Supplementary Information

A synthetic RNA editing factor edits its target site in

chloroplasts and bacteria

Santana Royan¹, Bernard Gutmann², Catherine Colas des Francs-Small², Suvi Honkanen^{2,3}, Jason Schmidberger¹, Ashley Soet¹, Yueming Kelly Sun², Lilian Vincis Pereira Sanglard², Charles S. Bond¹, Ian Small²

¹School of Molecular Sciences, The University of Western Australia, Crawley 6009, Australia

²Australian Research Council Centre of Excellence in Plant Energy Biology, School of Molecular Sciences, The University of Western Australia, Crawley 6009, Australia ³Synthetic Biology Future Science Platform, CSIRO, Australia

a

dsn3PLS-DYW

Сар	1	MG	INS	4					
P1	5	VF	SWNSMIRGYARNGQPEEALSLYSQMRRSGIKPD	40	P2	308	VVSWNAMISGYA	MHGHGKEALELFEEMQQSGIKPS	342
L1	41	NY	TFPFVLKACASLSSLKEGKQIHGHVIKSGFESD	74	L2	343	HVTFTGVLSACS	HAGLVDEGRQYFNSMKKDYGIEPF	378
S1	75	VY	VQSALIDMYAKCGELEDARKVFDEMPERN	105	S2	379	VEHYGCMVDLLG	RAGRLDEAYEFIESMPIEPN	410
P1	106	VV	SWNAMISGYAQNGQSEEALELFREMQQEGIKPS	140	E1	411	AVVWGALLGACF	RIHGNVELGERAAEKLFELEPES	444
L1	141	EF	TFCSVLSACASLGSLEMGKQIHGYVIKSGFESI	175	E2	445	SGNYVLLSNIYA	SAGRWDDVAKVRKMMKERGIKK	478
S1	176	VF	VGNALIDMYAKCGSIEDARKVFDEMPERT	206		479	EPGCSWIEVKNK	VHEFVAGDRSHPQSEEIYAKLE	512
P1	207	VV	SWTAMISGYAQNGQSEEALELFREMQREGVKPD	241		513	ELSEKMKEAGYV	PDTSFVLHDVEEEEKEQMLSYH	546
L1	242	ΕV	TLPSVLSACANLGALEQGKQIHAYVIKSGFESD	276	DYW	547	SEKLAIAFGLIS	TPPGTPIRIVKNLRVCGDCHTA	580
S1	277	VF	VGSALIDMYAKCGSIEDARKVFDKMPERD	307		581	IKFISKIVGREI	IVRDSNRFHHFKDGSCSCGDYW	614
b (PLS) ₃ -F	PR	17	P ₁ α-1 α-2 VVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPDEFTFS	α-1 SVLKA(L, CARLGA		α-2 KQIHGYVIKSGFEGN	S ₁ α-1 α-2 VVVYNALIDMYSKCGLLEEARKVFDEMPI	 EKD 117
b (PLS) ₃ -F dsn3PL	PR S-DYW	17 5	P ₁ α-1 α-2 VVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPDEFTFS I.::I:I.III.IIII.I.:IIIIII VFSWNSMIRGYARNGQPEEALSLYSQMRRSGIKPDNYTFP	α-1 SVLKAC . FVLKAC	L, CARLGA I.I.: CASLSS	1 ALELG : I : . I SLKEG	α-2 KQIHGYVIKSGFEGN : : KQIHGHVIKSGFESD	S ₁ α-1 α-2 VVVYNALIDMYSKCGLLEEARKVFDEMPI I.I.:IIIIII:IIIIIII VYVQSALIDMYAKCGELEDARKVFDEMPI	EKD 117 I : : ERN 105
b (PLS) ₃ -F dsn3PL (PLS) ₃ -F	PPR 8-DYW PPR	17 5 118	α-1 α-2 VVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPDEFTFS I.::I::I.IIIIII.I:IIIIII.III. VFSWNSMIRGYARNGQPEEALSLYSQMRRSGIKPDNYTFP α-1 α-2 VVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPDEFTFS	α-1 SVLKAC . FVLKAC α-1 SVLKAC	L, I.I.: CARLGA	ALELG	α-2 KQIHGYVIKSGFEGN : : KQIHGHVIKSGFESD α-2 KOIHGYVIKSGFESN	S ₁ α-1 α-2 VVVYNALIDMYSKCGLLEEARKVFDEMPI I.I.:IIIIII:IIIIIII VYVQSALIDMYAKCGELEDARKVFDEMPI α-1 α-2 VVVYNALIDMYSKCGLLEEARKVFDEMPI	EKD 117 I : : ERN 105 EKD 218
