A Rice Diversity Panel Evaluated for Genetic and Agro-Morphological Diversity between Subpopulations and its Geographic Distribution

M. Liakat Ali, Anna M. McClung, Melissa H. Jia, Jennifer A. Kimball, Susan R. McCouch, and Georgia C. Eizenga*

ABSTRACT

A diverse collection of 409 Asian rice (*Oryza sativa* L.) accessions originating from 79 countries was fingerprinted with 36 simple sequence repeat (SSR) markers and evaluated for 18 agro-morphological traits. Genetically, the accessions clustered into five ancestral groups (subpopulations), *indica*, *aus*, *aromatic* (Group V), *tropical japonica*, and *temperate japonica*, based on model-based structure analysis. Thirty-three accessions with less than 60% ancestry from any single group were identified as admixtures. Canonical discriminant analysis identified eight agro-morphological traits (panicle number per plant, panicle length, plant height, flag leaf width, grain length, width, length:width ratio, and volume) as the main discriminatory characters among the rice accessions and between the subpopulations. Both SSR allele- and phenotypic trait-based analyses indicated a close relationship between *aus* and *indica* and similarly between *temperate japonica* and *tropical japonica*. The *aromatic* (Group V) rice represents a distinct small group that is more closely related to *tropical japonica* based on SSR alleles but to *aus* and *indica* based on phenotype. A strong relationship between subpopulations and geographical distribution was observed. This rice diversity panel with the accompanying genetic and phenotypic information provides a valuable foundation for association mapping, understanding the basis of both genotypic and phenotypic differences within and between subpopulations, and rice improvement programs.

M.L. Ali, Univ. of Arkansas, Rice Research and Extension Center, 2900 Hwy. 130 E, Stuttgart, AR 72160; G.C. Eizenga, M.H. Jia, and A.M. McClung, USDA-ARS, Dale Bumpers National Rice Research Center, 2890 Hwy. 130 E, Stuttgart, AR 72160; J.A. Kimball and S.R. McCouch, Dep. of Plant Breeding and Genetics, 162 Emerson Hall, Cornell Univ., Ithaca, NY 14853-1901. Supplementary material is available in the online version of this article at www.crops.org/publications/cs. Received 10 Nov. 2010. *Corresponding author (georgia.eizenga@ars.usda.gov).

Abbreviations: AFLP, amplified fragment length polymorphism; AMOVA, analysis of molecular variance; ARS, Agricultural Research Service; CDA, canonical discriminant analysis; D^2, Mahalanobis distance; F_{ST}, Wright's fixation index; GRIN, Germplasm Resources Information Network; GSOR Genetic Stocks-*Oryza*; IRGC, International Rice Germplasm Collection; NSGC, National Small Grain Collection; PCR, polymerase chain reaction; PIC, polymorphism information content; RAPD, random amplified polymorphic DNA; RFLP, restriction fragment length polymorphism; SNP, single nucleotide polymorphism; SSR, simple sequence repeat.

Rice (*Oryza sativa* L.) is the third major cereal crop produced in the world after maize and wheat (FAO, 2008) and is the primary source of food for more than half of the world's human population (Kish, 2008). While the demand for rice is increasing steadily with the increase in human population, the land area available for rice production is decreasing due to urban growth and the expansion of transportation networks. New rice cultivars that combine high yield potential, good grain quality, and resistance to both biotic and abiotic stress are urgently needed to meet future consumer demands.

Published in Crop Sci. 51 (2011).
doi: 10.2135/cropsci2010.11.0641
Published online 21 June 2011.

The success of plant breeding is based on the availability of genetic variation, knowledge about desired traits, and efficient selection strategies that make it possible to exploit existing genetic resources. Understanding the population structure and range of diversity in available germplasm is fundamental for genetic improvement in all crop species. Genetic distance estimates are helpful in selecting parental combinations with sufficient genetic diversity in a segregating population (Becelaere et al., 2005) and for classifying germplasm into heterotic groups for hybrid crop breeding (Menz et al., 2004). The search for and establishment of heterotic groups can be based on geographical origin, agro-morphological traits, pedigree information, or molecular marker data (Melchinger, 1999). Before the advent of molecular genetic tools, genetic diversity was estimated from pedigree or agronomic and morphological characteristics. Estimates based on pedigree are generally inflated and often unrealistic (Cox et al., 1986; Souza and Sorrells, 1989; Almanza-Pinzon et al., 2003; Fufa et al., 2005), while phenotypic observations can provide a reasonable representation of overall genetic performance if they are based on sufficiently large sample sizes and if the traits measured show significant differences among populations (Humphreys, 1991).

Phenotypic (trait) variation can be studied using several methods including canonical discriminant analysis (CDA), a multivariate statistical technique. In CDA, unlike univariate analysis, all independent variables (traits) are considered simultaneously in the differentiation of cultivars or populations. This approach results in greater differentiation of populations compared to univariate analysis. Multivariate procedures, including CDA based on phenotypic (agronomic and morphological) characteristics, have been used in the assessment of genetic diversity in many plant species, such as perennial ryegrass (*Lolium perenne* L.) (Humphreys, 1991), tall fescue (*Festuca arundinacea* Schreb.) (Vaylay and van Santen, 2002), hairy vetch (*Vicia villosa* Roth) (Yeater et al., 2004), rice (Sanni et al., 2008; Singh et al., 2008), red clover (*Trifolium pratense* L.) (Dias et al., 2007), and groundnut (*Arachis hypogaea* L.) (Upadhyaya, 2003).

Various types of molecular markers are currently available for assessing genetic diversity in crop species. Simple sequence repeats (SSRs) are useful for analyzing the structure of germplasm collections as they are abundant, codominant, multiallelic, highly polymorphic, and chromosome-specific (Ahmed, 2002; Huang et al., 2002; Thomson et al., 2007). The high level of allelic diversity in SSRs facilitates detection of the fine structure of diversity more efficiently than an equal number of restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), or single nucleotide polymorphism (SNP) loci (Powell et al., 1996; Lorz and Wendel, 2005; Varshney et al., 2007). Simple sequence repeat markers have been extensively used in genetic diversity and population structure studies in many plants including rice (Ni et al., 2002; Garris et al., 2005; Lu et al., 2005; Eizenga et al., 2006; Thomson et al., 2007).

Asian cultivated rice (*O. sativa*) is now grown in many regions of the world. Its origin and range of diversity have attracted a great deal of interest from plant scientists. There are two distinct varietal groups of rice, often referred to as different subspecies: *indica* and *japonica*. The primary phenotypic traits that have been used to classify *indica* and *japonica* are grain shape, phenol reaction, sensitivity to potassium chlorate, leaf color, and apiculus hair length, although the variation for any individual trait may overlap between the two varietal groups (Oka, 1988). At the genome level, the *indica–japonica* divergence is readily detected using several types of molecular markers (Wang and Tanksley, 1989; Zhang et al., 1992; Garris et al., 2005; Caicedo et al., 2007). Based on only 15 isozyme markers, Glaszmann (1987) identified six varietal groups, namely *indica* (Group I), *aus* (Group II), *ashwina* (Group III), *rayada* (Group IV), *aromatic* (Group V), and *japonica* (Group VI) from a global sample of 1688 rice accessions. The *aus*, *rayada*, and *ashwina* are small groups that form part of the *indica* clade. In recent years, using SSR and SNP markers, the existence of five major groups (subpopulations) has been resolved: *indica*, *aus*, *tropical japonica*, *temperate japonica*, and *aromatic* (Group V) (Garris et al., 2005; Caicedo et al., 2007; Zhao et al., 2010); these five subpopulation groups have also been recognized in more recent genetic studies (Gross et al., 2010; Thurber et al., 2010).

Over the course of domestication *O. sativa* became highly diversified, so that today there are many different cultivars adapted to a wide range of conditions, including different water regimes, soils, topography, photoperiod regimes, and temperatures (Takahashi, 1984; IRRI, 1984; Garrity et al., 1986), and rice can be found growing in Asia, Africa, Oceania, and South, Central, and North America. A large number of landrace cultivars, developed under early agricultural conditions in Asia and Africa, were the products of both human and natural selection. Landraces evolved slowly in response to generations of stabilizing selection by farmers (Hoyt, 1988) and they are generally very heterogeneous, embodying a large amount of genetic variation both within and between populations (Oka, 1988). A recent study of genomewide SNP variation revealed an impressive amount of molecular diversity among a set of 20 diverse rice landraces and varieties (McNally et al., 2009), but there was no accompanying agro-morphological characterization of these varieties.

In the present research, we purified a diverse collection of 409 rice accessions, including both landrace and modern cultivars originating from 79 different countries, genotyped them with 36 SSR markers, and phenotyped them for 18 agro-morphological and grain characteristics. The specific goals of this study were to (i) examine the differences and relationships among the genetically defined five rice subpopulations (*aus*, *indica*, *tropical japonica*, *temperate japonica*, and

Table 1. Geographic distribution of accessions by region for each subpopulation.

Subpopulation	East Asia	Southeast Asia	South Asia	West Asia	Central Asia	Africa	Oceania	Europe	North America and the Caribbean	Central and South America	Unknown	Total
Aromatic (Group V)			4	5		1				2	3	15
Aus	1	5	45	3		2	1				2	59
Indica	34	21	15		1	6	1		3	7	2	90
Temperate japonica	37	7	4	1	4	7	2	32	6	8		108
Tropical japonica	3	15		1		18	1	3	37	19	7	104
Admixture of *indica*			2	1		6			1		1	11
Admixture of *japonica*	6	4	4		1	1	1		4	1		22
Total	81	52	74	11	6	41	6	35	51	38	14	409

aromatic [Group V]) at the genotypic and phenotypic levels, (ii) assess the genetic diversity in this collection based on SSR polymorphism, (iii) identify the phenotypic traits that discriminate individual rice accessions and the subpopulations, and (iv) examine the relationship between geographic distribution and subpopulation ancestry of Asian rice.

