We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK
CITATION
INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Association Mapping of Four Important Traits Using the USDA Rice Mini-Core Collection

Wengui Yan, Aaron Jackson, Melissa Jia, Wei Zhou, Haizheng Xiong and Rolfe Bryant

Additional information is available at the end of the chapter

1. Introduction

Classical QTL mapping reveals only a slice of the genetic architecture for a trait because only two alleles that differ between the two parental lines segregate. A comprehensive analysis of genetic architecture requires consideration of a diverse population that represents genetic variation in a species. Association mapping provides an effective method to identify QTL that have effects across a broad spectrum of germplasm (Yu et al. 2006). Many studies have used association mapping for important traits since it was introduced from human genetics (Yu et al. 2006; Kim et al. 2006; Huang et al. 2010; Kang et al. 2008). Genome-wide association scans are expected to be effective when linkage disequilibrium (LD) and marker density are sufficiently high, so that the random markers have a greater chance of being in disequilibrium with QTL across diverse genetic materials (Kim et al. 2006). A substantial number of QTL at close to gene resolution for important traits have been identified by genome-wide association studies (GWAS) in rice (Zhao et al. 2007). Recently, the USDA Rice Mini-Core (URMC) collection was developed and serves as a genetically diversified panel for mining genes of interest (Li et al. 2010). The URMC was derived from 1,794 accessions in the USDA rice core collection using PowerCore software based on 26 phenotypic traits and 70 molecular markers (Agrama et al. 2009). The core collection represents over 18,000 accessions in the USDA global genebank of rice (Yan et al. 2007). The URMC contains 217 accessions originating from 76 countries and covering 14 geographic regions worldwide. The Objective of this review is to analyze the genetic diversity and differentiation of the URMC for genome-wide association mapping of harvest index, grain yield, sheath blight resistance and hull silica concentration.

2. Materials and methods

2.1. Rice association panel

Of 217 accessions in the URMC, 203 belong to *sativa* whereas the remaining belong to other species in *Oryza*, including 8 to *O. glaberrima,* 2 each to *O. nivara* and *rufipogon*, and 1 each to *O. glumaepatula* and *latifolia* (Agrama et al. 2009). Pure seed of these accessions were provided by the Genetic Stock *Oryza* Collection (GSOR) (www.ars.usda.gov/spa/dbnrrc/gsor). In this study, 217 accessions were used for genetic structure and diversity analyses, but only 203 *O. sativa* accessions were used for association mapping analyses because the wild relatives, *O. glaberrima*, *nivara, rufipogon, glumaepatula* and *latifolia*, contain many rare alleles, and rare alleles are one of the factors that increase the risk of type I errors or spurious associations (Breseghello and Sorrells 2006).

2.2. Location and field experiment

Evaluations were conducted for 14 traits in two field locations, USDA-ARS Dale Bumpers National Rice Research Center near Stuttgart, Arkansas and USDA-ARS Rice Research Unit near Beaumont, Texas during the growing season of 2009. The Stuttgart test site is located at N $34^0 27'44''$ and W $91^0 24'59''$, representing a temperate climate with a 243 d frost free period and average temperature of 23.9 C during the growing season. The Beaumont test site is located at N $30^0 03'47''$ and W $94^0 17'45''$, representing a subtropical climate with a 253 d frost free period and an average temperature of 26.1 C during the growing season. The experiments at both locations utilized a randomized complete block design having three replications with nine plants spaced 0.3×0.6 m in each plot. Li et al. (2012) had a detail description of experimental methods and field managements.

2.3. Phenotyping

Data collection followed procedures described by Yan et al. (2005a; 2005b) with modifications. Fourteen characteristics were recorded using the methods described by Li et al. (2010; 2011; 2012), including heading days, plant height, plant weight, tiller per plant, grain yield per plant, harvest index, main panicle length, panicle branches, Grain per panicle, seed set percentage, 1000 grain weight, grains per cm panicle, grains per branch panicle, and grain weight per panicle.

2.4. Genotyping

Bulk tissue from five plants was collected from each accession as described by Brondani et al. (2006) and total genomic DNA was extracted using a rapid alkali extraction procedure (Xin et al. 2003) and a CTAB method as described in Hulbert and Bennetzen (1991). The bulked DNA allowed identification of the origin of heterogeneity, which can result from the presence of heterozygous individuals or from a mix of individuals with different homozygous alleles (Borba et al. 2005). A total of 155 molecular markers covering the entire rice genome, with approximately one marker per 10 cM on average, were used to genotype the URMC accessions.

Among the markers, 149 SSRs were obtained from the Gramene database (http://www.gramene.org/), and five SSRs (AP5652-1, AP5652-2, AL606682-1, con673 and LJSSR1) were identified by Li et al(2011). The remaining marker was an *indel* at the *Rc* locus, named *Rid 12* and is responsible for rice pericarp color (Sweeny et al. 2006). Polymerase chain reaction (PCR) marker amplifications were performed as described in Agrama et al. (2009).

2.5. Statistical analysis marker and phenotype profile

Genetic distance was calculated from 155 molecular markers using Nei's method (Nei and Takezaki 1983). Phylogenetic reconstruction was based on the UPGMA method implemented in *PowerMarker* version 3.25 (Liu and Muse 2005). *PowerMarker* was also used to calculate the average number of alleles, gene diversity, and polymorphism information content (PIC) values. The tree to visualize the phylogenetic distribution of accessions and ancestry groups was constructed using MEGA version 4 (Tamura et al. 2007).

Each of the 14 phenotypic traits was modeled independently with the MIXED procedure in SASv.9.2, where genotype, location and interaction of location with genotype were defined as fixed effects while replication within a location (block effect) was a random effect. Broad-sense heritability was calculated using formula $H^2 = \sigma_g{}^2/(\sigma_g{}^2 + \sigma_e{}^2/n)$, where $\sigma_g{}^2$ as the genotypic variance, $\sigma_e{}^2$ as the environmental variance and n as the number of replications (Wang et al. 2007). Spearman rank correlation coefficients between each pair of the 14 traits were calculated using the mean of 9 plants, 3 in each of three replications for an accession, using the CORR procedure in SASv.9.2. Correlation coefficients were graphically displayed based on unweighted pair-group method using arithmetic average (UPGMA) by *NTSYSpc* software version 2.11V (Rohlf 2000).

2.6. Population structure

The model-based program *STRUCTURE* (Prichard et al. 2000) was used to infer population structure using a burn-in of 100,000, a run length of 100,000, and a model allowing for admixture and correlated allele frequencies. The number of groups (K) was set from 1 to 10, with ten independent runs each. The most probable structure number of (K) was calculated based on Evanno et al. (2005) using an *ad hoc* statistic *D(K)*, assisted with *L(K)*, *L'(K)* and *(L''K)*. The *D(K)* perceives the rate of change in log probability of the data between successive (K) values rather than just the log probability of the data. Determination of mixed ancestry (an accession unable to be clearly assigned to only one group) was based on 60% (Q) as a threshold to consider an individual with its inferred ancestry from one single group. Principal component analysis (PCA), that summarizes the major patterns of variation in a multi-locus data set, was performed with *NTSYSpc* software version 2.11 (Rohlf 2000). Two principal coordinates were used to visualize the dispersion of the mini core accessions graphically. *Fst* indicative of ancestral relationship between genetic groups was calculated using an AMOVA approach in Arlequin V2.000 (Weir 1996; Schneider and Excoffier 1999). The number of private alleles was estimated by Genetic Data Analysis (GDA) program (Lewis and Zaykin 2001).

Fourteen phenotypic characteristics were used to calculate Mahalanobis distance as a measurement of genetic differentiation among the groups (Kouame and Quesenberry 1993). The Mahalanobis distance and Canonical discriminant analysis were performed by the procedures PROC CANDISC of the *SAS* version 9.1 statistical packages. Eventually, the correlation of genetic structure differentiation resulting from the genotypic markers with phenotypic traits was assessed using the Mantel test (Mantel 1967) performed by PowerMarker.

2.7. Model comparisons and association analysis

The flexible mixed model (Yu et al. 2006) was used to control population structure. The methods for model comparisons and association mapping are referred to Li et al. (2012) for harvest index, Li et al. (2011) for grain yield, Jia et al. (2012) for sheath blight resistance and Bryant et al. (2011) for silica concentration in rice hulls.

3. Analysis of genetic structure and genetic diversity

3.1. Profile of DNA markers

Among 217 accessions in the URMC, the average number of alleles per locus was 13.5 ranging from 2 for RM338 to 57 for con673. PIC mean was 0.71 ranging from 0.30 for AP5625-1 to 0.97 for con673 among these markers. Since every accession was analyzed as a bulk of five plants, 54 (42.19%) loci showed heterozygosity and 38 (17.51%) accessions showed heterogeneity for at least one locus. Nei genetic distance (Nei and Takezaki 1983) was estimated for each pair of the 217 rice accessions which ranged from 0.021 to 1.000, with an average 0.752.

In previous studies, the average number of alleles per locus was 5.1 in Cho et al. (2000), 7.8 in Jain et al. (2004), 11.9 in Xu et al. (2004) and 11.8 in Garris et al. (2005). Recently, 13 alleles per locus were reported in the rice population studied by Thomson et al. (2007), 5.5 by Thomson et al. (2009) and 12.4 by Borba et al. (2009). The PIC in the URMC was 0.71, larger than it in the population studied by Cho et al. (2000) (0.56 PIC), Jain et al. (2004) (0.60), Xu et al. (2004) (0.66), Garris et al. (2005) (0.67), Thomson et al. (2007) (0.66), and Thomson et al. (2009) (0.45). The PIC was slightly less in our study than in the population studied by Borba et al. (2009) (0.75). Both the average allele number and PIC values are indicative of genetic diversity or gene richness in a germplasm collection. The higher the genetic diversity is in a collection, the greater the probability is for a gene of interest to be mined from the collection. Greater genetic diversity in the URMC is due to its global originations, multiple *Oryza* species, and the way of sampling with PowerCore software (Kim et al. 2007) based on 26 phenotypic traits and 70 SSR markers in order to capture the most diversity in the core collection (Agrama et al. 2009). Most other rice collections are either for a single country (Thomson et al. 2007), or for certain groups (Jain et al. 2004) and regions in a country (Thomson et al. 2009), or for special interests (Xu et al. 2004; Garris et al. 2005).

### 3.2. Genetic structure and differentiation derived from DNA markers

UPGMA tree showed that the accessions of *Oryza sativa* were classified to two main branch-es equivalent to *lowland* and *upland* cultivars, respectively. Ecogeographically, *indica* is pri-marily known as *lowland* rice and is grown throughout tropical Asia, while *japonica* often referred to as *upland* rice is typically found in temperate East Asia, upland geographic re-gions of Southeast Asia, and high elevations of South Asia (Garris et al. 2005). The *lowland* branch was further distinguished into two minor groups corresponding to AUS and IND ac-cessions, while the *upland* branched into three groups, TEJ, TRJ and ARO (Fig. 1a). Wild germplasm cluster separates from the two main branches. Eight accessions of *O. glaberrima* stayed together, distinguishable from *O. latifolia*, and *glumaepatula* on one side, and *O.nivara* and *O.sativa* (PI 430909) grouped together on the other side of the tree. Although PI 430909 from Pakistan was classified *O. sativa* in the Germplasm Resources Information Network (GRIN) at www.ars-grin.gov, it exhibited shattering, had a spreading plant type, black hulls with fulllong awns, and small red kernels; all of which are typical characteristics of wild rice. Surprisingly, PI 590422 from Myanmar in 1995 and PI 346371 from Brazil in 1969 were classified as *O. rufipogon* in GRIN, but the former was clustered with *indica* (Q-indica = 0.77) and the latter with an admixture of *aus* and *indica* (Q-aus = 0.59, Q-indica = 0.41). The disa-greement of cluster analysis in the study with traditional classification in GRIN is worthy of further attention.
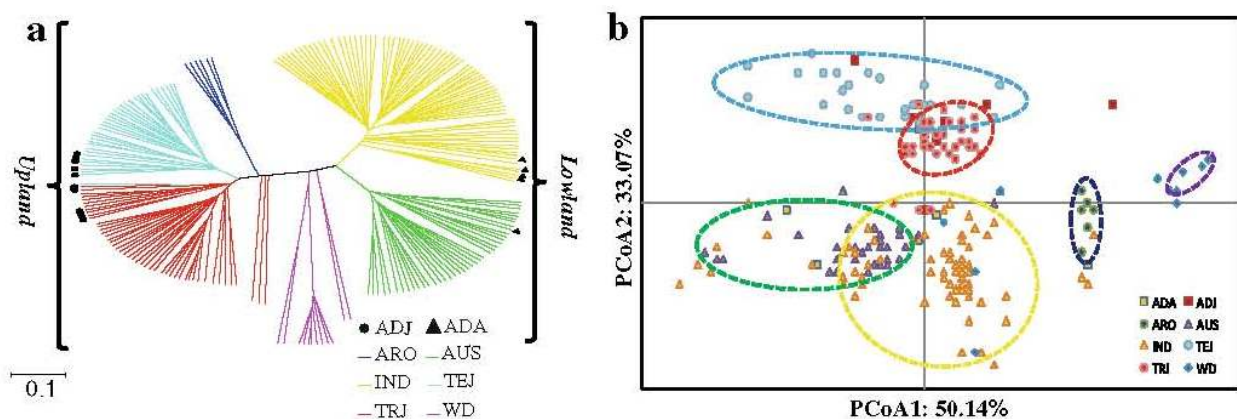
Figure 1. UPGMA tree (a) and principal coordinate analysis for 217 accessions in the USDA rice mini-core collection, both visualizing six main groups, AUS *aus*, IND *indica*, ARO *aromatic*, TEJ *temperate japonica*, TRJ *tropical japonica*, WD *wild rice*, ADA *admix of AUS and IND*, and ADJ *admix of TEJ and TRJ* (Li et al. 2010).

