

Genome-wide association study for yield and yield related traits under reproductive stage drought in a diverse indica-aus rice panel

Aditi Bhandari

International Rice Research Institute

Nitika Sandhu

International Rice Research Institute

Jérôme Bartholome

CIRAD

Tuong-Vi Cao-Hamadoun

CIRAD

Nourollah Ahmadi

CIRAD

Nilima Kumari

Banasthali Vidyapith

Arvind Kumar (✉ a.kumar@irri.org)

International Rice Research Institute <https://orcid.org/0000-0002-5488-9410>

Original article

Keywords: Reproductive-stage drought, rice, genetic diversity, association mapping, quantitative trait loci

Posted Date: February 24th, 2020

DOI: <https://doi.org/10.21203/rs.2.24272/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Rice on August 6th, 2020. See the published version at <https://doi.org/10.1186/s12284-020-00406-3>.

Abstract

Background Reproductive-stage drought stress is a major impediment to rice production globally. Conventional and marker-assisted breeding strategies for developing drought tolerant rice varieties are being optimized by mining and exploiting adaptive traits, genetic diversity; identifying the alleles and understanding their interactions with genetic backgrounds for contributing to drought tolerance. Field experiments were conducted in this study to identify marker-trait associations (MTAs) involved in response to yield under reproductive-stage drought. A diverse set of 280 indica-aus accessions was phenotyped for grain yield and nine yield-related traits under normal condition and under two managed drought environments. The accessions were genotyped with 215,250 single nucleotide polymorphism markers. Results The study identified a total of 220 significant MTAs and candidate gene analysis within 200kb window centred from GWAS identified SNP peaks detected these MTAs within/ in close proximity to 47 genes, 4 earlier reported major grain yield QTLs and 8 novel QTLs for 10 traits. The significant MTAs were majorly located on chromosomes 1, 2, 5, 6, 11 and 12 and the percent phenotypic variance captured for these traits ranged from 5 to 88%. The significant positive correlation of grain yield with yield-related traits, except flowering time, observed under different environments point towards their contribution in improving rice yield under drought. Seven promising accessions were identified for use in future genomics-assisted breeding program targeting grain yield improvement under drought. Conclusion These results provide a promising insight into the complex-genetic architecture of grain yield under reproductive-stage drought under different environments. Validation of major genomic regions reported in the study can be effectively used to develop drought tolerant varieties following marker-assisted selection as well as to identify genes and understanding the associated physiological mechanisms.

Background

Drought is the one of the most important limiting factors reducing rice yield in the Asian-Pacific region as the current rice production systems rely on an ample supply of water. Drought tolerance is a polygenic trait controlled by few large-effect and many small-effect QTLs/ genes involving various physiological and biochemical pathways affecting plants at different growth stages- seedling, vegetative or reproductive (Price and Courtois 1999; Tripathy et al. 2000; Xu et al. 2011; Nguyen and Bui 2008). Drought is particularly damaging in the reproductive stages, especially during flowering (Venuprasad et al. 2007; Serraj et al. 2009). To reduce yield losses of rice crop in rainfed lowland areas and to increase overall rice production, development of rice cultivars with improved drought tolerance is thereby essential. Epistatic interactions and the effects of genetic background in rice make molecular breeding even more complicated (Yadav et al. 2019). Earlier studies on improving yield under drought suggested use of secondary traits owing to low selection efficiency reported for improved grain yield under drought stress (Rosielie and Hamblin 1981; Blum 1988; Edmeades et al. 1989). During the last decade, studies have documented the effectiveness and the response of direct selection for grain yield under drought in upland rice (Venuprasad et al. 2007) and lowland rice (Kumar et al. 2008; 2009) proving the effectiveness of

direct selection for grain yield under drought and many varieties have been developed utilizing this approach in rice (Kumar et al. 2014).

Different breeding strategies have been developed, including marker assisted backcrossing (Bernardo and Charcosset 2006; Xu and Crouch 2008) with foreground and background selection (Hillel et al. 1990; Hospital et al. 1992), increasing favorable alleles in early breeding generations (Howes et al. 1998; Bonnett et al. 2005), selecting for numerous loci for quantitative traits (Fernando and Grossman 1989; Lande and Thompson 1990) and use of multiple cycles of selection for accumulation of positive alleles contributing to the traits of interest (Zhang et al. 1992). Although these breeding strategies have helped obtain yield gains and ensured both yield and grain quality improvements over existing varieties, there are still challenges to be addressed. Conventional breeding methods are extremely time consuming, on average, it takes up to ten years for elite varieties to be developed and released for cultivation by farmers and these methods do not always facilitate precise transfer of multiple traits like in molecular breeding methods.

Advances in genomics are providing novel tools, technologies and resources to understand the genotypic and phenotypic relationship of particular traits of interest, especially of complex traits. The development of DNA sequencing technologies has set the stage for the technological revolution in genomics by characterising the genomes in view of diversity and function, after significantly reducing the limitations in generation of sequence information. In particular, genomics-assisted breeding is now rapidly and effectively being utilized to improve breeding efficiency (Varshney et al. 2005; 2014; Abbai et al. 2019).

Genome-wide association study (GWAS) is an important tool in genomics-assisted breeding with enormous potential to accelerate breeding for stress tolerance as it shall enable breeders to make selection based on marker-trait associations (MTAs) as a response to combined effect of all favorable alleles. The transfer of well characterized genes/ QTLs in breeding programs for varietal development was initially low as the genomic regions of interest were being identified in biparental populations. Subsequently, identification of genomic regions associated with agronomic traits has been accelerated by association mapping in panels with larger genetic background allowing the use ancestral recombination events, which led to non-random association of alleles at different loci across the genome and that too at a higher mapping resolution than the biparental linkage analysis (Zhu et al. 2008).

Identification of significant associations between genetic markers and phenotypes is dependent on many factors including linkage disequilibrium between the genetic markers being tested for associations and the functional polymorphisms at the population level, population structure and kinship among individuals (Yang et al. 2010; Tian et al. 2011; Frazer et al. 2009; Vissecher et al. 2010). Controlling detection of false positives can be done by either assigning only population structure in General Linear Model (GLM) as covariate variable or by using information of both- population structure and each genotype's genetic effect as covariate variables in Mixed Linear Model (MLM). To tackle computational intensiveness of these models, different algorithms have been developed, namely EMMA (Kang et al. 2008) (Efficient Mixed-Model Association), EMMAx (Kang et al. 2010) (EMMA eXpedited), P3D (Zhang et al. 2010)

(Population parameters previously determined) and FaST-LMM (Lippert et al. 2011) (Factored Spectrally Transformed Linear Mixed Model). However for complex traits, especially under stress conditions, while these models with incorporation of population covariates do control p-value inflation but also weaken the real associations. Strategies have been developed to solve this problem to improve the statistical power of GWAS. These include-

- a. clustering of genotypes into groups by optimizing kinship between genotypes and using genetic values of groups (instead of genotypes) as random effect covariates in the MLM model. This is implemented in Compressed MLM (Zhang et al. 2010) (CMLM) and Enriched CMLM (Li et al. 2014) (ECMLM) models;
- b. using only the associated genetic markers (instead of all) as pseudo quantitative trait nucleotides (QTNs) to compute kinship between genotypes, using p-values from MLM model. Two algorithms based on this strategy assign QTN as correlated with the testing marker when it is either in the 2Mb interval of the testing marker (FaST_LMM) (Listgarten et al. 2012) or is in LD with the testing marker (SUPER – settlement of MLM under progressively exclusive relationship) (Wang et al. 2014);
- c. Farm-CPU is based on the above two strategies while reducing the residual in computing kinship selectively using pseudo QTNS (Liu et al. 2016).

Using the different methods described above, GWAS has been successfully employed in rice for a wide range of traits like yield and yield components (Agrama et al. 2007), harvest index (Li et al. 2012), flowering time (Ordonez et al. 2010) among others. GWAS in diversity panels (unrelated diverse germplasm) including locally adapted breeding material is highly advantageous to breeders (Bernardo 2008) as incorporation of detected beneficial alleles is direct benefit for stress tolerant varietal development (Pauli et al. 2014) as maximum allelic diversity contributing to agronomic traits are identified, exemplified by Huang et al. (2012) for flowering time and grain yield in worldwide rice germplasm collection; Zhao et al. (2011) and Yang et al. (2014) for revealing the rich genetic architecture and natural variants of complex traits .

In the present study, GWAS was performed on yield and nine yield related traits in a diverse set of 280 *indica-aus* accessions with the aim to identify the significant MTAs/ QTLs/ genes to be further used directly to breed reproductive stage drought tolerant rice varieties for different growing environments. In addition, we explored the effects of the model algorithm (MLM, CMLM, SUPER and Farm-CPU) and trait architecture in identifying genomic regions associated with traits of interest across seasons and environments.

Results

Phenotypic and genotypic characteristics of the population

Distribution, heritability and correlation of the measured phenotypic traits

For each experiment across seasons and environments, of the ten traits evaluated, some traits like plant height (PH), flag leaf area (FlgLA), biomass at maturity (BMDW) and number of effective panicles (NBP) were normally distributed while the remaining including days to 50% flowering (DTF), panicle length (PL), grain yield (GY), 100-seed weight (HSW) and spikelet fertility (SPKFT) exhibited skewed distribution owing to both, effect of environment and season (Additional file 1: Figure S1). Overall, the trait distribution range was more in 2015 dry season (DS) than 2014 wet season (WS) and the trait means exhibited seasonal variations (Table 1). Multidimensional analysis of phenotypic data for WS and DS was performed using experiment-wise data, with data projected on the space defined by the first two axes of FDA using the Y_{adj} values for the 10 phenotypic traits. In general, the phenotypic diversity was greater in DS than WS (Fig.1a). Fisher distances were highly significant ($p < 0.001$) between the experiments of WS and DS. The projection of the ten traits across populations revealed different degrees of relatedness between the traits measured at different growth stages of life cycle (Fig.1b). PH, PL, GY, SPKFT and NBP were the major factors affecting the 80% variance explained in both WS and DS. Using mean grain yield as criterion, each experiment was re-classified based on the grain yield reduction compared to the control-lowland-non-stress experiment (Kumar et al. 2009) (Additional file 1: Table S1). Higher grain yield reduction was observed in DS lowland stress experiment (98%) followed by upland stress experiment (94%). Significant effects of drought stress were observed on the DTF as reflected by early flowering. Heritability was in medium and high range for all the 10 traits in the two seasons, with relatively higher trait heritability in DS than WS (Table 1).

