# **Supporting Information**

### Sugimoto et al. 10.1073/pnas.0911965107



**Fig. S1.** Coarse- and high-resolution mapping of *Sdr4*. (*A*) Graphical genotype of chromosome 7 in NIL[*Sdr4*]. The introgressed segment from Kasalath is indicated by a black bar. The positions of RFLP markers and SSR markers are based on physical distance. (*B*) Chromosomal locations of *Sdr4*, based on genetic distance. (*C*) Coarse mapping of *Sdr4* using a population of n = 100. Numbers of recombinants between each set of markers are indicated. Phenotype "M" indicates mildly resistant to preharvest sprouting. By using this population, the *Sdr4* candidate region was delimited to 108 kbp. (*D*) High-resolution mapping using a population of n = 2,515. Numbers of recombinations between pairs of markers are shown. Black regions are chromosomal segments homozygous for the Kasalath allele; white regions are homozygous for the Nipponbare allele; hatched regions are heterozygous. Primer sequences of newly designed markers are listed in Table S2A.



**Fig. 52.** Genetic complementation of *Sdr4* using the genomic fragment encompassing the entire candidate genomic region and its derivative. (*A*) Candidate genomic region of *Sdr4* and predicted genes. Genes to the right of the black arrow are those predicted by RAP-DB in the Nipponbare genomic sequence around the *Sdr4* candidate region. The 8.7-kb *Sdr4* candidate region was defined on the basis of high-resolution mapping. Vertical bars indicate InDels and SNPs located in the 11.6-kb region from *Nrul*. The 11.6-kb fragment (KN) and 11.2-kb fragment with a 0.4-kb deletion around the first exon of the *Os07g0585800* (KNd) were used for complementation analysis. (*B*) Germination rates of T<sub>3</sub> seeds harboring the KN fragment, the KNd fragment, and empty vector (Vc). The presence of the KN fragment reduced germination rates, indicating that *Sdr4* was present in this fragment. The KNd fragment complemented the function of *Sdr4*, demonstrating that *Os07g0585800* was not the *Sdr4*.

