Geographic Description of Genetic Diversity and Relationships in the USDA Rice World Collection

WenGui Yan,* Hesham Agrama, Melissa Jia, Robert Fjellstrom, and Anna McClung

ABSTRACT
Knowledge of genetic diversity and relationships among global germplasms of rice collected by the USDA since 1866 is critical for utilization, conservation, and management of the collection. A core subset developed from this collection, including 1794 accessions obtained from 112 countries in 14 geographic regions, was genotyped with one indel and 71 simple sequence repeat (SSR) markers. Total alleles were 1005, averaging 14 alleles per locus. A great majority of genetic variance was due to within instead of among geographic regions and within instead of among countries. The regions and countries were highly and significantly differentiated using these markers. Germplasm accessions obtained from the southern Asia, Southeast Asia, and Africa were highly diversified, while those from North America and western and eastern Europe had the lowest diversity. Different measurements of genetic diversity, including average number of alleles per locus, polymorphism information content (PIC), Nei index, and average number of private alleles per locus uniformly reached this conclusion. Three main clusters were revealed by both analyses of principal coordinates and Nei genetic similarities for the 14 regions. Seventy-eight countries, from which five or more accessions were collected in the core subset, were differentiated into five main clusters. Germplasm accessions obtained from Myanmar, Indonesia, Cambodia, Malaysia, and Nepal were highly diversified, while those from France, Spain, Romania, Italy, and the United States were poorly diversified. This study proves that the USDA global collection effectively supports the U.S. rice industry with vast genetic diversity responsive to biotic and abiotic stresses.

Molecular marker are a powerful tool for the characterization and evaluation of genetic diversity, population structure, and relationship in plant germplasm collections and have been applied to cultivated rice (*Oryza sativa* L.) (Garris et al., 2005; Alvarez et al., 2007; Zeng et al., 2007; Thomson et al., 2007, 2009; Jayamani et al., 2007; Agrama et al., 2007; Agrama and Eizenga, 2008), weedy rice (*Oryza sativa* L.) (Gealy et al., 2009; Cao et al., 2006), bread wheat (*Triticum aestivum* L.) and wild emmer wheat [*Triticum turgidum* L. subsp. *dicoccoides* (Körn. ex Asch. & Graebn.) Thell.] (Balfourier et al., 2007; Roussel et al., 2004; Peleg et al., 2008), maize (*Zea mays* L.) (Hartings et al., 2008), soybean [*Glycine max* (L.) Merr.] (Li et al., 2008), robusta coffee (*Coffea canephora* Pierre ex A. Froehner) (Prakash et al., 2005), and *Globularia bisnagarica* (Honnay et al., 2007). Microsatellite (simple sequence repeat, or SSR) markers are particularly powerful and have become the markers of choice in recent years for germplasm diversity studies. The SSRs are abundant and dispersed throughout the entire genome and have high information content, codominant inheritance, reproducibility, and

Published in Crop Sci. 50:2406–2417 (2010).
doi: 10.2135/cropsci2010.02.0096
Published online 27 Sept. 2010.
© Crop Science Society of America | 5585 Guilford Rd., Madison, WI 53711 USA
All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.
locus specificity (Ellis and Burke, 2007). The high level of allelic polymorphism of SSRs facilitates the detection of genetic structure of diversity more efficiently than an equal number of restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), or single nucleotide polymorphism (SNP) markers (Garris et al., 2005).

Rice is one of the most important food crops because it feeds more than half of the world’s population (Khush, 1997). Molecular markers have been used to study genetic diversity and interrelationships for rice cultivars in the United States (Mackill, 1995; Xu et al., 2004; Garris et al., 2005; Caicedo et al., 2007), China (Zeng et al., 2007; Zhang et al., 2009), Indonesia (Thomson et al., 2007, 2009), Portugal (Jayamani et al., 2007), and Cuba (Alvarez et al., 2007). These studies focused on different areas: either a group of cultivars of specific interest, or germplasm accessions from a certain region, or a portion of a germplasm collection. Garris et al. (2005) studied 234 accessions genotyped with 169 nuclear SSRs and two chloroplast markers. Alvarez et al. (2007) conducted research on 50 Cuban accessions using 40 SSRs. Zeng et al. (2007) analyzed 692 core accessions in Yunnan, China, with 20 SSRs. Thomson et al. (2007) genotyped 330 Indonesian accessions with 30 SSRs, while Jayamani et al. (2007) genotyped 176 accessions from 19 countries in the Portuguese rice collection with 24 SSRs. To our knowledge, no study has been conducted on an entire germplasm collection including samples from almost every country in the world where rice is grown. Soybean (Li et al., 2008) and rice (Zhang et al., 2009) in China are the only collections assessed molecularly, but the two collections are limited to a single country. We studied 1794 accessions as a complete representation of the USDA rice world collection derived from 112 countries or districts in fourteen geographic regions using 71 SSRs plus an indel marker with an even coverage of the genome at a density of 30 cM.

