Seed Science Research

cambridge.org/ssr

Genetic markers associated with seed longevity and vitamin E in diverse Aus rice varieties

Jae-Sung Lee[1], Jieun Kwak[2] and Fiona R. Hay[3]

[1]T.T. Chang Genetic Resources Center, Strategic Innovation Platform, International Rice Research Institute, Los Baños, College, Laguna 4031, Philippines; [2]National Institute of Crop Science, Rural Development Administration, Suwon, Gyunggi-do, Republic of Korea and [3]Department of Agroecology, Aarhus University, Forsøgsvej 1, 4200 Slagelse, Denmark

Research Paper

Cite this article: Lee J-S, Kwak J, Hay FR (2020). Genetic markers associated with seed longevity and vitamin E in diverse Aus rice varieties. *Seed Science Research* **30**, 133–141. https://doi.org/10.1017/S0960258520000173

Received: 24 October 2019
Revised: 5 March 2020
Accepted: 11 May 2020
First published online: 8 July 2020

Key words:
Aus variety group; GWAS; harvest moisture content; *Oryza sativa* L; seed longevity; vitamin E

Author for correspondence:
Fiona R. Hay, E-mail: fiona.hay@agro.au.dk

Abstract

Vitamin E is known to scavenge lipid peroxy radicals and has a purported role in preventing seed deterioration during storage. In our previous studies using 20 rice varieties from different variety groups, the specific ratio of vitamin E homologues rather than total vitamin E content was associated with seed longevity. To validate this result, we extended the experiment to a rice panel composed of 185 Aus (semi-wild rice) varieties. Seed longevity values were determined through storage experiments at 45°C and 10.9% seed moisture content (MC). Eight types of vitamin E homologues (α-, β-, γ- and δ-tocopherol/tocotrienol) were quantified by ultra-performance liquid chromatography. The theoretical initial viability in NED, K_i, was positively correlated with γ- and δ-tocopherols and negatively correlated with α-tocotrienol. The time for viability to fall to 50% during storage at elevated temperature and relative humidity, p_{50}, was positively correlated with δ-tocopherol. The harvest MC was negatively correlated with all seed longevity traits. Taking this factor into account in a genome-wide association (GWA) analysis, we were able to correct false positives. A consistent major peak on chromosome 4 associated with $-\sigma^{-1}$ was detected with a mixed linear analysis. Based on rice genome annotation and gene network ontology databases, we suggest that RNA modification, oxidation–reduction, protein–protein interactions and abscisic acid signal transduction play roles in seed longevity extension of Aus rice. Although major GWA regions were not overlapped across traits, three genetic markers, on chromosomes 1, 3 and 4, were associated with both δ-tocopherol and K_i and two markers on chromosome 1 and 8 were associated with both δ-tocopherol and p_{50}.

Introduction

Some of the recent research on the International Rice Genebank Collection (IRGC) at the International Rice Research Institute (IRRI) has focused on understanding and improving the longevity of rice seeds in storage, to ultimately increase the efficiency and effectiveness of genebank operations. Some innovations include automated seed sorting to get seeds into the genebank cold rooms sooner after harvest; changed drying conditions from immediate cool drying (15°C) to high-temperature drying (40°C) prior to cool drying, which improved the longevity of seeds, particularly when harvested with high moisture content (MC) (Whitehouse et al., 2015, 2017, 2018; Timple and Hay, 2018); confirmation of 5-year monitoring intervals for seeds stored at 5°C (medium-term) (Hay et al., 2015); confirmed correlation between seed longevity and vitamin E homologue ratios, rather than total vitamin E content (Lee et al., 2017, 2019a) and through a genome-wide association (GWA) study of diverse Indica rice varieties, identification of eight loci that are related to longevity parameters for seeds of Indica rice varieties stored at 45°C and 10.9% MC (Hay et al., 2019; Lee et al., 2019b). Among these aspects, attention has been given to vitamin E α-, β-, γ- and δ-tocopherol/tocotrienol homologues, due to their antioxidant role. During seed storage under hot and humid conditions, oxidative stress and in particular lipid peroxidation occurs and results in membrane breakdown and cell ageing (Sattler et al., 2004). Vitamin E effectively scavenges lipid peroxy radicals, thus prolonging seed longevity. Indica is perhaps the largest variety group of cultivated rice, certainly the most represented within the IRGC, and widely grown in the lowlands of tropical Asia as well as some areas in the USA and Latin America (Londo et al., 2006; Wang et al., 2018). In contrast, Aus is a smaller variety group, mainly grown in Bangladesh and India during the March–August season (Khush, 1997). Aus varieties are considered more closely related to wild rice species (Londo et al., 2006) with unfavourable characteristics such as excessively tall plant height, small grain size and high shattering (Lee et al., 2018). However, they are a potentially important source of genes conferring tolerance to abiotic stresses such as drought and zinc deficiency (Bin Rahman and Zhang, 2018; Lee et al., 2018). The Aus group is less well described than the Indica and Japonica variety groups, and there is little information about the relative longevity of seeds from Aus varieties.

CAMBRIDGE UNIVERSITY PRESS

Table 1. Phenotypic traits measured for 185 Aus rice accessions

No.	Trait	Description	Range	Mean (SD)	Coefficient of variation (%)
1	K_i	Initial viability in normal equivalent deviates (NED)	0.84–8.36	3.28 (1.28)	38.98
2	$-\sigma^{-1}$	σ is the length of time (d) for viability to fall by 1 NED at 10.9% MC and 45°C	−0.72 – −0.06	−0.22 (0.09)	−40.94
3	p_{50}	Length of time for viability to fall to 50% at 10.9% MC and 45°C	2.43–54.20	16.90 (8.94)	52.90
4	Days to flowering	Days from sowing to flowering	66.00–121.00	80.01 (12.45)	15.56
5	Harvest moisture content	Seed moisture content at harvest (% fresh weight)	12.35–34.28	22.33 (3.26)	14.58
6	Grain colour	Whitish (1), gold (3), brown (5), red (7), purple (9), black (11)	1.00–11.00	3.49 (3.19)	91.25
7	Pericarp colour	Pure white (1), white (1.5), brown (2), pale red (2.5), red (3), dark red (3.5)	1.00–3.50	2.48 (0.72)	28.98
8	Vitamin E content	mg g^{-1} oil from brown rice	0.29–0.66	0.51 (0.08)	16.00

In this study, we first screened a diverse Aus rice panel for seed longevity and conducted GWA analysis for QTL detection. Secondly, we tested previous conclusions (Lee et al., 2019a) of a positive correlation between rice seed longevity and the proportion of γ-type vitamin E, and a negative correlation between seed longevity and the proportion of α- and β-types of vitamin E.

