Genome-wide association study identifies variation of glucosidase being linked to natural variation of the maximal quantum yield of photosystem II

Saber Hamdani, Hongru Wang, Guangyong Zheng, Shahnaz Perveen, Mingnan Qu, Naveed Khan, Waqasuddin Khan, Jianjun Jiang, Ming Li, Xinyu Liu, Xiaocen Zhu, Govindjee, Chengcai Chu, and Xin-Guang Zhu

Introduction

There are two photosystems in oxygenic photosynthesis: photosystem I (PSI) and photosystem II (PSII); the two together lead to water oxidation and NADP reduction during linear electron transport that was described by using the so-called Z-scheme (see Govindjee et al. 2017). PSII, the focus of this paper, is a membrane-bound protein complex with associated co-factors, which include chlorophyll a (Chl a), plastoquinone (PQ) and manganese (Mn). In cyanobacteria, algae and plants, PSII catalyzes light driven oxidation of water and reduction of PQ (Wydrzynski and Satoh 2005, Nelson and Yocum 2006). On the other hand, PSI, another membrane

The maximum quantum yield of photosystem II (as reflected by variable to maximum chlorophyll a fluorescence, $F_{v}/F_{m}$) is regarded as one of the most important photosynthetic parameters. The genetic basis underlying natural variation in $F_{v}/F_{m}$, which shows low level of variations in plants under non-stress conditions, is not easy to be exploited using the conventional gene cloning approaches. Thus, in order to answer this question, we have followed another strategy: we used genome-wide association study (GWAS) and transgenic analysis in a rice mini-core collection. We report here that four single-nucleotide polymorphisms, located in the promoter region of $\beta$-glucosidase 5 ($BGlu-5$), are associated with observed variation in $F_{v}/F_{m}$. Indeed, our transgenic analysis showed a good correlation between $BGlu-5$ and $F_{v}/F_{m}$. Thus, our work demonstrates the feasibility of using GWAS to study natural variation in $F_{v}/F_{m}$, suggesting that cis-element polymorphism, affecting the $BGlu-5$ expression level, may, indirectly, contribute to $F_{v}/F_{m}$ variation in rice through the gibberellin signaling pathway. Further research is needed to understand the mechanism of our novel observation.

Abbreviations – $BGlu-5$, $Oryza sativa$ 1 $\beta$-glucosidase 5; CaMV, cauliflower mosaic virus; CDS, coding sequence; Chl a, chlorophyll a; CK, cytokinin; ETR, electron transport rate; $F_{b}$, basal (initial, minimal) level of chlorophyll a fluorescence; $F_{m}$, maximal level of chlorophyll a fluorescence; $F_{v}$ ($F_{m}$ minus $F_{b}$), maximal variable chlorophyll a fluorescence; GA, gibberellin; GWAS, genome-wide association study; KO, knocked out; LD, linkage disequilibrium; MAF, minor allele frequency; M-PEA, multi-function plant efficiency analyzer; NPQ, non-photochemical quenching; OJIP transient, chlorophyll a fluorescence induction (where O is for $F_{b}$, P is for peak [equivalent to $F_{m}$ in saturating light] J and I are for inflections between O and P); PQ, plastoquinone; PSI, photosystem I; PSII, photosystem II; QQ, quantile–quantile; SNP, single-nucleotide polymorphism.
protein complex, oxidizes plastoquinol (PQH$_2$) via Cytochrome b$_{6}$f complex (Cramer and Kallas 2016), and reduces NADP (Golbeck 2006). After absorption of light, photosynthetic pigments, in the antenna complexes, are in their excited states, and they lose their energy by one of the three processes: (1) transfer of excitation energy to reaction center chlorophylls, where photochemistry occurs (Mamedov et al. 2015, Mirkovic et al. 2017), leading to photosynthesis (approximately 95%); and (2) loss of energy as heat (Demmig-Adams et al. 2014); and (3) as Chl a fluorescence (Papageorgiou and Govindjee 2004), the latter two together being approximately 5%. The sum of the probabilities of these three processes is equal to 1. The quantum yield Chl a fluorescence from PSI is weak and can be assumed to be constant, whereas that from PSII is strong and varies with time and light intensity. Thus, by measuring Chl a fluorescence (which is mostly from PSII), the yield of PSII photochemistry and heat dissipation in PSII antenna can be estimated (Papageorgiou and Govindjee 2004, Demmig-Adams et al. 2014). Thus, Chl a fluorescence measurements have been widely used in photosynthesis research, because many details related to the photosynthetic apparatus and performance can be obtained from this highly sensitive non-invasive probe (Krause and Weis 1984, Genty et al. 1989, Govindjee 1995, 2004, Xin et al. 2013, Hamdani et al. 2015).

In higher plants and algae, Chl a fluorescence induction curve, measured under high-intensity continuous light after a brief period of darkness, shows a characteristic OJIP induction curve (see e.g. Stirbet and Govindjee 2011). Several parameters have been derived from these fluorescence signals, among which the $F_v/F_m$ ratio, where $F_v$ (maximal variable fluorescence) is the difference between $F_m$, i.e. the maximal fluorescence level at the P level, and the $F_D$, i.e. the minimal fluorescence level, has been widely used in photosynthesis research to provide information on the overall performance of plants. The $F_v/F_m$ provides an estimate of the maximum quantum yield of PSII photochemistry (Butler 1978, Govindjee 2004). Therefore, any drop in this ratio, e.g. because of different stresses, can be interpreted as a decrease in PSII activity (Srivastava et al. 1997, Araus et al. 1998, Li et al. 2006, Faraloni et al. 2011). However, under unstressed conditions, $F_v/F_m$ is highly conserved and usually has a value around 0.8 (Björkman and Demmig 1987).