MATERIALS AND METHODS

Plant Materials

The 409 Asian rice (*O. sativa*) landraces and cultivars selected for genetic and agro-morphological evaluation in this study originated from 79 countries representing all the major rice growing regions of the world. The collection included 133 accessions from East and Southeast Asia, 74 from South Asia, 17 from West and Central Asia, 41 from Africa, 89 from the Americas, 35 from Europe, six from Oceania, and 14 were of unknown origin (Supplemental Table S1; summarized in Table 1). Out of the 409 accessions, 150 were a part of the genetic diversity study by Garris et al. (2005). Seed of 163 accessions, including the 150 accessions used in the Garris et al. (2005) study, were obtained from USDA-Agricultural Research Service (ARS) in Beaumont, TX, 159 accessions were from the USDA-ARS National Small Grains Collection (NSGC) rice core collection (Yan et al., 2007), and 87 were from the USDA-ARS NSGC rice collection (Yan et al., 2007). Twelve accessions were part of the *Oryza*SNP project (McNally et al., 2009), and 11 accessions were used as parental lines for the mapping populations evaluated as part of the USDA/CSREES/NRI Applied Plant Genomics Program entitled "RiceCAP: A coordinated research, education, and extension project for the application of genomic discoveries to improve rice in the United States" (USDA/CSREES grant 2004-35317-14867).

It should be noted that the 150 accessions included in the genetic study by Garris et al. (2005) were only evaluated for genetic diversity, not for phenotypic diversity, which is one of the main objectives of this study and the main reason for including this set of accessions. These accessions were also expected to provide a reference for the subpopulation classification in this study. With the hope of having a good representation of the *aus* group, a large number of accessions (74) were selected from the region in South Asia where this subpopulation is usually found. To capture maximum genetic and phenotypic diversity, and ensure representation from all rice growing regions of the world and all five subpopulations, accessions of both landraces and modern varieties were chosen from a large number of countries (79). Accessions from each country were chosen randomly from the NSGC core collection (1794 accessions), which represents the geographic and phenotypic variation in the NSGC rice collection. Seeds of accessions originating from the International Rice Germplasm Collection (IRGC) housed at the International Rice Research Institute (IRRI) in the Philippines were validated at IRRI by comparing seeds sent from the diversity panel with original seed stocks. Fourteen accessions did not match and their identity could not be determined, thus these accessions were identified by their National Science Foundation – Transgressive Variation ID numbers with an "Unknown" origin. Additional information about the accessions included in this panel (accession name, accession number, seed source, country of origin, geographic region, and subpopulation ancestry based on STRUCTURE) is given in Supplemental Table S1.

Purification

All accessions were purified for two generations (single seed descent) before DNA extraction, and DNA was extracted from single plant leaf tissue. Many of the accessions are landraces that are known to be heterogeneous in major germplasm repositories (Olufowote et al., 1997). All purified accessions in this diversity panel will be distributed as genetic stocks through the USDA-ARS Genetic Stocks-*Oryza* (GSOR) collection in Stuttgart, AR. All accessions not bound by intellectual property issues will be shared with IRRI and become part of the IRGC.

Phenotypic Evaluation

Rice accessions were planted in single-row plots of 5-m length with a spacing of 0.25 m between the plants and 0.50 m between the rows in two replications using a RCB design at Stuttgart, AR, in 2006 and 2007 (May–October). Check accessions selected based on maturity were Spring (early maturity), Cocodrie (intermediate maturity), Cybonnet (intermediate maturity), and Yang Dao 6 (late maturity). From the 2006 trial, three representative plants per replication were selected from each row for each accession and phenotypic data were recorded. Panicles from each selected plant were harvested separately. If possible, panicle and seed morphology was verified using the phenotypic descriptions and images available in the Germplasm Resources Information Network (GRIN) (USDA-ARS, 2010) and/or the T.T. Chang Genetic Resources Center of the International Rice Research Institute, Philippines. Seeds from a single representative panicle for each accession harvested in 2006

were planted in 2007. For further purification, three representative plants from each accession were sampled, as in 2006, and used for phenotyping. Data for 18 agro-morphological traits, including days to heading, plant height, flag leaf length, flag leaf width, panicle number per plant, panicle length, primary branch number per panicle, spikelets per panicle, number of unfilled grains per panicle, number of filled grains per panicle, seed length, seed width, seed length:width ratio, brown rice length, brown rice width, seed volume, 100-seed weight, and brown rice volume were recorded. The grain traits, length, width, and volume were measured on both hulled and dehulled grains using the WinSeedle Pro V2007 software and STD4800 scanner (Regent Instruments Inc., Quebec, QC, Canada).

Six accessions (Kaw Luyoeng [GSOR 301078], RTS 12 [GSOR 301127], Shuang-Chiang [GSOR 301137], Tam Cau 9A [GSOR 301150], JM70 [GSOR 301187], and FR13 A [GSOR 312015]) had very poor plant growth and five accessions (Baguamon 14 [GSOR 301382], Eh Ia Chiu [GSOR 301048], Pankhari 203 [GSOR 301390], Mudgo [GSOR 301102], and Sintane Diofor [GSOR 301139]) were photoperiod sensitive, thus these did not produce mature seed under field conditions. Several representative plants per accession were dug from the field and grown to maturity in the greenhouse with a view to harvest the seeds. None of these accessions were included in the reported statistical analysis of the phenotypic traits. (Unreported statistical analyses with data from some of these accessions included did not significantly alter the results).

A photograph of a reference panicle, along with hulled and dehulled seeds for each accession, will be available for identification purposes at the USDA-ARS GSOR web site (USDA-ARS, 2011) and at our project web site (Rice Diversity Project, 2010).

Genomic DNA Extraction

A high-throughput method (Xin et al., 2003) was followed for collecting leaf tissue and DNA extraction. A single leaf disk from a single plant of each accession was collected in a single well of a 96-well polymerase chain reaction (PCR) plate using a single hole punch from 4-wk-old seedlings. For each accession, leaf disks were sampled from five individual plants and each disk placed in the same location in each of five 96-well plates, to assess if there were any seed mixtures. Next, each well containing a leaf disk had 70 μL of Buffer A containing 2% Tween 20 and 100 mM NaOH added. The plate was incubated at 95°C for 10 min in a thermocycler with a PTC-DNA Engine (MJ Research, Waltham, MA). After incubation, 70 μL of Buffer B containing 100 mM Tris-HCl (pH 8.0) and 2 mM ethylenediaminetetraacetate (EDTA) was added to each well. This liquid solution containing DNA was suitable for PCR. For each accession, the DNA isolated from the five leaf disks in five separate 96-well plates was pooled together.

Simple Sequence Repeat Markers, Marker Amplification, and Allele Detection

A total of 36 nuclear SSR markers, further described in Supplemental Table S2, were used to fingerprint the accessions. Thirty-four SSR markers were chosen from the panel of 50 standard SSR markers recommended for rice diversity analysis by the Generation Challenge Programme sponsored by the Consultative Group on International Agricultural Research (CGIAR). The remaining two markers, RM208 (chromosome 2) and RM022 (chromosome 3), were chosen to ensure greater genome coverage for these two chromosomes. These markers represent a subset of those used by Garris et al. (2005). The genetic position of SSR markers was based on the Cornell SSR 2001 map (McCouch et al., 2002) except RM171, which was based on the CIAT SSR 2006 map (Orjuela et al., 2010). Information on the SSRs, including primer sequences, is available in the Gramene v21 database (Gramene, 2011).

Polymerase chain reaction was performed in a 25-μL reaction volume that contained 20 ng of genomic DNA, 1 U Taq DNA polymerase, 0.1% bovine serum albumin (BSA), 1% PVP40, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.5 mM $MgCl_2$, 200 μM of deoxyribonucleotide triphosphates (dNTPs), and 300 nM of each primer. For each marker, forward primers labeled with fluorescent dyes (6-FAM, VIC, HEX, or NED) were purchased from Applied Biosystems (Foster City, CA) or Integrated DNA Technologies, (Coralville, IA). The reverse primers were unlabeled. DNA was amplified under the following conditions: (i) initial denaturation at 94°C for 5 min; (ii) 35 cycles of 94°C for 1 min, 55°C to 67°C (marker dependent) for 1 min, and 72°C for 2 min; and (iii) 5 min final extension at 72°C. Polymerase chain reaction products were pooled based on color and size range of PCR fragments (three markers per run along with the ROX labeled size standard) and heated to 94°C for 5 min to denature the DNA. The samples were separated by capillary electrophoresis using an ABI Prism 3730 DNA Analyzer according to the manufacturer's instructions, and SSR fragments sizing and allele calls were performed with GeneMapper V3.7 (Applied Biosystems, Foster City, CA).

Population Structure and Cluster Analysis

Population structure was analyzed using the model-based program STRUCTURE (Pritchard et al., 2000; Falush et al., 2003) with the number of populations at $K = 5$. The analysis was run five times. Each run was performed using a burn-in of 100,000 with a run length of 10,000, and a model allowing for admixture and correlated allele frequencies. Following subpopulation structure analysis, the accessions belonging to each subpopulation group were identified and further cluster analysis was performed for the accessions included in each subpopulation using the PowerMarker V3.25 software (Liu and Muse, 2005). Pairwise genetic distances among all accessions were calculated following the Nei (1973) distance method. The neighbor-joining method (Saitou and Nei, 1987) was used for cluster analysis and phylogenetic reconstruction. MEGA V4.0 (Tamura et al., 2008) was used to view the trees. The PowerMarker software was used to calculate number of alleles, major allele frequency, gene diversity, and polymorphism information content (PIC) values. Molecular variation between and within the subpopulations and the Wright's fixation index (F_{ST}) correlation of alleles between the subpopulations were estimated using an analysis of molecular variance (AMOVA) approach in Arlequin V3.1 (Weir, 1996; Excoffier et al., 2005).

