Supporting Online Material for

Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield and plant biomass in rice

Xian-Jun Song[1], Takeshi Kuroha, Madoka Ayano, Tomoyuki Furuta, Keisuke Nagai, Norio Komeda, Shuhei Segami, Kotaro Miura, Daisuke Ogawa, Takamasa Suzuki, Tetsuya Higashiyama, Takumi Kamura, Masanori Yamasaki, Hitosh Mori, Yoshiaki Inukai, Jianzhong Wu, Hidemi Kitano, Hitoshi Sakakibara, Steven E. Jacobsen[1],

and Motoyuki Ashikari[1]

[1]To whom correspondence should be addressed. Email: songxj@ibcas.ac.cn (X.J.S.) jacobsen@ucla.edu (S.E.J.), ashi@agr.nagoya-u.ac.jp (M.A.)

This PDF file includes

SUPPORTING ONLINE MATERIAL

Materials and Methods

Plant populations and quantitative trait locus (QTL) analysis. A set of backcrossed inbred lines (BILs) containing 98 individual lines was grown in the paddy field of Nagoya University, Aichi Prefecture, Japan, in 2007, under standard cultivation conditions. An average of 1,000-grain weight of five individual plants from each of the BILs was obtained after harvesting and air-drying for around one month. The grain weight of the BILs was used as the phenotype for QTL detection.

A chromosomal segment substituted line, CSSL29, was chosen and crossed with the Nipponbare (Nipp) line to produce an F_2 population for QTL mapping. Markers xj112 and xj113 on the long arm of rice chromosome 6 were chosen as a result of the segregation of the desired genotype and grain weight phenotype on the F_2 population. The F_2 and correspondingly derived F_3 populations were used for marker-assisted QTL mapping, and *GW6* was mapped to the candidate region spanned by xj112 and xj113. To further map *GW6a* locus, progeny testing of homozygous recombinant plants was performed with the aid of newly developed molecular markers; and we selected the NIL(*OsglHAT1*) that has a fixed Nipp genotype at *GW6b* locus from a F5 generation by DNA marker assistance. Relevant marker sequences can be found in **Table S1.**

Transgenic assays in rice plants. We screened a Kasalath (Kasa) genomic DNA library using markers that define the *GW6a* locus (xj-6 and xj-11), and identified a positive BAC clone, BAC_K0242A07. Partially digested fragments of BAC_K0242A07 by the endogenous restriction enzyme *Hind*III were segregated, recovered and inserted into vector pYLTAC7 (1). We verified the vectors by sequence analysis and used them for transgenic assays in rice as described previously (2). The full-length *OsglHAT1* cDNA ORF was amplified from the CS tissue (see Text) of both Nipp and CSSL29 plants and cloned into the plant binary vector pHB (3) for over-expression of *OsglHAT1*, whereas down-regulation of the gene was obtained via the insertion of $OsglHAT1^N$ cDNA ORF in the

antisense orientation. Furthermore, we generated a series of amino acid swaps in *OsglHAT1* alleles (**Figure S4A**) by PCR amplification of mixed allele templates derived from restriction enzyme digestions, and then cloned them into the binary vector described above. We have a total of 16 OsglHAT1N-OE (7 of these showed significantly enlarged grains in T0 generation) and 11 OsglHAT1K-OE (4 of these produced enlarged grains in T0 generation) independent transgenic lines in rice plants, and we used typical transgenes (that were confirmed by RT-PCT experiments) in Figure C in the text.

Generation of transgenic *Arabidopsis* expressing *OsglHAT1*. The *OsglHAT1* coding region from Nipp and Kasa were amplified by RT-PCR using the primers 5'-caccatggtggagacgacgacg-3' and 5'-ttagaactcgcgggggtcgacg-3', ligated into the pENTR/D-TOPO vector (Invitrogen), and then integrated into the Gateway binary vector pBA002Gw-HA (a derivative of pBA002-HA) (4) using LR clonase (Invirtogen). These constructs were introduced into *Arabidopsis* plants by the floral dip method (5). T3 homozygous progeny were used for these experiments. We totally assayed 4 and 3 independent transgenic Arabidopsis lines of OsglHAT1N-OE and OsglHAT1K-OE, respectively, whose phenotypes are segregating in T2 generation.

RNA extraction, cDNA synthesis and RT-PCR. Total RNA was isolated by using the RNeasy Plant Mini Kit (Qiagene) and then digested by recombinant DNase I (RNase-free, Takara) to remove possible genomic DNA contamination, following the manufacturer's instructions; the resulting total RNA was quantified using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). For first-strand cDNA synthesis, 2 μg of total RNA for each sample was used for reverse transcription using Omniscript Reverse Transcriptase (Qiagene) according to the standard protocol of the manufacturer. The synthesized cDNA was then diluted 1:5 with milli-Q water and used directly for RT-PCR and qPCR reactions.

qPCR was performed on the thermal cycler CFX96 Real-time PCR System (Bio-Rad) using the SYBR Green PCR Master Mix (Bio-Rad) and the primers listed in **Table S1.** The

relative expression level was normalized to ubiquitin. Each analysis was performed in triplicate.

Protein preparation and assays for HAT activity. For the *in vitro* HAT assay, we cloned cDNA ORF encoding the *OsglHAT1* Nipp or Kasa alleles into pET32a (+). *Escherichia coli* BL21 (DE3) pLysS Rosetta-gami 2 (Novagen) was used as a host strain for the production of recombinant fusion HIS-OsglHAT1 proteins. The induction and purification of these proteins were performed as described in the manufacturer's protocol. We purchased a fluorescent HAT Assay Kit (Active Motif) and followed the manufacturer's instructions with the following modifications: the reaction mixture of 30 μl containing 5 × HAT assay buffer, 2 μl acetyl-CoA (0.5 mM), 1 μl *Xenopus* chromatin (treatment of nucleus exaction of 2×10^8 blood cell per milliliter) and the indicated volume of protein (purified fusion or HIS-tag only) was incubated at 30°C for 1.5 h. One third of each reaction mixture, 10 μl, was resolved in 15% SDS-PAGE for a Western blot probed for acetylation of Histone H4 (anti-H4Ac, Millipore).

For the *in vivo* HAT assay, we harvested 1.5 g of young panicle samples from both transgenic *OsglHAT1*-OE and vector control plants, ground them to powder in liquid nitrogen and suspended the samples in extraction buffer I (400 mM Sucrose, 10 mM Tris-Cl, pH 8.0, 10 mM $MgCl_2$, 5 mM β-mercaptoethanol, and complete protease inhibitor cocktail [Roche]). Nuclei preparations were prepared by using extraction buffer II (250 mM Sucrose, 10 mM Tris-Cl, pH 8.0, 10 mM $MgCl_2$, 1% Triton X-100, 5 mM β-mercaptoethanol, and complete protease inhibitor cocktail) and extraction buffer III (1.7 M Sucrose, 10 mM Tris-Cl, pH 8.0, 0.15% Triton X-100, 2 mM $MgCl_2$, 5 mM β-mercaptoethanol, and complete protease inhibitor cocktail). The pellets were suspended in nuclear lysis buffer (10 mM Tris-Cl, pH 8.0, 1% SDS, 10 mM EDTA, and complete protease inhibitor cocktail) for 30 minutes on ice. The reactions were stopped with 2× SDS-PAGE loading buffer (95°C, 5 min), and samples were analyzed by 15% SDS-PAGE.

In situ RNA hybridization. A cDNA fragment was amplified by RT-PCR using the primer-set specific to *OsglHAT1* listed in **Table S1** and cloned into both pBluescript II SK+ and pBluescript II KS+ vectors, linearized and used for making digoxygenin-labelled sense and anti-sense probes, respectively. Sample fixation, section and *in situ* hybridization were performed as described previously (6).

Subcellular localization and *OsglHAT1* promoter-GUS analysis. We made a GFP-OsglHAT1 (from Kasalath) in-frame fusion construct under the control of the CaMV 35S promoter and bombarded the construct into onion epidermal cells using the PDS-1000/He device (Bio-Rad). 4',6-diamidino-2- phenylindole (DAPI, pH 7.0) was used to stain nuclei of onion epidermal cells prior to examination of the transient expression of the bombarded samples using a Zeiss LSM700 confocal laser microscope. Using the primer set listed in **Table S1**, we amplified the *OsglHAT1* promoter segments from both parental genomic DNAs (*pOsglHAT1*^{N}: 1,681 base pairs and *pOsglHAT1*^{K}: 1,652 base pairs). We then inserted these segments into the binary vector pCAMBIA1300, generating transgenic rice plants carrying these constructs. GUS staining of tissues and organs of transgenic plants was carried out as described previously (7). The 20 day-old whole *pOsglHAT1*^{N}-GUS and *pOsglHAT1*^{K}-GUS transgenic plants were homogenized in an extraction buffer for crude protein extraction as described by Yamamoto *et al*. (8). For quantification of GUS activity, a MUG assay was conducted following the method described by Ge *et al*. (7).

Transient expression assays in maize leaf protoplasts. We inserted *pOsglHAT1*^{N} and *pOsglHAT1*^{K} fragments by a combination-digestion of *Xho*I and *Bam*HI into the *NBS-LUC* control reporter construct (9) in which the 35S minimal promoter was replaced by the insertions. Transient expression assays using maize leaf protoplasts were carried out according to the protocol described by Studer *et al*. (10). Reporter assays were performed more than three times with similar results, and each assay contained three technical replicates per construct.

