

Supplementary Materials

MISF2* encodes an essential mitochondrial splicing cofactor required for *nad2* mRNA processing and embryo development in *Arabidopsis thaliana

Tan-Trung Nguyen^{1,†}, Corinne Best^{2,†}, Sofia Shevtsov², Michal Zmudjak², Martine Quadrado¹, Ron Mizrahi², Hagit Zer², Hakim Mireau^{1,*}, Oren Ostersetzer-Biran^{2,*}

¹ Affiliation 1; Institut Jean-Pierre Bourgin INRAE, AgroParisTech, CNRS, Université Paris-Saclay, Versailles, France

² Affiliation 2; Department of Plant and Environmental Sciences, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Givat-Ram, Jerusalem 91904, Israel.

† These authors have equally contributed to the MS.

* Correspondence: Hakim Mireau, E-mail, hakim.mireau@versailles.inra.fr, and Oren Ostersetzer-Biran. E-mail, oren.ostersetzer@mail.huji.ac.il.

Supplementary data files:

Figure S1: The topology and structure of MISF2 protein.

Figure S2: MISF2 gene expression pattern during Arabidopsis development.

Figure S3: Nucleotide sequence of the *MISF2* gene and precise location of T-DNA insertion sites.

Figure S4: Steady-state level analysis of various mitochondrial proteins in wild type, *misf2.2* and functionally complemented *misf2.2* plants.

Figure S5: Phylogenetic analysis of homologous MISF2 in plants.

Figure S6: Transcription profiles of *RDM1* gene by RNA-seq analyses

Figure S7: Analysis of the 5' UTR of *RDM1* transcript by RT-PCR.

Table S1: List of oligonucleotides used in this study.

Table S2: List of antibodies used in this study.

a

MPP/ICP55 cleavage site

MLTKLRISKLVSYTLPRRIFQRRFLVTNNTAEESPIVAAESPELPSWIKDFLSNKPSSSS
SSVSKDDEDFVIPS LANWVESQKFSRQQVSEGNVVKPVEDIDKVCDFLNKKDTSHEDEVV
KELSKCDVVVTESLVLQVLRFRSNGWNQAYGFFIWANSQTGYVHS | GHTYNAMVDVLGKC
RNFDLMWELVNEMNKNEESKL | VTI | LDTMSKVMRRLAKSGKYNKAVDAFLEMEKSYGVK
TD | TIAMNSLMDALVKENSIEHAHEVFLKLFDTIKPD | ARTFNILIHGFCKARKFDDARA
MMDLMKVTEFTPD | VVTYTSFVEAYCKEGDFRRVNEMLEEMRENGCNPN | VVTYTIVMHS
LGKSKQVAEALGVYEKMKEDGCVPD | AKFYSSLIHILSKTGRFKDAAEIFEDMTNQGVRR
D | VLVYNTMISAALHHSRDEMALRLLKRMEDDEEGESCSPN | VETYAPLLKMCCHKKKMKL
LGILLHHMVKNDVSID | VSTYILLIRGLCMSGKVEEACLFEEAVRKGMVPR | DSTCKML
VDELEKKNMAEAKLKIQSLVQSKTMIDSHSPLSVS

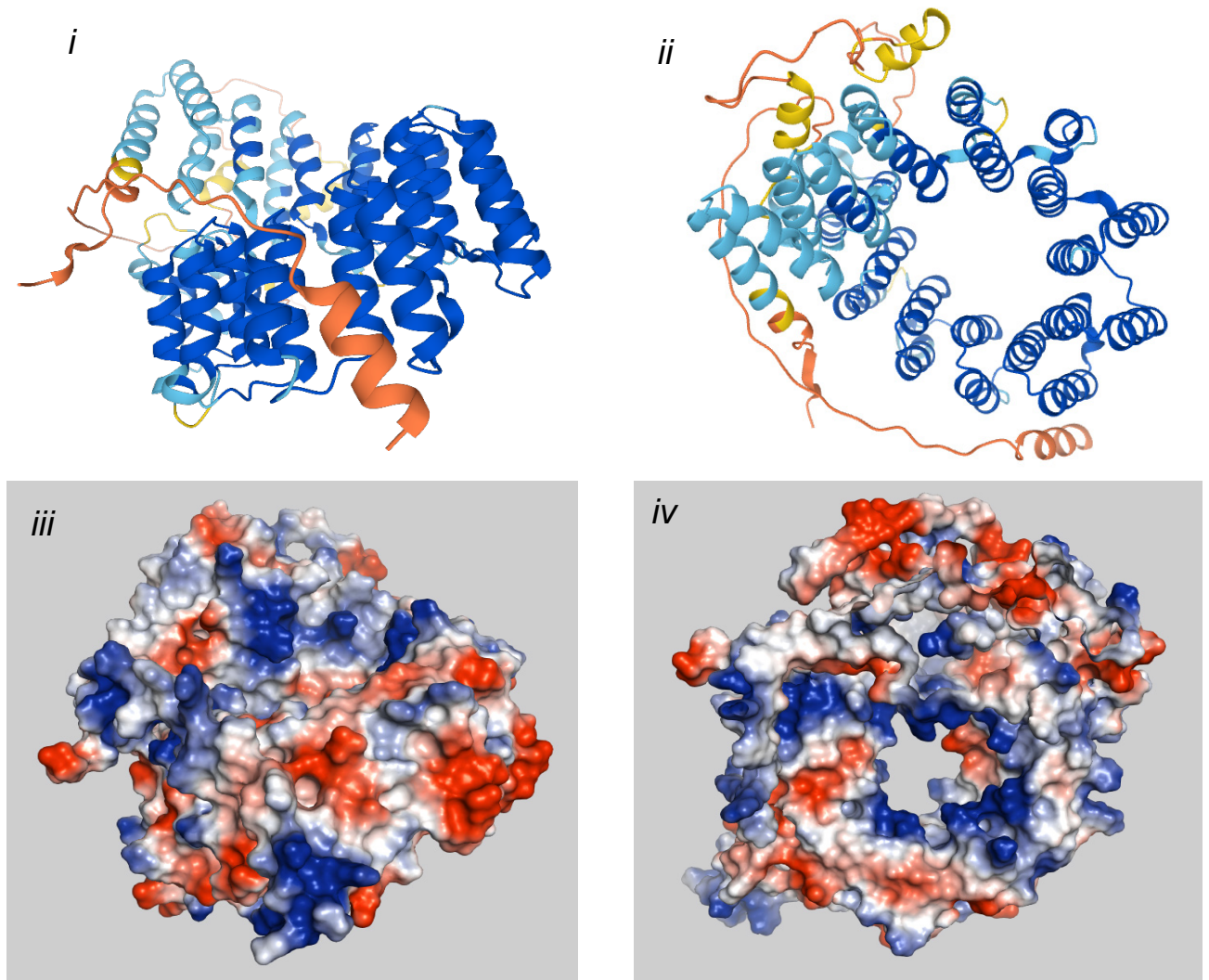
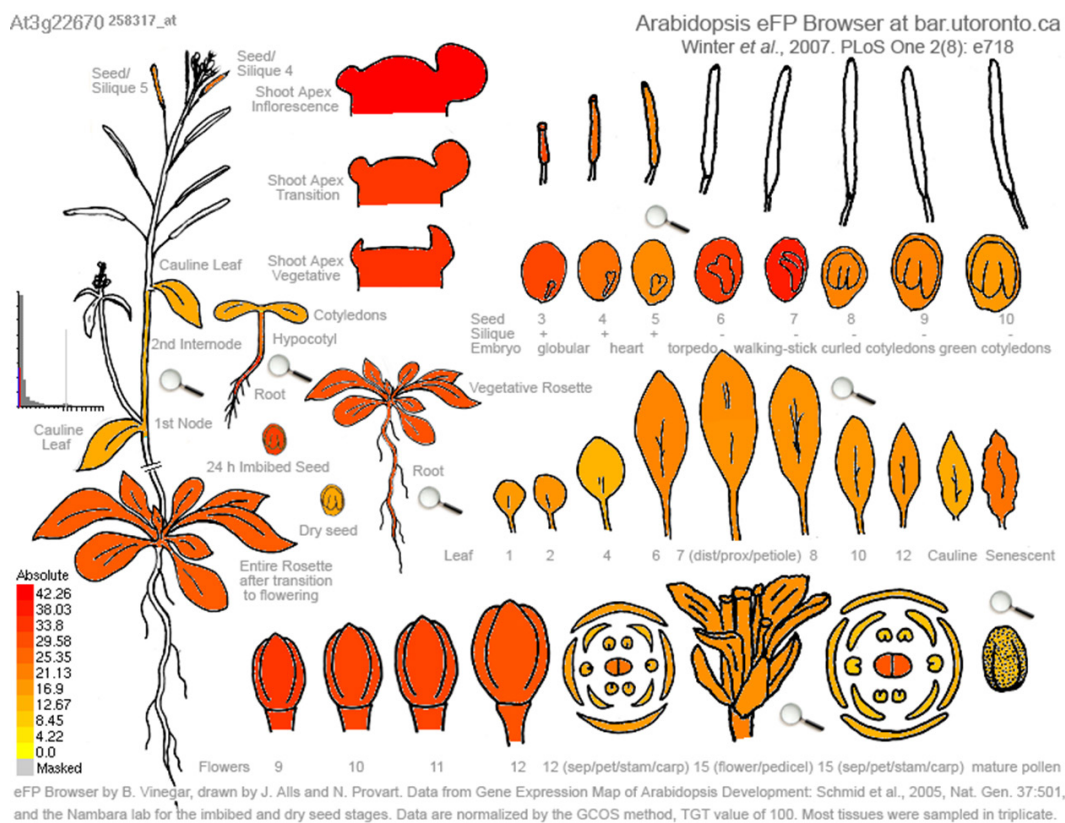
b