Statistical Analysis with Phenotypic Data

The SAS software (SAS Institute, 2003) was used for statistical analyses of the phenotypic data. The phenotypic differences among accessions and subpopulations (groups generated by structure analysis) were assessed via the canonical discriminant analysis using the CANDISC procedure. Canonical discriminant analysis

is a combination of principal component and canonical correlation analyses (Vaylay and van Santen, 2002). Canonical discriminant analysis derives canonical discriminant functions (linear combinations of the quantitative variables) that have the highest possible multiple correlation with groups and summarizes among-class variation in much the same way that principal component analysis summarizes total variation. Canonical variables are uncorrelated even though the measured traits may be highly correlated. In CDA, the differentiation of groups is done by taking into account the interrelationships among independent variables (measured traits) and their relationships with dependent variables (accessions) (Vaylay and van Santen, 2002). The mean value of the canonical discriminant function (variables) is referred to as group centroid. The difference between centroid values of the two groups is the Mahalanobis distance (D^2) and is calculated as:

$$D^2 = (\overline{X}_1 - \overline{X}_2)'S^{-1}(\overline{X}_1 - \overline{X}_2),$$

in which S^{-1} is the inverse of the pooled sample variance–covariance matrix and $\overline{X}_1$ and $\overline{X}_2$ are the respective vectors of the measurements on the groups 1 and 2. This measure has the distinct advantage of accounting for any correlations that might exist between the variables (Dillon and Goldstein, 1984). The D^2 statistic can be used as an indicator of the difference between populations (Loos, 1993).

An ANOVA was conducted using the ANOVA procedure to test the significance of variation among the genetically defined subpopulations for different measured phenotypic traits and among the geographic regions for different measured phenotypic traits. Using ANOVA, significant differences between pairs of groups (group means) and between pairs of geographic regions for individual phenotypic traits were identified by conducting *t* tests and estimating LSD values.

RESULTS

Allelic Diversity at Simple Sequence Repeat Loci

A total of 36 nuclear SSR markers distributed over the 12 chromosomes with an average distance between the markers of approximately 11.4 Mbp was used to genotype the 409 rice accessions. A total of 330 alleles were detected at 36 SSR loci across 409 rice accessions. The number of alleles revealed by each marker ranged from 2 (RM338) to 24 (RM154) with an average of 9.17 alleles per marker (Supplemental Table S2). The PIC value for the SSR loci ranged from 0.33 (RM338) to 0.88 (RM154) with an average of 0.63 across 36 loci and displayed a mean gene diversity of 0.68 per locus. Rare alleles, defined as those alleles with a frequency of less than 5%, accounted for an average of 4.6 alleles per locus and were identified at most loci. The frequency of the major (most common) allele at each locus ranged from 22 (RM154) to 70% (RM338). On average, 46% of the 409 accessions shared a common major allele at any given locus.

Population Structure Analysis

Population structure analysis grouped the 409 accessions into five subpopulations corresponding to *indica* (90), *aus* (59), *aromatic* (Group V) (15), *tropical japonica* (104), and *temperate japonica* (108) (Fig. 1). The ancestral composition of each accession is summarized in Supplemental Table S1. Accessions were assigned to a subpopulation when more than 60% of their inferred ancestry was derived from one of the model-based populations. Thirty-three accessions (8.0%) with less than 60% ancestry from any single group were identified as admixtures. Twenty-two of the admixtures had ancestry mainly from *temperate* and *tropical japonica* while 11 were admixtures mainly of *indica* and *aus*. Sixty-five percent of the 409 accessions (66 *indica*, 54 *aus*, 13 *aromatic*, 68 *tropical japonica*, and 67 *temperate japonica*) had more than 95% shared ancestry with their respective subpopulations.

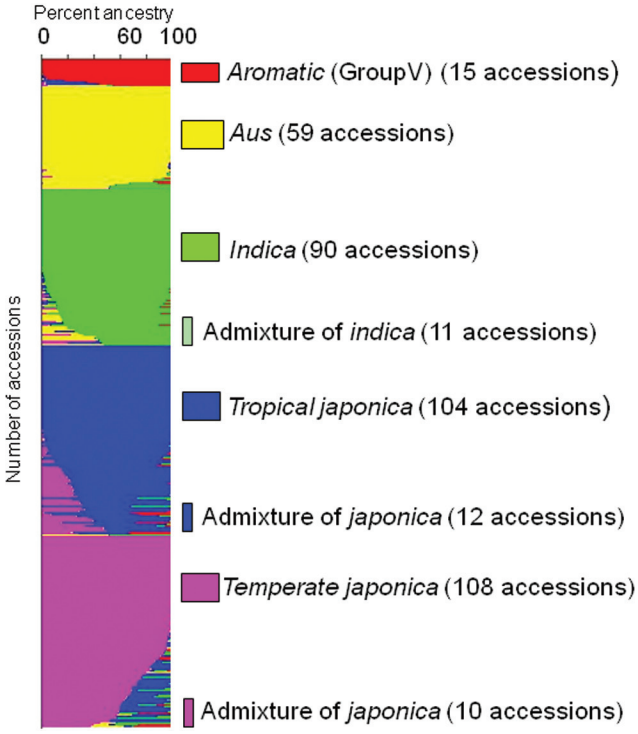

Figure 1. Model-based population structure of 409 rice accessions showing five subpopulation groups based on 36 simple sequence repeat (SSR) markers. The *y* axis represents each individual accession as a row, and the *x* axis indicates the percentage subpopulation ancestry of the given accession. The number of accessions in each group is given in parenthesis. Positions of the admixtures are indicated by the vertical solid bars.

Cluster Analysis

Cluster analysis based on 36 SSR markers using all 409 accessions was performed and the major five subpopulations clusters are visualized in a neighbor-joining tree (Supplemental Fig. S1). Cluster analysis was also performed to dissect the accessions within each subpopulation group and to visualize the relationship among accessions (Supplemental Fig. S2.1 to 2.5). Accessions with more than 40% mixed ancestry were excluded from these analyses. A close genetic relationship was observed among accessions originating from the same country or geographic region and these origins are noted on the figures.

In the *aromatic* (Group V) subpopulation, five accessions from West Asia including four from Iran (out of 6) were clustered in one subcluster (Cluster I) indicating a close genetic relationship (Supplemental Fig. S2.1). Four South Asian accessions, specifically from India and Pakistan, clustered together in another subcluster (Cluster II), indicating their close relationship.

In the *aus* subpopulation (Supplemental Fig. S2.2), 17 South Asian accessions, 12 from Bangladesh and five from India, grouped in the same subcluster (Cluster I) indicating their ancestral relationship and regional adaptation. Similarly, in another subcluster (Cluster V) of 21 accessions, 19 were from South Asia alone. Out of the 59 total *aus* accessions, 46 were from South Asia, suggesting that the *aus* cultivars are mainly grown in South Asia.

In the *indica* subpopulation, accessions were grouped in eight subclusters of varying sizes, each containing accessions originated from different geographic regions or countries (Supplemental Fig. S2.3). None of the subclusters comprised accessions originating exclusively from a single region or country. Accessions from the same region were spread over the subclusters. The highest number of accessions from a region was in Cluster V where 13 of the 18 accessions originated from East Asia (China and Taiwan) and in Cluster VI 10 of the 20 accessions originated from East Asia.

Although accessions in the *temperate japonica* group belonging to particular regions were spread over the different subclusters, within a subcluster, accessions from the same region tended to group together. In a large subgroup of 43 accessions (Cluster I), 24 of the 35 accessions originating from Europe were in Cluster I (Supplemental Fig. S2.4). Similarly, 25 accessions from East Asia were grouped in another major subgroup comprised of the 42 accessions (Cluster VIII) and 11 of these accessions originated from Japan. Also, 17 of the 19 accessions from Japan were in the *temperate japonica* subpopulation.

In the *tropical japonica* subpopulation, 21 U.S. long grain cultivars and one Caribbean cultivar grouped in one subcluster (Cluster I) (Supplemental Fig. S2.5) indicating their close genetic relationship. In Cluster II, 12 of the 19 accessions included in this group were from Southeast Asia. Eight cultivars from Indonesia (Southeast Asia) were included in the *tropical japonica* subpopulation and all were included in Cluster II. The large subcluster of 27 accessions (Cluster VIII) included accessions from four geographic regions and represented 19 countries. Four of the five accessions from Brazil grouped together in Cluster VIII. The clustering pattern in the *tropical japonica* subpopulation shows gene flow across the regions as well as the narrow gene pool characteristic of many individual breeding programs.

Genetic Diversity Between and Within Subpopulations

The overall AMOVA indicated that 48.13% of the variation was due to differences among the subpopulation groups while the remaining 51.87% due to differences within the subpopulation groups (Supplemental Table S3). Pair-wise estimates of F_{ST} values based on allelic differences revealed a high degree of differentiation among the five subpopulation groups with values ranging from 0.36 to 0.57 (Table 2). The greatest differentiation was observed between *aus* and the two *japonica* groups and also between *indica* and the two *japonica* groups. A lower level of differentiation was observed between *temperate* and *tropical japonica* and between *indica* and *aus*, while the *aromatic* (Group V) subpopulation was intermediate.

The highest genetic diversity was observed in the *indica* subpopulation (0.45) with a mean of 5.86 alleles per locus followed by *aus* (0.43) with a mean of 4.2 alleles per locus (Table 3). The *aromatic* (Group V) accessions exhibited the lowest level of genetic diversity (0.30) with an average of 2.3 alleles per locus. Both the *temperate* and *tropical japonica* groups had lower PIC values (0.29 and 0.30, respectively) compared to *aus* (0.38) and *indica* (0.42). The 90 *indica* and 59 *aus* accessions exhibited higher genetic diversity compared to the two *japonica* subgroups that had more than 100 accessions each. The *aromatic* (Group V) had the lowest genetic diversity but was represented by only 15 accessions.

