SUPPLEMENTARY FIGURES

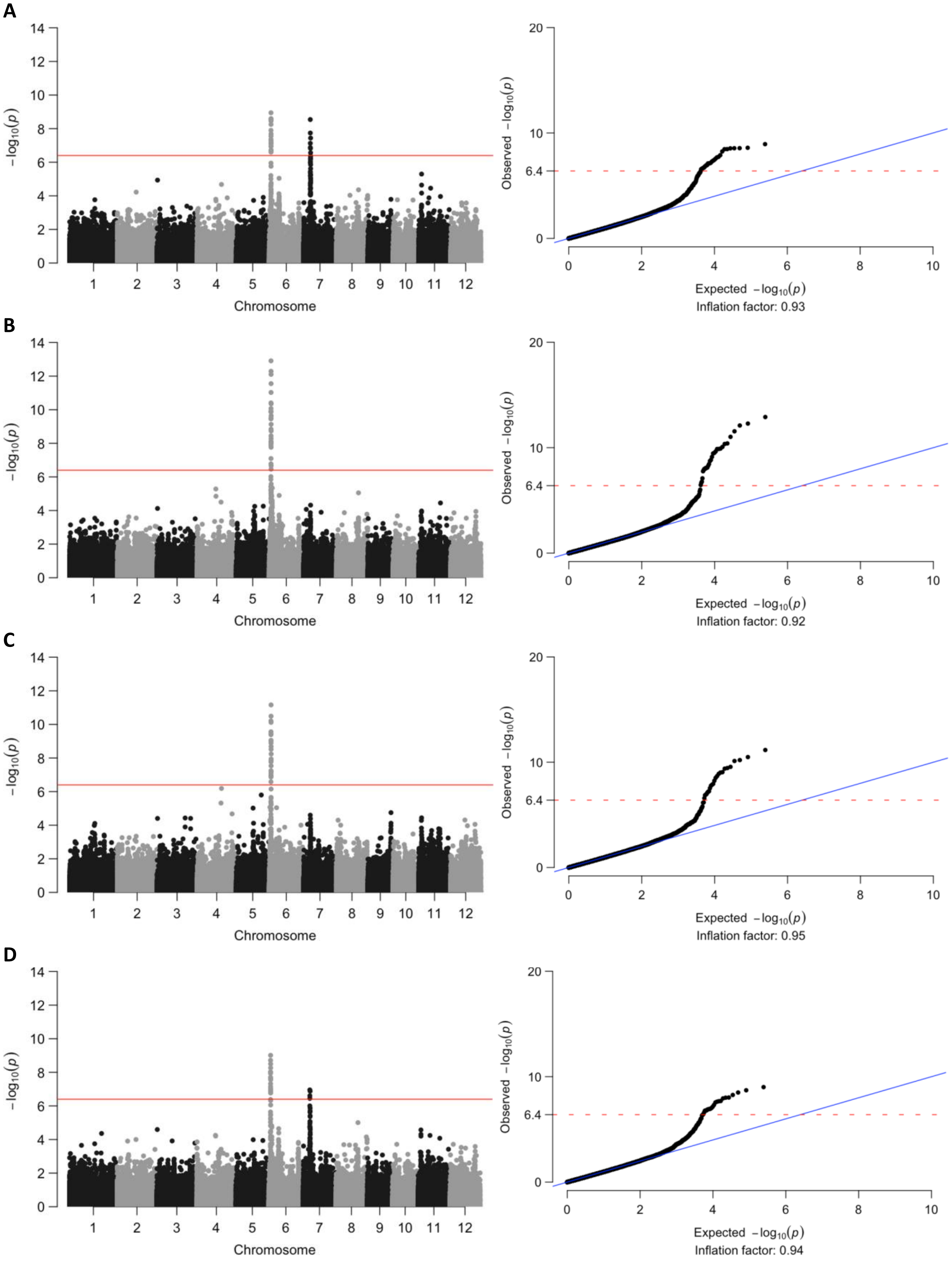

Figure S1. Genome-wide association study for (A) amylose peak 1 (AM1, DP >1,000), (B) amylose peak 2 (AM2, DP 1000-121), (C) medium chain amylopectin (MCAP, DP 120-37) and (D) short chain amylopectin (SCAP, DP 6-36) as depicted by Manhattan (left) and quantile-quantile (right) plots.