Although there have been a large number of physiological and biophysical studies on changes in $F_v/F_m$, especially under stress, there have been very few reports on genetic mechanisms affecting natural variation on $F_v/F_m$ (Yin et al. 2010). This can be attributed to the difficulty of efficiently pinpointing the genes affecting variation on $F_v/F_m$ using the traditional forward genetics approach. The genome-wide association study (GWAS), however, facilitated by the next-generation sequencing technology provides a unique opportunity to rapidly bridge the gap between the natural variation of phenotypic parameters with the associated genetic variation. This method has been successfully used for identifying genes that underlie natural variation of various ecological and agricultural traits (see e.g. Huang et al. 2010, Kump et al. 2011, Tian et al. 2011, Zhao et al. 2011, Tian et al. 2011, Huang et al. 2012). GWAS techniques have already been recognized as a useful tool to study natural variation of photosynthetic parameters (Flood et al. 2011, Lawson et al. 2012). Along this line, Flood et al. (2016) surveyed natural variation of photosynthetic parameters such as light use efficiency of PSII electron transport in *Arabidopsis thaliana* (Arabidopsis). In sorghum, Ortiz et al. (2017) have identified genetic regions and candidate genes that are related to the capacity of photosynthesis to cope with excess light energy under low temperature during the vegetative growth stage. Recently, Liu et al. (2018) identified potential genes related to net photosynthetic CO$_2$ uptake in soybean.

In the present study, by using a diverse rice collection, we first measured $F_v/F_m$ under different conditions. Then, using GWAS, we identified single-nucleotide polymorphisms (SNPs) that were associated with variation in $F_v/F_m$. Finally, using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 and overexpression technologies, we suggest here that *Oryza sativa* 1 β-glucosidase 5 (BGlu-5) may contribute (or be related) to natural variation in $F_v/F_m$ presumably through cis-element polymorphism affecting its expression level. We also propose a mechanism linking the gibberellin (GA) signaling pathway with the natural variation of $F_v/F_m$. Further experiments and models are needed to understand the full potential of our novel observation.

**Materials and methods**

**Growth conditions**

We used a rice mini-core panel, which includes 199 rice accessions of different geographic origins (Agrama et al. 2009). Seeds were grown into seedlings for approximately 30 days in a soil seed bed located in Beijing (China) (39°55′N, 116° 25′E), and then the seedlings were transplanted into 12-l plastic pots containing commercial peat soil (Pindstrup Substrate 4). Chl a fluorescence measurements ($F_v/F_m$) were made on leaves of rice plants from July to the 1st week of August 2013. Furthermore, we re-planted the entire rice population in the field from June to September 2013 and...
measured $F_v/F_m$ in the field again from early August till September of that year. Finally, we chose 16 rice accessions that showed either high or low $F_v/F_m$ values and grew them in controlled chambers under 14/10 h light/dark regime, at approximately 26°C, and a light intensity of approximately 500 μmol photons m$^{-2}$ s$^{-1}$ (during the day) given at the top of the plants. For these plants, leaves were sampled after 40 days of growth and used for quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis. Two rice GA (gibberelin) mutants, i.e. sd1 (GA-deficient, here called low GA) and eui1-4 (GA-accumulated, here called high GA), were obtained from the authors of Luo et al. (2006) and Tong et al. (2014), and grown in controlled growth chambers for 30 days under the above-mentioned conditions.

$F_v/F_m$ measurements

The ratio of variable fluorescence $F_v$ (the maximum, $F_m$, minus the minimum $F_0$) to $F_m$, i.e. $F_v/F_m$, was recorded using the multi-function plant efficiency analyzer (M-PEA) (Hansatech). For the $F_v/F_m$ measurements, plants were first kept overnight at 24°C in darkness, and then their detached leaves were kept under complete darkness for 10 min, using leaf clips; this was done to avoid any potential light that the plant may receive before experiments. After a 10-min dark adaptation, the upper side of the detached leaves was exposed for 0.5 s with a saturating orange-red (625 nm) actinic light (5000 μmol photons m$^{-2}$ s$^{-1}$), obtained from light-emitting diodes (LEDs). Measurements were repeated 3 to 4 times for each accession. In contrast to the above experiments on plants in pots, for field experiments, a pocket PEA chlorophyll fluorometer (Hansatech) was used to measure $F_v/F_m$. Leaves were first kept in darkness for 10 min, and then exposed for 1 s to white light (3000 μmol photons m$^{-2}$ s$^{-1}$) obtained from LEDs.

One caveat we need to emphasize here is that because the pocket PEA is not equipped with far-red light source, which is usually applied to completely oxidize the PQ pool and hence obtain the ‘correct’ $F_v/F_m$ value, we have used the basic setting for both ‘in-pot’ and ‘in-field’ experiments, i.e. without prior far-red light exposure. In order to test the effect of far-red pre-illumination on the $F_v/F_m$ value, the wild-type ZH11 was treated with and without far-red pre-illumination. Our result shows that although there was a difference in the absolute values of $F_v/F_m$ between measurements using these two conditions; however, the trend of $F_v/F_m$ variation remained relatively the same (Fig. S1).

