Contrasting genetic diversity among *Oryza longistaminata* (A. Chev et Roehr) populations from different geographic origins using AFLP


1International Plant Genetic Resources Institute (IPGRI) – Regional Office for sub-Saharan Africa, P.O. Box 30677, Nairobi, Kenya.
2School of Biological Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.
3International Center for Research in Agroforestry (ICRAF), P.O. Box 30677, Nairobi, Kenya.

Molecular markers have been used extensively in studying genetic diversity, genetic relationships and germplasm management. However, the understanding of between and within population genetic variation and how it is partitioned on the basis of geographic origin is crucial as this helps to improve sampling efficiency. The objective of this study was therefore to assess the intra-specific diversity in *Oryza longistaminata* and how the variation is partitioned within and between different geographic locations, using molecular markers. AFLP analysis generated 176 bands that revealed high levels of polymorphism (95.6%) and diversity within and between populations. The mean Nei’s genetic diversity for all the 176 loci in the 48 populations was 0.302 and diversity for populations within countries ranged from 0.1161 to 0.2126. Partitioning of between and within population diversity revealed that the mean allelic diversity at each polymorphic locus was HT = 0.3445. The within population diversity was (HS = 0.1755) and the between population diversity was (DST = 0.1688). Results of AMOVA revealed significant differences (p<0.05) in genetic variation among populations within different countries of the region. Genetic parameters estimated from AFLP data indicated that there are high levels of genetic diversity in the wild populations of *O. longistaminata* studied and that this diversity is higher within than between populations. Hierarchical partitioning also revealed that most of this diversity is found between populations within countries than among countries. Regional collection and conservation strategies therefore need to consider country differences while national strategies should consider population differences within countries.

Key words: Molecular markers, AFLP, *Oryza longistaminata*, populations, genetic diversity.

INTRODUCTION

The knowledge of the extent and structure of genetic variation in populations of wild plants is essential not only for understanding the processes of evolution but also the development of appropriate and efficient strategies for the collection and conservation of populations (Frankel and Bennet, 1970). The wild relatives of cultivated rice are an important constituent of the rice gene pool and have contributed significantly to breeding programmes. Evaluation of the populations continues to reveal new sources of resistance to diseases and pests. Interest in these species has increased as techniques to transfer genes from wild species to cultivated rice improve (Heinrichs et al., 1985; Khush et al., 1990; Brar and Khush, 1995). *O. longistaminata* A. Chev. et Roehr is
one of the wild rice species that is already making significant contributions to rice improvement programmes, particularly in Asia and Africa (Khush et al., 1990; Jones et al., 1996).

Molecular markers have a lot of potential for application in genetic diversity studies and the management of plant genetic resources as clearly shown by Virk et al. (1995; 2000), Ford-Lloyd et al. (1997) and Song et al. (2003). In particular, the use of AFLP in studying the rice gene pool is promising as they have proved to be reliable, produce a large number of bands and are more reproducible than RAPDs (Sensi et al., 1996; Xu et al., 1995; Virk et al., 2000). Using AFLPs, Mackill et al. (1996) generated more polymorphic bands than through any other molecular marker method previously used on rice. Cho-YongGu et al. (1996), Nandi et al. (1997) and Virk et al. (1998) have also used AFLP markers in cloning and mapping QTL. Zhu et al. (1998), Raven (1999) and Aggarwal et al. (2002) have clearly demonstrated their efficiency and reliability in analysis of genetic diversity of rice populations.