Materials and methods

Plant materials

A total of 185 Aus rice accessions that were included in 'The 3000 Rice Genomes Project' (2014) were grown at the Zeigler Experiment Station of the International Rice Research Institute (IRRI), Philippines, in the 2016 wet season. Twenty-day-old seedlings were transplanted to the field with 200 mm spacing of plants within and between rows. Seed lots were harvested at 35 d after peak flowering. Harvest MC was determined by weighing three replicate samples of ground seeds before and after 2 h oven-drying at 130°C followed by 1 h over silica gel at room temperature (ISTA, 2018). Grain colour and pericarp colour were scored visually (Table 1). After blowing and sorting to discard immature, damaged, off-type or diseased seeds, seed lots were dried at 15°C, 15% relative humidity (RH) and then transferred to foil-laminate bags and stored at −20°C until the seed storage experiments commenced.

Seed storage experiments

Foil bags were removed from −20°C storage and allowed to equilibrate to room temperature before opening and removal of seeds. Seed water content was adjusted at 60% RH, 25°C in a climate test chamber (Model VC3 0034-M; Vötschtechnik, Germany) to achieve a target MC of 10.9% (fresh weight basis) and then seeds were sealed inside foil-laminate bags and placed at 45°C. Samples were taken at 0, 7 and 14 d, then at 4 or 10 d intervals for up to 42 or 60 d of storage. For each sample, two replicates of 30 seeds were placed on two layers of Whatman No. 1 filter paper with 7 ml distilled water in 90 mm-diameter Petri dishes. Seeds were incubated at 30°C with 12 h light per day and germination was scored for up to 21 d after sowing (0 d-storage sample were scored daily and other storage treatments were scored at 5, 9, 14 and 21 d). MC was determined as above, at the outset, during and end of storage using the seeds from three further packets at each sampling time. Probit analysis of germination data was performed using GenStat version 18 (VSN International Ltd, Hemel Hempstead, UK). Seed longevity parameters from the Ellis and Roberts (1980) viability equation were estimated: K_i, the initial viability in normal equivalent deviates (NED); and $-1/\sigma$, the slope of the transformed survival curve (i.e. change in viability in NED d^{-1}). The p_{50}, the time (d) for viability to fall to 50% during storage at 10.9% MC and 45°C was also estimated.

Vitamin E analysis

Vitamin E in de-hulled (palea and lemma removed) seeds was analysed following Ko et al. (2003). Briefly, the ground sample (50 g) was mixed with 300 ml *n*-hexane for 2 h, then concentrated by evaporating hexane using nitrogen gas. The lipid extract (0.5 g) was mixed with 2 ml 5% pyrogallol solution in ethanol and 20 ml ethanol. After boiling at 70°C, 1 ml of 50% aqueous KOH was added for the 5 min saponification. The sample was extracted by 50 ml diethyl ether, washed with 20 ml distilled water, filtered through anhydrous sodium sulphate and evaporated at 30°C. The residue was diluted with 10 ml *n*-hexane and filtered through a Millipore 0.2 µm membrane. Individual vitamin E homologues were quantified by ultra-performance liquid chromatography (UPLC, H-Class System, Waters, Massachusetts, USA) at 298 nm excitation and 325 nm emission with a Lichrospher Si-60 column (250 × 4.6 mm i.d.; Merck Co., Gernsheim, Germany). Descriptive statistics of all traits and correlations were analysed using STAR v2.0.1 (International Rice Research Institute).

GWA analysis

GWA analysis was performed using TASSEL 5.2.7 (Bradbury et al., 2007) based on the 446k (filtered for 20% missing data and minor allele frequencies <5%) single nucleotide polymorphisms (SNP) marker data of The 3000 Rice Genomes Project (Wang et al., 2018). A mixed linear model (MLM) with a kinship matrix was applied and plots above a significance threshold ($P < 3.00 \times 10^{-5}$) were considered as major QTL associated with the

https://doi.org/10.1017/S0960258520000173 Published online by Cambridge University Press

traits. For haplotype analysis, SNP markers in the region of major QTL were extracted from the Rice SNP-Seek Database (http://snp-seek.irri.org) containing genotype data of The 3000 Rice Genomes Project (Wang et al., 2018) and effects of haplotype on the phenotype were determined.

Results

Due to variation in flowering date (66–121 d from sowing to flowering), accessions were harvested on different dates (between 6 August and 16 October) under different weather conditions. The harvest MC (% fresh weight) ranged between 12.4 and 34.3%.

The mean MC of the seeds during storage at 45°C in foil bags was 11.09% (SD 0.39) (three replicate observations per accession on each of three occasions) and did not vary significantly depending on sampling time.

Many accessions showed evidence of dormancy at the start of storage, with germination increasing during the early phase of storage, before declining as the seeds aged and lost viability (Supplementary Figs S1–S13). Because there were, in general, insufficient data to characterize the breaking of dormancy, the probit analysis to describe the loss of viability only used data in which germination was declining. Some accessions lost viability relatively quickly, with little or no germination after 20 d in storage (e.g. IRGC 120876, 120915, 125757, 126123, 127170, 127236, 127528, 128297); other seed lots showed some germination after 50 d in storage (e.g. IRGC 127130, 127228, 127481, 127652, 128369, 132306). Thus, there was a wide variation in the results of the probit analysis for each accession.