Figure S1. The topology and structure of MISF2 protein. (a) Amino acid sequence of the MISF2 protein (At3g22670). The region corresponding to the predicted mitochondrial targeting sequence is underlined and highlighted in red. The predicted cleavage sites of the Mitochondrial Processing Peptidase (MPP) or Intermediate Cleavage Peptidase 55 (ICP55) is indicated with a red arrow. The 10 PPR motifs (highlighted in blue letters) were predicted with the PPRfinder (Gutmann et al. 2020), PPRCODE (Shen et al. 2019), SMART (Letunic et al. 2012) and CDD (Marchler-Bauer et al. 2003) web-servers. Underlined letters indicate the 5 and 35 positions within each PPR repeat. (b) Atomic structural model of MISF2 predicted by the AlphaFold server (Jumper et al. 2021). Ribbon structures (i and ii) were generated by the PyMol software suite. Panels “iii” and “iv” represent charges on the surface of MISF2 from two different angles, with red, white, and blue for negatively, neutral, and positively charged residues, respectively.

a



b

Dataset: 10 developmental stages from data selection: AT_AFFY_ATH1-0

Showing 1 measure(s) of 1 gene(s) on selection: MSP2

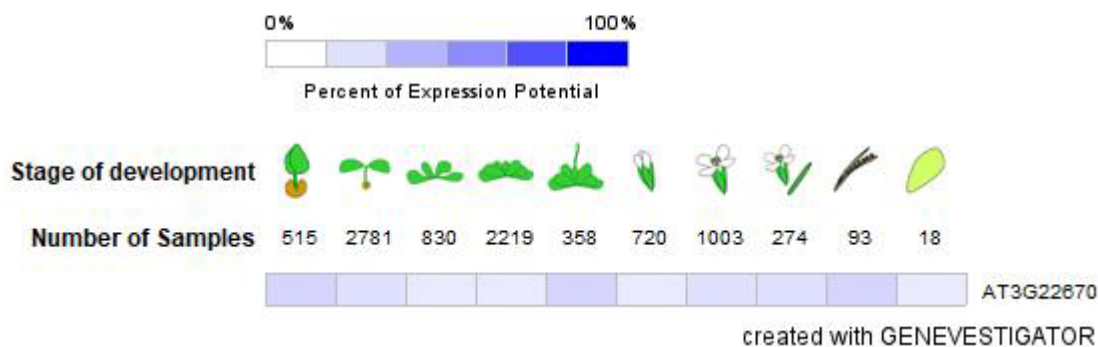


Figure S2. MISF2 gene expression pattern during Arabidopsis development. The expression pattern of MISF2 was analyzed using publicly available microarray and high-throughput RNAseq databases, including (a) 'The Arabidopsis Information Resource' (TAIR; <http://www.arabidopsis.org>), and (b) 'Genevestigator analysis toolbox' database (Hruz et al. 2008, Zimmermann et al. 2004).

a

5'...agatggaaaggaagattcacattttattgaataagaccctgcaattatacatataagcttcaatgcttgaacaagggtttctaa▲
aagcaaaagtagatacaaaaaaggatgggttcacacaattcaagaacttgtTCAATTTCTCAGGAAAGATTGGGTCAATGAAAGAGT
GATACATAGGATCAGATCTCCAAAGAAGAGCCATGTGAAGATGAGAAGGAGTGAGTCGATGAATCTCTTCAACAAGCCAGATAGTA
GCTTCAGCTTTTAGTAGGATGAATGATAGAATCCAACCTCTGTTCTTCAGAATCAGTACCGAATCGTGATTGATCCCATCGTTGTAA
GAGAACATTCGTGAGGTAATGTAGAGGCTGACCATAAAGCTGTTTCATTGAAATGCCTTAATCCAACCCAGCTTGTGAACGGAATCA
ACGAGCCACGATTCGTGCGGATTTGGTACTTGCTTCATGTAGTCTTGATACATCTCTGCTCTTCTCAGCAACGAACctgactcaaaa
aatttcagagcatttttgtaaaaatcaaaacttttgaatcaaaatggtgagaaaggaaagaaacaaggggaccattgaaatcgtgtga
cCTTCATCAGCGACGTCTCTGTGAGAAGTATCGAGCGGTGAAAAGCCATCGCTGATCTCAGCATCGACGTGAGATGAACCGGAATC
ACCTGATGGTCGTAGTTCCATTGTCATTGAGCTTTG[CAT]ggtcactgtcgcagcagtttagattttgacggcgtagaaagcagaaagg
tacgaagacgatgaccacacgaggttccttaaaattagggttttttagaggaagaagacgaagagaatgacttgtacacgtgtcagt
aaaaatcgtcacttgtcagatgtgggttaaaccaaaaatctaaacgctaaattaaattaaatcgaaccggggttgtaccaattaac
gaaccgacccgacccggcacaagcagcaagcttttaggtctaaaagattcgaactttgatcggaaaaacaaggagagaagtctgcgt
ttagattcaaacggaacaaa[ATG]CTCACTAAGCTTAGGATTAGTAAGCTTGTTTCGTATACTTTACCTCGGAGAATCTTCCAACGG
AGGTTTCTGGTAACGAATAACACTGCAGAGGAGTCGCCGATTGTTGCCGCAGAGTCACCGGAGCTACCGTCTTGGATTAAAGACTT
TCTATCAAACAAGCCTTCTTCTTCATCGAGTTCCGTTTCCAAAGATGATGAAGATTTTGTATCCCTTCTTTGGCGAATTGGGTGCG
AGAGCCAGAAGTTTAGTCGTCAACAAGTCTCAGAGGGAAATGTAGTGAAGAAACAGTGGAAGATATTGATAAAGTTTGCAGCTTT
▼TTGAATAAGAAGGACACGTCTCATGA▼GGATGTTGTTAAAGAGTTGAGTAAATGCGATGTGGTGGTTACAGAGAGTTTGGTTTTT
GCAAGTTCTAAGAAGGTTTAGTAATGGTTGGAATCAAGCTTACGGGTTTTTTCATTTGGGCAAATTCACAAACAGGTTATGTGCATT
CAGGTCATACTTATAATGCAATGGTTGATGTGTTGGGGAAATGTAGGAATTTTGATTTGATGTGGGAATTAGTCAATGAGATGAAC
AAGAATGAAGAATCTAAGCTTGTTACGCTTGATACGATGAGTAAGGTTATGAGGAGATTGGCCAAGTCAGGGAAATATAATAAAGC
TGTTGATGCTTTTTTTAGAGATGGAGAAGAGTTATGGTGTGAAAACAGATACTATTGCTATGAACAGTTTGATGGATGCACCTGTGA
AAGAGAATAGTATAGAGCATGCTCATGAAGTTTTTCTGAAGCTTTTCGATACAATCAAGCCGGATGCGCGGACATTCAACATTTTG
ATTACGGTTTTTGTAAAGCTCGGAAATTTGATGATGCTAGAGCGATGATGGATTTGATGAAGTTACTGAGTTCACCTCCTGATGT
TGTTACTTACACTAGCTTTGTGGAAGCTTATTGCAAGGAAGGGGATTTTAGGAGGGTTAATGAAATGTTGGAAGAGATGCGAGAGA
ATGGGTGTAATCCTAATGTTGTGACTTATACTATTGTGATGCATTCTTTGGGTAAATCTAAACAAGTAGCTGAAGCTTTAGGAGTG
TATGAGAAGATGAAAGAAGATGGTTGTGTTCCCTGATGCTAAGTTTTATAGCTCATTGATACATATTTTGTCTGAAGACTGGTAGGTT
TAAGGATGCAGCTGAGATATTTGAGGACATGACGAATCAAGGAGTTCGTCGTGATGTTTTAGTCTACAATACCATGATTTCCGCTG
CGCTTCATCACTCACGAGATGAGATGGCTCTACGTTTGCTTAAAAGGATGGAGGATGAGGAGGGGGAATCATGTAGTCCAAACGTT
GAGACTTATGCTCCATTGTTGAAGATGTGTTGCCATAAGAAGAAAATGAAGCTACTTGAATATTATTGCATCACATGGTGAAAAA
CGATGTTAGTATCGATGTTTCGACATATATTCTTCTCATTAGAGGGTTATGTATGAGCGGGAAAGTTGAAGAAGCTTGCTTGTTCT
TTGAGGAAGCAGTACGAAAAGGAATGGTTCCTAGAGATAGCACTTGCAAAATGTTGGTAGATGAGCTTGAGAAGAAAAATATGGCA
GAAGCAAAACTCAAGATTCAAGATTTGGTTCAATCGAAAACAATGATTGATTCTCACAGTCCTTTATCTGTTTCG[TAG]agattaga
gttttggagccttggctcaaagttttgttttcctttctgttgatgatactcttatactatgttagatctcaaatcatgtgacctt
tactaaaatgaaatgaatcaatttgtcacaagttttcatgtctt...3'