PNAS PNAS

+72 GCCATGGTGCAGCCGGTGGACATGGCCGTGAAGGCCAACGAGATCATGGCGAGGTTCAGG A M V Q P V D M A V K A N E I M A R F R

Sdr4-k	1	MAMVQPVDMAVKANE I MARFRP I APKPVLPAAAAGVTGGG	40	-1609	${\tt cgacgtacgcgcttgcgaattgaaagccttccttttcgtccgttccctgagaccctgacc}$
Sdr4-n	1	MAMVQPVDMAVKANE IMARFRP I APKPVL PAAAAGVTGGG	40	-1549	cgcgcccacccactcgccaagcgggagcgagacggagcggcgcctgcacccgcgcgcccg
Sdr4-k	41	DGAAAVAATNRVLCQLQSRPCRARKRGRPSVVPPVSPPAG	80	-1489	ccaatggcagggcggccactcgtcgcggcggtgctacccgctcgccacggcaggag
Sar4-n	41	KRKR	80	-1429	acatgccgtg <u>CATGCA</u> aaaagggtgggttcttataaaaaaaaa <mark>a</mark> agagagtaagaaa <u>c</u> gga
Sdr4-k Sdr4-n	81	AKRKRAPAYPEPVAPLRCAAVATATRARVSVVVVPAPGSA	120	-1369	gctgtagcgtcggtccaaaagtttgtttctcgagcagggagatcgaagagagag
our n	01		120	-1309	eccctectectectcceectgecceetttaaattcetectectecattceeccgeetagc
Sdr4-k Sdr4-n	121 121	GGVSALAPVSPSAGDSTRLSPTVVEVEDEDEERGVVLVER GGVSALAPVSPSAGDSTRLSPTVVEVEDEDEERGVVLVER	160 160	-1249	CE RY-repeat tagccggcccgtggttgggctccagtaggacacgdgaCATGCATGctccACGTGTgacgc
Cded Is	161		100		RY-repeat
Sdr4-n	161	DLLRKLLEPRKLLEPRAVRPVGSTIHVESVHIDVGRTTAA	200	-1189	tttgcctgctgtttcgtgtaccgcagdTGCATGtatgcgtaaagtttttctactactagt
				-1129	${\tt ttatttttattgggcggacaggtcgtcacagtagagcatttaccgtcgaggccgcgtac}$
Sdr4-k Sdr4-n	200	AAAAAPKTAEEVEAELESDSLPAVVSDSSNRVRLVNDAYK AAAAAPKTAEEVEAELESDSLPAVVSDSSNRVRLVNDAYK	239 240	-1069	$\tt tttttggaacagaagagttagatccacagtgttttacatatccagtcgatttatgtttgg$
				-1009	${\tt gtttgggaaatttcaccggtgtaaaaagcttctcacttcatcgggcaccgtcttctgtaa}$
Sdr4-k	240	RMVGQPECPWLDAVATAASRRISGEVALVVSEP	279	-949	atttttttcgatatgaactttgataaattttgctcctcgtccaggcgaagaaactccggc
3014-11	241	NINGGE COMEDAVA TAASAA TSGEVAEVVSEF	200	000	RY-repeat
Sdr4-k	280	PETCKGFSCSAKIAWERDGKWSSVHAPCDVTRLQCESRDY	319	-889	LILEge La Lagga a a cga Lge A loo A
Sdr4-n	281	PETCKGFSCSAKTAWERDGKWSSVHAPCDVTRLQCESRDY	320	-829	TICGCCCggICIIgaIgIgIgUGAUUIgCIggagaCcgICgICaccagCIIIIaaIICai
Sdr4-k Sdr4-n	320 321	VFAWRFRAAGDECNTHRRAAGDA	342 343	-769	cggaaagatagcaaagcgtttgtcgtcgtggtttacattgcatcgtctggtttccctccg <u>RY-repeat</u>
	0		0.0	-709	agattttgtatgtdCATGCATGtatagtctcgtagaccatacatgaccagcaaaaCCACG
				-649	Tgcggatatatcaacttttaggcggttttgtgctggctgaaacaatatgtacctcagctg RY-repeat
				-589	$a gaag taat {\tt ctcgaa} a g {\tt catgag ta} cacca {\tt ccggag tcggag tg} ta {\tt CATGCATGCATG} ca$
				-529	cggcaaggtatgaccaaaaagcagttttgacgcagattgatcattcgatctggtagtgaa
				-469	aaacagaaaaaaaaagtgatgtgtgtgtcccggtgtccgcctatacgctaaacttacc
				-409	gtggtgctagagtacaagcgagccggttggtggaggcgagcgcaaaagcgccacacggcc
				-349	${\tt ttctgccaatgaggcctcacgccgtcacgcggggcacgccagcca$
					RY-repeat
				-289	ggaacgcgccacgggggacgatgcgcccgagggggggggg
				-229	gagtacgcgacctcgcctcacctcacctc
				-169	$a \texttt{atccaccggagcgcgcgcgcgcgcgcgcgccgcgtcccgct} \\ \underline{\texttt{ACGTGT}} \texttt{cggccgcgtccccgc} \\ \\ \underline{\texttt{atccaccggagcgcgcgcgcgcgccgcgtccccgc}} \\ \\ \texttt{atccaccggagcgcgcgcgcgcgcgcgcgccgcgcgccgcgcgccgc$
				-109	gcggtcgc <u>CCACGT</u> accccgcccccgccctc <u>CCACGT</u> gcccctccccctgcgcgcatccg +1
				-49	attggccgcccacgccttcttaaccccaccaccaccactcgggccccAACGCATCCGT
				+11	TCACCACCTGCGAGAGCGAGCGTACCGTTCGTTCGTTCGT

Fig. S3. Deduced amino acid and promoter sequence comparison between Sdr4-n and Sdr4-k. (A) Amino acid sequences of predicted ORFs in Sdr4-k and Sdr4-n were compared. Putative NLS and deletion in sdr4 mutant (gray line) are shown. (B) Promoter sequences of Sdr4-k from -1,609 to +132. SNPs and InDels are indicated by black boxes, and RY repeats and ABRE-CEs are indicated by clear boxes. ABREs are indicated by double underlines.

