

Genome Wide Association Study on Development and Evolution of Glutinous Rice

Conghui Jiang

China Agricultural University

Muhammad Abdul Rehman Rashid

Government College University

Yan Zhao (✉ zhaoyan1216@163.com)

shandong Agricultural University <https://orcid.org/0000-0002-1011-033X>

Original article

Keywords: Glutinous rice, Genome-wide association study, *Oryza sativa*, Haplotype analysis, Evolution

Posted Date: August 20th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-60218/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: As a globally known staple food, rice is well domesticated in the world. Various regions have different food preferences, which lead to design the goals for breeding programs. Glutinous rice as a special endosperm type is also consumed as a staple food in East Asian countries by consumers' preference. Genetic studies on glutinous rice could be conducive to improve rice quality and understand its development and evolution. Genome-wide association study was performed to explore the associated loci/genes underlying glutinous rice by using 2108 rice accessions.

Results: A total of 399, 262 and 156 annotated genes with significant loci were identified to be associated with endosperm type. By combining the expression patterns analysis, 127, 81, and 48 candidate genes were screened in whole rice panel, *indica*, and *japonica* sub-populations with moderate expression in seeds. There were 32 genes, including three starch synthesis-related genes *Wx*, *SSG6*, and *OsSSIIa*, detected simultaneously in the whole rice panel and subpopulations, playing important role in determining glutinous rice. The main waxy haplotypes of three starch synthesis-related genes were identified by haplotype analysis, while the combined haplotype analyses, revealed that the waxy combined haplotype mainly distributed in Southeast Asia (SEA), SEA islands (SER) and East Asia islands (EAR). Through population structure and genetic differentiation, we proposed the evolutionary model of waxy haplotypes of the three genes from wild rice to *japonica*, and then to *indica*.

Conclusion: It was concluded that the waxy haplotypes of the three genes firstly generated (or were directly inherited from wild rice) in *japonica*, and then flowed into *indica* in SER, SEA and EAR. These results provide valuable information for further gene discovery and understanding the development and evolution for glutinous rice.

Background

Rice (*Oryza Sativa* L.) is one of staple food grain feeding more than half the world's population (Roy and Shil, 2020). High yield and good quality are two important goals of rice production (Fitzgerald et al., 2009). Since the green revolution, new farming methods and breeding techniques have greatly increased food production in many countries (Sasaki et al., 2002; Spielmeyer et al., 2002; Xing and Zhang, 2010; Cui et al., 2019), but the improvement of quality breeding have lagged. The demand of rice with good quality is more urgent for consumers and producers with the improvement of people's living standards. The yield and quality of rice largely determined by the starch content, the ratio of amylose to amylopectin, and the fine structure of amylopectin.

There are two unique subpopulations of rice, *japonica* and *indica*. But whether in *indica* or *japonica*, it can be divided into glutinous and non-glutinous rice. The type of endosperm in rice is mainly determined by the ratio of amylose and amylopectin (Yasunori et al., 2005; Pandey et al., 2012). Compared to non-glutinous rice, the texture of glutinous rice is very sticky. Asian preferences for amylose types have been reported and glutinous rice is also eaten as a staple food of East Asian countries, including Laos and

northern Thailand, known as the "center of the glutinous rice region"(Clarkson et al., 1968; Juliano and Villareal, 1993; Calingacion et al., 2014). Therefore, the formation of waxiness in rice is not only affected by natural selection but also human preference.

Starch biosynthesis is a complex system composed of synthesis of substrate adenosine diphosphate glucose, direct starch, and amylopectin (Bao et al., 2002). It involves 18 starch synthase enzymes related genes, and each gene plays a different role in various stages of starch synthesis (Nakamura, 2002; Martha et al., 2003; Vandeputte et al., 2003; Hannah and James, 2008). In rice grains, the *Waxy* (*Wx*) gene encodes granule-bound starch synthase (GBSS), is a major gene controlling amylose synthesis, and directly affects the amylose content (AC). The rice type (glutinous or non-glutinous) is mainly governed by two alleles (*wx* and *Wx*) of same gene (Yamanaka et al., 2004; Jeng et al., 2009). Recently, a study was reported to enrich the range of breeding materials by using a base editing system at the third, fourth, and fifth exon of *Wx^b* to create a series of mutants with AC of 1.4–11.9% (Xu et al., 2020). Rice starch content is a comprehensive trait contributed by a series of starch synthesis genes. A fine regulatory network that regulates the eating and cooking qualities (ECQs) in edible rice has been clarified by association analysis and transgenic verification experiments in the starch biosynthesis pathway (Tian et al., 2009). *Wx* and *SSII-3* are two major genes that determine ECQs by affecting AC, gel consistency (GC), and gelatinization temperature (GT). *Wx* is the only gene that has a major effect on AC and GC, and has a minimal effect on GT (Tian et al., 2009). Interaction of multiple pairs of genes has a significant effect on rice apparent amylose contents (AAC) (He et al., 2006). Owing to the interaction among starch-synthesis genes, mutation in a single gene will cause changes in the effects of multiple other genes.

Due to the limitations of traditional parental mapping and the special characteristics of waxiness, only *wx* for waxiness had been cloned (Yamanaka et al., 2004; Jeng et al., 2009). More genes related to waxiness are required to be urgently discovered to enrich the natural variation of waxiness and used for glutinous rice breeding. The information about evolutionary origin, domestication and adaptation of key genes of glutinous is conducive to study the formation and evolution of glutinous rice. With the development of sequencing technology, genome wide association study (GWAS) has become an effective mean to discover genes for complex trait (Huang et al., 2010; Huang et al., 2012; Si et al., 2016; Yano et al., 2016; Zhao et al., 2018). The sequencing of 3K core germplasm rice provides strong guarantee for the discovery of waxy genes (Wang et al., 2018).

In this study, 2108 rice germplasm were used in GWAS to identify the significant loci and candidate genes controlling the development of glutinous rice. Haplotype analysis was performed to identify the corresponding glutinous haplotypes of three key waxiness related genes. Combined haplotype analyses were carried out to reveal the genetic characteristics of glutinous rice in Southeast Asia (SEA), SEA islands (SER), and East Asia islands (EAR). Phylogenetic tree and population structure analysis for the origin and evolution of three key waxy genes. Our findings provide important information for further gene discovery and, to gain insight into the evolution and formation of glutinous rice in SEA, SER and EAR.

Results

Endosperm types analysis within different subgroups

A total of 2108 rice accessions, including 1965 non-glutinous (or non-waxy) and 143 glutinous (or waxy) rice accessions (<http://snp-seek.irri.org/>), were used to identify waxiness-related genetic loci and analyze the differentiation for development of both endosperm types (Table S1). Meanwhile, 17,132,232 SNPs of the rice panel were obtained from 3KRGP (Wang et al., 2018). Subsets of these data were further filtered and used in the subsequent analyses.

Reasonable assessment of population structure is conducive to detect the phenotypic differences and subsequent GWAS of natural population. Using Admixture software (Alexander and Lange, 2011), we calculated varying levels of K means within the rice population. The *indica* and *japonica* subpopulations appear clearly at $K = 2$ (Fig.1A). The principal component (PC) analysis indicated that top three PCs explained 17.0%, 6.1% and 2.3% of the genetic variation within the rice panel, which supported that there were two main subpopulations (Fig.1B). Referring to the recent results of 3,010 rice accessions (Wang et al., 2018), we classified the panel into two major subpopulations, 1298 *indica* and 810 *japonica*, although there were several atypical *indica* and *japonica* accessions (Table S1). Hence, the endosperm type of each rice subgroups was compared, and top three PCs were used as covariates to control for subgroup structure in GWAS.

Among the 143 glutinous rice accessions, there were 70 *indica* and 73 *japonica* (Table S1, Fig.1B), suggesting broad genetic variation of trait occurred in *indica* and *japonica*. To study the underlying external factors affecting glutinous differentiation, the geographic distribution of accessions with different glutinous traits was investigated. The vast majority of glutinous rice accessions are distributed in SEA, SER and EAR with 75, 31 and 19 accessions, respectively (Fig.1C). In contrast, non-glutinous rice, as a major endosperm type, was widely distributed in the whole rice growing area (Fig.1D). This geographic distribution was in consistent with previous research that reported the artificial selection of glutinous rice in Southeast Asia (Huang et al., 2012; Calingacion et al., 2014). Taken together, these results suggested that there were large genetic differentiation among glutinous rice accessions, although they were relatively geographically concentrated.