b (PLS) ₃ -f dsn3PL (PLS) ₃ -f	PPR 8-DYW PPR	17 5 118	P1 α-1 α-2 VVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPDEFTFS I.::I:I.I.I.I.IIIIIIIIIIIIIIIIIIIIIIII	α-1 SVLKAC . FVLKAC α-1 SVLKAC .	L, CARLGA I.I.: CASLSS CARLGA I.II:	ALELG : I : . I SLKEG ALELG : I I : I	α-2 KQIHGYVIKSGFEGN : : KQIHGHVIKSGFESD α-2 KQIHGYVIKSGFESN 	S1 α-1 α-2 VVVYNALIDMYSKCGLLEEARKVFDEMPI I.I.:IIIIII:IIIIIIIII VVVQSALIDMYAKCGELEDARKVFDEMPI α-1 α-2 VVVYNALIDMYSKCGLLEEARKVFDEMPI Ι.I.IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	EKD 117 I : : ERN 105 EKD 218 I : .
b (PLS) ₃ -f dsn3PL (PLS) ₃ -f dsn3PL	PPR S-DYW PPR S-DYW	17 5 118 106	P1 α-1 α-2 VVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPDEFTFS I.::I::I.II::IIII.I::IIIIII.:III. VFSWNSMIRGYARNGQPEEALSLYSQMRRSGIKPDNYTFP α-1 α-2 VVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPDEFTFS II::I.I.I.I.I.IIIIIIIIIIIIIIIIIIIIIIII	α-1 SVLKAC 	L, CARLGA I.I.: CASLSS CARLGA I.II: CASLGS	ALELG SLKEG ALELG	α-2 KQIHGYVIKSGFEGN IIIII:IIIIIII: KQIHGHVIKSGFESD α-2 KQIHGYVIKSGFESN IIIIIIIIIIIII KQIHGYVIKSGFESI	S1 α-1 α-2 VVVYNALIDMYSKCGLLEEARKVFDEMPI I.I.:IIIIII:IIIIIIIII VVVQSALIDMYAKCGELEDARKVFDEMPI α-1 α-2 VVVYNALIDMYSKCGLLEEARKVFDEMPI Ι.I.IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	EKD 117 I : : ERN 105 EKD 218 I : . ERT 206
b (PLS) ₃ -f dsn3PL (PLS) ₃ -f dsn3PL	PPR S-DYW PPR S-DYW	17 5 118 106	P1 α-1 α-2 VVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPDEFTFS I.::I:I.I.I.I.IIIIIIIIIIIIIIIIIIIIIIII	α-1 SVLKAC 	L, CARLGA I.I.: CASLSS CARLGA I.II: CASLGS	ALELG SLKEG ALELG	α-2 KQIHGYVIKSGFEGN : : KQIHGHVIKSGFESD α-2 KQIHGYVIKSGFESN . KQIHGYVIKSGFESI	α-1 α-2 VVVYNALIDMYSKCGLLEEARKVFDEMPI I.I.:IIIIII:IIIIIIIII VVVQSALIDMYAKCGELEDARKVFDEMPI α-1 α-2 VVVYNALIDMYSKCGLLEEARKVFDEMPI Ι.I.IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	EKD 117 I : : ERN 105 EKD 218 I : . ERT 206
b (PLS) ₃ -I dsn3PL (PLS) ₃ -I dsn3PL	PPR S-DYW PPR S-DYW	17 5 118 106	P1 α-1 α-2 VVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPDEFTFS I.::I::I.III::IIIII.I:IIIIIIIIIIIII	α-1 SVLKAC . FVLKAC α-1 SVLKAC α-1 SVLKAC	L, CARLGA I.I.: CARLGA I.II: CARLGA	ALELG SLKEG ALELG SLEMG	α-2 KQIHGYVIKSGFEGN : : KQIHGHVIKSGFESD α-2 KQIHGYVIKSGFESN KQIHGYVIKSGFESI α-2 κοτΗGYVIKSGFESN	S1 α-1 α-2 VVVYNALIDMYSKCGLLEEARKVFDEMPI 1.1.: !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!	EKD 117 I : : ERN 105 EKD 218 I : . ERT 206
b (PLS) ₃ -f dsn3PL (PLS) ₃ -f (PLS) ₃ -f	PPR 8-DYW PPR 8-DYW PPR	17 5 118 106 219	P1 α-1 α-2 VVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPDEFTFS I.::I:IIIIII:IIIIIIIIIIIIIIIIIIIIIIIII	α-1 SVLKAC . FVLKAC α-1 SVLSAC α-1 SVLKAC .	L, CARLGA I.I.: CASLSS CARLGA I.II: CASLGS	ALELG	α-2 KQIHGYVIKSGFEGN : : KQIHGHVIKSGFESD α-2 KQIHGYVIKSGFESI α-2 KQIHGYVIKSGFESI α-2 KQIHGYVIKSGFESI	S1 α-1 α-2 VVVYNALIDMYSKCGLLEEARKVFDEMPI 1.1.: 111111111111111111111111111111111111	EKD 117 I : : ERN 105 EKD 218 I : . ERT 206 EKD 319 I : I