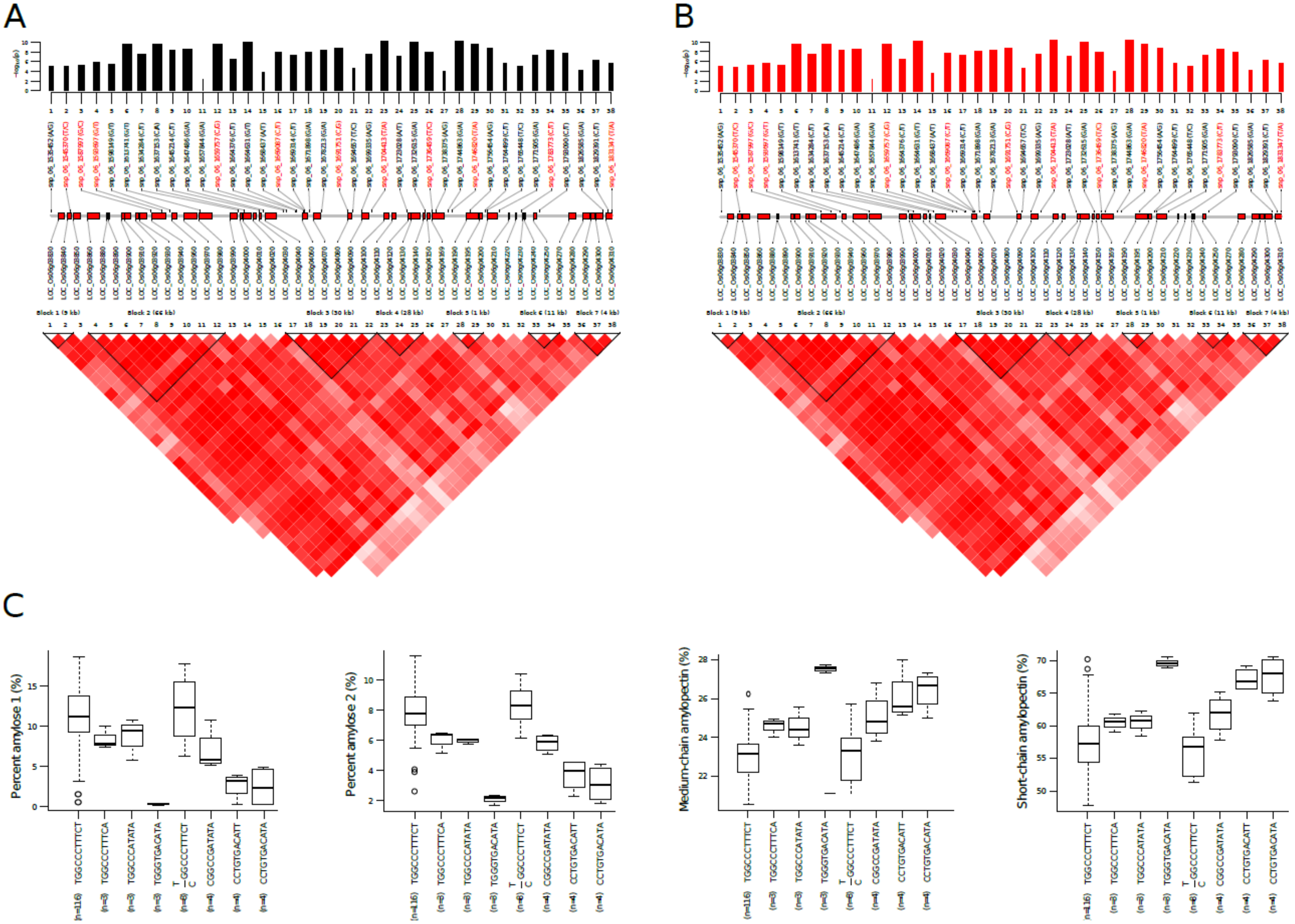

Figure S2. Mapping of the GWAS peak in Chromosome 6 associated with debranched starch structure showing the linkage disequilibrium plot of the most significant SNPs associated with four subfractions of starch corresponding to (A) amylose and (B) amylopectin measured using the SEC method. A scaled section of the chromosome underlying this region is shown where the positions of the genes are labelled with red boxes sized according to their annotation in the Nipponbare reference genome (release 7). The positions of the most significant SNPs are also indicated where the names of the tag SNPs are highlighted in red. The $\log_{10}$-scaled association p-values of these SNPs are shown in the bar plot where bars reflect their relative effect sizes. Also shown are haplotypes formed from the tag SNPs where specific phenotypic ranges for (C) % amylose 1 (AM1), % amylose 2 (AM2), medium chain amylopectin (MCAP) and short chain amylopectin (SCAP) are explained by specific haplotypes.