Geographic Distribution of Subpopulation Accessions

As summarized in Table 1, nine of the 15 *aromatic* (Group V) accessions (60%) originated from South and West Asia, while 76% of the *aus* accessions (45 of 59) originated in South Asia. *Indica* accessions originated from across the major rice growing regions of the world with 34 from East Asia, 21 from Southeast Asia, 15 from South Asia, six from Africa, seven from Central and South America, and three from North America and the Caribbean. *Temperate japonica* cultivars were spread across the continents with predominant representations from East Asia (37) and Europe (32) and fewer accessions from Southeast Asia (7), South Asia (4), Central Asia (4), Africa (7), North America and the Caribbean (6), and Central and South America (8). *Tropical japonica* accessions were mainly from Southeast Asia (15), Africa (18), North America and the Caribbean (37), and Central and South America (19). Most of the admixture of *indica* group was from Africa (6) while the majority of the admixture of *japonica* cultivars originated from four regions, East Asia (6), Southeast Asia (4), South Asia (4), and North America and the Caribbean (4).

Canonical Discriminant Analysis

The canonical loadings (coefficients) showing the contribution of each measured trait to the total variation are presented in Table 4. The first two canonical variates accounted for 77% of the variation among accessions. The significant ($p < 0.0001$) canonical correlations between the accessions and the first canonical variate ($r_c = 0.86$) and between the accessions and the second canonical variate

Table 2. Pair-wise comparisons using Wright's fixation index (F_{ST}) values to identify the subpopulations that had a significant number of different alleles between them.

	Aus	*Indica*	*Temperate japonica*	*Tropical japonica*	Admixture of *indica*	Admixture of *japonica*
Aromatic (Group V)	0.48**	0.53**	0.53**	0.45**	0.44**	0.36**
Aus		0.38**	0.57**	0.56**	0.25**	0.45**
Indica			0.56**	0.56**	0.17**	0.46**
Temperate japonica				0.36**	0.52**	0.19**
Tropical japonica					0.49**	0.12**
Admixture of *indica*						0.34**

**Significant at the 0.01 level.

Table 3. Summary of genetic diversity in all (409) accessions and genetically defined individual subpopulations (*aus*, *indica*, *aromatic*, *temperate japonica*, and *tropical japonica*).

	All	*Aromatic* (Group V)	*Aus*	*Indica*	*Temperate japonica*	*Tropical japonica*
Sample size	409†	15	59	90	108	104
Average no. of alleles per locus	9.19 (4.87)‡	2.3 (1.1)	4.22 (2.27)	5.86 (3.11)	4.53 (2.91)	4.33 (2.72)
Average gene diversity	0.68 (0.13)	0.30 (0.2)	0.43 (0.21)	0.45 (0.25)	0.32 (0.25)	0.32 (0.25)
Average PIC§ value	0.63 (0.14)	0.26 (0.2)	0.38 (0.19)	0.42 (0.24)	0.29 (0.24)	0.30 (0.23)

†Includes admixtures.

‡Standard deviations are indicated in the parenthesis.

§PIC, polymorphism information content.

(r_c = 0.77) indicate that the canonical variates explain the phenotypic differentiation of the rice accessions. The first canonical discriminant function (canonical variate 1) explained 50% of the variance and was dominated by large loadings from the plant morphological traits—plant height, panicle number per plant, and panicle length—and the grain traits—seed and brown rice width and seed and brown rice volume. Canonical variate 2 accounted for 27% of the variation and was dominated by large loadings from the plant morphological traits— flag leaf width, and panicle number per plant—and the grain traits—seed and brown rice length, seed and brown rice width, and seed length:width ratio. We also conducted CDA analysis excluding the brown (dehulled) rice grain characteristics and also found that the same grain traits, seed length, seed width, seed length:width ratio, and seed volume made some of the largest contribution to the canonical variates.

The centroid values of the first two canonical discriminant functions for the accessions were plotted as the *x* and *y* axis to visualize the subpopulation clusters (Fig. 2). The extent of separation between the subpopulation clusters was measured by D^2 (Table 5). All pairwise distances between the subpopulation clusters were significant ($p < 0.0001$). The greatest phenotypic distance was observed between *aromatic* (Group V) and *temperate japonica* (D^2 = 27.74) followed by *aus* and *temperate japonica* (D^2 = 20.20) while smallest distance was observed between *aus* and *indica* (D^2 = 6.71) followed by *temperate* and *tropical japonica* (D^2 = 11.83). Phenotypically, *tropical japonica* was closer to both *aus* (D^2 = 13.75) and *indica* (D^2 = 13.57) than to *aromatic* (Group V) (D^2 = 19.52). The D^2 values also indicated that both *aus* and *indica* were significantly different from both *temperate* and *tropical japonica*.

Table 4. The canonical loadings of the measured traits on the first two canonical variables of the rice accessions.

Phenotypic trait (trait ontology acronym†)	Canonical variate	
	1	2
Days to heading (DTHD)	0.24	0.13
Plant height (PTHT)	0.36	0.08
Flag leaf length (FLFLG)	0.11	0.13
Flag leaf width (FLFWD)	–0.04	0.49
Panicle no. per plant (PNNB)	0.44	–0.60
Panicle length (PNLG)	0.37	0.20
Primary branch no. (PBRNB) per panicle	–0.11	0.28
Spikelets per panicle (FLNBPPN)	–0.02	0.22
Unfilled grains per panicle (UNFILGRNB)	–0.03	0.19
Filled grain no. per panicle (FILGRNB)	0.04	0.17
Seed length (HULGRLG)	0.04	0.39
Seed width (HULGRWD)	–0.45	–0.27
Seed length:width ratio (HULGRLGWDRO)	0.28	0.40
Brown rice length (DHULGRLG)	0.05	0.47
Brown rice width (DHULGRWD)	–0.46	–0.28
Seed volume (HULGRVOL)	–0.39	–0.03
100-seed weight (HHULGRWT)	–0.31	0.12
Brown rice volume (DHULGRVOL)	–0.42	–0.04
Canonical correlation	0.86	0.77
p level of significance	0.0001	0.0001
Variance accounted for, %	0.50	0.27

†The trait ontology acronyms as listed in Gramene (Ilic et al., 2007).

Phenotypic Variation between the Subpopulations Based on Least Significance Difference

The variation observed among the five subpopulations and two admixture groups was highly significant for the agro-morphological traits evaluated, such as days to heading, plant height, panicles per plant, panicle length, primary branch number per panicle, flag leaf length, flag leaf

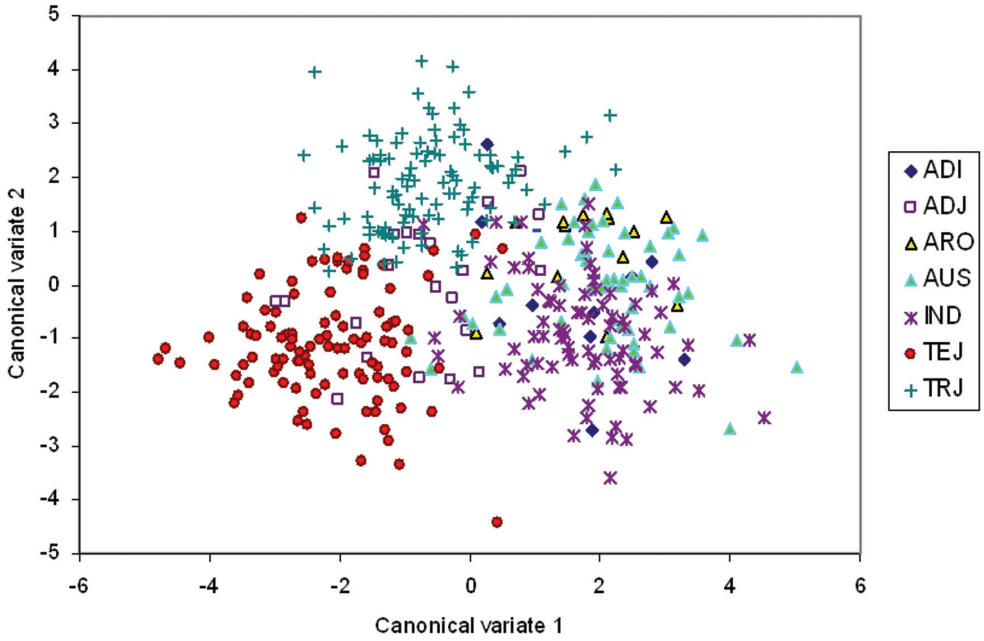

Figure 2. Scatterplot of the rice accessions belonging to different subpopulations based on the first two canonical discriminant functions (variates). Accessions are labeled with the abbreviations of their subpopulation names, which included ARO, *aromatic* (Group V) (14 accessions); AUS, *aus* (58); IND, *indica* (82); TEJ, *temperate japonica* (105); TRJ, *tropical japonica* (103); ADI, admixture of *indica* (11); and ADJ, admixture of *japonica* (22).

Table 5. Pairwise squared distance between the subpopulation clusters as calculated by Mahalanobis Distance (D^2). All distances between subpopulation clusters are significant ($p < 0.0001$). Mahalanobis distances measure the extent of phenotypic diversity between the subpopulation clusters.

Subpopulation	*Aus*	*Indica*	*Temperate japonica*	*Tropical japonica*	Admixture of *indica*	Admixture of *japonica*
Aromatic (Group V)	13.51	17.69	27.74	19.52	10.38	19.55
Aus		6.71	20.20	13.75	13.11	12.33
Indica			16.15	13.57	8.05	9.46
Temperate japonica				11.83	21.97	5.59
Tropical japonica					16.73	5.52
Admixture of *indica*						12.65

width, and spikelets per panicle and the grain traits, which included seed and brown rice length, seed and brown rice width, seed length:width ratio, seed and brown rice volume, and 100-seed weight (data not shown). The pairwise subpopulation means differences for the agro-morphological traits are presented in Supplemental Table S4 and the grain traits in Supplemental Table S5. The latest flowering accessions were classified genetically as belonging to the *indica* and admixture of *indica* groups with a mean of 94 and 101 d to heading, respectively, while the earliest flowering accessions were in the *temperate japonica* subpopulation with a mean of 79 d to flowering (Supplemental Table S4). The tallest plants were in the *aromatic* (Group V) subpopulation (143 cm) and the admixture of *indica* (142 cm) group, while the shortest were in *temperate japonica* (103 cm). *Tropical japonica* accessions had the widest flag leaves (1.4 cm) while the accessions in the *aromatic* (Group V) subpopulation had the narrowest flag leaves (0.98 cm). The mean number of panicles per plant was highest in the *indica* subpopulation (41.2) and lowest in *tropical japonica* (18.1). The longest panicles were found in the *aromatic* (Group V) and admixed *indica* accessions with a mean length of 29.4 and 29.9 cm, respectively, while *temperate japonica* had much shorter panicles (21.5 cm). *Tropical japonica* produced the highest number of spikelets as well as filled grains per panicle, whereas the *aromatic* (Group V) yielded the lowest number of spikelets and filled grains per panicle.

