El-Shami et al. Supplementary Material

Supplementary Methods.

Methods in sequence analysis

Sequence alignments were carried out using either the BLASTN or PSI-BLAST programmes (Altschul et al. 1997).

Methylation detection assays

Genomic DNA was extracted from leaves using the Wizard Genomic kit (Promega) and was digested with the methylation-sensitive enzyme *HaeIII*. The *AtSN1* PCR-based methylation assay was performed as described (Pontier et al. 2005). The 5S rDNA Southern was performed on 0.2 μg of genomic DNA digested by the methylation-sensitive enzyme *HaeIII* and separated on 0.9 % agarose gel. After blotting on Hybond N^+ membranes (Amersham Pharmacia Biotech Inc.), hybridization was performed in PerfectHyb solution (Sigma-Aldrich) following the supplier's instructions, with the transcribed region of the 5S rDNA gene.

Supplementary Table.

Table 1. Primers

Name	Sequence
204	GGATCCCCGACTTCTGACGTTTGGGG
205	GTCGACAGGCTGAGACTGTTCTTCCC
354	AAGCTTGGTATGATATCATCTGATACTTCATTTTGGAG
456	GTCGACATGGAGGAAGAATCTACATCAGAGATTCTTGAC

457	CCATGGGAGGCTGAGACTGTTCTTCCCC
458	CCATGGGCTGTCTCGTTGCTGTAAACATGGG
466	GGATCCGTCGACCCCTAACGAGCGAAAATCACACACCAGCG
518	GGATCCCCTAGGCCAGCTAGATCTGAG
543	GGATCCGTAAACAGAACTGACTTAGATCCACG
544	GTCGACGTTGTATTTGCTTGGATCTCCCCAAGC
545	GATCTTCGGACAAGAAGAATTCGGAAACTGAATCGGGTCCAGCTGCTTGGG
546	GATCCCCAAGCAGCTGGACCCGATTCAGTTTCCGAATTCTTCTTGTCCGAA
547	GATCTTCGGACAAGAAGAATTCGGAAACTGAATCGGGTCCAGCTGCTTTCG
548	GATCCGAAAGCAGCTGGACCCGATTCAGTTTCCGAATTCTTCTTGTCCGAA
557	GGATCCAGCTGGGATGCATTTCC
558	GTCGACTTGCAGTTTATCATTCCAGCC
579	GATCTTCGGACAAGAAGAATTCGGAAACTGAATCGGGTCCAGCTGCTTGGG
580	GATCCCCAAGCAGCTGGACCCGATTCAGTTTCCGAATTCTTCTTGTCCGAA
581	GGATCCCCATGGCTCGAGCCTAGGCCACCTAGATCTGAGG
582	ACATGTGAGGCTGAGACTGTTCTTCCCC
621	GGATCCCCATGGGTAAACAGAACTGACTTAGATCCACG
622	TCTAGGTGGCCTAGGGTTGTATTTGCTTGGATCTCCCCAAGC
623	CCAAGCAAATACAACCCTAGGCAACCTAGATCTGAGG
658	CCATGGTGTCGTGGATATGACCATTTGAAGG
661	GATCTAAGGACCAGGCTAATCCAGAAGACTCTTCTAAGACTGGAGGATGGTCTG
662	GATCCAGACCATCCTCCAGTCTTAGAAGAGTCTTCTGGATTAGCCTGGTCCTTA
663	GATCTAAGGACCAGGCTAATCCAGAAGACTCTTCTAAGACTGGAGGATTCTCTG
664	GATCCAGAGAATCCTCCAGTCTTAGAAGAGTCTTCTGGATTAGCCTGGTCCTTA

Supplementary Figures.