Trait correlation within DS and WS was studied for the 10 traits in each season. The correlation between traits within each environment across WS and DS revealed a similar trend of correlation between the measured traits. DTF was significantly negatively correlated to GY in LL_S and UL_S environments (in the range of 0.48-0.72) except for WS where the correlation was negligible (0.07) while this correlation was negligible for LL_N condition (-0.03 in WS and 0.09 in DS) . The grain yield related traits such as NBP, PL, BMDW, HSW and HI showed significant positive correlation with GY across seasons and experiments, for the LL_S and UL_S environments, this correlation was positive (0.06-0.91) while for LL_N condition, the range of this correlation was from 0-0.51 (Fig. 2). Overall, the correlation was strong in DS as compared to WS.

Population characteristics

The density, distribution of allele frequencies and heterozygosity of the working set of 215,250 loci is summarized in Additional file 1: Table S2. For this 215k SNP set, there is an uneven distribution of markers along the genome. Average density of markers per Mb of the genome is 503 SNPs. High-density marker regions were observed on chromosomes 2 and 4, with a magnitude of about 493 and 438 SNPs per Mb and on chromosome 11 with a magnitude of about 1,231 SNPs between 22-27 Mb region. The distribution of markers along each chromosome is depicted as heatmap in Additional file 1: Figure S2.

For the 215k marker set, average H_o at the accession level was 0.86% with a minimum and maximum of 0.4% and 4.81%, respectively. The distribution of observed heterozygous loci (H_o) varied among the 12

rice chromosomes in the working set and with an average of 0.36%, chromosomes with more heterozygous calls on chromosomes 7, 9, 10 and 12 (Additional file 1: Table S2).

Phylogenetic diversity illustrated by the unweighted NJ tree (Fig. 3a) validated the population structure analysis of diversity panel clustering into three main groups (Fig. 3b): Cluster-I predominant with *indica* background (ind1A, ind1B subgroups), Cluster-II *aus* background and Cluster-III with *indica* (ind2, ind3, indx subgroups) genetic background accessions. The ideal K value with the least cross-validation error detected by the population structure analysis was determined as 3 (Fig. 3c). PCA output of R/GAPIT illustrated accessions clustering in 3 distinct groups when plotted against the first two PC components (Fig. 3d). Rapid LD decay to 0.1 was observed around 100 bp (Additional file 1: Figure S3).

Effect of different algorithms on detection of significant trait associations

Four methods were tested for efficacy in detecting significant marker trait associations and Farm-CPU outperformed MLM, CMLM and SUPER in terms of controlling p-value inflation, particularly for severe stress condition, detecting novel MTAs and detecting MTAs overlapping with / in proximity of less than 1Mb with known major effect QTLs/ genes. Comparison of genomic inflation factor between 2 most contrasting algorithms- MLM and Farm-CPU for the data is summarized in Table 3. Genomic inflation factor (λ) and quantile-quantile (qq) plots were used to compare the genome-wide distribution of the p-values with the expected null distribution. The qq plot, a very useful tool to visualize the deviations of the observed p-values from the expected null distribution, is exemplified for grain yield under lowland and upland stress environments in DS (Additional file 1: Figure S4). Both, λ and qq plots reveal that for data from severe stress condition, Farm-CPU outperforms MLM, CMLM and SUPER models to identify the real associations from the 'noise' in the data. From comparison drawn between 2 contrasting algorithms in Table 3, it is evident that Farm-CPU had an efficacy of 72% in controlling the false positives as compared to other models. The λ values for Farm-CPU model range between 0.8-1.02 compared to higher average λ for other models. Results show that population stratification was effectively controlled by models as λ values ranged for all models between 0.8-1.07.

Effect of trait architecture and heritability on MTA identification across seasons and environments

The phenotypic variance (PV) captured for the ten traits reveal that severity of drought stress realized in the experiment and correlation of trait to grain yield underline differences across seasons. For DTF, high heritability and negative correlation to GY under reproductive-stage drought in upland environment and zero to negligibly positive correlation in lowland environment, the PV ranged from 15-29% in the wet season while in the dry season, it was between 19-42% across lowland and upland environments. Similarly, for PH which is another highly heritable trait, the PV ranged uniformly between 34-53% across seasons, environment and stress. However, for GY, highly heritable but polygenic trait, variation between wet and dry season was quite apparent. While PV ranged from 5-17% only in WS, it was in range of 12-55% for DS. Similarly, for some yield related traits, NBP (55-63%), PL (atleast 80% in LL_N and UL_S and 6-29% in LL_S) and FlgLA (atleast 80% in non-stress and WS and less than 15% in DS stress), the model explained significant variance for the traits. However, for traits with either low heritability or narrow range

of phenotypic values like BMDW (6-27%), HI (5-17%), HSW (5-16%) and SPFKT (5-30%), very minimal phenotypic variance was captured by markers across environments and seasons.

Significant MTAs

Several significant MTAs and QTLs were identified in the present study for the 10 studied traits. Among the 220 significant MTAs identified in the study, 95 were associated with grain yield across different environments, seasons and stress levels, 20 with DTF, 34 with PH, 8 with BMDW, 13 with NBP, 25 with HI, 10 with HSW and 20 with SPFKT while no significant association was detected for PL (Fig. 4,5,6; Additional file 1: Table S3, Figure S5 a-g). Among the 19 identified QTLs, 8 QTLs showed consistent effect across seasons and environments under both control non-stress and stress conditions. Three QTLs showed consistent effect under lowland stress, four QTLs under upland stress, four QTLs under both lowland and upland stress across seasons and eight QTLs under both lowland and upland stress in DS only. Significant MTAs were reported for GY on chr 1 and 12 under LL_N, on chr 1, 2, 5, 6, 7, 8, 11 and 12 under LL_S, while on chr 1, 2, 4, 5, 6, 7, 8, 11 and 12 under UL_S. Out of these, consistent across experiment level and combined level were on chr 10 for LL_N, on chr 1, 7, 8 and 12 for LL_S and on chr 2, 5, 7 and 8 for UL_S. In about 0.403 Mb interval region on long arm of chr 1 and 4.27 Mb interval region on long arm of chr 2 found to be associated with GY under non stress and reproductive stage drought stress conditions for both lowland and upland across seasons. Three SNPs in a region of 5.06 Mb interval on long arm of chr 11 reported to be linked with reproductive stage drought stress under lowland and upland conditions across seasons. The 0.941 Mb interval region below centromere on chr 12 showed association with GY under different level of stresses in upland environment. Under different environmental stresses, the MTAs for DTF were reported on rice chr 6 (7611279-7749410 bp, 9539728-10371528), chr 7 (19598023-20159780 bp), chr 11 (6525213-7215940 bp) and chr 12 (7712803-9203018 bp). Comparison of experiment level and combined analysis showed consistent effect of MTAs for DTF on chr 6 (7611279-7749410 bp) under lowland non-stress and on chr 11 (6525213-6602990 bp) for lowland stress. The long arm of chr 1 (33418648-34400345, 37960019-39044781 bp), chr 3 (33600040-33600989 bp), chr 6 (30802585-30807826 bp), chr 9 (13423222-16154337 bp) and chr 11 (20143839-24761315, 25597507-28789891 bp) observed to be associated with plant height trait under different environments, stress levels and seasons, with MTAs on chr 1 consistent at both individual experiment and combined levels. Some SNPs such as S5_352058, S5_4140355, S5_4266313, S8_857745, S9_19316065 and S9_20944019 with very high and almost similar levels of significance were associated with more than one grain yield and yield related traits. However, the SNP S1_3440034 and S12_1642245 were associated with PH and GY, respectively under different environmental stresses. The detailed description on the QTLs, marker-trait associations and the previously reported candidate genes/QTLs located in the genomic regions reported in the present study is presented in Additional file 1: Table S3.

Selection of promising accessions

Seven promising accessions viz Aus 329, Aus 344, Chungur Bali, Dangar, Lalsaita, Para Nelli and Simul Khuri possessing better yield and yield related traits across different seasons under lowland and upland

stresses in combination of the favourable allele for yield and yield related traits were identified (Table 2). The grain yield improvement in selected accessions ranged from 188 to 508 kg ha⁻¹ under lowland stress, to under 403 to 1645 kg ha⁻¹ lowland non-stress and 846 to 1800 kg ha⁻¹ under upland stress over the best performing check in DS.

Discussion

Phenotypic characterisation under different environments

In both the environments i.e. lowland and upland and across seasons, the yields of the accessions were lower under reproductive stage drought stress compared to the control non-stress indicating the severity of the drought stress imposed. Numerous studies point towards negative relation between yield potential and yield under drought and this has been used to establish response indices under different levels of stress severity (Raman et al. 2012; Kumar et al. 2014; Palanog et al. 2014). Positive correlation between moderate and severe stress response indices are informative of the genotypes with yield gain under all stress severities. In our study, different levels of stress severity were observed across seasons. Such differential levels of stress was useful in identifying the potential drought tolerant lines under variable growing environment. Verulkar et al. (2010) documented that yield reduction under reproductive stage drought is significant even at moderate stress severity and even lower under severe stress. The diversity panel used in the study was primarily *indica*-based to identify the donors/accessions that can be directly used in breeding programs targeting grain yield improvement under drought for South-Asian and South-East Asian region.

Effect of GWAS algorithm and trait architecture for MTA validation across seasons and environments

The difference between the different GWAS models— MLM, CMLM, SUPER and Farm-CPU was explored in addition to the effect of trait architecture. PV validation within the diversity panel in our study is affected by trait architecture and seasonal variation, where for example the range of variance captured by MLM-Farm-CPU is narrow for simple quantitative traits like 0.15-0.52 for flowering time and 0.34-0.63 for plant height as compared to that for a complex quantitative trait like GY with range of 0.02-0.55. In our study, the correlation among the two seasons (DS and WS) used in the current study is of lower magnitude which warrants the variable PV for traits across seasons for the methods. Our results draw similar interpretations with recent studies that conclude the effectiveness of multi-locus methods, especially Farm-CPU over single-locus methods (like MLM) for association analysis of traits with either high or low heritability by adequately controlling false positives and negatives, indicated by sharp deviations observed for p-value distribution in qq plots (Xu et al. 2018; Kaler and Purcell 2019).