AT3G22680 (RDM1)
AT3G22670 (MISF2)

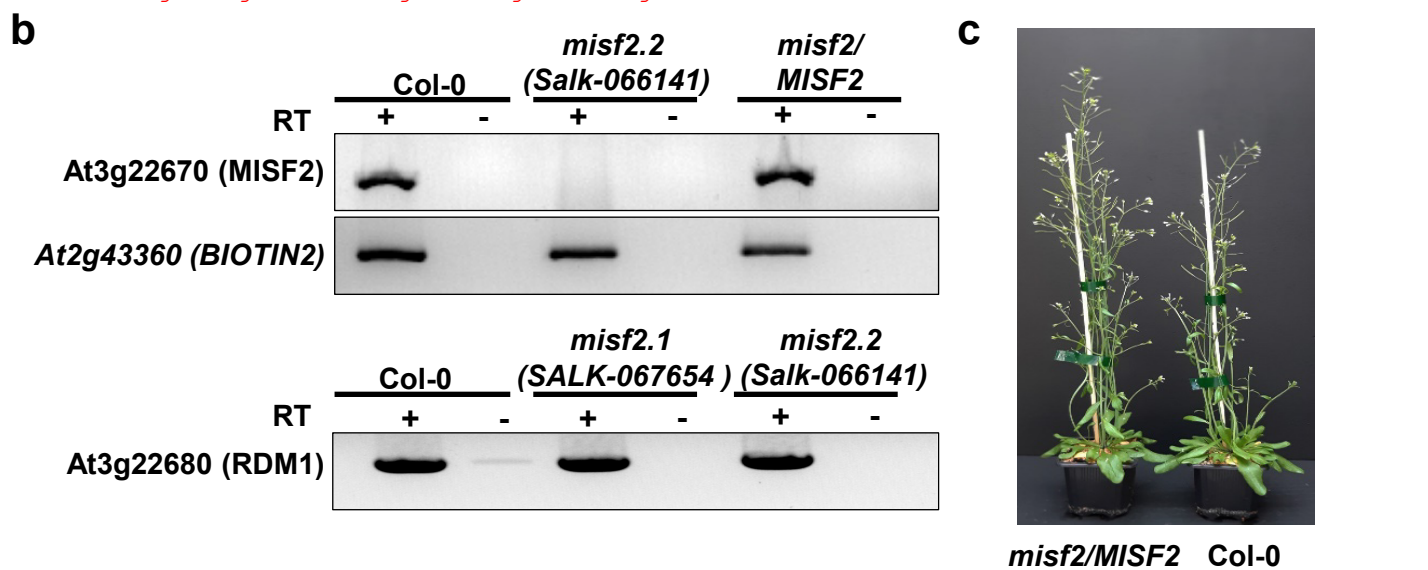


Figure S3. Nucleotide sequence of the MISF2 gene and precise location of T-DNA insertion sites. (a) Nucleotide sequence of MISF2 and RDM1, encoded by the At3g22670 and At3g22680 gene-loci, respectively. Lowercase letters colored in red indicate 5' and 3' untranslated regions (UTRs), as revealed by 5' and 3' RACE analysis, while uppercase letters correspond to the MISF2 open reading frame. Positions of T-DNA insertions in *misf2.1* (SALK-067654, 324 nts upstream to the AUG) and *misf2.2* (SALK-066141, 350 nts upstream to the AUG) lines are indicated. (b) RT-PCR analysis monitoring the expression of MISF2 in wild type (Col-0), homozygous *misf2.2* mutants (Salk_066141) and functionally complemented *misf2.2* plants (*misf2/MISF2*). (c) Photo of 5-week-old plants displaying the phenotypic recovery of *misf2.2* mutant by the expression of a complementary gene.

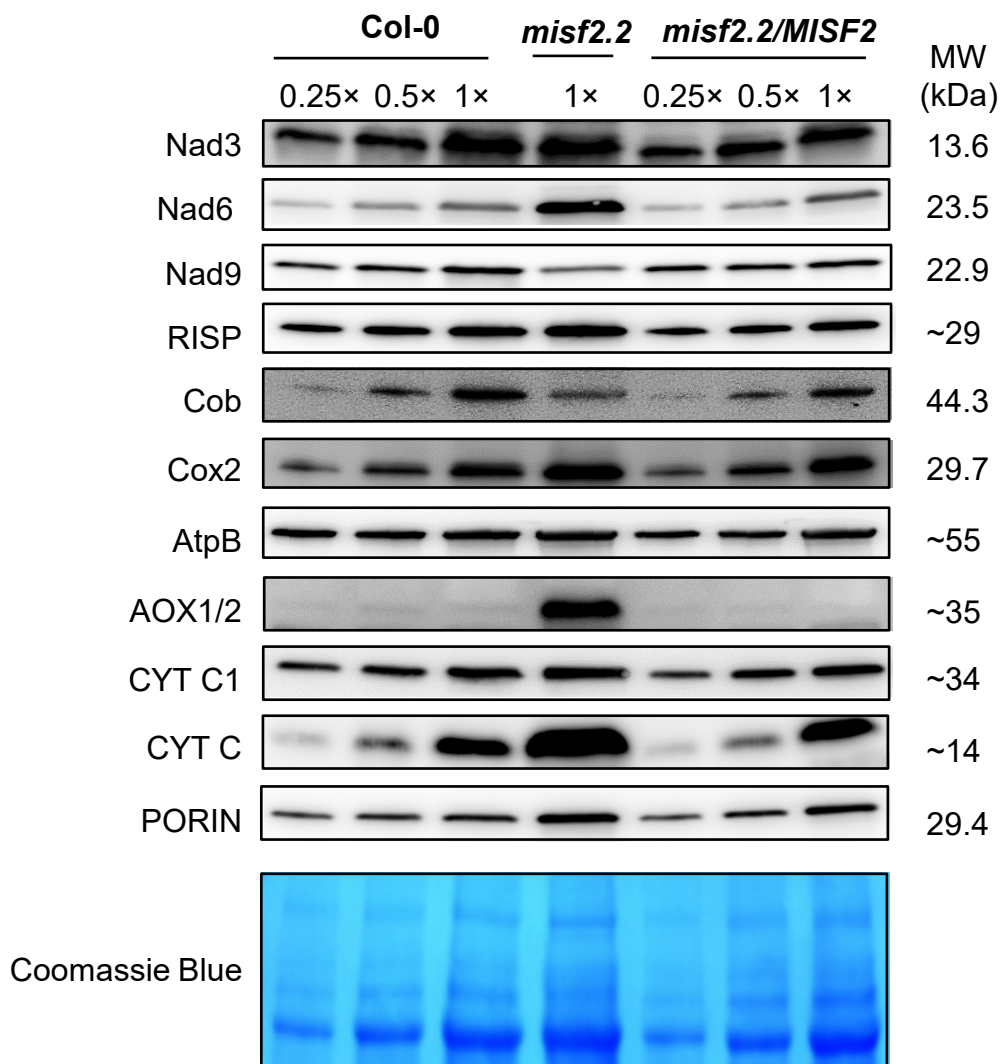


Figure S4. Steady-state level analysis of various mitochondrial proteins in wild type, *misf2.2* and functionally complemented *misf2.2* plants. Immunoblots performed with total proteins extracted from 3-week-old MS-grown wild type plants, rescued *misf2.1* plantlets, and complemented *misf2.2:35S-MISF2* plants. The blots were probed with polyclonal antibodies raised against the indicated mitochondrial proteins. Dilution series of proteins extracted from wild type and complemented plants were loaded for signal comparison with *misf2.1* extracts. Detection was carried out by chemiluminescence assays after incubation with HRP-conjugated secondary antibody. The original SDS-PAGE, following Coomassie-blue staining, is indicated below the immunoblot panels.