Fluorescence quenching analysis

Measurements for fluorescence quenching analysis were carried out using a Dual-PAM-100 instrument (Walz). We used plants that were moved from the growth chamber to the analysis chamber and kept overnight at 24°C in darkness. Then, after a 10-min dark adaptation period, the upper side of the detached leaves was exposed to a weak modulated measuring light for $F_0$ measurements. A saturating light pulse of 500 ms duration and a Photosynthetic Photon Flux Density of 20 000-μmol photons m$^{-2}$ s$^{-1}$, obtained from 620-nm LED arrays, were given to the sample to measure $F_m$. After the 1st saturating pulse, a red actinic light of 900 μmol photons m$^{-2}$ s$^{-1}$ was switched on for 10 min followed by a 10-min dark period. During this time, saturating pulses of light were applied every 20 s to measure $F_m'$ (i.e. $F_m$ in light). In order to obtain $F_m''$, red actinic light was switched off and far-red (720 nm) light was applied. The effective PSII quantum yield, $Y_{II}$, the non-photochemical quenching (NPQ), and the electron transport rate (ETR) were calculated, as described by Kramer et al. (2004) and Klughammer and Schreiber (2008).

RNA isolation and real-time RT-PCR analysis

Total RNA was extracted from mature rice leaves with a Purelink RNA Mini Kit (Invitrogen) following manufacturer’s instructions. RNA samples were reverse transcribed into cDNA using SuperScript III Reverse Transcriptase (Invitrogen) according to manufacturer’s instructions. The qRT-PCR was carried out using the LightCycle480 System (Roche Applied Science). Transcript levels were normalized to that of UBIQUITIN 5 (UBQ5, AK061988) (internal control). Each PCR reaction was performed in triplicate, and the error bars represent the standard deviation from three PCR reactions. Statistical analyses involved the 2$^{-\Delta\Delta C T}$ method (Livak and Schmittgen 2001). (For the primer sequences used, see Table S1).

GA measurement

Plants were sampled after 50 days of growth. About 0.5 g of leaves was collected for GA measurement (see Hedden and Sponsel 2015). Quantification of endogenous GA3 was performed, as described by Chen et al. (2012).

Genome-wide association study

The accessions of the rice mini-core diversity panel were genotyped by whole-genome sequencing, as described by Wang et al. (2016). In total, 2.3-M biallelic SNPs with minor allele frequency (MAF) ≥ 5% were used for association mapping. In our study, we first normalized
the phenotype data with the quantile-quantile (QQ) plots norm function, using the R software. Then, we performed the GWAS analysis employing linear mixed model implemented in Genome-wide Efficient Mixed Model Analysis (GEMMA) (Wang et al. 2016). To further correct for population structure and reduce false positives, we did Principal component analysis (PCA) analysis on the population, and supplied the 1st four principle components as covariates in the GWAS analysis, as described by Wang et al. (2009). We used a permutation strategy to decide on the genome-wide significance level. After 200 repetitions, a P-value of 10^{-6} was chosen as the genome-wide significance threshold; at this threshold, the false discovery rate is less than 5%. Linkage disequilibrium (LD) analysis was conducted using the Haploview software version 4.2, and the LD blocks were created when the upper 95% confidence bounds of D’ value exceeded 0.98 and the lower bounds exceeded 0.70, as described by Gabriel et al. (2002).

Plasmid construction for BGlu-5 overexpression

The promoter of cytosolic FBPase (FBPpro) is known to guide the expression of foreign genes only in the mesophyll cells of the leaf blades and leaf sheaths. This promoter was cloned into the binary plant transformation vector pHMS with single PsI clone site. The BGlu-5 gene was amplified from the Nipponbare rice variety using primers listed in Table S1. The product was then ligated downstream of the FBPpro with BamHII-Xmal clone sites. The generated construct was verified by DNA sequencing, and then transformed into the Agrobacterium tumefaciens strain EHA105. The Nipponbare was transformed by the Agrobacterium-mediated plant tissue culture method.

Plasmid construction using the CRISPR/Cas9 technique

The codon-optimized hSpCas9 (Cong et al. 2013) was linked to the maize ubiquitin promoter (UBI) and then inserted into binary vector pCAMBIA1300, which contains the hygromycin B phosphotransferase gene under the control of cauliflower mosaic virus (CaMV) 35S promoter. Then, a fragment containing an OsU6 promoter (Feng et al. 2013), a negative selection marker gene ccdB, flanked by two Bsal restriction sites, and a single-guide RNA derived from pX260 (Cong et al. 2013), was inserted into this vector using in-fusion cloning kit (Takara) to produce the CRISPR/Cas9 binary vector pBGK032. The sequence targetting the coding sequence (CDS) region of BGlu-5 was synthesized and annealed to form the oligo adaptor. Then, the vector pBGK032 was digested by Bsal and purified using DNA purification kit (Tiangen). Finally, a ligation reaction (10 μl) containing 10 ng of the digested pBGK032 vector and 0.05-mM oligo adaptor was carried out and directly transformed in E. coli competent cells to produce CRISPR/Cas9 plasmid (Fig. S2).

Plant transformation and mutation detection

The CRISPR/Cas9 plasmid was introduced into the A. tumefaciens strain EHA105, and the transformation of rice was done, as previously described (Nishimura et al. 2006). Genomic DNA was extracted from the transgenic plants, and the primer pairs flanking the designed target site were used for PCR amplification. The PCR products (300–500 bp) were sequenced and identified using the degenerate sequence decoding method (Ma et al. 2015).