Supplementary Figure S1. Comparison of the sequences of dsn3PLS-DYW and (PLS)3-PPR.

a: Sequence and motif structure of dsn3PLS-DYW. The catalytic E70 residue within the DYW domain is underlined in red. **b**: Alignment of the $(P1L1S1)_3$ regions of dsn3PLS-DYW and $(PLS)_3$ -PPR¹. Overall, throughout this region the two synthetic proteins are only 64% identical.



Supplementary Figure S2. Purification of dsn3PLS-DYW, dsn3PLS and MORF9.

a: Vector map of pETM20-dsn3PLS-DYW (left), pETM20-dsn3PLS (middle), and pETM11-MORF9 (right).
b: Elution profiles from nickel affinity chromatography using imidazole (dark blue trace) as a competitor.
Protein content was measured by absorbance at 255 nm (cyan trace) and 280 nm (dark green trace).
c: Protein profiles of selected elution fractions. L, molecular weight ladder; Lysis, unpurified bacterial lysate; Load, fraction loaded on HisTrap column; FT, flow-through from column; 6-32, specific elution fractions numbered as in b. d: Reverse nickel-affinity purification of dsn3PLS-DYW, dsn3PLS and MORF9 following TEV cleavage of the N-terminal Ni-binding His-tag. L, molecular weight ladder; -TEV, control with no TEV added; +TEV, TEV cleavage products; FT fractions, flow-through fractions containing the desired unbound product; Elution fractions, fractions eluted from the column showing the bound, cleaved His-tag. Expected molecular weights for dsn3PLS-DYW, dsn3PLS, and MORF9 with their purification tags: 82.88 kDa, 68.39 kDa, and 17.22 kDa respectively. Expected molecular weights for dsn3PLS, and MORF9 without their purification tags: 68.64, 54.14, and 14.22 kDa respectively.



Supplementary Figure S3. Analytical size-exclusion chromatography of complexes between dsn3PLS-DYW, MORF9 and their RNA target.

Fractions leaving the column were analysed for protein and RNA content by absorbance at 280 nm and 255 nm respectively. The peak containing dsn3PLS-DYW (vertical blue dashed line) was shifted to a larger molecular weight complex (vertical red dashed line) by the addition of MORF9. The addition of the rpoA-78691 target oligonucleotide raised the absorbance at 255 nm of this peak, indicating the incorporation of the RNA into the high molecular weight complex. The samples shown contain 4.88 nmol of dsn3PLS-DYW, with 19.92 nmol of MORF9, and 10 nmol RNA oligonucleotide. This equates to a 4:1 ratio of MORF9 to dsn3PLS-DYW and a 2:1 ratio of RNA oligonucleotide to dsn3PLS-DYW.



Supplementary Figure S4. Transcript abundance of editing factor transgenes and endogenous editing factor genes.

Transcript abundances were quantified as transcripts per million (TPM) by mapping RNA-seq reads to *Arabidopsis* transcripts supplemented with the sequences of the transgene constructs. For dsn3PLS-DYW and CLB19, the percentage editing at the *rpoA-78691* site is indicated for each of the transgenic lines to indicate how transcript abundance and editing are related.