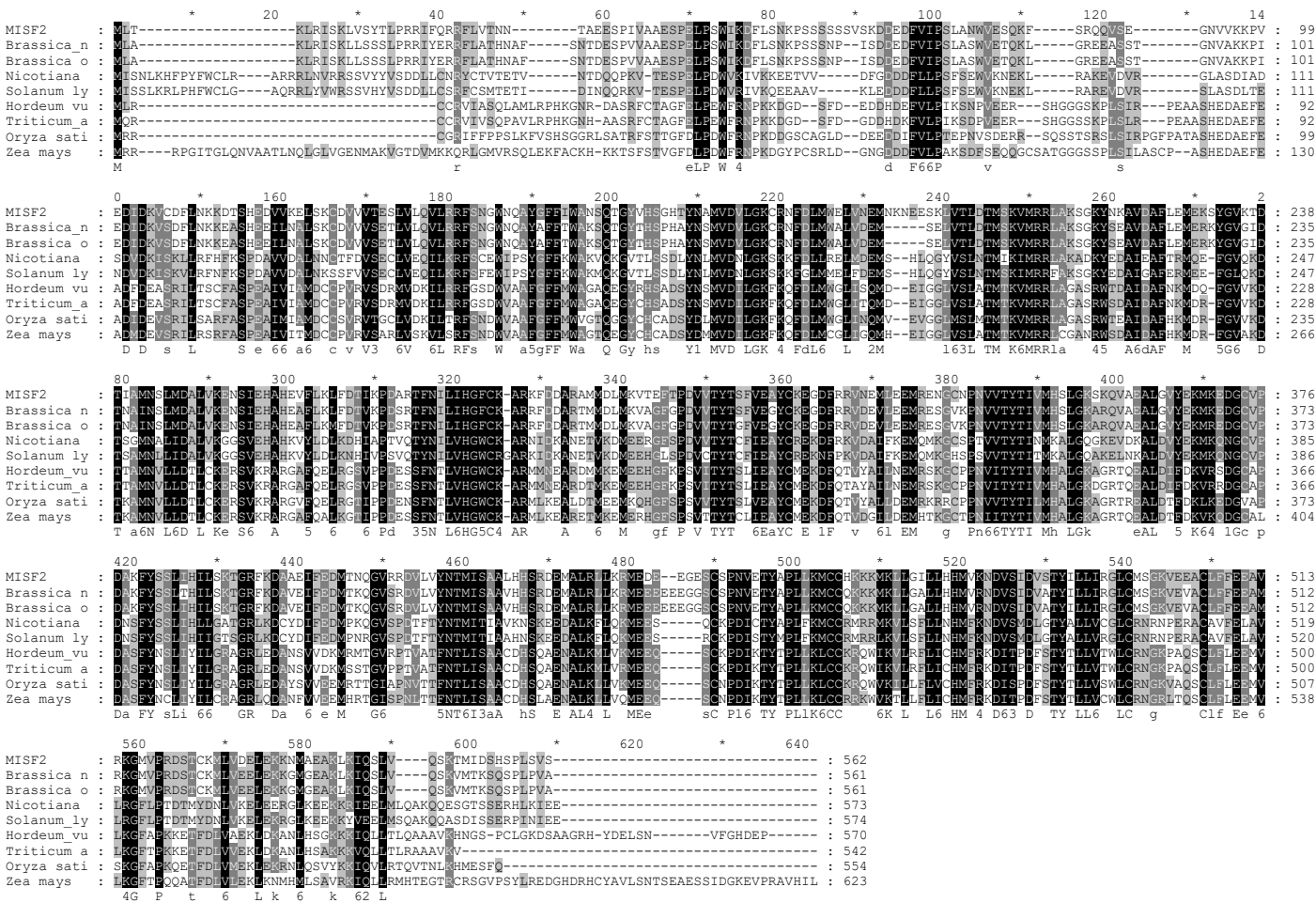
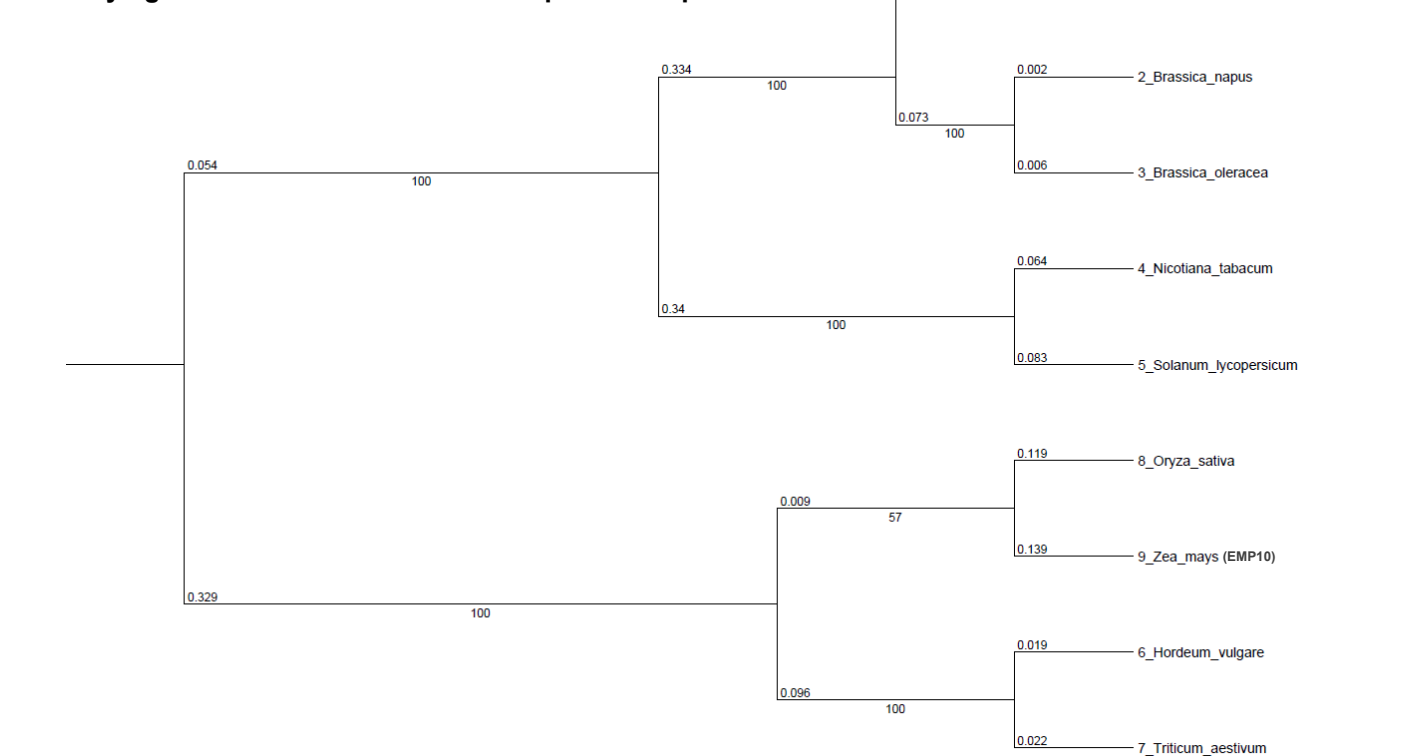
a**MISF2 homologous sequences in different angiosperm species****b****Phylogenetic three of MISF2/EMP10 sequences in plants**

Figure S5. Phylogenetic analysis of homologous MISF2 in plants. Homologues sequences to the MISF2 protein were obtained by the Basic Local Alignment Search Tool (BLAST), against the plant genome sequence databases. (a) Alignment of homologous MISF2 proteins from *A. thaliana* (NP_188906.1), *B. napus* (XP_013736563.2), *B. oleracea* (XP_013623013.1), *N. tabacum* (XP_016433260.1), *S. lycopersicum* (XP_004247584.1), *H. vulgare* (XP_044948743.1), *T. aestivum* (XP_044382896.1), *O. sativa* (XP_015628161.1) and *Z. mays* (EMP10, NP_001183846.1), was conducted with the MAFFT multiple sequence alignment server [83] and displayed using GeneDoc [82] with the conserved residue shading mode. A phylogeny tree (b) was constructed, using the MAFFT and Archaeopteryx tree viewer (<http://www.phylosoft.org/archaeopteryx>) (bootstrap values for 1,000 replicates).

