Evolution of the Wx genotype combination in hybrid rice

CURRENT STATUS: UNDER REVISION

bingran Zhao brzhao123@163.com
State Key Laboratory of Hybrid Rice, Hunan Hybrid Rice Research Center
Corresponding Author
ORCID: 0000-0002-2070-5695

Ye Shao
Hunan Agricultural University

Yan Peng
State Key Laboratory of Hybrid Rice, Hunan Hybrid Rice Research Center

Bigang Mao
State Key Laboratory of Hybrid Rice, Hunan Hybrid Rice Research Center

Qiming Lv
State Key Laboratory of Hybrid Rice, Hunan Hybrid Rice Research Center

Dingyang Yuan
State Key Laboratory of Hybrid Rice, Hunan Hybrid Rice Research Center

Xionglun Liu
Hunan Agricultural University

DOI: 10.21203/rs.2.18052/v1

SUBJECT AREAS
Evolutionary Genetics Plant Molecular Biology and Genetics

KEYWORDS
Hybrid Rice, Wx alleles, 3K RGP, KASP, Eating Quality, Germplasm
Abstract

Background: With increased interest in the quality characteristics of hybrid rice, its quality traits have received a certain extent of improvement. However, comparisons to high-quality conventional rice, especially its eating quality, have revealed that more work is still required. Amylose content is a key determinant of the eating quality of rice, and Wx, the major gene controlling amylose content, has many allelic variants. A complete understanding of the Wx allelic variations in cultivated rice and the evolution process of the Wx genotype in hybrid rice are the premises for the rational utilization of rich genetic resources in the quality breeding of hybrid rice. Based on the 3K RGP re-sequencing data, we sought to analyse the allelic variation at the Wx locus and develop a set of Kompetitive Allele-Specific PCR markers. Furthermore, we used the markers to analyse the evolution of a combination of the Wx genotype in hybrid rice.

Results: Eight known alleles existing globally were identified, and their evident regional preferences and Indica-japonica background differences were revealed. An additional five non-synonymous mutations were identified in the coding region of the Wx gene for the first time, including a new functional site located in the active centre of OsGBSS1. By genotyping the basic hybrid parents obtained from 1976 to 2018, we found that only three Wx allelic variations existed. Wx lv was widely used in the female parents of the three-line hybrid rice, and Wx a was used in the female parents of early two-line hybrid rice; however, they were gradually replaced by Wx b.

Conclusions: In this study, the Wx lv allele was found to be the cause of the poor eating quality of early hybrid rice. With the elimination of the Wx lv allele and
introduction of \( Wx_b \) in both parents, the eating quality of hybrid rice was generally improved. Only three allelic variations were present in the previous hybrid rice. For further eating quality improvement of hybrid rice, more \( Wx \) alleles should be introduced.

**Background**

Rice is one of the most important crops. In fact, its greatest success has been achieved in the application of heterosis. For many years, researchers have devoted much attention to high-yield hybrid rice. However, in China, research on the quality of rice began fairly late, especially in the southern areas. With continued improvement in the standard of living, the quality of rice is now of greater concern than ever before. In the last 15 years, the quality traits of *indica* hybrid rice was improved to a certain extent; however, further improvements are still critical compared to the high-quality conventional rice, especially its eating quality (Chen et al. 2015; Tang et al. 2016). As many hybrid parents widely used from the same basic materials, the genetic polymorphism of hybrid rice may be relatively low (Li et al. 2000). Understanding the diversity and evolution of genes associated with eating quality is thus the foundation needed to fully utilize the rich genetic resources to achieve improvements in the quality of hybrid rice.

The quality of rice is a complex characteristic, with amylose content (AC) serving as a key determinant of its eating and cooking qualities (Tian et al. 2009; Qian et al. 2016). The \( Wx \) gene, encoding a granule bound starch synthase (GBSS), is a major gene in the control of AC. Its RNA expression and the activity of the OsGBSS1 protein positively correlate with AC (Liu et al. 2014). To date, at least eight \( Wx \) alleles have been identified in rice—\( Wx^a, Wx^b, Wx^{in}, Wx^{op}, Wx^{mp}, Wx^{mw}, Wx^{lv} \), and
The AC of the eight alleles decreases in sequence from $Wx^a$ (>25%), $Wx^{lv}$ (20-25%), $Wx^{in}$ (18-22%), $Wx^b$ (15-18%), $Wx^{mw}$ (10-14%), $Wx^{mp}$ (8-12%), and $Wx^{op}$ (5-10%) to $wx$ (AC<2%) (Zhang et al. 2019; Yang et al. 2013; Liu et al. 2009; Sato et al. 1996; Mikami et al. 1999, 2000, 2008). Presently, eight $Wx$ alleles have been identified. However, whether there are new functional alleles in the cultivated rice resources worldwide are yet to be elucidated. Nonetheless, we hope to develop a set of molecular markers for high-throughput genotyping in the $Wx$ loci but whether the known alleles can cover most of the $Wx$ allelic variations and explain the AC variances remain unknown. A complete understanding of the $Wx$ allelic variations in cultivated rice worldwide is thus the premise for $Wx$ genotyping and the efficiency of utilizing hybrid rice.