The ancestry of each accession was inferred from the Q value and classified into one of the six groups which corresponded to *aromatic* (ARO), *aus* (AUS), *indica* (IND), *temperate japonica* (TEJ), *tropical japonica* (TRJ) and *wild rice* (WD) based on reference cultivars reported previously by Garris et al. (2005), Agrama and Eizenga (2008), Agrama and Yan (2009). The classification was clear for a single group when the Q value was greater than 60%, otherwise an accession of germplasm was considered admixture with another group. In total, 21 accessions (9.68%) in the URMC had admixed ancestry either between TEJ and TRJ (ADJ) or between AUS and IND (ADA) (Fig. 1a, b).

The first-two axes in PCoA with 83.2% of total variation sufficiently discriminated the six main groups and two admixture groups (Fig. 1b). Each main group was distinguishable from another, but overlaps existed either among *temperate* and *tropic japonica* and their admixtures, or among *indica, aus* and their admixtures. The PCoA visualization and UPGMA tree were in agreement, which demonstrates a correct division of genetic structure in the URMC.
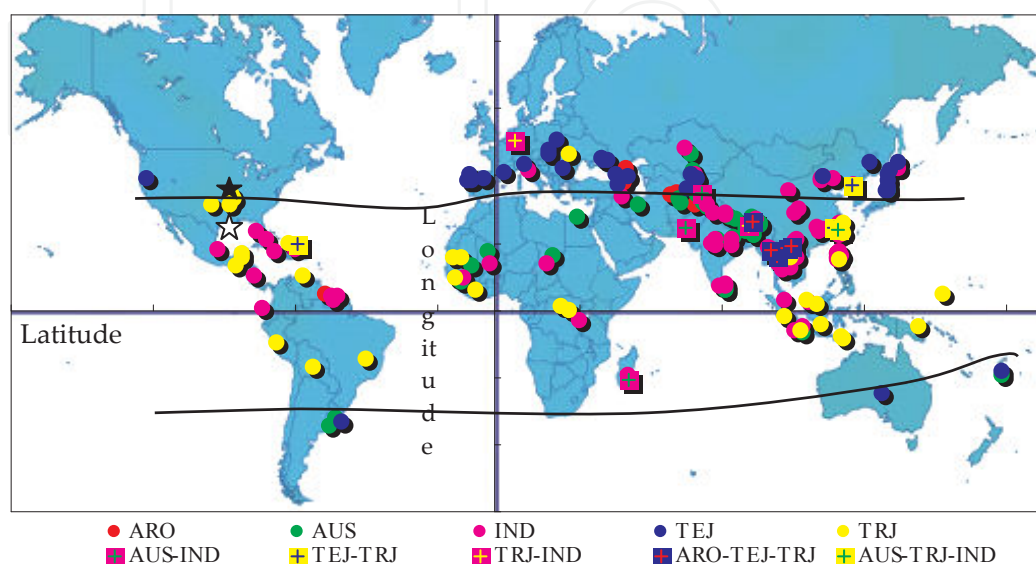

Figure 2. Geographic distribution of each mini-core accession with ancestral classification based on the latitude and longitude of germplasm origination, location of phenotypic evaluation ★: Stuttgart AR, ☆: Beaumont TX (Li et al. 2012).

Each accession with ancestry information was plotted on a world map using its latitude and longitude of geographic origin (Fig. 2). TEJ accessions were mainly distributed between latitudes 30 and 50 degrees north and south of the equator (i.e. temperate zone) while the other four groups scattered between latitude N 30 and S 30 degrees (i.e. tropical and subtropical zone).

In the URMC, the majority of accessions were IND (33%), followed by TRJ and AUS (18% each), TEJ (15%), WD (6%) and ARO with only six accessions (Fig. 3). All the marker loci were polymorphic in IND (Table 1). TRJ had 99% polymorphic loci, followed by WD, AUS, TEJ and ARO. IND had the most alleles per locus, TRJ and AUS the second most, TEJ and WD the third most and ARO had the fewest alleles. The largest number of private alleles per locus (alleles unique in one group and not found in another group) were found in WD (41.89%), followed by IND (23.78%) and AUS (17.66%). TRJ and TEJ had about equal private alleles, and the least was found in ARO. Gene diversity averaged 0.47 among the groups ranging from 0.37 in ARO to 0.52 in both IND and AUS. TRJ and WD had the same diversity (0.50), slightly greater than TEJ (0.43).

Results from the AMOVA showed that 37.92% of total variation was due to differences among groups, 61.21% within groups and 0.88% within individuals. Pair-wise estimates of *Fst* using

the AMOVA approach indicated a high degree of differentiation among the six main model-based groups (Fig. 3a). The mean *Fst* of all group pairs was 0.39 ranging from 0.24 between TRJ and TEJ to 0.48 between ARO and WD (Table 2). All pair-wise *Fst* values for the six groups were significant. The greatest genetic distance (0.990) among the 217 mini-core accessions was observed for PI 590413, an *O. glumaepatula* accession from the WD with 22 IND accessions and three accessions admixed AUS and IND, followed by the distance of 0.981 for PI 590413 with 9 AUS and 28 IND accessions, and the distance of 0.981 for PI 269727, an *O. latifolia* accession from the WD with 4 TEJ accessions. Two IND accessions, PI 202864 and PI 214077 had the shortest distance (0.021).
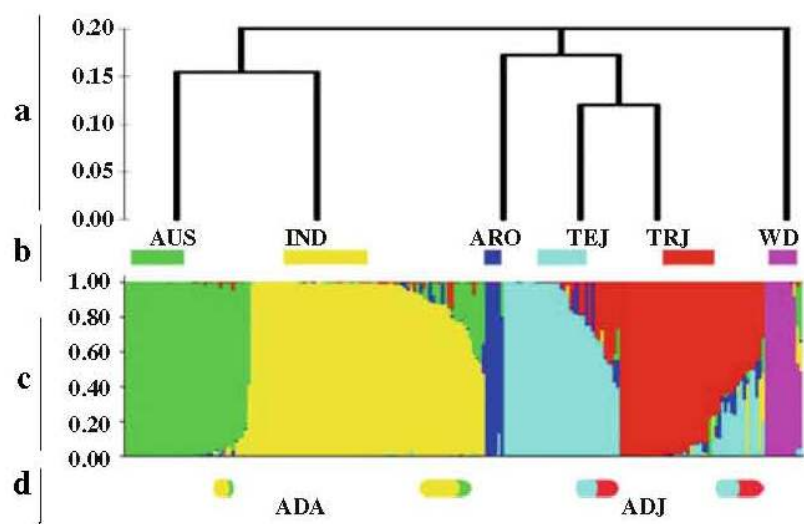
Figure 3. Dendrogram of genetic differentiation based AMOVA (Fst) (a) using DNA markers for six main groups (b), estimated group structure with each individual represented by *a horizontal bar* (c) and two admixture groups (d) (Li et al. 2010).

Group	Accessions*	Polymorphic loci %	Total alleles	Alleles/ locus	Private alleles	Private/total alleles %	Gene diversity
ARO	6	76	297	2.32	12	4.04	0.37
AUS	39	95	708	5.53	125	17.66	0.52
IND	71	100	900	7.03	214	23.78	0.52
TEJ	32	88	570	4.45	80	14.04	0.43
TRJ	40	99	758	5.92	103	13.59	0.50
WD	12	98	530	4.14	222	41.89	0.50

ARO, *aromatic*; AUS: *aus*; IND, *indica*; TEJ, *temperate japonica*; TRJ, *Tropic japonica*; WD, *wild rice*.

* Excluding admixed accessions.

Table 1. Analysis of genetic diversity among structural groups for 217 accessions in the USDA rice mini-core collection genotyped with DNA markers (Li et al. 2010).

3.3. Phenotypic analysis

Statistical analysis using a mixed model demonstrated that the differences due to genotypes and genotype × location interactions were highly significant at the 0.001 level of probability for all of the 14 traits (Table 2). The differences due to location were also significant for all traits except for panicle branches and seed set. Heritability was very high for all 14 traits. Heading had the highest heritability which was close to 100%. Although seed set had the lowest heritability, it was still above 70%. Heritability ranged from 77 to 97% among the other 12 traits. Harvest index had a heritability of 83% at Stuttgart and 90% at Beaumont. Correlation coefficients for each pair of the 14 traits were calculated using Spearman rank in each location for visualizing their relationships using PCA where the first two axes accounted for more than 50% phenotypic variation (Fig. 4a, b). At Stuttgart, 47 out of 91 correlations among the 14 traits were significant (<0.0001) (Fig. 4a), and 40 correlations were significant at Beaumont (Fig. 4b). Thirty four correlations were uniformly significant across two locations and their correlation directions (positive or negative) were also same across two locations.
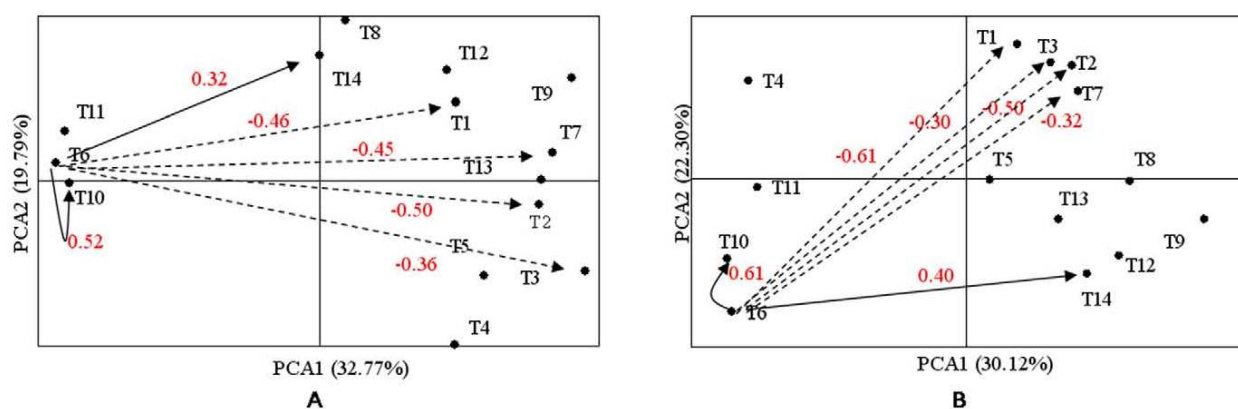
Figure 4. Relationship map constructed by PCA for 14 traits at Stuttgart, Arkansas (A) and Beaumont, TX (B). The distance between traits is inversely proportional to the sizes of the correlation coefficients. *Solid* and *dashed lines* indicate positive and negative correlations, respectively. Trait number corresponds to Table 2, and red numbers show the significant correlation coefficient with Trait 6: Harvest Index (Li et al. 2012).

3.4. Genetic structure and differentiation derived from phenotypic traits

Canonical discriminant analysis of 14 phenotypic traits for the mini-core accessions clearly separated the six plus two admixture model-based genetic groups derived from molecular data (Fig. 5). The first four significant ($P < 0.001$) canonical discriminant functions (CAN) explained 92.02% of the total variance, 54.87 % by the first CAN and 18.08% by the second CAN function, respectively. The accessions in group of AUS, ARO, IND, TEJ, TRJ and WD were clustered into their groups with various overlaps. The *upland* (ARO, TEJ and TRJ) were obviously discriminated from the *lowland* (AUS, IND), and the admixed groups ADA was scattered across AUS and IND while ADJ across TEJ and TRJ.

All 14 traits were significantly different among the eight (six plus two admixtures) model-based genetic groups. However, only three traits, plant weight (biomass), tillers and grain yield

| Trait	Location	Mean ± SD	Range	Heritability (%)	Genotype		Location (LOC)		Genotype*LOC	
					F value	Pr>F	F value	Pr>F	F value	Pr>F
1 Heading	Arkansas	99.33±21.31	34.00~181.67	98.08	341.53	0.0000	2634.77	0.0000	12.45	0.0000
(days)	Texas	87.55±22.63	38.00~182.00	98.64						
2 Plant height	Arkansas	109.73±20.20	61.08~153.92	97.11	127.48	0.0000	1676.50	0.0000	45.63	0.0000
(cm)	Texas	124.74±22.45	67.00~178.78	95.73						
3 Plant weight	Arkansas	168.71±79.88	27.83~548.42	86.33	30.87	0.0000	122.48	0.0004	9.94	0.0000
(g/plant)	Texas	219.02±87.70	35.93~558.02	86.83						
4 Tillers/plant	Arkansas	23.95±11.20	6.42~67.75	86.53	35.27	0.0000	818.76	0.0000	7.10	0.0000
	Texas	41.13±15.83	13.00~85.89	87.16						
5 Grain yield	Arkansas	60.02±25.51	8.54~127.27	87.05	29.06	0.0000	98.37	0.0006	8.33	0.0000
(g/plant)	Texas	76.67±30.05	5.64~165.97	84.03						
6 Harvest index	Arkansas	30.44±7.02	3.40~45.06	82.75	35.79	0.0000	2174.76	0.0000	6.10	0.0000
(%)	Texas	38.98±10.51	6.25~60.02	89.98						
7 Panicle length (main)	Arkansas	26.66±3.81	14.21~37.19	89.86	46.56	0.0000	293.26	0.0001	3.68	0.0000
(cm)	Texas	24.75±3.44	16.84~38.40	88.34						
8 Panicle branches (main)	Arkansas	10.97±2.15	5.44~17.78	85.65	29.97	0.0000	31.18	0.0046	2.40	0.0000
	Texas	10.64±2.06	5.56~16.33	81.68						
9 Kernels /panicle (main)	Arkansas	194.97±57.49	68.56~399.00	86.48	29.94	0.0000	367.90	0.0000	4.45	0.0000
	Texas	155.77±45.46	50.00~318.33	86.92						
10 Seed set (main)	Arkansas	78.15±15.23	25.48~96.97	78.39	15.39	0.0000	14.26	0.0191	4.39	0.0000
(%)	Texas	73.55±12.65	35.07~95.29	72.66						
11 1000 Seed weight	Arkansas	25.77±5.07	11.17~44.74	91.79	69.00	0.0000	75.18	0.0005	3.94	0.0000
(g)	Texas	24.41±4.66	12.32~43.86	95.52						
12 Kernels/cm panicle	Arkansas	7.30±1.80	3.25~14.61	84.71	28.72	0.0000	218.17	0.0001	3.60	0.0000
	Texas	6.31±1.63	2.80~12.27	87.02						
13 Kernels/branch panicle	Arkansas	17.88±4.24	11.56~37.10	82.66	19.90	0.0000	353.27	0.0001	4.31	0.0000
	Texas	14.67±2.98	9.61~23.23	77.42						
14 Grain weight/panicle	Arkansas	3.79±1.18	0.68~8.62	82.29	21.86	0.0000	241.69	0.0001	3.94	0.0000
(g)	Texas	2.75±0.95	0.63~6.27	80.72						

Table 2. Statistical analysis of 14 traits generated at Stuttgart, Arkansas and Beaumont, Texas in 2009 among 203 *O. sativa* accessions in the USDA rice mini-core collection (Li et al. 2011).

per plant, had larger variation among groups than within groups. Therefore, they are considered the main discriminatory characters ($r^2 \geq 0.49$) in differentiating these genetic groups. The first canonical loading was 0.81 for grain yield and tillers and 0.78 for plant weight. The second canonical loading was dominated by panicle length (0.59), heading days (0.55) and seed weight (0.51).

