

# Genome Wide Association Study on Development and Evolution of Glutinous Rice

**Conghui Jiang**

Shandong Rice Research Institute, Shandong Academy of Agricultural Sciences

**Muhammad Abdul Rehman Rashid**

Government College University

**Yanhong Zhang**

Xinjiang Academy of Agricultural Sciences

**Yan Zhao**

Shandong Agricultural University

**Yinghua Pan** (✉ [panyinghua2008@163.com](mailto:panyinghua2008@163.com))

Guangxi Academy of Agricultural Sciences

---

## Research Article

**Keywords:** Glutinous rice, Genome-wide Association Study, Oryza Sativa L., Haplotype analysis, Evolution

**Posted Date:** July 6th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-649622/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** Glutinous rice as a special endosperm type is 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. Therefore, we sought to explore more genes related to glutinous by genome wide association study and research the formation history for glutinous .

**Results:** Here, genome-wide association study was performed to explore the associated loci/genes underlying glutinous rice by using 2108 rice accessions. Combining the expression patterns analysis, 127, 81, and 48 candidate genes were identified to be associated with endosperm type in whole rice panel, *indica*, and *japonica* sub-populations. 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 combined haplotype analyses revealed that the waxy haplotypes combination of three genes mainly distributed in Southeast Asia (SEA), SEA islands (SER) and East Asia islands (EAR). Through population structure and genetic differentiation, we suggest that waxy haplotypes of the three genes firstly evolved or were directly inherited from wild rice in *japonica*, and then infiltrated into *indica* in SER, SEA and EAR.

**Conclusions:** The cloning and natural variation analysis of waxy-related genes are of great significance for the genetic improvement of quality breeding and comprehend the history in glutinous rice. This work provides valuable information for further gene discovery and understanding the evolution and formation for glutinous rice in SEA, SER and EAR.

## Background

Rice (*Oryza Sativa* L.) is one of the most important crop grain feeding more than half the world's population [1]. High yield and good quality are two important goals of rice production [2]. Since the green revolution, new farming methods and breeding techniques have greatly increased food production in many countries [3–6], 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 [7, 8]. 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" [9–11]. 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 [12]. It involves 18 starch synthase enzymes related genes, and each gene plays a different role in various stages of starch synthesis [13–16]. 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 [17, 18]. 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<sup>b</sup>* to create a series of mutants with AC of 1.4–11.9% [19]. 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 [20]. *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 [20]. Interaction of multiple pairs of genes has a significant effect on rice apparent amylose contents (AAC) [21]. 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 [17, 18]. 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 [22–26]. The sequencing of 3K core germplasm rice provides strong guarantee for the discovery of waxy genes [27].

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 [27]. 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 [28], 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 [27], 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 consistence with previous research that reported the artificial selection of glutinous rice in Southeast Asia [11, 29]. 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-specific 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 have 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 *OsSSIIa* (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 *OsSSIIa* (*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 [30-34]. The other three genes affect the morphology and amylose content of starch [35-37]. 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

The 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 *OsSSIIa*) 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 [25, 26, 38]. A total of 37 significant SNPs were identified within *Wx* gene, including two non-synonymous SNP, one SNP at split site, and 34 in promote 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* [30]. The mutant of fifth exons of *Wx<sup>b</sup>* 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 in 124 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 [29], we suggested that glutinous haplotypes of the three genes in *japonica* rice firstly evolved or were directly inherited from wild rice, and then flowed into *indica* rice in SER, SEA and EAR.

## Discussion

### Waxiness is a complex polygenic trait

The endosperm type is an important characteristic of rice quality. Glutinous rice is one of the traditional grains loved by mankind. There are various reports of research on waxiness in the past decades, but only *Wx* was cloned. In order to analyze the genetic composition more accurately, it is necessary to identify more genes that regulate this trait. Through genome wide association analysis in this study, a large number of waxiness-related loci were identified, indicating that the waxiness trait is a complex quantitative trait controlled by multiple genes. The *Wx*, *SSG6* and *OsSSIIa* as the starch synthesis controlling genes were identified in full population and subpopulations with strong correlation signals in GWAS result, are proved as key genes controlling waxiness. Sequence analysis showed several alleles of *Wx* for AC [30-34, 39]. Previous studies showed that the base substitutions in coding sequence of *ALK* may cause the alteration in gelatinization temperature [35-37], and the mutation in *ssg6* introduced a premature stop codon developed enlarged starch grains in endosperm [37]. In this study, natural variations for waxiness in three key genes were investigated on the basis of 2108 rice accessions. The 26, 9, and 5 haplotypes of *Wx*, *SSG6*, and *OsSSIIa* were found, respectively. Further analysis showed that *Wx-9*, *SSG6-5/7* and *OsSSIIa-1* were considered as the waxy haplotypes. The research provides important breeding variations for glutinous rice breeding.

### Three key genes cooperate in evolution of glutinous rice

In previous studies, *Wx* was observed as the key gene for waxiness. Other widely conducted evolutionary studies and the manipulation of *Wx* in rice breeding [33, 39] indicated that the *Wx* cooperated with others starch synthesis gene to form a fine regulating network that controls the eating and cooking quality [20]. In our research, through the combination of haplotypes, the interaction between the three genes (*Wx*, *SSG6*, *OsSSIIa*) was studied, which provided the reference for molecular pyramid breeding. The combined haplotype analysis of *Wx*, *SSG6* and *OsSSIIa* showed *Wx-9* / *SSG6-5(7)* and *OsSSIIa-1* genotypes may be the best allele combination for waxy rice and quality breeding, because the combination of these three alleles accounted for the largest proportion in the main distribution area of waxy materials. This study revealed the potential gene combination types of glutinous rice that are popular in Southeast Asia. Population structure and genetic differentiation of three waxiness-related genes showed 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. Although the mechanism of these genes to control waxiness is still not clear, elucidating their molecular characteristics and evolutionary patterns in rice germplasm will help to promote glutinous rice breeding. These cloned genes and ongoing gene cloning work will provide a comprehensive understanding of mechanism behind the waxiness, which can then be applied to design varieties of the desired quality

## Conclusions

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. But 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 *OsSSIa*, detected simultaneously in the whole rice panel and subpopulations, playing important role in determining glutinous rice. Combined haplotype analysis revealed that the waxy combined haplotype of three genes mainly distributed in Southeast Asia (SEA), SEA islands (SER) and East Asia islands (EAR). 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) [27, 40], 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 [29].

