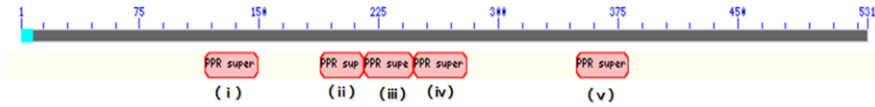


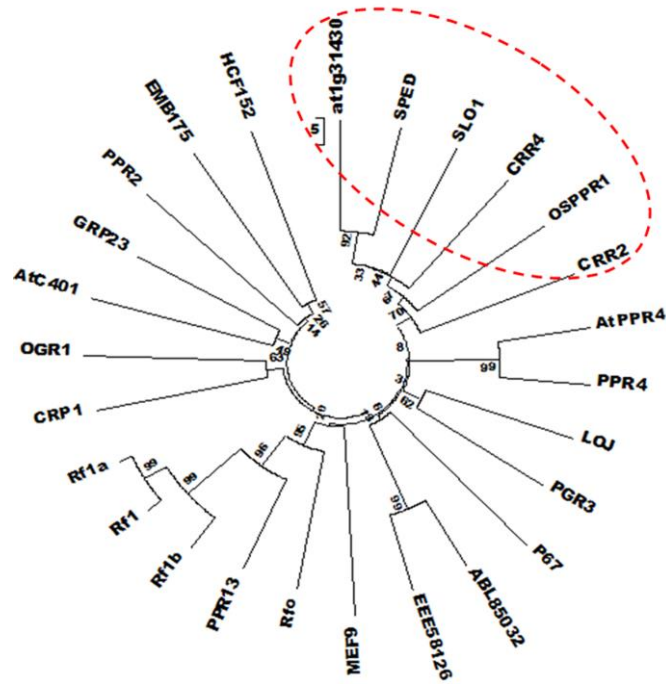
Figure S4

Cleaved sequence (34aa)

MAMAAARRGHGMPWECNVLIRTLARRGSFARYMAVYYDLRARGLVADSFTYPFVLR
AVGVLKLSVEGRKAHAAAVKTGFRWDAYTGSSLMEMYTMLGRVDIARKVFDEMPSRALVLVW
 * NMMVRCYIRCGWYSAAVALSEQMERSGV (i)
 TPDRVTLVTAVTACSRARDLSLGRRIHVYMDNVFGENLPVANALLDMYTKNDCLEEAVKLFEO
MPARNIISWITLVSGYGLAGQDKARVLFNQCKEKDLILWTAMINACVQHGCFEEALTLFFRDM
 A) QMQRV (ii)
 EPDRFTVVTLTCCANLGALDQGEWIHQYAEQRKMKIDAVLGTALIDMYSKCGHIEKSLEVFW
RMQGRD (iii)
ATAWTAIICGLATNGOAGRALELFQDMQRSKVKPDGVTFIGVLSACCHGGLVDEGRKQFHAM
REV (iv)
 YQIEPRVEHYSCLVNLLGRAGLLDEAERLIGDV (v)
 PINKDAMPLFGALLTACKAHGNVEMSERLTKRICEQDSQITDVNLLMSNVYATASRWEDVIRV
 RGKMAHPTVKKNAGCSLIEVKGY.



B)



E motif

C)

sped	LFGALLTACKAHGNVLSERLTKRICEQDSQITDVNLLMSNVYATASRWEDVIRVRGKMA
Slo1	VWGALLFCRMHGNVELGEKAKKILELDPSSDGIYVLLDGMYSANMWEDAKRAARMAN
MEF9	VWGALLDACRIYNNVGLAHVAEAASRLPESSIPYVLLYNNMYADMGLWDEASQVRMNE
At1g31430	LFGALLTACKAAGNVLSERLTKRICEQGYQIPDVNLLMSNVYATASRWEDAIRVRSKMA
CRR4	IWRITLITACKSHKEFETGELVAKHILIQAGYNPSSYVLLSNMYASFGMWRIVRRVRTMMK

sped	HPTVKKNAGCSLIEVKGY-----
Slo1	ERGVFKIFGCSLIEVNGIVCFEIVRDKSRPESEKIYDRLHCLGRHMRSLSLVLFSEYEIT
MEF9	SKRIKKERGSWVDSST-----
At1g31430	HPTLKKTAGCSLIEVKGH-----
CRR4	ERKIPKIFGCSWIEIDGRVHEFFVDSIEVSSTL-----

