OPEN ACCESS Freely available online

PLOS ONE

Identification of Drought Tolerance Markers in a Diverse Population of Rice Cultivars by Expression and Metabolite Profiling

Thomas Degenkolbe, Phuc T. Do, Joachim Kopka, Ellen Zuther, Dirk K. Hincha, Karin I. Köhl*

Max Planck Institute of Molecular Plant Physiology, Potsdam, Brandenburg, Germany

Abstract

Rice provides about half of the calories consumed in Asian countries, but its productivity is often reduced by drought, especially when grown under rain-fed conditions. Cultivars with increased drought tolerance have been bred over centuries. Slow selection for drought tolerance on the basis of phenotypic traits may be accelerated by using molecular markers identified through expression and metabolic profiling. Previously, we identified 46 candidate genes with significant genotype × environment interaction in an expression profiling study on four cultivars with contrasting drought tolerance. These potential markers and in addition GC-MS quantified metabolites were tested in 21 cultivars from both *indica* and *japonica* background that varied in drought tolerance. Leaf blades were sampled from this population of cultivars grown under control or long-term drought condition and subjected to expression analysis by qRT-PCR and metabolite profiling. Under drought stress, metabolite levels correlated mainly negatively with performance parameters, but eight metabolites correlated positively. For 28 genes, a significant correlation between expression level and performance under drought was confirmed. Negative correlations were predominant. Among those with significant positive correlation was the gene coding for a cytosolic fructose-1,6-bisphosphatase. This enzyme catalyzes a highly regulated step in C-metabolism. The metabolic and transcript marker candidates for drought tolerance were identified in a highly diverse population of cultivars. Thus, these markers may be used to select for tolerance in a wide range of rice germplasms.

Citation: Degenkolbe T, Do PT, Kopka J, Zuther E, Hincha DK, et al. (2013) Identification of Drought Tolerance Markers in a Diverse Population of Rice Cultivars by Expression and Metabolite Profiling. PLoS ONE 8(5): e63637. doi:10.1371/journal.pone.0063637

Editor: Girdhar K. Pandey, University of Delhi South Campus, India

Received November 29, 2012; **Accepted** April 4, 2013; **Published** May 22, 2013

Funding: The work was supported by the German Ministry for Education and Research (BMBF) (grant BMBF 0312854) and the Max Planck Society. Phuc Thi Do was supported by fellowships from the Vietnamese government (MOET), the German Academic Exchange Council (DAAD) and the Max Planck Society. 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.

* E-mail: koehl@mpimp-golm.mpg.de

Introduction

Rice (*Oryza sativa* L.) is one of the world's most important staple foods with 720 million tons harvested in 2011 (www.fao.org 24.07.2012). In Asia, its main cultivation area, rice provides 35–60% of the calories consumed [1]. Rice was domesticated at least twice independently, which resulted in the two subspecies *indica* and *japonica* [2]. Centuries of breeding furthermore yielded a wide range of cultivars adapted to different watering regimes from irrigated, deep-water cultures to rain-fed lowland and upland cultivars [3]. About 50% of the rice acreage is rain-fed and not irrigated [4]. In these areas, drought is the major environmental factor that reduces productivity by 13–35% [5,6]. Drought stress causes yield loss not only in rice, but in many other crops like potato, wheat and maize. The situation will aggravate in future as agriculture competes with others consumers for limited water supplies. Thus, more food will have to be produced with less water to provide for the increasing world population [7].

Therefore, strategies to identify drought-tolerant germplasms are of major interest. Traditionally, breeding of drought-tolerant cultivars relied on selection based on phenotypic and physiological traits observed under drought stress [8,9], namely leaf rolling [10–12], cell membrane stability [13], carbon isotope discrimination, gas exchange and chlorophyll fluorescence measurements [14–19], stomatal conductance and water use efficiency [19,20], root traits [21] and yield [22,23]. However, this selection process is labour-intensive and slow as it requires cultivation of breeding populations under drought conditions [24]. The phenotypic evaluation can, however, be replaced by the use of molecular markers such as DNA polymorphisms or chemical tags [25] associated with the trait. Marker-assisted selection (MAS) is cheaper and more convenient than phenotype-based selection and it presently is the only option to combine traits by gene pyramiding [25]. DNA based markers can be derived from quantitative trait loci (QTL) and allow selection already in the seedling stage. QTLs for drought tolerance traits have been identified in the last decade in rice [12,13,23,26–30], wheat [31,32], maize [33,34] and other crops. New breeding markers based on transcript or metabolite abundance can be derived from multi-parallel methods like expression and metabolic profiling [3]. For complex traits like drought tolerance, studies have shown that markers will indicate traits contributing to drought tolerance rather than overall tolerance [25]. Thus, ideally, the concentration of a marker transcript or metabolite will correlate with one or several traits contributing to drought tolerance in a wide range of cultivars.

For our marker search, we used metabolite and expression profiling. Metabolite were measured on the Golm Metabolomics platform [35]. In a previous microarray study, we have identified

genes that were differentially expressed in four rice genotypes of contrasting drought tolerance [36] and thus could be marker candidates for drought tolerance. As the ideal marker should correlate positively with drought tolerance in a wide range of genetic backgrounds, we tested the potential markers in an association type study. We choose *indica* and *japonica* cultivars that originated from a variety of Vietnamese agro-ecosystems and had been selected into a breeding program for drought and salt stress. Some well characterised cultivars from the International Rice Research Institute (IRRI, Manila, Philippines) were included additionally. All cultivars have been characterised for several traits related to drought tolerance under control and long-term drought stress in a parallel study (Do et al. PLOS ONE 10.1371/journal.pone.0060325). In the present study, we checked, which RNA and metabolite levels allowed prediction of drought tolerance related traits. These transcripts and metabolites may be drought tolerance markers in rice.

Results

Genotyping of Cultivars

In our study, 17 of the 21 cultivars (Table 1) originated from a Vietnamese breeding program for drought stress resistance. As information on the pedigree of these cultivars was limited, six subspecies-specific sequence tagged sites (STS) markers located on four chromosomes [37] were chosen to determine to which subspecies (*japonica* or *indica*) the cultivars belong (Table 1). Cultivars with known pedigree (*japonica* cultivars 50, 51 and 54, *indica* cultivars 52, 55 and 62) were included as references. Based on the results, three of the Vietnamese cultivars were classified as *japonica* and eight as *indica*. Four cultivars (3, 13, 15 and 17) were mainly *japonica* with some *indica* introgression. Two cultivars (14 and 18) were classified as mainly *indica* with some *japonica* introgression. Thus, the studied genotypes represent both the *indica* and *japonica* gene pools.

Identification of Potential Gene Expression-based Markers

Potential marker genes had been identified in an expression profiling study on four rice genotypes of contrasting drought tolerance [36]. Genes had been selected as marker candidates, when their expression response to drought stress differed between tolerant and sensitive cultivars. This response pattern was identified in an analysis of variance by a significant interaction effect of the factors condition (drought, control) and tolerance group (sensitive cultivars, tolerant cultivars) on gene expression [36]. To reduce the list of candidates to those with agricultural relevance, we have compared the position of the candidate gene to the positions of published QTL for drought tolerance in rice [36]. From 108 candidate genes with significant interaction and location

Table 1. Rice cultivars in study population belong to indica or japonica subspecies.

		Marker Name					
Cultivar	**ID**	**S01022**	**S03020**	**S03136**	**S04128**	**S07011**	**S07103**
CR203	1	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*
DR2	2	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*
Loc	3	**jap.**	**jap.**	**jap.**	*ind.*	**jap.**	**jap.**
C70	4	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*
C71	5	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*
K.lua nuong	13	**jap.**	**jap.**	**jap.**	*ind.*	**jap.**	**jap.**
Cuom	14	**jap.**	*ind.*	*ind.*	*ind.*	*ind.*	**jap.**
Khau cham	15	**jap.**	**jap.**	**jap.**	*ind.*	**jap.**	**jap.**
Khau hom	16	**jap.**	**jap.**	**jap.**	**jap.**	**jap.**	**jap.**
Khau non	17	**jap.**	**jap.**	**jap.**	*ind.*	**jap.**	**jap.**
Tra linh	18	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*	**jap.**
Nep men	19	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*
Loc dau	20	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*
Lua man	22	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*
LC-93-1	29	**jap.**	**jap.**	**jap.**	**jap.**	**jap.**	**jap.**
LC-93-2	30	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*
LC-93-4	31	**jap.**	**jap.**	**jap.**	**jap.**	**jap.**	**jap.**
Nipponbare	50	**jap.**	**jap.**	**jap.**	**jap.**	**jap.**	**jap.**
Taipei309	51	**jap.**	**jap.**	**jap.**	**jap.**	**jap.**	**jap.**
IR57311-95-2-3	52	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*
Zonghua	53	**jap.**	**jap.**	**jap.**	**jap.**	**jap.**	**jap.**
CT 9993-5-10-1	54	**jap.**	**jap.**	**jap.**	**jap.**	**jap.**	**jap.**
IR 62266-42-6-2	55	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*
IR 64	62	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*	*ind.*

Genotyping of rice cultivars based on the amplification of six subspecies-specific sequence tagged sites (STS) markers. *ind.* – *indica*, **jap.** – *japonica*. Bold print of cultivar name: reference cultivars with known genotype.

