rRNA (shown) and *A. thaliana* P5SM (**Fig. 1c**). Nucleotides in bold are identical between moss 5S rRNA and P5SM, and were used to calculate the percentage of the 5S rRNA sequence maintained in the P5SM sequence.

Supplementary Figure 3. In-line probing analysis of *Oryza sativa* P5SM RNAs



The structure of two *O. sativa* representatives of P5SM were analyzed by in-line probing ("1" consists of nucleotides 903-1109 and "2" consists of 1500-1693 of OS05G0121400 gene, NCBI gi 115465852). Nucleotides spontaneously cleaved under in-line reaction conditions are circled in the secondary structure models of the rice P5SM elements. Labeled G nucleotides correspond to the same positions labeled in the *A.thaliana* representative of P5SM (**Fig. 1b**).

Supplementary Figure 4. Sequences of TFIIIA gene, protein, and splice products

LEGEND: Primer annealing site, UTR, coding sequence, cassette sequence, stop codon, splice site, zinc finger, cassette protein sequence, **!** = translation termination, XhoI

AT1G72050: NCBI gi 42592260 / NC_003070 region 27118686..27121132

Precursor mRNA sequence | **Protein sequence**

```
   1 GUGCGGCGUC UUGAUGGAGG AGAUAAACCC UAGUUCUUCU GUAGACAAUA
  51 AGAGAGACAU GGCGGAAGAA GCUAAAGUUG AUGUGAAGAC UUCGGCGAAG      MAEEAKVDVKTSAK
 101 AAGGAUAUAC GCAAUUAUCU AUGCCAGUAU UGCGGAAUCA GCAGAUCUAA      KDIRNYLCQYCGISRSK
 151 AAACUAUCUC AUCACUAAAC ACAUCCAAUC UCAUCAUCAG GUUUGAGAUC      NYLITKHIQSHHQ
 201 UUCCUCCAAU UCGAUUCCAA UUUCUCAUCC GUGGCAUCUU GAUUUGUUUU
 251 CAAUGAAAUU GAGAGUUGAG UCUGUAGAAU CGGCGAUGGU UUGUUGAAUU
 301 GAGAGUUUCU AUGAUUCGUU UGUUUAGAUG GAACUUGAAG AGGAAAGAGA      MELEEERD
 351 UGAUGAAGCU UGUGAGGUUG AUGAGGAGUC UUCAAGUAAU CAUACUUGUC      DEACEVDEESSSNHTCQ
 401 AAGAAUGUGG UGCUGAGUUU AAGAAACCUG CUCACUUGAA GCAGCAUAUG      ECGAEFKKPAHLKQHMQ
 451 CAGAGUCAUU CGCUCGAGGU AGAUUUAUGC AUCCUCUUGU CAUGAGAAGU      SHSLE
 501 CGAAUUGUUC CCAUUCUGUG UGUUGCAGCU ACAGAUGGAG AUACAUAGAG
 551 AUACUCGUGG AUUUUGCUUA GUGUUGAGUU UUGUUCUGGU UGUGAACUAA
 601 AAGUUUAUAC AUUUGCAGGA AAUAAAUAGC CUUUUGUUUA AAUCAAAAGG      EINSLLFKSK
 651 UCUUACCUAU GUUAUUGCGU GAGGCAUUGG AUCCCAAAGA GAGAACUCCA      GLTYVIA!
 701 AAAUGCGAGG CUACAUGUUA UGGACUAGUA UCAGGUUGGG AGACCUCCUG
 751 AGAAGCUCCA GCAAGUAAGC CUCGAUCACG CAAAAUGUUU GAGGUCUGAU
 801 GUUCAAUAGC UUGUUUUGUU UCACUUUGCU UUGGACUUUC UUUUCGCCAA
 851 UGAGCUAUGU UUCUGAUGGU UUUCACUCUU UUGGUGUGUA GAGAUCUUUU      RSF
 901 ACUUGCUAUG UGGAUGAUUG UGCUGCUAGC UAUAGGAGGA AGGAUCAUCU      TCYVDDCAASYRRKDHL
 951 CAAUAGGCAU CUUCUUACAC AUAAAGGGAA GCUCUUUAAG UGUCCGAAGG      NRHLLTHKGKLFKCPKE
1001 AGAACUGCAA GAGUGAAUUC UCAGUACAGG GAAAUGUUGG UAGGCAUGUU      NCKSEFSVQGNVGRHVK
1051 AAGAAAUAUC AUAGUAAUGA CAACCGUGAU AAGGACAAUA CUGGUUUGGG      KYHSNDNRDKDNTGLGD
1101 CGAUGGUGAU AAGGACAAUA CUUGUAAGGG GGAUGAUGAU AAGGAAAAAU      GDKDNTCKGDDDKEKSG
1151 CUGGUAGUGG CGGUUGUGAG AAGGAAAAUG AAGGGAAUGG CGGAAGUGGU      SGGCEKENEGNGGSGKD
1201 AAGGACAAUA AUGGUAAUGG CGAUUCUCAG CCUGCGGAGU GUUCAACUGG      NNGNGDSQPAECSTGQK
1251 UCAGAAGCAG GUUGUCUGCA AAGAAAUUGG UUGUGGAAAA GCCUUUAAGU      QVVCKEIGCGKAFKYPS
1301 AUCCUUCACA GCUUCAAAAG CAUCAGGAUU CUCAUGGUAA GUGCACCUUC      QLQKHQDSH
1351 CUACCCUUAC UUUCCCUCUA GUUUAGUAUC CUGGGCAUAU GAAGAUUUCU
1401 ACGUUUCCUC UCUAUGUGCU UUGUUAUAAA UUAAAGACAG UUGUUUGUUA
1451 AAGCUUGAUA GAUUUUCAAU CUCUGAAGGU UUAGAUUUAC AUUUGCAGUG      V
1501 AAAUUAGACU CUGUGGAGGC AUUUUGUUCC GAGCCUGGGU GUAUGAAGUA      KLDSVEAFCSEPGCMKY
1551 CUUUACCAAC GAAGAAUGCC UCAAGUCACA CAUAAGAUCC UGUCAUCAGC      FTNEECLKSHIRSCHQH
1601 ACAUCAACUG UGAGAUAUGU GGUUCUAAGC AUUUGAAAAA GAACAUCAAG      INCEICGSKHLKKNIKR
1651 AGACAUCUAC GGACUCAUGA UGAAGAUUCC UCACCAGGAG AAAUCAAGUG      HLRTHDEDSSPGEIKCE
1701 UGAAGUUGAG GGUUGCUCUU CGACUUUCUC CAAGGUAAAG AAACAUCCUG      VEGCSSTFSK
1751 AGCUACGUCA AACUUAUAUA GUCCAAAACA AGUUUCGUUU CCAGAUUAUU
1801 CAACAUCACU AAUUACAUUA CGAUUAUUUC UCAGGCUUCU AAUCUUCAGA      ASNLQ
1851 AACACAUGAA AGCAGUGCAC GAUGAUAUAC GUCCCUUUGU CUGUGGCUUU      KHMKAVHDDIRPFVCGF
1901 CCCGGUUGUG GCAUGAGAUU UGCUUACAAA CAUGUCAGAA ACAAGCACGA      PGCGMRFAYKHVRNKHE
1951 GAAUUCCGGG UAUCACGUAU AUACCUGCGU AAGUUCAUCC AACCUACAUA      NSGYHVYTC
2001 CUAUCGUGUU UUUCUUACAA ACUCAAAAGA CUAGAAUCUC AUGUAAAACU
2051 GAAUGUGGUU UCAGGGUGAU UUUGUCGAAA CUGAUGAAGA UUUCACUUCA      GDFVETDEDFTS
2101 AGACCGAGAG GUGGACUAAA GAGGAAACAA GUUACUGCGG AAAUGCUGGU      RPRGGLKRKQVTAEMLV
2151 ACGAAAGAGA GUCAUGCCUC CUCGGUUUGA UGCAGAAGAA CACGAAACUU      RKRVMPPRFDA EEHET
2201 GCUAGUAGUG UCCAAGCCUU AAUUAUAUUU UCUGUCUUAA GAUAAGUGAA      C!
2251 GUAGUUUUGU GUAAAGUUCU UUUUGUUUGU GUGUUGGUAG GAAGAAAUAU
2301 AGAACUACAA UAGUAGGUAG UAAUAUAAGU AAUGUUGUGC UUAGAAUCUA
2351 UGUUCGUUUA ACCUUUUAUC UCCCACGGCU UUAAUGUAUU GAACCCAACG
2401 UUUGAGAUAU AAGUAAUGUU GUGCUUAUAA UCUAUGUUCG UUUAACC
```

RT-PCR products (from primers a and b) were cloned into TOPO vector and clones were sequenced. Primer sites are underlined. Overhang and restriction sites in primers are in lowercase.

Primer a: 5´-atgcggatccGTGCGGCGTCTTGATGGA
Primer b: 5´-ACTCTTGCAGTTCTCCTTCG

AT1G72050 PRECURSOR

```
   1 GUGCGGCGUC UUGAUGGAGG AGAUAAACCC UAGUUCUUCU GUAGACAAUA
  51 AGAGAGACAU GGCGGAAGAA GCUAAAGUUG AUGUGAAGAC UUCGGCGAAG
 101 AAGGAUAUAC GCAAUUAUCU AUGCCAGUAU UGCGGAAUCA GCAGAUCUAA
 151 AAACUAUCUC AUCACUAAAC ACAUCCAAUC UCAUCAUCAG GUUUGAGAUC
 201 UUCCUCCAAU UCGAUUCCAA UUUCUCAUCC GUGGCAUCUU GAUUUGUUUU
 251 CAAUGAAAUU GAGAGUUGAG UCUGUAGAAU CGGCGAUGGU UUGUUGAAUU
 301 GAGAGUUUCU AUGAUUCGUU UGUUUAGAUG GAACUUGAAG AGGAAAGAGA
 351 UGAUGAAGCU UGUGAGGUUG AUGAGGAGUC UUCAAGUAAU CAUACUUGUC
 401 AAGAAUGUGG UGCUGAGUUU AAGAAACCUG CUCACUUGAA GCAGCAUAUG
 451 CAGAGUCAUU CGCUCGAGGU AGAUUUAUGC AUCCUCUUGU CAUGAGAAGU
 501 CGAAUUGUUC CCAUUCUGUG UGUUGCAGCU ACAGAUGGAG AUACAUAGAG
 551 AUACUCGUGG AUUUUGCUUA GUGUUGAGUU UUGUUCUGGU UGUGAACUAA
 601 AAGUUUAUAC AUUUGCAGGA AAUAAAUAGC CUUUUGUUUA AAUCAAAAGG
 651 UCUUACCUAU GUUAUUGCGU GAGGCAUUGG AUCCCAAAGA GAGAACUCCA
 701 AAAUGCGAGG CUACAUGUUA UGGACUAGUA UCAGGUUGGG AGACCUCCUG
 751 AGAAGCUCCA GCAAGUAAGC CUCGAUCACG CAAAAUGUUU GAGGUCUGAU
 801 GUUCAAUAGC UUGUUUUGUU UCACUUUGCU UUGGACUUUC UUUUCGCCAA
 851 UGAGCUAUGU UUCUGAUGGU UUUCACUCUU UUGGUGUGUA GAGAUCUUUU
 901 ACUUGCUAUG UGGAUGAUUG UGCUGCUAGC UAUAGGAGGA AGGAUCAUCU
 951 CAAUAGGCAU CUUCUUACAC AUAAAGGGAA GCUCUUUAAG UGUCCGAAGG
1001 AGAACUGCAA GAGU
```

AT1G72050 SPLICE PRODUCT I

```
   1 GUGCGGCGUC UUGAUGGAGG AGAUAAACCC UAGUUCUUCU GUAGACAAUA
  51 AGAGAGACAU GGCGGAAGAA GCUAAAGUUG AUGUGAAGAC UUCGGCGAAG
 101 AAGGAUAUAC GCAAUUAUCU AUGCCAGUAU UGCGGAAUCA GCAGAUCUAA
 151 AAACUAUCUC AUCACUAAAC ACAUCCAAUC UCAUCAUCAG AUGGAACUUG
 201 AAGAGGAAAG AGAUGAUGAA GCUUGUGAGG UUGAUGAGGA GUCUUCAAGU
 251 AAUCAUACUU GUCAAGAAUG UGGUGCUGAG UUUAAGAAAC CUGCUCACUU
 301 GAAGCAGCAU AUGCAGAGUC AUUCGCUCGA GAGAUCUUUU ACUUGCUAUG
 351 UGGAUGAUUG UGCUGCUAGC UAUAGGAGGA AGGAUCAUCU CAAUAGGCAU
 401 CUUCUUACAC AUAAAGGGAA GCUCUUUAAG UGUCCGAAGG AGAACUGCAA
 451 GAGU
```

AT1G72050 SPLICE PRODUCT II

```
   1 GUGCGGCGUC UUGAUGGAGG AGAUAAACCC UAGUUCUUCU GUAGACAAUA
  51 AGAGAGACAU GGCGGAAGAA GCUAAAGUUG AUGUGAAGAC UUCGGCGAAG
 101 AAGGAUAUAC GCAAUUAUCU AUGCCAGUAU UGCGGAAUCA GCAGAUCUAA
 151 AAACUAUCUC AUCACUAAAC ACAUCCAAUC UCAUCAUCAG AUGGAACUUG
 201 AAGAGGAAAG AGAUGAUGAA GCUUGUGAGG UUGAUGAGGA GUCUUCAAGU
 251 AAUCAUACUU GUCAAGAAUG UGGUGCUGAG UUUAAGAAAC CUGCUCACUU
 301 GAAGCAGCAU AUGCAGAGUC AUUCGCUCGA GGAAAUAAAU AGCCUUUUGU
 351 UUAAAUCAAA AGGUCUUACC UAUGUUAUUG CGUGAGGCAU UGGAUCCCAA
 401 AGAGAGAACU CCAAAAUGCG AGGCUACAUG UUAUGGACUA GUAUCAGGUU
 451 GGGAGACCUC CUGAGAAGCU CCAGCAAGUA AGCCUCGAUC ACGCAAAAUG
 501 UUUAGAUCUU UUACUUGCUA UGUGGAUGAU UGUGCUGCUA GCUAUAGGAG
 551 GAAGGAUCAU CUCAAUAGGC AUCUUCUUAC ACAUAAAGGG AAGCUCUUUA
 601 AGUGUCCGAA GGAGAACUGC AAGAGU
```

