Output2 Introduction

Breeding materials to improve grain yield and nutrient use efficiency under low-input and low-fertility soil conditions are developed

Output 2 aimed to develop breeding materials with superior nutrient uptake and grain yields under poor nutrient conditions using our genetic resources such as Pup1. We also identify new QTLs and candidate genes that contribute to the nutrient efficiency of rice. For this, we set three targets:

1. Develop at least two breeding lines to improve nutrient acquisition capacity and utilization efficiency by 20% relative to the conventional varieties in the target area (or recurrent parents).
2. Develop at least two DNA markers for QTLs related to nutrient acquisition and utilization in rice.
3. Identify at least two genes and their mechanisms related to nutrient acquisition and utilization in rice.

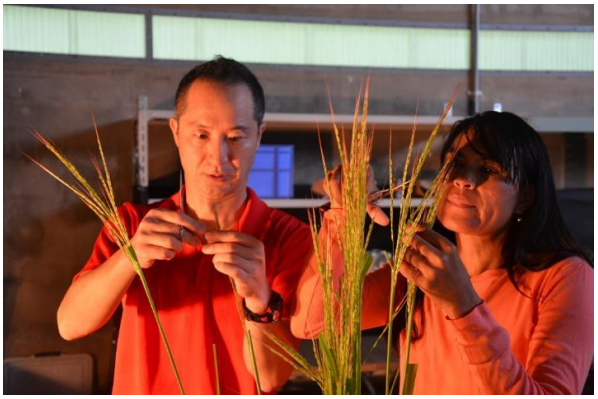
The highlights of this output include the official release of new lowland rice varieties in Madagascar and new genetic resources, such as QTLs (quantitative trait loci) and candidate genes that should contribute to sustainable rice production under nutrient deficiency in Madagascar and elsewhere in Africa. The development of a new laboratory and greenhouse for crop breeding and training young researchers is another important achievement to accelerate advanced rice breeding in Madagascar.

We were able to import two populations of early generation breeding lines to Madagascar for on-farm selection, including farmer participatory variety selection (PVS), due to the prior work done in Japan on developing new rice varieties. Advanced breeding lines subsequently entered the variety registration process supervised by the Malagasy seed certification agency, SOC. One breeding line from each population was released as varieties FyVary32 and FyVary85. Breeder seeds of these varieties have been distributed to local seed producer groups through FOFIFA and in collaboration with JICA technical cooperation Project (Papriz). These varieties have wider adaptation outside the FyVary project target region as their registration for the coastal Boeny region has been approved with additional trials being conducted in Southern and North-Eastern Madagascar.

The identification of a plausible mechanism of relevant genes to control tillering of rice and its possibility to increase rice yields under phosphorus-deficient and tillering-restricted soil conditions is a novel and practical achievement of the project. The QTL was introgressed

into X265 for evaluation. Evaluation of 350 gene bank accessions from IRRI in farmers' fields under zero-input conditions led to the identification of two QTLs associated with grain yield. The one with the stronger association is currently targeted in marker-assisted selection in a population derived from the local variety X265 crossed with QTL donor GP1103. Advanced breeding lines are being handed over to FOFIFA for further variety development. The above study provided the data that allowed us to venture into genomic selection as a new tool for exploring variation hidden in gene banks. We developed and applied genomic selection models to predict yield at zero-input in Madagascar and confirmed the good performance of several gene bank accessions, one of which is used as a donor in breeding. Following the same approach and collaborating with HarvestPlus and CIRAD, we developed the model to predict grain Zn concentration. This activity was extended to breeding for Zn biofortification—a much-needed activity given that Zn malnutrition is a major health concern in Madagascar and local varieties are Zn deficient in grain.

A memory of Output2 activities

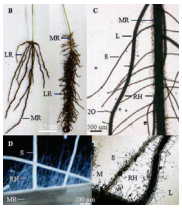
FyVary 32 - FyVary 85

Voka-bary ambany sy tany tsy ampy singa faosifaoro

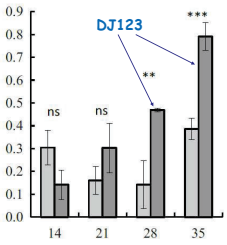
Maro ny olana mahatonga ny voka-bary ho ambany eto Madagasikara: ny tsy fahampian'ny rano (haintany), ny hatsiaka, ny tambiazina, ny sira izay hita matetika eny amoron-tsiraka (Boeny). Fa ny lehibe indrindra dia ny tsy fahampian'ny tsirontany indrindra ny singa Faosifaoro. Io farany izay voageja mafy ao anaty tany ka tsy afaka trohan'ny vary. Manampy izany ny taham-pampiasana zezika ambany izay antony mampihena fatratra ny voka-bary (2.5-3 t/ha) *Saito et al. 2019*. Eo ihany koa ny fahazaran'ny tantsaha mampiasa masomboly nentim-paharazana izay tsy dia mahavokatra.

Ny tanjona imasoan'ny tetikasa FyVary dia ny hamoaka vary vaovao ahazoana vokatra ambony na dia ambolena amin'ny tany manta aza.

Sary 1: Ny IR64-Pup1 izay nahazoana ny **FyVary 32** dia manana foto-tarazo (gène) PUP-1. Io PUP-1 io dia manosika ny fitrohana singa faosifaoro ao anaty tany.
Noho izany dia manana an'io tarazo io koa ny **FyVary 32** ka afaka misintona faosifaoro tsara, indrindra eny amin'ny tany manta.



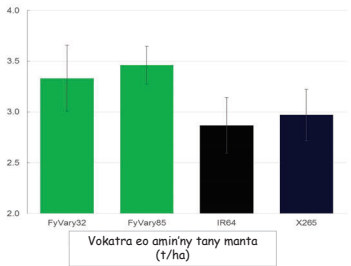
Faka jerena amin'ny mikraoskaopy



Fitrohana faosifaoro (µg P/ habean'ny faka)

Sary 2: Ny DJ123 izay nahazoana ny **FyVary 85** kosa dia manana faka miavaka amin'ny fahafahany misintona faosifaoro ao anaty tany ka afaka mitroka sy mampiasa tsara ny faosifaoro mihoatra noho ny vary hafa izay mety manana faka lehibe kokoa aza.

Ny fifantenana ireto vary ireto dia natao teny anivon'ny tantsaha. Sady tsy nisy ny fampiasana zezika no samihafa ny toe-tany nanaovana ny andrana.
Ireto avy ny toerana nanaovana ny fifantenana ireto vary ireto: ny faritra Vakinankaratra (Ankazomiriotra, Antohobe), ny faritra Alaotra Mangoro (Anjiro) ary ny faritra Boeny (Marovoay)



Miisa 14 ny andrana izay notontosaina nandritry ny roa taona. Hita mazava fa manatombo ny vokatra azo avy amin'ny ireo vary FyVary 32 sy FyVary 85 raha ampitahaina amin'ny X265 sy IR64 (izay malaza eran-tany amin'ny fahavokarany).
Ny fyVary 85 no mamokatra indrindra ka ahazoana vokatra 3.5t/ha eny amin'ny tany manta.

Toe-tany	Anarana	Elanelam-bokatra mitaha amin'ny X265 (%)	Elanelan'ny fiakaran'ny vokatra mitaha amin'ny X265 (andro)
Manerana ireo andrana rehetra	FyVary 85	17.0	4.9
	FyVary 32	10.0	-4.2
Tany manta	FyVary 85	22.3	1.3
	FyVary 32	5.5	-5.3
Tany masaka	FyVary 85	13.6	8.5
	FyVary 32	12.9	-3

Hita taratra fa eny amin'ny tany manta dia nahazo vokatra 22.3% ny FyVary 85 raha oharina amin'ny X265. Eo amin'ny tany masaka kosa dia nahitana fiakarany 13% izy roa (FyVary 32 sy FyVary 85).
Manatombo 2 ka hatramin'ny 9 andro ny tsingerim-piainan'ny FyVary 85 raha oharina amin'ny X265.

Ny FyVary 32 dia mailaka kokoa noho ny X265.

Ny famoahana ireto vary ireto dia vokatry ny fiaraha-miasa akaiky tamin'ny tantsaha.
Nanomboka tamin'ny fifantenana ka hatramin'ny fanandramana ny tsirony.

FyVary 85

FyVary 32

JIRCAS 国際農研
MINISTERE DE L'AGRICULTURE ET DE L'ELEVAGE
FOFIFA
JICA
MESUPRES

FYVARY 32

Nom commun	Nom botanique	**Dénomination**	Nature génétique	Origine (code)	Obtenteur	Référence (collection FOFIFA)	Année d'introduction (collection)	Mainteneur
Riz	*Oryza sativa*	**FYVARY32**	Lignée Issue de croisement	Japon	JIRCAS FOFIFA IRRI	7254	2021	FOFIFA JIRCAS

Grains de paddy

Grains décortiqués (riz cargo)

Grains usinés

CARACTERES MORPHOLOGIQUES

PANICULE		
• Longueur	:	Longue
• Type	:	Intermédiaire
FEUILLE		
• Angle de la feuille paniculaire	:	Oblique
GRAIN		
• Couleur du paddy	:	Jaune paille
• Couleur de l'apex	:	Jaune
• Longueur du paddy	:	9,97 mm
• Couleur du caryopse	:	Blanc
• Longueur du caryopse	:	7,02 mm
• Longueur du grain usiné	:	6,96mm
• Translucidité	:	Moyennement translucide

CARACTERES AGRONOMIQUES

• Région de culture	:	Alaotra Mangoro, Boeny, Vakinankaratra
• Saison de culture	:	Saison sèche et ou pluvieuse
• Aptitude culturale	:	Irriguée
• Hauteur du plant	:	Demi-naine
• Rendement à l'usinage	:	70 %
• Cycle (à maturité)	:	145-157 Jours
• Poids de 1000 grains (paddy)	:	24,4 g
• Tallage	:	Bon
• Rendement moyen	:	4,5 – 5,5 t/ha

COMPORTEMENT VIS-A-VIS DES BIOAGRESSEURS

• Résistance à la pyriculariose	:	Résistante
• Résistance aux insectes	:	Sensible aux insectes de stockage

Autres traits spécifiques

- Gonflement a la cuisson : oui
- Gout : apprécié
- Egrenage : facile
- Verse: résistante
- Insensible a la photopériode
- Faible attractivité au Borer
- Forte adaptation aux conditions de fertilité faible

FYVARY 85

Nom commun	Nom botanique	**Dénomination**	Nature génétique	Origine (code)	Obtenteur	Référence (collection FOFIFA)	Année d'introduction (collection)	Mainteneur
Riz	*Oryza sativa*	**FYVARY85**	Lignée Issue de croisement	Japon	JIRCAS FOFIFA	7255	2021	FOFIFA JIRCAS

Grains de paddy

Grains décortiqués (Riz cargo)

Grains usinés

CARACTERES MORPHOLOGIQUES

PANICULE		
• Longueur	:	Moyenne
• Type	:	Intermédiaire
FEUILLE		
• Angle de la feuille paniculaire	:	Oblique
GRAIN		
• Couleur du paddy	:	Jaune paille
• Couleur de l'apex	:	jaune
• Longueur du paddy	:	8,91 mm
• Couleur du caryopse	:	Blanc
• Longueur du caryopse	:	6,49 mm
• Longueur du grain usiné	:	6,41 mm
• Translucidité	:	Moyennement translucide

CARACTERES AGRONOMIQUES

• Région de culture	:	Alaotra Mangoro, Boeny, Vakinankaratra
• Saison de culture	:	Saison sèche et ou pluvieuse
• Aptitude culturale	:	Irriguée
• Hauteur du plant	:	Demi-naine
• Rendement à l'usinage	:	70 %
• Cycle (à maturité)	:	157-161 Jours
• Poids de 1000 grains (paddy)	:	26,13 g
• Tallage	:	Bon
• Rendement moyen	:	4,5 – 5,5 t/ha

COMPORTEMENT VIS-A-VIS DES BIOAGRESSEURS

• Résistance à la pyriculariose	:	Résistante
• Résistance aux insectes	:	Sensible aux insectes de stockage

Autres traits spécifiques

- Gonflement a la cuisson : oui
- Gout : apprécié
- Egrenage : facile
- Verse: résistante
- Insensible a la photopériode
- Faible attractivité au Borer
- Forte adaptation aux conditions de fertilité faible

frontiers in Plant Science

ORIGINAL RESEARCH
published: 15 January 2020
doi: 10.3389/fpls.2019.01668

Identification of Loci Through Genome-Wide Association Studies to Improve Tolerance to Sulfur Deficiency in Rice

Juan Pariasca-Tanaka[†], *Cedric Baertschi*[†] *and Matthias Wissuwa*[*]

Crop, Livestock and Environment Division, Japan International Research Center for Agricultural Sciences (JIRCAS), Tsukuba, Japan

Sulfur (S) is an essential nutrient for plant growth and development; however, S supply for crop production is decreasing due to reduced inputs from atmospheric deposition and reduced application of S-containing fertilizers. Sulfur deficiency in soil is therefore becoming a widespread cause of reduced grain yield and quality in rice (*Oryza sativa* L). We therefore assessed the genotypic variation for tolerance to S deficiency in rice and identified loci associated with improved tolerance. Plants were grown in nutrient solution with either low (0.01 mM) or high (1.0 mM) supply of S. Plants grown under low-S treatment showed a reduction in total biomass, mainly due to a marked reduction in shoot biomass, while root biomass and root-to-shoot ratio increased, relative to plants under high-S treatment. Genome-wide association studies (GWAS) identified loci associated with root length (*qSUE2-3, qSUE4,* and *qSUE9*), and root (*qSUE1, qSUE2-1, and qSUE3-1 and qSUE3-2*) or total dry matter (*qSUE2, qSUE3-1,* and *qSUE11*). Candidate genes identified at associated loci coded for enzymes involved in secondary S metabolic pathways (sulfotransferases), wherein the sulfated compounds play several roles in plant responses to abiotic stress; cell wall metabolism including wall loosening and modification (carbohydrate hydrolases: beta-glucosidase and beta-gluconase) important for root growth; and cell detoxification (glutathione S-transferase). This study confirmed the existence of genetic variation conferring tolerance to S deficiency among traditional aus rice varieties. The advantageous haplotypes identified could be exploited through marker assisted breeding to improve tolerance to S-deficiency in modern cultivars in order to achieve sustainable crop production and food security.

Keywords: rice, sulfur, deficiency tolerance, quantitative trait locus, genome-wide association studies, *aus*

OPEN ACCESS

Edited by:
Jianjun Chen,
University of Florida, United States

Reviewed by:
Satoru Ishikawa,
National Agriculture and Food Research Organization (NARO), Japan
Khurram Bashir,
RIKEN, Japan
Suri Jiao,
BASF (United States), United States

***Correspondence:**
Matthias Wissuwa
wissuwa@affrc.go.jp

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *06 June 2019*
Accepted: *27 November 2019*
Published: *15 January 2020*

Citation:
Pariasca-Tanaka J, Baertschi C and Wissuwa M (2020) Identification of Loci Through Genome-Wide Association Studies to Improve Tolerance to Sulfur Deficiency in Rice. Front. Plant Sci. 10:1668. doi: 10.3389/fpls.2019.01668

INTRODUCTION

Sulfur (S) is an essential element for growth and development in living organisms; however, animal digestion, including that of humans, is unable to metabolize sulfur directly. Therefore, animals depend on the consumption of plant biomass for the supply of organic S, such as sulfur-containing amino acids (methionine and cysteine), vitamins (thiamine and biotin), and many other secondary compounds (Maruyama-Nakashita et al., 2003; Lunde et al., 2008; Kopriva et al., 2015).

Sulfur-containing amino acids are found in very small amounts in legumes, while most cereals contain high levels for example rice (*O. sativa L.*) contains 190% of the reference value, egg protein (Mariotti, 2017). Rice is a staple food for more than half of the world, providing more than 50% of the caloric supply for low-income persons. A reduction in rice yield and grain quality, due to inadequate amount of S supply, could therefore have a negative impact on the world's food security.

In plants, S is also an essential macronutrient needed in similar amounts to Phosphorus for growth and development. Apart from being a vital component of amino acids and vitamins, S is also a component of coenzymes, cofactors, lipoic acid, glutathione, sufloquinovosyl diacylglycerol, and many other secondary compounds (Suzuki, 1978; Lunde et al, 2008; Kopriva et al., 2015).

Sulfur is abundant in nature as sulfate (SO_4^{2-}); however, to utilize this abundant sulfate, plants must invest energy for its reduction. Plants can incorporate sulfate by both primary and secondary sulfate assimilation pathways. In the primary pathway, S is taken up as sulfate, then reduced to sulfite (SO_3^{2-}), sulfide (S^{2-}), and then incorporated into the amino acid skeleton of O-acetylserine to form cysteine. Subsequently, methionine and other compounds containing reduced forms of S are produced. In the secondary pathway, sulfate is used as a donor for the sulfation reaction, for synthesis of larger compound such as brasssinosteroids, sulfoflavonoids, and glucosinolates (Kopriva et al., 2015).

Symptoms of S deficiency in plants during vegetative stages includes initial chlorosis of young leaves, which spread to the entire plant, reduced tillering, and stunted growth. Symptoms of S deficiency are similar to that of N deficiency. Yellowing of older leaves is seen in N-deficient plants, while yellowing of young leaves occurs in S-deficient plants, since S does not move readily in the plant (Chandel et al., 2003).

Sulfur deficiency in soil was rare in the past; however, it has become increasingly inadequate in recent history, due to factors like increased removal of nutrients by intensive cropping (estimated at 1.8 kg S t^{-1} grain) (Dobberman and Fairhurst, 2000), reduced application of S-containing N and P fertilizers (ammonium sulfate and single superphosphate) (Yoshida and Chaudhry, 1979), and reduced atmospheric inputs (from reduced industrial pollution). Furthermore, strong solar irradiation increases the S requirement of plants, which could further exacerbate the effect of S deficiency in soil (Resurrection et al., 2001). Therefore, in the last 2 decades, reports of S deficiency in crops has increased worldwide (Scherer, 2001), including some Asian countries like China, India, Indonesia, Philippines, Sri Lanka, and Thailand (Blair et al, 1978). Similarly, the problem of S deficiency in crops has also been reported in many African countries (Yamaguchi, 1999; Tsujimoto et al., 2013).

Although the advantage of fertilizer application to alleviate S deficiency has been widely reported (Blair et al., 1980), the applicability of this practice by low-resource farmers is uncertain. Therefore, the identification and selection of rice genetic resources with enhanced tolerance to S-deficiency in soil, and improved S utilization efficiency would be very desirable to attain sustainable rice production and food security.

Recently, genome-wide association studies (GWAS) based on single nucleotide polymorphism (SNP) has emerged as an alternative method to study association mapping using genetically diverse rice populations. GWAS can determine the strength of the association between a genotype and phenotype and identify loci, genes and alleles that are associated with specific traits. In rice, GWAS has been successfully applied to dissect the genetic bases of several traits (Zhao et al., 2011), including seedling vigor (Wang et al., 2018), root morphology (Biscarini et al, 2016), flowering time (Zhao et al, 2011), yield and grain size (Feng et al., 2016; McCouch et al., 2016), as well as ozone and aluminum tolerance (Famoso et al, 2011; Ueda et al., 2015), phosphorus-utilization efficiency (Wissuwa et al., 2015), and root efficiency (Mori et al., 2016).

The rice *aus* group is a minor sub-population which originated from Bangladesh and India (Klush, 1997; Norton et al., 2018). It harbors novel loci that, confer tolerance to abiotic stress (Famoso et al., 2011; Lee et al., 2018), and enhance root development (Mori et al, 2016). Therefore, the *aus* group would be an excellent genetic reserve in the search for new sources of tolerance to S deficiency in rice.

Despite the growing problem of S deficiency in soil and its negative effects on crop yield and food security, the genetic aspect of S deficiency tolerance in rice is not yet well understood. We therefore aimed to use GWAS to identify candidate loci, conferring tolerance to S deficiency in an *aus* rice panel. We hypothesize that such association mapping will identify several loci associated to tolerance, which could be used for the improvement of sensitive modern cultivars, to attain sustainable crop production and food security of an important staple food, rice.

MATERIALS AND METHODS

Three experiments were conducted as part of this study. Since nutrient solutions typically contain large amounts of S as a companion-ion for major nutrients (K, Mg, Zn), a low-S nutrient solution was tested in experiment 1. This was followed by screening a rice GWAS panel (experiment 2) and a validation trial with contrasting genotypes (experiment 3).

EXPERIMENT 1

Seeds of different genotypes (IR64, DJ123, Nerica4, Tsipala, WAB56-104) were sown and grown in modified Yoshida nutrient solution (using de-ionized water) under different S-treatments: no-S (0-S), low-S (0.01 mM S) or high-S (1.0 mM S). The full-strength Yoshida solution (1X) is composed of: N, 2.86 mM (as NH_4NO_3); P, 0.05 mM; K, 1mM; Ca, 1mM; Mg, 1mM; Mn, 9 µM; Mo, 0.5 µM; B, 18.5 µM; Cu, 0.16 µM; Fe, 36 µM; Zn, 0.15 µM (modified from Yoshida et al., 1971). To obtain the low-S solution, all sources of S were substituted as follows: KCl for K_2SO_4, $MgC1_2$ for $MgSO_4$, $ZnC1_2$ for $ZnSO_4$, and $CuCl_2$ for $CuSO_4$ (**Supplementary Table S1 in Data Sheet 1**). A stock solution of 10 mM $MgSO_4$ was used as S source for

the low-S treatment. The experiment was set in a factorial (Sulfur treatment and genotype, 2 × 5) arrangement in a randomized complete block design (RCDB), with three replications.

EXPERIMENT 2

Plant Material and Growth Condition

A GWAS panel of 98 accessions belonging to the *aus* sub-species of rice was selected from the High-Density Rice Array (HDRA; McCouch et al., 2016). De-hulled seeds were obtained from the International Rice Research Institute (IRRI, Philippines), and multiplied at a JIRCAS experimental field on Ishigaki Island, Japan. Seeds from each genotype were disinfected using 1% bleach solution, germinated in petri dishes at 30°C for 2 days, transferred to a floating mesh, and grown for 8 days in a solution containing 0.1 mM Ca ($CaCl_2$) and 0.012 mM Fe (Fe-EDTA). Subsequently, plants were grown either under low-S or high-S conditions, as described above.

In screening for the ability of rice plants to efficiently utilize a limited amount of S, we employed a similar approach to that used by Rose et al. (2011) in screening for P utilization efficiency (PUE). If all accessions are grown individually in a container with a known amount of the limiting nutrient, and if it is assured that this amount was fully taken up, then the biomass produced can be directly used to estimate nutrient use efficiency as follows: biomass (g) per nutrient applied (mg). Thus, for the low-S treatment 10-day old plants of each accession were transferred to 1-L black bottles (two plants per bottle). Bottles contained 0.3× Yoshida solution with 0.01 mM S (total 0.32 mg S available for uptake by 2 plants; as for Experiment 1). Plants were grown for 10 days in these bottles and during this period it was expected that they deplete the low amount of S in the solution. This was confirmed by drawing a 10 mL aliquot from the bottles and analyzing its S concentration using ICP-AES (inductively coupled plasma optical-emission spectroscopy). Sulfur concentrations in aliquots were similar to that of the blank, confirming complete S uptake (data not shown). Having ensured that S was fully taken up, plants were transferred to a 200-L container and grown in a no-S Yoshida solution until harvest (30 days after sowing, DAS).

For the high-S treatment (control), the 10-day old plants were directly transferred and grown in a 200 L container containing high-S solution (1.0 mM S), until harvest. In general, the pH was frequently monitored and adjusted to a range of 5.7 - 5.9. A flowchart of the experiment is shown in **Supplementary Figure S1 in Presentation 1**.

The experiment was conducted in a controlled environment green house at JIRCAS, Tsukuba, Japan. Growth conditions were as follows: temperature fluctuated from 28 to 33°C day time and around 25°C during night time, and relative humidity ranged from 30-50%. Plants were exposed to natural light during the entire growth period. The experiment was performed in a factorial (Sulfur treatment and genotype, 2 × 98) arrangement in a randomized complete block design (RCDB), with three replications (block) carried out in consecutive periods from March to Aug of 2017.Parameters evaluated during harvests were: plant height, number of leaves (green and dead leaves), number of tillers, root length, and root and shoot dry matter.

Sulfur Determination

Sulfur concentration in aliquots of nutrient solution was determined using inductively coupled plasma atomic-emission spectrometry, (ICP-AES; ICPE-9000, Shimadzu, Japan) using an S standard solution (ICP grade, Wako, Japan) to prepare the standard curve. This measurement was performed to confirm the complete uptake of S in the low-S treatment. Three independent replications were taken per each sample (total number of samples: 5)

Association Mapping

For the phenotype dataset, the overall performance of all rice accessions for each trait was calculated as the best linear unbiased prediction (BLUP, Henderson, 1975) using the R software (R Core Team., 2013), with the following simple mixed model:

$$y_{ij} = \mu + g_i + \ r_j + e_{ij},$$

in which y_{ij} refers to phenotype, μ overall mean, g_i: accession effect considered as random, r_j: replicate effect considered as fixed, and e_{ij}: residual effect considered as random.

The association analysis was then performed using: a) the resulting phenotypic BLUP values; b) the HDRA genotypic dataset (McCouch et al., 2016) composed of 700 K SNPs evenly distributed across the rice genome; and, c) the software Trait Analysis by aSSociation, Evolution and Linkage 5.0 (TASSEL, Bradbury et al., 2007).

The genotype dataset was filtered as follows: heterozygotes were set as missing values, minor allele frequency (MAF) was set to 0.03, and minimum count to 105 (0.82). The resulting dataset after curation was then used for association mapping using mixed linear model procedure (MLM), 3 principal components (PCA) and a kinship matrix. The associated loci were determined based on significant threshold of –log(10)(p) >5 (arbitrary threshold), and peaks having at least three consecutive SNP (above the threshold) as described by Wang et al. (2018). The marker effects at the significant position were extracted from the output of the MLM procedure for both alleles.

Linkage Disequilibrium and Haplotype Analysis

Linkage disequilibrium (LD) analysis to define LD blocks surrounding the significant SNPs was performed by Haploview 4.2 (Barret et al., 2005), using confidence intervals (Gabriel et al., 2002).

Selection of Putative Candidate Genes

Gene models and their respective annotations were obtained from the Rice Annotation Project Database (RAP-DB, https://rapdb.dna.affrc.go.jp/) for each significant peak and their surrounding linkage block. However, genes annotated as '(retro)transposon', 'hypothetical' or 'unknown' were excluded

from the analysis. Putative candidate genes were then selected based on their annotated function and gene ontology (http://www.geneontology.org/). Gene expression pattern was obtained from the Rice XPro database (http://ricexpro.dna.affrc.go.jp/). Sequence for the genotype DJ123 was obtained from the Schatz Lab-Johns Hopkins University (http://schatzlab.cshl.edu/data/rice/).

Statistical Analysis

The effects of sulfur treatment, genotype and their interaction on different traits were estimated using a two-way ANOVA, and multiple comparisons were performed using Tukey's honest significant difference (HSD) *post hoc* test (Statistix 9.0 Software). Box plots were generated by R software (R Core Team, 2013).

EXPERIMENT 3

This experiment was performed to confirm the phenotypic effects identified in experiment 2, and to investigate the expression pattern of selected candidate genes. Seeds of a selected set of contrasting genotypes were grown and evaluated as described in experiment 2. The experiment was conducted with a factorial (Sulfur treatment and genotype, 2 × 2) arrangement in a randomized complete block design (RCDB) with four replications.

The selected contrasting genotypes were divided into two groups: a) Santhi Sufaid 207, Kangro, Juma and DJ123; and b) Surja Mukhi, Harbhoondi, Khadasiya3, and Andikulan. The first group harbors the favorable haplotypes for total dry matter. The Asian mega variety IR54 was included as reference.

Tissue Sampling, RNA Extraction and PCR Conditions

At harvest (30 das), four independent biological replications for each sample were quickly taken, flash-frozen in liquid nitrogen, and stored at -70°C until analysis. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen), following the manufacturer instruction manual. Total RNA (around 400 ng) was then reverse transcribed (RT) using the PrimeScript RT Enzyme Mix I (Takara, Japan). RT-PCR was performed using RT (first strand-cDNA) as template, gene-specific primers, and Taq polymerase enzyme (Takara, Japan).

Quantitative PCR (qPCR) was performed using 1 ng RT template and SYBR Premix ExTaq (Perfect Real Time, Takara, Japan), using the Mini Opticon Real-Time PCR system (BioRad, USA) as previously described by Pariasca-Tanaka et al. (2009). Serial dilutions of RT product were used to determine the efficiency of each primer. A set of rice genes including Elongation factor (ELF-1), Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and Ubiquitin (Ubi) was used as internal controls. Relative expression levels between samples were calculated using the delta–delta comparison and expressed as fold changes. The normalized data were analyzed by ANOVA. The list of primers used in this study is shown in **Supplementary Table S2 in Data Sheet 1**.

RESULTS

Experiment 1

Plants were grown in a modified Yoshida solution where all sources of S were substituted (refer to material and method) to test the response of a set of rice cultivars to S-deficiency treatments.

Plants grown without any supply of S (0-S) showed a very limited growth, and their leaves spots rapidly turned from chlorotic to necrotic (**Supplementary Figure S2 in Presentation 1**). In contrast, plants grown with 1 mM S (high S) showed normal growth while those grown in low-S treatment (0.01 mM) showed slow growth and their leaves exhibited the characteristic symptoms of deficiency for this nutrient (interveinal chlorotic spots). In addition, plants grown in low-S condition showed significant root elongation and increased root:shoot ratio when compared to those grown under high-S treatment (**Figures 1A-C**). Among the studied genotypes, DJ123 belonging to the *aus* sub-species, showed better adaption to low-S conditions when compared to modern cultivars IR64 and Nerica 4.

Experiment 2

Since DJ123 was the best performer in experiment 1, the *aus* panel was chosen for screening to identify loci associated with tolerance to S deficiency. Plants were grown in either low S (0.01 mM) or high S (1 mM) treatment. Result showed that low-S treatment caused a significant reduction in shoot and total biomass, where in the larger effect was the significant increase in root length and root biomass (**Table 1**). The initial characteristic chlorotic symptoms lead to an increase in the number of dead leaves. Moreover, significant reduction in plant height and number of leaves and tillers was also observed (**Table 1**).

A box plot shows the phenotypic variation in each evaluated trait as affected by S treatment (**Figure 2**). The relative difference between treatments ([low-S value minus high-S value]/[high-S value]) was largest for root:shoot ratio and root biomass, followed by root length and number of dead leaves (high-S condition), while number of leaves showed the least change between treatments. Shoot biomass, plant height and number of tillers values were higher in high-S condition.

A correlation analysis between phenotypic traits indicated that despite the increase in root:shoot ratio, the total dry matter remained most closely correlated with shoot dry matter, and moderately correlated with plant height and number of leaves, and weakly correlated with dead leaves (**Table 2**). Root dry matter and root length were moderately correlated, despite that both significantly increased under S-deficiency condition.

Association Mapping

The genotyping dataset (from HDRA 700 K SNPs) and the phenotypic BLUP values were used for association mapping in each S treatment. For low-S treatment, a Mixed Linear Model (MLM) identified several quantitative trait loci (QTLs) associated with root length in chromosome 2 (*qSUE2-3*), 4 (*qSUE4*) and 9 (*qSUE9*); root dry matter in chromosome 1 (*qSUE1)*, 2 (*qSUE2-1*) and 3 (*qSUE3-1 and qSUE3-2)*; and total dry matter in chromosome 2 (*qSUE2-2*), 3 (*qSUE3-1*), 6 (*qSUE6*), and 11 (*qSUE11*) (**Table 3**). Moreover, for the high-S (control) treatment several QTLs were

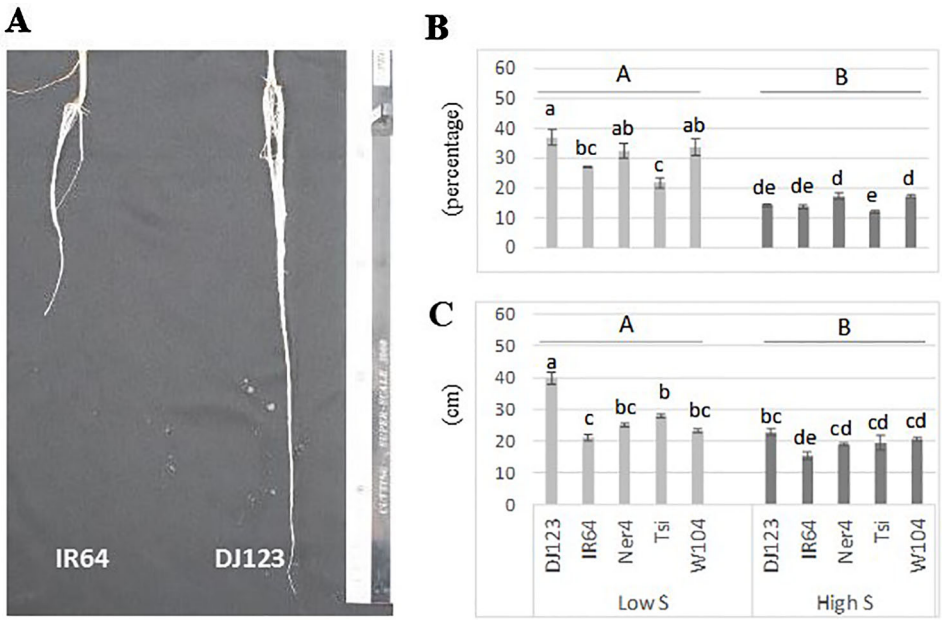


FIGURE 1 | Root elongation **(A)**, root:shoot ratio **(B)** and root length **(C)** of rice plants grown under low-S and high-S treatment. Data represent the mean (± standard error, SE) of three independent replications (n=3) and analyzed by two-way ANOVA. Different letters indicate significant difference within either factor: S treatment (capital letter) or genotype (small letter) according to Tukey's Honest Significant Difference test ($P < 0.05$). ANOVA results are provided in **Supplementary Table S4-1 in Data Sheet 1**. Cultivars, DJ123; IR64; Ner4, NERICA4; Tsi, Tsipala; W104, WAB56-104.

TABLE 1 | Summary of two-way analysis of variance (ANOVA) from phenotypic traits measured in plants grown under low-S and high-S treatment.

	Plant height (Ht)	Maximum root length (RLmax)	Number of tillers (Till)	Number of leaves (Lvs)	Number of dead leaves (deadL)	Root dry matter (RDM)	Shoot dry matter (ShDM)	Total dry matter (TotDM)	Root/shoot ratio (r/sh)
Genotype (G)	***	***	***	***	***	***	***	***	***
Sulfur treatment (S)	***	***	***	***	***	***	***	***	***
LS	61.9	52.5	1.5	9.0	3.2	290.1	456.4	746.5	66
HS	82.1	24.9	2.3	11.4	1.5	141.6	618.1	759.7	23.3
GxS	***	***	*	***	***	*	***	***	ns
n	364	364	364	364	364	364	364	364	364
Mean	71.9	38.7	1.9	10.2	2.4	214.8	533.5	748.3	44.6
SE	0.7	0.7	0.1	0.1	0.1	5.5	10.3	12.8	1.2
Min	34	13	1	1	1	47	123.5	193.5	13.2
Max	106	69	4	31	6	541	1313.5	1552	95.83

*ns, non-significant (P > 0.05), *significant (P ≤ 0.05), **(P ≤ 0.01), ***(P ≤ 0.01). LS, Low-S; HS, high-S treatment.*

identified; though in different chromosomes compared to low-S treatment. The QTLs were found in chromosome 4 and 6 for root length; in chromosome 7 and 12 for root dry matter; and in chromosome 2 and 7 for total dry matter (**Table 3**).

The minor allele frequency and effect of alleles were also determined in each locus (**Table 3**). Minor alleles exhibited favorable effects on root dry matter in *qSUE1,* and *qSUE3-1*. Likewise, for total dry matter the minor alleles had higher effect in *qSUE2-2, qSUE3-1, qSUE6,* and *qSUE11*. In the opposite, for root length (*qSUE2)* the minor allele had a small effect compared to the major allele.

The Manhattan plots and QQ plots for each associated QTL are shown in **Figure 3** (Low-S treatment) and **Supplementary Figure S3-1 in Presentation 1** (High-S treatment). A complementary GWAS analysis was performed using the phenotypic ratio of low S/high S result showed that several QTLs were consistently found in both analysis approaches (low-S and ratio Low/high S) (**Supplementary Figure S3-2 in Presentation 1**).

Selection of Putative Candidate Genes

The LD block for several significant loci is shown in **Supplementary Figure S4 in Presentation 1**. LD block was

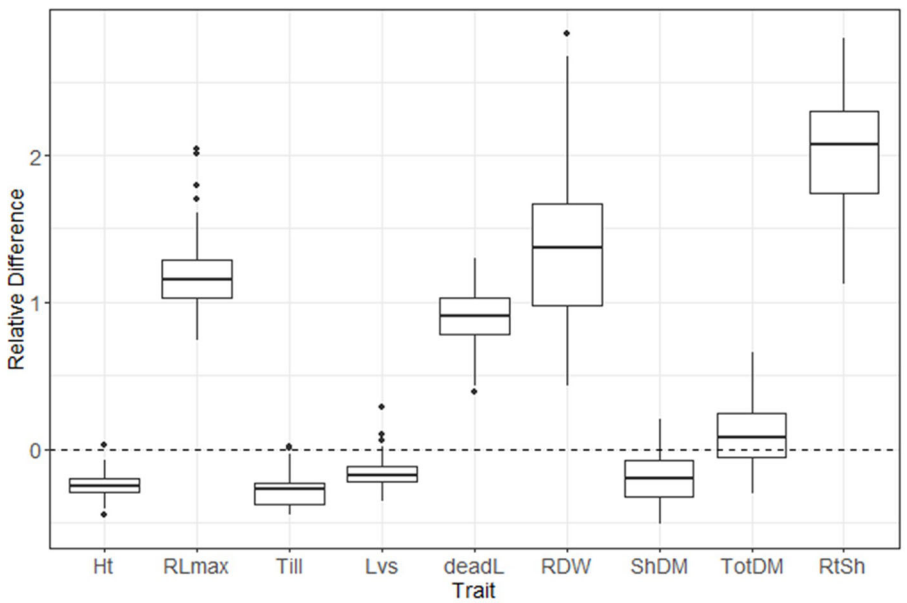


FIGURE 2 | Distribution of relative difference ([low S - high S]/high S) values from evaluated phenotypic parameters. Thick lines inside the box represent the median of the distribution, while the lower and upper boundaries represent the first and third quartile, respectively. Relative difference was calculated as: (low-S minus high-S)/high-S. Ht, plant height; RLmax, maximum root length; Till, number of tillers; Lvs, number of leaves; deadL, number of dead leaves; RDM, root dry matter; ShDM, shoot dry matter; totDM, total dry matter; RtSh, root:shoot ratio.

TABLE 2 | Correlation coefficient between phenotypic traits measured in plants grown in low-S treatment.

	Ht	RLmax	Till	Lvs	deadL	RDM	ShDM	TotDM
Plant height (Ht)								
Maximum root length (RLmax)	**0.288** **							
Number of tillers (Till)	**0.419** ***	**0.169** *						
Number of leaves (Lvs)	**0.420** ***	**0.125** *ns*	**0.786** ***					
Number of dead leaves (deadL)	**-0.049** *ns*	**0.127** *ns*	**0.180** *ns*	**0.370** ***				
Root dry matter (RDM)	**0.659** ***	**0.285** **	**0.359** ***	**0.337** ***	**-0.155** *ns*			
Shoot dry matter (ShDM)	**0.800** ***	**0.208** *	**0.592** ***	**0.566** ***	**-0.078** *ns*	**0.836** ***		
Total dry matter (TotDM)	**0.778** ***	**0.244** *	**0.528** ***	**0.503** ***	**-0.112** *ns*	**0.932** ***	**0.978** ***	
Root/shoot ratio (r/sh)	**-0.461** ***	**-0.051** *ns*	**-0.510** ***	**-0.512** ***	**-0.171** *	**0.032** *ns*	**-0.486** ***	**-0.308** **

*ns, non-significant (P > 0.05), *significant (P ≤ 0.05), **(P ≤ 0.01), ***(P ≤ 0.01).*

clearly delineated for most peaks, except for that of *qSUE3*. Gene models contained in the LD block were retrieved from RAP-DB, and those annotated as unknown and hypothetical proteins were removed. Putative candidate genes were therefore selected based on their functional annotation and GO molecular function (**Table 4**, and **Supplementary Table S3 in Data Sheet 1**).

Candidates genes for root length, locus (*qSUE2*), would be glutathione S-transferase (Os02g0564000), the BHLH transcription factor (Os02g0564700), and the PTO kinase interactor (Os02g0565500). In the case of *qSUE4* the candidate is a *protein containing Leucine rich repeat domain (Os04g0647900),* while for *qSUE9*, candidate genes are ubiquitin conjugating enzyme

TABLE 3 | Quantitative trait loci (QTL) associated to root length, root, and total dry matter in GWAS using a mixed linear model (MLM).

Trait	Loci name	Chr	SNP denomination	SNP position (bp)	P	Minor allele	
						[1]MAF	[2]Effect
Root length							
LS	***SUE2-3***	2	SNP-2.21491708	21,497,578	3.40E-06	0.17	5.4
	SUE4	4	SNP-4.32776574	32,961,688	5.39E-06	0.35	2.4
	SUE9	9	SNP-9.15018718	15,019,720	3.63E-07	0.34	5.2
HS		4	SNP-4.24473685	24,658,824	7.62E-06	0.47	3.05
		6	SNP-6.21637246	21,638,244	7.38E-06	0.06	5.29
Root Dry Matter							
LS	***SUE1***	1	SNP-1.12111994	12,113,021	6.20E-06	0.14	49.1
	SUE2-1	2	SNP-2.4411478	4,411,482	9.23E-06	0.41	79.8
	SUE3-1	3	SNP-3.19346088	19,347,578	2.32E-07	0.10	54.3
	SUE3-2	3	SNP-3.28010215	28,017,162	6.38E-06	0.41	76.3
HS	***SUE7***	7	SNP-7.22383573	22,384,567	4.75E-06	0.08	32.9
	SUE12	12	SNP-12.22908597	22,942,139	9.63E-06	0.05	46.7
Total Dry Matter							
LS	***SUE2-2***	2	SNP-2.16027627	16,033,498	1.40E-06	0.08	335.4
	SUE3-1	3	SNP-3.19346088	19,347,578	2.75E-06	0.10	299.1
	SUE6	6	SNP-6.292614	293,615	5.35E-06	0.06	442.8
	SUE11	11	SNP-11.17486608	17,952,751	7.06E-06	0.04	480.1
HS		2	SNP-2.22318649	22,324,519	6.90E-06	0.28	181.8
		7	SNP-7.7306443	7,307,439	2.94E-06	0.05	325.9

[1]MAF, minor allele frequency. [2]Marker effect.

binding group (Os09g0419500) and CBL-interacting protein kinase 16 (Os09g0418000).

For root dry matter, the candidates for locus *qSUE1-1* would be genes encoding pectin acetylesterase (Os01g0319000) and ABC1 protein (Os01g0318700). Starch debranching enzyme (Os02g0164900) and peptidase S26A protein (Os02g0165600) are candidates for *qSUE2-1,* while genes encoding interferon-related developmental regulator (Os03g0448700) and serine/threonine protein kinase (Os03g0556600) are candidates for *qSUE3*-1. For *qSUE3*-2, the candidate encodes for lipoxigenases (Os03g0700700)

For total dry matter, candidates for *qSUE2-2* are a protein phosphatase (Os02g0471500) and protein kinase (Os02g0472700), while a pectinesterase protein (Os06g0104300) for *qSUE6.* For locus *qSUE11*, there were two genes encoding sulfotransferase (Os11g0503900 and Os11g0505300) flanking the LD block (**Table 4**).

Experiment 3

An additional experiment was carried out to confirm the performance of a set of genotypes with contrasting haplotypes associated with total dry matter. The genotypes: Santhi Sufaid 207, Kangro, Juma and DJ123, which harbor the favorable haplotypes for root and total dry matter, showed a significant increase in root length, root and total dry matter, and root:shoot ratio compared to the unfavorable haplotype group: Surja Mukhi, Harbhoondi, Khadasiya3, and Andikulan. The Asian mega variety IR64 was included as reference (although its haplotype was not determined) (**Figure 4** and **Supplementary Table S4 in Data Sheet 1**, ANOVA). A complementary analysis using relative values for low S/high S is presented in **Supplementary Figure S5 in Presentation 1**.

Gene Expression

The *qSUE11* is surrounded by two genes: *qSUE11-1*, Os11g0503900 (cds: 987 nuc, amino acids: 328), and *qSUE11-2*, Os11g0505300 (cds: 1188 nuc, amino acids: 395). Both genes encode the sulfotransferase family protein since they contain the Sulfotransfer_1 domain, PF00685 (Hirschmann et al., 2014). An alignment comparison between both genes showed 55.2 and 62% similarity at the nucleotide and amino acid sequence, respectively (**Supplementary Figure S6-1 in Presentation 1**). The gene Os11g0503900 was partially sequenced in Surja Mukhi and Santhi Sufaid 207 (harboring the favourable- and unfavorable haplotype, respectively). Result showed no difference in their nucleotide alignment (**Supplementary Figure S6-2 in Presentation 1**). In addition, a BLAST search for both genes against the genome sequence of accession DJ123 (Schatz Lab-Johns Hopkins University, Schatz et al., 2014) showed the

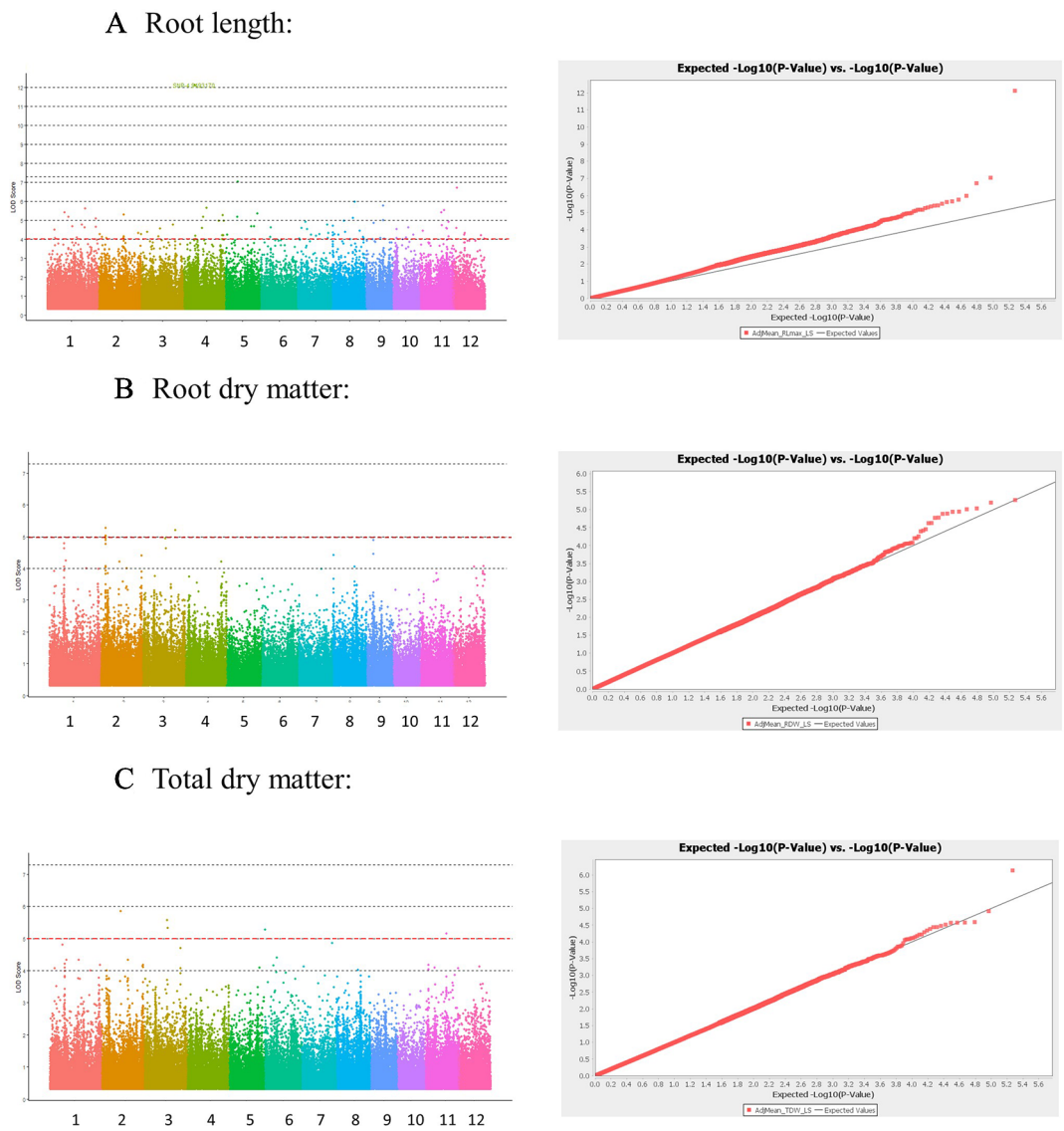


FIGURE 3 | Manhattan and Quantile-quantile (QQ) plots derived from GWAS analysis for traits: **(A)** Root length, **(B)** Root dry matter, and **(C)** total dry matter. Manhattan plot shows negative logarithmic ((-log10 (*P*)) values of association for each SNP (Y axis), and SNP location along the 12 chromosomes (colored bar in X axis). Red line indicates a -log10 (P value) threshold of 5. QQ plots X shows the expected -log10 (p value) on X-axis, and measured -log10 (p value) on Y-axis.

presence of SNPs and indel for both cases: Os11g0503900 (98.6% similarity, 3 indel and 16 SNP) and Os11g0505300 (98.7% similarity, with single 6 indel, and 2 short indel) (**Supplementary Figures S6-3**, **S6-4 in Presentation 1**).

RT-qPCR analysis indicated that the transcript-level abundance of both sulfotransferase genes Os11g0503900 and Os11g0505300 was consistently low (**Figure 5**), as it was reported by Chen et al. (2012) and Rice X-Pro. The transcript level of both genes was not induced by S treatment; however, higher transcript level was found for Os11g0503900 in shoot tissue of the favorable-haplotype group. ANOVA result is presented in **Supplementary Table S4 in Data Sheet 1**.

DISCUSSION

To date, most work related to S deficiency has focused on Arabidopsis; this study thus aimed to elucidate the genetic aspect of tolerance to S deficiency in rice, an agronomically important cereal crop model. Our initial experiment (Experiment 1) testing a modified Yoshida solution, with all sources of S substituted (**Supplementary Table S1 in Data Sheet 1**) indicated that the low-S supply (0.01 mM) was enough to cause measurable symptoms in 30-day-old plants, which could be used to screen and identify plants that are tolerant to S-deficiency.

TABLE 4 | List of candidate genes in each QTL.

RAPdb	MSU (LOC)	Annotation	Reference
Root length - *qSUE2*-3			
Os02g0564000	Os02g35590	Similar to glutathione S-transferase.	Kumar and Trivedi, 2018
Os02g0564400	Os02g35630	Similar to CLPX (Clp protease regulatory subunit X) 3B ATPase.	
Os02g0564700	Os02g35660	Similar to BHLH transcription factor.	Hudson and Hudson, 2015
qSUE4			
Os04g0647300	Os04g55360	Similar to H0811D08.10 protein.	
Os04g0647900	Os04g55420	Leucine-rich repeat 2C N-terminal domain containing protein	
qSUE9			
Os09g0418000	Os09g25090	Similar to CBL-interacting protein kinase 16.	
Os09g0419500	Os09g25190	Similar to ubiquitin-protein ligase/zinc ion binding protein.	Shu and Yang, 2017
Root dry matter - *qSUE1*			
Os01g0318700	Os01g21610	Similar to ABC1 protein (Fragment).	
Os01g0319000	Os01g21630	Similar to pectin acetylesterase	
Os01g0318000	Os01g21580	Similar to esterase/lipase/thioesterase family protein.	
qSUE2-1			
Os04g0164900		Similar to starch debranching enzyme	
Os04g0165600	Os04g08340.1	Peptidase S26A signal peptidase I family protein	
***qSUE3*-1**			
Os03g0448700	Os03g33590	Interferon-related developmental regulator domain containing protein.	
Os03g0556600	Os03g35600	Serine/threonine protein kinase-related domain containing protein.	Diédhiou et al., 2009
qSUE3-2			
Os03g0700700	Os03g49380	Similar to lipoxygenase	
Total dry matter - *qSUE2-2*			
Os02g0471500	Os02g27220	Protein phosphatase 2C domain containing protein	
Os02g0472700	Os02g27310	Similar to receptor-like serine-threonine protein kinase	
qSUE6			
Os06g0104300	Os06g01490	Similar to pectinesterase-like protein	
Os06g0104200	Os06g01480.1	Similar to OsNAC7 protein	
qSUE11			
Os11g0503900	Os11g30810	Sulfotransferase family protein.	Chen et al., 2012
Os11g0505300	Os11g30910	Sulfotransferase 2C resistance to rice stripe virus	

In the same experiment, the genotype DJ123 (belonging to the *aus* sub-species) showed better performance compared to modern rice cultivars (IR64 and Nerica 4), indicating that the *aus* group could be a good target for further screening. The next step of our study, therefore, focused on the screening of the *aus* group, which is known for their adaptability to unfavorable environments. For example, the genotype FR13A harbors the submergence tolerance gene Sub1 (Xu et al., 2006), and Kasalath harbors the phosphorus starvation tolerance gene *Pstol1* (Gamuyao et al., 2012). In addition, DJ123 was shown to have high PUE (Wissuwa et al., 2015) and root efficiency (Mori et al., 2016).

Rice plant response to S deficiency started with leaf yellowing and visible reduction of shoot growth, followed by physiological and morphological modification of the roots, resulting in an increase in the root:shoot ratio, as deficiency became more severe. This S-deficiency response is related to adaptations within the S metabolic pathway first, and then related to the interaction with metabolic pathways of other nutrients (Lewandowska and Sirko, 2008; Kopriva et al., 2015).

In our experiment, S deficiency differentially induced root elongation and enhanced root growth in 30-day old plants (**Figure 2**). Similar responses have been reported for N, P and Magnesium (Mg) starvation (Hermans et al., 2006; Wissuwa et al., 2015), indicating that plants share a common mechanism to favor root growth to improve their ability to search and acquire scarce mineral nutrients from the soil. However, the two-way ANOVA revealed significant interaction effects on GxS (genotype by environment) for several traits including root length and root biomass, indicating that response to Sulfur treatment varies among the members of the aus group. Depending upon the breeding purpose, GxS interaction would

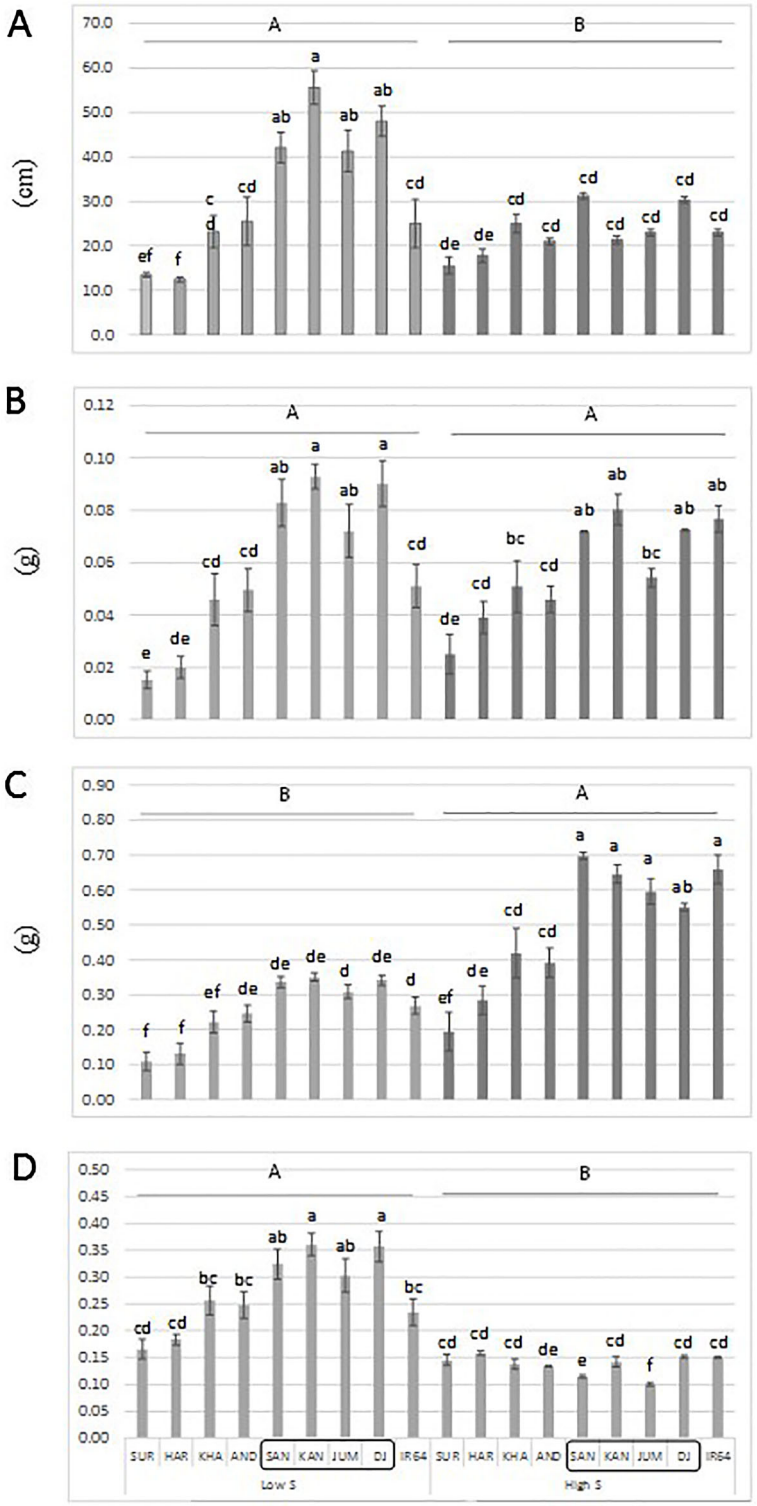


FIGURE 4 | Root length **(A)**, root dry matter **(B)**, total dry matter **(C)** and root:shoot ratio **(D)** of a set of *aus* genotypes grown under low-S or high-S condition. Genotypes included in the box harbor the favorable haplotype for total dry matter. Data represent the mean (± standard error, SE) of three independent replications (n = 3) and analyzed by two-way ANOVA. Different letters indicate significant difference within either factor: S treatment (capital letter) or genotype (small letter) according to Tukey's Honest Significant Difference test ($P < 0.05$). ANOVA results are provided in **Supplementary Table S4-2 in Data Sheet 1**. SUR, Surja Mukhi; HAR, Harbhoondi; KHA, Khadasiya3; CHU, Chungur Bali; AND, Andikulan; NP, NP 125; SAN, Santhi Sufaid207; KAN, Kangro; JUM, Juma; NCS, NCS 160; DJ, DJ123.

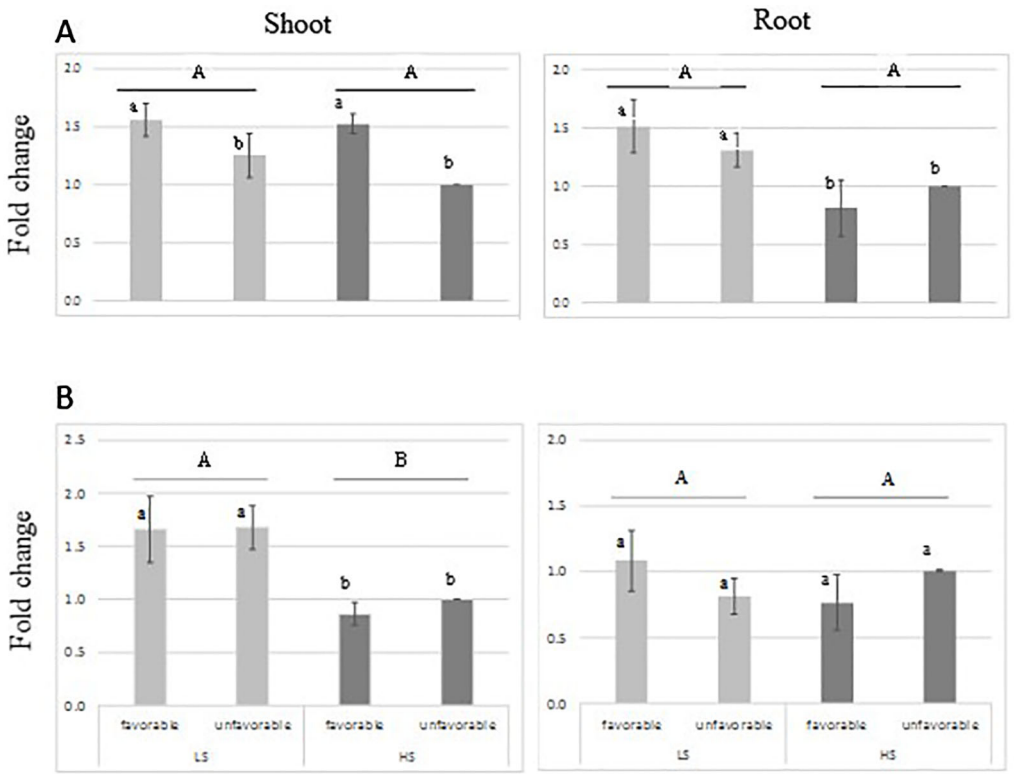


FIGURE 5 | Relative expression of sulfotransferase genes: Os11g0503900 **(A)** and Os11g0505300 **(B)** in root and leaves tissue from genotypes harboring favorable/unfavorable haplotypes for total dry matter. Plants were grown either under low-S or high-S condition. Data represent the mean (± standard error, SE) of three independent replications (n=4) and analyzed by two-way ANOVA. Different letters indicate significant difference within either factor: S treatment (capital letter) or haplotype group (small letter) according to Tukey's Honest Significant Difference test ($P < 0.05$). ANOVA results are provided in **Supplementary Table S4-3 in Data Sheet 1**. Favorable haplotype group: Santhi Sufaid 207, Kangro, Juma and DJ123. Unfavorable group: Surja Mukhi, Harbhoondi, Khadasiya3 and Andikulan.

be the very helpful to identify genotypes with good adaptation to specific S-deficiency conditions or with good ability to respond to broader nutritional problems.