In 2015, a total of 3,000 germplasm accessions from 89 different countries/regions were selected for genome-wide resequencing, and a comprehensive SNP and InDel sub-database was established for the Rice Functional Genomics-based Breeding (RFGB) Database (Li et al. 2014). From the samples, 2,466 accessions were retrieved from a core collection established from more than 101,000 rice accessions in the IRGC while 534 accessions were selected from a core collection established from 61,470 rice accessions preserved in the CNCGB (Zhang et al., 2011). Thus, the 3000 germplasm accessions represent a panel with abundant genetic diversity in the global cultivation of rice. By utilizing the currently available genome information from the 3 K RGP sequencing data, different types of allelic variations in the $Wx$ loci can be comprehensively analysed, and new allelic variants that are responsible for different AC classes can be discovered to improve future quality of hybrid rice.

A prior research revealed that a significant correlation exists between the number
of (CT)n and AC (Larkin et al. 2003) and related microsatellite markers have also been developed. However, some of the AC variations cannot be accurately explained by these markers (Dobo et al. 2010). A series of Wx alleles has been recently identified in rice (Zhu et al. 2015), and numerous SNP genotyping methods have been used for polymorphism analysis of the Wx loci. Kompetitive Allele-Specific PCR (KASP) is a high throughput method that can type SNPs and InDels at specific sites. Based on terminal fluorescence reading, different genotypes at a single site can be detected by two-colour fluorescence (He et al. 2014) Therefore, owing to advantages such as efficiency, accuracy, and high throughput, establishing KASP to distinguish functional SNP would greatly accelerate the molecular design for breeding good quality rice.

Herein, to make better use of Wx allele resources for improving the eating-quality breeding of hybrid rice, we sought to understand the diversity of Wx alleles in cultivated rice and the evolution of the Wx genotype combination in hybrid rice. By using the 3K RGP re-sequencing data, we could identify not only the eight known alleles that widely existed in the germplasm resources, but also a potential novel allele. A set of KASP markers based on the Wx alleles was developed for high throughput genotyping of cultivated rice and the basic parent lines of hybrid rice obtained from 1976 to 2018 were selected for genotyping the Wx loci. As a result, only three Wx allelic variations were found in the main parents of the hybrids, and the allelic combination of hybrids changing from Wxlv/Wxb to Wxa/Wxb and then to Wxb/Wxb with improvements in the quality of hybrids. Our results revealed the evolution of the eating quality of hybrid rice at the molecular level and provide a foundation for future breeding of hybrid rice with good eating quality. The present
study also revealed new approaches for improving other quality traits in future hybrids.

Materials and Methods

Plant materials

By referring to previously described data (Hu et al. 2016), the Hybrid Rice Variety Resources Database (http://www.hybridrice.com.cn/), and China Rice Data Centre (http://www.ricedata.cn/), we selected 36 basic parent lines of hybrid rice obtained between 1976 and 2018 (Hu et al. 2016) for Wx genotyping. All combinations of these parents are hybrid rice varieties that have occupied the largest planting areas for several years. The hybrid parents included 18 female parents, namely Erjiunan 1A, Zhenshan 97A, V20A, Gang 46A, Longtefu A, Bo A, II-32A, Xieqingzao A, Jin 23A, Zhong 9A, Tianfeng A, Annong-S1, Peiai 64S, Guangzhan 64-4S, Y58S, Zhu 1S, C815S, and Longke 638S, and 18 male parents, namely, IR24, IR26, Minghui 63, Ce 64-7, Duoxi 1, Fuhui 838, Xianhui 207, Shuhui 527, CDR22, Miyang 46, R402, Minghui 86, Mianhui 725, Gui 99, Guanghi 3550, Yangdao 6, Bing 4114, and Huazhan. The control group included Zhenshan 97A (Wx\(^{(v)}\)), Nipponbare (Wx\(^{(b)}\)), Basmati (Wx\(^{(in)}\)), Haopi (Wx\(^{(op)}\)), and Nanjing 46 (Wx\(^{(mp)}\)). All hybrid parent varieties and control materials for genotyping were stored in our lab. The four germplasm resources of IRIS_313-9445, IRIS_313-10892, IRIS_313-10866, and IRIS_313-8956 were obtained from the Crops Research Institute, Chinese Academy of Agricultural Sciences. All the varieties were cultivated under normal growing conditions in the experimental field of the Hunan Hybrid Rice Research Centre in Changsha.