Subcellular localization of BGlu-5

To determine the localization of BGlu-5, we generated BGlu-5-GFP fusion protein by tagging a green fluorescent protein (GFP) in the C-terminus and transiently transformed the wild tobacco (Nicotiana benthamiana). The CDS region of BGlu-5 was first amplified from a rice (Nipponbare) cDNA, and then inserted into the binary vector pCAMBIA1302 with a CaMV 35S promoter via a homologous recombination-based seamless cloning method (GBClonart). The final plasmid was verified by sequencing. Then, the vector was transformed into A. tumefaciens strain GV3101. The Agrobacterium cells were first cultured and resuspended to an Optical Density (at 600 nm) of approximately 1.0 in a buffer (10 mM MES, 10 mM MgCl2 and 200 mM acетosyngone, pH 5.7), and then infiltrated into young leaves by a syringe. After 36 to 48 h, the fluorescence signals in leaf pavement cells were analyzed with a Zeiss LSM880 two-photon confocal microscope. Chl a fluorescence was detected in the 650- to 720-nm range and used as a marker for the chloroplast.

Results

Fv/Fm measurement

In this study, a diversity panel of rice (O. sativa) mini-core accessions collected from different geographic regions was used, as grown under two different conditions from May to September 2013 (in Beijing, China). These two conditions were field growth (i.e. in-field) and outside pot growth (in-pot) condition. As mentioned in the Material and methods, Fv/Fm values were measured, using an M-PEA (Hansatech) for in-pot condition, and a pocket PEA chlorophyll fluorometer (Hansatech) for in-field condition. Under both the conditions, rice leaves at the
late vegetative stage to early booting stage were used. Fig. 1A shows that $F_{v}/F_{m}$ values, measured under an in-pot condition, followed a normal distribution with a median of 0.835 (at around 65% of the population) and varied from 0.82 to 0.85 across the entire rice mini-core collection. The *Japonica* varietal group (TRJ, tropical; TEJ, temperate; and ARO, aromatic) exhibited slightly higher $F_{v}/F_{m}$ as compared with the *Indica* varietal group (IND, Indica; and AUS, aus) and Admix (Fig. 1B). In order to test the accuracy of our measurements, we re-planted the mini-core collection under in-field condition from June to September 2013 (in Beijing, China). Fig. 1C shows that around 80% of the population exhibited higher $F_{v}/F_{m}$ than the median ($F_{v}/F_{m} = 0.725$). However, the box plot for the variation of $F_{v}/F_{m}$ across subpopulations showed a highly similar trend as compared to those under in-pot condition ($R = 0.39$; ***$P < 0.001$), i.e. the $F_{v}/F_{m}$ ratios observed in *Japonica* remained slightly higher than in *Indica* and Admix (Fig. 1D).

**GWAS on $F_{v}/F_{m}$**

In order to identify candidate genes contributing to natural variation in $F_{v}/F_{m}$, the accessions in the rice mini-core diversity panel were genotyped by whole-genome sequencing, and a total of 2.3-M biallelic SNPs with MAF $\geq 5\%$ were used for association mapping. Using a linear mixed model implemented in GEMMA, we identified several SNPs that were highly associated with $F_{v}/F_{m}$ variation ($P$-value $<10^{-6}$) (Fig. 2A). The QQ plot analysis indicates that the identified SNPs represent true positive discoveries, because the observed $-\log_{10} P$-values were higher than the expected distribution, especially in the most significant $P$-value range (Fig. 2B). The most significantly associated SNP (called here SNP1), which explains approximately 28% of the total phenotypic variation, was found at Chr 1: 40840545, with a $P$-value of 7.3E$^{-08}$, although there were also a number of other SNPs showing high significance levels ($P$-values $<10^{-6}$) (Table S2). In the current study, we focused only on SNP1 and the genes in its vicinity because it shows the highest statistical significance level (Table S2). All the candidate genes in the vicinity of this SNP and also other potential SNPs are listed in Table S2 according to the Michigan State University (MSU) rice genome database (http://rice.plantbiology.msu.edu/).

Using LD analysis covering 100-kb interval centered around SNP1, we identified one LD block of 11 Kb (Chr1:
Fig. 2. Legend on next Page.
40835717 to 40846995) harboring SNP1. In addition, three significant SNPs, i.e. SNP2 (Chr1: 40841210), SNP3 (Chr1: 40841265) and SNP4 (Chr1: 40842087) with a P-value of 8.32E-07 were located in the vicinity of SNP1 in the same LD block (Fig. 2C), providing additional evidence for this region to be a major locus associated with \( F_r/F_m \) variation. Interestingly, only one gene encoding a BGlu-5 (LOC_01g70520) was found in this LD block (Fig. 2C), suggesting that this gene is a strong candidate contributing to variation in \( F_r/F_m \), without excluding other candidate genes not located in this LD block. These other genes will be examined in the future. Furthermore, the allelic diversity of the above-mentioned SNPs across rice mini-core collection revealed two different BGlu-5-haplotypes, i.e. BGlu-5-haplotype 1 (containing major alleles: CTTA), which was represented in approximately 93% of the population, and BGlu-5-haplotype 2 (containing minor alleles: TGCG), which occurred in approximately 7% of the population (Table S3). Interestingly, we found that BGlu-5-haplotype 1 was only detected in Japonica; however, BGlu-5-haplotype 2 was detected in both Indica and Admix (Fig. 2D and Table 1). In order to check the haplotype frequency and distribution of large-scale rice genome, we further analyzed the 3000 rice genomes project. Our result shows that BGlu-5-haplotype 1 was represented in approximately 97% of the population, while BGlu-5-haplotype 2 occurred in approximately 1% of the population. Consistent with the result obtained with the rice mini-core collection, BGlu-5-haplotype 2 was only detected in the Indica varietal group (not in Japonica) and Admix (Tables S4 and S5). Despite the wide geographic distribution of the rice mini-core accessions, accessions with BGlu-5-haplotype 1 showed an independent origin as compared with those with BGlu-5-haplotype 2. In fact, accessions with haplotype 2 mostly exist in the Indian subcontinent (South Asia), while accessions with haplotype 1 exist in all the continents (Fig. S3). This localized distribution of two BGlu-5-haplotypes may reflect local adaptation of photosynthetic systems to variation in environmental factors across different latitudes or altitudes. On the other hand, based on the exact chromosomal position of the above-mentioned SNPs related to BGlu-5, provided by MSU rice genome database, we found that these SNPs were located in the promoter region of BGlu-5 at 1700, 1035, 980 and 158 bp from the ATG starting codon (Fig. 2E). Furthermore, we found that the correlation coefficients (\( R^2 \)) between these SNPs across the mini-core diversity panels varied from 92 to 100%, suggesting that these SNPs are likely inherited together (Fig. 2E).