The initial seed lot viability, K_i, ranged between 0.84 (IRGC 127796) and 8.36 NED (IRGC 127652) with a mean of 3.28 NED (Table 1). The slope of the survival curves, $-\sigma^{-1}$ ranged between −0.72 (IRGC 127253) and −0.06 (IRGC 127903) NED d^{-1} with a mean of −0.22 NED d^{-1}. The estimate of p_{50} ranged between 2.43 (IRGC 127454) and 54.20 (IRGC 127652) d with a mean of 16.90 d. A total of 90 accessions composed of 30 high, 30 intermediate and 30 poor seed longevity were selected for vitamin E analysis. Total vitamin E content in de-hulled seeds ranged between 0.29 and 0.66 mg g^{-1} oil (accessions IRGC 128442 and 127139, respectively), with a mean of 0.51 mg g^{-1} oil.

There was no significant correlation between total vitamin E content and seed longevity in this (data not shown) and previous studies (Lee et al., 2017, 2019b). We further tested correlations between proportions of each vitamin E homologue and seed longevity. K_i was positively correlated with γ- and δ-tocopherols (correlation coefficient, $r = 0.254$ ($P < 0.05$) and 0.282 ($P < 0.01$), respectively) and negatively correlated with α-tocotrienol ($r = -0.245$; $P < 0.01$) (Table 2). The parameter p_{50} was positively correlated with δ-tocopherol ($r = 0.233$; $P < 0.05$). The parameter $-\sigma^{-1}$ did not correlate with any vitamin E homologues. Pericarp colour (scores between 0 and 3.5 from pure white to dark red) was positively correlated with all seed longevity values ($P < 0.001$ for K_i and $-\sigma^{-1}$ or $P < 0.0001$ for p_{50}). Seed harvest MC (% fresh weight) was strongly negatively correlated with all seed longevity values, especially p_{50} ($r = -0.600$; $P < 0.0001$) (Table 2 and Fig. 1A). Seed harvest MC was also negatively correlated with days to flowering ($r = -0.334$; $P < 0.0001$), which indicates the tendency of decline in harvest MC in later-maturing seeds.

GWA analysis was conducted for seed longevity and other associated traits such as α-tocotrienol, δ-tocopherol and pericarp colour (Fig. 2; Supplementary Fig. S14). The initial Manhattan plot on p_{50} showed one significant single marker on chromosome 5 and a large consistent peak on chromosome 6 (Fig. 1B). As harvest MC was an environmental factor strongly affecting seed longevity, we took this factor into account in the analysis to correct false positives (Figs 1C and 2). Three of the most significant single markers (P-value $< 3.00 \times 10^{-5}$) on chromosomes 4 and 5 were associated with K_i (Fig. 2 and Table 3). Eight significant single markers on chromosomes 2, 3, 5, 6, 7, 9 and 11 were associated with p_{50}. Two single markers on chromosome 2 and a consistent major peak on chromosome 4 were associated with $-\sigma^{-1}$. These significant markers were located in genes or in the range of 3549 basepair (bp) downstream and 6505 bp upstream of the nearest genes (Table 3). Major loci associated with each trait differed (Fig. 2). However, a number of minor loci (P-value $> 3.00 \times 10^{-5}$) were associated with multi-traits (Table 4). Seven markers on chromosomes 2, 6, 8, 10, 11 and 12 associated with both K_i and p_{50}, with P-values between 7.78×10^{-4} and 1.71×10^{-5} (Table 4). These markers were located in the genes or in the range of 7740 and 289 bp downstream of the nearest gene. Three markers on chromosomes 1, 3 and 4 were associated with both δ-tocopherol and K_i (Fig. 3). Two markers on chromosomes 1 and 8 were associated with both δ-tocopherol and p_{50}. The allelic effects of those markers on phenotype values were estimated through haplotype mapping. The presence of favourable haplotypes (green shaded) was associated with enhancements of δ-tocopherol and K_i. Genotype no. 1 with favourable haplotypes on all markers had higher δ-tocopherol (120%) and K_i (67%) compared with genotype no. 5 which lacked favourable haplotypes (Fig. 3A). The favourable haplotypes with the two markers on chromosomes 1 and 8 (genotype no. 1) enhanced both δ-tocopherol and p_{50}, by 100 and 116%, respectively, when compared with the negative group (genotype no. 3) (Fig. 3B).

Discussion

Seed longevity is a complex trait, with high phenotypic plasticity (Leprince et al., 2017). This makes it difficult to identify the molecular basis of this trait. For example, our recent study using seeds of diverse Indica rice varieties produced during the 2015 dry season at IRRI identified that harvest MC, which depends on the environmental conditions at the time of harvest once seeds are in the late maturation phase, was more strongly correlated with p_{50} (r-value −0.344; $P < 0.0001$) than any genetic trait (Lee et al., 2019b). The current study, using seeds of Aus varieties produced in the 2016 wet season, showed a stronger correlation ($r = -0.600$; $P < 0.0001$) between harvest MC and p_{50} (Table 2). This led to the identification of loci which were a consequence of harvest MC and hence progression through the late maturation phase, rather than being associated with longevity *per se* (Fig. 1B). Another factor that significantly affects seed longevity is the post-harvest drying conditions. Whitehouse et al. (2018) reported that subsequent seed longevity at 10.9% MC and 45°C is improved for seeds that are harvested with a high MC, if seeds are initially dried at a higher temperature (45–60°C) prior to routine cool (15°C) drying. Therefore, the effect of seed drying condition on subsequent seed longevity may also interfere with the genetic analysis of this trait. For accurate QTL analysis, such environmental factors must be controlled or well addressed in genetic analysis. Thus, while we have attempted to identify genes with a role in Aus seed longevity, for gene validation purposes, it would be important to screen accessions produced

https://doi.org/10.1017/S0960258520000173 Published online by Cambridge University Press

Table 2. Correlation coefficients (Spearman's) for the traits potentially associated with seed longevity in a diverse Aus rice panel