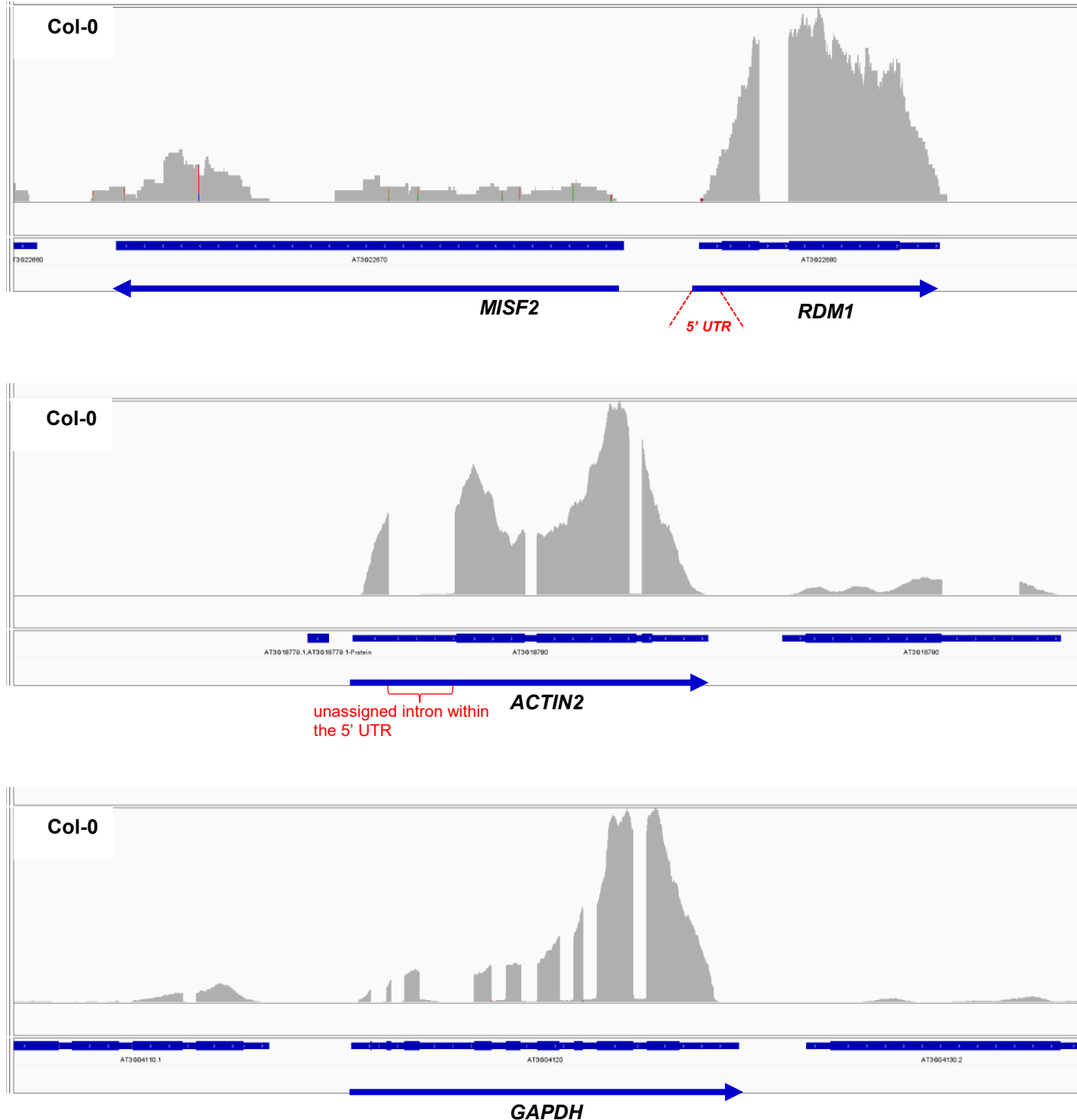


Figure S6. Transcription profiles of *RDM1* gene by RNA-seq analyses. Total RNA extracted from Col-0 plantlets was used to generate Illumina-sequencing (NovaSeq 6000, 150 bp paired-end) libraries (BioProject: PRJNA472433) [84]. Based on these datasets we analyzed the read coverage of *RDM1* (At3g22680) and *MISF2* (At3g22670), as well as *ACTIN2* (At3g18780) and *GAPDH* (At3g04120) genes, in *Arabidopsis thaliana* (Col-0) plants. Thicker boxes indicate to coding sequences (exons), while thinner regions indicate UTR and intron regions in Col-0 plants. The RNA-seq data indicated that *RDM1* harbors a 5'UTR of about 50 to 80 nucleotides long. The data also revealed to unannotated introns within the 5' UTR of *ACTIN2* gene.

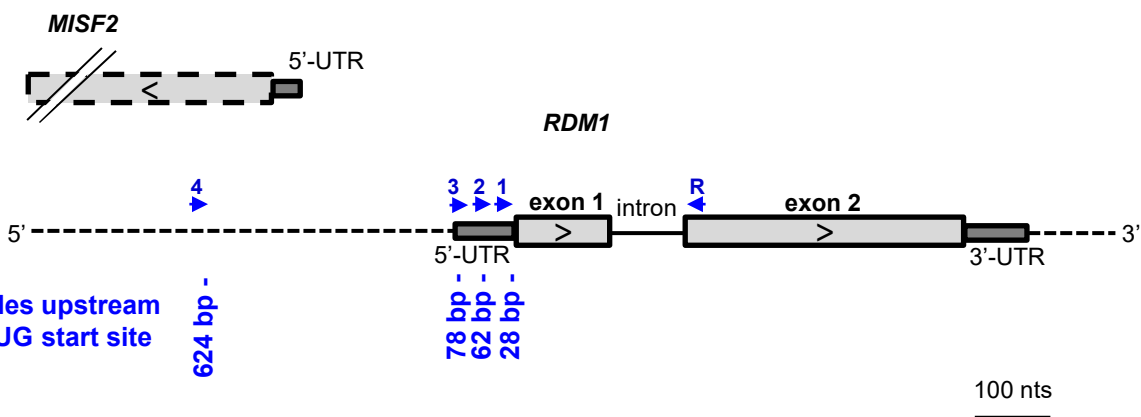
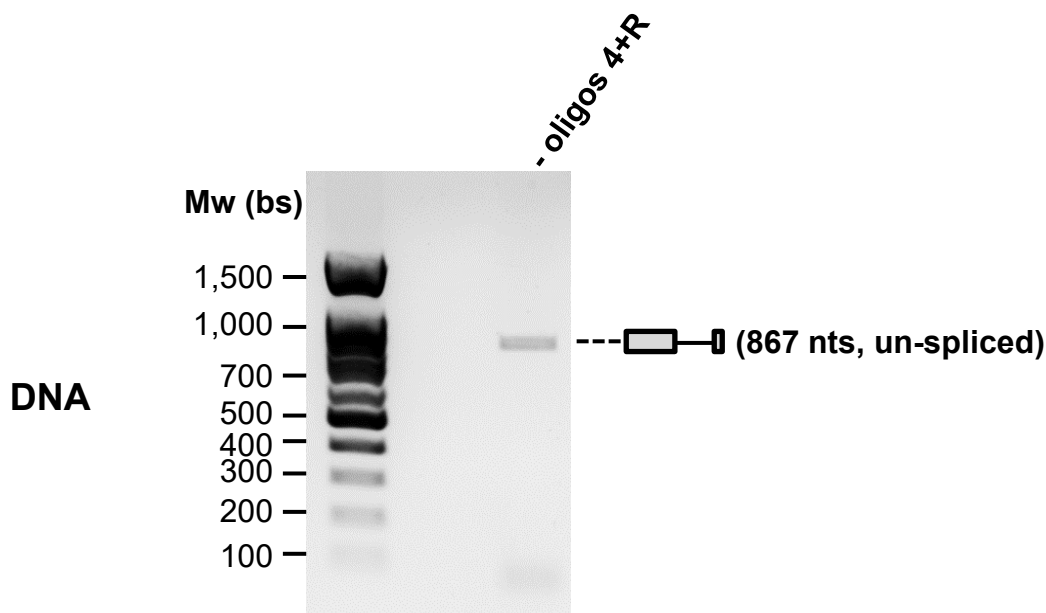
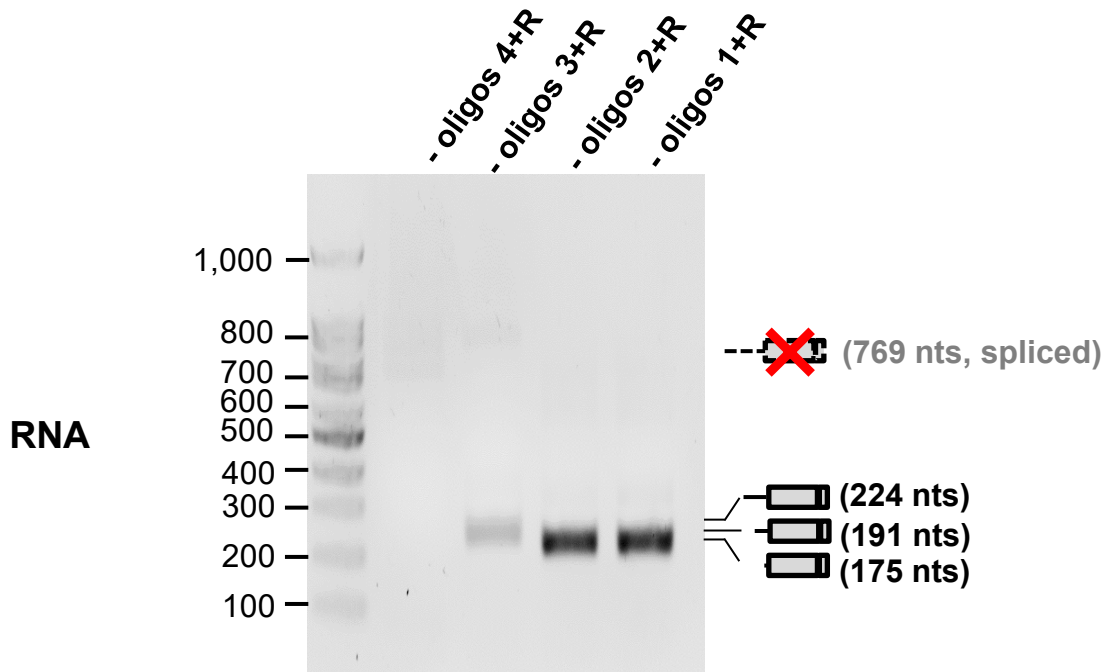
a**b****c**

Figure S7. Analysis of the 5' UTR of *RDM1* transcript by RT-PCR. (a) Scheme of the At3g22670 (*MISF2*) and At3g22680 (*RDM1*) gene-loci and gene structure. The two exons (1 and 2) consist of 124 and 368 nts, respectively, are separated by a 98 bp intron region. The position of the oligonucleotides used in PCR (panel 'b', DNA) and RT-PCR (RNA, panel 'c') are indicated. Oligonucleotides 1, 2 (28 and 62 nts upstream to the initiation codon, respectively) and the reverse primer (R, found at the very 5' of exon 2) are located inside the coding region of *RDM1* gene. Oligonucleotide 3 is found 78 nts upstream to the AUG start codon, while oligonucleotide 4 is found 624 nts upstream to the AUG site (inside the *MISF2* coding region).