### 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,  $R^2 \geq 0.3$ ) [41] were used to population structure and admixture patterns analysis by Admixture and GAPIT [42, 43]. 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 [44].

## 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 [45, 46]. Here, the top three principal components (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 [46, 47], the threshold to control the type I error rate was defined at  $-\log(p) = 5.6$  after Bonferroni-adjusted correction [47].

## Candidate genes expression 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 data and Material

The endosperm 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>). 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

Authors declare that they have no competing interests

## Funding

The present study was supported by the Project of Shandong Natural Science Foundation (Grant No. ZR201910280138), the National Natural Science Foundation of China (Grant Nos. NSFC32060452, NSFC32060454, U20A2032), the key project of Guangxi Natural Science Foundation (Grant Nos. 2018GXNSFDA281053, 2019GXNSFBA245006), the project of the central government guides the development of local science and technology (GuikeZY20198015, GuikeZY19183020, GuikeZY18057004), the project of the base and talent (Guike AD17129064 and Guike AD20297093, GuiKe2021JM50, GuiKe AD18281069, 2018-15-Z06-KF16), the Base Business Project of Guangxi Academy of Agricultural Sciences (Grant Nos. 2018 QN21), and the Project of Guangxi Rice High-quality Breeding Research (Grant No. Talent small highland for rice 160-380-16-2), Special Cultivation Program for Scientific and Technological Innovation of Xinjiang Academy of Agricultural Sciences (xjkcpy-005) and the funding bodies had no role in the design of the study; collection, analysis, and interpretation of data; and in writing the manuscript.

## Authors' Contributions

C.J. and Y.Z. contributed for conception and design of the study, and performed most of the experiments; C.J. analyzed data and wrote the initial draft; M.A.R.R. and Y.Zhang reviewed the study, edited and

constructed the final manuscript; Y.Z. and P.Y.H. supervised and helped in each step of the study. The final manuscript was read and approved by all authors.

## Acknowledgments

Not applicable.

## References

1. Roy SC, Shil P: **Assessment of genetic heritability in rice breeding lines based on morphological traits and caryopsis ultrastructure.** *Scientific reports* 2020, **10**(1):1–17.
2. Fitzgerald MA, McCouch SR, Hall RD: **Not just a grain of rice: the quest for quality.** *Trends in plant science* 2009, **14**(3):133–139.
3. Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Ishiyama K, Saito T, Kobayashi M, Khush GS: **A mutant gibberellin-synthesis gene in rice.** *Nature* 2002, **416**(6882):701–702.
4. Spielmeier W, Ellis MH, Chandler PM: **Semidwarf (sd-1), “green revolution” rice, contains a defective gibberellin 20-oxidase gene.** *Proceedings of the National Academy of Sciences* 2002, **99**(13):9043–9048.
5. Xing Y, Zhang Q: **Genetic and molecular bases of rice yield.** *Annual review of plant biology* 2010, **61**:421–442.
6. Cui Y, Li R, Li G, Zhang F, Zhu T, Zhang Q, Ali J, Li Z, Xu S: **Hybrid breeding of rice via genomic selection.** *Plant biotechnology journal* 2020, **18**(1):57–67.
7. Nakamura Y, Francisco PB, Hosaka Y, Sato A, Sawada T, Kubo A, Fujita N: **Essential amino acids of starch synthase IIa differentiate amylopectin structure and starch quality between japonica and indica rice varieties.** *Plant molecular biology* 2005, **58**(2):213–227.
8. Pandey MK, Rani NS, Madhav MS, Sundaram R, Varaprasad G, Sivaranjani A, Bohra A, Kumar GR, Kumar A: **Different isoforms of starch-synthesizing enzymes controlling amylose and amylopectin content in rice (*Oryza sativa* L.).** *Biotechnology advances* 2012, **30**(6):1697–1706.
9. Clarkson JD: **Glutinous Rice in Northern Thailand.** By Tadayo Watabe. Translated from the Japanese by Sumi Hiramoto Gluck. Reports on Research in Southeast Asia, Natural Science Series, N-2. The Center for Southeast Asian Studies, Kyoto University, 1967. xv, 160 pp. Illustrations, Maps, Bibliography, Notes, np. *The Journal of Asian Studies* 1968, **27**(4):922–922.
10. Juliano BO, Villareal C: **Grain quality evaluation of world rices:** Int. Rice Res. Inst.; 1993.
11. Calingacion M, Laborte A, Nelson A, Resurreccion A, Concepcion JC, Daygon VD, Mumm R, Reinke R, Dipti S, Bassinello PZ: **Diversity of global rice markets and the science required for consumer-targeted rice breeding.** *PLoS one* 2014, **9**(1):e85106.

