A

```
NA93-11  1    ATGAGGACGGCGGCCACGCCGCCTCT...CGCCGCCGCCGCCGCCGCCGTCGCGGCAGTG
OIL31    1    ATGAGGACGGCGGCCACGCCGCCTCTCGCCGCCGCCGCCGCCGCCGCCGTCGCGGCAGTG

NA93-11  58   TTCCTCTCTGCGTTGCTGCTCGCCTCCGCCTCCGCCTCCGCCTCCAGGCTCCCTCCTCCT
OIL31    61   TTCCTCTCTGCGTTGCTGCT......CGCCTCCGCCTCCGCCTCCAGGCTCCCTCCTCCT

NA93-11  118  CGCCGTCTTCTTCCCCTGGTACGCGTACGCCCTCGCCGCCGCCGTTCTTGATCGACCTGA
OIL31    115  CGCCGTCTTCTTCCCCTGGTACGCGTACGCCCTCGCCGCCGCCGTTCTTGATCGACCTGA

NA93-11  178  GTGTTTTTTAAAAATTTTGTTCGGTTAGTTGATGCGCTGTGTGTTCGTCCATGGCGATGC
OIL31    175  GTGTTTTTTAAAAATTTTGTTCGGTTAGTTGATGCGCTGTGTGTTCGTCCATGGCGATGC

NA93-11  238  CGATGCAGGTTGGTGGCGAGGTGGCGGTGGCGGTGGTGGCTGGGGAGGAGGAGAAGGTGC
OIL31    235  CGATGCAGGTTGGTGGCGAGGTTGCGGTGGCGGTGGTGGCTGGGGAGGAGGAGAAGGTGC

NA93-11  298  GGCTGGGGTCGAGCCCGCCGAGCTGCTACAGCAAGTGCTACGGGTGCAGCCCGTGCGTCG
OIL31    295  GGCTGGGGTCGAGCCCGCCGAGCTGCTACAGCAAGTGCTACGGGTGCAGCCCGTGCGTCG

NA93-11  358  CGGTGCAGGTGCCCACCTTGTCCGCCCCGTCCGTCCCCGCCGCCGCCGCCGCCGCCGCAC
OIL31    355  CGGTGCAGGTGCCCACCTTGTCCGCCCCGTCCGTCCCCGCCGCCGCCG.......CGCAC

NA93-11  418  GACGCCGCGCCGCTCGTGGCGACGTTCACCAACTACAAGCCGCTAGGGTGGAAGTGCCAG
OIL31    408  GACGCCGCGCCGCTCGTGGCGACGTTCACCAACTACAAGCCGCTAGGGTGGAAGTGCCAG

NA93-11  478  TGCCGCGACCGCCTGTTCGACCCCTGA
OIL31    468  TGCCGCGACCGCCTGTTCGACCCCTGA
```

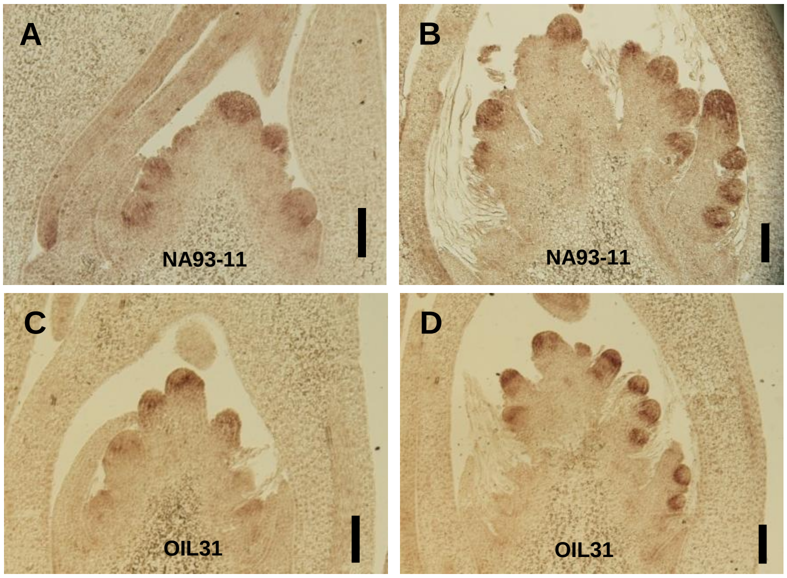
B

```
NA93-11  1    MRTAATPPL.AAAAAAVAAVFLSALLLASASASASRLPPPRRLLPLVGGEVAVAVVAGEE
OIL31    1    MRTAATPPLAAAAAAAVAAVFLSALLL..ASASASRLPPPRRLLPLVGGEVAVAVVAGEE

NA93-11  60   EKVRLGSSPPSCYSKCYGCSPCVAVQVPTLSAPSVPAAAAAAARRRAARGDVHQLQAARV
OIL31    59   EKVRLGSSPPSCYSKCYGCSPCVAVQVPTLSAPSVPAAAAHDAAPLVATFTNYKPLGWKC

NA93-11  120  EVPVPRPPVRPLTLRRARPVARRGVAWRVHGGARARALAVNYGVCGRVACPAAHGAALLL
OIL31    119  QCRDRLFDP

NA93-11  180  MLVVVESLSSRRAERDC
```

## Supplemental Figure 1

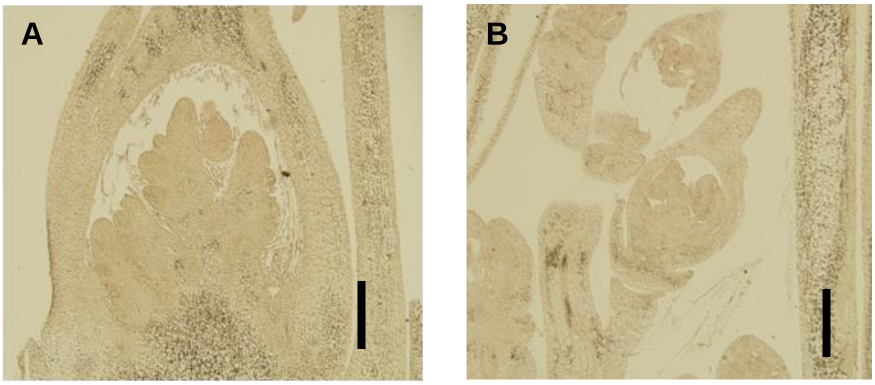
*GAD1* sequence comparison between OIL31 and NA93-11. **(A)** Comparison of the genomic sequences in the *GAD1* coding region between NA93-11 and OIL31. **(B)** Comparison of the GAD1 amino acid sequences between NA93-11 and OIL31.