The significant and positive correlation among the grain yield and yield related traits except DTF and the colocation of MTAs associated with these traits indicates the contribution of grain yield related traits in contributing to yield improvement under drought stress. Most of the important economic traits such as grain yield, grain quality, biotic and abiotic stresses in different crop species are polygenic in nature.

These complex quantitative traits being the focal for the breeding programs, genome wide analysis has proven to be advantageous in capturing the genetic variance of the diverse germplasm, subsequently contributing to improving crop productivity. Identification of marker-trait associations, QTLs, haplotypes, candidate genes and the functional characterization of the identified candidate genes underlying QTLs/genes will help plant breeders to design and develop drought tolerant rice varieties. In the present study, among the detected significant marker-trait associations, some were novel while the others were located near or co-located with the previously reported genes/QTLs.

Recently, GWAS studies conducted on 180 Vietnamese rice landraces identified a total of 17 QTLs associated with vegetative stage drought tolerance under greenhouse conditions (Hoang et al. 2019). Different significant MTAs in the two subpanels of the study, *indica* and *japonica* were detected using mixed model approach with structure control and kinship among the studied landraces. GWAS performed by Subedi et al (2019) reported 37 highly significant MTAs for 20 traits including plant and root morphological traits, nutrient uptake, yield and its components in MAGIC population of 5 diverse parents for increased adaptability in dry direct seeded rice (DDSR) system.

MAS optimization based on significant genomic regions identified

The QTLs; qGY_{1-2} and qPH_{1-2} and the MTAs (S1_37770897 for NBP) mapped on chromosome 1; qGY_{2-2} on chromosome 2; qPH_{9-1} and $qSPKFT_{9-1}$ on chromosome 9 and qGY_{12-1} mapped on chromosome 12 in both the years and environments were located near the earlier reported major grain yield QTLs namely $qDTY_{1.1}$, $qDTY_{2.3}$, $qDTY_{9.1}$ and $qDTY_{12.1}$ respectively. These findings indicate the valid consistency of the effects of drought grain yield QTLs across diverse germplasm. It is important to take note here that the $qDTY_{1.1}$ was reported to have significant effect on the grain yield under control non-stress and reproductive stage drought stress in different genetic backgrounds such as Swarna, IR64, MTU1010 under lowland and upland environments (Vikram et al. 2011; Sandhu et al. 2014; Sandhu et al. 2015). The qGY_{2-1} and qGY_{2-2} reported in the present study were found to be present in the upstream and downstream region of earlier reported $qDTY_{2.3}$, respectively (Sandhu et al. 2014; Palanog et al. 2014). Interestingly, the qGY_{5-2} reported in the present study was located near the earlier reported genes *OsRPK1* gene (Chen et al. 2013 for root development), *OsCCaMK* (Bao et al. 2014 for microbial symbiosis), *OsHAP3B*, *OsTPS1* (Miyoshi et al. 2003 for chloroplast biogenesis), *OsSTN8* (Nath et al. 2013) for protein phosphorylation of photosystem II) and MTAs for nutrient uptake (Sandhu et al. 2019). The colocation of identified QTLs in the present study with the earlier reported genes for root development, photosynthetic traits, and the stress-responsive genes further indicate the complex nature of grain yield traits in addition to the contribution of these traits/genomic regions in enhancing yield under drought. After validation, the identified significant marker-trait associations and the selected promising accessions possessing the QTLs/MTAs could be used further in genomics-assisted breeding program. These selected accessions may provide novel donors in developing drought tolerant rice varieties for variable growing environment.

Conclusions

The diverse indica-aus panel possessing wide range of phenotypic variability combined with the already available genomic information was exploited to identify the MTAs/QTLs associated with grain yield improvement under reproductive stage drought. A total of 220 significant MTAs were detected in the present study. Candidate gene analysis within 200 kb window centred from GWAS identified SNP peaks detected these MTAs within/ in close proximity to 47 reported genes, 4 earlier reported major grain yield QTLs and 8 novel QTLs for 10 traits. 8 QTLs for plant height and grain yield showed consistent effect across seasons and environments under both control non-stress and stress conditions. The significant positive correlation of the grain yield with grain yield related traits was further confirmed with the colocation of QTLs/MTAs associated with these traits. The introgression of the identified QTLs into elite genetic background, functional characterization of candidate genes identified in or near QTLs regions would be the next step in improving grain yield of rice under reproductive stage drought stress conditions. The identified promising accessions may serve as novel donors in drought breeding program targeting grain yield improvement.

Materials And Methods

Plant material and genotypic data

The study used data evaluated for a diverse indica-aus rice panel of 280 accessions, of which 245 represent the four major genetic subgroups belonging to indica genetic background and 35 to aus genetic background (Additional file 1: Table S4). They were selected from the 3000 accessions recently re-sequenced within the framework of the Rice Genome Project (Li et al. 2014), for their potential to breeding programs targeting rainfed lowland and upland drought environments in South and South-East Asia. In the selected panel, 215 accessions are landraces originating mainly from Asia and 65 accessions are improved lines. Seeds of the accessions were obtained from the International Rice Research Institute (IRRI) gene bank.

The genotypic data for the 280 accessions was obtained from the International Rice Informatics Consortium (IRIC) database for the 3,000 rice genomes project (<http://oryzasnp.org/iric-portal>). The raw genotypic data extracted from the database contained 962 k SNPs. The filtering for missing data ($\leq 20\%$), minor allele frequency (MAF) $\geq 2\%$ and rate of heterozygosity (H_o) $\leq 5\%$ led to working set of 215,250 SNPs, referred to 215 k set. This panel and the associated genotypic data were previously described in Bhandari et al. (2019).

Phenotyping of population

Experimental design and Crop Management

Six experiments (Additional file 1: Table S1) were conducted in the 2014 wet season (WS) and 2015 dry season (DS) at IRRI (14.18°N , 121.25°E), Philippines. In each season, the experiment was conducted under control conditions or non-stress experiment (LL_N) in lowland (under flooded, puddled, transplanted and anaerobic conditions) while the reproductive-stage drought-stress experiments were conducted in lowland and upland (under direct-sown, non-puddled, non-flooded and aerobic conditions in leveled

fields) environments, referred as LL_S and UL_S, respectively. The LL_N experiments were established in augmented randomized complete block design in single row plots with 5 m row length. The LL_S and UL_S experiments were established in an α -lattice design with two replications in single or two- row plots with 5 m row length in lowland and 2–3 m row length in upland. The crop management practices were as described in Kumar et al. (2014).

Drought application procedure

RS-drought phenotyping was as described in Kumar et al. (2014). Briefly, in the LL_S experiments, the field was drained 30 days after transplantation and irrigation was withheld to impose the RS-drought stress. Stress was continued until severe leaf rolling was observed in at least 75% of the accessions and water table depth remained below 100 cm for more than 2 weeks. Fields were thereafter re-irrigated (flash-flooding -WS and sprinklers - DS) and the water was drained after 24 hours to impose a subsequent cycle of drought stress. This cyclic pattern was implemented till harvest. In the UL_S experiments, where crop was established by direct-seeding, RS-drought stress was initiated 45 days after sowing, by withholding sprinkler irrigation until the soil water tension fell below – 50 kPa at 30 cm depth. Thereafter, sprinkler-irrigation and subsequent drainage after 24 hours for imposition of drought stress were done in a cyclic pattern till harvest.

Traits measured

For each experiment, days to 50% flowering (DTF, in days), plant height (PH, in cm, average for 3 measurements per plot), panicle length (PL, in cm, average for 3 measurements per plot), flag leaf area (FlgLA, in cm^2 , average for 3 measurements per plot), dry biomass at maturity (BMDW, in kg ha^{-1}), number of effective panicles (NBP), grain yield (GY, in kg ha^{-1}), 100-seed weight (HSW, in g) and spikelet fertility (SPKFT, in percentage) were measured in individual plots and harvest index (HI) was calculated as GY/BMDW. Details of measurement procedures of each trait are given in Additional file 1: Table S5.

Analysis of phenotypic data for each individual trait

For each trait from each of the six experiments, best linear unbiased predictors (BLUP) were estimated using the REML option in the PROC MIXED procedure of SAS v9.0 (SAS Institute Inc., 2002). Within a season, the performance of a genotype was modeled as $Y_{ij} = \mu + \beta_i + c_j + a_i + \epsilon_{ij}$ for augmented randomized complete block design where Y_{ij} is the phenotype of the i^{th} genotype in j^{th} block, μ the overall mean, β_i the block effect which was considered as random, c_j the checks effect in j^{th} block which was considered as fixed, a_i the random effect of the i^{th} genotype and ϵ_{ij} is the residual considered as a random effect. We constructed two variables- “checks” and “genotypes” variables in both WS and DS. Checks refer to the control genotypes included additionally in the experiment to compare the performance of genotypes being tested and were used to recover the block effects. For α -lattice design, genotype performance was modeled as $Y_{ijk} = \mu + a_i + r_j + b_{kj} + \epsilon_{ijk}$ where Y_{ijk} is the phenotype of the i^{th} genotype in k^{th} block of j^{th} replicate, μ the overall mean, a_i is the genotype effect considered as random, r_j is the replicate effect considered as fixed, b_{kj} is the random effect of the k^{th} block within j^{th} replicate and ϵ_{ijk} is the residual considered as a random effect.

An initial diagnostic analysis was performed on each experiment for all ten traits using the “influence” option of the model statement to detect and filter out the outliers among the individual data sets at the plot level. These outliers were considered as missing phenotypes in the subsequent steps of data analysis. We assessed the genetic variance of the genotypes being tested by keeping them as a random factor in the model and obtained the variance components. BLUP and $Y_{adj} (\mu + Y_{ij(k)})$ values were calculated for each of the genotype, the latter was used in GWAS for analysis at both individual experiment level and combined analysis for each environment- lowland non-stress, lowland stress and upland stress, to detect genomic regions associated with traits of interest. For each of the studied trait, the broad sense heritability was estimated using the formula

$$H^2 = \sigma^2_g / \sigma^2_p$$

where σ^2_g is the genotypic variance obtained from the experimental data (assuming only additive genetic variance among accessions) and the phenotypic variance is $\sigma^2_p = \sigma^2_g + \sigma^2_e/r$, where σ^2_e is the residual variance obtained from the ANOVA and r is the number of replication.

The corrplot package in R (R. v.1.2.5001) was used to estimate the correlation among the measured traits.