doi:10.1371/journal.pone.0063637.t001

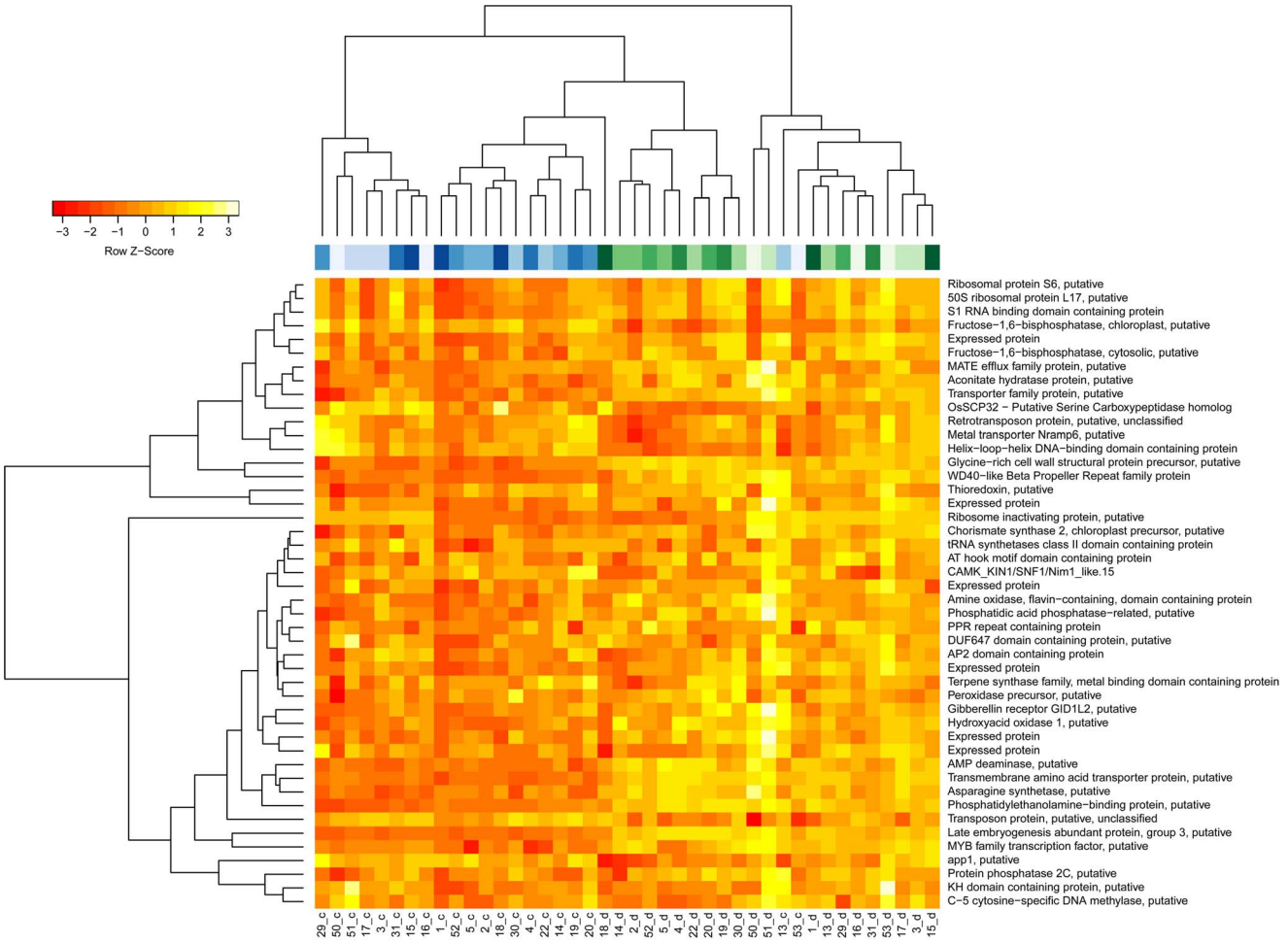

Figure 1. RNA expression response to drought differs between rice cultivars. Hierarchical clustering and heatmap of expression levels in leaves of rice plants grown under control conditions (c) or drought stress (d). The code below the heatmap indicates the line id (see Table 1) and the condition (blue: control = c, green: drought stress = d).
doi:10.1371/journal.pone.0063637.g001

in a published QTL, we have chosen the 46 genes with the lowest p-values for further analysis in the present study. Expression levels of these genes were measured by qRT-PCR in leaf blades of 21 rice cultivars grown under control and drought stress condition identical to those in the previous microarray study (see Table S1 for all qRT-PCR results). Hierarchical clustering of expression patterns in biological samples (Figure 1) separated samples from well-watered and drought treated plants neatly. The few exceptions were from extreme cultivars. The highly sensitive cultivar 53 showed a drought stress expression pattern already under control conditions. In contrast, under drought stress the expression pattern of the highly tolerant cultivar 18 resembled the expression pattern found in other cultivars under well-watered conditions. Most marker candidates were more highly expressed under drought stress than under control conditions.

To identify markers, correlations between relative expression levels of the candidate genes and physiological parameters contributing to performance under drought were determined. These traits were assessed on vegetative plants after 18 days of growth under control or drought stress conditions (see the accompanying paper Do et al. PLOS ONE 10.1371/journal.pone.0060325). Under drought stress, the parameters drought score (representing the stay-green trait) and mean water use efficiency (WUE) were determined. Additionally, correlations to shoot fresh and dry weight, total fresh and dry weight and photosynthetic yield (measured as photosystem II quantum use efficiency by chlorophyll fluorescence spectroscopy) under control and drought stress were analysed (Figure 2). Parameters were mathematically transformed to ensure that high parameter values indicate good performance (see Methods, 'Correlation analyses). Ideally, correlation to all performance parameters should show the same direction if the candidate gene expression is a good predictor for several tolerance traits. Under drought stress, expression levels of 28 of the 46 candidate genes correlated significantly ($p<0.05$) with physiological performance parameters under drought stress (Figure 2).

For 11 genes, expression under drought stress correlated negatively with several phenotypic parameters, indicating high expression levels in cultivars with poor performance. The expression of genes coding for an aconitate hydratase, an AMP deaminase and an asparagine synthetase correlated negatively with most performance parameters under drought stress (s. Figure 3). Also under control conditions, expression of asparagine synthase and AMP deaminase genes correlated negatively with performance parameters under drought. Both genes were thus more strongly expressed in drought-sensitive cultivars than in tolerant cultivars under water-sufficient and water-deficient conditions, suggesting a constitutive increase in the expression of

Annotation (MSU Osa Release 7)	P	Locus ID	Physiological data d vs. qPCR d							Physiological data d vs. qPCR c							Physiological data c vs. qPCR c				
			Negative rank	WUE	Yield	FWS	DWS	FWT	DWT	Negative rank	WUE	Yield	FWS	DWS	FWT	DWT	Yield	FWS	DWS	FWT	DWT
DUF647 domain containing protein, putative	P1	LOC_Os01g04860			0.43																
Glycine-rich cell wall structural protein precursor, putative	P2	LOC_Os01g06310	0.34	0.30		0.39	0.41		0.31	-0.28											
Ribosome inactivating protein, putative	P3	LOC_Os01g06740			0.30					-0.31								-0.29		-0.29	-0.29
KH domain containing protein, putative	P5	LOC_Os01g60260																			
Phosphatidic acid phosphatase-related, putative	P6	LOC_Os01g63060		-0.37		-0.28		-0.37	-0.34												
Fructose-1,6-bisphosphatase, putative	P8	LOC_Os01g64660	0.32	0.29	0.34		0.31		0.32												
Helix-loop-helix DNA-binding domain containing protein	P9	LOC_Os01g72370																			
tRNA synthetases class II domain containing protein	P10	LOC_Os02g41470																			
CAMK_KIN1/SNF1/Nim1_like.15	P11	LOC_Os03g03510		-0.37		-0.33		-0.36	-0.32			-0.30									
Aconitate hydratase protein, putative	P12	LOC_Os03g04410		-0.39	-0.51	-0.29	-0.31	-0.39	-0.42												
Expressed protein	P13	LOC_Os03g04710		0.28		0.31	0.35		0.3		0.31	0.33			0.32	0.30					
Transporter family protein, putative	P14	LOC_Os03g11900																			
Chorismate synthase 2, chloroplast precursor, putative	P15	LOC_Os03g14990		-0.3		-0.31		-0.29	-0.29												
Fructose-1,6-bisphosphatase, putative	P16	LOC_Os03g16050			0.32																
Asparagine synthetase, putative	P17	LOC_Os03g18130		-0.38	-0.41	-0.46	-0.43	-0.45	-0.43		-0.46	-0.48	-0.32	-0.33	-0.43	-0.43			-0.32	-0.31	-0.33
S1 RNA binding domain containing protein	P18	LOC_Os03g20100	0.34																		
Expressed protein	P19	LOC_Os03g21370				0.28	0.32														
Terpene synthase family, metal binding domain containing protein	P20	LOC_Os03g22620																			
MATE efflux family protein, putative	P21	LOC_Os03g37490		-0.34	-0.48			-0.33	-0.35												
PPR repeat containing protein	P22	LOC_Os03g40020										-0.35									
Expressed protein	P23	LOC_Os03g44810																			
app1, putative	P24	LOC_Os03g56930				-0.29															
Gibberellin receptor GID1L2, putative	P25	LOC_Os03g57640			-0.30																
C-5 cytosine-specific DNA methylase, putative	P26	LOC_Os03g58400																			
50S ribosomal protein L17, putative	P27	LOC_Os03g60100	0.31																		
Ribosomal protein S6, putative	P28	LOC_Os03g62630																			
Transmembrane amino acid transporter protein, putative	P29	LOC_Os04g38680		-0.41	-0.53		-0.3	-0.36	-0.4	-0.28	-0.37					-0.34		-0.28	-0.34	-0.32	-0.37
AP2 domain containing protein	P30	LOC_Os04g52090											0.32		0.29						
Hydroxyacid oxidase 1, putative	P31	LOC_Os04g53210			-0.33																
Expressed protein	P32	LOC_Os04g55600																			
Transposon protein, putative, unclassified	P34	LOC_Os04g55710	0.28	0.46	0.59	0.49	0.5	0.51	0.53					0.28	0.29	0.29					
Amine oxidase, flavin-containing, domain containing protein	P35	LOC_Os04g57550		-0.31	-0.34				-0.31												
Phosphatidylethanolamine-binding protein, putative	P37	LOC_Os05g39250			-0.31																
Late embryogenesis abundant protein, group 3, putative	P38	LOC_Os05g46480		-0.31	-0.46			-0.29	-0.29												
OsSCP32 - Putative Serine Carboxypeptidase homolog	P39	LOC_Os06g08720			0.35																
Protein phosphatase 2C, putative	P40	LOC_Os07g02330																			
Expressed protein	P41	LOC_Os07g02710																			
Retrotransposon protein, putative, unclassified	P42	LOC_Os07g04930																			
Thioredoxin, putative	P43	LOC_Os07g08840																			
Metal transporter Nramp6, putative	P44	LOC_Os07g15460																			
WD40-like Beta Propeller Repeat family protein	P46	LOC_Os07g44410																			
Expressed protein	P47	LOC_Os07g47590		-0.34	-0.41			-0.3	-0.32		-0.41	-0.42			-0.33	-0.37		-0.28		-0.31	-0.3
Peroxidase precursor, putative	P48	LOC_Os07g47990	0.38										0.36	0.32	0.32	0.3					
AMP deaminase, putative	P49	LOC_Os07g49270		-0.44	-0.59	-0.38	-0.41	-0.44	-0.47	-0.37	-0.31	-0.41		-0.31	-0.33	-0.36					
AT hook motif domain containing protein	P50	LOC_Os08g02490		-0.37		-0.4	-0.32	-0.44	-0.39		-0.36	-0.29			-0.31	-0.31					
MYB family transcription factor, putative	P53	LOC_Os12g37690																			

Figure 2. Correlation of physiological data with candidate gene expression. Annotation, primer (P), locus identifier of and correlation coefficients for candidate genes with significant ($p<0.05$) positive (blue) or negative (red) correlation of log-transformed expressions levels with physiological data under drought (d) or control (c) conditions. Data of 21 different cultivars with 2 to 3 replicates per cultivar and condition, 51 data pairs in total. Negative rank - mean scoring rank multiplied with -1; WUE - water use efficiency; yield - chlorophyll-*a* fluorescence yield; FWS - fresh weight shoot; DWS - dry weight shoot; FWT - total fresh weight; DWT - total dry weight. Sorted by LocusID.
doi:10.1371/journal.pone.0063637.g002

these genes in drought-sensitive compared to drought-tolerant cultivars. Expression of the asparagine synthetase gene and genes encoding a transmembrane amino acid transporter protein and an expressed protein at LOC_Os07g47590 also correlated negatively with performance parameters under control conditions. High expression levels of these genes thus indicated slow growth rather than poor performance especially under drought. For the other genes, expression under control conditions did not correlate with growth or photosynthesis under control conditions. This suggests that the difference in gene expression between cultivars was not linked to general differences in growth rates. High expression levels of these genes indicated drought sensitivity.

Significant positive correlations between expression levels under drought and drought score were found for six candidate genes (Figure 2). The expression of these genes could serve as tolerance markers. For three of these genes, expression levels were also positively correlated with shoot and total dry weight under drought conditions (Figure 2 and Figure S1). These genes encode a cytosolic fructose-1,6-bisphosphatase, a glycine-rich cell wall structural protein, and a transposon protein. For these genes, high expression levels under drought stress indicated high drought tolerance. Their expression levels under control conditions showed only a few significant correlations to performance parameters under drought stress. Thus, in contrast to the constitutive sensitivity markers (e.g. asparagine synthase gene and an AMP deaminase encoding gene, see above), expression of most tolerance marker genes seemed to be drought-induced, i.e. their expression levels could only be used as markers under drought stress. Both, the cytosolic fructose-1,6-bisphosphatase and the plastidial precursor of fructose-1,6-bisphosphatase correlated significantly with the photosynthetic yield (Figure 4).