The seed length or brown rice grain length was similar in the admixture of *indica*, *aromatic* (Group V), and *tropical japonica* groups and were significantly longer than the mean seed and brown rice grain length from the *indica*, *aus*, and *temperate japonica* subpopulations (Supplemental Table S5). On the other hand, *temperate japonica* had the widest seeds and brown rice grains with the lowest seed length:width ratio (2.3). The *aromatic* (Group V) seeds and brown rice were the narrowest (thinnest) with the highest seed length:width ratio (3.6). Seed and brown rice volume and 100-seed weight of *tropical* and *temperate japonica* accessions were higher than those of the *indica*, *aus*, and *aromatic* (Group V) accessions.

Phenotypic Variation between the Geographic Regions Based on Least Significance Difference

The variation observed among the accessions divided into the ten geographic regions from which they originated was significant for the agro-morphological traits (days to heading, plant height, panicles per plant, panicle length, primary branch number per panicle, flag leaf length, flag leaf width, spikelets per panicle), and grain traits (seed and brown rice length, seed and brown rice width, seed length:width ratio, seed and brown rice volume, and 100-seed weight) (data not shown). The phenotypic variation between the geographic regions is summarized in Supplemental Table S6 (agro-morphological traits) and Supplemental Table S7 (grain traits). The late flowering accessions were from Southeast Asia and South Asia with a mean of 94.6 and 93.1 d to flowering, respectively, while the early flowering accessions were from Europe and Central Asia with a mean of 74.5 and 76.4 d to flowering, respectively (as evaluated in Stuttgart, AR, located at 34.5004° N and 91.5526° W). Accessions with an intermediate flowering time were grown in other regions, including North America and Caribbean, Central and South America, and Oceania. Plants grown in South Asia, West Asia, Southeast Asia, and Oceania are generally taller with a mean plant height from 122 to 131 cm while plants grown in Central Asia, Europe, North America and the Caribbean, and East Asia were generally shorter with a mean plant height of 101 to 111 cm. Those accessions originating from North America and the Caribbean, Africa, and Southeast Asia had wider flag leaves (1.28–1.31 cm) whereas accessions from in Central Asia, East Asia, and Oceania had narrower (1.00–1.10 cm) flag leaves. More panicles per plant (31–35) were produced by accessions from South Asia, East Asia, and West Asia while fewer panicles per plant (21–24) were produced by accessions from Central Asia, North America and the Caribbean, Africa, and Europe. Panicles produced by the accessions from Europe and Central Asia were significantly shorter than those produced by the accessions from other regions. Accessions from North America and the Caribbean and Southeast Asia produced more spikelets and filled grains per panicle as compared to the accessions from West Asia and Central Asia.

The seeds or brown rice produced by the accessions grown in Africa, Central and South America, and North America and the Caribbean were, in general, significantly longer than those produced by the accessions grown in East Asia, Central Asia, and South Asia (Supplemental Table S7). Accessions grown in Europe, Central Asia, and East Asia had significantly wider seed and brown rice than the seed and brown rice produced by the accessions grown in Southeast Asia, North America and the Caribbean, Oceania, and West Asia. Seed and brown rice volume in accessions from Europe, Africa, and Central Asia was larger than the seed and brown rice volume in accessions from South Asia, North America and the Caribbean, Oceania, and Southeast Asia. Based on 100-seed weight, European accessions produced the heaviest grains while South Asian accessions produced the lightest grains.

DISCUSSION

Population Structure Based on 36 Simple Sequence Repeat Markers

Thirty-four of the 36 SSR markers were chosen from the panel of 50 standard SSR markers recommended for rice diversity analysis by the Generation Challenge Programme sponsored by the CGIAR. These markers also represent a subset of the 169 SSR markers used by Garris et al. (2005) for rice population structure analysis, which identified the five ancestral subpopulations groups: *aus*, *indica*, *tropical japonica*, *temperate japonica*, and *aromatic* (Group V). With these 36 SSR markers, the 409 accessions grouped into nearly the same five subpopulations groups as reported by Garris et al. (2005) using 169 SSR markers. In fact, only six of the 150 accessions, which also were included in the Garris et al. (2005) study, did not group with the same subpopulation. These were Cuba 65 (from Cuba), JC149 (India), Kiang-Chou-Chiu (Taiwan), Gotak Gatik (Indonesia), Gyehwa 3 (South Korea), and Hagimonae Mochi (Japan). We suspect incorrect DNA was analyzed in the Garris et al. (2005) study because the seeds of these accessions were validated against original seed stocks at IRRI and the seed of these accessions matched the backup stocks retained from the aforementioned Garris et al. (2005) study. In addition, images of the seed for the first four accessions were available in GRIN and our seed matched these images. The admixture of *japonica* subpopulation group for Gyehwa 3 and Hagimonae Mochi, identified in this study, most likely is correct because all the accessions from South Korea and Japan were included in either the *temperate japonica* or the admixture of *japonica* subpopulation group.

The high level of allelic diversity in SSR markers facilitates detection of the fine structure of diversity more efficiently than an equal number of RFLP, RAPD, AFLP, or SNP loci (Powell et al., 1996; Lorz and Wendel, 2005; Varshney et al., 2007). A subset of 384 accessions from this diversity panel also grouped into these same five subpopulations groups based on genotyping with 1536 SNP markers (Zhao et al., 2010) and the subpopulation grouping of these accessions was very similar to the grouping reported in Supplemental Table S1 with 36 SSR markers. Differences were noted because Zhao et al. (2010) used a more stringent ancestry threshold of 80% to identify accessions belonging to a specific subpopulation, thus 85 accessions were identified as admixtures. (In other words, an accession with less than 80% of its ancestry in a particular subpopulation was considered an admixture.) In this study, the percent ancestry threshold was set at 60% and 33 admixtures were identified. When the ancestry threshold was set at 80% using the SSR markers, 68 accessions were identified as admixtures and these were part of the 85 accessions classified as

admixtures by Zhao et al. (2010). Based on this scenario, only 17 additional accessions were identified as admixtures using 1536 SNP markers. More importantly, none of the accessions deviated from the major subpopulation groupings between the two studies. In conclusion, this demonstrates that these 36 SSR markers are nearly as powerful as the 1536 SNP markers. The tremendous discriminatory power of SSR markers was also demonstrated recently at IRRI, in the Philippines, by McNally et al. (personal communication, 2011) when he reported that 1000 rice accessions were grouped into the five subpopulation groups, *aus*, *indica*, *tropical japonica*, *temperate japonica*, and *aromatic* (Group V), based on 45 SSR markers.

Genetic Diversity and Regional Adaptation of Subpopulations

The 409 accessions in our diversity panel represent five distinct subpopulations (*indica*, *aus*, *aromatic* [Group V], *temperate japonica*, and *tropical japonica*) that could be readily identified using 36 SSR markers. The current diversity panel included accessions from 79 countries representing all the rice growing areas in the world and revealed overall genetic diversity (average gene diversity of 0.68) similar to the diversity panel used by Garris et al. (2005) (average gene diversity of 0.70). Ni et al. (2002) reported a slightly lower genetic diversity (PIC value of 0.62) in a global rice collection that was smaller than the two mentioned above. Based on the results of the aforementioned studies using global accessions, it can be inferred that this rice diversity panel represents a large proportion of the genetic diversity that exists in Asian rice (*O. sativa*). Both *indica* and *aus* displayed higher levels of genetic diversity than *aromatic* (Group V) or *tropical* or *temperate japonica* accessions (Table 3) as observed in the diversity panel used by Garris et al. (2005). Other studies based on regional collections that included only *indica* and *japonica* accessions also report more genetic diversity in *indica* accessions than in *japonica* accessions (Gao et al., 2005; Lapitan et al., 2007; Yongwen et al., 2006). Based on DNA sequence variation, Caicedo et al. (2007) found a similar pattern of genetic diversity in that the *indica* subpopulation had the highest diversity while *temperate japonica* had the lowest. Lu et al. (2005) and Qi et al. (2009) reported similar patterns of genetic diversity, with more alleles in *indica* cultivars than in *japonica* cultivars. On the other hand, Thomson et al. (2007), genotyping a collection of nearly all Indonesian cultivars, noted a similar amount of genetic diversity in both the *indica* and *tropical japonica* subpopulations. The *indica* subpopulation occupies the largest rice growing area in the world and includes a variety of environments, soil types, and ecological conditions. The genetic diversity observed in our study is accompanied by a high level of phenotypic diversity related to agro-morphological traits and this diversity contributes to its broad adaptation.

In contrast to *indica*, the *aus* group is largely confined to South Asia and yet it exhibits a high level of genetic diversity relative to its sample size and area of distribution (Table 1). The majority of the *aus* accessions were from Bangladesh and India, where they are traditionally grown as short summer season crops under rain-fed conditions (Parsons et al., 1999). Although *aus* accessions occupy a smaller geographic area compared to *indica*, they are grown in extremely diverse agro-ecological conditions. They are grown on hilltops, in rainfed lowlands, in flood-prone and deep water areas, and in well-drained or poorly drained soils and are the source of salt, drought (Khush, 1997), and flood or submergence tolerance (Xu et al., 2006; Hattori et al., 2009). These accessions are generally tall and produce many tillers with short, slender grains in a variety of colors (Takano-Kai et al., 2009; Sweeney et al., 2007). Thus, the *aus* group is characterized by a high level of both genetic and phenotypic diversity.