### Table 1. Allelic diversity of BGlu-5-haplotypes across rice mini-core collection and the corresponding \( F_r/F_m \) trait.

<table>
<thead>
<tr>
<th>Haplotype Subgroup</th>
<th>SNP1</th>
<th>SNP2</th>
<th>SNP3</th>
<th>SNP4</th>
<th>( F_r/F_m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haplotype 1 (n = 186) Japonica</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>A</td>
<td>0.837 ± 0.005</td>
</tr>
<tr>
<td>Haplotype 2 (n = 13) Indica/Admix/Aus</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>0.829 ± 0.005</td>
</tr>
</tbody>
</table>

Fig. 2. GWAS analysis and characterization of \( F_r/F_m \) trait. (A) Manhattan plot for \( F_r/F_m \) variation in a population of rice mini-core grown under in-pots condition. x-axis shows the SNPs along each chromosome; y-axis is the \( \log_{10}(P\text{-value}) \). Blue horizontal dashed line indicates a genome-wide significance threshold of \( \log_{10}(P\text{-value}) = 6 \). (B) QQ plot of the linear mixed model for \( F_r/F_m \) variation. x-axis shows the expected \( \log_{10}(P\text{-value}) \); y-axis is the observed \( \log_{10}(P\text{-value}) \). (C) Zoom view in the Manhattan plot of the genomic region harboring the highest scoring SNP (SNP1), located at Chr1: 40840545. The positions of SNP1 (highlighted with red color), SNP2 (Chr1: 40841210), SNP3 (Chr1: 40841265) and SNP4 (Chr1: 40842087) are indicated by blue dashed arrows. The pair-wise LD between SNP markers, covering 100-kb interval centered on SNP1, is indicated as \( D' \) values. Dark red for a value of 1 and white for 0. The black dashed line indicates the LD block that contains significant SNPs. The location of the candidate gene (BGlu-5) is indicated by blue box. (D) Maximum quantum yield of PSII, \( F_{v}/F_{m} \), measured in rice mini-core accessions. Black histogram is for BGlu-5-haplotype 1, while blue histogram is for BGlu-5-haplotype 2. The population size (n) is shown in the figure. Values are shown as means ± SD. (n = 186 for haplotype 1; n = 13 for haplotype 2). The P-value from a t-test is shown in the figure. (E) Schematic diagram of the BGlu-5 gene structure. BGlu-5-haplotype 1, containing the major alleles (T, C, T and A) of SNPs is shown in the upper panel, while BGlu-5-haplotype 2, containing the minor alleles (T, G, C and G) is shown in the bottom panel. The distances between SNPs and the ATG start codon are indicated in the figure. The promoter is indicated by green box, the CDS is indicated by blue arrow and the terminator (term.) is indicated by red box. The correlation coefficient (\( R^2 \)) of each pair-wise comparison of SNPs is indicated. The number in the square shows the 100-fold value of \( R^2 \), which ranged from 0 to 100. The location of SNPs in the promoter is indicated by dashed blue line.
Transcription analysis of *BGlu-5*

The above results suggest that the SNPs in the promoter region of *BGlu-5* might be related to variation of Fv/Fm. One possible mechanism is that these variations of SNPs lead to different expression levels of *B-Glu5* and later changes in Fv/Fm. To test this possibility, we evaluated the correlation between the expression levels of the *BGlu-5* gene and Fv/Fm. Specifically, we chose eight rice accessions, which had drastically different Fv/Fm values under different growth conditions, i.e. in-pots, in-field and in-controlled-growth chambers (Table S3). Our data show a significant difference in the *BGlu-5* expression level (P-value = 4.8E-4) between *BGlu-5*-haplotypes: haplotype 1 showed a 3.4-fold higher expression as compared with haplotype 2 (Fig. 3A). Fig. 3B shows that the relative expression level of *BGlu-5* was significantly correlated with variation in Fv/Fm in rice grown in controlled growth chamber (R^2 = 0.51, P-value = 0.002). This result shows a correlation between *BGlu-5* transcription level and Fv/Fm, suggesting that this gene probably contributes to Fv/Fm variation in rice mini-core collection.