OS02G0116000

Precursor mRNA sequence

```
   1 AGUCCUCACU CCCCGCCGCC GCCGCCGCCG CUCCAUCCUC CUCACGGAGA
  51 AGCGCGGGAA AUCACACCUC GUCGCAGCCA AGAUCUCCUC CAGUCCAAAC
 101 GCAUCAAAGA UUCCAAGGUG GGCAUCCAAU CCUCGCUUGC AAACACAACC
 151 ACCAUGCCCA AAACCGAUGC CUACUUAGGA GUUAGGAAUA UGUUUCUUCG
 201 AUUUGGUUAU CGCGUUGGUA AUUCUUCGCU UUCGAUUUGA GCAGUUAAAA
 251 GCAUGAGAUA GUAUUAACCA AAUCGAUCCU GAUCUGUGCG UUUAGAUUCU
 301 GGGCUUGUGG UCAGUGUUCG UUGAAACAGC ACGAAAUUGU UAAUAGUCUA
 351 AGCAGUGUGG AUGUCGUAUG AGUUAAUUGU AGCUUUUUUU CCUAGUAAUU
 401 UUAGCUCUAA UUCCAGGUUG UGCUGUUCGG GAGAGCUUCU UGCGUUGAGG
 451 AGAGGGGGGA UGAAUAGAAG CGUUUGCCAC CACCUUCUCA CCCAGUGCAA
 501 AACCAUAAGA GAGCUCCAAA GAAUCCACGC CCAGGCCCUC ACACACGGCC
 550 UCCACCCCAA CCAGCAGUCC AUCUCCUGCA AAAUCUUCCG GUCCUAUGCC
 601 GAGUUUGGCC GACCGGCUGA CGCAGGCAGG UUGUUUGAUG AAAUCCCCCA
 651 CCCUGACAUU AUCUCCUUCA CCAGCCUCAU GUCCCUUCAU CUCAAGUUAG
 701 AUCACCACUG GAAAGCUAUA UCCGUGUUCU CUCACGCCAU UGCUUCUGGC
 751 CACCGCCCUG AUGGCUUUGC UGCUGUCGGU GCCCUCUCGG CCUCUGGUGG
 801 AUUAGGGGAU CAGCGUAUUG GUAGCGUGGU GCAUGGCCUG AUUUUUCGGU
 851 GUGGAUUGGA CUCUGAGUUG GUCGUUUGUA AUGCUUUGGU UGAUAUGUAU
 901 UGUCGAUGUG GCAAGUUUGA GCCUGCACGG ACUGUAUUUG AUAGAAUGCU
 951 UGUAAAAGAU GAGGUCACUU GGGGUAGCAU GCUGUAUGGU UACAUGAAGU
1001 GUGUUGGUGU GGAUUCAGCU UUGUCAUUCU UUUAUCAGAU GCCUAUGAAG
1051 AGCACUGUUU CUUGGACAGC ACUGAUCACU GGUCAUGUUC AAGACAAGCA
1101 GCCUAUCCAA GCUCUUGAGC UAUUUGGUAA GAUGCUUUUG GAGGGCCACC
1151 GUCCUAACCA UAUUACAAUU GUAGGGGUGC UAUCAGCUUG UGCUGAUAUU
1201 GGUGCUUUGG AUCUUGGACG UGCCAUUCAU GGAUAUGGAA GCAAAUCUAA
1251 UGCCACCACC AAUAUAAUUG UUACAAAUGC UCUGAUGGAU AUGUAUGCAA
1301 AGAGUGGAAG CAUUGCUAGU GCAUUUUCUG UAUUUGAAGA AGUUCAGAUG
1351 AAGGACGCAU UCACAUGGAC AACUAUGAUU UCAAGUUUCA CUGUCCAGGG
1401 UAAUGGGAGG AAAGCUGUUG AGCUCUUUUG GGAUAUGCUA AGGUCUGGGA
1451 UACUUCCAAA CAGUGUGACA UUCGUCUCAG UUUUGUCAGC AUGCAGCCAU
1501 GCUGGGUUAA UACAAGAAGG CAGAGAGUUA UUUGAUAAAA UGCGUGAAGU
1551 CUACCAUAUU GAUCCCCGGC UUGAGCACUA UGGAUGCAUG GUUGAUCUGU
1601 UAGGACGGGG UGGACUUCUA GAAGAAGCAG AAGCUCUGAU AGAUCAUAUG
1651 GAUGUCGAGC CUGAUAUUGU UAUUUGGAGG UCACUUCUUA GUGCAUGCCU
1701 AGCUCAUGGC AAUGAUAGAU UAGCUGAGAU UGCUGGAAUG GAAAUUAUAA
1751 AGAGAGAACC UGGGGAUGAU GGGGUUUAUG UGCUUCUUUG GAACAUGUAU
1801 GCCUUAUCAA ACAGAUGGAA AGAAGCUUUG GAUAUGAGGA AGCAGAUGUU
1851 GAGUAGGAAA AUAUACAAAA AACCGGGUUG CAGUUGGAUU GAAGUUGAUG
1901 GUGUCGUCCA UGAAUUUUUG GUAGAGGACA AGACACAUGA UGCUAGAAGG
1951 GAAAUUUAUG UCACCUUGGA AAUCAUGGCU AGGCAUCUCA AGAUGGAUCC
2001 CAUACCAUCU CCAUUAGUGU UAACUGAUGG AGAACACAUG CUCUAGAAAC
2051 ACAUAUUGUC CUUUUUUUUU CUUUCCUUUG AUGACCUCUA ACCGAACUGA
2101 GGAUGAUAUG CUGAUAUUUU UUUCUUUUUG UUAGUGACAA GGCACCAUAG
2151 UUUUGUUUUG AUUUGUUUGA UUAACCAAUU AGUGAUAGCC AGUUGUAACA
2201 CACUGAUGUU AUUUCAUUGA AAAUGCAGAA UUGUAGAGAU UUAUUUUAGA
2251 UGUCGACCUG CAGAUAUGCU AUACCUGUAU UUUGUUUGGU CACUCUGUCA
2301 AAAAAAGAGU AAAGACUGCA AUGCGUCCCA GUUCAUGUUA UGGACUUCUC
2351 CUCAUUUUAC AGCUAUCUUG CUGACUACAU CGCUACCUCU GAAUGUUUUA
2401 CUUGUUUACU UUUUUUUAUA UAUAAAUGAC UUAUUUAUGU GUGCCGAAUU
2451 UCUAACUAUC UGAUUUGCCU ACUAUUAGAU GUGCUCUGGA GAUGAUAUUG
2501 AUGGAGACAU GAGGGUUGAA GCAACACAAC ACAGGGAUAU AAGGCGCUAC
2551 AAGUGUGAAU UUUGCACGGU UGUUAGGUCC AAAAAAUGUC UCAUCCGAGC
2601 UCACAUGGUU GCUCAUCAUA AGGUAUAAAA CUACUUAGAU AUUUUAGGAA
2651 UAUGGGCCAU AUUAUGCCUU AUGUUUUAAA GACAAUGUGG ACUGUAUUAA
2701 UUAAAUUGAG UUAUUUGACC GUUAUGGUUC UCUGAUUGGA GUUUAGGUCC
2751 AUCAAUCAAU AUAUUUGUGC UUUUGUCAAU UCCUUCUGCU CUUAUCCAUU
2801 UGAAUCAUUA UAUGAAUAUC UGAUGCCCUC ACUUCAAAUU UAUCUAUAGG
2851 AAGAACUGGA CAAAUCGGAA AUCUACAAGU CAAAUGGGGA AAAGGUUGUU
2901 CAUGAAGGUG ACCACACCUG CCAAGAGUGU GGUGCUUCUU UCCAGAAGCC
```

Protein sequence

```
MCSGDDI
DGDMRVEATQHRDIRRY
KCEFCTVVRSKKCLIRA
HMVAHHK

E
ELDKSEIYKSNGEKVVH
EGDHTCQECGASFQKPA
```

```
2951 AGCUCAUCUG AAGCAGCAUA UGCAAAGUCA CUCUGAUGAG GUUGGCAUUU       HLKQHMQSHSDE
3001 CUCAUAUGUG CAUACAACAU GAUAUUGCUU UUUUGUUGGA AUGGUGCUGG
3051 GAUUAGUUUU UUUGGAAACA UUUAAACAUA ACAUUCUUGU AUGGUUUUUC
3101 UCCGUCCGCU GAUCUGGUUG UCAAAGAAAU UGUUCUUAUC UACUUCAAUU
3151 UGCUAAAAAC AAGUGUUUUG CAGGAUAUAG AUUCUUAACU GUGUGAAGCA       DIDS!
3201 UUGGAUCCCA AAGAACUCCA AAAUGCGAUG AGGCAUAUUU AAUCUUGUCU
3251 GGACUAGUAA CAGGUUGGGA UGACCACCUG UGAAGCUCCA ACAGGAUUGC
3301 CUCCUCACGC UCUUUCAGGA GGUUUGUCCU UGCUUGUAUU AGUUCUUUUC
3351 UUUUUUACUA UUUUGAUUAU UAUUCAACUU CAUGGGAUAG CCUGCAUCUC
3401 CUUGUUUCCA AUUAAGUACU AAUGUCUCCU UUGUUGUAUG CACAUGUGCU
3451 GGAAUUCAUU GUUUAUCAAU UUGCCUUGUG GUAUGCUAUA CAUGAUAUCU
3501 UUAAUUUUUA UUAGUAACAA GGUUUUUUAC AGCAUACAUA UACAUAAAUA
3551 GUGUUUUUUU ACAGCAUGCA UAAAUGGUGA AAACUGAUUA CUGCUUAAUU
3601 AUUGACUAGU UGCUGGCUAU UUUACAUUUU UUUGAUAUUG CAAAUCUUGU
3651 GAAAUGGAUC CAUGCAGUGA AUUUGGGAGG UUAUCUUCUC CUUUGUUCUA
3701 AUGUUCUCUC UCACUCUCUU UGGGCUUCUG GAACCUGGUU UUGGAGGUCC
3751 UUAGUUGGAA CAAAGAGGGG AUCCAAGAUC AAAAUACCAU CUCAGAAAGC
3801 UAGGGCCUUC UGUAAUAGGU UCAUUUAGCU CGCUAUGGUG UUGAUAUCUC
3851 CAUCAUAUAC AAAACUACAA UAUGUUCCGU UGCAAAUUAU AUAUGAAUUG
3901 UUUUAGCUGA UAAACACCUG UGCACAACUU UUUGCUGAUA AAUAGUACUA
3951 CUUCUAUUUC AGGUUAUAAG AUGUUUUGAC UUUGGUCAAA GUCAAACUGC
4001 UUUAGGUUUG ACUUAGUUUG UAGAAAAAAA UAAUAACAUU UUCAACCCAA
4051 GACAAAUUUA UUAUGAAAAU AUAUUAAAUU AUUGAUUUAA UGAAACUAAU
4101 UUGGUAUUAU AAAUAUUACU AUAUUUGCCU AUAAAAUUAU UCAAACUUGA
4151 AGUAGUCUGA CUUUGACUAA AGUCAAAACU UCUUAUAACC UGAAACGGAG
4201 GGAGUAGGCA UUGAAUAGAA UAGGUUAUGG AUCCUUUGGA GAAUAAGGUU
4251 AGCAAAUAUC CUUCAUUCAU CCUACUAUUU CUUUGGAUCU CCUUUUGGGA
4301 GCUUUGAGAA GAGGUGAGAG GAUCCAACAU CAAAAUGUCA CCUCAGAAAG
4351 CUAGGACUAU UCAUCAUUGG UUCUCAGCUC GCUAUGGUGA UUAUACUCCC
4401 UUUUAUCUCU AUCAUCUUAA UGGAAAACAC UGAACAUUUA GGUAUUCGUU
4451 GCAUAUUCCG UUAACCAGUU UACUAGAUAA UUUAUUUUAG AUAAUUUAUA
4501 AGUAUGAAAG AAAUGUGUAA GCUUACAGCA GCUUCAGUUU UGACUCGUUC
4551 CCUUCUUGAC UGUAGAGAUC GUUCAUUUGC CCACUUGAAG ACUGCCCUUU       RSFICPLEDCP
4601 CAGCUACAUU AGGAAAGAUC ACUUGAACCG UCAUAUGCUU AAGCAUCAAG       FSYIRKDHLNRHMLKHQ
4651 GGAAGUUGUU UACCUGCUCU AUGGAUGGUU GUGGUAGGAA GUUCAGUAUA       GKLFTCSMDGCGRKFSI
4701 AAGGCAAAUA UGCAAAGGCA UGUAAAGGAA AUCCAUGAGG AUGAAACUGC       KANMQRHVKEIHEDETA
4751 UACUAAAAGC AACCGGCAGU UUGUUUGCAA GGAGGAGGGC UGUAACAAGG       TKSNRQFVCKEEGCNKV
4801 UUUUCAAGUA UGCUUCAAAG AUGAAGAAAC AUGAAGAAUC ACAUGGUAUC       FKYASKMKKHEESH
4851 UCCUCAGUUA CUUAUAGUAG AUCAGGCGCU GCCUUGUAAA UUUCAGAACU
4901 AAAAAACUGC AGCUCAAUCC AUGCAUUAUA UAAUUAAUCA UGCUAAUGGU
4951 AUCCCCAGAA UUAAAUUAAU UUGUGCUGUU CUGGCAGUGA AAUUGGACUA       VKLD
5001 CGUGGAAGUA GUCUGCUGUG AACCAGGCUG CAUGAAGACA UUCACAAAUG       YVEVVCCEPGCMKTFTN
5051 UUGAAUGUUU GAGGGCUCAU AAUCAGGCUU GCCAUCAAUA UGUUCAGUGU       VECLRAHNQACHQYVQC
5101 GAUAUCUGUG GGGAAAAACA CCUGAAGAAG AACAUUAAAA GGCACCUACG       DICGEKHLKKNIKRHLR
5151 AGCCCAUGAG GAGGUGCCAU CUACUGAAAG GAUAAAAUGC AGUUUUGAGG       AHEEVPSTERIKCSFEG
5201 GCUGUGAAUG CUCUUUUUCC AAUGUAAGUC UAUUAUCCCC AAAUGUUCUG       CECSFSN
5251 AUGCAUUGUU UGUUGUCUUC CUCUUGAACU GUAGCUUUAG UCGUAUCUGG
5301 UGUGGAUUGA AACUUGUUAC CUGUCCCUCC UAUUGCAGAA AUCAAAUUUA       KSN
5351 ACCAAACAUA UAAAGGCGUC CCAUGAUCAA GUAAAACCUU UCGCAUGUCG       LTKHIKASHDQVKPFAC
5401 AUUCACGGGG UGUGAAAAGG UGUUUCCAUA CAAGCAUGUC AGGGACAACC       RFTGCEKVFPYKHVRDN
5451 AUGAGAAAUC CAGCGCUCAU GUAUACACUC AGGUGAGUUU UGCUAACCGA       HEKSSAHVYTQ
5501 UCUUUAGCUC UUCUGCGCGC GUCUUGUGCU UGCCCAAAUC AUUUGAGUGA
5551 CAAACUGACA AGCUUUAACC ACAUCCAGGC AAACUUCACG GAGAUGGACG       ANFTEMD
5601 AGCAGUUACU CUCGUGUCCG AGAGGUGGAC GGAAGAGGAA AGCUGUGACU       EQLLSCPRGGRKRKAVT
5651 GUCGAAACUC UUACGCGCAA GAGGGUGACC AUGCACGGCG ACGCUUCGUC       VETLTRKRVTMHGDASS
5701 GUUGGACAAU GGAACUGAGU ACCUGCGCUG GUUGCUUUCU GGUGGGGAUG       LDNGTEYLRWLLSGGDD
5751 AUGAUUCGAG CCAAACUCAU UAGCCGUAUC AUCUUUGCAU CUGAUUGUGA       DSSQTH!
5801 UGGUCAUGGU GUCAAUGGCU CAAUGCUUUG UCGUAAAGGG GUGUGCUAGC
5851 CCACUCACGU UGGUAGUUAG AUGGCCGUGG UGUUUUUGUU GUAAGUAGUU
5901 GGUCUGAAUU UAGAAAGUGC AGCAGUUUUA GAGGAAGCCU ACUAGUAGUA
5951 UAUAUUUGUA GCAUAAUGGC GCAGUUUGAG UUGUGAGUUG UAAGUAUUUU
6001 GAUUCUUUAC UGUUUUGACC ACUGUUCUUG GAACCUCAAA AUUAUUUGAC
6051 UUGUUGAAUG CAAAAAAUGA UGUGAUUACA AGUUUUGGUU CUUC
```

RT-PCR products (from primers c and d) were cloned into TOPO vector and clones were sequenced.