*O. longistaminata* belongs to the AA genome group of the sativa complex and hybridizes easily with the cultivated species, *O. glaberrima* and *O. sativa* (Bezançon et al., 1977; Vaughan, 1989). The species is allogamous, has a high rate of self-incompatibility and is thus predominantly outcrossing (Ghesquiere, 1987; Jones et al., 1996). Several studies based on morphology, cytogenetics, ecology and biochemical and molecular markers have significantly contributed in complementary ways to the current knowledge of genus *Oryza* (Vaughan, 1994). However, the understanding of between and within population genetic variation and how it is partitioned on the basis of geographic origin is crucial as this will help to improve sampling efficiency in order to have a wider representation of the *O. longistaminata* genetic diversity conserved. The objective of this study was therefore to assess the intra-specific diversity in *O. longistaminata* and determine how the variation is partitioned within and between different geographic locations, using the East and Southern African countries as a case study.

We analyzed AFLP markers to study the partitioning of diversity among and within countries, between and within populations using 48 bulked populations and 24 individuals from 8 populations as working samples. The material represented a wide geographical range.

**MATERIALS AND METHODS**

**Plant material**

Seed was collected from Kenya (KEN), Zimbabwe (ZIM), Mozambique (MOZ) and Tanzania (TZA) in collaboration with the national programmes. More seed and leaf samples originating from Botswana (BOT), Namibia (NAM), Mozambique, Madagascar (MAD) and Zambia (ZAM) were obtained from the International Rice Gene Bank in IRRI. Leaf material was harvested fresh, frozen and stored at -70°C. 48 populations of *O. longistaminata* were selected for analysis to represent the geographical range of the species. For material received as seeds, young leaves were harvested from 2-week-old seedlings of each accession for DNA extraction.

**DNA extraction and AFLP analysis**

The protocol for isolation of genomic DNA developed by Gawel and Jarret (1991) was followed with modifications, using 20 mg of leaf material from 5-10 seedlings. AFLP analysis was carried out using the methodology of Virk et al. (1998), as modified from Vos et al. (1995). EcoRI from Pharmacia and Msel from New England Biolabs, Inc. were used to restrict the genomic DNA by incubating the mix for 1 h at 37°C followed by ligation to EcoRI and Msel adaptors by incubation at 37°C for 3 h.

Pre-amplification was carried out using EcoRI-(EP; 5'-GACTGCGTACCAATTCA-3') and Msel-(MP; 5'-GATGAGTCCTGCGAAC TACA-3') adapter-directed primers each having a single selective base (Vos et al., 1995). It was run for 2 h in a Hybaid Omnimgene Thermocycler and the products then diluted 50 times. Amplification was carried out using four primers from Operon Technologies Inc. The EcoRI primers possessed three selective bases at the 3’ end and were end-labeled with [γ-33P] ATP using T4 polynucleotide kinase (TPNK—Pharmacia). The amplification was performed using Gene Amp PCR System 9700 and the following profile: one cycle of 95°C; one cycle of 94°C for 1 min., 65°C for 1 min., 72°C for 1 min followed by eight cycles with a stepwise reduction of 1°C in the annealing temperature to 57°C; 27 cycles of 94°C for 30 seconds, 56°C for 1 min, 72°C for 1 min. The reaction was stopped by adding 25 μL of formamide loading buffer containing 98% formamide, 10 mM EDTA pH 8.0, 1 mg/ml xylene cyanol and 1 mg/ml bromophenol blue. The samples were denatured at 90°C for 1 min and 2 μL of the mix was loaded on to a 5% denaturing polyacrylamide gel. The gel was cast in a Sequigen 38x50 cm gel apparatus (Bio-Rad) and electrophoresis run for 2 h at a constant temperature of 50°C. It was dried, exposed to a Kodak Biomax film for 4-5 days and then processed in a COMPACT X2 X-GRAPH.

**Data analysis**

Since AFLP are usually dominant markers, only two character states were scored: presence (1) or absence (0) of bands. An assumption was made that the co-migrating fragments in different populations are allelic and map to similar genomic regions (Waugh et al., 1997; Virk et al., 1998, 2000).