Re-sequencing data of the Wx alleles

Data for the SNPs and InDels at the Wx locus in 3,000 rice accessions were
downloaded from the Rice SNP-Seek Database18 (http://oryzasnp.org/iric-portal/index.zul).

**Sequence analysis**

Sequences were aligned using CLUSTAL X version 2.0 and adjusted manually with Microsoft Office Excel 2010 (Larkin et al. 2007). Haplotype diversity was calculated using DNASP v5.0 (Rozas et al. 2003) while the haplotype network was constructed using PopART 1.7 (Leigh et al. 2015). A geographical distribution map of the Wx alleles was generated using rworldmap V1.36 (South et al. 2011).

**Crystal structure analysis**

The crystal structure of the rice GBSSI catalytic domain in complex with ADP was downloaded from Protein Data Bank (http://www.rcsb.org/pdb/). PyMOL Molecular Graphics System (Schrödinger LLC) was used to display the structural features of the OsGBSSI protein (PDB: 3VUF), with a focus on the novel mutations identified.

**Detection of the five allelic variations by Sanger sequencing**

Genomic DNA was extracted from fresh leaves of the IRIS_313–9445, IRIS_313–10892, IRIS_313–10866, and IRIS_313–8956 varieties using a modified CTAB method. Briefly, a 266-bp sequence containing the mutation site of Ex2+160 was amplified using the primer pair, 5′-ATGTCGGCTCTCACCACG–3′ and 5′-CCGACGAACACGACGTTCATG–3′; a 297-bp sequence containing the mutation site of Ex4+73 was amplified using the primer pair, 5′-GATACCAGCGTTGTGGCTGAG–3′ and 5′-CAGTCCAACTGCTAAATGCACTG–3′; a 194-bp sequence containing the mutation site of Ex14+2 was amplified using the primer pair, 5′-GAGTGACAAATTTCAGGCAATCGAG–3′ and 5′-CCAGAAGAACGATCTGGACGTC–3′; a 174-bp sequence containing the mutation sites of Ex14+2 and Ex14+28 was amplified
using the primer pair, 5′-CAGAGATTACCTGTCTGATGCTG–3′ and 5′-TCAAGGAGCAGCCACGTTCTC–3′. PCR amplification was carried out as follows: initial DNA denaturation at 95 °C for 4 min; 30 cycles of denaturation at 95 °C for 30 s, annealing at 58 °C for 30 s, and extension at 72 °C for 30 s; and final extension at 72 °C for 5 min. After gel purification, the PCR products were sequenced by TsingKe Biology Technology.

**KASP genotyping**

The allele-specific primers were designed to carry the standard FAM (5′GAAGGTAACCAATGCT3′) and HEX (5′ GAAGGTCGGAGTCAACGGATT 3′) tails and the targeted SNP at the 3′ end. Assays were carried out in 384-well formats and 10-µl reactions (20–30 ng/µl DNA, 5 µl of 1× KASP master mixture, 0.14 µl of KASP assay mix, and 4.86 µl of water). PCR was conducted using the following protocol: hot start at 94 °C for 15 min, ten touchdown cycles (94 °C for 20 s; initial touchdown at 61 °C and then a decrease by −0.6 °C per cycle for 60 s), and 26 additional cycles of annealing (94 °C for 20 s; 55 °C for 60 s). Finally, the PCR product with fluorescent labelling was scanned using a Roche Light Cycler480 (37 °C for 1 min).

**AC measurement**

The AC of rice seeds was measured using the iodine colorimetry assay described previously (Williams et al. 1970).

**Results**

**Analysis of the Wx allelic variation among 3 K RGP**

By using the CDS of *Nipponbare* as a reference sequence in this study, allelic
variations in the coding region of the Wx gene were analysed according to the 3 K RGP sequencing data. After eliminating the alleles with heterozygous sequences in the coding region and the possible deletion caused by insufficient sequencing coverage, a total of 2,752 Wx alleles were obtained for variation analysis. Based on the G/T SNP at the splice site of the first intron (In1G/T SNP), two important alleles, Wx\textsuperscript{a} and Wx\textsuperscript{b}, were originally identified. Although this functional SNP does not belong to the coding region, it was considered for the variation analysis.