Histological examination by scanning electron microscopy (SEM). Spikelet hulls from NIL(*OsglHAT1*) and Nipp plants were collected before fertilization and fixed in FAA solution (50% ethanol, 5% glacial acetic acid and 5% formaldehyde). The inner epidermal cells of lemma of the spikelet hulls were observed by SEM (S-3000N, Hitachi, Tokyo, Japan). A central 4 mm^2 region of the lemma was photographed and > 50 cells per lemma were measured using ImageJ software (11).

RNA-seq and GO analysis. Total RNA was extracted from CS tissues containing shoot apical meristems of Nipp, *GW6a*-4.6 and *OsglHAT1*-OE plants as described above. Single-end libraries were constructed using the Tru-seq RNA library construction kit (Illumina), and sequencing was performed on an Illumina Genome Analyzer IIx Sequencer. A total of 33 base pair single-end reads were aligned to the transcript sequence of the Nipp genome from IRGSP (http://rapdb.dna.affrc.go.jp/download/archive/irgsp1/) using Bowtie (12). Differentially expressed genes were identified through a pair-wise comparison using EdgeR (normalized with TbT) (13). Two or three biological replicates were used in each genotype to identify transcripts showing significant differences (cut-off false discovery rate (FDR) < 0.05; fold change > 2) between wild type and *GW6a*-4.6 or *OsglHAT1*-OE lines. Functional annotation of significantly different transcripts and enrichment analysis were performed with agriGO (14). Fisher's exact test was conducted to reveal significantly enriched GO terms and a representative set of GO terms was used in **Fig. S18**. The differentially expressed genes are listed in **Table Database S1** and gene ontology analysis data is available in **Table Database S2.**

Sequence analysis of putative OsglHAT1 homologs. Using the $OsglHAT1^N$ (Nipp allele) amino acid sequence as a query string, we performed a sequence blast against the GenBank (NCBI) and RGP databases, identifying a total of 59 putative homologs of *OsglHAT1*. The phylogenetic tree shown in **Fig. S19** was constructed using GENETYX (Ver.10).

Genetic diversity and coalescent simulation analyses. We used a diverse set of rice accessions for the genetic diversity analysis in the *GW6a* region: 50 landraces of *indica*, 14 landraces of *japonica* (see information at http://www.gene.affrc.go.jp/databases-core_collections_wr.php#note02_f), and 34 accessions of *O.rufipogon* (**Table S2**). Accessions were sequenced at three *OsglHAT1* sites—the promoter region, 50 kb upstream and 60 kb downstream of the gene body; nucleotide diversity per site was estimated for landrace groups and for *O. rufipogon* using DnaSP 5.1 (15). We conducted coalescent simulations with a two-population model of domestication as described in Gao & Innan (16), in which we assumed $N_{rufipogon} = N_{sativa}$ =125,000. To estimate the timing of the domestication event, we tested several values ($T_{domestication}$ = {7500, 9000, 10000, 12000}). Selfing rates of landraces and *O. rufipogon* were estimated, respectively, to be 95% and 60% in our simulation, with a recombination rate of 4 cM/Mb across the genome. Selection and bottleneck caused a reduction of genetic diversity in landraces. The severity of the bottleneck for the *indica* and *japonica* domestication process was estimated to be k_{indica} = 1.5 and $k_{japonica}$ = 0.9 (16). To distinguish these two factors, based on a two-population model with bottleneck (as a neutral model), we collected 10,000 simulation replications. We tested whether the low nucleotide diversity observed in rice landraces could be explained by a population bottleneck alone because this would have caused a reduction in nucleotide diversity throughout the genome. Respective neutrality in these three sites was not rejected (**Table S2**).

Fig. S1

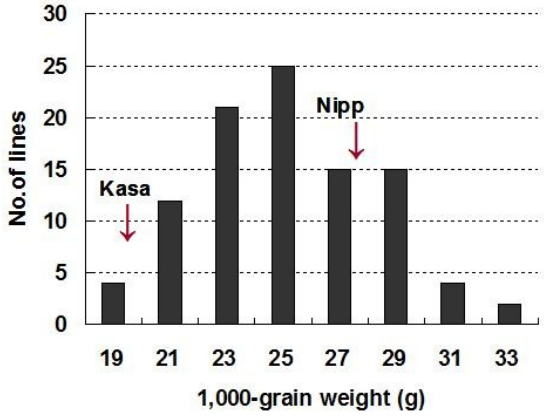

Fig. S1. Frequency distribution of grain weight in the BIL series derived from Nipp and Kasa. Arrows indicate the mean grain weight phenotype for two parents: Nipp and Kasa.

Fig. S2

Fig. S2. Transgenic plants containing GW6a-k-5 and GW6a-k-28 sub-BAC clones bore larger (***A***) and significantly heavier grains (***B***) than the vector control (Control). ***, $P < 0.001$. Student's *t*-test was used to generate the *P* values. Data are the means ± SD ($n = 3$).

Fig. S3

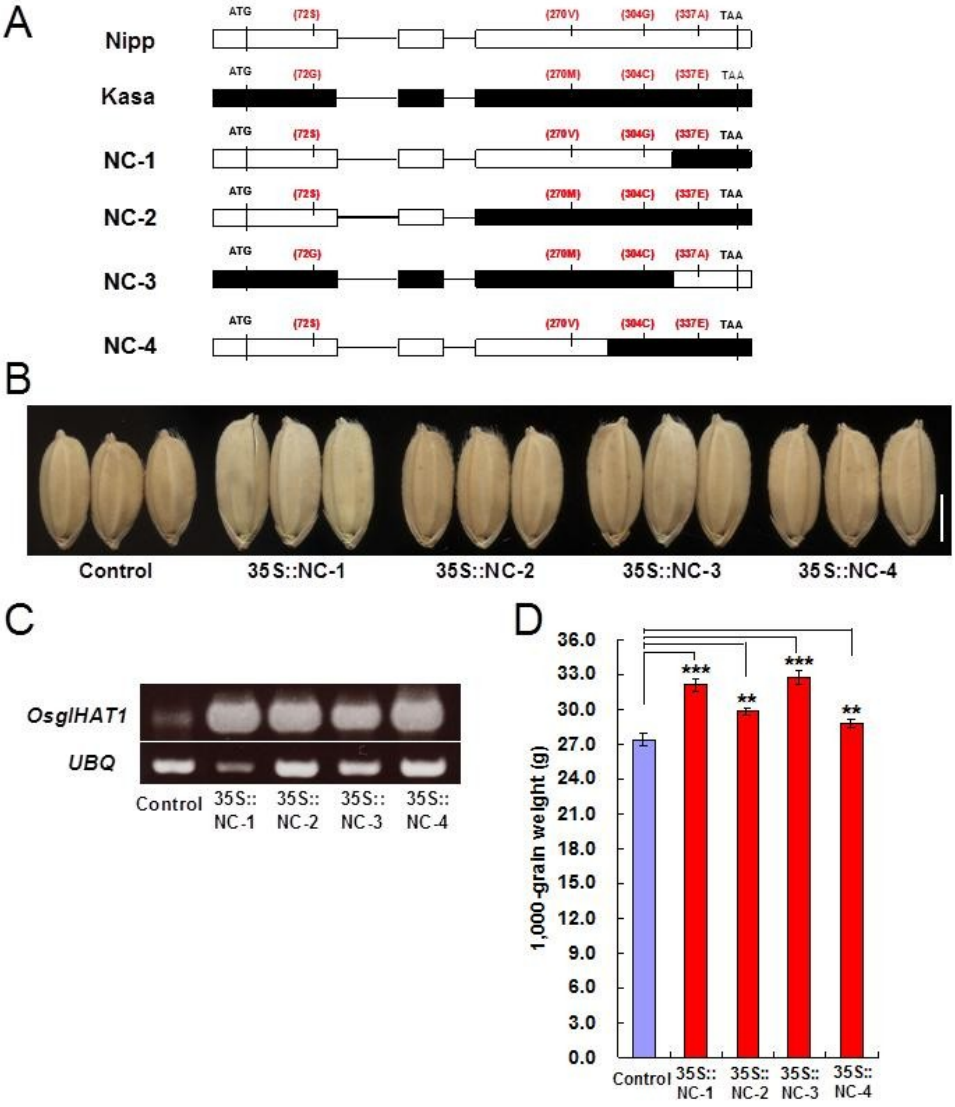

Fig. S3. Transgenic plants carrying amino acid-swapped *OsglHAT1* parental alleles (***A***) bore apparently larger (***B***) and significantly heavier grains (***D***) with increased *OsglHAT1* transcript expression as measured by RT-PCR (***C***). **, $P < 0.05$; ***, $P < 0.001$. Student's *t*-test was used to generate the *P* values. Data are the means ± SD ($n = 3$).

