Asian Journal of Plant Sciences

ISSN 1682-3974

science alert

ANSInet
an open access publisher
http://ansinet.com

Asian Journal of Plant Sciences 5 (3): 429-437, 2006
ISSN 1682-3974

Genetic Relationships among Rice Varieties Based On Expressed Sequence Tags and Microsatellite Markers

[1]Marie Noelle Ndjiondjop, [1]Kassa Semagn, [1]Mamadou Cissoko,
[2]Hirochi Tsunematsu and [3]Monty Jones
[1]Africa Rice Center (WARDA), 01 B.P. 2031, Cotonou, Benin
[2]Africa Rice Center (WARDA), PMB 5320, Ibadan, Nigeria
[3]Forum for Agricultural Research in Africa, PMB CT 173 Cantonments,
2 Gowa Close, Roman Ridge Accra, Ghana

Abstract: The genetic relationship and distance among rice varieties widely being used as parents by the African Rice Center (WARDA) are largely unknown but of great interest for breeding programs. The genetic relationship among 16 rice varieties was investigated using 83 EST and 174 SSR markers. Similarity among the 16 varieties varied from 21.9 to 91.2% for SSRs and 30.9 to 98.8% for ESTs. The extent of similarity of the four NERICA's with their donor (CG 14) and recurrent (WAB 56-104) parents ranged from 26.8 to 46.8% and 80.4 to 92.7%, respectively. Genetic similarity within the four NERICAs was the lowest between NERICA 1 and NERICA 6 (76.7 to 86.4%), which is in agreement with differences in agronomic traits. Cluster and principal component analyses performed on both marker types revealed three major groups for the cultivated rice species: the *glaberrima* group, *indica* group and NERICA and *japonica* group. The wild species, *O. longistaminata*, appeared to be close to the *glaberrima* than the other groups. Sixty four markers (22 EST and 42 SSR) were sufficient to clearly separate the 16 varieties into their respective group. Matrices correspondence tests demonstrated the presence of greater correspondence between the phenograms derived from SSRs and ESTs (r = 0.96).

Key words: Expressed sequence tag, microsatellite, NERICA, *Oryza glaberrima*, rice

INTRODUCTION

Genetic diversity is the raw material that farmers and plant breeders use to improve quality and productivity. It provides insurance against future adverse conditions and it is the basis of local selection of new desirable genotypes when the agricultural system changes. An understanding of the extent and distribution of genetic variation within and between populations of a species is therefore essential to (1) assist the selection of parents for breeding purpose, (2) choose populations for conservation programs, (3) estimate any possible loss of genetic diversity during conservation programs and (4) offer evidence of the evolutionary forces shaping natural populations (e.g., Thormann *et al*., 1994).

The study of genetic variation and structure has been greatly facilitated by the advent of DNA marker technology in the 1980s, which offered a large number of environmentally insensitive genetic markers that could be generated to follow the inheritance of important agronomic traits (Peleman and van der Voort, 2003). Restriction Fragment Length Polymorphic (RFLP) (Botstein *et al*., 1980), random amplified polymorphic DNA (RAPD) (Williams *et al*., 1990), sequence-tagged-sites (STS) (Talbert *et al*. 1994), amplified fragment length polymorphism (AFLP) (Vos *et al*., 1995) and microsatellite or simple sequence repeat (SSR) (Akkaya *et al*., 1992) are some of the most commonly used DNA marker types for a wide range of purposes. SSRs are short segments of DNA in which a specific motif of 1 to 6 bases is repeated in tandem (Litt and Luty, 1989). These motifs can be used as co-dominant DNA markers that detect higher levels of allelic variation than RFLP marker due to the variation in the number of tandem repeats between different genotypes (Becker and Heun, 1995). Much of this variation is probably caused by slippage events during replication (Schlötterer *et al*., 1991). SSRs are now considered as the most powerful Mendelian marker

Corresponding Author: Marie Noelle Ndjiondjop, Africa Rice Center (WARDA), 01 B.P. 2031, Cotonou, Benin
Tel: (229) 21 35 01 88 Fax: (229) 21 35 05 56