Identification of Potential Metabolic Markers

Metabolite levels were determined in leaves of control and drought-stressed plants from 21 cultivars (see Table S2). Hierarchical clustering for both metabolites and samples (cultivars × condition) are shown in Figure 5. The clustering of samples showed a complete separation of the metabolite pattern between samples from control and drought-treated material. The drought treatment was thus the main source of variance in the data, which indicates a complete change of metabolism under stress conditions in all cultivars. The changes induced by the treatment are predominantly larger than the differences between cultivars within a treatment. The upper cluster in the metabolite hierarchy shown in Figure 5 contains metabolites that increased under drought stress; it contains glutamine, glutamic acid and derivatives. In

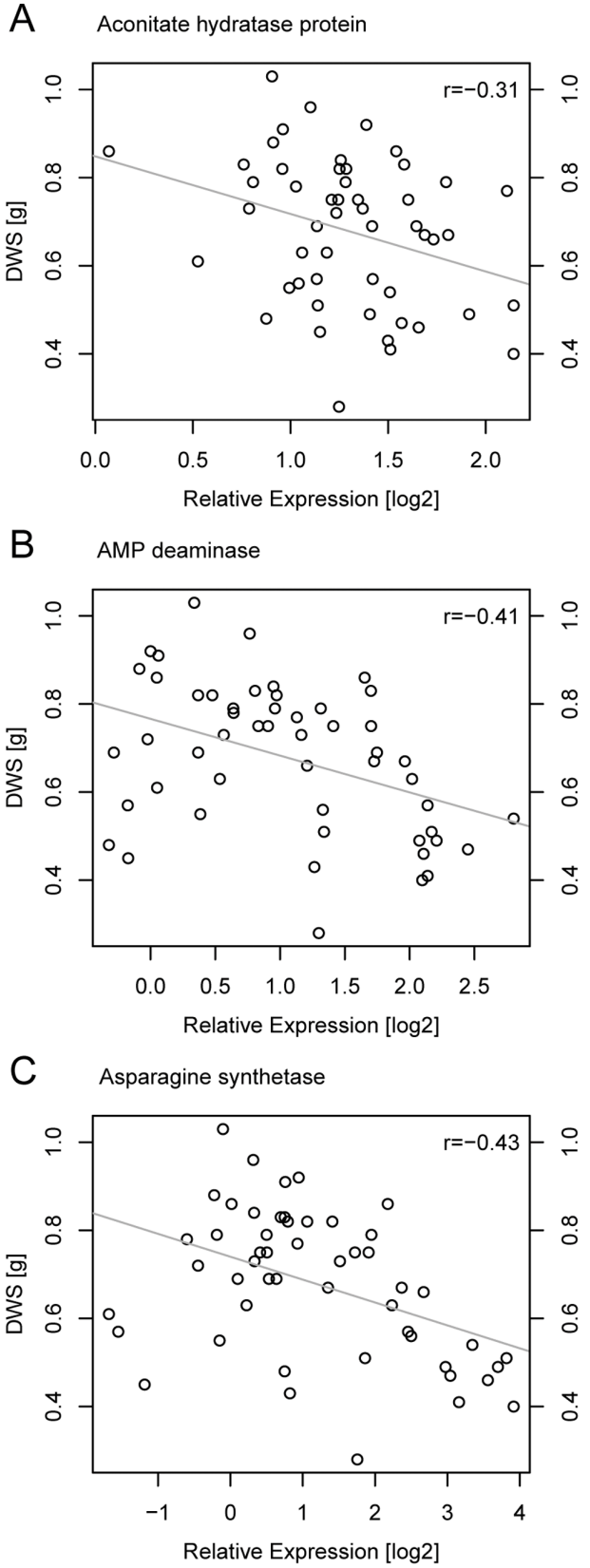

Figure 3. Expression of genes for aconitate hydratase (A), an AMP deaminase (B) and asparagine synthetase (C) correlated negatively with shoot dry weight. Relative expression levels of genes and average shoot dry weight measured in rice cultivars in three independent experiments. The regression coefficient r for the linear regression of shoot dry weight against expression level is shown in the upper right corner.
doi:10.1371/journal.pone.0063637.g003

contrast, metabolites grouping with sugar phosphates (lowest cluster in Figure 5) decreased under drought stress.

Most of the significant correlations between metabolite levels and performance parameters were negative under drought stress (Figure 6). Negative correlations were found for the concentration of the amino acids asparagine, glutamate, glutamine, glycine, serine and threonine, and for the organic acids erythronic, galactonic and threonic acid. Higher concentrations of these metabolites were connected with lower fresh and dry weight, lower photosynthetic quantum yield and lower water use efficiency. In sensitive cultivars, levels were 10 to 100 fold higher than in tolerant cultivars (see Figure 7 and Figure S2, notice log-10 scale for metabolite levels). Under control conditions, levels of asparagine, erythronic acid-1,4-lactone, serine and threonine correlated positively with performance under drought. Asparagine, threonine and serine levels were significantly higher ($p<0.05$) under drought than under control conditions. Those cultivars that accumulated asparagine more than most other cultivars (levels above mean plus one standard deviation) predominantly showed a below than average water use efficiency (WUE) (Figure 8). Thus, high asparagine levels indicated low WUE.

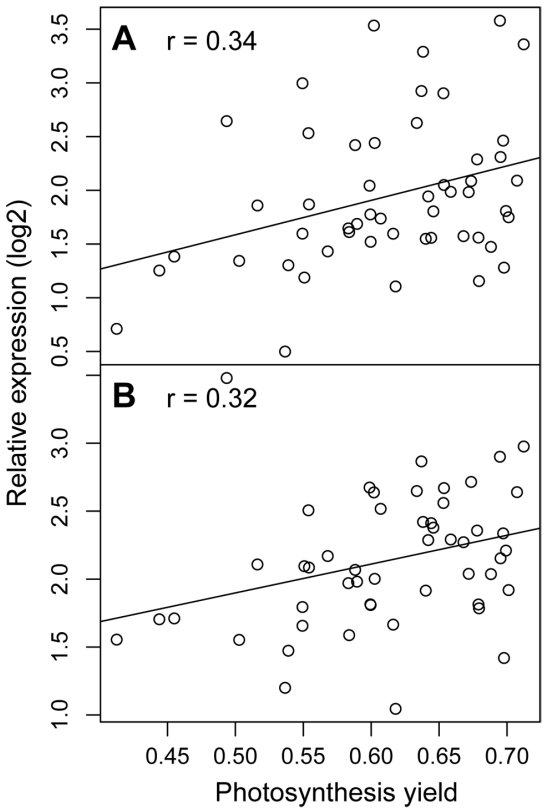

Figure 4. Expression of the fructose-1,6-bisphosphatase gene correlated with photosynthetic quantum yield. Correlation of relative expression levels of genes coding for a cytosolic (A) and a plastidial (B) fructose-1,6-bisphosphatase with photosynthetic quantum yield of leaves measured after 18 days of growth under drought stress.
doi:10.1371/journal.pone.0063637.g004

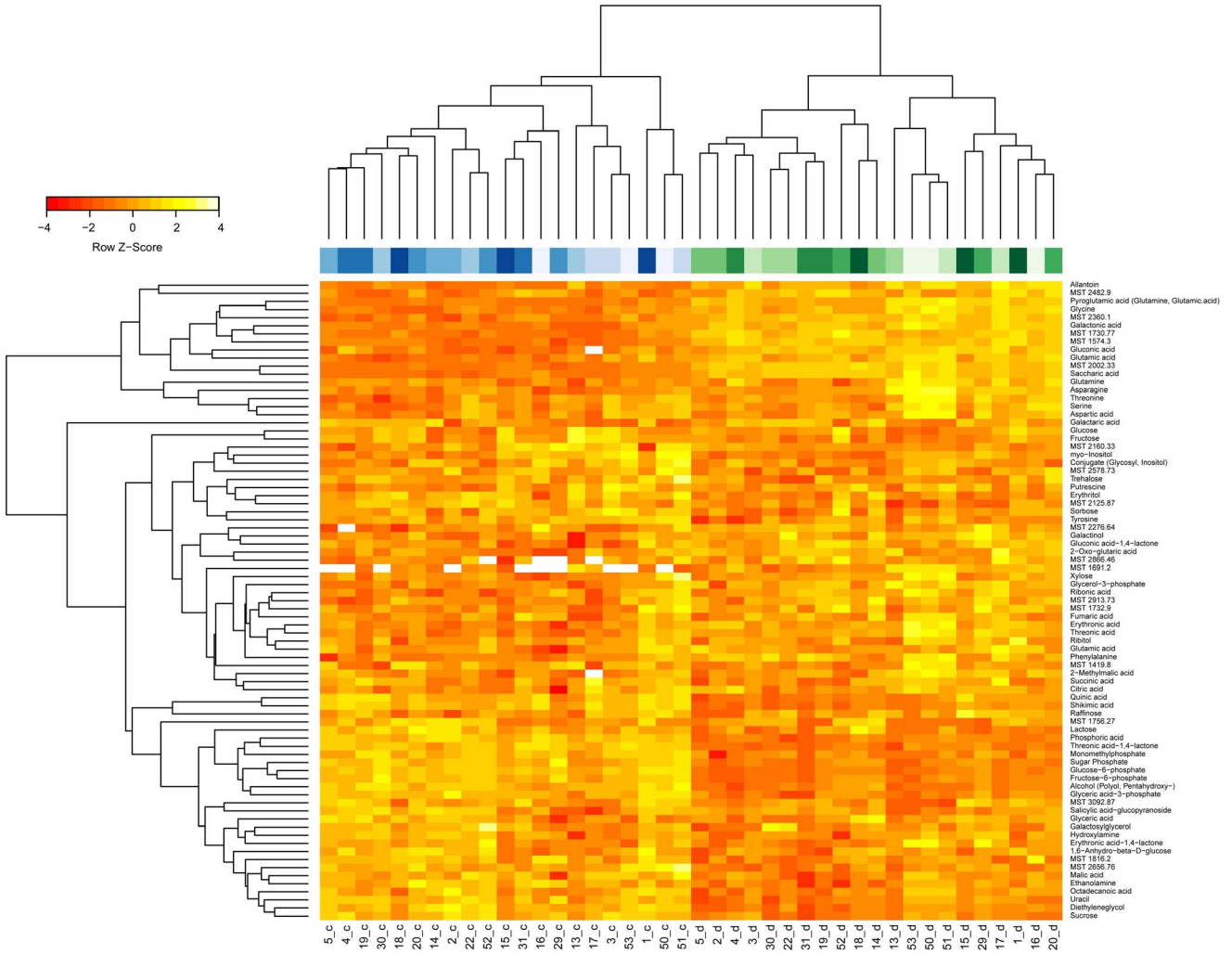

Figure 5. Metabolite response to drought differs between rice cultivars. Hierarchical clustering and heatmap of metabolite levels in leaves of rice plants grown under control conditions (c) or drought stress (d). Metabolite levels were normalised within an experiment by Z-transformation as indicated in Material and Methods. The code below the heatmap indicates the line id (see Table 1) and the condition (blue: control, green: drought stress).
doi:10.1371/journal.pone.0063637.g005

Positive correlations between metabolite levels and drought tolerance traits were identified for allantoin, galactaric acid, gluconic acid, glucose, a salicylic acid glucopyranoside and three unknown analytes with a retention time index of 1574.3, 1730.77 and 2482.9 (Figure 6, Figure 9 and Figure S2). Concentrations of these metabolites were 10 to 1000fold higher in tolerant plants than in sensitive plants. Under drought stress, levels of these metabolites were high in tolerant cultivars. However, for most of these metabolites no correlations between levels under control conditions and performance under drought were found. In contrast, galactaric acid concentrations under control conditions correlated positively with the performance under drought. As levels under control conditions correlated positively with growth under control conditions too, galactaric acid levels seemed to relate to growth rate rather than to drought tolerance. Glucose and gluconic acid, for which positive correlations between concentration and performance were restricted to drought stress conditions, were thus better marker candidates.