Fig. S4

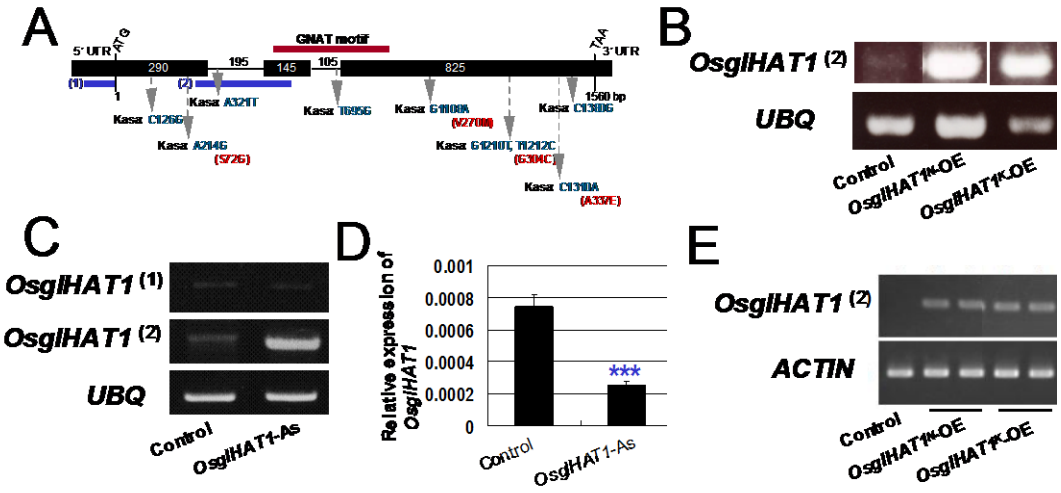

Fig. S4. Levels of *OsglHAT1* transcripts in the transgenic plants were probed. **(*A*)** Gene structure of *OsglHAT1* and relative PCR product locations (the numbered blue bars) for transcription analysis. (***B***) RT-PCR results showing that relative to the vector control, the expression of *OsglHAT1* transcripts was clearly elevated in rice plants containing the *OsglHAT1*^{N}- and *OsglHAT1*^{K}-OE transgenic constructs. (***C***) The enhanced exogenous expression of *OsglHAT1* $^{(2)}$ in the plant containing the *OsglHAT1*^{N}*-AS* transgenic construct indicated a successful transgenic assay, while the endogenous level of *OsglHAT1* transcripts in the same plant was actually reduced, as revealed by the amplification of primer set *OsglHAT1*$^{(1)}$. (***D***) The endogenous *OsglHAT1* transcription by qPCR analysis in the same *OsglHAT1-AS* transgenic plant as in (***C***) using primer set *OsglHAT1*$^{(1)}$ (see legend for ***A***). RNA was isolated and quantitated by qPCR, normalized to ubiquitin. ***, $P <$ 0.001. Student's *t*-test was used to generate the *P* values. Data are the means ± SD ($n = 3$). (***E***) The *OsglHAT1* transcript in *Arabidopsis* transgenic plants was clearly elevated.

Fig. S5

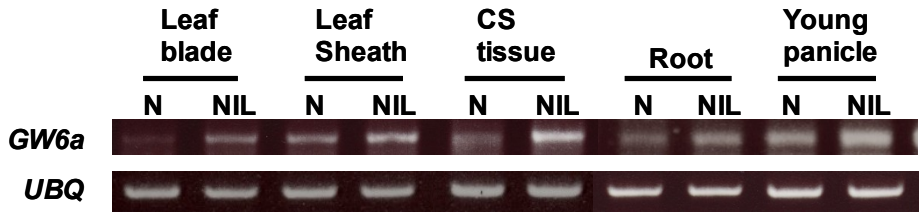

Fig. S5. The expression pattern of *OsglHAT1* was assayed using RT-PCR in the various organs and tissues indicated. N, Nipp; NIL, NIL(*OsglHAT1*); CS tissue, ~1cm-long culm tissue containing the shoot apical meristem.

Fig. S6

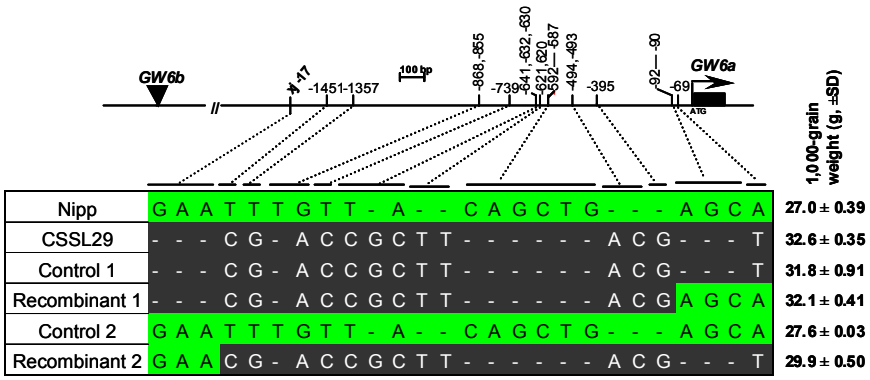

Fig. S6. A genotype map shows the altered SNPs of homozygous recombinants assayed by sequencing the genomic region between markers xj-17 and xj-20 with Nipp and CSSL29 as controls. Relative nucleotide distances from the translation start site (ATG) of the Nipp sequence are shown.

Fig. S7

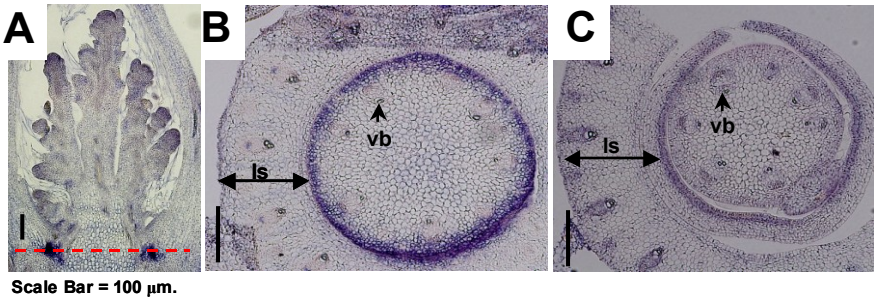

Fig. S7. The *OsglHAT1* mRNA is expressed in the basal part of the abaxial side of the bract shown by *in situ* hybridization of longitudinal (***A***) and transverse **(*B*)** sections compared to a negative control using a sense probe made from the *OsglHAT1* gene (***C***). ls, leaf sheath; vb, vascular bundle.

Fig. S8

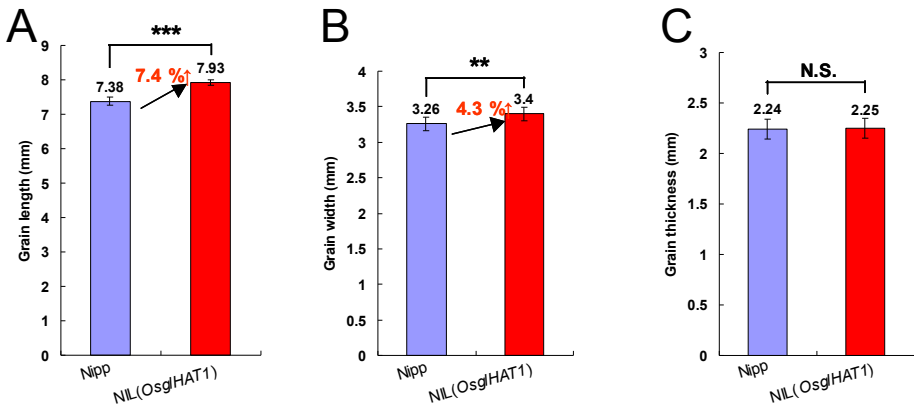

Fig. S8. Comparisons of grain shape components, including grain length (***A***), width (***B***), and thickness (***C***), in Nipp and NIL(*OsglHAT1*) plants. **, $P < 0.05$; ***, $P < 0.001$; N.S., not significant. Student's *t*-test was used to generate the *P* values. Data are the means ± SD ($n = 3$).

Fig. S9

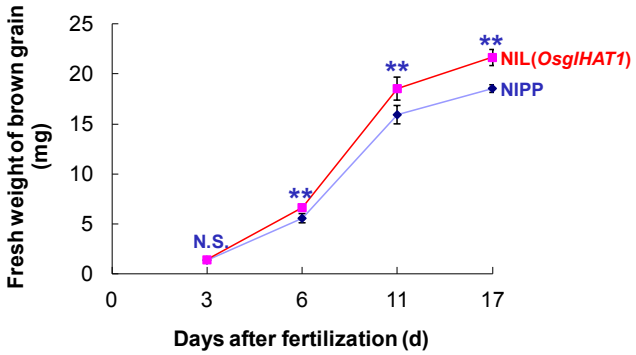

Fig. S9. Characterization of grain milk filling in Nipp and NIL(*OsglHAT1*) revealed the time course of the fresh weight increase of brown grains. Data are the means ± SD (n = ~3 to 5 plants). **, $P < 0.05$; N.S., not significant. Student's t-test was used to generate the P values.