**Transgenic analysis and Chl a fluorescence measurement**

To further evaluate the correlation between *BGlu-5* expression and Fv/Fm, both *BGlu-5* overexpression and *BGlu-5* knockout, using CRISPR/Cas9 technologies, were used. To do this, we first developed a T0 rice plant overexpressing *BGlu-5*. Fig. 3C shows that overexpressing lines exhibited different transcriptional levels of *BGlu-5* ranging from 10-fold downregulation to 8-fold upregulation as compared to the wild-type Nipponbare. The transcript levels of *BGlu-5* were positively correlated with changes in Fv/Fm (R^2 = 0.48, P-value = 0.05) (Fig. 3D). We further obtained homozygous mutant lines for which the *BGlu-5* gene was knocked out (KO), i.e. *bglu-5*. The
Characterization of rice bglu-5 knockout mutants. (A) Maximum quantum yield of PSII, $F_v/F_m$, measured in rice bglu-5 knockout mutants (blue histogram) and the wild-type ZH11 (black histogram). The P-value from a t-test, for $F_v/F_m$ measurements, is shown in the figure. (B) Effective quantum yield of PSII, $Y(II)$, measured in rice bglu-5 knockout mutants (blue) and the wild-type ZH11 (black). (Inset) Steady state of ETR measured in rice bglu-5 knockout mutants (blue histogram) and the wild-type ZH11 (black histogram). (C) Kinetics of maximum Chl fluorescence under light ($F_m'$) normalized to $F_m$, measured in rice bglu-5 knockout mutants (blue) and the wild-type ZH11 (black). Values are shown as means ± sd. (n = 9). The asterisks indicate that the comparison between two groups is significant by Student's t-test: *$P < 0.05$ and **$0.001 < P < 0.01$. Arrows indicate switching on (↓) and off (↑) of the red actinic light.

CDS of $BGlu-5$ was changed by a 1-bp insertion in all KO transgenic lines (Fig. S4), leading to a frame shift and subsequent loss of $BGlu-5$ function. As expected, we found that the KO transgenic lines exhibited a lower $F_v/F_m$ (0.77 ± 0.01) as compared to the wild-type plants (0.8 ± 0.02) (Fig. 4A). The small (~4%), but significant, decrease in $F_v/F_m$ was quite similar to that observed in $BGlu-5$-haplotypes in rice mini-core collection (Fig. 2D), providing additional evidence that $BGlu-5$ contributes, in some way, to the observed variation in $F_v/F_m$.

Furthermore, KO transgenic bglu-5 plants showed a decrease in the effective PSII quantum yield ($Y(II)$), during both light exposure and recovery in the dark, as well as in the ETR as compared to $Y(II)$ in the wild-type ZH11 plants (Fig. 4B and inset). This result suggests that $BGlu-5$ may influence the electron transport activity in PSII. Moreover, we showed that $F_m'$, i.e. $F_m$ under light, was slightly higher, in the bglu-5 lines as compared to that in the wild type, during light exposure (Fig. 4C); the speed of $F_m'$ recovery after leaves were switched to low light was also slower (Fig. 4C). All these results are consistent with the lower NPQ and slower NPQ induction during light, and slower rate of NPQ decrease in the KO transgenic lines as compared to that in the wild type (Fig. 4D). Such observed changes in fluorescence parameters suggest that the bglu-5 KO transgenic lines are deficient in both the induction and also the recovery of photoprotection.

Subcellular localization of $BGlu-5$

We examined the subcellular localization of $BGlu-5$ with a $BGlu-5$-GFP fusion protein using pavement cells from a wild tobacco ($N. benthamiana$) plant. Our data show that $BGlu-5$-GFP fusion protein is localized in both the nucleus and the cytoplasm (Fig. 5). One major feature conclusion that can be deduced from this result is that $BGlu-5$ can only regulate $F_v/F_m$ indirectly, because it is not in the chloroplast.

Role of GA in $F_v/F_m$ regulation

$BGlu-5$ is a member of the $BGlu$ family, which is involved in key plant developmental processes (Kleczykowski
Fig. 5. Localization BGlu-5. Subcellular localization of BGlu-5-GFP in the nucleus and the cytoplasm, shown in leaf pavement cells of Nicotiana benthamiana. Chlorophyll (middle panel) shows the chlorophyll a fluorescence detected at 650–720 nm, which is considered as a marker for the chloroplast. Scale bar, 50 μm.

and Schell 1995), and activates phytohormones through the hydrolysis of glucoside bonds of the hormones (Wiese and Grambow 1986). Among several phytohormones targeted by BGlu, we studied three major hormones [auxin (IAA, indole acetic acid), cytokinin (CK) and GA], which are known to regulate photosynthesis, albeit indirectly (Zubo et al. 2008, Chapman and Estelle 2009, Xie et al. 2016). Fig. 6A shows changes in $F_v/F_m$ in response to short-term hormone treatment in the bglu-5 mutant. Results show that $F_v/F_m$ in the bglu-5 was restored to the wild-type level by GA treatment; however, the $F_v/F_m$ in bglu-5 was not altered when plants were treated with either IAA or CK (Fig. 6A). To further investigate the role of GA in $F_v/F_m$ regulation, we measured $F_v/F_m$ in several rice GA mutants, including sd1 (GA-deficient, called here low GA) and eui1-4 (GA-accumulated, called here high GA) and the wild-type ZH11. Fig. 6B shows a significant (~12%) decrease in $F_v/F_m$ in low GA mutants (0.7 ± 0.05) as compared to that in the wild type (0.79 ± 0.02), while no change in $F_v/F_m$ was observed in high GA mutants.

Furthermore, we measured GA levels in different rice accessions carrying either functional or non-functional BGlu-5. Fig. 6C shows higher GA levels in plants with BGlu-5-haplotype 1 and the wild-type ZH11 (functional BGlu-5; BGlu-5-haplotype 1) [0.11 ± 0.001 ng g$^{-1}$ FW and 0.07 ± 0.001 ng g$^{-1}$ FW, respectively], as compared to BGlu-5-haplotype 2 and bglu-5 mutant (non-functional BGlu-5) [0.025 ± 0.002 ng g$^{-1}$ FW and 0.023 ± 0.0005 ng g$^{-1}$ FW, respectively]. Based on the above results, we suggest that variation of BGlu-5 gene expression and correspondingly GA variation must have contributed, although indirectly, to the natural variation in $F_v/F_m$ (Fig. 6D).