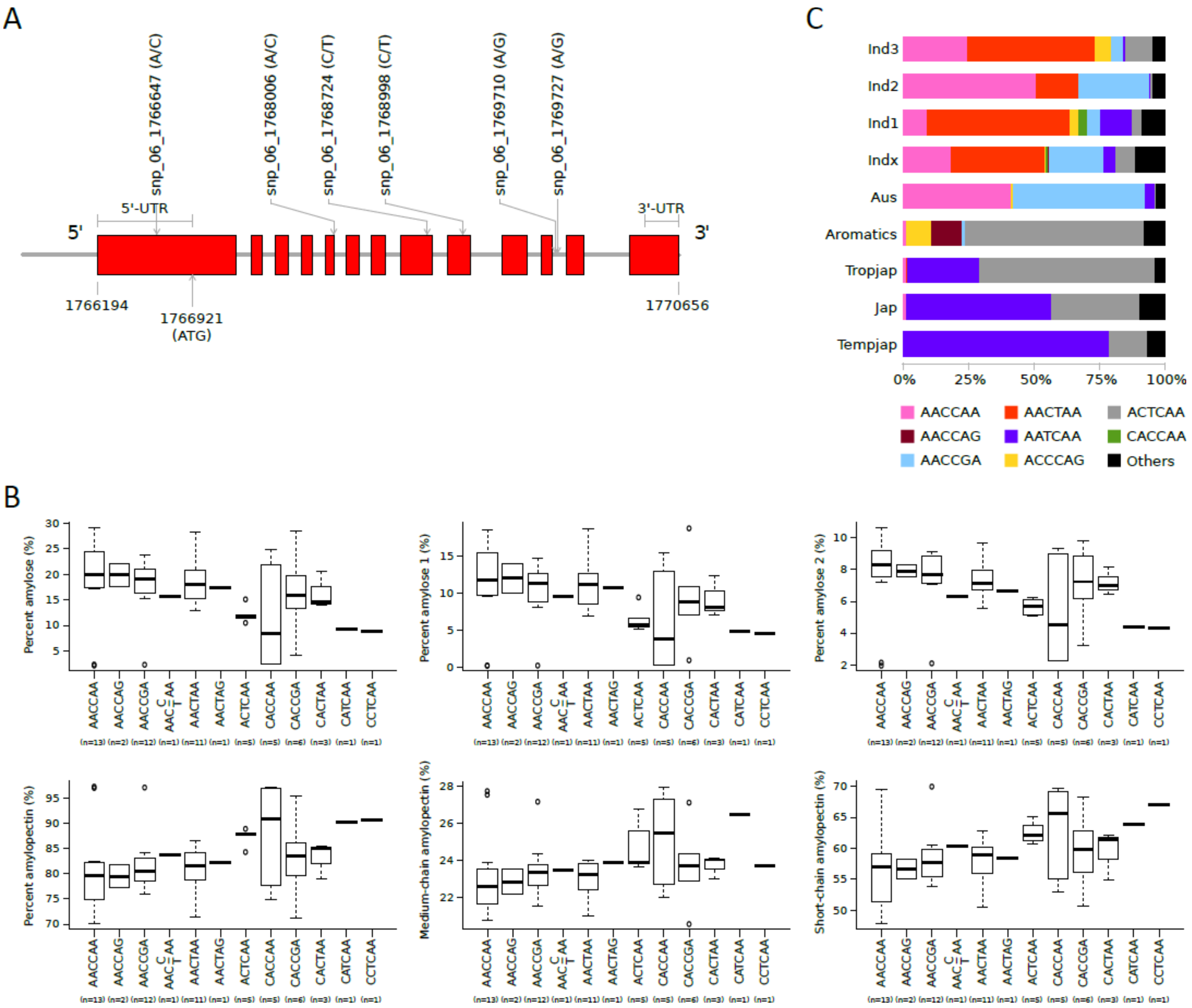

Figure S3. Targeted haplotyping of *waxy* gene (LOC_Os06g04200.3) coding for GBSSI showing SNPs detected from the 700K data (McCouch et al., 2016). (A) The structural gene for *GBSSI* is diagrammatically drawn to scale showing relative location of SNPs detected. (B) Boxplot of the haplotypes detected in a subset of the *indica* diversity panel (n=61) showing tight association with percent amylose (PAM), amylose 1 (AM1), amylose 2 (AM2), percent amylopectin (PAP), medium chain amylopectin (MCAP) and short chain amylopectin (SCAP). (C) Determination of haplotypic frequency distribution of the different *GBSSI* haplotypes among the major subpopulations of rice mined from The 3000 Rice Genomes Project (2014).