The most tillers were observed in AUS accessions PI 385697 (93) and 352687 (86), while the lowest were in TRJ PI 584567 (9) and PI 154464 (10). WD had the most tillers (60), followed by AUS (46), ADA (44), AUS (46), IND (38), ARO (27), TEJ (24), ADJ (21) and TRJ (18). The greatest plant weight was 731 g for PI 549215 (IND), and the lowest was 37g for PI 281630 (TEJ). Again, WD had the most plant weight (442g) and TEJ had the lowest (127g). PI 373335 (IND) had the highest grain yield per plant at 175g and PI 389933 (IND) had the lowest at 11g. ADA had the most grain yield per plant (127 g per plant), while TEJ had the lowest (55g).

3.5. Relationship between genetic and phenotypic differentiation

Both the dendrograms based on the Mahalanobis distance (D^2) using the 14 phenotypic traits (Fig. 5) and based on the *Fst* genetic differentiation from AMOVA using DNA markers (Fig. 3a) produced similar results. The two dendrograms differentiated the *lowland* including IND, AUS and their admixtures from the *upland* having TEJ, TRJ, ARO and their admixtures. The WD or non-*sativa* accessions remained independent from the others.
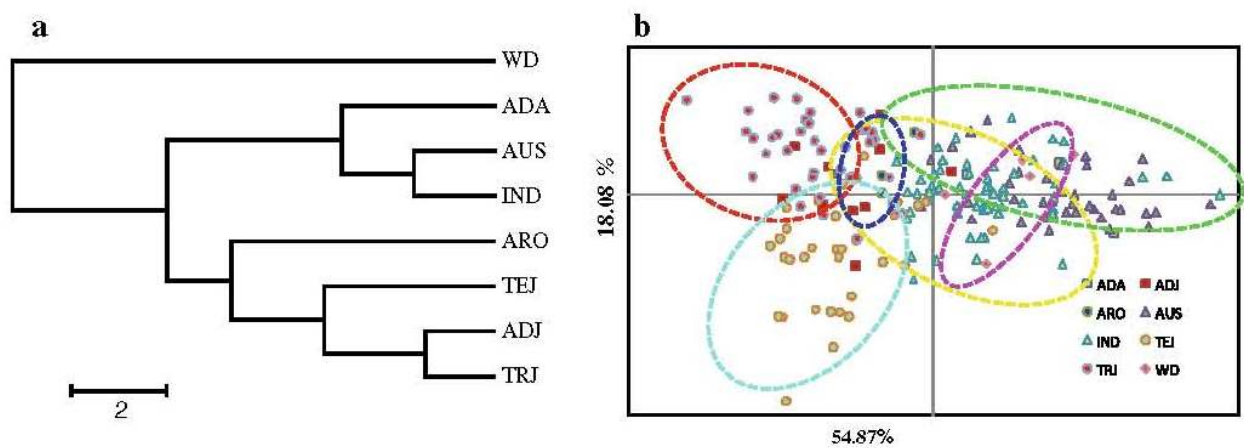
Figure 5. Dendrogram of differentiation based on Mahalanobis distance (left) and Canonical discriminant analysis (CDA) (right) using 14 phenotypic traits among structural groups for 217 accessions in the USDA rice mini-core collection (Li et al. 2010).

Group	ARO	AUS	IND	TEJ	TRJ	WD
ARO	-	0.37	0.41	0.38	0.31	0.48
AUS	12.90	-	0.31	0.44	0.40	0.38
IND	10.96	3.36	-	0.45	0.40	0.38
TEJ	13.08	15.92	9.59	-	0.24	0.46
TRJ	9.37	16.08	9.36	8.03	-	0.41
WD	21.47	14.90	17.10	22.70	22.57	-

All AMOVA-based Fst estimates from 110 permutations were significant (P<0.001), and all Mahalanobis distance (D^2) estimates were significant (P<0.001).

Table 3. Pairwise comparison of Fst values above the diagonal based on DNA markers and Mahalanobis distance (D^2) below the diagonal based on 14 phenotypic traits among structural groups for 217 accessions in the USDA rice mini-core collection (Li et al. 2010).

Analysis developed by Mantel (1967) is widely used to describe the genetic relationship between genotypic and phenotypic measurements (Gaudeul et al. 2000, Gizaw et al. 2007). In our study, genetic distance derived from the DNA markers among the six plus two admixture model-based groups was highly and significantly correlated with the distance derived from 14 phenotypic traits (r = 0.85, P =0.000<0.001). This explains the correspondence of the two dendrograms in Fig. 3a and Fig. 5(left), and similar pattern of D^2 and Fst in Table 3.

In rice ancestry, structure and genetic diversity of germplasm collections has been studied using a variety of molecular markers such as SNP (Zhao et al. 2011), SSR (Cho et al. 2000; Jain et al. 2004; Xu et al. 2004; Garris et al. 2005; Thomson et al. 2007; 2009; Borba et al. 2009), RAPD (Mackill 1995) and isozyme (Glaszmann 1987) markers. Phenotypic characteristics are rarely used to analyze genetic diversity or structure in rice germplasm collections. Zeng et al. (2003)

collected samples from each of six genetic groups for a diversity analysis using 31 phenotypic traits, but failed to reveal their genetic differentiations.

However, assessment of genetic diversity and structure using both genotypic and phenotypic characterization and relationship or accuracy between the genotypic and phenotypic assessments has long been attractive to the scientific community. Elias et al. (2001a) reported a significant positive association between genotypic and phenotypic distances ($r = 0.204$, $p = 0.054$) using eight SSRs and 14 traits for 38 accessions of cultivated cassava (*Manihot esculenta* Crantz). The association was improved ($r = 0.283$, $p < 0.01$) in a set of 29 cassava accessions genotyped with AFLP markers and phenotyped for 14 morphological and four agronomic traits (Elias et al. 2001b). A set of 68 sweet sorghum and four grain sorghum (*Sorghum bicolor* L.) accessions were genotyped with 41 SSRs and phenotyped for six traits (Ali et al. 2008). The genotypic analysis classified the 72 accessions in 10 clusters and the phenotypic variation among the clusters was described. Similarly, 15 morpho-physiological traits were used to describe four major groups of 61 tomato (*Solanum lycopersicum* L.) accessions classified by genotypic data of 29 SSRs (Mazzucato et al. 2008). In barley (*Hordeum vulgare* L.), based on five cultivars phenotyped for 18 traits and genotyped with 11 AFLP markers, trait relationships were demonstrated using simple correlation, path analysis and GGE biplot. The cultivars were clustered based on genetic dissimilarity estimated by the AFLP markers (Akash and Kang 2009).

We use both genotypic and phenotypic characterizations to analyze genetic differentiation in a plant germplasm collection. The present study in rice has a much greater association ($r = 0.85$, $P = 0.000 < 0.001$) of genetic distance derived from genotypic characterizations with phenotypic characterizations than the previous study in cassava.

4. Association mapping of harvest index and components

Harvest index is a ratio of grain yield to total biomass, which measures farming success in partitioning assimilated photosynthate to harvestable product (Hay 1995; Sinclair 1998). In cereal crops, dramatic improvements of harvest index during domestication have made commercial cultivars dramatically different from their wild ancestors (Gepts 2004). Rice (*Oryza sativa* L.) is one of the most important staple foods (Tyagi et al. 2004), and can be highly productive if high harvest index genotypes are grown with optimal management practices (Raes et al. 2009). Harvest index is one of the most complex traits in rice involving number of panicles per unit area, number of spikelets per panicle, percentage of fully ripened grains, kernel size (Terao et al. 2010) and plant height (Marri et al. 2005). Marri et al. (2005) found that harvest index was negatively correlated with plant height, but positively correlated with grain number per panicle, tiller number per plant, seed set, kernel size and grain yield per plant in rice. Similarly in maize, harvest index is negatively correlated with plant height, and positively correlated with grain yield (Can and Yoshida 1999). In sorghum, harvest index is negatively correlated with forage yield (Mohammad et al. 1993), but positively correlated with growth rate and grain filling rate (Soltani et al. 2001). The correlated traits are interrelated in most

cases, so that increases in one component may lead to either decreases or increases in others. Therefore, scientists aim to identify genes or QTL that increase one aspect of a target trait without affecting others, or improve the target trait indirectly through an improvement of its related traits.

In rice, previous studies on harvest index have identified numerous QTL all using a classic linkage-mapping strategy with two parents. Mao et al. (2003) reported four main QTL on chromosome (Chr) 1, 4, 8 and 11 and an epistatic interaction between two QTL respectively on Chr 1 and Chr 5. Sabouri et al. (1999) identified three QTL each on Chr 2, 3 and 5, and two QTL close to each other on Chr 4. Lanceras et al. (2004) described harvest index QTL on Chr 1 and 3. However, mapping populations developed from different parental combinations and/ or experiments conducted in different environments often result in partly or wholly non-overlapping sets of QTL (Hao et al. 2010).

4.1. Traits correlated with harvest index in our study

Six traits were significantly correlated with harvest index and these correlation directions were the same across the two locations. The correlations with harvest index were negative for heading (-0.46 at Stuttgart and -0.61 at Beaumont), plant height (-0.50 and -0.50), plant weight (-0.36 and -0.30), panicle length (-0.45 and -0.32), while positive for seed set (0.52 and 0.61) and grain weight/panicle (0.32 and 0.40) (Fig. 4a, b). In the PCA based on phenotypic traits of 203 mini-core accessions, four traits negatively correlated with harvest index were plotted on opposing axis from harvest index (Fig. 4a, b). Conversely, two traits positively correlated with harvest index were plotted in the same axis relatively close to harvest index.

4.2. Marker-trait associations

At Stuttgart, a total of 36 markers were significantly associated with harvest index traits at the 6.45×10^{-3} level of probability (the Bonferroni corrected significance level). Among 36 markers, seven were associated with harvest index, five with heading, three with plant height, six with plant weight, five with panicle length, nine with seed set and one with grain weight/panicle. Eight trait-marker associations have been reported by previous linkage mappings. Additionally, seven markers were associated with two or more harvest index traits, named "consistent" markers (Pinto et al. 2010). Out of the seven consistent markers, RM600, RM5 and RM302 were co-associated with harvest index and seed set, RM431 with heading and seed set, RM341 with plant height and panicle length, RM471 with heading and plant weight, and RM510 with three traits, plant height, harvest index and seed set.

At Beaumont, we identified 28 markers significantly associated with harvest index's traits. Among 28 markers, two were associated with harvest index, three with heading, nine with plant height, six with plant weight, four with panicle length, three with seed set and one with grain weight/panicle. Similarly with Stuttgart, 11 trait-marker associations have been identified in previous QTL studies. Two consistent markers were RM208 co-associated with harvest index and seed set, and RM55 co-associated with plant height and plant weight.