Primer c: 5´-TGATGGAGACATGAGGGTTG

Primer d: 5´-TGTTACAGCCCTCCTCCTTG

OS02G0116000 SPLICE PRODUCT I

```
  1 UGAUGGAGAC AUGAGGGUUG AAGCAACACA ACACAGGGAU AUAAGGCGCU
 51 ACAAGUGUGA AUUUGCACGG UUGUUAGGUC CAAAAAACGU CUCAUCCGAG
101 CUCACAUGGU UGCUCAUCAU AAGGAAGAAC UGGACAAAUC GGAAAUCUAC
151 AAGUCAAAUG GGGAAAAGGU UGUUCAUGAA GGUGACCACA CCUGCCAAGA
201 GUGUGGUGCU UCUUUCCAGA AGCCAGCUCA UCUGAAGCAG CAUAUGCAAA
251 GUCACUCUGA UGAGAGAUCG UUCAUUUGCC CACUUGAAGA CUGCCCUUUC
301 AGCUACAUUA GGAAAGAUCA CUUGAACCGU CAUAUGCUUA AGCAUCAAGG
351 GAAGUUGUUU ACCUGCUCUA UGGAUGGUUG UGGUAGGAAG UUCAGUGUAA
401 AGGCAAAUAU GCAAAGGCAU GUAAAGGAAA UCCAUGAGGA UGAAACUGCU
451 ACUAAAAGCA ACCGGCAGUU UGUUUGCAAG GAGGAGGGCU GUAACA
```

OS02G0116000 SPLICE PRODUCT II

```
  0 UGAUGGAGAC AUGAGGGUUG AAGCAACACA ACACAGGGAU AUAAGGCGCU
 51 ACAAGUGUGA AUUUUGCACG GUUGUUAGGU CCAAAAAAUG UCUCAUCCGA
101 GCUCACAUGG UUGCUCAUCA UAAGGAAGAA CUGGACAAAU CGGAAAUCUA
151 CAAGUCAAAU GGGGAAAAGG UUGUUCAUGA AGGUGACCAC ACCUGCCAAG
201 AGUGUGGUGC UUCUUUCCAG AAGCCAGCUC AUCUGAAGCA GCAUAUGCAA
251 AGUCACUCUG AUGAGGAUAU AGAUUCUUAA CUGUGUGAAG CAUUGGAUCC
301 CAAAGAACUC CAAAAUGCGA UGAGGCAUAU UUAAUCUUGU CUGGACUAGU
351 AACAGGUUGG GAUGACCACC UGUGAAGCUC CAACAGGAUU GCCUCCUCAC
401 GCUCUUUCAG GAGAGAUCGU UCAUUUGCCC ACUUGAAGAC UGCCCUUUCA
451 GCUACAUUAG GAAAGAUCAC UUGAACCGUC AUAUGCUUAA GCAUCAAGGG
501 AAGUUGUUUA CCUGCUCUAU GGAUGGUUGU GGUAGGAAGU UCAGUGUAAA
551 GGCAAAUAUG CAAAGGCAUG UAAAGGAAAU CCAUGAGGAU GAAACUGCCA
601 CUAAAAGCAA CCGGCAGUUU GUUUGCAAGG AGGAGGGCUG UAACA
```

OS05G0121400

First cassette sequence, second cassette sequence

Precursor mRNA sequence

```
   1 GGUUCGAGUC GAGGGUUUCU CCUUCUCCCA GUGCGCCGCC GUCGCCAACA
  51 UCUGGAGCCU CGCGGAGCCG CUGAUUCAGC GAAUUCCUCU GCAAUUAGAG
 101 GUAAUCCUCU GCUUAUUCAC GCCGCGUGUA GCGCGGAUUU CGGUGAUUUU
 151 GGUUGAUUUG GUUGUGAUUU UGCUCGUUAU GUGUGUUUUG UGACGCGAAU
 201 UUUGGGGGAU UGGUUUGGUC AGAGAUGGGG AGCGUGGAGC UCGGAGCGGA
 251 GGAAAGGGAG GUCGCCGGCG GGGAGGGGGG GAGCAAGGGG GCGGCGCCUC
 301 CUGCUAGGGA UAUUAGGCGG UACAAGUGCG AUUUCUGCAG UGUCGUCCGU
 351 UCCAAGAAGG GGUUGAUCCG UGCCCACGUC CUCGAACACC AUAAGGUCGC
 401 UCAUGGCGUC UCUCGCUUCA UAGUUCAGAA CCUCAGAUGC GUCUAGUGGU
 451 CAUAUUCUCC GUGCUUUGGC UGCUCGCCAC GGUGGUUGGU UACUUGGAUG
 501 GAUGAAUGAU UGACUGUUUG GUUUUGUCGG UGCUGCUAUU UGUGUUUUUA
 551 GGAUGAAGUG GAUGAUUUGG AUGAUUACUU GGGACGUGGU GGCGGCGAGA
 601 CGUGCAAAGA GAUGGACCAU GAUUGCAAGG UGUGCGGUGC UAGUUUUAAG
 651 AAGCCGGCCC AUUUGAGGCA ACACAUGCAG AGCCAUUCAC UUGAGGUAUC
 701 UGAAUUUCGA UACAUGUUCA CUGAAAUGAU CAAGUUCUUU GGCGCAUGUG
 751 AUAGAUGCUA UGUGUUUCUC UGUUCAGAGU AUUACUUGCA AUAGCUGUGA
 801 UAGAUGCUCU GUCCUAUGUA GGGACUAGGG AUGGUGAUAG UUAUGCUGUG
 851 UUGACAGCAA AUUUCCCUGG ACAGUUAAUA GUUGUCGUUU CACUCCUUUU
 901 UUGGUUGGGU AGGUUUGUUG AUCUAUGCUG AUUUUUGCAG ACUAUUUCAC
 951 UAUUUUGUUA CAUAUGGAUG UGUGAAGCAU UGGAUCCCAG AGAACUCCAA
1001 AAUGCGAGAA GGCCUUGAGU GGACUAGUAC CAGGAUGGGG AGACCCUUCU
1051 GGGAAGCUCC UUCAAGAAUG CCUCUUUCAC GCUCUUCUUC AGGUGAUUUC
1101 UUCUGUUGCA AAAAACUGCU UGCUCUAUGC UACUGCUGUU AUUGUUACGU
1151 GCGCUGCAUU CUUAGUUUUU CCAAGCUAUU GAAGGAUCUU UUCUUACGCG
1201 CAAGUCUACU GUUGAAGAAC UGCAUGCGUA GCGCUGUUGG UUGUAUCUAA
1251 UGCAAGCACC ACUUGAUGGU CUUUUGAUCA AACAGUUAAC CAAGUAGGUA
1301 GAGCAUGUAC ACUGAAGAGA UAGCUAUGCU UGUGAUGGAA UAACCAGCUC
1351 CUCGUAUUCC CCUUUAGUGU AAAAGACGUG UAUCUGUCCU GUUCAGUGAA
1401 UAUUCUACAC AACAACCAAA ACAAUUACUC AUUUGAAAGU UAAUUCAAGU
1451 AUAUAGGUAU AGGUGCCAUU UGGCUCUCUG UUUCGUUGGA UUAGUGCAUG
1501 GCCUAUUGUG GUUUCUGAUA UAUUUUGAUU UUGGCAGACU AUAUAUGACU
1551 GCGUGAAGCA UUGGAUCCCA AAGAACUCCA AAAUGCGUGA AGGCUUUGAA
1601 UGGACUAGUA UCAGGAUGGG GAGACUCUCC UGAGAAGCUC CAUCAAGAAU
1651 GCCUCGUUCA CACACUUUUU CAGGUUUGCU UGUUCAUUUC AAAUUGCUCA
1701 UUCUAAGACU UGAUUUUGUU CCAACUGUUG UUAUUACUUA CAGUGAUAUG
1751 UACGUACUGA GAUAUUGUUA UUCUCAAGCU ACUUAAGGAU CCUGUCUUAG
1801 AUGCAAUUAC UCAGUCAAGA UACACAUGUA UAUUUGUCAC UGGUAUUUGA
1851 CUAUUCAUGC AGCUGUCUCA UUUGCAGCAU AGUAGUGUGA UUAUCGGAAC
1901 AGUUUUAGCA UAUUGCACUG AAUGGAUAUG UGUGUACAUU GACCUACUUA
1951 UUUUAUGUCA AGUAAAUCAU AAAGUUCUGC UUUCAGUAUA ACAUUUAAGG
2001 UAUAUGCUUU UGCAGAGGCC CUUUUCCUGC CAUGUAGAUG GUUGCCCCUU
2051 CAGCUAUAGC AGGAAGGACC AUUUGAACCG GCAUCUACUU ACACAUCAAG
2101 GAAAGCUAUU UGCAUGCCCC AUGGAAGGAU GCAACCGUAA GUUCACUAUA
2151 AAGGGUAAUA UCCAAAGGCA UGUUCAGGAA AUGCAUAAAG AUGGCUCUCC
2201 UUGUGAAAGC AAGAAAGAGU UCAUCUGUCC UGAGGAAAAC UGUGGGAAGA
2251 CUUUUAAAUA UGCUUCUAAG UUACAGAAGC AUGAGGAAUC ACAUGGUGAG
2301 AUACUUGUGU GCCAUAUUUU UCUUACUGUC UUUAAAUAGC UCAGCAACCC
2351 UCCUAGAUCA AUUUUAUCAU AUAUAGACUG GUUAUGUCAU CUGACACUCU
2401 GACUUUUUCU AUUUCUGAGC UAGAAUACUG AUAUAAACUC UUCAAAGUAU
2451 CCUGCAAGAU ACUCCUGUUU CUUUUAAAGA UCAUUUAUCU CUUCAGAUCA
2501 AAUGCAUGAU UUACUUAUAC UGUGCAUGCG CUGUAUGCCU GCUUUUACUU
2551 CCUGUGAGAA GUUGAGAAGU GAGAAUUGGU UAAUUGUGCC GUGACAUACC
2601 UUGUUUUACU UGUGCAGUCA AGUUGGAUUA UAGUGAAGUU AUCUGCUGUG
2651 AACCAGGCUG CAUGAAGGCA UUUACCAACU UGGAAUGCCU CAAGGCCCAU
2701 AACAAAUCUU GUCAUCGACA UGUUGUCUGU GAUGUCUGUG GAACCAAACA
2751 GCUGAAGAAG AACUUCAAAC GUCAUCAGAG AAUGCAUGAA GGUUCCUGUG
2801 UUACUGAGAG GGUUCGAUGC CACCUCAAGG ACUGCAAGUU GUCAUUUUCA
2851 AAGGUAGGAU UUCAAUCUUU CAGGGUUGAU UCUGUUUCAC AGAAUUUUUU
```

Protein sequence

```
                                                    MGSVELGAE
                                                    EREVAGGEGGSKGAAPP
                                                    ARDIRRYKCDFCSVVRS
                                                    KKGLIRAHVLEHHK

                                                    DEVDDLDDYLGRGGGET
                                                    CKEMDHDCKVCGASFKK
                                                    PAHLRQHMQSHSLE

                                                    TIS
                                                    LFCYIWMCEALDPRELQ
                                                    NARRP!

                                                    TIYD
                                                    CVKHWIPKNSKMREGFE
                                                    WTSIRMGRLS!

                                                    RPFSCHVDGCPF
                                                    SYSRKDHLNRHLLTHQG
                                                    KLFACPMEGCNRKFTIK
                                                    GNIQRHVQEMHKDGSPC
                                                    ESKKEFICPEENCGKTF
                                                    KYASKLQKHEESH

                                                    VKLDYSEVIC
                                                    CEPGCMKAFTNLECLKA
                                                    HNKSCHRHVVCDVCGTK
                                                    QLKKNFKRHQRMHEGSC
                                                    VTERVRCHLKDCKLSFS
                                                    K
```

```
2901 UAUUUCUUUC CAAUAAAUUG CAAAUCGUUC CAUCCUUUUC AAUUUUUUUU
2951 UUUUGGGGGG GGGGGGGGGA CUUUGAACUG CAGAAAUCCA AUCUGGACAA        KSNLD
3001 GCAUGUGAAA GCAGUCCAUG AGCAGAAGCG ACCCUUUGUA UGUGGAUUCU        KHVKAVHEQKRPFVCGF
3051 CUGGGUGUGG CAAGAGUUUU UCGUACAAGC AUGUAAGGGA CAACCAUGAG        SGCGKSFSYKHVRDNHE
3101 AAAUCUAGUG CUCAUGUGUA CGUUCAGGUA ACACAAACAA CAAUUCUGCU        KSSAHVYVQ
3151 AUGAAAUAGC UGACCGAUGG UAUCUUCUUC UUGUUAAACC CUUAAGAUGA
3201 UUAUAAUUUU CCCAUCAUUA UGUGCAGGCU AAUUUCGAAG AAAUUGACGG        ANFEEIDG
3251 GGAACGACCA CGUCAAGCAG GUGGGAGGAA GAGGAAAGCC AUACCUGUGG        ERPRQAGGRKRKAIPVE
3301 AGAGUUUGAU GCGGAAGAGG GUAGCUGCUC CUGAUGAUGA UGCCCCAGCU        SLMRKRVAAPDDDAPAC
3351 UGUGAUGAUG GAACUGAGUA UUUGAGAUGG CUUCUCUCAG GCUGAUUCGU        DDGUEYLRWLLSG!
3401 CGUGAACCAG AAUACAGAAA UGUGAGAUCU UUGGAUAUCA AGGGUGUUGA
3451 UAUGGUUUGA GAUACUAGUA GCAACGUUCU UUUCCUUUGU GCAUGUAUAG
3501 GUAAAACAUG GAAGUAAUUU UGAUAAGACC UCUAUGUACU GUCUCAGACA
3551 UCAAGUUACC AGAAGUCCAG AAUGCAGAAA CGCCAGAUCU UUAUAUGCUC
3601 AAGGGUGUUU AUGUGCUAGU UAGGUAGCAA CGUUCUCUCU UUUUGUAUGU
3651 AUAGGAAAAG UGAAAGGAAU UUUAGUAGAC CCUAAUGUCU UGUGACAGGC
3701 AUGGUAAUUU AUCAAGGUUG GUUUGUUCUA CAGUGU
```

RT-PCR products (from primers f and g) were cloned into TOPO vector and clones were sequenced.