Promising accessions were selected from the population based on yield advantage over non-stress condition in WS for both lowland and upland stress environments and over checks in each environment in DS. The premise was to identify set of accessions which can be incorporated in breeding programs for drought tolerance under both lowland and upland environments with advantage of early flowering and short plant type under RS drought.

Methods for characterizing the population

Experimental evaluation

Multi-dimensional analysis of the phenotypic data by factorial discriminant analysis (FDA) was performed on phenotypic data (280 accessions x 10 trait variables x 6 experiments) to estimate the pairwise Fisher distance between the experiments using the XLSTAT package [Internet] 2012. Available: <http://www.xlstat.com/en/products-solutions/pro.html>.

Genetic Structure

The genetic diversity among the 280 accessions was studied with the working set of 215 k markers using the Neighbor-joining (NJ) clustering method in TASSEL 5 (Bradbury et al. 2007) and visualization using FigTree v1.4.3 (Rambaut and Drummond 2016). The population structure was assessed using ADMIXTURE v.1.3.0 (Alexander et al. 2009) and results visualized using R/pophelper (Francis 2017) package for 280 accessions and 215 k SNPs. Series of models for K value ranging from 2–8 were run with 5 fold cross validation to prime the main algorithm- QuasiNewton for convergence acceleration. Accuracy and precision were ensured by performing 20 runs for each value of K and the optimal number of clusters was determined by the K value with the least cross-validation (CV) error. Principal components

(PC) explaining genetic variation was estimated using R/GAPIT 3.0 package (Lipka et al. 2012). The estimated population structure covariates (Q) and kinship matrix (K) were used to improve the statistical power of the GWAS models used.

Pairwise linkage disequilibrium (LD)

LD between SNP loci at the individual chromosomal level was calculated and plotted by computing r^2 estimators between all pairs of SNP markers using the PopLDdecay (Zhang et al. 2019).

Methods for identifying associations at population level

In our study, we implemented GWAS with MLM, CMLM, SUPER and Farm-CPU methods using R/GAPIT 3.0 package and visualization of circular manhattan and qq plots using rMVP package (0.99.17) (<https://github.com/xiaolei-lab/rMVP>).

The false positives in GWAS study were corrected using “Bonferroni Correction” factor. After the Bonferroni multiple test correction (0.05/215,250; significance level of 5%/total number of markers used in analysis), the calculated threshold value was 2.32×10^{-7} . Only the MTAs that exceed the threshold value and which were consistent across different methods were reported. To detect seasonal variations, we explored two p-value thresholds (1e-6 and 1e-4).

The percent phenotypic variance (PV) explained by all significant SNPs detected in each environment and season was output from all models used in the study. PV explained by each significant SNP was calculated as the squared correlation between the phenotype and genotype of the SNP.

Genomic control

The chi-square test statistics were calculated from the p-values for 10 traits x 4 models x 6 experiments to compute the genomic inflation factor (λ). The genomic inflation factor was defined as the median of the observed chi-squared test statistics divided by the expected median of the corresponding chi-squared distribution, assuming 1 degree of freedom at the whole genome level.

Candidate genes discovery

The candidate genes were searched within 200-kb region around (100 kb upstream and 100 kb downstream) the detected significant SNP. The literature searches were also performed using QTARO and MSU databases (<http://qtaro.abr.affrc.go.jp> and <http://rice.plantbiology.msu.edu>) to identify the earlier reported QTLs present in the LD region.

Abbreviations

BMDW

dry biomass at maturity

DTF

days to 50% flowering

CMLM

compressed mixed linear model
Farm-CPU
fixed and random model for circulating probability unification
FlgLA
flag leaf area
GWAS
genome-wide association study
GY
grain yield
HI
harvest index
Ho
heterozygosity
HSW
100-seed weight
MAF
minor allele frequency
MLM
mixed linear model
MTA
marker-trait association
NBP
number of effective panicles
PH
plant height
PL
panicle length
QTL
quantitative trait locus
SNP
single nucleotide polymorphism
SPKFT
spikelet fertility
SUPER
settlement of MLM under progressively exclusive relationship

Additional File

Table S1. Field experiments conducted at IRRI, Philippines between the 2014 Wet-Season and 2015 Dry-Season. **Table S2.** Characterisation of the marker set of 215,250 SNPs. **Table S3.** Validation results for

the significant MTAs identified for ten traits in the study. **Table S4.** Details of the 280 diversity panel accessions. **Table S5.** Description of phenotypic data recording. **Figure S1.** Scatterplot matrices for the ten traits (a –DTF, b-PH, c-PL, d-FlgLA, e-NBP, f-BMDW, g-GY, h-HI, i-HSW, j-SPKFT) showing phenotypic distribution and correlation between environments and seasons. **Figure S2.** Heat map showing the uneven marker distribution along each of the 12 chromosomes using the 215,250 SNP working set. **Figure S3.** Pattern of rapid decay in linkage disequilibrium decay in the population of 280 accessions genotyped with 215,250 SNPs. **Figure S4.** Effect of different algorithms in controlling p-value inflation. **Figure S5.** Circular manhattan plots and qq-plots for each of the 6 experiments and comparison of season-wise analysis to combined analysis for each of the 3 growing environments for a. PL, b. FlgLA, c. NBP, d. BMDW, e. HI, f. HSW and g. SPKFT.

Declarations

Availability of Data and Materials

All relevant data are provided with the manuscript.

Funding

This work was funded by Agropolis Foundation (<http://www.agropolis-fondation.fr/>) and Cariplo Foundation (<http://www.fondazionecariplo.it/>), Grant no 1201-006. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The authors thank the Agropolis Foundation and Cariplo Foundation for providing financial support for the study.

Acknowledgments

The authors acknowledge the technical support received from Ma Teresa Sta Cruz, Paul C Maturan, Jocelyn Guevarra and Ruth Erica Carpio of the rainfed South Asia rice breeding team of IRRI throughout the course of the study.

Authors' Contributions

AK and NA conceptualized the study; AB carried out phenotyping studies and curated the data with JB and TVCH; AB performed the analysis and drafted the manuscript under the supervision of NS, JB and NK; AK, NS and JB reviewed and edited the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

The manuscript has been approved by all authors.

Competing Interests

The authors declare that they have no competing interests.

Author details

¹Rice Breeding Platform, International Rice Research Institute, DAPO Box7777, Metro Manila, Philippines.

²Banasthali University, Banasthali 304022, India. ³Punjab Agricultural University, Ludhiana 141004, India.

⁴CIRAD, UMR, AGAP, Montpellier, France. ⁵AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France. ⁶ IRRI South Asia Regional Centre, Varanasi 221006, India.

References

Abbai R, Singh VK, Nachimuthu VV, Sinha P, Selvaraj R, Vipparla AK, Singh AK, Singh

UM, Varshney RK, Kumar A (2019) Haplotype analysis of key genes governing
grain yield and quality traits across 3K RG panel reveals scope for the
development of tailor-made rice with enhanced genetic gains. Plant Biotechnol J
17:1612-1622

Agrama HA, Eizenga GC, Yan W (2007) Association mapping of yield and its
components in rice cultivars. Mol Breed 19(4):341–356

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry
in unrelated individuals. Genome Res 19(9):1655-1664

Bao Z, Watanabe A, Sasaki K, Okubo T, Tokida T, Liu D, Ikeda S, Imaizumi-Anraku H,
Asakawa S, Sato T, Mitsui H, Minamisawa K (2014) A rice gene for microbial
symbiosis, OsCCaMK, reduces CH₄ flux in a paddy field with low nitrogen
input. Appl Environ Microbiol 80(6): 1995-2003

Bernardo R, Charcosset A (2006) Usefulness of gene information in marker-assisted
recurrent selection: a simulation appraisal. Crop Sci 46(2):614–621

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning
from the last 20 years. Crop Sci 48(5):1649–1664

- Bhandari A, Bartholomé J, Cao-Hamadoun TV, Kumari N, Frouin J, Kumar A, Ahmadi N (2019) Selection of trait-specific markers and multi-environment models improve genomic predictive ability in rice. PLoS ONE 14(5): e0208871
- Blum A (1988) Plant breeding for stress environments. CRC Press Inc, Boca Raton, 223
- Bonnett DG, Rebetzke GJ, Spielmeyer W (2005) Strategies for efficient implementation of molecular markers in wheat breeding. Mol Breed 15:75–85
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633–2635
- Chen L, Xiong G, Cui X, Yan M, Xu T, Qian Q, Xue Y, Li J, Wang Y (2013) OsGRAS19 may be a novel component involved in the brassinosteroid signaling pathway in rice. Mol Plant 6(3): 988–991
- Edmeades GO, Bolanos J, Lafitte HR, Rajaram S, Pfeiffer W, Fischer RA (1989) Traditional approaches to breeding for drought resistance in cereals. In: Baker FWG (ed) Drought resistance in cereals. ICSU, Paris and CAB International, Wallingford, England, pp 27–52
- Fernando RL, Grossman M (1989) Marker assisted selection using best linear unbiased prediction. Genet Sel Evol 21:467-477
- Francis RM (2017) pophelper: an R package and web app to analyze and visualize population structure. Mol Ecol Resour 17(1):27-32
- Frazer KA, Murray SS, Schork NJ, Topol EJ (2009) Human genetic variation and its contribution to complex traits. Nat Rev Genet 10(4): 241–251
- Fukai S, Cooper M (1995) Development of drought-resistant cultivars using physiomorphological traits in rice. Field Crops Res 40(2):67-86

Hillel J, Sharp T, Haberfeld A, Jeffreys AJ, Plotzky Y, Cahaner A, Lavi U (1990) DNA fingerprints applied to gene introgression in breeding programs. *Genetics* 124(3):783–789

Hoang GT, Van Dinh L, Nguyen TT, Ta NK, Gathignol F, Mai CD, Jouannic S, Tran KD, Khuat TH, Do VN, Lebrun M, Courtois B, Gantet P (2019) Genome-wide association study of a panel of Vietnamese rice landraces reveals new QTLs for tolerance to water deficit during the vegetative phase. *Rice* 12: 4

Hospital F, Chevalet C, Mulsant P (1992) Using markers in gene introgression breeding programs. *Genetics* 132(4): 1199-1210

Howes NK, Woods SM, Townley-Smith TF (1998) Simulations and practical problems of applying multiple marker assisted selection and doubled haploids to wheat breeding programs. *Euphytica* 100:225–230\