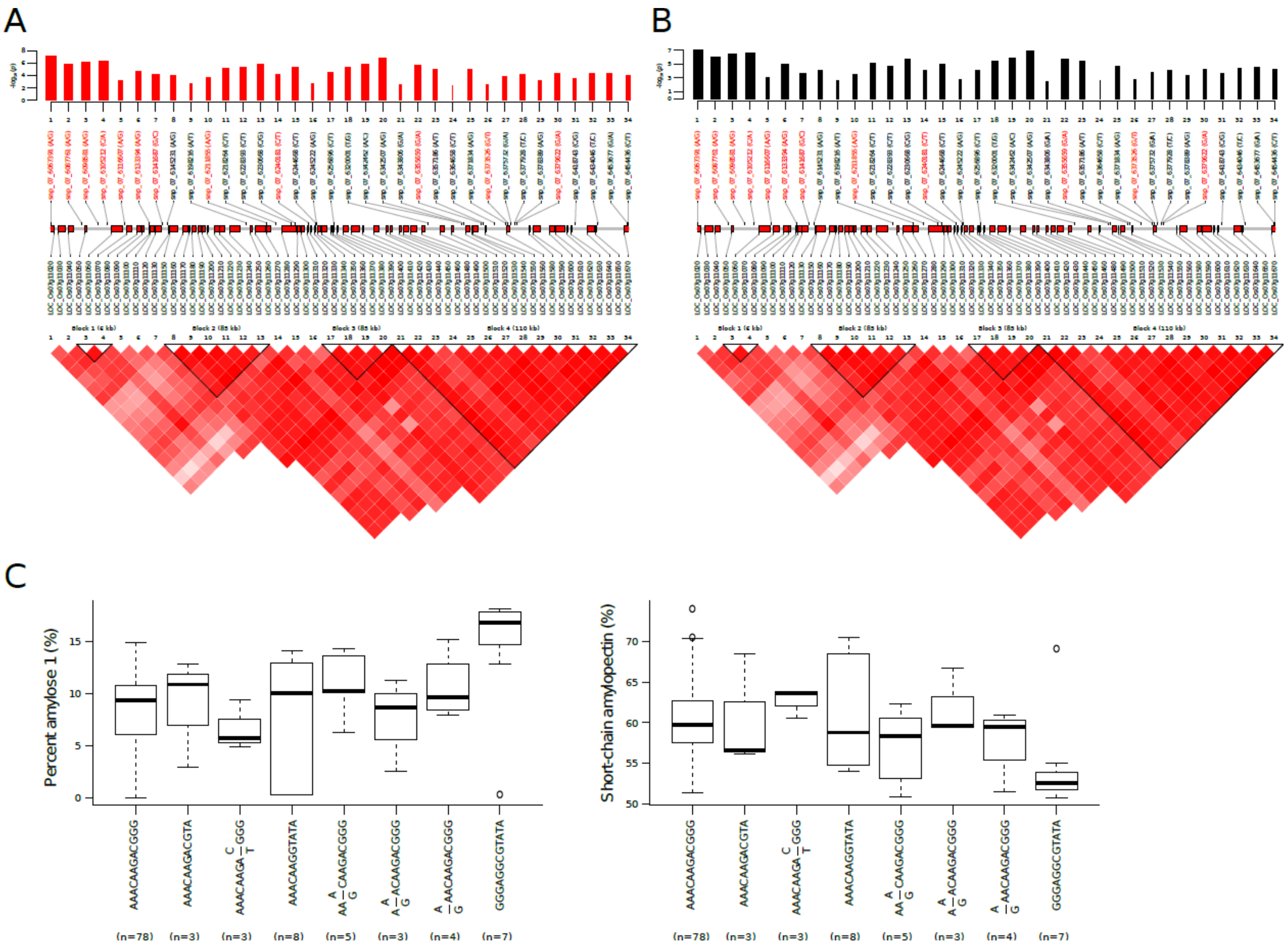

Figure S4. Mapping of the GWAS peak in Chromosome 7 associated with amylose 1 (AM1) and (SCAP) fractions. (A) The linkage disequilibrium blocks formed by the 34 significantly associated SNPs at the genetic interval flanked by LOC_Os07g11020 and LOC_Os07g11670 structural genes (shown as red horizontal bars). These SNPs are based on the Nipponbare reference genome (release 7). Also shown together with the SNP IDs are the alternative alleles. SNP IDs highlighted in red are the 12 tag SNPs from which the haplotypes were formed (B). The bar chart shows the log10 of the association p-values of the SNPs, and the relative thickness of the bars reflects the effect size of each SNP. The bars colored black and the bars colored red mean that these SNPs have positive and negative allelic effects, respectively. The haplotype distribution in the 3000 rice genomes is shown (C).