Primer f: 5´-ATGATTACTTGGGACGTGGT

Primer g: 5´-GTCTTCCCACAGTTTTCC

OS05G0121400 SPLICE PRODUCT I

```
   1 AUGAUUACUU GGGACGUGGU GGCGGCGAGA CGUGCAAAGA GAUGGACCAU
  51 GAUUGCAAGG UGUGCGGUGC UAGUUUUAAG AAGCCGGCCC AUUUGAGGCA
 101 ACACAUGCAG AGCCAUUCAC UUGAGAGGCC CUUUUCCUGC CAUGUAGAUG
 151 GUUGCCCCUU CAGCUAUAGC AGGAAGGACC AUUUGAGCCG GCAUCUACUU
 201 ACACAUCAAG GAAAGCUAUU UGCAUGCCCC AUGGAAGGAU GCAACCGUAA
 251 GUUCACUAUA AAGGGUAAUA UCCAAAGGCA UGUUCAGGAA AUGCAUAAAG
 301 AUGGCUCUCC UUGUGAAAGC AAGAAAGAGU UCAUCUGUCC UGAGGAAAAC
 351 UGUGGGAAGA C
```

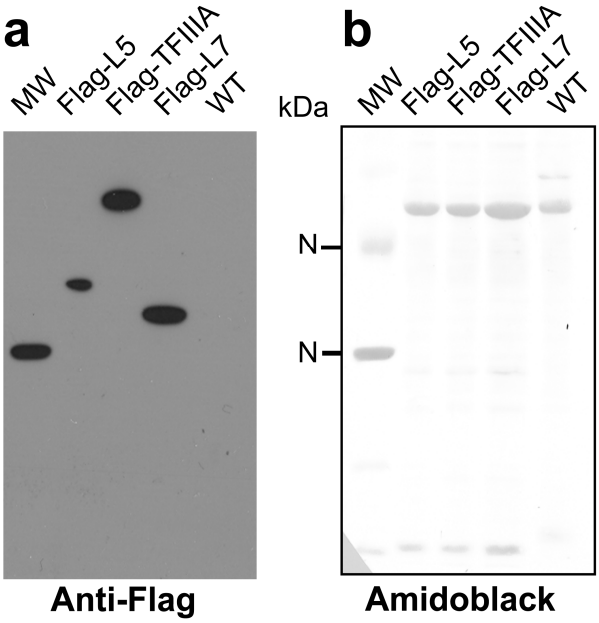
OS05G0121400 SPLICE PRODUCT II

```
   1 AUGAUUACUU GGGACGUGGU GGCGGCGAGA CGUGCAAAGA GAUGGACCAU
  51 GAUUGCAAGG UGUGCGGUGC UAGUUUUAAG AAGCCGGCCC AUUUGAGGCA
 101 ACACAUGCAG AGCCAUUCAC UUGAGACUAU UUCACUGUUU UGUUACAUAU
 151 GGAUGUGUGA AGCAUUGGAU CCCAGAGAAC UCCAAAAUGC GAGAAGGCCU
 201 UGAGUGGACU AGUACCAGGA UGGGGAGACC CUUCUGGGAA GCUCCUUCAA
 251 GAAUGCCUCU UUCACGCUCU UCUUCAGACU AUAUAUGGCU GCGUGAAGCA
 301 UUGGAUCCCA AAGAACUCCA AAAUGCGUGA AGGCUUUGAA UGGACUAGUA
 351 UCAGGAUGGG GAGACUCUCC UGAGAAGCUC CAUCAAGAAU GCCUCGUUCA
 401 CACACUUUUU CAGAGGCCCU UUUCCUGCCA UGUAGAUGGU UGCCCCUUCA
 451 GCUAUAGCAG GAAGGACCAU UUGAACCGGC AUCUACUUAC ACAUCAAGGA
 501 AAGCUAUUUG CAUGCCCCAU GGAAGGAUGC AACCGUAAGU UCACUAUAAA
 551 GGGUAAUAUC CAAAGGCAUG UUCAGGAAAU GCACAAAGAU GGCUCUCCUU
 601 GUGAAAGCAA GAAAGAGUUC AUCUGUCCUG AGGAAAACUG UGGGAAGAC
```

OS05G0121400 SPLICE PRODUCT III

```
   1 AUGAUUACUU GGGACGUGGU GGCGGCGAGA CGUGCAAAGA GAUGGACCAU
  51 GAUUGCAAGG UGUGCGGUGC UAGUUUUAAG AAGCCGGCCC AUUUGAGGCA
 101 ACACAUGCAG AGCCAUUCAC UUGAGACUAU AUAUGACUGC GUGAAGCAUU
 151 GGAUCCCAAA GAACUCCAAA AUGCGUGAAG GCUUUGAAUG GACUAGUAUC
 201 AGGAUGGGGA GACUCUCCUG AGAAGCUCCA UCAAGAAUGC CUCGUUCACA
 251 CACUUUUUCA GAGGCCCUUU UCCUGCCAUG UAGAUGGUUG CCCCUUCAGC
 301 UAUAGCAGGA AGGACCAUUU GAACCGGCAU CUACUUACAC AUCAAGGAAA
 351 GCUAUUUGCA UGCCCCAUGG AAGGAUGCAA CCGUAAGUUC ACUAUAAAGG
 401 GUAAUAUCCA AAGGCAUGUU CAGGAAAUGC AUAAAGAUGG CUCUCCUUGU
 451 GAAAGCAAGA AAGAGUUCAU CUGUCCUGAG GAAAACUGUG GGA
```

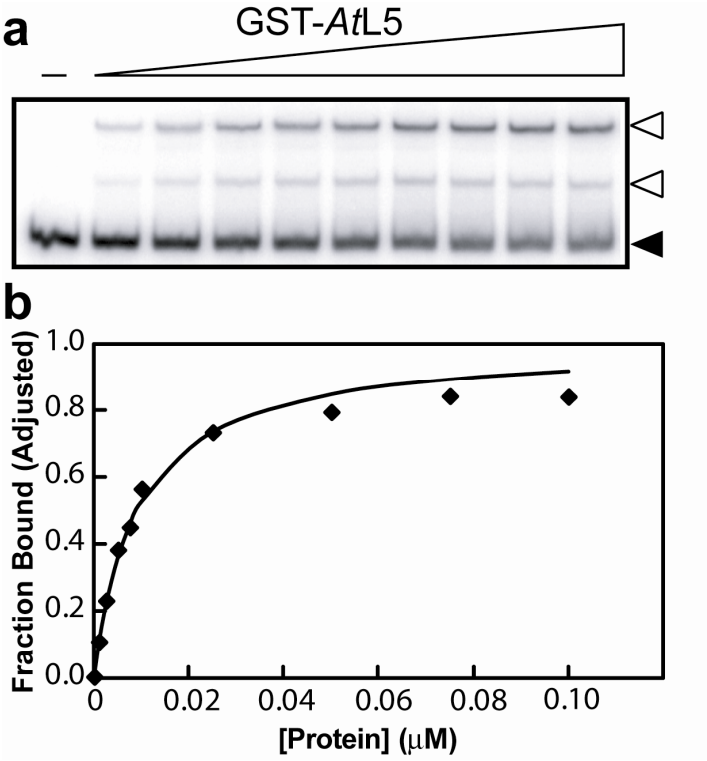
Supplementary Figure 5. Expression of Flag-tagged versions of L5, TFIIIA, and L7 proteins



(**a**) Western blot analysis of crude extracts from *N. benthamiana* infiltrated with Flag-L5, Flag-TFIIIA, Flag-L7 coding sequence constructs, or none (WT). A Flag peptide sequence was inserted at the N-terminus after the start codon and does not change the effect on Pre-EGFP reporter splicing (data not shown). The immunoblot was probed with anti-Flag antiserum. The 30 kDa molecular weight marker (MW) protein cross-reacts with the antibody. The molecular weights for the detected proteins were estimated using a standard curve for all marker bands (expected sizes in parentheses): Flag-L5 32 kDa (36 kDa), Flag-TFIIIA 50 kDa (48 kDa), Flag-L7 28 kDa (30 kDa).
(**b**) Amidoblack staining of the western blot shown in **a**, which visualizes all proteins present on the blot. The major band in the plant extracts corresponds to Rubisco.

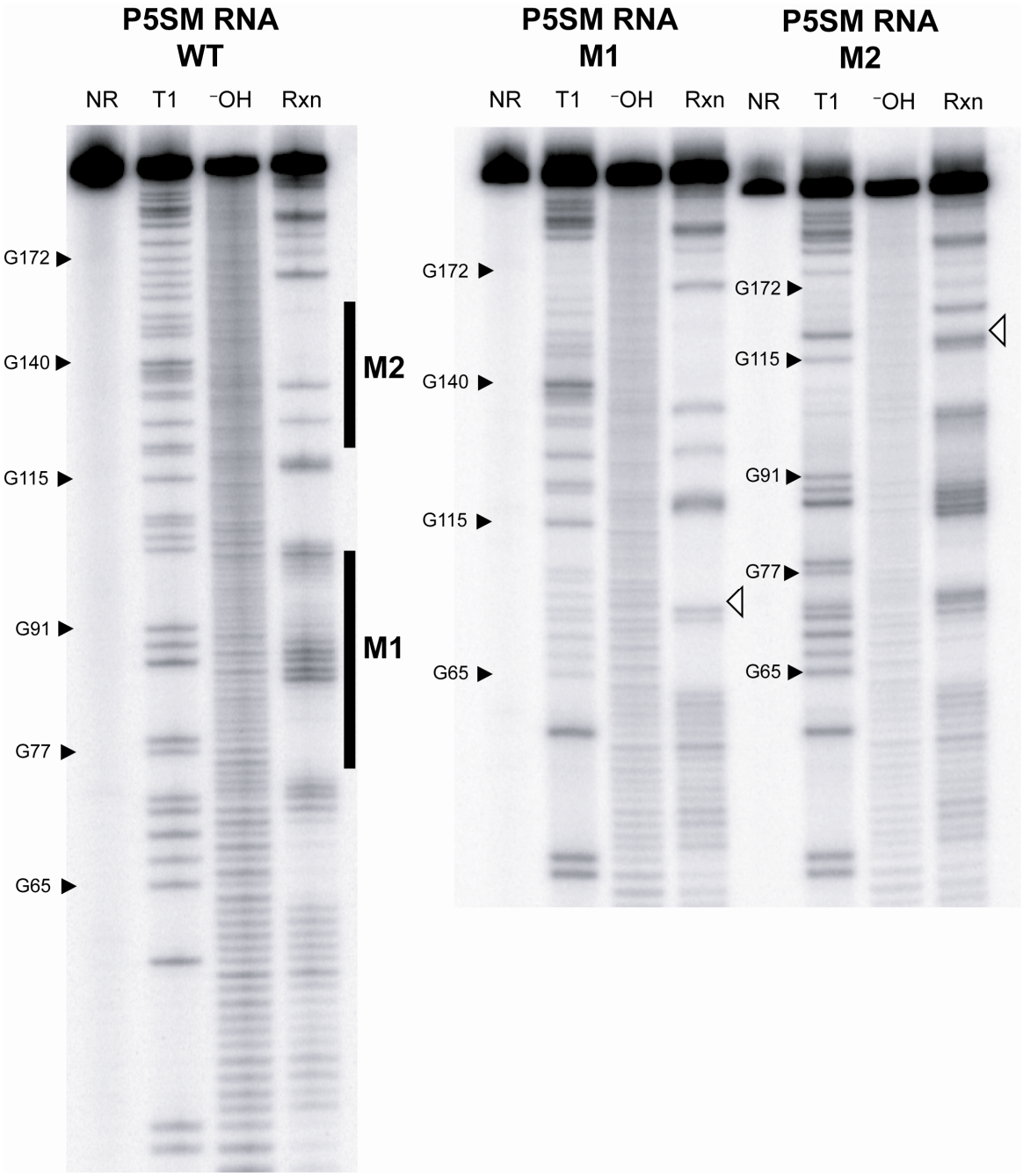
The expression of all three proteins tested for an effect on TFIIIA reporter was confirmed by western blot analysis. In general, somewhat less L5 is expressed than TFIIIA or L7, but it alone has an effect on reporter splicing and expression (**Fig. 3a, c**). No protein was detected by anti-Flag probing of crude extracts from wild-type *N. benthamiana* plants. Roughly equal loading of crude extracts in all lanes is shown by amidoblack staining and comparison of the main band corresponding to the highly abundant protein Rubisco (large subunit).