Fig. S10

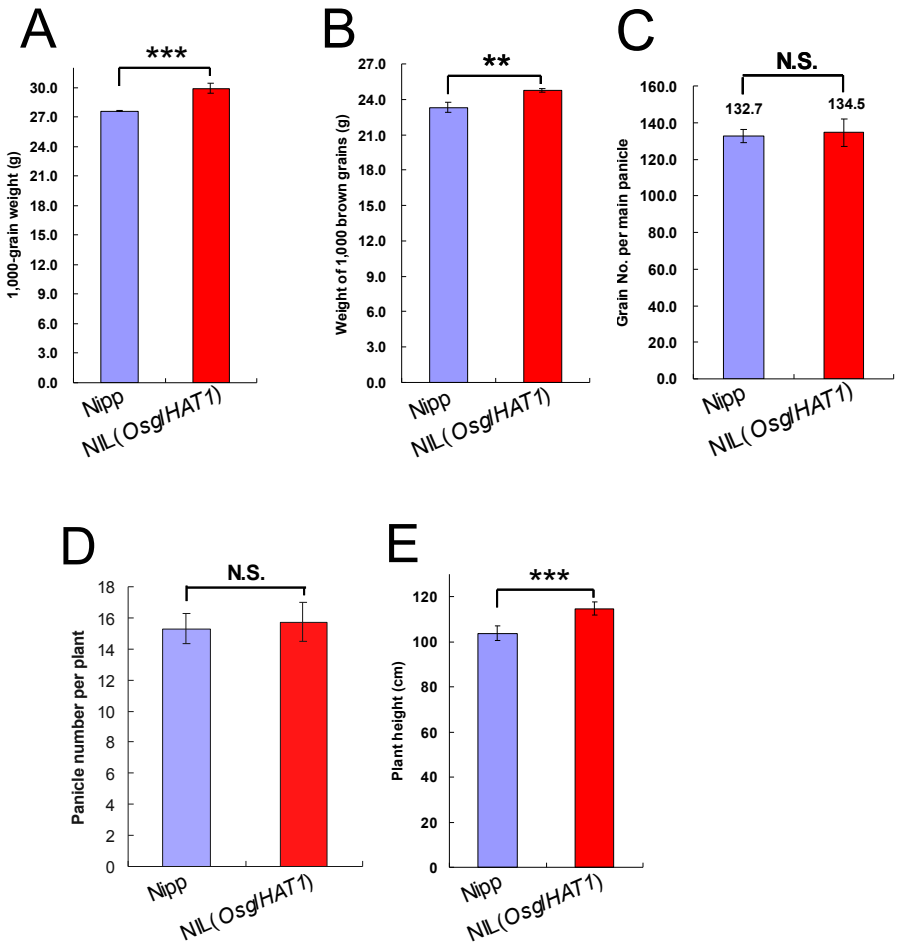

Fig. S10. Comparisons of agronomic traits between Nipp and NIL(*OsglHAT1*), including mean weight of 1,000 grains **(*A*),** mean weight of 1,000 brown grain **(*B*),** mean grain number per panicle **(*C*),** mean panicle number per plant **(*D*),** and mean plant height **(*E*). **,** $P < 0.05$; ***, $P < 0.001$; N.S., not significant. Student's *t*-test was used to generate the *P* values. Data are the means ± SD ($n > 20$ plants).

Fig. S11

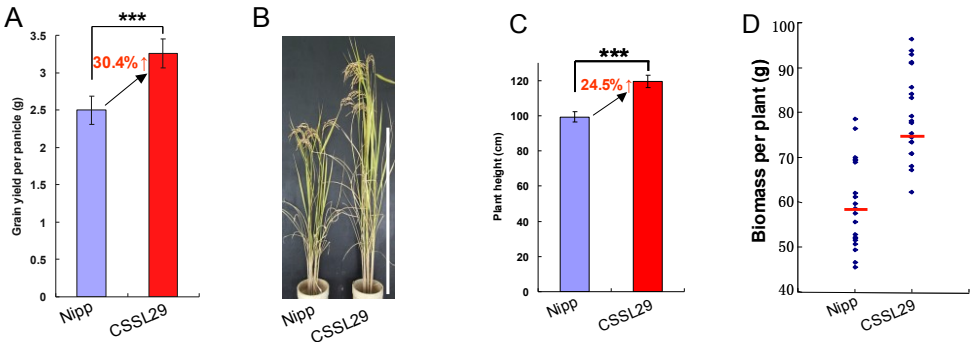

Fig. S11. *GW6* contributes to both grain yield and plant biomass. Comparison of grain yields per panicle (***A***). The plant phenotype of Nipp and CSSL29 (***B***), and accordingly, the quantification of plant height (***C***) and biomass per plant (***D***). ***, $P < 0.001$; N.S., not significant. Student's *t*-test was used to generate the *P* values. Data are the means ± SD ($n >$ 20 plants).

Fig. S12

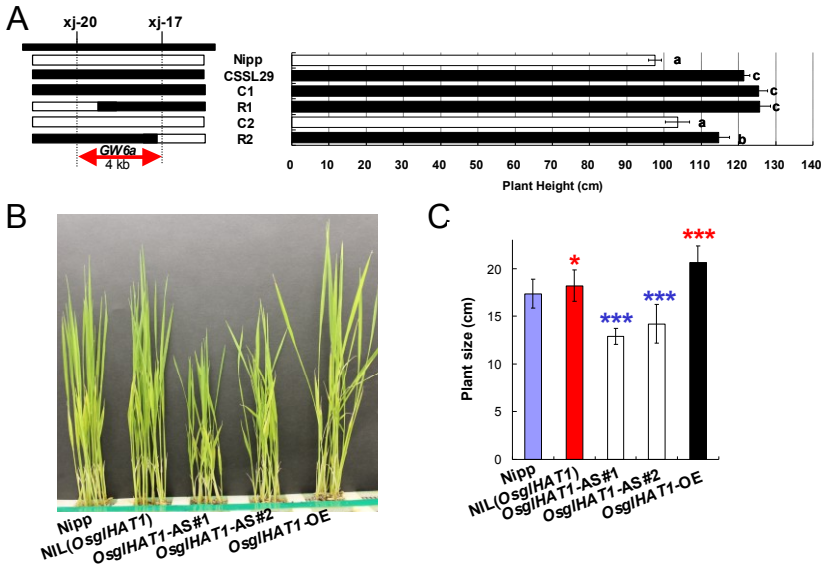

Fig. S12. *OsglHAT1* modulates plant height and vegetative growth. **(*A*)** Genetic evidence showing that the 4-kb region of *GW6a* is also responsible for plant height. (***B***) The early seedling stage phenotypes. (***C***) Quantification of the height of the plants shown in ***B***. *, $P <$ 0.1; ***, $P < 0.001$. Student's *t*-test was used to generate the *P* values. Data are the means ± SD ($n >$ 15 plants).

Fig. S13

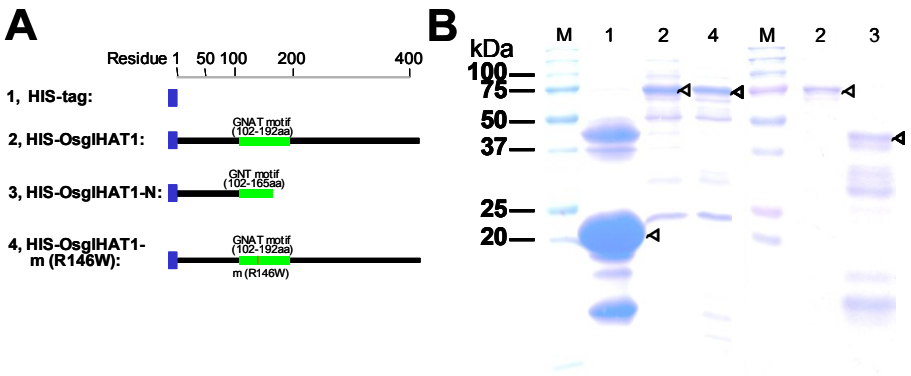

Fig. S13. Purification of the GNAT motif fragment of OsglHAT1. (***A***) Schematic of HIS-tag, the OsglHAT1 protein and derivatives for expression and purification from E.coli cells and for histone acetyltransferase activity assays. (***B***) SDS-PAGE analysis of the purified OsglHAT1 proteins from E.coli cells. Arrowheads indicate HIS-OsglHAT1 fusition proteins or HIS-tag alone.

Fig. S14

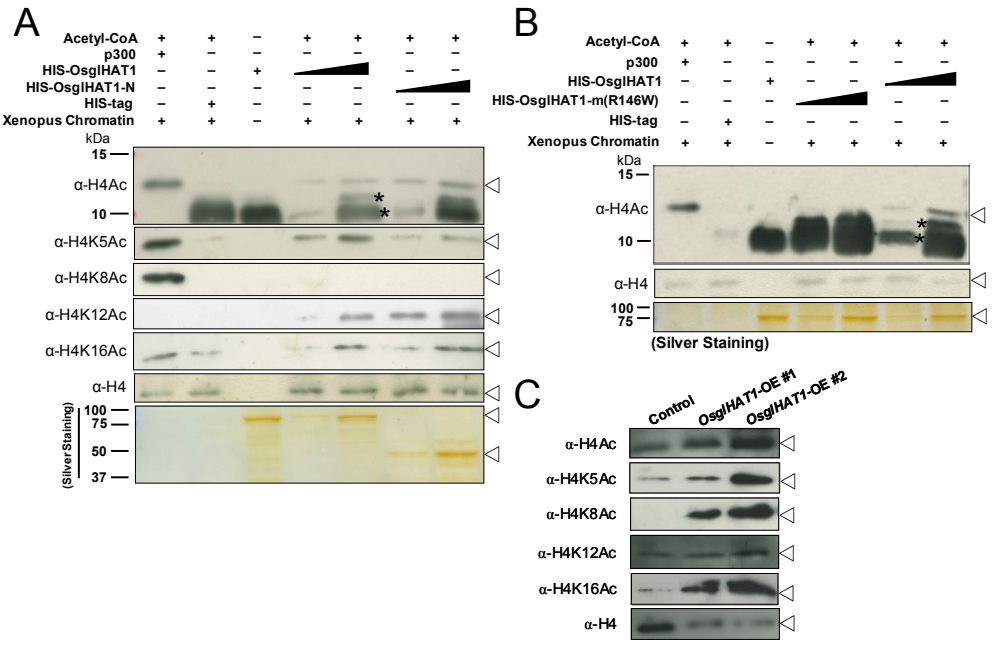

Fig. S14. OsglHAT1 is a histone H4 acetyltransferase. (***A***) *in vitro* HAT assay of OsglHAT1 proteins towards chromatin histone H4. Acetylation was detected by Western blot analysis using an antibody against acetylated histone H4 (H4Ac) or specific acetylation sites in the histone H4 N-terminal tail indicated on the left. (***B***) The R146W mutation of OsglHAT1 protein abolished its ability to acetylating chromatin histone H4 *in vitro* HAT assays. (***C***) The *in vivo* substrate specificity of OsglHAT1. Specific antibodies in Western blot analysis are indicated on the left. Asterisks in ***A*** and ***B*** denote nonspecific bands.