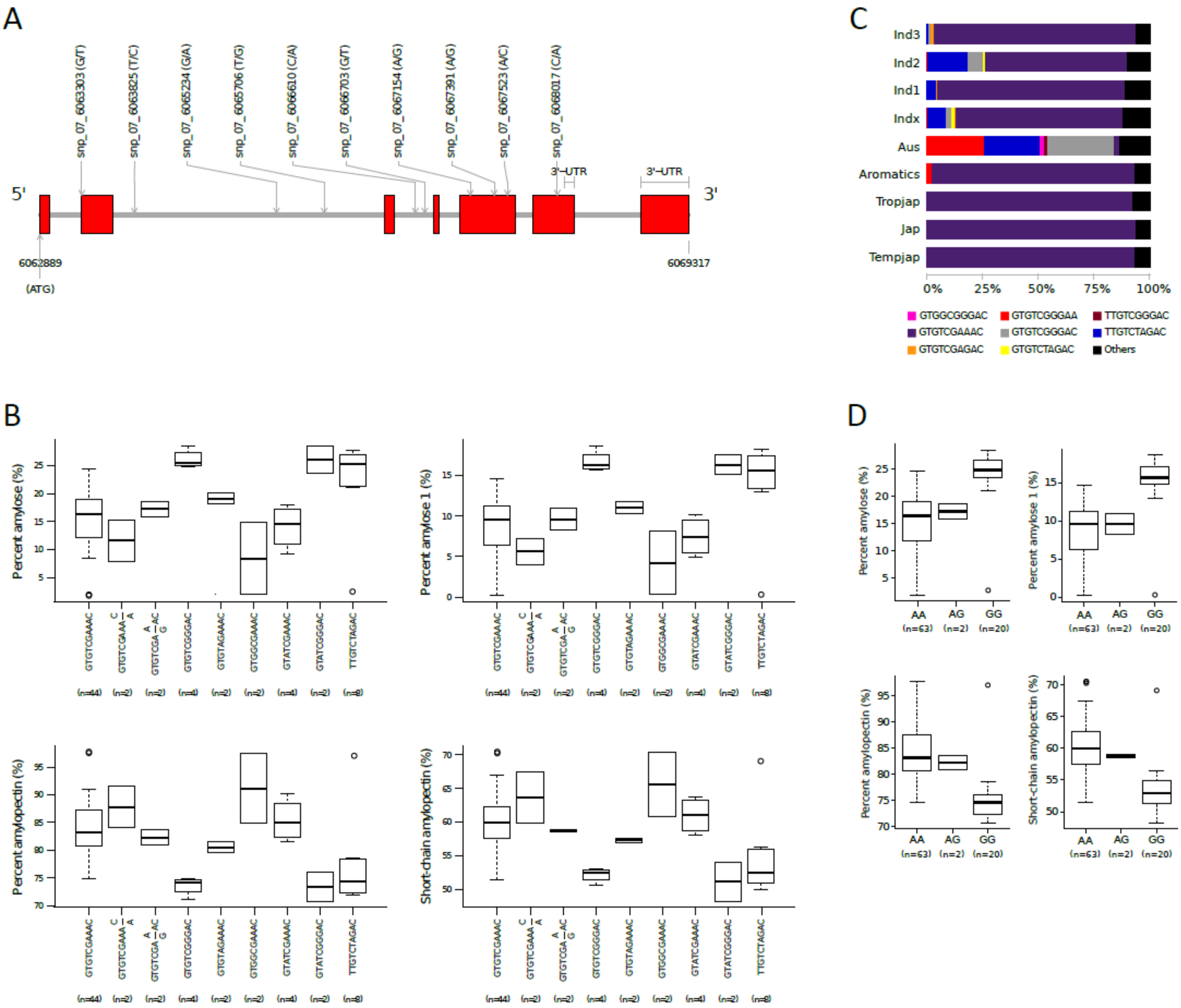

Figure S5. Targeted haplotyping for gene coding for bHLH transcription factor (LOC_Os07g11020) showing tagged SNPs detected from the 700K SNPs data (McCouch et al., 2016). (A) The structural gene is diagrammatically drawn to scale showing relative location of SNPs detected. (B) Boxplot of the haplotypes detected in a subset of the *indica* diversity panel (n=30) showing tight association with percent amylose (PAM), amylose 1 (AM1), amylose 2 (AM2), percent amylopectin (PAP), medium chain amylopectin (MCAP) and short chain amylopectin (SCAP). (C) Determination of haplotypic frequency distribution of the different *bHLH* haplotypes among the major subpopulations of rice mined from The 3000 Rice Genomes Project (2014). (D) The haplotype based on most significant SNP (snp_07_6067391) located in exon 5 of bHLH demonstrating strong influence on percent amylose (PAM) due to amylose 1 (AM1) and percent amylopectin (PAP) due to short chain amylopectin (SCAP) levels.