Although *temperate japonica* and *tropical japonica* are genetically closely related, they are adapted to very different regions. *Temperate japonica* rice is mainly distributed in the cooler regions of East Asia, Central Asia, Europe, North America, and South America, while *tropical japonica* is grown in tropical and subtropical climates across Southeast Asia, Africa, the southern United States, the Caribbean, and South America (Table 1). Despite their wide geographical distribution, the relatively low genetic diversity in both *temperate* and *tropical japonica* may be attributed to their origins from a narrow gene pool marked by a more severe domestication bottleneck than *indica* or *aus* (Gao and Innan, 2008). *Temperate japonica* includes the predominant rice accessions from East Asia while *tropical japonica* is widely grown in Southeast Asia, especially under upland conditions in Indonesia, Thailand, and the Philippines, regions hypothesized to be the location of early domestication (Chang, 1976; Oka, 1988; Izawa, 2008; Vaughan et al., 2008; Ikehashi, 2009).

The *aromatic* (Group V) subpopulation embodied the lowest level of genetic diversity of all the subpopulations examined, but our estimate of diversity may be biased by the small sample size. Traditional *aromatic* (Group V) accessions are grown in a small geographic region in South and West Asia, mainly in India, Pakistan, and Iran. Some *aromatic* (Group V) accessions, especially the basmati types, are known to perform well in specific regions but are very photosensitive and generally have poor adaptability across a wide range of environments.

Cultivars Used as Checks in Field Evaluation

The four cultivars used as standard checks were chosen from different maturity groups: Spring (early maturity), Cocodrie (intermediate maturity), Cybonnet (intermediate maturity), and Yang Dao 6 (late maturity). The first three checks are commercial cultivars adapted to the southern rice growing area of the United States and these check cultivars flowered as

expected with Spring at 81 d, Cocodrie at 86 d, Cybonnet at 89 d, and Yang Dao 6 at 94 d. This expected flowering patterns of check cultivars indicates that the flowering behavior of the other accessions included in this study is typical for this region even though the diversity panel was only grown at one location for two seasons. Consequently, there is good probability that the other agro-morphological traits, especially related to flowering behavior, measured at Stuttgart, AR, are reflective of the true genetic potential. The means of the two seasons (environments) for all traits, agro-morphological and grain traits, were used for all statistical analyses and thus it could be assumed that these are the true genetic potential of the rice accessions used in the study.

Traits Differentiating Rice Accessions and Subpopulations

The variability in morphological characteristics is a useful tool for classifying plants and extremely important in attempting to exploit the genetic diversity found in plant genetic resources (Rogers and Fleming, 1973; Maduakor and Lal, 1989). In this study, variation was found both between individual rice accessions and among the genetically defined subpopulations. The CDA showed that various traits contributed differently to the total spectrum of phenotypic variation, as indicated by their loadings on the canonical variables. Canonical loadings measure the simple linear correlation between an original independent variable (trait) and the canonical variate. Therefore, the canonical loading reflects the variance that the measured variable shares with the canonical variate and can be interpreted in assessing the relative contribution of each variable to each canonical variate function (Cruz-Castillo et al., 1994; Sanni et al., 2008). The first two canonical variables accounted for 77% of the diversity that exists among the 409 rice accessions and eight of the 18 traits contributed largely to the diversity captured by the first two canonical variates. These main discriminatory traits included four plant morphological characters (panicle number per plant, panicle length, plant height, and flag leaf width) and four grain characters (grain length, grain weight, grain length:width ratio, and grain volume). These canonical variate functions indicate that the genetic composition of the accessions or subpopulations are most different for these morphological and grain characters. Similar results were also reported in the studies with *O. sativa* accessions grown in Africa (Guei et al., 2005; Sanni et al., 2008).

The main discriminatory traits were also found to be significantly different between pairs of subpopulations as revealed by the *t* tests (Supplemental Tables S4 and S5). *Indica* rice produced the highest number of panicles per plant while *tropical japonica* produced the lowest. *Aus* and *aromatic* (Group V) had an intermediate numbers of panicles. The longest panicles were produced by the *aromatic* (Group V) while the shortest panicles were produced by the *temperate japonica* group. *Aus*, *indica*, and *tropical japonica* produced panicles of intermediate size. *Aromatic* (Group V) rice had the tallest plants and *temperate japonica* had the shortest plants. Subpopulation groups could also be distinguished by flag leaf width, with *tropical japonica* having the widest flag leaves and *aromatic* (Group V) the narrowest. Significant variation for the main grain traits that included grain length, width, length:width ratio, and volume was observed between pairs of subpopulations (Supplemental Table S5) and the variation in such grain traits is often used to differentiate varietal groups. The two main discriminatory traits, panicle number per plant and panicle length, are important yield components, while grain dimensions determine the ability to commercialize a cultivar within a market class. Because of the relationship between traits and subpopulations, breeders typically target their search for traits of interest by screening accessions within specific subpopulations. The deep subpopulation structure also means that there is a strong genetic background effect that impacts efforts to manipulate quantitative traits when working across subpopulations in a breeding program.

Geographic Distribution of Subpopulations Based on Traits

When all the traits are considered together, it is difficult to establish a relationship between subpopulations and their geographic distribution. However, individual traits are often predictive of a relationship between subpopulation and geographic distribution. A few examples indicating these relationships are followed. Early flowering accessions were mainly from Europe and Central Asia and included most accessions classified as *temperate japonica*. On the other hand, late flowering accessions were from South Asia and Southeast Asia. Most of the South Asian accessions were *indica* or *aus*, whereas about half of the Southeast Asian accessions were *indica*. Accessions with intermediate flowering were mainly from North America and the Caribbean and Central and South America, and they were mostly *tropical japonica*. Similarly, shorter accessions were mainly from Europe and Central Asia, and most were *temperate japonica*, while taller accessions were mainly from South Asia and West Asia, with most belonging to *aus*, *aromatic*, and admixture of *indica*. Based on the number of spikelets or filled grains per panicle, a similar relationship was observed. Accessions from North America and the Caribbean and Southeast Asia had more spikelets and filled grains per panicle, with most North American and the Caribbean accessions and many Southeast Asian accessions classified as *tropical japonica*. On the other hand, accessions from West Asia, Central Asia, Europe, and East Asia had fewer spikelets and filled grains per panicle, with many West Asian accessions classified as *aromatic* and most Central Asian and European and half of East Asian accessions classified as *temperate japonica*. The grain traits did not have a clear relationship between subpopulation and geographic distribution but in general longer grains were

produced by African, Central and South American, and North American and the Caribbean accessions with most accessions classified as *tropical japonica*. On the other hand, in general, shorter grains were produced by East Asian, Central Asian, and South Asian accessions with about half of the East Asian and more than half of the Central Asian accessions classified as *temperate japonica* and most short grain South Asian accessions classified as *aus*.

Comparison on the Assessment of Diversity as Revealed by Simple Sequence Repeats and Agro-Morphological Traits

The five subpopulations (*aus*, *indica*, *aromatic* [Group V], *temperate japonica*, and *tropical japonica*) discerned by model-based population structure analysis were analyzed for differences based on agro-morphological traits using CDA, D^2, and *t* tests. The scatterplot based on the first two canonical variates (of phenotypic traits) displayed distinct clustering for most of the accessions of each genetically defined group, in agreement with the subpopulation structure defined by SSRs. Both F_{ST} and D^2 values indicated a close genetic relationship between *aus* and *indica*, and these groups were also similar for all grain traits (Supplemental Table S5) and several morphological traits, especially flag leaf length, panicle length, and filled grain number per panicle (Supplemental Table S4), and this partly explains the closer relationship between the groups and overlapping of some accessions between the two subpopulations in the scatterplot (Fig. 2). Because these two subpopulations shared similar trait values for these important traits, some accessions from *aus* and *indica* are near each other on the scatterplot. The *aromatic* subpopulation is not a distinct group on the scatterplot because the *aromatic* subpopulation was represented by only 15 accessions and some of the accessions were near *aus* or *indica* accessions because they shared similar trait values. Both F_{ST} and D^2 indicated a close relationship between *temperate japonica* and *tropical japonica*, but there were more phenotypic differences between these two subpopulations than between the *aus* and *indica* subpopulations. This is supported by the fact that for many phenotypic traits the *temperate japonica* and *tropical japonica* groups were significantly different. Overall, both the genetic and phenotypic analyses validated the conclusion that both *indica* and *aus* are distantly related to the *japonica* types.

The D^2 values supported a close relationship between *aromatic* (Group V), *aus*, and *indica* accessions based on agro-morphological traits, while F_{ST} derived from SSR data indicated that *aromatic* (Group V) was genetically distant from *indica* and more closely related to the *tropical japonica* subpopulation. Both analyses supported a unique ancestry for *aromatic* (Group V), consistent with Jain et al. (2004).

The significant differences among the five subpopulations revealed by AMOVA were supported by the ANOVA among subpopulations for the 18 phenotypic traits examined in this study. Significant phenotypic variation between pairs of subpopulations was observed for almost all traits (identified by *t* tests and LSD), underscoring the distinctive genetic identities of the various rice subpopulations and the importance of controlling for population substructure in any study that seeks to examine genetic and/or phenotypic variation in *O. sativa*.

CONCLUSIONS

This study examines the patterns of diversity found in the 409 accessions that comprise our rice diversity panel. Within *O. sativa*, variation detected as either SSR polymorphism or agro-morphological traits is partitioned into clearly defined subpopulation groups. The genetic and phenotypic information on the subpopulations and accessions provide a valuable foundation for association mapping and selecting materials for specific genetic studies or as potential parents for use in plant improvement. Traits that robustly differentiate subpopulation groups could be used as the foundation for identifying and characterizing genes that distinguish evolutionary origins or domestication histories and thus understanding the basis of both genotypic and phenotypic differences within and between subpopulations. The inference of the *percentage of ancestry* for each accession based on population structure analysis represents an added value for rice breeders.