The red font indicates the different sites. Yellow rectangles represent Indels and SNPs between NA93-11 and OIL31.



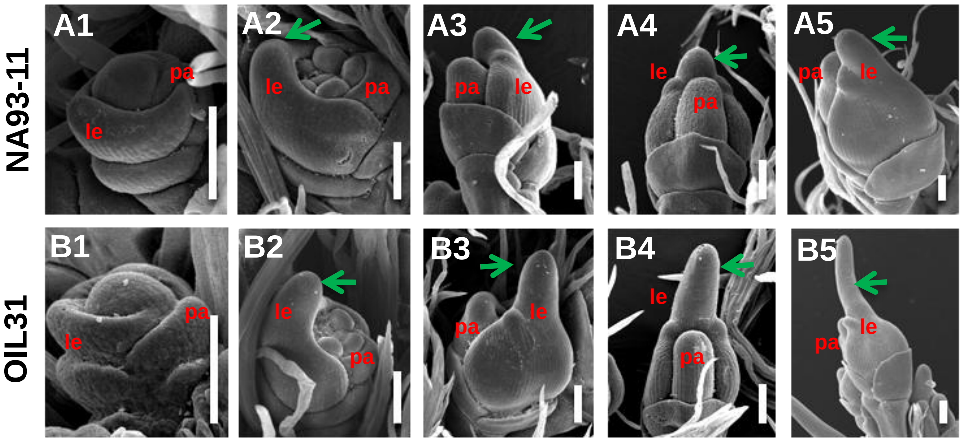
**Supplemental Figure 2**

RNA *in situ* hybridization of *GAD1* in inflorescence meristem. **(A)** RNA *in situ* hybridization of *GAD1* at the primary branch formation stage in NA93-11. Bar = 50 μm. **(B)** RNA *in situ* hybridization of *GAD1* at the secondary branch formation stage in NA93-11. Bar = 50 μm. **(C)** RNA *in situ* hybridization of *GAD1* at the primary branch formation stage in OIL31. Bar = 50 μm. **(D)** RNA *in situ* hybridization of *GAD1* at the secondary branch formation stage in OIL31. Bar = 50 μm.



**Supplemental Figure 3**

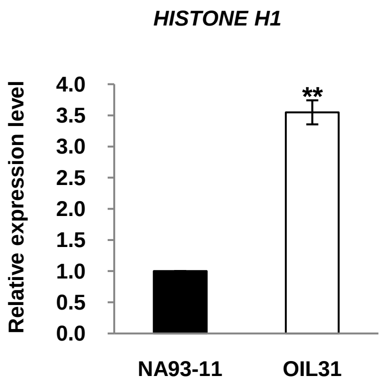
RNA *in situ* hybridization of the *GAD1* sense probe at different developmental stages. **(A)** Young panicle at the second branch formation stage. Bar = 100 µm. **(B)** Young panicle at the spikelet developmental stage. Bar = 100 µm.



**Supplemental Figure 4**

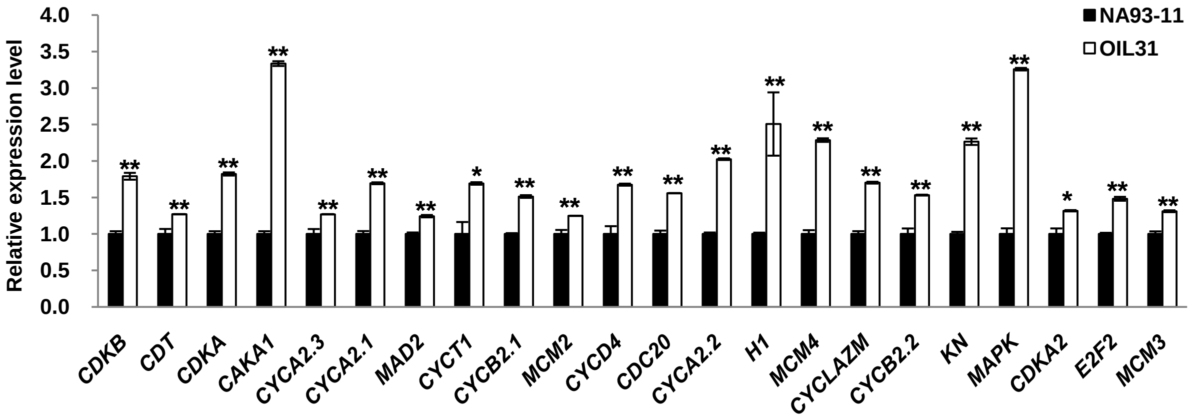
Scanning electron microscopy images of spikelets at different developmental stages. **(A1)** to **(A5)** Scanning electron microscopy images of spikelets at different developmental stages in NA93-11. Bar = 100 μm. **(B1)** to **(B5)** Scanning electron microscopy images of spikelets at different developmental stages in OIL31. Bar = 100 μm.

Green arrows point to the tip of lemmas.