Identification of waxy trait QTLs by GWAS

Under linear mixed model (LMM) with kinship matrix (K) and top three PCs (Q), GWAS was performed to study the genetic basis of endosperm types. Quantile-quantile (Q-Q) plot showed that LMM efficiently controlled population structure and relationships as there was no inflated P values and a majority (95%) of markers exhibited P value equal to or lower than the expected with accordance to null hypothesis (Fig.2A, B and C). Finally, a total 3,338 SNPs located in 399 annotated genes (including gene region and 2 kb promoter region) were identified to associate with endosperm type with threshold of $-\log(P) = 5.6$ (Table S2). Taking into account the large genetic differences between the glutinous accessions of *japonica* and *indica* (Fig.1A and B), we further conducted GWAS of *indica* and *japonica* to explore subpopulation-special waxy genes. According to the above criteria, a total of 2,670 and 1,034 associated SNPs were identified in *indica* and *japonica*, located in 262 and 157 annotated genes, respectively (Fig.2

B and C, Table S2). The GWAS detection efficiency of the whole panel was higher, and the most associated sites were found (Table S2). By comparing GWAS results of three populations, 1,424 significant SNPs (53.3%) of *indica* and 546 significant SNPs (70.3%) of *japonica* could be detected in the whole panel (Table S2). Interestingly, a certain degree of significant loci was detected simultaneously among the whole panel and subpopulations, including 244 common SNPs located in 32 annotated genes (Fig.2D and E, Table S3), indicating that these genes were important and conserved gene response for endosperm type between subpopulation.

Exploration of candidates for endosperm type in rice

The real genes related to rice endosperm type were required to be adequately expressed in seeds at grain filling stage such as *OsAGPL2*, *Wx* and *OsSSIa* (Fig.S1). To further screen candidate genes for endosperm types in QTL regions, we firstly analyzed the expression level of candidate genes in rice seeds at two periods (7-8 and 10-14 days after flowering) of seeds development in rice. Among them, 127 of 399 candidates in whole rice panel, 81 of 262 candidates in *indica*, and 48 of 156 candidates in *japonica* showed moderate expression at least one period (FPKM and RPKM > 10) (Table S4). To further verify the reliability of combined analysis of GWAS, and expression level, the comparison between GWAS detected candidate genes and the known waxy genes was performed. Three starch synthesis-related genes, *Wx* (*LOC_Os06g04200*), *SSG6* (*LOC_Os06g03990*) and *OsSSIa* (*LOC_Os06g12450*) were detected among three populations, respectively. As Manhattan plots showed, these known genes showed top signals in whole rice panel and subpopulations (Figure3A, B and C). Meanwhile, *OsSSI* (*LOC_Os06g06560*) showed association with endosperm types in the GWAS results of whole rice panel (Fig.3D). According to the functional reports of these genes, these genes are involved in starch synthesis, *Wx* is well known gene responsible for controlling amylose synthesis, and its multiple alleles had been identified (Cai et al., 1998; Isshiki et al., 1998; Wanchana et al., 2003; Mikami et al., 2008; Yang et al., 2013). The other three genes affect the morphology and amylose content of starch (Gao et al., 2003; Fujita et al., 2006; Matsushima et al., 2016). The comparison of the GWAS results and known genes indicated that the GWAS results for endosperm type were credible, the four known genes were key loci for natural variations of rice endosperm type.

Natural variation in three key genes responsible for rice endosperm type

Due to exploration of natural variation of key endosperm type is beneficial in breeding for high-quality rice. We performed haplotype analysis to identify their elite alleles of the three key genes (*Wx*, *SSG6* and *OsSSIa*) for rice endosperm types. Firstly, association analysis of candidate genes was performed between endosperm types and 537 SNPs with MAF > 0.01 located in three known genes. Of these, 100 SNPs were significant associated with rice endosperm type ($-\log(p) > 2$). Here, we focused on non-synonymous SNPs, SNPs at splice site and SNPs in promoter (Table S5), as these SNPs could be responsible for functional variation through changes in expression and protein sequence (Yano et al., 2016; Yu et al., 2018; Zhao et al., 2018). A total of 37 significant SNPs were identified within *Wx* gene, including two non-synonymous SNP, one SNP at splice site, and 34 in promoter or UTR regions. Twenty-six

haplotypes, named *Wx-1* to *Wx-26*, were identified in whole panel (Fig.4A). Twenty-four of 26 haplotypes were detected in *indica*, eight of which showed moderate frequencies ranging from 5% to 23.3%. By comparison, 47.8% and 24.1% *japonica* carried *Wx-8* and *Wx-9*, suggesting there were large genetic variation of *Wx* in *indica* than *japonica* (Fig.4A). Previous studies showed that Chr6_1765761 was a key functional SNP for post-transcriptional modification of *Wx* (Cai et al., 1998). The mutant of fifth exons of *Wx^b* induced to lower AC than that of glutinous rice. In our study, we did not detect a unique waxy haplotype of *Wx*. *Wx-9* (allele T at Chr6_1765761) considered as the waxy haplotype, with 37 of 108 in *indica* and 51 of 173 in *japonica* were glutinous rice (Fig.4A and Fig.S2). The results suggested that *Wx* was not the only key gene accounted for natural variation in proportion of amylopectin and amylose in rice. Waxiness of rice, as a physiological trait, is often the result of the continuous joint change of multiple biochemical processes of starch biosynthesis.

Based on 17 significant SNPs within *SSG6* (twelve in the promoter, one in the 5'UTR, 3 non-synonymous SNPs and one in the 3'UTR), nine major haplotypes, named *SSG6-1* to *SSG6-9*, were identified in whole panel. *SSG6-2*, *SSG6-3*, *SSG6-7* and *SSG6-8* were predominantly represented *indica* varieties, accounting for 15.2%, 43.4%, 10.1% and 22.3% of the total, respectively. Moreover, *SSG6-1*, *SSG6-3*, *SSG6-5*, *SSG6-6* and *SSG6-7* were predominant within *japonica*, accounting for 10.1%, 11.2%, 14.4%, 19.0% and 29.1% (Fig.4B). The results indicated the existence of a certain degree of genetic differentiation of *SSG6* between *indica* and *japonica*, although there were two shared haplotypes between *indica* and *japonica*. Further study showed that *SSG6-7* could be considered as main waxy haplotype, due to that 41 of 131 *indica* and 20 of 236 *japonica* carrying *SSG6-7* were glutinous rice. Additionally, a *japonica*-special glutinous haplotype *SSG6-5* was detected with 47 of 117 *japonica* carrying *SSG6-5* were glutinous rice (Fig.4B and Fig.S2). Meanwhile, we detected 5 haplotypes (named *OsSSIIa-1* to *OsSSIIa-5*) of *OsSSIIa* gene, based on 10 significant SNPs (five in promoter, one in 5'UTR, two non-synonymous SNPs and two in 3'UTR). *OsSSIIa-1*, *OsSSIIa-2* and *OsSSIIa-3* were predominant in whole panel (Fig.4C). There was no obvious genetic differentiation of *OsSSIIa* between *indica* and *japonica*. *OsSSIIa-1* could be considered as waxy haplotype, as 58 of 296 *indica* and 54 of 212 *japonica* accessions carrying *OsSSIIa-1* were glutinous rice (Fig.4C and Fig.S2).

Taken together, we identified the key glutinous rice haplotype of each gene (Fig.S2), although none of them completely determined the waxiness of rice. Furthermore, it provides an important message that waxiness of rice, as a physiological trait, is also determined by a complex network, rather than simple genes in the biochemical synthesis pathway in the traditional sense. To prove the above hypothesis, we first examined the geographical distribution of different haplotype combinations of the three genes. Totally, there were 27 haplotype combinations of three genes. Among 127 glutinous rice accessions, haplotype combinations with more than three accessions were listed (Fig.5A). Among 75 glutinous rice of SEA, 33 accessions carried the haplotype combination of *Wx-9*, *SSG6-7* and *OsSSIIa-1*, 12 accessions carried the haplotype combination of *Wx-9*, *SSG6-5* and *OsSSIIa-1*. The haplotype combination of *Wx-9*, *SSG6-5* and *OsSSIIa-1* was also the predominant in SER, while most glutinous accession of EAR carried the haplotype combinations of *Wx-10*, *SSG6-7* and *OsSSIIa-1* or *Wx-10*, *SSG6-5* and *OsSSIIa-1* (Fig.5B),

indicating haplotypes combining more glutinous alleles formation more glutinous rice and glutinous alleles were conducive to the formation of glutinous rice in SEA, SER and EAR.

Population structure and genetic differentiation of three key genes between both endosperm types

The sequence alignment of three key waxy genes and geographical distribution of their different haplotype combinations suggested that the genetic differences underlying waxiness trait among regions was greater than that between subpopulations in rice. To confirm the above hypothesis, we investigated the population structure and admixture patterns of each gene in the whole rice panel. We first estimated ancestry proportions of *Wx*, *SSG6* and *OsSSIIa* for individuals by Admixture. Population structure based on each of three genes showed different genetic structures from the whole genome. Admixture model using 202 SNPs within *Wx* gene indicated that 53 of 70 glutinous *indica* accessions clustered with glutinous *japonica* accessions (Fig.6A). Meanwhile, admixture model using 123 SNPs within *SSG6* gene indicated that 42 of 70 glutinous *indica* accessions clustered with glutinous *japonica* accessions, and one glutinous *japonica* accession clustered with other 28 glutinous *indica* accessions (Fig.6B). Additionally, admixture model using 194 SNPs within *OsSSIIa* gene showed that 66 of 70 glutinous *indica* accessions clustered with glutinous *japonica* accessions, and one glutinous *japonica* accession clustered with other 4 glutinous *indica* accessions (Fig.6C). The results confirmed that there was no obvious genetic differentiation of the three key waxy genes between *japonica* and *indica* distributed in SEA, SER and EAR, which was supported by further PC analysis (Fig.6C).