We have further analyzed the trait association using two approaches: comparing the GWAS result for low S (**Figure 3**) and high S (**Supplementary Figure S3-1 in Presentation 1**) separately, and by calculating the association for the ratio low S/high S of phenotypic values. The GWAS comparison in two contrasting nutritional conditions (low S and high S) showed no common response. The identified QTLs did not co-localize indicating that specific and independent genetic effects were controlling the phenotypic response in both conditions (Rebolledo et al., 2016). On the other hand, GWAS result from relative values (low S/high S) revealed the presence of QTLs in greater number than in low-S treatment. Although the Manhattan plots showed many significant SNPs, which were rather scattered with undefined peak shape, there were several SNPs that co-localized in both analysis (**Supplementary Figure S3-2 in Presentation 1**), indicating that both approaches could be used complementarily.

The GWAS result indicated that the *aus* group possesses enough genetic variations for tolerance to S deficiency as several novel QTLs associated with agronomic traits, such as root length, and root and total dry matter were identified (**Table 3**). Among the favorable alleles in the significant loci there were two minor alleles for root dry matter and four for total dry matter.

Subsequently, based on the delineated LD blocks (**Supplementary Figure S5 in Presentation 1**), candidate genes with annotations putatively related to S metabolism and/or related processes to plant growth and stress response were selected for each locus.

A common response to S deficiency in most *aus* accessions was an increase in root length, and GWAS analysis identified one locus for which the common major haplotype increased root length, while only 17% of accessions with the minor haplotype had roots about 5% increase in length. Though these alleles appear to be the common alleles, this locus would be of little practical interest (from the agronomical/breeding point of view), candidate genes at this locus indicate the involvement in: a) regulation of growth in roots and flowers, and response to various stresses (helix-loop-helix transcription factor, Pires and Dolan, 2010; Hudson and Hudson, 2015); b) plant detoxification reaction during abiotic stress, like cadmium and arsenic detoxification (glutathione S-transferase, Kumar and Trivedi, 2018) and plant development (Gong et al., 2005); c) protein ubiquitination which plays roles in several plant developmental

stages and several abiotic stress responses (ubiquitin-protein ligase enzymes; Sharma et al., 2016; Shu and Yang, 2017).

For root dry matter, main candidates belong to the family of a protein kinase (serine/threonine protein kinase) which is involved in the regulation of epidermal cell morphogenesis, root hair elongation, and possibly in plant growth and stress tolerance (Diédhiou et al., 2009).

For total dry matter, the sulfotransferase genes were selected as candidate genes because of their involvement in the secondary S-assimilation pathway. Sulfotransferase catalyzes the sulfation of several compounds to produce sulfate esters, sulfamates, and sulfate conjugates (Hirschman et al., 2014). In our study their expression pattern suggests that the one gene is differentially expressed in shoot tissue, but not induced by S-deficiency condition (**Figure 5**). Their transcript level abundance was consistently low, which agrees with the finding of Chen et al. (2012). The reported expression of Sulfotransferase genes suggests their involvement in plant defense, stress response, signaling and developmental regulation (Klein and Papenbrock, 2004; Hirschman et al., 2014), although their involvement in tolerance to sulfur tolerance is not yet understood. Further studies of gene expression at earlier stage of stress could contribute to elucidate their role.

The superior performance of a set of aus genotypes (Santhi Sufaid 207, Kangro, Juma and DJ123) harboring the positive haplotypes for root and total dry matter was confirmed in this study (**Figure 4**). Although, no difference among genotypes was found for total dry matter when relative values low S/high S was analyzed (**Supplementary Figure S4 in Presentation 1**), these genotypes with consistent superior performance would be agronomically desirable when selecting potential donors.

It is well known that S and N metabolism are highly integrated, for example, S deficiency can cause a reduction of synthesis of proteins, and accumulation of nitrogenous compounds, as well as lowering the utilization of available N in soil. Since the N level in low-input agrosystem could exacerbate the S deficiency problem, rice accessions/cultivars with S and N use efficiency should be considered in breeding program targeting yield and quality of rice.

To our knowledge, there are no reported QTLs for tolerance to S deficiency in rice, therefore this study represents a significant advance in current efforts to reduce S deficiency in agriculture. Furthermore, the rice accessions: Santhi Sufaid 207, Kangro, Juma and DJ123 could be used as donors to improve local cultivars (**Figure 4**). Although the *aus* group still has some negative agronomic traits, such as excessive plant height, shattering and lodging, the beneficial loci identified here can be transferred through conventional breeding and using the associated SNPs for marker assisted selection.

In summary, this study demonstrates the existence of genetic variability for tolerance to S-deficiency in the *aus* rice group. GWAS identified several QTLs associated with root and total dry matter where in the advantageous alleles were minor alleles, which could be exploited for improving modern rice cultivars, through marker assisted selection. In addition, the selected accessions, including DJ123 (which had already been identified as having high PUE and efficient P uptake), could be considered as donors in breeding for low-input conditions, in systems characterized by multiple nutrient deficiencies. This would allow for the development of new rice cultivars with tolerances to a range of nutrient deficiencies and overall improved nutrient-use efficiency. Thus, contributing towards efforts in achieving sustainable rice production and improved food security worldwide.

DATA AVAILABILITY STATEMENT

The SNP dataset used in this study (High Density Rice Array, HDRA) is available at the Rice diversity website: http://www.ricediversity.org/data. All phenotypic data produced in this study are included in the **Supplementary Material (Supplementary Table S5 in Data Sheet 2)**.

AUTHOR CONTRIBUTIONS

JP-T, CB, and MW designed the experiments. CB and JP-T perform the experiments. JP-T and MW wrote the manuscript.

FUNDING

This research was partially supported by the Science and Technology Research Partnership for Sustainable Development (SATREPS), Japan Science and Technology Agency (JST)/Japan International Cooperation Agency (JICA) Project (Grant No. JPMJSA1608).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01668/full#supplementary-material

REFERENCES

Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21 (2), 263–265. doi: 10.1093/bioinformatics/bth457

Biscarini, F., Cozzi, P., Casella, L., Riccardi, P., Vattari, A., Orasen, G., et al. (2016). Genome-wide association study for traits related to plant and grain morphology, and root architecture in temperate rice accessions. *PloS One* 11, e0155425. doi: 10.1371/journal.pone.0155425

Blair, G. J., Mamaril, C. P., and Momuat, E. (1978). Sulfur-nutrition of wetland rice. *IRRI Res. Pap.* 21, 29.

Blair, G. J. (1980). *Priorities for Alleviating Soil-Related Constraints to Food Production in the Tropics* (Los Banos, Philippines: IRRI), 251.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., et al. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635 doi: 10.1093/bioinformatics/btm308

Chandel, R. S., Singh, K., Singh, A. K., and Sudhakar, P. C. (2003). Effect of sulphur nutrition in rice (Oryza sativa L.) and mustard (Brassica juncea L. czern and coss.) grown in sequence. *Indian J. Plant Physiol.* 8, 155–159.

Chen, R., Jiang, Y., Dong, J., Zhang, X., Xiao, H., Xu, Z., et al. (2012). Genome-wide analysis and environmental response profiling of SOT family genes in rice (Oryza sativa). *Genes Genom.* 34, 549–560. doi: 10.1007/s13258-012-0053-5

Diédhiou, C. J., Popova, O. V., Dietz, K. J., and Golldack, D. (2009). The SNF1-type serine-threonine protein kinase SAPK4regulates stress-responsive gene expression in rice. *BMC Plant Biol.* 20088, 49. doi: 10.1186/1471-2229-8-49

Dobberman, A., and Fairhurst, T. (2000). *Rice: Nutrient Disorders & Nutrient Management* (Philippines: Potash & Phosphate Institute (PPI), Potash & Phosphate Institute of Canada (PPIC) and International Rice Research Institute (IRRI)), 94.

Famoso, A. N., Zhao, K., Clark, R. T., Tung, C. W., Wright, M. H., et al. (2011). Genetic architecture of aluminum tolerance in rice (Oryza sativa) determined through genome-wide association analysis and QTL mapping. *PloS Genet.* 7, e1002221. doi: 10.1371/journal.pgen.1002221

Feng, Y., Lu, Q., Zhai, R., Zhang, M., Xu, Q., Yang, Y., et al. (2016). Genome wide association mapping for grain shape traits in indica rice. *Planta* 244, 819–830. doi: 10.1007/s00425-016-2548-9

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy., J., Blumenstiel, B., et al. (2002). The structure of haplotype blocks in the human genome. *Science* 296, 2225–2229. doi: 10.1126/science.1069424

Gamuyao, R., Chin, J. H., Pariasca-Tanaka, J., Pesaresi, P., Catausan, S., Dalid, C., et al (2012). The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. *Nature* 488, 535–539. doi: 10.1038/nature11346

Gong, H., Jiao, Y., Hu, W. W., and Pua, E. C. (2005). Expression of glutathione-S-transferase and its role in plant growth and development in vivo and shoot morphogenesis in vitro. *Plant Mol. Biol.* 57, 53–66. doi: 10.1007/s11103-004-4516-1

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423–447. doi: 10.2307/2529430. JSTOR 2529430. PMID 1174616.

Hermans, C., Hammond, J. P., White, P. J., and Verbruggen, N. (2006). How do plants respond to nutrient shortage by biomass allocation? *Trends Plant Sci.* 11, 610–617. doi: 10.1016/j.tplants.2006.10.007

Hirschmann, F., Krause, F., and Papenbrock, J. (2014). The multi-protein family of sulfotransferases in plants: composition, occurrence, substrate specificity, and functions. *Front. Plant Sci.* 5, 556. doi: 10.3389/fpls.2014.00556

Hudson, K. A., and Hudson, M. E. (2015). A classification of basic helix-loop-helix transcription factors of soybean. *Int. J. Genomics* 603182, 10. doi: 10.1155/2015/603182

Klein, M., and Papenbrock, J. (2004). The multi-protein family of Arabidopsis sulphotransferases and their relatives in other plants species. *J. Exp. Bot.* 55, 1809–1820. doi: 10.1093/jxb/erh183

Klush, G. S. (1997). Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* 35, 25–34. doi: 10.1023/A:1005810616885

Kopriva, S., Calderwood, A., Weckopp, S. C., and Koprivova, A. (2015). Plant sulfur and Big Data. *Plant Sci.* 241, 1–10. doi: 10.1016/j.plantsci.2015.09.014

Kumar, S., and Trivedi, P. K. (2018). Glutathione S-Transferases: role in combating abiotic stresses including arsenic detoxification in plants. *Front. In Plant Sci.* 9, 751. doi: 10.3389/fpls.2018.00751

Lee, J. S., Wissuwa, M., Zamora, O., and Ismail, A. (2018). Novel sources of aus rice for zinc deficiency tolerance identified through association analysis using high-density SNP array. *Rice Sci.* 25, 293–296. doi: 10.1016/j.rsci.2018.08.004

Lewandowska, M. S., and Sirko, A. (2008). Recent advances in understanding plant response to sulfur-deficiency stress. *Acta Biochim. Polonica* 55, 457–471.

Lunde, C., Zygadlo, A., Simonsen, H. T., Nielsen, P. L., Blennow, A., and Haldrup, A. (2008). Sulfur starvation in rice: the effect on photosynthesis, carbohydrate metabolism, and oxidative stress protective pathways. *Physiol. Plant* 134, 508–521. doi: 10.1111/j.1399-3054.2008.01159.x

Mariotti, F. (2017). "Plant protein, animal protein and protein quality," in *Vegetarian and Plant-Based Diets in Health and Disease Prevention*. Ed. François Mariotti. (USA: Academic Press), 621–643. doi: 10.1016/B978-0-12-803968-7.00035-6

Maruyama-Nakashita, A., Inoue, E., Watanabe-Takahashi, A., Yarnaya, T., and Takahashi, H. (2003). Transcriptome profiling of sulfur-responsive genes in Arabidopsis reveals global effects of sulfur nutrition on multiple metabolic pathways. *Plant Physiol.* 132, 597–605. doi: 10.1104/pp.102.019802

McCouch, S. R., Wright, M. H., Tung, C. W., Maron, L. G., Mcnally, K. L., Fitzgerald, M., et al. (2016). Open access resources for genome-wide association mapping in rice. *Nat. Commun.* 7, 10532. doi: 10.1038/ncomms10532

Mori, A., Fukuda, T., Vejchasarn, P., Nestler, J., Pariasca-Tanaka, J., and Wissuwa, M. (2016). The role of root size versus root efficiency in phosphorus acquisition in rice. *J. Exp. Bot.* 67, 1179–1189. doi: 10.1093/jxb/erv557

Norton, G. J., Travis, A. J., Douglas, A., Fairley, S., Alves, E. D., Ruang-areerate, P., et al. (2018). genome wide association mapping of grain and straw biomass traits in the rice bengal and assam aus panel (BAAP) grown under alternate wetting and drying and permanently flooded irrigation. *Front. Plant Sci.* 9, 1223. doi: 10.3389/fpls.2018.01223.

Pariasca-Tanaka, J., Satoh, K., Rose, T., Mauleon, R., and Wissuwa, M. (2009). Stress response versus stress tolerance: a transcriptome analysis of two rice lines contrasting in tolerance to phosphorus deficiency. *Rice* 2, 167. doi: 10.1007/s12284-009-9032-0

Pires, N., and Dolan, L. (2010). Origin and diversification of basic-helix-loop-helix proteins in plants. *Mol. Biol. Evol.* 27, 862– 874. doi: 10.1093/molbev/msp288

R Core Team. (2013). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). http://www.R-project.org/

Rebolledo, M. C., Peña, A. L., Duitama, J., Cruz, D. F., Dingkuhn, M., Grenier, C., et al. (2016). Combining image analysis, genome wide association studies and different field trials to reveal stable genetic regions related to panicle architecture and the number of spikelets per panicle in rice. *Front. Plant Sci.* 7, 1384. doi: 10.3389/fpls.2016.01384

Resurrection, A. P., Makino, A., Bennett, J., and Mae, T. (2001). Effects of sulfur nutrition on the growth and photosynthesis of rice. *Soil Sci. Plant Nutr.* 47, 611–620. doi: 10.1080/00380768.2001.10408424

RiceXPro. (2019). http://riin/index.html

Rose, T., Rose, M., Pariasca-Tanaka, J., Heuer, S., and Wissuwa, M. (2011). The frustration with utilization: why have improvements in internal phosphorus utilization efficiency in crops remained so elusive? *Front. Plant Sci.* 2, 73. doi: 10.3389/fpls.2011.00073

Schatz, M. C., Maron, L. G., Stein, J. C., Hernandez, W. A., Gurtowski, J., et al. (2014). New whole genome de novo assemblies of three divergent strains of rice (O. sativa) documents novel gene space of aus and indica. *Genome Biol.* 15, 506–521. doi: 10.1186/s13059-014-0506-z

Scherer, H. W. (2001). Sulphur in crop production. *Eur. J. Agron.* 14, 81–111. doi: 10.1016/S1161-0301(00)00082-4

Sharma, B., Joshi, D., Yadav, P. K., Gupta, A. K., and Bhatt, T. K. (2016). Role of ubiquitin-mediated degradation system in plant biology. *Front. Plant Sci.* 7, 806. doi: 10.3389/fpls.2016.00806

Shu, K., and Yang, W. (2017). E3 ubiquitin ligases: ubiquitous actors in plant development and abiotic stress responses. *Plant Cell Physiol.* 58, 1461–1476. doi: 10.1093/pcp/pcx071

Suzuki, A. (1978). Sulfur nutrition and diagnosis of Sulfur deficiency of rice plants. *JARQ* 12, 7–11. doi: 10.1080/00380768.1979.10433152

Tsujimoto, Y., Yamamoto, Y., Hayashi, K., Zakaria, I., Inusah, Y., Hatta, T., et al. (2013). Topographic distribution of the soil total carbon content and sulfur deficiency for rice cultivation in a floodplain ecosystem of the Northern region of Ghana. *Field Crops Res.* 152, 74–82. doi: 10.1016/j.fcr.2012.11.007

Ueda, Y., Frimpong, F., Qi, Y., Matthus, E., Wu, L., Höller, S., et al. (2015). Genetic dissection of ozone tolerance in rice (Oryza sativa L.) by a genome wide association study. *J. Exp. Bot.* 66, 293–306. doi: 10.1093/jxb/eru419

Wang, F. M., Longkumer, T., Catausan, S. C., Calumpang, C. L. F., Tarun, J. A., Cattin, J., et al. (2018). Genome wide association and gene validation studies for early root vigor to improve direct seeding of rice. *Plant Cell Environ.* 41, 2731–2743. doi: 10.1111/pce.13400

Wissuwa, M., Kondo, K., Fukuda, T., Mori, A., Rose, M. T., Pariasca-Tanaka, J., et al. (2015). Unmasking novel loci for internal phosphorus utilization

efficiency in rice germplasm through Genome-Wide Association Analysis. *PloS One* 10 (4), e0124215. doi: 10.1371/journal.pone.0124215

Xu, K., Xu, X., Fukao, T., Canlas, P., Maghirang-Rodriguez, R., Heuer., S., et al. (2006). Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* 442, 705–708. doi: 10.1038/nature04920

Yamaguchi, J. (1999). Sulfur deficiency of rice plants in the Lower Volta area, Ghana. *Soil Sci. Plant Nutr.* 45, 367–383. doi: 10.1080/00380768.1999.10409351

Yoshida, S., and Chaudhry, M. R. (1979). Sulfur nutrition of rice. *Soil Sci. Plant Nutr.* 25, 121–134. doi: 10.1080/00380768.1979.10433152

Yoshida, S., Forno, D. A., and Cock, J. H. (1971). *Laboratory manual for physiological studies of rice* (Los Banos, Philippines: IRRI), 61.

Zhao, K., Tung, C. W., Eizenga, G. C., Wright, M. H., Ali, M. L., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. *Nat. Commun.* 2, 467. doi: 10.1038/ncomms1467

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

PLOS ONE

RESEARCH ARTICLE

Phenotyping of a rice (*Oryza sativa L.*) association panel identifies loci associated with tolerance to low soil fertility on smallholder farm conditions in Madagascar

Juan Pariasca-Tanaka[1], Mbolatantely Fahazavana Rakotondramanana[2], Sarah Tojo Mangaharisoa[2], Harisoa Nicole Ranaivo[2], Ryokei Tanaka[3], Matthias Wissuwa[1]*

1 Crop, Livestock and Environment Division, Japan International Research Center for Agricultural Sciences (JIRCAS), Tsukuba, Japan, **2** Rice Research Department, The National Center for Applied Research on Rural Development (FOFIFA), Antananarivo, Madagascar, **3** Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

* wissuwa@affrc.go.jp

Citation: Pariasca-Tanaka J, Rakotondramanana MF, Tojo Mangaharisoa S, Ranaivo HN, Tanaka R, Wissuwa M (2022) Phenotyping of a rice (*Oryza sativa L.*) association panel identifies loci associated with tolerance to low soil fertility on smallholder farm conditions in Madagascar. PLoS ONE 17(5): e0262707. https://doi.org/10.1371/journal.pone.0262707

Editor: Muhammad Abdul Rehman Rashid, Government College University Faisalabad, PAKISTAN

Received: August 13, 2021

Accepted: January 2, 2022

Published: May 18, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0262707

Abstract

Rice (*Oryza sativa L.*) is a staple food of Madagascar, where per capita rice consumption is among the highest worldwide. Rice in Madagascar is mainly grown on smallholder farms on soils with low fertility and in the absence of external inputs such as mineral fertilizers. Consequently, rice productivity remains low and the gap between rice production and consumption is widening at the national level. This study evaluates genetic resources imported from the IRRI rice gene bank to identify potential donors and loci associated with low soil fertility tolerance (LFT) that could be utilized in improving rice yield under local cultivation conditions. Accessions were grown on-farm without fertilizer inputs in the central highlands of Madagascar. A Genome-wide association study (GWAS) identified quantitative trait loci (QTL) for total panicle weight per plant, straw weight, total plant biomass, heading date and plant height. We detected loci at locations of known major genes for heading date (hd1) and plant height (sd1), confirming the validity of GWAS procedures. Two QTLs for total panicle weight were detected on chromosomes 5 (*qLFT5*) and 11 (*qLFT11*) and superior panicle weight was conferred by minor alleles. Further phenotyping under P and N deficiency suggested *qLFT11* to be related to preferential resource allocation to root growth under nutrient deficiency. A donor (IRIS 313–11949) carrying both minor advantageous alleles was identified and crossed to a local variety (X265) lacking these alleles to initiate variety development through a combination of marker-assisted selection with selection on-farm in the target environment rather than on-station as typically practiced.

Introduction

Rice (*Oryza sativa L.*) is a staple food for more than half of the world population, supplying about 35 to 60% of dietary calorie intake, micronutrients (Fe and Zn) and vitamins (B). In

Data Availability Statement: All relevant data are within the manuscript and its Supporting information files. The genotype dataset is publicly available at https://snp-seek.irri.org/_download.zul.

Funding: This research was funded by the Science and Technology Research Partnership for Sustainable Development (SATREPS), Japan Science and Technology Agency (JST)/Japan International Cooperation Agency (JICA) (Grant No. JPMJSA1608). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Madagascar, an island located off the southeast coast of Africa, rice has been introduced during the early migration from Asia [1,2] and remains the dominant staple in the Malagasy diet. The annual per-capita consumption of about 136 kg rice is among the highest in the world but unfortunately local production cannot meet the rising demand of an increasing population [3]. The situation is similar across the Sub-Saharan Africa (SSA) region where rice consumption is rapidly outpacing local production [4].

Rice productivity in most of the SSA region is limited by several biotic and abiotic stresses. Among the abiotic stresses, low soil fertility is the one of main concern, with phosphate (P) often being the most limiting nutrient [5]. This is certainly the case in Madagascar where P deficiency is widespread, possibly due to high levels of P-fixing element such as iron (Fe), aluminum (Al) and/or oxyhydroxides [6]. This problem is exacerbated by the continual removal of organic residues and limited application of organic matter and/or fertilizer inputs [7–9]. The nutritional deficiency in soil could be alleviated through fertilizer application; however, the cost of fertilizers is often higher in SSA compared to other regions and therefore access to fertilizers is limited for resource-poor farmers in small scale farming systems. Approximately 10% of the global population lives in Africa, however, only 0.8% (1.29 TM) of the total amount of applied fertilizer is used in Africa [10].

A cost-efficient partial solution to the soil fertility problem in SSA would be the improvement of nutrient acquisition and utilization efficiencies in local varieties [11]. Studies evaluating gene bank accessions concluded that ample variation for P acquisition and underlying root traits existed in the rice gene pool, with traditional varieties typically being superior to modern high-yielding varieties [12]. In comparison, less variation was observed within rice accessions for internal P utilization efficiency, but again traditional varieties tended to be more efficient [13]. The prevalence of traditional rice varieties throughout Madagascar [14] would confirm that modern varieties lack adaptation to low-input conditions and, furthermore, may suggest that plant breeding has not properly addressed the needs of the mostly resource-poor smallholder farmers.

Gene banks are considered a reservoir of untapped allelic variants waiting to be utilized for improving crop adaptation to biotic and abiotic stresses [15,16]. Through next generation sequencing (NGS) an increasing number of gene bank accession has been sequenced, leading to the detection of allelic variants mainly as single nucleotide polymorphisms (SNPs) where the genome sequence of two or more individuals differs by a single base. Genome-wide association studies (GWAS) detect associations between a genetic variant (SNPs throughout the genome) and trait variation (phenotype) for a large number of individuals. GWAS could identify loci, genes and alleles that contribute to specific traits, and therefore could accelerate breeding for targeted traits. For example, GWAS had been successfully applied in rice to dissect the genetic basis of nutrient-related traits such as aluminum tolerance [17], phosphorus utilization efficiency [12], manganese toxicity [18], sulfur deficiency [19], and of traits related to root development [20], and root efficiency [12].

Accessions combined in the GWAS panels of above studies typically comprise mostly traditional varieties from gene bank collections but may also include modern varieties and breeding lines. They may focus on accessions of just one rice sub-species or include representatives of all or most sub-species. Thus, GWAS represents a structured approach to assess the genetic diversity held at national/international gene banks (that would otherwise not be utilized), potentially identifying novel donors and alleles to be utilized in crop breeding to improve traits that lack genetic variability in the pool of currently used breeding lines.

The objective of this study is to follow such an approach with the goal to improve adaptation of rice to low soil fertility. A GWAS panel of 532 sequenced rice accessions was imported from IRRI and evaluated on-farm under low-input conditions in the central highlands of

Madagascar in order to identify novel loci associated with traits of relevance in such conditions. For loci detected we aim to identify suitable donors to initiate a local breeding program to improve grain yield for smallholder farmers.

Materials and methods

1. On-farm trials

1.1 Plant material and on-farm experiment. The 3000 Rice Genome Project (3KRGP) housed by the International Rice Research Institute, IRRI-Philippines (http://snp-seek.irri.org) [21], provides publicly available genotype information and seeds of sequenced accessions. A set of 532 rice accessions from this resource was imported from IRRI to Madagascar. The selected set predominantly included accessions from the *indica* subpopulation (81%), with minor proportions from the *japonica*, *aus* and *aromatic* subspecies (Fig 1A). The central focus in *indica* group was because of the preference for *indica*-type varieties by farmers and consumers in Madagascar. Accessions were then selected from different rice producing countries with similar conditions to the Central Highlands of Madagascar such such as India, Lao PDR, Thailand, Indonesia, Nepal, Sri Lanka, Bangladesh, Philippines, (Fig 1B). Several accessions originated from Madagascar were also included in the set.

Trials were conducted at two field sites in the highlands of Madagascar during the main rice growing season (November to April, 2017–2018): Anjiro-Moramanga (Latitude: -18˚ 56' 58.13" S, Longitude: 48˚ 13' 48.25"), at 950 m above sea level (masl), with average maximum temperature of 29˚C and average minimum temperature of 17˚C, and average precipitation of 220 mm; and Ankazo-Antsirabe (Latitude: -19˚ 51' 57.10" S, Longitude: 47˚ 01' 59.99", at 1150 masl, with average maximum temperature of 27˚C and average minimum temperature of 15˚C, and average precipitation of 210 mm.

All experiments were conducted on smallholder farms characterized by low-input cultivation. Field plots used had no history of mineral fertilizer application in the past. The soils used in these experiments were clay loam, with pH: 5.3–5.8 (1:5, H_2O), total N (g kg^{-1}): 0.4–0.6 g kg-1, Olsen P (mg kg^{-1}): 5.9–7.5, and organic C (g kg^{-1}): 10.7–15.3. Seeds were sown in elevated nursery beds (20 m L x 0.6 m W x 0.1 m H). Each accession was sown in a 40 cm row with 10 cm spacing between rows (S1 Fig). Seeds were covered with fine soil and a layer of non-rice straw mulch to maintain moisture and warmness during seed germination. Water was supplied (depending on availability) either by partially flooding the bed soil or manually using watering cans. Seedlings were raised for 4 weeks in this nursery, followed by manual transplanting of 1 plant per hill into 2-m long single-row plots with 20 cm spacing within and between rows. The experiment was conducted in a completely randomized block design (CBRD) with two replications per site. Agronomic practices such as manual weeding, watering, etc, were performed following the local practices.

1.2 Phenotypic data. Phenotypic parameters evaluated in this experiment included heading date, plant height and culm height, straw and panicle weight. Heading date were taken throughout the vegetative period (start, 50% and 100%). Plant height was defined as the distance from the base of the stem to the tip of the flag leaf, while culm height was measured up to the panicle base node. Harvests of panicles and straw were done continuously as plants reached maturity, using 5 plants per plot. Panicles were individually harvested by cutting at the basal node of the rachis. Straw weight was recorded as fresh weight directly at the field site. This was later adjusted to dry weight based on the moisture content determined for a sub-sample after drying samples in an oven for 3 days at 60˚C. Harvested panicles were brought to a green house and air dried in mesh bags before weights were taken.

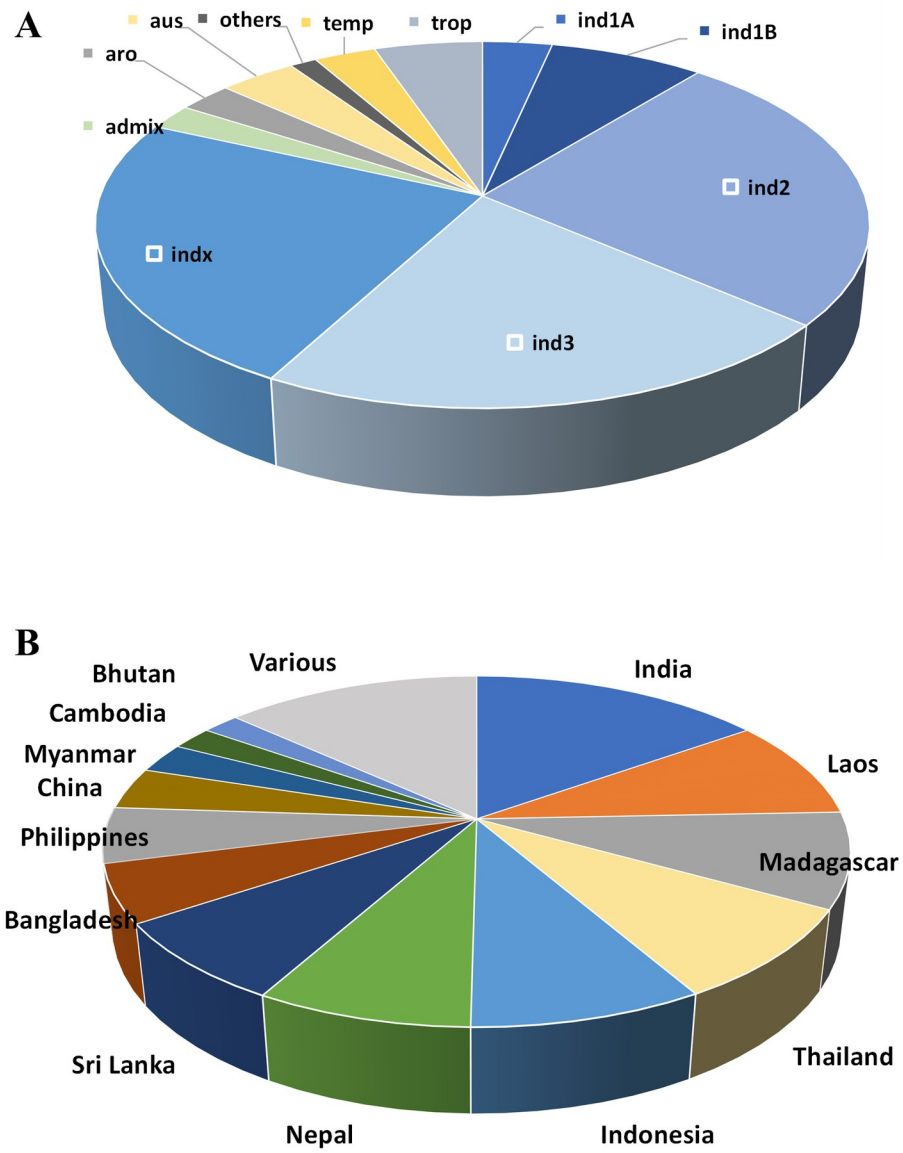


Fig 1. Distribution of subspecies (A) and country of origin (B) of selected rice accessions in the GWAS panel. aro: aromatic; aus: aus; ind: indica; trop: tropical; temp: temperate; admix: admixture.

https://doi.org/10.1371/journal.pone.0262707.g001

The phenotypic data was analyzed using a mixed linear model where the effects of genotype were considered as fixed effect, and those of locations and replicates per location as random effects. The analysis was performed using the R package Linear Mixed-Effects Models using 'Eigen' and S4 (lme4) [22], and an in-house script based on:

$$[\text{phenotypic value}]_{ilrj} = [\text{genotypic value}]_{i} + [\text{location}]_{l} + [\text{replication}]_{lr} + [\text{error}]_{ilrj}$$

This analysis provided the best linear unbiased estimator (BLUE) for genotypes across 2 sites and 2 replicates per site.

Heritability was estimated by using the following model:

$$y_{ijkl} = \mu + genotype_i + location_k + rep(\text{year} \times location)_{jkl} + e_{ijkl}$$

where y_{ijkl} is the phenotypic value of the *i*-th genotype in *j*-th year on the *l*-th replicate in the *k*-th location.

1.3 Association mapping. The 404K core SNP genotype dataset of the rice accessions was obtained from the 3000 Rice Genome Project (3KRGP), (https://snp-seek.irri.org). A matrix genotype file composed of 186,229 (187K) SNPs and 3026 accessions was prepared and reported in a previous study [23]. A subset containing the 532 accessions was filtered from the matrix prior to analysis. Association analysis was then performed using: a) BLUE values obtained for each trait; b) the 187K matrix genotype dataset; and, c) the GWAS function in the Ridge Regression and Other Kernels for Genomic Selection package (rrBLUP v.4.6) [24], using a simple mixed model, where in the phenotype was estimated by setting the accession and residual effects as random, while the replicate effect considered as fixed effect. The effect of population structure was controlled by using a genomic relationship matrix calculated in the A.mat function. The model was run using an in-house R script and the rrBLUP package returned a quantile-quantile plot and a Manhattan plot with a significant threshold set to a 5% FDR (false discovery rate) [20,23]. Loci were considered significantly associated with a trait based on a threshold of–log(10)(p) $>$ 5 for those peaks characterized by at least three consecutives Quantitative Trait Nucleotide (QTN) [19,20].

For the purpose of confirming detected associations, a second analysis was conducted with the software program Trait Analysis by association, Evolution and Linkage 5.0 (TASSEL) [25]. Prior to association analysis, the 404K coreset SNP genotype dataset was filtered as follows: heterozygotes and indels were set as missing values, and SNP having more than 5% missing data or minor allele frequency (MAF) below 0.03 were excluded [19]. The association mapping was then analyzed using the mixed linear model (MLM) procedure, with three principal components (PCA) and a kinship matrix. Adjusted p-values were calculated using the False Discovery Rate (FDR = 0.05) correction method in R.

The phenotypic effect of minor alleles at each locus was determined by calculating the average phenotypic values of all accessions carrying either allele, and a box-plot graph was generated for each locus using an in-house R script.

Linkage disequilibrium (LD) analysis to define LD blocks (non-random association of alleles at a defined region) surrounding the significant SNPs was performed by Haploview 4.2 [26]. To check whether identified regions corresponded to previously identified QTL, peak QTN positions were searched against the public QTL databases QTARO (http://qtaro.abr.affrc.go.jp) and GRAMENE (https://archive.gramene.org/qtl).

1.4 Selection of putative candidate genes. Gene models were obtained from the Rice Annotation Project Database (RAP-DB, https://rapdb.dna.affrc.go.jp/) for each significant peak and their surrounding LD block. Genes annotated as '(retro)transposon', 'hypothetical' or 'unknown' were excluded from further analysis. Putative candidate genes were then selected based on annotated function and gene ontology (http://www.geneontology.org), and expression pattern obtained from the Rice XPro database (http://ricexpro.dna.affrc.go.jp).

Furthermore, SNP variant effects were investigated using the Variant Effect Predictor (VEP, Ensembl, https://asia.ensembl.org/Tools/VEP), which predicts the potential effects of the SNP variant in terms of changes in protein sequences.

2. Validation of result

2.1 Using a different set of 3K accessions. To confirm the effects of positive alleles identified in the on-farm trials (experiment 1), a different set of 3K accessions was grown in the following year and TPW was measured. This set consisted of 52 newly imported rice accessions and 23 accessions repeated from year 1. Field experiments were carried out at the same sites as

in year 1 but in different small-holder farmer fields (under low fertility soil). Fields were not fertilized and had no history of mineral fertilizer application in the past. Experimental procedures were as reported for experiment 1.

2.2 Using water culture (low P and/or low N). A second confirmatory experiment was conducted in the greenhouse in Japan, evaluating accessions at the vegetative growth stage in hydroponic culture. Dehulled seeds from selected accessions had been imported into Japan where a seed multiplication step was necessary. Pregerminated seeds were sown onto a mesh floating over a solution containing 10% Yoshida solution without P. The full-strength Yoshida solution (1X) is composed of: N, 2.86 mM (as NH4NO3); P, 0.05 mM; K, 1mM; Ca, 1mM; Mg, 1mM; Mn, 9 μM; Mo, 0.5 μM; B, 18.5 μM; Cu, 0.16 μM; Fe, 36 μM; Zn, 0.15 μM [27].

Ten days after germination seedlings were transferred to 45-L hydroponic containers with 28 seedlings fixed to holes in the container lid using sponge strips. Four treatments were imposed in an otherwise modified Yoshida nutrient solution as described above: low P (LP, 5uM), low N (LN, 0.28 mM), a combination of low N and low P (LNP) and a control treatment (2.86 mM N, 50uM P). The experiment was conducted in a temperature-controlled (30˚C during daytime, and 25˚C nighttime) glass house under natural light at JIRCAS-Tsukuba (36˚12'0"N, 140˚6'0"E). The experiment was conducted in a randomized complete block design (RCDB) with four replications. Rice accessions were harvested 35 days after germination, root length was evaluated together with root and shoot dry matter, and root/shoot subsamples of four independent replications were flash-frozen in liquid nitrogen and stored at -70˚C until RNA extraction using the RNeasy Plant Mini Kit (Qiagen), following the manufacturer instruction manual.

Expression of candidate genes. Total RNA (400 ng) was then reverse transcribed (RT) using the PrimeScript RT Enzyme Mix I (Takara, Japan). Quantitative PCR (qPCR) was performed using 2 ng RT template and SYBR Premix ExTaq (Perfect Real Time, Takara, Japan), using the CFX96 Touch Real-Time PCR system (BioRad, USA). Primer efficiency was determined by serial dilutions of RT product. Elongation factor (ELF-1), Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and Ubiquitin (Ubi) was used as internal controls. Relative expression levels between treatment and shoot or root of control samples were calculated using the standard-curve method and expressed as fold changes. The normalized data was then analyzed by ANOVA. The list of primers used in this study is shown in S1 Table.

3. Development of a cross population

In order to utilize the main peaks associated with TPW on chromosome 11 in marker assisted selection, we designed a Kompetitive Allele Specific PCR (KASP) marker (qTLF11-1). Using this KASP marker we determined that the popular local Malagasy variety X265, also known as "Mailaka", carries the (major) unfavorable allele, and would therefore be a potential recipient benefitting from the introgression of the positive minor allele from donor accession GP1103.

X265 and GP1103 parent plants were grown under paddy condition and during flowering time, panicles of previously designated female plants were emasculated using heat treatment (immersion in water bath at 42˚C for 7 min) and cross-pollinated using pollen from the male parent. Successfully crossed F1 plants were identified with KASP markers using an in-house protocol [28] following the manufacturer instruction manual (LGC Genomics). In brief, KASP amplification was performed using allele-specific primers with FAM and HEX fluorophores, a common primer and master mix. The fluorescence signal was then recorded at 520 nm (FAM) and 556 nm (HEX) for 2 min at 25˚C, at the end of the thermal cycles.

Using a modified rapid generation advance (RGA) protocol, crossed individuals were advanced through the F2 and F3 generation and F4 seeds were sent to Madagascar for field evaluations in on-farm trials.

Statistical analysis. The effects of treatment, allele and their interaction on different traits were estimated using a one or two-way ANOVA, and mean comparisons were performed using Tukey's honestly significant difference (HSD) post hoc test (Statistix 9.0 Software). Correlation values were generated by "Hmisc" and visualized in scatter plots by "PerformanceAnalytics" R packages [29].

Results

On-farm field trials and GWAS analysis

A set of rice accessions of diverse origin but primarily belonging to the *indica* sub-species (Fig 1) was imported and evaluated in two rice-growing areas in the central highlands of Madagascar. Plant performance was evaluated by straw dry weight (STW), total panicle weight (TPW), which is the average total weight of all panicles per one plant (rather than the weight of an individual panicle) and total dry weight (TDW) (Fig 2). Biomass weights are given per plant and grain yield is estimated by TPW. STW ranged from 7.1 to 97.4 g plant^{-1} with a mean of 29.5 g plant^{-1}, TPW from 3.0 to 41.5 g plant^{-1} with a mean of 16.3 g plant^{-1}, and TDW from 10.1 to 121.1 g plant^{-1} with a mean of 45.2 g plant^{-1} (Fig 2, S2 Table). Traits STW and TPW showed greatest variation with a coefficient of variation (CV) of 38.8 and 33.7%, respectively, which indicates great accession variability in their adaptation to the new environment. The distribution for TPW was near-normal with the exception of three outliers with high values.

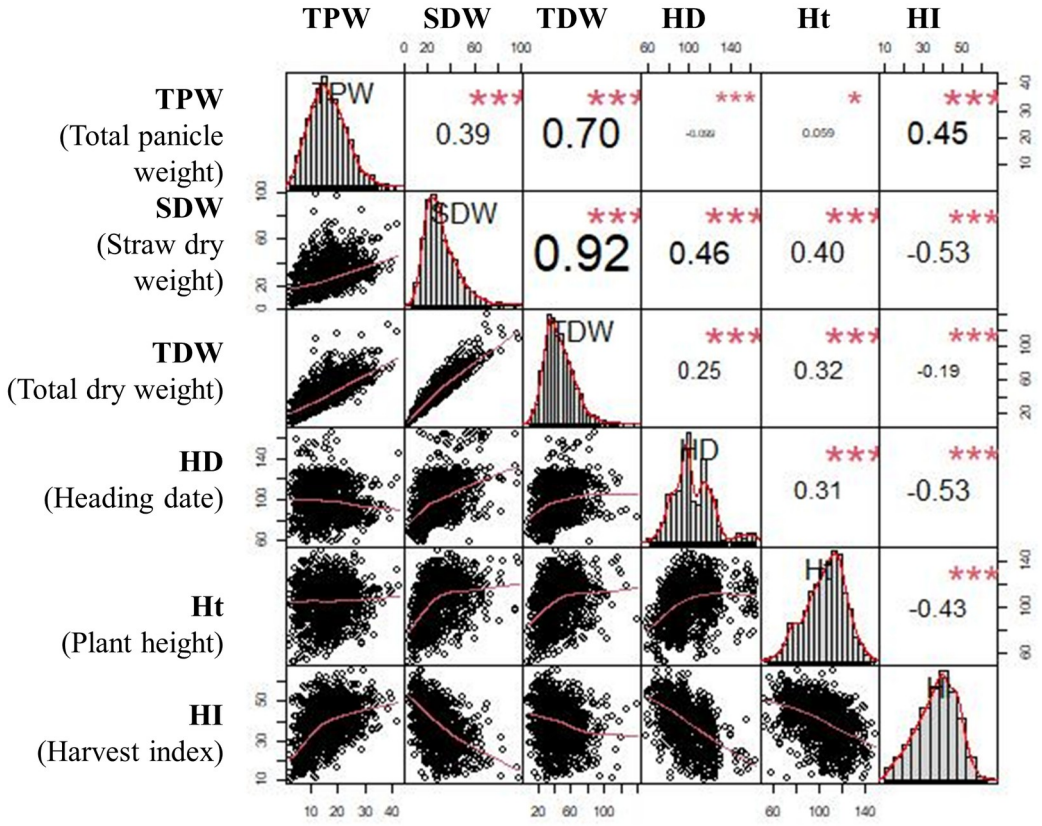


Fig 2. Scatterplots showing the relationship among all evaluated agronomic traits. The distribution of each variable is shown on the histogram, while the bivariate scatter plots with a fitted line, and correlation values and significance level are shown on the left and right side, respectively. Significance level: p-values ($p<0.001$: ***, $p<0.01$: **, $p<0.05$: *, $p>0.05$: "").

https://doi.org/10.1371/journal.pone.0262707.g002

The distribution for STW was slightly skewed towards smaller values with five accessions showing high STW.

Correlation coefficients between traits measured ranged from as high as r = 0.92 between TDW and STW, to as low as 0.06 between TPW and plant height (Fig 2). Straw and panicle biomass had a moderately positive correlation and variation in STW contributed more to TDW compared to TPW (r = 0.70). As expected, STW was positively correlated with plant height (r = 0.40) but plant height did not affect TPW. Similarly, late heading was associated with higher STW (r = 0.46) but not with TPW. Heading date (HD) was negatively correlated with harvest index (HI), presumably because late heading accessions produced more straw biomass (Fig 2).

Higher heritability values were found for SWT and HD with 0.54 and 0.40, respectively, while TPW showed a value of 0.30 (S2 Table).

The association analysis using the Mixed Linear Model (MLM) in rrBLUP identified several quantitative trait loci (QTL) associated with tolerance of low-fertility soils (*qLFT*) (Fig 3). Two QTLs associated with TPW were detected on chromosomes 5 (*qLFT-5*) and 11 (*qLFT-11*). These loci were represented by two significant Quantitative Trait Nucleotide (QTNs) at 14.496 and 14.827 Mbp on chromosome 5 and by 3 QTNs between 25.827–25.849 Mbp on chromosome 11 (S3 Table). For both loci the minor allele frequency (MAF) was below 10% and the minor allele had a positive effect, increasing TPW from 15.9 to 22.7 g plant^{-1} (+42.8%, chromosome 5) and from 15.8 to 22.0 g plant^{-1} (+39.0%, chromosome 11) (Tables 1 and 2).

Four QTLs were detected for STW on chromosomes 1, 3, 4 and 11 (Fig 3). The strongest effect was seen on chromosome 1 where four consecutive QTNs between 10.993 and 11.591 Mbp exceeded the significance threshold of 5 (S3 Table). This locus had a MAF of 5% and the minor allele was estimated to increase STW from 28.5 to 46.5 g plant^{-1} (+63.2%) (Tables 1 and 2). The second most significant QTL for STW was delineated by 3 QTN between 31,542 and 31,543 Mbp on chromosome 4 and the minor allele (MAF = 12%) increased STW by 42.7% (Table 1). The remaining two QTL on chromosomes 3 and 11 had lower significance but estimated phenotypic effects were large with 61% and 54% increase in STW due to the minor allele, respectively (Table 1). Two QTLs associated with TDW were detected on chromosome 5 and 11 but at different locations from QTL for TPW and STW and with lower significance (Fig 3). Unlike for above loci, the MAF was not low but above 30%. For the locus on chromosome 11 the minor allele increased TDW by 20% but a negative effect was associated with the minor allele on chromosome 5 (Table 1).

Additionally, QTLs for heading date (HD) were found in chromosome 1, 2, 3, 4, 6 and 7 (Table 1 and S3 Table) with the most significant association detected on chromosome 6 at 21,342 Mbp where the minor allele delayed heading by 30%. For plant height a highly significant locus was detected between 37,876–39,548 Mbp on chromosome 1. This interval contains the known semi-dwarf gene *sd1* and the minor allele reduced plant height by 30%.

The result was then validated using the software TASSEL with the 3K-400K SNP dataset. The resulting Manhattan and QQ plots for each evaluated trait are shown in Table 1, and S2 Fig.

The MAF values calculated for the studied panel were corroborated for the entire 3K dataset. For the two QTL associated with TPW (5@14,496,649 and 11@25827214) the frequency of minor alleles was below 10% in the entire 3K set (S4 Table) and therefore very similar to the subset phenotyped in Madagascar (Table 1). More than 90% of the accessions with the minor allele belong to the *indica* sub-species (ind1, ind2, ind3, and indx). For loci associated with STW (1@ 11039294, 3@ 20761847, and 4@31542322) a very similar situation was observed, the minor allele predominantly being detected in the indica group.

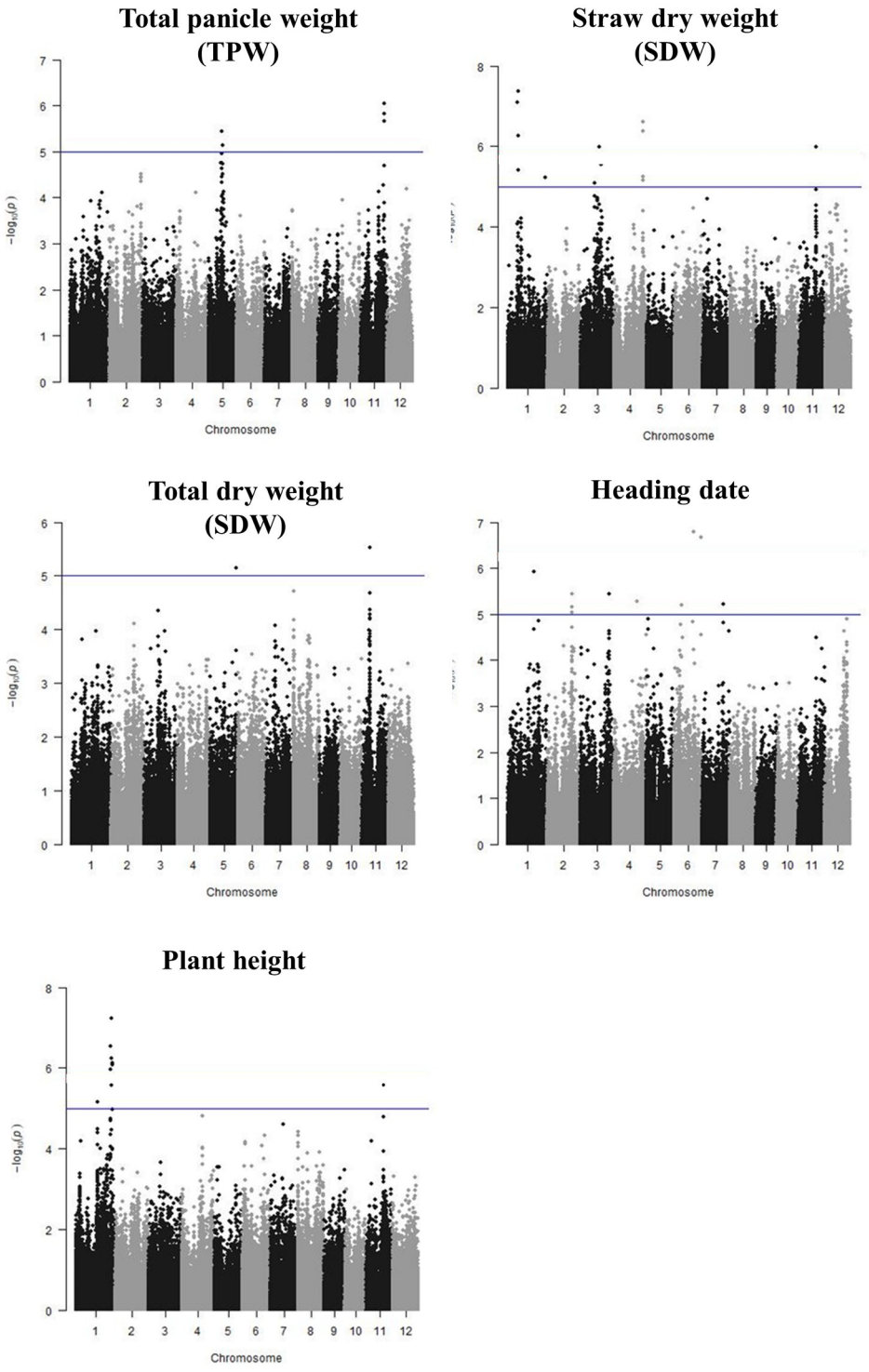


Fig 3. Manhattan plots derived from GWAS analysis for all evaluated traits. Y-axis shows the negative logarithm of the association ((-log10 (P-value)) for each SNP, while X-axis displays the SNP location along the 12 chromosomes. Vertical line indicates a -log10 (P-value) threshold of 5.

https://doi.org/10.1371/journal.pone.0262707.g003

Table 1. Summary of quantitative trait loci (QTL) associated with low fertility soil for several agronomic traits using a mixed linear model (MLM).

Trait	Loci name	Chr	SNP denomination	SNP position	P value		minor allele	
				(bp)	rrBLUP	[3]TASSEL	[1]MAF	[2]effect
Total panicle weight (TPW)								
	qLFT-5	5	5@14496649	14,496,649	3.5E-06	3.48E-05	0.04	41
	qLFT-11	11	11@25827214	25,827,214	2.1E-06	4.23E-05	0.08	39
Straw dry weight (SDW)								
	qLFT-1	1	1@11039294	11,039,294	4.2E-08	4.32E-05	0.05	60
	qLFT-3	3	3@20761847	20,761,847	1.0E-06	5.03E-05	0.06	60
	qLFT-4	4	4@31542322	31,542,322	4.3E-07	1.44E-05	0.11	40
	qLFT-11s	11	11@19334313	19,334,313	1.0E-06	3.95E-05	0.04	54
Total dry weight (TDW)								
	LFT-11t	11	11@8850567	8,850,567	3.0E-06	8.2E-05	0.34	20
Heading date (HD)								
	qHD-1	1	1@28488608	28,488,608	1.2E-06	2.2E-05	0.10	30
	qHD-2	2	2@27125471	27,125,471	3.5E-06	6.7E-05	0.08	23
	qHD-3	3	3@31256576	31,256,576	3.6E-06	5.1E-05	0.30	21
	qHD-4	4	4@25834920	25,834,920	5.4E-06	4.9E-05	0.05	30
	qHD-6	6	6@21342504	21,342,504	1.6E-07	3.4E-05	0.07	30
	qHD-7	7	7@22523621	22,523,621	6.2E-06	7.1E-05	0.32	21
Plant height (Ht)								
	qHt-1	1	1@38730952	38,730,952	6.0E-08	1.6E-05	0.11	-30

[1] MAF: minor allele frequency.

[2] allele effect: phenotypic value (((minor allele-major allele)/major allele)*100).

[3] values corrected by False discovery rate (FDR).

https://doi.org/10.1371/journal.pone.0262707.t001

For main QTL associated with TPW and STW, minor and major alleles were investigated in detail in relation to effects on other traits (Table 2). Accessions belonging to the minor allele group for TPW QTL *qLFT-5* and *qLFT-11* had similar HD and STW compared to the group with the major allele. Accessions carrying minor alleles at both loci (n = 9) showed a further improvement in TPW, being 82.1% superior to accessions lacking both loci. For STW QTL, the minor and major allele groups did not differ significantly for TPW, but the minor allele group showed significantly later heading (Table 2). We also calculated effects of having two of these loci simultaneously and while this led to further increases in STW, it caused additional delays in heading and several accessions were not yet mature at the end of the experimental period (data not shown).

Selection of putative candidate genes. To determine to what distance linkage would extend from the peak QTN, the relatedness of all SNP in the larger region surrounding the peak QTN were investigated (S3 Fig). Very distinct linkage blocks could not be identified but based on the decay in LOD between markers we identified likely regions to be considered for candidate gene identification for *qLFT-5* from 14.343 to 14.585 Mbp, and from 25.734 to 25.948 Mbp for *qLFT-11*. Potential candidate genes for TPW were selected based on their expression pattern in different tissues and environmental conditions (RiceXpro, S4 Fig) and their functional annotation is listed in S5 Table (excluding unknown genes and hypothetical proteins). Estimating functional consequences of SNPs in candidate genes using the Variant Effect Predictor (Ensembl) showed that most SNPs for TPW were located in the intergenic,

Table 2. Distribution and interaction of minor and major alleles across the main identified QTLs.

		minor allele		major allele		effect	
		mean	SD	mean	SD	%	
Total panicle weight (TPW)							
qLFT-5	TPW	22.7	8.3	15.9	5.1	42.8	***
	HD	95.3	15.3	100.1	16.6	-4.8	ns
	STW	31.4	10.7	28.4	10.4	10.6	ns
	n	27		461			
qLFT-11	TPW	22.0	7.7	15.8	5.2	39.2	***
	HD	97.4	18.6	100.0	16.5	-2.6	ns
	STW	29.3	10.8	28.5	10.4	2.8	ns
	n	32		447			
qLFT- 5 x 11	TPW	28.4	9.1	15.6	5.1	82.1	***
	HD	94.1	22.2	100.3	16.7	-6.2	ns
	STW	30.0	13.3	28.3	10.5	6.0	ns
	n	9		430			
Straw dry weight (STW)							
qLFT-1	STW	46.5	17.9	28.5	10.2	63.2	***
	HD	125.7	22.0	100.3	17.5	25.3	***
	TPW	14.8	4.4	16.3	5.58	-9.2	ns
	n	25		486			
qLFT-3	STW	45.4	16.6	28.2	10.1	61.0	***
	HD	124.5	22.4	99.9	17.3	24.6	***
	TPW	12.9	4.3	16.5	5.6	-21.8	ns
	n	30		466			
qLFT-4	STW	40.1	15.5	28.1	10	42.7	***
	HD	115.4	20.8	99.8	17.7	15.6	***
	TPW	14.6	5.6	16.5	5.5	-11.5	ns
	n	59		448			

Significance levels (***, **, *, ns: $p<0.001$, 0.01, 0.05, non-significant, respectively).

SD: standard deviation.

TPW (total panicle weight), STW (straw total weight), HD (heading date).

https://doi.org/10.1371/journal.pone.0262707.t002

and up/down stream region (more than 80%), while few existed in the intron or 5/3'UTR regions (S5 Fig). Two genes had either gained or lost a stop codon but none of these were considered functionally relevant.

Based on above criteria the following potential candidate genes for panicle weight at *qLFT-5* and *qLFT-11* were identified: 1-aminocyclopropane-1-carboxylic acid synthase (Os05g0319200), protein kinase (Os05g0319700), WRKY transcription factor (Os05g0322900), cytochrome P450 (Os05g0320700), and Zn finger protein (Os05g0316000), and NB-ARC domain (Os11g0645886), oxidoreductase (Os11g0645200), E3 ubiquitin-protein ligase EL5 (Os11g0649801), sugar transporter (Os11g0643800) (Table 3 and S5 Table). Candidates for STW would be galactose oxidase (Os01g0300900), and Chitinase precursor (Os01g0303100), while SAM dependent carboxyl methyltransferase family protein (Os11g0260100), polygalacturonase (Os05g0578600), and UDP-glucosyltransferase (Os03g0757000, Os06g0271000) were considered candidate genes for total weight.

Table 3. List of potential candidate genes in QTLs associated to total panicle weight (TPW), shoot dry weight (SDW) and total dry weight (TDW).

RAPdb	MSU (LOC)	Chr	PosMb	Annotation
Total panicle weight, TPW (*qLFS-5*, *qLSF-11*)				
Os05g0316000	Os05g25180	5	14.588	Zinc finger RING/FYVE/PHD-type domain
Os05g0319200	Os05g25490	5	14.825	1-aminocyclopropane-1-carboxylic acid synthase
Os05g0319700	Os05g25540	5	14.844	Protein kinase-like protein
Os05g0320700	Os05g25640	5	14.900	Similar to Cytochrome P450
Os11g0644800	Os11g42510	11	25.597	Tyrosine/nicotianamine aminotransferases family
Os11g0645200	Os11g42540	11	25.615	Oxidoreductase
Os11g0645886	Os11g42590	11	25.635	NB-ARC domain containing protein
Os11g0648400	Os11g42850	11	25.806	Protein of unknown function DUF3615 domain
Os11g0649801	None	11	25.911	Similar to E3 ubiquitin-protein ligase EL5
Straw dry weight (SDW)				
Os01g0300900	Os01g19480	1	11.059	Galactose oxidase
Os01g0301000	Os01g19490	1	11.065	Pentatricopeptide repeat domain
Os01g0301900	Os01g19610	1	11.110	Protein of unknown function DUF247
Os01g0302500	Os01g19694	1	11.167	Knotted1-type homeobox protein OSH6
Os01g0303100	Os01g19750	1	11.209	Chitinase precurso
Os01g0303600	Os01g19800	1	11.232	RING/FYVE/PHD-type domain
Os04g0618700	Os04g52780	4	31.421	Protein kinase
Os04g0619400	Os04g52840	4	31.463	Protein kinase
Os04g0620400	Os04g52940	4	31.532	SIT4 phosphatase-associated protein
Total dry weight (TDW)				
Os05g0578600	Os05g50260	5	28.802	Similar to Polygalacturonase PG2
Os05g0578900	Os05g50270	5	28.818	GAGA-type zinc finger transcription factor
Os11g0260100	Os11g15340	11	8.677	SAM dependent carboxyl methyltransferase
Os11g0260200	Os11g15370	11	8.702	Sulfotransferase domain containing protein

https://doi.org/10.1371/journal.pone.0262707.t003

Validation of the TPW QTL

A set of 75 accessions including 52 not previously phenotyped accessions was tested under similar condition as for Experiment 1. Of these 52 new accessions, 21 harbored the positive minor allele at 11_25827214 (*qLFT-11*). This group had significantly higher total panicle weight compared to the group with the major but disadvantageous allele (Table 4). Although both groups showed similar mean values for plant height and number of panicles.

A subset of rice accessions with contrasting alleles at *qLFT-11* was grown in hydroponics under low N (LN), low P (LP) or combined low N and P (LNP) conditions to simulate the low fertility of soils in Madagascar. All nutrient deficient treatments increased root biomass and

Table 4. Total panicle weight from rice accessions selected from within and outside the GWAS panel.