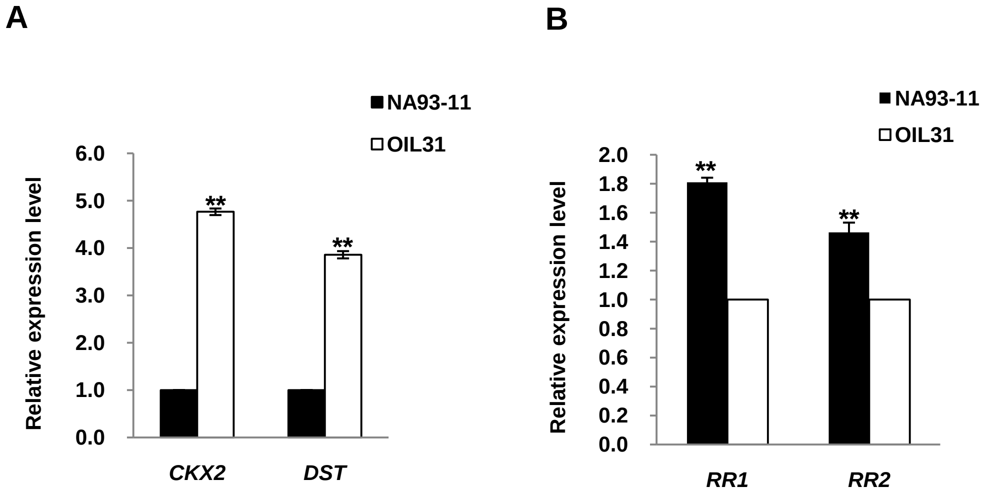
**Supplemental Figure 5**

qRT-PCR analysis of *HISTONE H1* in young panicles of NA93-11 and OIL31. The data represent the average of three independent biological replicates. The statistical significance was set at $P < 0.05$ based on a two-tailed Student's t test. Error bars represent the SD.



**Supplemental Figure 6**

Transcript level comparison of 22 cell cycle-associated genes between NA93-11 and OIL31. The apices of glumes were subjected to qRT-PCR. The data represent the average of three independent biological replicates. The statistical significance was set at $P < 0.05$ based on a two-tailed Student's t test. Error bars represent the SD.