Data analysis using PC-ORD Version 4.10 was performed with each primer combination individually. Correlation between distance matrices generated using Jaccard’s coefficient was tested with the standardized Mantel Statistics and a significance value assigned. Cluster analysis to show the relatedness of the populations according to the geographic origin (different countries) was performed and a dendrogram generated using the UPGMA algorithm of POPGENE Version 1.31. Jaccard’s co-efficient and Sequential Agglomerative, Hierarchical and Nested Clustering (SAHN) feature of NTSYS-pc Version 2.0 were used to generate dendrograms showing the clustering of the 48 populations and the 24 individuals from 8 populations. Principal component analysis was performed using the DECENTER and EIGEN procedures of NTSYS-pc Version 2.0 and both 3-D and 2-D scatterplots generated.

An analysis of molecular variance (AMOVA; Excoffier et al., 1992) was performed on a Euclidean distance matrix generated and analyzed using the ARLEQUIN 1.1 software package. Significance values were assigned to variance components based on the
random permutation (10,000 times) of individuals assuming no
geometric structure. The significance of pair wise $Fst$ values between
countries was calculated to check for significant differences.
Canonical discrimination analysis of POPGENE 1.31 was used to
reveal a 2-dimensional view of relationships between countries by
representing the multivariate data on orthogonal axes. POPGENE
Version 1.31 was used to calculate genetic diversity in each
country based on AFLP product frequencies. At each polymorphic
locus, the total allelic diversity is represented by $HT$, which is
partitioned into the mean allelic diversity within populations, $HS$ and
the allelic diversity among populations, $DST$. These quantities are
related by the expression $HT = HS + DST$. The proportion of total
allelic diversity found among populations, $GST$ was calculated as the
ratio $DST/HT$ (Hartl, 1987; Buso et al., 1998).

RESULTS
The four primers used (E1+M2 and E2+M1) in AFLP
analysis revealed a total of 184 bands that were 95.6% polymorphic. The 176 polymorphic bands ranged in size
from 24 to 726 bp and clearly distinguished all the 48
populations. 88 polymorphic bands clearly distinguished
between the 24 individuals in 8 populations. The value of
the standardized Mantel Statistic, $r$, was 0.82 (P<0.0002)
indicating a good correlation between primer combinations.

Between country relationships
From the pair wise difference distance method of
AMOVA, there were significant differences ($p<0.05$) of
populations among all the countries except BOT, as
shown in Table 1.

The results of cluster analysis using an arbitrary
similarity defined as 1- distance of 65% divided the
countries into four major groups that to a large extent,
reflect their geographic origin (Figure 1). One group
consisted of MOZ and NAM populations. The second
group comprised the BOT populations while the third
group revealed an affinity between TZA, ZAM, ZIM and
KEN populations.

The fourth group comprised MAD populations, which is
quite distinct from the other groups and separates at a
similarity of 57%. The mean genetic distances between
countries are given in Table 2.

Cluster 1 comprised MOZ and NAM populations with
14 sub-clusters ranging from 52 to 84% in similarity.
Cluster 2 comprises 2 of the 3 MAD populations with
similarity of 79%. The cluster is quite distinct, separating
from the other populations at 39% similarity. The third
cluster comprises TZA populations, which separate from
the other populations at 49%. Cluster 4 comprises some
populations from TZA, ZIM, ZAM and KEN and
separates from the other populations at 52% similarity.
Cluster 5 is a unique accession from ZAM, which
separates from the other populations at 42%. Cluster 6
comprises TZA, ZAM, ZIM, KEN, BOT and MAD
Table 2. Mean (Nei, 1973) genetic distance between countries.