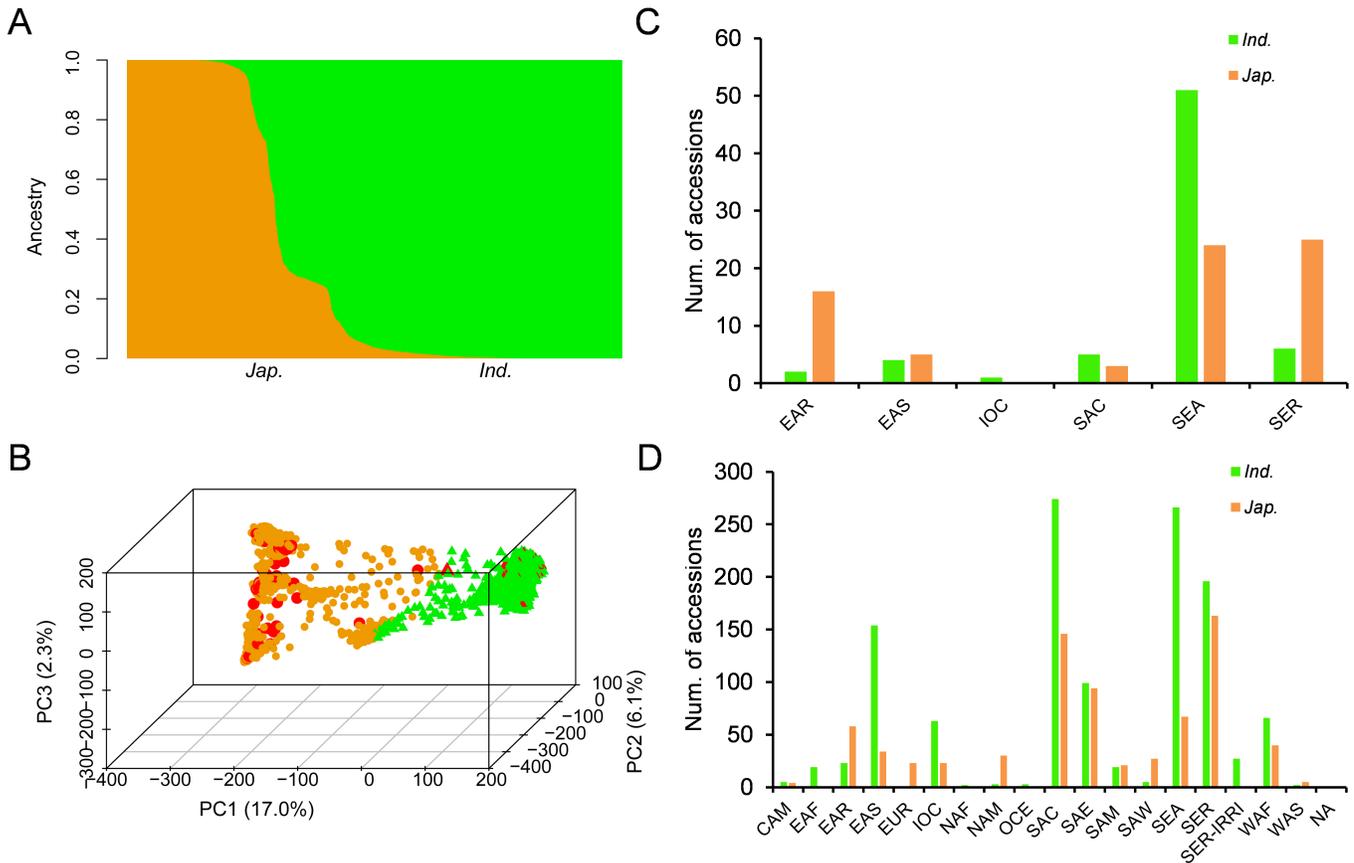
12. Bao J, Sun M, Corke H: **Analysis of the genetic behavior of some starch properties in indica rice (*Oryza sativa* L.): thermal properties, gel texture, swelling volume.** *Theoretical Applied Genetics* 2002, **104**(2):408–413.
13. Nakamura Y: **Towards a better understanding of the metabolic system for amylopectin biosynthesis in plants: rice endosperm as a model tissue.** *Plant Cell Physiology* 2002, **43**(7):718–725.
14. James MG, Denyer K, Myers AM: **Starch synthesis in the cereal endosperm.** *Current opinion in plant biology* 2003, **6**(3):215–222.
15. Vandeputte G, Vermeylen R, Geeroms J, Delcour J: **Rice starches. I. Structural aspects provide insight into crystallinity characteristics and gelatinisation behaviour of granular starch.** *Journal of Cereal Science* 2003, **38**(1):43–52.
16. Hannah LC, James M: **The complexities of starch biosynthesis in cereal endosperms.** *Current Opinion in Biotechnology* 2008, **19**(2):160–165.
17. Yamanaka S, Nakamura I, Watanabe KN, Sato Y-I: **Identification of SNPs in the waxy gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice.** *Theoretical Applied Genetics* 2004, **108**(7):1200–1204.
18. Jeng TL, Wang CS, Tseng TH, Wu MT, Sung JM: **Nucleotide polymorphisms in the waxy gene of NaN3-induced waxy rice mutants.** *Journal of Cereal Science* 2009, **49**(1):112–116.
19. Xu Y, Lin Q, Li X, Wang F, Chen Z, Wang J, Li W, Fan F, Tao Y, Jiang Y: **Fine-tuning the amylose content of rice by precise base editing of the Wx gene.** *Plant biotechnology journal* 2021, **19**(1):11.
20. Tian Z, Qian Q, Liu Q, Yan M, Liu X, Yan C, Liu G, Gao Z, Tang S, Zeng D: **Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities.** *Proceedings of the National Academy of Sciences* 2009, **106**(51):21760–21765.
21. He Y, Han Y, Jiang L, Xu C, Lu J, Xu M: **Functional analysis of starch-synthesis genes in determining rice eating and cooking qualities.** *Molecular Breeding* 2006, **18**(4):277–290.
22. Huang X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M: **Genome-wide association studies of 14 agronomic traits in rice landraces.** *Nature genetics* 2010, **42**(11):961.
23. Huang X, Zhao Y, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, Fan D: **Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm.** *Nature genetics* 2012, **44**(1):32–39.
24. Si L, Chen J, Huang X, Gong H, Luo J, Hou Q, Zhou T, Lu T, Zhu J, Shangguan Y: **OsSPL13 controls grain size in cultivated rice.** *Nature genetics* 2016, **48**(4):447–456.
25. Yano K, Yamamoto E, Aya K, Takeuchi H, Lo P-c, Hu L, Yamasaki M, Yoshida S, Kitano H, Hirano K *et al.*: **Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice.** *Nature Genetics* 2016, **48**(8):927–934.
26. Zhao Y, Zhang H, Xu J, Jiang C, Yin Z, Xiong H, Xie J, Wang X, Zhu X, Li Y *et al.*: **Loci and natural alleles underlying robust roots and adaptive domestication of upland ecotype rice in aerobic conditions.** *PLOS Genetics* 2018, **14**(8):e1007521.

27. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F: **Genomic variation in 3,010 diverse accessions of Asian cultivated rice.** *Nature* 2018, **557**(7703):43–49.
28. Alexander DH, Lange K: **Enhancements to the ADMIXTURE algorithm for individual ancestry estimation.** *BMC bioinformatics* 2011, **12**(1):1–6.
29. Huang X, Kurata N, Wang Z-X, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, Guo Y: **A map of rice genome variation reveals the origin of cultivated rice.** *Nature* 2012, **490**(7421):497–501.
30. Cai XL, Wang ZY, Xing YY, Zhang JL, Hong MM: **Aberrant splicing of intron 1 leads to the heterogeneous 5' UTR and decreased expression of waxy gene in rice cultivars of intermediate amylose content.** *The Plant Journal* 1998, **14**(4):459–465.
31. Isshiki M, Morino K, Nakajima M, Okagaki RJ, Wessler SR, Izawa T, Shimamoto K: **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.** *The Plant Journal* 1998, **15**(1):133–138.
32. Wanchana S, Toojinda T, Tragoonrung S, Vanavichit A: **Duplicated coding sequence in the waxy allele of tropical glutinous rice (*Oryza sativa* L.).** *Plant Science* 2003, **165**(6):1193–1199.
33. Mikami I, Uwatoko N, Ikeda Y, Yamaguchi J, Hirano H-Y, Suzuki Y, Sano Y: **Allelic diversification at the wx locus in landraces of Asian rice.** *Theoretical Applied Genetics* 2008, **116**(7):979–989.
34. Yang J, Wang J, Fan FJ, Zhu JY, Chen T, Wang CL, Zheng TQ, Zhang J, Zhong WG, Xu JL: **Development of AS-PCR marker based on a key mutation confirmed by resequencing of W x-mp in Milky P rincess and its application in japonica soft rice (*O ryza sativa* L.) breeding.** *Plant Breeding* 2013, **132**(6):595–603.
35. Gao Z, Zeng D, Cui X, Zhou Y, Yan M, Huang D, Li J, Qian Q: **Map-based cloning of the ALK gene, which controls the gelatinization temperature of rice.** *Science in China Series C: Life Sciences* 2003, **46**(6):661–668.
36. Fujita N, Yoshida M, Asakura N, Ohdan T, Miyao A, Hirochika H, Nakamura Y: **Function and characterization of starch synthase I using mutants in rice.** *Plant physiology* 2006, **140**(3):1070–1084.
37. Matsushima R, Maekawa M, Kusano M, Tomita K, Kondo H, Nishimura H, Crofts N, Fujita N, Sakamoto W: **Amyloplast membrane protein SUBSTANDARD STARCH GRAIN6 controls starch grain size in rice endosperm.** *Plant physiology* 2016, **170**(3):1445–1459.
38. Yu J, Miao J, Zhang Z, Xiong H, Zhu X, Sun X, Pan Y, Liang Y, Zhang Q, Muhammad Abdul Rehman R: **Alternative splicing of Os LG 3b controls grain length and yield in japonica rice.** *Plant biotechnology journal* 2018, **16**(9):1667–1678.
39. Zhang C, Zhu J, Chen S, Fan X, Li Q, Lu Y, Wang M, Yu H, Yi C, Tang S: **Wxlv, the ancestral allele of rice Waxy gene.** *Molecular plant* 2019, **12**(8):1157–1166.
40. Li Z, Fu B-Y, Gao Y-M, Wang W-S, Xu J-L, Zhang F, Zhao X-Q, Zheng T-Q, Zhou Y-L, Zhang G *et al.*: **The 3,000 rice genomes project.** *GigaScience* 2014, **3**:7.
41. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ: **PLINK: a tool set for whole-genome association and population-based linkage analyses.**

*The American journal of human genetics* 2007, **81**(3):559–575.