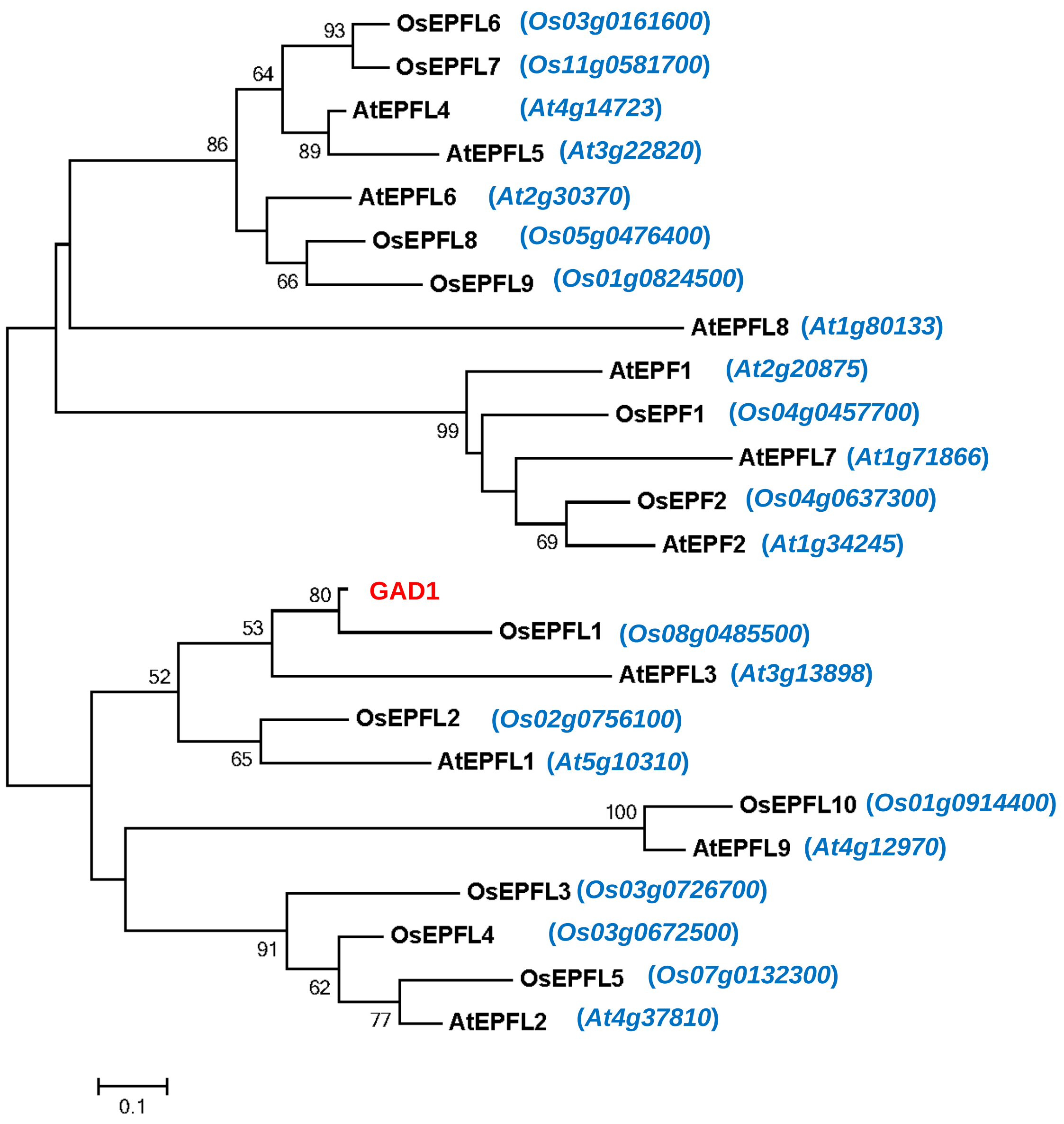
**Supplemental Figure 7**

Comparison of *CKX2*, *DST* and *RR* transcript levels between NA93-11 and OIL31. **(A)** Comparison of *CKX2* and *DST* transcript levels in the young panicles of the NA93-11 and OIL31. **(B)** Comparison of *RR1* and *RR2* transcript levels in the young panicles of the NA93-11 and OIL31. The data represent the average of three independent biological replicates. The statistical significance was set at $P < 0.05$ based on a two-tailed Student's t test. Error bars represent the SD.

```
1           attttgaccagaccacctcgctgcctctgctctgctcagtgctgctcactgcctcactcacactgtcacagt
73       gctcttctgggtataagtagctgggcgcgcggcgtcaccttccttggctgcccgtgagcttcccgcgcgcgccatggc
151      acggcctcgtgggcgtgggcgagcttcttggtaggcgaggtgtcgccgggatcgagttgtctcgagagctactcagct

229       ATGAGGACGGCGGCCACGCCGCCTCTCGCCGCCGCCGCCGCCGCCGCCGTCGCGGCAGTGTTCCTCTCTGCGTTG
1          M  R  T  A  A  T  P  P  L  A  A  A  A  A  A  A  V  A  A  V  F  L  S  A  L
304       CTGCTCGCCTCCGCCTCCGCCTCCAGGCTCCCTCCTCCTCGCCGTCTTCTTCCCCTGGTTGGTGGCGAGGTTGCG
26         L  L  A  S  A  S  A  S  R  L  P  P  P  R  R  L  L  P  L  V  G  G  E  V  A
379       GTGGCGGTGGTGGCTGGGGAGGAGGAGAAGGTGCGGCTGGGGTCGAGCCCGCCGAGCTGCTACAGCAAGTGCTAC
51         V  A  V  V  A  G  E  E  E  K  V  R  L  G  S  S  P  P  S  C  Y  S  K  C  Y
454       GGGTGCAGCCCGTGCGTCGCGGTGCAGGTGCCCACCTTGTCCGCCCCGTCCGTCCCCGCCGCCGCCGCGCACGAC
76         G  C  S  P  C  V  A  V  Q  V  P  T  L  S  A  P  S  V  P  A  A  A  A  H  D
529       GCCGCGCCGCTCGTGGCGACGTTCACCAACTACAAGCCGCTAGGGTGGAAGTGCCAGTGCCGCGACCGCCTGTTC
101        A  A  P  L  V  A  T  F  T  N  Y  K  P  L  G  W  K  C  Q  C  R  D  R  L  F
613       GACCCCTGAccctgaggcgcgcgcgccccgtggcgcgccgtggcgtggcgtggcgcgtgcatggcggggctcgcgc
126        D  P  *

680      tcgcgctctcgctgtgaattacggtgtgtgtggccgcgtcgcgtgcccggcagcacatggcgctgcactgctgctgat
758      gctggtggtggtggaatcgttgtcgtcgcgtcgggctgagagggattgttgatagattccgtgtaatatgccaggaca
836      aaattttgtcaccgctgctgctgcccatgcagctggatcggctcggctactctcaccacctatactgtatctttcact
914      ggcatctgctcgccgttttggaatctctgcggtgtggggttgcctcttgcatgtacacatgtttttcatgtatcgatc
992      catgcctcctccatggaatctaatgggaatccatcatcgttcatgctggatggatggatggatgtagtgaatggtagt
1070     ttttcttatttttgttggagatggatatttttttactttacatctaatcggatatatgttgtcttttaaattgagaatt
1148     tagc
```