<table>
<thead>
<tr>
<th></th>
<th>MAD</th>
<th>ZAM</th>
<th>BOT</th>
<th>KEN</th>
<th>TZA</th>
<th>ZIM</th>
<th>NAM</th>
<th>MOZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAD</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZAM</td>
<td>0.1540</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOT</td>
<td>0.2654</td>
<td>0.1408</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEN</td>
<td>0.1900</td>
<td>0.0549</td>
<td>0.1786</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TZA</td>
<td>0.1569</td>
<td>0.0491</td>
<td>0.1665</td>
<td>0.0459</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZIM</td>
<td>0.1872</td>
<td>0.0724</td>
<td>0.2129</td>
<td>0.06436</td>
<td>0.0846</td>
<td>****</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAM</td>
<td>0.2824</td>
<td>0.2502</td>
<td>0.3622</td>
<td>0.3195</td>
<td>0.2774</td>
<td>0.2862</td>
<td>****</td>
<td></td>
</tr>
<tr>
<td>MOZ</td>
<td>0.2126</td>
<td>0.2030</td>
<td>0.3700</td>
<td>0.2467</td>
<td>0.2192</td>
<td>0.2237</td>
<td>0.1299</td>
<td>****</td>
</tr>
</tbody>
</table>

Figure 2. Clustering of 48 populations using 176 polymorphic AFLPs, based on NTSYs-pc UPGMA cluster analysis (Rohlf, 1993) and Jaccard's coefficient.

When the PCA was performed, the first three eigenvectors explained 35.1% (1st 19.4%, 2nd 9% and 3rd 6.7%) of the variation. The 1st principal component separates Cluster 1 comprising MOZ and NAM populations. It also separates Cluster 2 comprising the two MAD populations. The 2nd principal component separates Clusters 4 and 6. The 3rd principal component populations. It separates at 49% similarity and is the most diverse cluster. The TZA and ZAM populations are spread over 3 out of the 6 clusters while KEN, MAD and ZIM populations are found in 2 clusters. NAM and MOZ populations are only found in 1 cluster. The spread of the clusters seems to reflect the within country diversity as indicated by Nei's diversity index.
Figure 3. 2-D principal component analysis based on genetic relationships of 48 populations of *O. longistaminata* using the DECENTER and EIGEN features of NTSYS-pc (Rohlf, 1993).

Figure 4. 3-D principal component analysis based on genetic relationships of 48 populations of *O. longistaminata* using the DECENTER and EIGEN features of NTSYS-pc (Rohlf, 1993).
Figure 5. Clustering of 24 individuals from 8 populations using 88 polymorphic AFLPs, based on UPGMA cluster analysis (Rohlf, 1993) using Jaccard’s coefficient.

(3-D) clarifies the separation of Cluster 2 and further separates Cluster 1 into MOZ and NAM populations with almost all the MOZ located much closer to the basement. Cluster 5 with one ZAM accession has been separated by the 2nd Principal component though not very distinctly. Cluster 3, having 3 TZA populations, has not been clearly separated by either 3-D or 2-D as shown in the scatter plots provided in Figures 3 and 4.

The visual grouping of 24 individuals from 8 populations using PCA and cluster analysis, indicated that 6 clusters of individuals generally reflecting the diversity were formed at an arbitrary similarity level of 40% as shown in Figure 5. The PCA revealed that the first three eigenvectors explained 49% (1st 28 %, 2nd 12 % and 3rd 9%) of the variation. The analysis shows that the individuals are generally clustering on the basis of the high within population genetic diversity as reflected by Nei’s diversity index. The most diverse populations (T7737MOZ and T7352NAM) with diversity indices of 0.268 and 0.257 respectively had the individuals spread out widely in the clusters. The analysis of similarities within populations revealed high within population diversity.

Genetic diversity

There was a clear genetic differentiation both among countries and populations within countries using significance tests based on 10,100 permutations. The examination of the hierarchical partitioning of genetic variation by AMOVA demonstrated that genetic differentiation amongst countries was significant at p<0.05 with a variance component of 9.75. The within countries differentiation was significant at p<0.05 with variance components of 23.14. Of the total diversity, 29.65% was attributed to country differences while 70.35% of the total variation was attributed to population differences within the countries and the FST value was 0.30.