The U.S. National Plant Germplasm System (NPGS), coordinated by the U.S. Department of Agriculture (USDA), Agricultural Research Service (ARS), is one of the largest germplasm systems (Bretting, 2007) and presently safeguards 535,202 accessions of 13,426 plant species (http://www.ars-grin.gov/npgs/stats/summary.html; verified 17 July 2010). Most of the accessions were acquired internationally at a time when genetic resources were freely exchanged. Because of computerized databases and communication, as well as reliable and rapid long-distance transport, the distribution volume of NPGS samples has expanded steadily to a current average of about 120,000 per year with about a quarter of these being distributed internationally free of charge or restriction. The rice component of the USDA NPGS is maintained by the National Small Grains Collection (NSGC) and includes over 18,000 accessions representing 12 Oryza species with 99% being O. sativa, the predominant cultivated species (National Plant Germplasm System, 2009). Although the NPGS rice germplasm collection started in 1866, the earliest accession ‘Ostiglia’ that remains today was collected from Italy in 1904 (Bockelman et al., 2003). Accessions in the NPGS rice collection were introduced from 115 countries, making it second only to the collection (130 countries) at the International Rice Research Institute (IRRI), the Philippines, in geographic diversity (K. McNally, personal communication, 2009).

The advantages of using a core subset strategy for characterizing and managing a large germplasm collection have been widely recognized (Grenier et al., 2001; Spagnoletti Zeuli and Quase, 1993) since it was first proposed by Brown in 1989 (Brown, 1989). The USDA rice core collection assembled with a stratified random sampling method has been described by Yan et al. (2004a) and proved to be representative of the whole collection with 88% certainty (Yan et al., 2007). Originally the entries in the core collection numbered 1801 from 115 countries (Yan et al., 2004a) but were reduced to 1790 from 114 countries because some entries lost viability (Yan et al., 2007) and then increased to 1794 (Agrama et al., 2009) after adding four reference cultivars that had been sequenced by the International Rice Functional Genomics Consortium (IRFGC) (McNally et al., 2006). The countries become 112 when origins of “Unknown” and “Uncertain” are not taken into account. This core collection has been characterized for 26 phenotypic traits including straighthead resistance (Yan et al., 2004b) with associated quantitative trait loci (QTL) (Agrama and Yan, 2009), agronomic (Yan et al., 2005a), and kernel (Yan et al., 2005b) characteristics.

The objective of this study was to characterize genetic diversity and relationships in the USDA rice world collection using the core subset of 1794 accessions and 72 genome-wide SSR markers. Information resulting from this study will help utilize, conserve, and manage the collection for more efficient service to international rice research communities.

**MATERIALS AND METHODS**

In the USDA rice core collection of 1794 accessions, nine have either “Unknown” or “Uncertain” origin, thus were replaced by either frequently used genetic materials ('Jasmine 85', 'Lemont', and 'Teqing' for sheath blight, 'Jing 185-7' and 'Shufeng 109' for straighthead, and 'Zhe 733') or popularly grown cultivars ('Cocodire', 'Francise', and 'Wells') because this study deals with geographic issues. Teqing, Jing 185-7, Shufeng 109, and Zhe 733 originated in China and the remainings are U.S. cultivars. In 2006, single plant selection was made to remove “heterogeneity” for each accession at Dale Bumpers National Rice Research Center (DBNRRRC) near Stuttgart, AR. Twenty plants of each accession were individually grown in...
0.09 m² spacing. At maturity, a single representative plant from each accession was selected based on previously recorded agronomic data (Yan et al., 2005a) and descriptions in the Germplasm Resources Information Network (GRIN) (http://www.ars-grin.gov/; verified 17 July 2010). Seeds harvested from the selected plant were planted in the greenhouse and leaf tissue was collected for DNA extraction.