Associations of RM431 with plant height, Rid12 and RM471 with plant weight and RM24011 with panicle length were found in both locations. The four markers that associated with the same trait across both locations are called "constitutive QTL" markers, while others that associated with a certain trait only at one location are called "adaptive QTL" markers (Mao et al. 2003).
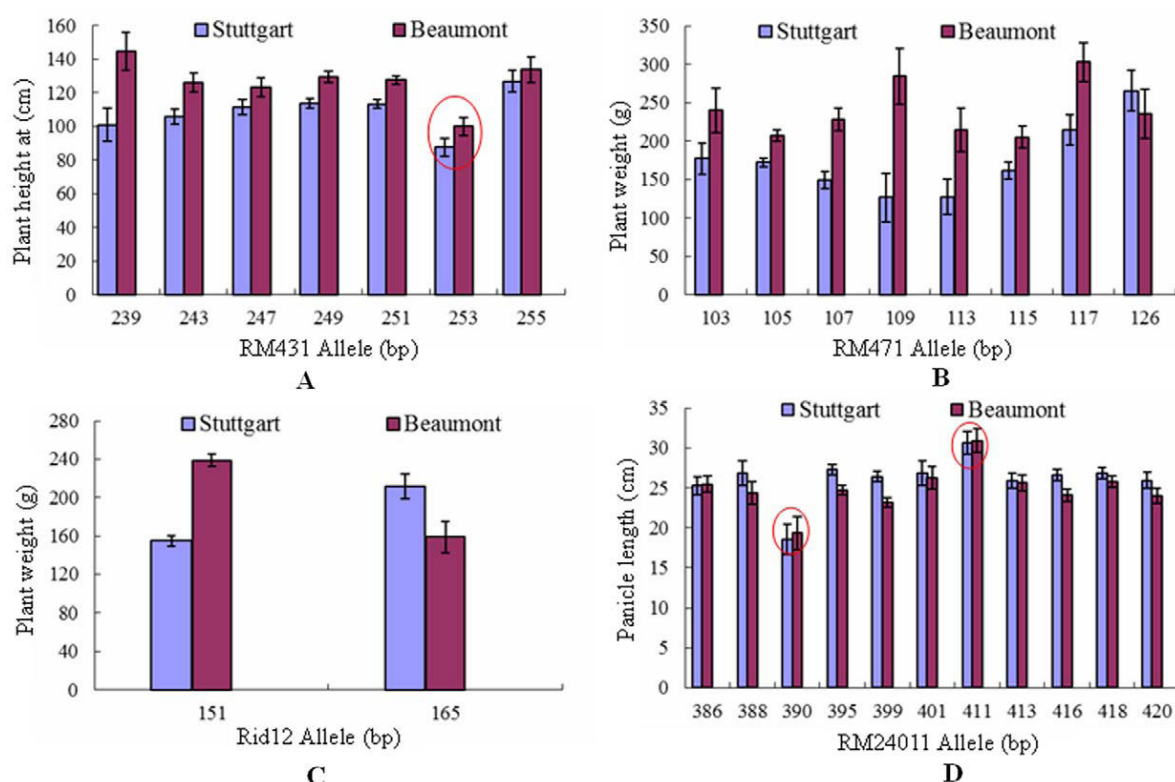
4.3. Allelic effects

The allelic effects of the constitutive markers associated with their traits were estimated with the least square mean (LSMEAN) of phenotypic value and presented in Fig. 6. Meanwhile, an algorithm was employed to generate a letter-based representation of all-pairwise comparisons for allelic effect. For RM431, allele 253bp had a significantly larger effect than all other 6 alleles at Beaumont and than 4 others at Stuttgart to reduce plant height. For RM24011, allele 390bp had the greatest effect on decreasing panicle length while allele 411bp had the largest effect on increasing panicle length at both locations. However, for Rid12, the allelic effects were opposite between two locations. Allele 151bp of Rid12 had a decreasing effect on plant weight at Stuttgart, but an increasing effect at Beaumont instead. The 165 allele of Rid12 had an opposite effect to 151bp on plant weight. For RM471, the allelic effects on plant weight were not consistent from one location to another. The 109bp allele had the largest effect on decreasing plant weight at Stuttgart, but a fairly larger effect on increasing plant weight at Beaumont.

4.4. Genetic dissection of harvest index

Harvest index is an integrative trait including the net effect of all physiological processes during the crop cycle and its phenotypic expression is generally affected by genes responsible for non-target traits, such as heading (Lanceras et al. 2004; Hemamalini et al. 2000), plant height (Lanceras et al. 2004) and panicle architecture (Ando et al. 2008). The magnitude and direction of these gene functions on different phenotypes would bear heavily on the utility of such genes for improvement of these traits. In the current study, the traits like heading, plant height, plant weight and panicle length had a strong negative correlation with harvest index, while seed set and grain weight/panicle were positively correlated with harvest index. These phenotypic correlations were consistently reflected in the identification of molecular markers associated with harvest index and related traits. For example, four consistent markers at Stuttgart, RM600, RM302, RM25, and RM431, were associated with not only harvest index itself, but also for one or more additional traits correlated with harvest index. Another consistent marker, Rid12, associated with both heading and plant weight was close to a reported QTL "qHID7-1" responsible for harvest index and the gene "Ghd7" which effects grains per panicle, plant height and heading in rice (Hittalmani et al. 2003). At Beaumont, the consistent marker RM55 associated with both plant height and plant weight was adjacent to a QTL "qHID3-2" for control of harvest index (Hemamalini et al. 2000). RM431 co-associated with plant height and harvest index in this study has been reported to be closely linked to gene "sd1" (Xue et al. 2008; Peng et al. 2009). sd1 is involved in gibberellic acid biosynthesis, decreases plant height and thus increases harvest index. The decreased height confered by sd1allows the plant to have a reduced risk of lodging,

be more tolerant to heavy doses of nitrogen fertilizer, and allows for planting increased stand densities. The *sd1* gene has greatlyimproved grain yield and has contributed to the Green Revolution in cereal crops including rice (Fu et al. 2010).

Other markers were associated with the traits correlated with harvest index, but not with harvest index directly in this study. These markers have been reported either nearby or flanking the QTL for harvest index. RM5, which was associated with plant height in the Stuttgart study, was close to a reported QTL for harvest index on Chr 1 (Marri et al. 2005). RM471 associated with plant weight was close to the reported *qHID4-1* and *qHID4-2* for harvest index (Hemamalini et al. 2000). Furthermore, RM257 and RM22559 associated with seed set were co-localized with a known QTL on Chr 9 (Marri et al. 2005), and with *qHID8-1* (Hemamalini et al. 2000) for harvest index, respectively. Similarly, at Beaumont, RM44 associated with plant height was close to *qHID8-1* (Hemamalini et al. 2000), and RM263 associated with heading was adjacent to *hi2.1* (Marri et al. 2005). The chromosomal regions where numerous correlated traits are mapped indicate either pleiotropy of a single gene or tight linkage of multiple genes. Fine-mapping of such chromosomal regions would help discern the actual genetic control of these congruent traits. Development of markers for such traits in specific regions could lead to a highly effective strategy of marker-assisted selection for improving harvest index.

Figure 6. Comparisons of allelic effects of four constitutive marker loci RM431 (a) associated with plant height, RM471 (b) and Rid12 (c) associated with plant weight, RM24011 (d) associated with panicle length constitutively at both Stuttgart, Arkansas and Beaumont, Texas (Li et al. 2012).

5. Association mapping of grain yield and components

Yield is one of the most important and complex traits in crops that does not evolve independently but shows correlations with other traits. Thus, breeders have to consider correlated traits in breeding programs. Yield and its related traits are quantitatively inherited and controlled by many genes with small effects subject to environmental effects (Inostroza et al. 2009; Shi et al. 2009). Many studies have focused on the improvement and inheritance of agronomically important yield-related traits for achieving greater yield (Gravois and McNew 1993; Samonte et al. 1998). Other traits such as biomass, plant architecture, adaptation, and resistance to biotic and abiotic constraints may also indirectly affect yield through yield components or other physical and physiological mechanisms. Hence, estimation of the positions and effects of quantitative trait loci (QTL) for traits related to yield is of central importance for marker-assisted selection for yield improvement. In rice genetics, most QTLs related to yield have been identified through classic linkage mapping approaches (Moncada et al. 2001; Brondani et al. 2002; Thomson et al. 2003; Jiang et al. 2004; Suh et al. 2005). With a few notable exceptions, most of these QTLs have not been successfully validated or consistently used in crop improvement (Bernardo 2008). The classic approaches are too simplistic to effectively model most of the genetic variation for complex traits because they are unable to reflect the genetic realities of these traits (Cooper et al. 2005; Holland 2007).

Trait	Locus	Chr. no.	Position (cM)	Prob F			Annotation
Grain yield	OSR13	3	36.0	0.0046	#		
(4 markers)	RM471	4	45.0	0.0004	#		
	RM247	12	26.7	0.0062			
	RM7003	12	41.0	0.0011	#	*	GY_QTL Marker (Thomson et al. 2003; Fu et al. 2010)
Plant height	RM5	1	98.5	0.0061		*	
(3 markers)	RM431	1	154.6	0.0001	#	*	GY_QTL Marker (Fu et al. 2010)
	RM509	5	59.0	0.0043	#		
Plant weight	RM279	2	14.0	0.0011	#		
(6 markers)	RM471	4	45.0	0.0008	#		
	RM527	6	59.0	0.0048			
	Rid12	7	41.0	0.0006	#	*	GY_QTL Marker (Xue et al. 2008)
	RM224	11	115.0	0.0033	#	*	
	RM7003	12	41.0	5.87×10^{-6}	#		GY_QTL Marker (Thomson et al. 2003; Fu et al. 2010)
Tillers	RM431	1	154.6	0.0033	#		GY_QTL Marker (Fu et al 2010)
(9 markers)	RM279	2	14.0	0.0014	#		

Trait	Locus	Chr. no.	Position (cM)	Prob F		Annotation
	RM341	2	70.0	0.0023	*	
	RM3558	4	69.8	0.0024		
	Rid12	7	41.0	0.0001	#	GY_QTL Marker (Xue et al. 2008)
	RM125	7	41.0	0.0012		GY_QTL Marker (Jiang et al. 2004; Fu et al. 2010; Borba et al. 2010)
	RM484	10	71.4	0.0005		
	RM287	11	68.0	0.0050		GY_QTL Marker (Moncada et al. 2001)
	RM224	11	115.0	0.0002	#	
Panicle length	RM509	5	59.0	0.0036	#	
(5 markers)	RM510	6	11.5	0.0003		
	RM24011	9	27.4	0.0033		
	RM245	9	91.8	0.0040		GY_QTL Marker (Suh et al. 2005)
	RM3739	12	97.0	0.0009		
Kernels/branch	OSR13	3	36.0	0.0016	#	
(3 markers)	RM471	4	45.0	0.0048	#	
□	RM1335	7	106.0	0.0005		

#: The marker was associated with two or more traits

*: The marker was reported to be associated with the corresponding trait previously

GY_QTL Marker: The markers were identified to be linked or close to yield QTL in previous studies

Table 4. Marker loci significantly associated with grain yield and its related traits mapped in the USDA rice mini-core collection (Li et al. 2011)

5.1. Traits correlated with grain yield per plant in our study

The traits significantly correlated with grain yield were plant height (0.43), plant weight (0.81), tillers (0.77), panicle length (0.30) and kernels/branch (0.40). All these traits were clustered into one branch except kernels/branch. This exploratory assessment showed that grain yield and the set of five correlated traits would serve as an appropriate base population for an association mapping application.

5.2. Marker-yield trait associations

Using the selected PCA model, a total of 30 marker loci were identified to have significant marker-trait associations at the 6.45×10^{-3} level of probability (the Bonferroni corrected significance level) for yield and its correlated traits (Table 4). Out of the 30 markers, four were associated with grain yield, three with plant height, six with plant weight, nine with tillers, five with panicle length and three with kernels/branch. Six markers were co-localized with

previous identified QTL (Thomson et al. 2003; Jiang et al. 2004; Xue et al. 2008; Fu et al. 2010; Borba et al. 2010; Moncada et al. 2001) (Table 4).

Most importantly, eight of the 30 markers were synchronously associated with two or more traits (Table 4). RM471 was co-associated with three traits, grain yield, plant weight and kernels/branch. Three markers Rid12, RM224 and RM279 were co-associated with plant weight and tillers. RM431 was co-associated with plant height and tillers; RM509 with plant height and panicle length; RM7003 with grain yield and plant weight; and OSR13 with grain yield and kernels/branch. Three markers, OSR13, RM471 and RM7003 were included for the allelic analysis because they were not only associated with grain yield directly, but also co-associated with other yield correlated traits (Fig. 7). The allelic effect of each loci associated with the traits was estimated with mean of phenotypic value for each allele. For marker locus RM471, allele 126bp had the highest effect on all three traits (93.48 for grain yield, 266.23 for plant weight and 25.36 for kernels/branch), while two other alleles 109bp and 113bp had the lowest effect on grain yield with 48.19 and 49.90, and plant weight with 17.54 and 19.82, respectively (Fig. 7a and b). For OSR13, allele 123bp had a large effect on both grain yield and kernels/branch with 66.37 and 19.91, respectively while allele 115 had the highest effect on kernels/branch and the lowest on grain yield (Fig. 7c). For RM7003, allele 108bp had the highest effect on both traits (66.37 for grain yield and of 228.05 for plant weight) while the allele 106bp had the lowest effect on both traits (43.19 for grain yield and with 154.48 for plant weight) (Fig. 7d).