The exceptional genetic similarity among glutinous rice revealed by PC and admixture analyses could be caused by a unique domestication process. The origin of waxy haplotypes of the three genes and how they spread in *japonica* and *indica* rice are two key issues to reveal the formation of glutinous rice. Here, we firstly examined haplotypes of three known genes in wild rice. There were 72, 64 and 52 haplotypes in *Wx*, *SSG6* and *SSIIa* of wild rice. The waxy haplotypes *Wx-9* of *Wx* gene could be detected in 3 wild rice accessions, which were from Thailand and China. The results indicated that the waxy haplotype *Wx-9* could be inherited from wild rice, but it is a very unlikely scenario that all waxy haplotype in both rice subpopulations originate directly from a small amount of wild rice (Fig.7A). Additionally, none of wild rice carried waxy haplotypes *SSG6-5* and *SSG6-7* of *SSG6* and waxy haplotype *OsSSIIa-1* of *OsSSIIa* (Fig.7B and C), suggesting that the waxy haplotypes of *SSG6* and *OsSSIIa* newly generated during rice domestication. Taken together, a more possible hypothesis for the exceptional genetic similarity among glutinous rice is substantial local gene flow of *Wx*, *SSG6* and *SSIIa* between *indica* and *japonica* in SEA, EAR, and EAR.

To further determine the hypothesis of gene flow and examine the direction of gene flow, we performed phylogenetic analyses using all haplotype types of each gene. For *Wx* gene, waxy haplotype *Wx-9* clustered with other *japonica* haplotypes and formed a monophyletic group (Fig.7A). Meanwhile, two waxy haplotypes *SSG6-5* and *SSG6-7* of *SSG6* clustered together with long genetic distance to other haplotypes of cultivated rice (Fig.7B). Additionally, waxy haplotype *SSIIa-1* clustered with *SSIIa-2* and two wild haplotypes (Fig.7C). Phylogenetic trees in cultivated rice indicated that the waxy haplotypes of each

gene were closer to their corresponding *japonica* haplotypes than *indica* haplotypes, such as *Wx-9* closed to *Wx-7/8/10*, *SSG6-5/7* closed to *SSG6-6*, and *OsSSIIa-1* mainly closed to *japonica* as the haplotype *OsSSIIa-2* account for 46.5% of total in *japonica* accessions (Fig.S3). Given that *japonica* was first domesticated from wild rice in southern China, and that *indica* was subsequently developed from crosses between *japonica* and local wild rice (Huang et al., 2012), we suggested that glutinous haplotypes of the three genes in *japonica* rice firstly generated (or were directly inherited from wild rice), and then flowed into *indica* rice in SER, SEA and EAR.

Conclusion

Glutinous rice as a special endosperm type is also consumed as a staple food in East Asian countries by consumers' preference. But the no genetic study on development and evolution of glutinous rice, specifically in this region has been reported. Here, 2108 rice germplasm were used in GWAS to identify the significant loci and candidate genes controlling the development of glutinous rice. Candidate genes were screened in whole rice panel, *indica*, and *japonica* sub-populations with transcriptome analysis. There were 32 genes, including three starch synthesis-related genes *Wx*, *SSG6*, and *OsSSIIa*, detected simultaneously in the whole rice panel and subpopulations, playing important role in determining glutinous rice. Combined haplotype analyses revealed that the waxy combined haplotype of three genes mainly distributed in Southeast Asia (SEA), SEA islands (SER) and East Asia islands (EAR). Phylogenetic tree and population structure analysis for the origin and evolution of three key waxy genes. This study provides valuable information for further gene discovery and understanding the evolution and formation for glutinous rice in SEA, SER and EAR

Methods

Plant material

A total of 2108 cultivated rice varieties were used in the present study, which were obtained from the 3000 Rice Genome Project (3KRGP) (Lenihan et al., 2014; Wang et al., 2018), including 1965 non-glutinous (or non-waxy) and 143 glutinous (or waxy) rice (<http://snp-seek.irri.org/>). Additionally, 446 wild rice accessions from a previous report were used to study evolutionary aspects (Huang et al., 2012).

Population structure and genetic differentiation

In the whole accessions panel, 5,039,852 independent SNPs across the whole genome determined by PLINK (window size 50, step size 50, $R^2 \geq 0.3$) (Purcell et al., 2007) were used to population structure and admixture patterns analysis by Admixture and GAPIT (Peter, 2016; Tang et al., 2016). For the population structure against the three known genes, SNPs located in gene region and 2 kb promoter region were used. Meanwhile, these SNPs were used to construct neighbor-joining tree in accessions panel with and/or without 446 wild rice accessions. Neighbor-joining tree were developed in MEGA version 7 with the bootstrap method and 1000 replicates (Kumar et al., 2016).

Genome-wide association study

Total of 5,039,852 SNPs with missing rates $\leq 50\%$ and minor allele frequencies $\geq 5\%$ obtained from 3000 Rice Genome Project (3KRGP) using an in-house perl script. GWAS were performed using GAPIT under the LMM model (Lippert et al., 2011; Yang et al., 2014). Here, the top three PCs were used to estimate population structure. Given that it was too stringent for significant association detection when the threshold was derived from the total number of markers (Li et al., 2012; Yang et al., 2014), the threshold to control the type I error rate was defined at $-\log(p) = 5.6$ after Bonferroni-adjusted correction (Li et al., 2012).

Candidate genes analysis

According to the results of the association analysis, genes with significant loci were screened, including gene region and 2 kb promoter region. To study the expression pattern of each gene, two sets of transcriptome data of rice seeds at two periods (GSE98924 for 7-8 days and GSE132303 for 10-14 days after flowering) were obtained from NCBI (<https://www.ncbi.nlm.nih.gov/guide/genes-expression>), Gene expression with the value of FPKM or RPKM > 10 in at least one set deemed as stable expression gene, then genes with stable expression in rice seeds were selected as candidate genes.

Haplotype analysis

Based on information on coding sequence (CDS) coordinates and the transcript from MSU RGAP 7, we separated non-synonymous SNPs, SNPs at splice site and SNPs in promoter from all SNPs across the 2108 accessions using an in-house Perl script. Non-synonymous SNPs, SNPs at splice site and SNPs in promoter significant associated with rice endosperm type ($-\log(p) > 2$) were used for haplotype analysis.

Abbreviations

GWAS

Genome-wide association study; AC (amylose content); ECQs (seating and cooking qualities); GC (gel consistency) ; GT (gelatinization temperature) AAC (apparent amylose contents); EAR (East Asia Islands); EAS (East Asia); SEA (Southeast Asia); SER (SEA islands); IRRI (Philippines); OCE (Oceania); SAE (South Asia - East); SAC (South Asia - Central); SAW (South Asia - West); WAS (West Asia); IOC (Indian Ocean); EAF (East Africa); WAF (West Africa); NAF (North Africa), SAM (South America); CAM (Central America and Caribbean); NAM (North America); EUR (Europe); NA (No region information)

Declarations

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Availability of supporting data

The endersperm type of 2108 cultivated rice varieties used in our study can be obtained from the 3K-RG dataset : <http://snp-seek.irri.org>. The 3K-RG sequencing data used for our analyses can be obtained via project accession PRJEB6180 from NCBI ([https://www.ncbi.nlm.nih.gov/sra/?term= PRJEB6180](https://www.ncbi.nlm.nih.gov/sra/?term=PRJEB6180)). The transcriptome data of rice seeds at 7-8 days and 10-14 days after flowering were obtained from NCBI (<https://www.ncbi.nlm.nih.gov/gds>), the series accession ID: GSE98924 and GSE132303

Competing Interests

The authors declare that they have no competing interests

Funding

Not applicable.

Authors' Contributions

CJ and YZ contributed for conception and design of the study, and performed most of the experiments; CJ analyzed data and wrote the initial draft; MARR reviewed the study, edited and constructed the final manuscript; YZ supervised and helped in each step of the study. The final manuscript was read and approved by all authors

Acknowledgments

Not applicable.