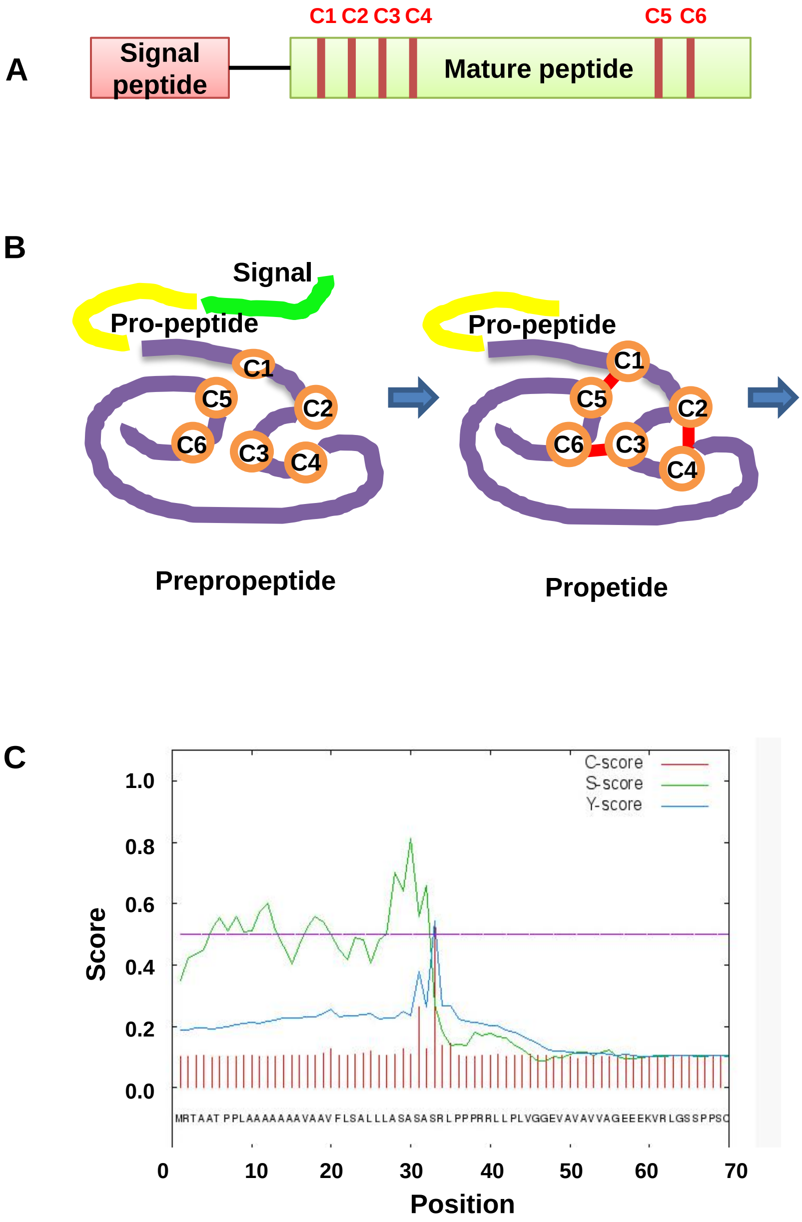
**Supplemental Figure 8**

Full-length cDNA of *GAD1* and its predicted amino acid sequence. *GAD1* cDNA in W2014 is 1,151-bp long, containing an ORF of 384-bp, with 228-bp and 539-bp 5′ and 3′ untranslated regions, respectively, encoding a 127-amino-acid protein.



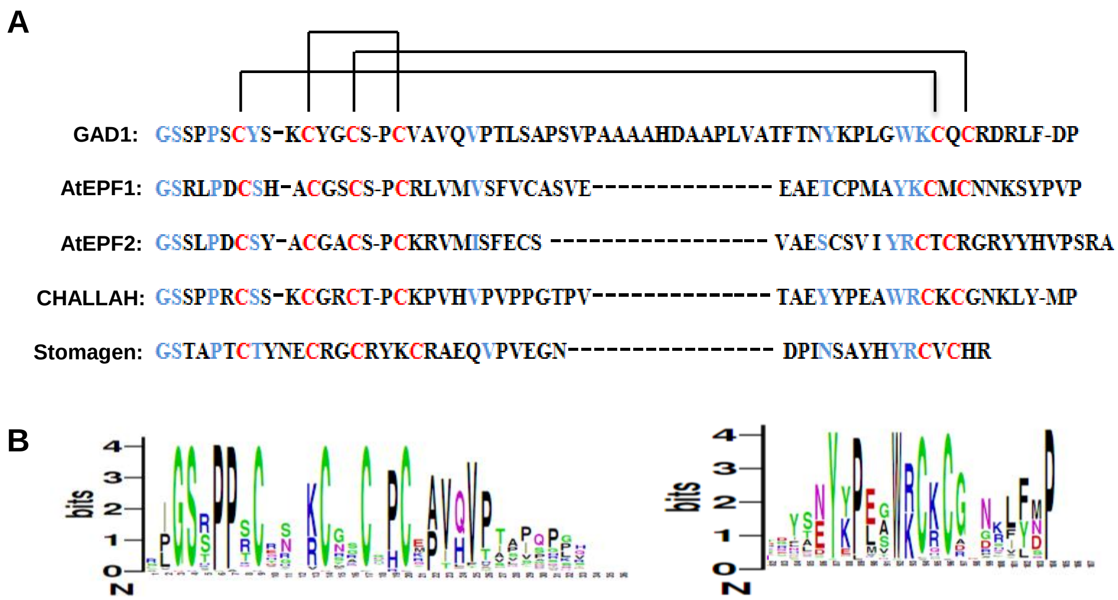
**Supplemental Figure 9**

Phylogenetic tree of *EPF/EPFL* family genes from rice and *Arabidopsis*. Amino acid sequences were aligned using the ClustalX program. The phylogenetic tree was constructed by the MEGA 5 program. The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (values 50% or greater are shown). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site.



## Supplemental Figure 10

Protein structure and maturation process predictions of GAD1. **(A)** Predicted protein structure of GAD1. The red rectangle represents the signal peptide; the green rectangle represents the mature peptide; red lines represent the consensus cysteine residues in GAD1. **(B)** Predicted GAD1 the maturation process. Red lines represent intramolecular disulfide bonds. **(C)** Analysis of the signal peptide in the GAD1 protein using SignalP4.1 (http://www.cbs.dtu.dk/services/SignalP).



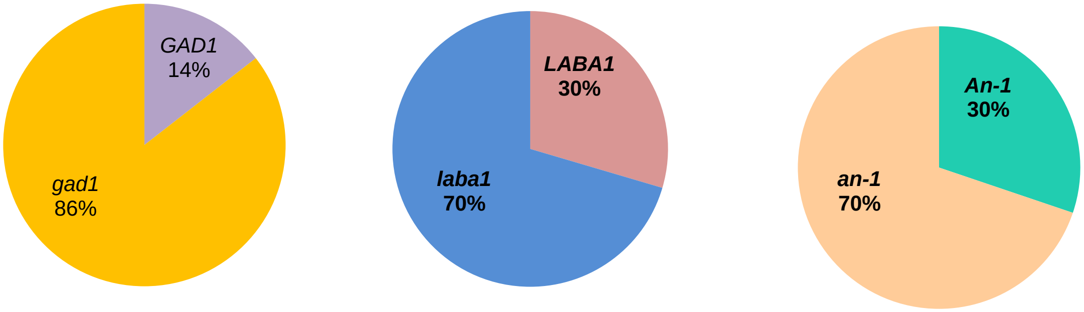
## Supplemental Figure 11

Analysis of the consensus sequence from the GAD1 protein. **(A)** GAD1 and other EPF/EPFL family member's mature protein sequences in *Arabidopsis* were aligned. The black lines represent the pairs of cysteine residues forming disulfide bonds predicted for the *EPF/EPFL* genes in *Arabidopsis.* The alignment uses putative active peptides based on the mature STOMAGEN sequence. AtEPF1: *At2g20875*; AtEPF2: *At1g34245*; CHALL: *At2g30370*; Stomagen*:At4g12970.* **(B)** Consensus sequences of the GAD1 protein. The consensus sequence graphically depicts conserved positions, and the height of the symbols indicates the relative frequency of each amino acid at that position.