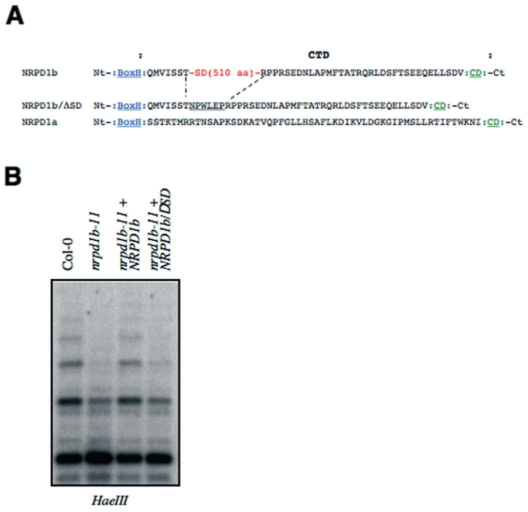
Figure S1. Functional and biochemical analysis of the truncated mutant NRPD1b/ΔSD. (*A*) The amino acid sequence of the linker separating the CD from the evolutionary conserved box H is indicated for NRPD1a/b and NRPD1b/ΔSD. (*B*) Transgene rescue of 5S rDNA methylation in *nrpd1b-11* mutant. Southern blot analysis of 5S rDNA digested with methylation-sensitive restriction enzyme *HaeIII* and hybridized to a 5S probe (DNA was extracted from pooled samples of the T1 transgenic lines shown in figure 1C).

Figure S2. The SD of AtNRPD1b is highly divergent and has a WG/GW repeat signature sequence. (*A*) Schematic representation of the Arabidopsis and spinach NRPD1b proteins. AtNRPD1b shows low sequence identity to SoNRPD1b in the SD (red) but not in the N-terminal RPB1-like region (blue) and the C-terminal CD (green). (*B*) Alignment of AtNRPD1b and SoNRPD1b CTDs. The WG/GW repeats present in the homologuous region are highlighted in red. The color code is the same as in A.

Figure S3. A WG motif is located in a highly conserved region of the imperfect 16-aa repeats of the extended CTD of AtNRPD1b. The repeated motifs of the AtNRPD1b CTD were aligned and the conserved WG motif is highlighted in grey.

Figure S4. The SD of AtNRPD1b is highly divergent and has a WG/GW repeat signature sequence. (*A*) Schematic representation of the Arabidopsis and rice NRPD1b proteins. AtNRPD1b shows low sequence identity at an amino acid level to OsNRPD1b in the SD (red) but not in the N-terminal NRPB1-like region (blue) and the C-terminal CD (green). (*B*) Alignment of AtNRPD1b and OsNRPD1b CTDs. The WG/GW repeats present in the homologuous region are highlighted in black. The color code is the same as in A.

Figure S5. Amino acid composition of GW-rich regions from AtNRPD1b and HsGW182. Amino acid content, as a percentage of molecular mass, was calculated for the GW/WG-rich regions described in the manuscript. Global amino acid percentages were obtained from the Codon Usage Database (http://www.kazusa.or.jp/codon/). Column 4 shows the number of residues for each amino acid, column 5 the percentage of molecular mass, column 6 the percentages for Arabidopsis or man from the Codon Usage Database (based on NCBI-GenBank Flat File Release 156.0 [October 15 2006]) and column 7 the ratio % in specific protein/global % of molecular mass. Amino acids are ordered by physico-chemical properties. Overrepresented residues in specific proteins are overlined in red, underrepresented residues in blue.