Total genomic DNA was extracted using a rapid alkali extraction procedure (Xin et al., 2003) from a bulk of five plants derived from the selected plant, representing each entry in the core collection. Seventy-two molecular markers, covering the entire rice genome, with an average of one marker per 30 cM, were used to genotype the 1794 accessions. Of the 71 SSRs, 69 were obtained from the Gramene (http://www.gramene.org/; verified 17 July 2010), and two, AP5652-1 and AP5652-2, were developed in-house from BAC AP5652. One indel marker at the Rf locus (Rid 12; Brooks et al., 2008) was also included. Fifty of these SSRs are in common with those being used by the Generation Challenge Program for rice diversity analysis (http://www.gramene.org/markers/microsat/50_ssr.html; verified 17 July 2010). The remaining 21 SSRs were arranged to fill the gaps on the rice genome for an even coverage. Polymerase chain reaction (PCR) amplification of the markers is described by Agrama et al. (2009). DNA samples were separated on an ABI Prism 3730 DNA analyzer according to the manufacturer’s instructions (Applied Biosystems, Foster City, CA). Fragments were sized and binned into alleles using GeneMapper version 3.7 software (Applied Biosystems, Foster City, CA).

The 112 countries or districts from which the 1794 accessions originated were classified into 14 geographic regions according to groupings of the United Nations Statistic Division (2009) (Table S1). Each accession was plotted on the global map using its latitude and longitude coordinates according to the GRIN passport database when available. If the coordinate is not available, it was inferred by the “State” or “Province” from which the accession was collected. Otherwise, the inference was based on the capital city of the origin country for an accession. The map was built using the “prcomp” procedure in the statistics module (version 2.8.1) of the R statistical package (http://cran.r-project.org/web/packages/RStatsPack.pdf; verified 4 Aug. 2010) including “spatial,” “maps,” and “fields” (Venables and Ripley, 1998; Venables et al., 2008).

PowerMarker software (Liu and Muse, 2005) was used to calculate allele frequencies and polymorphism information content (PIC) values (Botstein et al., 1980) for each marker, region, and country. On the country level, genetic diversity analysis was only applied to those countries from which five or more accessions were collected. Thirty-four countries were excluded for the analysis because they have less than five accessions. Analysis of molecular variance (AMOVA; Excoffier et al., 1992) was conducted for variance components within and among regions and countries of origin, respectively, using ARLEQUIN 3.0 software (Schneider et al., 2000). The AMOVA-derived Φst [the correlation of random haplotypes within populations, relative to that of random pairs of haplotypes drawn from the whole species (Excoffier et al., 1992)] (Weir and Cockerham, 1984) is analogous to Wright’s F statistics differing only in their assumption of heterozygosity (Paun et al., 2006). Φst provides an effective estimate of the amount of genetic divergence or structuring among populations (Excoffier et al., 1992). Significance of variance components was tested using a nonparametric procedure based on 1000 random permutations of individuals using the software ARLEQUIN 3.0 (Schneider et al., 2000). Genetic diversity was estimated using Nei diversity index for each accession according to Lynch and Milligan (1994). Geographical distribution of diversity index represented by Kriging methods was globally mapped using the R-script (François et al., 2008).

Principal coordinate analysis was conducted using GenAlex 6.1 (Peakall and Smouse, 2006) for 1794 accessions genotyped by 72 markers among 14 regions. Genetic relationships among accessions represented by regions and countries were determined by the unweighted pair-group method with an arithmetic mean (UPGMA) analysis based on Nei (Nei, 1973) genetic similarity estimated using the 72 markers. The clustering trees were constructed from 1000 bootstrap replicates using the software PowerMarker (Liu and Muse, 2005) and drawn with MEGA v. 3.1 (Kumar et al., 2004).