D)

```

61      70      80      90      100     110     120
|-----|-----|-----|-----|-----|
OsSPED1 VLKLSVEGRKAHAAAVKTGFRWDAYTGSSLMEHYTHLGRVDIARKVFOEMPSRALVLMNH
BdSPED1 VLKLSVEGRKAHAAAMKTGFWDAYTASSLMDHYTLGRADVARKLFDEMPHRAALVLMNH
HvSPED1 ALQISVEGRKAHAAATLKTGFRWDAYTASSLMDHYTHLGRLEAARKVFOEMPQRRAVVMNH
SsSPED1 VMKLSVEGRKAHAAAVKTGFRWDAYTASSLMDHYAHLGRADLARKVFOEMPRRFLVLMNH
SbSPED1 VMKLSVEGRKVHAAAVKTGFRWDAYTASSLMEHYTHLGRVDFARKVFOEMPQRFLVLMNH
ZnSPED1 THKLSVEGRKVHAAAVKTGFRWDAYTACSLMEHYTHLGRADSARRVFOEMPQRFLVLMNH
RtSPED1 RLKRVIEGEKVVHGYAVKAGLEFDQSYVSNLSLGMHYASLGKIEITHKVFDEMPQRDQVSHNG
Consensus .#kls!EGrKaHaaavKtGfrwDayTasSL$.MYtLGr.#.arkvFOEMPQR.LV.LMNH

121     130     140     150     160     170     180
|-----|-----|-----|-----|-----|
OsSPED1 HVRCYIRCGHYSRAVALSEQM-ERSGVTPDRVTLVTAVTACSRAROLSLGRRIHVYMDNV
BdSPED1 HIRCYVRCGRYTAALAEEM-ERSGLTPDKVTLVTSVTVCSRAGDLSLGRRIHAYMDGV
HvSPED1 MLRCYVRCGRNTEAVALAEEM-ERGRLTPDRVTLTALTACSRAGDLSLGRKIHAYMDGV
SsSPED1 HIRCYVRCGRFTAVALAEEM-ERSGATPDRVTLVTAVTACSRAGDLSLGRRIHAYMDAV
SbSPED1 HMRCYIRCGRFTAVALAEEM-ERSGATPDRVTLVTAVTACSRAGDLSLGRRIHAYMDGV
ZnSPED1 HMRCYIRCGRFTAVALAVQM-ESGGATPDRVTLVTAVTACSRAGDLSLGRRIHAYMDGV
RtSPED1 LISSYVGNRFDRIIGVFKRMSQESNLKFDGTYSTLSACSAKLNLEIGERIYRFVYTE
Consensus $ircY!rcGrftaA!alae.H.#rsgltpDrvtLvtavtaCSrag#LslGrrIh.%nd.v

181     190     200     210     220     230     240
|-----|-----|-----|-----|-----|
OsSPED1 FGFNLPVANALLDHYTKNDCLEEARVKLFEQMPARNIISWITILVSGYGLAGQLDKARVLFN
BdSPED1 FGFSLPVANALLDHYMKNGCLEEARVKLFEQMPARNIISWITILVSGYAFAGQLDKARVLFY
HvSPED1 TGFSLPVANALLDHYVKNKGCLEARVNLFEKMPARNVVSWITILVSGYAFAGQVQKARLLFH
SsSPED1 FGFNLPVANALLDHYTKNGCLEEARVKHFEQMPARNIISWITILVSGYALAGQLDKARALFY
SbSPED1 FGFSLPVANALLDHYTKNGYLEEARVKHFEQMPARNIISWITILVSGYAVAGQLDKARMLFY
ZnSPED1 FGFSLPVANALLDHYTKNGYLEEARVKHFEQMPARNIISWITILVSGYGIAGQLDKARVFFY
RtSPED1 FEHSVRIGNALVDMFCCKGCLDKARAVFDSHRDKNVCKWTSWVFGYVSTGRIDEARVLF
Consensus fgfs!p!aNALIDH%.Kngcl#eAvk.F#qHp.rN!isWTi$VsGY..aGqIDkARVLF.