	Number of accessions (n)	Plant height (Ht)		Number of Panicles		Total panicle weight (TPW)	
Allele (A)		ns		ns		*	
Advantageous	23	84.61	a	9.69	a	35.65	a
Disadvantageous	52	82.22	a	9.61	a	32.49	b

Plants were grown on-farm field, in the next cropping season, under low input condition, Madagascar (Experiment 2–1). The accessions were divided into two groups: harboring the advantageous or disadvantageous alleles for total panicle weight (TPW). Values are the mean of four independent replication. Statistical significance was determined by one-way ANOVA and Tukey's tests. Significance levels (***, **, *, ns: $p<0.001$, 0.01, 0.05, non-significant, respectively).

https://doi.org/10.1371/journal.pone.0262707.t004

this effect was more pronounced in the group harboring the positive minor allele at 11@25827214 (Fig 4, S6 Table). Both groups did not differ significantly for root biomass in the nutrient-replete control treatment but root biomass more than doubled for the minor allele group whereas it increased between 53–59% in the group with the major allele. Shoot biomass, on the other hand, decreased in all nutrient deficient treatments relative to the control (Fig 4). Differences between allelic groups were small and not specific to nutrient deficiency. However, significant differences between groups were seen in the root to shoot ratio, which increased

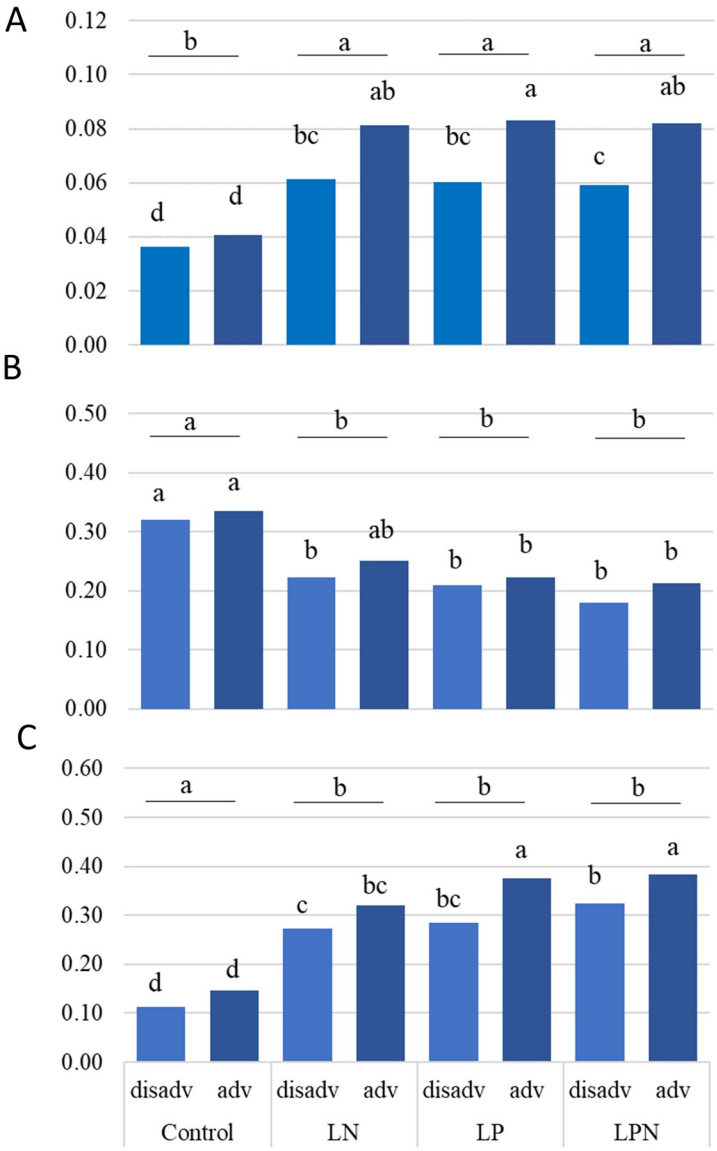


Fig 4. Root (A), shoot (B) and root to shoot ratio (C) from rice accessions harboring the advantageous or disadvantageous alleles for total panicle weight (TPW). Plants were grown under hydroponic condition with low Nitrogen and/or low Phosphorus (Experiment 2–2). Values are the mean of four independent biological replicates (n = 4). Statistical significance was determined using two-way ANOVA and Tukey's tests. Different letters represent distinct means within groups at $p < 0.05$ (***, **, *, and ns refers to $p<0.001$, 0.01, 0.05, non-significant, respectively). adv: advantageous allele (G-38, G-355, G-1103), disadv: disadvantageous allele (X265, IR64, G-61, G-97).

https://doi.org/10.1371/journal.pone.0262707.g004

significantly in all nutrient deficient treatments and for which allelic differences were significant in the two low-P treatments but not in the LN treatment.

Gene expression in shoot and root tissue of the allelic groups under LP, LN and LNP compared to control conditions (base = 1) are represented in a heatmap graph (Fig 5). A gene known to respond strongly to P deficiency (OsSPX) was included to corroborate and gauge the typical P response. This gene showed the highest transcript abundance in low P tissue with no difference between the allele group (Fig 5).

Candidate genes for *qLFT-11* exhibited differential expression across treatments. The genes encoding for oxidoreductase (Os11g0645200), and plant resistance (Os11g0645400) were more responsive to N than to P deficiency and expression tended to be higher in the advantageous allele group. The sugar transporter (Os11g0643800) was only differentially regulated in roots where highest expression was detected in response to P deficiency in the advantageous group (Fig 5). For Os11g0645800 (NB-ARC domain) patterns between groups were opposite in shoot and root and again, highest expression was detected in response to P deficiency in roots of the advantageous group. Similar strong responses to P deficiency in the advantageous allele group was seen in shoot tissue for two candidates at *qLFT-5*, WRKY (Os05g0322900) and Cytochrome P450 (Os05g0320700).

Development of a cross population

Rice accessions harboring both advantageous alleles for panicle weight at *qLFT-5 and qLFT-11* were identified and accession GP-1103 (IRIS 313–11949) with high average total panicle weight (30.0 g plant^{-1}), medium heading (92 days under P deficiency) and plant height (98 cm) was selected as candidate donor. Recommended Malagasy variety X265 did not harbor either advantageous allele for panicle weight (S7 Table) and was therefore selected as the recipient parent. A set of 350 F4 lines was phenotyped on-farm under low-input conditions and wide segregation for total panicle weight per plant was observed, ranging from 13 g plant^{-1} to 50 g plant^{-1}, which compares to 22.8 g plant^{-1} for local parent X265 (Fig 6).

Discussion

In lowland rice fields of Madagascar, the deficiency for P is typically the most serious yield-limiting factor [30], however, deficiencies for N and to a lesser extent for S and other nutrients are also common [31]. We have conducted all our field experiments on small-holder farms in fields that never received mineral fertilizer, and to which manure had not been applied at least in the two seasons preceding our experiments. Fields were therefore characterized by low fertility (Ferralsols containing very low available soil P) [7], and the average panicle weight of 16.3 g per hill, resulting in an estimated grain yield of about 3.6 t ha^{-1}, was just slightly above the national average of 2.9 t ha^{-1} [3]. Considering that yield estimated from single row measurements tend to overestimate achievable grain yields on a field-scale, we may conclude that our field experiments represented typical low-input field conditions for the country and that genotypic differences in yield may reflect adaptations to low soil fertility. We therefore chose to designate identified QTL as *qLFT* (Low Fertility Tolerance) to distinguish the present study from field experiments conducted specifically under P deficiency (with other nutrients supplied through fertilization) and, especially to distinguish from the many QTL identified in studies conducted under controlled conditions in low-P nutrient solution.

Loci associated with tolerance to low soil fertility

The GWAS analysis identified a highly significant locus for plant height at 38.7 Mbp on chromosome 1 (*qHt-1* in Table 1), which is only 0.3 Mbp from the position of the semidwarf gene

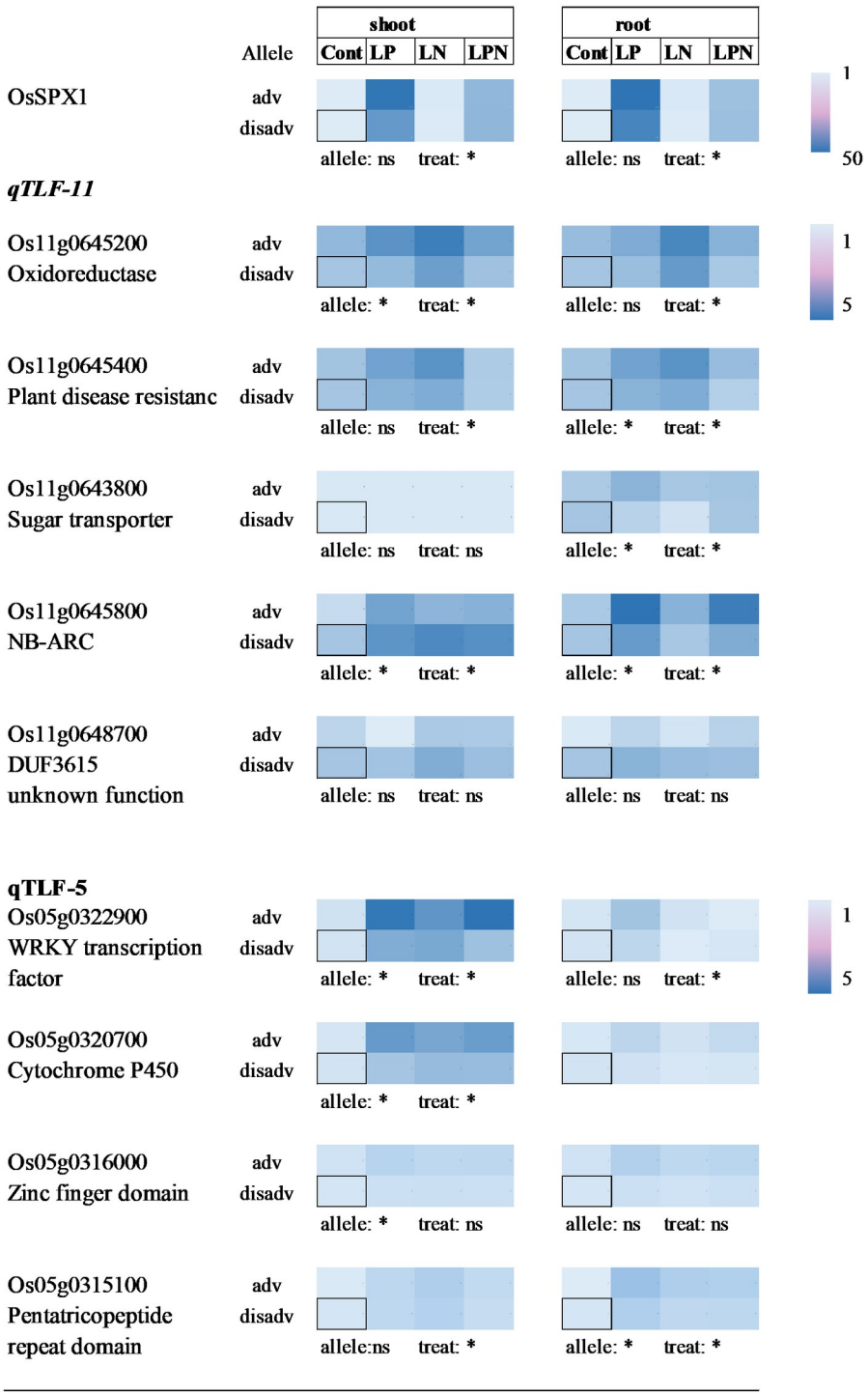


Fig 5. Relative expression pattern of potential candidate genes located within the total panicle weight (TPW) QTLs. Heatmap displays the differentially expressed candidate genes in root and shoot tissue from genotypes harboring advantageous/disadvantageous alleles, and under P, N, or both deficiency condition. Statistical significance was determined by two-way ANOVA and Tukey's tests. Asterisks indicates significance levels (***, **, *, ns: $p<0.001$, 0.01, 0.05, non-significant, respectively). Values were normalized to control treatment in accession with disadvantageous allele, in each tissue (square). adv: advantageous allele (G-38, G-355, G-1103), disadv: disadvantageous allele (X265, IR64, G-61, G-97).

https://doi.org/10.1371/journal.pone.0262707.g005

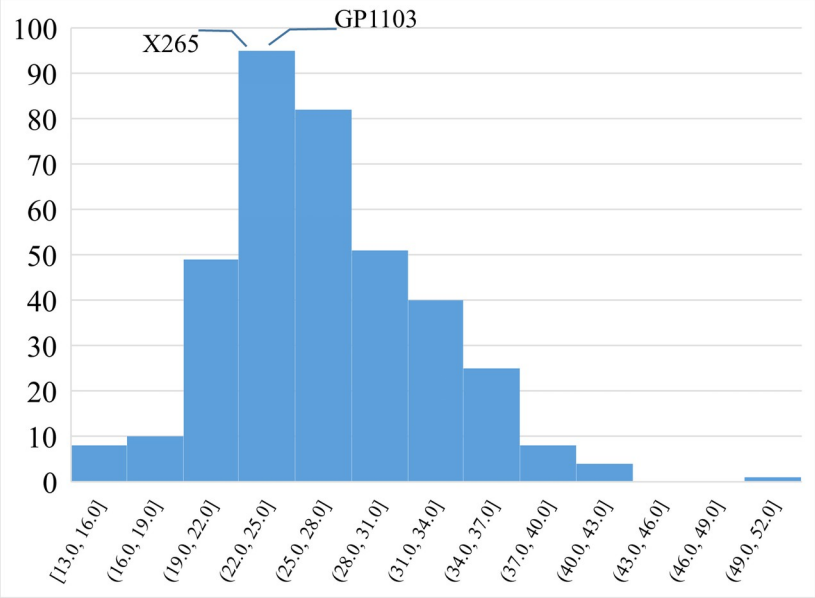


Fig 6. Histogram showing the frequency distribution for total panicle weight among 340 lines of the X265 x GP1103, F5 population. Plants were evaluated under low input condition in the central highland of Madagascar. The mean of GP1103 and X265 are shown in callouts.

https://doi.org/10.1371/journal.pone.0262707.g006

sd1 (Os01g0883800). As expected for a panel consisting of gene bank accession mostly exhibiting the plant habitus of traditional varieties, the minor allele (MAF = 0.11) reduced plant height by 30%, which would be consistent with most traditional varieties carrying the functional *SD1* allele. A second known locus identified in our panel through GWAS (*qHD-6* in Table 1) was within 0.3 Mbp of the known heading date gene *Hd1* (Os06g0275000) on chromosome 6. Having detected two known major genes corroborated the general suitability of the collected data for the purpose of identifying genetic determinants associated with traits of interest.

We detected two novel QTL associated with total panicle weight (TPW) and three associated with straw weight (STW) and in all cases a rare minor allele with MAF between 4–11% increased TPW and STW. These low MAF were not caused by some bias in the selection of the accessions to be phenotyped in Madagascar but could be verified among the entire 3K SNP-seek dataset (S4 Table), which showed that *qLFT-5* was even less frequent in the entire set (MAF = 3.6%) than in the phenotyped set (MAF = 5.5%). For *qLFT-11* both frequencies were around 7%. These results confirm the power of GWAS to identify rare but positive alleles in gene banks and of donors carrying such rare alleles. For the phenotyped accessions we investigated whether the origin of both rare alleles was linked to some country or region. For *qLFT*-5 all accessions belonged to the *indica* sub-species and overrepresented countries were India, Lao, and Indonesia (data not shown). For *qLFT*-11 accessions also belonged to the *indica* sub-species and overrepresented countries were China, the Philippines and Lao. Among accessions were few modern varieties such as PSBRC18, BR11 and four breeding lines from IRRI (data not shown).

Donors and their use in rice improvement

That *qLFT-5* and *qLFT-11* headed slightly earlier than the average whereas loci associated with STW caused a delay in heading and a slight decrease in TPW indicated that the utility of STW

loci identified here for rice breeding in our target environment (highlands of Madagascar) is very limited. Late heading exposes a crop to cold spells at the end of the cropping season and can severely reduce yields. Furthermore, late heading may increase vulnerability to climate change as rainfall patterns become less predictable. Thus, we only consider the TPW loci identified here as being of interest for rice improvement.

Among the nine potential donors carrying both positive alleles, accession Liu He Xi He from China (IRIS 313–11949; ind1A) combined high TPW with medium-early heading and the medium plant height preferred in the highlands of Madagascar. A cross population between this donor and the local cultivar X265 (lacking both positive alleles) is now being evaluated under P deficiency at several sites in Madagascar in order to select breeding lines combining high grain yield with local adaptation. Since TPW was affected by environment condition ($H^2 = 0.3$), breeding lines will be tested in multi-location trials characterized by multiple nutrient deficiencies. Selected elite breeding lines could thus contribute to achieve sustainable rice production and improved food security in Madagascar.

Putative candidate genes

The objective of this study was to evaluate a diverse panel of gene bank accessions to identify potential donors and markers to be used in rice breeding and a detailed analysis of candidate genes is beyond the scope of this study. However, patterns observed in our gene expression analysis provided some preliminary evidence suggestive of allelic differences, especially for the WRKY transcription factor and the member of the cytochrome P450 gene family on chromosome 5. For *qLFT-11* on chromosome 11 a higher proportion of differential regulation was seen in root tissue. Os11g0645800 containing the AB-ARC domain more typically associated with disease resistance [32] was strongly up-regulated by P deficiency in the advantageous allele group. Disease is unlikely to have played a role in our nutrient solution experiment, however, disease resistance would be achieved through triggered cell death [32], and one may speculate whether this could play a role in aerenchyma formation as more rapid aerenchyma formation in P efficient rice genotypes were previously reported [33].

The higher expression of sugar transporter Os11g0643800 under P deficiency in root but not shoot tissue of the positive allele group may corroborate results from the nutrient solution experiment that showed an increase root to shoot ratio of this group under P deficiency. To what extent this may be related to a shift in resource allocation to roots is a potential topic for further investigation.

Conclusions

Alleles absent from the modern rice breeding gene pool but present in traditional varieties housed in crop gene banks have the potential to improve crop yields in less favorable environments, thereby closing the yield gap that is so persistent in Africa. However, gene banks remain a largely untapped resource and here we have attempted to address this issue by testing 500 gene bank resources for which sequence information is available. With phenotyping done in the target environment, smallholder farms under typically practice low-input conditions, our results provide evidence of power of such a GWAS approach to identify rare positive alleles in gene banks. With agronomically acceptable donors for such rare alleles identified, they may (re)-enter the breeding gene pool and contribute to variety development that would specifically benefit resource-poor farmers that have not sufficiently profited from current mainstream rice breeding.

Supporting information

S1 Fig. Photo showing land preparation, indirect sowing, plant growth and harvest of on-farm experiments in the central highland of Madagascar.
(TIF)

S2 Fig. Manhattan plots derived from GWAS analysis for traits: A) Root length, B) Root dry matter, and C) total dry matter in ratio values of treatments: Low-S/high-S. Manhattan plot shows negative logarithmic ((-log10 (P)) values of association for each SNP (Y axis), and SNP location along the 12 chromosomes (colored bar in X axis). Red line indicates a -log10 (P value) threshold of 5.
(TIF)

S3 Fig. Linkage Disequilibrium (LD) block for total panicle weight (TPW). Blue lines indicate the delineated region.
(TIF)

S4 Fig. Heatmap showing differential gene expression during vegetative and development stages reported in RiceXPro. Veg: vegetative; repr: reproductive; LP: low Phosphorus; LN: Low Nitrogen condition.
(TIF)

S5 Fig. Effect of variants on genes, transcripts, and protein sequence, as well as regulatory regions determined by Variant Effect Predictor (VEP, ensemble).
(TIF)

S1 Table. List of primers used in this study.
(DOCX)

S2 Table. Descriptive statistics and summary of phenotypic traits in on-farm trials (Experiment1).
(DOCX)

S3 Table. Complete list of QTLs associated with tolerance to low fertility soil.
(DOCX)

S4 Table. Allele frequency of total panicle weight (PWT) and straw dry weight (SDW) in the 3K-Rice Genome Project (3KRGP).
(DOCX)

S5 Table. Complete list of gene models included in GWAS associated loci.
(DOCX)

S6 Table. Descriptive statistics and summary of validation trial—Phenotypic traits (Experiment 2–2).
(DOCX)

S7 Table. Allelic distribution between the donor and recipient for the total panicle weight (PWT) QTL.
(DOCX)

Acknowledgments

The author would like to thank the IRRI gene bank for providing seeds of the accession used in this study.

Author Contributions

Conceptualization: Juan Pariasca-Tanaka, Ryokei Tanaka, Matthias Wissuwa.

Data curation: Mbolatantely Fahazavana Rakotondramanana, Sarah Tojo Mangaharisoa, Harisoa Nicole Ranaivo.

Investigation: Juan Pariasca-Tanaka, Mbolatantely Fahazavana Rakotondramanana, Sarah Tojo Mangaharisoa, Harisoa Nicole Ranaivo, Matthias Wissuwa.

Writing – original draft: Juan Pariasca-Tanaka.

Writing – review & editing: Juan Pariasca-Tanaka, Ryokei Tanaka, Matthias Wissuwa.

References

1. Dewar RE, Wright HE (1993) The culture history of Madagascar. Journal of World Prehistory, 7, 417–466.
2. Vaughan DA, Miyazaki S, and Miyashita K (2004) The rice genepool and human migrations. *In* "Biological resources and migration", Werner D. (ed.), Springer, Berlin. p. 1–13.
3. FAOSTAT, retrieve from http://www.fao.org/faostat/.
4. USDA (2018). Production, supply and distribution online. https://apps.fas.usda.gov/psdonline/app/index.html#/app/home.
5. Saito K., Vandamme E., Johnson J.-M., Tanaka A., Senthilkumar K., Dieng I., et al. (2019). Yield-limiting macronutrients for rice in sub-Saharan Africa. Geoderma, 338, 546–554. https://doi.org/10.1016/j.geoderma.2018.11.036
6. Nishigaki, T., Tsujimoto, Y., Rinasoa, S. et al. (2019). Phosphorus uptake of rice plants is affected by phosphorus forms and physicochemical properties of tropical weathered soils.
7. Rabeharisoa L, Razanakoto OR, Razafimanantsoa MP, Rakotoson T, Amery F, Smolders E (2012). Larger bioavailability of soil phosphorus for irrigated rice compared with rainfed rice in Madagascar: results from a soil and plant survey. Soil Use and Management. https://doi.org/10.1111/j.1475-2743.2012.00444.x
8. Tsujimoto Y, Sakata M, Raharinivo V, Juan Pariasca Tanaka, Toshiyuki Takai (2020): AZ-97 (Oryza sativa ssp. Indica) exhibits superior biomass production by maintaining the tiller numbers, leaf width, and leaf elongation rate under phosphorus deficiency, Plant Production Science, https://doi.org/10.1080/1343943X.2020.1808026
9. Rakotoson T & Tsujimoto Y. 2020. Pronounced effect of farmyard manure application on P availability to rice for paddy soils with low total C and low pH in the central highlands of Madagascar, Plant Production Science, 23:3, 314–321, https://doi.org/10.1080/1343943X.2020.1740601
10. International Fertilizer Development Center (2013) https://ifdc.org/.
11. Vandamme E, Rose TJ, Saito K, Yeong K, Wissuwa M (2015) Integration of P acquisition efficiency, P utilization efficiency and low grain P concentrations into P-efficient rice genotypes for specific target environments. Nutrient Cycling in Agroecosystems https://doi.org/10.1007/s10705-015-9716-3
12. Mori A., Fukuda T., Vejchasarn P., Nestler J., Pariasca-Tanaka J., Wissuwa M. (2016). The role of root size versus root efficiency in phosphorus acquisition in rice. J. Exp. Bot. 67, 1179–89. https://doi.org/10.1093/jxb/erv557 PMID: 26842979
13. Wissuwa M., Kondo K., Fukuda T., Mori A., Rose M.T., Pariasca-Tanaka J., et al. (2015). Unmasking novel loci for internal phosphorus utilization efficiency in rice germplasm through Genome-Wide Association Analysis. PLOS ONE 10(4): e0124215. https://doi.org/10.1371/journal.pone.0124215 PMID: 25923470
14. Minten B and Barrett CB (2008) Agricultural technology, productivity, and poverty in Madagascar. World Development 36: 797–822. https://doi.org/10.1016/j.worlddev.2007.05.004
15. Brar D.S., Khush G.S. (2018) Wild Relatives of Rice: A Valuable Genetic Resource for Genomics and Breeding Research. In: Mondal T., Henry R. (eds) The Wild Oryza Genomes. Compendium of Plant Genomes. Springer.
16. McCouch S.R., McNally K.L., Wang W. and Sackville Hamilton R. (2012) Genomics of gene banks: A case study in rice. American Journal of Botany, 99: 407–423. https://doi.org/10.3732/ajb.1100385 PMID: 22314574

17. Famoso A.N., Zhao K., Clark R.T., Tung C.W., Wright M.H., et al. (2011) Genetic architecture of aluminum tolerance in rice (Oryza sativa) determined through genome-wide association analysis and QTL mapping. PLoS Genet. 7: e1002221. https://doi.org/10.1371/journal.pgen.1002221 PMID: 21829395

18. Shrestha A, Dziwornu AK, Ueda Y, Wu L-B, Mathew B, Frei M (2018) Genome-wide association study to identify candidate loci and genes for Mn toxicity tolerance in rice. PLoS ONE 13(2): e0192116. https://doi.org/10.1371/journal.pone.0192116 PMID: 29425206

19. Pariasca-Tanaka J., Baertschi C., Wissuwa M., 2020. Identification of Loci Through Genome-Wide Association Studies to Improve Tolerance to Sulfur Deficiency in Rice. Frontiers in Plant Science 10, 1668. https://doi.org/10.3389/fpls.2019.01668 PMID: 32010158

20. Wang F.M., Longkumer T., Catausan S.C., Calumpang C.L.F., Tarun J.A., Cattin J., et al. (2018). Genome wide association and gene validation studies for early root vigor to improve direct seeding of rice. Plant, Cell and Environment 41, 2731–2743. https://doi.org/10.1111/pce.13400 PMID: 29981171

21. Alexandrov N, Tai S, Wang W, Mansueto L, Palis K, Fuentes RR, et al. (2015) SNP-Seek database of SNPs derived from 3000 rice genomes. Nucleic Acids Res. 2015 Jan; 43(Database issue):D1023–7. https://doi.org/10.1093/nar/gku1039 PMID: 25429973

22. Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using lme4." Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

23. Tanaka R, Mandaharisoa ST, Rakotondramanana M, Ranaivo HN, Pariasca-Tanaka J, Kajiya-Kanegae H, et al. 2021. From gene banks to farmer's fields: Using genomic selection to identify donors for a breeding program in rice to close the yield gap on smallholder farms. Theoretical and Applied Genetics (in press?).

24. Endelman J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250–255. https://doi.org/10.3835/plantgenome2011.08.0024

25. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007 Oct 1; 23(19):2633–5. https://doi.org/10.1093/bioinformatics/btm308 PMID: 17586829

26. Barrett J.C., Fry B., Maller J., Daly M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005 Jan 15; 21(2):263–5. https://doi.org/10.1093/bioinformatics/bth457 PMID: 15297300

27. Yoshida S., Forno D.A., Cock J.H. (1971). Laboratory manual for physiological studies of rice. Los Banos, Philippines: International Rice Research Institute.

28. Pariasca-Tanaka J., Lorieux M., He C. et al. Development of a SNP genotyping panel for detecting polymorphisms in Oryza glaberrima/O. sativa interspecific crosses. Euphytica 201, 67–78 (2015). https://doi.org/10.1007/s10681-014-1183-4

29. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

30. Saito K, Dieng I, Toure AA, Somado EA, Wopereis MSC (2015) Rice yield growth analysis for 24 African countries over 1960–2012. Glob Food Secur 5:62–69.

31. Tsujimoto Y, Rakotoson T, Tanaka A, Saito. 2019. Challenges and opportunities for improving N use efficiency for rice production in sub-Saharan Africa. Agronomy & Crop Ecology. https://doi.org/10.1080/1343943X.2019.1617638

32. Wang, Yiqin and Tang, Jiuyou and Li, Hua and Teng, Zhenfeng and Wang, Wei and Xu, Fan et al (2020) A NB-ARC-CRRSP Signaling Module Triggers HR-Like Cell Death and Associated Disease Resistance in Rice by Suppressing Antioxidant Defense Systems. SSRN.

33. Puyol V and Wissuwa M (2018) Contrasting development of lysigenous aerenchyma in two rice genotypes under phosphorus deficiency. BMC Research Notes 11:60. https://doi.org/10.1186/s13104-018-3179-y PMID: 29357942

Rhizosphere 14 (2020) 100198

Contents lists available at ScienceDirect

Rhizosphere

journal homepage: www.elsevier.com/locate/rhisph

Phosphorus deficiency tolerance in *Oryza sativa*: Root and rhizosphere traits

Tovohery Rakotoson[a,*,1], Maire Holz[b,1], Matthias Wissuwa[c]

[a] Laboratoire des RadioIsotopes, Université d'Antananarivo, BP 3383, Route d'Andraisoro, 101, Antananarivo, Madagascar
[b] Group of Isotope Biogeochemistry and Gas Fluxes, Leibniz Centre for Agricultural Landscape Research (ZALF) e.V., Müncheberg, Germany
[c] Japan International Research Center for Agricultural Sciences (JIRCAS), 1-1 Ohwashi, Tsukuba, Ibaraki, 16 305-8686, Japan

ARTICLE INFO

Keywords:
Oriza sativa (rice)
P uptake efficiency
Phosphatase activity
Rhizodeposition
^{13}C labelling

ABSTRACT

We tested whether root exudation confers high P acquisition efficiency (PAE) in rice by quantifying ^{13}C allocation and rhizosphere phosphatase activity for three contrasting genotypes. Phosphatase activity per root surface was similar for all genotypes, indicating that it is not conferring PAE in rice. DJ123 allocated more ^{13}C from shoots to roots compared to Nerica4 and Chomrong Dhan in accordance to increased root biomass of this genotype. The allocation of ^{13}C from roots to the soil was similar between genotypes indicating no difference of root exudation. However, DJ123 showed lower PAE compared to the other two genotypes. We conclude that in the tested genotypes, root systems size rather than root exudation confers PAE.

Phosphorus uptake is in major parts governed by root architecture and exploration potential of soil volume by the root system particularly in highly P-deficient soils (Marschner and Rengel, 2012). Rice genotypes tolerant to P deficiency are known to expand more roots and roots hairs which are important traits for exploring greater soil volumes. However, for rice genotypes adapted to upland conditions, high exploration of soil volume by increased root surface area and root hair growth can only partly explain P acquisition efficiency (PAE: Total P uptake/Root dry weight (TP/RDW)) (Mori et al., 2016; Nestler and Wissuwa, 2016) and it can be assumed that root exudation partly explains high P uptake efficiency in the tested genotypes (Mori et al., 2016). Particularly under P-starving condition, plant roots can release diverse organic and inorganic products (root exudates) that can increase the availability of P as a response (Louw-Gaume et al., 2017). Phosphatases on one hand may mineralize organic P which is not available to the plant (George et al., 2002). On the other hand, the release of low molecular root exudates may enhance microbial turnover and herby increase P mineralization (Gunina and Kuzyakov, 2015). However, knowledge on root exudation related to PAE in upland rice is scarce and this study aims to assess the effect of root exudation on P uptake of rice genotypes tolerant and non-tolerant to P deficiency.

Rice plants (Nerica 4, Chomrong Dhan and DJ123) were grown in rhizoboxes with an inner size of $28 \times 28 \times 1$ cm. The rhizoboxes were filled with upland soil from the central highlands of Madagascar (Olsen extractable-P: 6.6 mg P kg^{-1}, Organic C: 25.8 g kg^{-1}, CEC: 3.5 $cmol_c$ kg^{-1}, pH-$CaCl_2$: 4.7, clay: 45.4%, loam: 28.7%, sand: 25.9%). The experiment comprised five replicates per treatment. Soil volumetric water content was kept at 23–25% during plant growth. The temperature was 30 °C during the day and 25 °C during night, relative humidity 50%/80% day/night and the photoperiod was 12 h. During the growth period, root elongation was traced by marking the root tips on the transparent plexiglas surface each day. The elongation rate was calculated using the Smart Root plugin (Lobet et al., 2011) in ImageJ (https://fiji.sc/). After 5 weeks, plants were labelled with ^{13}C to trace carbon allocation in plant and soil (Pausch and Kuzyakov, 2018). Labelling was conducted as described in Holz et al.(2018). Each plant received 0.167 g of 99 atom% $NaHCO_3$. 24 h after ^{13}C labelling, rhizoboxes were opened and photos of the root system were taken. Root surface area was segmented from these images, using the Roottracker 2D (Menon et al., 2007) and calculated in Matlab (The MathWorks). For soil zymography measurements, polyamide membranes (Tao Yuan, China) with a pore size of 0.45 µm were soaked in a solution containing 4-methylumbelliferyl-phosphate (MUF-P) and placed on the rhizobox surface for 60 min. After removal of the membrane, zymographs were taken at 360 nm wavelength. More details on the procedure can be found in Razavi et al.(2019). For calibration of zymographs, solutions with increasing µM 4-methylumbelliferone (MUF) were prepared and imaged (Holz et al., 2019; Razavi et al., 2016). The obtained grey values from the images were related to the phosphatase activity and the obtained calibration line was applied to the zymographs. To remove the background signal, the signal of a control membrane was subtracted from the image. Total phosphatase activity per rhizobox surface as well

* Corresponding author.
E-mail addresses: tovohery.rakotoson@gmail.com (T. Rakotoson), maire.holz@zalf.de (M. Holz), wissuwa@affrc.go.jp (M. Wissuwa).
[1] Joint first authors.

https://doi.org/10.1016/j.rhisph.2020.100198
Received 28 January 2020; Received in revised form 22 March 2020; Accepted 22 March 2020
Available online 29 March 2020

Table 1
Shoot/root dry weight, root elongation rate, shoot and root P concentration as well as total P uptake per plant and P acquisition efficiency (PAE). The values in the brackets are standard errors. Letters behind the values indicate statistical differences, $\alpha = 0.05$.

Genotype	SDW (g)	RDW (g)	Root elongation (cm d^{-1})	Shoot [P] (mg g^{-1})	Root [P] (mg g^{-1})	Total P (mg $plant^{-1}$)	PAE (mg P g^{-1})
Nerica4	0.98^{a} (0.08)	0.92^{a} (0.01)	2.60^{a} (0.34)	1.46^{a} (0.08)	1.06^{ab} (0.03)	2.37^{a} (0.19)	2.68^{ab} (020)
Chomrong Dhan	1.31^{a} (0.15)	0.92^{a} (0.02)	2.00^{a} (0.17)	1.34^{a} (0.06)	1.21^{a} (0.05)	2.81^{a} (0.32)	3.28^{a} (0.30)
DJ 123	1.78^{b} (0.09)	1.97^{b} (0.07)	3.97^{b} (0.23)	1.25^{a} (0.10)	0.95^{b} (0.03)	4.07^{b} (0.21)	2.08^{b} (0.13)

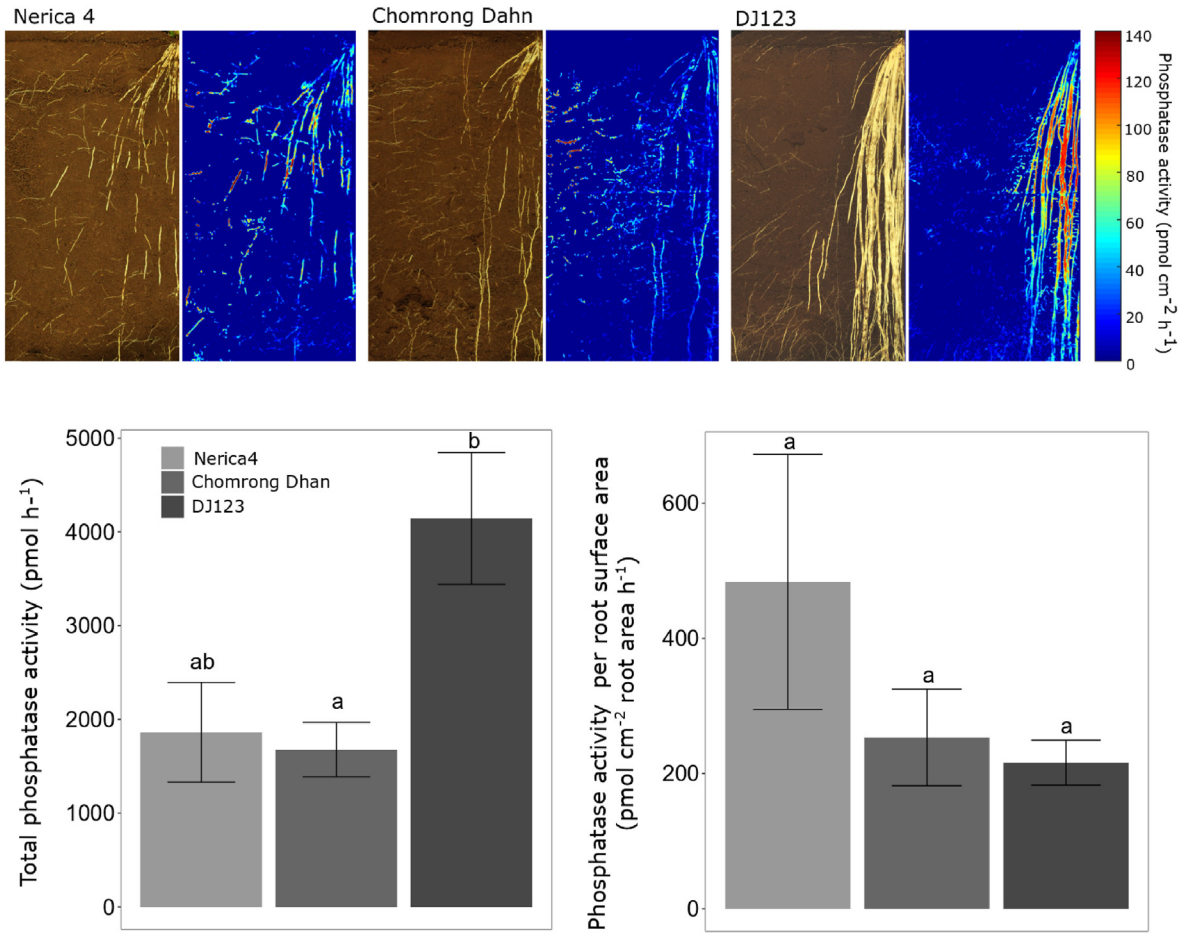


Fig. 1. Top: Exemplary zymographs of phosphatase activity for the three genotypes Nerica 4, Chomrong Dhan and DJ123. Bottom: Total phosphatase activity as well as phosphatase activity related to the root surface for the same genotypes. Error bars indicate standard error, n = 5. The letters above the bars indicate statistical significances, $\alpha = 0.05$.

as phosphatase activity per root surface area were then calculated in Matlab. Plant harvest was done 24 h after ^{13}C pulse labelling.

Shoot and root biomass for DJ123 was 1.7- and 2-fold higher than for Nerica 4 and Chomrong Dhan respectively (Table 1). Root biomass was highest for DJ123 which is reflected in increased root elongation rate being twice as high for this genotype compared to Nerica4 and Chomrong Dhan. Subsequently, DJ123 took up at least 45% more P in total than the two other genotypes. However, due to the even bigger difference in root growth, DJ123 had the lowest root P acquisition efficiency (PAE: TP/RDW), while highest PAE was found for the genotype Chomrong Dhan (Table 1). These results would indicate that root size is the main driver for genotypic differences in P uptake. For DJ123, our results on PAE contradict those by Mori et al. (2016) and Wissuwa et al. (2020) who found that this genotype had a higher PAE than Nerica4 under P deficient conditions. However, these studies were conducted under conditions where a larger root system would explore a larger soil volume, which may not have been the case in the narrow root boxes that severely restricted root growth (as seen by the dense roots growing along the glass plate in DJ123) and would disadvantage genotypes with larger root systems more. Estimates of PAE therefore need to be treated with caution in small rhizobox systems. Under non-limiting P availability in soil, Wissuwa et al. (2020) found no significant difference between shoot biomass and P content between DJ123 and Nerica4. In practice, a recent study conducted in Madagascar showed that DJ123 is the most tolerant to P deficiency followed by Chomrong Dhan and then Nerica4 with this latter as potentially the best under more fertile condition (Ranaivo et al., unpublished data).

Phosphatase activity was highest in the rhizosphere (Fig. 1) and relatively low in the bulk soil which corresponds to previous findings (Holz et al., 2019; Ma et al., 2018; Razavi et al., 2016). Total phosphatase activity was higher for DJ123 than for the other two genotypes but did not differ between genotypes when related to the root surface area indicating that it is not conferring increased P uptake efficiency in the tested rice genotypes (Fig. 1).

In contrast to phosphatase activity, allocation of ^{13}C from shoots to roots was higher for DJ123 than for Nerica4 and Chomrong Dhan

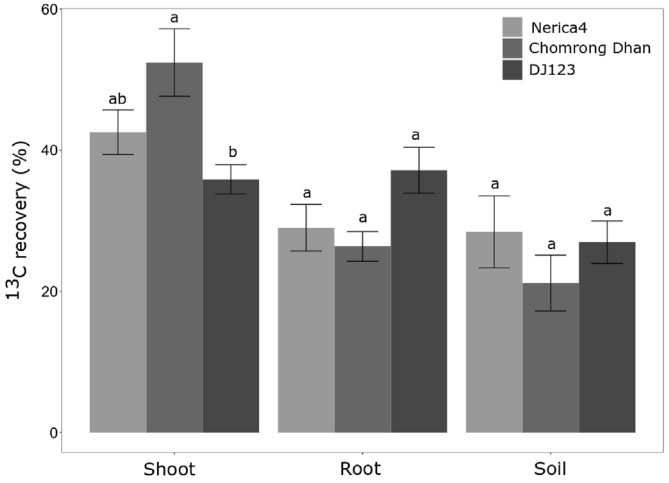


Fig. 2. 13C allocation (% of added) to shoots, roots and soil for the three genotypes Nerica 4, Chomrong Dahn and DJ123 and the proportion of ^{13}C allocated from shoot to root and from root to soil. Error bars indicate standard error, n = 5. The letters above the bars indicate statistical significances, $\alpha = 0.05$.

(Fig. 2) after 5 weeks of plant growth which is in accordance with the increased root biomass of this genotype. However, allocation of ^{13}C from roots to the soil did not differ between the genotypes which indicates that the genotypes did not differ in their root exudation at the time of measurement. Therefore, although increased rhizodeposition can potentially increase nutrient availability (Herman et al., 2006; Landi et al., 2006) it is unlikely that rhizodeposition affected P availability in our study. In contrast, our results indicate that increased P uptake for DJ123 compared to Nerica4 and Chomrong Dhan can be explained by increased root system size which may be the result of preferential C allocation to roots in DJ123. In less P-deficient soil as a difference to our soil used in our study, rice plants would have less root exudation as a response and could change soil P release as a consequence (Bhattacharyya et al., 2013).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was partly supported by the Science and Technology Research Partnership for Sustainable Development (SATREPS, Japan Science and Technology Agencyy (JST)/Japan International Cooperation Agency (JICA) (Grant No. JPMJSA1608). Maire Holz acknowledges the German Research Foundation for granting the projects CA 921/3–1 and KU 1184/33–1. The soils used were imported to Japan with permission of 30Y1175.

References

Bhattacharyya, P., Das, S., Adhya, T.K., 2013. Root exudates of rice cultivars affect rhizospheric phosphorus dynamics in soils with different phosphorus statuses. Commun. Soil Sci. Plant Anal. 44, 1643–1658. https://doi.org/10.1080/00103624.2013.769562.

George, T.S., Gregory, P.J., Wood, M., Read, D., Buresh, R.J., 2002. Phosphatase activity and organic acids in the rhizosphere of potential agroforestry species and maize. Soil Biol. Biochem. 34, 1487–1494.

Gunina, A., Kuzyakov, Y., 2015. Sugars in soil and sweets for microorganisms: review of origin, content, composition and fate. Soil Biol. Biochem. 90, 87–100. https://doi.org/10.1016/j.soilbio.2015.07.021.

Herman, D.J., Johnson, K.K., Jaeger, C.H., Schwartz, E., Firestone, M.K., 2006. Root influence on nitrogen mineralization and nitrification in rhizosphere soil. Soil Sci. Soc. Am. J. 70, 1504. https://doi.org/10.2136/sssaj2005.0113.

Holz, M., Zarebanadkouki, M., Carminati, A., Howind, J., Kaestner, A., Spohn, M., 2019. Increased water retention in the rhizosphere allows for high phosphatase activity in drying soil. SPlant Soil. https://doi.org/10.1007/s11104-019-04234-3.

Holz, M., Zarebanadkouki, M., Kaestner, A., Kuzyakov, Y., Carminati, A., 2018. Rhizodeposition under drought is controlled by root growth rate and rhizosphere water content. Plant Soil 423, 429–442. https://doi.org/10.1007/s11104-017-3522-4.

Landi, L., Valori, F., Ascher, J., Renella, G., Falchini, L., Nannipieri, P., 2006. Root exudate effects on the bacterial communities, CO2 evolution, nitrogen transformations and ATP content of rhizosphere and bulk soils. Soil Biol. Biochem. 38, 509–516.

Lobet, G., Pagès, L., Draye, X., 2011. A novel image-analysis toolbox enabling quantitative analysis of root system Architecture. Plant Physiol. 157 (29) LP – 39.

Louw-Gaume, A.E., Schweizer, N., Rao, I.M., Gaume, A.J., Frossard, E., 2017. Temporal differences in plant growth and root exudation of two Brachiaria grasses in response to low phosphorus supply. Trop. Grasslands-Forrajes Trop. 5, 103–116.

Ma, X., Zarebanadkouki, M., Kuzyakov, Y., Blagodatskaya, E., 2018. Spatial patterns of enzyme activities in the rhizosphere : effects of root hairs and root radius Spatial patterns of enzyme activities in the rhizosphere : E ff ects of root hairs and root radius. Soil Biol. Biochem. 118, 69–78. https://doi.org/10.1016/j.soilbio.2017.12.009.

Marschner, P., Rengel, Z., 2012. Chapter 12 - nutrient availability in soils. In: Marschner, P. (Ed.), Marschner's Mineral Nutrition of Higher Plants, third ed. Academic Press, San Diego, pp. 315–330. https://doi.org/10.1016/B978-0-12-384905-2.00012-1.

Menon, M., Robinson, B., Oswald, S.E., Kaestner, A., Abbaspour, K.C., Lehmann, E., Schulin, R., 2007. Visualization of root growth in heterogeneously contaminated soil using neutron radiography. Eur. J. Soil Sci. 58, 802–810.

Mori, A., Fukuda, T., Vejchasarn, P., Nestler, J., Pariasca-Tanaka, J., Wissuwa, M., 2016. The role of root size versus root efficiency in phosphorus acquisition of rice. J. Exp. Bot. 67, 1179–1189.

Nestler, J., Wissuwa, M., 2016. Superior root hair formation confers root efficiency in some, but not all, rice genotypes upon P deficiency. Front. Plant Sci. 7, 1935. https://doi.org/10.3389/fpls.2016.01935.

Pausch, J., Kuzyakov, Y., 2018. Carbon input by roots into the soil: quantification of rhizodeposition from root to ecosystem scale. Global Change Biol. 24, 1–12. https://doi.org/10.1111/gcb.13850.

Razavi, B.S., Zarebanadkouki, M., Blagodatskaya, E., 2016. Rhizosphere shape of lentil and maize : spatial distribution of enzyme activities. Soil Biol. Biochem. 96, 229–237. https://doi.org/10.1016/j.soilbio.2016.02.020.

Razavi, B.S., Zhang, X., Bilyera, N., Guber, A., Zarebanadkouki, M., 2019. Soil zymography: simple and reliable? Review of current knowledge and optimization of the method. Rhizosphere 11, 100161. https://doi.org/10.1016/j.rhisph.2019.100161.

Wissuwa, M., Gonzalez, D., Watts-Williams, S.J., 2020. The contribution of plant traits and soil microbes to phosphorus uptake from low-phosphorus soil in upland rice varieties. Plant Soil. https://doi.org/10.1007/s11104-020-04453-z.

PLANT PRODUCTION SCIENCE
2021, VOL. 24, NO. 1, 41–51
https://doi.org/10.1080/1343943X.2020.1808026

RESEARCH ARTICLE

OPEN ACCESS Check for updates

AZ-97 (*Oryza sativa* ssp. *Indica*) exhibits superior biomass production by maintaining the tiller numbers, leaf width, and leaf elongation rate under phosphorus deficiency

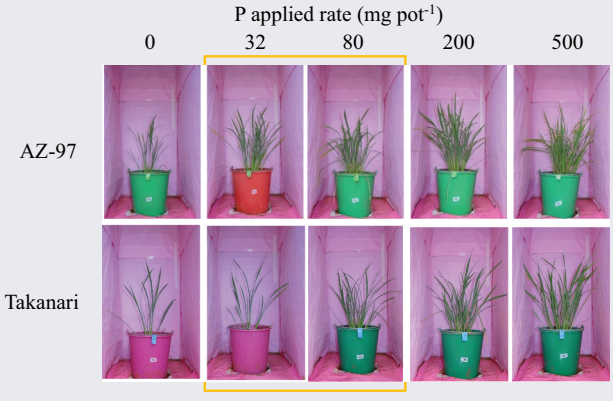
Yasuhiro Tsujimoto[a], Mitsukazu Sakata[b], Viviane Raharinivo[c], Juan Pariasca Tanaka[a] and Toshiyuki Takai[a]

[a]Japan International Research Center for Agricultural Sciences (JIRCAS), Tsukuba, Japan; [b]Faculty of Agriculture and Marine Science, Kochi University, Nankoku, Japan; [c]Department of Rice Research, Centre National de Recherche Appliquée au Développement Rural (FOFIFA), Antananarivo, Madagascar

ABSTRACT

Shoot morphology in rice plants changes in response to P deficiency. However, how genotypic variations in these morphological changes affect the canopy development and biomass production have hardly been explored. The study aimed to identify specific shoot morphological traits that confer biomass production under P deficiency. Four rice genotypes, including AZ-97 (WAS 63–22-5-9-10-1), which exhibits high performance under highly P-deficient fields in Madagascar, were grown for 56 days in flooded pots over several P application rates to simulate P-sufficient and moderately, severely, and extremely P-deficient conditions. AZ-97 had superior shoot P contents and biomass than Takanari, a high-yielding cultivar, and X265, a common high-yielding cultivar in Madagascar at severely to moderately P-deficient conditions. Shoot biomass was highly correlated with projected leaf area (PLA) from the early growth stage, and tiller number, leaf width, and leaf elongation rate explained the variations in PLA. These morphological traits reduced significantly with decreased P application rates, while reduction in AZ-97 was small relative to the other genotypes, even for equivalent shoot P contents. As the result, AZ-97 had greater PLA per unit of shoot P content at equivalent shoot P contents. The result indicates that lower sensitivity and degrees of change in shoot morphology when exposed to P deficiency stress could be a key trait facilitating the maintenance of captured radiation and subsequently influencing genotypic differences in external P uptakes and biomass production. AZ-97 is a potential donor with such traits that can offer an additional avenue for genetic improvement toward P-efficient rice production.

Abbreviations DAT: days after transplanting; LN: leaf number in the main stem; PLA: projected green leaf area; PUE: phosphorus use efficiency; SSA: Sub-Saharan Africa.



ARTICLE HISTORY
Received 20 March 2020
Revised 27 May 2020
Accepted 28 July 2020

KEYWORDS
Leaf elongation rate; leaf width; *Oryza Sativa*; P deficiency; projected leaf area; shoot morphology; tillering

Introduction

Existing rock phosphate reserves could be exhausted over the next 50–100 years and the quality of reserves is declining with the increasing costs of extraction, processing, and shipping (Cordell et al., 2009). Considering the finite nature of P fertilizer resources, it is vital to investigate potential sustainable crop production

CONTACT Yasuhiro Tsujimoto tsjmt@affrc.go.jp
Supplemental data for this article can be accessed here.

strategies that involve the efficient utilization of the supplied P and available P in soils. Such strategies are obviously required in high-input systems in developed countries where excess P fertilizer amounts applied in agricultural systems cause environmental challenges such as the eutrophication of lakes and marine estuaries (Vance et al., 2003). Moreover, efficient use of P is critical in low-input systems such as in rice production by smallholder farmers in Sub-Saharan Africa (SSA), where rice yield – despite the increasing demand of rice for food security – is restricted largely by P-deficient soils and inadequate P fertilizer inputs (Nziguheba et al., 2015; Saito et al., 2019; Tsujimoto et al., 2019).

Genetic improvement is one of the major strategies for the efficient utilization of the supplied P and available P in soils. For instance, the identification of a P starvation tolerance gene (*PSTOL1*), which enhances early root growth and promotes P uptake (Gamuyao et al., 2012), and the identification of certain genotypes with high root efficiency (P uptake per root surface area) (Mori et al., 2016) could facilitate the development of rice cultivars that produce high yields even under highly P-deficient soils with low rates of P fertilizer application. Wissuwa and Ae (2001) demonstrated that genotypic variations in P uptake in rice were largely explained by their root sizes when cultivated under P-deficient conditions. In addition, understanding root architectural traits, such as steep root angle (Uga et al., 2013), deep rooting (Obara et al., 2010), and shallow rooting (Uga et al., 2012), and the use of such quantitative trait loci (QTL) in breeding activities could also offer avenues for improving P-use efficiencies by enabling the cultivation of rice with root systems that are adapted to the P availability conditions in local soil profiles or localized P applications. Lynch and Brown (2001) reported the development of more surface roots as a typical plant response to P deficiency because most of the available P are accumulated in the topsoil. As such, the root system plays a primary role under P deficiency and has been studied extensively for rice genetic improvement to enhance P deficiency stress tolerance (Campos et al., 2018; Wissuwa, 2005).

In contrast, hardly any studies have explored aboveground morphological variations and adaptations of rice genotypes in response to P deficiency. Rice plants exhibit unique aboveground morphological changes under P deficiency stress, including reduced tillering, narrow leaves, erect leaves, and retarded leaf elongation rates (Dobermann & Fairhurst, 2000; Luquet et al., 2005). Such morphological changes are potentially linked to leaf area expansion and captured radiation, and, therefore, influence photosynthetic assimilation and biomass production under P-deficient conditions. The most well-understood physiological mechanism in shoot morphological change in rice is an increase in strigolactones in root exudates, which plays a dual role under P deficiency; (i) inducing hyphal branching of arbuscular mycorrhizal (AM) fungi for exploiting available P in soils and (ii) inhibiting tiller bud outgrowth in rice seedlings (Umehara et al., 2010). Then, it is considered that assimilates that are no longer required for new tiller growth are partitioned to maintain root growth, which results in high root to shoot mass ratio as an initial plant response to P deficiency (e.g., Mollier & Pellerin, 1999; Wissuwa et al. 2005). However, none of the studies above explored genotypic variations and how the inhibition of tillering under P deficiency influence canopy development and biomass production. Therefore, understanding genotypic variations in the aboveground morphological characteristics and their correlation with biomass production under P-deficient conditions could offer an additional avenue for genetic improvement that could facilitate P-efficient rice production. We hypothesized that canopy development is an important trait to maintain the positive chain cycle among photosynthetic carbon assimilation, allocation of assimilates to the root system, and external P uptakes, which in turn lead to high biomass under P deficiency.

In the current study, we targeted two rice accessions that demonstrated the highest grain yields among 300 accessions tested under severely P-deficient fields in the central highlands of Madagascar in our preliminary on-farm trials. The two accessions were AZ-78 (FACAGRO 64::IRGC 82,059, *indica* type originating from India) and AZ-97 (WAS 63–22-5-9-10-1, *indica* type originating from Senegal). The set of 300 accessions were selected from 3000 rice accessions that cover a wide range of genotypes and are available publicly at the Rice SNP-Seek Database (http://snp-seek.irri.org) (Mansueto et al., 2017). We conducted pot experiments to 1) verify the superior performance, and 2) identify specific aboveground morphological changes/adaptations of the two varieties and their relationship with biomass production under various P-deficient conditions.

Materials and methods

Experiment design

Pot experiments were conducted in a screenhouse at the National Center for Applied Research on Rural Development (FOFIFA) at Antananarivo, Madagascar (18° 52′S, 47°33′E, 1310 m alt.). The screenhouse was approximately 4 m high and all sides were covered with mesh nets that ensured more or less ambient temperature and natural sunlight conditions. The daily mean temperature

inside the screenhouse was recorded by Thermo Recorder (TR-72 U, T&D Corporation, Nagano, Japan) and ranged from 20.0°C to 25.3°C throughout the experimental period.

The experiment soil was collected from a farmer's field (0–20 cm). The major physico-chemical properties are summarized in Table 1. Briefly, the experimental soil was clay loam with a pH of 5.4 and with extremely low amounts of P and cations. The amount of oxalate-extractable P – a suitable indicator of P availability for lowland rice production in the region (Rabeharisoa et al., 2012) – is merely 40.7 mg kg^{-1}, which is the lowest level among many lowland and upland rice fields in Madagascar (Kawamura et al., 2019).

After being air-dried and sieved, this extremely P-deficient soil was put into 5-L plastic pots (20 cm in height, 20 cm in diameter) with 4 kg of soil per pot. Afterward, five P treatments were established by applying $NaH_2PO_4 \cdot 2H_2O$ at rates of 0, 32, 80, 200, and 500 mg P pot^{-1} (0P, 32P, 80P, 200P, and 500P, respectively, hereafter). The P application rates were determined based on the result of preliminary pot experiments that revealed a wide range of morphological changes in tiller number and leaf width by growing X265, a popular high-yielding variety in the region. The rates were approximately equivalent to 0, 10, 25, 64, and 159 kg P ha^{-1}, as estimated based on pot size. Nitrogen (N) and Potassium (K) were supplied adequately to all the pots using NH_4NO_3 and KCl, respectively, at the rates of 608 mg pot^{-1} for both N and K, to avoid any potential influence of N and K deficiency on rice growth. All the nutrients were mixed uniformly with soils and puddled a day prior to transplantation.

Four genotypes were used – the aforementioned two *indica* varieties (AZ-78 and AZ-97), X265, and Takanari, a high-yield *indica* variety developed in Japan. Twenty-day-old seedlings of the four varieties grown in P-free sand were transplanted one plant per hill and one hill per pot. Five levels of P treatments factorially combined with the four varieties were allocated in a randomized complete block design with six replicates and grown under continuously flooded conditions.

Table 1. Soil properties of the pot experiments.

Parameter	Unit	
Clay[a]	%	30.7
Silt[a]	%	35.9
Sand[a]	%	33.5
pH (1:5 H_2O)	-	5.4
Total N[b]	g kg^{-1}	1.5
Total C[b]	g kg^{-1}	18.5
Oxalate-P[c]	mg kg^{-1}	40.7
Available P[d]	mg kg^{-1}	8.0
CEC[e]	c mol kg^{-1}	7.4
Exchangeable Ca[e]	c mol kg^{-1}	0.33
Exchangeable K[e]	c mol kg^{-1}	0.17
Exchangeable Mg[e]	c mol kg^{-1}	0.34
Exchangeable Na[e]	c mol kg^{-1}	0.05

a: Sieving and pipetting method.
b: NC analyzer, Sumigraph NC-220F (SCAS, Tokyo, Japan).
c: Inductively coupled plasma mass spectrometer (ICPE-9000, Shimazu, Japan) after oxalate extraction.
d: Bray-I extraction method.
e: Ammonium acetate extract method at pH 7.0.

Measurements

Tiller number, leaf age or the number of leaves on the main stem (LN), leaf width of the open-top leaf, and projected green leaf area (PLA) in each pot were observed weekly from 19 days after transplanting (DAT). PLA was estimated using digital imagery analysis according to the method of Tsujimoto et al. (2016). Briefly, each pot was sequentially put into a 1.6 m-height carton box with a small lens hole on the top and one side open (the opposite side of the side illuminated by sun's rays) to avoid reflectance from direct sunlight, and then a digital photograph was taken virtually from the lens hole. The PLAs on individual images were determined using image processing software (Image J, NIH, USA). LN was determined by counting the number of leaves in the main stem starting from the incomplete (prophyll) leaf. If the top leaf had not expanded fully, LN was estimated based on the ratio of the length of the elongating top leaf to that of the preceding leaf.

At 56 DAT, shoot biomass was determined by sampling the whole plants at the soil surface and dried at 70°C for 3 days in a ventilated oven. Each shoot sample was ground to fine powder using a high-speed vibrating sample mill (T1–100, CMT Co. Ltd., Fukushima, Japan). Afterward, shoot P concentrations were determined using the molybdate blue method (Murphy & Riley, 1962) after dry-ashing at 550°C for 2 h and digesting with 0.5 M HCl. Shoot P content and internal PUE were determined as follows:

$$\text{Shoot P content (mg P pot}^{-1}) = \text{shoot biomass} \times \text{shoot P concentration}$$

$$\text{PUE (g biomass per mg P)} = \text{shoot biomass/shoot P content}$$

Based on the synchronous leaf and tiller development theory – rice has the potential to produce one tiller on the nth node when the new leaf on n + 3th node emerges (Katayama, 1951) –, potential tiller number as a function of LN was calculated at 41 DAT using the following equation:

$$\text{Number of potential tiller} = 0.5 \times LN^2 - 4.5 \times LN + 13 \quad LN > 4 \qquad (1)$$

based on the assumptions that 1) the first tiller emerged from the 1st node on the main stem and 2) only primary and secondary tillers emerged.

Table 2. F value and level of significance in the GLM test.

	Shoot Biomass	P uptake	PUE	PLA at 41 DAT	Tiller No. at 41 DAT	LN at 41 DAT	Leaf width at 41 DAT
Genotype (G)	7.1***	14.4***	5.6**	13.6***	63.8***	83.9***	5.6**
P level (P)	437.6***	1792.9***	442.8***	534.7***	668.0***	196.5***	442.8***
G×P	3.6*	13.3***	3.1*	3.8*	3.7*	ns.	3.1*

***P<0.001, **P<0.01, *P<0.05, ns. not significant.

Further, we assessed the efficiency in canopy development under P deficiency by dividing PLA by shoot P content. We used the PLA at 41 DAT before the start of shoot elongation and leaf senescence.

Statistical analysis

JMP v14 software (SAS Institute Inc., Tokyo, Japan) was used to perform the statistical analyses. A generalized linear model (GLM) was used to determine the single and interaction effects of genotype (G) and P application rate (P) on the measured variables. The replicate (n = 6) was treated as a random factor. F-values from the GLM tests for each variable are summarized in Table 3. Afterward, mean values were compared separately for each P application rate using Tukey's honestly significant difference test at a 5% significance level. The linear regression coefficients of shoot P content on biomass (Δ Shoot biomass/Δshoot P content) or the P-response slope parameters with the intercept at 0 were compared among genotypes by the analysis of covariance. A stepwise regression analysis was performed to identify significant explanatory factors of PLA using LN, tiller number, leaf width, and total number of leaves as candidate factors. In the stepwise process, the 'selection' and 'removal' of factors were controlled with an F-value of P < 0.05. Thereafter, a multiple regression model was developed using the selected parameters.