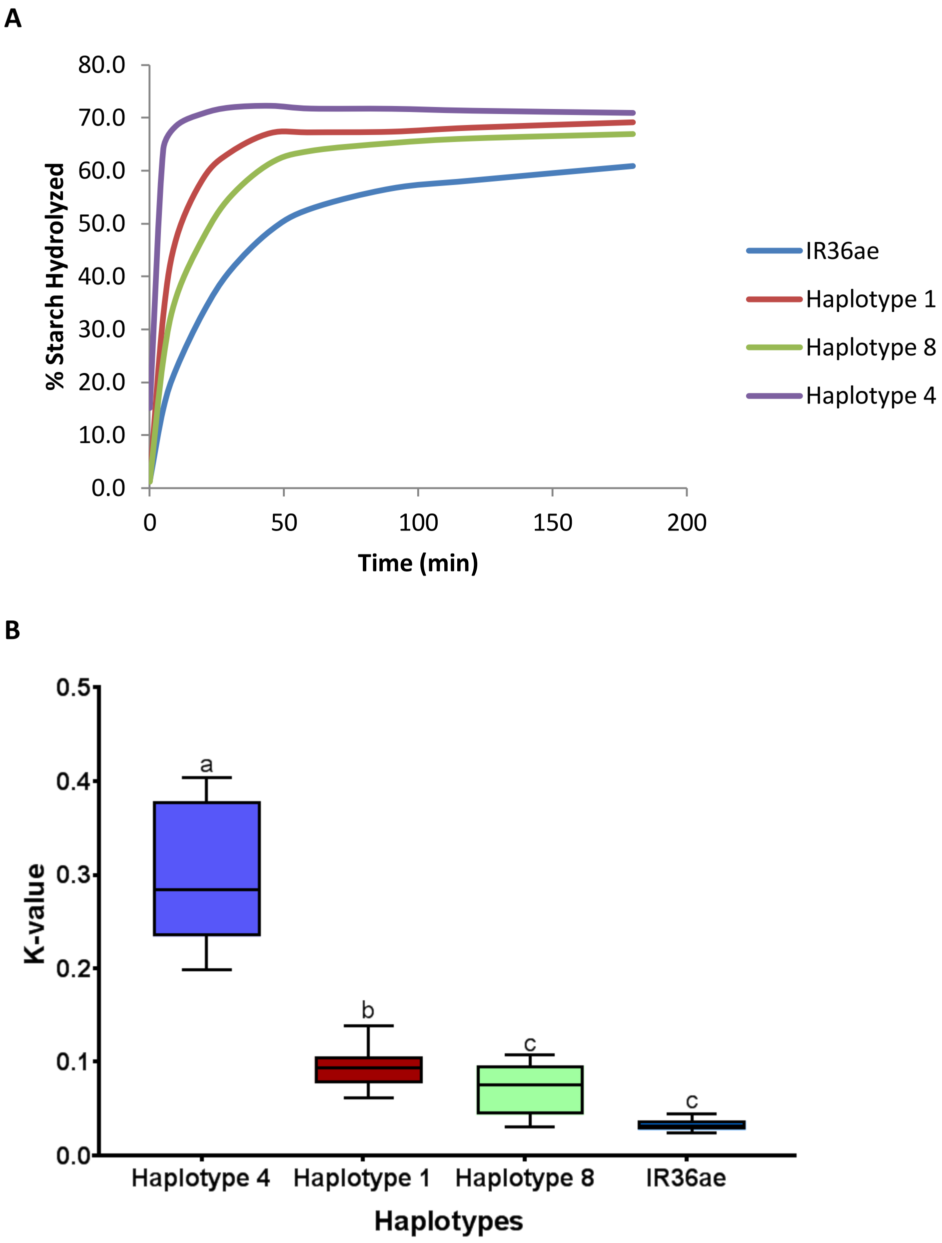

Figure S6. Determination of correlation of starch hydrolysis kinetics (A) and k-value estimations (B) of lines carrying different haplotypes using IR36 amylose extender (IR36ae) mutant as less digestible control.

SUPPLEMENTAL TABLES

Table S1. Amylose classification of *indica* diversity panel using iodine and SEC estimation methods.

Amylose Class	**Range**	**Iodine**	**SEC**
Waxy	0-2.0	6	2
Low	2.1-19.9	43	165
Intermediate	20.0-24.9	117	49
High	25.0-30.0	78	28
Total		**244**	**244**

Table S2. List of most significant (MS) and linked (tag) SNPs detected in Chromosome 6 and 7 and their significance and allelic effects based on 6 fractions defined for debranched SEC. Significance is measured using –log10(p) value while allele effects is measured using beta-coefficient. Three SNPs in GBSSI (LOC_Os06g04200) were added as reference for Chromosome 6, with the location of the MS included based on GBSI start codon in kilobasepairs (Kbp).

(See Table S2.xlsx)

Table S3. Novel SNPs identified at the structural gene of *Granule Bound Starch Synthase I* using a subset of the 700K SNPs data.

SNP Name	Location	SNP		Location	Type of Amino Acid Change	Exon: Amino Acid Change	% Amylose			
		Ref	Alt				R-sq	SE	p-value	-log10(p)
snp_06_1766647	1766647	A	C	5'-UTR	N/A	N/A	0.0443	1.9215	0.0566	1.2475
snp_06_1767387	1767387	G	A	exonic	synonymous SNV	Exon3: R118R	0.0182	4.8109	0.1515	0.8195
snp_06_1768006	1768006	A	C	exonic	nonsynonymous SNV	Exon5 or Exon6: Y224S	0.0504	2.8292	0.0452	1.3445
snp_06_1768724	1768724	T	C	exonic	synonymous SNV	Exon 8 or Exon9: P362P	0.0735	2.6111	0.0196	1.7085
snp_06_1768998	1768998	C	T	exonic	nonsynonymous SNV	Exon10: P415S	-0.0225	6.9040	0.9253	0.0337
snp_06_1769525	1769525	G	A	intronic	N/A	N/A	-0.0059	4.8697	0.4249	0.3718
snp_06_1769710	1769710	A	G	intronic	N/A	N/A	-0.0148	1.9097	0.7233	0.1407
snp_06_1769727	1769727	A	T	intronic	N/A	N/A	-0.0114	4.0210	0.5706	0.2437

Table S4. SNPs and other structural variations in GBSSI previously reported in literature. The major allelic variations are shaded in gray.