Fig. S15

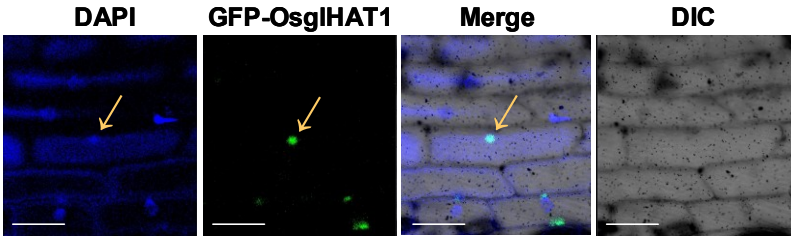

Fig. S15. GFP-OsglHAT1 was localized to the nucleus. DAPI staining indicates the nucleus of the onion epidermal cell. Scale bars: 100 μm.

Fig. S16

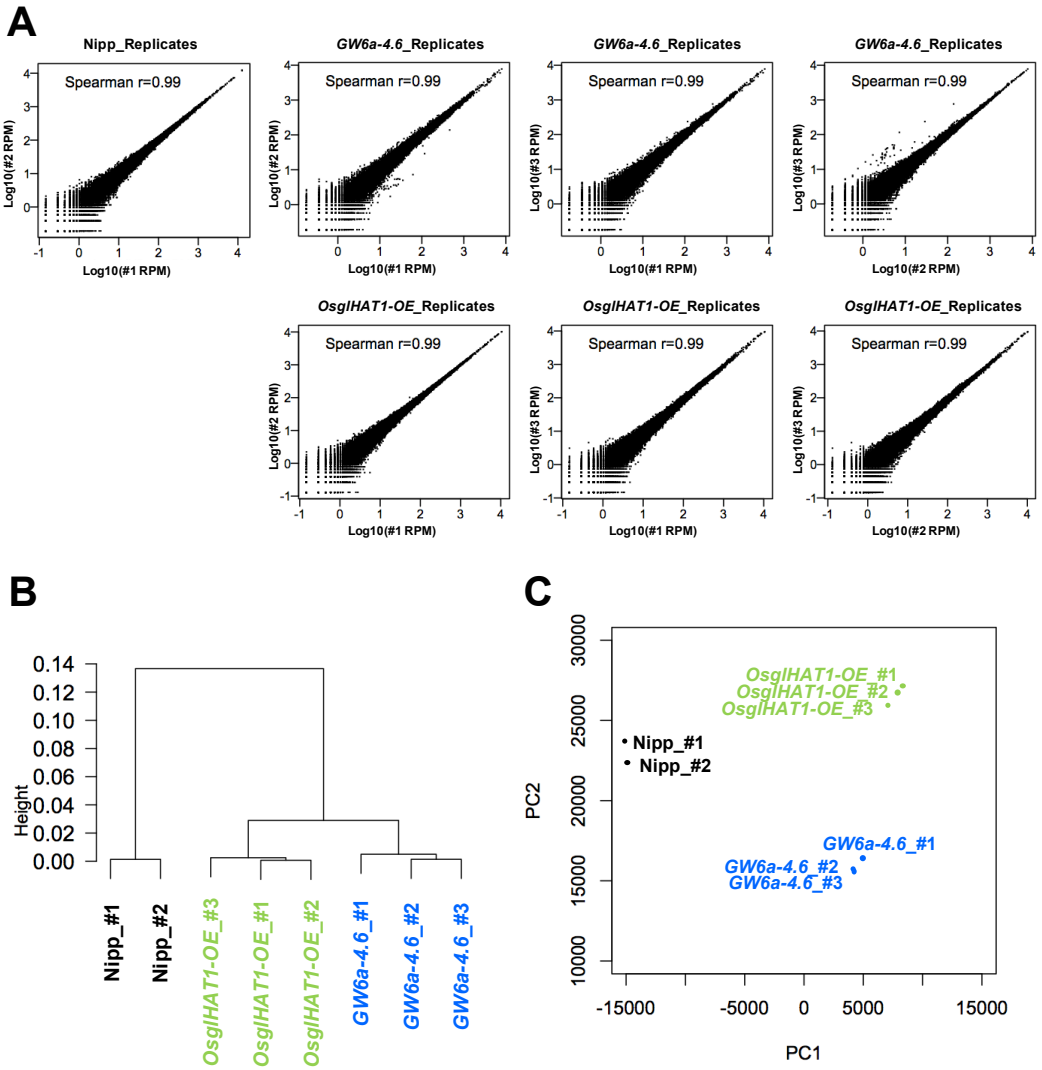

Fig. S16. Biological replicates of RNA-seq results are highly reproducible. (***A***) Correlation of RNA-seq from replicates in the wild type Nipp, GW6a-4.6 and *OsglHAT1*-OE samples. (***B***) Hierarchical clustering of all samples from the wild type Nipp, GW6a-4.6 and *OsglHAT1*-OE. (***C***) Principal component analysis of all samples from the wild type Nipp, GW6a-4.6 and *OsglHAT1*-OE.

Fig. S17

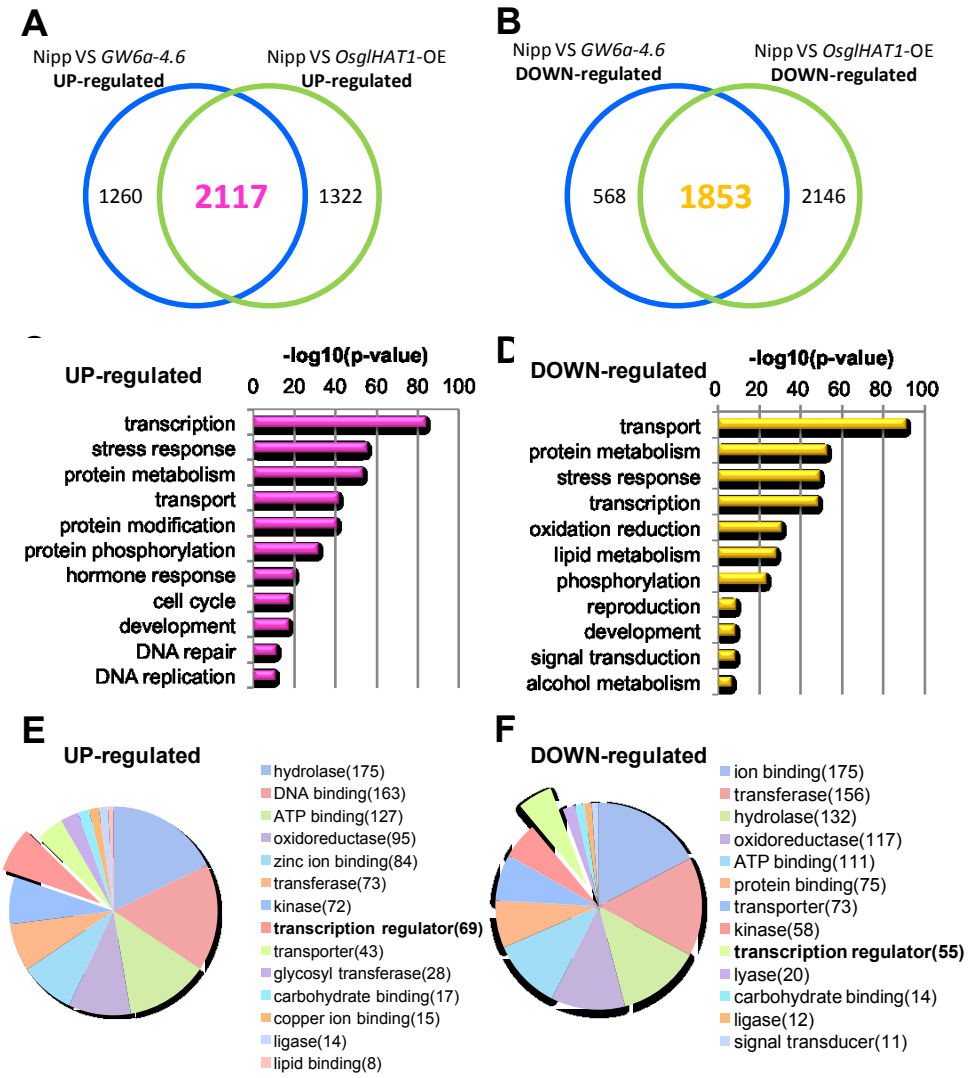

Fig. S17. RNA-seq analysis shows that changed *OsglHAT1* expression in transgenic plants alters transcription of a wide variety of biological processes and molecular functions. Venn diagram shows the numbers of up-regulated (***A***) and down-regulated genes (***B***). Significantly enriched GO terms show representative biological processes of up-regulated (***C***) and down-regulated genes (***D***). Significantly enriched GO -terms of representative molecular function categories of up-regulated (***E***) and down-regulated genes (***F*****)** identified in ***A*** and ***B*****,** respectively.