(Tautz, 1989) for genome mapping (Temnykh *et al*., 2001), varietal identification (Olufowote *et al*., 1997) and genetic diversity as they are highly polymorphic even between closely related lines (Semon *et al*., 2005). SSR require low amount of DNA, thus they can be easily automated and allow high-throughput screening, can be exchanged between laboratories and are highly transferable between populations (Gupta *et al*., 1999). In addition to being highly variable, microsatellite markers are also distributed relatively uniformly throughout the eukaryotic genomes. Expressed Sequence Tags (EST) markers resulting from transcriptional map provide a preliminary description of the organization of expressed genes and insights into genome evolution (Jermstad *et al*., 1998). They are small pieces of DNA sequence (usually 200 to 500 nucleotides long) that are generated by sequencing either one or both ends of an expressed gene (Jongeneel, 2000). The idea is to sequence bits of DNA that represent expressed genes and use these "tags" to fish a gene out of a portion of chromosomal DNA by matching base pairs. ESTs therefore provide researchers with a quick and inexpensive route for discovering new genes, for obtaining data on gene expression and regulation and for constructing genome maps (Wu *et al*., 2002).

Rice is not only an important food crop in the world but also a model plant for cereals (Havukkala 1996; Izawa and Shimamoto, 1996). It belongs to the genus *Oryza* which consisted of 22 wild and 2 cultivated species. The two cultivated species are the Asian rice (*O. sativa*) and the African rice (*O. glaberrima*). Long before the Asian rice reached to Africa, local farmers had domesticated the local species, *O. glaberrima*. Its local ancestry and numerous generations of selection *in situ* have made *O. glaberrima* to harbor a rich reservoir of genes that have allowed the species to survive and prosper in West Africa (Jones, 1997). Today, both *O. glaberrima* and *O. sativa* are commonly grown in mixtures by African farmers in upland and rain fed lowland environments (Semon *et al*. 2005). Recently, the African Rice Center (WARDA) developed New Rice for Africa (NERICA) through inter-specific cross between *O. glaberrima* and *O. sativa* (www.warda.org). However, the F1 offspring are male sterile and can only survive if pollinated by the mother species. When this occurs, inter-specific hybrid progenies can be very productive, as has been demonstrated by the NERICAs (Jones, 1997; www.warda.org).

Fourteen rice varieties have been widely used in breeding programs at WARDA. The genetic similarity or distance and relationship among these varieties are largely unknown but of great interest for breeding programs. Genetic similarities based on molecular markers are well suited for direct exploration of relationships within a germplasm pool. The objective of this study was therefore to investigate the genetic relationships among rice varieties widely used in WARDA's breeding programs using EST and SSR markers and assess the level of correspondence between the two DNA markers.

MATERIALS AND METHODS

Sampling and data collection: Fourteen of the varieties used in this study were the most widely used parents for breeding programs at WARDA and represent the two cultivated rice species and their inter-specific hybrids (Table 1). *O. sativa* was represented by a total of six genotypes with three *O. sativa* subsp. *indica* (Wita12, IR64, Tos145-19) and three *O. sativa* subsp. *japonica* (WAB 56-104, WAB 181-18, WAB 56-50) varieties. *O. glaberrima* consisted of a total of four land races adapted to both lowland (Tog 5681 and Tog 71062) and upland (CG 14 and CG 20) conditions. The four NERICAs (NERICA 1, NERICA 4, NERICA 5 and NERICA 6) are inter-specific hybrids between *O. glaberrima* and *O. sativa*. *O. longistaminata* and Nipponbare were included in this study for comparative purpose. The former is a wild species and has good potential for future

Table 1: Origin and characteristic of the 16 rice varieties used in the present study

Variety name	Species	Origin	Traits of interest*
O. longistaminata	Wild Africa rice	Senegal	Severalusefulgenes
Tog 5681	*O.glaberrima*	Nigeria	RYMVR
Tog 7106	*O.glaberrima*	Nigeria	ARGMR
CG 20	*O.glaberrima*	Senegal	DT, BT, WCA
CG 14	*O.glaberrima*	Senegal	DT, BT, WCA
NERICA 1	Inter-specific hybrid	WARDA	Aroma, GGQ, HYP
NERICA 4	Inter-specific hybrid	WARDA	GGQ, HYP
NERICA 5	Inter-specific hybrid	WARDA	GGQ, HYP
NERICA 6	Inter-specific hybrid	WARDA	GGQ, HYP
WAB56-104	*O. japonica*	WARDA	HYP, DT, BT
WAB181-18	*O. japonica*	WARDA	HYP, GGQ
WAB56-50	*O. japonica*	WARDA	HYP, DT, BT
Wita12	*O. indica*	WARDA/IITA	Iron toxicity
IR64	*O. indica*	IRRI	HYP
Tos14519	*O. indica*		AfRGMR
Nipponbare	*O. japonica*		Universal model for rice