References

1. Alexander D H, Lange K (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12: 246. doi:10.1186/1471-2105-12-246
2. Bao J S, Sun M, Corke H (2002). Analysis of the genetic behavior of some starch properties in *indica* rice (*Oryza sativa* L.): Thermal properties, gel texture, swelling volume. Theor Appl Genet 104: 408-413. doi:10.1007/s001220100688
3. Cai X, Wang Z, Xing Y, Zhang J, Hong 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.
4. Calingacion M, Laborte A, Nelson A, Resurreccion A, Concepcion J, Daygon V, Mumm R, Reinke R (2014). Diversity of global rice markets and the science required for Consumer-Targeted rice breeding. PLoS ONE 9: e85106. doi:10.1371/journal.pone.0085106
5. Clarkson J D, Watabe T, Gluck S H (1968). Glutinous rice in northern thailand. Reports On Research in Southeast Asia Natural Science 27: 922.

6. Cui Y, Li R, Li G, Zhang F, Zhu T, Zhang Q, Ali J, Li Z (2019). Hybrid breeding of rice via genomic selection. *Plant Biotechnology Journal* 18: 57-67. doi:10.1111/pbi.13170
7. Fitzgerald M, McCouch S, Hall R (2009). Not just a grain of rice: The quest for quality. *Trends in Plant Science* 14: 133-139. doi:10.1016/j.tplants.2008.12.004
8. Fujita N, Yoshida M, Asakura N, Ohdan T, Miyao A, Hirochika H, Nakamura Y (2006). Function and characterization of starch synthase I using mutants in rice. *Plant Physiology* 140: 1070-1084. doi:10.1104/pp.105.071845
9. Gao Z, Zeng D, Cui X, Zhou Y, Yan M, Huang D, Li J, Qian Q (2003). Map-based cloning of the ALK gene, which controls the gelatinization temperature of rice. *Science in China. Series C, Life Sciences* 46: 661. doi:10.1360/03yc0099
10. Hannah L C, James M (2008). The complexities of starch biosynthesis in cereal endosperms. *Curr.Opin.Biotechnol* 19: 160-165.
11. He Y, Han Y, Jiang L, Xu C, Lu J, Xu M (2006). Functional analysis of starch-synthesis genes in determining rice eating and cooking qualities. *Molecular Breeding* 18: 277-290. doi:10.1007/s11032-006-5505-7
12. Huang X, Kurata N, Wei X, Wang Z, Wang A, Zhao Q, Zhao Y, Liu K (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490: 497-501. doi:10.1038/nature11532
13. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics* 42: 961-967. doi:10.1038/ng.695
14. Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics* 44: 32-39. doi:10.1038/ng.1018
15. Isshiki M., Morino K., Nakajima M, Okagaki R, Wessler S, 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*.
16. Jeng T, Wang C, Tseng T, Wu M, Sung J (2009). Nucleotide polymorphisms in the waxy gene of NaN-induced waxy rice mutants. *Journal of Cereal Science* 49: 112-116.
17. Juliano B, Villareal C (1993). Grain quality evaluation of world rices..
18. Kumar S, Stecher G, Tamura K (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* 33: 1870-1874. doi:10.1093/molbev/msw054
19. Lenihan J, Kvist S, Fernández R, Giribet G, Ziegler A (2014). The 3,000 rice genomes project. *Gigascience* 3: 7.
20. Li M, Yeung J, Cherny S, Sham P (2012). Evaluating the effective numbers of independent tests and significant *p*-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human Genetics* 131: 747-756. doi:10.1007/s00439-011-1118-2
21. Lippert C, Listgarten J, Liu Y, Kadie C, Davidson R, Heckerman D (2011). FaST linear mixed models for genome-wide association studies. *Nature Methods* 8: 833-835. doi:10.1038/nmeth.1681

22. James M, And, Denyer K, Myers Alan (2003). Starch synthesis in the cereal endosperm. Current Opinion in Plant Biology.
23. Matsushima R, Maekawa M, Kusano M, Tomita K, Kondo H, Nishimura H, Crofts N, Fujita N (2016). Amyloplast membrane protein SUBSTANDARD STARCH GRAIN6 controls starch grain size in rice endosperm. Plant Physiology 170: 1445-1459. doi:10.1104/pp.15.01811
24. 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 & Applied Genetics 116: 979-989.
25. Nakamura Y (2002). Towards a better understanding of the metabolic system for amylopectin biosynthesis in plants: Rice endosperm as a model tissue. Plant and Cell Physiology 43: 718-725. doi:10.1093/pcp/pcf091
26. Pandey M, Rani N, Madhav M, Sundaram R, Varaprasad G, Sivaranjani A, Bohra A, Kumar G (2012). Different isoforms of starch-synthesizing enzymes controlling amylose and amylopectin content in rice (*Oryza sativa* L.). Biotechnology Advances 30: 1697-1706. doi:10.1016/j.biotechadv.2012.08.011
27. Peter B (2016). Admixture, population structure, and F-Statistics. Genetics 202: 1485-1501. doi:10.1534/genetics.115.183913
28. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, Maller J, Sklar P (2007). PLINK: A tool set for Whole-Genome association and Population-Based linkage analyses. The American Journal of Human Genetics 81: 559-575. doi:10.1086/519795
29. Roy S, Shil P (2020). Assessment of genetic heritability in rice breeding lines based on morphological traits and caryopsis ultrastructure. Scientific Reports 10. doi:10.1038/s41598-020-63976-8
30. Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Ishiyama K, Saito T (2002). A mutant gibberellin-synthesis gene in rice. Nature 416: 701-702. doi:10.1038/416701a
31. Si L, Chen J, Huang X, Gong H, Luo J, Hou Q, Zhou T, Lu T (2016). *OsSPL 13* controls grain size in cultivated rice. Nature Genetics 48: 447-456. doi:10.1038/ng.3518
32. Spielmeyer W, Ellis M H, Chandler P M (2002). Semidwarf (sd-1), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. Proceedings of the National Academy of Sciences - PNAS 99: 9043-9048. doi:10.1073/pnas.132266399
33. Tang Y, Liu X, Wang J, Li M, Wang Q, Tian F, Su Z, Pan Y (2016). GAPIT version 2: An enhanced integrated tool for genomic association and prediction. The Plant Genome 9: e2011-e2015. doi:10.3835/plantgenome2015.11.0120
34. Tian Z, Qian Q, Liu Q, Yan M, Liu X, Yan C, Liu G, Gao Z (2009). Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. Proceedings of the National Academy of Sciences 106: 21760-21765. doi:10.1073/pnas.0912396106
35. Vandeputte G, Vermeylen R, Geeroms J, Delcour J (2003). Rice starches. I. Structural aspects provide insight into crystallinity characteristics and gelatinisation behaviour of granular starch. Journal of Cereal Science 38: 43-52. doi:10.1016/S0733-5210(02)00140-6
36. 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: 1199.

37. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557: 43-49. doi:10.1038/s41586-018-0063-9
38. Xing Y, Zhang Q (2010). Genetic and molecular bases of rice yield. *Annual Review of Plant Biology* 61: 421-442. doi:10.1146/annurev-arplant-042809-112209
39. Xu Y, Lin Q, Li X, Wang F, Chen Z, Wang J, Li W, Fan F (2020). Fine-tuning the amylose content of rice by precise base editing of the Wx gene. *Plant Biotechnology Journal*. doi:10.1111/pbi.13433
40. Yamanaka S, Nakamura I, Watanabe K N, Sato Y (2004). Identification of SNPs in the waxy gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. *Theoretical and Applied Genetics* 108: 1200-1204. doi:10.1007/s00122-003-1564-x
41. Yang J, Wang J, Fan F, Zhu J, Chen T, Wang C, Zheng T, Zhang J (2013). Development of AS-PCR marker based on a key mutation confirmed by resequencing of Wx-mpin Milky Princess and its application in japonica soft rice (*Oryza sativa* L.) breeding. *Plant Breeding*.
42. Yang W, Guo Z, Huang C, Duan L, Chen G, Jiang N, Fang W, Feng H (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nature Communications* 5. doi:10.1038/ncomms6087
43. Yano K, Yamamoto E, Aya K, Takeuchi H, Lo P, Hu L, Yamasaki M, Yoshida S (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nature Genetics* 48: 927-934. doi:10.1038/ng.3596
44. Yasunori N, Perigio B F, Yuko H, Aya S, Takayuki S, Akiko K, Naoko F (2005). Essential amino acids of starch synthase IIa differentiate amylopectin structure and starch quality between *japonica* and *indica* rice varieties. *Plant Molecular Biology* 58. doi:10.1007/sl
45. Yu J, Miao J, Zhang Z, Xiong H, Zhu X, Sun X, Pan Y, Liang Y (2018). Alternative splicing of *OsLG3b* controls grain length and yield in *japonica* rice. *Plant Biotechnology Journal* 16: 1667-1678. doi:10.1111/pbi.12903
46. Zhao Y, Zhang H, Xu J, Jiang C, Yin Z, Xiong H, Xie J, Wang X (2018). Loci and natural alleles underlying robust roots and adaptive domestication of upland ecotype rice in aerobic conditions. *PLOS Genetics* 14: e1007521. doi:10.1371/journal.pgen.1007521