5.3. Trait-trait and marker-trait associations

Correlation among phenotypic traits is a common phenomenon in biology. Plant breeders need to consider trait correlations for either improving numerous correlated traits simultaneously or reducing undesirable side effects when their goal is only one of the correlated traits (Chen and Lubberstedt 2010). In this study, 34 of 91 pairs (37.36%) of 14 traits were observed to have significant correlation, and five traits were correlated with grain yield among 203 mini-core accessions. The correlations exhibited a complex network among these traits. Numerous researchers have concluded that rice yield is highly dependent on the number of productive tillers or panicles (Sharma and Choubey 1985; Dhanraj and Jagadish 1987), which is recently verified with a high correlation between tillers and yield (r=0.88; p < 0.01) by Borba et al. (2010). Panicle characters including panicle length, number of primary branches, secondary branches per primary branch, total kernels and seed set in a panicle, are reported to be tightly related to yield performance (Thomson et al. 2003; Ando et al. 2008; Terao et al. 2010). Although seed set and kernel weight per panicle were not directly correlated with yield in this study, they may be correlated in other panels of germplasm or may be indirectly contributable to yield. For example, seed weight per panicle, seed set and 1000 kernel weight are identified to be highly correlated with yield in wild rice (Oryza *rufipogon* Griff.) (Fu et al. 2010). Similarly, seed weight per panicle and seed set have correlations with yield in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson (Thomson et al. 2003). These different results could be expected since different materials were used in those studies.
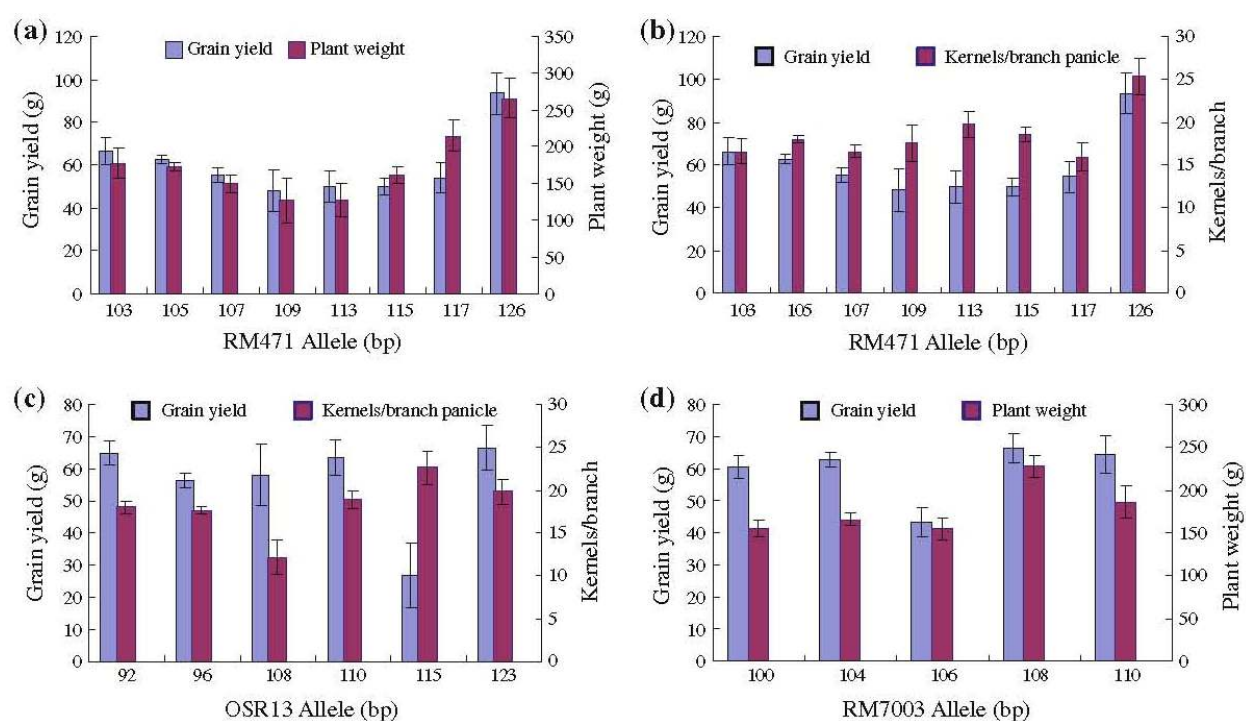
Figure 7. Comparisons of allelic effects of three marker loci on yield traits: RM471 (a, b), OSR13 (c) and RM7003 (d). RM471 co-associated with grain yield, plant weight and kernels/branch panicle; OSR13 co-associated with grain yield and kernels/branch; RM7003 co-associated with grain yield and plant weight; bars are the standard error (Li et al. 2011).

Morphological correlations could be explained by either pleiotropy or linkage disequilibrium. The former describes the impact of a single gene on multiple phenotypic traits. The latter deals with influence of two or more genes on multiple traits, where the genes are physically located so close to each other, that they cannot be practically separated (Chen and Lubberstedt 2010). Co-association of a single gene (or two linked genes) with multiple traits that are phenotypically correlated has occurred in numerous studies. Yan et al. (2009) reported five SSRs that were co-associated with two correlated traits affecting stigma exertion, another five SSRs with two traits correlated to spikelets, and one SSR with three correlated traits to spikelets in rice. Similarly, Terao et al. (2010) identified the gene of *APO1* that increases both the primary rachis branches and grains per panicle in rice. Gene *DEP1* increases both rachis branches and grain yield in rice (Huang et al. 2009). Gene *Ghd7* has major effects on grains per panicle, plant height and heading date in rice (Xue et al. 2008). Further, developmentally related traits (like number of tillers and roots) have been mapped to the same chromosome regions (Hemamalini et al. 2000; Brondani et al. 2002, Li et al. 2006; Thomson et al. 2003, Fu et al. 2010). In this study, eight markers were co-associated with two or more correlated traits and some QTLs related to yield and yield components have been reported to be near these regions. RM7003 co-associated with grain yield and plant weight is reported to flank a major yield QTL (*yld12.1*) (Thomson et al. 2003; Fu et al. 2010). Also, RM7003 is near the QTL *gpp12.1* which influences grains per panicle (Thomson et al. 2003), the QTL *pss12.1* which effects seed set (Fu et al. 2010) and another QTL *qFG12-2* which is involved with filled grain number (Li et al. 2002). Interestingly, five particular

markers were not associated with yield directly in this study, but they were all identified to be the markers flanking grain yield QTL in other studies. RM431, RM340, and RM245 were found to be associated with yield QTLs, *yld1.1* (Fu et al. 2010), *qYI-6-1* and *qYI-9* (Suh et al. 2005), respectively. Rid12 co-associated with tillers and plant weight was found to be very close to *Ghd7* that had major effects on grain yield, plant height and heading date (Xue et al. 2008) in addition to its function for rice pericarp color (Sweeney et al. 2006; Brooks et al. 2008). RM125 associated with tillers was also identified to have a strong association with yield (Borba et al. 2010, Jiang et al. 2004). RM431 co-associated with plant height and tillers in this study has been reported to be closely linked with a QTL "*sd1*" to decrease plant height and increase yield (Peng et al. 1999; Fu et al. 2010). The chromosomal regions where numerous traits are mapped indicate either pleiotropy resulting from a single gene or tight linkage of multiple genes. Fine-mapping of such chromosomal regions would help discern the actual genetic control of these congruent traits. Development of markers for such traits in these regions could lead to a highly effective strategy of marker-assisted selection.

Several genes for grain yield and its related traits have been recently cloned, and each of these genes has a clearly distinct biological function (Li et al. 2003; Ashikari et al. 2005; Fan et al. 2006; Song et al. 2007). Molecular cloning and functional analyses of several genes have shown that these genes are mostly related to the synthesis and regulation of the phytohormone gibberellin (Peng et al. 1999; Ashikari et al. 1999; Spielmeyer et al. 2002; Itoh et al. 2004). For example, a semidwarf QTL "*sd-1*" close to RM431 contains a defective gibberellin 20-oxidase gene responsible for height reduction. The shorter statured plants have a decreases lodging threat and tolerates higher dosags of nitrogen fertilization, thus dramatically increases grain yield. Furthermore, the photoperiod pathway controls flowering time or heading directly, thus affects plant weight and yield indirectly (Xue et al. 2008). Two other genes regulating heading have been identified. One is *Hd6* which encodes a subunit of protein kinase CK2 (Takahashi et al. 2001), and the other is *Ehd1* which encodes a B-type response regulator (Doi et al. 2004). Also, a gene *GHD7* has been identified to simultaneously control yield, plant height and heading in rice (Xue et al. 2008). This gene locates close to Rid12 and encodes a CCT (CO, Co-like and Timing of CAB1) domain protein. These findings demonstrate that genes regulating yield usually share some common pathways for traits that contribute to yield. Regions with either tightly linked QTLs or pleiotropic effects would become QTL hot spots, worth further investigation.

Comparison of the allelic effect among different alleles at the same locus could determine which specific alleles would be most informative for marker assisted selection. For example, allele 126bp of RM471 and 108bp of RM7003 were considered major alleles with a positive effecton increasing yield among all the alleles in the loci (Fig. 7). Howeve, the allele 106bp of RM7003 would be less desirable because it had a negetaive effect which is associated with a decrease of both grain yield and plant weight among accessions containing the allele. Results of the present study demonstrated that genome-wide association mapping in the USDA rice mini-core collection could complement and enhance the information from linkage-based QTL studies, and help increase yield through improvement of these related traits by marker-assisted selection either directly or indirectly.

6. Association mapping of resistance to Sheath Blight disease

Rice sheath blight (ShB), caused by the soil-borne fungal pathogen *Rhizoctonia solani* Kühn, is a major disease of rice that greatly reduces yield and grain quality worldwide (Savary et al. 2006). Due to the high cost of cultural practices and the phytotoxic influence associated with the application of fungicides, the use of ShB resistant cultivars is considered the most economical and environmentally sound strategy in managing this disease. Understandings of genetic control will facilitate cultivar improvement for this disease and secure global food production.

The necrotrophic ShB pathogen has a broad host range and no complete resistance has been identified in either commercial rice cultivars or wild related species (Mew et al. 2004; Eizenga et al. 2002). However, substantial differences in susceptibility to ShB among rice cultivars have been observed under field conditions (Jia et al. 2007). Differential levels of resistance and the associated resistance genes have been studied among rice germplasm accessions (Manosalva et al. 2009). Rice ShB resistance is believed to be controlled by multiple genes or quantitative trait loci (QTLs) (Pinson et al. 2005). Since Li et al. (1995) first identified ShB QTLs using restricted fragment length polymorphism (RFLP) markers under field conditions, over 30 resistant ShB QTLs have been reported using various mapping populations, such as F_2s (Sharma et al. 2009; Che et al. 2003), double haploid (DH) lines (Kunihiro et al. 2002), recombinant inbred lines (RILs) (Liu et al. 2009; Jia et al. 2007; Prasad and Eizenga 2008), near-isogenic introgression lines (NIL) (Loan et al. 2004) and backcross populations (Zuo et al. 2007; Sato et al. 2004). 'Teqing' and 'Jasmine 85' have been repeatedly involved in these studies as the ShB resistant parents. We are the first to map rice ShB QTLs using an association mapping strategy in a global germplasm collection (Jia et al. 2012).

6.1. Phenotypic evaluation of Sheath Blight resistance

The isolate RR0140-1 of *R. solani* was selected from 102 isolates collected state-wide from Arkansas rice fields due to its slow growing phenotype (Wamishe et al. 2007). Field evaluations have showed similar disease reactions between slow growing and fast growing isolates (Wamishe et al. 2007). Further, the RR0140-1 isolate has been adapted by numerous studies (Liu et al. 2009; Jia et al. 2007; Prasad and Eizenga 2008). Pathogen preparation and inoculation are referred to Jia et al. (2007; 2011; 2012).

Plant response to the sheath blight pathogen was measured using the ratio between the height of the pathogen growing up the plant and the height of the leaf collar on the last emerged leaf. Because mature plant height varied from 70 to 202 cm in this collection (Yan et al. 2007), the ratio excluded possible interference of plant height in scoring disease response. Therefore, the smaller the ratio, the greater the resistance was for an entry. Measurements were taken when the ratio reached 1.0 for 75% of the susceptible check plants, Lemont, so that maximum susceptibility was scored as 1.0.

ShB rating data were analyzed using the GLIMMIX procedure in SAS version 9.1.3. The experimental design of randomized incomplete block formed the basis of the statistical model,

where the accession is a fixed effect and block is treated as random effect. The LSMEANS option was used to calculate the least-square means (LSMs) from 18 plant scores in 6 replicates of each entry and the LSMs were used for the association mapping. The statistical differences of the accession to each check (Jasmine 85 and Lemont) were determined by a Dunnett's multiple comparison test, using the diff=control option.

6.2. Phenotypic variation of Sheath Blight severity ratings

The ShB severity ratings among the 217 entries were distributed normally, ranging from 0.256 ± 0.111 to 0.909 ± 0.096 with an average of 0.521 ± 0.008 (Fig. 8). The resistant check Jasmine 85 was rated 0.472 ± 0.021 and susceptible check Lemont was rated 0.946 ± 0.080. Twenty-four entries (11.1 %) were significantly more resistant to ShB than Jasmine 85 at the 5% level of probability while 54 others (24.9%) had similar resistance.
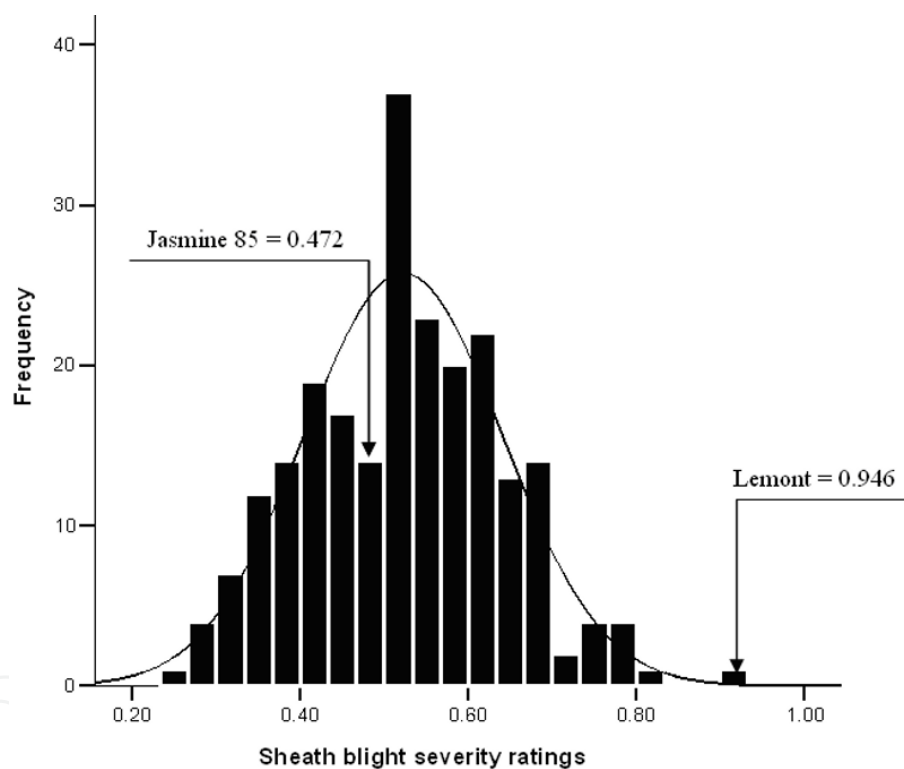
Figure 8. Distribution of sheath blight severity ratings among 217 mini-core accessions averaged over 18 plant scores, 3 in each of 6 replicates using a micro-chamber method with resistant check Jasmine 85 and susceptible Lemont (Jia et al. 2012).