The number of alleles private to a population, which does not exist in other populations, is especially informative when populations are studied with highly variable multiallelic markers, such as SSRs (Szpiech et al., 2008). Estimation of the private alleles depends heavily on sample size, which can be difficult to interpret when sample sizes differ across populations. The rarefaction approach has become an important strategy for producing estimates that are comparable in different populations (Hurlbert, 1971; Kalinowski, 2004). For a standardized size “g,” populations are compared by considering the estimates of “private allelic richness” that could be calculated when averaging across all subsamples of size “g.” The estimated private allelic richness is the number of private alleles expected in the population when random subsamples of size “g” are taken from each population under consideration (Kalinowski, 2005). The average number of private alleles per locus for core accessions originating in each of 14 geographic regions was estimated using ADZE (Allelic Diversity AnalyZEr) software (Szpiech et al., 2008) with the 72 molecular markers.

RESULTS
Allelic Diversity
A total of 1005 alleles were revealed by 72 molecular markers, averaging 14 alleles per locus and ranging from 2 for RM338 to 36 for RM11229 (Table S2). Polymorphic information content (PIC) measures the probability that two randomly chosen alleles from a population are distinguished. PIC values averaged 0.66 ± 0.02 ranging from 0.17 (AP5625-1) to 0.92 (RM11229) with the majority distributed between 0.50 and 0.90. Sixty markers (83%) were highly informative (PIC > 0.50), 10 (14%) reasonably informative (0.50 > PIC > 0.25), and 2 (3%) slightly informative (PIC < 0.25), demonstrating a high discriminatory power of these selected markers.

The 1794 accessions in the USDA rice core collection were introduced from 112 countries and distributed to 14 worldwide geographic regions (Fig. 1). Africa had 26 countries, the most among the regions, while North...
America included only two countries, the United States and Mexico (Table 1). The number of germplasm accessions for a region ranged from 57 in Oceania to 224 in South America. China was the country represented by the most accessions (135), while 34 countries had less than five accessions each. Most of these underrepresented countries (14 or 41%) were in Africa (Yan et al., 2004a). Thus, these 34 countries were not included in the genetic diversity analysis on a country basis but were included in the geographic region analysis. Analysis of molecular variance (AMOVA) showed that the majority (89%) of total genetic variance was due to differences within regions and the rest (11%) was due to variance among regions (Table 2). Genetic variations were characterized by significant differentiation among regions ($\Phi_{st} = 0.11, p < 0.001$) and by very high and significant differentiation within regions ($\Phi_{st} = 0.89, p < 0.001$). Likewise, when countries were taken into account, 82% of the total variation was due to the differences within countries, and the remaining portion of the variance was equally shared by both among regions and among countries. Genetic variations were significantly differentiated among regions ($\Phi_{st} = 0.10, p < 0.001$) and among countries ($\Phi_{st} = 0.12, p < 0.001$) and very highly and significantly differentiated within countries ($\Phi_{st} = 0.85, p < 0.001$).

### Genetic Diversity and Genetic Relationships among Geographic Regions

Rice accessions collected from southern Asia had the most number of alleles per locus before and after standardizing the sample size in each region, followed by Africa, Southeast Asia, China, South America, South Pacific, and Central America, while those in western and eastern Europe, North America, and Central Asia had the least (Table 1). As demonstrated by the PIC value, the accessions derived from Southeast Asia had the greatest diversity, followed by southern Asia, South Pacific, Africa, Middle East, South America, and Oceania, while those in western and eastern Europe and North America had the lowest diversity. Visualized by Nei genetic diversity index on the world map using the Kriging method, germplasm accessions collected from southern Asia, Southeast Asia, Central America, and Africa were most diversified, while those from North Pacific, Oceania, western and eastern Europe, and North America had the lowest diversity (Fig. 2).

Germplasm accessions that were introduced from southern Asia had the most private alleles per locus, followed by Africa, Central America, Southeast Asia, South Pacific, China, Oceania, and Middle East, while those in eastern Europe, Central Asia, North and South America,

![Figure 1. Global distribution of 1794 accessions in the USDA rice core collection among 14 geographic regions classified by the United Nations. Each dot represents an accession placed on the world map according to its latitude and longitude.](image-url)
and western Europe had the least private alleles per locus (Fig. 3).