							Proportion of vitamin E homologues							
	$-\sigma^{-1}$	p_{50}	Days to flowering	*Harvest moisture content*	Grain colour	Pericarp colour	α-T	β-T	γ-T	δ-T	α-T_3	β-T_3	γ-T_3	δ-T_3
K_i	−0.245***	0.551****	0.044^{ns}	*−0.365****	0.041^{ns}	0.254***	-0.173^{ns}	-0.135^{ns}	0.254*	0.282**	−0.245**	0.051^{ns}	0.102^{ns}	0.102^{ns}
$-\sigma^{-1}$		0.591****	0.224**	*−0.389****	0.152*	0.260***	0.133^{ns}	0.153^{ns}	-0.093^{ns}	-0.034^{ns}	0.119^{ns}	-0.018^{ns}	0.048^{ns}	0.098^{ns}
p_{50}			0.207**	*−0.600****	0.183*	0.415****	-0.025^{ns}	-0.010^{ns}	0.153^{ns}	0.233*	-0.104^{ns}	0.054^{ns}	0.079^{ns}	0.158^{ns}
Days to flowering				*−0.334****	0.107^{ns}	0.089^{ns}	0.066^{ns}	-0.110^{ns}	-0.028^{ns}	0.038^{ns}	-0.069^{ns}	-0.171^{ns}	0.084^{ns}	-0.014^{ns}
Harvest moisture content					*−0.025*^{ns}	*−0.243****	*−0.040*^{ns}	*−0.021*^{ns}	*0.116*^{ns}	*0.063*^{ns}	*−0.015*^{ns}	*0.165*^{ns}	*−0.036*^{ns}	*−0.184*^{ns}
Grain colour						0.315****	-0.170^{ns}	-0.027^{ns}	0.187^{ns}	0.170^{ns}	-0.142^{ns}	0.116^{ns}	0.030^{ns}	0.327**
Pericarp colour							-0.110^{ns}	0.091^{ns}	0.160^{ns}	0.224*	-0.101^{ns}	0.299**	-0.040^{ns}	0.245*
α-T								0.818****	−0.823****	0.635****	0.874****	0.048^{ns}	−0.862****	−0.518****
β-T									−0.672****	−0.388***	0.764****	0.227*	−0.815****	−0.229*
γ-T										0.849****	−0.903****	0.154^{ns}	0.496****	0.277**
δ-T											−0.752****	0.199^{ns}	0.300**	0.259*
α-T_3												0.110^{ns}	−0.722****	−0.365***
β-T_3													−0.385***	0.158^{ns}
γ-T_3														0.395***

Harvest moisture content (italics) was considered as an environmental factor for the GWAS.
Significant at *$P < 0.05$; **$P < 0.01$, ***$P < 0.001$; ****$P < 0.0001$; ns, not significant ($P > 0.05$)

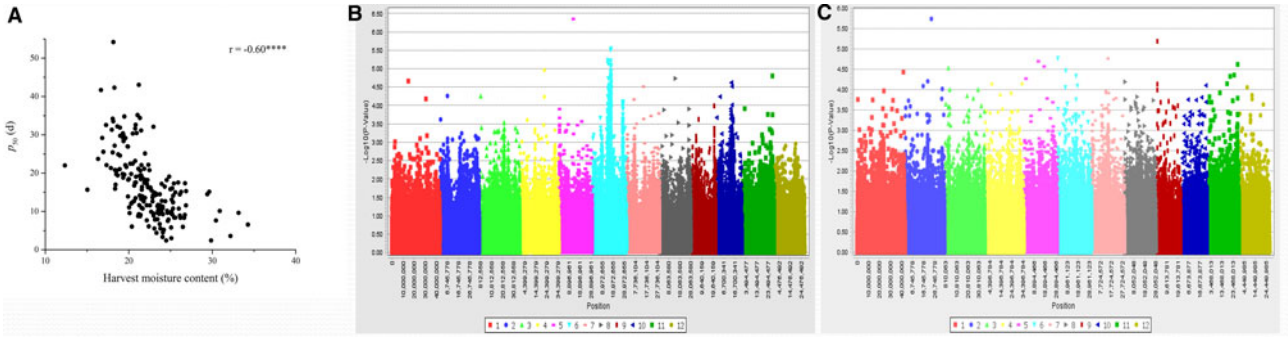

Fig. 1. (A) Correlation between harvest moisture content (%) and p_{50} (d) for 185 Aus rice accessions. Correlation coefficient (r) was significant at ****$P < 0.0001$; (B) GWA analysis (MLM) of seed longevity value (p_{50}) without considering harvest moisture content as an environmental factor. A large false positive peak appeared on chromosome 6; (C) GWA analysis (MLM) of seed longevity value (p_{50}) with including harvest moisture content as a covariate of p_{50}.