Supplemental Information Available

Supplemental material is available free of charge at http://www.crops.org/publications/cs.

Acknowledgments

This research was supported by National Science Foundation–Plant Genome Project: "Exploring the Genetic Basis of Transgressive Variation in Rice" (Award # 0606461). The superb technical assistance of Teresa Hancock and Daniel Wood and the assistance of student interns T. Grant Brock, Chancellor Jackson, and Jason London with this research and supported by this grant, is gratefully acknowledged. The assistance of USDA-ARS technician Quynh Grunden and USDA-ARS STEP students Robert Floyd and Ben Owens is acknowledged. Authors wish to express thanks with gratitude to Dr. Kathleen Yeater, USDA-ARS-SPA Statistician, for her guidance with the statistical analysis of the phenotypic data. Assistance from USDA-ARS personnel Lorie Bernhardt (GSOR technician), Dr. Wengui Yan (rice germplasm geneticist), and Dr. Harold Bockleman (NSGC curator) in managing the rice seed stocks is acknowledged. Validation of the seed stocks was done in collaboration with Dr. Kenneth McNally, Ma. Elizabeth Naredo, and other staff members at the International Rice Research Institute, Philippines.

References

Ahmed, M. 2002. Assessment of genomic diversity among wheat genotypes as determined by simple sequence repeats. Genome 45:646–651. doi:10.1139/g02-028

Almanza-Pinzon, M.I., M. Khairullah, P.N. Fox, and M.L. Warburton. 2003. Comparison of molecular markers and coefficients of parentage for the analysis of genetic diversity among spring bread wheat accessions. Euphytica 130:77–86. doi:10.1023/A:1022310014075

Becelaere, G.V., L.L. Edward, A.H. Paterson, and P.W. Chee. 2005. Pedigree- vs. DNA marker-based genetic similarity estimates in cotton. Crop Sci. 45:2281–2287. doi:10.2135/cropsci2004.0715

Caicedo, A.L., S.H. Williamson, R.D. Hernandez, A. Boyko, A. Fledel-Alon, T.L. York, N.R. Polato, K.M. Olsen, R. Nielsen, S.R. McCouch, C.D. Bustamante, and M.D. Purugganan. 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genet. 3:1745–1756. doi:10.1371/journal.pgen.0030163

Chang, T.T. 1976. The origin, evolution, cultivation, dissemination, and diversification of Asian and African rice. Euphytica 25:435–44.1

Cox, T.S., J.P. Murphy, and D.M. Rodgers. 1986. Changes in genetic diversity in red winter wheat regions of the United States. Proc. Natl. Acad. Sci. USA 83:5583–5586. doi:10.1073/pnas.83.15.5583

Cruz-Castillo, J.G., S. Ganeshanandam, B.R. MacKay, G.S. Lawes, C.R.O. Lawoco, and D.J. Woolley. 1994. Application of canonical discriminant analysis in horticultural research. HortScience 29:1115–1119.

Dias, P.M.B., B. Julier, J.P. Sampoux, P. Barre, and M.D. Agnol. 2007. Genetic diversity in red clover (*Trifolium pretense* L.) revealed by morphological and microsatellite (SSR) markers. Euphytica 160:189–205. doi:10.1007/s10681-007-9534-z

Dillon, W.R., and M. Goldstein. 1984. Multivariate analysis methods and applications. John Wiley & Sons, New York, NY.

Eizenga, G.C., H.A. Agrama, F.N. Lee, W. Yan, and Y. Jia. 2006. Identifying novel resistance genes in newly introduced blast resistant rice germplasm. Crop Sci. 46:1870–1878. doi:10.2135/cropsci2006.0143

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin version 3.0. An integrated software package for population data analysis. Evol. Bioinform. Online 1:47–50.

Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567–1587.

FAO. 2008. Food outlook. Global market analysis. November 2008. pp 14–27. Available at http://www.fao.org/docrep/011/ai474e/ai474e05.htm (verified 25 May 2011). FAO, Rome, Italy.

Fufa, H., P.S. Baenziger, B.S. Beecher, I. Dweikat, R.A. Graybosch, and K.M. Eskridge. 2005. Comparison of phenotypic and molecular marker-based classifications of hard red winter wheat cultivars. Euphytica 145:133–146. doi:10.1007/s10681-005-0626-3

Gao, L., and H. Innan. 2008. Nonindependent domestication of the two rice subspecies, *Oryza sativa* ssp. Indica and ssp. Japonica, demonstrated by multilocus microsatellites. Genetics 179:965–976. doi:10.1534/genetics.106.068072

Gao, L., C.H. Zhang, L.P. Chang, J.Z. Jia, Z.E. Qiu, and Y.S. Dong. 2005. Microsatellite diversity within *Oryza sativa* with emphasis on indica-japonica divergence. Genet. Res. Camb. 85:1–14. doi:10.1017/S0016672304007293

Garris, A.J., T.H. Tai, J.R. Coburn, S. Kresovich, and S. McCouch. 2005. Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1931–1638.

Garrity, D.P., L.R. Olderman, R.A. Morris, and D. Lenka. 1986. Rainfed lowland rice ecosystems: Characterization and distribution. p. 3–23. *In* Progress in rainfed lowland rice. International Rice Research Institute, Manila, Philippines.

Glaszmann, J.C. 1987. Isozymes and classification of Asian rice varieties. Theor. Appl. Genet. 74:21–30. doi:10.1007/BF00290078

Gramene. 2011. Microsatellite markers. Gramene (release # 33, April 2011). Available at http://www.gramene.org//markers/microsat/ (verified 25 May 2011). Cold Spring Harbor Laboratory, Laurel Hollow, NY, and Cornell University, Ithaca, NY.

Gross, B.L., M. Reagon, S.H. Hsu, A.L. Caicedo, Y. Jia, and K.M. Olsen. 2010. Seeing red: The origin of pigmentation in US weedy rice. Mol. Ecol. 10.1111/j.1365–294X.04707.x.

Guei, R.G., K.A. Sanni, F.J. Abamu, and I. Fawole. 2005. Genetic diversity of rice (*Oryza sativa* L.). Agronomie Africaine 5:17–28.

Hattori, Y., K. Nagai, S. Furukawa, X.J. Song, R. Kawano, H. Sakakibara, J. Wu, T. Matsumoto, A. Yoshimura, H. Kitano, M. Matsuoka, H. Mori, and M. Ashikari. 2009. The ethylene response factors SNORKEL1 and SNOKKEL2 allow rice to adapt to deep water. Nature 460:1026–1030. doi:10.1038/nature08258

Hoyt, E. 1988. Conserving the wild relatives of crops. IBPGR/IUCN/WWF, Rome, Italy.

Huang, H.Q., A. Borner, M.S. Roder, and M.W. Ganal. 2002. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor. Appl. Genet. 105:699–707. doi:10.1007/s00122-002-0959-4

Humphreys, M.O. 1991. A genetic approach to the multivariate differentiation of perennial ryegrass (*Lolium perenne* L.) cultivars. Heredity 66:437–443. doi:10.1038/hdy.1991.53

Ikehashi, H. 2009. Why are there Indica type and Japonica type in rice? – History of the studies and a view for origin of two types. Rice Sci. 16:1–13. doi:10.1016/S1672-6308(08)60050-5

Ilic, K., E.A. Kellogg, P. Jaiswal, F. Zapata, P.F. Stevens, L.P. Vincent, S. Avraham, L. Reiser, A. Pujar, M.M. Sachs, N.T. Whitman, S.R. McCouch, M.L. Schaeffer, D.H. Ware, L.D. Stein, and S.Y. Rhee. 2007. The plant structure ontology, a unified vocabulary of anatomy and morphology of a flowering plant. Plant Physiol. 143:587–599. doi:10.1104/pp.106.092825. Also available at http://www.gramene.org/plant_ontology/index.html (verified 20 May 2011).

IRRI. 1984. Terminology for rice growing environments. International Rice Research Institute, Manila, Philippines.

Izawa, T. 2008. The process of rice domestication: A new model based on recent data. Rice 1:127–134. doi:10.1007/s12284-008-9014-7

Jain, S., R. Jain, and S. McCouch. 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. doi:10.1007/s00122-004-1700-2

Kish, S. 2008. Determining rice gene function. Available at http://www.csrees.usda.gov/newsroom/impact/2009/nri/02171_rice_gene.html (verified 25 May 2011). USDA-CSREES, Washington, DC.

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

Lapitan, V.C., D.S. Brar, T. Abe, and E.D. Redona. 2007. Assessment of genetic diversity of Philippine rice cultivars carrying good quality traits using SSR markers. Breed. Sci. 57:263–270. doi:10.1270/jsbbs.57.263

Liu, K., and S.V. Muse. 2005. PowerMarker: Integrated analysis environment for genetic marker data. Bioinformatics 21:2128–2129.

Loos, B.P. 1993. Morphological variation in *Lolium* (Poaceae) as a measure of species relationships. Plant Syst. Evol. 188:87–89. doi:10.1007/BF00937838

Lorz, H., and G. Wenzel. 2005. Molecular marker systems in plant breeding and crop improvement. Springer, Heidelberg, Germany.

Lu, H., M.A. Redus, J.R. Coburn, J.N. Rutger, S.R. McCouch, and T.H. Tai. 2005. Population structure and breeding patterns of 145 US rice cultivars based on SSR marker analysis. Crop Sci. 45:66–76. doi:10.2135/cropsci2005.0066

Maduakor, H.O., and R. Lal. 1989. Root system top-growth and yield of cassava (*Manihot esculenta* Crantz) as affected by plant population in an arid soil. J. Root Crops 15:115–122.