Principal Component Analysis

Improved prediction of drought tolerance might be gained from derived variables e.g. from linear combinations of gene expression or metabolite concentration values. To that end, we checked whether the variation in drought tolerance can be resolved by a combination of principle components on metabolite concentrations. In a dataset from leaves of control and drought-stressed plants grown in two experiments, component 1 (PC1) separated control and drought-stressed plants (Figure 10A). This component explained 31% of the variance. PC1 was a linear combination of many metabolites without obvious overrepresentation of metabolites from a single pathway (Table S3). A combination of PC2 and PC3, explaining 16% and 9% of the variance, respectively, separated *japonica* and *indica* cultivars (Figure 10B). In PC2, erythronic acid-1,4-lactone and three amino acids (aspartate, serine and threonine) had loadings higher than 0.2. However, tolerance differences between the cultivars could not be resolved in one of the PC-plots. Likewise, multiple regression approaches (data not shown) yielded poorly reproducible results that were highly dependent on the regression method and the quality

	Physiological data d vs. metabolites d							Physiological data d vs. metabolites c							Physiological data c vs. metabolites c				
Metabolite	Negative rank	WUE	Yield	FWS	DWS	FWT	DWT	Negative rank	WUE	Yield	FWS	DWS	FWT	DWT	Yield	FWS	DWS	FWT	DWT
Allantoin		0.42	0.38			0.41	0.35												
Asparagine		-0.64	-0.71	-0.61	-0.64	-0.68	-0.68		0.39	0.42			0.42	0.4				0.38	0.34
Erythronic acid		-0.36		-0.47	-0.46	-0.33	-0.38												
Erythronic acid-1,4-lactone		-0.56	-0.43	-0.56	-0.60	-0.55	-0.59		0.45	0.45		0.34	0.43	0.44				0.4	0.34
Ethanolamine	-0.43	-0.39		-0.51	-0.48	-0.41	-0.42												
Galactaric acid	0.32			0.37	0.34			0.4				0.35				0.36	0.38	0.36	0.39
Galactonic acid		-0.64	-0.57	-0.50	-0.59	-0.55	-0.63												
Galactosylglycerol			-0.34						0.39			0.35	0.36	0.39		0.42	0.4	0.53	0.47
Gluconic acid		0.35				0.32	0.33		-0.59	-0.72	-0.51	-0.45	-0.66	-0.57					
Glucose		0.35	0.38	0.35		0.36	0.34												
Glutamate		-0.38		-0.32	-0.35	-0.31	-0.37												
Glutamine		-0.38	-0.41	-0.45	-0.44	-0.41	-0.43												
Glycine	-0.34		-0.32	-0.38	-0.37	-0.31	-0.33												
Glycosyl inositol conjugate		-0.34	-0.33	-0.38	-0.37	-0.36	-0.37	-0.61											
NA 1574.3	0.38			0.36	0.41		0.33												
NA 1691.2	-0.33	-0.39	-0.47	-0.49	-0.46	-0.47	-0.43												
NA 1730.77	0.45	0.37	0.53	0.36	0.41	0.4	0.44												
NA 2482.9		0.44	0.48	0.47	0.49	0.48	0.48		-0.44			-0.36	-0.37	-0.41		-0.33	-0.36	-0.37	-0.38
Salicylic acid glucopyranoside		0.38	0.39	0.45	0.39	0.4	0.37												
Serine				-0.44	-0.43	-0.32	-0.32		0.49	0.41			0.45	0.44					
Threonic acid		-0.49	-0.32	-0.51	-0.56	-0.48	-0.52												
Threonine	-0.31				-0.31				0.37	0.35			0.34	0.33					
Xylose	-0.37	-0.37	-0.44	-0.5	-0.48	-0.43	-0.42												

Figure 6. Correlation of physiological data with metabolite levels. Correlation coefficients for selected metabolites with significant ($p<0.05$) positive (blue) or negative (red) correlation between log-transformed metabolite levels with physiological data under drought (d) or control (c) conditions. Data of 21 different cultivars grown in two experiments. Mean values of three to five replicates per cultivar and condition and experiment were correlated. Negative rank - mean scoring rank multiplied with -1; WUE – water use efficiency (g DW/g H_2O per day); yield - chlorophyll-*a* fluorescence yield; FWS - fresh weight shoot [g]; DWS - dry weight shoot [g]; FWT - total fresh weight [g]; DWT - total dry weight [g].
doi:10.1371/journal.pone.0063637.g006

criterion. Thus, no combined markers based on multiple metabolite levels could be gained from the data.

Discussion

Drought tolerance in crops is an increasingly relevant trait, as water availability is the limiting factor for plant production especially in those parts of the world, in Asia and in Africa, where malnutrition is a major issue. However, drought tolerance is a quantitative agricultural trait that is very difficult and labour-intensive to determine. In the past, drought tolerance has been assessed in field trials to measure either final yield or physiological parameters that are predictive for yield under stress. Yield itself is the most relevant parameter, but its heritability is regrettably low. Additionally, drought tolerance depends very much on the target environment. Thus, marker search concentrates on features that predict traits contributing to drought tolerance in a defined environment [8,25]. One of these traits is the stay-green trait that estimates the degree of leaf chlorosis and necrosis [9] under stress. The ability to maintain a high biomass under drought stress at the juvenile stage enhances plant survival after transplanting as well as rapid recovery after drought. Both features increase yield.

We used the stay-green trait measured as drought score categories as the main trait for the quantification of drought tolerance in our test population and tested the predictability of this from transcript or metabolite data. Additionally, we checked whether we can predict further traits relevant for drought tolerance, such as chlorophyll fluorescence yield [38], water use efficiency, or total and shoot biomass. In QTL studies, the association between genes or genomic markers and various proxy-parameters for drought tolerance is not consistent [32,39]. In our study, some metabolites like asparagine concentrations and some transcript levels e.g. of asparagine synthetase correlated closely to several traits, whereas others marker candidates correlated specifically to e.g. water use efficiency or chlorophyll fluorescence.

To speed up breeding by marker assisted selection (MAS), markers should allow tolerance prediction from features that can be measured on young plants, ideally without the need of prior stress treatment [19,22]. In contrast to genomic markers, metabolite and transcript levels vary with the environmental conditions, the plant organ and the developmental stage. We performed our analysis in the juvenile growth stage on fully expanded leaves as easily accessible organ. Additionally, we tested for correlations between tolerance traits and metabolite or transcript levels measured under drought stress and control conditions to find markers that are independent of the water supply.

The Test Population

For MAS, the correlation between marker and tolerance must hold in a wide range of genetic backgrounds. We therefore tested potential expression and metabolite markers in 21 rice cultivars. Most of the cultivars were selected from a Vietnamese breeding program to gain a test population comparable to the breeding material, for which the markers are intended. Our test population represented the two major subspecies of rice, *indica* and *japonica*. The substructure in our data set needs to be taken into account, as otherwise, like in association mapping, false-positive associations between genotype – or in our case marker - and phenotype may result [40]. For association mapping, statistical approaches are available to control the influence of the substructure [41]. In our study, cultivars belonging to the *indica* subspecies tended to be

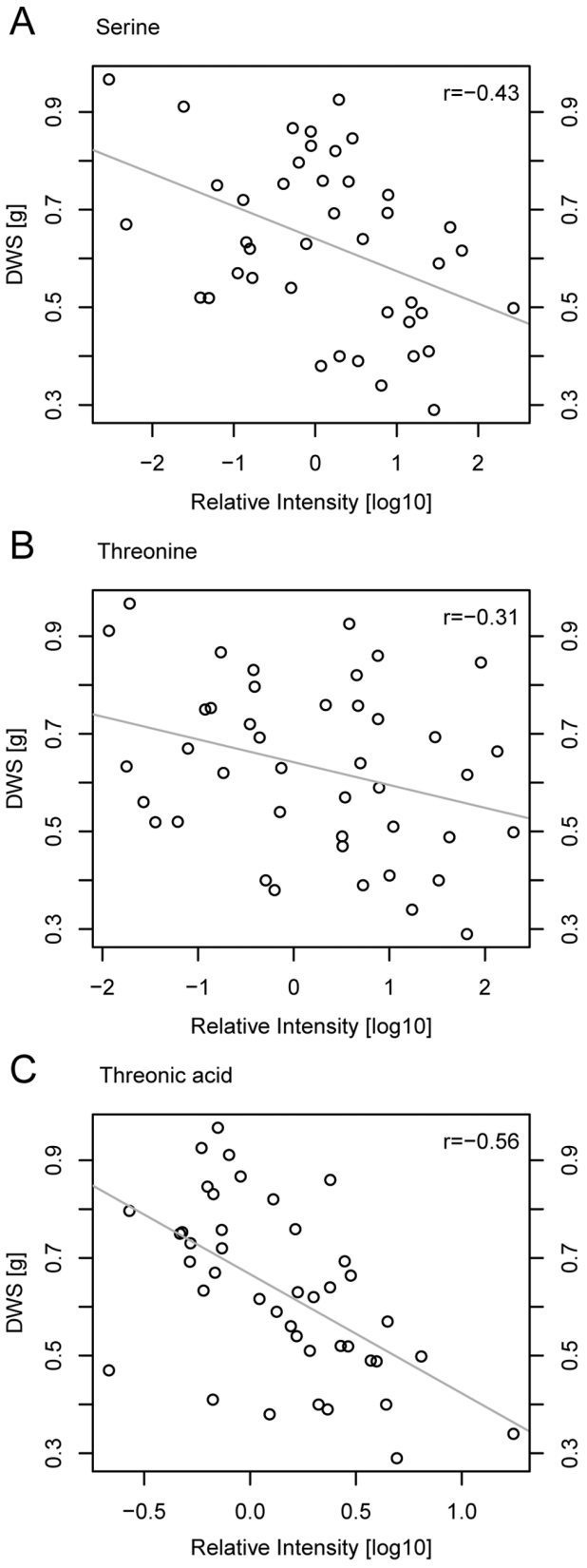

Figure 7. Levels of serine (A) threonine (B) and threonic acid (C) correlated negatively with shoot dry weight of rice plants under drought stress. Average metabolite levels and average shoot dry weight of 20 rice cultivars from two independent experiments.
doi:10.1371/journal.pone.0063637.g007