Supplementary Figure 6. GST-*At*L5 fusion protein binds to 5S rRNA *in vitro*

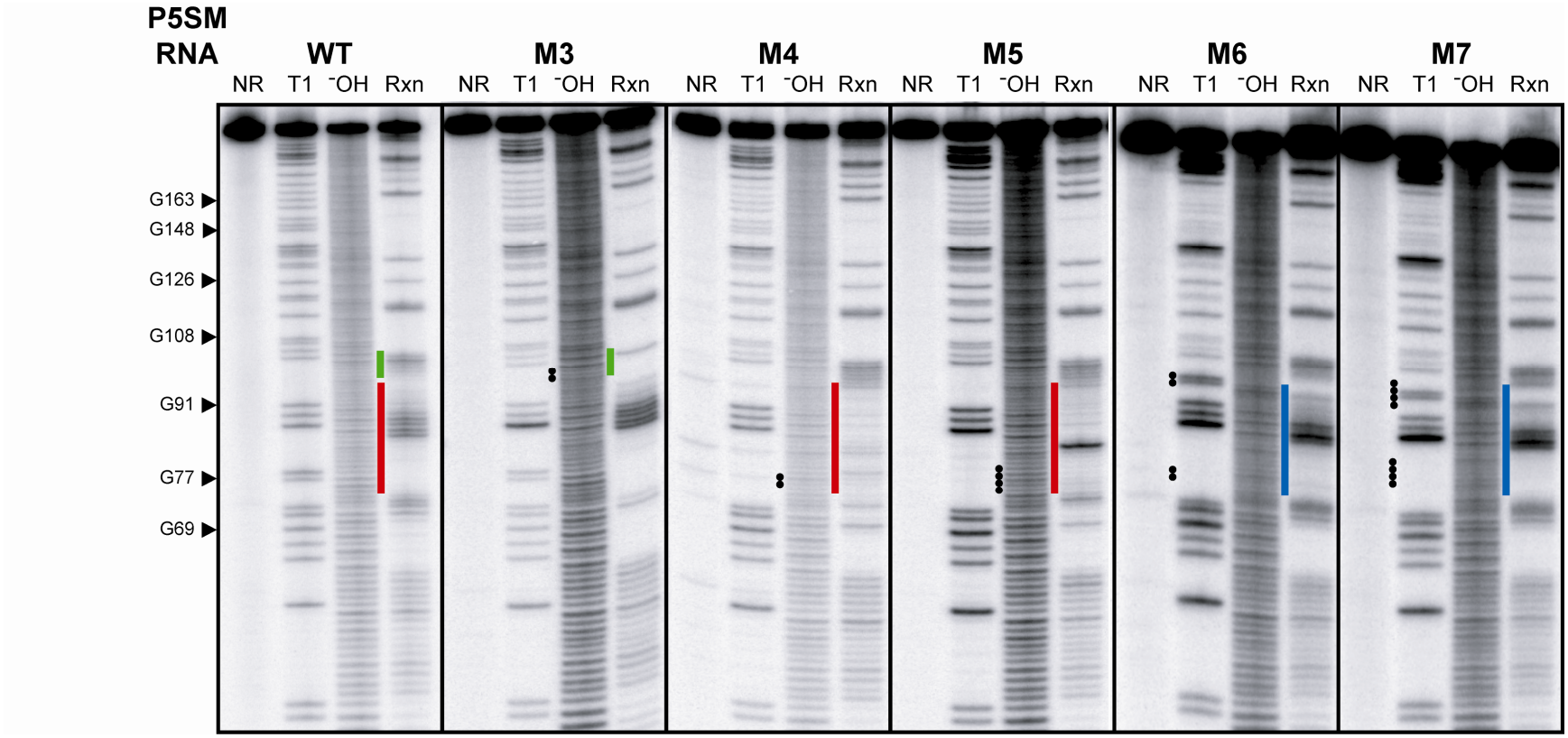


(**a**) *In vitro* binding analysis for 5S rRNA with GST-*At*L5. Sequence used is the major 5S transcript identified in *A. thaliana*[8]. Radiolabeled RNA was incubated at 25ºC in binding buffer in the absence or presence of protein (0-0.1 μM). RNA-protein complex formation was analyzed by non-denaturing PAGE. Unbound RNA (filled arrowhead) and RNA-protein complexes (open arrowhead) are indicated. Similar to previous data[9], two bands corresponding to RNA-protein complexes are observed which have similar binding characteristics. This suggests that the RNA fold may be heterogenous but still bind the protein. Additionally, ~50% of the RNA remains unbound at saturating concentrations of protein, even with optimized renaturation protocols.
(**b**) Representative plot used to determine the apparent K_D for the interaction between 5S rRNA and GST-*At*L5 protein. Maximal binding observed at 0.5 μM protein was normalized to 1. Graphed line corresponds to the best-fit curve for a two-state binding model with 1:1 stoichiometry and K_D of 9 nM. Analysis of either of the two RNP signals or a summation of both give similar results for the apparent K_D.

Supplementary Figure 7. In-line probing of *A. thaliana* P5SM RNA mutants



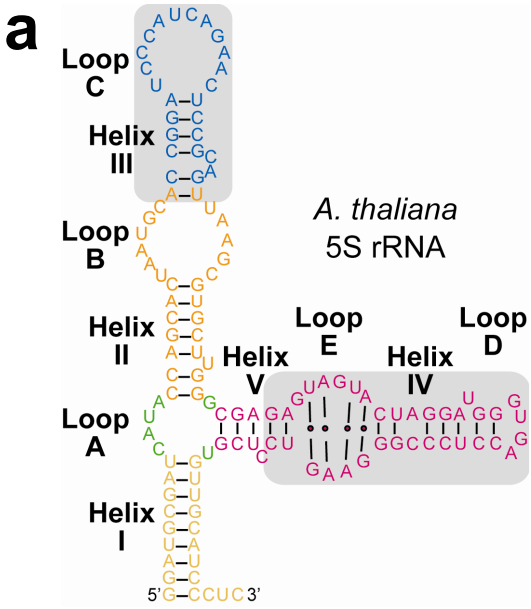
Comparison of in-line patterns between wild-type P5SM RNA (nucleotides 603-810 of *At*TFIIIA gene, NCBI gi 42592260) and truncated constructs show that deletion of the designated regions (M1 and M2, **Fig. 5a**) of the RNA do not perturb the other parts of the RNA. A white arrowhead indicates the location of each truncation. The unperturbed portions of the RNAs remain well-folded and have similar patterns of spontaneous cleavage compared to WT except missing the indicated deletion region.

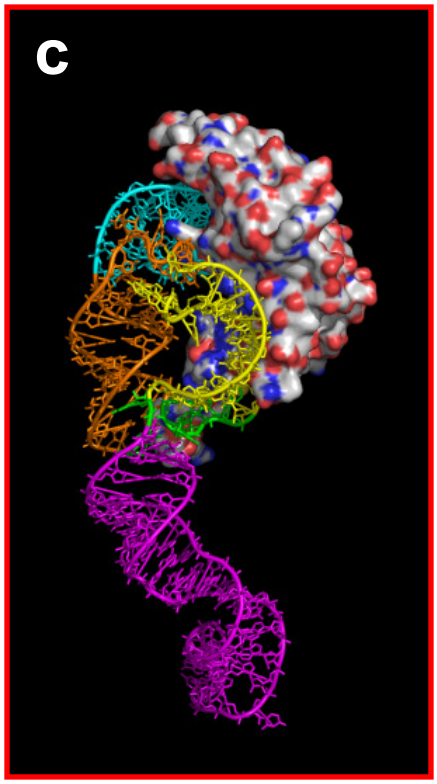


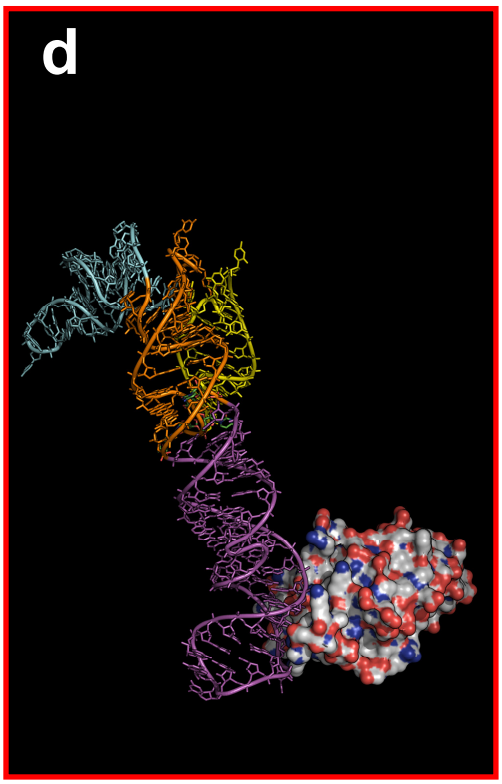
Comparison of in-line patterns between wild-type P5SM RNA and mutant constructs with nucleotide changes in the P2 stem (M3 through M7, **Fig. 5a**) show expected changes in structure confined to the P2 stem. Positions of labeled G nucleotides are the same as shown for the shorter construct in **Fig. 1b**. Nucleotide deletions or substitutions are mapped by black dots. Colored bars indicate regions for comparison between WT and a corresponding mutant.

Deletion of the dinucleotide bulge (M3) results in a more stable hairpin structure, indicated by reduced spontaneous phosphoester cleavage in the stem region surrounding the deleted AA (green bars). In contrast, nucleotide substitutions that disrupt base pairing (M4, M5) destabilize the normal P2 stem structure, indicated by including increased spontaneous cleavage overall in the stem region and alterations in the cleavage pattern (red bars). Finally, nucleotide substitutions that compensate for base pairing (M6, M7) restore very similar in-line patterns to WT (blue bars).

Supplementary Figure 8. The interaction of 5S rRNA with ribosomal proteins L18 (L5 homolog) and L30 (L7 homolog) in the *Haloarcula marismortui* large ribosomal subunit





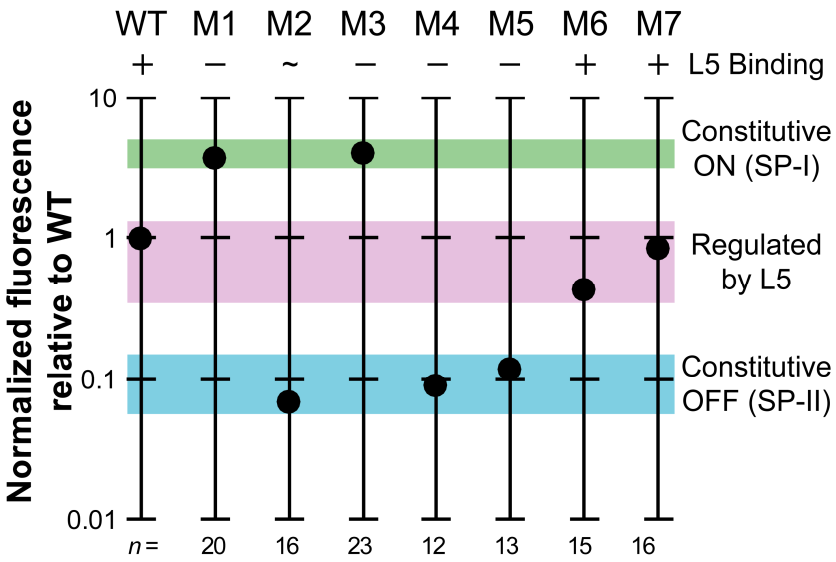


Atomic coordinates used to make figures are from the crystal structure of the large ribosomal subunit (pdb accession code 1QVG, chain 9 = 5S rRNA, chain M = L18P, the homolog to L5, chain V = L30P, the homolog to L7)[10].

(**a**) Conventional 5S rRNA structure model colored in accordance with panels and with regions homologous to P5SM boxed in grey; (**b**) view highlighting the 5S rRNA helix III interaction with L5; (**c**) second view highlighting the 5S rRNA helix I interaction with L5 and showing that helix II is not in contact with the protein; (**d**) third view highlighting the 5S rRNA loop E interaction with L7.

P5SM does not have a region strongly homologous to helix I, which may account for lower affinity in the P5SM-L5 interaction. Other structural features, such as the conserved but distinct P1 (**Fig. 1c**), may partially substitute for helix I, whose length has been shown to be important for complex formation in yeast[11], but allow L5 to bind P5SM in the context of a long pre-mRNA.

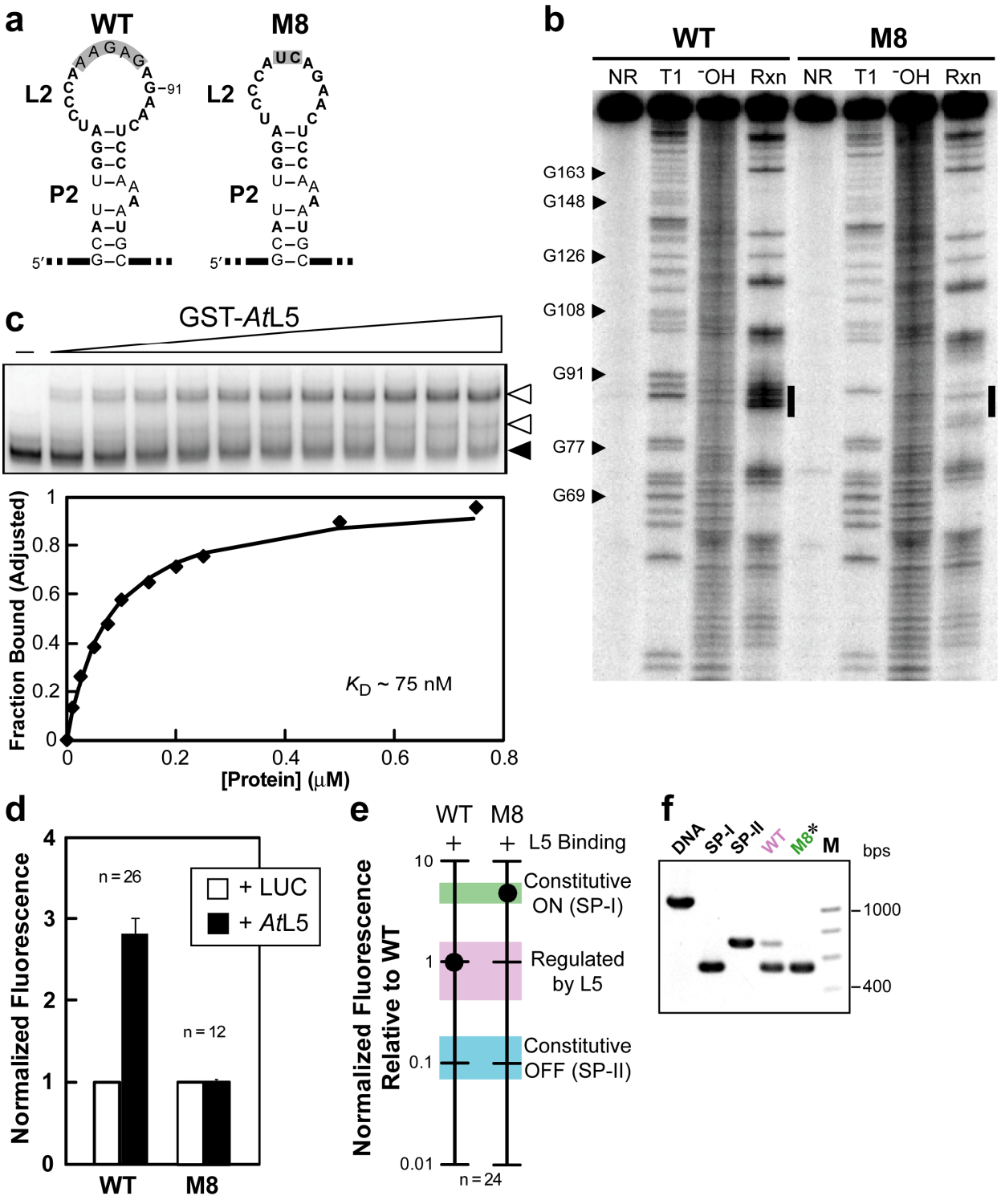
Supplementary Figure 9. The relative protein expression of P5SM mutant reporters reflects their splicing patterns



Reporter fluorescence for mutants normalized to WT Pre-EGFP are shown on a semi-log plot. WT and each mutant reporter construct was transformed on half of the same leaf to ensure near identical conditions for comparison. Effects from varying endogenous L5 levels were minimized by measuring reporter fluorescences upon constitutive *At*L5 expression. Thus, the data from this figure corresponds to the filled bars in **Fig. 5d**, except that the former is normalized to WT + *At*L5 and the latter is normalized to + LUC for each construct separately. Numbers of independent leaf samples (*n*) measured are shown. Error bars representing SEM are within the diameter of the symbols.