Supplementary Figure S5. Splicing of *ycf*3 intron 2 is

unaffected by editing of *ycf3-43350.* The extent of *ycf3* intron 2 splicing was estimated from RNA-seq data by counting reads crossing the donor (exon-intron) junction, the acceptor (intron-exon) junction and the spliced (exon-exon) junction. Percentage splicing was calculated as spliced/(spliced + (donor + acceptor)/2). The percentage splicing in plants expressing dsn3PLS-DYW (no editing of *ycf3-43350*) and CLB19 (~20% editing of *ycf3-43350*) is not statistically distinguishable (p-value 0.33, t-test, n=3 biological replicates).

			functional					
dsn3PLS-DYW	CLB19	position	annotation	effect of editing				
23/34518	31384/36063	69942	<i>clpP1</i> coding sequence	C AU	(His)	->	U AU	(Tyr)
471/1071	1279/1575	78691	rpoA coding sequence	U C U	(Ser)	->	U U U	(Phe)
2/3622	829/4255	43350	<i>ycf</i> 3 intron 2					
0/3810	64/3940	59002	intergenic (<i>accD-psal</i>)					
113/7898	0/9704	114860	ccsA coding sequence	CUC	(Leu)	->	UUC	(Phe)
160/11446	99/12332	118393	ndhG coding sequence	U C U	(Ser)	->	U U U	(Phe)
340/28513	6/33229	116203	ndhD coding sequence	U C U	(Ser)	->	U U U	(Phe)
1/1647	19/1709	9538	3' atpA					
0/2842	37/3501	111658	ndhF coding sequence	CU C	(Leu)	->	CU U	(Leu)
5/19019	139/22668	14647	atpl coding sequence	CUU	(Leu)	->	U UU	(Phe)
76/13655	1/16473	49646	ndhK coding sequence	CU	(Pro)	->	UCU	(Ser)
149/26854	1/30538	97693	rps7 coding sequence	C AU	(His)	->	U AU	(Tyr)
4/24888	136/29717	13561	5' atpH					
49/9909	1/11722	119998	ndhA coding sequence	CUU	(Leu)	->	U UU	(Phe)
1/10196	45/11786	97277	3' rps7					
6/63955	345/92791	38871	psaB coding sequence	CCC	(Pro)	->	UCC	(Ser)
0/36770	122/39249	66041	petG coding sequence	C C U	(Pro)	->	C U U	(Leu)
150/45178	7/46938	69992	clpP1 coding sequence	U C C	(Ser)	->	U U C	(Phe)
105/58902	11/79032	39020	psaB coding sequence	U C A	(Ser)	->	U U A	(Leu)
20/105478	233/107140	13258	3' atpH					
69/102074	25/114170	73588	psbB coding sequence	CU C	(Leu)	->	CU U	(Leu)

Supplementary Table S1. Editing events detected in this study and their predicted functional impact.

The primary targets of CLB19 are indicated first, in bold. The other sites are presumed 'off-target' events. The values in the first two columns are editing events/read counts. Ratios in bold were statistically significant in comparison to the equivalent values from the negative control, i.e. clb19 plants expressing dsn3PLS-DYW (E70A). Where editing leads to a predicted amino acid change in the encoded protein, this is indicated in the final column, with the altered nucleotide in bold.

Oligonucleotide name	Sequence (5' - 3')	Modification	Purpose
rpoA_L	UAUUACACGUGCAAAAUCUG	None	Analytical size- exclusion chromatography
rpoA_REMSA	/5Cy5/AUGUAUUACACGUGCAAAAUCUGAGA	5'-Cy5 labelled	REMSA experiments
<i>clpP1_</i> REMSA	/5Cy5/CAGCAACAGAAGCCCAAGCUCAUGGA	5'-Cy5 labelled	REMSA experiments
• • •			

Supplementary Table S2. RNA oligonucleotides used in this study.