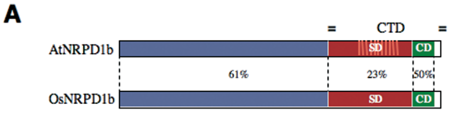
A
:
CTD
:
NRPD1b
Nt-:BoxH:QMVISST-SD(510 aa)-RPPRSEDNLAPMFTATRQRLDSFTSEEQELLSDV:CD:-Ct
NRPD1b/ΔSD
Nt-:BoxH:QMVISSTNPWLEPRPPRSEDNLAPMFTATRQRLDSFTSEEQELLSDV:CD:-Ct
NRPD1a
Nt-:BoxH:SSTKTMRRTNSAPKSDKATVQPFGLLHSAFLKDIKVLDGKGIPMSLLRTIFTWKNI:CD:-Ct
B
Col-0
nrpd1b-11
nrpd1b-11 + NRPD1b
nrpd1b-11 + NRPD1b/ΔSD
HaeIII

```
AtD1b 1215 -SWGKRVDVGTGSQFELLWNQKETGLDDKEETDVYSFLQMVISTTNADAFVSSPGFDVTEE
            SWGKRV +GTG++F+LLW  KE  + DK  TDVY+FL +V S    +      G D+
SoD1b 1015 -SWGKRVSIGTGAKFDLLWETKEIEMADK-PTDVYNFLHLVSSANEEEVDSGGLGEDI---

AtD1b 1275  EMAEWAESPERDSALGEPKFEDSADFQNLHDEGKPSGANWEKSSSWDNGCSGGSEWGVSK
                + +    + AL  P+ E+ A F+   + G  S       SSWD   S G+ W  +K
SoD1b 1071  --ESFEKDVYMEPAL-SPEQENKAVFEETLEIGVDSDITGADESSWDAFPSSGTGWNANK

AtD1b 1335  STGGEANPESNW----EKTTNVEKEDA-----WSSWNTRKDAQESSKSDSGG-----AWG
               G  + E  W     K      ED+     WSS  +++  Q    S SGG     +WG
SoD1b 1128  IDTGSGSAEGGWSSWGSKKDQANPEDSSKTGGWSSGGSKQKPQPEDSSKSGGWDASKSWG

AtD1b 1381  IKTKDADADTTPNWETSPAPKDSIVPENNEPTSDV----WG---HKSVSDKSWDKKNWGT
               +    D +P W       + I  EN+  +       W   G K    K  +    G
SoD1b 1188  GSNQ---GDPSPVWGQPVKATNDISIENDHGSGSAEGGGWANSGMKKDLSKQENSSTAGG

AtD1b 1434  ESAPAAWGST----DAAVWGSSDK-------KNSETESDAAAWGSRDKNNSDVGSGAGVL
              A  +W  +     ++ WG+  K       K S+++ D A+    D   S  G    +L
SoD1b 1245  WDASKSWSGSKPKDPSSAWGAGKKTDDNNGWKKSDSKKDLASGSVEDGGCSGWGPKKDLL

AtD1b 1483  GP--------WNKKSSETESNGATWGS-----------------SDKTKSGAAAWN---S
             P        W    S+++   + WG                    +   G + WN
SoD1b 1305  QPEDSAGENGWGASKSKSKEPSSAWGKPAQETDNIGWKKNNPQRDSENLEGTSGWNDKLQ

AtD1b 1515  WDKKNIETDSEPAA---WGSQG---------------------------KKNSETESGPA
             + K+    S+PA+   W S G                           KK++  ++G
SoD1b 1365  KENKSFSKQSQPASSKDWDSTGNITAGSTGFGVEKGNEKPWDVASNVSVKKSTWGQTGGN

AtD1b 1545  AWGAWDKKKSETEPGPAGWGMGDKKNSETELG-----PAAMG------------------
            +W   ++ + + +P    WG   K +     G     P + G
SoD1b 1425  SWKKNEQDEKDGDPQGLPWGKSHKSSDSWTSGQGNQHPVSQGVSEKQGTLSSWGQPRDSS

AtD1b 1582  --------------------NWDKKKSDT-------KSGPAAWG-------------STD
                                +WD KK ++       + G + W              ST+
SoD1b 1485  QKNNNENGVSSNFNRQGAGKSWDSKKKESNVQSSWAQQGDSTWKDSKEARSSVKANNSTN

AtD1b 1602  --------------AAAWGSSDKNNSETESDAAAWGSRNKKTSEIESGAGAWG-------
                          +++WGS  ++  + +S+  + G  N      E G  +W
SoD1b 1545  SGGWSTGKALVDGVSSSWGSQKEDRPQPKSNDRSVGDGNFDKDAKEEGLSSWDAKKVERK

AtD1b 1640  ---SWGQPSPT--------------------------AEDKDTNEDD--RNPW----VSL
               SWGQPS +                          +ED +  +D   ++ W    VS
SoD1b 1605  TQSSWGQPSESKNSAQSSADHWGSDKSNQPGKSSGWGSEDTNAGKDSEKQDSWGKSNVST

AtD1b 1666  KETKSREK---DDKERSQWGNPA-----KKFPSSG-GWSNGGGADWKGNRNHTP------
             + +S EK    D  +S WG P      KK P  G GW +    +WK  +N
SoD1b 1665  WKKESGEKLHGSDDSQSPWGQPGGSGWNKKQPEGGRGWGSSNTGEWKSRKNQNQNQNQNQ

AtD1b 1711  -RPPRSEDNLAP--MFTATRQRLDSFTSEEQELLSDVEPVMRTLRKIMHPSAYPDGDPIS
             RPPR  ++ +P    TATR+R+D F +EE+++LS+VE +M+++R+IMH S   DG+P+
SoD1b 1725  NRPPRGPNDDSPRVALTATRKRMDEFPTEEKDVLSEVESLMQSIRRIMHQSGCVDGEPLL

AtD1b 1768  DDDKTFVLEKILNFHPQKETKLGSGVDFITVDKHTIFSDSRCFFVVSTDGAKQDFSYRKS-
             DD+T++++ ILN+HP K  K+G+GVDFITV KH+ F +SRCF+VVSTDG   DFSY K
SoD1b 1785  PDDQTYLIDNILNYHPDKAAKIGAGVDFITVKKHSNFQESRCFYVVSTDGKDTDFSYIKC-
```