The mean Nei’s (1973) genetic diversity for all the 176 loci in the 48 populations was 0.302. Nei’s (1973) mean genetic diversity for populations within countries ranged from 0.1161 to 0.2126. The highest mean diversity was found in TZA (H = 0.2126) followed by ZAM and MAD. The lowest diversity was found in NAM (H = 0.1161) followed by MOZ, ZIM and KEN. The partitioning of between and within population diversity using 24 individuals from 8 populations revealed that the mean allelic diversity at each polymorphic locus was HT = 0.3445, the within population diversity was higher (HS = 0.1755) than the between population diversity, which was DST = 0.1688. The proportion of total allelic diversity found among populations was GST = 0.4905. The within population diversity varied considerably from one population to another even in the same countries and ranged from 0.041 to 0.268 in three MOZ populations, 0.069 to 0.257 in three NAM populations,
Table 3. Genetic diversity parameters of populations from different countries.

<table>
<thead>
<tr>
<th>Country of origin</th>
<th>Mean Nei’s genetic diversity Value</th>
<th>Rank</th>
<th>% of polymorphic loci</th>
<th>% of rare bands at frequency of p&lt;0.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAD</td>
<td>0.2032</td>
<td>3</td>
<td>49.14</td>
<td>38</td>
</tr>
<tr>
<td>ZAM</td>
<td>0.2069</td>
<td>2</td>
<td>53.14</td>
<td>6</td>
</tr>
<tr>
<td>BOT</td>
<td>0.0000</td>
<td>8</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>KEN</td>
<td>0.1981</td>
<td>4</td>
<td>58.29</td>
<td>28</td>
</tr>
<tr>
<td>TZA</td>
<td>0.2126</td>
<td>1</td>
<td>69.71</td>
<td>25</td>
</tr>
<tr>
<td>ZIM</td>
<td>0.1816</td>
<td>5</td>
<td>45.14</td>
<td>3</td>
</tr>
<tr>
<td>NAM</td>
<td>0.1161</td>
<td>7</td>
<td>29.21</td>
<td>0</td>
</tr>
<tr>
<td>MOZ</td>
<td>0.1728</td>
<td>6</td>
<td>49.71</td>
<td>0</td>
</tr>
</tbody>
</table>

0.206 in 1 MAD population and 0.17 in 1 ZAM population. Table 3 shows the diversity parameters while Figures 6 and 7 show the level of polymorphism, country differences in diversity and rare bands as revealed by AFLP markers.

**DISCUSSION**

A study of the cluster analysis of relationships between countries reveal four major groups as shown in Figure 1 and the among country differentiation is significant at p<0.05. These major groups seem to be formed on the basis of a) the level of diversity according to Nei’s indices b) the % of polymorphic loci and, c) the % of rare fragments at an arbitrary frequency of ≤0.02. A close analysis of the MAD populations shows that they are clearly distinct from the other groups. Compared to the other countries, the group had a high level of diversity, good level of % polymorphism, high % of rare fragments and the highest number of bands. Madagascar, where the populations originated, is an offshore island and the clear distinction from populations from the other countries may be attributed to environmental differences in addition to likely influences of the ‘Equilibrium theory of Island biogeography’ (MacArthur and Wilson, 1963; Brown and Lomolino, 1998) as reported by Kiambi et al. (2001). This is also consistent with the work on genetic variation in *Prunus Africana* showing that genetic distinctness and differentiation of populations may arise from geographic and ecological isolation (Dawson et al., 1999).