Primer name	Sequence (5' – 3')	Purpose			
SynthPLS.FOR	GAGAATCTTTATTTTCAGGGCGCCATGGGTAAT TCTGTTTTCTC	Cloning designed PPR proteins from first PLS triplet			
SynthDYW.REV	TCAGTGGTGGTGGTGGTGGTGCTCGAGTTACC AATAATCTCCACAAG	Cloning designed DYW-type PPR proteins			
DYW_E70A.FOR	GTTGTCTTACCATTCTGCGAAACTTGCTATCGC TTTC	Mutagenesis of catalytic E70 in the designer DYW domain			
DYW_E70A.REV	GAAAGCGATAGCAAGTTTCGCAGAATGGTAAGA CAAC	Mutagenesis of catalytic E70 in the designer DYW domain			
SynthSeq.FOR	ACTGTTGTTTCTTGGACAG	Sequencing designed PPR proteins			
SynthSEQ2.FOR	GGTAGACTTGATGAGGCTTATGAG	Sequencing designed PPR protein cDNA from the S2 motif			
SynthSEQ3.FOR	CTGGATATGTTCCTGATACTTC	Sequencing editing site in dsn3PLS- DYW			
SynthSeq.REV	TCCGTAATGCTCAACTCTTG	Sequencing designed PPR proteins			
T7short	TATGCTAGTTATTGCTCAG	Sequencing cDNA from the T7 terminator			
<i>rpoA_</i> GA.FOR	AGGAACATGTATTACACGTGCAAAATCTGAGAG CTGAGCAATAACTAGCATAACCCCTTG	Insertion of <i>rpoA</i> editing site cassette			
rpoA_GA.REV	GATTTTGCACGTGTAATACATGTTCCTTCTATTT CTCGAGTTACCAATAATCTCCACAAG	Insertion of <i>rpoA</i> editing site cassette			
<i>clpP1_</i> GA.FOR	TTATGTCAGCAACAGAAGCCCAAGCTCATGGAG CTGAGCAATAACTAGCATAACCCCTTG	Insertion of <i>clpP1</i> editing site cassette			
<i>clpP1_</i> GA.REV	GAGCTTGGGCTTCTGTTGCTGACATAAAAACAT CCTCGAGTTACCAATAATCTCCACAAG	Insertion of <i>clpP1</i> editing site cassette			
clb19-3_RP	TTGATTTCTTGGACGGCTATG	Genotyping clb19-3 mutants			
clb19-3_LP	GAGGAAGCGACAATGGTTATG	Genotyping clb19-3 mutants			
LBb1.3	ATTTTGCCGATTTCGGAAC	Genotyping SALK T-DNA lines			
MORF9_75.FOR	GGTGGTCCATGGAGCAGAGAGAGACGATTATG CTTCCTGGTTCCGATTAC	Cloning MORF9 75-196, introducing C85S mutation			
MORF9_196.REV	GGTGGTCTCGAGTCACTGCTTTGGTTGGTATGT TGGGTATGTGGACGGGATAATCTC	Cloning MORF9 75-196, introducing C187S mutation			
rpoA_ES.FOR	ATGGTTCGAGAGAAAGTC	Amplifying A. thaliana rpoA cDNA			
rpoA_ES.REV	GATAAGGTCTTCTTGACTG	Amplifying A. thaliana rpoA cDNA			
rpoA_PS.FOR	CAGTGGAAGTGTGTTGAATC	Sequencing A. thaliana rpoA editing site			
<i>clpP1_</i> ES.FOR	TGCCTATTGGCGTTCCAAAAG	Amplifying A. thaliana clpP1 cDNA			

<i>clpP1_</i> ES.REV	TGAACCGCTACAAGATCAAC	Amplifying A. thaliana clpP1 cDNA
<i>clpP1_</i> PS.FOR	GTATAGCATTCCCTCACGCTAG	Sequencing A. thaliana clpP1 editing site
MORF2_73.FOR	GGTGGTCCATGGCGACGGAGATGGCTCCTTTG TTTCCGGGATCCGATTATG	Cloning MORF2 73-193, introducing C82S mutation
MORF2_193.REV	GGTGGTCTCGAGTCACTCAACCCGCCTCTGCC TCTCTGGTGATCGTTGGACTATC	Cloning MORF2 73-193

Supplementary Table S3. DNA primers used in this study.

Supplementary References

1. Yan, J. *et al.* MORF9 increases the RNA-binding activity of PLS-type pentatricopeptide

repeat protein in plastid RNA editing. Nat Plants 3, 17037 (2017).