Fig. S18

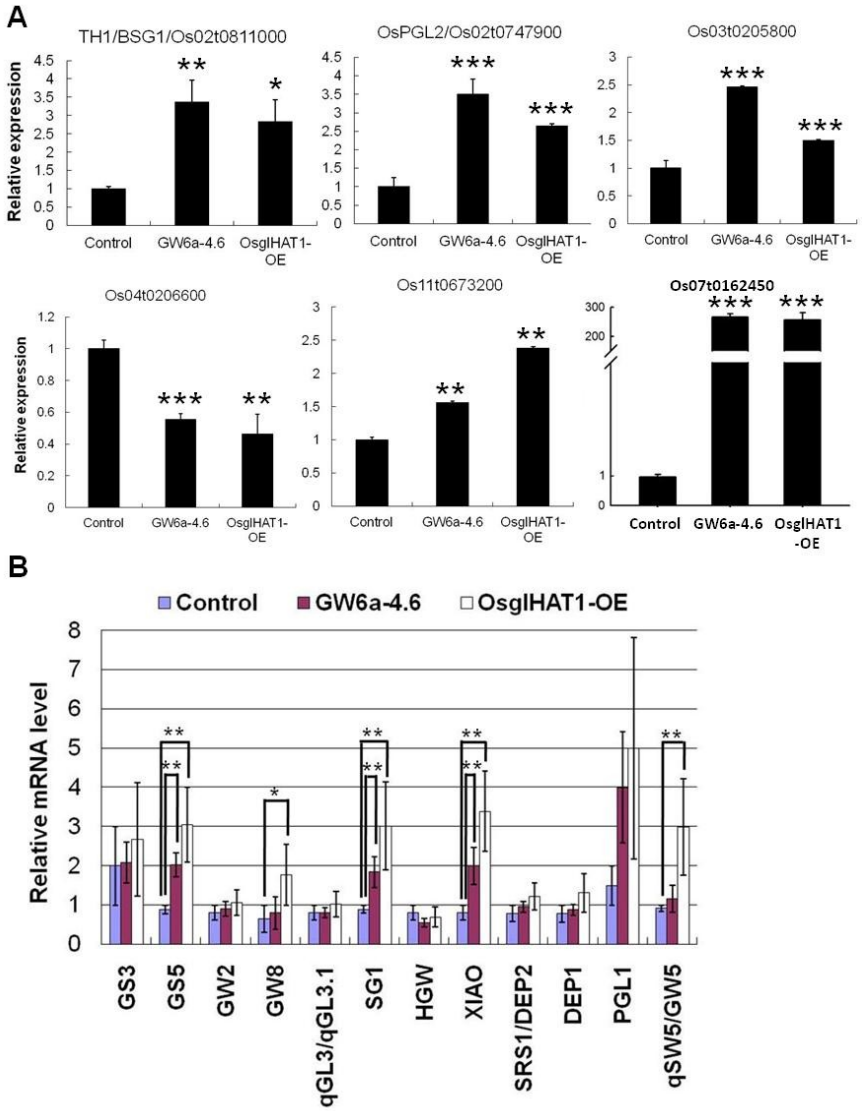

Fig. S18. qPCR analysis of indicated gene expressions (***A***). RNA was isolated from the indicated young panicle tissues, and these RNAs quantitated by qPCR, normalized to *ACTIN*. *, $P < 0.01$; **, $P < 0.001$. Student's *t*-test was used to generate the *P* values. Graph shows comparisons of read counting among the control, *GW6a*-4.6, and *OsglHAT1*-OE genotypes in the RNA-seq experiments (***B***). *, $P < 0.01$; **, $P < 0.001$. We used EdgeR with TbT normalization to find differentially expressed genes and calculate FDR values as described in the **Materials and Methods**.

Fig. S19

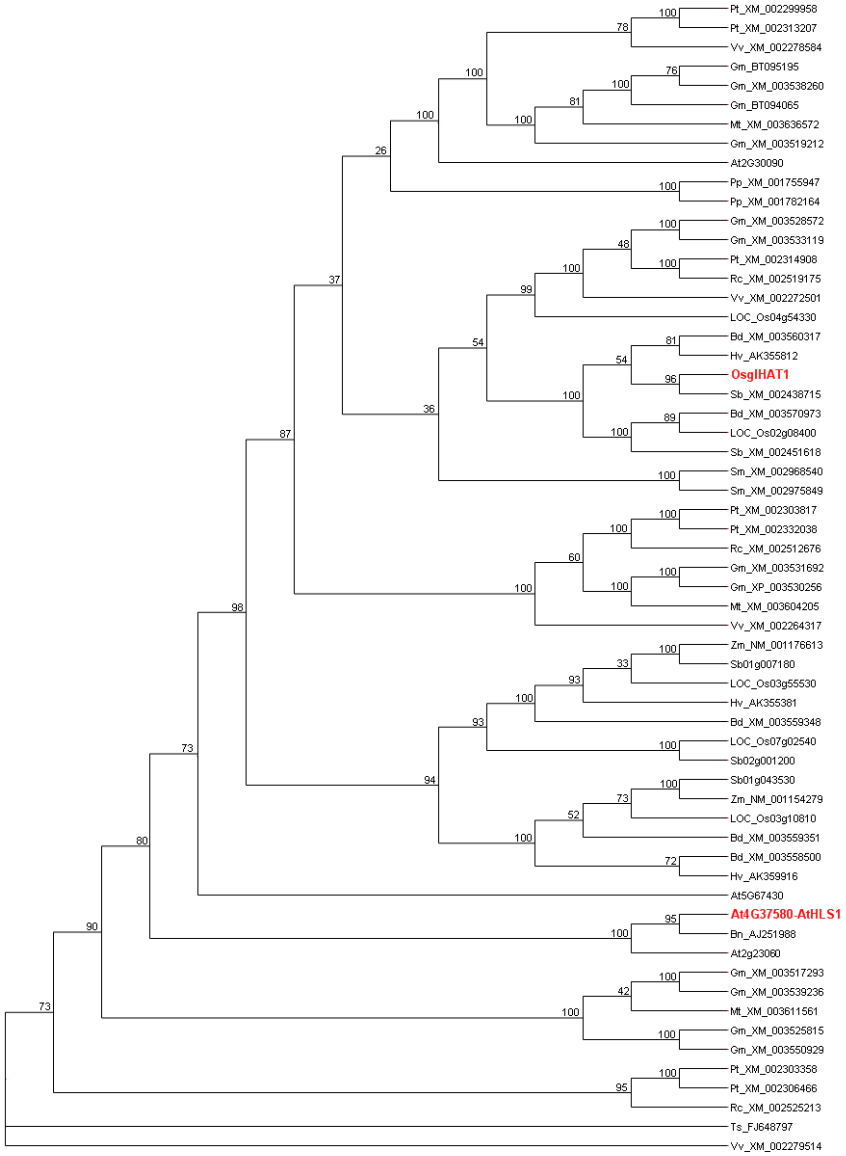

Fig. S19. A phylogenetic view of putative *OsglHAT1* homologs. Fifty-nine *OsglHAT1* homologs were obtained from database searches. At, *Arabidopsis thaliana*; Bd, *Brachypodium distachyon*; Bn, *Brassica napus*; Gm, *Glycine max*; Hv, *Hordeum vulgare*; Mt, *Medicago truncatula*; Pp, *Physcomitrella patens*; Pt, *Populus trichocarpa*; Rc, *Ricinus communis*; Sb, *Sorghum bicolor*; Sm, *Selaginella moellendorffii*; Ts, *Turnera subulata*; Vv, *Vitis vinifera*; Zm, *Zea mays*.