Mutation	Genotype and Phenotype Effects	Publication
$(CT)_n$ Repeat at 5'-UTR of Waxy exon 1	Also known as RM190, possibly a closely-linked marker rather than the causal mutation for AAC variation	Bligh et al. (1995) Ayres et al. (1997) Bergman et al. (2001)
23-bp duplication at Exon 2	Present only in waxy varieties, 23 bp duplication 100 bp downstream the ATG causing a premature stop codon which inactivates the *Waxy* gene *wx* = null frameshift mutation coding non-functional GBSSI leading to glutinous rice grain	Wanchana et al. (2003) Mikami et al. (2008) Fitzgerald et al. (2009)
Exon 2 A/G SNP	Wx^{hp} = Low AAC Yunnan landraces A → G Exon 2 at position +497 Asp → Gly substitution leading to reduction in GBSSI activity	Liu et al. (2009)
Exon1-Intron 1 Boundary G/T SNP*	Wx^{a} = G allele = intermediate to high amylose (indica) Wx^{b} = T allele = glutinous to low amylose (japonica) Only distinguishes glutinous and low amylose types from those with intermediate and high amylose content AG**G**T → AG**T**T SNP Reduction in pre-mRNA splicing efficiency and promotion of alternative splicing at cryptic sites in exon 1 = decreased production of functional enzymes = glutinous and low amylose phenotypes	Wang et al. (1995) Ayres et al. (1997) Bligh et al. (1998) Cai et al. (1998) Isshiki et al. (1998) Larkin and Park (1999)
Exon 4 A/G SNP	A → G at +715 Asp → Gly Wx^{op} = indica varieties from India, Nepal, Indonesia and China with opaque and chalky endosperm with very low amylose content	Larkin and Park (1999) Mikami et al. (1999)
Exon 4 G/A and Exon 5 T/C SNP	Wx^{mq} = low AAC rice cv. Milky Queen Two base changes within the coding region:	Sato et al. (2002)

	G → A Exon 4 T → C Exon 5 Missense amino acid substitution	
Exon 6 A/C SNP*	*Wx*^{in} = discriminates intermediate from high amylose Non-conservative mutation at +1,083 Tyr → Ser substitution at the active site of the enzyme reducing its specific activity • Accessions belonging to aromatic group and tropical japonica with intermediate AAC • Non-conservative, changes polarity and function of enzyme • Impacts enzyme activity and amylose content	Cai et al. (1998) Larkin and Park (2003) Chen et al. (2008) Dobo et al. (2010) Kharabian-Masouleh et al. (2011)
Exon 9 T/C	Silent mutation (does not change amino acid sequence)	
Exon 10 C/T SNP*	Proline → Serine Distinguishes high amylose	Larkin and Park (2003) Dobo et al. (2010)

*Most significant impact on starch (Cai et al., 1998)

Table S5. Identification of putative *cis*-regulatory elements binding to the 5'-promoter region of the structural gene of *GBSSI* using online JASPAR 2016 Plant database search tool (Mathelier et al., 2016). 100% relative profile score threshold was used for the search.

Model ID	Model name	Score	Relative score	Start	End	Strand	Predicted Site Sequence
MA1049.1, MA1034.1	ERF094, Os05g0497200	13.294, 14.061	1.000006908, 1.000006685	176	183	-1	CGCCGCCG
MA1053.1, MA0980.1,MA1052.1	ERF109,RAP2-10, RAP2-6	14.122,14.349,14.712	0.999992949, 0.999990699, 1.000001819	177	184	-1	GCGCCGCC
MA0121.1	ARR10	9.98	1.00001178	214	221	1	AGATCCGC
MA1063.1,MA1019.1, MA1066.1	TCP19,Glyma19g26560.1, TCP23	14.648, 14.509, 15.171	1.000002482, 0.9999955, 1.0000007	333	342	1	TGGGGCCCAC
MA1019.1,MA1063.1,MA1066.1	Glyma19g26560.1,TCP19,TCP23	14.509,14.648, 15.171	0.999995573, 1.000002482, 1.00000072	343	351	-1	GGGGCCCAC
MA0053.1, MA0064.1	MNB1A, PBF	8.115, 8.062	1.000019324, 0.999992556	678	682	-1	AAAGC
MA0053.1, MA0064.1	MNB1A, PBF	8.115, 8.062	1.000019324, 0.999992556	901	905	1	AAAGC
MA0959.1, MA0959.1, MA0958.1	AIB, BHLH13	14.865, 14.188	1.000000943, 0.999996191	1444	1451	-1, 1	GCACGTGC
MA0020.1, MA0053.1, MA0064.1	Dof2,MNB1A, PBF	8.76, 8.115, 8.062	0.999993793, 1.000019324, 0.999992556	1547	1552	1	AAAGCA
MA0020.1	Dof2	8.76	0.999993793	1809	1814	-1	AAAGCA
MA0053.1	MNB1A	8.115	1.000019324	1810	1814	-1	AAAGC
MA0064.1	PBF	8.062	0.999992556	1810	1814	-1	AAAGC

Table S6. Motif analysis for starch-related genes and bHLH done using locally installed Meme package (Bailey et al., 2009) with Jaspar Plant database (Mathelier et al., 2016). Those motifs were not showing any hits with the local version of Jaspar database was removed from the table.

(See Table S6.xlsx)

Table S7. List of differentially-expressed genes identified from contrasting lines of Haplotype 8 (high amylose) and Haplotype 4 (waxy).