241     250     260     270     280     290     300
|-----|-----|-----|-----|-----|
OsSPED1 QCKEKDLILHTAMINACVQHGCFFEEALTLFRDQHQRVPEPDRFTVVTLTCCANLQALDQ
BdSPED1 QCSEKDLIMHTAMINACVQHGCFFEEALSLFREHQHQRVPEPDRFTIVTLTCCANLQALDQ
HvSPED1 QCTEKDLIMHTAMINAYVQHGCFFIEALSLFRDQHQHQIEPDRFTVVTLTCCANLQALDQ
SsSPED1 QCTEKDLILHTAMINACVQHGGSFFEEALSLFRDQHLQRVPEPDRFTVVTLTCCANIGTLDQ
SbSPED1 QCTQKDLILHTAMINACVQHGSFFEEALTLFRDQHLQRVPEPKFTVVTLTCCANIGALDQ
ZnSPED1 QCTQKDLILHTSMINACVQHGSFFEEALILFRDQHLQRVPEPKFTVVTLTCCANIGALDQ
RtSPED1 RSPVKDQVVLHTAMHNGYVQFNRFDEALELFRCHQTAGIRPNFVLSLLTGCAQTGALEQ
Consensus qc.eKDL!$HTaHINacVQhg.FeEAL.LFRdHQ.qr!ePDrFvVtLLTcCA#.GaL#Q

301     310     320     330     340     350     360
|-----|-----|-----|-----|-----|
OsSPED1 GEWIHQYAEQRKMKIDAVLGTALIDMYSKCGHIEKSLSEVFRHQGRDATAHTAICGLAT
BdSPED1 GEWIHQFAVDRKMKVDAVLGTALIDMYAKCGHVKKSHEVFEHQGRDTTANTSIICGLAT
HvSPED1 GQLIHQFAEGRNHNKIDAVLGTALIDMYAKCGHVEKSVVFERHNEGROTKAHTAICGLAT
SsSPED1 GEWIHQYAEGRKMKIDAVLGTALIEHYSKCGHVOKALDVFERNQGRDAAHTAICSLAT
SbSPED1 GEWIHQYAEGRNHNKIDAVLGTALIEHYSKCGHVOKSLQIFGRHQGRDAAHTAICGLAT
ZnSPED1 GEWIHQYAESRNHNKIDAVLGTALIEHYSKCGHVOKSLQIFGRHQGRDAAHTAICGLAT
RtSPED1 GKWIHGYINENRVTVQKVVGTALVDMYAKCGCIETALEVFEYIEKERDTASHTSLIYGLAM
Consensus GewIHq$ae.r.nk.DaVlGTAL!$HYaKCGh!eksl#!F.rnqgrDtaaHTaiIcGLat

361     370     380     390     400     410     420
|-----|-----|-----|-----|-----|
OsSPED1 NGQAGRALELFDHQRSKVKPDGVTFIGVLSACCHGGLVDEGRKQFHAMREYVQIEPRVE
BdSPED1 NGQAGRALELFDHRSKVKPDSITFIGVLSACCHGGLVDEGRKQFHAMKOVYRIIPRIE
HvSPED1 NGQAGRALELFDHERSEAKPDSITFIGVLSACCHGGLVDEGRKQFHAMKEYVRIIPRIE
SsSPED1 HGQASRALELFEHQISKVKPDGITFIGVLSACCHGGLVDEGRKQFHAMKEYVRIIPRIE
SbSPED1 NGQASKALELFEHQRSKTKPDGITFIGVLSACCHGGLVDEGRRHFFQAMKEYVQIEPRIE
ZnSPED1 NGQASKALELFEHQRSKTKPDGITFIGVLSACCHGGLVDEGRQRYFQAMKEYVRIIPRIE
RtSPED1 NGSGRALDLYEHENVGVRLDITFVAVLTACNHGGFVAEGRKIFHSHTERHNVPKSE
Consensus nGqagrAL#L$e#H#rskvkpD.!TF!gVLsACcHGGlVdEGrk.Fhamk#vy.!Pr.E