6.3. Marker loci and their alleles associated with Sheath Blight resistance

Ten marker loci were identified to be significantly associated with ShB resistance at the probability level of 5% or lower, three on chromosome (Chr) 11, two on Chr1, and one each on Chr2, 4, 5, 6 and 8 (Table 5). RM237 on Chr1 at 27.1 Mb had the highest significance rating for ShB at the 0.002 level of probability. RM11229 on the long arm of Chr1 explained the most

phenotypic variation (9.5%) with significance at the 0.044 level of probability. RM11229 and 1233 each had six alleles, the most among the 217 mini-core entries, followed by RM341 and 254 (five alleles), RM237, 8217,146 and 408 (four), RM133 (three) and RM7203 (two) (Table 5).

Among the six alleles of RM11229, allele 158 was present in 18 entries that had the lowest average ShB rating (0.414), and thus, it was designated as the 'putative resistant allele' of this marker locus. Accordingly, ten alleles, one each from the ten associated marker loci, were noted as the putative resistant allele in Table 5 because they had the greatest effect to decrease ShB among all the alleles for their respective loci (Table 5). ShB rating was the smallest for putative resistant allele 158 of RM11229 among the ten putative resistant alleles. Of the other five putative resistant alleles, 139 of RM341 (present in 17 entries), 340 of RM146 (28 entries), 88 of RM7203 (120 entries), 169 of RM254 (12 entries) and 177 of RM1233 (35 entries), had lower ShB means ranging 0.447 - 0.470 than the resistant check Jasmine 85 (0.472), suggesting a stronger effect for resistance to ShB than Jasmine 85. The remaining four putative resistant alleles had similar ShB ratings with Jasmine 85, suggesting a similar effect for the level of ShB control.

Marker	Chr.	Position (Mb)	P value	Rsq_Marker [a]	Allele (bp)	Number of Entries	ShB Mean [b]
RM11229	1	22.4	0.044	9.5%	158*	18	0.414
					192	21	0.515
					195	21	0.473
					198	13	0.532
					207	14	0.608
					224	12	0.466
RM237	1	26.8	0.002	6.9%	122	19	0.526
					128*	32	0.473
					130	105	0.515
					132	20	0.635
RM341	2	19.3	0.041	4.1%	135	89	0.558
					138	39	0.545
					139*	17	0.447
					141	15	0.579
					171	39	0.461
RM8217	4	32.7	0.044	3.2%	178	67	0.581
					182	19	0.534
					184	65	0.482
					186*	49	0.476
RM146	5	18.1	0.021	3.8%	330	26	0.539

|					332	127	0.512
					340*	28	0.463
					344	28	0.591
RM133	6	0.2	0.043	2.4%	228	89	0.557
					230*	104	0.479
					232	22	0.577
RM408	8	0.1	0.023	4.0%	117	18	0.577
					119*	105	0.478
					125	18	0.498
					127	57	0.591
RM7203	11	1.1	0.033	1.9%	88*	120	0.470
					104	80	0.589
RM254	11	23.8	0.030	5.3%	159	25	0.580
					161	36	0.564
					163	54	0.511
					167	50	0.480
					169*	12	0.463
RM1233	11	26.5	0.036	5.1%	158	97	0.538
					164	12	0.543
					168	15	0.593
					173	12	0.520
					177*	35	0.451
□	□	□	□	□	179	12	0.524
Resistant check 'Jasmine 85'							0.472

[a] Rsq_Marker - total explained phenotypic variation.

[b] The mean of ShB severity rating for the entries with the allele.

Allele*: Putative resistant allele which had the lowest ShB mean at the marker locus.

Table 5. Marker loci significantly associated with sheath blight resistance, their physical locations on chromosomes (Chr), allele size in 217 mini-core entries, number of entries with the allele, and their mean sheath blight (ShB) rating (Jia et al. 2012)

Among the ten putative resistant alleles, allele 88 of RM7203 was the most prevalent and existed in 120 (55%) of 217 entries in the mapping panel, followed by allele 230 of RM133 and 119 of RM408 (48% of the lines), allele 186 of RM8217 (23%), allele 340 of RM146, 128 of RM237 and 177 of RM1233 (13-16%), allele 139 of RM341 and 158 of RM1229 (8%), and allele 169 of RM254 (6%).
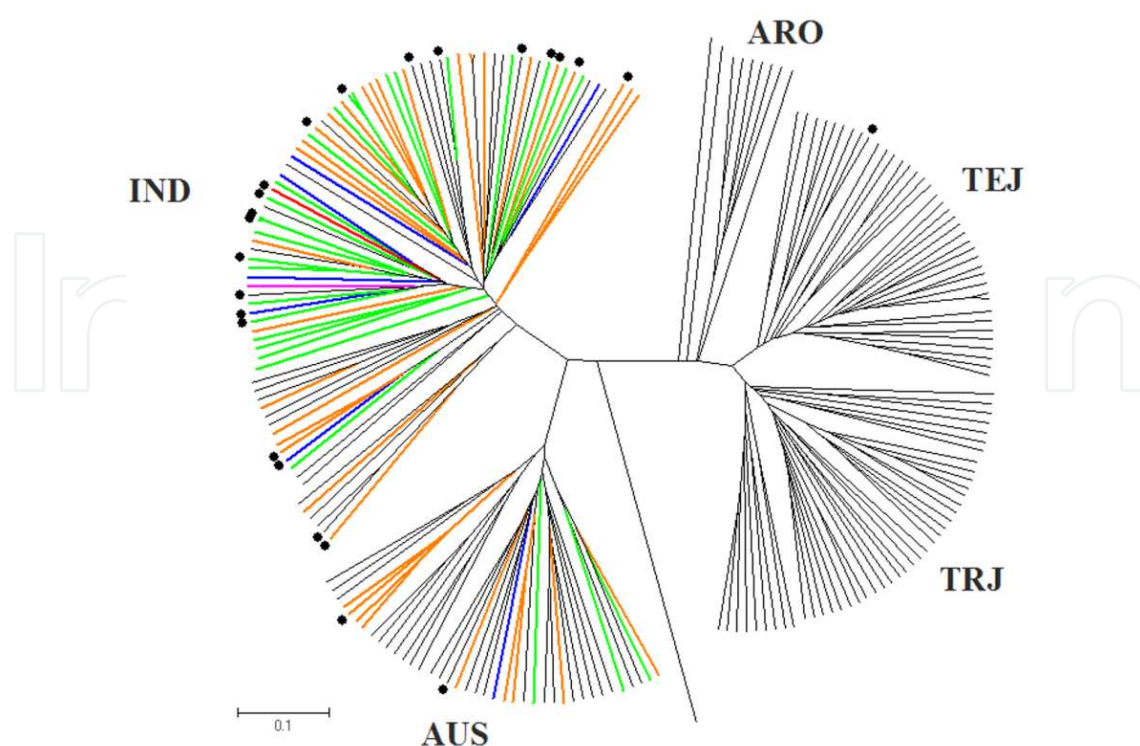
Figure 9. UPGMA tree based on Nei genetic distance for 217 mini-core entries where 24 marked with ● were significantly more resistant to sheath blight than the resistant check 'Jasmine 85'. Presence of 'putative resistant alleles' is distinguished by branch color: Red= eight putative resistant alleles, Pink= seven, Blue= six, Green= five, and Orange = four (Jia et al. 2012).

6.4. Number of putative resistant alleles and Sheath Blight resistance

The number of putative resistant alleles increased along with an increase of sheath blight resistance in an accession of rice germplasm. GSOR 310389 from Korea contained the most putative resistant alleles, eight out of ten, and had a ShB rating of 0.351 which was significantly more resistant than the resistant check Jasmine 85 which contained three putative resistant alleles and had a ShB rating of 0.472. Among seven entries containing six putative resistant alleles with a mean of 0.386 ShB, GSOR 310475 and 311475 were more resistant than Jasmine 85 and had ShB ratings of 0.324 and 0.336, respectively. Among 28 entries having five putative resistant alleles with a mean ShB rating of 0.444, seven were significantly more resistant than Jasmine 85. Seven, out of 35 entries which carried four putative resistant alleles and had a mean ShB of 0.466, were identified to be significantly more resistant than Jasmine 85. The mean ShB ratings for entries containing three, two, one and zero putative resistant alleles were 0.483, 0.535, 0.582 and 0.598, respectively. There was a strong and negative correlation between the ShB severity rating and number of putative resistant alleles (r = -0.535, p<0.0001).

Our mapping results showed that most entries containing a large number of putative resistant alleles were IND (Fig. 9). All entries with six or more putative resistant alleles were IND with only one exception of AUS. Among 28 entries with five putative resistant alleles, 25 were IND and the remaining three were AUS. There were 35 entries with four putative resistant alleles,

nine were AUS, one was admix of TRJ, AUS and IND, and the remaining 25 were IND. Among 35 entries with three putative resistant alleles, 18 were IND, eight AUS, seven TRJ and two admixes of IND. However, among 51 entries without a single putative resistant allele, 26 were TEJ, 18 TRJ, four ARO and two admixes of TRJ-TEJ-ARO, and one IND. Among 72 entries that carried four or more putative resistant alleles, 58 (81%) were IND and 13 AUS (18%) plus admix of TRJ-AUS-IND.

6.5. Putative resistant alleles and ancestry background for Sheath Blight resistance

Jia et al. (2011) reported 52 entries that are significantly more resistant to ShB than Jasmine 85. The resistant entries were identified from 1,794 entries of the USDA rice core collection that has 35% *indica*, 27% *temperate japonica*, 24% *tropical japonica*, 10% *aus* and 4% *aromatic* genotypes (Agrama et al. 2010). Based on the ancestry classification, there are 621 *indica* entries in the core and 45 of them are included in the resistant list, making a resistance frequency of 7.2% for *indica* germplasm. Accordingly, the resistance frequency is 2.8% for *aromatic*, 1.7% for *aus*, and 0.2% each for *temperate japonica* and *tropical japonica*. In a study conducted by Zuo et al. (2007), *japonica* cultivars showed higher sheath blight severity than *indica* cultivars. They describe a general observation that *japonica* rice is more susceptible than *indica* rice. Furthermore, Jasmine 85, Tetep and Teqing, used as parents in many studies on mapping ShB resistance, are all *indica*.

Our study demonstrated that: 1) a majority of the ShB putative resistant alleles existed in *indica* germplasm, 2) most of the resistant entries with a large number of putative resistant alleles were *indica*, conversely 3) only a very small portion of putative resistant alleles existed in *japonica*, and 4) the most susceptible entries with very few or no putative resistant alleles were *japonica* (Fig. 8). Entry GSOR 310389 is an example which had eight out of ten putative resistant alleles, showed a high level of resistance to ShB, and is *indica*. The results from association mapping match well with the phenotypic observation that most resistant genotypes are *indica* and resistant germplasm is rare in *japonica*.

7. Association mapping of silica concentration in rice hulls

Rice (*Oryza sativa* L.) accumulates silicon (Si) in various tissues including hulls. Although Si is not an essential nutrient, it plays an important role in the growth and health of rice plants. Silicic acid is actively taken up by rice roots, which is then translocated in the form of mono-silicic acid (silica gel) through the xylem (Mitani et al. 2005; Ma and Yamaji 2006) to the leaves, stems, hulls and grains of the plant where it converted to silica (SiO_2) (Ma et al. 2007). Often, rice hulls are burned in the mills to produce steam or electricity. However, disposal of the rice hull ash is difficult due to the high silica content (70-95%) (Marshall 2004). Unused hulls and ash are taken to a landfill where they remain for years due to their chemical stability. Another approach for reducing the amount of hulls and ash going to the landfill is to use the silica for value-added products. Rice hulls have been used to produce particle board, poultry bedding, brick making, package cushioning, and absorbents. Due to the high silicon content, rice hulls and ash are good raw materials in the production of silicon-based industrial materials with

high economic value, including silicon carbide, silica, silicon nitride, silicon tetrachloride, pure silicon, and zeolite (Sun and Gong 2001). Since the Si in rice hulls is amorphous, it can be extracted at lower temperatures than Si derived from other conventional sources, thus reducing the cost of Si production (Kalapathy et al. 2002). Understanding the genetic control of Si content in rice will facilitate the development of new varieties with either high or low Si content. Varieties with high silica content in their hulls would be useful for raw material for silica based industrial compounds, while varities with low silica hulls would be more biode-gradable and better suited for energy producing purposes (i.e. cleaner energy production at the mills and possible use in bioenergy production).

7.1. Chemical analysis of silica concentration in rice hulls

The rough rice samples from test plots were dehulled with a Satake Rice Machine (Satake Engineering Co., LTD, Ueno Taito-Ku, Tokyo). After drying at 80°C for 2 hr, the hulls (~3g) were stored in 50 ml polypropylene tubes (Cat. # 05-539-5, Fisher Scientific, Houston, TX) at room temp. (22°C) until analyzed. Silica was determined using the molybdenum yellow method described by Saito et al. (2005) and Bryant et al. (2011).

7.2. Variation of silica concentration in the USDA rice mini-core collection

Si content averaged 200 mg g^{-1} and ranged from 118 mg g^{-1} for ACNO 430909, an Admixture of *aus* (AUS), *indica* (IND) and *wild rice* (WD) from the Punjab region of Pakistan, to 249 mg g^{-1} for ACNO 353722, an AUS accession from Assam, India. The non-Admix accession with the lowest Si was ACNO 439683, a TEJ from Eastern Europe, having a Si of 147 mg g^{-1}. Wide variation of Si was seen in all genetic groups. Mean Si of the TRJ (219 mg g^{-1}) and AUS (208 mg g^{-1}) was greater while other groups were less than the overall mean. All the accessions native to Central America region (n = 9), except a TRJ ACNO 2169 from Guatemala, were above the Mini-Core mean value, whereas the Si contents of accessions native to the Mideast (n = 5), Eastern Europe (n = 8), Central Asia (n = 9) and North America (n = 3) were below the Mini-Core mean with a few exceptions. The variation due to genetics (accessions) accounted for 32.4% of the total Si variation in the Mini-Core. The silica content of samples grown in Beaumont, TX (186±1.3 mg g^{-1}) was lower than those grown in Stuttgart, AR (211±1.2 mg g^{-1}), with Location accounting for 19.5% of the silica content variation.