Pt551773 Pt572029 RcEEF52271 VtJOC100262679 AtSFL1 Sb02g037770 ZmLOC100279098 ZmLOC100278263 Sdr4-n Sdr4-k	1 1 1 1 1 1 1 1 1 1 1	MIKTLSPCSNTARTAE INSPYPPIAAPR EGS-ASSMDESSSMOK I RESPYLRTLINPOMDARTIE I RKRBRAVSPPNIKAPR MIKTLSPCSNTARTAE INSPYPPIAAPR EGS-ASSMDESSSMOK I RESPYLRTLINPOMDARTIE I RKRBRAVSPPNIKAPR MIKTLSPCSTTARTAE INSPYPPIAPR EGSSTDESSMDF	82 73 82 90 81 79 67 78 78
Pt551773 Pt572029 RcEEF52271 VtLOC100262679 AtSFL1 Sb02g037770 ZmLOC100279098 ZmLOC100279098 Sdr4-n Sdr4-k	83 74 83 83 91 82 80 68 79 79	THLLGLSSPSHVTSPAKHLSLOGFVHGIPOLPVPNLVGVNSGLENSVTMSSN-LVTUPLLOSPT-TVVVMANAAVPELSOME   THLLGLSSPSHATVSAKHLSLOGFAHGITOLPVPNLVGVNSGLENSVTMSSN-LVTUPLLOSPT-TVVVENAAAVPELSOME   THUGISSPSHATVSAKHLSLOGFAHGITOLPVPNLVGVNSGLENSVTMSSN-LVTUPLLOSPT-TVVVENAAAVPELSOME   THUGISSPSHATVSAKHLSLOGFAHGITOLPVPNLVGVNSGLENSVTMSSN-LVTTELLOSPT-TVVVENAAAVPELSOME   THUGISSPSHATVSAKHLSLOGFAHGITOLPVPAPSFAPLNOGFERAISTAGLAMTUPLLPOPPPOLPVVENAAAVPELSOME   SGASPAKRPSHATVSAKNLSMHGTHSPLOLPVPAPSFAPLNOGFERAISTAGLAMTUPLLPOPPPOLPVVNDENCLIP   SGASPAKR	164 154 163 159 140 134 126 141 141
Pt551773 Pt572029 RcEEF52271 Vt.OC100262679 AtSFL1 Sb02g037770 ZmLOC100279098 ZmLOC100278263 Sdr4-n Sdr4-k	165 155 165 164 160 141 135 127 142 142	NRDKV1 D.NTVAEFPE 0.01.000 0VPPTNNV1APOPLR. VGSSTS1AC ISEDPSF1PLVRVPK4-PEEVELS   NRDKV1 D.NTVAEFPE 0.01.000 0VPTNNV1APOPLR. VGSSTS1AC ISEDPSF1PLVRVPK4-PEEVELS   NRDKV1 D.NTVAEFPE 0.000 0VPTNNV1APOPLPVGSSTSVSSTS1AS (SST-SSTP-VKVPK4-PEEVELS   GGNV1 D.NTVAEFPE 0.000 0VPTNNV1APOPLPVGSSTSVSSTSSTPVVVPK4-PEEVELS   CGGNV1 FD.NTSVVAAIPE SKUL000 0APAMIT	239 229 239 244 203 203 188 218 217
Pt551773 Pt572029 RcEEF52271 VvLOC100262679 AtSFL1 Sb02g037770 ZmLOC100279088 ZmLOC100278263 Sdr4-n Sdr4-k	240 230 245 245 204 204 189 219 218	EVETU ISSNINKURLANS AVKEM OPECSIN DISMITIGDOSFAGRSCKI I GEVELIHLSDLRVPASSNGFS AVKTEMONIKUS EVETU ISSNINKURLANS AVKEM OPECSIN DISMITISDORFAGSSCKI I GEVELIHLSDRVPASSNGFS AVKTEMONIKUS ELITATU ISSNIKURLANS AVKEM OPECSIN DISMITISDORFAGSSCKI I GEVELIHLSDRVPASS	324 314 325 325 322 281 281 266 301 300
Pt551773 Pt572029 RcEEF52271 Vt.OC100262679 AtSFL1 Sb02g037770 ZmLOC100279098 ZmLOC100279098 ZmLOC100278263 Sdr4-n Sdr4-k	325 315 325 326 323 282 282 267 302 301	NV I I TE OV UR S & SRDV E SINE HTRGRKDEQSKTNA NV I I TE OV UR S & SKDV E PARE HT REGRKDEQSKTNA STATATER VI I NE SK SKDV E ANNE HTI SREGSGESKTKA	363 353 361 354 325 327 312 343 342
		deletion in sdr4	