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, Fan D, Lu Y, Weng Q, Liu K, Zhao T, Jing Y, Si L, Dong G, Huang T, Lu T, Feng Q, Qian Q, Li J, Han B (2012) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat Genet* 44: 32-39

Huang X, Han B (2014) Natural variations and genome-wide association studies in crop plants. *Annual review of plant biology* 65:531-551

Kaler AS, Purcell LC (2018) Estimation of a significance threshold for genome-wide association studies. *BMC Genomics* 20:618

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. *Genetics* 178: 1709–1723

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E

(2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42(4): 348– 354

Kumar A, Bernier J, Verulkar S, Laffite HR, Atlin GN (2008) Breeding for drought tolerance: direct selection for yield, response to selection and use of drought-tolerant donors in upland and lowland-adapted populations. *Field Crops Res* 107(3):221-231

Kumar A, Verulkar SB, Dixit S, Chauhan B, Bernier J, Venuprasad R, Zhao D, Shrivastava MN (2009) Yield and yield-attributing traits of rice (*Oryza sativa* L.) under lowland drought and suitability of early vigor as a selection criterion. *Field Crops Res* 114(1): 99–107

Kumar A, Dixit S, Ram T, Yadaw RB, Mishra KK, Mandal NP (2014) Breeding high-yielding drought-tolerant rice: genetic variations and conventional and molecular approaches. *J Exp Bot* 65(21): 6265-6278

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124(3): 743–756

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

Li M, Liu X, Bradbury P, Yu J, Zhang YM, Todhunter RJ, Buckler ES, Zhang Z (2014) Enrichment of statistical power for genome-wide association studies. *BMC Biol* 12:73

Li X, Yan W, Agrama H, Jia L, Jackson A, Moldenhauer K, Yeater K, McClung A, Wu D (2012) Unraveling the complex trait of harvest index with association mapping in rice (*Oryza sativa* L.). *PLoS ONE* 7 (1): e29350

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z

(2012) GAPIT: genome association and prediction integrated tool.

Bioinformatics 28(18):2397– 2399

Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D (2011) FaST linear

mixed models for genome-wide association studies. Nat Methods 8(10):833–

835

Listgarten J, Lippert C, Kadie CM, Davidson RI, Eskin E, Heckerman D (2012) Improved

linear mixed models for genome-wide association studies. Nat Methods 9(6):

525–526

Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016) Iterative usage of fixed and

random effect models for powerful and efficient genome-wide association

studies. PLoS Genet 12(3): e1005957

MSU database (<http://rice.plantbiology.msu.edu>). Accessed on 19 Nov 2019

Miyoshi K, Ito Y, Serizawa A, Kurata N (2003) OsHAP3 genes regulate chloroplast

biogenesis in rice. Plant J 36(4):532–540

Nath K, Poudyal RS, Eom JS, Park YS, Zulfugarov IS, Mishra SR, Tovuu A, Ryoo N, Yoon

HS, Nam HG, An G, Jeon JS, Lee CH (2013) Loss-of-function of OsSTN 8

suppresses the photosystem II core protein phosphorylation and interferes with

the photosystem II repair mechanism in rice (*Oryza sativa*). Plant J 76(4):675–

686

Nguyen TL, Bui CB (2008) Fine mapping for drought tolerance in rice (*Oryza sativa* L.).

Omonrice 16:9–15

Ordonez Jr SA, Silva J, Oard JH (2010) Association mapping of grain quality and

flowering time in elite *japonica* rice germplasm. J Cereal Sci 51(3):337–343

Palanog AD, Swamy BPM, Shamsudin NAA, Dixit S, Hernandez JE, Boromeo

TH, Sta Cruz PC, Kumar A (2014) Grain yield QTLs with consistent-effect under reproductive-stage drought stress in rice. *Field Crops Res* 161:46–54

Pauli D, Muehlbauer GJ, Smith KP, Cooper B, Hole D, Obert DE, Ullrich SE, Blake TK (2014) Association mapping of agronomic QTLs in US spring barley breeding germplasm. *Plant Genome* 7: 3

Price A, Courtois B (1999) Mapping QTLs associated with drought resistance in rice: progress, problems, and prospects. *Plant Growth Regulation* 29: 123-133
QTARO database (<http://qtaro.abr.affrc.go.jp>). Accessed on 19 Nov 2019

R/MVP package (<https://github.com/xiaolei-lab/rMVP>)_Accessed on 19 Aug 2019

Raman A, Verulkar S, Mandal NP, Variar M, Shukla V, Dwivedi J, Singh B, Singh O, Swain P, Mall A, Robin S, Chandrababu R, Jain A, Ram T, Hittalmani S, Haefele S, Piepho HP, Kumar A (2012) Drought yield index to select high yielding rice lines under different drought stress severities. *Rice* 5:31

Rambaut A, Drummond A (2016) <http://tree.bio.ed.ac.uk/software/figtree/> Accessed on 9 May 2019

Rosielle AA, Hamblin, J (1981) Theoretical aspects of selection for yield in stress and non-stress environment. *Crop Sci* 21 (6):943-946

Sandhu N, Singh A, Dixit S, Sta Cruz MT, Maturan PC, Jain RK, Kumar A (2014) Identification and mapping of stable QTL with main and epistasis effect on rice grain yield under upland drought stress. *BMC Genet* 15:63

Sandhu N, Torres RO, Sta Cruz MT, Maturan PC, Jain R, Kumar A, Henry A (2015) Traits and QTLs for development of dry direct-seeded rainfed rice varieties. *J Exp Bot* 66(1):225–244

Sandhu N, Subedi SR, Singh VK, Sinha P, Kumar S, Singh SP, Ghimire SK, Pandey M,

Yadaw RB, Varshney RK, Kumar A (2019) Deciphering the genetic basis of root morphology, nutrient uptake, yield, and yield-related traits in rice under dry direct-seeded cultivation systems. *Sci Rep* 9(1): 9334

Serraj R, Kumar A, McNally KL, Slamet-Loedin I, Bruskiewich RM, Mauleon R, Cairns J, Hijnmans RJ (2009) Improvement of drought resistance in rice. *Adv Agron* 103:41–98

Subedi SR, Sandhu N, Singh VK, Sinha P, Kumar S, Singh SP, Gimire SK, Pandey M, Yadaw RB, Varshney RK, Kumar A (2019) Genome-wide association study reveals significant genomic regions for improving yield, adaptability of rice under dry direct seeded cultivation condition. *BMC Genomics* 20(1): 471

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* 43(2): 159–162

Tripathy JN, Zhang J, Robin S, Nguyen TT, Nguyen HT (2000) QTLs for cell-membrane stability mapped in rice (*Oryza sativa* L.) under drought stress. *Theor Appl Genet* 100:1197–1202

Varshney RK, Graner A, Sorrells ME (2005) Genomics-assisted breeding for crop improvement. *Trends Plant Sci* 10(12):621–630

Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biol* 12(6):e1001883

Venuprasad R, Lafitte HR, Atlin GN (2007) Response to direct selection for grain yield under drought stress in rice. *Crop Sci* 47: 285–293

- Verulkar SB, Mandal NP, Dwivedi JL, Singh BN, Sinha PK, Mahato RN, Dongre P, Singh ON, Bose LK, Swain P, Robin S, Chandrababu R, Senthil S, Jain A, Shashidhar HE, Hittalmani S, Vera Cruz S, Paris T, Raman A, Haefele S, Serraj R, Atlin G, Kumar A (2010) Breeding resilient and productive genotypes adapted to drought-prone rainfed ecosystem of India. *Field Crops Res* 117: 197–208
- Vikram P, Swamy BPM, Dixit S, Ahmed HU, Sta Cruz MT, Singh AK, Kumar A (2011) *qDTY_{1.1}*, a major QTL for rice grain yield under reproductive-stage drought stress with a consistent effect in multiple elite genetic backgrounds. *BMC Genet* 12: 89
- Visscher PM, Yang J, Goddard ME (2010) A commentary on “common SNPs explain a large proportion of the heritability for human height” by Yang J. *Twin Res Human Genet* 13(6): 517–524
- Wang Q, Tian F, Pan Y, Buckler ES, Zhang Z (2014) A super powerful method for genome wide association study. *PLoS ONE* 9(9):e107684
- XLSTAT (<http://www.xlstat.com/en/products-solutions/pro.html>). Accessed on 22 July 2019
- Xu Q, Yuan XP, Yu HY, Wang YP, Tang SX, Wei X (2011) Mapping QTLs for drought tolerance at seedling stage in rice using doubled haploid population. *Rice Sci* 18(1):23–28
- Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci* 48(2):391–407
- Xu Y, Yang T, Zhou Y, Yin S, Li P, Liu J, Xu S, Yang Z, Xu C (2018) Genome-Wide Association Mapping of Starch Pasting Properties in Maize Using Single-Locus and Multi-Locus Models. *Front Plant Sci* 9:1311
- Yadav S, Sandhu N, Majumder RR, Dixit S, Kumar S, Singh SP, Mandal NP, Das SP, Yadaw

RB, Singh VK, Sinha P, Varshney RK, Kumar A (2019) Epistatic interactions of major effect drought QTLs with genetic background loci determine grain yield of rice under drought stress. *Sci Rep* 9(1): 2616

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 42(7): 565– 569

Zhang C, Dong SS, Xu JY, He WM, Yang TL (2019) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35(10):1786-1788

Zhang L, Cui X, Schmitt K, Hubert R, Navidit W, Arnheim N (1992) Whole genome amplification from a single cell: implications for genetic analysis. *Proc Natl Acad Sci* 89(13): 5847-5851

Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* 42(4): 355- 360

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun* 2: 467

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. *Plant Genome* 1(1):5–20

Tables

Table 1 Summary Statistics, ANOVA results and heritability for the 10 phenotypic traits of the 280 lines