**Genetic structure and diversity**

Averaging of all loci, the total genetic diversity of HT = 0.3445 in the *O. longistaminata* populations is quite high and compares well with that of HT = 0.37 found in *Oryza glumaepatula* by Buso et al. (1998) using RAPDs. In their study of the South American populations of this species, they found much lower levels of within population diversity (0.081 and 0.089) but higher levels of between population diversity (0.28 and 0.29). This may be expected as they were working with an inbreeding species. In this study, the relatively high means of level of diversity over all the loci and the higher within (0.1755) than between (0.1688) population diversity may be attributed to the mating system (Maki and Horie, 1999) since *O. longistaminata* is allogamous, has a partial self-incompatibility and cross pollination is favoured by reproductive organs especially anther and stigma (Chu et al., 1969). The average diversity over all the populations found by Parsons et al. (1999) in Bangladesh rice landraces was 0.200, which is lower than in the present study. This however may be expected since *O. longistaminata* is a wild species. In addition, more genetic variation is likely to be detected in a widely spreading taxon, particularly when the sampling is carried out over a wide geographical range (Maki and Horie, 1999), as was the case in this study.

Wright (1978) cited by Hartl (1987) suggested that an FST range of 0-0.05 indicates little differentiation, 0.05-0.15 moderate, and 0.15-0.25 large differentiation and above 0.25 indicates very large differentiation. In this study, the level of population differentiation was FST = 0.30 which is very large using the suggested parameters. Hamrick and Godt (1989) reported that selfing species are characterized by a relatively high HT value (0.33), a low HS value (0.15) and a high GST value (0.51). From the studies, *O. longistaminata* has a relatively high FST of 0.29, high HT value of 0.343 high GST value of 0.49 and a HS value of 0.175. With the exception of the HS value, all the other values are characteristic of selfing species although the difference in genetic diversity between selfing and outcrossing species at the species level are not statistically significant (Hamrick and Godt 1989). However, the higher within than between population diversity is characteristic of out-crossing species (Hamrick and Godt, 1989; Buso et al., 1998). Using RAPDs, Buso et al. (1998) working with *O. glumaepatula* found the following values: FST = 0.64 and
Implications for conservation of wild rice genetic diversity

The results of AMOVA revealed significant differences (p<0.05) in genetic variation among populations within different countries of the region. Other genetic structure and diversity parameters including mean genetic diversity, band frequency and % of polymorphism have also indicated clear country differences in the structure of diversity.

The results also point to the need of developing sampling strategies that consider the population dynamics. For instance the NAM and MOZ populations have comparatively low between population diversity but show considerably higher within population diversity of 0.157 and 0.185 respectively, averaged from three populations. In these countries, it may be more cost effective to sample fewer populations but more individuals within those populations. In ZAM where the populations are high in diversity and show comparatively lower levels of within population diversity (0.175), it may be better to sample more populations but fewer individuals in those populations. These results are not conclusive and further studies are necessary for better understanding of the population dynamics, especially the mating system in O. longistaminata as this may influence the population structure and diversity.

In conclusion, the genetic parameters estimated from AFLPs data have indicated that there are high levels of genetic diversity in the wild populations of O. longistaminata studied and that this diversity is higher within than between populations. The hierarchical partitioning also revealed that most of this diversity is found between populations within countries than among countries. Regional collection and conservation strategies therefore need to consider country differences while national strategies should consider population differences within countries.

The study has clearly demonstrated the usefulness of AFLP in studying diversity in rice populations and its power of resolution in discriminating between populations and individuals within populations. It is clear that the study has also added significant knowledge and understanding of the population structure and genetic diversity of O. longistaminata in East and Southern Africa and its partitioning among countries, between and within populations.

ACKNOWLEDGEMENTS

We wish to express our gratitude to the International Plant Genetic Resources Institute and the International Rice Research Institute for their financial support. We also thank John Wasswa, Herta Kolberg, Michael Nawa, Margaret Nkya, Joseph Kemei, A. Mafa, Charles Mhazo, Paulino Munisse and Edwin Chiwona for facilitating the field sample collection and Dr. Mike Jackson for facilitating the acquisition of additional rice samples from the International Rice Gene Bank. We are indebted to Dr. Perminder Virk, Dr. Mark Raven and Ms. Fay Hughes for
their assistance in refining the protocols and other laboratory related activities.

REFERENCES