Coordinate 1 and 2 jointly explained 71% of variation generated by 72 markers among 14 regions (Fig. 4). The two coordinates classified the 1794 accessions into three distinctive groups among 14 regions of germplasm origination. Similarly, three main clusters were resulted from the UPGMA analysis based on Nei (Nei 1973) genetic similarity (Fig. 5). In cluster 1, germplasm accessions from South America were mostly related to Central America and then to Africa, Oceania, and North America. Two subgroups of the originating region among rice accessions obviously existed in cluster 2, while eastern Europe and western Europe were in subgroup 1 and Central Asia, Middle East, and North Pacific in subgroup 2. In cluster 3, germplasm accessions originating in Southeast Asia were closest to those in South Pacific and then to China and southern Asia. Cluster 1 was closer to cluster 2 than to cluster 3.

Genetic Diversity and Genetic Relationships among Countries

Among the 78 countries from which 5 or more accessions were introduced in the core collection, Myanmar had the most diversification indicated by the highest PIC (0.65) (Fig. 6). The PICs measuring genetic diversities ranged from 0.60 to 0.63 in 13 countries: four in Africa, three in Southeast Asia, and two each in South America, South Pacific, and southern Asia. The PICs ranged from 0.50 to 0.60 in 27 countries: four each in Africa and Central America, three each in South America and southern Asia, two each in Central Asia, China, Middle East, and North Pacific, and one each in eastern Europe, North America, Oceania, Southeast Asia, and South Pacific. There were 22 countries with the PICs ranging from 0.40 to 0.50: five in South America, four each in Africa and Central America, two each in Middle East and Oceania, and one each in Central Asia, China, North Pacific, southern Asia, and western Europe. France and Spain in western Europe and Romania in eastern Europe had the lowest PIC value.

Table 1. Allelic analysis of 1794 accessions in the USDA rice core collection originating from 14 geographic regions and genotyped with 72 DNA markers.

<table>
<thead>
<tr>
<th>Geographic region</th>
<th>Countries</th>
<th>Accessions</th>
<th>Major allele frequency</th>
<th>Avg. alleles per locus</th>
<th>Adjusted alleles per locus†</th>
<th>Availability</th>
<th>Gene diversity</th>
<th>Polymorphic information content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>26</td>
<td>198</td>
<td>0.46</td>
<td>9.32</td>
<td>8.21</td>
<td>0.97</td>
<td>0.68</td>
<td>0.64</td>
</tr>
<tr>
<td>Central America</td>
<td>12</td>
<td>116</td>
<td>0.50</td>
<td>8.01</td>
<td>7.50</td>
<td>0.97</td>
<td>0.63</td>
<td>0.59</td>
</tr>
<tr>
<td>Central Asia</td>
<td>7</td>
<td>61</td>
<td>0.50</td>
<td>6.71</td>
<td>6.57</td>
<td>0.97</td>
<td>0.62</td>
<td>0.59</td>
</tr>
<tr>
<td>China</td>
<td>4</td>
<td>212</td>
<td>0.51</td>
<td>8.58</td>
<td>7.43</td>
<td>0.97</td>
<td>0.62</td>
<td>0.58</td>
</tr>
<tr>
<td>Eastern Europe</td>
<td>7</td>
<td>102</td>
<td>0.65</td>
<td>6.96</td>
<td>6.68</td>
<td>0.98</td>
<td>0.47</td>
<td>0.45</td>
</tr>
<tr>
<td>Middle East</td>
<td>6</td>
<td>91</td>
<td>0.46</td>
<td>7.47</td>
<td>7.27</td>
<td>0.97</td>
<td>0.66</td>
<td>0.62</td>
</tr>
<tr>
<td>North America</td>
<td>2</td>
<td>75</td>
<td>0.64</td>
<td>6.06</td>
<td>5.88</td>
<td>0.98</td>
<td>0.49</td>
<td>0.46</td>
</tr>
<tr>
<td>North Pacific</td>
<td>3</td>
<td>108</td>
<td>0.59</td>
<td>7.50</td>
<td>7.11</td>
<td>0.98</td>
<td>0.55</td>
<td>0.52</td>
</tr>
<tr>
<td>Oceania</td>
<td>6</td>
<td>57</td>
<td>0.48</td>
<td>6.79</td>
<td>6.68</td>
<td>0.98</td>
<td>0.64</td>
<td>0.61</td>
</tr>
<tr>
<td>South America</td>
<td>12</td>
<td>224</td>
<td>0.45</td>
<td>8.44</td>
<td>7.39</td>
<td>0.97</td>
<td>0.66</td>
<td>0.62</td>
</tr>
<tr>
<td>South Pacific</td>
<td>4</td>
<td>120</td>
<td>0.44</td>
<td>8.42</td>
<td>7.96</td>
<td>0.97</td>
<td>0.68</td>
<td>0.64</td>
</tr>
<tr>
<td>Southeast Asia</td>
<td>6</td>
<td>114</td>
<td>0.42</td>
<td>8.86</td>
<td>8.43</td>
<td>0.97</td>
<td>0.70</td>
<td>0.66</td>
</tr>
<tr>
<td>Southern Asia</td>
<td>7</td>
<td>215</td>
<td>0.45</td>
<td>10.06</td>
<td>8.72</td>
<td>0.97</td>
<td>0.67</td>
<td>0.64</td>
</tr>
<tr>
<td>West Europe</td>
<td>10</td>
<td>101</td>
<td>0.69</td>
<td>6.00</td>
<td>5.85</td>
<td>0.98</td>
<td>0.42</td>
<td>0.39</td>
</tr>
<tr>
<td>Total</td>
<td>112</td>
<td>1794</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td>0.52</td>
<td>7.80</td>
<td>7.26</td>
<td>0.97</td>
<td>0.61</td>
<td>0.57</td>
</tr>
</tbody>
</table>