```

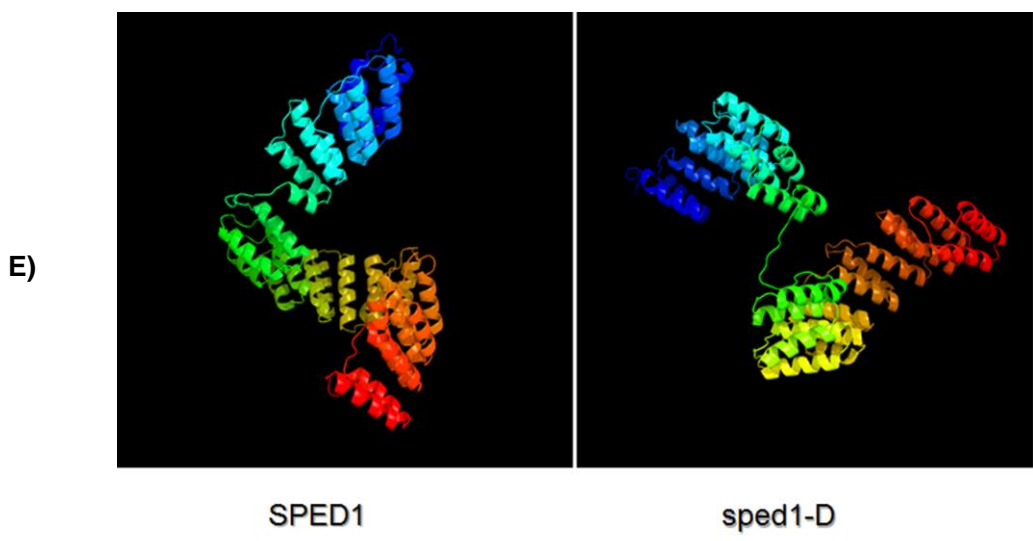


Figure S4 Sequence and structure of SPED 1

A) SPED 1 is a PPR protein containing 12 PPR motifs. The positions of the two changed amino acid, K70T and 414L in sped1-D are indicated by asterisks.

B) Phylogenetic analysis showed that SPED or sped has the highest sequence similarity with at1g31430 and SLO1 in Arabidopsis. Sequences used in this analysis were as follows: At1g31430, MEF9 (At1g62260), SLO1 (At2G22410), CRR4 (At2g45350), CRR2-At3g46790, AtPPR4 (At5g04810), LOJ (At2g39230), PGR3 (At4g31850), P67 (At4g16390), AtC401 (At5g21222), GRP23 (At1g10270), EMB175 (At5g03800) and HCF152 (At3g09650) from Arabidopsis; sped (NP_001057989), OSPPR1 (AAS93059), Rf1a (ABC42330), Rf1b (ABC42331), EEE58126 and OGR1 (ACL79585) from rice; PPR4 (NP_001105869), CRP1(NP_001105879) and PPR2 (ACG30078) from maize; ABL85032 from *Brachypodium sylvaticum*; Rfo (ACJ70132) from *Brassica napus*; PPR13 (XP_002436844) from sorghum; and Rf1 (AAM52341) from petunia.

C) Sequence alignment of sped 1, Slo1, MEF9, At1g31430 and CRR4 also showed that the sped 1 contains a truncated E motif.

D)Protein sequence alignment of SPED1 homologs from Arabidopsis (AtSPED1, At1g31430), *Setaria italic* (SiSPED1, XP_004967201.1), *Sorghum bicolor* (SbSPED1, XP_002438633), *Zea mays* (ZmSPED1, AFW76438) and *Brachypodium distachyon* (BdSPED1, XP_003563682).The two changed amino acid,K70T and I414L in sped1-D are both located in a conservative domain. Their positions are indicated by asterisks.

E) Prediction of 3-D structures of SPED1 and sped1-D proteins by using phyre2.