Results

Genotype differences in shoot biomass production, shoot phosphorus content, and phosphorus-use efficiency under various phosphorus deficiency status

Shoot biomass ranged from 5.0 to 36.4 g pot^{-1} at 56 DAT, with significant effects of and interaction between genotype and P application rates (Figure 1, S1 Fig). The various

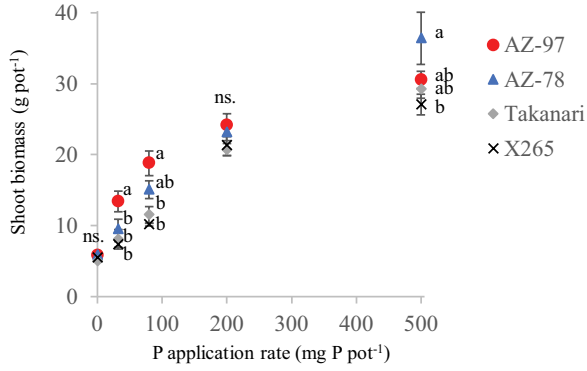


Figure 1. Genotype comparison in shoot biomass at 56 DAT at difference levels of P fertilizer applied.

P application rates in the present study produced diverse P-deficiency statuses. Shoot biomass in the P-sufficient condition (500P) more or less plateaued against increased shoot P content (Figure 2). In the moderately P-deficient condition (200P), biomass production was 63–79% of the rate of production under 500P and curved against increased shoot P content. Under the severely P-deficient conditions at 80P and 32P, the biomass production was 26–43% (61% for AZ-97 at 80P) of the production at 500P, and responded linearly to an increase in shoot P content. In the extremely P-deficient condition at 0P, biomass production was <20% of the production under 500P. No significant genotypic differences were observed at 0P. However, AZ-97 produced significantly greater biomass than the other three genotypes, by 41–83%, at 32P, and, significantly greater biomass than X265 and Takanari, by 63–85%, at 80P. Genotypic differences in biomass became less apparent at 200P, while AZ-97 retained slightly higher biomass than Takanari and X265. At the P-sufficient condition of 500P, AZ-78 produced the greatest biomass, which was significantly greater than that of X265 by 35%.

The linear regression coefficients of shoot P content on biomass (intercept = 0) at the low P application rates from

Table 3. Genotype comparison in shoot Phosphorus content at 56 DAT at different levels of P fertilizer applied.

	P application rate									
Genotype	0P		32P		80P		200P		500P	
AZ-97	5.4	a	14.0	a	20.8	a†	31.1	a	60.9	b
AZ-78	5.5	a	9.5	b	16.9	a†	35.5	a	76.9	a
Takanari	4.1	a	7.7	b	12.0	b	27.8	a	54.6	b
X265	4.4	a	5.9	b	10.5	b	31.1	a	56.2	b

Within each column, the same alphabets indicate no significant mean differences at P<0.05.
†Significant at *P*=0.10.

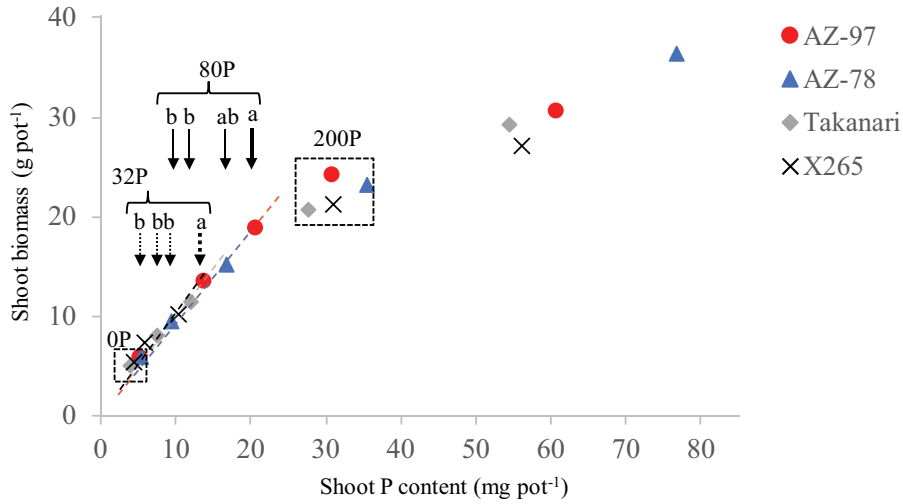


Figure 2. Relationship between shoot P content and shoot biomass at 56 DAT.

0P to 80P were 0.93 g mg^{-1} for AZ-97, 0.93 g mg^{-1} for AZ-97, 0.99 g mg^{-1} for Takanari, and 1.04 g mg^{-1} for AZ-97, showing no significant genotypic differences (Figure 2). The result indicates that greater biomass production of AZ-97 at 32P and 80P treatments was attributable primarily to the greater shoot P contents. The shoot P content of AZ-97 was significantly greater than the other three genotypes by 47–136% at 32P and by 73–99% than X265 and Takanari at 80P (Table 2).

It should be noted, however, that AZ-97 had equivalent PUE with the other genotypes despite its greater shoot P content at 80P, and it retained relatively high PUE at 200P (tended to be greater than AZ-78 and greater than X265 at $P = 0.09$), when shoot P content was not significantly different among the genotypes (Figure 3(a), Table 2). Therefore, AZ-97 tended to have higher PUE than the other varieties when PUE was compared at similar shoot P contents, approximately 20–40 mg plant^{-1}, as estimated from the PUE reduction curves plotted against shoot P content (Figure 3(b)).

Genotype variations in morphological responses to different phosphorus deficiency status

Projected green leaf area (PLA) and PLA per unit of shoot P content

PLA was closely correlated with shoot biomass from the early growth days at 19 DAT irrespective of genotype

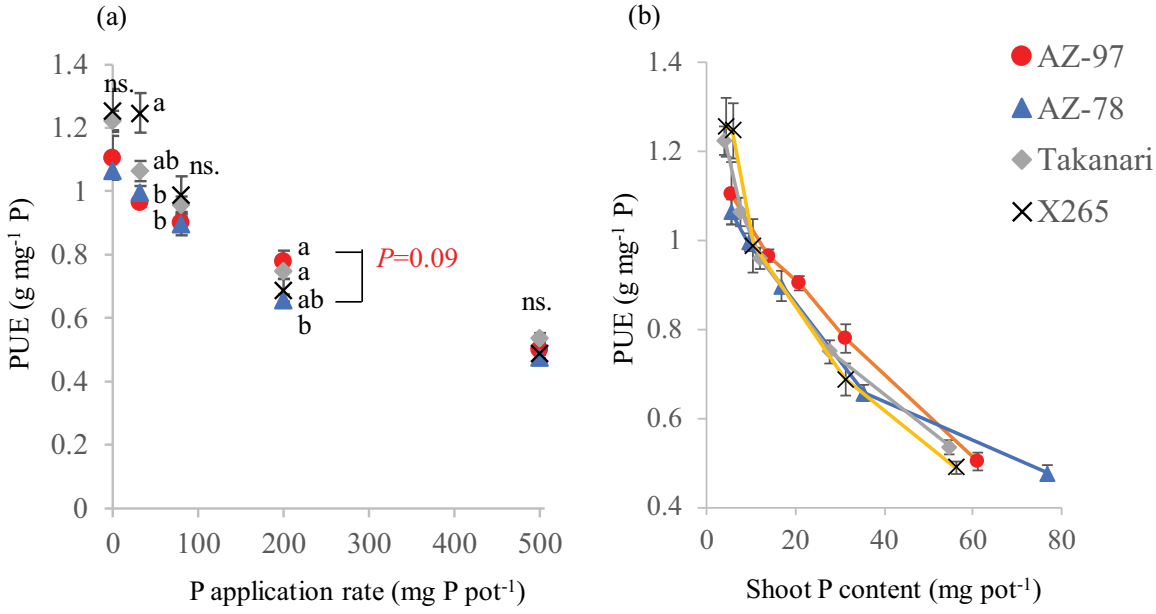


Figure 3. A comparison of PUE changes among genotypes against (a) P application rate (b) shoot P content.

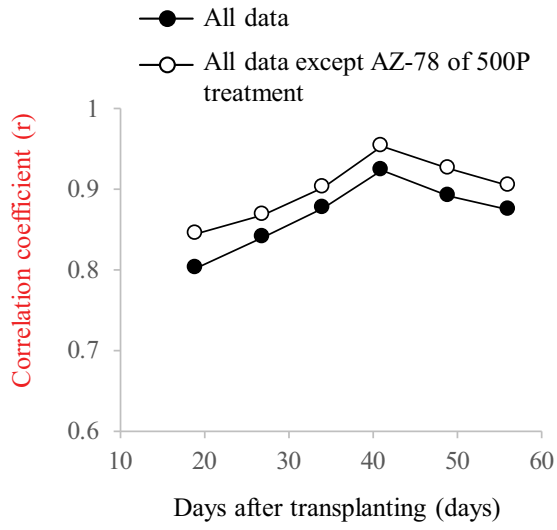


Figure 4. Changes in coefficient of determinations between PLA (projected green leaf area) and shoot biomass at different growing days after transplanting.

(Figure 4). AZ-78 at 500P tended to have relatively low PLA despite its high biomass production rate (potentially because AZ-78 had erect leaves as illustrated in S1 Fig.). When AZ-78 data at 500P were excluded, the correlation coefficient (r) in the simple linear regression between the PLA and shoot biomass ranged from 0.85 to 0.95 at different DATs, with the peak value at 41 DAT.

AZ-97 had significantly greater PLA, by 57–169% at 32P, and 46–102% at 80P, than the other three genotypes at 41 DAT (Figure 5). AZ-97 retained slightly greater PLA than X265 and AZ-78, by 8–11%, at 200P. In addition, AZ-97 had significantly greater PLA per unit of shoot P content at 13.3 cm^2 mg^{-1} than Takanari (10.0 cm^2 mg^{-1}) and AZ-97 (10.8 cm^2 mg^{-1}), and than X265 (11.9 cm^2 mg^{-1}) at the significance level of $P = 0.10$ when compared at equivalent shoot P contents at 200P. The greater PLA of AZ-97 under low P application rates from 32P and 200P were consistently observed from the early stages of plant growth (S2 Fig). Genotype differences were less apparent at 0P and 500P, excluding in the case of Takanari, which had consistently lower PLA than the other three genotypes.

Tiller number and leaf elongation rate

AZ-97 consistently produced higher tiller numbers than the other three genotypes under all the P application rates (Figure 6(a)). The maximum number of tillers in AZ-97 was 32.5 at 500P, which was 46–60% higher than the other genotypes. The number of tillers decreased gradually in all genotypes under lower P application rates while this reduction rate was relatively small for AZ-97. The proportion in the number of tillers relative to 500P for AZ-97 vs. other genotypes were 75% vs. 64–72% at 200P, 61% vs. 35–46% at 80P, and 46% vs. 27–34% at 32P (Figure 6(b)). Consequently, the differences in the number of tillers between AZ-97 and the other three genotypes were particularly large at 200P (24.2 vs. 14.2–16.2), at 80P (19.7 vs. 8.5–9.5), and at 32P (15.0 vs. 6.0–7.0). Higher tiller numbers in AZ-97, particularly from 32P to 200P, were observed consistently from the early plant growth stages (data not shown) Tiller number was highly restricted even in AZ-97 at 0P, and the differences among genotypes became less significant.

AZ-97 also demonstrated significantly more rapid leaf elongation rate or shorter phyllochrons than in the other three genotypes under all the P application rates (Figure 7

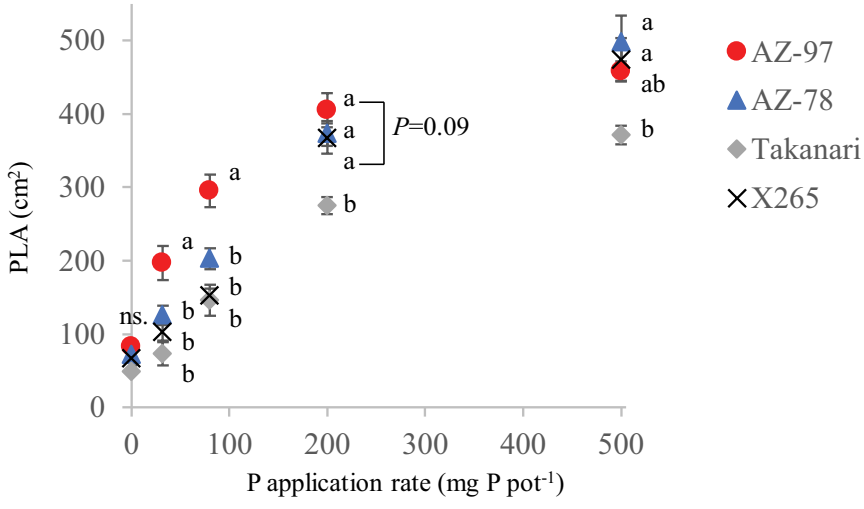


Figure 5. Genotype comparison in PLA (projected green leaf area) at 41 DAT at difference levels of P fertilizer applied.

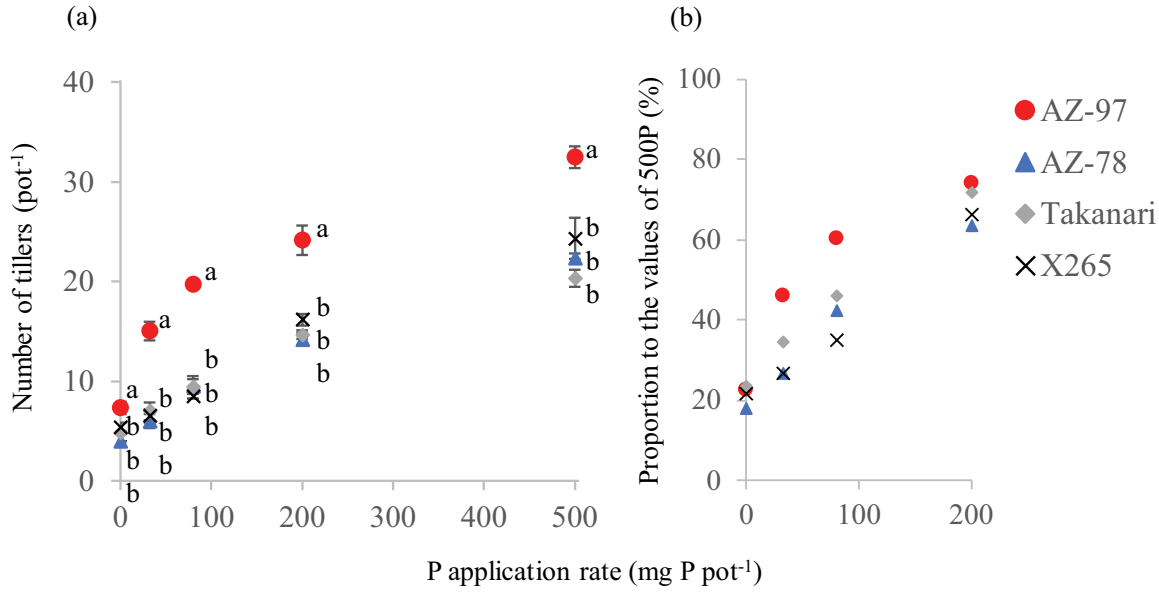


Figure 6. Genotype comparison in (a) the number of tillers at 41 DATP at difference levels of P fertilizer applied and (b) proportion in the number of tilers to the values of 500P treatment.

(a)). The superior leaf elongation rate of AZ-97 was consistently observed from the early stages of plant growth (data not shown). More importantly, change in LN was low in AZ-97 relative to the other genotypes, while leaf elongation rate was gradually retarded with a decrease in P application rates in all the genotypes. The reductions in LN when compared with the case in the 500P treatment were 0.06 leaves for AZ-97 vs. 0.13–0.55 leaves for the other three genotypes at 200P, 0.37 vs. 0.93–1.00 leaves at 80P, 0.87 vs. 1.27–1.44 leaves at 32P, and 1.37 vs. 1.67–-1.93 leaves at 0P (Figure 7(b)). Therefore, the differences in LN between AZ-97 and the other genotypes were apparently large under P deficient conditions.

Figure 8 illustrates the relationship between LN and number of tillers for each genotype based on different P application rates, demonstrating that low P supply suppressed the number of tillers through (1) reduced potential tiller numbers as a result of slower leaf elongation rates (Equation 1) and (2) reduced number of developed tillers in comparison to the potential tiller number. For instance, the tiller number decreased from 24.3 at 500P to 5.3 at 0P, in the case of X265, which was related to the reduction in

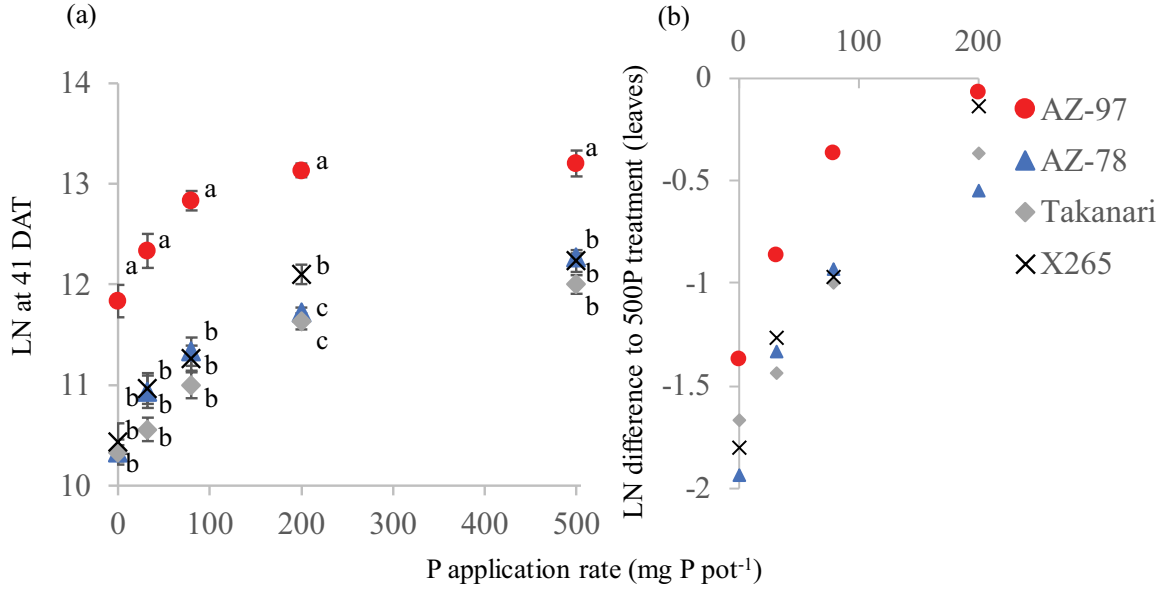


Figure 7. Genotype comparison in (a) LN (the number of leaves on the main stem) at 41 DAT at difference levels of P fertilizer applied and (b) the LN reduction against the 500P treatment.

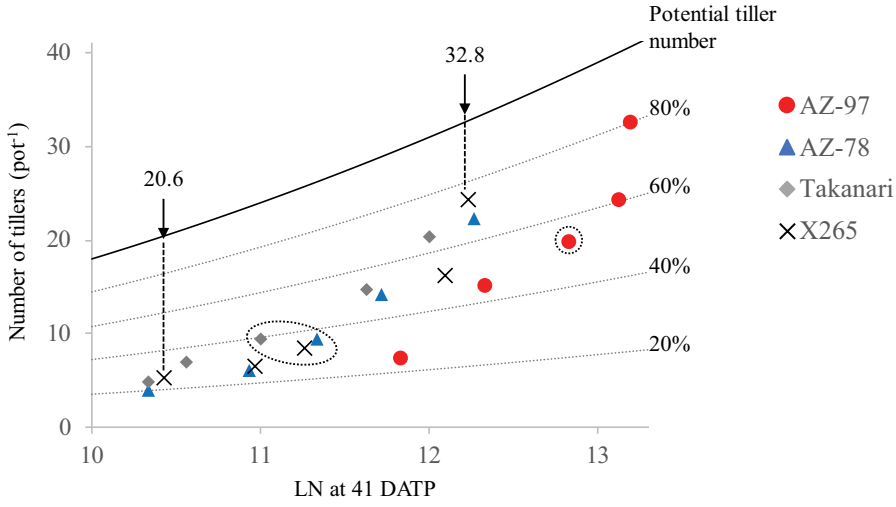


Figure 8. Relationship between the LN (number of leaves on the main stem) and the number of tillers at 41 DAT for each genotype by P application rate.

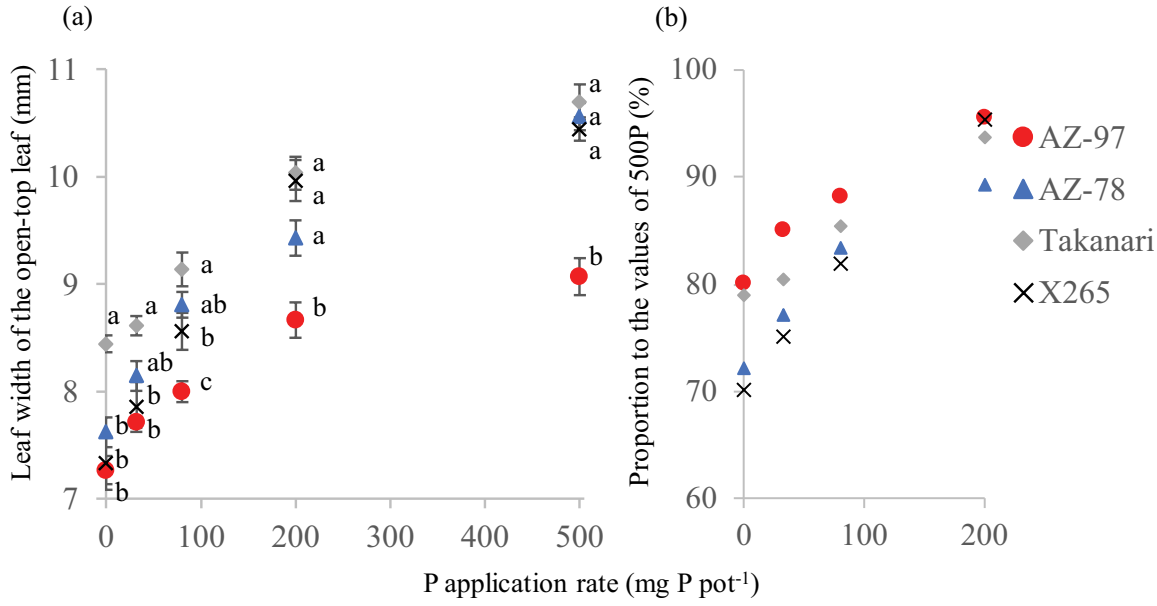


Figure 9. Genotype comparison in (a) the leaf width of the open-top leaf at the main stem at 41 DAT at difference levels of P fertilizer applied and (b) proportion in the leaf width to the values of 500P treatment.

potential tiller number from 32.8 to 20.6 due to lower LN and the concomitant reduction in the proportion of developed tillers against the potential numbers from 74% to 26%. The proportion of developed tillers against the potential tiller numbers reduced at lower P supply similarly for the other genotypes. However, this proportion remained relatively high for AZ-97 vs. the other genotypes for all the P application rates excluding 0P; 80% vs. 66–74% at 500P, 60% vs. 49–52% at 200P, 52% vs. 33–39% at 80P, and 45% vs. 25–33% at 32P. At 0P, the proportion was similarly restricted to 20% across all the genotypes while the AZ-97 still had slightly higher tiller number due to the greater LN, by 1.4–1.5 leaves, than the other three genotypes.

Leaf width

Contrary to the observed tiller number, AZ-97 had narrower leaves as measured from the open-top leaf in the main stem, while there was a significant interaction between genotype and P application rate (Table 3). The leaf width of AZ-97 was narrower than those of the other three genotypes at 500P by 1.4–1.6 mm (Figure 9(a)). The difference in leaf width became less considerable with

a decrease in P application rates. In particular, the leaf width in AZ-78 and X265 decreased sharply at the lower P application rates. The leaf widths at 0P and 32P were 70–72% and 75–77% relative to the leaf widths at 500P, respectively, for both AZ-78 and X265, whereas these values were retained at 86% and 88%, respectively, for AZ-97 (Figure 9(b)). As a result, there were no significant differences in leaf width among AZ-97, AZ-78, and X265 in the 0P and 32P treatments. Takanari had wide leaves consistently relative to the other genotypes under all the P rates.

A multiple regression model after a stepwise selection confirmed that the number of tillers ($P < 0.001$), leaf width ($P < 0.001$), and LN ($P < 0.05$) were closely and positively correlated with PLA and, in turn, with shoot biomass, irrespective of genotype or P application rate. The standardized partial regression coefficients indicated that the number of tillers, leaf width, and LN at 41 DAT explained 55%, 32%, and 13% of the variations in PLA, respectively, with the determinant coefficient (R^2 after being df adjusted) of the regression model at 0.91.

Discussion

Superior growth of AZ-97 on severe to moderate P-deficiency stresses

AZ-97 had superior shoot biomass production relative to the common high-yielding *indica* cultivars, i.e., Takanari and X265, under severely to moderately P-stressed conditions (32P, 80P, 200P). The varieties had equivalent productivity when P was fully supplied (500P) (Figure 1), implying that the genotype variations in treatments with lower P supply rates originated from their different responses to P deficiency stress.

With regard to P uptake and PUE, the superior shoot biomass production in AZ-97 was attributed primarily to greater shoot P contents in response to the low amounts of P applied (Figure 2) and secondly to the maintenance of relatively high PUE even when the shoot P contents increased (Figure 3). Such AZ-97 traits with both high P acquisition capacity and high efficiency in converting acquired P into shoot biomass are highly relevant to rice breeding under P deficiency stress (Wang et al., 2010). Rose et al. (2015) highlighted the importance of comparing PUE among genotypes at equal shoot P contents to avoid the risk of screening false genotypes (or related QTLs) whose PUE are apparently high but only because the genotypes have low shoot P contents – see X265 at 0P and 32P in Figure 3(a). Similarly, excess allocation of assimilates to root systems for P acquisition can be a cost and a trade-off against efficient shoot biomass production (Wissuwa et al., 2009). AZ-97 may have prospective traits to maintain a balance between external P acquisition and shoot biomass production following exposure to P deficiency stress. Notably, there were significant genotypic differences neither in shoot P contents nor in shoot biomass at 0P, which could be because the genetic traits related to P acquisition capacity or PUE are hardly exerted not only when available P is abundant but also when available P is extremely low in soils.

Key morphological traits of AZ-97 facilitating superior growth under phosphorus deficiency stress

PLA or canopy coverage are key parameters that influence the amount of radiation intercepted by a canopy, and, in turn, biomass production (Monteith, 1977). The present study confirmed a close correlation between PLA from early growth stages and subsequent shoot biomass production, irrespective of genotype or P application rate (Figure 4). The result indicated that rapid canopy development or the maintenance of high leaf area following exposure to P deficiency conferred an advantage to biomass production. In this regard, AZ-97 had a trait to retain the leaf expansion even when the plants were exposed to the same level of P deficiency status (superior PLA per unit of shoot P content at 200P). In addition, shoot morphological observations revealed that the variation in PLA can be mostly explained by the tiller number, leaf width, and LN, which is anticipated because tiller number and LN influence the number of leaves per plant, and single leaf size is largely dependent on the changes in leaf width while P deficiency does not much influence leaf length of rice (Luquet et al., 2005).

At the P-sufficient condition (500P), no genotypic differences in PLA were observed. This is probably because an advantage of AZ-97 having a large number of tillers and LN could be counteracted by its significantly narrower leaves. However, AZ-97 gradually achieved a major advantage in PLA relative to the other genotypes under lower P application rates as AZ-97 had less significant changes in all of the parameters. The trends in morphological changes across the genotypes under P deficiency – AZ-97 exhibited lower degrees of change – were already observed at 200P, at which the final shoot P contents were more or less equivalent among genotypes. The result implied that AZ-97 tended to retain shoot morphology, and in turn, canopy coverage relative to the other genotypes even when exposed to the same level of P deficiency status.

According to Umehara et al. (2010), the inhibition of the production of new tillers is a potential adaptive strategy in rice for saving the limited P resources for

existing tillers. Conversely, our results implied that 'too sensitive shoot morphological changes' may not be advantageous for biomass production, at least under moderate to severe P-deficient conditions.

Interrelationship among canopy development, phosphorus uptake and biomass production

Less significant changes in shoot morphology of AZ-97 might be an advantageous trait to eventually produce large biomass under P deficiency by retaining captured radiation, consistent supply of carbon assimilates to the root system, and external P uptake. It is generally understood there is a close interaction among PLA or leaf area expansion, root growth, and external P uptakes via photosynthetic assimilation and allocation of assimilates into the root system. Such interactions could gradually amplify genotypic differences in P uptakes and biomass production. For instance, Mollier and Pellerin (1999) reported that P deficiency first reduces leaf expansion to maintain root growth, and the morphological response subsequently inhibits root production due to smaller leaf area and inadequate carbohydrate supply. This study indicates that excess allocation of assimilates to roots at a cost of leaf expansion may eventually cause a negative effect on biomass production under P deficiency. Wissuwa (2005) demonstrated that P deficiency-tolerant genotypes in rice maintained root to shoot ratios even under P stress conditions, while intolerant genotypes disproportionately increased root to shoot ratio. Less sensitive shoot morphological responses to P deficiency as observed in AZ-97 could be partly linked to such genotypic variations in P-deficiency tolerance.

In addition, Luquet et al. (2005) noted that rapid canopy development should be advantageous in fields by facilitating the acquisition of limited nutrient resources in the presence of weed competitors. Saito et al. (2015) also pointed out that rapid initial growth is an important trait for rice breeding in P-deficient and low-input production systems in SSA. Such aspects could also be related to the results of our preliminary field trial in which AZ-97 had the highest yield among 300 accessions under low-input and poor-nutrient soils in Madagascar. In summary, since AZ-97 exhibited superior shoot biomass production, it is a potential donor for traits such as high P acquisition capacity and high leaf expansion capacity from early growth stages based on less significant morphological changes in the number of tillers, leaf width, and leaf elongation rate, even under limited P fertilizer input amounts under severely P-deficient soils.

Acknowledgments

The authors are grateful to Dr Tomohiro Nishigaki and Ms Mayumi Yonemura for chemical analysis of soils and plants, and Ms Mihoko Enami for imagery analysis. The authors thank Dr Matthias Wissuwa, Japan International Research Center for Agricultural Sciences, for providing us with valuable comments and suggestions to improve the draft of this manuscript.

Disclosure statement

The authors declare no conflict of interest in this paper.

Funding

This research was funded by the JIRCAS research program 'Development of sustainable agricultural productivity and improve food security in Africa' and by the Science and Technology Research Partnership for Sustainable Development (SATREPS), Japan Science and Technology Agency (JST)/Japan International Cooperation Agency (JICA) (Grant No. JPMJSA1608).

ORCID

Yasuhiro Tsujimoto http://orcid.org/0000-0001-7738-9913
Juan Pariasca Tanaka http://orcid.org/0000-0002-9488-4729
Toshiyuki Takai http://orcid.org/0000-0002-6498-610X

References

Campos, P., Borie, F., Cornejo, P., Lopez-Raez, J. A., Lopez-Garcia, A., & Seguel, A. (2018). Phosphorus acquisition efficiency related to root traits: is mycorrhizal symbiosis a key factor to wheat and barley cropping? *Frontiers in Plant Science*, *9*, 752. https://doi.org/10.3389/fpls.2018.00752

Cordell, D., Drangert, J.-O., & White, S. (2009). The story of phosphorus: Global food security and food for thought. *Global Environmental Change*, *19*(2), 292–305. https://doi.org/10.1016/j.gloenvcha.2008.10.009

Dobermann, A., & Fairhurst, T. H. (2000). *Rice: Nutrient disorders & nutrient management*. Potash & Phosphate Institute (PPI), Potash & Phosphate Institute of Canada (PPIC) and International Rice Research Institute (IRRI).

Gamuyao, R., Chin, J. H., Pariasca-Tanaka, J., Pesaresi, P., Catausan, S., Dalid, C., Slamet-Loedin, I., Tecson-Mendoza, E. M., Wissuwa, M., & Heuer, S. (2012). The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. *Nature*, *488*, 535–539. https://doi.org/10.1038/nature11346

Katayama, T. 1951. *A study on tillering systems in rice and wheat*. Yokendo Ltd. (in Japanese).

Kawamura, K., Tsujimoto, Y., Nishigaki, T., Andriamananjara, A., Rabenarivo, M., Asai, H., Rakotoson, T., & Razafimbelo, T. (2019). Laboratory visible and near-infrared spectroscopy with genetic algorithm-based partial least squares regression for assessing the soil phosphorus content of upland and lowland rice fields in madagascar. *Remote Sensing*, *11*(5), 506. https://doi.org/10.3390/rs11050506

Luquet, D., Zhang, B. G., Dingkuhn, M., Dexet, A., & Clément-Vidal, A. (2005). Phenotypic plasticity of rice seedlings: Case of phosphorus deficiency. *Plant Production Science*, *8*(2), 145–151. https://doi.org/10.1626/pps.8.145

Lynch, J. P., & Brown, K. M. (2001). Topsoil foraging – An architectural adaptation of plants to low phosphorus availability. *Plant and Soil*, *237*, 225–237. https://doi.org/10.1023/A:1013324727040

Mansueto, L., Fuentes, R. R., Borja, F. N., Detras, J., Abriol-Santos, J. M., Chebotarov, D., Sanciangco, M., Palis, K., Copetti, D., Poliakov, A., Dubchak, I., Solovyev, V., Wing, R. A., Hamilton, R. S., Mauleon, R., McNally, K. L., & Alexandrov, N. (2017). Rice SNP-seek database update: New SNPs, indels, and queries. *Nucleic Acids Research*, *45*(D1), 1075–1081. https://doi.org/10.1093/nar/gkw1135

Mollier, A., & Pellerin, S. (1999). Maize root system growth and development as influenced by phosphorus deficiency. *Journal of Experimental Botany*, *50*(333), 487–497. https://doi.org/10.1093/jxb/50.333.487

Monteith, J. L. (1977). Climate and the efficiency of crop production in Britain. Philosophical Transactions of the Royal Society of London. *B, Biological Sciences*, *281*, 277–294. https://doi.org/10.1098/rstb.1977.0140

Mori, A., Fukuda, T., Vejchasarn, P., Nestler, J., Pariasca-Tanaka, J., & Wissuwa, M. (2016). The role of root size versus root efficiency in phosphorus acquisition in rice. *Journal of Experimental Botany*, *67*(4), 1179–1189. https://doi.org/10.1093/jxb/erv557

Murphy, J., & Riley, J. P. (1962). A modified single method for the determination of phosphates in natural waters. *Analytica Chimica Acta*, *27*, 31–36. https://doi.org/10.1016/S0003-2670(00)88444-5

Nziguheba, G., Zingore, S., Kihara, J., Merckx, R., Njoroge, S., Otinga, A., Vandamme, E., & Vanlauwe, B. (2015). Phosphorus in smallholder farming systems of sub-Saharan Africa: Implications for agricultural intensification. *Nutrient Cycling in Agroecosystems*, *104*, 321–340. https://doi.org/10.1007/s10705-015-9729-y

Obara, M., Tamura, W., Ebitani, T., Yano, M., Sato, T., & Yamaya, T. (2010). Fine-mapping of qRL6.1, a major QTL for root length of rice seedlings grown under a wide range of NH4(+) concentrations in hydroponic conditions. *Theoretical and Applied Genetics*, *121*(3), 535–547. https://doi.org/10.1007/s00122-010-1328-3

Rabeharisoa, L., Razanakoto, O. R., Razafimanantsoa, M. P., Rakotoson, T., Amery, F., & Smolders, E. (2012). Larger bioavailability of soil phosphorus for irrigated rice compared with rainfed rice in Madagascar: Results from a soil and plant survey. *Soil Use and Management*, *28*, 448–456. https://doi.org/10.1111/j.1475-2743.2012.00444.x

Rose, T. J., Mori, A., Julia, C. C., & Wissuwa, M. (2015). Screening for internal phosphorus utilisation efficiency: Comparison of genotypes at equal shoot P content is critical. *Plant and Soil*, *401*, 79–91. https://doi.org/10.1007/s11104-015-2565-7

Saito, K., Vandamme, E., Johnson, J.-M., Tanaka, A., Senthilkumar, K., Dieng, I., Akakpo, C., Gbaguidi, F., Segda, Z., Bassoro, I., Lamare, D., Gbakatchetche, H., Abera, B. B., Jaiteh, F., Bam, R. K., Dogbe, W., Sékou, K., Rabeson, R., Kamissoko, N., & Wopereis, M. C. S. (2019). Yield-limiting macronutrients for rice in sub-Saharan Africa. *Geoderma*, *338*, 546–554. https://doi.org/10.1016/j.geoderma.2018.11.036

Saito, K., Vandamme, E., Segda, Z., Fofana, M., & Ahouanton, K. (2015). A screening protocol for vegetative-stage tolerance to phosphorus deficiency in upland rice. *Crop Science*, *55*, 1223–1229. https://doi.org/10.2135/cropsci2014.07.0521

Tsujimoto, Y., Pedro, J. A., Boina, G., Murracama, M. V., Tobita, S., Oya, T., Nakamura, S., Cuambe, C. E., & Martinho, C. (2016). An application of digital imagery analysis to understand the effect of N application on light interception, radiation use efficiency, and grain yield of maize under various agro-environments in Northern Mozambique. *Plant Production Science*, *20*(1), 12–23. https://doi.org/10.1080/1343943X.2016.1240013

Tsujimoto, Y., Rakotoson, T., Tanaka, A., & Saito, K. (2019). Challenges and opportunities for improving N use efficiency for rice production in sub-Saharan Africa. *Plant Production Science*, *22*(4), 413–427. https://doi.org/10.1080/1343943X.2019.1617638

Uga, Y., Hanzawa, E., Nagai, S., Sasaki, K., Yano, M., & Sato, T. (2012). Identification of qSOR1, a major rice QTL involved in soil-surface rooting in paddy fields. *Theoretical and Applied Genetics*, *124*(1), 75–86. https://doi.org/10.1007/s00122-011-1688-3

Uga, Y., Sugimoto, K., Ogawa, S., Rane, J., Ishitani, M., Hara, N., Kitomi, Y., Inukai, Y., Ono, K., Kanno, N., Inoue, H., Takehisa, H., Motoyama, R., Nagamura, Y., Wu, J., Matsumoto, T., Takai, T., Okuno, K., & Yano, M. (2013). Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. *Nature Genetics*, *45*(9), 1097–1102. https://doi.org/10.1038/ng.2725

Umehara, M., Hanada, A., Magome, H., Takeda-Kamiya, N., & Yamaguchi, S. (2010). Contribution of strigolactones to the inhibition of tiller bud outgrowth under phosphate deficiency in rice. *Plant and Cell Physiology*, *51*(7), 1118–1126. https://doi.org/10.1093/pcp/pcq084

Vance, C. P., Uhde-Stone, C., & Allan, D. L. (2003). Phosphorus acquisition and use: Critical adaptations by plants for securing a nonrenewable resource. *New Phytologist*, *157*, 423–447. https://doi.org/10.1046/j.1469-8137.2003.00695.x

Wang, X., Shen, J., & Liao, H. (2010). Acquisition or utilization, which is more critical for enhancing phosphorus efficiency in modern crops? *Plant Science*, *179*(4), 302–306. https://doi.org/10.1016/j.plantsci.2010.06.007

Wissuwa, M. (2005). Combining a modelling with a genetic approach in establishing associations between genetic and physiological effects in relation to phosphorus uptake. *Plant and Soil*, *269*, 57–68. https://doi.org/10.1007/s11104-004-2026-1

Wissuwa, M., & Ae, N. (2001). Genotypic variation for tolerance to phosphorus deficiency in rice and the potential for its exploitation in rice improvement. *Plant Breeding*, *120*(1), 43–48. https://doi.org/10.1046/j.1439-0523.2001.00561.x

Wissuwa, M., Gamat, G., & Ismail, A. M. (2005) Is root growth under phosphorus deficiency affected by source or sink limitations? *Journal of Experimental Botany*, *56*(417), 1943–1950. https://doi.org/10.1093/jxb/eri189

Wissuwa, M., Mazzola, M., & Picard, C. (2009). Novel approaches in plant breeding for rhizosphere-related traits. Plant and Soil, 321, 409–430. https://doi.org/10.1007/s11104-008-9693-2

Received: 31 March 2020 | Accepted: 17 September 2020 | Published online: 9 December 2020

DOI: 10.1002/csc2.20344

Crop Science

ORIGINAL RESEARCH ARTICLE

Crop Breeding & Genetics

Effects of quantitative trait locus *MP3* on the number of panicles and rice productivity in nutrient-poor soils of Madagascar

Toshiyuki Takai[1] | **Mitsukazu Sakata**[2] | **Njato M. Rakotoarisoa**[3] | **Ny Toky Razafinarivo**[3] | **Tomohiro Nishigaki**[1] | **Hidetoshi Asai**[1] | **Takuma Ishizaki**[4] | **Yasuhiro Tsujimoto**[1]

[1] Japan International Research Center for Agricultural Sciences, Tsukuba, Ibaraki 3058686, Japan

[2] Graduate School of Integrated Arts and Sciences, Kochi Univ., Nankoku, Kochi 7838502, Japan

[3] Rice Research Dep., National Center of Applied Research on Rural Development, Tsimbazaza, Antananarivo BP1690, Madagascar

[4] Tropical Agriculture Research Front, Japan International Research Center for Agricultural Sciences, Ishigaki, Okinawa 9070002, Japan

Correspondence
Toshiyuki Takai, Japan International Research Center for Agricultural Sciences, Tsukuba, Ibaraki, 3058686, Japan.
Email: toshi336@affrc.go.jp

Assigned to Associate Editor Esten Mason.

Funding information
Science and Technology Research Partnership for Sustainable Development, Grant/Award Number: JPMJSA1608; Japan Science and Technology Agency (JST)/Japan International Cooperation Agency (JICA)

Abstract

The majority of paddy fields in sub-Saharan Africa (SSA) are characterized by nutrient-poor soils. In such fields, tillering in rice (*Oryza sativa* L.) plants is severely restricted, which results in a reduced number of panicles and thus a decrease in grain yield. In this study, we evaluated the effects of a quantitative trait locus (QTL), associated with number of panicles and referred as *MP3*, on rice growth and grain yield in nutrient-poor soils in Madagascar. We used a high-yielding rice cultivar, Takanari, and its near-isogenic line bearing the *MP3* allele (NIL-*MP3*). A pot experiment with various P application rates demonstrated vigorous tillering in NIL-*MP3* compared to Takanari from the early vegetative stage even under low P levels. This led to enlarged leaf area and thus increased biomass. We then conducted multiple field trials with a total of 12 experimental conditions using the two varieties. The experiments led to a range of grain yield from 1.3 to 4.1 t ha^{-1} and a range in number of panicles from 107 to 270 m^{-2}. The results revealed that NIL-*MP3* produced a greater number of panicles and spikelets m^{-2} (19 and 12%, respectively) than Takanari across all 12 experiments. Grain yield increased in NIL-*MP3* under some experimental conditions. This study is the first of its kind to demonstrate that *MP3* increased number of panicles and spikelets and grain yield in the nutrient-poor and low-yielding soils of SSA. Thus, we conclude that *MP3* could become a prominent genetic resource for the improvement of rice yields in SSA.

1 | INTRODUCTION

Rice (*Oryza sativa* L.) is increasingly becoming an important food source in sub-Saharan Africa (SSA). Consumption is steadily increasing due to population growth and a shift in consumer preference for rice (Balasubramanian, Sie, Hijmans, & Otsuka, 2007). This situation requires a

Abbreviations: CSSL, chromosome segment substitution line; DAT, days after transplanting; FOFIFA, the National Center for Applied Research on Rural Development; IRRI, International Rice Research Institute; *MP3*, *MORE PANICLES 3*; NIL, near-isogenic line; NUE, nitrogen use efficiency; PUE, phosphorus use efficiency; QTL, quantitative trait locus; SSA, sub-Saharan Africa.

further increase in rice production through both arable land expansion and grain yield enhancement in SSA (AfricaRice, 2017).

It is generally recognized that poor soil fertility and nutrient availability are major limiting factors in the production of rice, as well as other crops, in SSA (Tittonell & Giller, 2013). Among the three major nutrients (N, P, and K) which are essential for plant growth and development (Lines-Kelly, 1994), N is the least abundant in the soils of SSA, followed by P (Saito et al., 2019). Although P is present in the soil, it is predominantly fixed by active Al and Fe and is not readily available for absorption by plants (Nishigaki et al., 2018, 2019). Rice plants deficient in N display stunted growth, small and yellowish-green leaves, reduced tillering, and reduced grain numbers (Dobermann & Fairhurst, 2000). Similarly, P deficiency leads to stunted rice plants with greatly reduced tillering, narrow and erect leaves, and retarded development (Dobermann & Fairhurst, 2000; Ye et al., 2019). Under deficiencies of N, P, or both, the simple solution is to apply appropriate amounts of fertilizers to enhance nutrient-poor soils. However, the majority of local SSA farmers lack the finances to purchase sufficient fertilizer (Vanlauwe et al., 2014). Therefore, it is necessary to develop effective fertilizer application techniques with the least amount of fertilizer (Tsujimoto, Rakotoson, Tanaka, & Saito, 2019), and to genetically improve rice varieties in favor of high nutrient-use efficiencies (Ismail, Heuer, Thomson, & Wissuwa, 2007).

Fertilizer management practices that entail localized forms of application or dipping seedling roots into a P-soil slurry are promising, cost-effective techniques to improve fertilizer profitability (De Datta, Biswas, & Charoenchamratcheep, 1990; Ros, White, & Bell, 2015). Promising genetic routes to improved tolerance of N and P deficiencies are suggested by many sources, including Hu et al. (2015), who identified the nitrate-transporter gene NRT1.1B as the causal gene of the quantitative trait locus (QTL) that controls nitrate uptake. The *NRT1.1B-indica* allele enhances nitrate uptake, root–shoot transport, nitrogen-use efficiency (NUE), and final grain yield compared to rice types without that allele. Zhang et al. (2015) cloned QTL-*TOND1* as a single gene which confers tolerance to N deficiency. The *TOND1* allele increased total N uptake and biomass under N-deficient conditions. The *PISTOL1* gene, which was identified as the causal P-starvation-tolerance gene on the *PUP1* locus, is an enhancer of early crown root development and root growth, which enables plants to take up more P and other nutrients under nutrient-deficient conditions (Gamuyao et al., 2012). Recently discovered is the *SPDT* gene, which controls the allocation of P to grains (Yamaji et al., 2017). This gene may be useful in increasing phosphorus-use efficiency (PUE), because the knockdown of *SPDT* could reduce the amount of P in grains, and therefore the amount which is removed from the soil. Although these techniques are novel and promising, more genetic resources and new techniques need to be identified in order to cost-effectively enhance rice production under conditions of poor soil fertility (McCouch et al., 2013). It should be noted that none of these genetic resources have been tested in nutrient-poor fields.

Core Ideas

- NIL-*MP3* promoted tillering from the early vegetative stage even under low P levels.
- NIL-*MP3* produced a greater number of panicles and spikelets m^{-2} than Takanari in field tests.
- NIL-*MP3* increased grain yield under some experimental condtions.

While several studies have focused on nutrient uptake and transport and root growth under nutrient deficiency (Wissuwa, Kretzschmar, & Rose, 2016), only a few studies have assessed adaptive aboveground morphological variations to overcome nutrient deficiency (Luquet, Zhang, Dingkuhn, Dexet, & Clément-Vidal, 2005; Tsujimoto, Sakata, Raharinivo, Tanaka, & Takai, 2020). The variations can directly influence rice productivity by changing the leaf area expansion or radiation capture ability. As mentioned above, nutrient deficiency, especially that of P, severely restricts tillering. Because the number of panicles, which is one of the yield components, is determined by the number of effective tillers, vigorous tillering may be an important trait in poor-fertility soils. Recently, Takai et al. (2014) detected a QTL associated with the number of panicles using chromosome segment substitution lines (CSSLs) under relatively fertile lowland conditions in Japan. One of the CSSLs that carries the QTL produced 19% more panicles without a decrease in number of spikelets per panicle, and generated a brown rice yield of 7 t ha^{-1}. However, the effect of this QTL on tillering has not been elucidated under nutrient-poor and low-yielding environments such as in SSA. Here, we examined whether the QTL contributes to increased rice productivity through enhancement of tillering under nutrient-poor soil conditions in SSA.

The objectives of this study were (a) to elucidate the effect of the QTL on tillering and biomass at the vegetative stage under various concentrations of P in a pot experiment, (b) to evaluate the effect of the QTL on yield and yield components in the irrigated lowland fields in the central highland of Madagascar where nutrient-poor soils predominate (Nishigaki et al., 2019; Tsujimoto et al., 2019), and (c) to determine under which conditions the QTL can be most useful in enhancing rice productivity.

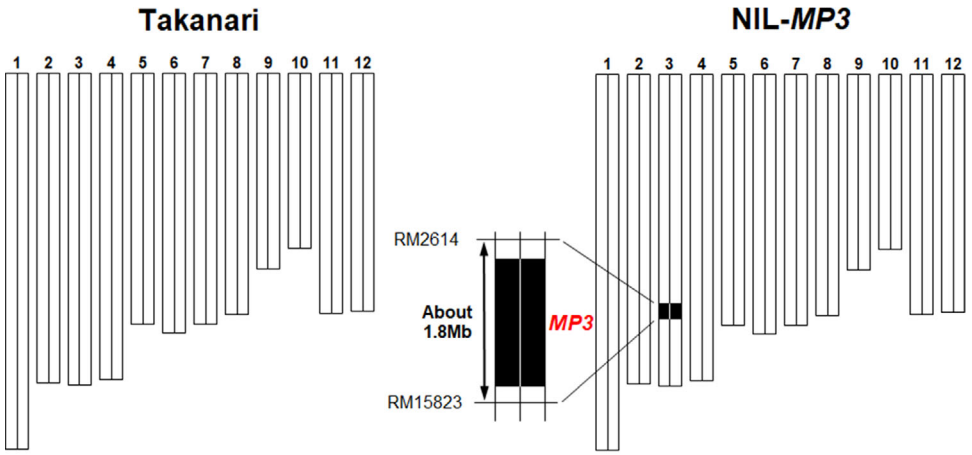


FIGURE 1 Graphical genotype of Takanari and NIL-*MP3*. The bars show chromosomes. Chromosome numbers are indicated above each bar. The white bars denote regions homozygous for Takanari and the black segment denotes a region homozygous for Koshihikari. *MP3* is located within a 1.8-Mb genomic region between DNA markers RM2614 and RM15823 on chromosome 3

2 | MATERIALS AND METHODS

2.1 | Plant materials

A previous study developed reciprocal CSSLs between Takanari and Koshihikari through four generations of backcrossing and marker-assisted selection (Takai et al., 2014). Takanari is a high-yielding *indica* cultivar (Imbe et al., 2004) and Koshihikari is a leading *japonica* cultivar with good eating quality, but it is not high yielding (Kobayashi, Hori, Yamamoto, & Yano, 2018); both were developed in Japan. One of the CSSLs, SL1310, which carries the Koshihikari genomic segment on chromosome 3 in the Takanari genetic background, contained the QTL that increased the number of panicles (Takai et al., 2014). Therefore, we named the QTL *MP3* (*MORE PANICLES 3*). To develop a near-isogenic line carrying the *MP3* (NIL-*MP3*), SL1310 was backcrossed with Takanari. Subsequently, the F_2–F_4 progenies containing as the small Koshihikari chromosome segment with *MP3* as possible were selected using DNA markers. Consequently, we developed NIL-*MP3* that carries a 1.8-Mb Koshihikari genomic segment between RM2614 and RM15823, bearing *MP3* on chromosome 3 (Figure 1). Takanari and NIL-*MP3* were used in the pot and field experiments.

2.2 | Pot experiments

A pot experiment was conducted in a screenhouse at the National Center for Applied Research on Rural Development (FOFIFA) at Antananarivo in Madagascar (18°52′ S, 47°33′ E, 1310 m altitude; Figure 1) from November 2018 to January 2019. We sampled extremely P-deficient soil with no history of fertilization from an upland field in Antohobe

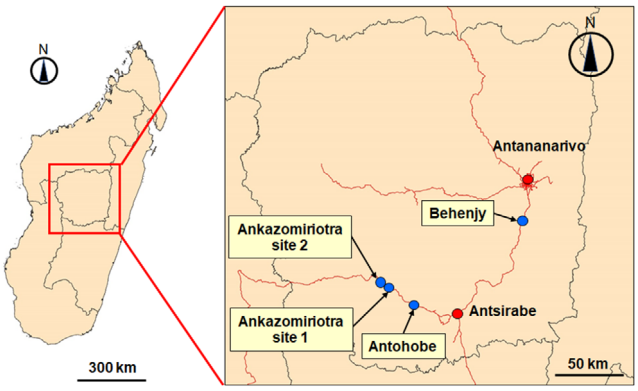


FIGURE 2 Maps of field experiment sites conducted in the central highland of Madagascar

(19°46′ S, 46°41′ E, 1260 m altitude; Figure 2). Major characteristics of the soil are summarized in Table 1. After being air-dried, 4 kg of soil was placed into a 5-L plastic pot. Five P treatments were established, with $NaH_2PO_4 \cdot 2H_2O$ applied at different concentrations: 0, 8, 20, 50, and 125 mg P per kg soil (0P, 8P, 20P, 50P, and 125P, hereafter). All pots were supplemented with N and K (NH_4NO_3 and KCl, respectively) each at a concentration of 152 mg kg^{-1} soil, in order to avoid any potential effects of N or K deficiencies. All nutrients were uniformly mixed with soil, and the soil was flooded 1 d before transplanting. One 20-d-old seedling of either Takanari or NIL-*MP3*, grown in P-free sand, was transplanted to each pot. We employed a randomized block design with five levels of P treatments, two varieties, and six replicates; the plants were grown under continuously flooded conditions.

The number of tillers in each pot was counted weekly from 21 to 56 days after transplanting (DAT). At 56 DAT, leaf age and the number of leaves were measured and the leaves were

TABLE 1 Soil properties of each study site in Madagascar

Experimental site	pH[a]	Total C[b]	Total N[b]	Oxalate P[c]
		g C kg^{-1}	g N kg^{-1}	mg P kg^{-1}
Pot experiments at Antananarivo	4.9	22.8	2.1	40.7
Antohobe	5.2	13.2	1.1	44.7
Behenjy	5.1	15.4	1.3	116.3
Ankazomiriotra site 1	5.2	17.5	1.6	48.3
Ankazomiriotra site 2	5.9	15.7	1.6	45.5

[a]Soil pH was measured by a 1:2.5 ratio of soil/water.
[b]Total C and N were measured by the combustion method.
[c]Oxalate P was extracted following the method of Courchesne and Turmel (2008).

sampled for the picture-taking with a digital camera (SX610 HS, Canon). The total leaf area was measured for each pot using image processing software (Image J, NIH). On the same day, aboveground plant parts were harvested at the soil surface and combined with the leaves to determine aboveground biomass after drying at 70 ˚C for 3 d in an oven.

2.3 | Field experiments

A total of six field experiments were conducted in farmers' paddy fields in the central highland of Madagascar during the 2017–2018 (hereafter, 2018) and 2018–2019 (hereafter, 2019) growing seasons, respectively. In 2018, Antohobe (19˚46′ S, 46˚41′ E, 1250 m altitude) and Behenjy (19˚10′ S, 47˚29′ E, 1370 m altitude) were chosen as experimental sites (Figure 2). In 2019, Ankazomiriotra Site 1 (19˚41′ S, 46˚35′ E; 1,100 m altitude) and Site 2 (19˚40′ S, 46˚34′ E; 1,110 m altitude) were used as well as Antohobe and Behenjy (Figure 2). In all of the experimental fields, farmers had continuously cultivated rice once a year without crop rotation and with no mineral fertilizer inputs for at least 10 yr prior to the start of the experiments. The major soil characteristics of the four experimental fields are summarized in Table 1. Levels of oxalate P were below 200 mg P kg^{-1} in the four sites, indicating the highly P-deficient character of the soils in the central highland of Madagascar (Nishigaki et al., 2019). Takanari and NIL-*MP3* were used in the field experiments. A local rice cultivar, X265, was also grown as a reference. Seeds were sown in nursery beds constructed at the paddy fields, and 20- to 30-d-old seedlings were transplanted into the fields, with two seedlings per hill. The planting density was 25 hills m^{-2}, with 20 cm between hills and between rows. The experimental plots (8–12 m^2) were laid out in a split-plot arrangement, with the main plot consisting of fertilizer treatments (with and without fertilizer) and rice types in the subplots, with four replications. Basal fertilizer was applied to the fertilizer application plot at transplanting at the rates of 20 kg N ha^{-1}, 17.5 kg P ha^{-1}, and 24.1 kg K ha^{-1} as a N–P–K compound fertilizer (11–22–16). Plots were top-dressed with urea (20 kg N ha^{-1}) at 30 and 60 DAT.

The number of tillers in each plot was counted weekly until heading in 2019. Heading was defined as the date when half of the panicles in each plot had emerged. Maturity was defined as the date when 95% of the spikelets had turned from green to yellow. Days to heading and days to maturity were defined as the number of days from sowing to heading and maturity, respectively.

At maturity, the number of panicles was counted for the hills covering 2.8–4.0 m^2 (70–100 hills) in each plot. The hills were then harvested to investigate yield and yield components. After the panicles were threshed, the whole-grain weight was measured, and the moisture content was measured with a grain moisture tester (Riceter f512, Kett). Approximately three sets of 40-g grains were selected as subsamples and counted with an electronic seed counter (WAVER IC-VA, Aidex Co. Ltd.). The number of spikelets m^{-2} was calculated by multiplying the grain number per unit weight in subsamples by the total grain weight m^{-2}. Spikelet number per panicle was calculated as the number of spikelets m^{-2} divided by the number of panicles m^{-2}. The subsamples of grain were submerged in tap water; grains that sank to the bottom were considered filled. The filled grains were then oven-dried at 37 ˚C to constant weight, counted, and weighed. Finally, the moisture content of the filled grains was measured. The filled spikelet percentage was calculated as the number of filled spikelets divided by the number of the whole spikelets in the subsamples. Single-grain weight was calculated by dividing the filled spikelet weight by the number of filled spikelets. Grain yield was determined by multiplying each yield component. Grain yield and single-grain weight were adjusted to 14% moisture content. To determine aboveground biomass, 6–10 hills were sampled at the soil surface from each plot, dried at 70 ˚C for 72 h, and weighed.

Weather data such as solar radiation and air temperature were recorded by installing Watchdog 1525 micro station (Spectrum Technologies Inc.) at each experimental site.

2.4 | Statistics

Statistical analysis was performed using a general linear model in SPSS 23.0 software (IBM). Plant variety, P application rate, and their interaction were considered fixed effects, and replication was considered a random effect in the pot experiment. Two-way ANOVA was conducted to test the effect of variety, P application rate, and their interaction on measured variables. In field experiments, two-way ANOVA was conducted to test the effects of variety, environment, and their interaction on yield, its

TABLE 2 Two-way ANOVA for plant growth-related traits at 56 d after transplanting in Takanari and NIL-*MP3* grown in the pots with various P application rates

P application rate	Variety	No. of tillers (pot^{-1})	Leaf age	No. of leaves (pot^{-1})	Leaf area	Biomass
mg P $kg\ soil^{-1}$					$cm^2\ pot^{-1}$	$g\ pot^{-1}$
0		8.8	11.9	31.0	473	5.5
8		12.0	12.4	42.6	649	8.0
20		15.7	12.8	56.8	917	13.7
50		25.8	13.2	91.1	1,341	21.8
125		31.1	13.6	114.8	1,944	30.9
	Takanari	15.2	13.0	59.6	808	14.9
	NIL-*MP3*	22.2	12.6	74.9	1,322	17.0
ANOVA						
P application rate		***	***	***	***	***
Variety		***	***	***	***	*
Variety × P application rate		ns	ns	ns	*	ns

Note. ns, Not significant. *P* values are based on ANOVA.
*Significant at $P < .05$.
***Significant at $P < .001$.

components, and biomass across 12 conditions (6 experiments × 2 fertilizer treatments). Variety, environment, and their interaction were considered as fixed effects and replication was considered as a random effect.

3 | RESULTS

3.1 | Performance of Takanari and NIL-*MP3* in pot experiment

A two-way ANOVA showed that application rate and variety had statistically significant effects on the number of tillers at 56 DAT (Table 2). The interaction between P application rate and variety was not significant. The mean number of tillers of both Takanari and NIL-*MP3* increased from 8.8 to 31.1 per pot at 56 DAT under increasing P application rates from the 0P to 125P treatments (Table 2). From 30 DAT, the number of tillers in NIL-*MP3* plants was greater than in Takanari under all P application rates (Figure 3). By 56 DAT, the number of tillers produced by NIL-*MP3* was significantly greater (33–63%; average 46%) than Takanari under all P application rates (Table 2).

With an increase in P application rate, leaf age also increased from 11.9 to 13.6 (Table 2). However, leaf age in NIL-*MP3* at 56 DAT was significantly lower than that of Takanari. The number of leaves, leaf area, and biomass increased from 31 to 114.8 per pot; 473–1,944 cm^2 per pot; and 5.5–30.9 g per pot, respectively, under increasing P application rates. By 56 DAT, NIL-*MP3* had produced 26% more leaves, 64% greater leaf area, and 14% greater biomass than Takanari; all differences are significant.

3.2 | Climate conditions and phenology in field experiments

The solar radiation and the daily mean temperature during the rice growing period ranged from 17.2 to 21.4 MJ $m^{-2}\ d^{-1}$ and from 20.3 to 22.6 °C, respectively, among the six experiments (Table 3). Solar radiation and daily mean temperature at Ankazomiriotra Site 2 were the highest in 2019, while those at Behenjy were the lowest in 2018.