Figures

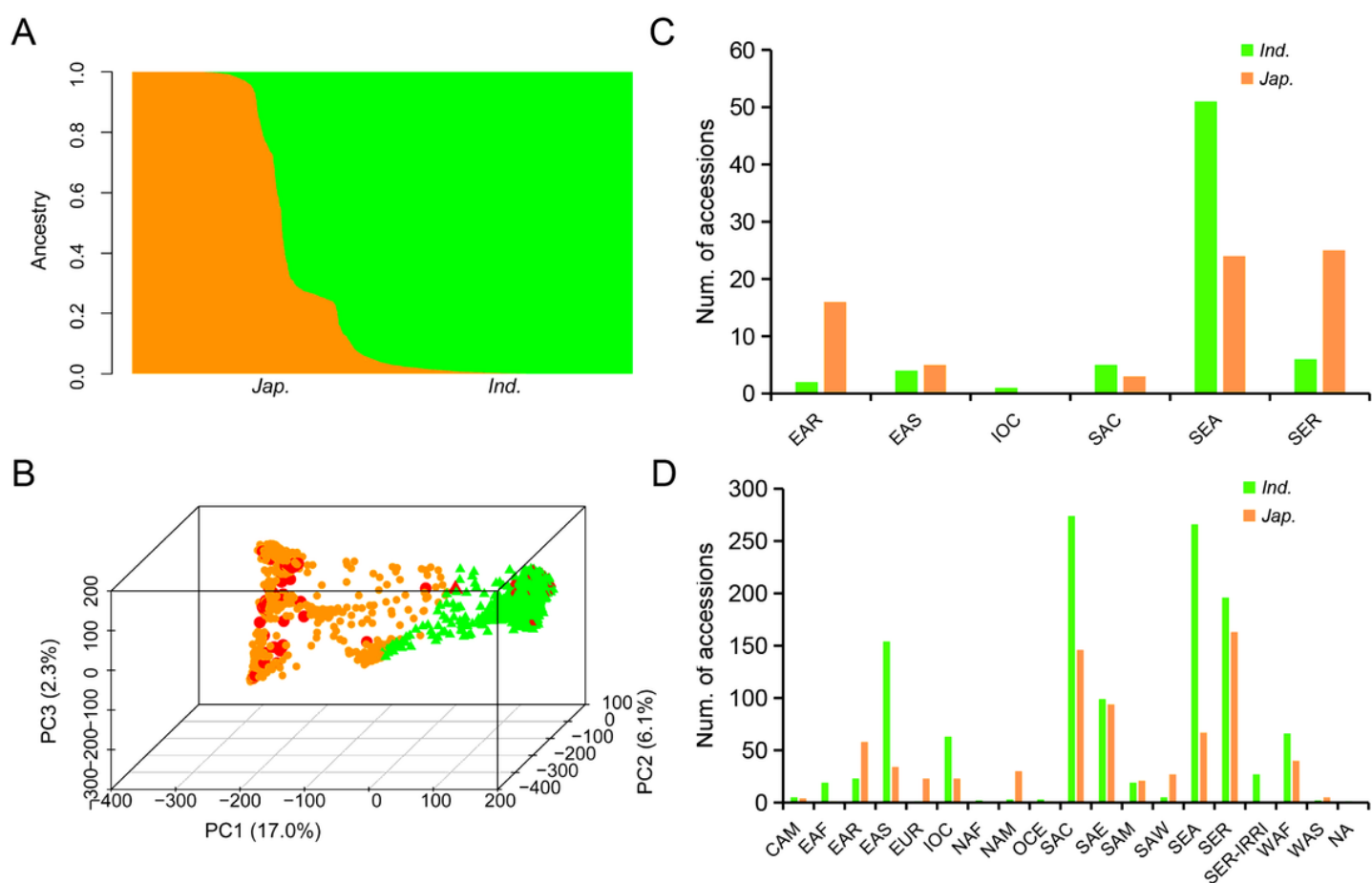


Figure 1

Population structure and geographic distribution of accessions in association panel. (A) Genetic structure of the panel based on ADMIXTURE for $K = 2$. (B) Principal components analysis with first 3 principal components explaining 25.4% of the genetic variation within the panel. Orange and Green dots show japonica and indica, respectively. Red dot indicates glutinous rice accessions. (C) and (D) Geographic distribution of glutinous and non-glutinous rice accessions. Geographical information reference to 3010 rice accession (DOI: <https://doi.org/10.1038/s41586-018-0063-9>), EAR (East Asia Islands) EAS (East Asia), SEA (Southeast Asia), SER (SEA islands), SER-IRRI (SER-Philippines), OCE (Oceania), SAE (South Asia - East), SAC (South Asia - Central), SAW (South Asia - West), WAS (West Asia), IOC (Indian Ocean), EAF (East Africa), WAF (West Africa), NAF (North Africa), SAM (South America), CAM (Central America and Caribbean), NAM (North America), EUR (Europe), NA (No region information)

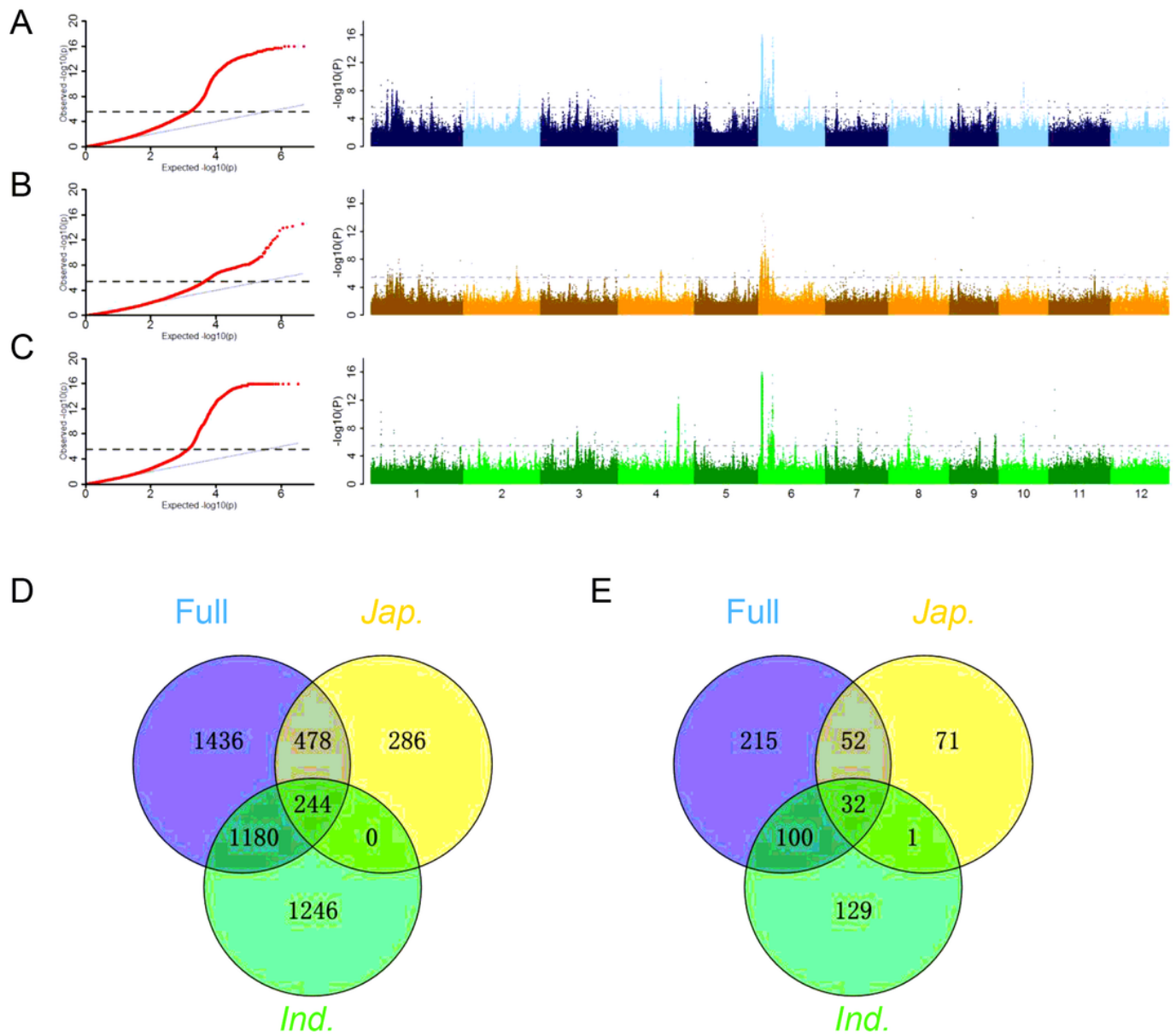


Figure 2

Identification of waxiness QTLs by GWAS. Quantile-quantile (Q-Q) plots and Manhattan plots in the whole panel (A), japonica (B) and indica (C). (D) Venn diagram of significant loci detected in different populations. (E) Venn diagram of genes with significant loci detected in different populations.