Table S1. List of oligonucleotides used in this study.

List of oligonucleotides used for the analysis of the mRNA profiles of wild-type and mutant plants by RT-qPCR experiments.

| gene target | oligo name | sequence (5'-to-3') |
|----------------|----------------|---------------------------|
| <i>AOX1A</i> | <i>aox1aF</i> | AGCATCATGTTCCAACGACGTTTC |
| | <i>aox1aR</i> | GCTCGACATCCATATCTCCTCTGG |
| <i>AOX1B</i> | <i>aox1bF</i> | GGACCGTGAAATCTCTTCGATGGC |
| | <i>aox1bR</i> | TCTAGCATCATTGCTCTGCATCCG |
| <i>AOX1C</i> | <i>aox1cF</i> | TCTTCCAGAGGAGGTATGGTTGCC |
| | <i>aox1cR</i> | AGTGCATAAGCATCCCTCCAACC |
| <i>AOX1D</i> | <i>aox1dF</i> | TTTGCTCGAAGAGGCTGAGAACG |
| | <i>aox1dR</i> | CTCGTTCGTACCATTGGGTTGTG |
| <i>AOX2</i> | <i>aox2F</i> | ACGGTGATTTCGTGCTGATGAAGC |
| | <i>aox2R</i> | TCCTTGATTGCGAATGTCAGAAGC |
| <i>NDA1</i> | <i>nda1F</i> | GTATCCAACCGCGATTTCACG |
| | <i>nda1R</i> | AGTTACAGTCTCACAATGCACCTC |
| <i>NDA2</i> | <i>nda2F</i> | TGGTGTGGTCTTCTCCTTTCG |
| | <i>nda2R</i> | TCCATTTCGTCAATGCCAATCCTTC |
| <i>NDB1</i> | <i>ndb1F</i> | TAACACATTGGCACTCCTGGTG |
| | <i>ndb1R</i> | CTCTGTGCATCCTCTACTTCCTTG |
| <i>atp1</i> | <i>atp1F</i> | TCACTTCGACACGCTTTTGC |
| | <i>atp1R</i> | GGAATGGCCTTGAATCTTGA |
| <i>atp6</i> | <i>atp6-1F</i> | TCTTTTGCGAGTCAATGCAC |
| | <i>atp6-1R</i> | TCTCGCGTATCTCACATTGC |
| <i>atp8</i> | <i>atp8F</i> | CCGTCGACTTATTGGGAAAA |
| | <i>atp8R</i> | TTCCTTGGCCATGTACAACA |
| <i>atp9</i> | <i>atp9F</i> | CATTCCCTCTGACGTGCAAT |
| | <i>atp9R</i> | TCGTCGATTCTTACCCTCGT |
| <i>atp4</i> | <i>atp4F</i> | GGATCAGCTTGCGAATTTGT |
| | <i>atp4R</i> | GCAAATTGCTTCCCCACTAA |
| <i>ccmb</i> | <i>ccmBF</i> | TCTTGGAATCACATCCAGCA |
| | <i>ccmBR</i> | CGAGACCGAAATTGGAAAAA |
| <i>ccmc</i> | <i>ccmCF</i> | AGCTACGCGCAAATTCTCAT |
| | <i>ccmCR</i> | GCCGTGGCGATATAAACAAT |
| <i>ccmfc</i> | <i>ccmFcF</i> | CACATGGAGGAGTGTGCATC |
| | <i>ccmFcR</i> | GTGGGTCCATGTAAATGATCG |
| <i>ccmfn-1</i> | <i>ccmFN1F</i> | AGCTCTTGGCATTGCTTTGT |
| | <i>ccmFN1R</i> | AGTGCCACAATCCCATTCTAT |
| <i>ccmfn-2</i> | <i>ccmFN2F</i> | CGTGTCGTTCTGTAATGGAAA |
| | <i>ccmFN2R</i> | TGATAAGCCCACCAACTTCC |
| <i>cob</i> | <i>cobF</i> | TGCCGGAATGGTATTTCCTA |
| | <i>cobR</i> | GCCAAAAGCAACCAAAACAT |
| <i>cox1</i> | <i>cox1F</i> | GTAGCTGCGGTGAAGTAGGC |
| | <i>cox1R</i> | CTGCCTGGATTTCGGTATCAT |
| <i>cox2</i> | <i>cox2F</i> | TGATGCTGTACCTGGTCGTT |

| | | |
|-----------------------|------------------------|-----------------------------|
| | <i>cox2R</i> | TGGGGGATTAATTGATTGGA |
| <i>cox3</i> | <i>cox3F</i> | CCGTAACTTGGGCTCATCAT |
| | <i>cox3R</i> | AAACCATGAAAGCCTGTTGC |
| <i>mttb</i> | <i>mttBF</i> | GGGGTCTTTCTTTGGAAACC |
| | <i>mttBR</i> | TCTCCCTCATTCCACTCGTC |
| <i>nad1 exons a-b</i> | <i>nad1 1-2F</i> | GACCAATAGATACTTCATAAGAGACCA |
| | <i>nad1 1-2R</i> | TTGCCATATCTTCGCTAGGTG |
| <i>nad1 exons b-c</i> | <i>nad1 2-3F</i> | ATTCAGCTTCCGCTTCTGG |
| | <i>nad1 2-3R</i> | TCTGCAGCTCAAATGGTCTC |
| <i>nad1 exons c-d</i> | <i>nad1 3-4F</i> | AAAAGAGCAGACCCCATTTGA |
| | <i>nad1 3-4R</i> | TCCGTTTGATCTCCCAGAAG |
| <i>nad1 exons d-e</i> | <i>nad1 4-5F</i> | AGCCCGGGATCTTCTTGA |
| | <i>nad1 4-5R</i> | TCTCAATGGGGTCTGCTC |
| <i>nad2 exons a-b</i> | <i>nad2 exons a-bF</i> | GCGAGCAGAAGCAAGGTTAT |
| | <i>nad2 exons a-bR</i> | GGATCCTCCCACACATGTTT |
| <i>nad2 exons b-c</i> | <i>nad2 exons b-cF</i> | AAAGGAACTGCAGTGATCTTGA |
| | <i>nad2 exons b-cR</i> | AATATTTGATCTTAGGTGCATTTTC |
| <i>nad2 exons c-d</i> | <i>nad2 exons c-dF</i> | GCGCAATAGAAAGGAATGCT |
| | <i>nad2 exons c-dR</i> | CTATGGGTCTACTGGAGCTACCC |
| <i>nad2 exons d-e</i> | <i>nad2 exons d-eF</i> | CAAAGGAGAGGGGTATAGCAA |
| | <i>nad2 exons d-eR</i> | TATTGTCTCTCGCCGCTTT |
| <i>nad3</i> | <i>nad3F</i> | CGAATGTGGTTTCGATCCTT |
| | <i>nad3R</i> | GCACCCCTTTTCCATTGATA |
| <i>nad4 exons a-b</i> | <i>nad4 exons a-bF</i> | ATTCTATGTTTTTCCCGAAAGC |
| | <i>nad4 exons a-bR</i> | GAAAAACTGATATGCTGCCTTG |
| <i>nad4 exons b-c</i> | <i>nad4 exons b-cF</i> | AATACCCATGTTTCCCGAAG |
| | <i>nad4 exons b-cR</i> | TGCTACCTCCAATTCCTGT |
| <i>nad4 exons c-d</i> | <i>nad4 exons c-dF</i> | TTCCCTCCATAAATTCTCCGATT |
| | <i>nad4 exons c-dR</i> | TGAAATTTGCCATGTTGCAC |
| <i>nad4L</i> | <i>nad4L-F</i> | GGGGAATCCTCCTTAATAGACG |
| | <i>nad4L-R</i> | AACGAAAATGGCTAACCCAATA |
| <i>nad5 exons a-b</i> | <i>nad5 exons a-bF</i> | TGGACCAAGCTACTTATGGATG |
| | <i>nad5 exons a-bR</i> | CCATGGATCTCATCGGAAAT |
| <i>nad5 exons b-c</i> | <i>nad5 exons b-cF</i> | TACCTAAACCAATCATCATATC |
| | <i>nad5 exons b-cR</i> | CTGGCTCTCGGGAGTCTCTT |
| <i>nad5 exons c-d</i> | <i>nad5 exons c-dF</i> | AACTCGGATTCGGCAAGAA |
| | <i>nad5 exons c-dR</i> | GATATGATGATTGGTTTAGGTA |
| <i>nad5 exons d-e</i> | <i>nad5 exons d-eF</i> | AACATTGCAAAGGCATAATGA |
| | <i>nad5 exons d-eR</i> | GTTCTCGGTTTCGGATATG |
| <i>nad6</i> | <i>nad6F</i> | TATGCCGGAAGGTACGAAG |
| | <i>nad6R</i> | GTGAGTGGGTCAGTCGTCTT |
| <i>nad7 exons a-b</i> | <i>nad7 exons a-bF</i> | ACCTCAACATCCTGCTGCTC |
| | <i>nad7 exons a-bR</i> | AAGGTAAAGCTTGAAGATAAGTTTTGT |
| <i>nad7 exons b-c</i> | <i>nad7 exons b-cF</i> | GAGGGACTGAGAAATTAATAGAGTACA |
| | <i>nad7 exons b-cR</i> | TGGTACCTCGCAATTCAAAA |
| <i>nad7 exons c-d</i> | <i>nad7 exons c-dF</i> | ACTGTCACTGCACAGCAAGC |
| | <i>nad7 exons c-dR</i> | CATTGCACAATGATCCGAAG |