Days to heading and days to maturity were similar between Takanari and NIL-*MP3* across all 12 experimental conditions (Table 4). Days to heading and days to maturity were the lowest in the Ankazomiriotra Site 2 plot with fertilizer in 2019 (97 and 134 d, respectively), while those in the Behenjy plot without fertilizer were the highest in 2019 (133–134 and 187 d, respectively). The treatment without fertilizer delayed days to heading and days to maturity by 4–23 d, respectively, compared with the fertilizer treatment.

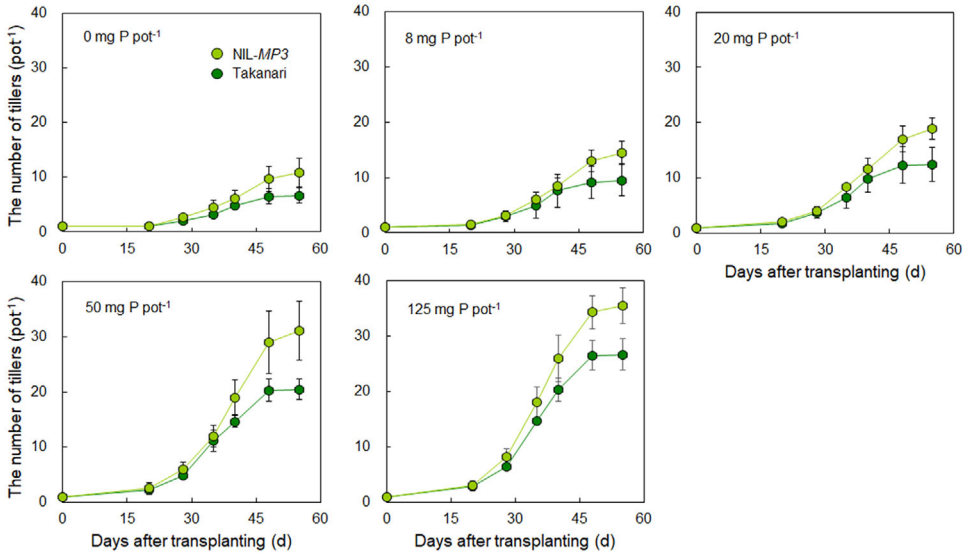


FIGURE 3 Changes in tiller number after transplanting in Takanari and NIL-*MP3* at various P application rates in a pot experiment. Error bars show the standard deviation

TABLE 3 Solar radiation and daily mean temperature of each experimental site during rice growing period

Experimental site	Year	Solar radiation	Daily mean temperature
		MJ m^{-2} d^{-1}	°C
Antohobe	2018	20.1	21.7
	2019	20.8	22.1
Behenjy	2018	16.4	20.3
	2019	18.9	20.7
Ankazomiriotra site 1	2019	17.2	22.5
Ankazomiriotra site 2	2019	21.4	22.6

3.3 | Yield, yield components, and biomass in field experiments

A two-way ANOVA showed that the environment had a statistically significant effect on grain yield (Table 5). Mean grain yield in both Takanari and NIL-*MP3* widely varied across the 12 experimental conditions, ranging from 1.3 t ha^{-1} at Antohobe without fertilizer in 2018 and 2019, to 4.1 t ha^{-1} at Behenjy and Ankazomiriotra Site 2 with fertilizer in 2019. The two-way ANOVA also revealed that interaction between variety and environment significantly affected grain yield, although the effect of variety was not significant by itself (Table 5). A comparison between Takanari and NIL-*MP3* in each of the 12 conditions revealed a higher grain yield in NIL-*MP3* at Behenjy under both fertilizer treatments in 2018 and 2019 (Table 6). However, this trend was not observed in other environments.

Each yield component was significantly affected by environment and variety (Table 5). The mean number of panicles in both Takanari and NIL-*MP3* varied from 107 to 270 m^{-2} across the 12 experimental conditions (Table 5). Overall panicle production of NIL-*MP3* (221 panicles m^{-2}) was 19% higher than that of Takanari (186 panicles m^{-2}). The mean number of spikelets m^{-2} in both Takanari and NIL-*MP3* ranged from 8,295 to 26,348 across the 12 experimental conditions. Overall spikelet production of NIL-*MP3* (19,067 spikelets m^{-2}) was 12% higher than that of Takanari (16,962 spikelets m^{-2}).

Just as for grain yield, the interaction between variety and environment had a significant effect on the percentage of filled spikelets (Table 5). A comparison between Takanari and NIL-*MP3* in each of the 12 experimental conditions showed a higher percentage of filled spikelets in NIL-*MP3* at Behenjy in 2018 and a small difference at Behenjy in 2019 (Table 6). The percentage of filled spikelets was reduced in NIL-*MP3* compared to Takanari in most of the other experimental conditions.

The effect of environment on biomass at maturity was statistically significant (Table 5). The mean biomass at maturity ranged from 2.3 t ha^{-1} at Antohobe without fertilizer in 2019 to 8.8 t ha^{-1} at Ankazomiriotra Site 1 in 2019. Neither variety nor interaction between variety and environment had significant effects on biomass at maturity.

4 | DISCUSSION

The field trials employing various nutrient regimes in this study clearly demonstrated that *MP3* can increase the number of panicles and the number of spikelets m^{-2}, even under

TABLE 4 Days to heading and days to maturity for Takanari and NIL-*MP3* grown at each experimental site

Environment			Days to heading		Days to maturity	
Experimental site	Year	Fertilizer	Takanari	NIL-*MP3*	Takanari	NIL-*MP3*
			d			
Antohobe	2018	+	106	106	147	147
		0	114	110	159	162
	2019	+	111	112	148	149
		0	133	134	187	187
Behenjy	2018	+	109	109	164	164
		0	117	117	164	164
	2019	+	117	117	165	165
		0	130	129	180	180
Ankazomiriotra site 1	2019	+	114	112	148	148
		0	114	112	149	149
Ankazomiriotra site 2	2019	+	97	97	134	134
		0	101	113	147	156

TABLE 5 Two-way ANOVA for grain yield, its components, and biomass at maturity in Takanari and near-isogenic line (NIL)-*MP3* grown at each experimental site

Environment										
Experimental site	Year	Fertilizer	Variety	Grain yield	No. of panicles (m^{-2})	No. of spikelets (Panicle^{-1})	No. of spikelets (m^{-2})	Filled spikelets	Single-grain weight	Biomass
				t ha^{-1}				%	mg	t ha^{-1}
Antohobe	2018	+		3.9	267	100	26,348	63.4	23.6	8.2
		0		1.3	108	81	8,828	63.2	22.5	4.0
	2019	+		3.4	254	91	23,177	65.1	23.0	6.8
		0		1.3	107	76	8,295	71.8	21.8	2.3
Behenjy	2018	+		3.3	212	98	20,768	65.5	24.5	7.5
		0		2.0	175	83	14,597	54.7	24.4	5.4
	2019	+		4.1	236	92	21,610	78.1	24.3	7.5
		0		3.1	206	78	15,801	82.5	24.2	5.2
Ankazomiriotra site 1	2019	+		3.2	270	81	21,836	67.8	21.9	8.8
		0		2.3	195	80	15,507	67.4	21.8	5.2
Ankazomiriotra site 2	2019	+		4.1	232	106	24,709	71.2	23.6	7.7
		0		2.2	175	78	13,610	73.0	22.7	3.9
			Takanari	2.9	186	90	16,962	70.0	24.0	6.0
			NIL-*MP3*	2.9	221	84	19,067	67.3	22.4	6.1
ANOVA										
Environment				***	***	***	***	***	***	***
Variety				ns	***	***	***	**	***	ns
Variety × Environment				*	ns	**	ns	***	ns	ns

Note. ns, not significant. *P* values are based on ANOVA.

*Significant at $P < .05$.

**Significant at $P < .01$.

***Significant at $P < .001$.

TABLE 6 Comparisons of grain yield and filled spikelets for Takanari and near-isogenic line (NIL)-*MP3* at each experimental site

Environment			Grain yield			Filled spikelets		
Experimental site	Year	Fertilizer	Takanari (a)	NIL-*MP3* (b)	Benefit of NIL-*MP3* (b) − (a)	Takanari (a)	NIL-*MP3* (b)	Benefit of NIL-*MP3* (b) − (a)
			t ha^{-1}			%		
Antohobe	2018	+	4.27	3.58	−0.69	67.1	59.8	−7.3
		0	1.38	1.15	−0.23	62.3	64.1	1.7
	2019	+	3.45	3.43	−0.02	68.1	62.0	−6.0
		0	1.50	1.16	−0.35	75.3	69.1	−6.2
Behenjy	2018	+	3.08	3.56	0.48	61.9	69.1	7.1
		0	1.56	2.35	0.79	47.2	62.2	15.0
	2019	+	3.80	4.35	0.55	79.6	76.6	−3.0
		0	3.01	3.25	0.24	84.1	80.9	−3.2
Ankazomiriotra site 1	2019	+	3.21	3.24	0.03	69.0	66.6	−2.4
		0	2.27	2.28	0.00	68.0	66.7	−1.3
Ankazomiriotra site 2	2019	+	4.16	4.00	−0.17	79.1	63.3	−15.8
		0	2.33	2.15	−0.18	79.1	66.9	−12.2

nutrient-poor conditions, to achieve yield levels ranging from 1.3 to 4.1 t ha^{-1}. This yield range is consistent with most of the average yields typically found in the lowland rice production systems in SSA (Saito et al., 2019; Tsujimoto et al., 2019). The present study is one of only a few to demonstrate the effect of QTLs and its interaction with field environments on important yield components in farmers' fields. To the best of our knowledge, this study is the first of its kind to explore breeding-based options to improve rice yields under nutrient-poor conditions in the tropics.

The Koshihikari *MP3* allele increased the number of panicles m^{-2} by 19% on average and the number of spikelets m^{-2} by 12% on average across the 12 experimental conditions (Table 5). These results are comparable to a previous report by Takai et al. (2014), which showed that one of the CSSLs (SL1310) carrying the Koshihikari *MP3* allele in the Takanari genetic background produced a greater number of panicles m^{-2} and spikelets m^{-2} (19 and 20%, respectively) than Takanari under fertile soil conditions, yielding 7 t ha^{-1}. These results suggest that *MP3* may enhance the number of panicles by approximately 20% and thus produce 10–20% more spikelets m^{-2} irrespective of soil nutrient conditions. However, it should be noted that differences in the numbers of panicles and spikelets m^{-2} between Takanari and NIL-*MP3* tend to be small under extremely low-yielding environments (at the level of 1.3 t ha^{-1}; Supplemental Figure S1). At the yield level, Takanari produced approximately 100 panicles m^{-2} which was equal to 4 panicles per hill at a planting density of 25 hills m^{-2}. Because a 20% increase for four panicles per hill would increase by only 0.8 panicle per hill, the contribution of *MP3* may be small under such extremely low-yielding environments. Nevertheless, *MP3* could be useful when applied to a large number of fields in SSA, as demonstrated by the wide range in yield we obtained in this study (1.3–4.1 t ha^{-1}), which may represent an improvement over the current average rice yield of approximately 2.1 t ha^{-1} (Tsujimoto et al., 2019). Further studies are necessary to clarify the interaction between planting density and the effect of *MP3*, as planting density typically affects the number of panicles (tillers) per hill (Nakano, Morita, Kitagawa, Wada, & Takahashi, 2012).

Although *MP3* had the effect of increasing the number of panicles and spikelets m^{-2} in this study, it did not lead to a yield increase under all 12 experimental conditions (Table 6). While the increase in grain yield was observed at Behenjy, there was no increase in grain yield at other sites. No increase in grain yield could be primarily associated with a reduction in the percentage of filled spikelets in NIL-*MP3* compared to Takanari (Table 6). Solar radiation and daily mean temperature at Behenjy were the lowest among experimental sites during the rice-growing period (Table 3). Because NIL-*MP3* was developed in a temperate region in Japan, it may not be well adapted to environments with high solar radiation and air temperature, such as in Madagascar. At present, however, we are unable to fully explain all of the factors that affect the percentage of filled spikelets in NIL-*MP3*. In general, we expect that local rice varieties bred or selected in Madagascar should exhibit better performance than Takanari and NIL-*MP3*. In this study, we grew one of the promising local cultivars, X265 (Diagne, Kinkingninhoun-Medagbe, Amovin-Assagba, Nakelse, & Toure, 2015), as a reference. Grain yield in X265

was higher than that of Takanari and NIL-*MP3* under most of 12 experimental conditions (Supplemental Figure S2). This cultivar was bred by the International Rice Research Institute (IRRI) and was selected by FOFIFA in Madagascar (Diagne et al., 2015), suggesting that X265 is an *indica* cultivar and does not carry the positive *MP3* allele. Therefore, introducing the Koshihikari *MP3* allele into the X265 genome may lead to a further increase of the number of panicles and thus grain yield. We are currently breeding X265 with the Koshihikari *MP3* allele by backcrossing and marker-assisted selection.

The results of the pot experiment are similar to those of the field experiments; the Koshihikari *MP3* allele increased the number of tillers by 46% on average across all P application rates at 56 DAT (Table 2). In addition, we observed vigorous tillering in NIL-*MP3* compared to Takanari from 30 DAT under all P application rates (Figure 3). Similar tillering patterns were observed in most of the field experiments in this study (Supplemental Figure S3). The correspondence between the pot and field experiment results corroborates that the Koshihikari *MP3* allele can promote tillering from the early vegetative stage, sustain a greater number of tillers until heading, and thus produce more panicles than the Takanari *MP3* allele, even in low-P soils. The pot experiment also demonstrates greater biomass in NIL-*MP3* than in Takanari (Table 2). This can be explained by the fact that the Koshihikari *MP3* allele led to enlarged leaf area due to a greater number of leaves. The latter trait derives from a greater number of tillers rather than faster leaf emergence rate. Large biomass production in rice was reported as the pleiotropic effect of QTLs controlling culm length, such as *SD1* and *GW6a* (Okuno et al., 2014; Song et al., 2015). To our knowledge, this is the first study using genetically sophisticated materials, such as NILs in rice, to demonstrate biomass increase via QTL-induced tillering enhancement. We are currently cloning a causal gene of *MP3*, which would clarify its molecular function and thus extend its utility in breeding programs designed to improve rice yields under nutrient-poor conditions in the tropics as well as under fertile conditions.

In summary, this study demonstrated that *MP3* enhances the number of panicles and spikelets m^{-2} by 19 and 12%, respectively, across 12 nutrient conditions. The enhanced panicle number resulted from vigorous tillering from the early vegetative stage. This also led to enlarged leaf area and thus increased biomass. The increase in grain yield in plants carrying the *MP3* allele could be achieved in environments where the percentage filled spikelets was not affected. We conclude that *MP3* can improve rice yield by increasing the number of panicles in nutrient-poor soils in SSA.

ACKNOWLEGMENTS

We thank Mr. Patrick Rakotoarivelo, Mr. Henintsoa Fenitra, Ms. Bakoly Laurencia, Ms. Anny Ranjakason, Ms. Aina Andriatsiorimanana, and local farmers for their generous research support in the field experiments. We would like to extend our gratitude to Ms. Marie Paul Razafimanantsoa and Ms. Mayumi Yonemura for conducting soil analysis. This research was financially supported by the Science and Technology Research Partnership for Sustainable Development (SATREPS), Japan Science and Technology Agency (JST)/Japan International Cooperation Agency (JICA) (Grant no. JPMJSA1608).

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ORCID

Toshiyuki Takai https://orcid.org/0000-0002-6498-610X

REFERENCES

AfricaRice. (2017). *AfricaRice Annual Report 2016: Towards rice self-sufficiency in Africa*. Côte d'Ivoire: Abidjan.

Balasubramanian, V., Sie, M., Hijmans, R. J., & Otsuka, K. (2007). Increasing rice production in sub-Saharan Africa: Challenges and opportunities. *Advances in Agronomy*, *94*, 55–133. https://doi.org/10.1016/S0065-2113(06)94002-4

Courchesne, F., & Turmel, M. C. (2008). Extractable Al, Fe, Mn, and Si. In M. R. Carter & E. G. Gregorich (Eds.), *Soil sampling and methods of analysis* (2nd ed., pp. 307–316). Boca Raton, FL: Canadian Society of Soil Science, CRC Press.

De Datta, S. K., Biswas, T. K., & Charoenchamratcheep, C. (1990). Phosphorus requirements and management for lowland rice. In S. J. Banta (Ed.), *Phosphorus requirements for sustainable agriculture in Asia and Oceania* (pp. 303–323). Los Baños, Philippines: IRRI.

Diagne, A., Kinkingninhoun-Medagbe, F. M., Amovin-Assagba, E., Nakelse, T., & Toure, A. (2015). Evaluating the key aspects of the performance of genetic improvement in priority food crops and countries in sub-Saharan Africa: The case of rice. In T. S. Walker & J. Alwang (Eds.), *Crop improvement, adoption and impact of improved varieties in food crops in sub-Saharan Africa* (pp. 183–205). Montpellier, France: CGIAR and Wallingford, UK: CABI.

Dobermann, A., & Fairhurst, T. (2000). *Rice: Nutrient disorders and nutrient management*. PPI, PPIC, and Los Baños, Philippines: IRRI.

Gamuyao, R., Chin, J. H., Pariasca-Tanaka, J., Pesaresi, P., Catausan, S., Dalid, C., … Heuer, S. (2012). The protein kinase *Pstol1* from traditional rice confers tolerance of phosphorus deficiency. *Nature* (London), *488*, 535–539. https://doi.org/10.1038/nature11346

Hu, B., Wang, W., Ou, S., Tang, J., Li, H., Che, R., … Chu, C. (2015). Variation in *NRT1.1B* contributes to nitrate-use divergence between rice subspecies. *Nature Genetics*, *47*, 834–838. https://doi.org/10.1038/ng.3337

Imbe, T., Akama, T., Nakane, A., Hata, T., Ise, K., Ando, I., … Koga, Y. (2004). Development of a multipurpose high-yielding rice variety "Takanari." (In Japanese with English abstract.) *Bulletin of the National Institute of Crop Science*, *5*, 35–51.

Ismail, A. M., Heuer, S., Thomson, M. J., & Wissuwa, M. (2007). Genetic and genomic approaches to develop rice germplasm for problem soils. *Plant Molecular Biology*, *65*, 547–570. https://doi.org/10.1007/s11103-007-9215-2

Kobayashi, A., Hori, K., Yamamoto, T., & Yano, M. (2018). Koshihikari: A premium short-grain rice cultivar - its expansion and breeding in

Japan. *Rice*, *11*, article number 15. https://doi.org/10.1186/s12284-018-0207-4

Lines-Kelly, R. (Ed.). (1994). Soil sense: Soil management for NSW north coast farmers. North coast soil management working party, New South Wales, Australia. Wollongbar, NSW: New South Wales Agriculture.

Luquet, D., Zhang, B. G., Dingkuhn, M., Dexet, A., & Clément-Vidal, A. (2005). Phenotypic plasticity of rice seedlings: Case of phosphorus deficiency. *Plant Production Science*, *8*, 145–151. https://doi.org/10.1626/pps.8.145

McCouch, S., Baute, G. J., Bradeen, J., Bramel, P., Bretting, P. K., Buckler, E., … Zamir, D. (2013). Agriculture: Feeding the future. *Nature* (London), *499*, 23–24. https://doi.org/10.1038/499023a

Nakano, H., Morita, S., Kitagawa, H., Wada, H., & Takahashi, M. (2012). Grain yield response to planting density in forage rice with a large number of spikelets. *Crop Science*, *52*, 345–350. https://doi.org/10.2135/cropsci2011.02.0071

Nishigaki, T., Sugihara, S., Kobayashi, K., Hashimoto, Y., Kilasara, M., Tanaka, H., … Funakawa, S. (2018). Fractionation of phosphorus in soils with different geological and soil physicochemical properties in southern Tanzania. *Soil Science and Plant Nutrition*, *64*, 291–299. https://doi.org/10.1080/00380768.2018.1436406

Nishigaki, T., Tsujimoto, Y., Rinasoa, S., Rakotoson, T., Andriamananjara, A., & Razafimbelo, T. (2019). Phosphorus uptake of rice plants is affected by phosphorus forms and physicochemical properties of tropical weathered soils. *Plant and Soil*, *435*, 27–38. https://doi.org/10.1007/s11104-018-3869-1

Okuno, A., Hirano, K., Asano, K., Takase, W., Masuda, R., Morinaka, Y., … Matsuoka, M. (2014). New approach to increasing rice lodging resistance and biomass yield through the use of high gibberellin producing varieties. *PLOS ONE*, *9*, e86870. https://doi.org/10.1371/journal.pone.0086870

Ros, C., White, P. F., & Bell, R. W. (2015). Nursery fertilizer application increases rice growth and yield in rainfed lowlands with or without post-transplanting crop stress. *American Journal of Plant Sciences*, *6*, 2878–2892. https://doi.org/10.4236/ajps.2015.618285

Saito, K., Vandamme, E., Johnson, J. M., Tanaka, A., Senthilkumar, K., Dieng, I., … Wopereis, C. S. (2019). Yield-limiting macronutrients for rice in sub-Saharan Africa. *Geodema*, *338*, 546–554. https://doi.org/10.1016/j.geoderma.2018.11.036

Song, X. J., Kuroha, T., Ayano, M., Furuta, T., Nagai, K., Komeda, N., … Ashikari, M. (2015). Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 76–81. https://doi.org/10.1073/pnas.1421127112

Takai, T., Ikka, T., Kondo, K., Nonoue, Y., Ono, N., Arai-Sanoh, Y., … Yamamoto, T. (2014). Genetic mechanisms underlying yield potential in the rice high-yielding cultivar Takanari, based on reciprocal chromosome segment substitution lines. *Bmc Plant Biology [Electronic Resource]*, *14*, article number 295. https://doi.org/10.1186/s12870-014-0295-2

Tittonell, P., & Giller, K. E. (2013). When yield gaps are poverty traps: The paradigm of ecological intensification in African smallholder agriculture. *Field Crops Research*, *143*, 76–90. https://doi.org/10.1016/j.fcr.2012.10.007

Tsujimoto, Y., Rakotoson, T., Tanaka, A., & Saito, K. (2019). Challenges and opportunities for improving N use efficiency for rice production in sub-Saharan Africa. *Plant Production Science*, *22*, 413–427. https://doi.org/10.1080/1343943X.2019.1617638

Tsujimoto, Y., Sakata, M., Raharinivo, V., Tanaka, J. P., & Takai, T. (2020). AZ-97 (*Oryza sativa* ssp. *Indica*) exhibits superior biomass production by maintaining the tiller numbers, leaf width, and leaf elongation rate under phosphorus deficiency. *Plant Production Science* (In press.) https://doi.org/10.1080/1343943X.2020.1808026

Vanlauwe, B., Wendt, J., Giller, K. E., Corbeels, M., Gerard, B., & Nolte, C. (2014). A fourth principle is required to define conservation agriculture in sub-Saharan Africa: The appropriate use of fertilizer to enhance crop productivity. *Field Crops Research*, *155*, 10–13. https://doi.org/10.1016/j.fcr.2013.10.002

Wissuwa, M., Kretzschmar, T., & Rose, T. J. (2016). From promise to application: Root traits for enhanced nutrient capture in rice breeding. *Journal of Experimental Botany*, *67*, 3605–3615. https://doi.org/10.1093/jxb/erw061

Yamaji, N., Takemoto, Y., Miyaji, T., Mitani-Ueno, N., Yoshida, K. T., & Ma, J. F. (2017). Reducing phosphorus accumulation in rice grains with an impaired transporter in the node. *Nature* (London), *541*, 92–95. https://doi.org/10.1038/nature20610

Ye, T., Li, Y., Zhang, J., Hou, W., Zhou, W., Lu, J., … Yamamoto, T. (2019). Nitrogen, phosphorus, and potassium fertilization affects the flowering time of rice (*Oryza sativa* L.). *Global Ecology and Conservation*, *20*, e00753. https://doi.org/10.1016/j.gecco.2019.e00753

Zhang, Y., Tan, L., Zhu, Z., Yuan, L., Xie, D., & Sun, C. (2015). *TOND1* confers tolerance to nitrogen deficiency in rice. *Plant Journal*, *81*, 367–376. https://doi.org/10.1111/tpj.12736

SUPPORTING INFORMATION

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

How to cite this article: Takai T, Sakata M, Rakotoarisoa NM, et al. Effects of quantitative trait locus *MP3* on the number of panicles and rice productivity in nutrient-poor soils of Madagascar. *Crop Science*. 2021;61:519–528. https://doi.org/10.1002/csc2.20344

Received: 31 May 2017 | Accepted: 31 January 2018

DOI: 10.1111/pbr.12575

ORIGINAL ARTICLE

WILEY Plant Breeding

Identification of a novel QTL for the number of spikelets per panicle using a cross between *indica*- and *japonica*-type high-yielding rice cultivars in Japan

Toshiyuki Takai | Hiroshi Nakano | Satoshi Yoshinaga | Motohiko Kondo

NARO Institute of Crop Science, Tsukuba, Ibaraki, Japan

Correspondence
Toshiyuki Takai, Japan International Research Center for Agricultural Sciences, Tsukuba, Ibaraki, Japan.
Email: toshi336@affrc.go.jp

Present addresses
Toshiyuki Takai, Japan International Research Center for Agricultural Sciences, Tsukuba, Ibaraki, Japan.
Hiroshi Nakano, NARO Kyushu Okinawa Agricultural Research Center, Chikugo, Fukuoka, Japan.
Satoshi Yoshinaga, NARO Central Region Agricultural Research Center, Tsukuba, Ibaraki, Japan.
Motohiko Kondo, Nagoya University, Nagoya, Aichi, Japan.

Funding information
Science and Technology Research Partnership for Sustainable Development (SATREPES), Japan Science and Technology Agency (JST)/Japan International Cooperation Agency (JICA); the NARO Institute of Crop Science, Japan

Communicated by: Sang Nag Ahn

Abstract

Yield is a complex trait. To improve it, the accumulation of the favourable alleles of valuable genes is required for each yield-related trait. In this study, we used two high-yielding rice cultivars developed in Japan, *indica*-type 'Takanari' and *japonica*-type 'Momiroman', for a genetic analysis of the sink capacity-related traits. An F_2 population showed transgressive segregation for the number of spikelets per panicle. Quantitative trait locus (QTL) analysis detected four QTLs for the trait. Two of the QTLs were most likely identical to previously cloned *GN1a* and *APO1*, and their Takanari alleles had positive effects. The Momiroman alleles of the other two QTLs had positive effects, and one of these QTLs was most likely identical to *SPIKE*/*GPS*. The QTL on the long arm of chromosome 3 appeared to be novel; it clustered with QTLs for grain length and days-to-heading. Substitution mapping revealed that the close linkage of QTLs caused the clustering. These results suggest that the combination of the favourable alleles of detected QTLs could lead to greater sink capacity than that of the parental cultivars.

KEYWORDS
QTL linkage, quantitative trait locus, rice, sink capacity, transgressive segregation

1 | INTRODUCTION

Yield is the most important agronomic trait in various crops. Because of the continuous population increase worldwide and the unpromising prospects for the expansion of arable land area, crop production must be boosted through further increases in the maximum attainable yield per unit land area, namely, yield potential (Long, Marshall-Colon, & Zhu, 2015). Yield potential is defined as the maximum yield that a crop cultivar can achieve in a given environment in the absence of biotic or abiotic stresses with non-limiting nutrients (Evans & Fischer, 1999). Because yield is a complex trait, to improve the yield potential, a wide diversity of germplasms must be used to accumulate favourable alleles of valuable genes for each yield-related trait in breeding programs (McCouch et al., 2013).

Rice is a staple food in Asia. Asian rice (*Oryza sativa* L.) is grouped into *indica*, *temperate japonica*, *tropical japonica*, *aus* and *aromatic* (Huang et al., 2012). In Japan, *temperate japonica* was likely introduced from China 2200–3000 years ago and has been primarily cultivated as a staple food (Matsuo, 1997). In Japan, *indica* and *tropical japonica* have also been used to develop high-yielding cultivars (Horie et al., 2005).

An *indica*-type high-yielding cultivar, 'Takanari', was developed in 1990 (Imbe et al., 2004). Takanari originates from high-yielding *indica*

cultivars including 'IR8'; it has the *sd1* gene and is shorter in plant stature than normal *temperate japonica* cultivars in Japan (Takai et al., 2014). Recent yield trials reported that Takanari produced the highest brown rice yield (11.7 t ha^{-1}) on record in Japan (Nagata, Sasaki, Ohdaira, & Yoshinaga, 2009; Takai et al., 2014). The high yield was achieved because of large panicles, a high leaf photosynthetic ability, the high accumulation of non-structural carbohydrates in the culm and leaf sheath at the full heading stage, and the translocation of many of these carbohydrates into panicles during grain filling (Kanemura, Homma, Ohsumi, Shiraiwa, & Horie, 2007; Ohsumi et al., 2007; Takai et al., 2006). At least two quantitative trait loci (QTLs), *GN1a* and *APO1*, contribute to the large panicles (Takai et al., 2014), and another QTL, *GPS*, underlies the high chlorophyll content and, therefore, the high leaf photosynthetic ability in Takanari (Takai et al., 2013). The causal gene of *GPS* is identical to that of *SPIKE*, which controls the number of spikelets per panicle, which indicates an allelic relationship between *GPS* and *SPIKE* (Fujita et al., 2013; Takai et al., 2013, 2017).

A *japonica*-type high-yielding cultivar, 'Momiroman', was developed in 2008 from the New Plant Type (NPT) line IR65598-112-2 as the donor parent (Hirabayashi et al., 2010). NPT lines, which descended from *tropical japonica*, were developed in 2008 on the basis of a newly designed plant type characterized by a small number of large panicles (200–250 spikelets per panicle), thick and sturdy stems, and dark green erected leaves (Peng, Khush, Virk, Tang, & Zou, 2008). The NPT allele of *SPIKE/GPS* increases the number of spikelets per panicle in the genetic backgrounds of *indica* cultivars (Fujita et al., 2013). Momiroman produces large panicles with large grains, which lead to a larger sink capacity than that in Takanari, although the grain filling in Momiroman is inferior (Yoshinaga, Takai, Arai-Sanoh, Ishimaru, & Kondo, 2013). Sink capacity is defined as the product of single grain weight and the number of spikelets per unit land area (Horie et al., 2003). Although whether *SPIKE/GPS* causes large panicles or large sink capacity in Momiroman remains unclear, identifying the genetic factors underlying sink capacity in Momiroman and combining these factors with high leaf photosynthesis rate and good grain filling in Takanari may be a promising strategy to increase rice yield potential in Japan.

In this study, we conducted genetic analysis of traits related to sink capacity using materials derived from a cross between Takanari and Momiroman. Because promising QTLs were clustered in a specific genomic region, we next developed backcross progeny with recombination close to the QTL cluster. Using this progeny, we assessed whether QTL clustering was due to the pleiotropic effect of a single QTL or linkage of several QTLs.

2 | MATERIALS AND METHODS

2.1 | Plant materials and cultivation

Two high-yielding cultivars, 'Takanari' and 'Momiroman', and a leading cultivar, 'Koshihikari', from Japan, were used. Takanari is *indica*-type (Imbe et al., 2004), whereas Momiroman is *japonica*-type (Hirabayashi et al., 2010) (Figure 1a). Koshihikari is a *temperate japonica*. Takanari was crossed with Momiroman, and 200 self-pollinated F_2 progeny and the parents were grown for QTL analysis in a paddy field at the NARO Institute of Crop Science, Tsukubamirai (36°02′N, 140°04′E), Ibaraki, Japan, in 2010. Twenty-day-old seedlings were transplanted (one seedling per hill) on June 1. F_2 progeny were planted in 10 rows of 20 hills, with a spacing of 15 cm between hills and 30 cm between rows. The parents were planted in three rows with the same plant density. Basal fertilizer was applied at a rate of 6 g N/m^2 as controlled-release fertilizer (2 g of LP40, 2 g of LPS100 and 2 g of LP140), 5.2 g P/m^2 and 7.5 g K/m^2. LP40 and LP140 release 80% of their total nitrogen content at a uniform rate for up to 40 and 140 days after application, respectively, at 20–30°C. LPS100 releases 80% of its total nitrogen content at a sigmoid rate for up to 100 days after application at 20–30°C.

On the basis of initial analysis of F_2 progeny, an F_1 plant was backcrossed to Takanari for two generations to obtain a BC_2F_1 plant with heterozygous parts of chromosomes 3, 7 and 11 in the Takanari background. To confirm QTLs, 174 self-pollinated BC_2F_2 progeny and the parents were grown in a paddy field at NARO in 2012; 23-day-old seedlings were transplanted into the field on June 8. Plant density and fertilizer treatment were as in 2010. For substitution mapping of the candidate QTLs, three of the 174 BC_2F_2 plants with recombination near the QTLs were selected, and BC_2F_3 seeds were harvested. From each of the BC_2F_3 lines, we selected one plant that was homozygous for the Takanari allele and one that was homozygous for the Momiroman allele of the recombinant chromosome segment near the QTLs. The BC_2F_3 plants were self-pollinated, BC_2F_4 seeds were harvested, and three pairs of BC_2F_4 lines were grown in the paddy field at NARO in 2015; 22-day-old seedlings were transplanted into the field on June 4 in three rows with the same plant density as in 2010. Fertilizer treatment was also as in 2010.

2.2 | Phenotyping

Days-to-heading from sowing was recorded when the first panicle headed on each plant. At full heading stage, the soil–plant analysis development (SPAD) value, an index of leaf chlorophyll content, of the fully extended flag leaf on the main stem was measured with a SPAD meter (SPAD-502; Konica-Minolta, Tokyo, Japan). At maturity, a panicle on the main stem was harvested, and the number of spikelets on the panicle was counted. Seeds were threshed and their images were taken with a digital camera (D7000; Nikon, Tokyo, Japan). Seed length and width were measured with the SmartGrain grain shape analysis software (Tanabata, Shibaya, Hori, Ebana, & Yano, 2012). Phenotyping was performed for 153 of the 200 F_2 plants and 102 of the 174 BC_2F_2 plants (border plants were excluded). Five plants were measured for each parent. For substitution mapping using three pairs of BC_2F_4 lines, 10 plants in the middle of each line were phenotyped.

2.3 | QTL mapping in F_2 and BC_2F_2 populations

For QTL analysis of the F_2 population, 105 genomewide SSR markers (International Rice Genome Sequencing Project 2005; McCouch et al., 2002) and one InDel marker for *GN1a* (Takai et al., 2014)

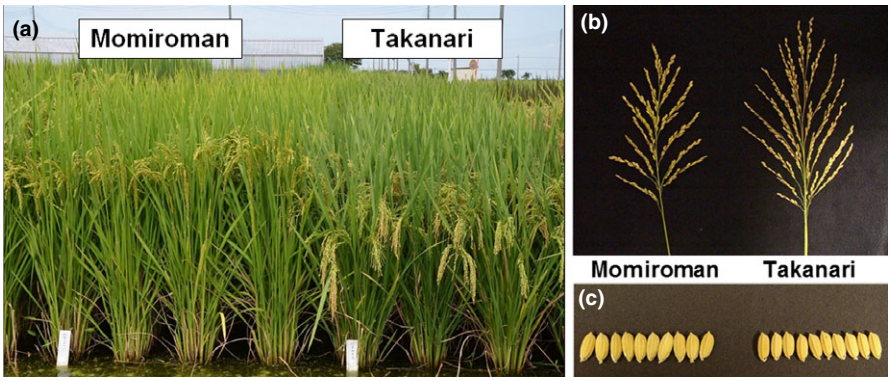


FIGURE 1 Characteristics of the high-yielding rice cultivars Takanari and Momiroman. (a) Plant morphology, (b) panicle and (c) grains

were used. An additional eight SSR markers on the long arm of chromosome 3 were used for QTL analysis of the BC_2F_2 population and the development of BC_2F_4 lines. Total DNA of each plant was extracted from a small piece of young leaf by the simple DNA extraction method according to Takeuchi et al. (2008). Linkage maps were constructed in MAPMAKER/EXP 3.0 software (Lander et al., 1987). The chromosomal positions and effects of putative QTLs were determined by composite interval mapping in QTL Cartographer 2.0 software (Basten, Weir, & Zeng, 2002). The threshold of QTL detection was based on 1,000 permutation tests at the 5% level of significance. The additive and dominant effects and phenotypic variance (R^2) explained by each QTL were estimated from the peak LOD score. To address the effect of QTL combinations on the number of spikelets per panicle in the F_2 population, plants that were homozygous for the 'Takanari' allele or the Momiroman allele for all of the detected QTLs were selected, and the mean values of the number of spikelets per panicle in each combination were calculated. To investigate DNA polymorphisms in *GN1a*, *APO1* and *SPIKE/GPS* among 'Takanari', Momiroman and Koshihikari, the forward and reverse primers in Table S1 were used.

2.4 | Substitution mapping in BC_2F_4 chromosomal substitution lines

For substitution mapping, the significance of the differences in traits within a pair of BC_2F_4 lines was determined by Student's *t* test (JMP 6.0.3 software; SAS Institute, USA).

3 | RESULTS

3.1 | QTL detection in the F_2 population

Both parental cultivars, 'Takanari' and 'Momiroman', produced approximately 250 spikelets per panicle (Figures 1b and 2), whereas grain length and width values were significantly higher ($p < .001$) in Momiroman (8.2 mm and 3.6 mm, respectively) than in 'Takanari' (7.6 mm and 3.3 mm, respectively) (Figures 1c and 2). Takanari headed 6 days earlier (86 days-to-heading) than Momiroman (92 days-to-heading), and at full heading, Takanari had a significantly higher ($p < .001$) SPAD value of the flag leaves (40.8) than that for Momiroman (32.9) (Figure 2). In the F_2 population, each trait showed continuous distribution with transgressive segregation (Figure 2). In particular, the transgressive segregation was remarkable for the number of spikelets per panicle, which ranged from 166 to 488 (Figure 2).

Quantitative trait locus analysis of the F_2 population detected four QTLs for the number of spikelets per panicle on the short arm of chromosome 1 and on the long arms of chromosomes 3, 4 and 6 (Figure 3). Each QTL explained 6.9% to 17% of R^2. The Takanari alleles of the QTLs on chromosomes 1 and 6 and the Momiroman alleles of the QTLs on chromosomes 3 and 4 increased the number of spikelets per panicle (Table 1). Combination analysis showed that plants with positive alleles for the four QTLs produced the highest number of spikelets per panicle (397) (Table 2). The number of spikelets per panicle decreased as the number of positive alleles decreased (322–236), and plants with negative alleles for the four QTLs produced the fewest spikelets per panicle (212). Three QTLs were detected for grain length on the long arms of chromosomes 3, 6 and 7; each of them explained from 8.8% to 17.9% of R^2. The Momiroman alleles of QTLs on chromosomes 3 and 6 and the Takanari allele of the QTL on chromosome 7 increased grain length. For grain width, one QTL was detected on the long arm of chromosome 2, which explained 17.4% of R^2, and the Momiroman allele increased grain width. For SPAD value, two QTLs were detected on the long arms of chromosomes 4 and 6; they explained 33% and 10.3% of R^2, respectively. The Takanari allele of the QTL on chromosome 4 and the Momiroman allele of the QTL on chromosome 6 increased the SPAD value. For days-to-heading, four QTLs were detected on the short arms of chromosomes 1 and 6 and on the long arms of chromosomes 3 and 10; each of these QTLs explained 5.9–33.9% of R^2. The Takanari alleles of QTLs on chromosomes 3 and 6 and the Momiroman alleles of QTLs on chromosomes 1 and 10 increased days-to-heading.

Three of the four QTLs for the number of spikelets per panicle were detected near *GN1a*, *APO1* and *SPIKE/GPS* (Figure 3), each of which was previously cloned and found to be a single gene controlling the number of spikelets per panicle or chlorophyll content (Ashikari et al., 2005; Fujita et al., 2013; Takai et al., 2013; Terao, Nagata, Morino, & Hirose, 2010). The QTL for SPAD value on chromosome 4 was also near *SPIKE/GPS*. To reveal whether the detected

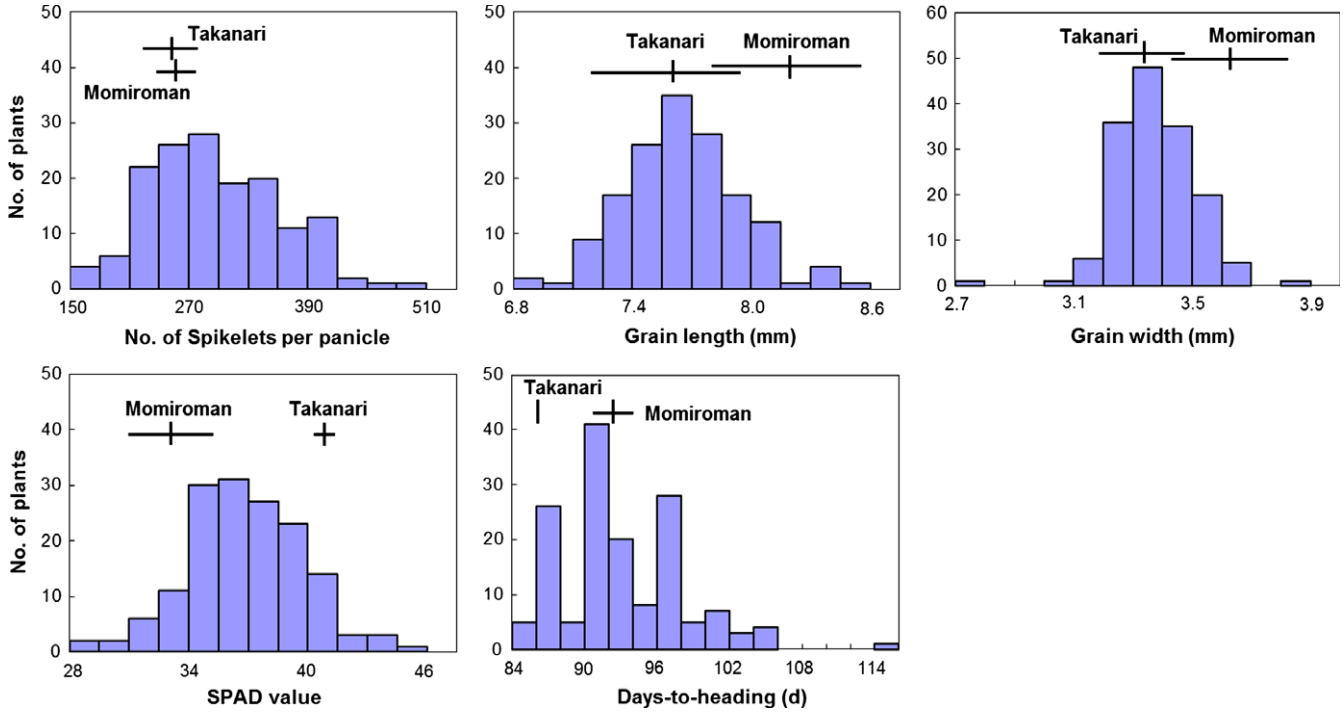


FIGURE 2 Frequency distribution of three sink capacity-related traits (number of spikelets per panicle, grain length and grain width), soil–plant analysis development (SPAD) values of flag leaves at full heading, and days-to-heading in 153 F_2 plants derived from a cross between Takanari and Momiroman. Vertical lines denote mean parental values; horizontal lines denote *SD* [Colour figure can be viewed at wileyonlinelibrary.com]

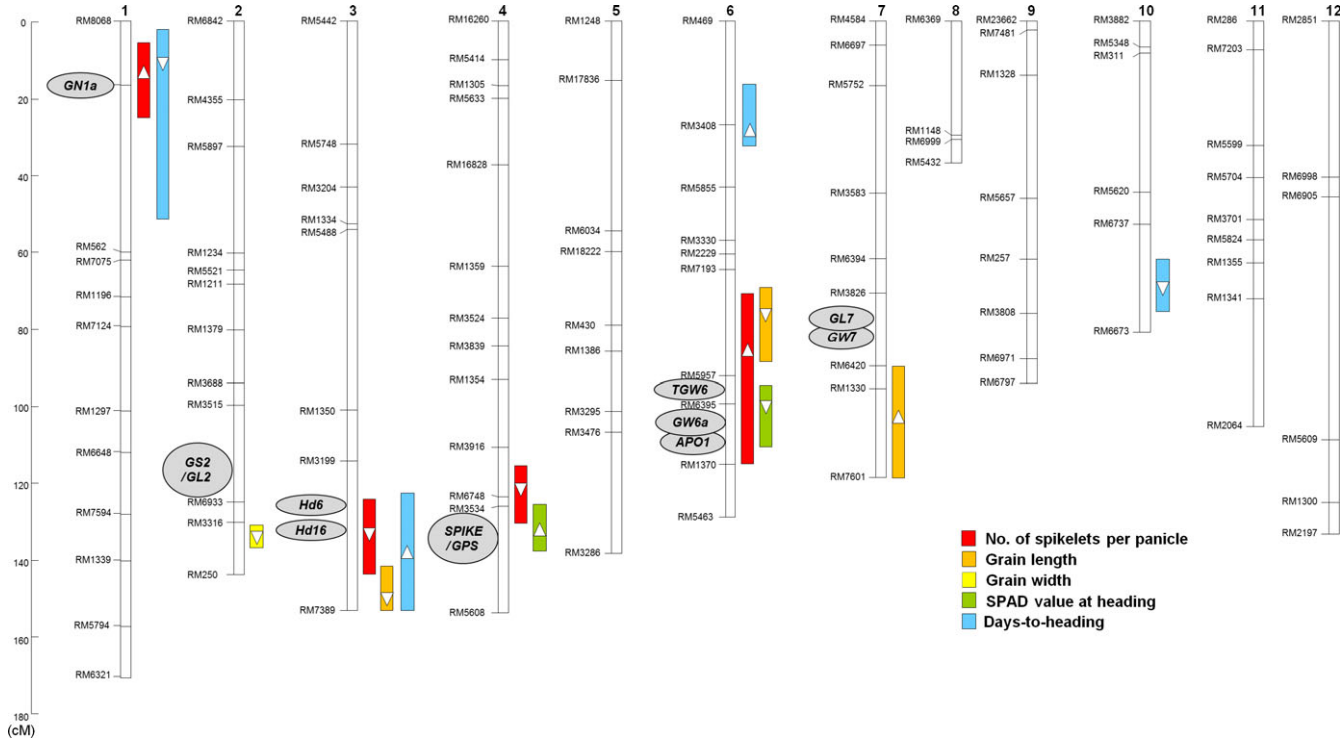


FIGURE 3 Chromosomal locations of quantitative trait locus (QTLs) for three sink capacity-related traits (number of spikelets per panicle, grain length and grain width), soil–plant analysis development (SPAD) value of flag leaves at full heading, and days-to-heading in 153 F_2 plants. Chromosome numbers are indicated above each linkage map. Marker names and QTLs previously identified as single genes are indicated to the left of each linkage map. Logarithm of odds (LOD) peaks of putative QTLs (triangles) and their 1-LOD support intervals (boxes; van Ooijen, 1992) are shown to the right of each linkage map. Trait values increased by Takanari alleles (upward triangles) and by Momiroman alleles (downward triangles) are shown [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Putative QTLs controlling three sink capacity-related traits, SPAD value of flag leaves at full heading stage, and days-to-heading in 153 F_2 plants derived from a cross between Takanari and Momiroman

Trait	Chr.	Flanking markers	LOD[a]	A[b]	D[c]	R^2[d]
No. of spikelets per panicle	1	RM8068–*GN1a*	5.9	29.5	−9.0	11.7
	3	RM3199–RM7389	4.5	−36.1	−19.6	16.6
	4	RM3916–RM6748	8.6	−38.7	−16.7	17.0
	6	RM7193–RM5957	3.1	20.2	−17.6	6.9
Grain length	3	RM3199–RM7389	9.4	−0.21	−0.03	21.9
	6	RM7193–RM5957	6.5	−0.16	−0.09	17.9
	7	RM1330–RM7601	3.7	0.08	−0.12	8.8
Grain width	2	RM3361–RM250	6.0	−0.07	−0.04	17.4
SPAD value at heading	4	RM3534–RM5608	13.7	2.2	1.6	33.0
	6	RM6395	4.1	−1.3	0.2	10.3
Days-to-heading	1	RM8068–*Gn1*	3.1	−1.4	1.6	5.9
	3	RM3199–RM7389	3.1	2.2	−0.4	7.9
	6	RM3408–RM5855	8.7	2.8	−1.1	15.1
	10	RM6737–RM6673	13.5	−4.1	1.2	33.9

LOD, logarithm of odds; QTL, quantitative trait locus; SPAD, soil–plant analysis development.
[a]Logarithm of odds.
[b]Additive effect of the Takanari allele compared with the Momiroman allele.
[c]Dominance effect of the Takanari allele compared with the Momiroman allele.
[d]Percentage of phenotypic variance explained by each QTL.

TABLE 2 The effect of four QTLs combinations on the number of spikelets per panicle in the F_2 population derived from a cross between Takanari and Momiroman

Chr.1 (*GN1a*)[a]	Chr.3 (RM3199)	Chr.4 (RM6748)	Chr.6 (RM5957)	No. of spikelets per panicle	No. of F_2 plants
+[b]	+	+	+	397	2
+	+	+	–	322	1
+	–	+	+	311	2
+	+	–	–	311	1
+	–	+	–	352	1
–	–	+	+	228	1
–	+	–	–	257	1
–	–	–	+	236	2
–	–	–	–	212	2
+	–	–	+	251 (Takanari)	
–	+	+	–	257 (Momiroman)	

[a]Nearest markers of QTLs for the number of spikelets per panicle.
[b]+ and – indicate homozygous genotype of QTLs increasing and decreasing the number of spikelets per panicle, respectively.

QTLs were *GN1a*, *APO1* and *SPIKE/GPS*, we tested DNA polymorphisms in these QTLs with PCR-based gene markers and found that Momiroman differed from Takanari, but not from Koshihikari, at all three QTLs (Figure 4).

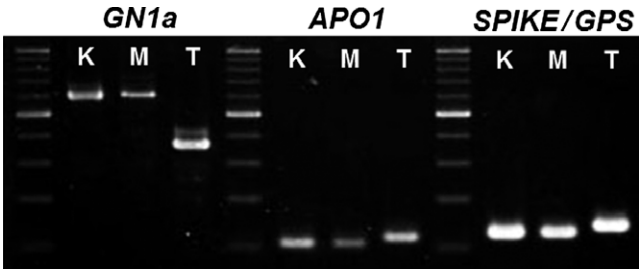


FIGURE 4 DNA polymorphisms in *GN1a*, *APO1* and *SPIKE/GPS* detected with PCR-based gene markers. K, Koshihikari; M, Momiroman; T, Takanari

3.2 | Substitution mapping of the QTL on the long arm of chromosome 3

One QTL for the number of spikelets per panicle on the long arm of chromosome 3 was not detected near any previously cloned QTL

TABLE 3 Putative QTLs controlling two sink capacity-related traits and days-to-heading in 102 BC_2F_2 plants in the Takanari genetic background

Trait	Chr.	Flanking markers	LOD[a]	A^b	D^c	R^{2d}
No. of spikelets per panicle	3	RM1350-1–RM3199	2.8	−11.3	−16.9	13.7
Grain length	3	RM6135	19.7	−0.3	0.03	57.9
Days-to-heading	3	RM7389	3.6	0.7	0.6	13.2

LOD, logarithm of odds; QTL, quantitative trait locus.
[a]Logarithm of odds.
[b]Additive effect of the Takanari allele compared with the Momiroman allele.
[c]Dominance effect of the Takanari allele compared with the Momiroman allele.
[d]Percentage of phenotypic variance explained by each QTL.

but was clustered with QTLs for grain length and days-to-heading (Figure 3). To confirm the QTLs and to verify the position of each QTL, we selected a BC_2F_1 plant with heterozygous parts of chromosomes 3, 7 and 11 in the Takanari background (Figure 5a) and used self-pollinated BC_2F_2 progeny for genetic analysis. The number of spikelets in the BC_2F_2 population was 198–345, grain length was 7.4–8.3 mm, and days-to-heading was 88–94 (Figure 5b). The BC_2F_2 population headed earlier than the parental cultivars. For each trait, a QTL on the long arm of chromosome 3 was confirmed (Figure 5c). The peak of LOD score for the number of spikelets per panicle was near RM1350-1, that for grain length was near RM6135 and that for days-to-heading was near RM7389. The QTL for the number of spikelets per panicle explained 13.7% of R^2, and the Momiroman allele increased the number of spikelets per panicle (Table 3). The QTL for grain length explained 57.9% of R^2, and the Momiroman allele increased grain length. The QTL for days-to-heading explained

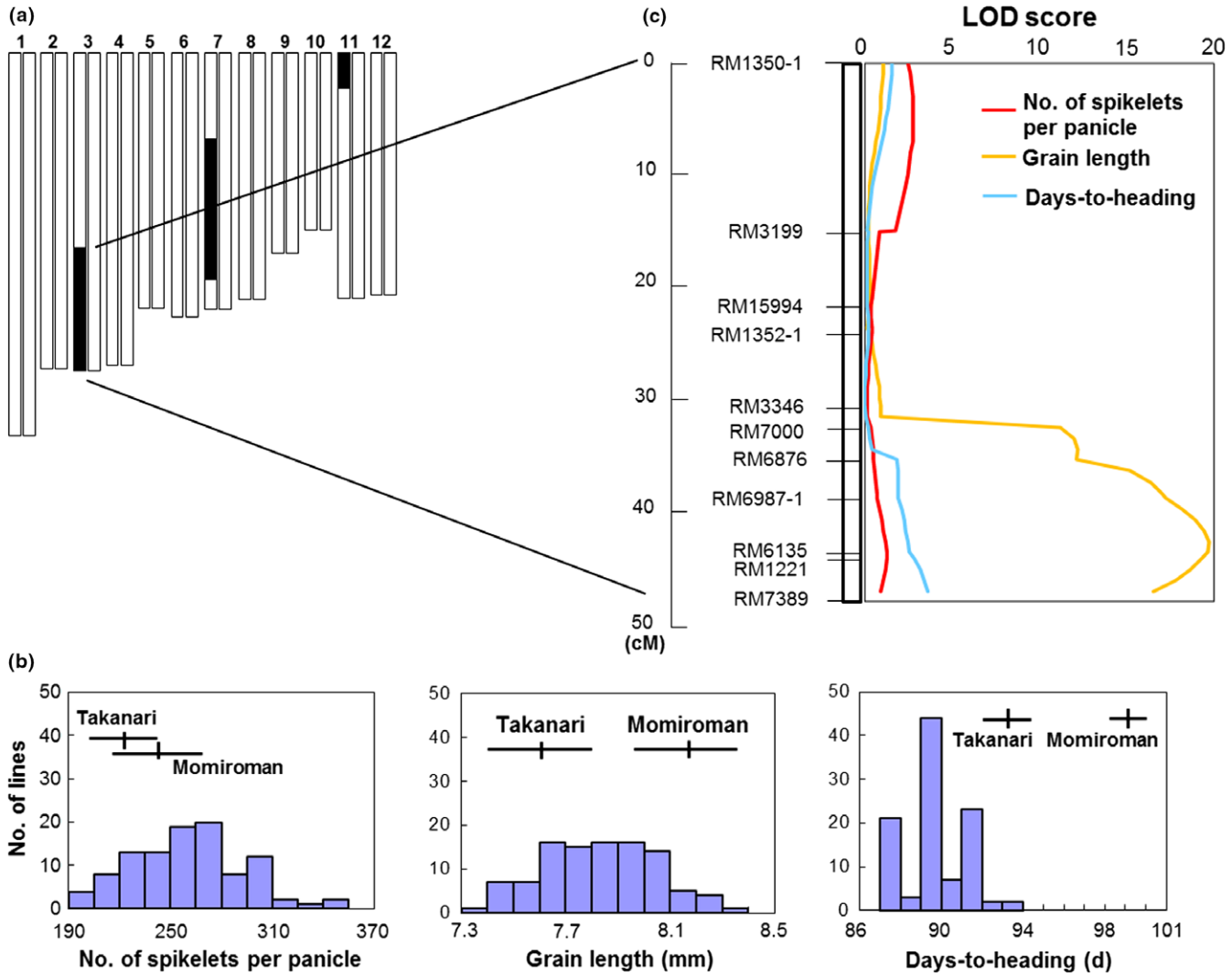


FIGURE 5 QTL analysis for the number of spikelets per panicle, grain width and days-to-heading using 102 BC_2F_2 plants. (a) Graphical genotype of the selected BC_2F_1 plant. White bars, regions from Takanari; black bars, regions from Momiroman. (b) Frequency distribution of the traits. Vertical lines denote mean parental values; horizontal lines denote *SD*. (c) Logarithm of odds (LOD) scores of QTLs for the number of spikelets per panicle (red line), grain length (orange line) and days-to-heading (light blue line) on the long arm of chromosome 3. LOD, logarithm of odds; QTL, quantitative trait locus [Colour figure can be viewed at wileyonlinelibrary.com]

13.2% of R^2, and the Takanari allele increased the number of days-to-heading.

To further delimit the candidate genomic regions of the QTLs, we selected three BC_2F_2 plants with recombination close to the detected QTLs and obtained pairs of BC_2F_4 recombinant lines (one line homozygous for Takanari and the other one for Momiroman) for each plant through self-pollination (Figure 6). Significant differences were observed for all three traits in pair No. 130, for days-to-heading in pair No. 159 and for the number of spikelets per panicle in pair No. 29 (Figure 6). Based on the phenotype and genotype data, we mapped the QTLs for the number of spikelets per panicle between RM5488 and RM3199, for grain length between RM6897-1 and RM7389, and for days-to-heading between RM1350-1 and RM6135 (Figure 6).

4 | DISCUSSION

The balance among sink capacity, source strength and carbohydrate translocation determines rice yield. One of the strategies to improve yield potential in rice is to reveal the genetic factors underlying these components and to pyramid the favourable alleles in the current high-yielding cultivar backgrounds. In this study, we focused on the traits related to sink capacity and conducted genetic analysis using two high-yielding cultivars, Takanari (*indica*-type) and Momiroman (*japonica*-type).

We detected four QTLs for the number of spikelets per panicle by QTL analysis of the F_2 population. The Takanari alleles of QTLs on chromosomes 1 and 6 and the Momiroman alleles of the QTLs on chromosomes 3 and 4 increased the number of spikelets per panicle (Figure 3, Table 1). The F_2 plants with positive alleles for the four QTLs produced the highest number of spikelets per panicle (397) in the QTL combination analysis (Table 2). These results suggest that larger panicles (more spikelets per panicle) than those of the parental cultivars can be achieved by combining the favourable alleles of these QTLs. Three of these four QTLs were detected near *GN1a*, *APO1* and *SPIKE/GPS* (Figure 3). Takanari has positive alleles of *GN1a* and *APO1* for the number of spikelets per panicle in comparison with those of Koshihikari (Takai et al., 2014). However, *GPS/SPIKE* has pleiotropic effects, namely, the allele from Takanari increases chlorophyll content but decreases the number of spikelets per panicle compared with that from Koshihikari (Takai et al., 2013). Because the effects of the three QTLs detected between Takanari and Momiroman in this study were similar to the effects of QTLs detected between Takanari and Koshihikari, we tested DNA polymorphisms in these genes among Takanari, Momiroman and Koshihikari. The test showed that Momiroman had the same DNA polymorphisms as Koshihikari in all three genes (Figure 4). These results suggest that three of the four QTLs detected in this study are identical to *GN1a*, *APO1* and *SPIKE/GPS*.

Because another QTL for the number of spikelets per panicle on the long arm of chromosome 3 was not close to any previously cloned QTLs and was clustered with QTLs for grain length and days-to-heading, we focused on this genomic region. QTL clustering may be caused by the pleiotropy of a single gene or close linkage of several genes (Cai & Morishima, 2002). The days-to-heading often pleiotropically affects agronomic traits (Keurentjes et al., 2007; Takai et al., 2009). To investigate whether pleiotropic effects or multifactorial linkage caused the QTL cluster, we conducted substitution mapping using recombinant homozygous lines with recombination within the QTL regions. We found that the three QTLs were located in different regions of the long arm of chromosome 3 (Figure 6), indicating that QTL clustering was caused by the linkage of the three QTLs. The QTL for the number of spikelets per panicle was mapped within 9 Mb between RM5488 and RM3199. To our knowledge, no QTL for the number of spikelets per panicle has been cloned as a

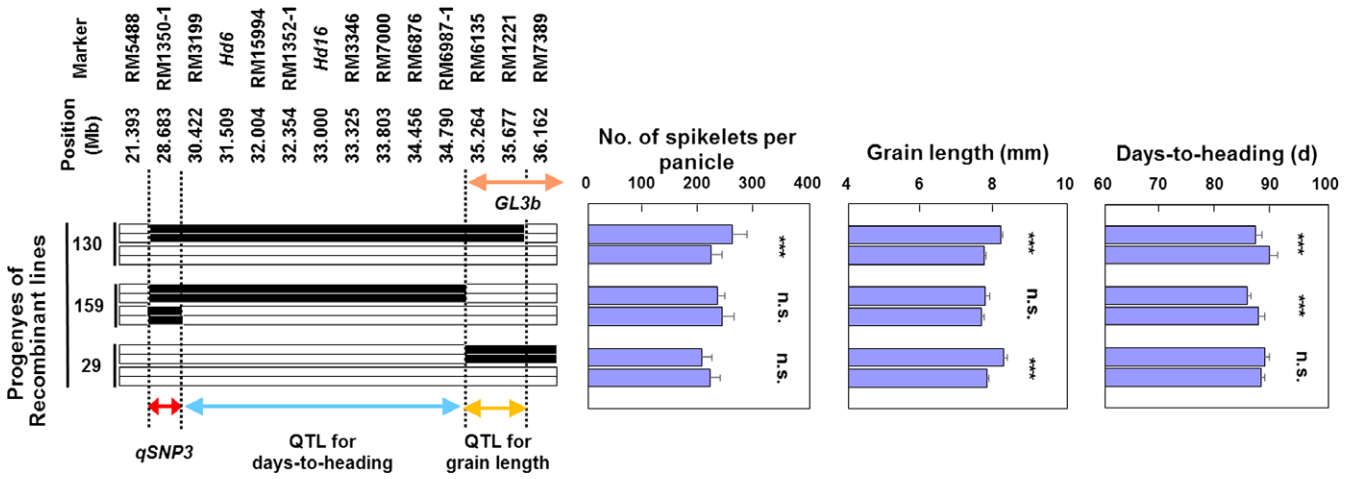


FIGURE 6 Substitution mapping of QTLs for the number of spikelets per panicle, grain length and days-to-heading on the long arm of chromosome 3 based on three pairs of BC_2F_4 lines. White bars, regions from Takanari; black bars, regions from Momiroman. Each bluish bar in the trait graphs represents the mean ± *SD* of the genotype shown on the left ($n = 10$). $^{**}p < .001$; n.s., not significant within pairs of recombinant lines (Student's *t* test) [Colour figure can be viewed at wileyonlinelibrary.com]

single gene or fine-mapped in this region; therefore, this QTL is likely to be novel and we named the QTL *qSNP3* (*Spikelet Number per Panicle 3*). The days-to-heading QTL was mapped within 6.6 Mb between RM1350-1 and RM6135. Previously, *Hd6* was identified in this region between *indica* and *japonica* cultivars, and *Hd16* was identified between *japonica* cultivars (Hori et al., 2013; Takahashi, Shomura, Sasaki, & Yano, 2001). Further study is necessary to reveal whether the QTL detected in this study is *Hd6*, *Hd16* or neither of them. The QTL for grain length was narrowed down to the 1.4-Mb region between RM6987-1 and RM7389. The region partly overlapped with that of *GL3b*, a QTL for grain length recently mapped within a 1.2-Mb region on the long arm of chromosome 3 (Segami et al., 2016). Further study is also required to confirm whether our QTL is identical to *GL3b*.