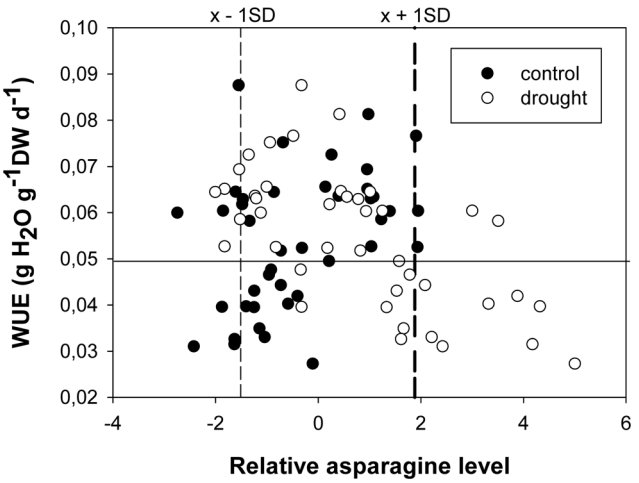

Figure 8. High asparagine levels are predominantly found in cultivars with low water use efficiency. Water use efficiency (g water per g final dry weight per day) of rice cultivars grown under control or drought conditions plotted against the relative asparagine level (Z score of log2 transformed values) in their leaves. The vertical reference lines indicate the average asparagine level of all samples minus (left) or plus (right) one standard deviation.
doi:10.1371/journal.pone.0063637.g008

more tolerant than the *japonica* cultivars. Subspecies-specific differences in a metabolite level can thus lead to pseudocorrelations. The consequence of the substructure in the population is clearly visible in the PCA on the metabolite data set, where the third component separated *indica* and *japonica* cultivars. For the transcript data, the substructure was broken by the pre-selection of the candidate genes [36]. The expression of these genes in two sensitive (both *japonica*) and two tolerant cultivars (one *indica* and one *japonica*) was significantly affected by condition × tolerance group interaction. The risk of pseudocorrelations was therefore much lower for the selected candidate genes than for the metabolite markers.

The samples for the marker search were taken in the early vegetative growth phase of the cultivars before flower initiation. Under the climate chamber conditions employed in the experiment, the cultivar with the shortest live cycle (Nipponbare) flowered 55 days after sowing, most of the other cultivars flowered about 100 days after sowing, some considerably later (Köhl, unpublished data). By precise definition of the sampling time in the vegetative growth phase, we reduced the effect of differences in the live cycle term between cultivars on the validity of the marker. The cultivars showed considerable variation in height and tiller number (see accompanying paper Do et al. PLOS ONE 10.1371/journal.pone.0060325) and cultivars with short shoots generally grew more tillers than cultivars with high shoots. Thus, the selected population represented the variance in growth patterns found in rice.

Multi-parallel Methods for Marker Search

Metabolite and expression profiling both allow multi-parallel measurements of several hundreds to thousands of parameters with predictive capability. Each method has their relative merits. Expression profiling by microarray hybridization is by now well established for several crops. Based on such analyses, PCR based tests can be established for candidate genes. If a linked genetic marker can be identified, a genome-based test can be designed. This allows fast screening at an early growth stage, independent of environmental conditions. If the functions of the proteins encoded

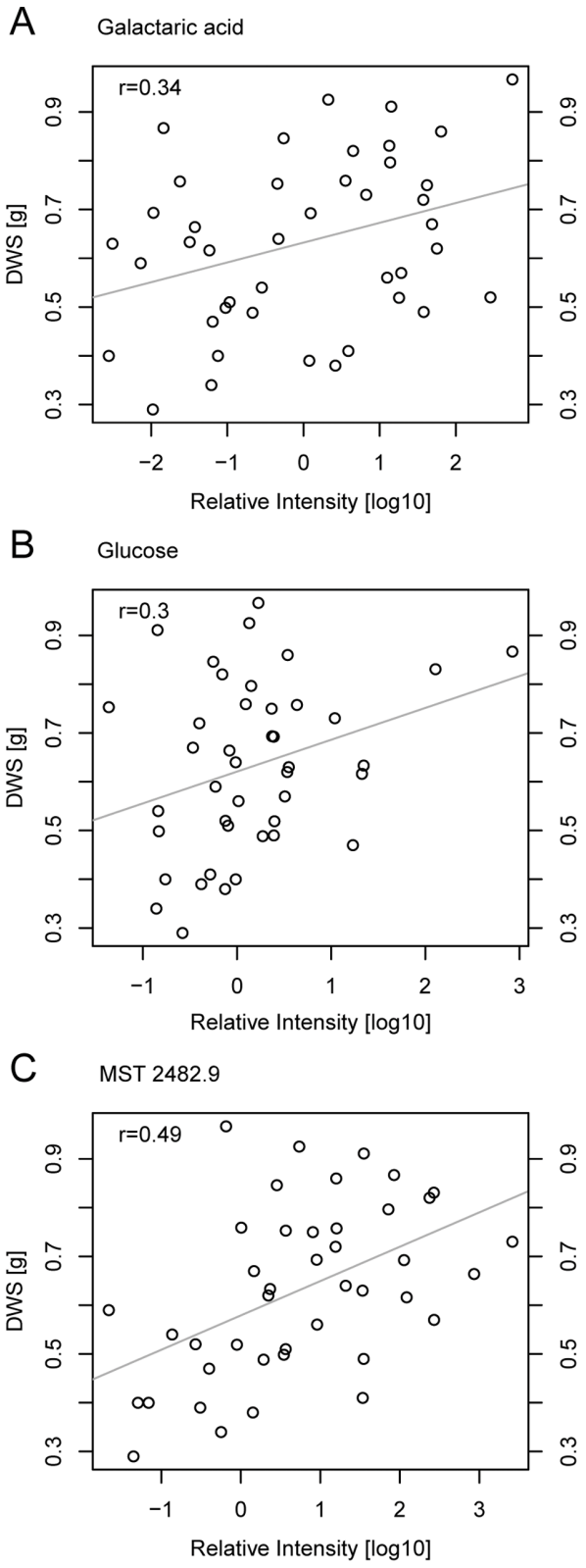

Figure 9. Levels of galactaric acid (A), glucose (B) and MST 2482.9 (C) correlate positively with shoot dry weight of rice plants under drought stress. Average metabolite levels and average shoot dry weight of 20 rice cultivars from two independent experiments.
doi:10.1371/journal.pone.0063637.g009

by the genes showing altered expression are known, regulatory or metabolic pathways that affect drought tolerance can be identified. The gained insight into drought tolerance mechanisms can then be used to increase tolerance by altering the expression of a key gene [24,42].

Metabolite profiling generally yields less response parameters than expression profiling and the ratio between found analytes and known metabolites is generally worse than the ratio between genes of known and unknown function. In rice, for which metabolite profiling is still in an early stage, GC-MS profiling yields 50–150 known metabolites [43]. In spite of this limitation, metabolomics is becoming a major tool to study plant stress responses [44–48] and will become a key factor in molecular breeding [49–51]. A major advantage of metabolite profiling is the huge body of reference data available from more than a hundred years of biochemical research compared to only thirty years of genomic research. If a metabolite is found to correlate with stress tolerance, relevant pathways, in which the metabolite is involved, can thus be rapidly identified and the mechanism of tolerance unravelled. In contrast to most genes, metabolite markers provide condensed information over several processes [52]. Thus, metabolite and gene expression markers both have their advantages.

The main disadvantage of metabolite and expression markers is their lower stability compared to genomic markers. Metabolite and RNA concentrations can be influenced by diurnal rhythm, environmental conditions and developmental stage of the plant. This can be taken into account by standardising the sampling conditions and choosing developmental stages that are metabolically relatively stable (e.g. vegetative growth in Poaceae) and time intervals in the diurnal cycle, in which metabolite and transcript concentrations change but slowly [53]. Another approach is to choose metabolite or transcript markers, in which concentration differences between tolerant and sensitive cultivars are large compared to the changes caused by environmental factors or diurnal rhythms. In contrast to transcript and metabolite markers, genomic DNA markers like SSR or SNPs are independent of environmental conditions and developmental stages. The identification of genomic markers by association or QTL mapping requires phenotyping and genotyping of a sufficiently large segregating mapping population [54] and is thus much more labour-intensive than the identification of metabolite or transcript markers. However, both approaches can be combined. Instead of doing a genome-wide association study, the region of interest can be narrowed down to the location of candidate genes from transcript profiling. In contrast to genomic markers, transcript and metabolite markers can be preselected based on their response to the stress, for which tolerance markers are to be identified.

Marker Identification by Correlation Analysis

To test the value of potential markers, we first characterised a population of 21 rice cultivars for drought tolerance and phenotyped them for traits that had been used to predict drought tolerance. Levels of 46 candidate genes and 79 metabolites were measured in leaves of 21 cultivars, which had been grown under control and drought conditions. Potential drought tolerance markers were identified by analysing correlations of expression and metabolite levels with the phenotypic traits. Significant positive correlations of metabolite or transcript levels with phenotypic traits indicate a high expression or metabolite level in tolerant cultivars, while negative correlations indicate high levels in sensitive cultivars. As high levels of a metabolite or transcript can be ascertained more reliably than low levels or absence, we focus on the prediction from high levels. In the first case (positive correlation), the metabolite or gene expression would be a

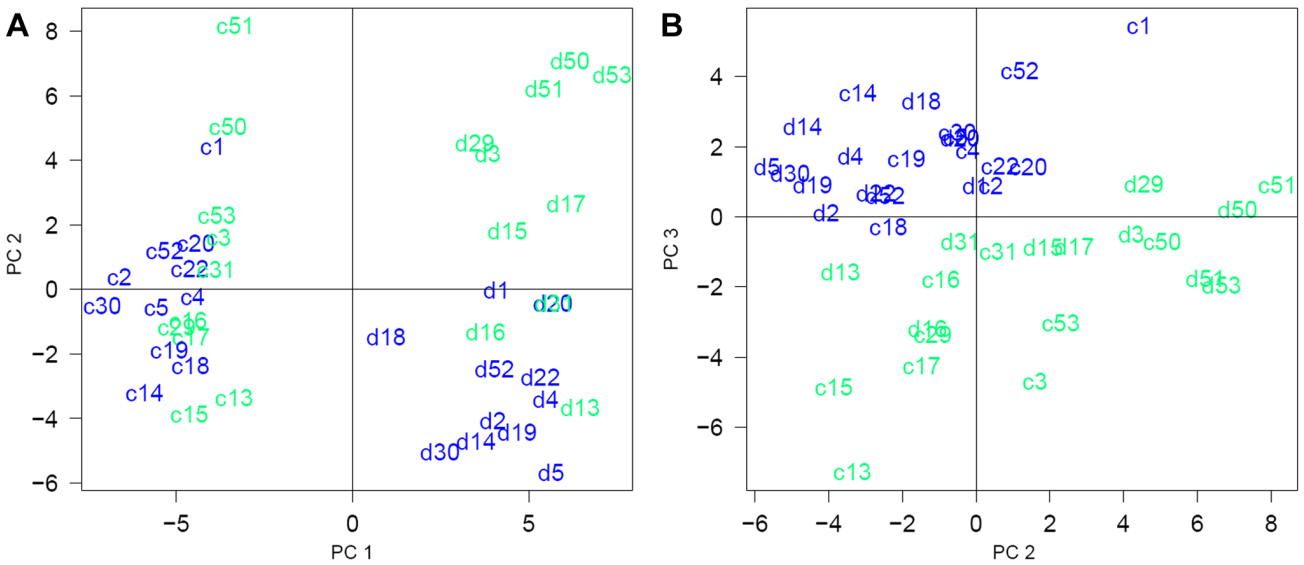

Figure 10. Principal component analysis (PCA) separated samples by treatment and genetic origin of rice cultivars. PCA plots on normalised metabolite levels with PC1 and PC2 (A) separated samples of plants grown under control (c) and drought conditions (d). PC2 and PC3 (B) separated samples of *indica* (blue) and *japonica* cultivars (green). Numbers for cultivars see Table 1.
doi:10.1371/journal.pone.0063637.g010

tolerance marker, as high levels indicate tolerance. Gene expression or metabolites with negative correlation are sensitivity markers.