*DT: Drought tolerance; BT: Blast Tolerance; WCA: Weed Competitiveness Ability; AfRGMR: Resistant to Africa Rice Gall Midge; RYMVR: Resistant to Rice Yellow Mottle Virus; HYP: High yield potential; GGQ: Good Grain Quality

breeding programs as a donor parent for some useful traits (e.g., resistance to Bacterial leaf blight and perenniality) while the latter is a universal model variety for the International Rice Genome Sequencing Project.

DNA was extracted from 4 weeks old seedlings using the CTAB protocol as described by Murray and Thompson (1980). The quality and quantity of the extracted DNA was checked on 1% agarose gel and working dilutions were prepared for PCR reactions. Rice is one of the most widely studied species with more than 2740 SSRs and 6590 ESTs available in public database (http://www.gramene.org). A total of 213 SSR primers (Table 2) were selected to cover all rice chromosomes according to their position on the microsatellite framework map published by Chen *et al.* (1997), which is available at the Gramene database. One hundred and thirty EST primers distributed throughout rice genome were also selected based on Wu *et al.* (2002) publication.

Amplification reactions for both SSR and EST markers were performed in a total reaction volume of 20 μL that consisted of 20 ng DNA, 1X PCR buffer (50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin, 50% glycerol and 10 mM Tris-HCl, pH 8.3), 20 pmol of each of the forward and reverse primers, 200 μM each dNTP (Boerhinger Mannheim) and 2 U Taq DNA polymerase (Perkin-Elmer). Amplifications were carried out in MJ PTC 100/96 thermal cycler using the following programs: 5 min at 94°C followed by 35 cycles of 1 min 94°C, 1 min annealing (55°C for SSRs and 60°C for ESTs) and 2 min at 72°C, with a final extension of 5 min at 72°C. The amplified products were separated on 3% agarose gel using 0.5X TBE buffer and 90V.

Statistical analyses: Three data sets were used for the statistical analyses. The first data set consisted of a 16×174 matrix (16 rows representing varieties and 174 columns corresponding to the total number of polymorphic SSR loci); the second data set contained 16×83 matrix (16 rows representing varieties and 83 columns corresponding to total the number of polymorphic EST loci); and a third data set (16×257 matrix) was created by combining all polymorphic SSR and EST markers. Simple matching coefficient was calculated as a measure of genetic similarity between varieties and used to generate phenograms using the UPGMA method of SAHN clustering. In order to obtain estimates of the magnitudes of differences among phenograms produced by each marker type, a cophenetic matrix was computed for each phenogram and compared using the Mantel matrix-comparison test (Mantel, 1967). For our second multivariate analyses, the data sets were used to derive correlation matrices from which the principal components were extracted and projected in two dimensions (EIGEN, PROJ and 2D PLOT programs). Similarities, cluster and principal component analyses and Mantel test were performed using NTSYS-pc for windows, version 2.0, Exeter Software (Rohlf, 1998). UNSCRAMBLER software version 9.2 (Computer-Aided Modelling, CAMO, Trondheim, Norway) was used to select markers that contributed most to the separation of the 16 varieties into their respective groups.

RESULTS

Among the 213 SSR and 130 EST primers initially used to characterize and evaluate the genetic relationship of the 16 rice varieties, 174 SSRs (81.9%) and 83 ESTs (63.8%) showed polymorphism among the study materials. The observed number of alleles in the two marker types varied from 1 to 7 for SSRs and from 1 to 4 for ESTs. The total number of polymorphic markers used in this study varied from 12 on chromosome 9 to 30 on both chromosomes 2 and 3 and the overall average per chromosomes was about 21 markers (Table 2). All polymorphic markers were single copy sequences and most showed co-dominant banding pattern; only 8 markers (2%) were dominant (data not shown).