†Adjust the accessions in each geographic region to the same number

Table 2. Analysis of molecular variance (AMOVA) in 14 regions for 1794 accessions in the USDA rice core collection genotyped with 72 DNA markers.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>Φst†</th>
<th>p value</th>
<th>Estimated variance</th>
<th>Percentage of total variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among regions</td>
<td>13</td>
<td>18956.7</td>
<td>1458.2</td>
<td>0.11</td>
<td>&lt;0.001</td>
<td>10.8</td>
<td>11</td>
</tr>
<tr>
<td>Within regions</td>
<td>1780</td>
<td>164044.8</td>
<td>92.2</td>
<td>0.89</td>
<td>&lt;0.001</td>
<td>92.2</td>
<td>89</td>
</tr>
<tr>
<td>Total</td>
<td>1793</td>
<td>183001.5</td>
<td>103.0</td>
<td></td>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Among regions</td>
<td>13</td>
<td>19674.3</td>
<td>1513.4</td>
<td>0.10</td>
<td>&lt;0.001</td>
<td>8.9</td>
<td>8.6</td>
</tr>
<tr>
<td>Among countries</td>
<td>65</td>
<td>17106.2</td>
<td>263.2</td>
<td>0.12</td>
<td>&lt;0.001</td>
<td>9.2</td>
<td>8.9</td>
</tr>
<tr>
<td>Within countries</td>
<td>1672</td>
<td>142286.8</td>
<td>84.8</td>
<td>0.85</td>
<td>&lt;0.001</td>
<td>84.8</td>
<td>82.5</td>
</tr>
<tr>
<td>Total</td>
<td>1750</td>
<td>179067.2</td>
<td>102.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

†Φst, the correlation of random haplotypes within populations, relative to that of random pairs of haplotypes drawn from the whole species (Excoffier et al., 1992).
Figure 2. Geographic diversity of rice germplasm demonstrated by Nei genetic diversity index in the USDA rice core collection genotyped with 72 DNA markers. The deeper the red color is, the greater the genetic diversity is for the area. The deeper the blue color is, the smaller the genetic diversity is for the area. Each dot represents an accession placed on the world map according to its latitude of Y axis and longitude of X axis.

Figure 3. Mean number of private alleles per locus as a function of standardized sample size (g) for 14 geographic regions arranged from high on the top to low on the bottom for 1794 accessions in the USDA rice core collection.
Cluster analysis of 78 countries from which 5 or more accessions were present in the core collection formed five distinctive groups (Fig. 7). Fourteen countries were placed in Cluster 1, six in Central America, four in South America, three in Africa, and one in North America, which is the United States. Cluster 2 contained 20 countries, six in eastern Europe, four in western Europe, three in Middle East, two each in North Pacific and South America, and one each in Africa, Central Asia, and Oceania. Cluster 3 included 19 countries, seven in Africa, three in South America, two each in South Pacific and Southeast Asia, and one each in Central Asia, China, North America, North Pacific, and Oceania. Cluster 4 had 18 countries, four in southern Asia, three each in Central America and Southeast Asia, two each in Africa, China, and South America, and one each in Oceania and South Pacific. Cluster 5 was the smallest, including five countries, two each in Middle East and southern Asia, and one in Central Asia. Two countries each with five accessions were independent of these clusters. Haiti in Central America was between Cluster 4 and 5, while Guinea-Bissau in Africa was between Cluster 1 and 5.