SeasonTrait		LL_N				LL_S				UL_S	
		Adjusted	Entry (significance level)	h2	Adjusted	Entry (significance level)	h2	Adjusted	Entry (significance level)	h2	
		(Std.Err)									
VS	DTF	90.4 (0.58)	22.55 ^{NS}	0.2	79.66 (0.64)	37.55***	0.9	80.28 (0.58)	53.63***	0.72	
	PH	113.79 (1.57)	610.47***	0.9	102.42 (1.74)	580.83***	0.9	110.23 (1.76)	443.24 ^{NS}	0.43	
	PL	21.75 (0.21)	11.43***	0.9	20.05 (0.19)	7.26***	0.8	20.41 (0.19)	1.08 ^{NS}	0.16	
	FlgLA	36.0 (0.66)	83.89***	0.7	40.19 (0.77)	95.67***	0.7	37.77 (0.68)	16.55 ^{NS}	0.19	
	BMDW	15.57E+03	3.13E+07***	0	36.13E+02	4.23E+07***	0.4	54.61E+02	5.87E+07 ^{NS}	0.35	
			(568.08)		(146.23)			(409.17)			
	NBP	46.99 (1.12)	247.35***	0.7	20.38 (1.29)	15.81 ^{NS}	0.2	23.32 (1.08)	19.22***	0.13	
	GY	16.66E+02	7.94E+05***	0.8	5.73E+02	2.41E+04 ^{NS}	0.1	9.42E+02	9.96E+03 ^{NS}	0.03	
			(62.44)		(46.04)			(47.59)			
	HSW	22.1 (0.57)	17.63***	0.9	14.97 (0.39)	3.64 ^{NS}	0	17.79 (0.79)	4.26 ^{NS}	0.01	
	HI	0.17 (0.01)	0.001 ^{NS}	0.3	0.15 (0.01)	5E-03 ^{NS}	0.2	0.11 (0.01)	8E-05 ^{NS}	0.01	
	SPKFT	86.81 (0.59)	261.11***	0.8	57.49 (1.59)	99.48 ^{NS}	0.2	61.19 (2.59)	86.21 ^{NS}	0.15	
DS	DTF	87.83 (0.63)	76.31***	0.7	88.09 (2.11)	205.92***	0.9	93.07 (1.87)	352.44***	0.93	
	PH	108.63 (1.42)	538.47***	1	69.09 (1.02)	214.30***	0.8	71.03 (2.27)	187.20***	0.88	
	PL	22.15 (0.21)	9.29***	0.8	17.87 (0.24)	5.07***	0.6	16.45 (0.33)	1.35***	0.46	
	FlgLA	36.73 (0.63)	98.77***	0.9	29.85 (0.58)	43.68***	0.7	22.96 (0.38)	17.11***	0.61	
	BMDW	14.78E+03	4.23E+07***	0.8	36.50E+02	7.12E+05***	0.5	55.29E+02	8.79E+05***	0.42	
			(474.24)		(153.15)			(347.86)			
	NBP	54.81 (1.19)	130.23 ^{NS}	0.4	14.99 (1.44)	108.80***	0.8	24.84 (1.49)	202.87***	0.84	
	GY	52.11E+02	3.91E+06***	0.7	1.58E+02	4.14E+04***	0.7	3.36E+02 (31)	4.54E+09***	0.99	
			(157.17)		(17.64)						
	HSW	21.97 (0.55)	14.20***	0.9	11.49 (0.44)	5.76 ^{NS}	0.3	8.23 (0.54)	24.60***	0.77	
	HI	0.39 (0.01)	0.006***	0.6	0.05 (0.01)	0.003***	0.8	0.05 (0.01)	5E-03***	0.82	
	SPKFT	96.41 (0.38)	6.83***	0.8	47.03 (4.31)	452.50***	0.6	25.19 (3.29)	709.67***	0.81	

F-days to 50% flowering, PH-plant height, PL-panicle length, GY-grain yield, BMDW-biomass dry weight at maturity, FlgLA-flag leaf a, NBP-number of effective panicles, HI-harvest index, HSW-100 seed weight and SPKFT-spikelet fertility from each of the six experiments (lowland non stress - LL_N, lowland stress - LL_S and upland stress - UL_S) in the two seasons (WS and DS). NS- Non-significant and ***-significant at <0.05 level.

Table 2 Selected promising accessions showing significant MTAs for the traits of interest across seasons and environments

S.No	Taxa Name	DTF	PH	GY							
				LL_N	LL_N	LL_S	UL_S				
		WS	SDS	WS	DS	WS	SDS				
1	AUS 329::IRGC 29116-1	89	8311411611714966163139751224								
2	AUS 344::IRGC 29131-1	87	84105125137919435753359941907								
3	CHUNGUR BALI::IRGC 25855-186	771241151083710592569871537									
4	DANGAR::IRGC 76296-1	86	75125105178586592569801059								
5	LALSAITA::IRGC 43915-1	88	8310911234654895904519651778								
6	PARA NELLU::IRGC 50009-1	87	751441273719592582633975953								
7	SIMUL KHURI::IRGC 35154-1	91	86120118117448445743409941399								
Check1	IRRI 154	-	93	-	93	-	5462	-	7	-	65
Check2	MTU1010	-	93	-	94	-	4323	-	2	-	60
Check3	Sabitri	-	90	-	101	-	3761	-	124	-	107
Trial Mean		91 901141091655214595128707322									
LSD		1 2 3 3 140 280 82 32100 62									

Table 3 Comparison of models based on inflation factor (λ) for detecting marker trait associations

Trait	Model	LL_N_WSLL_N_DSLL_S_WSLL_S_DSUL_S_WSUL_S_DS						
DTF	MLM	0.968	0.932	0.983	1.004	1.009	1.014	
	Farm-CPU	0.926	0.905	0.96	0.886	1.013	0.943	
PH	MLM	0.897	0.96	0.938	1.005	0.91	0.954	
	Farm-CPU	0.838	0.921	0.995	0.951	0.97	0.947	
PL	MLM	0.999	1.031	1.02	1.061	0.996	1.029	
	Farm-CPU	0.933	0.942	0.973	1.042	1.004	1.022	
FlgLA	MLM	1.022	1.008	1.023	0.999	1.01	0.998	
	Farm-CPU	0.96	0.973	1.01	0.977	1.032	0.938	
NBP	MLM	1.032	1.016	0.913	0.992	1.029	1.044	
	Farm-CPU	0.951	0.834	1.013	0.879	1.028	0.944	
BMDW	MLM	0.943	0.964	1.011	1.008	0.996	1.037	
	Farm-CPU	0.973	0.945	0.962	0.973	0.966	0.857	
GY	MLM	0.979	0.995	1	0.992	0.878	0.897	
	Farm-CPU	0.981	0.897	1.01	0.93	0.873	0.913	
HI	MLM	1.022	0.958	0.969	0.983	1.047	0.964	
	Farm-CPU	0.931	0.83	1.02	0.871	0.982	0.908	
HSW	MLM	1.011	0.958	0.958	0.874	0.927	0.998	
	Farm-CPU	0.875	0.94	1	0.87	1.071	0.949	
SPKFT	MLM	0.978	0.858	0.901	1.019	1.005	0.996	
	Farm-CPU	0.921	0.83	1.03	1.023	1.028	0.885	

Figures

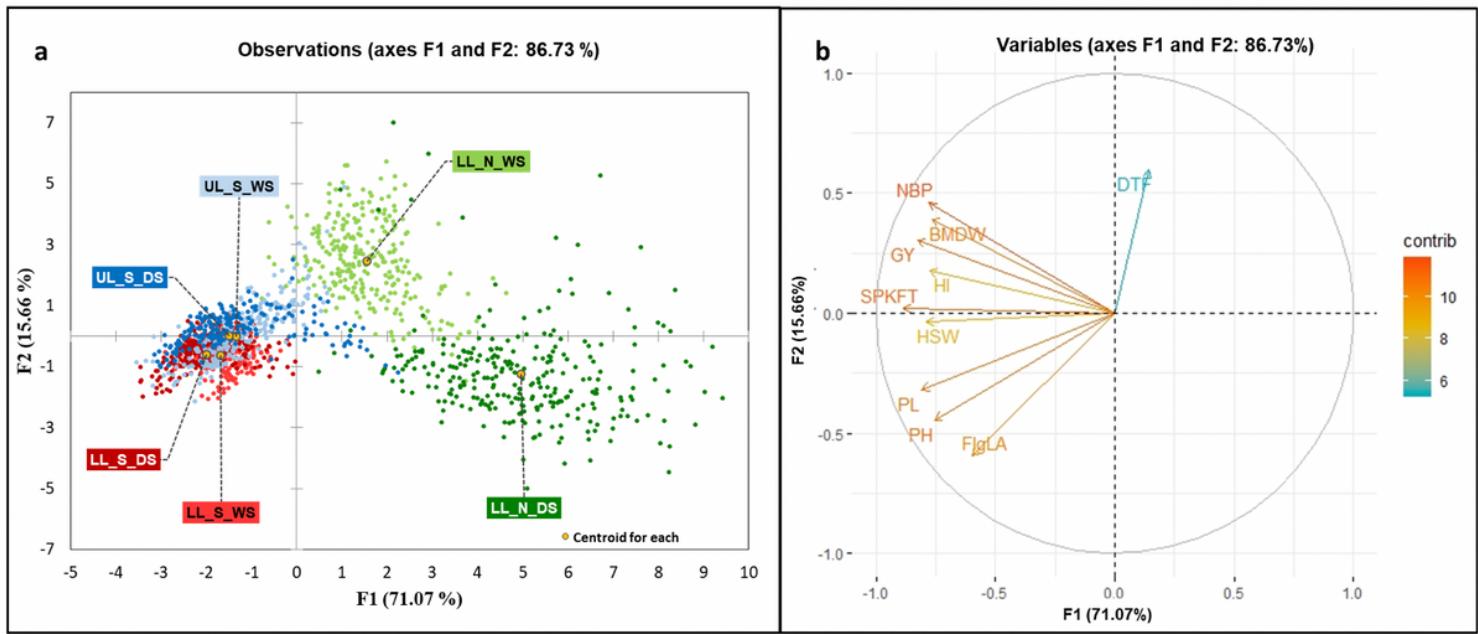


Fig.1 **a** Projection of 280 lines of the *indica-aus* diversity panel on the first plane of factorial discriminant analysis using phenotypic data for ten traits (Days to 50% flowering – DTF, Plant height – PH, Panicle length – PL, Grain yield – GY, Biomass – BMDW, Flag leaf area – FlgLA, Number of effective panicles– NBP, Harvest index – HI, 100 seed weight – HSW and Spikelet fertility – SPKFT) from each of the six experiments (lowland non stress – LL_N, lowland stress – LL_S and upland stress – UL_S) in the two seasons (WS and DS). The first two components, PC1 and PC2 explain 86.73% of the variance between the accessions in different experiments. **1b** Among the traits, DTF explains the least variance and is overall negatively correlated to other traits.