Genetic similarities derived from SSR markers varied from 0.219 between CG 20 and Nipponbare to 0.912 between NERICA 4 and WAB 56-104 (Table 3). In ESTs, similarity was lowest between IR64 and Nipponbare

Table 2: Chromosomal distribution of the SSRs and ESTs used to study the 16 rice varieties and the number of polymorphic primers detected

	No. of primers screened for polymorphism			No. of polymorphic primers (% polymorphism)		
Chromosome	SSRs	ESTs	Total	SSRs	ESTs	Total
1	17	22	39	14(82.4)	13(59.1)	27(69.2)
2	23	15	38	20(87.0)	10(66.7)	30(78.9)
3	26	14	40	21(80.8)	9(64.3)	30(75.0)
4	16	10	26	13(81.3)	7(70.0)	20(76.9)
5	19	14	33	17(89.5)	10(71.4)	27(81.8)
6	12	8	20	11(91.7)	5(62.5)	16(80.0)
7	17	10	27	16(94.1)	7(70.0)	23(85.2)
8	16	12	28	15(93.8)	5(41.7)	20(71.4)
9	14	7	21	7(50.0)	5(71.4)	12(57.1)
10	23	1	24	15(65.2)	1(100.0)	16(66.7)
11	17	8	25	14(82.4)	5(62.5)	19(76.0)
12	13	9	22	11(84.6)	6(66.7)	17(77.3)
Total	213	130	343	174(982.5)	83(806.2)	257 (895.6)
Average	17.8	10.8	28.6	14.5(81.9)	6.9 (63.8)	21.4(74.6)