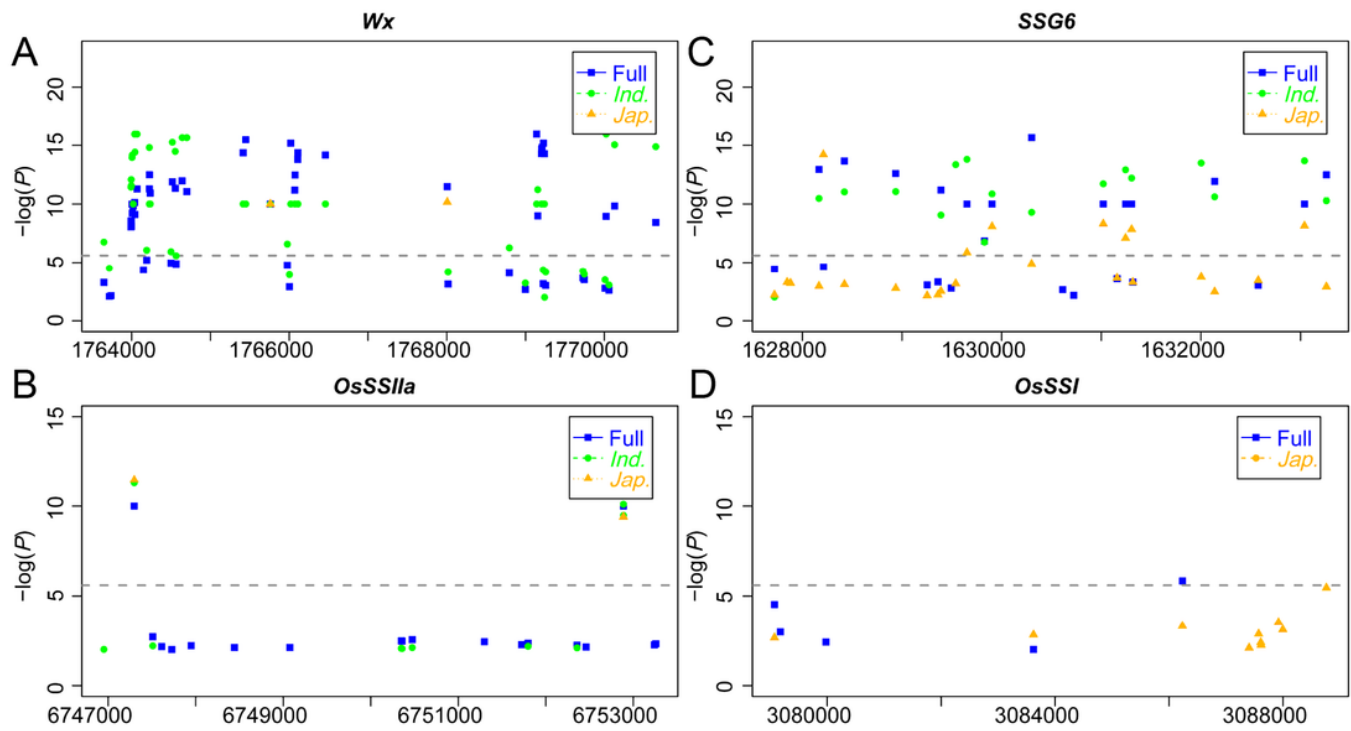


Figure 3

Gene-based association analysis of four starch synthesis-related genes in different population

A	HAP	Promoter and 5'UTR																												Nonsvs		Num of accessions										
		1763647	1763716	1763739	1763992	1763994	1763995	1764004	1764008	1764019	1764038	1764042	1764068	1764149	1764178	1764191	1764224	1764225	1764235	1764499	1764516	1764553	1764564	1764641	1764700	1765415	1765448	1765761	1765976	1766003	1766020	1766071	1766074	1766108	1766110	1766460	1768006	1768998	Ind./ Glu ind	Jap./ Glu jap		
	W _x -1	A	A	A	A	T	T	G	T	C	A	G	T	C	G	A	C	G	T	C	G	C	G	T	A	G	T	G	A	C	G	T	C	T	C	C	A	C	6/0	20/0		
	W _x -2	A	A	A	A	T	T	G	T	C	A	G	T	C	G	C	C	G	T	T	G	C	A	T	A	G	T	G	A	C	G	T	C	T	C	C	A	C	7/0	20/0		
	W _x -3	A	A	A	A	T	T	G	T	C	A	G	T	T	G	A	C	G	T	C	G	C	G	T	A	G	T	G	A	C	G	T	C	T	C	C	A	C	10/1	28/0		
	W _x -4	A	A	A	A	T	T	G	T	C	A	G	T	T	G	C	C	G	T	C	G	C	G	T	A	G	T	G	A	C	G	T	C	T	C	C	A	C	4/0	7/0		
	W _x -5	A	A	A	A	T	T	G	T	C	A	G	T	T	G	C	C	G	T	T	G	C	A	T	A	G	T	G	A	C	G	T	C	T	C	C	A	C	9/0	19/0		
	W _x -6	A	A	A	A	C	G	G	T	A	C	G	A	C	C	G	C	T	A	T	T	T	A	C	G	A	C	G	A	C	A	C	A	C	T	T	C	C	0/0	10/0		
	W _x -7	A	A	A	A	C	G	G	T	A	T	G	A	C	C	A	C	T	A	C	T	T	T	A	C	G	A	C	T	A	C	A	C	A	C	T	T	A	C	5/0	37/0	
	W _x -8	A	A	A	A	C	G	G	T	A	T	G	A	C	C	G	C	T	A	C	T	T	T	A	C	G	A	C	G	A	C	A	C	A	C	T	T	C	C	65/0	343/5	
	W _x -9	A	A	A	A	C	G	G	T	A	T	G	A	C	C	G	C	T	A	C	T	T	T	A	C	G	A	C	T	A	C	A	C	A	C	T	T	A	C	108/37	173/51	
	W _x -10	A	A	A	A	C	G	G	T	A	T	G	A	C	C	G	C	T	A	C	T	T	T	A	C	G	A	C	T	A	C	A	C	A	C	T	T	C	C	0/0	21/7	
	W _x -11	G	A	A	A	C	G	G	T	A	T	G	A	C	C	G	C	T	A	C	T	T	T	A	C	G	A	C	T	A	C	A	C	A	C	T	T	A	C	23/9	0/0	
	W _x -12	G	G	A	A	T	T	G	T	C	A	G	T	T	G	A	C	G	T	C	G	C	G	T	A	G	T	G	A	C	G	T	C	T	C	C	A	C	12/0	1/0		
	W _x -13	G	G	A	A	T	T	G	T	C	A	G	T	T	G	A	C	G	T	C	G	C	G	T	A	G	T	G	A	C	G	T	C	T	C	C	A	C	38/0	1/0		
	W _x -14	G	G	A	A	T	T	G	T	C	A	G	T	T	G	A	C	G	T	C	G	C	G	T	A	G	T	G	A	C	G	T	C	T	C	C	A	T	90/0	0/0		
	W _x -15	G	G	A	A	T	T	G	T	C	A	G	T	T	G	A	C	G	T	C	G	C	G	T	A	G	T	G	A	C	G	T	C	T	C	C	A	C	65/1	4/0		
	W _x -16	G	G	A	A	T	T	G	T	C	A	G	T	T	G	A	C	G	T	C	G	C	G	T	A	G	T	G	A	C	G	T	C	T	C	C	A	C	20/0	0/0		
	W _x -17	G	G	A	A	T	T	G	T	C	A	G	T	T	G	A	C	G	T	C	G	C	G	T	A	G	T	G	A	C	G	T	C	T	C	C	A	T	142/3	1/0		
	W _x -18	G	G	A	A	T	T	G	T	C	A	G	T	T	G	A	C	G	C	G	T	T	G	C	A	T	A	G	T	G	A	C	G	T	C	T	C	C	A	C	9/0	3/0
	W _x -19	G	G	A	A	T	T	G	T	C	A	G	T	T	G	A	C	G	C	C	G	T	T	G	C	A	T	A	G	T	G	A	C	G	T	C	T	C	A	C	13/0	1/0
	W _x -20	G	G	A	A	T	T	G	T	C	A	G	T	T	G	A	C	G	T	C	G	C	G	T	A	A	C	T	A	C	A	C	A	C	A	C	T	T	A	C	11/5	0/0
	W _x -21	G	G	A	A	T	T	G	T	C	A	G	T	T	G	A	C	G	T	C	G	C	G	T	A	G	T	G	A	C	G	T	C	T	C	C	A	C	113/2	13/0		
	W _x -22	G	G	A	A	T	T	G	T	C	A	G	T	T	G	A	C	G	T	C	G	C	G	T	A	G	T	G	A	C	G	T	C	T	C	C	A	C	64/0	1/0		
	W _x -23	G	G	A	A	T	T	G	T	C	A	G	T	T	G	A	C	G	T	C	G	C	G	T	A	G	T	G	A	C	G	T	C	T	C	C	A	T	273/3	4/0		
	W _x -24	G	G	A	A	T	T	G	T	C	A	G	T	T	G	C	C	G	T	T	G	C	A	T	A	G	T	G	A	C	G	T	C	T	C	C	A	C	45/0	1/0		
	W _x -25	G	G	A	A	T	T	G	T	C	A	G	T	T	G	C	C	G	T	T	G	C	A	T	A	G	T	G	A	C	G	T	C	T	C	C	C	C	28/1	6/0		
	W _x -26	G	G	A	A	T	T	G	T	C	A	G	T	T	G	C	C	G	T	T	G	C	A	T	A	G	T	G	A	C	G	T	C	T	C	C	A	C	11/0	4/0		