The observed protein expression relative to WT corresponds to the three splicing pattern types labeled in the different colors, which are also shown for a representative leaf sample in **Fig. 5d**.

Supplementary Figure 10. Replacement of the purine-rich loop sequence in L2 causes constitutive exon skipping without loss of L5 binding



(a) Partial sequence of the L2 mutant (M8) compared to wild-type (WT) P5SM RNA with the altered nucleotides shaded. Bolded nucleotides are identical to the *A. thaliana* 5S rRNA sequence.

(b) Comparison of RNA cleavage patterns from in-line probing between WT and M8 P5SM constructs show that changes in structure are confined to L2. The labeled G nucleotides correspond to the same positions labeled for the shorter construct in **Fig. 1b**. Black bars indicate the location of nucleotide changes. Note that the pattern for M8 is shifted down from the adjacent pattern for WT above this region because these 5′ cleavage products include the deletion, which shortens the length of M8 relative to WT.

(c) *In vitro* binding analysis for the M8 P5SM RNA with GST-*At*L5. The same methods were used as described in **Fig. 4**. As observed for 5S rRNA in **Supplementary Fig. 6**, two bands corresponding to RNA-protein complexes are observed, suggesting that the RNA fold may be heterogenous but still binds the protein. Also shown is a representative plot used to determine the apparent K_D for the interaction between M8 P5SM RNA and GST-*At*L5 protein. Maximal binding observed at 1 μM was normalized to 1. Graphed line corresponds to the best-fit curve for a two-state binding model with 1:1 stoichiometry and K_D of 75 nM.

(d) *In vivo* expression analysis for the Pre-EGFP reporter construct incorporating the M8 mutation in P5SM with co-expression of AtL5. Data for the Pre-EGFP WT construct (from **Fig. 3c**) is shown for reference. For each construct, the EGFP fluorescence measured with expression of luciferase (LUC) was set to a value of 1. Number of independent leaf samples (*n*) measured are shown. Error bars represent SEM.

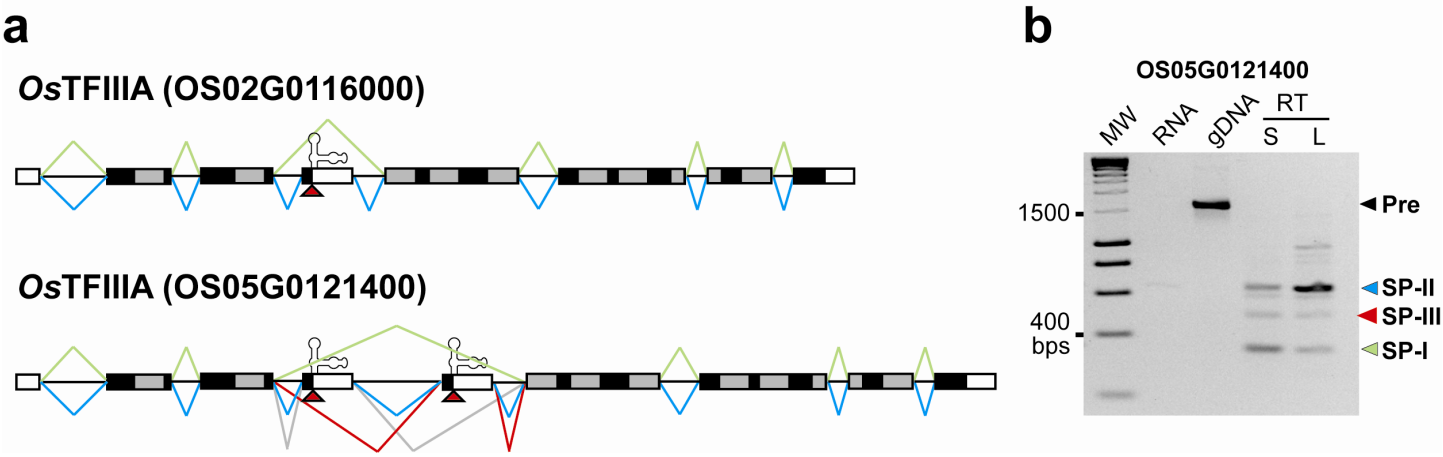
(e) Reporter fluorescence for the M8 mutant normalized to WT Pre-EGFP shown on a semi-log plot. WT and M8 were transformed on half of the same leaf to ensure near identical conditions for comparison. Effects from varying endogenous L5 levels were minimized by measuring reporter fluorescences upon constitutive *At*L5 expression. Number of independent leaf samples (*n*) measured are shown. Error bars representing SEM are within the diameter of the symbols.

(f) RT-PCR detection of splice products arising from splicing of WT versus M8* Pre-EGFP reporter constructs. The M8* construct is identical in sequence to the M8 construct analyzed in **d** and **e**, except for two single-nucleotide mutations in the constitutive intronic region. These additional mutations do not affect reporter splicing, as the protein expression results for M8* were the same as for M8. This set consists of WT and M8* from a representative leaf sample, in which *At*L5 was over-expressed. The color coding corresponds to the splicing pattern types labelled in **e**. Also shown are PCR products corresponding to unspliced precursor, SP-IE, and SP-IIE derived from DNA templates.

Based upon a displacement model for splicing regulation by ribosomal protein L5, the purine-rich insertion in the L2 loop of P5SM was postulated to be an exon splicing enhancer (ESE) that binds a splice factor to favor exon definition and splicing to SP-II. To test its proposed role as an ESE with minimal perturbations to other functions of the RNA, the five purine nucleotides were replaced with the UC sequence from Loop C of 5S rRNA. As expected, the replacement mutant M8 still binds L5, with comparable affinity as the WT P5SM and changes to the structure are confined to the L2 region. However, unlike all other previously tested P5SM reporter constructs that bound L5 (WT, M6, M7), loss of the purine-rich sequence causes loss of L5 activation and constitutive splicing to SP-I, as observed by reporter fluorescence and RT-PCR. These results support a role for this sequence in exon definition, possibly as an ESE that recruits an exonic splice factor.

Supplementary Figure 11. The two TFIIIA genes in *O. sativa*, *Zea mays*, and *Triticum aestivum* may be differentially regulated by a single or tandem arrangement of the P5SM element

(**a**) Two TFIIIA genes (OS02G0116000 and OS05G0121400) were identified in the *O. sativa* genome, shown is the comparison of the annotated splicing models; (**b**) RT-PCR analysis of OS05G0121400, the rice TFIIIA gene containing the tandem P5SM arrangement detected three main splice types (SP-I, SP-II, and SP-III) but not the fourth predicted splice type (splice reaction in grey); see Figure S4 for sequences of the gene, the translated protein, and the splice products detected by RT-PCR for OS02G0116000 and OS05G0121400; (**c**) sequences of SP-II type transcripts for the homologous two TFIIIA genes in *Zea mays* and *Triticum aestivum* (from EST or cDNA data). Cassette exon is highlighted in blue; P5SM sequences are underlined.



c

Tae_20437489 (corresponds to SP-II type transcript)
Contains single P5SM

```
   1 CCACGCGTCC GCCCACGCGT CCGCACGAGG GAGCGGCGTC ATCCCCAGCT CCGTTCCCCA
  61 CGGCCGCACC CTCCGCCTCC GCCTCCGAAC CCACGATGTC TTCTGGAGAT GGGATCGATG
 121 GAGATGAGAA GTCTGAAGAG ACACATGGCA AAGATATTAG ACGCATCAAG TGTGAATTTT
 181 GCACTGTTGT TAGGTCCAAG ATGTATCTCA TACGAGCTCA CATGGTAGCT CAGCACAAGG
 241 ATGAGTTGGA CGCATCAGAA ATCTATGACT CAAATGGTGA AAAGGTTGTT TATGGGGTTG
 301 GACACACATG TGAAGAGTGT GGAGCTTGTT TCCGGAAGCC AGCTCATCTG AAGCAGCATA
 361 TGCAAAGTCA TTCCAAAGAG GGTTTAGATA ATAGATACTT AACAGCGTGA AGCATTGGAT
 421 CCCATTACAG AACTCCAAAA TGCATGGAGG CACATTTTAA TCTTGGTTGG ACTAGTAACA
 481 GGTTGGGATG ACCTCCTGTG AAGCTCCAAC AAGATTGCCT CCTTCACGCT TACTCAGGAG
 541 AGATCCTTCG CCTGCCCACT TGAAGACTGC CCCTTCAGCT ATATAAGAAA AGATCACTTG
 601 AACCGGCATA TGCTTACACA TGAGGGCAAG TTGTTTACAT GCCCTC
```

Tae_141663088 (corresponds to SP-II type transcript)
Contains tandem P5SM

```
   0 GCAGCGTCGT CGGAGCGACG GAGACCTCGA CCTCGCCGGC GGCCGCGGCG
  50 GCCGCCCCTG TGAGGGACAT CAGGCGGTAC AAGTGCGAGT TCTGCGACGT
 100 CGTGCGCTCG AAGAAGCGGC TGATCCGAGA CCACGTCCTC GAGCACCATA
 150 AGGACGAAGT GGATGGTCTG AATGAGTACA ACGTAGGTGG TGGTGGCGGC
 200 AGTGCGCCGC CGGGCAAGGA GATCGGCCAT GATTGCAAGG AGTGCGGCGC
 250 GAGGTTTAAG AAGCCGGCGC ATCTGAAGCA GCATATGCAG AGCCATTCGC
 300 CCGAGGCTGT TTAATTGTGT GAAGCATTGG ATCCCAAAGA ACTCCAAAAT
 350 GCGAGAGGGC CCTGAATGGA CTAGTACCAG GATGGGGAGA CCCTCCTGGG
 400 AAGCTCCTTC AGGAAGCCTC TTTCACGCCC TTCTTCAGAC TACATGTTCA
 450 GATATGAATG TGTGAAGCAT TGGATCCCAA GAACTCCAAA ATGCCAGAAG
 500 GCCTTGAATG GACCAGTATC AGGAAGGGGA GACCCTCCTG AGAAGCTCCA
 550 TCAAGAATGC CTCATTCACG CGCTTCCTCA GAGACCCTTT GCCTGCCATG
 600 TAGATGGTTG CCCTTTCAGC TATAGCAGGA AGGACCATTT GAACCGCCAT
 650 CTGCTTACTC ATCAAGGAAA ACTATTCATG TGCCCTATGG AAGGATGCAA
```

```
       700 CCGTAAGTTC AGTATCAAGG GTAATATCCA GAGACATGTT GAGGAATTTC
       750 ACGAGGATGG CCCTCGTTGT GGAGGCAAGA AAGAGTTCAT CTGCCCAGAG
       800 GCTAACTGTG GGAAG
```