7.3. Marker loci associated with silica concentration

We identified four associated markers in AR, and they were different from the four identified in TX (Table 6). Three out of four AR markers were among seven associated markers mapped in the combined location, whereas none of the TX markers were in. The 19.5% of the total silica content variation due to the difference of AR from TX might be responsible for the mapping results. It is known that the amount of silica present in the soil, the presence of other elements and/or nutrients, the amount of light, and temperature are all factors that affect silica concen-trations in the plant (Ma and Takahashi 2002; Ma et al. 2002). RM263 from AR, RM6544 from both AR and Combined location and RM5371 from TX are all within a 1.5 Mb region where additive by additive QTL effects were previously identified by Dai et al. (2005). In summary,

five of the marker-trait associations found in this study are within 1.5 Mb of the reported QTLs for silica concentrations from linkage mapping studies, and one marker-trait association (RM5371 on chromosome 6 at 25.83 Mb) overlaps with a QTL involved in grain arsenic concentration as well as silica concentration (Dai et al. 2005). The present study demonstrates that association mapping of the diverse germplasm in the USDA rice Mini-Core collection is an effective method for identifying new genetic markers and validating previously reported marker regions associated with silica concentration.

Location	Locus	Chr.	Position (Mb)	*P Value*	-log₁₀P	Major allele (bp)	Effect	# of Observed Genotypes
Arkansas	RM23869	9	6.3	2.30E-04	3.64	182	-14.6	37
	RM6544[ab]	11	3.9	3.10E-03	2.51	165	20.2	11
						170	-17.9	47
	RM5953	4	9.4	3.30E-03	2.48	111	15.3	35
	RM263[b]	2	25.9	5.20E-03	2.28	176	-22.8	6
						158	22.0	30
Texas	RM1335[a]	7	28.3	3.70E-04	3.44	173	22.8	7
						163	-22.3	10
						165	21.1	6
						171	-19.1	13
	RM5371[bc]	6	25.8	6.30E-04	3.2	170	-18.6	7
	RM1186	7	9	1.40E-03	2.85	120	-17.5	18
	RM178[a]	5	25.1	4.40E-03	2.36	114	-23.6	72
Combined	RM5953	4	9.4	8.60E-04	3.07	111	14.6	35
	RM23869	9	6.3	2.30E-03	2.64	182	-9.9	37
	RM283[a]	1	4.9	5.10E-03	2.29	150	-19.7	5
	RM6544[ab]	11	3.9	6.80E-03	2.17	165	14.9	11
	RM489	3	4.3	7.90E-03	2.1	269	15.2	25
	RM171	10	19.1	8.40E-03	2.08	328	-18.8	42
	RM484	10	21.1	8.80E-03	2.06	290	-22.4	4

[a] Markers that occur within 1.5 Mb or less of previously identified silica QTL's.

[b] Markers that occur within 1.5 Mb or less of previously identified additive-by-additive QTL regions.

[c] Marker that occur within 1.5 Mb or less of previously identified arsenic QTL's.

Table 6. Marker loci associated with hull silica content at less than 0.01 probability mapped among 174 mini-core accessions genotyped with 164 SSR markers and phenotyped at Stuttgart, Arkansas (AR) and Beaumont, Texas (TX) (Bryant et al. 2011)

Acknowledgements

Author details

Wengui Yan[1], Aaron Jackson[1], Melissa Jia[1], Wei Zhou[1,2], Haizheng Xiong[1,3] and Rolfe Bryant[1]

1 United States Department of Agriculture, Agricultural Research Service, Dale Bumpers National Rice Research Center, Stuttgart, Arkansas, USA

2 University of Arkansas at Pine Bluff, Pine Bluff, Arkansas, USA

3 Zhejiang University, State Key Lab of Rice Biology, Institute of Nuclear-Agricultural Sciences, Hangzhou, Zhejiang, China

USDA is an equal opportunity provider and employer

References

[1] Agrama HA, Eizenga GC (2008) Molecular diversity and genome-wide linkage disequilibrium patterns in a worldwide collection of *Oryza sativa* and its wild relatives. Euphytica 160:339–355

[2] Agrama HA, Yan WG, Lee F, Fjellatrom R, Chen MH, et al. (2009) Genetic assessment of a mini-core subset developed from the USDA rice genebank. Crop Sci. 49:1336-1346

[3] Akash WM, Kang SM (2009) Molecular Clustering and Interrelationships among Agronomic Traits of Jordanian Barley Cultivars. J. Crop Impro. 24: 28 – 40

[4] Ali ML, Rajewski JF, Baenziger PS, Gill KS, Eskridge KM (2008) Assessment of genetic diversity and relationship among a collection of US sweet sorghum germplasm by SSR markers. Mol Breed 21:497-509

[5] Ando T, Yamamoto T, Shimizu T, Ma XF, Shomura A, et al. (2008) Genetic dissection and pyramiding of quantitative traits for panicle architecture by using chromosomal segment substitution lines in rice. Theor Appl Genet 116: 881-890

[6] Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, et al. (2005) Cytokinin oxidase regulates rice grain production. Science 309:741-745

[7] Ashikari M, Wu J, Yano M, Sasaki T, Yoshimura A (1999) Rice gibberellin-insensitive dwarf mutant gene Dwarf 1 encodes the alpha-subunit of GTP-binding protein. Proc Natl Acad Sci USA 96: 10284-10289

[8] Bernardo R (2008) Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Sci 48:1649-1664

[9] Borba TCO, Brondani RPV, Breseghello F, Coelho ASG, Mendonça JA, et al. (2010) Association mapping for yield and grain quality traits in rice (*Oryza sativa* L.). Genet Mol Biol 33:515-524

[10] Borba TCO, Brondani RPV, Rangel PHN, Brondani C (2009) Microsatellite marker–mediated analysis of the EMBRAPA Rice Core Collection genetic diversity. Genetica 137:293–304

[11] Borba TCO, Brondani RPV, Rangel PHN, Brondani C (2005) Evaluation of the number and information content of fluorescent-labeled SSR for rice germplasm characterization. Crop Breed Appl Biotechnol 2: 157-165

[12] Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635

[13] Breseghello F, Sorrells ME (2006) Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci 46: 1323-1330

[14] Brondani C, Borba TCO, Rangel PHN, Brondani RPV (2006) Determination of traditional varieties of Brazilian rice using microsatellite markers. Genet Mol Biol 29: 676-684

[15] Brondani C, Rangel N, Brondani V, Ferreira E (2002) QTL mapping and introgression of yield-related traits from *Oryza glumaepatula* to cultivated rice (*Oryza sativa*) using microsatellite markers. Theor Appl Genet 104:1192-1203

[16] Brooks SA, Yan W, Jackson AK, Deren CW (2008) A natural mutation in *rc* reverts white-rice-pericarp to red and results in a new, dominant, wild-type allele:*Rc-g*. Theor Appl Genet 117:575-580

[17] Bryant R, Proctor A, Hawkridge M, Jackson A, Yeater K, et al. (2011) Genetic variation and association mapping of silica concentration in rice hulls using a germplasm collection. Genetica 139:1383-1398

[18] Can ND, Yoshida T (1999) Genotypic and phenotypic variances and covariances in early maturing grain sorghum in a double cropping. Pl Prod Sci 2: 67-70

[19] Che KP, Zhan QC, Xing QH, Wang ZP, Jin DM, et al. (2003) Tagging and mapping of rice sheath blight resistant gene. Theor Appl Genet 106: 293-297

[20] Chen Y, Lubberstedt T (2010) Molecular basis of trait correlations. Trends in Plant Sci 15:454-461

[21] Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, et al. (2000) Diversity of microsatellites derived from genomic libraries and genbank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

[22] Cooper M, Podlich DW, Smith OS (2005) Gene-to-phenotype models and complex trait genetics. Aust J Agric Res 56:895-918

[23] Dai WM, Zhang KQ, Duan BW, Zheng KE, Zhuang JY, et al. (2005) Genetic dissection of different organs in rice. Crop Sci 45:1345-1352

[24] Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55: 997-1004

[25] Dhanraj A, Jagadish CA (1987) Studies on character association in the F2 generation of ten selected crosses in rice (*Oryza sativa* L.). Journal of Research-A FAU 15:64-65

[26] Doi K, Izawa T, Fuse T, Yamanouchi U, Kubo T, et al. (2004) *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls *FT-like* gene expression independently of *Hd1*. Genes Dev 18:926-936

[27] Eizenga GC, Lee FN, Rutger JN (2002) Screening *Oryza* species plants for rice sheath blight resistance. Plant Dis 86: 808-812

[28] Elias M, Penet L, Vindry P, McKey D, Panaud O, et al. (2001a) Unmanaged sexual reporoduction and the dynamics of genetic diversity of a vegetatively propagated crop plant, cassava (*Manihot esculenta* Crantz), in a traditional farming system. Mol Eco 10:1895-1907

[29] Elias M, McKey D, Panaud O, Anstett MC, Robert T (2001b) Traditional management of cassava morphological and genetic diversity by the Makushi Amerindians (Guyana, South America): perspectives for on-farm conservation of crop genetic resources. Euphytica 120:143-157

[30] Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611-2620

[31] Falush D, Stephens M, Prithchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567-1587

[32] Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q (2006) *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet 112: 1164-1171

[33] Fu Q, Zhang P, Tan L, Zhu Z, Ma D, et al. (2010) Analysis of QTLs for yield-related traits in Yuanjiang common wild rice (*Oryza rufipogon Griff.*). J Genet Genomics 37: 147-157

[34] Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch SR (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638

[35] Gaudeul M, Taberlet P, Till-Bottraud I (2000) Genetic diversity in an endangered alpine plant, Eryngium alpinum L. (Apiaceae), inferred from amplified fragment length polymorphism markers. Molecular Ecol 9: 1625–1637

[36] Gepts P (2004) Crop domestication as a long-term selection experiment. Plant breeding reviews 24: 1-44

[37] Gizaw S, Van Arendonk JAM, Komen H, Windig JJ, Hanotte O (2007) Population structure, genetic variation and morphological diversity in indigenous sheep of Ethiopia. Animal Genetics 38: 621-628

[38] Glaszmann JC (1987) Isozymes and classification of Asian rice varieties. Theor Appl Genet 74: 21–30

[39] Gravois KA, McNew RW (1993) Genetic relationships among and selection for rice yield and yield components. Crop Sci 33:249-252

[40] Hao Z, Li X, Liu X, Xie C, Li M, et al. (2010) Meta-analysis of constitutive and adaptive QTL for drought tolerance in maize. Euphytica 174: 165–177

[41] Hay RKM (1995) Harvest index: a review of its use in plant breeding and crop physiology. Annu Appl Biol 126: 197-216

[42] Hemamalini GS, Shashidhar HE, Hittalmani S (2000) Molecular marker assisted tagging of morphological and physiological traits under two contrasting moisture regimes at peak vegetative stage in rice (*Oryza sativa* L.). Euphytica 112: 69-78

[43] Hittalmani S, Huang N, Courtois B, Venuprasad R, Shashidhar HE, et al. (2003) Identification of QTL for growth- and grain yield-related traits in rice across nine locations of Asia. Theor Appl Genet 107: 679-690

[44] Holland J (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10:156-161

[45] Huang X, Wei X, Sang T, Zhao Q, Feng Q, et al. (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genetics 42: 961-969

[46] Huang X, Qian Q, Liu Z, Sun H, He S, et al. (2009) Natural variation at the *DEP1* locus enhances grain yield in rice. Nat Genet 41:494-497

[47] Hulbert SH, Bennetzen JL (1991) Recombination at the Rp1 locus of maize. Mol Gen Genet 226:377–382

[48] Inostroza L, Pozo AD, Matus I, Castillo D, Hayes P, et al. (2009) Association mapping of plant height, yield, and yield stability in recombinant chromosome substitution lines (RCSLs) using *Hordeum vulgare* subsp. spontaneum as a source of donor alleles in a *Hordeum vulgare* subsp. Vulgare background. Mol Breed 23:365-376

[49] Itoh H, Tatsumi T, Sakamoto T, Otomo K, Toyomasu T, et al. (2004) A rice semi-dwarf gene, *Tan-Ginbozu (D35)*, encodes the gibberellin biosynthesis enzyme, *ent*-kaurene oxidase. Plant Mol Biol 54:533-547

[50] Jain S, Jain RK, McCouch SR (2004) Genetic analysis of Indian aromatic and quality rice (*Oryza sativa* L.) germplasm using panels of fluorescently–labeled microsatellite markers. Theor Appl Genet 109:965–977

[51] Jia LM, Yan WG, Zhu C, Agrama HA, Jackson A, et al. (2012) Allelic analysis of sheath blight resistance with association mapping in rice. PLoS ONE 7(3): e32703. doi:10.1371

[52] Jia L, Yan W, Agrama HA, Yeater K, Li X, et al. (2011) Searching for germplasm resistant to sheath blight from the USDA rice core collection. Crop Sci 51: 1507-1517

[53] Jia Y, Correa-Victoria F, McClung A, Zhu L, Liu G, et al. (2007) Rapid determination of rice cultivar responses to the sheath blight pathogen *Rhizoctonia solani* using a micro-chamber screening method. Plant Dis 91: 485-489

[54] Jiang GH, Xu CG, Li XH, He YQ (2004) Characterization of the genetic basis for yield and its component traits of rice revealed by doubled haploid population. Acta genetica Sinicavol 31:63-72

[55] Kalapathy U, Proctor A, Shultz J (2002) An improved method for production of silica from rice hull ash. Bioresource Technol 85:285-289

[56] Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, et al. (2008) Efficient control of population structure in model organism association mapping. Genetics 178: 1709-1723

[57] Kim KW, Chung HK, Cho GT, Ma KH, Chandrabalan D, et al. (2007) PowerCore: A program applying the advanced M strategy with a heuristic search for establishing core sets. Bioinformatics 23:2155–2162.