Discussion

In this study, we surveyed natural variation of $F_v/F_m$ in a rice mini-core diversity panel, which includes accessions collected from different geographic origins. The variable to maximum chlorophyll a fluorescence, $F_v/F_m$, representing the maximal quantum yield of PSII photochemistry, is widely used in photosynthesis research, providing information on photosynthetic systems under both normal and stress conditions (see e.g. Yamane et al. 1998, Garg et al. 2002). However, given the high level of conservation of the ‘light reactions’ in higher plants (El-Lithy et al. 2005, Leister 2012), only a small variation of $F_v/F_m$ exists under unstressed (normal) conditions, which hinders the progress of genetic research aiming at identifying new genes affecting changes in $F_v/F_m$. However, our current study shows that GWASs provide a unique opportunity to ‘mine’ genes controlling $F_v/F_m$. In this study, we have shown that there is a small but significant variation in $F_v/F_m$ across different rice subpopulations; specifically, we found a slightly higher $F_v/F_m$ in Japonica as compared to Indica and Admix varieties (Fig. 1A, B). This difference in $F_v/F_m$ between the subpopulations was confirmed in three separate experiments, i.e. in-pots, in-field and in-growth Physiol. Plant. 2019
BGlu-5 regulates the stability of the photosynthesis capacity in plants (Fig. 4). We found that BGlu-5 was slightly lower in the bglu-5 KO transgenic plants as compared to the wild type (Fig. 4A), showing similar trend as expected for the BGlu-5-haplotypes in the rice mini-core accessions. The homozygous bglu-5 KO plants were sprayed with 40 μM of cytokinins, and their growth was measured (n = 8 for the remaining samples). The qRT-PCR analysis of mini-core accessions and overexpressing lines showed that BGlu-5 had different transcriptional levels, which were correlated with the values of Fv/Fm (Fig. 3), suggesting that BGlu-5 must have contributed, although indirectly, to natural variation of Fv/Fm.

To further confirm the correlation between BGlu-5 and Fv/Fm, variation, we had used a CRISPR/Cas9 technique to knock out BGlu-5 in ZH11, a line commonly used in rice genetics studies. The homogygous bglu-5 KO plants contain a single-nucleotide insertion in the CDS region (Fig. S4), leading to a frame shift and subsequent loss of BGlu-5 function. As expected, Fv/Fm was slightly lower in the bglu-5 KO transgenic plants compared to the wild type (Fig. 4), showing similar trend as expected for the BGlu-5-haplotypes in the rice mini-core accessions. The above-mentioned SNPs were strongly linked together, forming two BGlu-5-haplotypes across rice mini-core accessions. Furthermore, we found that all the four SNPs mentioned above were located in the promoter region of BGlu-5 (Fig. 2E), suggesting that the observed polymorphism can influence the expression level of this gene and later influence the values of Fv/Fm. The qRT-PCR analysis of mini-core accessions and overexpressing lines showed that BGlu-5 had different transcriptional levels, which were correlated with the values of Fv/Fm (Fig. 3), suggesting that BGlu-5 must have contributed, although indirectly, to natural variation of Fv/Fm.

To further confirm the correlation between BGlu-5 and Fv/Fm, variation, we had used a CRISPR/Cas9 technique to knock out BGlu-5 in ZH11, a line commonly used in rice genetics studies. The homozygous bglu-5 KO plants contain a single-nucleotide insertion in the CDS region (Fig. S4), leading to a frame shift and subsequent loss of BGlu-5 function. As expected, Fv/Fm was slightly lower in the bglu-5 KO transgenic plants compared to the wild type (Fig. 4), showing similar trend as expected for the BGlu-5-haplotypes in the rice mini-core accessions. The above-mentioned SNPs were strongly linked together, forming two BGlu-5-haplotypes across rice mini-core accessions. Furthermore, we found that all the four SNPs mentioned above were located in the promoter region of BGlu-5 (Fig. 2E), suggesting that the observed polymorphism can influence the expression level of this gene and later influence the values of Fv/Fm. The qRT-PCR analysis of mini-core accessions and overexpressing lines showed that BGlu-5 had different transcriptional levels, which were correlated with the values of Fv/Fm (Fig. 3), suggesting that BGlu-5 must have contributed, although indirectly, to natural variation of Fv/Fm.