McCouch, S., L. Teytelman, Y. Xu, K. Lobos, K. Clare, M. Walton, B. Fu, R. Maghirang, Z. Li, Y. Xing, Q. Zhang, I. Kono, M. Yano, R. Fjellstrom, G. DeClerck, D. Schneider, S. Cartinhour, D. Ware, and L. Stein. 2002. Development of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res. 9:199–207. doi:10.1093/dnares/9.6.199

McNally, K.L., K.L. Childs, R. Bohnert, R.M. Davidson, K. Zhao, V.J. Ulat, G. Zeller, R.M. Clark, D.R. Hoen, T.E. Bureau, R. Stokowski, D.G. Ballinger, K.A. Frazer, D.R. Cox, B. Padhukasahasram, C.D. Bustamante, D. Weigel, D.J. Mackill, R.M. Bruskiewich, G. Rätsch, C.R. Buell, H. Leung, and J.E. Leach. 2009. Genome-wide SNP variation reveals relationships among landraces and modern varieties of rice. Proc. Natl. Acad. Sci. USA 106:12273–12278. doi:10.1073/pnas.0900992106

Melchinger, A.E. 1999. Genetic diversity and heterosis. p. 99–118. *In* The genetics and exploitation of heterosis in crops. ASA, CSSA, SSSA, Madison, WI.

Menz, M.A., R.R. Klein, N.C. Unruh, W.L. Rooney, P.E. Klein, and J.E. Muller. 2004. Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. Crop Sci. 44:1236–1244. doi:10.2135/cropsci2004.1236

Nei, M. 1973. The theory and estimation of genetic distance. p. 45–54. *In* N.E. Morton (ed.) Genetic structure of populations. Univ. Press of Hawaii, Honolulu, HI.

Ni, J., P.M. Colowit, and D.J. Mackill. 2002. Evaluation of genetic diversity in rice subspecies using microsatellite markers. Crop Sci. 42:601–607. doi:10.2135/cropsci2002.0601

Oka, H.I. 1988. Origin of cultivated rice. Japan Sci. Soc. Press, Tokyo, Japan.

Olufowote, J.O., Y. Xu, X. Chen, W.D. Park, H.M. Beachell, R.H. Dilday, M. Goto, and S.R. McCouch. 1997. Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RELP markers. Genome 40:370–378. doi:10.1139/g97-050

Orjuela, J., A. Garavito, M. Bouniol, J.D. Arbelaez, L. Moreno, J. Kimball, G. Wilson, J.F. Rami, J. Tohme, S.R. McCouch, and M. Lorieux. 2010. A universal core genetic map for rice. Theor. Appl. Genet. 120:563–572. doi:10.1007/s00122-009-1176-1

Parsons, B.J, H.J. Newbury, M.T. Jackson, and B.V. Ford-Lloyd. 1999. The genetic structure and conservation of aus, aman and boro rices from Bangladesh. Genet. Resour. Crop Evol. 46:587–598. doi:10.1023/A:1008749532171

Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalsky. 1996. The comparison on RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breed. 2:225–238. doi:10.1007/BF00564200

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Interference of population structure using multilocus genotype data. Genetics 155:945–965.

Qi, Y., H. Zhang, D. Zhang, M. Wang, J. Sun, L. Ding, F. Wang, and Z. Li. 2009. Assessing *indica-japonica* differentiation of improved rice varieties using microsatellites markers. J. Genet. Genomics 36:305–312. doi:10.1016/S1673-8527(08)60119-8

Rice Diversity Project. 2010. Seed photo library. Available at http://www.ricediversity.org/photolibrary/ (verified 25 May 2011), Cornell University, Ithaca, NY.

Rogers, D.J., and H.S. Fleming. 1973. A monograph of *Manihot esculenta* with exploitation of taximetric methods used. Econ. Bot. 27:5–113. doi:10.1007/BF02862222

Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

Sanni, K.A., I. Fawole, R.G. Guei, D.K. Ojo, E.A. Somado, D.T. Tia, S.A. Ogunbayo, and I. Sanchez. 2008. Geographical patterns of phenotypic diversity in *Oryza sativa* landrace of Cote d'Ivoire. Euphytica 160:389–400. doi:10.1007/s10681-007-9549-5

SAS Institute. 2003. The SAS system for windows. Release 9.1. SAS Institute, Cary, NC.

Singh, S., S.K. Pradhan, and P. Virk. 2008. Genetic divergence in new plant type rice under shallow lowland ecosystem. SABRAO J. 40:1–8.

Souza, E., and M.E. Sorrells. 1989. Pedigree analysis of North American oat cultivars released from 1951 to 1985. Crop Sci. 29:595–601. doi:10.2135/cropsci1989.0011183X002900030008x

Sweeney, M.T., M.J. Thomson, Y.G. Cho, Y.J. Park, S.H. Williamson, C.D. Bustamante, and S.R. McCouch. 2007. Global dissemination of a single mutation conferring white pericarp in rice. PLoS Genet. 3:e133. doi:10.1371/journal.pgen.0030133

Takahashi, N. 1984. Seed germination and seedling growth. Chapter 3. p. 71–88. *In* S. Tsunoda and N. Takahashi (ed.) Biology of rice, Vol. 7. Elsevier, Amsterdam, The Netherlands.

Takano-Kai, N., H. Jiang, T. Kubo, M. Sweeney, T. Matsumoto, H. Kanamori, B. Padhukasahasram, C.D. Bustamante, A. Yoshimura, A.K. Doi, and S. McCouch. 2009. Evolutionary history of GS3, a gene conferring grain length in rice. Genetics 182:1323–1334. doi:10.1534/genetics.109.103002

Tamura, K., J. Dudley, N. Masatoshi, and K. Sudhir. 2008. The MEGA V4.0 (molecular evolutionary genetic analysis program). Biodesign Institute, Arizona State University, AZ.

Thomson, M.J., E.M. Septiningsih, F. Suwardjo, T.J. Santoso, T.S. Silitonga, and S.R. McCouch. 2007. Genetic diversity analysis of traditional and improved Indonesian rice (*Oryza sativa* L.) germplasm using microsatellite markers. Theor. Appl. Genet. 114:559–568. doi:10.1007/s00122-006-0457-1

Thurber, C.S., M. Reagon, B.L. Gross, K.M. Olsen, Y. Jia, and A.L. Caicedo. 2010. Molecular evolution of shattering loci in U.S. weedy rice. Mol. Ecol. 10.1111/j.1365–294X.2010.04708.x.

Upadhyaya, H.D. 2003. Phenotypic diversity in groundnut (*Arachis hypogaea* L.) core collection assessed by morphological and agronomic evaluations. Genet. Resour. Crop Evol. 50:539–550. doi:10.1023/A:1023980713848

USDA-ARS. 2010. Rice. Available at http://www.ars-grin.gov/cgi-bin/npgs/html/crop.pl?75 (verified 25 May 2011). USDA, Agricultural Research Service, National Plant Germplasm System, Beltsville, MD.

USDA-ARS. 2011. Genetic stocks – *Oryza* (GSOR) collection. Available at http://www.ars.usda.gov/Main/docs.htm?docid=8318 (verified 25 May 2011). USDA-Agric. Res. Serv., Dale Bumpers

National Rice Research Center, Stuttgart, AR.

Varshney, R.K., K. Chabane, P.S. Hendre, R.K. Aggarwal, and A. Graner. 2007. Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. Plant Sci. 173:638–649. doi:10.1016/j.plantsci.2007.08.010

Vaughan, D.A., B.R. Lu, and N. Tomooka. 2008. Was Asian rice (*Oryza sativa*) domesticated more than once? Rice 1:16–24. doi:10.1007/s12284-008-9000-0

Vaylay, R.E., and E. van Santen. 2002. Application of canonical discriminant analysis for the assessment of genetic variation in tall fescue. Crop Sci. 42:534–539. doi:10.2135/cropsci2002.0534

Wang, Z.Y., and S.D. Tanksley. 1989. Restriction fragment length polymorphism in *Oryza sativa* L. Genome 32:1113–1118.

Weir, B.S. 1996. Genetic data analysis II: Methods for discrete population genetic data. Sinauer Associates, Sunderland, MA.

Xin, Z., J.P. Velten, M.J. Oliver, and J.J. Burke. 2003. High-throughput DNA extraction method suitable for PCR. Biotechniques 34:820–826.

Xu, K., X. Xu, T. Fukao, P. Canlas, R.M. Rodriguez, S. Heuer, A.M. Ismail, J.B. Serres, P.C. Ronald, and D.J. Mackill. 2006. Sub1A is an ethylene-response-like gene that confers submergence tolerance to rice. Nature 442:705–708. doi:10.1038/nature04920

Yan, W.G., J.N. Rutger, R.J. Bryant, H.E. Bockelman, R.G. Fjellstrom, M.G. Chen, T.H. Tai, and A.M. McClung. 2007. Development and evaluation of a core subset of the USDA rice (*Oryza sativa* L.) germplasm collection. Crop Sci. 47:869–878. doi:10.2135/cropsci2006.07.0444

Yeater, K.M., G.A. Bollero, D.G. Bullock, A.L. Rayburn, and S. Rodriguez-Zas. 2004. Assessment of genetic variation in hairy vetch using canonical discriminant analysis. Crop Sci. 44:185–189. doi:10.2135/cropsci2004.0185

Yongwen, Q.I., Z. Dongling, Z. Hongliang, W. Meixing, S. Junli, W. Xinghua, Q. Zongen, T. Shengxian, W. Xiangkun, and L. Zichao. 2006. Genetic diversity of rice cultivars (*Oryza sativa* L.) in China and the temporal trends in recent fifty years. Chin. Sci. Bull. 51(6):681–688. doi:10.1007/s11434-006-0681-8

Zhang, Q.F., M.A.S. Maroof, T.Y. Lu, and B.Z. Shen. 1992. Genetic diversity and differentiation of *Indica* and *Japonica* rice detected by RFLP analysis. Theor. Appl. Genet. 83:495–499.

Zhao, K., M. Wright, M. Kovach, A. Reynolds, W. Tyagi, J. Kimball, G. Eizenga, A. McClung, M.L. Ali, C.D. Bustamante, and S. McCouch. 2010. Genomic diversity and introgression in *O. sativa* reveal impact of domestication and breeding on the rice genome. PLoS ONE 5(5):E10780 10.1371/journal.pone.0010780. doi:10.1371/journal.pone.0010780