Based on 1 insertion and 29 nucleotide polymorphic sites (SNP), a total of 30 haplotypes were identified for the Wx alleles according to the 1830-bp coding region and the In1G/T SNP of the 2,752 Wx alleles (Fig. 1). Among them, Int1+1, Ex2+88, Ex4+53, Ex4+77, Ex6+62, and Ex10+115 were the six known functional SNPs, thereby forming the eight alleles and representing 99.9% of the allelic variation among the 2,752 Wx alleles. Five non-synonymous mutations, namely Ex2+160, Ex4+73, Ex10+101, Ex14+2, and Ex14+28 were identified for the first time, and they resulted in amino acid changes from Ala to Thr(54), Ile to Val(165), Glu to Gly(410), Gly to Glu(572), and Leu to Phe(581), respectively. Ex2+160 and Ex4+73 coexisted in the IRIS_313-9445 material, with the Wx\textsuperscript{a} genetic background; Ex10+101 existed in IRIS_313-10892, with the Wx\textsuperscript{a} genetic background; Ex14+2 existed in IRIS_313-10866, with the Wx\textsuperscript{in} genetic background; and Ex14+28 existed in IRIS_313-8956, with the Wx\textsuperscript{a} genetic background. By determining the AC in the mature seeds of these four materials, we found that the AC of IRIS_313-9445 and IRIS_313-8956 was 25 and 24, respectively, a finding consistent with Wx\textsuperscript{a}. In addition, the AC of IRIS_313-10866 was 18, aligning with Wx\textsuperscript{in}. Such findings suggest that Ex10+101 in IRIS_313-10892 may be a new functional allele with an
AC of 5 (Fig. 2-F). By analysing the crystal structure of OsGBSS1, we found that Glu410Gly (Ex10+101) is located in the active centre of the enzyme, beside the ligand ADP of OsGBSS1, while the remaining 4 mutation sites were located at a distant from the active centres (Fig. 3-B). In addition, Ala54Thr did not exist in the truncated body of OsGBSS1.

**Construction of the Wx haplotype flowchart**

To comprehensively understand the genetic background of each haplotype and the relationships among them, a haplotype flowchart was constructed to describe the mutational steps of the 30 haplotypes. In Fig. 4, Hap1, Hap3, Hap4, and Hap9, which represent Wxb, Wxa, Wxl, and Wxin, respectively, demonstrated absolute predominance compared to others in cultivated rice. Wxl is an ancestral allele of Wx, which markedly affects the mouthfeel of rice grains by modulating the size of the amylose molecules (Zhang et al. 2019). In this study, the nucleotide diversities of Wxl were much lower than those of the other three alleles and had only one haplotype. This finding may indicate that Wxl has fewer positive selection than others because of its relatively poor taste. With the exception of wx, an evident indica-japonica background difference existed among the remaining seven alleles. Notably, the Wxa, Wxl, and Wxo alleles belonged to the indica background while Wxb, Wxin, Wxm, and Wxmp belonged to the japonica background. Wxa was recognized as the predominant allele of the cA subgroup while Wxin, which is the predominant allele of cB, had the famous Basmati and Sadri aromatic varieties. Altogether, these findings indicate that each subpopulation contained significantly dominant Wx alleles, which suggest that Wx alleles might have regional preference and a corresponding environmental adaptability.
Worldwide distribution of different Wx alleles

Based on the geographic distribution of the eight alleles (Fig. 5), $Wx^a$, $Wx^b$, $Wx^{in}$, and $Wx^{iv}$ are the four predominant alleles that display a wide distribution in most rice producing regions of the world. However, evident regional differences exist among them; $Wx^{iv}$ was among the most widely distributed in the world, $Wx^b$ was widely distributed in higher latitudes, $Wx^{in}$ was widely distributed in the tropical areas of low latitudes, $wx$ was commonly distributed in the Southeast and East Asia, and $Wx^a$ was widely distributed in the areas of SAC and SAE. Such findings are consistent with the eating habits as high amylose varieties are popular in Myanmar, Sri Lanka, provinces of Indonesia, and many states of India (Calingacion et al. 2014).