B

HAP	Promoter											5'UTR	Nonsys		3'UTR	Num of accessions			
	1627720	1627847	1627881	1628167	1628211	1628422	1628937	1629252	1629361	1629391	1629494	1629652	1629828	1629904	1630301	1633040	1633260	<i>Ind./Glu ind</i>	<i>Jap./Glu jap</i>
<i>SSG6-1</i>	A	C	T	C	T	C	T	G	T	A	T	A	T	G	C	G	C	1/0	82/0
<i>SSG6-2</i>	A	T	G	C	A	C	T	G	T	A	T	A	T	G	C	G	C	197/6	20/0
<i>SSG6-3</i>	A	T	G	C	T	C	T	G	T	A	T	A	T	G	C	G	C	563/18	91/0
<i>SSG6-4</i>	A	T	T	C	A	C	T	G	T	A	T	A	T	G	C	G	C	13/0	0/0
<i>SSG6-5</i>	G	T	G	A	A	G	C	G	T	T	T	T	G	A	T	A	T	3/0	117/47
<i>SSG6-6</i>	G	T	G	A	T	G	C	A	A	T	A	A	G	G	C	G	T	10/0	154/5
<i>SSG6-7</i>	G	T	G	A	T	G	C	G	T	T	T	T	G	A	T	A	T	131/41	236/20
<i>SSG6-8</i>	G	T	G	C	A	C	T	G	T	A	T	A	T	G	C	G	C	289/4	11/0
<i>SSG6-9</i>	G	T	G	C	T	C	T	G	T	A	T	A	T	G	C	G	C	57/0	75/1

C

HAP	Promoter												5'UTR	Nonsys			3'UTR	Num of accessions	
	1627720	1627847	1627881	1628167	1628211	1628422	1628937	1629252	1629361	1629391	1629494	1629652	1629828	1629904	1630301	1633040	1633260	<i>Ind./Glu ind</i>	<i>Jap./Glu jap</i>
<i>SSG6-1</i>	A	C	T	C	T	C	T	G	T	A	T	A	T	G	C	G	C	1/0	82/0
<i>SSG6-2</i>	A	T	G	C	A	C	T	G	T	A	T	A	T	G	C	G	C	197/6	20/0
<i>SSG6-3</i>	A	T	G	C	T	C	T	G	T	A	T	A	T	G	C	G	C	563/18	91/0
<i>SSG6-4</i>	A	T	T	C	A	C	T	G	T	A	T	A	T	G	C	G	C	13/0	0/0
<i>SSG6-5</i>	G	T	G	A	A	G	C	G	T	T	T	T	G	A	T	A	T	3/0	117/47
<i>SSG6-6</i>	G	T	G	A	T	G	C	A	A	T	A	A	G	G	C	G	T	10/0	154/5
<i>SSG6-7</i>	G	T	G	A	T	G	C	G	T	T	T	T	G	A	T	A	T	131/41	236/20
<i>SSG6-8</i>	G	T	G	C	A	C	T	G	T	A	T	A	T	G	C	G	C	289/4	11/0
<i>SSG6-9</i>	G	T	G	C	T	C	T	G	T	A	T	A	T	G	C	G	C	57/0	75/1

Figure 4

Haplotype analyses for three key cloned genes conferring waxiness in rice. Haplotypes filled in pink-color are waxy haplotypes. Number in the two columns on right, represents the number of total accessions and glutinous accessions in indica and japonica, respectively. SNP in red represents the splice-site of Wx

A

Group	Genotype			Num. of glutinous accessions	
				<i>Ind.</i>	<i>Jap.</i>
I	<i>Wx-9</i>	<i>SSG6-7</i>	<i>OsSSIIa-1</i>	27	14
II	<i>Wx-9</i>	<i>SSG6-5</i>	<i>OsSSIIa-1</i>	0	25
III	<i>Wx-10</i>	<i>SSG6-7</i>	<i>OsSSIIa-1</i>	0	4
IV	<i>Wx-11</i>	<i>SSG6-7</i>	<i>OsSSIIa-1</i>	9	0
V	<i>Wx-9</i>	<i>SSG6-5</i>	<i>OsSSIIa-2</i>	0	9
VI	<i>Wx-10</i>	<i>SSG6-5</i>	<i>OsSSIIa-1</i>	0	3
VII	<i>Wx-20</i>	<i>SSG6-3</i>	<i>OsSSIIa-1</i>	4	0
VIII	<i>Wx-8</i>	<i>SSG6-5</i>	<i>OsSSIIa-2</i>	0	4
IX	<i>Wx-9</i>	<i>SSG6-7</i>	<i>OsSSIIa-2</i>	2	2
X	<i>Wx-9</i>	<i>SSG6-8</i>	<i>OsSSIIa-1</i>	3	0

B

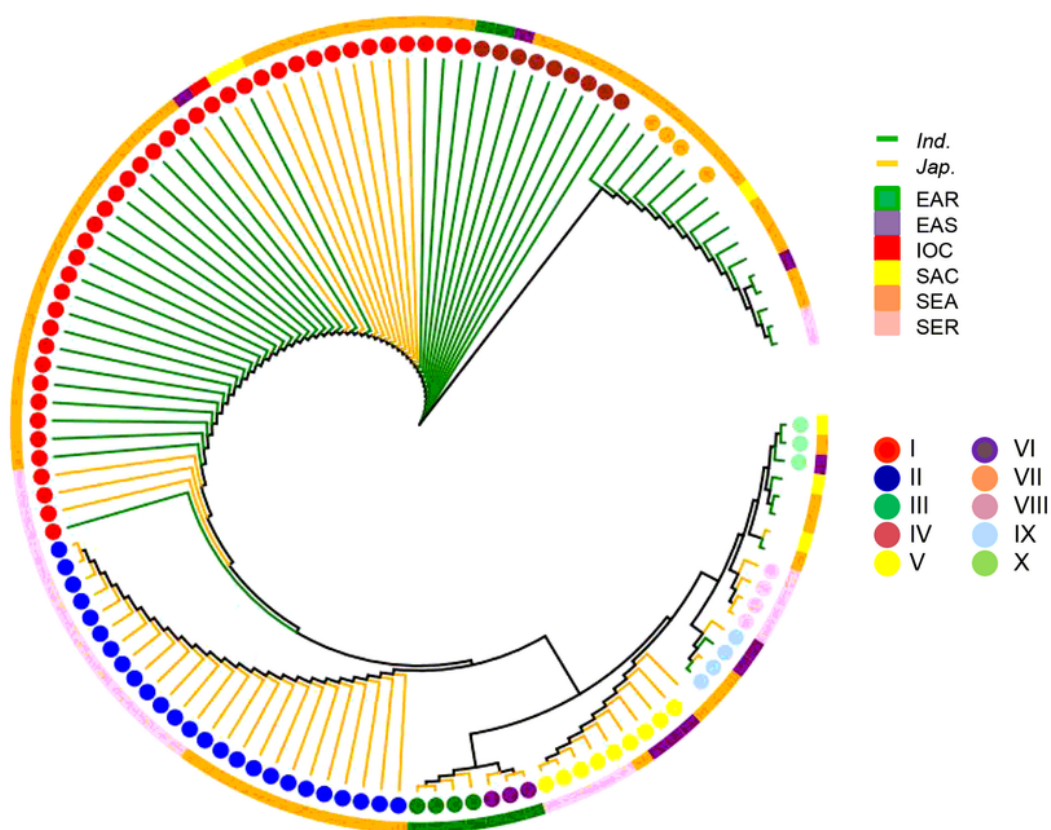


Figure 5

Combined haplotypes and phylogenetic tree of haplotype combinations based on the haplotype of Wx, SSG6 and OsSSIIa. (A) Combined haplotypes of Wx, SSG6 and OsSSIIa in 124 glutinous accessions. The haplotype combinations with more than 3 accessions are listed. Haplotypes filled in color are waxy haplotypes. (B) The phylogenetic tree based on combined haplotypes of Wx, SSG6 and OsSSIIa in 124

glutinous accessions. Colored lines represent subpopulations. Colored circles represent different combined haplotypes. Outermost colored squares represent different geographical regions.