**DISCUSSION**

The majority of genetic variance exists within instead of among geographic regions and within instead of among countries in the USDA rice world collection. Germplasm
cross introductions among countries in different regions might explain this phenomenon. As we know, IR is the prefix for all the cultivars released from the International Rice Research Institute (IRRI) in the Philippines of South Pacific. However, some germplasm accessions named with the IR prefix in the collection are listed as originating in Columbia of South America, Cuba of Central America, India, Nepal, and Pakistan of southern Asia, Korea of North Pacific, Myanmar of Southeast Asia, Taiwan of China, and Zimbabwe of Africa. Furthermore, over 40% of accessions introduced from Cote D’Ivoire of Africa have the name prefix IRAT and some accessions introduced from Burkina Faso, Madagascar, and Mali of Africa and the Philippines of South Pacific are also named with the IRAT prefix. Obviously, these countries collected those accessions from the originating sources first, and the United States made the second introductions from them later.

Germplasm accessions in the USDA rice collection that originated in southern Asia, Southeast Asia, and Africa are highly diversified, while those in North America and western and eastern Europe have the lowest diversity. Different measurements of genetic diversity, such as average number of alleles per locus, PIC, Nei index, and average number of private alleles per locus have identically reached this conclusion. Different rice growing environments, availability of genetic stocks, and diverse rice consumption behaviors may be responsible for diversity differences among the regions.

Southern Asia, including India, Bangladesh, Pakistan, Nepal, Sri Lanka, Bhutan, and Afghanistan, is a major player in global rice production and accounts for about 40% of total world rice acreage in 2007 (FAO, 2008). Southeast Asia, including Thailand, Myanmar, Vietnam, Cambodia, and Laos, is another major producer with about 20% of global rice area in 2007. Rice grown in such vast areas of the regions must adapt to complex environmental conditions in terms of temperatures, water availabilities, elevations, soil types, and field managements associated with diverse cultures in these countries. The adaptation forces rice to diversify itself by rule of natural selection for survival of the fittest proposed by C. Darwin in the 19th century. This could explain higher genetic diversity among germplasm accessions that were introduced from these regions than those from other regions in the USDA rice world collection.

Association of germplasm diversity with rice planting area is demonstrated in a major rice growing country. India has the largest rice area in the world (44 million ha in 2007, 28% of world total) and Indian accessions are highly diversified with PIC of 0.61. China is ranked the second by country for rice area (29 million ha in 2007, about 19% of world total); however, Chinese accessions have much less diversity (PIC = 0.52) than the Indian accessions as well as those originating in countries that have much less rice area, that is, Myanmar, Indonesia, Cambodia, Malaysia, Nepal, Thailand, etc. Availability of rice germplasm for...
introduction to the USDA NPGS may partly explain the low PIC value for Chinese accessions. There are 2082 Chinese accessions in the NPGS and introduction within a 3-yr period accounts for 73% of the total: 636 accessions in 1947 (31%), 603 accessions in 1974 (29%), and 276 accessions in 1975 (13%). Introduction in such a concentrated way would obviously reduce genetic diversity because the introduced accessions could not be well designed and prepared. Conversely, germplasm introduction from other countries has occurred more consistently. For example, the biggest introduction from India included 250 accessions in 1975, which is 17% of current holdings (1432). No other introductions from India in a single year were over 200 accessions except 1972 with 214 accessions.