AtNRPD1b repeats

```
1427-DKKNWGTESAPAAWGS-1442
1452-DKKNSETESDAAAWGS-1467
1486-NKKSSETESNGATWGS-1501
1516-DKKNIETDSEPAAWGS-1531
1533-GKKNSETESGPAAWGA-1548
1550-DKKKSETEPGPAGWGM-1565
1567-DKKNSETELGPAAMGN-1582
1584-DKKKSDTKSGPAAWGS-1599
1609-DKNNSETESDAAAWGS-1624
1626-NKKTSEIESGAGAWGS-1641
```

Consensus DKKNSETESgPAAWGS



B

```
AtD1b 1188 -LIAPRKCFEKAAEKCHTDSLSTVVGSCSWGKRVDVGTGSQFELLWNQKETGLDDKEETDV
            LI P KCFEKAAEKCH+DSL  VV SCSWGK    GTGS F++LWN+ +   + +    +
OsD1b 1199 -LITPMKCFEKAAEKCHSDSLGCVVSSCSWGKHAASGTGSSFQILWNESQLKSNKEYGDGL

AtD1b 1248  YSFLQMVISTTNADAFVSSPGFD-VTEEEMAEWAESPERDSALGEPKFEDSADFQNLH--
            Y +L +V +      +      D + EE  A+   SPE D  +G+P F+D+ + Q++
OsD1b 1259  YDYLALVRTDEEKARYTFFDDVDYLAEENEADVCLSPELDGTIGQPIFDDNLEEQDVQNN

AtD1b 1305  ---DEGKPSGANWEKSSS-------WD--NGCSGGSEWGVSKSTGGEANPESNWEKTTNV
               D G  + A+WE++ S       W   N  + G++ GV+K         S W+    V
OsD1b 1319  SSWDNGTTTNASWEQNGSAGNDSDKWGGWNDAAAGADTGVTKPAN---QGNSCWDVPATV

AtD1b 1353  EKEDA-WSSWNTRKDAQESSKSDSGG---AWGIKTKDADADTTPNWETSPAPKDSIVPEN
            EK  + W  W T K  ++   S+      AW ++      D   +W+     K S    +
OsD1b 1376  EKSSSDWGGWGTEKAKEKEKISEEPAQHDAWSVQGPKRATDGGASWK-----KQSSTQND

AtD1b 1409  NEPTSDVWGHKSVSDKSWDKKNWGTESAPAAWGSTDAAVWGSSDKKNSETESDAA-----
                 + G  S +  SW+K N    +   +WG  +     ++D +N   E+ AA
OsD1b 1431  GNSWKENKGRGS-NGGSWEKDN----AQKGSWGRGNDEAENNNDVQNKSWETVAADAHAS

AtD1b 1464  ---AWGSRDKNNSDVGSGAGVLGPWNKKSSETESNGATWGSSDKTKSGAAAWNSWDKKNI
               +WG+   + SD    A  +   N  SS+T+ + +  G        A   +     N+
OsD1b 1486  TEKSWGNVTASPSDNAWSAAPVSQGN-GSSDTKQSDSWDGWKSAGVDKAINKDKESLGNV

AtD1b 1521  ETDSEPAAWG----SQGKKNSETESGPAAWGAW---------DKKKSE-----TEPGPAG
                  +AW     SQG + S+ +    +W  W         +K K         P  +
OsD1b 1545  PASPSFSAWNASPVSQGNERSDAKQSD-SWDGWKSAGVDKAINKDKESLGNVPASPSFSA

AtD1b 1563  WGMGDKKNSETELGPAAMGNWDKKKS----DTKSGPAAWGSTDA----AAWGSSDKNNSE
            W           L      +WD  KS    D+     +WG+  A    +AW ++   +
OsD1b 1604  WNAAPVSQGNERLDAKQSDSWDGWKSAGVDDSVKDKESWGNVPASPSDSAWNAAPVSQGN

AtD1b 1615  TESDAA------AWGSRNKKTSEIESGAGAWGSWG-QPSPTAEDKDTNEDDRNPWVSLKE
              SDA        W S     S       +WG+    PS +A +        + W S +
OsD1b 1664  ESSDAKQSDSWDGWKSAGVDAS-TNKDKESWGNVPASPSDSAWNAAPVSQGDDVWNSAEA

AtD1b 1668  TKSREKDDKERSQWGNPAKKFPSSGGWSNGGGADWKGNRNHTPRPPRSEDNLA----PMF
             +SR KD K              S GW   GG +W+G RN+  RPPR  D       P
OsD1b 1723  NESRNKDWK--------------SDGWGARGG-NWRGQRNNPGRPPRKPDGRGLPRRPDE

AtD1b 1724  TATRQRLDSFTSEEQELLSDVEPVMRTLRKIMHPSAYPDGDPISDDDKTFVLEKILNFHP
                +R    T+EE+++L ++EP + ++RKI   S   D   +S +D+ F+ E +L  HP
OsD1b 1768  RGPPRRHFDLTAEEEKILGEIEPTVLSIRKIFRESI--DSIKLSPEDEKFIKENVLEHHP

AtD1b 1784  QKETKLGSGVDFITVDKHTIFSDSRCFFVVSTDGAKQDFSYRKSLNNYLMKKYPDRAEEF
            +K++K+   +D I VDKH +F DSRC FVVS+DG + DFSY K + N++ K YP+  + F
OsD1b 1826  EKQSKVSGEIDHIMVDKHQVFQDSRCLFVVSSDGTRSDFSYLKCMENFVRKTYPEHGDSF

AtD1b 1844  IDKYFTKPRPS-GNRDRNNQDATPPGEEQS-
              KYF + R      D      TP G  QS
OsD1b 1886  CKKYFKRRRDQPPAADGGTAPGTPAGATQS-
```