Zma_157151872 (Chr 8 contig)
Contains single P5SM

Chr 8 Contig 27601..37898
Exons are based upon manual alignment of available ESTs to genomic sequence
Sequences highlighted in red are inconclusive from available ESTs

```
     34021 TGTCAGATGT GTTCTGAAGT TCATGTTGAG GGAGATGCAA GTGTCGGACA GAAGGGTTGC
     34081 AAGGATATAA GGCGTTACAA ATGTGAATTT TGCACAGTTG TCAGGTCAAA GAAATGCCTA
     34141 ATTCAAGCTC ACATGGTTGC ACATCATAAG GTATAAACAT TTTGTGATAT ATTGACACAT
     34201 TAGTCGATAT ATATCCTGTT TGATTGCATA GCTATGCACC ATCTATATTA TGACACATTA
     34261 TTTACCTTTT GATGTCTGCC TTCATATTAT TGTTGTTTTG GAGCAATACA TAGCATCTCT
     34321 AATGTTTTCT GAAAATTATC TGCAGGATGA ACTTGATAAA TCAGAAATAT ACAACTCAGA
     34381 TGGGGAAAAG ATTGTTCATG AGCAGGAACA TAGGTGCCAA GAGTGTGGCC GTTGTTTCCA
     34441 GAAACCAGCT CATCTGAAGC AGCACATGCA AAGTCACTCA CATGAGGTTG GTATTCATCT
     34501 CCTCAAATTA AGCATGAGCT TGCTCTTTTT ATTTTGTTGG CACTGGTTGC CCCCAATTTC
     34561 ACTTTTGAAG TATACAGACA TACAGACATG ATGATTTGCA TTCATGACTT ATGAGGCTGG
     34621 TTGTAATAGT GATTTTCCCA CTATGATTTG CTAAATGCAG TTGTTTTGCA GGGCATAGAT
     34681 AGTTAACAGC GTGAAGCATT GGATCCCATT TAAAGAACTC CAAAATGCGA CGAGGTATTT
     34741 TTTATTCTTG CCTGAACTAG TAACAGGTTG GGATGACCTC CTGTGAAGCT TCAACAAGAT
     34801 TGCCTCAGTC ACGCTCACTG AGTAGGTTTG TTCTTGTACA AAATTAGTTT TTTATCATGA
     34861 TGCTCATAGT CTAAATTCTG CATTTCATGT CATACATGAT TGTTTTCAGT GTTAGTATGT
     34921 ATTGTGCCTT CTCATATGGG CCAATTCAGT GAATTATGGA GTTTTTTTTA TTTCTTTGTT
     34981 CTTCTGACTT GGGGGCCTTT AGTTGTGTCA ATCTGGTGAT CAAACATCAA AATATTTCAG
     35041 AAAGCTTGGT CTTCACTTTA TAAGTTCCTT TATTACCTCT AATTTTCTTT AGTGCTGTTA
     35101 GGATTAGATG GGGTATGGAT CCTGGAGATA AGGCTGAAAA ATATCCTCCC ATCATTCTGC
     35161 CATGTCTCCA TGGGCTTCCT TATGGGAGCT CTGATGTAAT GCTTGAGGAT CCAATATCAA
     35221 AATGTCATCT CAGAAAGCTA GGCCCATTCT TCATTGGCTC TCAGCTCGCT ATGGTGACAA
     35281 TAGTCCTAAA ACCGTGTTAT TTTGCAATTA AAAAATGATG AAGTGAACTT GCTATAATCC
     35341 TGTAAGGATT ATATGGGTTA TGGATCTTTT GGAGATAAGG GTGGAAATAT CCTCCCTTCA
     35401 TTCTGCCATG TCATAAAGGG CTTCCTAATG GGAGCTTCCA GAAGATGCTT TAGGATCCAA
     35461 TATCAAAATG TCGTCTCAGA AAGCTGTGGC CATTCTGCAT TGGTTCTCAG CTCGCTATGG
     35521 TGACAATAGT CCTGTATCGT CTCATTATGC AAGAACAGAT GTACTGAATG GAAGTGCTGA
     35581 ACTTGTTATA GTTTGTTGAG AAGGCTTTAG CATTTAGTTT TTCTTTATTA TAGATTCATA
     35641 GTTCAAAAGT AGATGCATTT ACCTTATAAT TTTAGCAGGT TTTTTTCCCA TTATATGCGT
     35701 CATAACATTT CTCCTAAAAC ACACGTTTAA CAAGTTGGCA CATTTGTTTA GTTATTAGGT
     35761 AAAAATTGTG TCACTTTTGT AGGTCCTTAT CATGATCCAG TACAGTTTAT AATTCTGAGT
     35821 TCTGATAATC GTTTGGTTAA TTATTTGAAG CATTTAGCAC AATGTTAAAC AAAGAGCATT
     35881 ATAAATATTT CGGTCGTAAC ACATATCCTT GTTTGACTGT AGAGACTCTT TATTTGCCCA
     35941 CTGGAAGACT GCCCCTTCAG CTATAAAAGA AAAGATCACT TGAACCGACA CATGCTTAAG
     36001 CATGAGGGCA AATTGTTTAG TTGCACCCTG GATGGTTGTG ATATGAGGTT CAGTATGAAG
     36061 GCAGATATGC AACGGCATGT TAAGGAAATC CATGATAGTA AGAACCCCCA TCAATTTGTT
     36121 TGTAAGGAGG ACGGCTGTAA CAAGGTTTTC AGGTATTCAT CAAAGCTGAA GAAACATGAG
     36181 GAATCTCATG GTAGGTCAGC ATTAACCTGA TAATTAGGAA ATTAATAAAT GCAGAGTGTT
     36241 CATATTACTC ACATCATGTT CATGTACTAG TACAAGTTTT TCATTTGTCC AGCATATTGG
     36301 TTCTCCAAAT TTGAGTTTGT TTCCTTGCAG TTAAATTGGA CTATGTGGAG GTAGTGTGCT
     36361 GTGAGCCAGG TTGCATGAAG ATGTTCAGAA ATGTTGAATA TCTGAGGGCT CATAACCAAT
     36421 CTTGCCATCA GTATGTTCAG TGTGAGATAT GTGGAGAAAA GCACCTGAAG AAGAACATCA
     36481 AGCGGCATCT GCAATCACAT GACAAGGTGC CCTCTGATGG AAGGATGAAA TGCACTTTTG
     36541 ATGGTTGTGA ACACTTCTTT TCCAACGTAA GCAAGATTGG TAGATTTCTG AATGCGCCTA
     36601 TTCAGTTTTT CCAATGATTT CTTAAACCAA GCGTATTCCT ATTTTCTATT GCAGAAATCA
     36661 AATCTGACCA AGCATGTCAA GGCATGCCAT GACCAGCTGA AACCTTTCAA GTGCGGAATC
     36721 GTTGGGTGTG GCAAGGCGTT CACCTACAAG CATGTCAGGG ATAACCATGA GACATCTGGT
     36781 GCTCATGTAT ACATGGAAGT GAGTTGTGCT ACCTGATCTA TGTTAATTCC GAGTTTGTGT
     36841 TTTTTGGCCT ATTAGGTCAA ACTATCAGCT CCTGATTTTC GTACGTATCT CTAATTTAAA
     36901 CGTATGTTCA CATCTGCTTT GTGATCAAAC CGTGTGAACC TGTGGATAAC AAAAATCTGC
     36961 TTCATGCAGG GAGACTTTGA AGAGTTGGAT GAGCAGTTGC GTGCCCGGCC AAGAGGTGGT
     37021 CGGAAGAGGA AAGCATTGAC GGTGGAAACA CTGACACGCA AGAGGGTGAC CATCTCCGGC
     37081 GAGGCTTCAT CCTTGGATGA TGGAGTGGAG TATCTTCGCT GGTTGCTATC TGGTGGGGAT
```

```
    37141 GATTCAATTG ATTGTCAGTA GCCGGCGTCT TTTTGCTATC ACCGCTGCTT TGTACATGGG
    37201 TTAGATTATG TAATGTAAGT TGTTTGGTAG ACTACGATGG TGTTCATGAC GCGTTTGATA
    37261 TATTGTCAGC AGTATGGTAG TGCAGGGCTC TCATAGTAAA GCTTGCCATT TTTGTTGTAT
    37321 CTTTAACCCA CGTGGGTCCA GATGATGTGT TTATTCCTAT GTATTTACTT ATCTACATGT
    37381 ATCCACATGT GCGACTTTGG CGTTTGTTTG AGATTATAAT CTGTTAAAAT ATGTTGGATT
    37441 ATATAATTTG GGCAGATTAT AATTCCAAAC AAACACTGAC TAATGCTCTA TTTGTTTACC
```

Zma_126352889 (Chr 6 contig)
Contains tandem P5SM

Chr 6 Contig 74014..86960
Exons are based upon manual alignment of available ESTs to genomic sequence

```
    76021 TCGTTTATTT CGCTGTGGTT TTGGGGTTAG TCAGACTCAG ATGGGGGAAG GCACAGAGAG
    76081 CGGCGGCGGG GCGACCACAG CGCCGCCTCT GAGGTATATA AGGTGCTACA AGTGCGGGTT
    76141 CTGCGACGTG GTCCGGTCCA AGAAGTGCTT GCTCCGCGCC CACGTGCTTG AGCACCACAA
    76201 GGTACCAGCT GTGCTGCTCG CTTGCTCGCC GCCGCTTGGT TCAGACTTCA GATTTGCGCG
    76261 CGTTCTCCTT TGTATTGTTC TACAGGGGTA ACACTTATGG GGTGTGTGTA GGATGAGGTG
    76321 GATGCTTTGG GGGGTTACTG GGAAGGTGGA GATGCTGGTC CGCGCAAAGA GATCAGCCGT
    76381 GCGTGCGAAC AGTGCGGGAT GAGCTTCAAG AAGCAGGCGC ATCTGAAGCA ACATATGCAG
    76441 AGCCATTCGC TCGAGGTACT GATCTGGCTG CTTATGTTGG GTTTTTTGGT ATATTGCTCA
    76501 GTGAATGCTC AGGCTACTAG TGCTCGCCTG GTTTGTTACT ATCCGCTTAA TGTTTATTGT
    76561 GTGAGAAAAG GCCGAAAGCG AGATACATGC TGGTTTGACT CTTTTGGTGT GGTTGCTACA
    76621 ACCCTGTTAT AATAACAATT GGTTCGATGG ATGACTCTGT ATACCTCATG GTCTTTCTTT
    76681 AGTCCATTGT ATGTTATTAT TAATTATGCT GACTTTTGCA GTCTGTTTGT TCAGATACAG
    76741 ATGTGTGAAG CATTGGATCT CAAAGAACTC CAAAATGCGA GAAGGCCTTG AACGGACTAG
    76801 TACCAGGATG GGGAGACCCT CCTGAGAAGC TCCATCAAGA GTTCCTTCAT TCACGCTATT
    76861 CTTCAGGTTG TGTGTTCTGA GCTATTCTGA GCTGCTTGCT GCTTTTTTTG TTATTATGTT
    76921 CCTGTCACAT GACATATGCT AAAGGTCTAG TCTGAGATTG CGATGCCACT TTTTTCAACA
    76981 AGTTGGGCGT CTGCTGTTCT AGTCATATTT GATAATTGAT GAATTTACAC TGCAATGCTG
    77041 TTTGAGTATA TCTAGCTGTC ACCGTTTGAT GGCCCTCATT ATTGACAGGA TAGATTTAGC
    77101 AGGCATAGAG AGACACAATT CCTGTTCTGC ATAGTTTATG TGCTCGACTT TAGTTATGTG
    77161 TTCACTGAAA TGATTGGATC ATGATCTTCC TTAGTTGAAT TATACATTTC ATGCAAAAAT
    77221 CAAATTAATT TTCATTTGAA GGTGAGTTAT GCGGATACGG TGTGCTGCTA TTTGCCCCTG
    77281 TATTTGGGGA CTGGTGTACG ACATTTGATT TGCTGGTTTT TGTAGGCAAT TTGTTGACAT
    77341 ATGAATGCGT GAAGCATTGG ATCCCAAAGA ACTCCAAAAT GCGATAAGGC CTTGAATGGA
    77401 CTAGTATCAG GATTGGGAGA CCATCCTGAG AAGCTCCATC AAGCATGCCT TGTTCACGCG
    77461 TTTCTTAAGG TTTGCTCATT TCGTTCCAGC CTTCTCATTC TAAGATTTTT CCTTTTATGT
    77521 TATTATTTAC TTGCACTTTT ATAATTGTAT TTTCTATTAT TTTCACACCA CTTAAAGCTC
    77581 ATGTTGTAGC ATGCCTGGTT GCACTTTGTG GTACTGTTTG CATCTCTATT ACTTTATTTG
    77641 TATCAAACCT TTTTTTAGGT ACTATATAGT GATACATGTC TTTCCAAAAC TACTGCAGAG
    77701 GCCCTTTTCC TGCACTGTTG ATGGTTGCCC CTTTAGCTAT AGCAGGAAGG ATCATTTGAA
    77761 CAGACATTTA CTTACTCATG AAGGGAAACT ATTTGTGTGC CCTGTCGAAG GATGTGGCCG
    77821 TAAGTTCAAT ATCAAGGGTA ATATGCAGAG ACATGTTCAG GAAATCCACA AAGATGTCTC
    77881 TCCTTGTGAA AGCAAGAAAG AATTCATCTG TCCAAAGGTT AACTGTGGGA AGGCTTTCAA
    77941 ATATGCTTCT AAGTTAAAGA AGCACGAAGA ATCACATGGT GAGAGGCACA CTTTATAACC
    78001 ATTTCCATAT TTCTGACACC AAGTTTAAAT AACACACTTT GAAAAGCCTG TTGATATAGT
    78061 ATATTGATGA TGTTAGCATG TAGTTCTGGT CTTGTTATCC TCTGAGCCAT ATTTATCAAT
    78121 ATGAACTCTC ACCTTTACCA ATTGTTTTTG TCTTTCTGAT ATGATTATTT TATGTGTTCT
    78181 GCCTTAGTGC ATGCTTTTCA TCTATTCTGC ATAATATTAT GTATGCAAGT GCTTATTATC
    78241 AACACTGACA GTAGTTTCTC TGTCATGATA AACTTTTGGT ACCTGTGGTG CAGTCCAGCT
    78301 GGAATACACA GAAGTTATGT GCTGTGAGCC AGGCTGCATG AAGTTCTTTT CAAACATGGA
    78361 ATGCCTGAAG GCGCATAACC AATCTTGCCA TCAATATGTT CAGTGTGATA TCTGTGGTAC
    78421 GAAACAGCTA AAGAAGAATT TCCAGCGCCA TCATCGGATG CATGAAGGTT CCTGCGTCAC
    78481 TGAGAGGGTT AAATGCCACA TCGAGGACTG CAAATGTTCG TTCTCGAAGG TTGTGGTTTCT
    78541 AATCTTAACA CTATTTGATT CTCTTATGGT TTGTTTCCGT GGAATTTATC AAGAAAAGCA
    78601 CTCAGTTATT CATTCTTGGT GAACATGAAA CTGCAGAAAT CCAATTTGGA CAAGCATGTT
    78661 AAGGCGGTAC ATGAGCAGCG TCGTCCTTTT GTGTGCCAAT TCTCTGGGTG TGGCAAGAGG
    78721 TTTTCTTACA AGCACGTAAG GGACAATCAT GAGAAGTCAA GCGCTCATGT GCACACTGAG
    78781 GTATCACAAA TGATCACCTG ATAACAATTT CCTTGGCATC CTACATTGTG CTTTCTCCTT
    78841 CAGAAATGCA TCTTGTGAGG GGGATGTTTA TGATTTCCCA TCAATTGTTT GCAGGGTGAT
    78901 TTTGTCGAGG CTGATGAGCA GCGGCCACGT TCAGTAGGTG GGTGCAAGAG GAAACCCGTA
    78961 TCTGTTGAGA GTTTGATGCG TAAGAGGGTA GCCGCTCCTG ACGATGGGCC TGCTCATGCT
```

```
79021 GATGCAACTG AATATTTGAG ATGGCTTCTG TCGGTTTGAT TGGACGTTCA AGTGCACTTT
79081 GTGTAGTTGG AAACATCAGC GATTTACTTG GCATGTATAG GAAGATGAAT GTCTCTAGTA
79141 TGTTATGTAA ATGTTTATGT ACGGTATCAG AAATACAGAT CATTGGACCT CAATTATAAA
79201 GGCAATATGC TGTATTGGCA AGATCTTGAA TGTAAGAGCG GTTGCTCATG ATCATTAGAT
79261 CTATGTGTCT TCAGCACAGC TTGGCAACAT GTTTTGGTAT GTTCGCAGTT GCAGTGTCTT
79321 TGCCTTTTTT TGCCTCCTGC TGGTTGCAGG TGAAAATTGG ATTCATGATT CTTGAATGCT
79381 CTTATGTAAC AACAAAAATA TCTGTTCAGT TCTCATTGTG CAGAACCACA GCACTCTTCA
79441 TTCCGATTTG CAGATGCACG TGACCAAAAA AGATAATTTG CAGATGCACC ATGAACCCGT
79501 GGCACCAGAT GCCCGTCAGC AGCTGGCACC CGTCCAGCTC GCGCCGGAGA GCCTCCAGGC
79561 GGACCAGCTC CGCCGCCCCG CGGATCGATT GGACTAGCGA CGAGGCCGAG ATGCGCACGG
79621 ACGTCATCGG CGGCGGCGAT GGCTCTTCTG CCTCGGAACT CGGATGCCTC AGTGTGCGTG
79681 CGCGTGTCAG TGTGATCTCT TGCCTCCAAT GGATCGCGTG CTTCCTTTCT ACGGCGATGA
79741 TGGCTTAGTT AAAGCCCAGT AGTTTGCAGT AGTGAGCTCA GAGCCCAATA AACTCCAGTA
79801 ATGGGCTCCG AAGATTAAAA GGCCGTTAGC AAATTAATCA CTTCATTTGG ATGGCATGGG
79861 GATCCTCCCA AAAAAATTGA ATGGGCTGAG AAACGCTTGT TGCATGGACC TTCGACAGCC
79921 CCAGGACGTT TCTCATGAAA TTTCTATTAT ATGAAAGGCG GTTCAAAACG ACTCCATTCG
79981 TGGTGTCTAA GTCGTTCTTT TTATTTAGAC CGTTCGATCC CATCTCAATG GTATGCATCG
```