Fig. S4. Deduced amino acid sequence comparison between Sdr4 and homologs. Amino acid sequences of predicted ORFs were compared between Sdr4 and homologs. Amino acid substitutions between Sdr4-n and Sdr4-k are marked with as asterisk followed by a number. Putative NLS and deletion in sdr4 mutants are indicated by clear and hatched boxes, respectively.



Fig. S5. Tissue specificity of Sdr4 expression. Sdr4 mRNA levels in various tissues of plants or seeds 14 days after heading of Nipponbare (N) or NIL[Sdr4] (S) were detected by RNA gel blot analysis. Ca, callus; L, leaf; Rt, root; Fl, flower; Sd, whole seed; Emb, embryo; Cot, seed coat; End, endosperm; Glu, outer glume.

SANG SANG



**Fig. 56.** Complementation of *sdr4* mutation, mutations in *Osvp1* mutants, and linkage analysis of M101 viviparous phenotype and *Osvp1-2* mutation. (*A*) Complementation of *sdr4* mutant (M100) was performed with a 3.3-kb *Sdr4-k* genomic fragment (Fig. 1 *C* and *D*). The germination rates of transgenic lines with *Sdr4-k* (KNa) were lower than those of the vector control lines (Vc), indicating that the viviparous phenotype of the M100 line was due to the mutation in *Sdr4*. Germination rates of NIL[*Sdr4*] (Sdr4) and the *sdr4* mutant (M100) are shown together. (*B*) Mutations in the *OsVP1* gene are summarized. (*C*) Tight linkage of *Osvp1-2* mutation and viviparous phenotype shown by genetic mapping using a mapping population derived from F<sub>2</sub> seeds selected by preharvest sprouting at 4 weeks after heading from M101 × Kasalath cross F<sub>1</sub> plants. This result shows that the *Osvp1-2* mutation and the causal mutation of the viparous phenotype lay within the indicated 131-kb region. The sequences of the PCR primers for markers were 5'-CGT CAG TGC ACC GAA GTT T-3' and 5'-CTG GCT TTT GGA GAA GAT CG-3' for IND11, and 5'-GAT TAA GCA CGG CTC ACC AC-3' and 5'-GGT TAA CGG GTC TCC ACA CA-3' for IND7.