The accumulation of a metabolite or transcript under stress is not necessarily functionally connected with an increase in the tolerance level or with tolerance differences between genotypes. Metabolite levels can increase as a result from an accelerated degradation or a reduced biosynthesis of another metabolite without any protective effect. Likewise, not all candidate genes, for which expression levels are significantly correlated with physiological data, are necessarily connected with drought tolerance. It could, for example, also be generally correlated with the growth rate not only under drought but also under control conditions. To identify such false positives, we studied the correlation between expression and metabolite levels under control conditions and phenotypic traits measured under these conditions. Indeed, expression levels of, for example, the genes coding for an asparagine synthetase and an AMP deaminase were negatively correlated with shoot and total fresh and dry weight under drought conditions as well as under control conditions. This suggests that high expression of those genes indicated slow growth rather than specific performance under drought.

Metabolite Markers

Among metabolite levels, potential sensitivity markers were identified that correlated significantly and negatively with phenotypic traits under drought stress. This group contains many amino acids (asparagine, glutamate, glutamine, glycine, serine and threonine). Their concentrations were high in drought-sensitive cultivars with low biomass under drought stress. This agrees with findings in *Arabidopsis thaliana* grown under optimal conditions where intermediates of the central metabolism were mostly negatively correlated with biomass production [48,55]. High amino acid levels observed in sensitive cultivars reflect the increase in protein degradation and the decrease in protein synthesis under drought stress [36,48,56,57]. Accordingly, high amino acid levels have been observed in plants subjected to other stresses and in senescing leaves [58–61].

In contrast, eight metabolites were identified, whose levels under drought stress were higher in tolerant than in sensitive cultivars. This pattern was found for allantoin, galactaric and gluconic acid, glucose and salicylic acid glucopyranoside plus three unidentified metabolites. These metabolites are promising candidates for drought tolerance markers. Especially mid-day glucose level in young leaves is an interesting marker candidate, as levels of glucose were already shown to be increased during drought in Eucalyptus [61,62], during salt stress in Lotus [48] and to correlate significantly with acclimated freezing tolerance in *Arabidopsis thaliana* [48,63,64]. In all these stresses, glucose may be part of a C-based osmotic adjustment [65,66]. In contrast to sucrose and starch, glucose concentration is not negatively correlated with biomass production under unstressed conditions [67]. The positive correlations were restricted to plants under stress conditions and were not found under control conditions. Thus, no constitutive metabolite markers could be identified.

Gene Expression Markers

Among the 46 candidate genes selected from a previous study [36], more than 20% showed significant correlation between expression and plant performance under drought stress in the test population. Similar to the metabolite markers, negative correlations dominated. For many of these sensitivity markers, the correlation with performance under drought could also be found for expression levels measured under control conditions. These sensitivity markers seem to be constitutive and may thus be useful to exclude germplasms from a breeding population at an early stage. In the case of asparagine synthetase, the increased gene expression in sensitive cultivars was matched by an increased asparagine level in these cultivars.

Many of the genes with a positive correlation between expression and performance under drought showed this correlation exclusively under drought conditions. These genes are scientifically interesting, but of limited value as breeding markers. These tolerance markers code for proteins involved in several pathways, which fits the general assumption of a multigenic nature of drought tolerance. Remarkably, expression levels of a cytosolic

(cFBPase) and a plastidial fructose-1,6-bisphosphatase (pFBPase) were positively correlated with drought tolerance or photosynthetic yield under drought stress, respectively. In sensitive cultivars, the cFBPase was slightly down-regulated under drought stress, whereas it was significantly induced in tolerant cultivars [36]. Both enzymes are physiologically and biochemically very well studied [68–70]. The plastidial enzyme catalyzes a rate-limiting, highly regulated step in the Calvin-Benson cycle [71] towards regeneration of ribulose-bis-phosphate and starch production in the chloroplast. The cytosolic enzyme promotes a highly regulated step in the conversion of triose phosphates into sucrose, which then may be exported to sink organs such as roots. A down-regulation of the genes for the cytosolic enzyme has been observed before in water-stressed sunflower plants [72]. The authors suggest that the down-regulation of this enzyme could be involved in the non-stomatal limitation of photosynthesis [72]. Additionally, it is discussed that rates of sucrose synthesis and rates of photosynthesis may be coordinated by changes in the activity of this enzyme [73]. In the Arabidopsis mutant (*hcef*) with increased cyclic electron flow around photosystem I, the mutation has been mapped to the pFBPase [74]. Antisense repression of cFBPase reduced sucrose synthesis in Arabidopsis [75] and potato [76]. When cFBPase is overexpressed together with the triose phosphate/phosphate transporter, photosynthetic CO_2 assimilation rates are enhanced and glucose levels increased compared to the wildtype [77]. The increased expression of both *FBPase* genes in tolerant rice cultivars under stress is counterintuitive, as the enzymes are crucial parts of competing pathways. However, a similar situation was revealed in a detailed metabolic analysis of Arabidopsis under drought stress [65]. The activity of AGPase, the rate limiting enzyme in plastidial starch synthesis, was increased in severely drought-stressed plants compared to control plants during the entire diurnal cycle. At the same time, the amount of sucrose exported to the roots increased under drought stress. A possible explanation is that a higher expression level of both genes may allow a higher turnover of the enzymes and thus an increased regulatory capacity for the switch between photosynthetic CO_2 fixation and starch storage in the chloroplasts and carbohydrate exports to the sinks, especially the roots. The adaptive value of modifications in the source/sink relationship has been shown in rice. Increase in cytokinin synthesis by genetic modification improved grain yield under drought [78]. A regulation switch with a high capacity is obviously only needed, if triose phosphates are produced by photosynthesis. In sensitive cultivars, where chlorosis and necrosis reduced photosynthesis under drought stress, this regulatory capacity may not be needed, thus gene expression levels may be decreased.

Altogether, the application of expression and metabolic profiling methods on rice cultivars subjected to long-term drought stress revealed several marker candidates for drought tolerance. The most promising markers were glucose, high levels of which indicated high tolerance and high expression levels of the *cFBPase* and *pFBPase* genes. Their elevated expression in tolerant cultivars may contribute to the adjustment of photosynthesis and source-sink relationships under drought. The test population, in which these marker candidates were identified, was highly diverse in drought tolerance and genetic background. This makes it likely that the markers are useful for breeder's selection in a wide range of rice germplasm. As the correlations between transcript and metabolite levels and drought tolerance were found in a controlled-environment drought stress system, the next step required would be the validation in field experiments. These experiments would also give insight into the effect of environmental factors other than water supply on the potential markers.

Materials and Methods

Plant Material and Stress Treatment

Twenty-one rice (*Oryza sativa* L.) cultivars (Table 1) originating either from the IBT (Institute of Biotechnology, Hanoi, Vietnam) or from the IRRI (International Rice Research Institute, Manila, Philippines) were grown under water sufficient and water limiting conditions in three independent experiments (#1–3) in a controlled climate chamber as described by Degenkolbe et al. [36]. The design was a split-plot design with five blocks per drought or control treatment. Each treatment and cultivar was represented by five replicate pots with one plant per pot. Pots were randomized within the blocks. Block position was rotated daily. Plants were cultivated in 10 cm pots on a 7.5 cm deep layer of an artificial substrate. The shallow substrate level was chosen to reduce the effect of differences in rooting patterns between cultivars on the result. Pots were positioned in polypropylene boxes filled with water to the level of the substrate surface. Rice plants were grown in 12 h day (600 µE m^{-2} s^{-1}) with 26°C and 75% relative humidity in the light and 22°C and 70% relative humidity at night. Twenty-six days after sowing, water was removed from half of the boxes and plants were left to dry for four days, until the soil water content had reached the permanent wilting point (PWP) for 50% of the plants. Thereafter, the soil water content was kept constant to the fixed PWP value over a period of 14 days by weighing each pot at the end of the light period and adding the amount of water lost during the last 24 hours.

After 18 days of drought stress, plants were harvested four to six hours after the beginning of the light period. Samples for expression and metabolic profiling were harvested from the middle section of the blades of fully expanded green leaves, weighted and immediately frozen in liquid nitrogen, and stored at -80°C until use.

Cultivars were genotyped for seven subspecies-specific sequence tagged site (SS-STS) markers [37] as described before [36].

RNA Isolation and cDNA Synthesis

Frozen leaf material was homogenized in a ball mill for 90 sec at 28 Hz. Plant material from five replicate plants per cultivar and condition was pooled and 60 to 90 mg were used for total RNA isolation using the NucleoSpin RNA plant kit (Macherey-Nagel, Düren, Germany) following manufacturer's instruction. RNA concentration was determined with the Nanodrop N-1000 Spectrophotometer (Nanodrop Technologies, Wilmington, DE). To remove remaining genomic DNA, samples were treated with Baseline-ZERO DNase. The absence of genomic DNA contamination was subsequently confirmed by quantitative RT-PCR (qRT-PCR) with primers for an intron (LOC_Os01g01840). The integrity of total RNA was checked on an 1.7% (w/v) agarose gel. cDNA synthesis from 4 µg of total RNA with Superscript III reverse transcriptase was performed following the manufacturer's instruction (Invitrogen, Karlsruhe, Germany). Quality of synthesized cDNA was checked by qRT-PCR with two primer pairs binding to the 3′ and 5′ ends of the actin 1 (LOC_Os03g50890) transcript, respectively.

Quantitative RT-PCR

Expression levels of 46 candidate genes (Table S1) were measured in leaf material from 21 rice cultivars grown under control and drought stress conditions in three experiments. Primers for qRT-PCR (Table S4) were designed on the published *japonica* sequence with PrimerExpress 2.0 (Applied Biosystems, Darmstadt, Germany) and checked with NetPrimer (www.

premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html). Sequences were blasted on the databases of GRAMENE (www.gramene.org) and the Beijing Genomics Institute to ensure specific amplification in both *japonica* and *indica* cultivars. Correct size of the amplified region for each primer pair was checked by agarose gel electrophoresis.

qRT-PCR was performed with the ABI Prism 7900HT (Applied Biosystems, Foster City, CA) using SYBR Green Master Mix (Eurogentec, Köln, Germany) with standard thermal cycling conditions (50°C for 2 min, 95°C for 10 min, 40 cycles of 95°C for 15 sec and 60°C for 1 min). Dissociation curves were checked with the SDS 2.2.1 software (Applied Biosystems) for shoulders or additional peaks. The expression values were normalised to the expression of the housekeeping genes actin 1 and cyclophilin and the primer efficiency as described before [36]: 'Normalised expression of the genes of interest was calculated by dividing the average relative expression (primer efficiency P to the power of cycle number Ct) of the two housekeeping genes (H1 and H2) by the relative expression of the gene of interest (GOI): $((P_{H1}\text{^}Ct_{H1}+P_{H2}\text{^}Ct_{H2})/2)/P_{GOI}\text{^}Ct_{GOI}$. Primer efficiency was calculated using LinRegPCR [79].