In addition to the four QTLs for the number of spikelets per panicle and the QTL for grain length on the long arm of chromosome 3, we detected two QTLs for grain length on the long arms of chromosomes 6 and 7 and a QTL for grain width on the long arm of chromosome 2 (Figure 3). The position of each QTL slightly differed from those of QTLs previously cloned as single genes, namely, *TGW6* and *GW6a* on chromosome 6 (Ishimaru et al., 2013; Song et al., 2015), *GL7* and *GW7* on chromosome 7 (Wang, Li et al., 2015; Wang, Xionget al., 2015), and *GS2*/*GL2* on chromosome 2 (Che et al., 2015; Duan et al., 2015). Although we did not confirm these QTLs by substitution mapping, the positive effects of Momiroman alleles of the QTLs on the long arms of chromosomes 2 and 6 suggest their utility for enlarging grains in the Takanari genetic background.

Our study detected four QTLs for the number of spikelets per panicle of which three were most likely identical to *GN1a*, *APO1* and *SPIKE*/*GPS*. The fourth QTL-*qSNP3* was linked to QTLs for grain length and days-to-heading but was mapped in a different region on the long arm of chromosome 3. Although the trade-off effect of the QTLs on the number of panicles per m^2 must be examined, which is a component of sink capacity, our results suggest that larger panicles might be possible with the combination of the favourable alleles of these QTLs, thereby increasing the sink capacity above that of the current high-yielding cultivars. We are also conducting genetic analysis of grain filling using the materials derived from the same cross between Takanari and Momiroman. Combining improved sink capacity with good grain filling in the Takanari genetic background can shed light on the possible ways to increase rice yield potential in future breeding programs.

ACKNOWLEDGEMENTS

We thank the staff of the research support section at the Central Region Agricultural Research Center, NARO. This work was supported by a grant from the NARO Institute of Crop Science, Japan, and partly supported by Science and Technology Research Partnership for Sustainable Development (SATREPES), Japan Science and Technology Agency (JST)/Japan International Cooperation Agency (JICA). We declare that we have no conflict of interest.

ORCID

Toshiyuki Takai http://orcid.org/0000-0002-6498-610X

REFERENCES

Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., … Matsuoka, M. (2005). Cytokinin oxidase regulates rice grain production. *Science*, *309*, 741–745. https://doi.org/10.1126/science.1113373

Basten, C.J., Weir, B. S., & Zeng, Z. B. (2002). *QTL cartographer. Version 1.16. A reference manual and tutorial for QTL mapping*. Raleigh: North Carolina State University.

Cai, W., & Morishima, H. (2002). QTL clusters reflect character associations in wild and cultivated rice. *Theoretical and Applied Genetics*, *104*, 1217–1228.

Che, R., Tong, H., Shi, B., Liu, Y., Fang, S., Liu, D., … Chu, C. (2015). Control of grain size and rice yield by *GL2*-mediated brassinosteroid responses. *Nature Plants*, *2*, 15195. https://doi.org/10.1038/nplants.2015.195

Duan, P., Ni, S., Wang, J., Zhang, B., Xu, R., Wang, Y., … Li, Y. (2015). Regulation of *OsGRF4* by *OsmiR396* controls grain size and yield in rice. *Nature Plants*, *2*, 15203. https://doi.org/10.1038/nplants.2015.203

Evans, L. T., & Fischer, R. A. (1999). Yield potential: Its definition, measurement, and significance. *Crop Science*, *39*, 1544–1551. https://doi.org/10.2135/cropsci1999.3961544x

Fujita, D., Trijatmiko, K. R., Tagle, A. G., Sapasap, M. V., Koide, Y., Sasaki, K., … Kobayashi, N. (2013). *NAL1* allele from a rice landrace greatly increases yield in modern indica cultivars. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 20431–20436. https://doi.org/10.1073/pnas.1310790110

Hirabayashi, H., Nemoto, H., Ando, I., Kato, H., Oota, H., Satou, H., … Kaji, R. (2010). "Momiroman", a new rice cultivar for feed use. *Bulletin of National Institute of Crop Science*, *11*, 31–47.

Hori, K., Ogiso-Tanaka, E., Matsubara, K., Yamanouchi, U., Ebana, K., & Yano, M. (2013). *Hd16*, a gene for casein kinase I, is involved in the control of rice flowering time by modulating the day-length response. *Plant Journal*, *76*, 36–46.

Horie, T., Lubis, I., Takai, T., Ohsumi, A., Kuwasaki, K., Katsura, K., & Nii, A. (2003). Physiological traits associated with high yield potential in rice. In T. W. Mew, D. S. Brar, S. Peng, D. Dawe, & B. Hardy (Eds.), *Rice science: innovations and impact for livelihood* (pp. 117–145). Philippines: International Rice Research Institute, Chinese Academy of Engineering, and Chinese Academy of Agricultural Sciences.

Horie, T., Shiraiwa, T., Homma, K., Katsura, K., Maeda, S., & Yoshida, H. (2005). Can yields of lowland rice resume the increases that they showed in the 1980s?. *Plant Production Science*, *8*, 259–274. https://doi.org/10.1626/pps.8.259

Huang, X., Kurata, N., Wei, X., Wang, Z. X., Wang, A., Zhao, Q., … Han, B. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature*, *490*, 497–501. https://doi.org/10.1038/nature11532

Imbe, T., Akama, Y., Nakane, A., Hata, T., Ise, K., Ando, I., … Koga, Y. (2004). Development of a multipurpose high-yielding rice variety "Takanari". *Bulletin of National Institute of Crop Science*, *5*, 35–51.

International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome. *Nature*, *436*, 793–800.

Ishimaru, K., Hirotsu, N., Madoka, Y., Murakami, N., Hara, N., Onodera, H., … Katoh, E. (2013). Loss of function of the IAA-glucose hydrolase gene *TGW6* enhances rice grain weight and increases yield. *Nature Genetics*, *45*, 707–711. https://doi.org/10.1038/ng.2612

Kanemura, T., Homma, K., Ohsumi, A., Shiraiwa, T., & Horie, T. (2007). Evaluation of genotypic variation in leaf photosynthetic rate and its associated factors by using rice diversity research set of germplasm. *Photosynthesis Research*, *94*, 23–30. https://doi.org/10.1007/s11120-007-9208-7

Keurentjes, J. J. B., Bentsink, L., Alonso-Blanco, C., Hanhart, C. J., Blankestijn-De Vris, H., Effgen, S., … Koornneef, M. (2007). Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics*, *175*, 891–905. https://doi.org/10.1534/genetics.106.066423

Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daley, M. J., Lincoln, S. E., & Newburg, L. (1987). MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, *1*, 174–181. https://doi.org/10.1016/0888-7543(87)90010-3

Long, S. P., Marshall-Colon, A., & Zhu, X. G. (2015). Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell*, *161*, 56–66. https://doi.org/10.1016/j.cell.2015.03.019

Matsuo, A. (1997). Origin and differentiation of cultivated rice. In T. Matsuo, Y. Futsuhara, F. Kikuchi, & H. Yamaguchi (Eds.), *Science of the rice plant, vol. III, genetics* (pp. 69–88). Tokyo, Japan: Food and Agriculture Policy Research Center.

McCouch, S., Baute, G. J., Bradeen, J., Bramel, P., Bretting, P. K., Buckler, E., … Zamir, D. (2013). Agriculture: Feeding the future. *Nature*, *499*, 23–24. https://doi.org/10.1038/499023a

McCouch, S. R., Teytelman, L., Xu, Y., Lobos, K. B., Clare, K., Walton, M., … Stein, L. (2002). Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Research*, *9*, 199–207. https://doi.org/10.1093/dnares/9.6.199

Nagata, K., Sasaki, R., Ohdaira, Y., & Yoshinaga, S. (2009). Growth, yield and dry matter production of high-yielding rice in the warmer region of Japan. *Japanese Journal of Crop Science*, *78*(Extra1), 240–241.

Ohsumi, A., Hamasaki, A., Nakagawa, H., Yoshida, H., Shiraiwa, T., & Horie, T. (2007). A model explaining genotypic and ontogenetic variation of leaf photosynthetic rate in rice (*Oryza sativa*) based on leaf nitrogen content and stomatal conductance. *Annals of Botany*, *99*, 265–273. https://doi.org/10.1093/aob/mcl253

van Ooijen, J. W. (1992). Accuracy of mapping quantitative trait loci in autogamous species. *Theoretical and Applied Genetics*, *84*, 803–811. https://doi.org/10.1007/BF00227388

Peng, S., Khush, G. S., Virk, P., Tang, Q., & Zou, Y. (2008). Progress in ideotype breeding to increase rice yield potential. *Field Crops Research*, *108*, 32–38. https://doi.org/10.1016/j.fcr.2008.04.001

Segami, S., Yamamoto, T., Oki, K., Noda, T., Kanamori, H., Sasaki, H., … Miura, K. (2016). Detection of novel QTLs regulating grain size in extra-large grain rice (*Oryza sativa* L.) lines. *Rice*, *9*, 34. https://doi.org/10.1186/s12284-016-0109-2

Song, X. J., Kuroha, T., Ayano, M., Furuta, T., Nagai, K., Komeda, N., … Ashikari, M. (2015). Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 76–81. https://doi.org/10.1073/pnas.1421127112

Takahashi, Y., Shomura, A., Sasaki, T., & Yano, M. (2001). *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 7922–7927. https://doi.org/10.1073/pnas.111136798

Takai, T., Adachi, S., Fujita, D., Arai-Sanoh, Y., Okamura, M., Kondo, M., & Kobayashi, N. (2017). Effects of yield-related QTLs *SPIKE* and *GPS* in two *indica* rice genetic backgrounds. *Plant Production Science*, *20*, 467–476. https://doi.org/10.1080/1343943X.2017.1385404

Takai, T., Adachi, S., Taguchi-Shiobara, F., Sanoh-Arai, Y., Iwasawa, N., Yoshinaga, S., … Yamamoto, T. (2013). A natural variant of *NAL1*, selected in high-yield rice breeding programs, pleiotropically increases photosynthesis rate. *Scientific Reports*, *3*, 2149. https://doi.org/10.1038/srep02149

Takai, T., Ikka, T., Kondo, K., Nonoue, Y., Ono, N., Arai-Sanoh, Y., … Yamamoto, T. (2014). Genetic mechanisms underlying yield potential in the rice high-yielding cultivar Takanari, based on reciprocal chromosome segment substitution lines. *BMC Plant Biology*, *14*, 295. https://doi.org/10.1186/s12870-014-0295-2

Takai, T., Matsuura, S., Nishio, T., Ohsumi, A., Shiraiwa, T., & Horie, T. (2006). Rice yield potential is closely related to crop growth rate during late reproductive period. *Field Crops Research*, *96*, 328–335. https://doi.org/10.1016/j.fcr.2005.08.001

Takai, T., Ohsumi, A., San-oh, Y., Laza, M. R. C., Kondo, M., Yamamoto, T., & Yano, M. (2009). Detection of a quantitative trait locus controlling carbon isotope discrimination and its contribution to stomatal conductance in *japonica* rice. *Theoretical and Applied Genetics*, *118*, 1401–1410. https://doi.org/10.1007/s00122-009-0990-9

Takeuchi, Y., Hori, K., Suzuki, K., Nonoue, Y., Takemoto-Kuno, Y., Maeda, H., … Ando, I. (2008). Major QTLs for eating quality of an elite Japanese rice cultivar, Koshihikari, on the short arm of chromosome 3. *Breeding Science*, *58*, 437–445. https://doi.org/10.1270/jsbbs.58.437

Tanabata, T., Shibaya, T., Hori, K., Ebana, K., & Yano, M. (2012). SmartGrain: High-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiology*, *160*, 1871–1880. https://doi.org/10.1104/pp.112.205120

Terao, T., Nagata, K., Morino, K., & Hirose, T. (2010). A gene controlling the number of primary rachis branches also controls the vascular bundle formation and hence is responsible to increase the harvest index and grain yield in rice. *Theoretical and Applied Genetics*, *120*, 875–893. https://doi.org/10.1007/s00122-009-1218-8

Wang, S., Li, S., Liu, Q., Wu, K., Zhang, J., Wang, Y., … Fu, X. (2015). The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nature Genetics*, *47*, 949–954. https://doi.org/10.1038/ng.3352

Wang, Y., Xiong, G., Hu, J., Jiang, L., Yu, H., Xu, J., … Qian, Q. (2015). Copy number variation at the *GL7* locus contributes to grain size diversity in rice. *Nature Genetics*, *47*, 944–948. https://doi.org/10.1038/ng.3346

Yoshinaga, S., Takai, T., Arai-Sanoh, Y., Ishimaru, T., & Kondo, M. (2013). Varietal differences in sink production and grain-filling ability in recently developed high-yielding rice (*Oryza sativa* L.) varieties in Japan. *Field Crops Research*, *150*, 74–82. https://doi.org/10.1016/j.fcr.2013.06.004

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Takai T, Nakano H, Yoshinaga S, Kondo M. Identification of a novel QTL for the number of spikelets per panicle using a cross between *indica*- and *japonica*-type high-yielding rice cultivars in Japan. *Plant Breed*. 2018;137:109–117. https://doi.org/10.1111/pbr.12575

Breeding Science 71: 615–621 (2021)
doi: 10.1270/jsbbs.21045

Note

Development and evaluation of pyramiding lines carrying early or late heading QTLs in the *indica* rice cultivar 'IR64'

Toshiyuki Takai*[1,2], Patrick Lumanglas[2], Daisuke Fujita[3], Kazuhiro Sasaki[1], Njato Michael Rakotoarisoa[4], Yasuhiro Tsujimoto[1], Nobuya Kobayashi[5] and Eliza Vie Simon[2]

[1] *Japan International Research Center for Agricultural Sciences*, Tsukuba, Ibaraki 305-8686, Japan
[2] *International Rice Research Institute*, DAPO Box 7777, Metro Manila, Philippines
[3] *Faculty of Agriculture, Saga University*, Saga, Saga 840-8502, Japan
[4] *Rice Research Department, National Center of Applied Research on Rural Development*, Tsimbazaza, Antananarivo BP1690, Madagascar
[5] *Institute of Crop Science, National Agriculture and Food Research Organization (NARO)*, Tsukuba, Ibaraki 305-8518, Japan

The heading date is an important trait for determining regional and climatic adaptability in rice. To expand the adaptability of the *indica* rice cultivar 'IR64', we pyramided multiple early or late heading quantitative trait locus (QTLs) in the 'IR64' genetic background by crossing previously developed near-isogenic lines (NILs) with a single QTL for early or late heading. The effects of pyramiding QTLs were observed in three different climatic zones of the Philippines, Madagascar, and Japan. The early heading pyramiding lines (PYLs) headed 6.2 to 12.8 days earlier than 'IR64' while the late heading PYLs headed 18.8 to 27.1 days later than 'IR64'. The PYLs tended to produce low grain yield compared to 'IR64'. The low yield was not improved by combining *SPIKE*, which is a QTL that increases the number of spikelets per panicle. Conversely, 'IR64-PYL(7+10)' carrying *Hd5* and *Hd1* headed earlier, produced more tillers, and more panicles per m^2 than 'IR64', and mitigated the yield decrease in early heading. These results suggest that the effects of pyramided QTLs on heading date were consistent across various environments and PYLs could be used to enhance the adaptation of 'IR64' in other rice growing environments.

Key Words: adaptability, heading date, pyramiding lines, quantitative trait loci (QTL), rice.

Introduction

Crop adaptation in an environment is crucial for maximum growth and productivity in the environmental conditions (Chloupek and Hrstkova 2005). Flowering time is an important determinant of regional and climatic adaptability in crops (Izawa 2007). An appropriate flowering time enables crops to fully utilize light and temperature resources in the given environment (Zhang *et al.* 2015).

Rice is a short-day plant that is grown widely in Asia and Africa as a staple food. The *indica* high-yielding variety 'IR64' was developed by the International Rice Research Institute (IRRI) in the 1980s. This variety has a wide adaptability and has been distributed in Southeast Asia, South Asia, and West Africa, and was grown in over 10 million hectares of paddy fields by the end of last century (Mackill and Khush 2018). The diverse success was partly due to early maturation (approximately 116 days in the tropics) compared with traditional varieties, which enabled double rice cropping in the tropics (Khush and Virk 2005, Peng and Khush 2003). However, increased versatility is necessary in IR64 to keep up with demand in growing and diverse markets and with producers' demands in changing climatic conditions. With further earlier heading, three rice crops a year is possible and producers can escape the risks of drought or low temperature stresses at the end of the growing period (Farooq *et al.* 2009, Shavrukov *et al.* 2017). Alternatively, a longer growing period may improve the one-time grain yield and reduce labor costs compared with multiple rice cropping. In general, a longer growth duration contributes to a yield increase through high biomass production when grown in adequate climatic conditions (Zhang *et al.* 2009) and higher temperature regions of the future, where late-season low temperatures are problematic (van Oort and Dingkuhn 2021).

The flowering time (heading date) in rice is controlled by multiple genes that respond to photoperiod and temperature (Yano *et al.* 2001). The progress in rice genomics over the past two decades has elucidated the genetic and molecular

Communicated by Mallikarjuna Swamy
Received June 8, 2021. Accepted September 8, 2021.
First Published Online in J-STAGE on November 17, 2021.
*Corresponding author (e-mail: toshi336@affrc.go.jp)

mechanisms underlying heading date (Hori *et al.* 2016, Matsubara *et al.* 2014). Recently, Wei *et al.* (2016) revealed that 'IR64' obtained early maturity with an insensitivity to photoperiod due to the loss of functional alleles of the quantitative trait loci (QTLs), *Hd1* (Yano *et al.* 2000) and *Ehd1* (Doi *et al.* 2004). Using new plant type (NPT) rice varieties as donors, Fujita *et al.* (2009) detected QTLs for heading date in the 'IR64' genetic background. Fujita *et al.* (2011) developed five near-isogenic lines (NILs) carrying the QTLs in the 'IR64' genetic background. Three NILs showed a three to five day earlier heading than 'IR64' while two NILs showed an eight to ten day later heading than 'IR64' in the tropics. Therefore, we hypothesized that a wider variation in heading date is possible by pyramiding the QTLs in the 'IR64' genetic background to expand the adaptability in other rice-growing environments and contribute to an increase in rice production.

The objectives of this study were to develop breeding materials using 'IR64' with modified days to heading. The experiment will be undertaken by pyramiding QTLs for heading date using the NILs in the 'IR64' genetic background. The effects of the pyramided QTLs on heading date and productivity will be evaluated in three different climatic zones. A short growth duration generally decreases grain yield, therefore, we also use *SPIKE*, a QTL that increases the number of spikelets per panicle (Fujita *et al.* 2013), to the QTLs for heading date to evaluate whether *SPIKE* can compensate for the yield decrease.

Materials and Methods

Development of pyramiding lines (PYLs)

Six NILs for heading date and spikelet number per panicle were previously developed by Fujita *et al.* (2011) as 'IR64-NIL7' to 'IR64-NIL11' and 'NIL-*SPIKE*' in the 'IR64' genetic background using more than 200 genome-wide SSR markers (**Fig. 1**). The six NILs were used to develop PYLs according to the procedure summarized in **Supplemental Fig. 1**. 'IR64-NIL7', 'IR64-NIL10', and 'IR64-NIL11' are early heading NILs carrying *Hd5*, *Hd1*, and *qDTH11*[yp7] from each NPT or *temperate japonica* donor variety, respectively. The NILs, 'IR64-NIL8' and 'IR64-NIL9' are late heading NILs carrying *Hd3a* and *qDTH11*[YP6] from each NPT donor variety (**Fig. 1**) (Fujita *et al.* 2011). We first conducted crossings using 'IR64-NIL7', 'IR64-NIL10', and 'IR64-NIL11' as well as between 'IR64-NIL8' and 'IR64-NIL9'. Using DNA markers for the target regions, we selected F_2 or F_3 progenies with the two target segments homozygous for the donor varieties as PYLs; 'IR64-PYL(7+10)', 'IR64-PYL(7+11)', 'IR64-PYL(10+11)', and 'IR64-PYL(8+9)'. In the same way, we then crossed 'IR64-PYL(10+11)' with 'IR64-NIL7' as well as 'NIL-*SPIKE*'. We selected the PYLs with the three target segments homozygous for the donor varieties; 'IR64-PYL(7+10+11)' and 'IR64-PYL(10+11+*SPIKE*)'. Finally, we crossed 'IR64-PYL(7+10+11)' with 'IR64-PYL(10+11+*SPIKE*)', and selected the 'IR64-PYL(7+10+11+*SPIKE*)' that were homozygous with the donor varieties containing the four target segments.

Filed experiments

Field experiments were conducted in the tropical regions in the International Rice Research Institute, Los Baños, Philippines (14°17N, 121°26E), in the wet seasons (WS) of 2018, in farmers' paddy fields in Ankazomiriotra, Madagascar (19°40S, 46°34E) in the rice growing season of 2019–2020, and in the temperate region in the Japan International Research Center for Agricultural Sciences, Tsukuba, Japan (36°05N, 140°08E) in 2020. Rice was grown conventionally in each environment. Seeds were sown in seedling nurseries. The 3- to 4-week-old seedlings were transplanted into the experimental paddy fields with

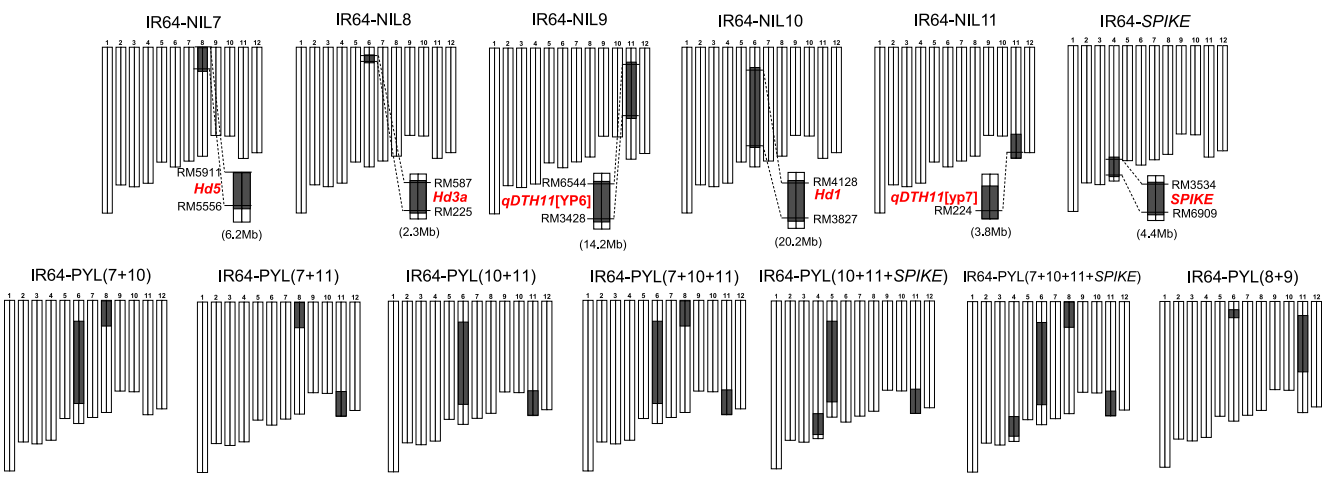


Fig. 1. Graphical genotypes of the near-isogenic lines (NILs) for heading quantitative trait locus (QTLs) and *SPIKE* and also the pyramiding lines (PYLs). The bars represent chromosomes. Chromosome numbers are provided above each bar. The white bars denote regions homozygous for 'IR64' and the black segment denotes a region homozygous for the donors. Mega-base (Mb) in the parentheses shows the size of the introgression segment.

one or two seedlings per hill. The planting densities were: in Los Baños, 20 hills m^{-2} in plots of 7 rows with 21 hills; In Ankazomiriotra, 25 hills m^{-2} in 4 rows with 20 hills; and in Tsukuba, 18.5 hills m^{-2} with a plot size of 4 rows with 13 hills. To evaluate the yield, the experimental plots were arranged in a randomized complete block design with four and three replicates in Ankazomiriotra and Tsukuba, respectively. Chemical fertilizers were applied at a rate of: In Los Baños, 150 kg N ha^{-1}, 45 kg P_2O_5 ha^{-1}, and 45 kg K_2O ha^{-1}; in Ankazomiriotra, 60 kg N ha^{-1}, 40 kg P_2O_5 ha^{-1}, and 29 kg K_2O ha^{-1}; and in Tsukuba, 48 kg N ha^{-1}, 64 kg P_2O_5 ha^{-1}, and 32 kg K_2O ha^{-1}. A total of 14 lines were grown in Tsukuba and 12 lines were grown in Los Baños and Ankazomiriotra. In Los Baños and Ankazomiriotra the excluded lines were 'IR64-PYL(7+10+11+*SPIKE*)' and 'NIL-*SPIKE*'.

Evaluation of days to heading, aboveground biomass, and grain yield

Days to heading was defined as the number of days from sowing to the first panicle heading in each plant. Ten to twenty plants were evaluated for each line in each experiment. In Ankazomiriotra and Tsukuba, six hills were sampled from each plot at the soil surface. The samples were dried at 70°C for 72 h, and weighed to determine the aboveground biomass at heading. At maturity, the number of panicles was counted for 14 hills in Ankazomiriotra and 10 hills in Tsukuba. The hills were then harvested, and the yield and yield components were determined according to the methods described by Takai *et al.* (2021). In Ankazomiriotra, the yield and yield components were not obtained for the late heading lines of 'IR64-NIL8', 'IR64-NIL9', and 'IR64-PYL(8+9)' because farmers accidentally harvested and consumed them. In Tsukuba, the number of tillers was counted using nine hills of 'IR64' and 'IR64-PYL(7+10)' for the periods from transplanting to heading.

Statistics

Statistical analysis was performed using SPSS 23.0 software (IBM). In all analyses a probability value less than 0.05 was considered statistically significant ($P < 0.05$).

Results

Genotypes of PYLs

A total of seven PYLs were developed, and the genotype of each PYL is provided in **Fig. 1**. Using DNA markers for the target segments, we confirmed that 'IR64-PYL(7+10)', 'IR64-PYL(7+11)', and 'IR64-PYL(10+11)' carried the donor alleles for two of the three QTLs (*Hd5*, *Hd1*, and *qDTH11*[yp7]) that accelerate heading, respectively. We also confirmed 'IR64-PYL(7+10+11)' carried the donor alleles for the three QTLs. Both 'IR64-PYL(10+11+*SPIKE*)' and 'IR64-PYL(7+10+11+*SPIKE*)' carried *SPIKE* as well as the two and three early heading QTLs from the donor varieties, respectively. Similarly, 'IR64-PYL(8+9)' carried the donor alleles for *Hd3a* and *qDTH11*[YP6] that delayed heading.

Climate conditions at experimental fields

Daylength from sowing to heading was the shortest in Los Baños (13.0 to 12.0 h), intermediate in Ankazomiriotra (13.5 to 12.2 h), and the longest in Tsukuba (14.6 to 12.8 h) (**Supplemental Fig. 2**). The mean temperature during the growing period was the highest in Los Baños at around 28°C, intermediate in Ankazomiriotra with a constant temperature of approximately 23.3°C, and the lowest in Tsukuba at approximately 21.8°C with a gradual increase until heading in mid-August (13.2°C to 29.5°C) followed by a gradual decrease during grain filling until maturity (29.5°C to 14.3°C). The average solar radiation during the growing period in Ankazomiriotra was 21.3 and in Tsukuba was 15.1 MJ m^{-2} d^{-1}. Solar radiation records were not available in Los Baños.

Variation in heading date among NILs and PYLs

The 'IR64' headed 82.1, 86.1, and 99.3 days after sowing in Los Baños, Ankazomiriotra, and Tsukuba, respectively (**Fig. 2**). Wide variations were observed for days to heading among NILs and PYLs. The days to heading ranged from 69.3 to 100.9 in Los Baños, from 75.7 to 106.7 in Ankazomiriotra, and 93.1 to 126.4 in Tsukuba (**Fig. 2**). Thus, the earliest PYL headed 12.8, 10.4, and 6.2 days earlier than 'IR64' while the latest PYL headed 18.8, 20.6, and 27.1 days later than 'IR64' in Los Baños, Ankazomiriotra, and Tsukuba, respectively. The differences were based on the effects of pyramiding the two heading QTLs. The 'IR64-PYL(7+10)', 'IR64-PYL(7+11)', and 'IR64-PYL(10+11)' headed earlier than the parental NILs, whereas 'IR64-PYL(8+9)' headed later than the parental NILs. However, additional variations by the three QTLs was not observed; 'IR64-PYL(7+10+11)' did not head earlier than 'IR64-PYL(7+10)', 'IR64-PYL(7+11)', and 'IR64-PYL(10+11)' in any experimental field.

Biomass, yield, and yield components among NILs and PYLs

Days to heading was closely correlated with aboveground biomass at heading in both Ankazomiriotra ($r = 0.88$) and Tsukuba ($r = 0.95$) (**Fig. 3**). In contrast, no significant correlation was observed between days to heading and grain yield in either experimental field (**Fig. 4**). 'IR64' produced 3.5 and 5.0 t ha^{-1} of grain yield in Ankazomiriotra and Tsukuba, respectively (**Table 1**). NILs and PYLs with early heading tended to produce low grain yield compared to 'IR64' at both experimental fields. Similarly, the late heading lines showed significantly lower grain yield than 'IR64' in Tsukuba. This result is probably due to the low temperature during the grain filling period. *SPIKE* did not improve the grain yield in the PYLs for early heading in either experimental field. Despite the low yield tendency compared to 'IR64', in Ankazomiriotra,

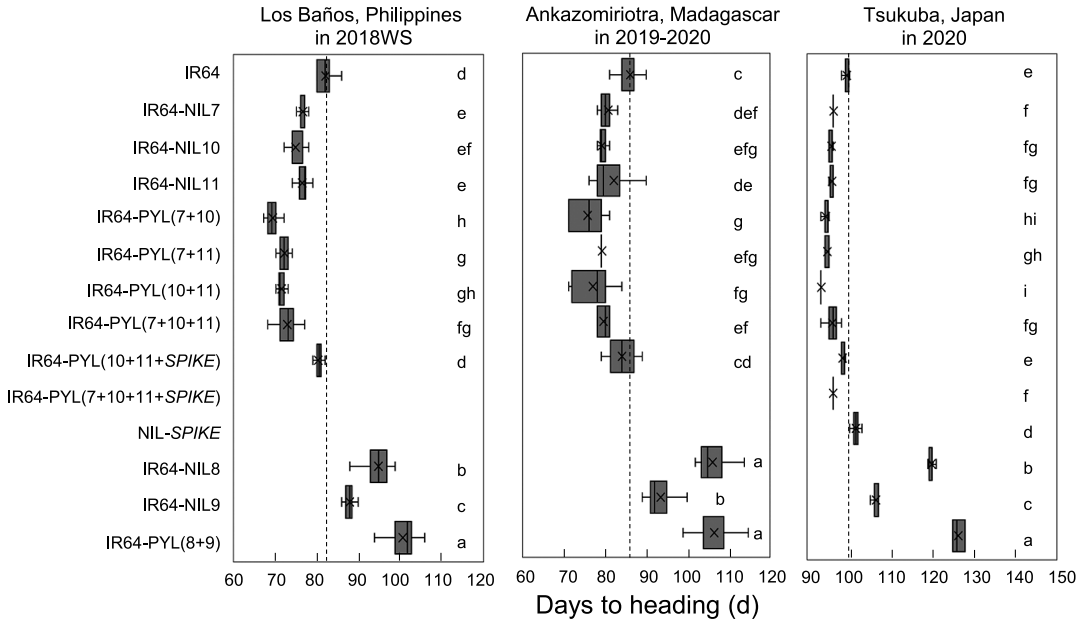


Fig. 2. Comparisons of days to heading among NILs and PYLs with the 'IR64' background in Los Baños, Philippines in 2018WS, Ankazomiriotra, Madagascar in 2019–2020, and Tsukuba, Japan in 2020. Days to heading in each line is represented by a box plot. The dotted line exhibits days to heading of 'IR64'. Different letters indicate significant differences ($P < 0.05$, Tukey's HSD test).

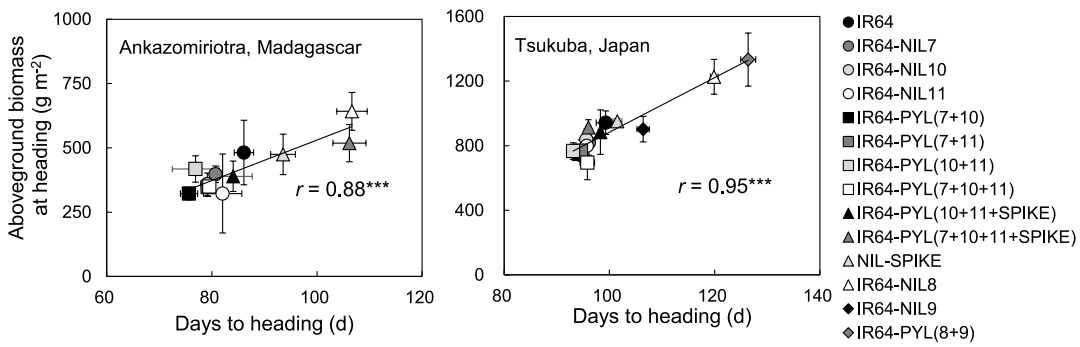


Fig. 3. Relationship between days to heading and aboveground biomass at heading among NILs and PYLs grown in Ankazomiriotra, Madagascar and Tsukuba, Japan. Bars represent standard deviation. *** denotes significance at the 0.1% level.

'IR64-PYL(7+10)' headed earliest and did not decrease in grain yield among NILs and PYLs (**Fig. 4**). To elucidate why 'IR64-PYL(7+10)' maintained the yield levels, we investigated the yield components and identified the greatest number of panicles in 'IR64-PYL(7+10)'. The higher number of panicles resulted in a similar number of spikelets m^{-2} to the other NILs and PYLs (**Table 1**). The high number of panicles were derived from vigorous tillering. 'IR64-PYL(7+10)' had already produced more tillers than 'IR64' at 12 days after transplanting and maintained this difference until the heading stage (**Fig. 5**).

Discussion

Using the PYLs developed in this study, we demonstrated that pyramiding two heading QTLs expanded the variation in days to heading in the 'IR64' genetic background in three different experimental fields. The results suggest that PYLs could be used to expand the adaptation of 'IR64' into new rice-growing environments.

Many genetic and molecular studies have elucidated the mechanisms underlying rice heading date and revealed that the combination of *Ghd7*, *Hd5/Ghd8*, and *Hd1* resulted in high natural variation in rice (Zhang *et al.* 2015, 2019). We also identified that pyramiding two QTLs among *Hd5*, *Hd1*, *qDTH11[yp7]* accelerated heading in the 'IR64' genetic background. However, pyramiding the three QTLs did not accelerate heading. Previous studies reported that flowering time was not always determined by additive effects of multiple genetic factors (Lin *et al.* 2000, Reeves and Coupland 2001). The results, therefore, imply an interaction among *Hd5*, *Hd1*, and *qDTH11*[yp7], although the mechanisms involved remains unclear. Or there may be other masked QTLs or genes on heading date in the donor

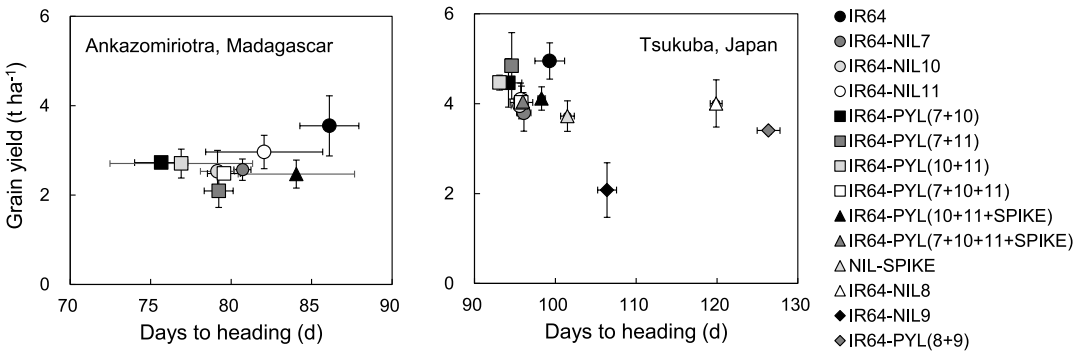


Fig. 4. Relationship between days to heading and grain yield among NILs and PYLs grown in Ankazomiriotra, Madagascar and Tsukuba, Japan. Bars represent standard deviation.

Table 1. Grain yield and yield components for the near-isogenic lines (NILs) and pyramiding lines (PYLs) for days to heading in the 'IR64' genetic backgrounds grown in Ankazomiriotra, Madagascar and Tsukuba, Japan

Lines	Grain yield[a] (t ha^{-1})	No. of panicles (m^{-2})	No. of spikelets (panicle^{-1})	No. of spikelets (m^{-2})	Filled spikelets (%)	Single-grain weight (mg)
Ankazomiriotra, Madagascar in 2019–2020						
IR64	3.5 a[b]	215 b	60 a	13,046 a	85.5 ab	31.7 a
IR64-NIL7	2.6 ab	207 b	53 ab	10,820 a	78.7 b	30.2 b
IR64-NIL10	2.5 ab	267 ab	39 c	10,450 a	82.4 ab	29.6 bc
IR64-NIL11	3.0 ab	227 ab	51 ab	11,551 a	84.7 ab	30.2 b
IR64-PYL(7+10)	2.7 ab	289 a	37 c	10,752 a	90.0 a	28.3 c
IR64-PYL(7+11)	2.1 b	201 b	46 bc	9,099 a	76.0 b	30.3 ab
IR64-PYL(10+11)	2.7 ab	234 ab	47 bc	11,095 a	83.4 ab	29.4 bc
IR64-PYL(7+10+11)	2.5 b	270 ab	38 c	10,240 a	82.3 ab	29.5 bc
IR64-PYL(10+11+*SPIKE*)	2.5 b	222 ab	46 bc	10,308 a	83.7 ab	28.6 c
Tsukuba, Japan in 2020						
IR64	5.0 a	254 ef	106 abc	26,934 bc	69.2 abc	26.8 abc
IR64-NIL7	3.8 ab	270 de	83 defgh	22,510 cde	66.0 abcd	25.6 cdef
IR64-NIL10	4.0 ab	312 bcd	69 gh	21,554 cde	68.5 abc	26.8 abc
IR64-NIL11	4.1 ab	273 e	87 defg	23,735 cde	66.8 abc	26.0 bcdef
IR64-PYL(7+10)	4.5 ab	344 ab	67 h	23,268 cde	72.0 ab	26.9 ab
IR64-PYL(7+11)	4.8 a	354 a	74 fgh	26,226 bcd	71.9 ab	26.2 bcde
IR64-PYL(10+11)	4.5 ab	306 cd	74 fgh	22,689 cde	73.2 ab	27.1 a
IR64-PYL(7+10+11)	4.0 ab	278 cde	76 efgh	21,111 de	70.3 ab	27.1 a
IR64-PYL(10+11+*SPIKE*)	4.1 ab	226 f	90 cdef	20,248 e	78.4 a	25.9 cdef
IR64-PYL(7+10+11+*SPIKE*)	4.0 ab	227 f	93 cde	21,087 de	73.2 ab	26.2 abcd
NIL-*SPIKE*	3.7 ab	220 f	119 a	26,280 bcd	56.4 bcde	25.2 efg
IR64-NIL8	4.0 ab	285 cde	120 a	34,058 a	48.2 cde	24.5 g
IR64-NIL9	2.1 c	231 f	98 bcd	22,748 ce	36.6 e	25.2 fg
IR64-PYL(8+9)	3.4 b	269 de	116 ab	31,373 ab	45.3 de	24.4 g

[a] Data represented by means; n = 3 to 4 replications. [b] Different letters indicate significant differences ($P < 0.05$, Tukey's HSD test) in each environment.

segments or in the IR64 genetic background.

Our study also observed that the absolute differences in days to heading between the earliest PYLs and 'IR64' reduced (12.8 > 10.4 > 6.2 days) in the order of Los Baños, Ankazomiriotra, and Tsukuba. In contrast, the differences between the latest PYL and 'IR64' increased (18.8 < 20.6 < 27.1 days) in the order of Los Baños, Ankazomiriotra, and Tsukuba. These results suggest that shorter day lengths and higher air temperatures in the tropical region can enhance the effects of the pyramided QTLs on early heading, while longer daylengths and lower air temperatures of the temperate region can enhance the effects on late heading.

Contrary to our expectation, *SPIKE* did not improve grain yield in the pyramiding lines with early heading. While *SPIKE* increased the number of spikelets per panicle (Fujita *et al.* 2013), it also decreased the number of panicles per plant, depending on soil fertility (Takai *et al.* 2017, 2019). In this study, 'IR64-PYL(10+11+*SPIKE*)' and

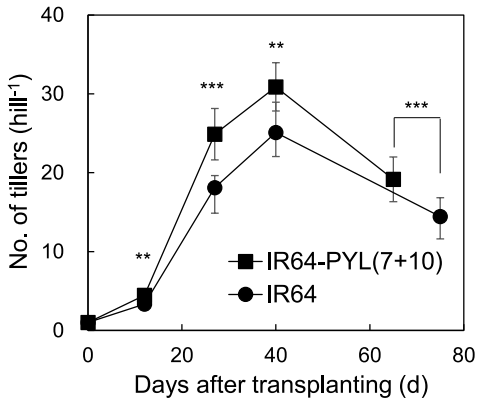


Fig. 5. Changes the number of tillers per hill in 'IR64' and 'IR64-PYL(7+10)' between transplanting and heading when grown in Tsukuba, Japan. Bars represent standard deviation. ** and *** denote significance at the 1% and 0.1% level analyzed with Student's t-test.

'IR64(7+10+11+*SPIKE*)' decreased the number of panicles m^{-2} compared with 'IR64-PYL(10+11)' and 'IR64-PYL(7+10+11)', respectively. This decrease caused no increase in the number of spikelets m^{-2} (**Table 1**). These results suggest that *SPIKE* may not be useful in combination with early heading QTLs. Conversely, 'IR64-PYL (7+10)' headed earlier, produced more tillers, and thus more panicles m^{-2} than 'IR64'. This result indicates the pleiotropic effect of heading QTLs on rice tillering. Previous studies have reported pleiotropic effects with some flowering genes (Tsuji *et al.* 2015, Wang *et al.* 2020). It should be noted that pyramiding *Hd5* and *Hd1* promoted tillering greater than single *Hd5* or *Hd1*. Although further studies are necessary, an enhancement of the flowering signal by the two QTLs may further induce the outgrowth of tiller buds. Consequently, pyramiding *Hd5* and *Hd1* could be useful in mitigating yield decreases due to early heading.

Author Contribution Statement

TT, DF, KS, and NK designed the study. TT, PL, DF, KS, NMR, YT, and EVM performed the experiments. TT and EVM analyzed the data. TT wrote the paper.

Acknowledgments

We are grateful to Dr. Hiroki Saito for his valuable advice on the manuscript. We thank the research technicians and contract workers for their research support at the International Rice Research Institute (IRRI). We also thank research technicians and local farmers for their generous support in the field experiments in Madagascar. This study was financially supported by the Collaborative Breeding Project within the Japan International Research Center for Agricultural Sciences-International Rice Research Institute (JIRCAS-IRRI), and partly funded by the Science and Technology Research Partnership for Sustainable Development (SATREPS), Japan Science and Technology Agency (JST)/Japan International Cooperation Agency (JICA) (Grant no. JPMJSA1608). We would like to thank Editage (https://www.editage.com) for English language editing.

Literature Cited

Chloupek, O. and P. Hrstkova (2005) Adaptation of crops to environment. Theor Appl Genet 111: 1316–1321.

Doi, K., T. Izawa, T. Fuse, U. Yamanouchi, T. Kubo, Z. Shimatani, M. Yano and A. Yoshimura (2004) *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls *FT-like* gene expression independently of *Hd1*. Genes Dev 18: 926–936.

Farooq, M., A. Wahid, D.J. Lee, O. Ito and K.H.M. Siddique (2009) Advances in drought resistance of rice. CRC Crit Rev Plant Sci 28: 199–217.

Fujita, D., R.E. Santos, L.A. Ebron, M.J. Telebanco-Yanoria, H. Kato, S. Kobayashi, Y. Uga, E. Araki, T. Takai, H. Tsunematsu *et al.* (2009) Development of introgression lines of an Indica-type rice variety, IR64, for unique agronomic traits and detection of the responsible chromosomal regions. Field Crops Res 114: 244–254.

Fujita, D., R.E. Santos, L.A. Ebron, Y. Fukuta and N. Kobayashi (2011) Characterization of quantitative trait locus for days to heading in near-isogenic lines with genetic background of Indica-type rice variety IR64 (*Oryza sativa*). Plant Breed 130: 526–532.

Fujita, D., K.R. Trijatmiko, A.G. Tagle, M.V. Sapasap, Y. Koide, K. Sasaki, N. Tsakirpaloglou, R.B. Gannaban, T. Nishimura, S. Yanagihara *et al.* (2013) *NAL1* allele from a rice landrace greatly increases yield in modern *indica* cultivars. Proc Natl Acad Sci USA 110: 20431–20436.

Hori, K., K. Matsubara and M. Yano (2016) Genetic control of flowering time in rice: integration of Mendelian genetics and genomics. Theor Appl Genet 129: 2241–2252.

Izawa, T. (2007) Adaptation of flowering-time by natural and artificial selection in *Arabidopsis* and rice. J Exp Bot 58: 3091–3097.

Khush, G.S. and P.S. Virk (2005) IR varieties and their impact. International Rice Research Institute, Philippines, p. 163.

Lin, H.X., T. Yamamoto, T. Sasaki and M. Yano (2000) Characterization and detection of epistatic interactions of 3 QTLs, *Hd1*, *Hd2*, and *Hd3*, controlling heading date in rice using nearly isogenic lines. Theor Appl Genet 101: 1021–1028.

Mackill, D.J. and G.S. Khush (2018) IR64: a high-quality and high-yielding mega variety. Rice (N Y) 11: 18.

Matsubara, K., K. Hori, E. Ogiso-Tanaka and M. Yano (2014) Cloning of quantitative trait genes from rice reveals conservation and divergence of photoperiod flowering pathways in Arabidopsis and rice. Front Plant Sci 5: 193.

Peng, S. and G.S. Khush (2003) Four decades of breeding for varietal improvement of irrigated lowland rice in the International Rice Research Institute. Plant Prod Sci 6: 157–164.

Reeves, P.H. and G. Coupland (2001) Analysis of flowering time control in Arabidopsis by comparison of double and triple mutants. Plant Physiol 126: 1085–1091.

Shavrukov, Y., A. Kurishbayev, S. Jatayev, V. Shvidchenko, L. Zotova, F. Koekemoer, S. de Groot, K. Soole and P. Langridge (2017) Early flowering as a drought escape mechanism in plants: How can it aid wheat production? Front Plant Sci 8: 1950.

Takai, T., S. Adachi, D. Fujita, Y. Arai-Sanoh, M. Okamura, M. Kondo and N. Kobayashi (2017) Effects of yield-related QTLs *SPIKE* and *GPS* in two *indica* rice genetic backgrounds. Plant

Prod Sci 20: 467–476.
Takai, T., D. Fujita, P. Lumanglas, E.V. Simon, K. Sasaki, T. Ishimaru, H. Asai and N. Kobayashi (2019) *SPIKE*, a quantitative-trait locus, increases rice grain yield under low-yield conditions. Euphytica 215: 102.
Takai, T., M. Sakata, N.M. Rakotoarisoa, N.T. Razafinarivo, T. Nishigaki, H. Asai, T. Ishizaki and Y. Tsujimoto (2021) Effects of quantitative trait locus *MP3* on the number of panicles and rice productivity in nutrient-poor soils of Madagascar. Crop Sci 61: 519–528.
Tsuji, H., C. Tachibana, S. Tamaki, K. Taoka, J. Kyozuka and K. Shimamoto (2015) *Hd3a* promotes lateral branching in rice. Plant J 82: 256–266.
van Oort, P.A.J. and M. Dingkuhn (2021) Feet in the water and hands on the keyboard: A critical retrospective of crop modelling at AfricaRice. Field Crops Res 263: 108074.
Wang, F., T. Han, Q. Song, W. Ye, X. Song, J. Chu, J. Li and Z.J. Chen (2020) The rice circadian clock regulates tiller growth and panicle development through strigolactone signaling and sugar sensing. Plant Cell 32: 3124–3138.
Wei, F.J., Y.C. Tsai, H.P. Wu, L.T. Huang, Y.C. Chen, Y. F. Chen, C.C. Wu, Y.T. Tseng and Y.C. Hsing (2016) Both *Hd1* and *Ehd1* are important for artificial selection of flowering time in cultivated rice. Plant Sci 242: 187–194.
Yano, M., Y. Katayose, M. Ashikari, U. Yamanouchi, L. Monna, T. Fuse, T. Baba, K. Yamamoto, Y. Umehara, Y. Nagamura *et al.* (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. Plant Cell 12: 2473–2483.
Yano, M., S. Kojima, Y. Takahashi, H. Lin and T. Sasaki (2001) Genetic control of flowering time in rice, a short-day plant. Plant Physiol 127: 1425–1429.
Zhang, B., H. Liu, F. Qi, Z. Zhang, Q. Li, Z. Han and Y. Xing (2019) Genetic interactions among *Ghd7*, *Ghd8*, *OsPRR37* and *Hd1* contribute to large variation in heading date in rice. Rice (N Y) 12: 48.
Zhang, J., X. Zhou, W. Yan, Z. Zhang, L. Lu, Z. Han, H. Zhao, H. Liu, P. Song, Y. Hu *et al.* (2015) Combinations of the *Ghd7*, *Ghd8* and *Hd1* genes largely define the ecogeographical adaptation and yield potential of cultivated rice. New Phytol 208: 1056–1066.
Zhang, Y., Q. Tang, Y. Zou, D. Li, J. Qin, S. Yang, L. Chen, B. Xia and S. Peng (2009) Yield potential and radiation use efficiency of "super" hybrid rice grown under subtropical conditions. Field Crops Res 114: 91–98.

Theoretical and Applied Genetics (2021) 134:3397–3410
https://doi.org/10.1007/s00122-021-03909-9

ORIGINAL ARTICLE

From gene banks to farmer's fields: using genomic selection to identify donors for a breeding program in rice to close the yield gap on smallholder farms

Ryokei Tanaka[1] · Sarah Tojo Mandaharisoa[2] · Mbolatantely Rakotondramanana[2] · Harisoa Nicole Ranaivo[2] · Juan Pariasca-Tanaka[3] · Hiromi Kajiya-Kanegae[1] · Hiroyoshi Iwata[1] · Matthias Wissuwa[3]

Received: 20 January 2021 / Accepted: 6 July 2021 / Published online: 15 July 2021

Abstract

Key message **Despite phenotyping the training set under unfavorable conditions on smallholder farms in Madagascar, we were able to successfully apply genomic prediction to select donors among gene bank accessions.**

Abstract Poor soil fertility and low fertilizer application rates are main reasons for the large yield gap observed for rice produced in sub-Saharan Africa. Traditional varieties that are preserved in gene banks were shown to possess traits and alleles that would improve the performance of modern variety under such low-input conditions. How to accelerate the utilization of gene bank resources in crop improvement is an unresolved question and here our objective was to test whether genomic prediction could aid in the selection of promising donors. A subset of the 3,024 sequenced accessions from the IRRI rice gene bank was phenotyped for yield and agronomic traits for two years in unfertilized farmers' fields in Madagascar, and based on these data, a genomic prediction model was developed. This model was applied to predict the performance of the entire set of 3024 accessions, and the top predicted performers were sent to Madagascar for confirmatory trials. The prediction accuracies ranged from 0.10 to 0.30 for grain yield, from 0.25 to 0.63 for straw biomass, to 0.71 for heading date. Two accessions have subsequently been utilized as donors in rice breeding programs in Madagascar. Despite having conducted phenotypic evaluations under challenging conditions on smallholder farms, our results are encouraging as the prediction accuracy realized in on-farm experiments was in the range of accuracies achieved in on-station studies. Thus, we could provide clear empirical evidence on the value of genomic selection in identifying suitable genetic resources for crop improvement, if genotypic data are available.

Introduction

The demand for rice in sub-Saharan Africa (SSA) is increasing steadily, outpacing local supply, and forcing many Africa countries to import increasing amounts of rice from Asia (USDA 2018). This growing shortage in local supply is due to the much lower average yields (2.3 t ha^{-1}) achieved across Africa compared to Asia (4.8 t ha^{-1}) (FAOSTAT 2018). The low grain yields in rice in SSA are caused by a combination of low fertilizer application rates with generally low soil fertility of the highly weathered soils which are typical throughout the region (Saito et al. 2019). Soils like the commonly encountered Oxisols are known to bind phosphorous (P) in forms that are not plant available, causing P to be the most frequent limiting nutrient in rice production in SSA (Saito et al. 2019). VanDamme et al. (2015) highlighted that a cost-efficient partial solution to the soil fertility problem in

Communicated by Andreas Graner.

✉ Matthias Wissuwa
wissuwa@affrc.go.jp

[1] Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-8657, Japan

[2] Rice Research Department, The National Center for Applied Research On Rural Development (FOFIFA), 101 Antananarivo, Madagascar

[3] Crop, Livestock and Environment Division, Japan International Research Center for Agricultural Sciences (JIRCAS), 1-1 Ohwashi, Tsukuba, Ibaraki 305-8686, Japan

SSA would be the development of varieties with improved P acquisition and utilization efficiencies.

The abovementioned general trends for rice production and consumption also apply to Madagascar, the second biggest rice producer in SSA. Fertilizer applications have remained very low (Tsujimoto et al. 2019) despite standard recommendations for NPK fertilizers being in existence for decades. As a result, the national average yield remains below 3 t ha^{-1}, whereas achievable on-farm yields can exceed 7 t ha^{-1} (Saito et al. 2017) and are therefore comparable to tropical regions elsewhere. This large yield gap highlights that conventional on-station breeding approaches that seek to select breeding lines with high yield potential under "ideal" high-input conditions may not produce desired results. The prevalence of traditional rice varieties throughout Madagascar (Minten and Barrett 2008) is a further sign that plant breeding has not properly addressed the needs of the mostly resource-poor smallholder farmers. It is furthermore indicative of specific adaptations to lower soil fertility being present in such traditional varieties, which were found to be more efficient in P acquisition (Mori et al. 2016) and internal P utilization (Wissuwa et al. 2015) or may even show a combination of both desirable traits (Rose et al. 2015). A future breeding program targeting to close this yield gap may move the selection process from highly fertilized breeding stations to fields representing conditions a crop may experience in farmers' fields and should attempt to utilize any adaptive traits of traditional varieties.

That traditional varieties may contain useful genes and alleles for certain traits and may therefore serve as donors to improve such traits in modern breeding populations which have been a chief reason to collect and preserve such varieties in crop gene banks. The largest collection of rice genetic resources with more than 130,000 accessions is stored in the gene bank of the International Rice Research Institute (IRRI). One potential problem of exploiting this resource is its sheer size. Phenotyping thousands of lines will require resources in terms of land and labor that few projects and institutes can manage. It is therefore of importance for the utility of such resources to have enough information associated with accessions to allow for targeted selections of smaller sub-sets of accessions for more detailed phenotypic evaluations. One invaluable step in this regard was the establishment of the publicly available SNP-Seek database (https://snp-seek.irri.org) providing sequence variants and passport data of more than 3000 rice accessions (3 K accessions) of diverse origin and genetic background (Mansueto et al. 2017).

Such genotyping efforts allow for the implementation of genomic selection (Meuwissen et al. 2001) as a tool to identify promising accession from gene banks. Based on phenotypic values available for a subset of the 3 K accessions, their genotypic values can be predicted applying a statistical model that establishes relations between SNP genotype and phenotype. Using this genomic selection model to predict genotypic values of the untested accessions would allow for the identification of potentially promising donor accessions among the entire 3 K set without the need for testing all accessions in the field. This concept was first proposed by Pace et al. (2015) and demonstrated in a large-scale experiment using more than 1,000 sorghum accessions (Yu et al. 2016) and is becoming one of the important applications of genomic selection (Crossa et al. 2017). In a typical genomic selection, genotypes are selected based on having higher predicted genotypic values (PGVs). Tanaka and Iwata (2018) recently proposed an alternative selection criterion termed "expected improvement" (EI). Using simulations based on a diversity panel of rice, they demonstrated that the proposed EI-based selection can identify superior accessions more efficiently compared to the standard PGV-based selection.

Genomic selection has been evaluated in rice (e.g., Onogi et al. 2015; Spindel et al. 2015) and in other crops such as wheat (e.g., Crossa et al. 2010; Heffner et al. 2011) or maize (e.g., Zhao et al. 2012; Cui et al. 2020). Typically, phenotypic evaluations were conducted in well-managed breeding stations under high-input conditions where efforts had been made to maximize heritability by minimizing the environmental variation, as was the case in the proof-of-concept study of Yu et al. (2016). To what extent genomic prediction would be capable of reliably identifying superior accessions under less uniform conditions and in the presence of abiotic stresses, as one may encounter in farmers' fields, has not been investigated.

The logistics of conducting experiments in smallholder farmer fields can be challenging and typically place a size limit on experiments as a result of factors such a field size, presence of gradients, absence of machinery for precise land leveling, distance from laboratories or even electricity. While the entire 3 K set has been phenotyped on the IRRI research farm, such a task would be daunting on a smallholder rice farm in Madagascar. The potential of using a small subset from the 3 K accessions to build a genomic selection model in order to predict the performance of the entire 3 K set would therefore be a game-changer as the genotypic variation present in much larger sets of gene bank accessions would become accessible to applied breeding programs. The present study tests to what extent this approach would be feasible under the low-input conditions in smallholder rice farms in Madagascar where typically no fertilizers other than organic manures or composts are applied. Specifically, our objectives were: (1) to develop a GS model based on field data obtained in farmer's fields in Madagascar, (2) to use the model to predict best accessions among the 3 K set and to import them from IRRI for confirmatory tests in the field, (3) compare the selection based on the "expected improvement" (EI) criterion to the standard PGV-based selection,

and (4) to investigate whether a second cycle of improving the model and prediction accession performance increases the precision of this approach.

Materials and methods

Field experiments and selection methods

Field experiments were conducted in the central highlands of Madagascar over a 3-year period from 2016 to 2019. Experimental field sites used were at elevations between 950 and 1350 m, and experiments were conducted during the rainy season with sowing in November, transplanting in late November–December and harvests in April–May. A common characteristic of all experimental sites was that they were located in farmers' fields, that no fertilizer was applied, which is the predominant farmer's practice in the region, and that no supplementary irrigation was provided beyond what was available from rainfall and rainfed small creeks. Briefly, the flow of experiments was as follows: In the 2016–17 season (year 1), a diversity panel (set 1, $n = 359$) was selected from the set of 3,024 sequenced rice accessions (3 K panel; https://snp-seek.irri.org) publicly available at IRRI and phenotyped at two sites (Anjiro and Behenjy). Based on the phenotypic data, a genomic prediction model was developed and utilized to predict the performance of all 3 K panel accessions. Accessions predicted to perform well were selected together with a random control group, and both ($n = 234$) were imported from the IRRI gene bank and grown together with the original panel at two sites (Anjiro and Ankazo) in the 2017–18 season (year 2). Phenotypic data from these trials were used to update the prediction model, and best-performing accessions not already used previously were selected ($n = 52$) and imported from IRRI for year 3 confirmatory trials in the 2018–19 season.

Genotypes used were selected from the 3 K panel that is publicly available at IRRI. We targeted accessions of the *indica* subpopulation and first selected based on geographical origin. All available accessions from the following countries were selected: Madagascar (57), Sri Lanka (44), Nepal (37), Bhutan (18). The reason why we focused on the latter two countries was to include accessions with possible adaptation to highland environments, while Sri Lankan accessions may contribute some adaptation to problems of soils (e.g., iron toxicity). The remaining 203 accessions were selected to obtain a representative sample of the *indica* gene pool, avoiding closely related accessions using phylogenetic analyses tools provided in SNP-Seek website. From the total 359 accessions imported from IRRI into Madagascar in year 1, 22 belonged to the *japonica* subspecies (from Madagascar and Bhutan), 15 or 12 accessions belonged to the *aus* and *aromatic* subpopulations (from Nepal, Madagascar, Bhutan, respectively), and 2 were classified as *admix*. The remaining 298 accessions belonged to the *indica* group, and in total 29 countries were represented. Hereinafter, experimental details will be provided year by year and a diagram depicting the flow of materials used in evaluations and predictions in different years is given in supplementary Figure S1.