| | | |
|---------------------------------------|------------------------|--------------------------|
| <i>nad7</i> exons d-e | <i>nad7</i> exons d-eF | GATCAAAGCCGATGATCGTAA |
| | <i>nad7</i> exons d-eR | AGGTGCTTCAACTGCGGTAT |
| <i>nad9</i> | <i>nad9F</i> | GGATGACCCTCGAAACCATA |
| | <i>nad9R</i> | CACGCATTCTGTACAAACC |
| <i>rpl2</i> | <i>rpl2F</i> | CCGAAGACGGATCAAGGTAA |
| | <i>rpl2R</i> | CGCAATTCATCACCATTTTG |
| <i>rpl5</i> | <i>rpl5F</i> | AAGGGGTTCGACAGGAAAGT |
| | <i>rpl5R</i> | CGTATTTTCGACCGGAAAATC |
| <i>rpl16</i> | <i>rpl16F</i> | GAGCATTTGCCAAACTCACA |
| | <i>rpl16R</i> | CGGACACTTTCATCGTGCTA |
| <i>rps3</i> | <i>rps3F</i> | CCGATTTTCGGTAAGACTTGG |
| | <i>rps3R</i> | AGCCGAAGGTGAGTCTCGTA |
| <i>rps4</i> | <i>rps4F</i> | ACCCATCACAGAGATGCACA |
| | <i>rps4R</i> | TCACACAAACCCTTCGATGA |
| <i>rps7</i> | <i>rps7F</i> | CTCGAACTGAACGCGATGTA |
| | <i>rps7R</i> | AAGCTGCTTCAAGGATCCAA |
| <i>rps12</i> | <i>rps12F</i> | AGCCAAAGTACGGTTGAGCA |
| | <i>rps12R</i> | TTTGGGTTTTTCTGCACCAT |
| <i>matR</i> | <i>matR-F</i> | AATTTTTGCGAGAGCTGGAA |
| | <i>matR-R</i> | TTGAACCCCGTCTGTAGAC |
| <i>rrn18</i> | <i>rrn18F</i> | CGTCACCTGGGTCAAAAACCT |
| | <i>rrn18R</i> | GCTTGAAAACCGAAGTGAGC |
| <i>rrn26</i> | <i>rrn26F</i> | GACGAGACTTTCGCCTTTTG |
| | <i>rrn26R</i> | CTTGAGCGAATTGGATGAT |
| <i>rrn5</i> | <i>rrn5F</i> | CCGACCTCGATATGTGGAATCGTC |
| | <i>rrn5R</i> | TGGACCATGTCTCCCGAACAATC |
| <i>18S rRNA</i> (nuclear) | <i>18S nucl-F</i> | AAACGGCTACCACATCCAAG |
| | <i>18S nucl-R</i> | ACTCGAAAGAGCCCGGTATT |
| <i>ACTIN2</i> (At3g18780, nuclear) | <i>actin2-F</i> | GGTAACATTGTGCTCAGTGGTGG |
| | <i>actin2-R</i> | AACGACCTTAATCTTCATGCTGC |
| <i>GAPDH</i> (At3g04120) | <i>GAPDH-F</i> | TCTCGATCTCAATTCGCAAAA |
| | <i>GAPDH-R</i> | CGAAACCGTTGATTCCGATTC |

List of oligonucleotides used for the analysis of the splicing profiles of wild-type and mutant plants by RT-qPCR experiments.

| Gene | Forward primer | Reverse primer |
|----------------------------|-----------------------------|-----------------------|
| <i>rpl2</i> | CCGAAGACGGATCAAGGTAA | CGCAATTCATCACCATTTTG |
| <i>rpl2</i> intron exon2 | TTAGGAAGAGCCGTACGAGG | CGCAATTCATCACCATTTTG |
| <i>rps3</i> | AGCCGAAGGTGAGTCTCGTA | CCGATTTTCGGTAAGACTTGG |
| <i>rps3</i> intron1 exon2 | AGCCGAAGGTGAGTCTCGTA | TCTACGGCGGGGTCACTAT |
| <i>cox2</i> | TGGGGGATTAATTGATTGGA | TGATGCTGTACCTGGTCGTT |
| <i>cox2</i> intron1 exon2 | TGGGGGATTAATTGATTGGA | AGCAGTACGAGCTGAAAGGC |
| <i>ccmFc</i> | GTGGGTCCATGTAAATGATCG | CACATGGAGGAGTGTGCATC |
| <i>ccmFc</i> intron1 exon1 | CCCGGATCGAATCAGAGTT | CACATGGAGGAGTGTGCATC |
| <i>nad1</i> exon1-2 | GACCAATAGATACTTCATAAGAGACCA | TTGCCATATCTTCGCTAGGTG |
| <i>nad1</i> intron1 exon2 | GACCAATAGATACTTCATAAGAGACCA | CGTGCTCGTACGGTTCATAG |
| <i>nad1</i> exon2-3 | ATTCAGCTTCCGCTTCTGG | TCTGCAGCTCAAATGGTCTC |

| | | |
|---------------------------|-----------------------------|-----------------------------|
| <i>nad1</i> intron2 exon2 | GGTTGGGTTAGGGGAACATC | TCTGCAGCTCAAATGGTCTC |
| <i>nad1</i> exon3-4 | AAAAGAGCAGACCCCATTGA | TCCGTTTGATCTCCCAGAAG |
| <i>nad1</i> intron3 exon4 | AAAAGAGCAGACCCCATTGA | GGGAGCTGTATGAGCGGTAA |
| <i>nad1</i> exon4-5 | AGCCCGGGATCTTCTTGA | TCTTCAATGGGGTCTGCTC |
| <i>nad1</i> intron4 exon5 | AGCCCGGGATCTTCTTGA | ACGGAGCTGCATCCCTACT |
| <i>nad2</i> exon1-2 | GCGAGCAGAAGCAAGGTTAT | GGATCCTCCCACACATGTTT |
| <i>nad2</i> intron1 exon2 | GCGAGCAGAAGCAAGGTTAT | CCCATTCCTAACCAGTGGAG |
| <i>nad2</i> exon2-3 | AAAGGAACTGCAGTGATCTTGA | AATATTTGATCTTAGGTGCATTTTC |
| <i>nad2</i> intron2 exon2 | CCCGATCCGATAGTTTACAA | AATATTTGATCTTAGGTGCATTTTC |
| <i>nad2</i> exon3-4 | GCGCAATAGAAAGGAATGCT | CTATGGGTCTACTGGAGCTACCC |
| <i>nad2</i> intron3 exon4 | GCGCAATAGAAAGGAATGCT | GGCGAATTTCAAACCTTGTTG |
| <i>nad2</i> exon4-5 | CAAAGGAGAGGGGTATAGCAA | TATTTGTTCTTCGCCGCTTT |
| <i>nad2</i> intron4 exon4 | CTTATTCGTGGCAACCTTCC | TATTTGTTCTTCGCCGCTTT |
| <i>nad4</i> exon1-2 | ATTCTATGTTTTTCCCGAAAGC | GAAAAACTGATATGCTGCCTTG |
| <i>nad4</i> intron1 exon2 | CCGTATGATGCGGAAGTCTC | GAAAAACTGATATGCTGCCTTG |
| <i>nad4</i> exon2-3 | AATACCCATGTTTCCCGAAG | TGCTACCTCCAATTCCCTGT |
| <i>nad4</i> intron2 exon3 | GCGGAACGACCAGAAAAATA | TGCTACCTCCAATTCCCTGT |
| <i>nad4</i> exon3-4 | TTCTCCATAAATTCTCCGATT | TGAAATTTGCCATGTTGCAC |
| <i>nad4</i> intron3 exon4 | TCTAGCTTGGTTTCGGAGAGC | TGAAATTTGCCATGTTGCAC |
| <i>nad5</i> exon1-2 | TGGACCAAGCTACTTATGGATG | CCATGGATCTCATCGGAAAT |
| <i>nad5</i> intron1 exon2 | TGGACCAAGCTACTTATGGATG | TTCGCAAATAGGTCCGACT |
| <i>nad5</i> exon2-3 | TACCTAAACCAATCATCATATC | CTGGCTCTCGGGAGTCTCTT |
| <i>nad5</i> intron2-exon2 | GTACGATCGTGTCGGGTGA | CTGGCTCTCGGGAGTCTCTT |
| <i>nad5</i> exon3-4 | AACTCGGATTCGGCAAGAA | GATATGATGATTGGTTTAGGTA |
| <i>nad5</i> intron3-exon4 | AACTCGGATTCGGCAAGAA | GCCGTGTAATAGGCGACCA |
| <i>nad5</i> exon4-5 | AACATTGCAAAGGCATAATGA | GTTCTGCGTTTCGGATATG |
| <i>nad5</i> intron4 exon5 | AACATTGCAAAGGCATAATGA | CCTGTAAACCCCATGATGT |
| <i>nad7</i> exon1-2 | ACCTCAACATCCTGCTGCTC | AAGGTAAAGCTTGAAGATAAGTTTTGT |
| <i>nad7</i> intron1 exon2 | ACGGTTTTTAGGGGGATCTG | AAGGTAAAGCTTGAAGATAAGTTTTGT |
| <i>nad7</i> exon2-3 | GAGGGACTGAGAAATTAATAGAGTACA | TGGTACCTCGCAATTCAAAA |
| <i>nad7</i> intron2 exon3 | AGTGGGAGAGCCGTGTTATG | TGGTACCTCGCAATTCAAAA |
| <i>nad7</i> exon3-4 | ACTGTCACTGCACAGCAAGC | CATTGCACAATGATCCGAAG |
| <i>nad7</i> intron3 exon4 | TAAAGTGAAGTGTTGGGCCT | CATTGCACAATGATCCGAAG |
| <i>nad7</i> exon4-5 | GATCAAAGCCGATGATCGTAA | AGGTGCTTCAACTGCGGTAT |
| <i>nad7</i> intron4 exon5 | CGGCCAAATGACTACAGGAT | AGGTGCTTCAACTGCGGTAT |