GC-ToF-MS

From 120 mg of frozen, ground leaf material from experiment #1 and 2, a fraction enriched in polar primary metabolites was prepared and processed as described previously [80]. Gas chromatography coupled to electron impact ionization-time of flight-mass spectrometry (GC/EI-TOF-MS) was performed on an Agilent 6890N24 gas chromatograph attached to a Pegasus III mass spectrometer, LECO, St. Joseph, USA [81]. Chromatograms were pre-processed with ChromaTOF software 1.00, Pegasus driver 1.61 (Leco; http://www.leco.de). Selective peak heights representing arbitrary mass spectral ion currents were normalised by sample dry weight and to an internal standard that was added upon extraction of the polar metabolite fraction. Data were subsequently processed with TagFinder [82]. Analytes that were detected in less than 50% of control and 50% of drought-stressed plants were excluded from the dataset. Clusters of at least three corresponding mass fragments were selected for relative metabolite quantification. Metabolites were identified by matching to references in the Golm Metabolome Database [83]. The matching process was manually supervised for a match factor >500 and retention index deviations <1% [84].

Outlier samples were detected in plots of a principal component analysis (PCA; R package pcaMethods; [80,85] of raw data and were removed from further analysis. Log-transformed metabolite levels were normalised by subtracting the median metabolite level for each experiment and metabolite to remove the effect of experiment and GC-MS run. Mean values of normalised metabolite levels were calculated for each cultivar, condition and experiment and analysed by PCA with the settings mean centred matrix and unit variance scale (R package pcaMethods). Euclidean distance of scaled data was used for hierarchical clustering.

Correlation Analysis

Expression of candidate genes and metabolites levels were analysed for Pearson correlations (cor.test function, R) with physiological data that are indicative of drought tolerance, namely shoot and total fresh and dry weight, mean scoring rank, mean water use efficiency and photosynthesis yield. Expression and metabolite data were log-transformed. Scoring ranks were multiplied with -1. The correlation analysis was performed on three variable combinations, namely correlating (1) expression/metabolite data from drought-stressed plants with performance parameters from drought-stressed plants, (2) expression/metabolite data from control plants with performance parameters under stress and (3) expression/metabolite data from control plants with performance parameters under control conditions.

Supporting Information

Figure S1 Average shoot dry weight in rice cultivars plotted against the relative expression of 46 genes. The regression coefficient r for the linear regression of shoot dry weight against expression level is shown in the upper left corner. The primer number and the gene name are indicated in the title of each figure. The complete name for each gene can be retrieved from Table S4. File SupportingFigure2.pdf, Format pdf.
(PDF)

Figure S2 Average shoot dry weight in rice cultivars plotted against the level (= signal intensity) of 79 metabolites. The regression coefficient r for the linear regression of shoot dry weight against metabolite level is shown in the upper left corner. Unidentified metabolites are labelled with their retention time index MST. File SupportingFigureS4.pdf. Format pdf.
(PDF)

Table S1 Mean normalised expression values for 45 genes measured in 21 cultivars under drought and control conditions. Cultivar identifiers see manuscript Table 1. Condition d = drought, c = control. Primer identifiers indicated in column heading see Table S4. File SupplementalTableS1.xls, Format xls.
(XLS)

Table S2 Mean normalised metabolite levels measured in rice cultivars under drought and control conditions. Cultivar identifiers see Table S1. Condition d = drought, c = control. Metabolite identifiers indicated in column heading see Table S3. NA = not detected. File Supplemental Table S3.xls, Format xls.
(XLS)

Table S3 Metabolite identifiers (Mid), retention times, metabolite names and loadings of the first five principal components. File Supplemental Table S5.pdf, Format pdf.
(PDF)

Table S4 List of qPCR primer sequences that were used for quality checks and RT-PCR together with their Primer Identifier (PId), TIGR Locus identifier (Locus ID), OligoID from the gene chip (see Degenkolbe et al. 2009), and direction (FW = forward, RV = reverse). File Supplemental Table S6.pdf, Format pdf.
(PDF)

Acknowledgments

We thank Le Tran Binh from the Institute of Biotechnology, Hanoi, Vietnam and Ruaraidh Sackville Hamilton and Gary Atlin from the IRRI for the seed material. At the MPI of Molecular Plant Physiology, Golm, we thank Dirk Walther, Matt Hannah and Björn Usadel for assistance with software tools and discussions about R, Alexander Erban and Ines Fehrle for skilled technical assistance and Camila Caldana for testing primers of rice housekeeping genes.

Author Contributions

Conceived and designed the experiments: TD PD EZ DH KK. Performed the experiments: TD PD JK EZ DH KK. Analyzed the data: TD JK KK. Contributed reagents/materials/analysis tools: TD PD JK DH KK. Wrote the paper: TD DH KK.

References

1. Khush G (2003) Productivity improvements in rice. Nutr Rev 61: S114–S116.
2. Cheng CY, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, et al. (2003) Polyphyletic origin of cultivated rice: Based on the interspersion pattern of SINEs. Mol Biol Evol 20: 67–75.
3. Bernier J, Atlin GN, Serraj R, Kumar A, Spaner D (2008) Breeding upland rice for drought resistance. J Sci Food Agric 88: 927–939.
4. Hanson AD, Peacock WJ, Evans LT, Arntzen CJ, Khush GS (1990) Drought resistance in rice. Nature 345: 26–27.
5. Jongdee B, Fukai S, Cooper M (1998). Genotypic variation for grain yield of rice under water-deficit conditions. In: Michalk D, Pratley J, editors; Proceedings of 9th Australian Agronomy Conference. Wagga Wagga. 403–406.
6. Jongdee S, Mitchell J, Fukai S (1997). Modeling approach for estimation of rice yield reduction due to drought in Thailand. In: Fukai S, Cooper M, editors. ACIAR Proceedings; Ratchathani, Thailand. 65–73.
7. Serraj R, McNally KL, Slamet-Loedin I, Kohli A, Haefele SM, et al. (2011) Drought resistance improvement in rice: an integrated genetic and resource management strategy. Plant Prod Sci 14: 1–14.
8. Reynolds M, Tuberosa R (2008) Translational research impacting on crop productivity in drought-prone environments. Curr Opin Plant Biol 11: 171–179.
9. Fischer KS, Lafitte R, Fukai S, Atlin GN, Hardy B (2003) Breeding rice for drought-prone environments. Los Banos (Philippines): International Rice Research Institute. 98.
10. Manickavelu A, Nadarajan N, Ganesh SK, Gnanamalar RP, Babu RC (2006) Drought tolerance in rice: morphological and molecular genetic consideration. Plant Growth Regul 50: 121–138.
11. Price AH, Tomos AD, Virk DS (1997) Genetic dissection of root growth in rice (Oryza sativa L).1. a hydrophonic screen. Theor Appl Genet 95: 132–142.
12. Price AH, Cairns JE, Horton P, Jones HG, Griffiths H (2002) Linking drought-resistance mechanisms to drought avoidance in upland rice using a QTL approach: progress and new opportunities to integrate stomatal and mesophyll responses. J Ex Bot 53: 989–1004.
13. Tripathy JN, Zhang J, Robin S, Nguyen TT, Nguyen HT (2000) QTLs for cell-membrane stability mapped in rice (*Oryza sativa* L.) under drought stress. Theor Appl Genet 100: 1197–1202.
14. Havaux M, Lannoye R (1985) Invivo chlorophyll fluorescence and delayed light-emission as rapid screening techniques for stress tolerance in crop plants. J Plant Breed 95: 1–13.
15. Dingkuhn M, Cruz RT, O'Toole JC, Doerffling K (1989) Net photosynthesis water use efficiency leaf water potential and leaf rolling as affected by water deficit in tropical upland rice. Aust J Agric Res 40: 1171–1182.
16. Laza MR, Kondo M, Ideta O, Barlaan E, Imbe T (2006) Identification of quantitative trait loci for {delta}^{13}C and productivity in irrigated lowland rice. Crop Sci 46: 763–773.
17. Centritto M, Lauteri M, Monteverdi MC, Serraj R (2009) Leaf gas exchange, carbon isotope discrimination, and grain yield in contrasting rice genotypes subjected to water deficits during the reproductive stage. J Ex Bot 60: 2325–2339.
18. Gonzalez A, Bermejo V, Gimeno BS (2010) Effect of different physiological traits on grain yield in barley grown under irrigated and terminal water deficit conditions. J Agric Sci 148: 319–328.
19. Richards RA, Rebetzke GJ, Watt M, Condon AG, Spielmeyer W, et al. (2010) Breeding for improved water productivity in temperate cereals: phenotyping, quantitative trait loci, markers and the selection environment. Funct Plant Biol 37: 85–97.
20. Munns R, James RA, Sirault XRR, Furbank RT, Jones HG (2010) New phenotyping methods for screening wheat and barley for beneficial responses to water deficit. J Ex Bot 61: 3499–3507.
21. Gouda PK, Varma CMK, Saikumar S, Kiran B, Shenoy V, et al. (2012) Direct selection for grain yield under moisture stress in *Oryza sativa* cv. IR58025B x *Oryza meridionalis* population. Crop Sci 52: 644–653.
22. Babu RC, Nguyen BD, Chamarerk V, Shanmugasundaram P, Chezhian P, et al. (2003) Genetic analysis of drought resistance in rice by molecular markers: Association between secondary traits and field performance. Crop Sci 43: 1457–1469.
23. Kumar R, Venuprasad R, Atlin GN (2007) Genetic analysis of rainfed lowland rice drought tolerance under naturally-occurring stress in eastern India: Heritability and QTL effects. Field Crop Res 103: 42–52.
24. Salekdeh GH, Reynolds M, Bennett J, Boyer J (2009) Conceptual framework for drought phenotyping during molecular breeding. Trends Plant Sci 14: 488–496.
25. Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science 327: 818–822.
26. Nie Y-Y, Zou G-H, Li Y, Liu G-L, Cai Y-H, et al. (2012) Fine mapping of drought tolerance QTL on chromosome 2 in rice. Acta Agron Sin 38: 988–995.
27. Vikram P, Swamy BPM, Dixit S, Ahmed HU, Cruz MTS, et al. (2011) qDTY(1.1), a major QTL for rice grain yield under reproductive-stage drought stress with a consistent effect in multiple elite genetic backgrounds. BMC Genet 12.
28. Salunkhe AS, Poornima R, Prince KSJ, Kanagaraj P, Sheeba JA, et al. (2011) Fine mapping QTL for drought resistance traits in rice (*Oryza sativa* L.) using bulk segregant analysis. Mol Biotechnol 49: 90–95.
29. Bernier J, Kumar A, Ramaiah V, Spaner D, Atlin G (2007) A large-effect QTL for grain yield under reproductive-stage drought stress in upland rice. Crop Sci 47: 507–516.
30. Xu JL, Lafitte HR, Gao YM, Fu BY, Torres R, et al. (2005) QTLs for drought escape and tolerance identified in a set of random introgression lines of rice. Theor Appl Genet 111: 1642–1650.
31. Maccaferri M, Sanguineti MC, Corneti S, Ortega JLA, Salem MB, et al. (2008) Quantitative trait loci for grain yield and adaptation of durum wheat (*Triticum durum* Desf.) across a wide range of water availability. Genetics 178: 489–511.
32. Kumar S, Sehgal SK, Kumar U, Prasad PVV, Joshi AK, et al. (2012) Genomic characterization of drought tolerance-related traits in spring wheat. Euphytica 186: 265–276.
33. Messmer R, Fracheboud Y, Baenziger M, Vargas M, Stamp P, et al. (2009) Drought stress and tropical maize: QTL-by-environment interactions and stability of QTLs across environments for yield components and secondary traits. Theor Appl Genet 119: 913–930.
34. Vargas M, van Eeuwijk FA, Crossa J, Ribaut JM (2006) Mapping QTLs and QTL x environment interaction for CIMMYT maize drought stress program using factorial regression and partial least squares methods. Theor Appl Genet 112: 1009–1023.
35. Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, et al. (2005) GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics: 1635–1638.
36. Degenkolbe T, Do PT, Zuther E, Repsilber D, Walther D, et al. (2009) Expression profiling of rice cultivars differing in their drought tolerance to long-term drought stress. Plant Mol Biol 69: 133–153.
37. Chin J-H, Kim J-H, Jiang W, Chu S-H, Woo M-O, et al. (2007) Identification of subspecies-specific STS markers and their association with segregation distortion in rice (*Oryza sativa* L.). J Crop Sci Biotech 10: 175–184.
38. Liu H, Zou G, Liu G, Hu S, Li M, et al. (2005) Correlation analysis and QTL identification for canopy temperature, leaf water potential and spikelet fertility in rice under contrasting moisture regimes. Chinese Sci Bull 50: 317–326.
39. Yu S, Liao F, Wang F, Wen W, Li J, et al. (2012) Identification of rice transcription factors associated with drought tolerance using the ecotilling method. PLoS One 7 doi 0030765.
40. Pritchard JK, Donnelly P (2001) Case-control studies of association in structured or admixed populations. Theor Popul Biol 60: 227–237.
41. Yu JM, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotech 17: 155–160.
42. Reguera M, Peleg Z, Blumwald E (2012) Targeting, metabolic pathways for genetic engineering abiotic stress-tolerance in crops. BBA Gene Regul Mech 1819: 186–194.
43. Zuther E, Köhl KI, Kopka J (2007) Comparative metabolome analysis of the salt response in breeding cultivars of rice. In: Jenks MA, Hasegawa PM, Jain SM, editors. Advances in molecular breeding toward drought and salt tolerant crops: Springer. 285–315.
44. Cramer GR, Ergul A, Grimplet J, Tillett RL, Tattersall EAR, et al. (2007) Water and salinity stress in grapevines: early and late changes in transcript and metabolite profiles. Funct Integr Genomic 7: 111–134.
45. Rizhsky L, Liang HJ, Shuman J, Shulaev V, Davletova S, et al. (2004) When Defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. Plant Physiol 134: 1683–1696.
46. Guy C, Kaplan F, Kopka J, Selbig J, Hincha DK (2008) Metabolomics of temperature stress. Physiologia Plantarum 132: 220–235.
47. Shulaev V, Cortes D, Miller G, Mittler R (2008) Metabolomics for plant stress response. Physiol Plantarum 132: 199–208.
48. Sanchez DH, Schwabe F, Erban A, Udvardi MK, Kopka J (2012) Comparative metabolomics of drought acclimation in model and forage legumes. Plant Cell Environ 35: 136–149.
49. Vinocur B, Altman A (2005) Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. Curr Opin Biotech 16: 123–132.
50. Oikawa A, Matsuda F, Kusano M, Okazaki Y, Saito K (2008) Rice metabolomics. Rice 1: 63–71.
51. Fernie AR, Schauer N (2009) Metabolomics-assisted breeding: a viable option for crop improvement? Trends Genet 25: 39–48.
52. Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, et al. (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44: 217–220.
53. Gibon Y, Usadel B, Blaesing OE, Kamlage B, Hoehne M, et al. (2006) Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in Arabidopsis rosettes. Genome Biol 7: doi:10.1186.
54. Stich B, Utz HF, Piepho H-P, Maurer H, Melchinger A (2010) Optimum allocation of resources for QTL detection using a nested association mapping strategy in maize. Theor Appl Genet 120: 553–561.
55. Meyer RC, Steinfath M, Lisec J, Becher M, Witucka-Wall H, et al. (2007) The metabolic signature related to high plant growth rate in Arabidopsis thaliana. P Natl Acad Sci USA 104: 4759–4764.
56. Munne-Bosch S, Alegre L (2004) Die and let live: leaf senescence contributes to plant survival under drought stress. Funct Plant Biol31: 203–216.
57. Roy-Macauley H, Zuily-Fodil Y, Kidric M, Thi ATP, Da Silva JV (1992) Effect of drought stress on proteolytic activities in phaseolus and Vigna leaves from sensitive and resistant plants. Physiol Plantarum 85: 90–96.