SUPPLEMENTARY TABLE 1. Sequences of DNA primers

DNA templates for *in vitro* transcription		
	A. thaliana P5SM, minimal (nucleotides 649-793 of AT1G72050)	
DNA1	5′-*TAATACGACTCACTATA*GGTCTTACCTATGTTATTGCG	For, T7 promoter
DNA2	5′-CTCAAACATTTTGCGTGATCG	Rev
	A. thaliana P5SM RNA (nucleotides 603-810 of AT1G72050)	
DNA3	5′-*TAATACGACTCACTATAG*GTTTATACATTTGCAGGAAATAA	For, T7 promoter
DNA4	5′-GCTATTGAACATCAGACCTC	Rev
	O. sativa P5SM RNA_1 (nucleotides 903-1109 of Os05g0121400)	
DNA5	5′-*TAATACGACTCACTATA*GGTTGGGTAGGTTTGTTGATCTA	For, T7 promoter
DNA6	5′-GCAACAGAAGAAATCACCTGAAG	Rev
	O. sativa P5SM RNA_2 (nucleotides 1500-1693 of Os05g0121400)	
DNA7	5′-*TAATACGACTCACTATA*GGCCTATTGTGGTTTCTGATATATT	For, T7 promoter
DNA8	5′-TTTGAAATGAACAAGCAAACCTG	Rev
	A. thaliana 5S gene (CIC YAC 6A1, [8])	
DNA9	5′-CGTGATTTGGGCTATATTACG	For
DNA10	5′-CAGTCTACAAGTTATCGAGTCATA	Rev
	A. thaliana 5S rRNA	
DNA11	5′-*TAATACGACTCACTATA*GGATGCGATCATACCAGC	For, T7 promoter
DNA12	5′-GAGGGATGCAACACGAG	Rev
RT-PCR analysis		
	TFIIIA (AT1G72050) from *A. thaliana*	
a (DNA13)	5´-atgc**ggatcc**GTGCGGCGTCTTGATGGA	For, BamHI
b (DNA14)	5´-ACTCTTGCAGTTCTCCTTCG	Rev
e (DNA15)	5´-GGCACGGGCAGCTTACCGGTGGTGCATATGAACTTCAGGGT	Rev (EGFP)
	TFIIIA (Os02G0116000) from *O. sativa*	
c (DNA16)	5´-TGATGGAGACATGAGGGTTG	For
d (DNA17)	5´-TGTTACAGCCCTCCTCCTTG	Rev
	TFIIIA (Os05G0121400) from *O. sativa*	
DNA18	5′-ATGATTACTTGGGACGTGGT	For
DNA19	5′-GTCTTCCCACAGTTTTCCTC	Rev
qRT-PCR analysis		
	TFIIIA transcripts retaining exon (SP-II and unspliced*)	
DNA20	5′-TTATTGCGTGAGGCATTGGA	For
DNA21	5′-TCTCAGGAGGTCTCCCAACCT	Rev
	TFIIIA transcripts skipping exon (SP-I)	
DNA22	5′-TGTCAAGAATGTGGTGCTGA	For
DNA23	5′-GTAAAAGATCTCTCGAGCGAATG	Rev
	DsRED transcripts (reference)	
DNA24	5′-AGACCCACAAGGCCCTGAA	For
DNA25	5′-CAGCTGCACGGGCTTCTT	Rev
Cloning of reporter constructs		
	5' fragment of TFIIIA (Pre-EGFP, I-EGFP, II-EGFP)	
DNA26	5′-atgc**ggtacc**GTGCGGCGTCTTGATGGA	For, KpnI
DNA27	5′-agct**tctaga**ATCCACATAGCAAGTAAAAGA	Rev, XbaI
	EGFP without start codon	

DNA28	5′-agct**tctaga**GTGAGCAAGGGCGAGGA	For, XbaI
DNA29	5′-agct**gtcgac**TTACTTGTACAGCTCGTCCATGC	Rev, SalI
	Flag tag insertion in frame within N-terminal coding region	
DNA30	atgc**ctcgag**GACTACAAAGACGATGATGACAAG**ctcgag**atgc	For, XhoI, XhoI
DNA31	gcat**ctcgag**CTTGTCATCATCGTCTTTGTAGTC**ctcgag**gcat	Rev, XhoI, XhoI
Cloning of protein coding sequences		
	TFIIIA cDNA (AT1G72050) from *A. thaliana*	
DNA32	5′-gac**ggatcc**ATGGCGGAAGAAGCTAAAG	For, BamHI
DNA33	5′-gac**gtcgac**CTAGCAAGTTTCGTGTTCTTC	Rev, SalI
	L5 cDNA (AT3G25520) from *A. thaliana*	
DNA34	5′-gac**ggatcc**ATGGTGTTTGTGAAGTCCACC	For, BamHI**
DNA35	5′-gac**gtcgac**TAAAGAAGGCTTGACTGATTTACTCTTC	Rev, SalI
DNA36	5′-atg**cagatc**tATGGTGTTTGTGAAGTCCACC	For, BglII
	L7A cDNA (AT1G80750) from *A. thaliana*	
DNA37	5′-atgc**ggatcc**ATGGCTGAGGAAGAAGCTAA	For, BamHI
DNA38	5′-atgc**gtcgac**CTAATTCATTTTGCTGATGAGA	Rev, SalI
	L7B cDNA (AT2G01250) from *A. thaliana*	
DNA39	5′-atgc**ggatcc**ATGGTTGAGTCAAAGGTTGT	For, BamHI
DNA40	5′-atgc**gtcgac**CTAATTCATCCTCCTGATAAGC	Rev, SalI
	N-terminal Flag tag w/ start codon and BamHI overhangs	
DNA41	**gatcc**ATGGACTACAAAGACGATGATGACAAG**g**	For, BamHI, Flag
DNA42	**gatcc**CTTGTCATCATCGTCTTTGTAGTCCAT**g**	Rev, BamHI, Flag
	Primer containing N-terminal Flag tag w/ start codon and overlap 5' end of L5 cDNA (internal BamHI site prevents use of above primers)	
DNA43	atgc**agatct**ATGGACTACAAAGACGATGATGACAAG ATGGTGTTTGTGAAGTCC	For, BglII, Flag
P5SM mutant constructs by two-piece PCR ligation		
	M1 (template TOPO-DNA26/27 PCR)	
DNA44	5′-GTAGCCTCGCCTCACGCAATAACATAGG	5' seg, w/ DNA3
DNA45	5′-GCGTGAGGCGAGGCTACATGTTATGGAC	3' seg, w/ DNA4
	M2 (template TOPO-DNA26/27 PCR)	
DNA46	5′-GCTTACTTGCTATAACATGTAGCCTCGCATTTTGG	5' seg, w/ DNA3
DNA47	5′-GCTACATGTTATAGCAAGTAAGCCTCGATCACG	3' seg, w/ DNA4
	M3 (template TOPO-DNA26/27 PCR)	
DNA48	5′-CGCATTGGAGTTCTCTCTTTGGGATCC	5' seg, w/ DNA3
DNA49	5′-CTCCAATGCGAGGCTACATGTTATGGAC	3' seg, w/ DNA4
	M4 (template TOPO-DNA26/27 PCR)	
DNA50	5′-GGGATGGAATGCCTCACGCAATAACATAGG	5' seg, w/ DNA3
DNA51	5′-CATTCCATCCCAAAGAGAGAACTCCAAAATGC	3' seg, w/ DNA4
	M5 (template TOPO-DNA26/27 PCR)	
DNA52	5′-GGGAAGGTATGCCTCACGCAATAACATAGG	5' seg, w/ DNA3
DNA53	5′-CATACCTTCCCAAAGAGAGAACTCCAAAATGC	3' seg, w/ DNA4
	M6 (template TOPO-M4)	
DNA54	CATTTTCCAGTTCTCTCTTTGGGATGGAATGC	5' seg, w/ DNA3
DNA55	GAACTGGAAAATGCGAGGCTACATGTTATGG	3' seg, w/ DNA4
	M7 (template TOPO-M5)	
DNA56	CATTTACCTGTTCTCTCTTTGGGAAGGTATGC	5' seg, w/ DNA3

DNA57	GAACAGGTAAATGCGAGGCTACATGTTATGG	3’ seg, w/ DNA4
	M8 (template TOPO-DNA26/27 PCR)	
DNA58	GAGTTCTGATGGGATCCAATGCCTCACG	5’ seg, w/ DNA3
DNA59	GATCCCATCAGAACTCCAAAATGCGAGGC	5’ seg, w/ DNA4

Italicized sequence includes T7 promoter sequence for *in vitro* transcription; lowercase sequence includes overhang and restriction digest sites (bold) indicated

*The unspliced pre-mRNA is not observed by RT-PCR, and so should contribute negligibly to the transcript abundance measured by qRT-PCR.
**The L5 cDNA sequence contains a BamHI site at nucleotide position 104; a partial digest with BamHI could be performed and the larger DNA product isolated, or the restriction site could be switched to BglII (DNA34), which results in a digested end compatible with the BamHI cloning site.

SUPPLEMENTARY METHODS

Cloning of DNA constructs

Reporter constructs containing the 5' region of either the unspliced pre-mRNA (Pre-EGFP) or the two splice variants (I-EGFP and II-EGFP) of TFIIIA from *A. thaliana* fused to the cDNA of enhanced green fluorescent protein (EGFP) were cloned into the binary vector pBinAR[12]. The TFIIIA 5' regions starting from the 5' UTR and extending to the exon downstream of P5SM were PCR amplified with primers DNA26 and DNA27 from *A. thaliana* genomic DNA or cDNA, then the resulting DNA products were subjected to restriction digest with KpnI and XbaI. EGFP cDNA was PCR amplified with primers which skip the start codon of EGFP, and the resulting DNA product was subjected to restriction digest with XbaI and SalI. To generate in-frame fusion constructs, the different TFIIIA fragments were ligated with EGFP cDNA through the common XbaI ends. The resulting products were cloned into the KpnI/SalI sites of pBinAR. For insertion of a Flag tag into these reporter constructs, two complementary oligonucleotides encoding the Flag peptide sequence (DYKDDDDK) flanked by XhoI sites were annealed, subjected to restriction digest, and cloned in-frame into a natural XhoI site within the 5' part of the TFIIIA coding sequence (annotated in **Supplementary Fig. 4**).

For cloning of Pre-EGFP reporter constructs with mutations in P5SM, first the 5' region of wild-type TFIIIA gene was PCR amplified with primers DNA26 and DNA27. The resulting product was cloned into EcoRV linearized vector pBluescript II SK (Stratagene) via blunt end ligation. PCR mutagenesis was performed on the plasmid containing the TFIIIA fragment and, after sequence confirmation, the fragment was released by restriction digest with KpnI and SalI for use in cloning with EGFP into pBINAR as described above.

Various proteins were co-expressed with TFIIIA reporter constructs to assess effects on reporter activity. Their respective sequences were amplified by PCR from *A. thaliana* cDNA and after restriction digest, cloned into the BamHI/SalI sites of pBINAR. Cloning of control constructs containing luciferase (LUC) or DsRED2 were described previously[13]. For introduction of an N-terminal Flag tag, two complementary oligonucleotides encoding a start codon followed by the Flag peptide sequence flanked by sequences with BamHI-compatible overhangs were annealed and cloned in-frame into the BamHI site immediately 5' of the protein coding region. Since the L5 coding region contains a BamHI site, the Flag tag was instead introduced as an extended overhang to the 5' primer complemenary to L5 (DNA43).

Additional notes on non-denaturing gel shift assays

Renaturation of the RNA before performing the assay was found to be important to reduce the amount of alternatively folded forms, as sometimes observed by the appearance of multiple RNA bands in the absence of added protein. Gels below 6% acylamide are not able to resolve these multiple bands, and it is possible that even at 10% acrylamide there are still unresolved bands. Some of these alternatively folded P5SM RNAs apparently do not bind GST-*At*L5 and therefore are unaffected by addition of protein. We observe with the renaturation protocol that saturated binding of the RNA by protein usually is reached with ~60-70% of P5SM RNA bound and shifted. To calculate the dissociation constant for the RNA-protein interaction, the percentage bound was normalized to this empirically observed maximal binding. We and others[9] have observed a similar effect for *A. thaliana* 5S rRNA, in which saturated binding is reached with some of the RNA remaining unshifted.

References

1. Weinberg, Z. et al. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res* **35**, 4809-19 (2007).

2. Swarbreck, D. et al. The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res* **36**, D1009-14 (2008).

3. Tanaka, T. et al. The Rice Annotation Project Database (RAP-DB): 2008 update. *Nucleic Acids Res* **36**, D1028-33 (2008).

4. Jaillon, O. et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463-7 (2007).

5. Lang, D., Eisinger, J., Reski, R. & Rensing, S.A. Representation and high-quality annotation of the *Physcomitrella patens* transcriptome demonstrates a high proportion of proteins involved in metabolism in mosses. *Plant Biol (Stuttg)* **7**, 238-50 (2005).

6. Chenna, R. et al. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**, 3497-500 (2003).

7. Katoh, K., Hori, H. & Osawa, S. The nucleotide sequences of 5S ribosomal RNAs from four *Bryophyta* species. *Nucleic Acids Res* **11**, 5671-5674 (1983).

8. Cloix, C. et al. Analysis of the 5S RNA pool in *Arabidopsis thaliana*: RNAs are heterogeneous and only two of the genomic 5S loci produce mature 5S RNA. *Genome Res* **12**, 132-144 (2002).

9. Mathieu, O. et al. Identification and characterization of transcription factor IIIA and ribosomal protein L5 from *Arabidopsis thaliana*. *Nucleic Acids Res* **31**, 2424-2433 (2003).

10. Schmeing, T.M., Moore, P.B. & Steitz, T.A. Structures of deacylated tRNA mimics bound to the E site of the large ribosomal subunit. *RNA* **9**, 1345-52 (2003).

11. Lee, Y. & Nazar, R.N. Terminal structure mediates 5 S rRNA stability and integration during ribosome biogenesis. *J Biol Chem* **278**, 6635-41 (2003).

12. Hofgen, R. & Willmitzer, L. Biochemical and genetic analysis of different patatin isoforms expressed in various organs of potato (*Solanum tuberosum*). *Plant Sci* **66**, 221-230 (1990).

13. Wachter, A. et al. Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cell* **19**, 3437-50 (2007).