| Haplotype | The variation sites of *GAD1* coding region | | | | | | | | | | | | | | | | | | | No. of cysteine in protein | Awn phenotype | *Oryza rufipogon* | *indica* | *japonica* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Indel 1 | | | Indel2 | | | | | | SNP 1 | Indel3 | | | | | | | | | | | | | |
| **H1** | G | C | C | - | - | - | - | - | - | T | - | - | - | - | - | - | - | - | - | 6C | No awn | 0 | 1 | 3 |
| | | | | | | | | | | | | | | | | | | | | | Awn | 5 | 1 | 4 |
| **H2** | - | - | - | - | - | - | - | - | - | G | - | - | - | - | - | - | - | - | - | | Awn | 8 | 0 | 0 |
| **H3** | - | - | - | G | C | C | T | C | C | G | - | - | - | - | - | - | - | - | - | | Awn | 1 | 0 | 0 |
| **H4** | - | - | - | G | C | C | T | C | C | T | - | - | - | - | - | - | - | - | - | | Awn | 1 | 0 | 0 |
| **H5** | - | - | - | G | C | C | T | C | C | G | C | G | C | C | G | C | C | G | C | 6C | No awn | 0 | 0 | 3 |
| | | | | | | | | | | | | | | | | | | | | | Awn | 20 | 3 | 8 |
| **H6** | - | - | - | - | - | - | - | - | - | G | C | G | C | C | G | C | C | G | C | | Awn | 4 | 0 | 0 |
| **H7** | G | C | C | - | - | - | - | - | - | G | C | G | C | C | G | C | C | G | C | | Awn | 4 | 0 | 0 |
| **H8** | (G | C | C)2 | G | C | C | T | C | C | G | C | G | C | C | G | C | C | G | C | | Awn | 3 | 0 | 0 |
| **H9** | - | - | - | G | C | C | T | C | C | G | C | G | C | C | G | C | C | - | - | 7C | No awn | 0 | 93 | 34 |
| | | | | | | | | | | | | | | | | | | | | | Awn | 6 | 2 | 8 |
| **H10** | - | - | - | G | C | C | T | C | C | G | - | G | C | C | G | C | - | - | - | 4C | No awn | 0 | 13 | 21 |
| | | | | | | | | | | | | | | | | | | | | | Awn | 1 | 5 | 4 |
| | | | | | | | | | | | | | | | | | | | | | Total | 53 | 118 | 85 |

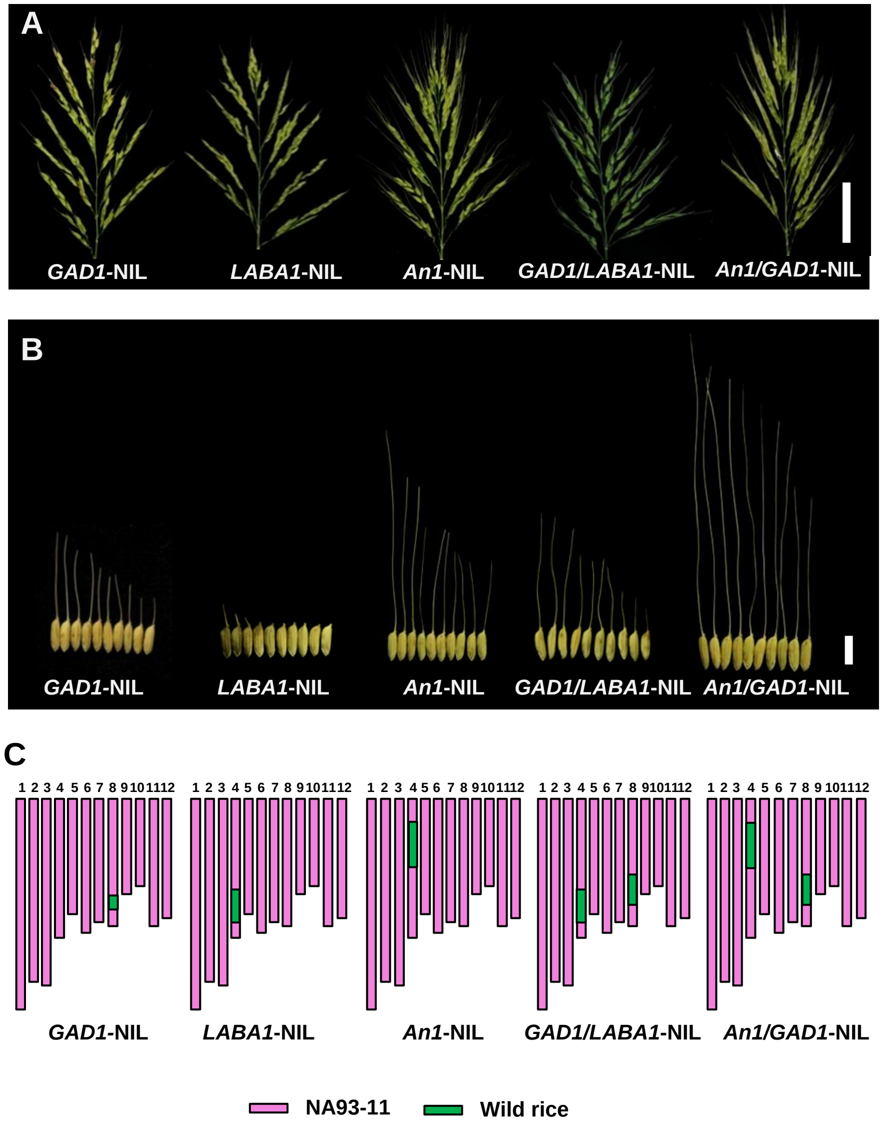
**Supplemental Figure 12**

Haplotypes of *GAD1* in diverse wild and cultivated rice. The numbers of accessions that harbored each haplotype are indicated in the right-hand table.