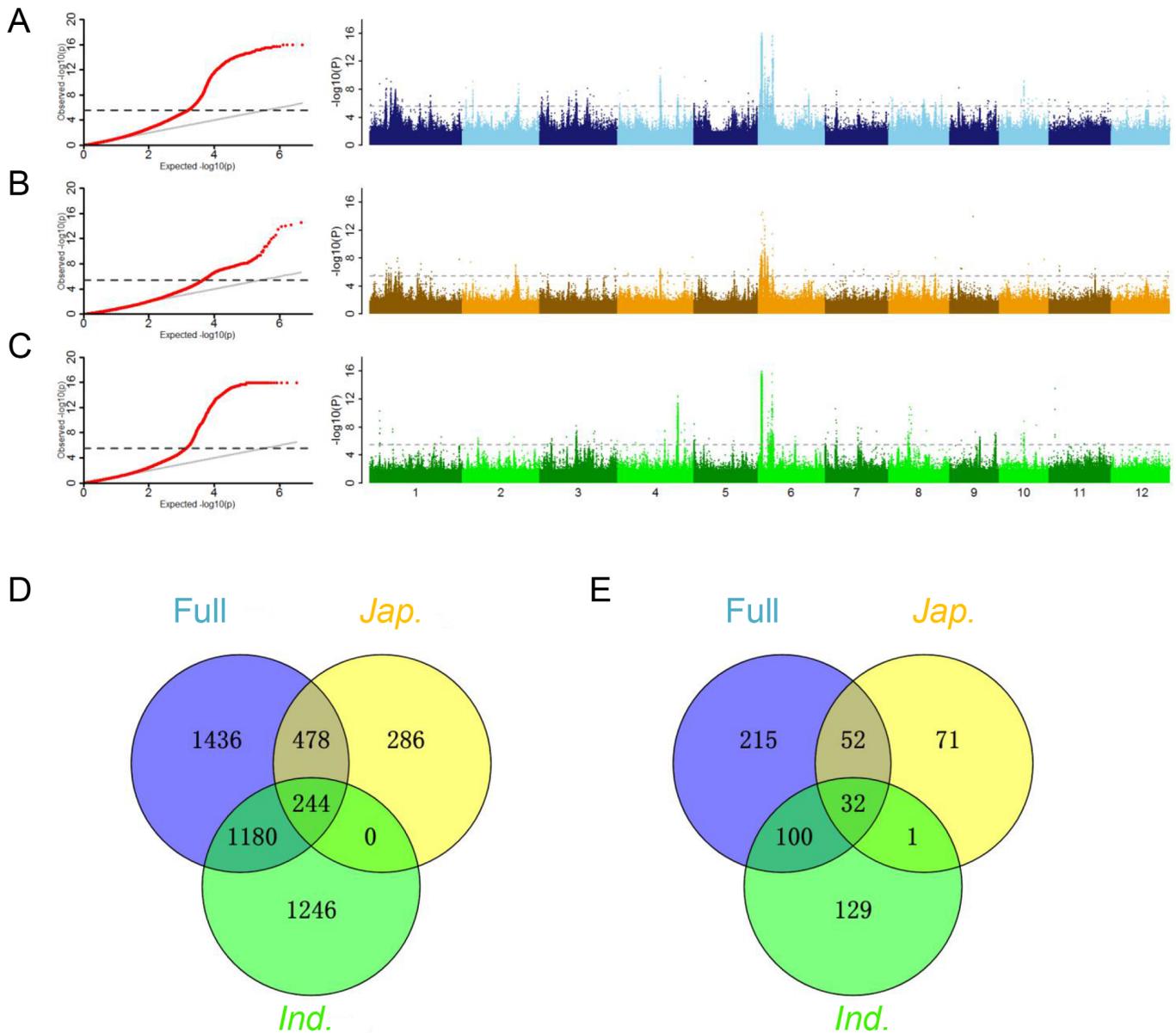
42. Martinez-Cortes G, Salazar-Flores J, Fernandez-Rodriguez LG, Rubi-Castellanos R, Rodríguez-Loya C, Velarde-Félix JS, Muñoz-Valle JF, Parra-Rojas I, Rangel-Villalobos H: **Admixture and population structure in Mexican-Mestizos based on paternal lineages.** *Journal of Human Genetics* 2012, **57**(9):568–574.
43. Tang Y, Liu X, Wang J, Li M, Wang Q, Tian F, Su Z, Pan Y, Liu D, Lipka AE: **GAPIT version 2: an enhanced integrated tool for genomic association and prediction.** *The plant genome* 2016, **9**(2):plantgenome2015.2011.0120.
44. Kumar S, Stecher G, Tamura K, evolution: **MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets.** *Molecular biology* 2016, **33**(7):1870–1874.
45. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D: **FaST linear mixed models for genome-wide association studies.** *Nature methods* 2011, **8**(10):833–835.
46. Yang W, Guo Z, Huang C, Duan L, Chen G, Jiang N, Fang W, Feng H, Xie W, Lian X: **Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice.** *Nature communications* 2014, **5**(1):1–9.
47. Li M-X, Yeung JM, Cherny SS, Sham PC: **Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets.** *Human genetics* 2012, **131**(5):747–756.