Correlation analysis between the Wx alleles and CT repeats

In most prior studies, GBSS alleles were primarily defined according to the CT repeat in exon 1. Generally, varieties with 17 or 18 CT repeats are considered to be “low amylose” types, while those with 10 or 11 CT repeats are generally classified as “high amylose” types (Dobo et al. 2010). Although previous studies found a significant correlation between the number of CT repeats and AC, some related molecular markers have been developed. Nonetheless, a considerable part of AC variation could not be explained by the CT repeats. By analysing the number of CT repeats that corresponds to different Wx alleles in the 3K data, we found that most low amylose varieties had long CT repeats while high amylose varieties had short CT repeats, a result that aligns with that of a previous study. However, some high amylose alleles, such as $Wx^a$ and $Wx^{iv}$, also contain long CT repeats, and some low
amylose alleles, such as Wx$^{op}$, only have short CT repeats (Fig. 6). These exceptions help to explain the AC variation, which cannot be explained by microsatellite markers.

**KASP genotyping of the Wx alleles in hybrid parents**

Owing to the importance of the eight Wx alleles that account for 99.9% of the Wx allele resources, a set of KASP markers was developed based on these Wx alleles (Fig. 7). Thereafter, 36 basic parent lines, including 18 female parents and 18 male parents of hybrid rice from 1976 to 2018 (Hu et al. 2016), were selected for genotyping (Fig. 8). Fig. 9 indicates that Wx$^{lv}$ is the major allele in the male sterile line of the three-line hybrid rice, such as zhenshan97A, which is the female parent of many hybrids that are widely used in the early stages of developing three-line hybrid rice. Conversely, for the two-line hybrid rice, Wx$^a$ is the major allele of the early sterile line, such as Annon 6–1. Additionally, its appearance indicates the beginning of two-line hybrid rice, and Peiai 64S, which is the most influential sterile line, has the largest two-line hybrid combination planting area. Owing to the development of the two-line hybrid rice, Wx$^a$ was replaced by Wx$^b$. Wx$^b$ is a major allele for nearly all restorer lines, except Ce64–7 and R402. The genotyping results revealed that the major allelic combinations changed from Wx$^{lv}$/Wx$^b$ to Wx$^a$/Wx$^b$ and then to Wx$^b$/Wx$^b$, ultimately indicating the evolution process, at the molecular level, for the quality improvement of hybrid rice in China.