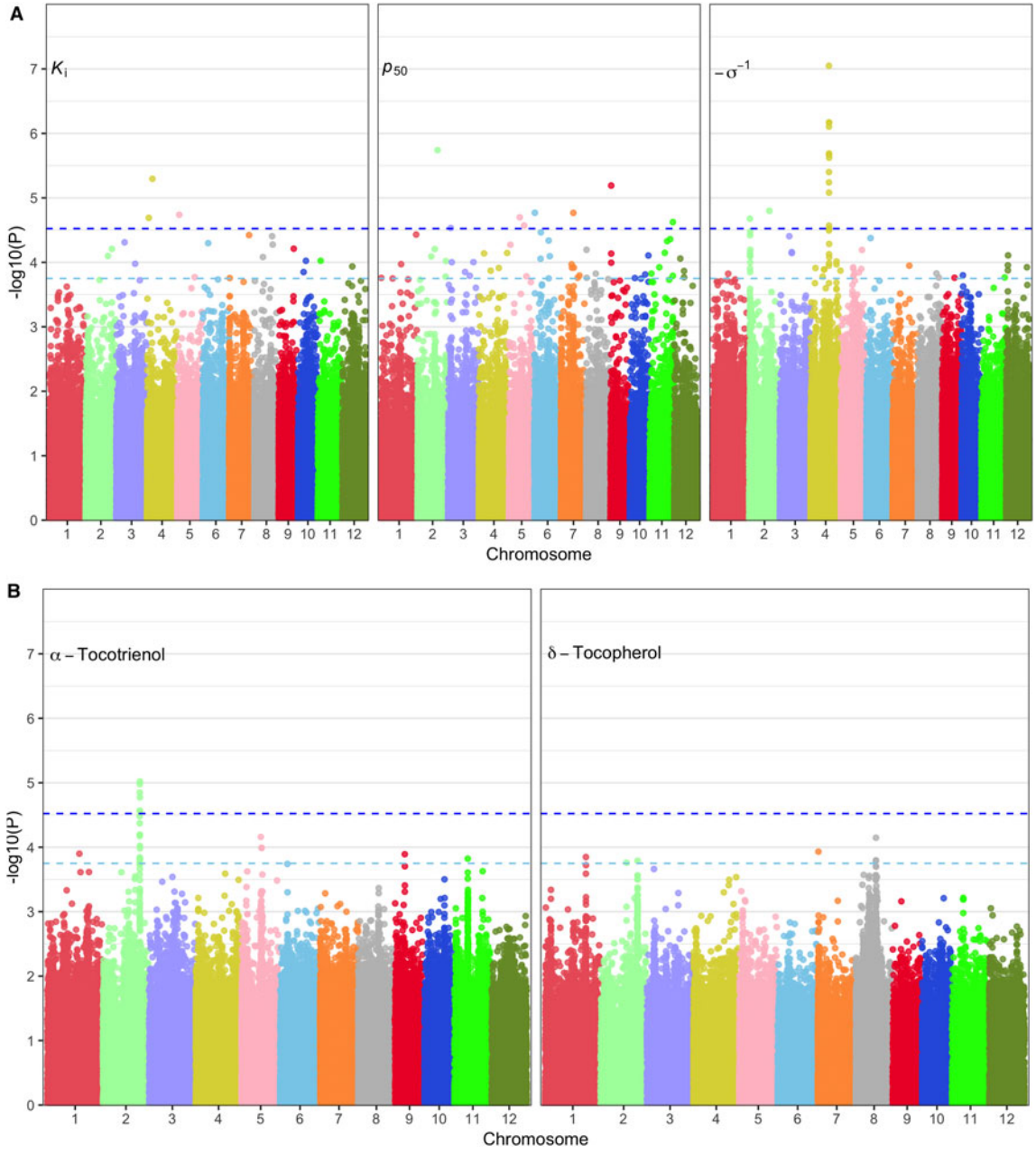

Fig. 2. GWA analysis (MLM) of seed longevity traits (K_i, $-\sigma^{-1}$ and p_{50}) and other traits (α-tocotrienol and δ-tocopherol) significantly correlated with seed longevity.

https://doi.org/10.1017/S0960258520000173 Published online by Cambridge University Press

Table 3. SNP markers most significantly (*P*-value < 3.00×10^{-5}) associated with seed longevity traits

No.	Associated trait	Marker ID	Chr.	Position	*P*-value	R^2 (%)	Gene nearest to marker	Annotation	Location of markers
1	K_i	117684625	4	2062633	2.04×10^{-5}	11	LOC_Os04g04380	Retrotransposon protein	3549 bp downstream
2		121895448	4	6273456	5.07×10^{-6}	12	LOC_Os04g11470	Retrotransposon protein	2790 bp upstream
3		153510853	5	2386167	1.83×10^{-5}	11	LOC_Os05g04950	Protein-binding protein	1st exon
4	$-\sigma^{-1}$	44113975	2	843052	2.10×10^{-5}	11	LOC_Os02g02410	DnaK family protein	380 bp upstream
5		67039297	2	23768374	1.59×10^{-5}	10	LOC_Os02g39384	Expressed protein	52 bp downstream
6		137281281	4	21659289	8.94×10^{-8}	15	LOC_Os04g35550	Expressed protein	2035 bp downstream
7		137310536	4	21688544	5.76×10^{-6}	12	LOC_Os04g35570	Regulator of chromosome condensation domain-containing protein	238 bp downstream
8		137347478	4	21725486	7.86×10^{-7}	13	LOC_Os04g35650	Pentatric opeptide	UTR 3′
9		137355302	4	21733310	3.98×10^{-6}	13	LOC_Os04g35660	No apical meristem protein	313 bp upstream
10		137360590	4	21738598	6.74×10^{-7}	13	LOC_Os04g35670	Retrotransposon protein	102 bp upstream
11		137370348	4	21748356	2.69×10^{-5}	9	LOC_Os04g35680	U-box domain-containing protein	799 bp downstream
12		137386114	4	21764122	2.13×10^{-6}	11	LOC_Os04g35710	Hypothetical protein	430 bp downstream
13		137397326	4	21775334	2.13×10^{-6}	11	LOC_Os04g35730	Transposon protein	4671 bp upstream
14		137400975	4	21778983	8.30×10^{-6}	10	LOC_Os04g35739	Retrotransposon protein	6505 bp upstream
15		137448650	4	21826658	2.05×10^{-6}	11	LOC_Os04g35810	Expressed protein	439 bp downstream
16		137514998	4	21893006	6.83×10^{-7}	13	LOC_Os04g35900	Translocon Tic40	19 bp upstream
17		137531898	4	21909906	2.38×10^{-6}	11	LOC_Os04g35930	OsFBX134 - F-box domain-containing protein	1035 bp upstream
18		137565627	4	21943635	2.97×10^{-5}	9	LOC_Os04g35990	OsFBX136 - F-box domain-containing protein	5th intron
19	p_{50}	67091629	2	23820706	1.82×10^{-6}	9	LOC_Os02g39480	Protein phosphatase 2C	535 bp downstream
20		82483859	3	3275686	2.93×10^{-5}	6	LOC_Os03g06520	Sulphate transporter	2nd intron
21		163889393	5	12764707	2.00×10^{-5}	7	LOC_Os05g22480	Retrotransposon protein	2nd exon
22		169096796	5	17972110	2.69×10^{-5}	8	LOC_Os05g30950	MSP domain-containing protein	1406 bp upstream
23		181572233	6	489113	1.71×10^{-5}	9	LOC_Os06g01870	Expressed protein	803 bp downstream
24		227031602	7	14699695	1.71×10^{-5}	8	LOC_Os07g25640	Retrotransposon protein	6th exon
25		271536787	9	1064237	6.44×10^{-6}	9	LOC_Os09g02560	Expressed protein	2nd exon
26		344302514	11	27609957	2.38×10^{-5}	6	LOC_Os11g45620	Rust-resistance protein Lr21	814 bp upstream

https://doi.org/10.1017/S0960258520000173 Published online by Cambridge University Press