To further confirm the correlation between BGlu-5 and Fv/Fm, variation, we had used a CRISPR/Cas9 technique to knock out BGlu-5 in ZH11, a line commonly used in rice genetics studies. The homozygous bglu-5 KO plants contain a single-nucleotide insertion in the CDS region (Fig. S4), leading to a frame shift and subsequent loss of BGlu-5 function. As expected, Fv/Fm was slightly lower in the bglu-5 KO transgenic plants compared to the wild type (Fig. 4), showing similar trend as expected for the BGlu-5-haplotypes in the rice mini-core accessions. The above-mentioned SNPs were strongly linked together, forming two BGlu-5-haplotypes across rice mini-core accessions. Furthermore, we found that all the four SNPs mentioned above were located in the promoter region of BGlu-5 (Fig. 2E), suggesting that the observed polymorphism can influence the expression level of this gene and later influence the values of Fv/Fm. The qRT-PCR analysis of mini-core accessions and overexpressing lines showed that BGlu-5 had different transcriptional levels, which were correlated with the values of Fv/Fm (Fig. 3), suggesting that BGlu-5 must have contributed, although indirectly, to natural variation of Fv/Fm.
BGluS are ubiquitous enzymes involved in multiple physiological functions, including biomass degradation, and hydrolysis of glycolipids and oligosaccharides leading to the release of glucose. In plants, BGluS play an additional role in phytohormone activation by releasing glucose from the conjugated form of a hormone (Brzobohaty et al. 1993), which is considered to be a crucial step controlling many plant growth and developmental processes. We ask: does this mechanism potentially contribute to the variation in \( F_{v}/F_{m} \)? To test the hypothesis that the slight decrease in \( F_{v}/F_{m} \) in \( bglu-5 \) KO transgenic lines was due to decreased level of BGlu-5, which results in lower free-hormone levels, we had supplied exogenous IAA, CK and GA on \( bglu-5 \) for a short-term treatment. Our results show that only GA treatment can recover the \( F_{v}/F_{m} \) value in the \( bglu-5 \) to the WT level (Fig. 6A). Our results are consistent with those of Xie et al. (2016) who had observed that short-term GA treatment in \textit{Populus tomentosa} induces an increase in photosynthesis activity, accompanied by 2- to 7.4-fold up-regulation of genes involved in the ‘light reactions’. Furthermore, the free-GA content was significantly higher in rice accessions with BGlu-5 haplotype 1, as compared to accessions with BGlu-5 haplotype 2 and \( bglu-5 \) KO transgenic lines (Fig. 6C), which is consistent with the hypothesis that the free-GA level might be related to the variation in \( F_{v}/F_{m} \).

Does the observed natural variation of BGlu-5, and correspondingly that of \( F_{v}/F_{m} \), affect plant performance? To answer this question, we studied the relationship between \( F_{v}/F_{m} \) and several agronomic traits, such as biomass, \( CO_{2} \) assimilation and grain yield, using our rice mini-core collection (data not shown). Our results show that \( F_{v}/F_{m} \) was significantly negatively correlated with biomass (Table 2). In other words, the slight decrease in \( F_{v}/F_{m} \), shown in BGlu-5 haplotype 2, was significantly associated with an increase in biomass. It is worth mentioning here that during the green revolution, the mutated SD1 gene, which results in decreased GA level and decreased \( F_{v}/F_{m} \), was artificially selected in rice and barley (Sasaki et al. 2002, Spielmeyer et al. 2002, Jia et al. 2009). Detailed study is needed in the future to elucidate the mechanistic linkage between the decrease in \( F_{v}/F_{m} \) and the increase in biomass production. It is worth mentioning that in the rice accessions used in this study, both in the 199 mini-core accessions and also the 3000 accessions, the BGlu-5 haplotype 1 is the majority of the haplotype (Tables S4 and File S5).

### Conclusions

In the present study, we have applied GWAS as well as a transgenic approach to investigate potential genetic mechanisms underlying natural variation in \( F_{v}/F_{m} \) in a rice mini-core diversity panel. Our data suggest that cis-element polymorphism affecting the \( BGlus \) transcript level might have contributed to variation in \( F_{v}/F_{m} \) through a GA signaling pathway. Although we have focused on SNP1 in this study, the influence of other SNPs and genes on natural variation of \( F_{v}/F_{m} \) cannot be eliminated and will be examined in future studies. Here we emphasize that the GWAS technique can be used as a tool to mine genetic mechanisms controlling extremely conserved physiological parameters, as demonstrated here by the ability of this approach in mining genes controlling \( F_{v}/F_{m} \), which is usually considered to be invariable across higher plants. Therefore, this study provides new support for using GWAS to study large arrays of physiological parameters and processes in plant biology, not only those with large genetic variation, but also those with high level of conservation between plants.

### Author contributions

S.H. and X.-G.Z. conceived and conducted the experiments; S.H. performed most of the experiments; S.P., M.Q., N.K., J.J., M.L., X.L. and X.Z. provided technical assistance to S.H.; S.H., H.W. and G.Z. analyzed the data; G., C.C. and X.-G.Z. supervised the experiments; S.H. wrote most of the article; G., C.C. and X.-G.Z. supervised and complemented the writing.

### Acknowledgement

The authors would like to thank the Chinese Academy of Sciences Strategic Research Project for financial support. Govindjee would like to thank the Department of Plant Biology, of the University of Illinois at Urbana-Champaign (UIUC) for providing office and Lab space. He gives special thanks to all the members of the staff of Information Technology, School of Integrative Biology and Molecular & Cell Biology, UIUC, for their help with the use of computers.
References


study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet 44: 32–39


Wiese G, Grambow HJ (1986) Indole-3-methanol-b-D-glucoside and indole-3-carboxylic acid-b-glucoside are products of indole-3-acetic acid degradation in wheat leaf segments. Phytochemistry 25: 2451–2455


Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Effect of far-red pre-illumination in $F_0$ and $F_v/F_m$ values.

Fig. S2. Structure of the CRISPR/Cas9 binary vector pBGK032.

Fig. S3. Map showing geographical distribution for our rice mini-core collection (199 accessions).

Fig. S4. Sequence alignment of target gene between homozygous mutants (obtained using CRISPR/Cas 9 method) and wild type.

Table S1. Primers used for gene amplification and qRT-PCR analysis.

Table S2. List of significant SNPs and the surrounding genes.

Table S3. Allelic diversities of SNPs and the corresponding $F_v/F_m$.

Table S4. The allelic variations of SNP in 3000 rice accessions.

Table S5. The number of accessions in different haplotypes in the 3000 rice accessions.