**Discussion**

Rice quality, especially its eating quality, is a complex characteristic, with Wx
serving as one of the major genes affecting its eating quality and consisting of many allelic variations. In the present study, resequencing data from 3,000 germplasm resources from 89 countries worldwide were used to comprehensively analyse the $Wx$ alleles. Through further analyses, 99.9% of the alleles were recognized to be the eight known functional alleles. Thus, a set of KASP markers was developed, and the basic hybrid parent varieties from 1976 to 2018 were selected for $Wx$ genotyping. Based on the typing results, the allelic combinations changed from $Wx^{lv}/Wx^{b}$ to $Wx^{a}/Wx^{b}$ then to $Wx^{b}/Wx^{b}$ as quality improved. Previously, the $Wx^{a}$ allele was widely assumed to be the main reason for the poor taste of the early three-line hybrid rice. However, the findings of this study demonstrate that $Wx^{lv}$, an ancestral allele of $Wx$ (Zhang et al. 2019), is responsible for the considerable effects of the mouthfeel of rice grains and was thus gradually selected in the subsequent breeding process. $Wx^{a}$ was used in male parents of the early two-line hybrid rice and could gradually be replaced with $Wx^{b}$. As $Wx^{b}$ was introduced in both parents, the quality of the two-line hybrid rice was generally improved. Quality improvement of hybrid rice might thus be caused by $Wx^{lv}$ elimination. In addition, the homozygosis of the $Wx^{b}$ allele in hybrids may serve as another important factor. Genotype segregation is known to exist in the endosperm characters of hybrid rice. With homozygosis of the $Wx$ allele in hybrids, the AC of each grain in the hybrid combination tend to be more homogeneous, and the eating quality was identified to be better. For the future breeding of hybrid rice with high-quality, more favourable $Wx$ allelic variations should be identified, introduced, and effectively used. By evaluating the genetic effects of different allelic combinations, we can derive the most appropriate genotype combination to achieve high-quality.
By analysing the molecular evolution of the Wx allele in the *indica* hybrid rice, we found that only three Wx allelic variations existed in the main parents of hybrids. Thus, expanding the genetic background of hybrid parents might be key for improving the quality character. Our analysis showed that rich allelic variations of Wx exist in cultivated rice. However, the mechanism whereby these alleles affect the eating quality of hybrids has not been comprehensively evaluated, and most are not well utilized in quality breeding. Nonetheless, we hope to introduce the remaining four non-glutinous alleles to accommodate different consumption demands in the quality breeding of *indica* hybrid rice. With the introduction of Wx\textsuperscript{b} in both parents in recent years, the AC of *indica* hybrid rice was found to generally reduce to 15% or lower, with potential improvements in its quality traits. Some studies have confirmed that a high temperature at the milky stage of grain filling has the greatest influence on rice quality, with the panicle identified as the organ most sensitive to high temperature (Yamakawa et al. 2007). As high temperatures are frequent in the southern regions of China, the AC of hybrids in these regions was significantly reduced and their appearance and eating qualities might be greatly affected. Wx\textsuperscript{in} is a major allele with intermediate AC (18–20%). Thus, when it is compared to Wx\textsuperscript{b}, the effect of its response to abnormally high temperature on the quality of AC may not be evident. Based on our analysis, Wx\textsuperscript{in} is the predominant allele in the circum-Basmati group (cB) that many high-quality rice varieties, such as Basmati and Sadri aromatic varieties. Most of the Wx\textsuperscript{in} alleles were distributed in tropical areas with low latitudes compared to the Wx\textsuperscript{b} alleles distributed in high latitudes. Thus, we concluded that Wx\textsuperscript{in} may be the preferred choice for improving *indica*-type rice quality in southern China.
In the present study, five new alleles were identified for the first time. By determining the AC in four resources, including the allelic variations, we found that with the exception of the significant decline in the AC of IRIS_313-10892 compared to that of the $Wx^a$ genetic background, the AC of the other three resources was generally unchanged. By analysing the crystal structure of OsGBSS1, we found that Lys413 and Arg408 are two important amino acids that directly interact with ADP. In addition, Glu410 was found to be linked to both amino acids via simultaneous hydrogen bonds. Previously, the effects of amino acid substitutions on OsGBSS1 activity were revealed (Liu et al. 2014). Glu410Gly is a natural allelic variation where the acidic amino acid is replaced by a non-polar amino acid without any R group; this type of change must have a greater effect on the conformation of Lys413 and Arg408. With the exception of wx in glutinous rice, $Wx^{op}$ is presently the allelic variation with the lowest AC in non-glutinous rice. In this study, we found a novel allele that has an AC between wx and $Wx^{op}$, which ultimately cause more diversity in the variation of AC in cultivated rice. Thus, in future rice breeding, varieties with different ACs could be employed to meet different consumer demands and industrial applications.

Conclusions

Previously, replacing $Wx^b$ with $Wx^a$ in hybrid rice was widely believed to improve its eating quality. However, in the present study, the $Wx^a$ allele in early hybrid rice was in fact $Wx^{iv}$, which may have been responsible for the poor taste of early hybrid rice. By eliminating the $Wx^{iv}$ allele and introducing $Wx^b$ in both parents, the eating-quality of hybrid rice can be generally improved. As only three allelic variations
were present in the previous hybrid rice, for further eating quality improvement of hybrid rice, more $Wx$ alleles should be introduced.

List of Abbreviations

- amylose content, AC
- circum-Aus, cA
- circum-Basmati, cB
- derived cleaved amplified polymorphic sequence, dCAPS
- granule bound starch synthase, GBSS
- non-template control, NTC
- Kompetitive Allele-Specific PCR, KASP
- Rice Functional Genomics-based Breeding, RFGB
- East Asia Islands, EAR
- East Asia, EAS
- Southeast Asia, SEA
- SEA islands, SER
- SER-IRRI
- Oceania, OCE
- South Asia—East, SAE
- South Asia—Central, SAC
- South Asia—West, SAW
- West Asia, WAS
- Indian Ocean, IOC
- East Africa, EAF
- West Africa, WAF
North Africa, NAF
South America, SAM
Central America and Caribbean, CAM
North America, NAM
Europe, EUR
Xian/Indica, XI
Geng/Japonica, GJ

Declarations

Ethical approval and consent to participate
Not applicable

Consent for publication
Not applicable

Availability of data and material
All data generated or analysed during this study are included in this published article and its supplementary information files

Competing interests
The authors declare that they have no competing interests

Funding
This research was financially supported by funds from the Pre-cultivation Project of National Biological Seed Industry Technology Innovation Center (2018XK2005) and the Key Project of Hunan Province (2018NK1020-1).

Author’s contributions
XL and BZ designed the project. YS and YP performed all the experiments and wrote the manuscript together. YS responsible for data analysis. BM, QL and DY
responsible for material selection and planting. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Chen LY, Lei DY, Tang WB, Deng HB, Xiao YH, Zhang GL (2015) Challenges and strategies of hybrid rice development in China. Hybrid rice 30(5):1-4.