Table 4. SNP markers associated with multi-traits

No.	Marker ID	Chr.	Position	Associated trait	*P*-value	Gene nearest to marker	Annotation	Location of marker
1	59648488	2	16377565	K_i	1.88×10^{-4}	LOC_Os02g27640	Retrotransposon protein	3th intron
				p_{50}	8.11×10^{-5}			
2	181572233	6	489113	K_i	7.38×10^{-4}	LOC_Os06g01870	Expressed protein	803 bp downstream
				p_{50}	1.71×10^{-5}			
3	199290474	6	18207354	K_i	1.85×10^{-4}	LOC_Os06g31290	Expressed protein	2224 bp downstream
				p_{50}	8.03×10^{-5}			
4	252150287	8	10120759	K_i	8.28×10^{-5}	LOC_Os08g16530	Retrotransposon protein	4th exon
				p_{50}	2.20×10^{-4}			
5	315490397	10	22005127	K_i	6.78×10^{-4}	LOC_Os10g40990	Flavonol synthase	7740 bp downstream
				p_{50}	7.84×10^{-5}			
6	320077635	11	3385078	K_i	9.46×10^{-5}	LOC_Os11g06900	Amidase family protein	8th intron
				p_{50}	7.78×10^{-4}			
7	352879193	12	7165530	K_i	2.11×10^{-4}	LOC_Os12g12920	Hypothetical protein	289 bp downstream
				p_{50}	8.75×10^{-5}			

A

Position	493394	22170562	28835836					
Chr.	1	3	4					
Genotype #				No. of accession	Mean δ-T (mg g^{-1} oil from brown rice)	Phenotypic effect (%)	Mean K_i (NED)	Phenotypic effect (%)
1	C	A	T	3	0.011	120	5.283	67
2	C	G	T	3	0.009	80	3.608	14
3	T	A	T	6	0.007	40	3.829	21
4	T	G	T	42	0.006	20	3.409	8
5	T	G	C	36	0.005	-	3.158	-

B

Position	33253508	6060064					
Chr.	1	8					
Genotype #			No. of accession	Mean δ-T (mg g^{-1} oil from brown rice)	Phenotypic effect (%)	Mean p_{50}	Phenotypic effect (%)
1	A	C	3	0.010	100	38.13	116
2	G	C	3	0.008	60	11.61	-34
3	G	T	78	0.005	-	17.67	-

Fig. 3. Haplotype effects on δ-tocopherol and seed longevity traits in a diverse Aus rice panel. The accessions with the absence of genotype data on the regions were not included in this figure.

in different seasons and perhaps with different drying treatments, depending on harvest MC.

Longevity parameters (K_i, $-\sigma^{-1}$ and p_{50}) observed in this study for seeds of Aus varieties were similar to those observed for seeds of Indica varieties stored in the same way (Lee et al., 2019b). For example, p_{50} ranged between 5.41 and 59.12 d for Indica accessions (maximum for each accession across sequential harvests) with a mean of 24.24 and coefficient of variation (CV) of 44.32%, compared with a range of 2.43 to 54.20 for Aus accessions, with a mean of 16.90 and CV of 52.90% (Table 1). The variation in the slope, $-\sigma^{-1}$, in both studies, again demonstrates that the Ellis and Roberts (1980) viability equation should be used with some caution to make predictions of longevity in storage, at least across very diverse rice accessions (Whitehouse et al., 2018).

Seed longevity loci of Indica (Lee et al., 2019b) and Aus (current) rice panels did not coincide. This was possibly due to the intraspecific variation in genomic structure (Wang et al., 2018).

https://doi.org/10.1017/S0960258520000173 Published online by Cambridge University Press

Among 19 sub-groups of 3010 diverse rice accessions, there was a large gap in genomic structures between Indica and Aus groups. McCouch et al. (2016) also discovered subgroup-specific alleles associated with grain traits in 1953 diverse rice accessions. Famoso et al. (2011) identified clear SNP variation in three exon regions of the *Nrat1* Al tolerance gene between Aus and Indica varieties. With this evidence, we speculate that different rice groups may have different mechanisms conferring high seed longevity. TPP7, a trehalose-6-phosphate phosphatase played an important role in seed longevity extension in near-isogenic lines (NIL) derived from the cross between temperate Japonica (short survival time) and Aus (long survival time) parents (Sasaki et al., 2015). However, in many Indica accessions, there was a large deletion in the TPP7-region, thus this gene is not functional in Indica group (Kretzschmar et al., 2015). In the Indica panel, a major mechanism conferring high seed longevity was related to DNA repair and transcription rather than TPP7 (Lee et al., 2019b).

Among 26 candidate genes where the most significant SNP markers located on or nearby, 13 genes are expressed genes, hypothetical genes or (retro-)transposon proteins (Table 3). The function of these genes has not been examined in detail, mainly due to the lack of technologies elucidating their expressions and functions. However, recent biological studies revealed that those genes are potentially important in biological processes as well as stress tolerance (Ijaq et al., 2019; Yang et al., 2019). Among seven SNP markers associated with multi-traits (K_i and p_{50}), two, one and two markers were on or nearby expressed genes, hypothetical genes and retrotransposon proteins, respectively; the other two were located on annotated genes (Table 4). Further integrative approaches combining systems-genetics and networking functions will be key to unravelling the role of those genes.