List of other oligonucleotides used for the analysis of *MISF2* and *RDM1* genes, the construction of MISF2-GFP and RACE analyses.

| Gene | Forward primer | Reverse primer |
|---|--|--|
| <i>MISF2</i> (AT3G22670) | ATGCTCACTAAGCTTAGGATTA | ATTATAAGTATGACCTGAATGCA |
| <i>MISF2.1</i> (SALK_067654) | ATTTTGCCGATTTTCGGAAC | ATTATAAGTATGACCTGAATGCA |
| <i>MISF2.2</i> (SALK_066141) | ATTTTGCCGATTTTCGGAAC | ATTATAAGTATGACCTGAATGCA |
| <i>MISF2-GFP</i> (<i>misf2</i> -B1+ <i>misf2</i> -B2) | GGGGACAAGTTTGTACAAAAAAGC AGGCTAAAATGCTCACTAAGCTTAG GATTAGT | GGGGACCACTTTGTACAAGAAAG CTGGGTCAAGCGTAACAAGCTTA GATTCTTC |
| <i>MISF2-RACE_S1</i> | AACAATCGGCGACTCCTCTG | AGGATGGAGGATGAGGAGGG |
| <i>MISF2-promo-B1</i> (F) <i>MISF2-Cpl-B2</i> (R) | GGGGACAAGTTTGTACAAAAAAGCAGGCT AAAGCTTCAATGCTTGAACAAGGTTC | GGGGACCACTTTGTACAAGAAAG CTGGGTCCGAAACAGATAAAGGA CTGTGAG |
| <i>MISF2-B1</i> (F) <i>MISF2-Cpl-B2</i> (R) | GGGGACAAGTTTGTACAAAAAAGCAGGCT AAAATGCTCACTAAGCTTAGGATTAGT | GGGGACCACTTTGTACAAGAAAG CTGGGTCCGAAACAGATAAAGGA CTGTGAG |
| <i>RDM1-GSP</i> | GATTACGCCAAGCTTCAGCGATGGC TTTTCACCGCTCGAT | GATTACGCCAAGCTTTCAACGAG CCACGATTCGTCGGGAT |
| <i>RDM1-NGSP</i> | GATTACGCCAAGCTTCTTCTCACAG AGACGTCGCTGATGAAG | GATTACGCCAAGCTTGCTTAATCC AACCAGCTTGTGAACGGA |
| <i>RDM1-RT-PCR1</i> | GTCAAAATCTAACTGCTCGACAG | CTCTGCTCTTCTCAGCAACGAAC |
| <i>RDM1-RT-PCR2</i> | GTCATCGTCTTCGTACCTTTCTGC | CTCTGCTCTTCTCAGCAACGAAC |
| <i>RDM1-RT-PCR3</i> | AAGGAACCTCGTGTGGTCATC | CTCTGCTCTTCTCAGCAACGAAC |
| <i>RDM1-RT-PCR4</i> | CCACTGGTTTCTTCACTACATTTC | CTCTGCTCTTCTCAGCAACGAAC |

Table S2. List of antibodies used in this study.

| Antibody | Protein I.D. | origin | serum | dilution | Reference / source |
|--------------|--|------------------------------------|---------------------|----------|---|
| AtpA | Mitochondrial ATP-synthase subunit α | <i>Zea mays</i> | Mouse (monoclonal) | 1/500 | Thomas Elthon collection, PM014 |
| AtpB | Mitochondrial ATP-synthase subunit β | <i>Zea mays</i> | Mouse (monoclonal) | 1/5,000 | (Michael <i>et al.</i> 1993) |
| CA2 | γ -carbonic anhydrase-like subunit 2 | <i>Arabidopsis thaliana</i> | Rabbit (polyclonal) | 1/1,000 | (Perales <i>et al.</i> 2005, Sunderhaus <i>et al.</i> 2006) |
| Cox2 | Cytochrome oxidase subunit-2 | <i>Arabidopsis thaliana</i> | Rabbit (polyclonal) | 1/5,000 | Agrisera antibodies, AS04 053A |
| Nad1 | NADH-dehydrogenase complex subunit-1 | <i>Arabidopsis thaliana</i> | Rabbit (polyclonal) | 1/1000 | Gift of Dr. Etienne Meyer, Halle U. |
| Nad9 | NADH-dehydrogenase complex subunit-9 | <i>Triticum spp.</i> | Rabbit (polyclonal) | 1/50,000 | (Lamattina <i>et al.</i> 1993) |
| RISP | Rieske iron-sulfur protein | <i>Arabidopsis thaliana</i> | Rabbit (polyclonal) | 1/5,000 | Gift of Prof. Ian Small, UWA |
| VDAC (PORIN) | Mitochondrial membrane-associated β -barrel proteins | <i>Zea mays</i> | Mouse (monoclonal) | 1/1000 | Thomas Elthon collection, PM035 |
| AOX1/2 | Alternative oxidase | <i>Sauromatum guttatum</i> | Rabbit (polyclonal) | 1/500 | Agrisera antibodies, AS04 054 |
| Nad6 | NADH-ubiquinone oxidoreductase chain 6 | <i>Arabidopsis thaliana</i> | Rabbit (polyclonal) | 1:1000 | Agrisera (AS15 2926) |
| Cob | Apocytochrome B | <i>Beta vulgaris ssp. maritima</i> | Rabbit (polyclonal) | 1:4000 | Meyer et al. 2018 |
| CYTC1 | Cytochrome C1 | <i>Arabidopsis thaliana</i> | Rabbit (polyclonal) | 1:1000 | Giégé et al. 2005 |
| CYTC | Cytochrome C | <i>Arabidopsis thaliana</i> | Rabbit (polyclonal) | 1:1000 | Agrisera (AS08 343A) |

References

- Giege, P, Sweetlove LJ, Cognat V, Leaver CJ. 2005.** Coordination of nuclear and mitochondrial genome expression during mitochondrial biogenesis in Arabidopsis. *Plant cell*, **17**, 1497-1512.
- Lamattina, L., Gonzalez, D., Gualberto, J. and Grienenberger, J.-M.** (1993) Higher plant mitochondria encode an homologue of the nuclear-encoded 30-kDa subunit of bovine mitochondrial complex I. *Eur J Biochem*, **217**, 831-838.
- Meyer EH, Lehmann C, Boivin S, Brings L, De Cauwer I, Bock R, Kuhn K, Touzet P.** (2018) CMS-G from Beta vulgaris ssp. maritima is maintained in natural populations despite containing an atypical cytochrome c oxidase. *Biochem J*, **475**, 759–773.
- Michael, H.L., Arnost, H. and Elthon, T.E.** (1993) Monoclonal Antibodies to the a- and b- Subunits of the Plant Mitochondrial F₁-ATPase. *Plant Physiology*, **101**, 931-937.
- Perales, M., Eubel, H., Heinemeyer, J., Colaneri, A., Zabaleta, E. and Braun, H.-P.** (2005) Disruption of a nuclear gene encoding a mitochondrial gamma carbonic anhydrase reduces complex I and supercomplex I+III2 levels and alters mitochondrial physiology in Arabidopsis. *J Mol Biol*, **350**, 263-277.
- Sunderhaus, S., Dudkina, N.V., Jansch, L., Klodmann, J., Heinemeyer, J., Perales, M., Zabaleta, E., Boekema, E.J. and Braun, H.P.** (2006) Carbonic anhydrase subunits form a matrix-exposed domain attached to the membrane arm of mitochondrial complex I in plants. *J Biol Chem*, **281**, 6482-6488.