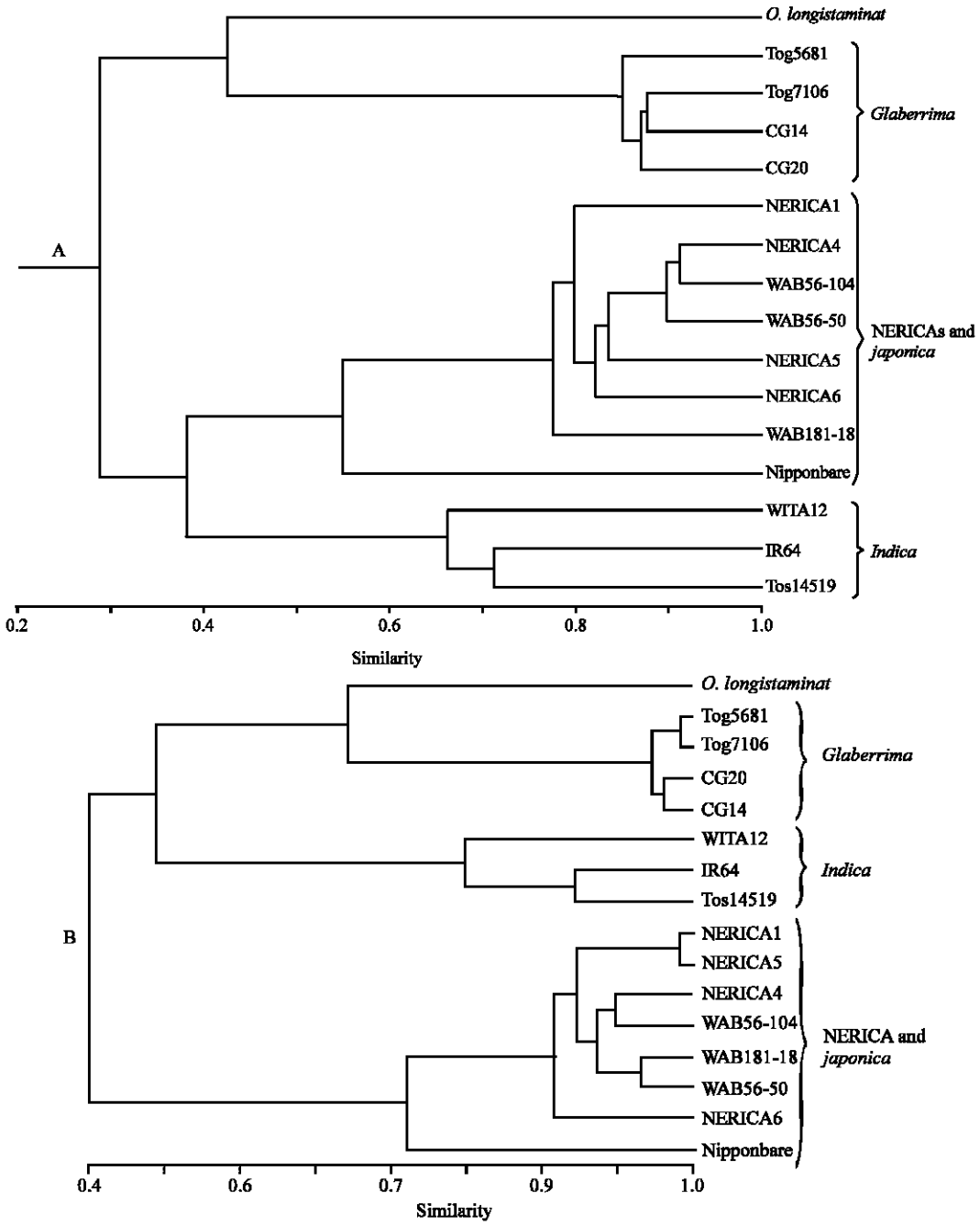

Fig. 1: UPGMA phenograms of 16 rice varieties based on similarity matrices calculated from (A) 174 SSRs and (B) 83 ESTs markers

(0.309) and highest between NERICA 1 and NERICA 5 (0.988). At an arbitrarily defined similarity value of 57% for SSRs and 70% for ESTs, cluster analyses performed using similarity matrices shown in Table 3 revealed three distinct groups (Fig. 1). The *glaberrima* group consisted of all varieties derived from *O. glaberrima* (Tog5681, Tog7106, CG14 and CG20); the *indica* group consisted of varieties derived from *O. sativa* subsp. *indica* (WITA12, IR64 and Tos14519) while the *japonica* group consisted of varieties derived from *O. sativa* subsp. *japonica* (Nipponbare,