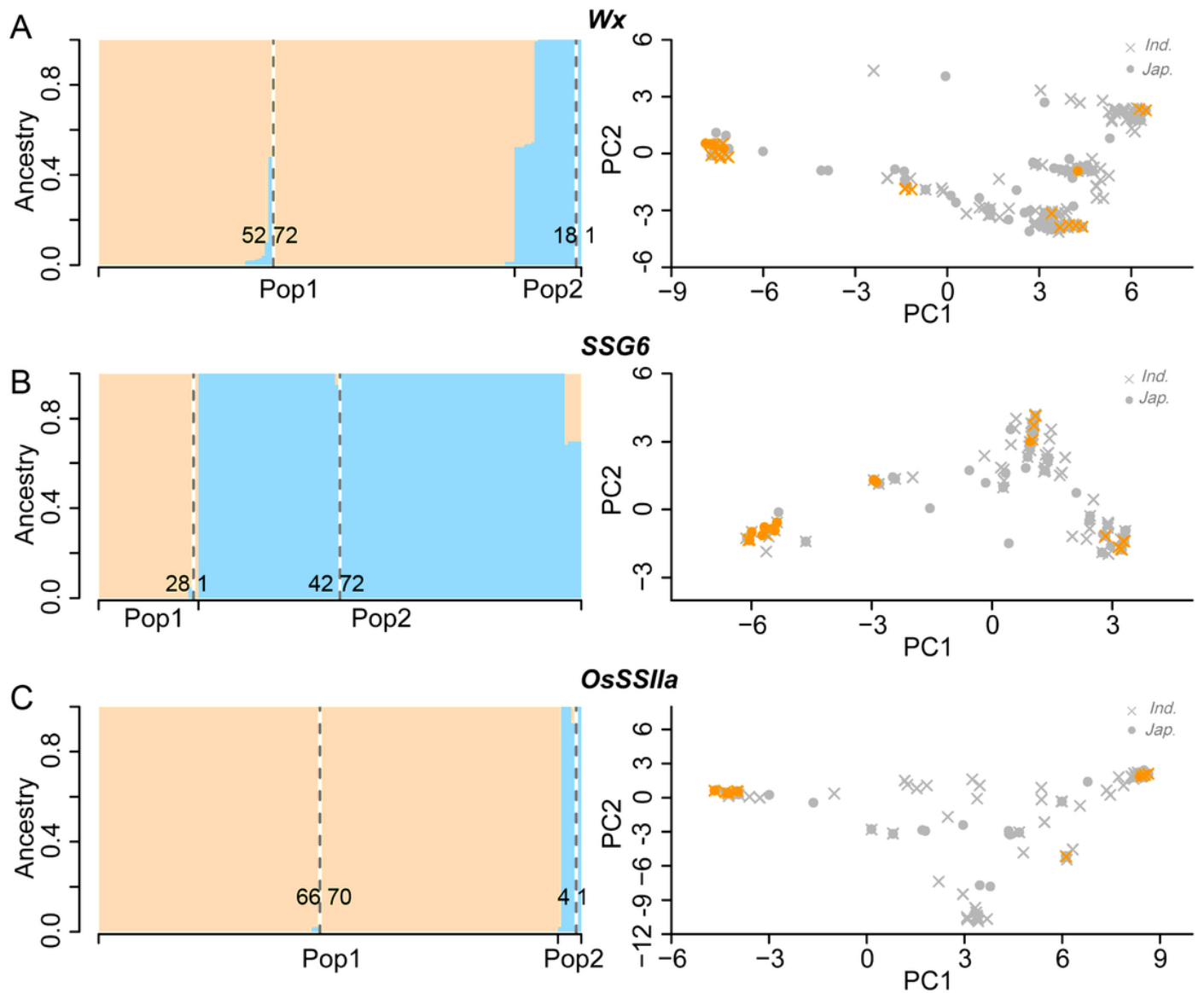


Figure 6

Admixture patterns and principal component analysis for three key genes for waxiness based on SNPs within gene region. Navajowhite and skyblue represent Panel1 and Panel2 analyzed by the Admixture software. Here only 143 glutinous rice accessions were shown. The numbers on the left and right of lines in Pop1 or Pop2 represent the number of glutinous indica and japonica, respectively. For PC analysis, grey and orange represent non-glutinous and glutinous rice

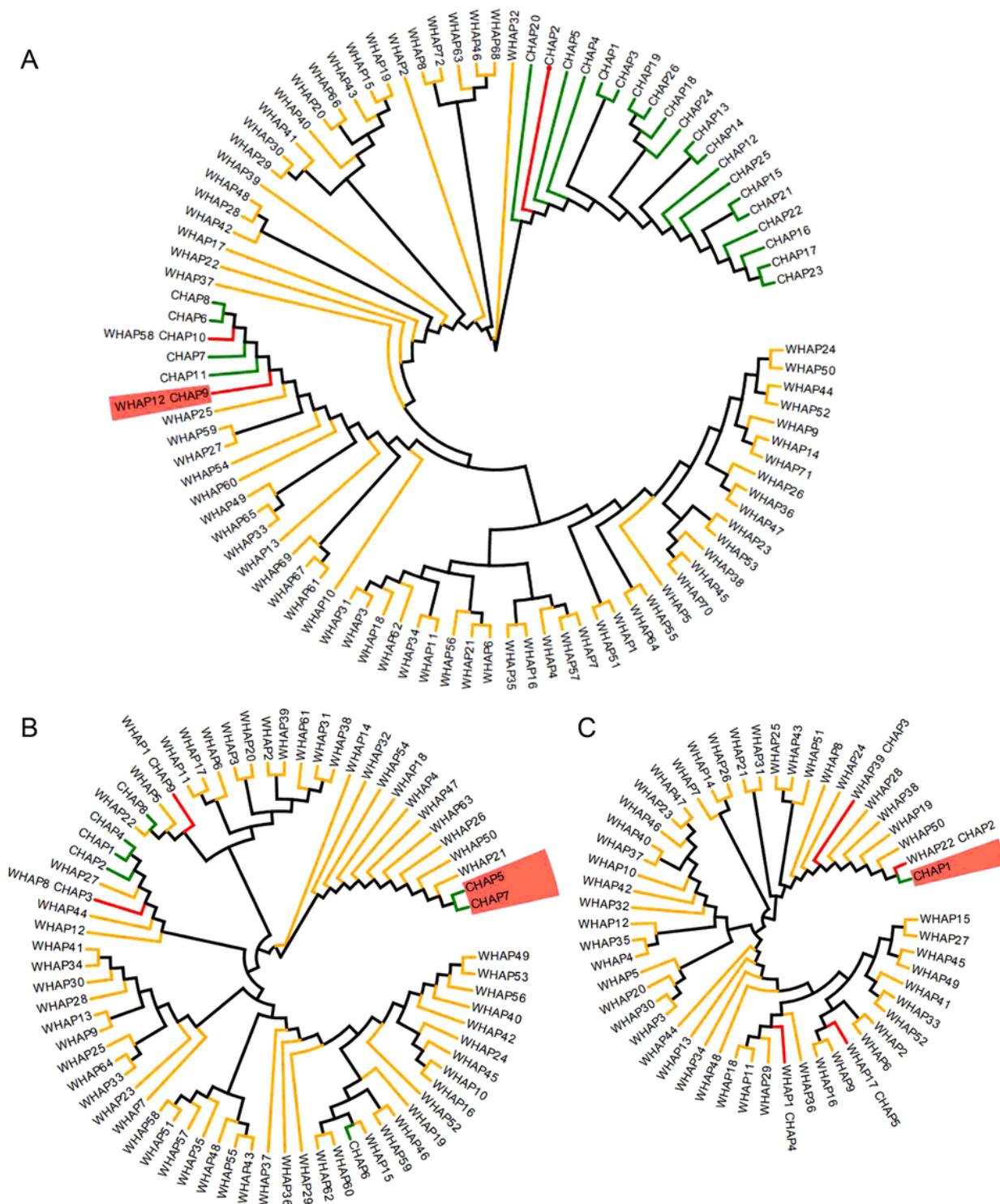


Figure 7

Phylogenetic tree of three keys gene based on haplotypes on cultivated rice and wild rice. (A), (B) and (C) represent Wx, SSG6 and OsSSIIa. The prefix W stands for wild rice and the prefix C stands for cultivated rice. Green, orange and red lines represent haplotypes in cultivated, wild rice and shared between them, red Taxa name with red background represents waxy haplotypes of each gene.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryinformation.docx](#)
- [TableS5.xlsx](#)
- [TableS4.xlsx](#)
- [TableS3.xlsx](#)
- [TableS2.xlsx](#)
- [FigureS3.tif](#)
- [FigureS2.tif](#)
- [FigureS1.tif](#)
- [TableS1.xlsx](#)