Table S1

Primer sets used in this study

Primer/ gene name	Forward (5'-3')	Reverse (5'-3')	Primer type /usage
xj112	CAC TAA TCA AGC CAC TTC GG	CGA AAC TTG TTT TCC TTC CC	SSR
xj113	AGG AAA ACC GTA GCG TAG AC	GGC TTT CAG CAA TTC ACT GG	SSR
xj-14	GTG AGG GTG TTG ACG ATT TTC	TCC GTT TCC TTA TAG GTT TTG	STS
xj-6	AGC CAA GAA GCA AGA ACT CA	ACC TCA ACC TGT CGC TCA A	STS
xj-11	AGA TAG CTT TAC GGC CTG TT	CAT CGG ATA TGC GGA CAC	STS
xj-20-5	ATA GAG TAT CAT TCC GTT GG	GAG TGG CTC CAT TTC TTG	STS
xj-19-7	TCT GTT GGC AGC ACG ATT TG	CTG TGA ATG CGG CTG TTT GC	STS
xj-5769	ACT GGC AGG ATG AGT GGT A	GGG CCG TTG ATA GTA AAG AT	STS
xj-7	AGG TGG GGC ATG TCG GTG	CGG AAG GCG CAG CAG AGT	STS
xj-16a	TGG ACA CGA ATG AAA AGG	ATA CAG AGA GAG GGG GGA	STS
xj-20	ATC ATT GCC ACC GAT GCT	TTG ACC GGC CAA ATC ACT	CAPS/TaqI
xj-17	ATG TTC GTT CTG GTC TTG	CTG TCC TCT TTT TTC TTC	STS
OsglHAT1	ATG GTG GAG ACG ACG ACG ATG	TTA GAA CTC GCG GGG GTC G	ORF cloning
UBQ	GA CGGA CGCA CCCT GGCT GA CTA C	TGCTGCCAATTACCATATACCA CGA C	RT-(q)PCR
OsglHAT1(1)	CGT GTA TAA ATG CGC CAC AC	GGC CGA TCT CAC CAG CTA C	qPCR
OsglHAT1(2)	GAA GGC GAG CAT GTC TCT CTG CG	GGC GAC CTT CAC GAA TGG CTT C	RT-(q)PCR
pOsglHAT1	GCtctagaG GCC GAT CTC ACC AGC TAC	ACGCgtcgacCGCTGCCAATTCACATTAC	promoter cloning
Up-50K	ATTATGGCACCGGAGTGGTT	GAGCAGGCTAGGACATGGGT	Domestication analysis
Down-60K	GGAAAATGATCCGGCAAG	GCCCGCAAGGAAAGAAAT	Domestication analysis
ACTIN	GTT GGG ATG AAC CAG AAG GA	GAA CCA CCG ATC CAG ACA CT	RT-PCR
	GCT GTT CAC GGG GAG GTT	TGA GGT TCT TGA TGC ACC AG	in situ hybridization
R146W	GGT GTC GCC ATC TCA C**T**G GCG GCT GGG GAT CGG G		Making of construct of
	CCC GAT CCC CAG CCG CC**A** GTG AGA TGG CGA CAC C		OsglHAT1-m (R146W)

Table S2

Sequence diversity in *Oryza sativa* and *rufipogon* around *OsglHAT1* region and results of the tests of selection.

	Oryza rufipogon							*Oryza sativa spp. indica*								*Oryza sativa spp. japonica*							
Gene/region	***N***	***L***	***S***	***h***	***π***	***θ***	***Tajima's D***	***N***	***L***	***S***	***h***	***π***	***θ***	***Tajima's D***	**Coalescent Simulation *rufipogon* VS *indica***	***N***	***L***	***S***	***h***	***π***	***θ***	***Tajima's D***	**Coalescent Simulation *rufipogon* VS *japonica***
up-stream	34	620	40	22	0.01012	0.01578	-1.29410	50	622	14	5	0.00570	0.00538	0.17771	$P > 0.51$	14	659	7	2	0.00152	0.00334	-2.01359*	$P > 0.34$
promoter	34	619	8	8	0.00230	0.00316	-0.80339	50	623	6	5	0.00262	0.00215	0.55712	$P > 0.90$	14	623	1	2	0.00023	0.00050	-1.15524	$P > 0.21$
down-stream	34	520	8	7	0.00424	0.00376	0.37493	50	520	6	5	0.00224	0.00258	-0.33344	$P > 0.47$	14	520	1	2	0.00085	0.00060	0.84228	$P > 0.40$

N, number of sampled sequences; *L*, length of the core alignments in which all sequences contain bases, excluding gaps; *S*, total number of segregating sites; *h*, number of unique sequences (haplotypes); π, average proportion of pairwise differences per base pair at all sites (17); θ, a function of both the number of polymorphic sites and the number of sampled sequences at all sites (18); *Tajima's D*, statistics of neutrality at all sites (19). *, $P < 0.05$.

Table S3. *pOsglHAT1* alleles and the nucleotide polymorphisms in a set of 50 *Oryza sativa* ssp. *indica* cultivars, 14 *Oryza sativa* ssp. *japonica* cultivars, and 34 *Oryza rufipogon* accessions.

				Position in *pOsglHAT1* sequence												
				6	11	22	27	49	54	134	143	293–296	403	407	450	510	
Accession name	Accession no.	Origin	Group	SNP	SNP	SNP	SNP	SNP	SNP	SNP	SNP	InDel	SNP	SNP	SNP	SNP	*pGNAT* type
Kasalath	WRC 02	India	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
ARC 5955	WRC 35	India	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Badari Dhan	WRC 39	Nepal	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Nepal 555	WRC 40	India	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Jena 035	WRC 04	Nepal	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
ARC 7291	WRC 34	India	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Shoni	WRC 31	Bangladesh	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Tupa 121-3	WRC 32	Bangladesh	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Ratul	WRC 36	India	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
ARC 7047	WRC 37	India	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
ARC 11094	WRC 38	India	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Muha	WRC 25	Indonesia	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Jhona 2	WRC 26	India	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Nepal 8	WRC 27	Nepal	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Surjamukhi	WRC 33	India	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Jarjan	WRC 28	Bhutan	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Kalo Dhan	WRC 29	Nepal	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Anjana Dhan	WRC 30	Nepal	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Local Basmati	WRC 42	India	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Kaluheenati	WRC 41	Sri Lanka	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Neang Menh	WRC 58	Cambodia	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Naba	WRC 05	India	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Hakphaynhay	WRC 60	Laos	*Oryza sativa* ssp. *indica*	T	T	G	T	C	A	C	G	Insertion	C	C	C	A	
Radin Goi Sesat	WRC 61	Malaysia	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Kemasin	WRC 62	Malaysia	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Puluik Arang	WRC 06	Indonesia	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Bleiyo	WRC 63	Thailand	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Padi Kuning	WRC 64	Indonesia	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Davao 1	WRC 07	Philippines	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Ryou Suisan Koumai	WRC 09	China	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Shuusoushu	WRC 10	China	*Oryza sativa* ssp. *indica*	T	T	G	T	C	A	C	G	Insertion	C	C	C	A	
Keiboba	WRC 17	China	*Oryza sativa* ssp. *indica*	T	T	G	T	C	A	C	G	Insertion	C	C	C	A	
Bingala	WRC 66	Myanmar	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Rambhog	WRC 65	India	*Oryza sativa* ssp. *indica*	T	T	G	T	C	A	C	G	Insertion	C	C	C	A	
Asu	WRC 13	Bhutan	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	A	
Jinguoyin	WRC 11	China	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Co 13	WRC 15	India	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Vary Futsi	WRC 16	Madagascar	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
IR 58	WRC 14	Philippines	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Milyang 23	WRC 57	South Korea	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Qingyu (Seiyu)	WRC 18	Taiwan	*Oryza sativa* ssp. *indica*	T	T	G	T	C	A	C	G	Insertion	C	C	C	A	
Deng Pao Zhai (Toufutsusai)	WRC 19	China	*Oryza sativa* ssp. *indica*	T	T	G	T	C	A	C	G	Insertion	C	C	C	A	
Tadukan	WRC 20	Philippines	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Shwe Nang Gyi	WRC 21	Myanmar	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Chin Galay	WRC 97	Myanmar	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Deejiaohualuo	WRC 98	China	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Hong Cheuh Zai	WRC 99	China	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Vandaran	WRC 100	China	*Oryza sativa* ssp. *indica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Basilanon	WRC 44	China	*Oryza sativa* ssp. *indica*	C	T	G	G	T	C	C	T	Insertion	C	C	C	G	

Table S3, continued.