- LOC_Os05g04950, associated with K_i, was annotated to protein-binding proteins (Rice Genome Annotation Project; accessed on 22 September 2019). Based on gene ontology with networking genes (RiceNet v2, accessed on 22 September 2019; Lee et al., 2015), this gene involved protein modification process, lipid metabolic process and RNA modification.
- LOC_Os04g35650 associated with $-\sigma^{-1}$ was annotated to pentatric opeptide which is a kind of RNA binding protein mediating gene expression in the nucleus as well as organelles (Manna, 2015). This gene was connected with LOC_Os06g30380 and LOC_Os08g29110 which are related to RNA modification and oxidation reduction, respectively.
- LOC_Os04g35930 and LOC_Os04g35990, associated with $-\sigma^{-1}$, were annotated to OsFBX134 - F-box domain-containing protein which is known to mediate protein–protein interactions (Ho et al., 2006).
- LOC_Os02g39480, associated with p_{50}, was annotated to protein phosphatase 2C which is a central component in abscisic acid signal transduction (Rodriguez, 1998).
- LOC_Os11g45620, associated with p_{50}, was annotated to rust-resistance protein Lr21. There was no other network gene, but since this gene was highly expressed in mature seeds (Rice Genome Annotation Project), we speculate that it may play a role in the seed defence mechanism, i.e. oxidative stress occurring during seed storage.

Further gene validation studies are needed to confirm the roles of these highlighted genes.

Previous (Lee et al., 2017, 2019a) and current studies consistently showed that high proportions of γ- and δ-types of vitamin E homologues were positively correlated with seed longevity (Table 2). This could be due to more efficient and synergistic antioxidant activity of γ- and δ-homologues when compared with α- and β-types (Kadoma et al., 2006; Jiang, 2014; Kim, 2014). Hence, we recommend that breeding programmes for improving seed longevity of elite rice varieties focus on increase in proportions of γ- and δ-vitamin E homologues rather than total vitamin E content. Pericarp colour was positively correlated with p_{50}, which indicated stronger correlations than that of γ- and δ-vitamin E homologues (Table 2). Pericarp colour is known to contain various antioxidants such as flavonoids (Goufo and Trindade, 2013). Marker 315490397 on chromosome 10 at position 22005127 associated with both K_i and p_{50} located at 7740 bp downstream of the flavonol synthase gene (Table 4). With this evidence, there is a possibility that in Aus rice, other types of antioxidants, besides vitamin E, that are abundant in pericarp colour can play an important role in extending seed shelf life. Our complementary studies on seed metabolite profiling also suggest the specific types of flavonoids are relevant to high germination rate after seed storage (not published).

Conclusion

In line with previous studies, we confirmed that a high proportion of γ- and δ-type homologues rather than total vitamin E content was associated with seed longevity in a large rice panel. This confirmation revisited a theory of prolonged seed longevity via an antioxidant mechanism. Red pericarp colour containing other types of antioxidants was also correlated with high seed longevity.

Based on the comparative genetic analysis, we observed totally different QTL regions between GWA mapping with and without taking into account harvest MC. This environmental factor, which depends on ambient humidity of the harvest date, strongly interfered with genetic analysis. We, therefore, concluded that along with accurate phenotyping methods, interpretation of data in the context of environmental factors is extremely important in discovering the correct genes associated with the trait. Identified gene functions such as RNA modification and oxidation reduction, and/or SNP markers associated with both high δ-tocopherol and seed longevity traits (K_i and p_{50}) would be useful information in further vitamin E-seed longevity studies.

Supplementary material. To view supplementary material for this article, please visit: https://doi.org/10.1017/S0960258520000173.

Acknowledgements. We would like to thank IRRI colleagues: Renato Reaño for seed multiplication, Nora Kuroda for seed storage experiments, Dmytro Chebotarov for technical advice on genotype data and Ruaraidh Sackville Hamilton for supporting this research.

Financial support. This research was funded by grants from the Rural Development Administration, Korea (RDA) as RDA-IRRI cooperative research project (Grant Ref. PJ012723) and Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH. BMZ Funding of Genebank Uplift (Grant Ref. 17.7860.4-001.00).

Conflicts of interest. None declared.

References

Bin Rahman ANMR and Zhang J (2018) Preferential geographic distribution pattern of abiotic stress tolerant rice. *Rice* **11**, 10. doi:10.1186/s12284-018-0202-9

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y and Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635.

Ellis RH and Roberts EH (1980) Improved equations for the prediction of seed longevity. *Annals of Botany* **45**, 13–30.

Famoso AN, Zhao K, Clark RT, Tung C-W, Wright MH, Bustamante C, Kochian LV and McCouch SR (2011) Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS Genetics* **7**, e1002221. doi:10.1371/journal.pgen.1002221.

Goufo P and Trindade H (2013) Rice antioxidants: phenolic acids, flavonoids, anthocyanins, proanthocyanidins, tocopherols, tocotrienols, c-oryzanol, and phytic acid. *Food Science and Nutrition* **2**, 75–104.

Hay FR, de Guzman F and Hamilton NRS (2015) Viability monitoring intervals for genebank samples of *Oryza sativa*. *Seed Science and Technology* **43**, 218–237.

Hay FR, Valdez R, Lee J-S and Sta. Cruz PC (2019) Seed longevity phenotyping: recommendations on research methodology. *Journal of Experimental Botany* **70**, 425–434.

Ho M, Tsai P and Chien C (2006) F-box proteins: the key to protein degradation. *Journal of Biomedical Science* **13**, 181–191.

Ijaq J, Malik G, Kumar A, Das PS, Meena N, Bethi N, Sundararajan VS and Suravajhala P (2019) A model to predict the function of hypothetical proteins through a nine-point classification scoring schema. *BMC Bioinformatics* **20**, 14.

ISTA (2018) *International rules for seed testing 2018*. Bassersdorf, Switzerland, The International Seed Testing Association.

Jiang Q (2014) Natural forms of vitamin E: metabolism, antioxidant, and anti-inflammatory activities and their role in disease prevention and therapy. *Free Radical Biology and Medicine* **72**, 76–90.

Kadoma Y, Ishihara M, Okada N and Fujisawa S (2006) Free radical interaction between vitamin E (alpha-, beta-, gamma- and delta-tocopherol), ascorbate and flavonoids. *In Vivo* **20**, 823–828.