2. Calingacion M, Laborte A, Nelson A, et al (2014) Diversity of global rice markets and the science required for consumer-targeted rice breeding. PloS one 9(1): e85106.

3. Dobo M, Ayres N, Walker G, Park WD (2010) Polymorphism in the GBSS gene affects amylose content in US and European rice germplasm. Journal of cereal science 52(3): 450-456.

4. He C, Holme J, Anthony J (2014) SNP genotyping: the KASP assay. Crop breeding Humana Press, New York, NY, 75-86.

5. Hu ZX, Tian Y, Xu QS (2016) Review of extension and analysis on current status of hybrid rice in China. Hybrid rice 31(2): 1-8.

6. Li YH, Qian Q, Zeng DL, Sun ZX (2000) Identification by RAPD analysis and studies on genetic relationship of main parents of hybrid rice in China. Zuo wu xue bao 26(2): 171-176.

7. Liu DR, Wang W, Cai XL (2014) Modulation of amylose content by structure-based modification of OsGBSS1 activity in rice (Oryza sativa L). Plant biotechnology journal 12(9): 1297-1307.

8. Larkin PD, Park WD (2003) Association of waxy gene single nucleotide
polymorphisms with starch characteristics in rice (*Oryza sativa*). Molecular Breeding 12(4): 335-339.

9. Liu LL, Ma XD, Liu SJ, Zhu CL, Jiang L, Wang YH, Shen Y, Ren YL, Dong H, Chen LM, Liu X, Zhao ZG, Zhai HQ, Wan JM (2009) Identification and characterization of a novel Waxy allele from a Yunnan rice landrace. Plant molecular biology 71(6): 609-626.

10. Li JY, Wang J, Zeigler RS (2014) The 3,000 rice genomes project: new opportunities and challenges for future rice research. GigaScience 3(1): 8.

11. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. bioinformatics 23: 2947-2948.

12. Leigh JW, Bryant D (2015) popart: full-feature software for haplotype network construction. Methods in Ecology and Evolution 6(9): 1110-1116.

13. Mikami I, Uwatoko N, Ikeda Y, Yamaguchi J, Hirano HY, Suzuki Y, Sano Y (2008). Allelic diversification at the wx locus in landraces of Asian rice. Theoretical and Applied Genetics 116(7): 979-989.

14. Mikami I, Aikawa M, Hirano HY, Sano Y (1999) Altered tissue-specific expression at the Wx gene of the opaque mutants in rice. Euphytica 105(2): 91-97.

15. Mikami I, Dung L V, Hirano HY, Sano Y (2000) Effects of the two most common Wx alleles on different genetic backgrounds in rice. Plant breeding 119(6): 505-508.

16. Nevame AYM, Emon RM, Malek MA, Hasan MM, Alam M, Muharam FM, Aslani F, Rafii MY, Ismail MR (2018) Relationship between High Temperature and Formation of Chalkiness and Their Effects on Quality of Rice. BioMed research
17. Qian Q, Guo L, Smith SM, Li JY (2016) Breeding high-yield superior quality hybrid super rice by rational design. National Science Review 3(3): 283-294.

18. Rozas J, Sánchez-DelBarrio JC, Meseguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19(18): 2496-2497.

19. Sato H, Suzuki Y, Sakai M, Imbe T (2002) Molecular characterization of Wx-mq, a novel mutant gene for low-amylose content in endosperm of rice (Oryza sativa L). Breeding Science 52(2): 131-135.

20. Suto M (1996) Breeding of low amylose content paddy rice variety “Milky Queen” with good eating quality. Jpn. J. Breed 46(1): 221.

21. South A (2011) rworldmap: A new R package for mapping global data. R Journal, 3(1).

22. Tang WB, Zhang GL, Xiong YD, Ming XQ, Ding XC, Yi GL (2016) Reflection on rice breeding strategy in transitional period of China - developing high quality hybrid rice. Hybrid rice 31(1): 1-5.

23. Tian ZX, Qian Q, Liu QQ, Yan MX, Liu XF, Yan CJ, Liu GF, Gao ZY, Tang SZ, Zeng DL, Wang YH, Yu JM, Gu MH, Li JY (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(51): 21760-21765.

24. Tian ZX, Yan CJ, Qian Q, Yan S, Xie HL, Wang F, Xu JF, Liu GF, Wang YH, Liu QQ, Tang SZ, Li JY, Gu MH (2010) Development of gene-tagged molecular markers for starch synthesis-related genes in rice. Chinese Science Bulletin 55(33): 3768-3777.