Figure 1

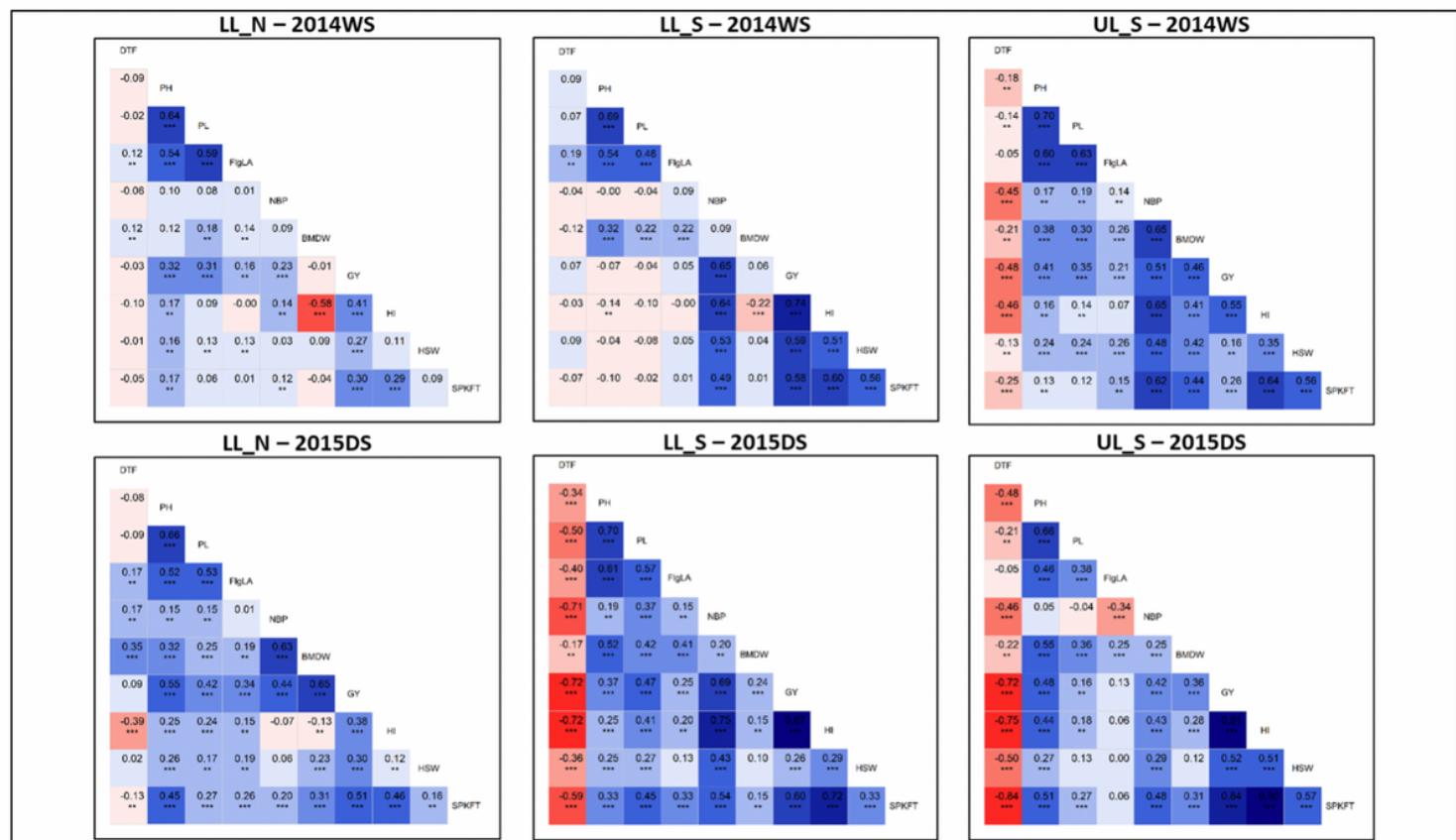


Fig.2 Plots of Pearson's r -values showing correlation between each of the ten traits in each of the six experiments (LL_N-2014WS – lowland non-stress 2014WS, LL_S-2015DS – lowland non-stress 2015DS, LL_S-2014WS – lowland stress 2014WS, LL_S-2015DS – lowland stress 2015DS, UL_S-2014WS – upland stress 2014WS and UL_S-2015DS – upland stress 2015DS). Blue color indicates positive correlation and red color indicates negative correlation among different traits, with color intensity depicting the strength of correlation. *significant at <0.01 level, **significant at <0.05 level, *** significant at <0.001 level, blank for non-significant.

Figure 2

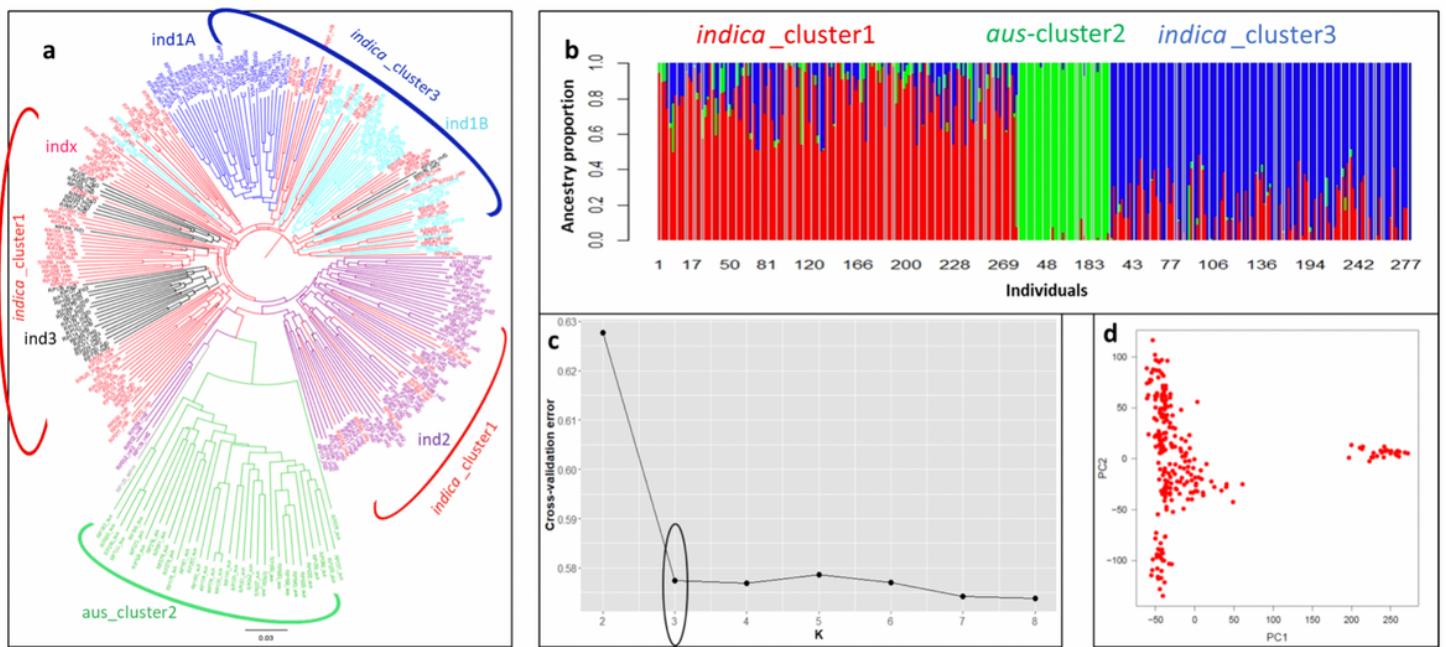


Fig.3 Genetic relatedness and population structure of the diversity panel; **3a** Genetic diversity depicted through the unweighted neighbour-joining tree method within the population as *indica* (*ind1A*, *ind1B*, *ind2*, *ind3* and *indx*) and *aus* accessions, as established in the 3k rice genome project; **3b** Ancestry proportions from ADMIXTURE analysis represented for $K = 3$, the optimal with the lowest cross-validation error for $K=2-8$ and visualized using R/pophelper package; **3c** Cross-validation error for $K=2-8$ from ADMIXTURE analysis ; **3d** Variation captured by the first two PCs using R/GAPIT shows 3 cluster distribution along the first two PC components.