Khush GS (1997) Origin, dispersal, cultivation and variation of rice. *Plant Molecular Biology* **35**, 25–34.

Kim HJ (2014) Effect of α-, β-, γ-, and δ-tocotrienol on the oxidative stability of lard. *Journal of the American Oil Chemists' Society* **91**, 777–782.

Ko SN, Kim CJ, Kim H, Kim CT, Chung SH, Tae BS and Kim IH (2003) Tocol levels in milling fractions of some cereal grains and soybean. *Journal of the American Oil Chemists' Society* **80**, 585–589.

Kretzschmar T, Pelayo MAF, Trijatmiko KR, *et al.* (2015) A trehalose-6-phosphate phosphatase enhances anaerobic germination tolerance in rice. *Nature Plants* **1**, 15124.

Lee J-S, Kwak J, Cho J-H, Chebotarov D, Yoon M-R, Lee J-S, Hamilton NRS and Hay FR (2019a) A high proportion of beta-tocopherol in vitamin E is associated with poor seed longevity in rice produced under temperate conditions. *Plant Genetic Resources* **4**, 375–378.

Lee J-S, Kwak J, Yoon M-R, Lee JS and Hay FR (2017) Contrasting tocol ratios associated with seed longevity in rice sub-populations. *Seed Science Research* **27**, 273–280.

Lee T, Oh T, Yan S, Shin J, Hwang S, Kim CY, Kim H, Shim H, Shim JE, Ronald PC and Lee I (2015) RiceNet v2: an improved network prioritization server for rice genes. *Nucleic Acids Research* **43**, W122–W127.

Lee J-S, Valdez R, Punzalan M, Pacleb M, Kretzschmar T, McNally K, Ismail AM, Sta Cruz PS, Hamilton NRS and Hay FR (2019b) Variation in seed longevity among diverse Indica rice varieties. *Annals of Botany* **124**, 447–460.

Lee J-S, Wissuwa M, Zamora OB and Ismail AM (2018) Novel sources of *aus* rice to zinc deficiency tolerance identified through association analysis using high-density SNP array. *Rice Science* **25**, 293–296.

Leprince O, Pellizzaro A, Berriri S and Buitink J (2017) Late seed maturation: drying without dying. *Journal of Experimental Botany* **68**, 827–841.

Londo J, Chiang Y, Hung K, Chiang T and Schaal B (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences USA* **103**, 9578–9583.

Manna S (2015) An overview of pentatricopeptide repeat proteins and their applications. *Biochimie* **113**, 93–99.

McCouch SR, Wright MH, Tung C-W, Maron LG, McNally KL, Fitzgerald M, Singh N, DeClerck G, Agosto-Perez F, Korniliev P, Greenberg AJ, Naredo MEB, Mercado SMQ, Harrington SE, Shi Y, Branchini DA, Kuser-Falcaõ PR, Leung H, Ebana K, Yano M, Eizenga G, McClung A and Mezey J (2016) Open access resources for genome-wide association mapping in rice. *Nature Communications* **7**, 10532.

Rice Genome Annotation Project http://rice.plantbiology.msu.edu/

Rice SNP-Seek Database http://snp-seek.irri.org/

Rodriguez PL (1998) Protein phosphatase 2C (PP2C) function in higher plants. *Plant Molecular Biology* **38**, 919–27.

Sasaki K, Takeuchi Y, Miura K, Yamaguchi T, Ando T, Ebitani T, Higashitani A, Yamaya T, Yano M and Sato T (2015) Fine mapping of a major quantitative trait locus, *qLG-9*, that controls seed longevity in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **128**, 769–778.

Sattler SE, Cahoon EB, Coughlan SJ and DellaPenna D (2003) Characterization of tocopherol cyclases from higher plants and cyanobacteria. evolutionary implications for tocopherol synthesis and function. *Plant Physiology* **132**, 2184–2195.

The 3,000 Rice Genomes Project (2014) The 3,000 rice genomes project. *Giga Science* **3**, 7.

Timple SE and Hay FR (2018) High-temperature drying of seeds of wild Oryza species intended for long-term storage. *Seed Science and Technology* **46**, 107–112.

Wang WS, Mauleon R, Hu ZQ, Chebotarov D, Tai SS, Wu ZC, Li M, Zheng TQ, Fuentes RR, Zhang F, Mansueto L, Copetti D, Sanciangco M, Palis KC, Xu JL, Chen S, Fu BY, Zhang HL, Gao YM, Zhao XQ, Shen F, Cui X, Yu H, Li ZC, Chen ML, Detras J, Zhou YL, Zhang XY, Zhao Y, Kudrna D, Wang CC, Li R, Jia B, Lu JY, He XC, Dong ZT, Xu JB, Li YH, Wang M, Shi JX, Li J, Zhang DB, Lee SH, Hu WS, Poliakov A, Dubchak I, Ulat VJ, Borja FN, Mendoza JR, Ali J, Li J, Gao Q, Niu YC, Yue Z, Naredo MEB, Talag J, Wang XQ, Li JJ, Fang XD, Yin Y, Glaszmann JC, Zhang JW, Li JY, Hamilton RS, Wing RA, Ruan J, Zhang GY, Wei CC, Alexandrov N, McNally KL, Li ZK and Leung H (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**, 43–49.

Whitehouse KJ, Hay FR and Ellis RH (2015) Increases in the longevity of desiccation-phase developing rice seeds: response to high-temperature drying depends on harvest moisture content. *Annals of Botany* **116**, 247–259.

Whitehouse KJ, Hay FR and Ellis RH (2017) High-temperature stress during drying improves subsequent rice (*Oryza sativa* L.) seed longevity. *Seed Science Research* **27**, 281–291.

Whitehouse KJ, Hay FR and Ellis RH (2018) Improvement in rice seed storage longevity from high-temperature drying is a consistent positive function of harvest moisture content above a critical value. *Seed Science Research* **28**, 332–339.

Yang Z, Zeng X and Tsui SK-W (2019) Investigating function roles of hypothetical proteins encoded by the *Mycobacterium tuberculosis* H37Rv genome. *BMC Genomics* **20**, 394.