25. Wang CL, Chen T, Zhang YD, Zhu Z, Zhao L, Lin J (2009) Breeding of a new rice
variety with good eating quality by marker assisted selection Chinese Journal of Rice Science 23(1): 25-30.

26. Wang WS, Mauleon R, Hu ZQ, et al (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature 557(7703): 43-49.

27. Yang J, Wang J, Fan FJ, Zhu JY, Chen T, Wang CL, Zheng TQ, Zhang J, Zhong WG, Xu JL (2013) Development of AS-PCR marker based on a key mutation confirmed by resequencing of Wx-mp in Milky P rincess and its application in japonica soft rice (Oryza sativa) breeding. Plant Breeding 132(6): 595-603.

28. Yamakawa H, Hirose T, Kuroda M, Yamaguchi T (2007) Comprehensive expression profiling of rice grain filling-related genes under high temperature using DNA microarray. Plant physiology 144(1): 258-277.

29. Zhang CQ, Zhu JH, Chen SJ, Fan XL, Li QF, Lu Y, Wang M, Yu HX, Yi CD, Tang SZ, Gu MH, Liu QQ (2019) Wxlv, the Ancestral Allele of Rice Waxy Gene. Molecular plant 12(8): 1157-1166.

30. Zhu JH, Zhang CQ, Gu MH, Liu QQ (2015) Progress in the Allelic Variation of Wx Gene and Its Application in Rice Breeding. Chinese Journal of Rice Science 29(4): 431-438.

31. Zhang HL, Zhang DL, Wang MX, Sun JL, Qi YW, Li JJ, Wei XH, Han LZ, Qiu ZE, Tang SX, Li ZC (2011) A core collection and mini core collection of Oryza sativa in China. Theoretical and Applied Genetics 122(1): 49-61.

32. Williams PC, Kuzina FD, Hlynka I (1970). Rapid colorimetric procedure for estimating the amylose content of starches and flours. Cereal chemistry.

Figures
Figure 1

Analysis of the allelic variation among the 2,752 Wx alleles. Each vertical line wit
DNA sequence validation and AC measurement. (A-E) Sequencing validation of five non-synonymous mutations.

Figure 3

Crystal structure of rice OsGBSS1. (A) Crystal structure of the OsGBSSI catalytic c
Flowchart of the 30 Wx haplotypes. Different colours represent different varietal groups.
Figure 5

Worldwide distribution of different Wx alleles. Each colour represents a unique Wx allele.
Figure 6

Relationship between the Wx alleles and number of CT repeats. The size of each grey dot represents the amount of each allele that corresponds to different CT repeats.
**Figure 7**

Primer sequences for KASP genotyping of different Wx alleles. Primer X and Primer Y are two allele-specific primers. Allele X and Allele Y represent SNP or InDel at the 3′ end of Primer X and Primer Y, respectively.

| Kasp marker | alleleX/alleleY | PrimerX | PrimerY | PrimerC |
|-------------|----------------|---------|---------|---------|
| K1-Wx<sup>a</sup>/Wx<sup>b</sup> | G/T | 5' CGAGCTGCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' | 5' CGAGCTGCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' | 5' GCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' |
| K2-Wx | G | 5' CGAGCTGCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' | 5' CGAGCTGCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' | 5' GCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' |
| K3-Wx<sup>op</sup> | G/A | 5' CGAGCTGCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' | 5' CGAGCTGCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' | 5' GCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' |
| K4-Wx<sup>op</sup> | A/G | 5' CGAGCTGCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' | 5' CGAGCTGCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' | 5' GCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' |
| K5-Wx<sup>1</sup>/Wx<sup>mw</sup> | A/C | 5' CGAGCTGCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' | 5' CGAGCTGCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' | 5' GCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' |
| K6-Wx<sup>iv</sup> | C/T | 5' CGAGCTGCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' | 5' CGAGCTGCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' | 5' GCCAAGCTGGATTT TCTCGAGGGAAAGAAGACGATG 3' |

**Figure 8**

KASP genotyping of the different Wx alleles in basic parent lines. Scatter plots of
Genotyping of Wx in the hybrid parents. The left column represents the sterile lin

Supplementary Files

This is a list of supplementary files associated with the primary manuscript. Click to download.

Wx allelic variation analysis of 3K RGP sequencing data.xls
Genotyping materials.xlsx
Primer sequence for KASP genotyping.xlsx