AtNRPD1b

code	aa	name	n	NRPD1b (%)	global (%)	D1b/global
M	met	methionine	2	0,5	2,45	0,2
I	ile	isoleucine	4	1,1	5,26	0,21
L	leu	leucine	3	0,8	9,35	0,09
V	val	valine	9	2,5	6,74	0,37
F	phe	phenylalanine	1	0,3	4,25	0,07
Y	tyr	tyrosine	-	-	2,83	0
W	trp	tryptophan	29	7,9	1,25	6,32
G	gly	glycine	39	10,7	6,58	1,63
A	ala	alanine	40	11	6,51	1,69
S	ser	serine	54	14,8	8,93	1,66
T	thr	threonine	26	7,1	5,12	1,39
C	cys	cysteine	-	-	1,77	0
P	pro	proline	19	5,2	4,88	1,07
D	asp	aspartic acid	31	8,5	5,38	1,58
E	glu	glutamic acid	31	8,5	6,65	1,28
N	asn	asparagine	24	6,6	4,32	1,53
K	lys	lysine	42	11,5	6,35	1,81
R	arg	arginine	6	1,6	5,4	0,3
Q	gln	glutamine	4	1,1	3,46	0,32
H	his	histidine	1	0,3	2,25	0,13

HsGW182

code	aa	name	n	GW182 (%)	global (%)	GW182/global
M	met	methionine	5	0,9	2,21	0,41
I	ile	isoleucine	13	2,3	4,42	0,52
L	leu	leucine	15	2,7	10,02	0,27
V	val	valine	16	2,9	6,08	0,48
F	phe	phenylalanine	3	0,5	3,79	0,13
Y	tyr	tyrosine	2	0,4	2,74	0,15
W	trp	tryptophan	35	6,3	1,32	4,77
G	gly	glycine	75	13,5	6,6	2,05
A	ala	alanine	30	5,4	6,97	0,77
S	ser	serine	84	15,1	8,1	1,86
T	thr	threonine	42	7,6	5,31	1,43
C	cys	cysteine	4	0,7	2,31	0,3
P	pro	proline	40	7,2	6,11	1,18
D	asp	aspartic acid	34	6,1	4,7	1,3
E	glu	glutamic acid	31	5,6	6,84	0,82
N	asn	asparagine	45	8,1	3,6	2,25
K	lys	lysine	34	6,1	5,62	1,09
R	arg	arginine	17	3,1	5,68	0,55
Q	gln	glutamine	27	4,9	4,64	1,06
H	his	histidine	4	0,7	2,59	0,27