(See Table S7.xlsx)

Supplementary Table 8. The turquoise module showing the most distinct expression pattern consisting of 441 nodes which codes for most of the genes of interest.

(See Table S8.xlsx)

ADDITIONAL REFERENCES USED:

Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellites and a single-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theoretical and Applied Genetics **94:** 773–781

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren JY, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research **37:** W202-W208

Bergman CJ, Delgado JT, McClung AM, Fjellstrom RG (2001) An improved method for using a microsatellite in the rice *waxy* gene to determine amylose class. Cereal Chemistry **78:** 257–260

Bligh HFJ, Larkin PD, Roach PS, Jones CA, Fu H, Park WD (1998) Use of alternate splice sites in granule-bound starch synthase mRNA from low-amylose rice varieties. Plant Molecular Biology **38:** 407–415

Bligh HFJ, Till RI, Jones CA (1995) A microsatellite sequence closely linked to the *Waxy* gene of *Oryza sativa*. Euphytica **86:** 83–85

Cai XL, Wang ZY, Xing YY, Zhang J-L, Hong M-M (1998) Aberrant splicing of intron 1 leads to the heterogeneous 5 ' UTR and decreased expression of waxy gene in rice cultivars of intermediate amylose content. Plant Journal **14:** 459–465

Chen MH, Bergman C, Pinson S, Fjellstrom R (2008) *Waxy* gene haplotypes: Associations with apparent amylose content and the effect by the environment in an international rice germplasm collection. Journal of Cereal Science **47:** 536-545

Dobo M, Ayres N, Walker G, Park WD (2010) Polymorphism in the *GBSS* gene affects amylose content in US and European rice germplasm. Journal of Cereal Science **52:** 450–456

Fitzgerald MA, Bergman CJ, Resurreccion AP, Moller J, Jimenez R, Reinke RF, Martin M, Blanco P, Molina F, Chen M-H, Kuri V, Romero MV, Habibi F, Umemoto T, Jongdee S, Graterol E, Reddy KR, Bassinello PZ, Sivakami R, Rani NS, Das S, Wang Y-J, Indrasari SD, Ramli A, Ahmad R, Dipti SS, Xie L, Lang NT, Singh P, Toro DC, Tavasoli F, Mestres C (2009) Addressing the dilemmas of measuring amylose in rice. Cereal Chemistry **86:** 492–498

Isshiki M, Morino K, Nakajima M, Okagaki RJ, Wessler SR, Izawa T, Shimamoto K (1998) A naturally occurring functional allele of the rice waxy locus has a GT to TT mutation at the 5' splice site of the first intron. Plant Journal **15:** 133–138

Kharabian-Masouleh A, Waters DLE, Reinke RF, Henry RJ (2011) Discovery of polymorphisms in starch-related genes in rice germplasm by amplification of pooled DNA and deeply parallel sequencing. Plant Biotechnology Journal **9:** 1074–1985

Larkin PD, Park WD (1999) Transcript accumulation and utilization of alternate and non-consensus splice sites in rice granule-bound starch synthase are temperature-sensitive and controlled by a single-nucleotide polymorphism. Plant Molecular Biology **40:** 719–727

Larkin PD, Park WD (2003) Association of *waxy* gene single nucleotide polymorphisms with starch characteristics in rice (*Oryza sativa* L.). Molecular Breeding **12:** 335–339

Liu L, Ma X, Liu S, Zhu C, Jiang L, Wang Y, Shen Y, Ren Y, Dong H, Chen L, Liu X, Zhao Z, Zhai H, Wan J (2009) Identification and characterization of a novel *Waxy* allele from a Yunnan rice landrace. Plant Molecular Biology **71:** 609–626

Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, Shi WQ, Shyr C, Tan G, Worsley-Hunt R, Zhang AW, Parcy F, Lenhard B, Sandelin A, Wasserman WW (2016) JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Research **44:** D110-D115

Mikami I, Aikawa M, Hirano H-Y, Sano Y (1999) Altered tissue-specific expression at the *Wx* gene of the opaque mutants in rice. Euphytica **105:** 91–97

Mikami I, Uwatoko N, Ikeda Y, Yamaguchi J, Hirano H-Y, Suzuki Y, Sano Y (2008) Allelic diversification at the *wx* locus in landraces of Asian rice. Theoretical and Applied Genetics **116:** 979–989

Sato H, Suzuki Y, Sakai M, Imbe T (2002) Molecular characterization of Wx^{mq}, a novel mutant gene for low-amylose content in endosperm of rice (*Oryza sativa* L.). Breeding Science **52:** 131–135

Wanchana S, Toojinda T, Tragoonrung S, Vanavichit A (2003) Duplicated coding sequence in the waxy allele of tropical glutinous rice (*Oryza sativa* L.). Plant Science **165:** 1193–1199

Wang ZY, Zheng F-Q, Shen G-Z, Gao J-P, Snustad DP, Li M-G, Zhang J-L, Hong M-M (1995) The amylose content in rice endosperm is related to the post-transcriptional regulation of the *Waxy* gene. Plant Journal **7:** 613–622