**Supplemental Figure 13**

Distribution of *GAD1/gad1, LABA1/laba1* and *An-1/an-1* in 159 varieties of cultivated rice.



**Supplemental Figure 14**

Phenotypes of near-isogenic and pyramiding lines containing *GAD1*, *An-1* and *LABA1*. **(A)** Panicle comparison among *GAD1*-NIL, *LABA1*-NIL, *An1*-NIL, *GAD1*/*LABA1*-NIL and *An1*/*GAD1*-NIL. Bar = 50 mm. **(B)** Awn length comparison among *GAD1*-NIL, *LABA1*-NIL, *An1*-NIL, *GAD1*/*LABA1*-NIL and *An1*/*GAD1*-NIL. Bar = 10 mm. **(C)** Genotypes of *GAD1*-NIL, *LABA1*-NIL, *An1*-NIL, *GAD1*/*LABA1*-NIL and *An1*/*GAD1*-NIL.

**Supplemental Table 1.**The signal peptide and cleavage site predictions in the GAD1 protein.

| Measure | Position | Value | Cutoff | signal peptide |
|---|---|---|---|---|
| max.C | 33 | 0.542 | | |
| max.Y | 33 | 0.544 | | |
| max. S | 30 | 0.812 | | |
| mean S | 1-32 | 0.517 | | |
| mean D | 1-32 | 0.533 | 0.5 | Yes |

The prediction was performed by SignalP4.1 (http://www.cbs.dtu.dk/services/SignalP). The GAD1 protein cleavage site was predicted to occur between positions of 32 and 33; D = 0.533; D-cutoff = 0.500.

**Supplemental Table 2.** *GAD1* nucleotide diversity and Tajima's *D* test.

| Taxon | N | L | S | H | $\pi$ | $\theta w$ | Tajima's *D* |
|---|---|---|---|---|---|---|---|
| *O. sativa* | 203 | 579 | 20 | 23 | 0.00184 | 0.00663 | -2.14422 ($P<0.01$) |
| *O. rufipogon* | 53 | 540 | 25 | 30 | 0.00664 | 0.01289 | -1.64418 ($P>0.1$) |

N, total number of sequences; L, average length (bp) of the sequences per taxon; S, number of polymorphic (segregating) sites; H, number of haplotypes; $\pi$, average number of pairwise nucleotide differences per site calculated based on the total number of polymorphic sites; $\theta w$, Watterson's estimator of per base pair calculated based on the total number of polymorphic sites.