## Figures



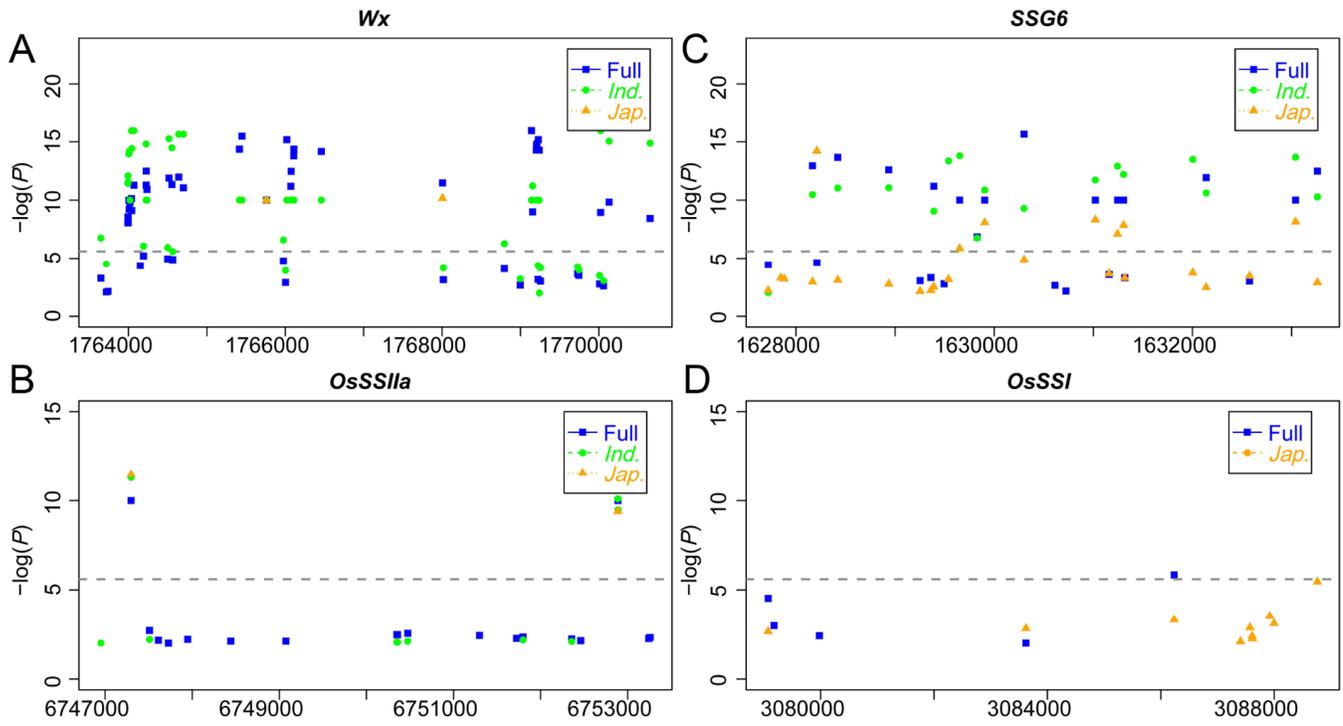
**Figure 1**

Population structure and geographic distribution of accessions of the association panel. (A) Genetic structure of the panel based on ADMIXTURE for  $K = 2$ . (B) Principle components analysis reveals that the first 3 principle components explain 25.4% of the genetic variation within the panel. Green and yellow dot show *i indica* and *japonica*, respectively. Red dot indicates glutinous rice accession. (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, 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 r waxy trait QTLs by GWAS. Quantile-quantile (Q-Q) plots and Manhattan plots in the whole population (A), japonica (B) and indica (C). (D) Venn diagram of significant loci detected in different populations. Black four known genes identified in this study. (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	Nonsys	Num.of accessions	
			<i>Ind./Glu ind</i>	<i>Jap./Glu jap</i>
<i>Wx-1</i>	AAAATTGTCAGTCGACGTCGCGTAGTGACGTCCTCC	AC	6/0	20/0
<i>Wx-2</i>	AAAATTGTCAGTCGCGGTTGCATAGTGGTGTCTCC	AC	7/0	20/0
<i>Wx-3</i>	AAAATTGTCAGTTGACGTCGCGTAGTGACGTCCTCC	AC	10/1	28/0
<i>Wx-4</i>	AAAATTGTCAGTTGCCGTCGCGTAGTGACGTCCTCC	AC	4/0	7/0
<i>Wx-5</i>	AAAATTGTCAGTTGCCGTTGCATAGTGGTGTCTCC	AC	9/0	19/0
<i>Wx-6</i>	AAACGGTACGACCCGCTATTTACGACGACACACTT	CC	0/0	10/0
<i>Wx-7</i>	AAACGGTATGACCACTACTTTACGACTACACACTT	AC	5/0	37/0
<i>Wx-8</i>	AAACGGTATGACCCGCTACTTTACGACGACACACTT	CC	65/0	343/5
<i>Wx-9</i>	AAACGGTATGACCCGCTACTTTACGACTACACACTT	AC	108/37	173/51
<i>Wx-10</i>	AAACGGTATGACCCGCTACTTTACGACTACACACTT	CC	0/0	21/7
<i>Wx-11</i>	GAACGGTATGACCCGCTACTTTACGACTACACACTT	AC	23/9	0/0
<i>Wx-12</i>	GGAAATTGTCAGTTGACGTCGCGTAGTGACGTCCTCC	AC	12/0	1/0
<i>Wx-13</i>	GGAAATTGTCAGTTGACGTCGCGTAGTGGTGTCTCC	AC	38/0	1/0
<i>Wx-14</i>	GGAAATTGTCAGTTGACGTCGCGTAGTGGTGTCTCC	AT	90/0	0/0
<i>Wx-15</i>	GGGATTGTCAGTCGACGTCGCGTAGTGACGTCCTCC	AC	65/1	4/0
<i>Wx-16</i>	GGGATTGTCAGTCGACGTCGCGTAGTGGTGTCTCC	AC	20/0	0/0
<i>Wx-17</i>	GGGATTGTCAGTCGACGTCGCGTAGTGGTGTCTCC	AT	142/3	1/0
<i>Wx-18</i>	GGGATTGTCAGTCGCCGTTGCATAGTGGCGTCTCC	AC	9/0	3/0
<i>Wx-19</i>	GGGATTGTCAGTCGCCGTTGCATAGTGGTGTCTCC	AC	13/0	1/0
<i>Wx-20</i>	GGGATTGTCAGTTGACGTCGCGTAACCTACACACTT	AC	11/5	0/0
<i>Wx-21</i>	GGGATTGTCAGTTGACGTCGCGTAGTGACGTCCTCC	AC	113/2	13/0
<i>Wx-22</i>	GGGATTGTCAGTTGACGTCGCGTAGTGGTGTCTCC	AC	64/0	1/0
<i>Wx-23</i>	GGGATTGTCAGTTGACGTCGCGTAGTGGTGTCTCC	AT	273/3	4/0
<i>Wx-24</i>	GGGATTGTCAGTTGCCGTTGCATAGTGGCGTCTCC	AC	45/0	1/0
<i>Wx-25</i>	GGGATTGTCAGTTGCCGTTGCATAGTGGCGTCTCC	CC	28/1	6/0
<i>Wx-26</i>	GGGATTGTCAGTTGCCGTTGCATAGTGGTGTCTCC	AC	11/0	4/0

B

HAP	Promoter										5'UTR	CDS		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