Experiments in year 1 were conducted with 359 accessions at two locations, Anjiro (elevation 950 m, 18°54′01.7″S 47°58′12.4″E) and Behenjy (elevation 1350 m, 19°10′48.3″S 47°29′46.3″E). In addition to low soil fertility and especially P deficiency, growth at the Behenjy site was limited by low temperatures and mild iron toxicity. At both sites, accessions were transplanted in 2-row micro-plots of 2 m length with spacing of 20 cm between and within rows (22 single plant hills per plot). Heading date (HD) was recorded at 50% heading rate for each accession. During harvest, five representative plants per plot were cut, and tiller number and plant height (in cm from the base to the tip of the flag leaf) were recorded. Panicles were separated from straw and straw weight (STW) determined. Panicles were taken to the laboratory and air-dried for a week before total panicle dry weight (TPW) was determined. Straw and panicle weights are given in grams per plant for all experiments.

Experiments in year 2 were conducted in Anjiro, and instead of the cold-affected Behenjy, the warmer site Ankazo at elevation 1150 m (19°40′07.9″S 46°33′53.9″E) was chosen. Plot size was as in year 1, but each site had a second replication in a randomized complete block design. Procedures were similar as in year 1, with the exception that straw of one of the 5 representative plants sampled was taken to the laboratory to determine straw dry weight. The measured moisture content in the straw of this plant was then used to estimate STW of the entire 5-plant sample. The new set of 234 accessions was selected using the predicted genotypic values (PGV) and expected improvement (EI) values from the GBLUP genomic prediction model (Eqs. 5–7) built on the phenotypic record obtained in the year 1. Out of the 234 accessions, 79 were selected for screening of superior accessions for STW or TPW. For this screening, the top 20 accessions were independently listed for each trait (STW and TPW) by using each criterium (PGV and EI) for each environment (Anjiro and Behenjy) and then merged into one list. Details of the selection result, including the overlap between the two criteria, are described in the Result section. Of the remaining 155 new accessions, 68 were chosen randomly as a control group and the remaining 87 were included to enhance the training data for genomic selection, considering the predicted STW and HD (supplementary Figure S2). In short, accessions were selected for early/middle/late predicted HD, and high/middle/low predicted STW. Furthermore, 289 out of the initial 359 accessions were evaluated again in the second year (the remaining 70 accessions were removed due to low adaptation in Madagascar in the year 1

experiment). In total, 523 accessions were evaluated in the field in year 2.

The experiment in year 3 was conducted on a smaller number of accessions, mainly aiming at screening accessions predicted to be superior for TPW. A genomic prediction model for TPW was built on the observed phenotypic values from Anjiro (year 1) and Ankazo (year 2) (Eq. 3). The observations from year 2 of Anjiro were not used for the selection because that field was flooded after a cyclone, and thereafter, the plants were no longer under P-limited condition. Using PGV or EI from the model, 42 accessions were selected by PGV and 45 accessions by EI from the untested subset of the 3 K panel (i.e., non-tested in our previous two years of experiment). Due to an overlap of 35 accessions between the 42 and 45 selected sets, a total of 52 accessions were selected in year 3. In addition to those 52 newly selected accessions, 23 accessions were evaluated as a control group. These 23 were chosen from accessions evaluated in both two previous years. In order to ensure a wide phenotypic (and genetic) variation, these check accessions were selected at equally spaced intervals (every 22nd–23rd accession) from the distribution of estimated genotypic values for TPW (i.e., fitted genotypic values calculated in Eq. 3). Predicted HD was taken into account to avoid selecting very late maturing accessions (not suitable during very dry or cool years). Thus, in total 75 accessions were evaluated in year 3 in Anjiro and Ankazo.

For replicated experiments in year 2 and 3, broad-sense heritability of each trait was estimated based on the model:

$$y_{ij} = g_i + r_j + e_{ij} \tag{1}$$

where y_{ij} is the phenotypic value of the ith genotype in the jth replication, g_i is the genotypic value of the ith genotype (modeled as a random effect), r_j is the effect of the jth replication (modeled as a fixed effect), and e_{ij} is the residual.

Heritability was estimated using

$$H^2 = \frac{V_g}{V_g + \frac{V_e}{R}} \tag{2}$$

where V_g is the estimated genotypic variance, V_e the estimated residual variance, and R the number of replications (therefore $R = 2$ in our case).

Genomic prediction

For each trait, the GBLUP model (VanRaden 2008) was used for predicting or estimating the genotypic values of untested or tested accessions:

$$y = X\beta + Zu + e, u \sim MVN\left(0, G\sigma_u^2\right), e \sim MVN\left(0, I\sigma_e^2\right) \tag{3}$$

where y is a vector of phenotypic values, X is a design matrix for the fixed effect, β is a vector of fixed effects, Z is a design matrix for the random effect, u is a vector of genotypic values, e is a vector of residuals, G is a genomic relationship matrix, σ_u^2 is a genotypic variance, and σ_e^2 is a residual variance. A zero vector and an identity matrix were denoted as 0 and I, respectively. Note that the fixed effect-related terms varied in cases. When applied to the year 1 phenotype data (remember that genomic prediction was separately performed for each location) to select accessions for the year 2 experiment, only an intercept μ was modeled as a fixed effect in the equation. Therefore, the following equation was used:

$$y = 1\mu + Zu + e \tag{4}$$

where 1 represents vector of ones. On the other hand, when applied to the year 1 Anjiro and year 2 Ankazo phenotypes to select accessions for the year 3 experiment, we had two fixed effects, and therefore, Eq. (3) with $\beta^T = \left[\beta_1 \ \beta_2\right]$ was used, where β_1 is the intercept for year 1 Anjiro and β_2 is the intercept for year 2 in Ankazo. The predictions of the genotypic values (u) were then used for the PGV-based genomic selection.

When there were more than two environments, the observed phenotypic values were scaled with mean zero and variance one for each environment, before applying the model. This is because the raw observations were distributed in largely different ranges even for the same trait, and therefore, single variance component should not be assigned. This procedure might imply that the two location has similar heritability, which is unlikely but acceptable assumption for keeping our model simple.

A genomic relationship matrix (G matrix) was calculated by using the A.mat function implemented in rrBLUP package, without using the shrinkage option (Endelman 2011). The 404 K core SNPs dataset was downloaded from the IRRI SNP-Seek website (https://snp-seek.irri.org/_download.zul). SNP having more than 5% missing data or a minor allele frequency below 2.5% was removed, and finally 186,229 SNPs for 3,024 accessions were extracted. The remaining missing states were imputed by using Beagle v.4.1 (21Jan17.6 cc; Browning and Browning 2016).

Prediction accuracy was evaluated by tenfold cross-validation (*i.e.*, validation based on 90% of randomly chosen accessions as training data and 10% of remaining accessions as test data) within each year and location, and results are shown as the Pearson's correlation coefficient between predicted and observed values. Each tenfold validation was repeated 10 times, and the average prediction accuracy and its standard deviation were estimated. Note that this validation was not performed for year 3 phenotype data because the number of accessions was too small.

In addition to the cross-validation, the prediction accuracy for year 1 and year 2 experiments was also evaluated. For this validation, year 1 phenotype data were used to predict year 2 phenotypes. Accuracies were evaluated separately for overlapping accessions (*i.e.*, 289 accessions planted in both years) and non-overlapping accessions (*i.e.*, 234 newly added accessions in year 2).

Expected improvement (EI)

In the previous study by Tanaka and Iwata (2018), EI of the *i*th accession was defined on the GBLUP model as follows:

$$EI_i = \left(m_i - M\right)\Phi\left(z_i\right) + s_i\phi\left(z_i\right) \tag{5}$$

$$z_i = \frac{m_i - M}{s_i} \tag{6}$$

where m_i is the mean predicted genotypic value (i.e., PGV), s_i is the standard deviation of the predicted genotypic value, and M is the maximum estimated genotypic value among the observed genotypes. Further, ϕ and Φ are the probability density function and the cumulative density function of the standard normal distribution, respectively.

As its definition, this EI criterion becomes larger when PGV or its standard deviation (i.e., prediction uncertainty) is larger, as higher standard deviation increases the chance of large genetic gain from the current best accession. Because EI is proportional to PGV given a prediction uncertainty, EI and PGV are highly related and top accessions in terms of EI or PGV may overlap.

There would be a few different ways to approximate or estimate the standard deviation of the predicted genotypic values. In this study, Eq. (4) used in the year 1 to year 2 selection was implemented using the BGLR package with 6000 iterations and 1200 burn-in period by sampling once per five iterations (Perez and de los Campos 2014). When using the Bayesian inference via the BGLR package, posterior standard deviation was returned in the summary object, calculated by using the MCMC samples after the burn-in period. Meanwhile, when Eq. (3) is used for the selection from year 2 to year 3, the mixed.solve function in the rrBLUP package was executed for solving the model via the restricted likelihood-based approach. By considering a conditional distribution of genotypic values given phenotypic values ignoring the fixed effects, standard deviations of the genotypic values were approximated as follows (Bishop 2006):

$$s = \left[\, s_1 \ \cdots \ s_{3024} \,\right]^T \approx diag\left[\left(Z^T Z\frac{1}{\hat{\sigma}_e^2} + G^{-1}\frac{1}{\hat{\sigma}_u^2}\right)^{-1}\right] \tag{7}$$

where $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$ are the estimated genetic and residual variance, respectively, and the *diag* operator takes diagonal elements and makes a vector whose *i*th element is the (*i*, *i*)-element of the matrix inside the bracket. The above two methods are mathematically not equivalent (since the former is defined from a Bayesian perspective and the latter is from a frequentist perspective) but having a similar meaning, and therefore they will result in very close values. The former MCMC-based method is theoretically straightforward, as EI had been inspired by the Bayesian optimization algorithm. However, the latter was used when there were more than two fixed effects (i.e., when using year 1 + year 2 data) because calculation was faster when using rrBLUP.

Phenotypic and genetic correlation analysis

As the objective of our study was to select accessions showing better performance across low-fertility fields in Madagascar, the GBLUP model was applied to predict genotypic values across sites. Thus, the genotype-by-site (GxS) interaction was not explicitly accounted for in the model (*i.e.*, it is included in the residuals). Nonetheless, it is of interest to evaluate the strength of GxS and genotype x year (GxY) observed in our experiment and for that purpose phenotypic and genetic correlations were calculated for every pair of year-site combination in year 1 and 2. Data from year 3 were excluded from this analysis because only 23 accessions overlapped between year 3 and other datasets.

Phenotypic correlations were calculated using Pearson's correlation coefficients, by using the accessions whose phenotypic observations are available in both year-site combinations. For the genetic correlation, the GBLUP model was fitted in each year-site combination by using Eq. 4, to calculate estimated genotypic values of the 3 K accessions. For the subset of accessions tested in each year-site combination, we then calculated genetic correlations as Pearson's correlations.

Results

Selection result for each trait

For the selection from year 1 to year 2, the GBLUP model (Eq. 4) was applied to each site (Anjiro and Behenjy) and for each trait (STW and TPW) independently, and PGV and EI were calculated for all untested accessions to select top 20 accessions for each criterion-site-trait combination. As there was a large overlap in the top accessions between the two criteria, in total 98 accessions was listed as final selection candidates. However, due to unavailability of seeds and other technical issues, the final number of newly

selected and phenotyped accessions were 41 for TPW and 40 for STW. Similarly, the selection based on PGV and EI from year 2 to year 3 showed a large overlap, and in total 52 accessions were selected for TPW. We did not consider STW at that time. In this section, results are shown by merging the two criteria and the two sites of the training data. Overlap and difference among criteria and sites will be described later.

The selected group had a predicted average STW that was 78% (Anjiro) and 45% (Behenjy) higher compared to the average 3 K predicted values (Fig. 1a). For TPW, the predicted differences were smaller, about 21% and 19% higher in Anjiro and Behenjy, respectively. The control group ("Control" in Fig. 1a) medians were slightly higher compared to the 3 K average, but the overall distribution was similar (Fig. 1a; supplementary Figure S2). The group of accessions repeated in year 2 ("Repeated" in Fig. 1a) had similar predicted STW compared to the control and 3 K average, while their predicted TPW were slightly higher, particularly in Anjiro.

Figure 1b shows the actual performance in year 2 of the selected group *versus* the control and repeated groups. At both sites, TPW and STW were significantly higher in the selected ("Selected" in Fig. 1b) compared to the control group ("Control" in Fig. 1b) and this difference was more pronounced for STW. The repeated group of accessions ("Repeated" in Fig. 1b) performed similarly to the random control group for STW, while for TPW it contained a high number of well-performing accessions. In Anjiro, this led to significantly higher TPW compared to the control group (Fig. 1b).

Selections from year 2 to year 3 were made for superior TPW using an updated prediction model that took year2 data from Ankazo ($n=523$) into consideration in addition to year 1 data in Anjiro ($n=359$) (supplementary Figure S1). The selected group ("Selected" in Fig. 2a) significantly outperformed the control group ("Control" in Fig. 2a) in Ankazo showing 22% higher observed TPW (Fig. 2a). In Anjiro, the average TPW was 14% higher, but due to large variation observed within selected and control groups this difference was not significant.

Difference between selection criteria and sites to train the model

Results from prediction in year 1 validated in year 2, in which two sites (Anjiro and Behenjy) of training data were separately considered, were used to explain the difference among the selection criteria and sites. PGV and EI calculated from our model were positively correlated within a site and trait (supplementary Figure S3), and therefore the top accessions selected by these criteria largely overlapped. From year 1 to year 2, we listed the top 20 accessions for each trait and site predicted by the two criteria. If no overlap existed between the two criteria, 40 accessions would have been listed in each site for each trait. However, the overlap between criteria reduced accessions to 24 for STW in Anjiro (*i.e.*, 16 overlaps between PGV and EI) and to 28 in

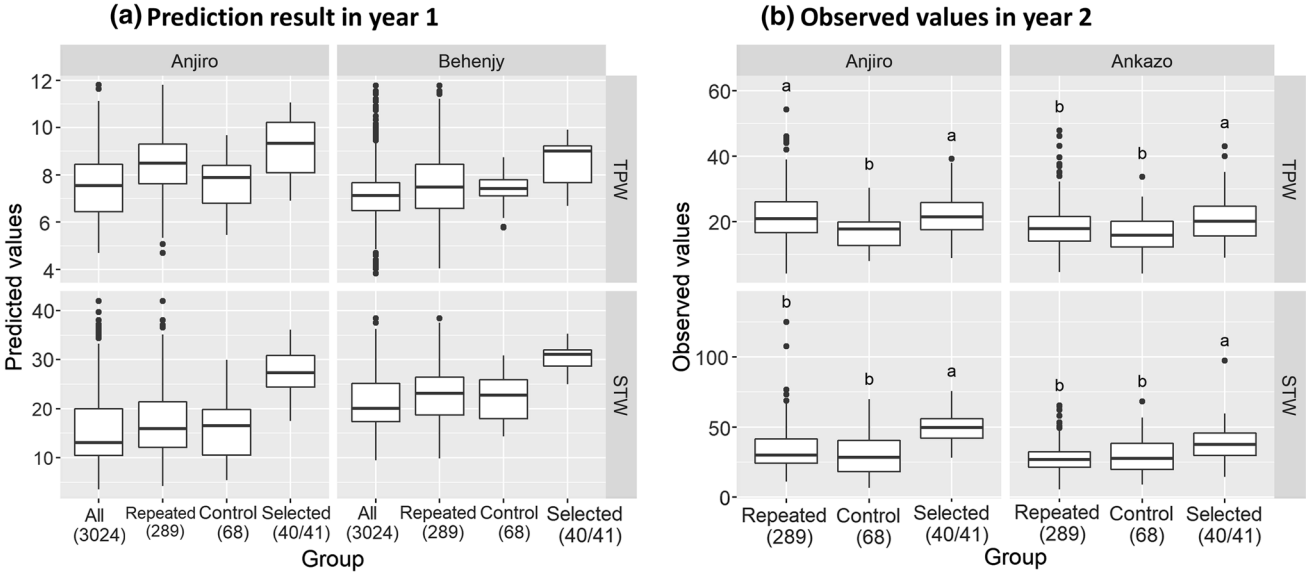


Fig. 1 Selection result from year 1 to year 2. **a** Predicted genotypic values based on year 1 phenotypic values. **b** Observed phenotypic values in year 2. Selected accessions ($n=40$ for total straw weight, STW; $n=41$ for panicle dry weight, TPW) includes all accessions selected by predicted genotypic values (PGV) and expected improvement (EI), based on both Anjiro and Behenjy phenotypic values. For the observed values, Tukey HSD was applied for each combination of trait and site. Group with label "a" has significantly larger average than the group with label "b"

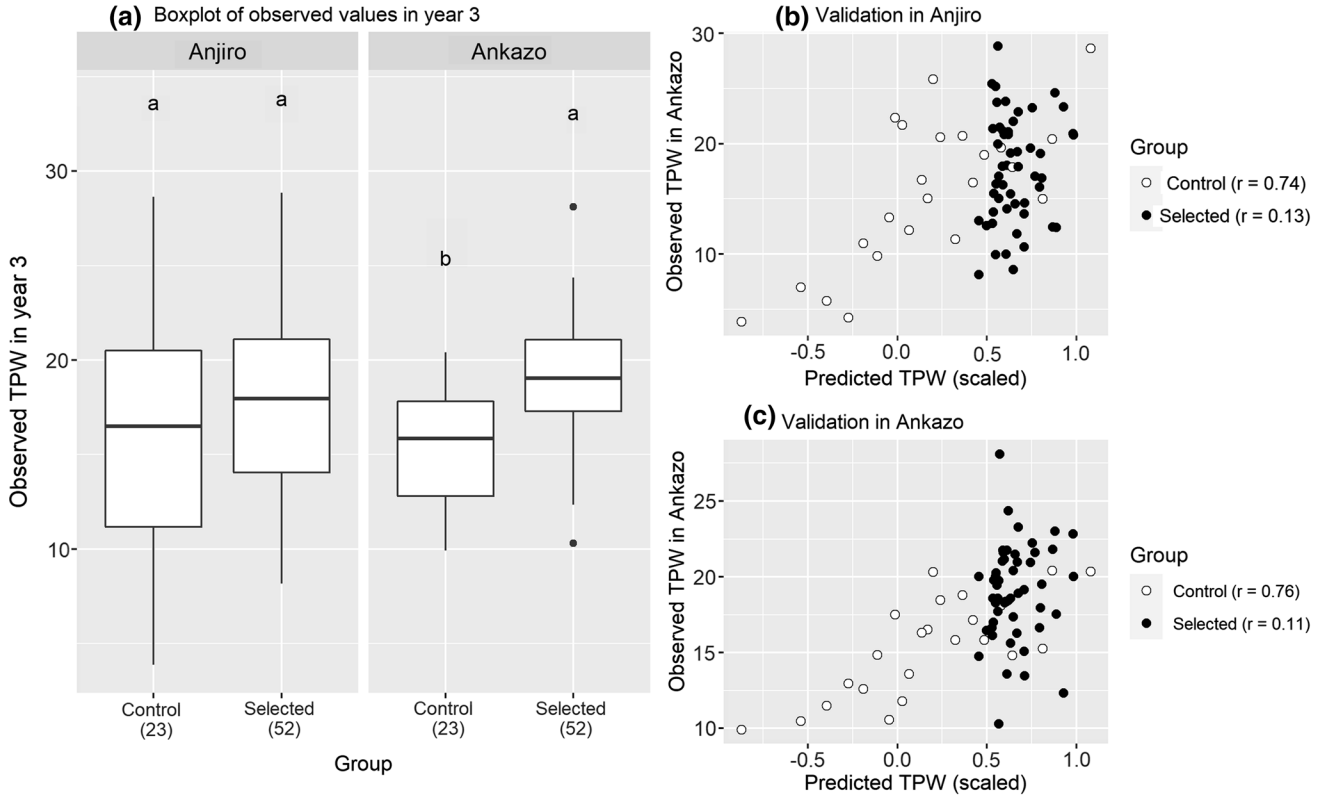


Fig. 2 Selection result for total panicle dry weight (TPW) in year 3. Selection was based on TPW phenotype data in Anjiro year 1 and Ankazo year 2. **a** Boxplot of the observed TPW grouped by check ($n=23$) and selected ($n=52$) accessions. **b** Predicted and observed values in Anjiro. **c** Predicted and observed values in Ankazo

Behenjy (12 overlaps between PGV and EI), giving a total of 49 accessions across sites (3 overlaps between locations). Similarly, for TPW, 25 accessions were selected in Anjiro (15 overlaps between PGV and EI) while 26 accessions in Behenjy (14 overlaps), giving a total of 51 accessions (*i.e.*, no overlap between locations). There were 2 overlaps between the 49 and 51 accessions for STW and TPW.

Because some of those 98 accessions were not available for distribution at IRRI, the final number of accessions tested was 41 for TPW and 40 for STW. Figure 3 summarizes the overlap of those 40 and 41 accessions. For example, there were 15 accessions newly evaluated according to the selection by PGV based on the Anjiro phenotype in year 1, while 9 accessions out of the 15 were also selected by EI based on the same phenotype data. Different accessions were selected when a different site was used as a training data. For the above example, only 1 common accession was selected based on Anjiro and Behenjy phenotype data, even though the same criterium (PGV) was used for the selection. Note that the total number of newly selected accessions based on PGV or EI in year 2 was 79, because there were 2 overlaps between the 41 and 40 accessions for TPW and STW.

Figure 4 shows observed phenotypic values in year 2, grouped by selection criteria and sites of the training phenotypic data. Significant differences were not detected for any trait and testing environment (Tukey's HSD at $P<0.05$). However, there was a consistent trend suggesting selection based on Behenjy TPW and STW phenotype data resulted in better performance in Ankazo (Fig. 4b, d). For the second site (Anjiro), this trend was not consistent: Higher TPW was realized when the Anjiro phenotype was used for the selection, while higher STW was realized from selections based on Behenjy phenotype data (Fig. 4a, c).

The PGV and EI for year 2–3 selection for TPW were compared as well. As described in the Materials and Methods section, 42 and 45 accessions had been selected based on PGV and EI, respectively, with 35 accessions overlapping. Due to this large overlap, no difference between the two criteria was identified (supplementary Figure S4).

Prediction accuracy

Prediction accuracy is an important factor for the success of genomic selection. Table 1 shows prediction accuracies estimated using tenfold cross-validation on year 1 or 2 phenotype data. Accuracy ranged from 0.29 for TPW in year 1 Anjiro to 0.75 for HD in year 2 Anjiro. The average accuracy for each trait across four year-by-site combinations was

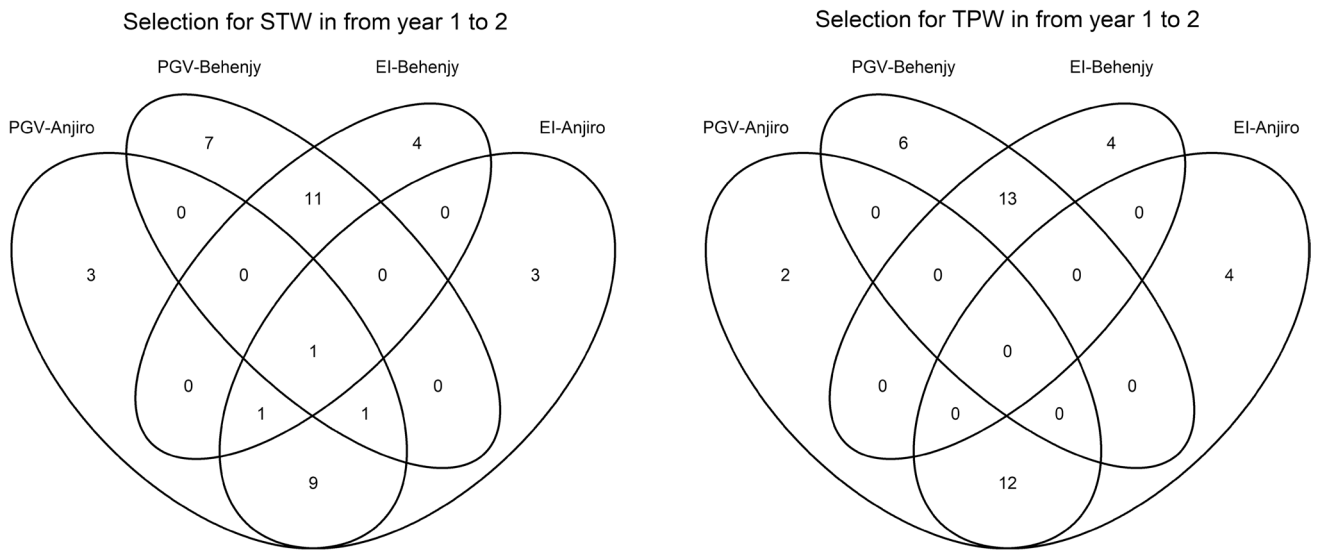


Fig. 3 Overlap among the combination of selection methods and sites of the training data for total panicle dry weight (TPW) and PGV weight (STW), respectively. There was a large overlap between the two selection methods when applied to the same site

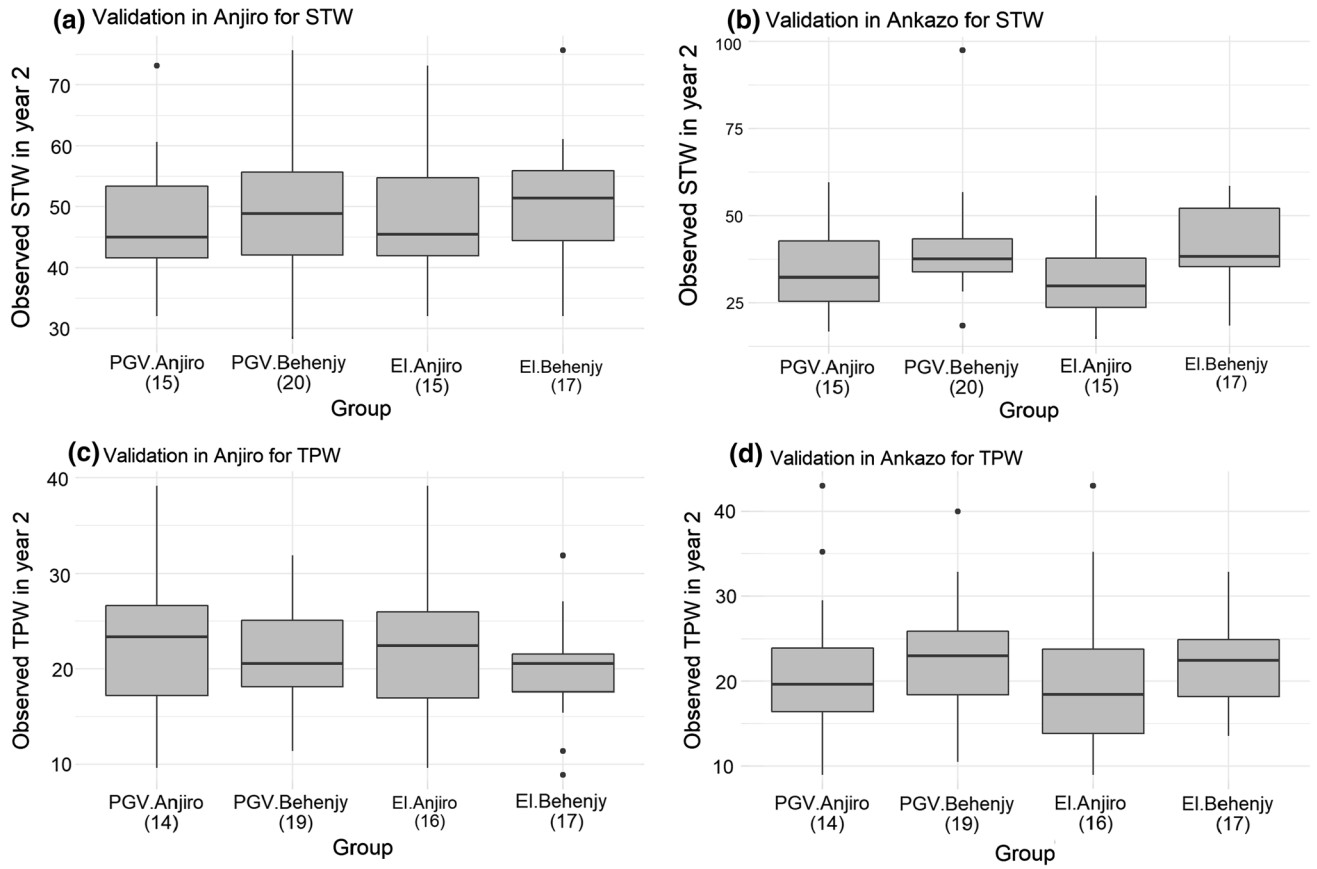


Fig. 4 Observed phenotypic values grouped by selection criteria and site of the training phenotype data. There was no significant difference among the four groups in any figure

Table 1 Estimated accuracy using tenfold cross-validations. Numbers in brackets are standard deviations based on 10 replications

Trait	Year 1		Year 2	
	Anjiro	Behenjy	Anjiro	Ankazo
STW	0.61 (0.011)	0.47 (0.016)	0.57 (0.005)	0.50 (0.045)
TPW	0.29 (0.018)	0.43 (0.016)	0.37 (0.006)	0.45 (0.008)
HD	0.67 (0.011)	0.62 (0.008)	0.75 (0.005)	0.75 (0.003)

lowest for TPW ($r=0.38$), intermediate for STW ($r=0.53$), and highest for HD ($r=0.70$). It is interesting that the accuracies for STW did not show a clear improvement from year 1 to 2, although the size of the dataset changed from $n=359$ to $n=523$ as new accessions had been added in year 2.

Although cross-validation is a common way to estimate the accuracy, it is more interesting and important to validate the prediction result from year 1 using the phenotypic values in year 2. Table 2 summarizes the prediction accuracies using year 1 phenotype as training data and year 2 phenotype as test data. The accuracies for accessions not included in the training dataset were as high or even higher than for repeatedly phenotyped accessions. For example, accuracy for STW in year 2 Anjiro based on the Anjiro phenotype in year 1 was 0.46 for accessions having been repeatedly phenotyped in year 1 and year 2 compared to 0.63 for newly included accessions. This is a surprising but promising result, because it indicates that predictions were accurate even for accessions lacking phenotypic data in a training dataset. The ranking of the prediction accuracy for different traits was identical (HD > STW > TPW) to the cross-validation result.

The prediction accuracy of TPW realized in year 3 based on year 1 and year 2 phenotype data (Eq. 3) was also tested even though the number of accessions evaluated in year 3 was small. It was found that the accuracy for the control group was high ($r=0.74$ in Anjiro, $r=0.76$ in Ankazo), while for the newly selected accessions it was low ($r=0.13$ in Anjiro, $r=0.11$ in Ankazo) (Fig. 2b, c). This may be because newly selected accessions were genetically less diverse compared to the control group, which was selected (from the training set) to contain a wide range of predicted performance. In fact, the accession with the highest predicted TPW was from the control set and it performed well in Anjiro but was surpassed by many newly selected accessions in Ankazo.

Phenotypic and genetic correlation

Phenotypic correlations for TPW ranged from $r=0.10$ (year 1 at Anjiro with year 2 at Ankazo) to $r=0.31$ (year 1 at Behenjy with year 2 at Ankazo) and on average $r=0.21$ (supplementary Table S 2). Correlation within sites ($r=0.23$ between year 1 and 2 at Anjiro) were generally not higher than between sites (*e.g.*, $r=0.30$ between Anjiro and Ankazo in year 2), possibly indicating that GxS effects were not stronger than GxY (genotype-by-year) effects. Similarly, phenotypic correlations for STW within sites between years and between sites within years appeared to be similar (e.g., $r=0.42$ between years 1 and 2 at Anjiro compared to $r=0.42$ between Anjiro and Behenjy at year1 or $r=0.43$ between Anjiro and Ankazo in year 2). Phenotypic correlations for HD were moderate to high ranging from $r=0.64$–0.92, confirming that HD is a highly heritable trait for which site and year effects are expected to be low.

As expected, genetic correlations were higher than phenotypic correlations. Genetic correlations ranged from $r=0.22$–0.64 for TPW, from $r=0.44$–0.69 for STW, and from $r=0.67$–0.93 for HD (supplementary Table S 2). Interestingly, for all three traits, the lowest genetic correlation was observed between Anjiro year 1 and year 2 at Ankazo, and the highest genetic correlation was observed between Anjiro and Ankazo in year 2. This coheres with phenotypic correlations, suggesting that GxS may have been weak and that some GxY interaction affected outcomes.

Table 2 Prediction accuracy of year 2 performance from year 1 training data in accessions that were either present in both years (repeated) or newly added (new) in year 2 based on their predicted superior performance

Trait	Test set	From Anjiro to Anjiro	From Anjiro to Ankazo	From Behenjy to Anjiro	From Behenjy to Ankazo
STW	Repeated ($n=289$)	0.46	0.25	0.40	0.26
	New ($n=234$)	0.63	0.46	0.63	0.49
TPW	Repeated ($n=289$)	0.20	0.13	0.24	0.30
	new ($n=234$)	0.22	0.10	0.24	0.24
HD	Repeated ($n=289$)	0.73	0.65	0.73	0.72
	New ($n=234$)	0.68	0.69	0.74	0.76

Best combination of training and testing site was highlighted by underbars

Discussion

Gene banks are considered a reservoir of untapped allelic variants that could improve the adaptation of crops to biotic and abiotic stresses (Brar and Khush 2018; McCouch et al. 2012). One of the biggest single crop gene banks is housed by IRRI, containing 132,000 accessions of cultivated *Oryza sativa, Oryza glaberrima*, and their wild relatives (https://www.irri.org/international-rice-genebank). Possibly the largest obstacle in exploiting these genetic resources in crop improvement is the lack of a systematic approach in evaluating this resource (McCouch et al. 2012). For most applications, it is simply not feasible to test all stored accessions, and sufficient information allowing for the selection of a representative subset typically does not exist. To overcome this limitation, IRRI and partners have selectively sequenced a cross section of 3024 of their *O. sativa* resources and made sequence data and seed publicly available (Mansueto et al. 2017).

Our general objective is to utilize this resource in breeding for adaptation to low soil fertility as one would encounter in small holder farmers' fields in Madagascar and sub-Saharan Africa, where low soil fertility is compounded by insufficient fertilizer inputs (Tsujimoto et al. 2019; Saito et al. 2019). Most breeding programs are conducted on-station under high-fertility conditions, and this difference between target and selection environment is a potential obstacle in identifying new donors, loci, and alleles improving on-farm grain yield. Screening for new donors should therefore be conducted under target environment conditions; however, phenotyping all 3 K accessions in farmer's fields in several locations is a formidable task that may not be an ideal use of resources, given the inherent variability in unfertilized farmer's fields. Instead, one may test a subset of accessions and develop a genomic prediction model to predict the performance of the untested accessions, then only further test the best predicted accessions in the field. To what extent such an approach is feasible in selection for superior performance in off-station trials conducted in farmer's fields was the specific objective of this study.

Predicting and selecting superior gene bank accessions

The prediction accuracy of year 2 performance from year 1 training data differed considerably between traits with HD having the highest accuracy with an average of around 0.71 (Table 2) which was followed by the accuracy for shoot weight (ranging from 0.25 to 0.63) and TPW (0.10–0.30). This order matched the order for estimated broad-sense heritability, which was $H^2 = 0.95–0.98$ for HD, $H^2 = 0.62–0.73$ for STW and $H^2 = 0.56–0.61$ for TPW (data not shown). Such a positive effect of increasing heritability on prediction accuracy is to be expected (Daetwyler et al. 2008) and has indeed been reported for rice in an empirical study (Spindel et al. 2015).

Compared to the difference in prediction accuracy between traits, the effect of sites was minor. Generally training data from the central Behenjy location was better able to predict TPW in Anjiro (East) and Ankazo (mid-West) compared to data from Anjiro, but for shoot weight and HD, differences between sites were small and inconsistent. The poor predictability of Ankazo TPW from year 1 Anjiro training data could have been due to high GxS or GxY effects. Although the low phenotypic and genetic correlations suggested the presence of GxS interactions for TPW, correlations between sites were in a similar range compared to correlations between years within sites. This implies that the performance of the accessions was under a shared genetic control among sites, particularly at Anjiro and Ankazo and that year effects may have been strong in same cases as for Anjiro year1, which was an untypically dry year during the early part of the season. Consistent GxS effects on the other hand appear to have been weak and this was expected considering all three sites in this study were P deficient low-input sites. Behenjy in year 1 was additionally limited by low temperatures, especially during the grain filling phase of late-maturing accessions. This known effect was purposely eliminated in year 2 by a) omitting very late maturing accessions as not adapted to Malagasy conditions, and b) selecting screening sites (Anjiro and Ankazo) for years 2 and 3 that do not experience this low temperature limitation.

Prediction accuracies were estimated for the set of accessions repeated in both years in comparison with the newly evaluated set (selected based on predictions). It may be expected that prediction accuracies would be higher for the repeated set as these accessions made up the training set; however, for HD and panicle weight no difference was observed, while the newly evaluated set had higher accuracies for shoot weight (Table 2). This may reflect a slightly narrower phenotypic variation within the repeated set compared to the newly evaluated set (supplementary Figure S5), which could have resulted from having omitted accessions having shown poor adaptation to local conditions in the first year.

Realized prediction accuracies for total PWT were in the range of 0.24–0.30 for training data from Behenjy and of 0.20–0.22 for Anjiro (in Anjiro). Comparable accuracies of $r = 0.29$ for panicle weight were reported from on-station trials conducted with a similar size training set at CIAT, Colombia, by Grenier et al. (2015). Similarly, on-station evaluations done at IRRI, Philippines, reported accuracies of up to $r = 0.31$ for grain yield (Spindel et al. 2015), whereas

higher accuracies of $r = 0.54$ and 0.63 were detected for days to flowering in the wet and dry seasons, respectively. Conducting the phenotypic evaluations in farmers' fields without fertilizer addition, mechanic land leveling, nor reliable irrigation therefore was able to provide a training dataset that could be used to predict field performance.

Prediction criteria EI versus PGV

A simulation study showed that EI-based selection can discover superior accessions more efficiently than the usual PGV-based selection (Tanaka and Iwata 2018). We intended to test whether this proved correct in this study; however, no significant difference between EI and PGV in the ability to identify superior accessions was detected. One possible reason was the large training dataset ($n = 359$ in year 1, increasing to $n = 523$ in year 2) used here. Tanaka and Iwata (2018) argued that the difference between EI and PGV may decrease as the size of the training dataset increases, although their simulation did not show that trend. Since minimizing the training set size was not an objective of the current study, one likely outcome of employing a sizable training set was that prediction uncertainties were rather low, and as a result, EI and PGV selected a very similar set of accessions. Given our experimental/selection pipeline, it was difficult to make inference regarding the power of EI to provide an efficient genomic selection based on uncertain prediction result generated from smaller field phenotyping experiments.

A similar question related to maximizing selection gain given a limited budget is related to benefits of replicated phenotyping experiments. In year 1, field phenotyping was done without replication while year 2 experiments were replicated once at each site. Estimating the accuracy in both years through tenfold cross-validations indicated improvements of accuracies for TPW and HD in the year 2 dataset but these were minor, while the standard deviations of the accuracies were wider in year 1. Thus, predictions based on a non-replicated experiment in year 1 were not markedly inferior to those from the repeated year 2 experiments. This observation would match simulations of the benefits of replicated experiment versus increasing the size of the training set which concluded that adding replications may not be as effective as increasing training set size (Lorenz 2013).

Optimization of the training set and the field experiment

In this study, we selected 359 initial accessions based on their origin, assuming a focus on accessions whose origins are in potentially similar environments to low-input medium elevation lowland paddy fields in Madagascar may increase chances of identifying superior performing one. To what extent this may have materialized is summarized in the form of group means by country of origin in supplementary Table S1. For TPW in Behenjy, the Madagascar group was superior followed by the group from Lao, whereas Indonesian and Sri Lankan accessions performed rather poorly. This is likely attributable to sensitivity to lower temperatures at the higher elevation site of the latter groups, which was also evident from the delayed heading of between 14.1 and 18.8 days in Behenjy relative to Anjiro. At the warmer Anjiro site, the country of origin did not have an effect on TPW. At both sites, the groups from Indonesia and Lao had highest straw weights, which could indicate that Indonesian accessions may have some adaptation to low soil fertility despite their low TPW.

A different approach in assembling the initial training set would be to aim for maximizing the prediction accuracy based on statistical methods such as PEV (prediction error variance) or CD (coefficient of determination) (Rincent et al. 2012; Akdemir et al. 2015; Akdemir and Isidro-Sanchez 2019). Further, recent studies empirically showed that accessions having higher *U*-value (upper bound of reliability) tend to show an extreme predicted value which can improve prediction accuracy (Yu et al. 2016, 2020) and contribute to selection of superior genotypes.

This optimization of the training set becomes a more complex but important problem when planning field experiments in multiple locations (i.e., multi-environmental trial). We tested the same set of accessions in multiple locations. This is a straightforward experimental design which allows us to calculate phenotypic correlations among sites to evaluate the strength of GxS interaction. However, when genomic selection is used, it is no longer necessary to evaluate the same set of accessions across sites. Even though no accession is shared among environments, both GBLUP model and multi-environment genomic prediction models can be applied. Previous studies showed that multi-environment genomic prediction models can significantly improve the prediction accuracy from one environment to another (Burgueño et al. 2012; Cuevas et al. 2017; Guo et al. 2013; Lopez-Cruz et al. 2015; Mageto et al. 2020). This implies that it might be better to evaluate different sets of accessions in different environments, which is suggested as a "sparce test" in Jarquin et al. (2020).

In practice, accuracy in predicting the overall population (in our case, all 3 K accessions) is not necessarily the most important criterion for choosing an approach in genomic selection. For example, little is gained by a model predicting poorly performing accessions with high accuracy if this accuracy is not equally matched at the desired end of the distribution. The accuracy of predicting best performers has been assessed by a few studies (Ornella et al. 2014; Blondel et al. 2015), but the approach considered by these studies has not been commonly employed. It is therefore

not resolved whether assembling a training set to maximize prediction accuracy is a better approach compared to assembling a training set based on anticipated adaptation due to geographical similarity.

While our study was not designed to answer above question, one may consider the evidence based on progress made in selecting superior performing accessions over the 2-year selection period. In predictions of the performance of 3 K accessions based on the model developed from year 1 data, a very high proportion of accessions from the year 1 training set were among the top accessions. Of the 60 top accessions (2% of the 3 K set) for STW in Anjiro, 29 had been part of the training set (8.1% of the training set). This increased to 42 accessions (11.7% of the training set) for predicted total panicle weight at Behenjy. This higher than expected proportion of top predicted accession in the training set may be caused by a shrinkage of the genomic prediction (i.e., predicted values of the untested accessions tend to be biased toward the population mean), however, Fig. 1b indicates that many of the best-performing accessions in year 2 were from the repeat set. The performance of accessions in year 3 (Fig. 2b, c) further indicated that we were able to successfully predict and select superior accessions despite having a low prediction accuracy overall. It is interesting to note that two accessions have been chosen as donors in the Malagasy lowland rice breeding program. Donor IRIS 313–11949 (subpopulation: ind1A, origin: China) was part of the accessions newly selected in year 2. It showed 6th highest predicted TPW based on the Behenjy site data in year 1, and it was recommended by both EI and PGV. This accession subsequently proved to be among the top 3% performers for TPW across both sites in year 2, while having average STW and medium-early maturity. The second donor (IRIS 313–7832) was included in the initial set of 359 accessions (subpopulation: ind1B, origin: unknown) and was among the top 3 accessions at Ankazo but performed less well in Anjiro.

Conclusions

How to accelerate the utilization of gene bank resources in crop improvement is an unresolved question and here we empirically tested whether a genomic prediction approach could aid in the selection of promising donors. Despite having conducted phenotypic evaluations under challenging conditions on smallholder farms, our results are encouraging as the prediction accuracy realized in these on-farm studies was in the range of accuracies achieved in on-station or even pure simulation studies. Thus, we could provide clear empirical evidence on the value of genomic selection in identifying suitable genetic resources for crop improvement, if genotypic data are available. The approach taken by IRRI and partners of making such data publicly available for at least 3,024 of their gene bank accessions represents an important step toward more efficient utilization of these gene bank resources.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-021-03909-9.

Acknowledgements This research was supported by the Science and Technology Research Partnership for Sustainable Development (SATREPS), Japan Science and Technology Agency (JST)/Japan International Cooperation Agency (JICA)–Grant No. JPMJSA1608. The contribution of James DM King in conceiving this study is gratefully acknowledged. Authors thank the IRRI gene bank for providing seeds of accessions used.

Author contribution statement RT and MW designed the study and wrote the manuscript, RT implemented the GP model, RT, HI, and MW discussed results, HKK processed SNP data, and STM, MR, HNR, JPT conducted field experiments and collected and summarized all data.

Funding This research was supported by the Science and Technology Research Partnership for Sustainable Development (SATREPS), Japan Science and Technology Agency (JST)/Japan International Cooperation Agency (JICA)—Grant No. JPMJSA1608.

Availability of data and material Genotypic data used in the study are publicly available at https://snp-seek.irri.org/_snp.zul.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Akdemir D, Sanchez JI (2019) Design of training populations for selective phenotyping in genomic prediction. Sci Rep 9:1446

Akdemir D, Sanchez JI, Jannink JL (2015) Optimization of genomic selection training populations with a genetic algorithm. Genet Sel Evol 47:38

Blondel M, Onogi A, Iwata H, Ueda N (2015) A ranking approach to genomic selection. PLoS ONE 10:e0128570

Brar DS, Khush GS (2018) Wild relatives of rice: a valuable genetic resource for genomics and breeding Research. In: Mondal T, Henry R (eds) The Wild oryza genomes compendium of plant genomes. Springer, pp 1–25

Bishop CM (2006) Pattern recognition and machine learning, vol 128. Springer
Browning BL, Browning SR (2016) Genotype imputation with millions of reference samples. Am J Hum Genet 98:116–126
Burgueño J, de Los CG, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci 52:707–719
Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de Los CG, Burgueño J, González-Camacho JM, Pérez-Elizalde S, Beyene Y, Dreisigacker S, Singh R, Zhang X, Gowda M, Roorkiwal M, Rutkoski J, Varshney RK (2017) Genomic selection in plant breeding: methods, models and perspectives. Trends Plant Sci 22:961–975
Crossa J, Campos Gde L, Pérez P, Gianola D, Burgueño J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan J, Arief V, BanzigerM, Braun HJ (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics186(2):713–24. https://doi.org/10.1534/genetics.110.118521
Cuevas J, Crossa J, Montesinos-López OA, Burgueño J, Pérez-Rodríguez P, de Los CG (2017) Bayesian genomic prediction with genotype × environment interaction kernel models. 3 7:41–53
Cui Z, Dong G, Zhang A, Ruan Y, He Y, Zhang Z (2020) Assessment of the potential for genomic selection to improve husk traits in maize. G3 10:3741–3749
Daetwyler HD, Villanueva B, Woolliams JA (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS ONE 3:e3395
Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250–255
FAOSTAT. Retrieved from http://www.fao.org/faostat/en/
Grenier C, Cao TV, Ospina Y, Quintero C, Châtel MH, Tohme J, Courtois B, Ahmadi N (2015) Accuracy of genomic selection in a rice synthetic population developed for recurrent selection breeding. PLoS ONE 10:e0136594
Guo Z, Tucker DM, Wang D, Basten CJ, Ersoz E, Briggs WH, Lu J, Li M, Gay G (2013) Accuracy of across-environment genome-wide prediction in maize nested association mapping populations. G3 3:263–272
Heffner EL, Jannink JL, Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome 4:65–75
Jarquín D, Howard R, Crossa J, Beyene Y, Gowda M, Martini JWR, Pazaran GC, Burgueño J, Pacheco A, Grondona M, Wimmer V, Prasanna BM (2020) Genomic prediction enhanced sparse testing for multi-environment trials. G3 10:2725–2739
Lopez-Cruz M, Crossa J, Bonnett D, Dreisigacker S, Poland J, Jannink JL, Singh RP, Autrique E, de Los Campos G (2015) Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3 5:569–582
Lorenz A (2013) Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: a simulation experiment. G3 3:481–491
Mageto EK, Crossa J, Pérez-Rodríguez P, Dhliwayo T, Palacios-Rojas N, Lee M, Guo R, San Vicente F, Zhang X (2020) Hindu V. G3 10:2629–2639
Mansueto L, Fuentes RR, Borja FN, Detras J, Abriol-Santos JM, Chebotarov D, Sanciangco M, Palis K, Copetti D, Poliakov A, Dubchak I, Solovyev V, Wing RA, Hamilton RS, Mauleon R, McNally KL, Alexandrov N (2017) Rice SNP-seek database update: new SNPs, indels, and queries. Nucleic Acids Res 45(1):D1075–D1081. https://doi.org/10.1093/nar/gkw1135
McCouch SR, McNally KL, Wang W, Sackville Hamilton R (2012) Genomics of gene banks: a case study in rice. Am J Bot 99:407–423. https://doi.org/10.3732/ajb.1100385
Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genomewide dense marker maps. Genetics 157:1819–1829
Minten B, Barrett CB (2008) Agricultural technology, productivity, and poverty in Madagascar. World Dev 36:797–822. https://doi.org/10.1016/j.worlddev.2007.05.004
Mori A, Fukuda T, Vejchasarn P, Nestler J, Pariasca-Tanaka J, Wissuwa M (2016) The role of root size versus root efficiency in phosphorus (P) acquisition of rice. J Exp Bot. https://doi.org/10.1093/jxb/erv557
Onogi A, Ideta O, Inoshita Y, Ebana K, Yoshioka T, Yamasaki M, Iwata H (2015) Exploring the areas of applicability of whole-genome prediction methods for Asian rice (*Oryza sativa* L.). Theor Appl Genet 128:41–53
Ornella L, Pérez P, Tapia E, González-Camacho JM, Burgueño J, Zhang X, Vicente FS, Bonnett D, Dreisigacker S, Singh R, Long N, Crossa J (2014) Genomic-enabled prediction with classification algorithms. Heredity 112:616–626
Pace J, Yu X, Lübberstedt T (2015) Genomic prediction of seedling root length in maize (*Zea mays* L.). Plant J 83(5):903–912. https://doi.org/10.1111/tpj.12937
Perez P, de Los CG (2014) Genome-wide regression and prediction with the BGLR statistical package. Genetics 198:483–495
Rincent R, Laolë D, Nicolas S, Altman T, Brunel D, Revilla P, Rodríguez VM, Moreno-Gonzalez J, Melchinger A, Bauer E, Schoen CC, Giauffret C, Bauland C, Jamin P, Laborde J, Monod H, Flament P, Charcosset A, Moreau L (2012) Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). Genetics 192:715–728
Rose TJ, Mori A, Julia CC, Wissuwa M (2015) Screening for internal phosphorus utilisation efficiency: comparison of genotypes at equal shoot P content is critical. Plant Soil. https://doi.org/10.1007/s11104-015-2565-7
Saito K, van Oort P, Tanaka A, Dieng I, Senthilkumar K, Vandamme E, Nanfumba D (2017) Yield gap analysis towards meeting future rice demand. In: Sasaki T (ed) Achieving sustainable cultivation of rice, vol 2. Cultivation, pest and disease management. Burleigh Dodds Science Publishing, Cambridge, UK, pp 157–182
Saito K, Vandamme E, Johnson JM, Tanaka A, Senthilkumar K, Dieng I, Akakpo C, Gbaguidi F, Segda Z, Bassoro I, Lamare D, Gbakatchetche H, Abera BB, Jaiteh F, Bam RK, Dogbe W, Sékou K, Rabeson R, Kamissoko N, Mossi IM, Tarfa BD, Bakare SO, Kalisa A, Baggie I, Kajiru GJ, Ablede K, Ayeva T, Nanfumba D, Wopereis MCS (2019) Yield-limiting macronutrients for rice in sub-Saharan Africa. Geoderma 338:546–554. https://doi.org/10.1016/j.geoderma.2018.11.036
Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redona E, Atlin G, Jannink JL, McCouch SR (2015) Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genet 11:e1004982
Tanaka R, Iwata H (2018) Bayesian optimization for genomic selection: a method for discovering the best genotype among a large number of candidates. Theor Appl Genet 131(1):93–105
Tsujimoto Y, Rakotoson T, Tanaka A, Saito K (2019) Challenges and opportunities for improving N use efficiency for rice production in sub-Saharan Africa. Plant Prod Sci 22(4):413–427. https://doi.org/10.1080/1343943X.2019.1617638

USDA (2018). Production, supply and distribution online. Retrieved from https://apps.fas.usda.gov/psdonline/app/index.html#/app/home

Vandamme E, Rose TJ, Saito K, Yeong K, Wissuwa M (2015) Integration of P acquisition efficiency, P utilization efficiency and low grain P concentrations into P-efficient rice genotypes for specific target environments. Nutr Cycling Agroecosys. https://doi.org/10.1007/s10705-015-9716-3

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Wissuwa M, Kondo K, Fukuda T, Mori A, Rose MT, Pariasca-Tanaka J, Kretzschmar T, Haefele SM, Rose TJ (2015) Unmasking novel loci for internal phosphorus utilization efficiency in rice germplasm through genome-wide association analysis. PLoS ONE 10:e0124215. https://doi.org/10.1371/journal.pone.0124215

Yu X, Leiboff S, Li X, Guo T, Ronning N, Zhang X, Muehlbauer GJ, Timmermans MCP, Schnable PS, Scanlon MJ, Yu J (2020) Genomic prediction of maize microphenotypes provides insights for optimizing selection and mining diversity. Plant Biotechnol J 18:2456–2465

Yu X, Li X, Guo T, Zhu C, Wu Y, Mitchell SE, Roozeboom KL, Wang D, Wang ML, Pederson GA, Tesso TT, Schnable PS, Bernardo R, Yu J (2016) Genomic prediction contributing to a promising global strategy to turbocharge gene banks. Nat Plants 2:16150

Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin FH, Ranc N, Reif JC (2012) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet 124:769–776

Theoretical and Applied Genetics (2022) 135:2265–2278
https://doi.org/10.1007/s00122-022-04110-2

ORIGINAL ARTICLE

Genomic prediction of zinc-biofortification potential in rice gene bank accessions

Mbolatantely Rakotondramanana[1] · Ryokei Tanaka[2] · Juan Pariasca-Tanaka[3] · James Stangoulis[4] · Cécile Grenier[5] · Matthias Wissuwa[3]

Received: 24 November 2021 / Accepted: 19 April 2022 / Published online: 26 May 2022

Abstract

Key message **A genomic prediction model successfully predicted grain Zn concentrations in 3000 gene bank accessions and this was verified experimentally with selected potential donors having high on-farm grain-Zn in Madagascar.**

Abstract Increasing zinc (Zn) concentrations in edible parts of food crops, an approach termed Zn-biofortification, is a global breeding objective to alleviate micro-nutrient malnutrition. In particular, infants in countries like Madagascar are at risk of Zn deficiency because their dominant food source, rice, contains insufficient Zn. Biofortified rice varieties with increased grain Zn concentrations would offer a solution and our objective is to explore the genotypic variation present among rice gene bank accessions and to possibly identify underlying genetic factors through genomic prediction and genome-wide association studies (GWAS). A training set of 253 rice accessions was grown at two field sites in Madagascar to determine grain Zn concentrations and grain yield. A multi-locus GWAS analysis identified eight loci. Among these, QTN_11.3 had the largest effect and a rare allele increased grain Zn concentrations by 15%. A genomic prediction model was developed from the above training set to predict Zn concentrations of 3000 sequenced rice accessions. Predicted concentrations ranged from 17.1 to 40.2 ppm with a prediction accuracy of 0.51. An independent confirmation with 61 gene bank seed samples provided high correlations ($r = 0.74$) between measured and predicted values. Accessions from the *aus* sub-species had the highest predicted grain Zn concentrations and these were confirmed in additional field experiments, with one potential donor having more than twice the grain Zn compared to a local check variety. We conclude utilizing donors from the *aus* sub-species and employing genomic selection during the breeding process is the most promising approach to raise grain Zn concentrations in rice.

Introduction

Zinc (Zn) is an essential element for plants and humans alike, because Zn is a component of thousands of enzymes and a key regulator of gene expression and protein synthesis (Broadley et al. 2007; Galetti 2018). Zinc malnutrition is a global health problem that is particularly serious in infants where it impairs immune system function and delays infant development, causing stunting as the most visible symptom (Roohani et al. 2013; Galetti 2018). Alleviating human malnutrition for Zn has been included as one of the top priorities in the Sustainable Development Goals (SDG 2.2: End all forms of malnutrition). In 2003 the Consultative Group on International Agricultural Research (CGIAR) initiated a program to breed crops with higher concentrations of Zn, Fe and Pro-vitamin A carotenoids in the edible parts of a crop, an approach termed biofortification (Bouis and Saltzman 2017). Biofortification of crops represents an alternative to

Communicated by Huihui Li.

✉ Matthias Wissuwa
wissuwa@affrc.go.jp

1 Rice Research Department, The National Center for Applied Research on Rural Development (FOFIFA), 101 Antananarivo, Madagascar

2 Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-8657, Japan

3 Crop, Livestock and Environment Division, Japan International Research Center for Agricultural Sciences (JIRCAS), 1-1 Ohwashi, Tsukuba, Ibaraki 305-8686, Japan

4 College of Science and Engineering, Flinders University, Bedford Park, SA 5042, Australia

5 CIRAD, INRAE, Institut Agro, UMR AGAP Institut, Univ Montpellier, 34398 Montpellier, France

food fortification and while both approaches are important in alleviating malnutrition, it is believed that crop biofortification is a very efficient tool to reach rural communities that are largely food self-sufficient (Virk et al. 2021). Programs to develop Zn biofortified rice varieties have been successful in Asia (Swamy et al. 2016; Rao et al. 2020) and Latin America (Harvest Plus 2021) but concerted efforts to do so in Africa are still non-existing.

Madagascar remains a low-income country with a high level of malnutrition (The World Bank 2016; WFP 2010). In rural areas, 50% of children suffer from stunting and are underweight (Stewart et al. 2020), one of the highest rates in the world (UNICEF 2019). In the central highlands, the highest levels of stunting (60%) are found and recent surveys by JIRCAS and partners estimated that 80% of the population consume inadequate amounts of Zn (Shiratori et al. 2018). Rice is essential in Malagasy diets; it is eaten three times a day and represents 50 percent of the daily caloric intake with per capita consumption being above 120 kg annually. Rice, having such a pre-eminent position for food supply, is naturally a target for intervention. Consequently, Madagascar has the 3rd highest Biofortification Prioritization Index (BPI) for Zn in rice for Africa and the 13th highest globally (Harvest Plus 2021).

Zn concentrations in polished rice are typically too low to supply a sufficiently high proportion of the daily required intake of Zn (Bouis and Welch 2010), thus where rice is the main staple and households cannot afford to diversify their meal by adding mineral-rich fruits, vegetables and meat, Zn deficiency is prevalent (Harvest Plus 2021). To overcome this deficiency, grain Zn concentration in rice needs to be increased by 50% or more to significantly alleviate Zn malnutrition (Bouis and Welch 2010). Developing rice varieties with increased grain Zn concentrations therefore remains an important global objective (Rao et al. 2020) that offers a low-cost and long-lasting solution to the persisting problem of Zn malnutrition (Bouis et al. 2011).

Rice grain Zn concentrations are strongly affected by factors such as genotype and environment, with soil properties being the main source of environmental variation. For a given genotype, grain Zn concentrations may vary by a factor 2–3 depending on soil type and related Zn bio-availability for plant uptake (Wissuwa et al. 2008; Goloran et al. 2019; Rao et al. 2020). Low Zn bio-availability in paddy soils is commonly associated with alkalinity (high soil pH and excess bicarbonate) and very low soil redox potentials (Johnson-Beebout et al. 2016). Both factors trigger the formation of Zn-complexes with soil constituents and in consequence the soluble Zn fraction that is removed by the plant will be replenished too slowly to assure high Zn uptake rates (Broadley et al. 2007). The effect of a decreasing soil redox potential after flooding tends to cause Zn bio-availability to be lowest toward the end of the cropping season and thus reduces Zn uptake during the reproductive phase when Zn taken up may be directly transported to reproductive organs. For this reason, basal Zn fertilizer application has often very limited effects on increasing grain Zn concentrations (Johnson-Beebout et al. 2016) and it would explain the observation that grain Zn concentrations tend to be lower during the rainy season compared to the dry season (Goloran et al. 2019).

The genotypic variation in grain Zn concentrations is similar in magnitude to the environmental variation with 2–3 fold differences having been detected repeatedly (Norton et al. 2014; Swamy et al. 2018; Zhang et al. 2018). Since grain Zn concentrations are influenced at many levels, starting with Zn uptake by the root, followed by transport and reallocation of Zn within the plant, to Zn loading into the grain (Swamy et al. 2016), it is likely the genotypic differences at each of these levels exist. Which of these factors contribute most to genotypic differences in grain Zn concentrations remains uncertain. Some high-Zn genotypes appear to rely mostly on Zn remobilization, whereas others maintain high Zn uptake rates during grain filling (Johnson-Beebout et al. 2016). At the same time, increased root uptake does not necessarily result in enhanced Zn accumulation in rice grains, suggesting Zn loading into the endosperm to be the main limiting step for which genotypic differences fortunately exist (Jiang et al. 2008).

Nicotianamine (NA) is a ubiquitous chelator of metal cations, such as Fe^{2+} and Zn^{2+}. Biosynthetic precursor of phytosiderophore secretion from roots, NA is responsible for Fe internal metal transport and maintenance of metal homeostasis. In rice, three NA synthase genes were identified (*OsNAS1*, *OsNAS2* and *OsNAS3*) that have been largely studied to demonstrate their role in increased bioavailable Fe levels in rice grains (Higuchi et al. 2001; Inoue et al. 2003). Through transgenic approaches overexpressing the *OsNAS* gene, it has been possible to significantly increase both grain Fe and Zn concentration, indicating Zn transport processes to be of additional importance (Johnson et al. 2011).