**Supplemental Table 3.** Primers used in this study.

| Purpose | Name | Forward primer (5'-3') | Reverse primer (5'-3') |
|---|---|---|---|
| Mapping of *GAD1* | M12 | AAGTTCTGAGCCCGTGAATG | CATCCGATCAACATCCATTG |
| | M10 | GAATACTGCAGCTGGGGTTG | TGACTTGATTCATCATCAAATGTTC |
| | M3 | AAAAATGTGCACCAACTTCT | TAGAGGGCTCAAATTGTGTT |
| | M62 | AATGTGTATTTAGATGCGATAGTTGTC | TTCGATATCCGAAGCAATCC |
| | M67 | TCCAATCAAGAAACCAAACCA | CCCTTCTTCACGTGTAGTCTCA |
| | MX15 | CTCTTGCCTCTTGGAAGTTG | GCAAATTGAGTTATGGCAATG |
| | MX14 | CTTCAAATTTATCTTTTTATTTCTACG | TCTGTTTTACCCTTCGATTCC |
| | MX16 | ATCTTCCACGCACGTTACAT | GCTACAAGCCCAAATGCTAC |
| | MX12 | GCCCATCTGCTACAGTTTGA | CCCACTCTCACTGACAGCTT |
| | MX10 | CAGCAAACTGCAAAGTGTAGC | GCAATTATTTATAGCTAGGAATTTAGA |
| | MX9 | TTCAGGAGTAACCCAATGTC | AACACCTCTGGTTTCTTGC |
| | RM5485 | ATGATTGCATCTGCATCACTGC | ATACCTGTTTCCAATGCGTAGC |
| | RM256 | GACAGGGAGTGATTGAAGGC | GTTGATTTCGCCAAGGGC |
| 5'RACE | 5'RACE inner | CAACGCAGAGAGGAACACTG | |
| | 5'RACE outer | CACTTGCTGTAGCAGCTCG | |
| 3'RACE | 3'RACE inner | CGAGCTGCTACAGCAAGTG | |
| | 3'RACE outer | CAGTGTTCCTCTCTGCGTTG | |
| pCPL construct | 1300-3 | GGGGTACCTAGGCCCATGAAACAAGAAA | CGCGTCGACTTGCAGCGATCTGAATAACA |
| pRNAi construct | RNAi3 | ATCACTAGTATGAGGACGGCGGCCACGCC | ATCGAGCTCCTAGCGGCTTGTAGTTGGTGA |
| qRT-PCR | Actin | TGCTATGTACGTCGCCATCCAG | AATGAGTAACCACGCTCCGTCA |
| | *GAD1*-QRT | ATGCTGGTGGTGGTGGAATC | TACATGCAAGAGGCAACCCC |
| | *OsCKX2*-QRT | CGGCAACAAGTGGGACAGTAA | CAGGGCGATGTAGGAAAGC |
| | *DST*-QRT | ATCCAAGAAGGCAAGGTCAATC | ACACACGAGGAGGAATTGGAA |
| | *OsRR1*-QRT | AGGATCAGCAGATGCATGAATG | GAGACGCTGTACGTCCTTGCTT |
| | *OsRR2*-QRT | ACGATCTTCTCAAAGCCATCAAG | TGAGAGGCTTAAGGATGAAATCCT |
| | *OsRR3*-QRT | GTTACGACGGTGGATAGTGG | ATTCTCCGACGACATTAGCA |
| | *OsRR4*-QRT | CTGATCGACAGGAAGCTCAT | CAGGCATGCAGTAGTCAGTG |
| | *MCM2* | AAGTTGGCAAAAGATCCACGG | CCCCCAAACATAGCTAGTGCAA |
| | *MCM3* | TTCATGCGTCACTAAATGCGAG | TGAATCTGGAAGCCCAATGTTC |
| | *MCM4* | CCCGAATGCGATTCTCTGAA | ACCAGTGGCATGATCAGTTGC |
| | *CYCB2.2* | CTCAAGGCTGCACAATCTGACA | GCATTGACGGCTGGAATTTG |
| | *CYCA2.2* | AGGTTGTCAAGATGGAGAGCGA | CGCTTTTTGTCTTCCTGGCA |
| | *MAPK* | ACAGAGCAGCCGAATTTTGAGA | TTCAGCGAAGCTCACACTTGG |
| | *KN* | CACCAGCTTCAAGAGATCGTGA | CCGGAATTGAGACACAACTGC |
| | *CDC20* | TCGAATCACCTGTTTGTTGGC | TGGAGACAATCCAACGCAAAG |
| | *CDKB* | AAGTTTGGCCAGGAGTGAGCA | TCAAGAGCATCAGCGTCGAGA |
| | *CDKA2* | CGAGATTTGAAGCCCCAGAA | TCCGCGAGCTTCAATGAGTT |
| | *CDT2* | AACCGCACCAAACACTGGAA | GCAATTCACCATCTGCACTGG |
| | *CDKA1* | GGTTTGGACCTTCTCTCTAAAATGC | AGAGCCTGTCTAGCTGTGATCCTT |
| | *CAK1A* | GACCGACAAGGGTTTCAGCAT | CCAGCATGTTCAGGAAGATACAAT |
| | *CYCA2.3* | GTTTCGGTTGACGAGACGATGT | CGCTGCAAGGAACCTAGAACTG |
| | *CYCT1* | GCATTTGTTGCAGCTCAAG | TCACCACTTCGCTGACTTATTG |
| | *CYCB2.1* | AAGTTTGGCCAGGAGTGAGCA | TCAAGAGCATCAGCGTCGAGA |
| | *CYCD4* | GCCATGGAGTTGATACATCCAA | CCAGTAGGGCTCCGTGGAAT |
| | *H1* | GCAAGGCACCTGCAGCTT | AGGCAGCCTTTGTACAGATCCT |
| | *CYCIaZm* | CACTCTCAAGCACCACACTGGA | ACAACCCTCAGCTTGCTCTCAG |
| | *CYCA2.1* | AGGTTGTCAAGATGGAGAGCGA | CGCTTTTTGTCTTCCTGGCA |
| | *E2F2* | TGTTGGTGGCTGCCGATAT | CGCCAGGTGCACCCTTT |
| | *MAD2* | GAGCCATGCATATTCGACGTG | GGTGTCGAAGGAATGCAGCTT |
| | *GAD1*-CX | TTCTTGGTAGGCGAGGTGTC | CCTGGCATATTACACGGAATC |
| RNA *in situ* probe | *GAD1-in situ* | TTCTTGGTAGGCGAGGTGTC | CCTGGCATATTACACGGAATC |
| | *OsCKX2-in situ* | GTCCACGACGGCGAGCTCAA | CACCTCGCAGAAACCCAAGAT |
| | *OsRR1-in situ* | TGCTTGACGAATGAAGCAAC | AGGGCAAGGAAAAAGCTCTC |
| | *OsHistoneH4-in situ* | CTCCGCGACAACATCCAG | TCCAAACCTCCATATTTTCGGT |
| Selective sweep analysis | sweep1 | TCACCTTTCTCCCCAATTGC | GAGTGTTCAACATGTGCAGC |
| | sweep2 | TGGGCTTACGTTTTGGTTCA | GCCAAGCAATTAGGGCCAAT |
| | sweep3 | AATCTGGTTGGTCGTTGGTT | GGTATTTGGGTTAGGGCTGG |
| | sweep4 | CCTGGTGAATACTTGTCGGG | AATTTACTCCTCCCGCGTTC |
| | sweep5 | ACAAGCGGCTAATCGAATGT | GTGGCTGGAAAATCTCGTCT |
| | sweep6 | ACCCTTCATGATGTTCTTGTGA | CAGAGCCTCAGAATTGCAGT |
| | sweep7 | CTGCAAAAGATGGACATGC | GACTGCTGCCACATCTAACA |
| | sweep8 | GCTAGCACAATCGATCACCT | AGAAAGAGATGGATGCAGCG |
| | sweep9 | GTTTACGGGCGGATTCTGTA | CCGTACGTACCTGTCCATTG |
| | sweep10 | TGGGGTACTACTTAGGTCACT | ATTGTACTGGGCGTGTAACC |
| | sweep11 | GGCGGTTTGAATGTTGTGAG | ACCTCAAGCAAGCAAACAGA |
| | sweep12 | CATTGATGTGGGGTCTCTGG | GGAGAACACCACCAAACGTA |
| | sweep13 | TGCTCGCACAAGTCAATTTG | GGAATGAGCATCCCTACTGC |
| | sweep14 | AAATCAACCGGGTCAGACAC | GACAACGGCAACACTGGATA |
| | sweep15 | GGCTACATGGGCATTGAGAA | AGTCTGTTCTGAGGTGAGCT |