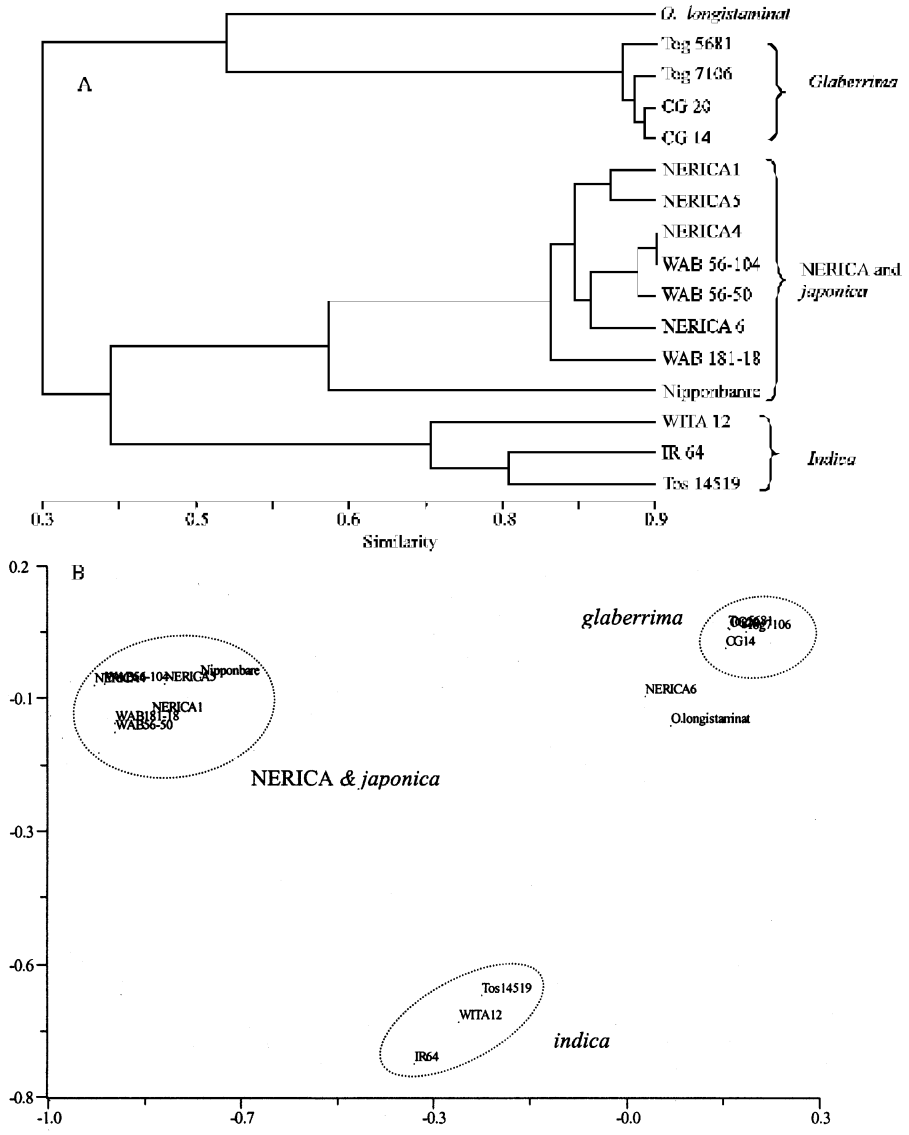

Fig. 2: Cluster and principal component analysis performed by combining 83 EST and 174 SSR markers: (A) UPGMA phenogram of 16 rice varieties based on similarity matrices and (B) score plot of 16 rice varieties from principal component analysis

WAB56-50, WAB56-104 and WAB181-18) and the inter-specific hybrids (NERICA 1, NERICA 4, NERICA 5, NERICA 6). The overall pattern of groupings of the 16 rice varieties obtained from either of the two marker types was identical with the combined data set of the two markers (Fig. 2). In both marker types and the combined data set, *O. longistaminata* appeared to be much closer to *O. glaberrima* than *O. sativa*.

The patterns of cluster analyses were also confirmed by principal component analyses (PCA). The first seven principal components (PC) from Principal Component Analyses (PCA) explained 99.2% of the SSR and 96.1% of the EST variations (Table 4). As shown in the 2 dimensional plots of the first two PCS (Fig. 3), the patterns of grouping for both marker types appeared to be basically similar to the cluster analyses. The loading scores derived from the Unscrambler program indicated the contribution of the individual markers for the separation of the 16 varieties in the first two axes. A total of 22 EST and 42 SSR informative markers were needed to separate the NERICA and *japonica* group from the two other groups (data not shown). When these selected