continued

Accession name	Accession no.	Origin	Group	Position in *pOsgIHAT1* sequence													*pGNAT* type
				6 SNP	11 SNP	22 SNP	27 SNP	49 SNP	54 SNP	134 SNP	143 SNP	293–296 InDel	403 SNP	407 SNP	450 SNP	510 SNP	
Kasalath	WRC 02	India	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Dahonggu	WRC 12	China	*Oryza sativa* ssp. *indica*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
W2265	W2265	Laos	*Oryza rufipogon*	T	T	G	G	C	C	C	T	Insertion	C	C	C	G	
W2014	W2014	India	*Oryza rufipogon*	T	T	G	G	C	C	C	T	Insertion	C	C	C	G	
AS067	AS067	Thailand	*Oryza rufipogon*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
W0106	W0106	India	*Oryza rufipogon*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
AS049	AS049	Sri Lanka	*Oryza rufipogon*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
AS052	AS052	Nepal	*Oryza rufipogon*	T	T	G	G	T	C	C	T	Insertion	C	T	C	G	
AS081	AS081	Vietnam	*Oryza rufipogon*	T	T	G	G	C	C	C	T	Insertion	C	C	C	G	
W0108	W0108	India	*Oryza rufipogon*	T	T	G	G	C	C	C	T	Insertion	C	C	C	G	
IRGC105402	IRGC105402	China	*Oryza rufipogon*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
W1294	W1294	Philippines	*Oryza rufipogon*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
W1943	W1943	China	*Oryza rufipogon*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
AS085	AS085	China	*Oryza rufipogon*	T	T	G	G	C	C	T	G	Insertion	A	C	C	G	
W1681	W1681	India	*Oryza rufipogon*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
W0593	W0593	Malaysia	*Oryza rufipogon*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
W1715	W1715	China	*Oryza rufipogon*	T	T	T	G	T	C	C	T	Insertion	C	C	C	G	
W1865	W1865	Thailand	*Oryza rufipogon*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
W0137	W0137	India	*Oryza rufipogon*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
W1944	W1944	China	*Oryza rufipogon*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
W1685	W1685	India	*Oryza rufipogon*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
IRGC101508	IRGC101508	India	*Oryza rufipogon*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
AS062	AS062	Laos	*Oryza rufipogon*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
W2264	W2264	Vietnam	*Oryza rufipogon*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
W1551	W1551	Thailand	*Oryza rufipogon*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
W2003	W2003	India	*Oryza rufipogon*	T	T	G	G	T	C	C	T	Insertion	C	C	T	G	
AS059	AS059	Bangladesh	*Oryza rufipogon*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
W1852	W1852	Thailand	*Oryza rufipogon*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
W1669	W1669	India	*Oryza rufipogon*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
AS051	AS051	Nepal	*Oryza rufipogon*	T	T	T	G	T	C	C	T	Insertion	C	C	C	G	
W0574 (IRGC105491)	W0574 (IRGC105491)	Malaysia	*Oryza rufipogon*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
IRGC105908	IRGC105908	Thailand	*Oryza rufipogon*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
W1981	W1981	Indonesia	*Oryza rufipogon*	T	A	G	G	T	C	C	T	Insertion	C	C	C	G	
W0107	W0107	India	*Oryza rufipogon*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
W2266	W2266	Laos	*Oryza rufipogon*	T	T	G	G	C	C	C	G	Deletion	C	C	C	G	
W0610	W0610	Myanmar	*Oryza rufipogon*	T	T	G	G	C	C	C	G	Insertion	C	C	C	G	Kasalath-type allele
Tupa729	WRC 55	Bangladesh	*Oryza sativa* ssp. *japonica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Dianyu 1	WRC 43	China	*Oryza sativa* ssp. *japonica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Tima	WRC 53	Bhutan	*Oryza sativa* ssp. *japonica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Padi Perak	WRC 49	Indonesia	*Oryza sativa* ssp. *japonica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Phulba	WRC 67	India	*Oryza sativa* ssp. *japonica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Khau Mac Kho	WRC 48	Vietnam	*Oryza sativa* ssp. *japonica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Rexmont	WRC 50	United States	*Oryza sativa* ssp. *japonica*	C	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Jaguary	WRC 47	Brazil	*Oryza sativa* ssp. *japonica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Ma sho	WRC 45	Myanmar	*Oryza sativa* ssp. *japonica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Khao Nok	WRC 46	Laos	*Oryza sativa* ssp. *japonica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Urasan 1	WRC 51	Japan	*Oryza sativa* ssp. *japonica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Khao Nam Jen	WRC 68	Laos	*Oryza sativa* ssp. *japonica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Khau Tan Chiem	WRC 52	Vietnam	*Oryza sativa* ssp. *japonica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	
Nipponbare	WRC 01	Japan	*Oryza sativa* ssp. *japonica*	T	T	G	G	T	C	C	T	Insertion	C	C	C	G	

Table S4

Table S4. *O. rufipogon* accession list

Sequence ID	Accession No.	Origin	Group	Annual	Perennial	Allele
W1	W2265	Laos	*O. rufipogon*	Annual	-	Different allele from Kasalath
W2	W2014	India	O. rufipogon	-	-	Different allele from Kasalath
W3	AS067	Thailand	*O. rufipogon*	-	-	Different allele from Kasalath
W4	W0106	India	*O. rufipogon*	-	-	Kasalath allele
W5	AS049	Sri Lanka	*O. rufipogon*	Annual	-	Different allele from Kasalath
W8	AS052	Nepal	*O. rufipogon*	Annual	-	Different allele from Kasalath
W9	AS081	Vietnam	*O. rufipogon*	-	-	Different allele from Kasalath
W10	W0108	India	*O. rufipogon*	-	Perennial	Different allele from Kasalath
W11	IRGC105402	China	*O. rufipogon*	-	-	Different allele from Kasalath
W12	W1294	Philippines	*O. rufipogon*	-	Perennial	Different allele from Kasalath
W13	W1943	China	*O. rufipogon*	Annual	-	Different allele from Kasalath
W14	AS085	China	*O. rufipogon*	-	-	Different allele from Kasalath
W15	W1681	India	*O. rufipogon*	-	-	Kasalath allele
W16	W0593	Malaysia	*O. rufipogon*	-	-	Different allele from Kasalath
W17	W1715	China	*O. rufipogon*	-	Perennial	Different allele from Kasalath
W18	W1865	Thailand	*O. rufipogon*	-	-	Kasalath allele
W19	W0137	India	*O. rufipogon*	-	-	Different allele from Kasalath
W20	W1944	China	*O. rufipogon*	-	-	Kasalath allele
W21	W1685	India	*O. rufipogon*	Annual	-	Kasalath allele
W22	IRGC101508	India	*O. rufipogon*	-	-	Different allele from Kasalath
W23	AS062	Laos	*O. rufipogon*	-	-	Kasalath allele
W25	W2264	Vietnam	*O. rufipogon*	-	Perennial	Different allele from Kasalath
W26	W1551	Thailand	*O. rufipogon*	-	-	Kasalath allele
W28	W2003	India	*O. rufipogon*	-	-	Different allele from Kasalath
W29	AS059	Bangladesh	*O. rufipogon*	-	-	Kasalath allele
W30	W1852	Thailand	*O. rufipogon*	Annual	-	Kasalath allele
W31	W1669	India	*O. rufipogon*	-	-	Kasalath allele
W32	AS051	Nepal	*O. rufipogon*	-	-	Different allele from Kasalath
W34	W0574 (IRGC105491)	Malaysia	*O. rufipogon*	-	-	Different allele from Kasalath
W35	IRGC105908	Thailand	*O. rufipogon*	-	-	Kasalath allele
W36	W1981	Indonesia	*O. rufipogon*	-	-	Different allele from Kasalath
W37	W0107	India	*O. rufipogon*	-	-	Kasalath allele
W38	W2266	Laos	*O. rufipogon*	-	Perennial	Different allele from Kasalath
W39	W0610	Myanmar	*O. rufipogon*	-	-	Kasalath allele

Table database S1. Differentially expressed gene list. The database contains a list of significantly (FDR < 0.05) up- or down-regulated genes with 2-fold or 1/2-fold change in both *GW6a*-4.6 and *OsglHAT1*-OE compared to Nipp. Fold change is indicated as a $\log_{10}$ value.

Table database S2. Enriched GO term. Genes listed in Database 1 were subjected to GO enrichment analysis. Database 2 includes significantly enriched GO terms (FDR < 0.05) for biological process (P), molecular function (F) and cellular component (C). Genes annotated with each enriched GO term are listed in the "entries" column. "bgitem", the background number of genes annotated with the GO term; "querytotal", the number of genes annotated with GO terms in the genes subjected to analysis; "queryitem", the number of genes annotated with the GO term in the genes subjected to analysis.

Supplementary references

1. Liu Y, et al. (1999) Complementation of plant mutants with large genome DNA fragments by a transformation-completent artificial chromosome vector accelerates positional cloning. *Proc. Natl. Acad. Sci. USA* 96: 6535-6540.
2. Nishimura A, Aichi I, Matsuoka M (2006) A protocol for Agrobacterium-mediated transformation in rice. *Nat Protocols* 1: 2796-2802.
3. Song X, et al. (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet* 39: 623-630.
4. Jang I, Yang J, Seo H, Chua N (2005) HFR1 is targeted by COP1 E3 ligase for post-translational proteolysis during phytochrome A signaling. *Genes Dev* 19: 593-602.
5. Clough S, Bent A (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* 16: 735-743.
6. Kouchi H, Hata S (1993) Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. *Mol Gen Genet* 238: 106-119.
7. Ge L, et al. (2010) Arabidopsis ROOT UVB SENSITIVE2/WEAK AUXIN RESPONSE1 is required for polar auxin transport. *Plant Cell* 22:1749-1761.
8. Yamamoto Y, et al. (2010) A rice gid1 suppressor mutant reveals that gibberellin is not always required for interaction between its receptor, GID1, and DELLA proteins. *Plant Cell* 22: 3589-3602.
9. Yanagisawa S, Yoo S, Sheen J (2003) Differential regulation of EIN3 stability by glucose and ethylene signalling in plants. *Nature* 425: 521-525.
10. Studer A,et al. (2011) Identificaiton of a functional transposon insertion in the maize domestication gene tb1. *Nat Genet* 43: 1160-1163.
11. Abràmoff M, Magalhaes P, Ram S (2004) Image processing with ImageJ. *Biophotonics Int.* 11: 36-41.
12. Langmead B, Trapnell C, Salzberg S (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25.

13. Kadota K, Nishiyama T, Shimizu K (2012) A normalization strategy for comparing tag count data. *Algorithms for molecular biology* 7: 5.
14. Du Z, et al. (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* 38: W64-70.
15. Librado P, Rozas (2009) J DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.
16. Gao L, Innan H (2008) Nonindependent domestication of the two rice subspecies, Oryza sativa ssp. Indica and ssp. Japonica, demonstrated by multilocus microsatellites. *Genetics* 179: 965-976.
17. Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437-460.
18. Watterson G (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7: 256-276.
19. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.