High genetic diversity among germplasm accessions that were collected from Africa could be explained by its large number of countries and two species of Oryza sativa and O. glaberrima Steud. while other regions have sativa only. Diversity analysis was based on germplasm accessions collected from 26 countries in Africa, which is double or triple number of countries in other regions, much more than other regions, that is, only two countries (the United States and Mexico) in North America. Rice cultivars originating in so many countries should be diversified because they have to adapt to so many diversified environments in those countries. Cultivated rice belongs to two species in genus Oryza: sativa and glaberrima (Panaud, 2008; Vaughan et al., 2003, 2008; Izawa, 2008). Oryza sativa that is cultivated worldwide was domesticated in Asia and its closet wild relative is O. rufipogon Griff., a species found throughout Asia. However, O. glaberrima was domesticated in West Africa and grown exclusively in Africa, and its closet wild relative is O. barthii A. Chev., an endemic species of West Africa that is often found as a weed in rice fields. There are 14 accessions of O. glaberrima in the rice core collection, and all of them originated in Africa except one that was introduced from El Salvador, Central America. These glaberrima accessions may contribute greatly to high diversity among 198 accessions introduced from 26 countries in Africa. For example, PIC

Figure 7. Cluster analysis of countries having five or more accessions in the USDA rice core collection genotyped with 72 DNA markers.
dropped from 0.64 to 0.61 in Africa when *O. glaberrima* accessions are taken out of the calculation.

With the association between rice genetic diversity and size of growing area, it is easy to understand that germplasm accessions collected from western and eastern Europe and North America have the lowest diversity among the 14 regions. Only two countries grow rice in North America, the United States and Mexico. Rice area in Mexico is 6.4% of the U.S. rice area, that is, 0.7% of world rice area in 2007 (FAO, 2008). United States rice breeding has focused on two major grain types, medium (or short) and long (Mackill and McKenzie, 2003). The former is mainly grown in one state on the west coast, California, and the latter in the southern rice belt including Arkansas, Louisiana, Mississippi, Missouri, and Texas. The long grain type belongs to tropical *japonica* and the medium (or short) to temperate *japonica*, two groups of a subspecies *japonica* in *O. sativa* (Mackill, 1995). Another subspecies, *indica*, has greater genetic diversity than *japonica* (Gao et al., 2005; Mackill, 1995; Ni et al., 2002; Negrao et al., 2008; Caicedo et al., 2007; Komishi et al., 2008), probably because *indica* is grown on about 80% of the world’s total rice area and *japonica* on the remaining 20% (Mackill, 1995). Furthermore, intensive breeding efforts greatly reduce genetic diversity of U.S. cultivars (Dilday, 1990). Pedigree analysis demonstrates that rice cultivars can be traced back to 22 accessions of germplasm in the southern U.S. and 23 germplasm introductions in California (Dilday, 1990; Bockelman et al., 2003). Two prominent long grain cultivars, Lebonnet and Lemont, have more than 72% of their genes in common, and two medium grain cultivars, Calrose and Caloro, have over 90% of their genes in common. As a result, small rice areas concentrated in two belts under intensive breeding and *japonica* subspecies that has low genetic diversity may be factors contributable to low genetic diversity of rice germplasm in the United States as well as North America.

Germplasm accessions originating in South and Central America, Africa, and Oceania are highly diversified. These regions are grouped in the same cluster with North America, which could prove to be valuable gene pools to diversify U.S. germplasm. The USDA rice world collection provides security for the U.S. rice industry when a need occurs. For example, germplasm accessions resistant to a race of blast (*Magnaporthe grisea*) were identified (Eizenga et al., 2006) about a year after this race caused significant yield losses on a commercial cultivar (Lee et al., 2005). Also, this diversified collection is internationally available for research purposes in the Genetic Stock–Oryza (GSOR) Collection at http://www.ars.usda.gov/Main/docs.htm?docid=8318 (verified 4 Aug. 2010) for the core collection, and in the National Small Grains Collection (NSGC) at http://www.ars.usda.gov/main/docs.htm?docid=2884 (verified 4 Aug. 2010) for the whole collection, free of charge and restrictions.

### Supplemental Information Available

Supplemental information associated with this article is available free of charge online at www.crops.org/publications/cs.

### Acknowledgments

The authors thank Olivier François and Eric Durand (Dep. Mathematical Biology, TIMC-IMAG, Faculty of Medicine, La Tronche, France) and Doug Nychka (Institute for Mathematics Applied to Geosciences, Colorado) for help with representation of diversity index using Kriging methods, Xueyan Sha, Christopher Deren, and Ellen McWhirter for critical review, and TIFFANY Sookasern, Tony Beaty, Aaron Jackson, Yao Zhou, BiaoLin Hu, XiaoBai Li, and LiMeng Jia for technical assistance.

### References