Figure 3

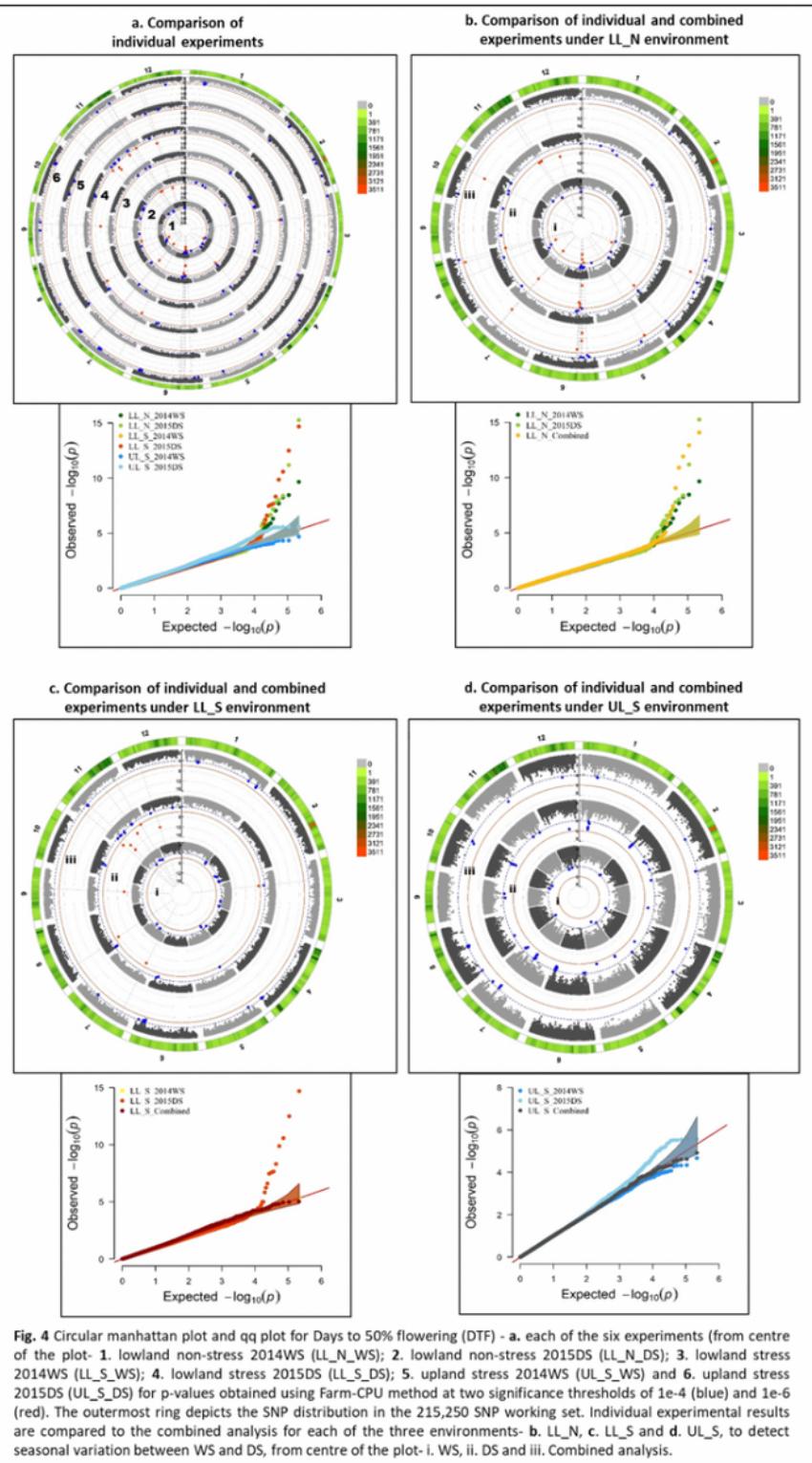


Fig. 4 Circular manhattan plot and qq plot for Days to 50% flowering (DTF) - a. each of the six experiments (from centre of the plot- 1. lowland non-stress 2014WS (LL_N_WS); 2. lowland non-stress 2015DS (LL_N_DS); 3. lowland stress 2014WS (LL_S_WS); 4. lowland stress 2015DS (LL_S_DS); 5. upland stress 2014WS (UL_S_WS) and 6. upland stress 2015DS (UL_S_DS) for p-values obtained using Farm-CPU method at two significance thresholds of 1e-4 (blue) and 1e-6 (red). The outermost ring depicts the SNP distribution in the 215,250 SNP working set. Individual experimental results are compared to the combined analysis for each of the three environments- b. LL_N, c. LL_S and d. UL_S, to detect seasonal variation between WS and DS, from centre of the plot- i. WS, ii. DS and iii. Combined analysis.

Figure 4

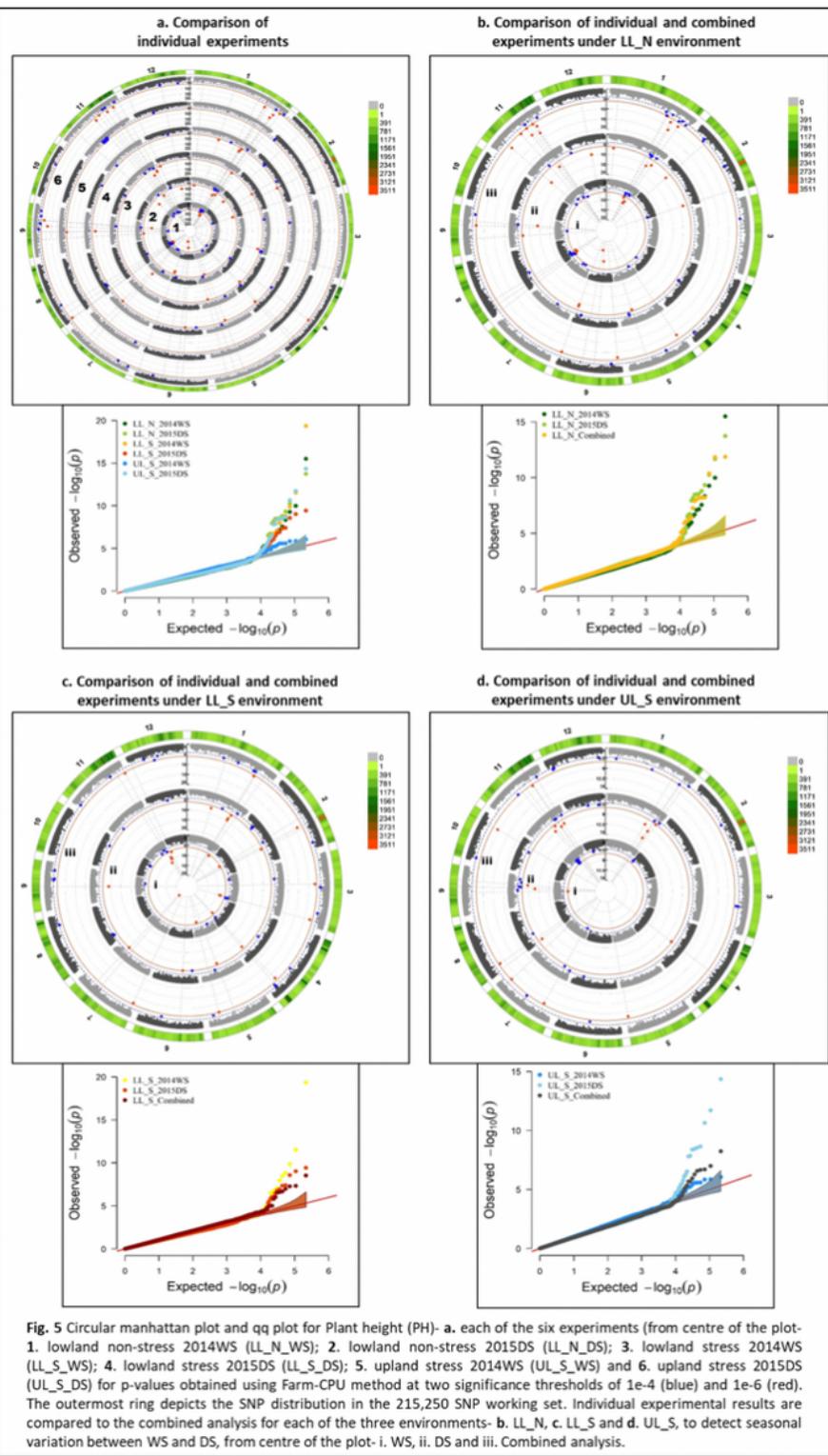


Figure 5

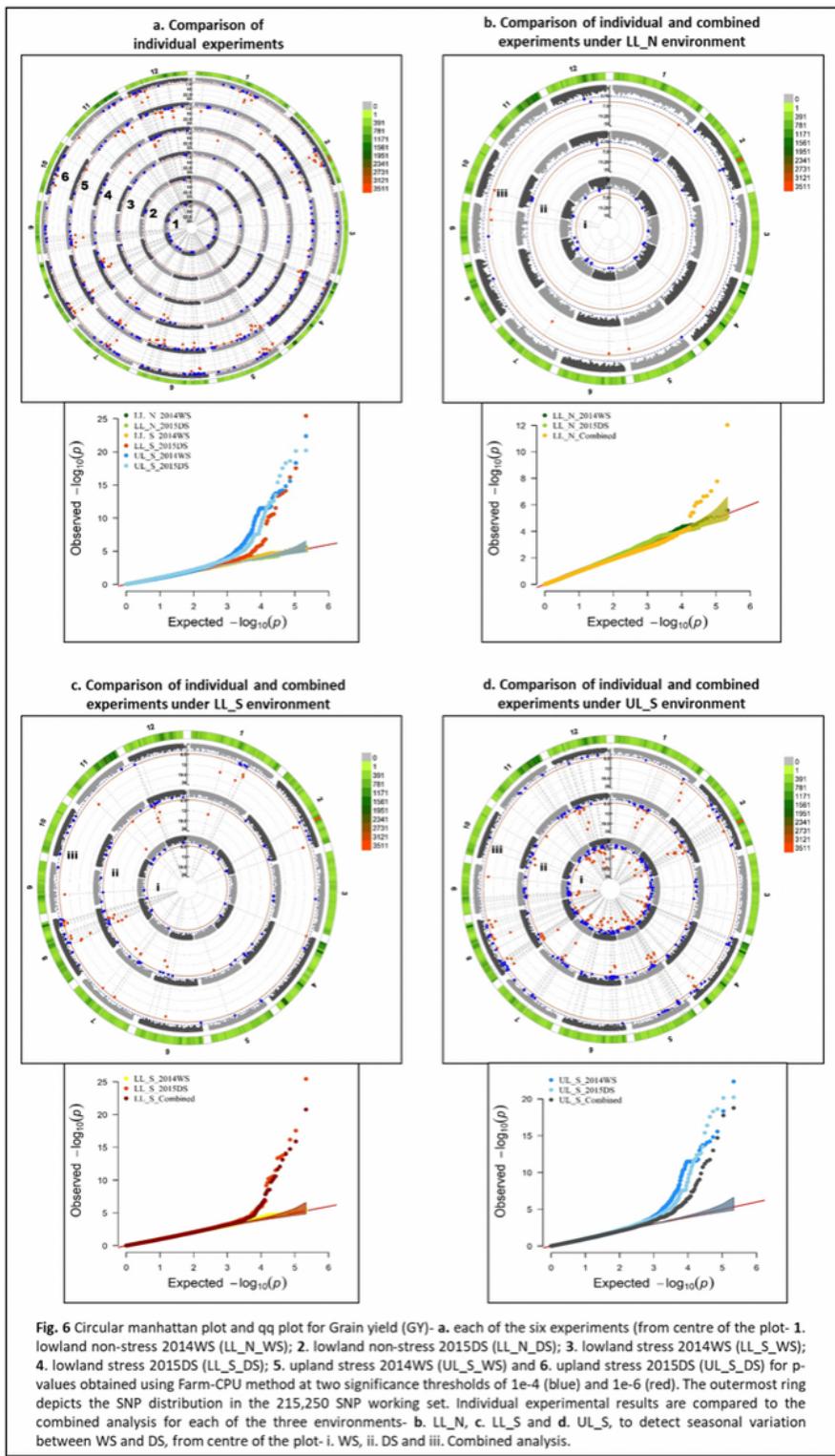


Figure 6

Supplementary Files

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

- AdditionalfileBhandarietal.docx