**Fig. 57.** Phylogenetic analysis of *DOG1 homologs* and expression analysis of *DOG1 like-3* and *PIP2;2*. (A) DOG1 homologs were searched using the BLASTP program of NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against the reference protein sequence database. Those with *E* values  $<e^{-10}$  were selected for phylogenetic analysis. Sequences were aligned using Genetyx version 9 software (Genetyx). Evolutionary history was inferred using the neighbor-joining method (1). The optimum tree with a sum of branch lengths of 6.23632813 is shown. The percentages of replicate trees in which the 27 associated amino acid sequences clustered together in the bootstrap test (1,000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset. The final dataset contained a total of 88 positions. Phylogenetic analyses were conducted using MEGA4 software (2). (*B*) Temporal changes (DAF) in mRNA levels in embryos of the *OsDOG1 like-3* gene was monitored by real-time PCR in Nipponbare (Npb), the *sdr4* mutant (M100), and NIL[*Sdr4*] (NIL), with three biological repeats. Expression levels are normalized to *Actin-1* expression. (*C*) Temporal changes (DAF) in mRNA levels in embryos of the aquaporin gene (*PIP2;2*) monitored by real-time PCR.

1. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425.

2. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599.

	SNP name	SNP1	SNP2	Sdr4	SNP3	SNP4	SNP5	
	Distance (Mb)	-1.6	-0.1	0	+0.1	+0.3	+1.2	
Code	Chr. position (Mb)	22.9	24.4	24.5	24.6	24.8	25.7	Sub Group
	Nipponbare	C	Т	n	Т	A	Т	
	Kasalath	Т	Ċ	k	G	Т	Ċ	
WRC01	Ninnonhare	Ċ	т	n	т	Δ	т	In
WRC43	Dianyu 1	C C	T	n	+ ÷		+ +	Jp
WRC47	Jaquary	<u> </u>	T	n	+ +		+ ÷	Jp
WRC47	Tima	<u> </u>	- <u>-</u>	n	+		+	Jp
WRC51	Urasan 1	C	т	n	ι -		+ +	Jp
WRCJI	Khao Nok	т	T		<u>+</u>		- T	Jp
WPC48	Khau Mac Kho	T	T	n	+ +		+	Jp
WRC40	Padi Perak	T	т Т		++		T	Jp
WRC49	Khao Nam Jen	T	T	n	+		<u>+</u>	Jp
WRC00	Khau Tan Chiem	T	т Т		<u>+</u>		<u>+</u>	Jp
WRC52	Ma cho				+ +	A	++	Jp
WRC45	Dhulho			n	++	A	+	Jp
WRC67	Priuba			n	++	A	+	Jp
WRC50	Acu		 	n		A		Jp
WRC13	Asu			n	<u> </u>	A	0	In-2
WRC42	Local Basmati	-	-	n	<u> </u>	A	C	In-1
WRC29	Kalo Dhan	T	T	n	T	A	C	In-1
WRC28	Jarjan	Т	T	n	Т	A	C	In-1
WRC21	Shwe Nang Gyi	Т	Т	n	Т	A	C	In-2
WRC66	Bingala	Т	Т	n	Т	A	С	In-2
WRC27	Nepal 8	Т	Т	n	Т	A	С	In-1
WRC04	Jena 035	Т	Т	n	Т	A	С	In-1
WRC26	Jhona 2	Т	Т	n	Т	A	С	In-1
WRC34	ARC 7291	Т	Т	n	Т	A	С	In-1
WRC25	Muha	Т	Т	n	Т	A	С	In-1
WRC44	Basilanon	Т	Т	n	Т	Т	С	In-1
WRC09	Ryousuisannkoumai	Т	Т	n	Т	Т	С	In-2
WRC55	Tupa729	Т	Т	n	G	A	С	Jp
WRC03	Bei Khe	Т	T	n	G	A	С	In-2
WRC07	Davao 1	Т	С	n	Т	A	С	In-2
WRC38	ARC 11094	Т	С	n	Т	A	С	In-1
WRC62	Kemasin	С	С	n	G	A	С	In-2
WRC64	Padi Kuning	С	С	n	G	Т	С	In-2
WRC06	Puluik Arang	С	С	n	G	Т	С	In-2
WRC39	Badari Dhan	Т	Т	n	G	Т	С	In-1
WRC36	Ratul	Т	Т	n	G	Т	С	In-1
WRC11	Kinkagin	Т	С	n	G	Т	С	In-2
WRC18	Seiyu	Т	С	n	G	Т	С	In-2
WRC19	Touchusai	Т	С	n	G	Т	С	In-2
WRC20	Tadukan	Т	С	n	G	Т	С	In-2
WRC57	Mitusyou 23	Т	С	k	G	Т	С	In-2
WRC05	Naba	Т	С	k	G	Т	С	In-2
WRC10	QuiZhaoZong	Т	С	k	G	Т	С	In-2
WRC37	ARC 7047	Т	С	k	Т	Т	С	In-1
WRC40	Nepal 555	Т	С	k	G	Т	С	In-1
WRC33	Surjamukhi	Т	Т	k	G	Т	С	In-1
WRC35	ARC 5955	Т	С	k	G	Т	C	In-1
WRC30	Anjana Dhan	Т	C	k	G	Т	C	In-1
WRC31	Shoni	Т	C	k	G	Т	C	In-1
WRC02	Kasalath	Т	C	k	G	T	C	In-1
WRC32	Tupa 121-3	Т	C	k	G	T	C	In-1
WRC58	Neang Menh	т	C	k	G	T	C	In-2
WRC59	Neang Phtong	Т	C	k	G	Т	C	In-2
WRC60	Hakphavnhav	Ť	C	k -	G	T	C	In-2
WRC61	Radin Goi Sesat	T	<u> </u>	k	G	T	C	In-2
WRC62	Bleivo	T	C	k	G	T	C	In-2
WPC65	Rambhog	T	0	ĸ	G		0	III-2
VVRC05	ramonog		U U	ĸ	6	A		10-2