58. Mohammadi M, Kav NNV, Deyholos MK (2007) Transcriptional profiling of hexaploid wheat (*Triticum aestivum* L.) roots identifies novel, dehydration-responsive genes. Plant Cell Environ 30: 630–645.
59. Morcuende R, Bari R, Gibon Y, Zheng W, Pant BD, et al. (2007) Genome-wide reprogramming of metabolism and regulatory networks of Arabidopsis in response to phosphorus. Plant Cell Environ 30: 85–112.
60. Gibon Y, Pyl E-T, Sulpice R, Lunn JE, Höhne M, et al. (2009) Adjustment of growth, starch turnover, protein content and central metabolism to a decrease of the carbon supply when Arabidopsis is grown in very short photoperiods. Plant Cell Environ 32: 859–874.
61. Rivero RM, Ruiz JM, García PC, López-Lefebre LR, Sánchez E, et al. (2001) Resistance to cold and heat stress: accumulation of phenolic compounds in tomato and watermelon plants. Plant Sci 160: 315–321.
62. Warren CR, Aranda I, Javier Cano F (2012) Metabolomics demonstrates divergent responses of two Eucalyptus species to water stress. Metabolomics 8: 186–200.
63. Hannah MA, Wiese D, Freund S, Fiehn O, Heyer AG, et al. (2006) Natural genetic variation of freezing tolerance in Arabidopsis. Plant Physiol 142: 98–112.
64. Zuther E, Schulz E, Childs LH, Hincha DK (2012) Clinal variation in the non-acclimated and cold-acclimated freezing tolerance of *Arabidopsis thaliana* accessions. Plant Cell Environ 35: 1860–1878.
65. Hummel I, Pantin F, Sulpice R, Piques M, Rolland G, et al. (2000) Arabidopsis plants acclimate to water deficit at low cost through changes of carbon usage: an integrated perspective using growth, metabolite, enzyme, and gene expression analysis. Plant Physiol 154: 357–372.
66. Kim JY MA, Brangeon J, Prioul JL (2000) A maize vacuolar invertase, IVR2, is induced by water stress. Organ/tissue specificity and diurnal modulation of expression. Plant Physiol 124: 71–84.
67. Sulpice R, Pyl ET, Ishihara H, Trenkamp S, Steinfath M, et al. (2009) Starch as a major integrator in the regulation of plant growth. P Natl Acad Sci USA 106: 10348–10353.
68. Herzog B, Stitt M, Heldt HW (1984) Control of Photosynthetic Sucrose Synthesis by Fructose 2,6-Bisphosphate.3. Properties of the Cytosolic Fructose 1,6-Bisphosphatase. Plant Physiol 75: 561–565.
69. Hedrich R, Stitt M, Raschke K (1984) Fructose-2,6-bis-phosphate in guard cells presence and effects on cytosolic phospho-fructo-kinase and fructose-1,6-bis-phosphatase. Plant Physiol (Rockville) 75: 130.
70. Stitt M, Wirtz W, Heldt HW (1983) Regulation of sucrose synthesis by cytoplasmic fructosebisphosphatase and sucrose phosphate synthase during photosynthesis in varying light and carbon-dioxide. Plant Physiol 72: 767–774.
71. Stitt M, Lunn J, Usadel B (2010) Arabidopsis and primary photosynthetic metabolism - more than the icing on the cake. Plant J 61: 1067–1091.
72. Kiani SP, Grieu P, Maury P, Hewezi T, Gentzbittel L, et al. (2007) Genetic variability for physiological traits under drought conditions and differential expression of water stress-associated genes in sunflower (*Helianthus annuus* L.). Theor Appl Genet 114: 193–207.
73. Daie J (1993) Cytosolic fructose-1,6-bisphosphatase - a key enzyme in the sucrose biosynthetic-pathway. Photosynth Res 38: 5–14.
74. Livingston AK, Kanazawa A, Cruz JA, Kramer DM (2010) Regulation of cyclic electron flow in C-3 plants: differential effects of limiting photosynthesis at ribulose-1,5-bisphosphate carboxylase/oxygenase and glyceraldehyde-3-phosphate dehydrogenase. Plant Cell Environ 33: 1779–1788.
75. Strand A, Zrenner R, Trevanion S, Stitt M, Gustafsson P, et al. (2000) Decreased expression of two key enzymes in the sucrose biosynthesis pathway, cytosolic fructose-1,6-bisphosphatase and sucrose phosphate synthase, has remarkably different consequences for photosynthetic carbon metabolism in transgenic *Arabidopsis thaliana*. Plant J 23: 759–770.
76. Zrenner R, Krause KP, Apel P, Sonnewald U (1996) Reduction of the cytosolic fructose-1,6-bisphosphatase in transgenic potato plants limits photosynthetic sucrose biosynthesis with no impact on plant growth and tuber yield. Plant J 9: 671–681.
77. Cho M-H, Jang A, Bhoo SH, Jeon J-S, Hahn T-R (2012) Manipulation of triose phosphate/phosphate translocator and cytosolic fructose-1,6-bisphosphatase, the key components in photosynthetic sucrose synthesis, enhances the source capacity of transgenic Arabidopsis plants. Photosynth Res 111: 261–268.
78. Peleg Z, Reguera M, Tumimbang E, Walia H, Blumwald E (2011) Cytokinin-mediated source/sink modifications improve drought tolerance and increase grain yield in rice under water-stress. Plant Biotechnol J 9: 747–758.
79. Ramakers C, Ruijter J, Lekanne Deprez R, Moorman A (2003) Assumption-free analysis of quantiative real-time polymerase chain reaction (PCR) data. Neurosci Lett 339: 62–66.
80. Siahpoosh MR, Dehghanian E, Kamgar A (2011) Drought tolerance evaluation of bread wheat genotypes using water use efficiency, evapotranspiration efficiency, and drought susceptibility index. Crop Sci 51: 1198–1204.
81. Erban A, Schauer N, Fernie A, Kopka J (2007) Non-supervised construction and application of mass spectral and retention time index libraries from time-of-flight GC-MS metabolite profiles. In: Weckwerth W, editor. Metabolomics: methods and protocols. Totowa: Humana Press. 19–38.
82. Luedemann A, Strassburg K, Erban A, Kopka J (2008) TagFinder for the quantitative analysis of gas chromatography - mass spectrometry (GC-MS) based metabolite profiling experiments Bioinformatics 24: 732.
83. Hummel J, Strehmel N, Selbig J, Walther D, Kopka J (2010) Decision tree supported substructure prediction of metabolites from GC-MS profiles. Metabolomics 6: 322–333.
84. Strehmel N, Hummel J, Erban A, Strassburg K, Kopka J (2008) Estimation of retention index thresholds for compound matching using routine gas chromatography-mass spectrometry based metabolite profiling experiments. J Chromatogr B 871: 182–190.
85. Stacklies W, Redestig H, Scholz M, Walther D, Selbig J (2007) pcaMethods - a Bioconductor package providing PCA methods for incomplete data. Bioinformatics 23: 1164–1167.