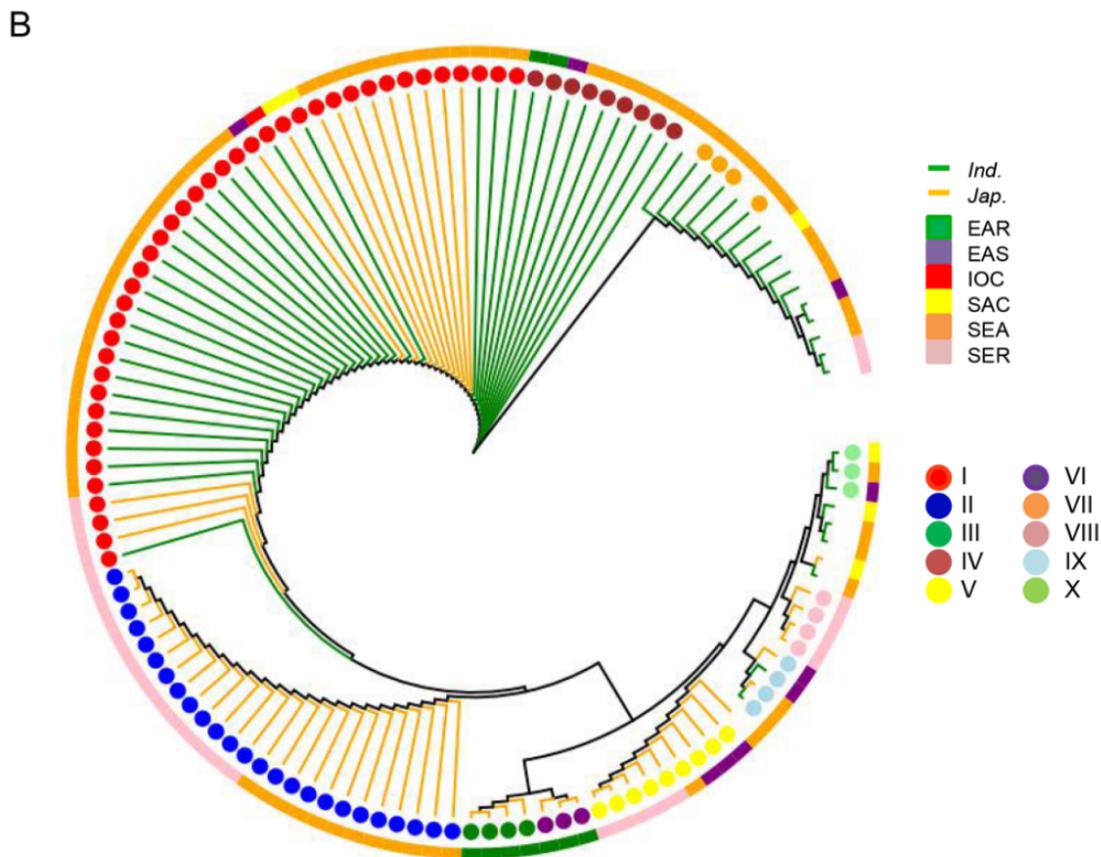
HAP	Promoter					5'UTR	CDS		3'UTR		Num.of accessions	
	6747298	6747510	6747612	6747727	6747948	6748443	6752357	6752888	6753241	6753257	<i>Ind./Glu ind</i>	<i>Jap./Glu jap</i>
<i>OsSSIIa-1</i>	A	A	T	C	G	A	A	T	A	A	296/58	212/54
<i>OsSSIIa-2</i>	G	A	T	C	G	A	A	C	A	A	384/7	375/18
<i>OsSSIIa-3</i>	G	A	T	C	G	A	A	C	G	G	12/0	0/0
<i>OsSSIIa-4</i>	G	G	C	A	C	T	G	C	G	G	550/4	215/1
<i>OsSSIIa-5</i>	G	G	C	A	G	A	G	C	G	G	35/0	4/0

Figure 4

Haplotype analyses for three key cloned genes conferring waxy traits. Haplotypes filled in color are waxy haplotypes. Number on the left and right represent the number of total accessions and glutinous accessions in indica and jaпонica, respectively

**A**

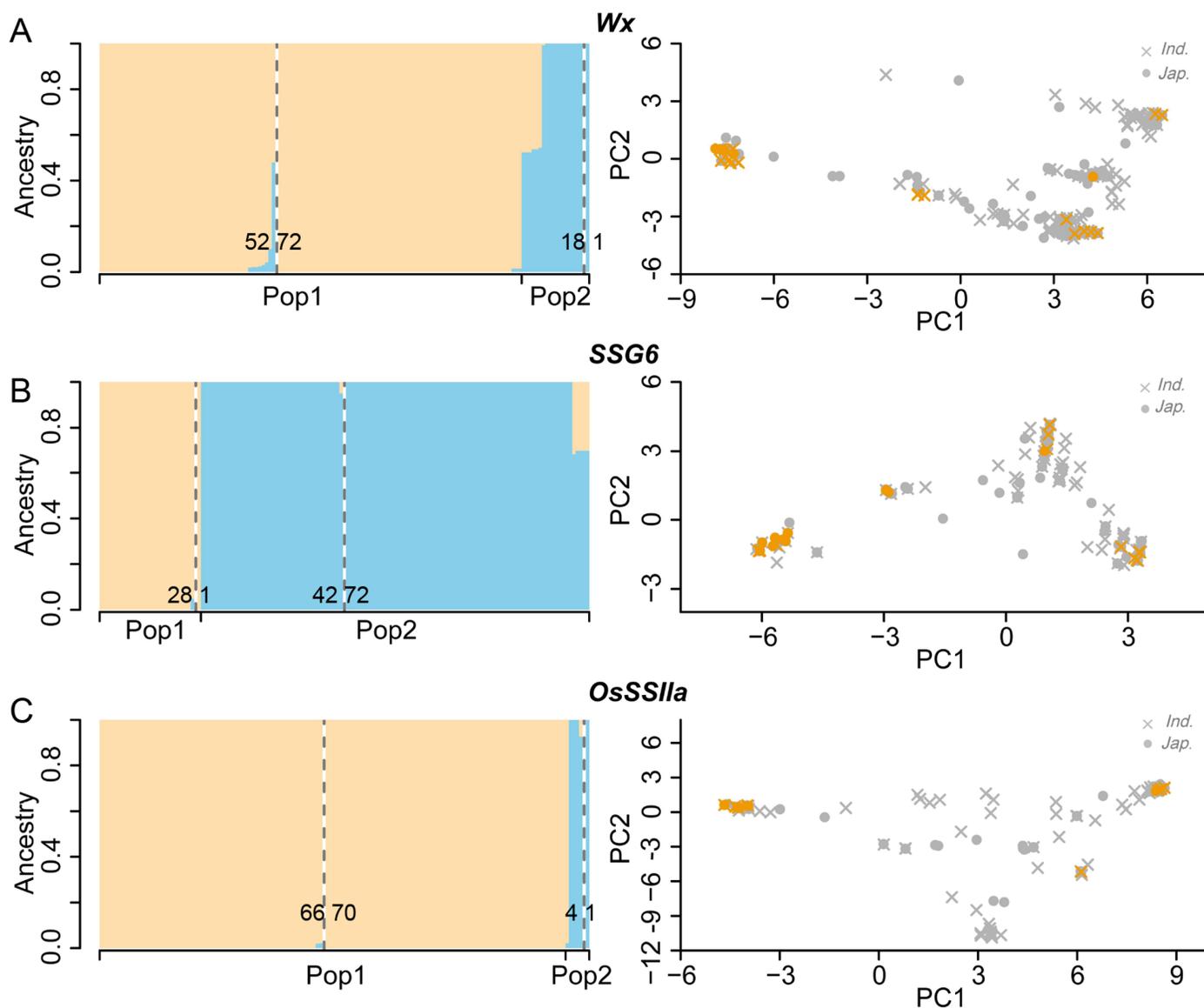
Group	Genotype			Num. of glutinous accessions	
				<i>Ind.</i>	<i>Jap.</i>
I	<i>Wx-9</i>	<i>SSG6-7</i>	<i>OsSSIIa-1</i>	27	14
II	<i>Wx-9</i>	<i>SSG6-5</i>	<i>OsSSIIa-1</i>	0	25
III	<i>Wx-10</i>	<i>SSG6-7</i>	<i>OsSSIIa-1</i>	0	4
IV	<i>Wx-11</i>	<i>SSG6-7</i>	<i>OsSSIIa-1</i>	9	0
V	<i>Wx-9</i>	<i>SSG6-5</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