The genetic control of grain Zn concentration in rice has been widely studied, using bi-parental mapping populations (reviewed in Swamy et al. 2016), diversity panels (Norton et al. 2014; Zhang et al. 2018) or double-haploid (DH) derived biparental populations (Swamy et al. 2018). Consistently, these studies reported a large number of genetic regions controlling Zn concentration, each with relatively minor effects. This may be expected given that grain Zn concentrations are likely the result of multiple interacting physiological processes. For Zn uptake alone, it has been shown that at least two distinct processes, root proliferation and rhizosphere Zn mobilization, are causative of genotypic differences in plant Zn uptake (Mori et al. 2016). While none of the identified loci appear to be currently used in marker-assisted breeding, loci on chromosomes 7, 11 and 12 have

been identified consistently (Swamy et al. 2016). Of these, the QTL on chromosome 7 co-localizes with *OsNAS3* and may therefore be of particular interest (Cu et al. 2021).

The complex nature of a trait like grain Zn concentration, which depends on multiple physiological mechanisms, each potentially controlled by multiple underlying genes, may necessitate a genome-wide rather than a single marker selection approach. Genomic Prediction (GP; Meuwissen et al. 2001) for mineral content has already proven efficient in maize and wheat where the predictive ability (PA) for grain Zn content was between 0.43 and 0.73 in maize (Mageto et al. 2020) and between 0.33 and 0.69 for wheat (Velu et al. 2016) depending on the population and environment chosen. A similar PA of 0.51 was achieved for grain Zn concentration improvements in a rice synthetic population managed through recurrent selection when multi-site data were considered for the calibration model (Baertschi et al. 2021).

Given that environmental factors strongly affect grain Zn concentrations, it is of interest to determine to what extent GP can be employed in target environments that are less homogenous compared to the well-managed trials conducted on research stations. While the polygenic nature of grain Zn may favor a GP approach, it is possible that main effect single loci are more stable across environments and therefore possess greater predictive power in less controlled environments. We have grown a set of 253 rice gene bank accessions sampled from the 3 K genome project (Mansueto et al. 2017) in two farmer's fields in Madagascar and determined the variation for grain Zn concentrations and grain yield. Using this dataset, the objectives of this study were to:

(i) Conduct genome-wide association studies (GWAS) in an attempt to detect alleles associated with high grain Zn concentrations,
(ii) Develop a GP model for grain Zn concentrations based on above training set and employ this model to predict grain Zn concentrations among the 3000 sequenced rice accessions available at the IRRI gene bank,
(iii) Identify potential donors with high grain Zn concentrations and confirm their suitability through confirmatory experiments.

Materials and methods

Field phenotyping

Field experiments were conducted at two sites in the central highlands of Madagascar, Anjiro (elevation 950 m, 18°54′01.7 ″S 47°58′12.4 ″E) and Ankazomiriotra (1150 m, 19°40′07.9 ″S 46°33′53.9 ″E). The experiments were carried out in farmers' fields under flooded lowland conditions during the 2017–18 rainy season with sowing in November, transplanting in late November to December and harvests in April to May. Following the typical farmer's practice in the region, chemical fertilizer was not applied and neither did fields receive organic manure. At each site, 523 accessions selected from the set of 3 K sequenced accessions available at IRRI were grown with two replications in a randomized complete block design. Several sub-sets selected to represent extreme variation for grain yield, maturity or plant height existed within these 523 accessions, (Tanaka et al. 2021) and only those considered adapted to our field sites were used in the present study (see below).

At both sites, accessions were transplanted in 2-row micro-plots of 2 m length with spacing of 20 cm between and within rows (22 single plant hills per plot). Heading date (HD) was recorded at 50% heading for each accession. During harvest, five representative plants per plot were cut, panicles were separated from straw, placed in paper bags to avoid contamination by soil or dust, and taken to the laboratory where they were air-dried for a week before total panicle dry weight was determined. Grain yield (GY) was estimated from the panicle weight of these five plants, assuming a realized density of 22 hills per m^2 and expressed in kg per ha. Straw weight (SWT) was determined on the same five harvested plants, first as fresh weight which was then adjusted for moisture content after oven-drying a subsample for 48 h at 70 °C.

Grain processing and grain Zn analysis

Grain Zn concentrations were determined for a subset of 253 accessions from the 523 grown at field sites. All accessions considered poorly adapted to experimental sites were omitted, which included accessions with very early or late maturity and all accessions with low GY or that had lodged and had been contaminated by soil. A focus on well-adapted accessions was meant to prevent the potential confounding effect of high grain Zn being the result of poor grain yield and very low harvest index. A random sub-sample of the harvested grain from ten panicles per plot was dehulled to obtain brown rice and these whole grain samples were sent to Flinders University, Australia for further analysis.

For inductively coupled plasma mass spectrometry (ICP-MS) analysis, 0.3 g of whole rice seed, which had been oven-dried at 80 °C for 4 h to remove remaining moisture, was acid-digested in a closed tube as described in Wheal et al. (2011). Elemental concentrations of samples were measured using ICP-MS (8900; Agilent, Santa Clara, CA) according to the method of Palmer et al. (2014). The grain Zn concentration is given in μg g^{-1} on a dry weight basis. In each of 10 digestion batches, a blank and a certified reference material (CRM; NIST 1568b rice flour) were added for quality assurance. Samples with Al present at > 5 μg g^{-1} were considered

to have unacceptable levels of purported soil contamination (Yasmin et al. 2014), thus they were eliminated from the dataset.

Statistical analysis for phenotypic values

Given the experimental design, the following linear model was fitted for each trait:

$$y_{ijk} = \mu + g_i + s_{ij} + \alpha_j + \beta_{jk} + e_{ijk}$$

where y_{ijk} is the phenotypic value (*i.e.*, observed Zn concentration) of i-th genotype evaluated at the k-th block in j-th site, μ is an intercept, g_i is the genotypic value of i-th genotype, s_{ij} is the interaction effect between the i-th genotype and the j-th site, α_j is the effect of j-th site, β_{jk} is the effect of k-th block in the j-th site, and e_{ijk} is the residual. The interaction effects s_{ij} and the block effects β_{jk} were modeled as random effects, and the other model terms were modeled as fixed effects. This model was implemented in *lmer* function in the lme4 package. To test the statistical significance between two sites, site effect (α_j) for each trait was tested based on the type-III analysis of variance with Satterthwaite's method using the *anova* function in the lmerTest package (Kuznetsova et al. 2017). The estimated values (best linear unbiased estimates; BLUEs) of g_i were used in the subsequent association mapping and the genomic prediction analyses.

Heritability was calculated based on the same linear model but treating the genotypic values g_i as random effects. Using the estimated variance components of the genotypic values (s_{ij}), genotype-by-site interaction effects (σ_s^2) and residuals (σ_e^2), a broad-sense heritability (H^2) was calculated as follows (Holland et al. 2003);

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_s^2}{n_{\text{site}}} + \frac{\sigma_e^2}{n_{\text{block}} \times n_{\text{site}}}}$$

where n_{block} is the number of blocks per site, n_{site} is the number of sites (*i.e.*, $n_{\text{block}} = n_{\text{site}} = 2$ given our experimental design).

Phenotypic correlation among sites and traits was calculated based on the Pearson' correlation after averaging the observed phenotypic values over the two blocks for each accession (if an accession did not have an observed value in one of the two blocks, the available phenotypic value was used instead of the average).

Genomic data and Genome-Wide Association (GWA) analysis

The 404 K core SNPs dataset was downloaded from the IRRI SNP-Seek website (https://snp-seek.irri.org/_download.zul). SNP having more than 5% missing data or a minor allele frequency below 2.5% were removed, retaining 186,229 SNPs for 3,024 accessions. Remaining missing states were imputed using Beagle v.4.1 (21Jan17.6 cc; Browning and Browning 2016).

Without further filtering, GWA analysis was performed on the 253 accessions with the BLUE values for grain Zn concentrations and the 186 k genotype matrix using the multi-locus random-SNP-effect mixed linear model (mrMLM) software package, which includes the mrMLM, FASTmrMLM, FASTmrEMMA, pkWmEB, pLARmEB, ISIS EM-BLASSO methods (https://cran.r-project.org/web/packages/mrMLM/index.html). A kinship matrix was calculated by mrMLM by default using the method of Kang et al. (2008) and default values were used for the parameters in all methods. To account for additional population structure a set of principal components (PC) was calculated using TASSEL (v5.2.75). PCs explaining more than 5% of the variation were included in the GWA analysis by indicating the type of population structure (PopStrType) = "PCA." An output of Quantitative Trait Nucleotides (QTN) exceeding a threshold LOD value > 3 at each of the six multi-locus models was generated as the last step of the analysis and visualized in a combined Manhattan plot. QTN exceeding this LOD threshold in at least three of the six models were considered significant and evaluated further.

The allele effect at each locus was determined by calculating the average phenotypic values of all accessions carrying either allele and a box-plot graph was generated using an in-house R script. A graphical representation of subset of SNPs surrounding the significant QTNs was generated by the Haploview 4.2 software (Barret et al. 2005). Linkage disequilibrium (LD) blocks were then identified and manually delignated based on the recombination rate, which is displayed using the standard color scheme: *D*'/LOD (wherein red color reveals linkage disequilibrium between two genetic markers, $D' = 1$ and LOD > 2).

Genomic prediction

Genomic prediction was performed with the GBLUP model (Bernardo 1994) using the rrBLUP package (Endelman 2011):

$$\mathbf{g} = 1\mu + \mathbf{Zu} + \mathbf{e}, \quad \mathbf{u} \sim \text{MVN}\left(0, \mathbf{G}\sigma_u^2\right), \quad \mathbf{e} \sim \text{MVN}\left(0, \mathbf{I}\sigma_e^2\right)$$

where $\mathbf{g}$ is the Zn BLUE values, 1 is a vector of ones, μ is the grand mean, $\mathbf{Z}$ is the design matrix, $\mathbf{u}$ is the vector of genotypic values, $\mathbf{e}$ is the vector of residuals, 0 is a vector of zeros, $\mathbf{G}$ is a genomic relationship matrix, σ_u^2 is genetic variance, $\mathbf{I}$ is an identify matrix, and σ_e^2 is the residual variance. The genotypic values and residuals are assumed to follow a

multivariate normal distribution (MVN). The genomic relationship matrix was calculated by using the A.mat function in rrBLUP package, as described in Tanaka et al. (2021).

By using the above GBLUP model, Zn concentrations of the all 3 K accessions were predicted with three slightly different training sets: (1) all phenotyped accessions ($n=253$), (2) all phenotyped accessions excluding IRIS_313_9368, which had extremely high Zn concentrations and may therefore be highly influential ($n=252$) and (3) excluding the six members of the *aus* subspecies ($n=247$). In addition, tenfold cross-validation was repeated 10 times to evaluate the prediction accuracy within the phenotyped accessions. Prediction accuracy was evaluated by taking correlation between the BLUE values (g) and the predicted genotypic values from the GBLUP model (u).

Confirmatory experiments

Two subsequent experiments were conducted to confirm results obtained with the training set. To independently confirm the reliability of the GP model, 61 additional accessions from the 3 K set available at IRRI were selected for the determination of grain Zn concentrations. These accessions were selected based on their predicted grain Zn concentrations being either high ($n=19$), intermediate ($n=24$) or low ($n=18$). Accessions had not been grown in Madagascar but seed produced at IRRI and distributed to JIRCAS was used directly for the determination of grain Zn concentrations. After acid-digestion of 0.25 g dehulled seed, elemental concentrations in samples were measured using ICP Emissions Spectrometer ICPE- 9000 (Shimadzu, Kyoto, Japan).

A further experiment was conducted to investigate whether the high grain Zn concentrations identified in potential high-Zn donor accessions were repeatable and stable across several field sites in Madagascar. Three high-Zn accessions of the training set (IRIS313-9368, IRIS313-10,114 and CX158) were grown in a multi-location trial together with local (X265) and international (IR64) check varieties. The experiments were conducted in Anjiro, Ankazo and Behenji (elevation 1428 m, 19°14′44.92″S 47°28′45.38″E) villages in the central highlands of Madagascar during the 2018–19 rainy season, using five farmer's fields with two fertilizer treatments (zero input versus fertilization with NPK) and three replications. Plot sizes were 2 m^2. Grain samples were processed as for the training set and sent to Flinders University for the determination of elemental concentrations as outlined above.

Results

Phenotypic variation

The average grain Zn concentrations (Zn) for the 253 tested accessions ranged from 16.6 to 48.4 µg g^{-1} at the Anjiro site and from 14.2 to 35.6 µg g^{-1} at the Ankazo site (Fig. 1) and highest values at both sites were detected in accession IRIS_313_9368. In addition to having wider variation, the Anjiro site average of 25.4 µg g^{-1} was significantly higher ($p<0.01$) compared to Ankazo with 21.6 µg g^{-1}. Despite these differences, grain Zn concentrations at both sites had a tighter correlation ($r=0.65$) compared to other traits with the exception of days to heading ($r=0.86$). The average GY was 4.4 t ha^{-1}at Anjiro and 4.2 t ha^{-1} at Ankazo and respective SWT means were 29.7 and 28.2 g $plant^{-1}$ (Fig. 1). Neither trait differed significantly between sites.

Correlations between the two sites were low for GY ($r=0.15$; $p<0.05$) and slightly higher for SWT ($r=0.28$; $p<0.001$). However, at each site GY was positively correlated to SWT ($r=0.55$ and $r=0.61$ for Anjiro and Ankazo, respectively; $p<0.001$ for both). Accessions showed large variation for HD, ranging from 60 to 127 days at Ankazo and from 64 to 129 days at Anjiro. The similar range and high correlation of $r=0.86$ indicated that site effects were very small for HD. Late heading was associated with increased SWT at both sites but the effect of late heading on GY was site-specific, with a low but significantly positive effect at Anjiro ($r=0.32$; $p<0.001$) compared to a non-significant (negative) effect ($r=-0.11$; ns) in Ankazo (Fig. 1).

Interestingly, Zn concentrations showed weak correlations with other traits, except for a low and negative correlation with GY in Anjiro ($r=-0.33$; $p<0.001$). Furthermore, the broad-sense heritability for Zn concentrations across sites was high ($H^2=0.79$; Fig. S1), implying that genotype by site interaction effects for Zn concentrations were small. For that reason, the association mapping and GP were conducted with the across-site BLUE values to analyze the common genetic control across two sites.

The 253 accessions tested belonged primarily to the *indica* sub-species of rice, the second biggest group were *japonica* accessions while other sub-populations were represented by only 6–10 individuals (Fig. 2a). The focus on mostly *indica* accessions was due to the preference for *indica*-type varieties by lowland rice growers and consumers in Madagascar. The *aus* sub-species group had the highest average grain Zn concentration (33.5 µg g^{-1}), compared with an average of 25.1 µg g^{-1} for the *indica* group and 28.4 µg g^{-1} for the *japonica* group.

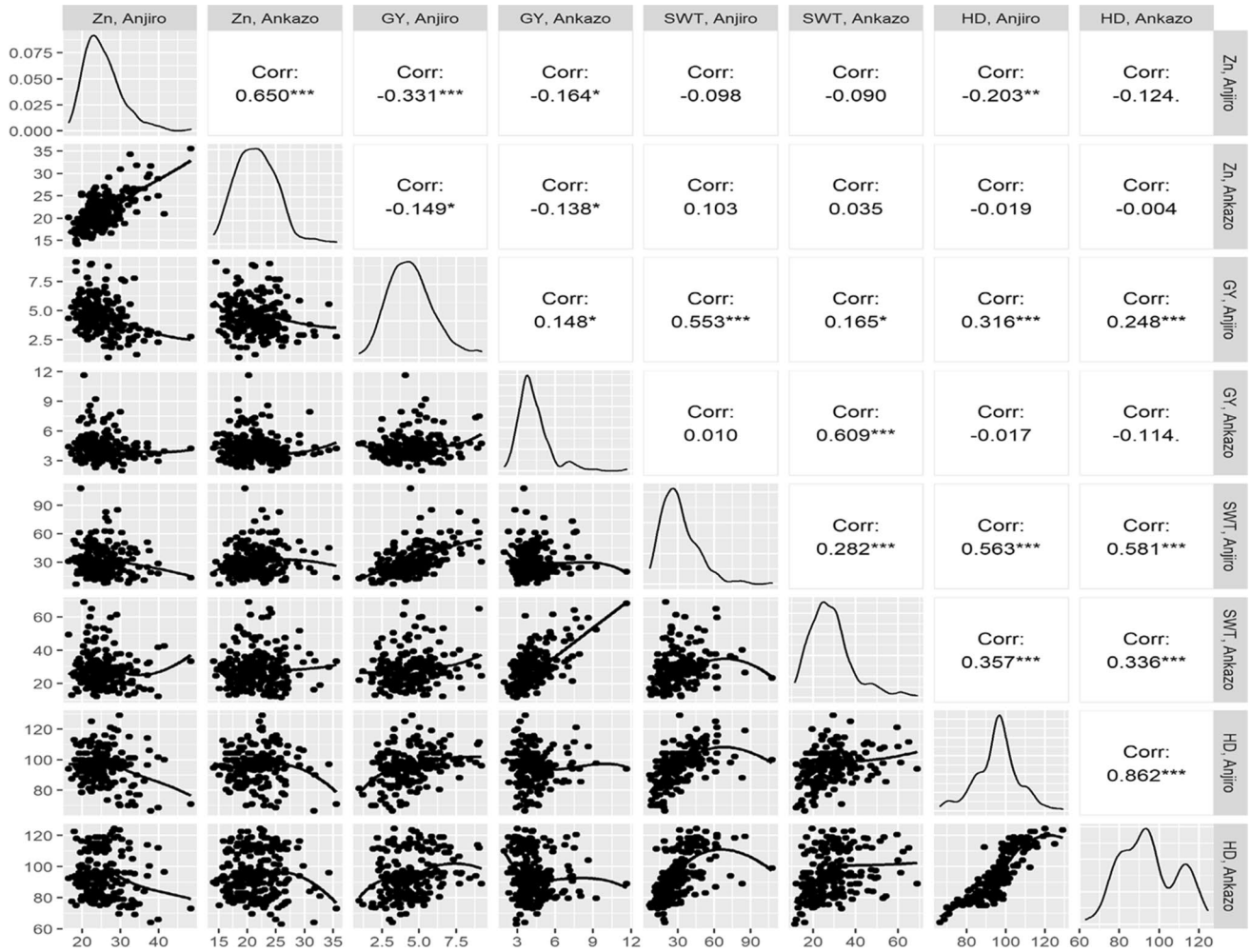


Fig. 1 Repartition and correlation of zinc concentration (Zn), grain yield (GY), shoot weight (SWT) and heading date (HD) at sites Ankazomiriotra (AZ) and Anjiro (AJ)

Genome-wide associations for grain Zn concentrations

A multi-locus GWAS approach considering results from six multi-locus analysis methods was employed to identify genetic associations with grain Zn concentrations and associations were considered significant when the estimated LOD surpassed the threshold of 3.0 in at least three methods. Based on these criteria, eight loci associated with grain Zn concentrations were detected on chromosomes 2, 4, 8, 10, 11 and 12 (Table 1). The full list of all QTN with a LOD > 3 in any of the six methods is shown in Table S1 and the corresponding Manhattan plots in Fig. S2.

The strongest peak identified in terms of maximum LOD value (16.4), number of approaches identifying the locus (5) and consistently high QTN effect estimates (1.69–3.69 µg Zn g^{-1}) was QTN_11.3 at 28,757,650 bp on chromosome 11 (Table 1; Table S1). The minor allele frequency (MAF) at this locus was 4.4% and the R^2 was 15.4. The remaining seven loci had comparatively minor effects with R^2 estimates of 2.1–5.9 and maximum QTN effect estimates between 0.9 and 1.9 µg Zn g^{-1} (Table 1).

At each QTN we investigated to what extent the minor allele and the allele increasing grain Zn concentrations was associated with different rice sub-populations (Fig. 3). For QTN_2.1 allelic variation was detected in all sub-species except for *japonica* but differences were only significant in the *indica* group. QTN_8.1 had allelic variation within the *aus* and *indica* groups but mean differences were not significant. QTN_11.2 allelic variants existed within all sub-species and the minor allele significantly increased grain Zn in the *aus* and *indica* groups. For the strongest QTN_11.3, the minor allele was associated with higher grain Zn concentrations in the *aus*, *indica* and *japonica* sub-species (Fig. 3). At QTN on chromosomes 10 and 12 allelic variation was only detected within the *indica* group and the minor allele

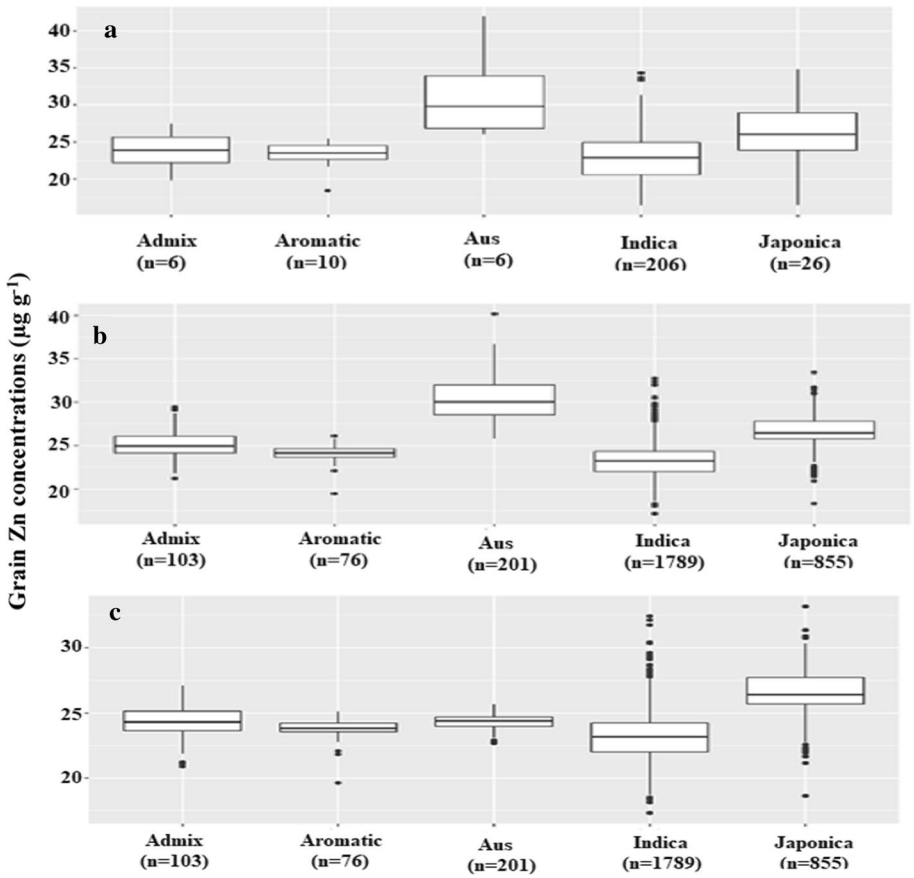


Fig. 2 Variation in grain Zn concentrations in accessions from five rice sub-populations: **a** measured data of the training set ($n=253$); **b** predicted values of the entire 3 K set using the full training set ($n=253$); **c** predicted values of the entire 3 K set using a training set excluding six aus accessions ($n=247$)

Table 1 Associations for grain Zn concentrations detected (with LOD > 3.0) in at least three of the six multi-locus approaches employed. For R^2 and QTN effects, highest estimates by any of the significant approaches are shown

QTN	Chromosome	Position	mrMLM	FASTmrMLM	FASTmrEMMA	pkWmEB	pLARmEB	ISIS EM-BLASSO	R^2	MAF (%)	QTN effect (µg Zn g^{-1})
2.1	2	18,697,369	5.1	3.2	4.9	3.2			5.5	45.9	1.2
4.1	4	20,025,747		7.8		3.7	4.2		5.9	20.6	1.2
8.1	8	26,505,039		7.5	6.5	6.7	5.6		3.0	10.7	1.9
10.1	10	14,217,374	4.4	3.6		7.6	4.1	3.9	3.5	5.5	1.5
11.1	11	26,546,816	4.3	6.2		8.6	6.2		4.9	13.0	1.2
11.2	11	27,604,708		6.0	3.2	7.4			3.2	15.6	− 1.4
11.3	11	28,757,650	15.2	10.4		13.6	16.4	3.6	15.4	4.4	3.7
12.1	12	7,184,806	3.1	4.8		3.9			2.1	9.7	0.9

increased grain Zn (Fig. S3). In only one case (chromosome 11: 27,604,708), the minor allele was associated with reduced grain Zn and the allelic difference was only pronounced in the *aus* and *japonica* groups where it reduced grain Zn by 24 and 19%, respectively.

For the most influential locus QTN_11.3, we delimited the linkage block surrounding significant QTNs. Strong linkage that would define a clear block was not detected but similarities between SNP genotypes were suggestive of a linkage block extending from 28.681 to 28.798 Mbp (Fig. S4). This region contained 26 gene models of which 18 were functionally annotated (Table S2). One gene family was strongly overrepresented at this locus as 11 genes were annotated as glucosyl hydrolases or, more specifically, as either class III chitinase homologs or xylanase inhibitors. In addition, two "thaumatin family domain containing proteins" and two Zinc finger proteins were annotated in the target region.

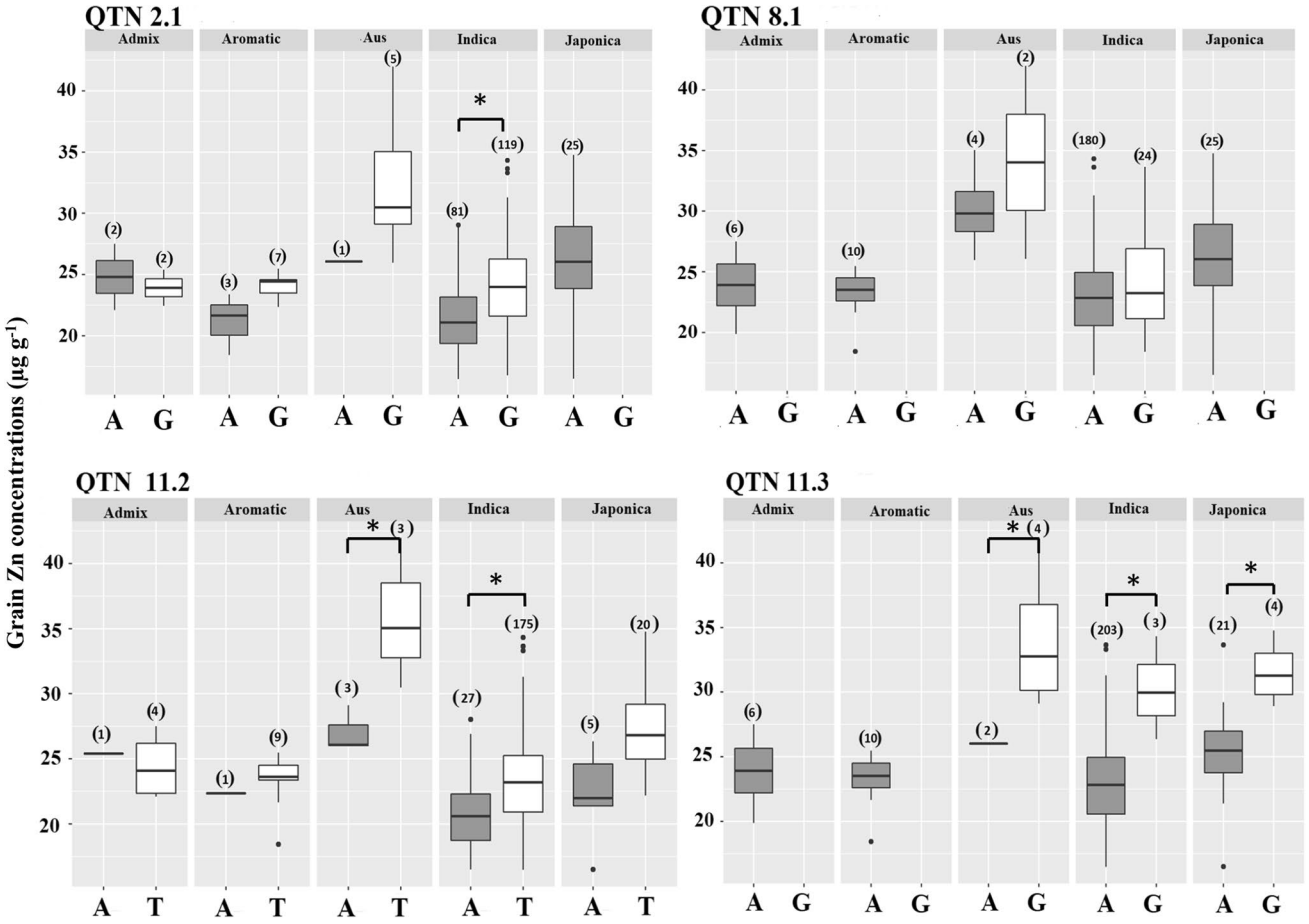


Fig. 3 Grain Zn concentrations for the two allelic groups in the five sub-populations for QTN_2.1, QTN_8.1, QTN_11.2 and QTN_11.3. Numbers above bars indicate the number of accessions in the respective group

Genomic prediction for grain Zn concentrations

Utilizing the same GWAS dataset as a training set, a GP model was developed to predict grain Zn concentrations of the entire set of 3 K accessions. The full model including all 253 training accessions predicted grain Zn concentrations to range from 17.1 µg Zn g^{-1} to as high as 40.2 µg Zn g^{-1} (Fig. 2b). Differences between subpopulations were pronounced, with the *aus* group having highest predicted values and an average of 30.3 µg Zn g^{-1}. The second highest average was predicted for the *japonica* group (26.7 µg Zn g^{-1}) and the lowest for the *indica* group (23.2 µg Zn g^{-1}). In fact, only members of the *aus* sub-species were found among the top 20 predicted accessions (Table S3) and IRIS_313_9368, which had the highest grain Zn concentrations in the training set (42.0 µg Zn g^{-1}), was also the highest predicted accession (40.2 µg Zn g^{-1}).

The training set contained only six *aus* accessions, among which IRIS_313_9368 may have been highly influential. To test to what extent the small number of *aus* accessions may have skewed predictions, two additional GP models were tested, a 2nd model excluding IRIS_313_9368 and a 3rd model excluding all six *aus*. Predicted values of the 3rd model are shown in comparison to the full model in Fig. 4. Excluding *aus* from the training set did not have major effects on predicted values for the four non-*aus* subpopulations, but strongly decreased the predicted grain Zn concentrations of the 201 *aus* accessions among the 3 K set. Their average decreased from 30.3 to only 24.3 µg Zn g^{-1}, which was lower than the predicted average of the *japonica group* (26.6 µg Zn g^{-1}) using the same training set (Fig. 2c). Only excluding IRIS_313_9368 did not have comparably strong effects, though the predicted Zn concentrations in *aus* accessions decreased from the full model (Fig. S5).

Tenfold cross-validation was performed to evaluate the accuracy of predictions for the full model and the one excluding *aus* accessions. When using all phenotyped accessions, average prediction accuracy of the ten replications was $r = 0.51$ with the standard deviation (SD) of 0.02. This dropped slightly to $r = 0.49$ (SD = 0.01) for the 2nd model and to $r = 0.48$ (SD = 0.01) for the 3rd model with *aus* excluded.

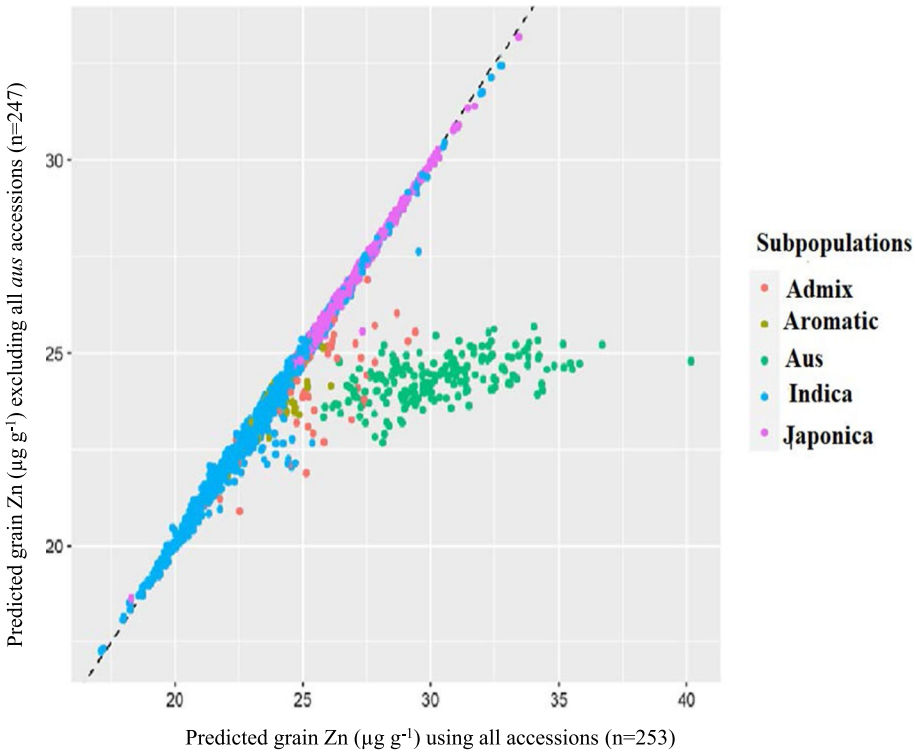


Fig. 4 Predicted grain Zn concentrations of the entire 3 K set of accessions available at the IRRI gene bank based on two different training sets. Predictions shown on the x-axis are based on the entire set of 253 accessions tested in Madagascar, whereas predictions on the y-axis are based on only 247 accessions with all six members of the aus sub-population omitted

For an additional and independent confirmation of prediction results, grain Zn concentrations were determined in a different subset of accessions selected from the 3 K set. These belonged to the *aus* ($n=30$), *indica* ($n=24$) and *japonica* ($n=7$) subpopulations (Table S4). Seeds analyzed had been imported directly from the IRRI genebank and were thus not grown in Madagascar (though there was a small overlap with accessions in the field in Madagascar, $n=11$). The correlation between measured and predicted values was $r=0.74$ (Fig. 5) and correlations with the 2nd model ($r=0.66$) or 3rd model ($r=0.35$) were lower (data not shown). Predicted mean values separated the *aus* group (average 31.3 µg Zn g^{-1}) from the *indica* (23.6 µg g^{-1}) and *japonica* (25.1 µg g^{-1}) groups and measured means were within 1.0 µg Zn g^{-1} of predicted group means (31.7, 23.4 and 24.2 µg Zn g^{-1}, respectively). For the main locus identified on chromosome 11 in GWAS (QTN_11.3), the positive minor allele was present in 12 of the 61 accessions. The allelic effect appeared to be significant with an average Zn concentration of 32.4 µg Zn g^{-1} for the minor compared to 26.3 µg Zn g^{-1} for the major allele (Table S5); however, all accessions with the minor allele belonged to the *aus* group and within that sub-species, allelic groups did not differ (32.4 vs. 31.1 µg Zn g^{-1}).

The second confirmatory experiment was conducted with potential high-Zn donors at five field sites and with two fertilizer treatments in Madagascar. The ANOVA indicated that genotypic differences were the dominant source of variation in this dataset (Fig. 6) and potential high-Zn

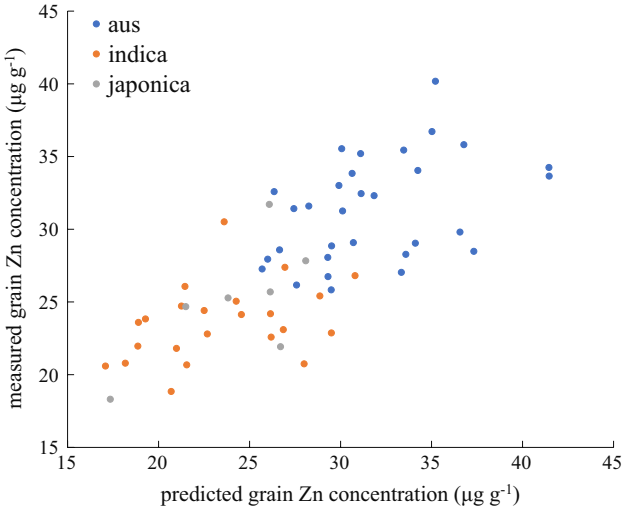


Fig. 5 Independent validation of predicted grain Zn concentrations in a set of 61 accessions imported from the IRRI gene bank. Predicted values are based on the training set of 253 accessions grown in Madagascar and measured values are from seed distributed by the IRRI gene bank

donor IRIS313-9368 (*aus*) was consistently superior to other accessions, irrespective of sites and fertilizer treatments. With an average of 42.5 µg Zn g^{-1}, it surpassed its predicted grain Zn concentration of 40.2 µg Zn g^{-1}. Furthermore, it had almost twice the Zn concentration of IR64 (21.7 µg g^{-1}) and more than twice compared to the Malagasy check X265 (18.6 µg g^{-1}). Other tested accessions were the top predicted

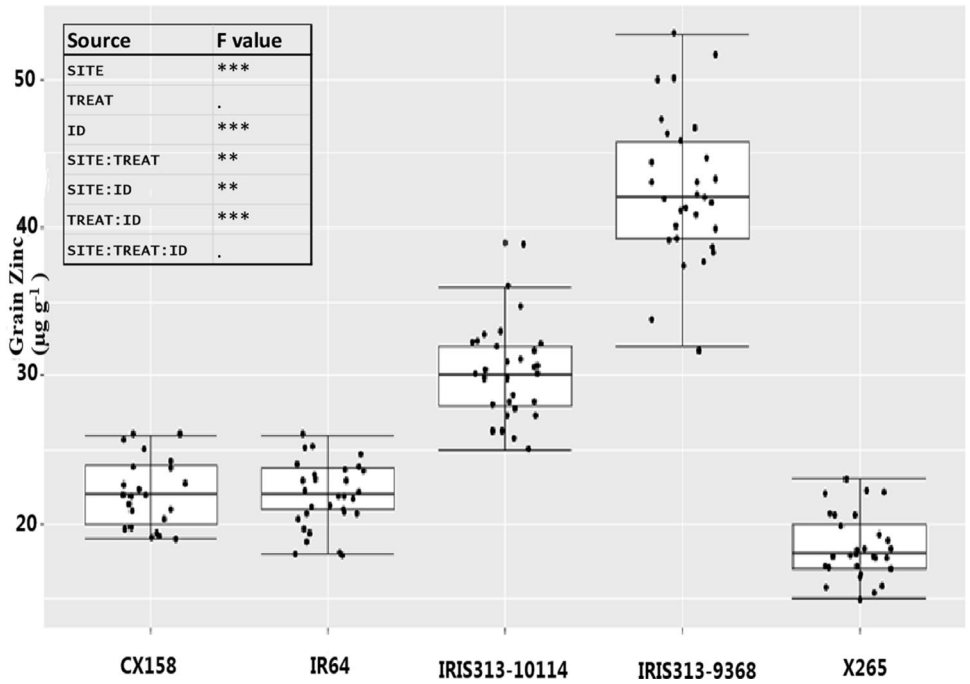


Fig. 6 Variation in grain Zn concentration of five rice accessions across five field sites and two fertilizer treatments (no input and 120 kg ha^{-1} NPK compound fertilizer). Differences between accessions were the dominant source of variation, followed by differences between sites (see inlet of ANOVA table)

indica CX158 and IRIS313-10,114 and while CX158 failed to reach its predicted value of 32.0 µg Zn g^{-1}, IRIS313-10,114 matched its predicted value of 30.5 µg g^{-1} to within 1% and was consistently superior to the check varieties, irrespective of site or fertilizer effects (Fig. 6). As in the training set, plants had higher average grain Zn concentrations at both field sites in Anjiro (28.6–29.6 µg Zn g^{-1}) compared to Ankazo village (26.5 µg Zn g^{-1}), with Behenji (25.0–25.9 µg Zn g^{-1}) being lowest. Applying NPK fertilizer had a small positive effect, increasing average grain Zn concentrations from 26.6 to 27.7 µg Zn g^{-1}) but this effect was only significant in two of the five sites (data not shown).

Discussion

Experiments were conducted with a diverse set of gene bank accessions and more than twofold variation in grain Zn concentrations were detected. Other traits such as GY or HD also varied considerably, but correlations between this variation and grain Zn concentrations were low (Fig. 1). Large differences in GY could have affected grain Zn concentrations, due to a possible dilution of Zn in the greater grain biomass of higher-yielding accessions and of the reverse effect in very low-yielding accessions. While excluding high-yielding varieties was no option because such material should be the target of any breeding program, we had omitted accessions with extremely low yield at any of the two sites as we considered these to be not sufficiently adapted to local conditions to provide reliable data. Possibly as a result of this precaution the correlation between GY and grain Zn concentrations remained very weak ($r = -0.27$; Fig. S1) and likely did not affect outcomes of the GWAS and GS studies. In addition, the heritability for grain Zn concentrations ($H^2 = 0.79$) was larger than for GY ($H^2 = 0.40$). High heritability (> 0.70) for grain Zn has also been reported elsewhere (Swamy et al. 2016, 2018; Baertschi et al. 2021).

A rather high heritability and good correlations for grain Zn concentrations between different sites did, however, not mean site effects were absent. Grain Zn concentrations differed significantly and consistently between sites, with samples from Anjiro village (25.4 µg g^{-1}) having significantly higher grain Zn concentrations compared to the Ankazo site (21.6 µg g^{-1}) and this may have been due to poor drainage in Ankazo, which could lower Zn availability due to a more reduced soil state. Nevertheless, neither site can be described as Zn deficient considering our observed ranges were comparable to or higher than in similar studies (Norton et al. 2014; Swamy et al. 2018; Rao et al. 2020).

All rice grain analyzed was unpolished brown rice, which was primarily due to the fact that high-quality milling equipment capable of polishing rice without contaminating samples with Zn during the milling process was not available in Madagascar. To what extent our analysis of brown rice could have affected results and conclusions were briefly assessed by polishing a small sub-sample of grain and results indicated that the average grain Zn concentrations decreased by 18% from 33.2 µg g^{-1} in brown rice to 27.2 µg g^{-1} in polished rice (Fig. S6), which is similar to reductions reported elsewhere (Suman et al. 2021). Interestingly, the reduction was larger in low-Zn accessions (− 24.5%) compared to high-Zn accessions (− 14.2%), which would indicate that

using brown rice samples would not have induced a bias in favor of high-Zn accessions in our analyses.

Differences between rice sub-populations

The association panel used was predominantly of the *indica* sub-population with smaller additions from the *aus*, *japonica*, *aromatic* and *admix* groups, and measured grain Zn concentrations indicated significantly higher concentrations in the *aus* group (Fig. 2). The average of the *aus* group was 35.6% higher compared to the *indica* group and this superiority could subsequently be substantiated (+35.4%) in the confirmatory set including a much larger proportion of *aus* accessions. Considering that *indica* are the predominant group of varieties grown by lowland rice farmers of Madagascar and many other biofortification target countries, *aus* accessions identified here experimentally or through GP should be considered as potential donors in biofortification programs. Norton et al. (2014) reached a similar conclusion as 3 of the 5 high-Zn donor accessions identified in that study belonged to the *aus* group.

The inclusion or omission of *aus* accessions did not affect the ability to predict grain Zn in non-*aus* groups but omitting the *aus* led to a strong under-estimation of grain Zn in that group (Figs. 2 and 4). Thus, some largely *aus*-specific genetic factors must exist that lead to the superior grain Zn in that group and a GP approach appears to accurately consider these. A similar conclusion about the need to include all sub-populations in the training set was reached by Grenier et al. (2015). Considering that many component traits lead to high grain Zn, it is likely that some of these component traits are only or at least predominantly expressed in the *aus* group. Such traits would make ideal breeding targets and physiological studies need to investigate if such *aus*-specific traits exist and which limiting step in the processes between Zn uptake, transport, retranslocation and endosperm loading they affect.

Understanding such bottleneck traits and the underlying genetic control may be key to increasing grain Zn concentrations in the predominantly *indica* modern cultivars. One possible aspect to study further in this regard is the tendency of *aus* accessions like IRIS313-9368 to rapidly senesce at maturity, whereas modern *indica* varieties have the tendency to remain comparatively green at maturity. More rapid senescence could favor Zn remobilization and translocation to grains.

Donors from the *aus* group have been used repeatedly to move alleles for tolerance to many abiotic stresses into the modern rice breeding pool (Heredia et al. 2021). It is furthermore interesting to note that several *aus* accessions possess high tolerance to Zn deficiency (Lee et al. 2018) due to their efficient Zn uptake capacity from highly reduced soil. While it is not known whether a link between this Zn uptake efficiency from Zn deficient soil and high grain Zn concentrations under normal Zn availability exists, it would be very interesting to pursue such a possibility further. The use of *aus* donors in rice breeding has typically involved the transfer of major genes or QTL through their marker-assisted introgression into modern breeding lines and to what extent this is a likely approach to improve grain Zn remains to be resolved.

GWA and GP analysis

Results of Norton et al. (2014) and Swamy et al. (2018) suggest that grain Zn is a polygenic trait controlled by multiple small to medium effect loci. Employing a multi-locus GWAS approach should therefore be more suitable in identifying loci controlling grain Zn concentrations compared to single-locus models that test one locus at a time without considering interactions between loci (Xu et al. 2018). A weakness of the single-locus GWAS analysis is the problem of false positives and negatives and this is better balanced in the multi-locus association analysis employed here, which eliminates the need for a Bonferroni correction in multi-locus GWAS (Wang et al. 2016).

This study identified eight QTN of which seven had minor effects while the QTN on chromosome 11 (28,757,650) can be considered a medium-effect locus. It is attributed to a rare allele present in accessions of the *aus*, *japonica* and *indica* sub-species and the difference in average grain Zn concentration between the minor (32.1 µg Zn g^{-1}) and major (23.2 µg Zn g^{-1}) allele groups at this QTN is 8.9 µg Zn g^{-1} (+38%). This contrasts with the estimated QTN effect of 3.69 µg Zn g^{-1} obtained by the multi-locus analysis. This would indicate a strong over-estimation of QTN effects if individual loci are investigated in isolation and that the multi-locus model may provide lower but more realistic estimates of QTN effects.

Other QTL or GWA studies have identified loci on chromosome 11 (summarized by Swamy et al. 2016, 2018) but these do not overlap with QTN_11.3 identified here. Conversely, we did not detect an otherwise commonly identified locus on chromosome 7 that is potentially linked to the *OsNAS3* gene considered a prime candidate for increasing grain mineral concentration (Johnson et al. 2011). Predicted gene models for the candidate region at QTN_11.3 (28.681–28.798 Mbp) did not contain genes previously associated with Zn metabolism or transport. Instead, a cluster of 11 genes belonging to the glucosyl hydrolase family (class III chitinase homologs or xylanase inhibitors) were present. However, it is beyond the scope of this paper to further analyze any potential role of these genes.

Whereas several QTL and GWAS studies have been reported in the literature this is only the second report applying GP for grain Zn concentrations in rice. In

contrast to the work by Baertschi et al. (2021) that focused on assessing the potential of GP models for early selection of families to improve upland rice synthetic populations, the aim of the GP approach taken here is to predict grain Zn concentrations of gene bank accessions in lowland rice. If successful, this would allow for a very efficient search of potential new donors for high grain Zn concentrations. The prediction accuracy of 0.51 achieved in this study is similar to the PA ranging from 0.33 to 0.69 in wheat (Velu et al. 2016), of 0.43–0.73 reported for maize (Mageto et al. 2020) and of 0.51 in upland rice (Baertschi et al. 2021). The present study was conducted in two low-input farmers' fields rather than under the more controlled conditions one may encounter on research farms and that the PA achieved here is comparable to PAs reported from research farms is further suggestive of GP being a suitable approach for identifying potential donors from gene banks. It is furthermore noteworthy that the GP model based on field data from Madagascar was able to reliably predict ($r = 0.74$) grain Zn concentrations of the confirmation set that had been grown on the IRRI farm in the Philippines, which represents a more favorable environment compared to the low-input farmers' fields providing the data for the training set. Thus, the confirmation outside the training environment lends further credibility to the predictive ability of GP for grain Zn concentrations.

Such robustness across environments would offer options to further economize resources through sparse testing of only part of the entire training set at each site or environment. Baertschi et al. (2021) suggested optimizing the GP scheme by evaluating small training sets and using phenotypic correlation between sites to calibrate the model, and in their case the phenotypic correlation for grain zinc concentration between sites ($r^2 > 0.41$) was similar to what was achieved here.

In a review of genomic approaches to biofortification of cereals, Koç and Karayigit (2021) concluded that conventional breeding would be the most sustainable, low cost and easily adoptable strategy. Our results concur inasmuch none of the QTN identified would be influential enough to be rapidly employed in marker-assisted selection. However, the success of GP in predicting grain Zn concentrations, here of gene bank accessions, but elsewhere in a rice breeding population (Baertschi et al. 2021), may offer opportunities, especially where genomic selection of other traits is already practiced. It should furthermore facilitate utilizing the high-Zn donors identified here, as breeders may be reluctant to employ such exotic material in a conventional elite breeding program. As efforts are under way to mainstream biofortification traits in crop breeding (Virk et al. 2021), it seems worthwhile to include grain Zn concentrations as one of the traits targeted in genomic selection.

Conclusions

Data obtained from field experiments conducted in Madagascar enabled us to successfully predict grain Zn concentrations among a set of gene bank accessions, thereby identifying potential donors for use in Zn biofortification breeding. The most promising donors all belonged to the *aus* sub-species of rice and to significantly increase Zn concentrations in the lowland rice breeding pool, which is predominantly belonging to the *indica* sub-species, it appears necessary to rely on *aus* introgressions. Donor of the *aus* group has been used repeatedly for the introgression of major abiotic stress tolerance loci through their marker-assisted introgression. This approach is less likely to be successful for the improvement of grain Zn concentrations as none of the identified loci identified here or elsewhere appear strong enough to raise grain Zn concentrations by the targeted 50% or more. Being a polygenic trait, the improvement of grain Zn concentrations would likely require the transfer of many small-effect loci simultaneously. Since we have shown the suitability of GP in identifying high-Zn donors, it can be expected that breeding populations developed from such donors could achieve target grain Zn concentrations if a similar genomic selection approach was used during the variety development process.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-022-04110-2.

Acknowledgements This research was supported by the Science and Technology Research Partnership for Sustainable Development (SATREPS), Japan Science and Technology Agency (JST)/Japan International Cooperation Agency (JICA)—Grant No. JPMJSA1608; and the CGIAR Research Program on Rice Agri-Food Systems (RICE). The authors thank the IRRI gene bank for providing seeds of accessions used.

Author contribution statement MR, RT, CG and MW designed the study and wrote the manuscript, RT implemented the GP model, MR, RT, CG and MW discussed results, JS provided the Zn analysis data, JPT processed SNP and GWAS data, and MR, JPT, MW conducted field experiments and collected and summarized all data.

Funding This research was supported by the Science and Technology Research Partnership for Sustainable Development (SATREPS), Japan Science and Technology Agency (JST)/Japan International Cooperation Agency (JICA)—Grant No. JPMJSA1608. The grain Zn analysis was supported by HarvestPlus funding to JS.

Availability of data and materials Genotypic data used in the study are publicly available at https://snp-seek.irri.org/_snp.zul.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Baertschi C, Cao T-V, Bartholome J, Ospina Y, Quintero C, Frouin J, Bouvet JM, Grenier C (2021) Impact of early genomic prediction for recurrent selection in an upland rice synthetic population. G3. https://doi.org/10.1093/g3journal/jkab320

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21(2):263–265. https://doi.org/10.1093/bioinformatics/bth457

Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci 34:20–25

Bouis HE, Hotz C, McClafferty B, Meenakshi JV, Pfeiffer WH (2011) Biofortification: a new tool to reduce micronutrient malnutrition. Food Nutr Bull 32:S31–S40. https://doi.org/10.1177/15648265110321S105

Bouis HE, Saltzman A (2017) Improving nutrition through biofortification. a review of evidence from HarvestPlus, 2003 through 2016. Glob Food Secur 12:49–58

Bouis HE, Welch RM (2010) Biofortification— a sustainable agricultural strategy for reducing micronutrient malnutrition in the global south. Crop Sci 50:20–32

Broadley MR, White PJ, Hammond JP, Zelko I, Lux A (2007) Zinc in plants. New Phytol 173:677–702. https://doi.org/10.1111/j.1469-8137.2007.01996.x

Browning BL, Browning SR (2016) Genotype imputation with millions of reference samples. Am J Hum Genet 98:116–126. https://doi.org/10.1016/j.ajhg.2015.11.020

Cu ST, Warnock NI, Pasuquin J, Dingkuhn M, Stangoulis JCR (2021) A high-resolution genome-wide association study of the grain ionome and agronomic traits in rice Oryza sativa subsp indica. Sci Rep. https://doi.org/10.1038/s41598-021-98573-w

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250–255

Galetti V (2018) Zinc Deficiency and Stunting. In: Preedy V, Patel V (eds) Handbook of famine starvation, and nutrient deprivation. Springer, Cham. https://doi.org/10.1007/978-3-319-40007-5_93-1

Goloran JB, Johnson-Beebout SE, Morete MJ, Impa SM, Kirk GJD, Wissuwa M (2019) Grain Zn concentrations and yield of Zn-biofortified versus Zn-efficient rice genotypes under contrasting growth conditions. Field Crop Res 234:26–32

Grenier C, Cao TV, Ospina Y, Quintero C, Châtel MH, Tohme J, Courtois B, Ahmadi N (2015) Accuracy of genomic selection in a rice synthetic population developed for recurrent selection breeding. PLoS One 10:e0136594

Harvest Plus (2021) Biofortification prioritization index. https://www.harvestplus.org/knowledge-market/BPI. Accessed 27 Jul 2021

Heredia MC, Kant J, Prodhan A, Dixit S, Wissuwa M (2021) Breeding rice for a changing climate by improving adaptations to water saving technologies. Theor Appl Genet. https://doi.org/10.1007/s00122-021-03899-8

Higuchi K, Takahashi M, Nakanishi H, Kawasaki S, Nishizawa NK, Mori S (2001) Analysis of transgenic rice containing barley nicotianamine synthase gene. Soil Sci Plant Nutr 47:315–322

Holland JB, Nyguist WE, Cervantes-Martínez CT (2003) Estimating and interpreting heritability for plant breeding: an update. Plant Breed Rev 22:11–112

Inoue H, Higuchi K, Takahashi M, Nakanishi H, Mori S, Nishizawa NK (2003) Three rice nicotianamine synthase genes, *OsNAS1*, *OsNAS2*, and *OsNAS3* are expressed in cells involved in long-distance transport of iron and differentially regulated by iron. Plant J 36:366–381

Jiang W, Struik PC, van Keulen H, Zhao M, Jin LN, Stomph TJ (2008) Does increased Zn uptake enhance grain Zn mass concentration in rice? Ann Appl Biol 153:135–147. https://doi.org/10.1111/j.1744-7348.2008.00243.x

Johnson AAT, Kyriacou B, Callahan DL, Carruthers L, Stangoulis JCR, Lombi E, Tester M (2011) Constitutive overexpression of the *OsNAS* gene family reveals single gene strategies for effective iron- and zinc-biofortification of rice endosperm. PLoS One 6(9):e24476. https://doi.org/10.1371/journal.pone.0024476

Johnson-Beebout SE, Goloran JB, Rubianes FHC, Jacob JDC, Castillo OB (2016) Zn uptake behavior of rice genotypes and its implication on grain Zn biofortification. Sci Rep 6:38301. https://doi.org/10.1038/srep38301

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. Genet 178:1709–1723. https://doi.org/10.1534/genetics.107.080101

Koç E, Karayiğit B (2021) Assessment of biofortification approaches used to improve micronutrient-dense plants that are a sustainable solution to combat hidden hunger. J Soil Sci Plant Nutr. https://doi.org/10.1007/s42729-021-00663-1

Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmertest package: tests in linear mixed effects models. J Stat Softw 82:1–26. https://doi.org/10.18637/jss.v082.i13

Lee JS, Wissuwa M, Zamora OB, Ismail AM (2018) Novel sources of aus rice for zinc deficiency tolerance identified through association analysis using a high-density SNP array. Rice Sci 25:293–296. https://doi.org/10.1016/j.rsci.2018.08.004

Mageto EK, Crossa J, Rodríguez PP, Dhliwayo T, Palacios-Rojas N, Lee M, Guo R, San Vicente F, Zhang X, Hindu V (2020) Genomic prediction with genotype by environment interaction analysis for kernel zinc concentration in tropical maize germplasm. G3 10:2629–2639. https://doi.org/10.1534/g3.120.401172

Mansueto L, Fuentes RR, Borja FN, Detras J, Abriol-Santos JM, Chebotarov D, Sanciangco M, Palis K, Copetti D, Poliakov A, Dubchak I, Solovyev V, Wing RA, Sackville Hamilton R, Mauleon R, McNally KL, Alexandrov N (2017) Rice SNP-seek database update: new SNPs, indels, and queries. Nucleic Acids Res 45:D1075–D1081. https://doi.org/10.1093/nar/gkw1135

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Mori A, Kirk GJD, Lee JS, Morete MJ, Nanda AK, Johnson-Beebout SE, Wissuwa M (2016) Rice genotype differences in tolerance of zinc-deficient soils: evidence for the importance of root-induced changes in the rhizosphere. Front Plant Sci 6:1160. https://doi.org/10.3389/fpls.2015.01160

Norton GJ, Douglas A, Lahner B, Yakubova E, Guerinot ML, Pinson SRM et al (2014) Genome wide association mapping of grain arsenic, copper, molybdenum and zinc in rice (Oryza sativa L.) grown at four international field sites. PLoS One 9(2):e89685. https://doi.org/10.1371/journal.pone.0089685

Palmer LJ, Palmer LT, Rutzke MA, Graham RD, Stangoulis JCR (2014) Nutrient variability in phloem: examining changes in K,

Mg, Zn and Fe concentration during grain loading in common wheat (*Triticum aestivum*). Physiol Plant 152:729–737

Rao SD, Neeraja CN, Madhu Babu P, Nirmala B, Suman K, Rao LVS, Surekha K, Raghu P, Longvah T, Surendra P, Kumar R, Babu VR, Voleti SR (2020) Zinc biofortified rice varieties: challenges, possibilities, and progress in India. Front Nutr 7:26. https://doi.org/10.3389/fnut.2020.00026

Roohani N, Hurrell R, Kelishadi R, Schulin R (2013) Zinc and its importance for human health: an integrative review. J Res Med Sci 18:144–157

Shiratori S, Nishide A (2018) Micronutrient supply based on the food balance sheet and the prevalence of inadequate intakes in madagascar. Proc Nutr Soc 77(OCE3):E70

Stewart CP, Fernald LCH, Weber AM, Arnold C, Galasso E (2020) Lipid-based nutrient supplementation reduces child anemia and increases micronutrient status in sadagascar: a multiarm cluster-randomized controlled trial. J Nutr 150:958–966. https://doi.org/10.1093/jn/nxz320

Suman K, Neeraja CN, Madhubabu P, Rathod S, Bej S, Jadhav KP et al (2021) Identification of promising rils for high grain zinc through genotype × environment analysis and stable grain zinc QTL using SSRS and SNPS in rice (Oryza sativa L.). Front Plant Sci 12:587482

Swamy BPM, Rahman MA, Inabangan-Asilo MA, Amparado A, Manito C, Chadha-Mohanthy P, Reinke R, Slamet-Loedin IH (2016) Advances in breeding for high grain zinc in rice. Rice 9:49

Swamy BPM, Kaladhar K, Anuradha K, Batchu AK, Longvah T, Sarla N (2018) QTL analysis for grain iron and zinc concentrations in two O. nivara derived backcross populations. Rice Sci 25(4):197–207

Tanaka R, Mandaharisoa ST, Rakotondramanana M, Ranaivo HN, Pariasca-Tanaka J, Kanegae HK, Iwata H, Wissuwa M (2021) From gene banks to farmer's fields: using genomic selection to identify donors for a breeding program in rice to close the yield gap on smallholder farms. Theor Appl Genet 134:3397–3410. https://doi.org/10.1007/s00122-021-03909-9

The World Bank (2016) Addressing chronic malnutrition in Madagascar. malnutrition-in-Madagascar. https://www.worldbank.org/en/programs/sief-trust-fund/brief/addressing-chronic-malnutrition-in-madagascar. Accessed 20 July 2021

UNICEF (2019) The state of the world's children 2019: children, food and nutrition. Unicef, New York

Velu G, Crossa J, Singh RP, Hao Y, Dreisigacker S, Perez-Rodriguez P, Joshi AK, Chatrath R, Gupta V, Balasubramaniam A, Tiwari C, Mishra VK, Sohu VS, Mavi GS (2016) Genomic prediction for grain zinc and iron concentrations in spring wheat. Theor Appl Genet 129:1595–1605. https://doi.org/10.1007/s00122-016-2726-y

Virk PS, Andersson MS, Arcos J, Govindaraj M, Pfeiffer WH (2021) Transition from targeted breeding to mainstreaming of biofortification traits in crop improvement programs. Front Plant Sci 12:703990

Wang SB, Feng JY, Ren WL, Huang B, Zhou L, Wen YJ et al (2016) Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. Sci Rep 6:19444. https://doi.org/10.1038/srep19444

Wheal MS, Fowles TO, Palmer LT (2011) A cost-effective acid digestion method using closed polypropylene tubes for inductively coupled plasma optical emission spectrometry (ICP-OES) analysis of plant essential elements. Anal Methods 3:2854–2863

Wissuwa M, Ismail AM, Graham RD (2008) Rice grain zinc concentrations as affected by genotype native soil-zinc availability and zinc fertilization. Plant Soil 306:37–48

World Food Programme (2010) Fighting hunger worldwide. https://documents.wfp.org/stellent/groups/public/documents/communications/wfp220666.pdf. Accessed 27 Jul 2021

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. https://doi.org/10.3389/fpls.2018.01311

Yasmin Z, Paltridge N, Graham R, Huynh B-L, Stangoulis J (2014) Measuring genotypic variation in wheat seed iron first requires stringent protocols to minimize soil iron contamination. Crop Sci 54:255–264

Zhang Y, Liu P, Zhang X, Zheng Q, Chen M, Ge F et al (2018) Multi-locus genome-wide association study reveals the genetic architecture of stalk lodging resistance-related traits in maize. Front Plant Sci 9:611. https://doi.org/10.3389/fpls.2018.00611