[58] Kim S, Zhao K, Jiang R, Molitor J, Borevitz JO, et al. (2006) Association mapping with single-feature polymorphisms. Genetics 173: 1125-1133

[59] Kouame CN, Quesenberry KH (1993) Cluster analysis of a world collection of red clover germplasm. Genet Res Crop Evo 40:39-47

[60] Kunihiro Y, Qian Q, Sato H, Teng S, Zeng D, et al. (2002) QTL analysis of sheath blight resistance in rice (*Oryza sativa* L.). Chin J Genet 29: 50-55

[61] Lanceras JC, Pantuwan GP, Jongdee B, Toojinda T (2004) Quantitative trait loci associated with drought tolerance at reproductive stage in rice. Plant Physiol 135: 384–399

[62] Li, X., Yan W, Agrama H, Jia L, Jackson AK, et al. 2012. Unraveling the complex trait of harvest index with association mapping in rice (Oryza sativa L.). PLoS One. 7(1): e29350. doi:10.1371/journal.pone.0029350

[63] Li X, Yan W, Agrama H, Jia L, Shen X, et al. (2011) Mapping QTLs for improving grain yield using the USDA rice mini-core collection. Planta 234: 347-361

[64] Li X, Yan W, Agrama H, Hu B, Jia L, et al. (2010) Genotypic and phenotypic characterization of genetic differentiation and diversity in the USDA rice mini-core collection. Genetica 138: 1221-1230

[65] Li C, Zhou A, Sang T (2006) Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara.* New Phytologist 170:185-194

[66] Li X, Qian Q, Fu Z, Wang Y, Xiong G, et al. (2003) Control of tillering in rice. Nature 422:618-621

[67] Li DJ, Sun CQ, Fu YC, Chen L, Zhu ZF, et al. (2002) Identification and mapping of genes for improving yield from Chinese common wild rice (*O. rufipogon* Griff.) using advanced backcross QTL analysis. Chin Sci Bull 18:1533-1537

[68] Li Z, Pinson SRM, Marchetti MA, Stansel JW, Park WD (1995) Characterization of quantitative trait loci (QTLs) in cultivated rice contributing to field resistance to sheath blight (*Rhizoctonia solani*). Theor Appl Genet 91: 382-388

[69] Liu G, Jia Y, Correa-Victoria FJ, Prado GA, Yeater KM, et al. (2009) Mapping quantitative trait loci responsible for resistance to sheath blight in rice. Phytopathology 99: 1078-1084

[70] Liu K, Muse SV (2005) Powermarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21: 2128-2129

[71] Lewis PO, Zaykin D (2001) Genetic data analysis: computer program for the analysis of allelic data version 1.0. http://lewis.eeb.uconn.edu/lewishome/software.html. Cited 21 Nov 2001

[72] Loan LC, Du PV, Li Z (2004) Molecular dissection of quantitative resistance of sheath blight in rice (*Oryza sativa* L.). Omonrice 12: 1-12

[73] Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). Am J Bot 82: 1420-1425

[74] Ma JF, Yamaji N (2006) Silicon uptake and accumulation in higher plants. Trends Plant Sci 11: 392-397

[75] Ma JF, Yamaji N, Tamai K, Mitani N (2007) Genotypic difference in silicon uptake and expression of silicon transporter genes in rice. Plant Physiology 145:919-924

[76] Ma JF, Takahashi E (2002) Soil, fertilizer, and plant silicon research in Japan. Elsevier Science

[77] Ma JF, Tamai K, Ichii M, Wu GF (2002) A rice mutant defective in Si uptake. Plant Physiol 130:2111–2117

[78] Mackill DJ (1995) Plant genetic resources: Classifying japonica rice cultivars with RAPD markers. Crop Sci 35: 889-894

[79] Manosalva PM, Davidson RM, Liu B, Zhu X, Hulbert SH, et al. (2009) A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. Plant Physiol 149: 286-296

[80] Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res., 27:209–220

[81] Mao B-B, Cai W-j, Zhang Z-h, Hu Z-L, Li P, et al. (2003) Characterization of QTLs for Harvest Index and Source-sink Characters in a DH Population of Rice (*Oryza sativa* L.). Acta Genetica Sinica 30: 1118-1126

[82] Marri PR, Sarla N, Reddy LV, Siddiq EA (2005) Identification and mapping of yield and yield related QTLs from an Indian accession of *Oryza rufipogon*. BMC Genet 13: 33-39

[83] Marshall WE (2004) Utilization of rice hull and rice straw as adsorbents. In: Champagne ET (ed), Rice Chemistry and Technology. American Association of Cereal Chemists, Inc., St. Paul, MN, pp. 611-630

[84] Mazzucato A, Papa R, Bitocchi E, Mosconi P, Nanni L, et al. (2008) Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (Solanum lycopersicum L.) landraces. Theor Appl Genet 116:657-669

[85] Mew TW, Leung H, Savary S, Cruz CMV, Leach JE (2004) Looking ahead in rice disease research and management. Cr Rev Plant Sci 23: 103-127

[86] Mitani N, Ma JF, Iwashita T (2005) Identification of the silicon form in xylem sap of rice (*Oryza sativa* L.). Plant Cell Physiol 46:279-283

[87] Mohammad D, Cox PB, Posler GL, Kirkham MB, Hussain A, et al. (1993) Correlation of characters contributing to grain and forage yields and forage quality in sorghum (*Sorghum bicolor*). Indian J Agric Sci 63: 92–95

[88] Moncada P, Martinez CP, Borrero J, Chatel M, Gauch H, et al. (2001) Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theor Appl Genet 102:41-52

[89] Nei M, Takezaki N (1983) Estimation of genetic distances and phylogenetic trees from DNA analysis. Proc 5th World Cong Genet Appl Livstock Prod 21: 405–412

[90] Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, et al. (1999) 'Green revolution' genes encode mutant gibberellin response modulators. Nature 400: 256-261

[91] Pinson SRM, Capdevielle FM, Oard JH (2005) Confirming QTLs and finding additional loci conditioning sheath blight resistance in rice using recombinant inbred lines. Crop Sci 45: 503-510

Association Mapping of Harvest Index, Grain Yield, Sheath Blight Resistance and... 139
http://dx.doi.org/10.5772/56830

[92] Pinto RS, Reynolds MP, Mathews KL, McIntyre CL, Olivares-Villegas JJ, et al. (2010) Heat and drought adaptive QTL in a wheat population to minimize confounding agronomic effects. Theor Appl Genet 121: 1001-1021

[93] Prasad B, Eizenga GC (2008) Rice sheath blight disease resistance identified in *Oryza* spp. accessions. Plant Dis 92: 1503-1509

[94] Prichard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. AM J Hum Genet 67: 170-181

[95] Raes D, Steduto P, Hsiao TC, Fereres E (2009) AquaCrop-The FAO Crop Model to Simulate Yield Response to Water: II. Main Algorithms and Software Description. Agron J 101: 438-447

[96] Rohlf F (2000) FNTSYS-PC numerical taxonomy and multivariate analysis system ver 2.11L. Applied Biostatistics, NY

[97] Saito K, Yamamoto A, Sa T, Saigusa M (2005) Rapid, micro-methods to estimate plant silicon content by dilute hydrofluoric acid extraction and spectrometric molybdenum method: I. Silicon in rice plants and molybdenum yellow method. Soil Sci Plant Nutr 51:29-36

[98] Samonte SOP, Wilson LT, McClung AM (1998) Path analyses of yield and yield-related traits of fifteen diverse rice genotypes. Crop Sci 38:1130-1136

[99] Sato H, Ideta O, Ando I, Kunihiro Y, Hirabayashi H, et al. (2004) Mapping QTLs for sheath blight resistance in the rice line WSS2. Breeding Sci 54: 265-271

[100] Savary S, Teng PS, Willocquet L, Nutter FW (2006) Quantification and modeling of crop losses: a review of purposes. Annu Rev Phytopathol 44: 89-112

[101] Schneider S, Excoffier L (1999) Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics 152:1079–1089

[102] Sharma A, McClung AM, Pinson SRM, Kepiro JL, Shank AR, et al. (2009) Genetic mapping of sheath blight resistance QTLs within *tropical japonica* rice cultivars. Crop Sci 49: 256-264

[103] Sharma RS, Choubey SD (1985) Correlation studies in upland rice. Indian Jour Agro 30:87-88

[104] Shi J, Li R, Qiu D, Jiang C, Long Y, Morgan C, Bancroft I, Zhao J, Meng J (2009) Unraveling the complex trait of crop yield with quantitative trait loci mapping in *Brassica napus*. Genetics 182:851-861

[105] Sinclair TR (1998) Historical changes in harvest index and crop nitrogen accumulation. Crop Sci 38: 638-643

[106] Soltani A, Rezai AM, Pour MRK (2001) Genetic variability of some physiological and agronomic traits in grain sorghum (*Sorghum bicolor* L.). J Sci Tech Agric Nat Resources 5: 127–137

[107] Song XJ, Huang W, Shi M, Zhu MZ, Lin HXA (2007) QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat Genet 39: 623-630

[108] Spielmeyer W, Ellis MH, Peter M (2002) Semidwarf (*sd-1*), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. Proc Natl Acad Sci USA 99:9043-9048

[109] Suh JP, Ahn SN, Cho YC, Kang KH, Choi IS, et al. (2005) Mapping of QTLs for yield traits using an advanced backcross population from a cross between *Oryza sativa* and *O. glaberrima*. Korean J Breed 37:214-220

[110] Sun L, Gong K (2001) Silicon-based materials from rice husks and their applications. Ind Eng Chem Res 40:5861-5877

[111] Sweeny MT, Thomson MJ, Pfeil BE, McCouch S (2006) Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. Plant Cell 18: 283-294

[112] Takahashi Y, Shomura A, Sasaki T, Yano M (2001) *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. Proc Natl Acad Sci USA 98:7922-7927

[113] Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Bio Evol 24: 1596-1599.

[114] Terao T, Nagata K, Morino K, Hirose T (2010) A gene controlling the number of primary rachis branches also controls the vascular bundle formation and hence is responsible to increase the harvest index and grain yield in rice. Theor Appl Genet 120: 875-893

[115] Thomson MJ, Polato NR, Prasetiyono, Trijatmiko KR, Silitonga TS, et al. (2009) Genetic diversity of isolated populations of Indonesian landraces of rice (*Oryza sativa* L.) collected in east Kalimantan on the island of Borneo. Rice 2:80-92

[116] Thomson MJ, Septiningsih EM, Suwardjo F, Santoso TJ, Silitonga TS, et al. (2007) Genetic diversity analysis of traditional and improved Indonesian rice (*Oryza sativa* L.) germplasm using microsatellite markers. Theor Appl Genet 114:559–568

[117] Thomson MJ, Tai TH, McClung AM, Lai XH, Hinga ME, et al. (2003) Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. Theor Appl Genet 107:479-493

[118] Tyagi A, Khurana JP, Khurana P, Raghuvanshi S, Gaur A, et al. (2004) Structural and functional analysis of rice genome. J Genet 83: 79-99

[119] Wamishe YA, Jia Y, Singh P, Cartwright RD (2007) Identification of field isolates of *Rhizoctonia solani* to detect quantitative resistance in rice under greenhouse conditions. Front Agric China 4: 361-367

[120] Wang LQ, Liu WJ, Xu Y, He YQ, Luo LJ, et al. (2007) Genetic basis of 17 traits and viscosity parameters characterizing the eating and cooking quality of rice grain. Theor Appl Genet 115: 463-476

[121] Weir BS (1996) Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sinauer Assoc., Inc., Sunderland, MA, USA

[122] Xin Z, Velten JP, Oliver MJ, Burke JJ (2003) High throughput DNA extraction method suitable for PCR. Biotechniques 34: 820-826.

[123] Xu YB, Beachell H, McCouch SR (2004) A marker–based approach to broadening the genetic base of rice in the USA. Crop Sci 44:1947–1959

[124] Xue W, Xing Y, Weng X, Zhao Y, Tang W, et al. (2008) Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. Nat Genet 143: 1-7

[125] Yan WG, Li Y, Agrama HA, Luo D, Gao F, et al. (2009) Association mapping of stigma and spikelet characteristics in rice (*Oryza sativa* L.). Mol Breed 24:277-292

[126] Yan WG, Rutger JN, Bryant RJ, Bockelman HE, Fjellstrom RG, et al. (2007) Development and evaluation of a core subset of the USDA rice (*Oryza sativa* L.) germplasm collection. Crop Sci 47: 869-878

[127] Yan WG, Rutger JN, Bockelman HE, Tai TH (2005a) Agronomic evaluation and seed stock establishment of the USDA rice core collection. In: Norman RJ et al, editor BR Wells Rice Research Studies 2004 Univ of Arkansas, Agri Exp Sta Res Ser 529: 63-68

[128] Yan WG, Rutger JN, Bockelman HE, Tai TH (2005b) Evaluation of kernel characteristics of the USDA rice core collection. In: Norman RJ et al, editor BR Wells Rice Research Studies 2004 Univ of Arkansas, Agri Exp Sta, Res Ser 529: 69-74

[129] Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38: 203-208

[130] Zeng Y, Shen S, Li Z, Yang Z, Wang X, et al. (2003) Ecogeographic and genetic diversity based on morphological characters on indigenous rice (Oryza sativa L.) in Yunnan, China. Genet Res Crop Evo 50:567-577

[131] Zhang YM, Mao Y, Xie C, Smith H, Luo L, et al. (2005) Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.). Genetics 169: 2267-2275

[132] Zhao K, Tung C, Eizenga GC, Wright MH, Ali M, et al. (2011) Genome-wide association mapping reveals rich genetic architecture of complex traits in Oryza sativa. Nature Communications. Available: DOI: 10.1038/ncomms1467.

[133] Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, et al. (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genet 19: 3(1):e4

[134] Zuo S, Yin Y, Zhang L, Zhang Y, Chen Z, et al. (2007) Breeding value and further mapping of a QTL qSB-11 conferring the rice sheath blight utilized resistance. Chin J Rice Sci 21: 136-142