**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

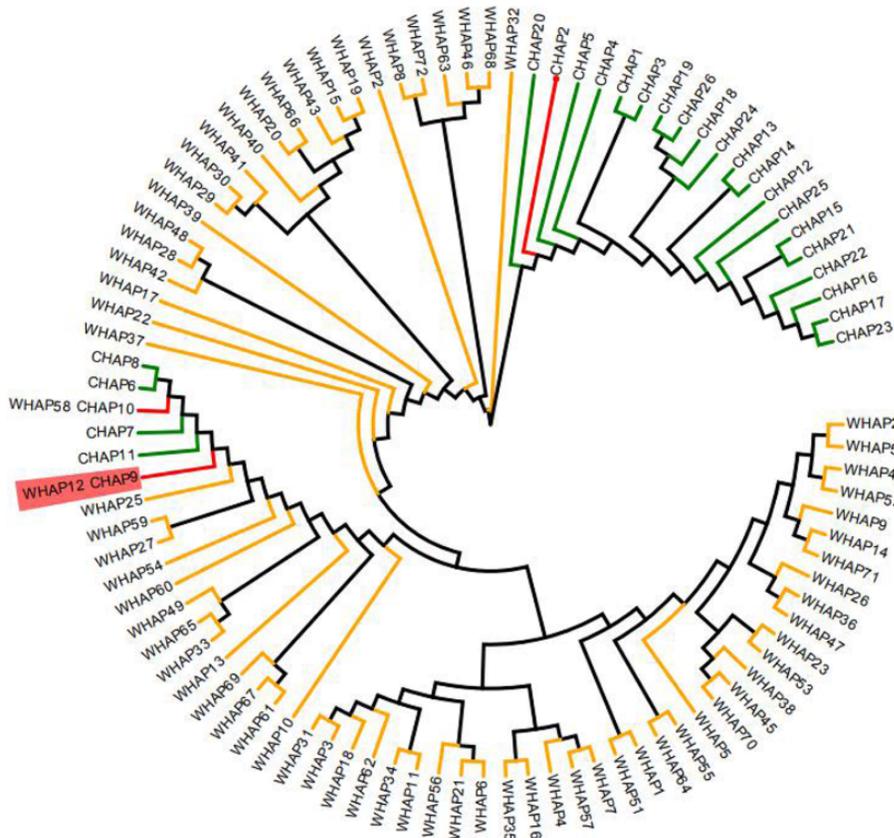
glutinous accessions. Linear represent subpopulation. Colored squares represent different regions. Colored circles represent different combined haplotypes.



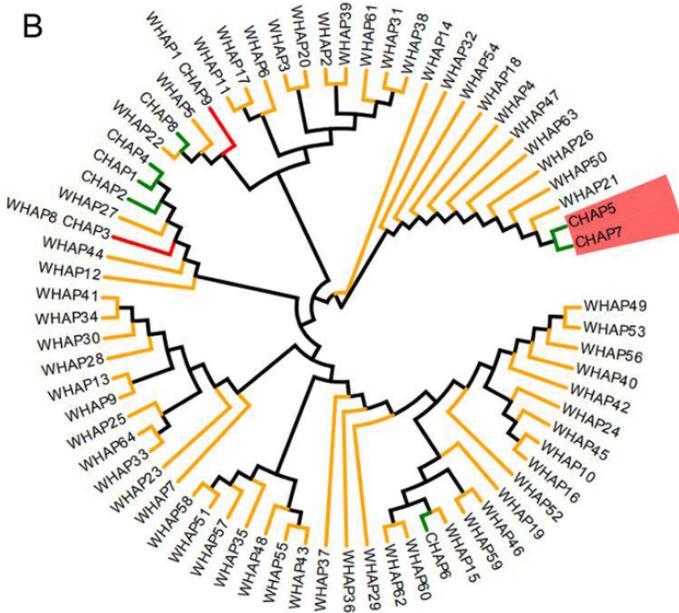
**Figure 6**

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

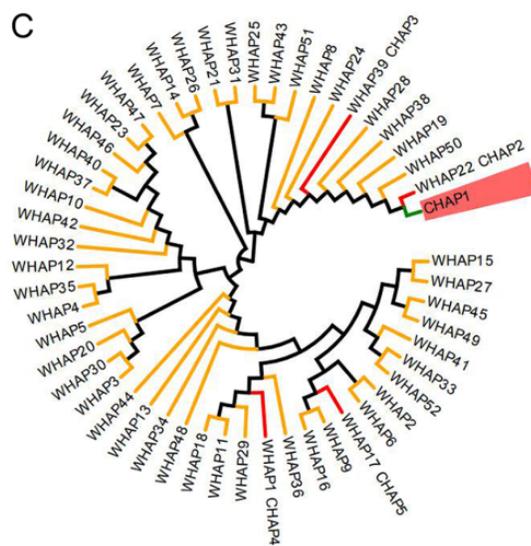
A



B



C



**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 linear represent haplotypes in cultivated ,wild rice and shared between them, 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.

- [TableS11.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [S1.png](#)
- [S2.png](#)
- [S3.png](#)