**Fig. S8.** SNPs in the peripheral region of *Sdr4*. Typing of world rice core collection around the *Sdr4* locus was performed as follows. Chromosomal regions located –1.6, –0.3, +0.4, +1.0, and +1.2 Mb from *Sdr4* in the world rice core collection (positions given in Table S2C) were amplified by PCR. They included introns and intergenic regions. PCR fragments were treated with ExoSap nuclease (GE Healthcare UK) and used as templates for sequencing reactions. Primers for this experiment are listed in Table S2C. SNPs were then summarized. Chromosomal positions of SNPs are based on the RAP-DB Build4 sequence. The *Sdr4* alleles *Sdr4-k* and *Sdr4-n* are represented by k and n, respectively. The *Indica*(In-1 or In-2)/*japonica* subgroups of each cultivar are shown as In-1, In-2, or Jp (*japonica*).

## **Other Supporting Information Files**

Table S1. SNPs and InDels in the Sdr4 gene in O. sativa cultivars and 46 accessions of O. rufipogon

#### Table S1 (DOC)

PNAS

DNAS DNAS Amino acid substitutions relative to Sdr4-n are shown in the first line. Positions from the first ATG of each substitution are shown in the next line. SNPs and InDels of Sdr4-n, Sdr4-k, and Sdr4-k' are been added as Nipponbare, Kasalath, and Co13, respectively. Sequences of large insertions and deletions are given in the bottom line.

Table S2. Primer sequences of newly designed DNA markers used in the linkage mapping (A), primer sequences used to monitor mRNA levels by semiquantitative RT-PCR or real-time PCR (B), and primer sequences used for SNP typing by sequencing (C)

#### Table S2 (DOCX)

In A, for InDel markers, amplified fragment sizes of Nipponbare and Kasalath are shown. For CAPS markers, target SNPs, restriction enzymes used to detect polymorphisms, and amplified sizes of larger fragments are shown. For SNP markers, primers for preamplification, acylCo primers, target SNPs, and amplified fragment sizes are shown. RAP-ID, RAP-DB accession number.