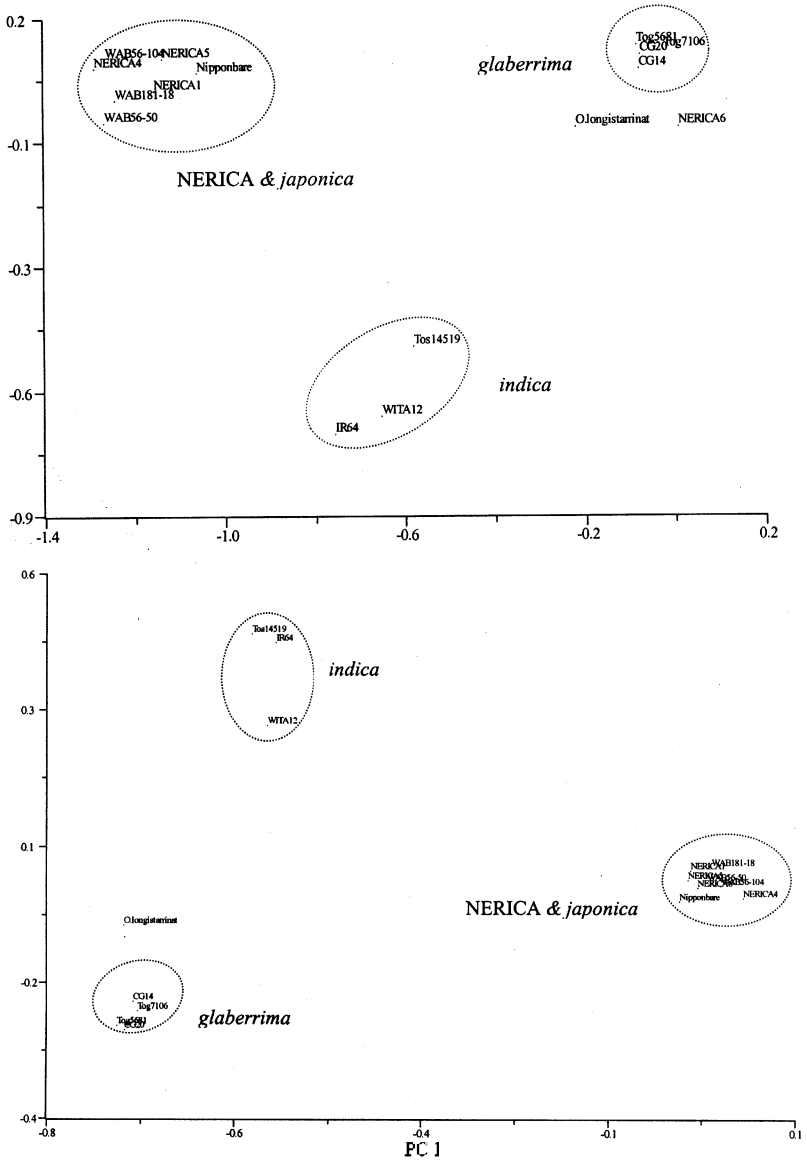

Fig. 3: Score plots of 16 rice varieties using (A) 174 SSRs and (B) 83 ESTs markers. The proportion of variance explained by each principal component is given in Table 4

markers were used for PCA, instead of all polymorphic ones for each marker type, the total explained variance by both PC1 and PC2 increased from 64.9 to 90.1% for ESTs and from 62.7 to 83.4% for SSRs.

DISCUSSION

Genetic relationships among varieties: The fourteen varieties used in this study have been extensively used at WARDA's breeding programs. To cite few examples, (a) CG 20 have been used as a parent in the development of over 30 varieties which are resistant to drought and/or blast; (b) WAB 56-104 and CG 14 have been used as parents in the development of a wide range of NERICAs. Several authors (Bhatt, 1970; Ariyo, 1987; Peeters and Martinelli, 1989; Souza and Sorrells, 1991) pointed out that crosses designed between genetically distant genotypes within a major cluster should produce higher variances in segregating populations than crosses between related genotypes. The high genetic distance between CG 14 and WAB 56-104 (58.2% for ESTs and 73.6% for SSRs) and their clustering into different groups confirm the presence

Table 3: Similarity among 16 rice varieties. Values above and below diagonal are based on 83 EST and 174 SSR markers, respectively

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
O.longistaminat	**1.000**	0.671	0.671	0.662	0.676	0.370	0.311	0.365	0.400	0.373	0.360	0.351	0.419	0.452	0.427	0.387
Tog5681	0.429	**1.000**	0.987	0.974	0.948	0.467	0.382	0.447	0.455	0.403	0.416	0.421	0.553	0.440	0.481	0.390
Tog7106	0.417	0.859	**1.000**	0.961	0.962	0.461	0.377	0.442	0.449	0.397	0.410	0.416	0.553	0.434	0.481	0.385
CG20	0.424	0.838	0.864	**1.000**	0.974	0.447	0.364	0.429	0.449	0.397	0.410	0.416	0.566	0.447	0.494	0.385
CG14	0.425	0.859	0.878	0.877	**1.000**	0.468	0.385	0.449	0.468	0.418	0.430	0.436	0.584	0.455	0.500	0.405
NERICA1	0.299	0.289	0.294	0.313	0.315	**1.000**	0.901	0.988	0.864	0.864	0.889	0.914	0.519	0.468	0.475	0.667
NERICA4	0.310	0.255	0.235	0.247	0.268	0.810	**1.000**	0.915	0.866	0.927	0.890	0.915	0.425	0.375	0.383	0.768
NERICA5	0.310	0.269	0.255	0.304	0.302	0.818	0.843	**1.000**	0.854	0.854	0.878	0.902	0.512	0.463	0.469	0.683
NERICA6	0.347	0.345	0.327	0.364	0.321	0.767	0.817	0.783	**1.000**	0.880	0.867	0.878	0.506	0.407	0.402	0.735
WAB56-104	0.300	0.235	0.227	0.239	0.261	0.804	0.912	0.825	0.885	**1.000**	0.904	0.927	0.457	0.407	0.390	0.771
WAB181-18	0.276	0.256	0.230	0.248	0.252	0.719	0.781	0.758	0.787	0.800	**1.000**	0.951	0.506	0.469	0.451	0.711
WAB56-50	0.299	0.241	0.239	0.264	0.280	0.788	0.887	0.834	0.797	0.905	0.808	**1.000**	0.512	0.463	0.457	0.744
WITA12	0.290	0.303	0.346	0.329	0.318	0.416	0.384	0.434	0.362	0.418	0.407	0.416	**1.000**	0.775	0.790	0.407
IR64	0.246	0.338	0.342	0.357	0.333	0.387	0.366	0.394	0.356	0.358	0.387	0.410	0.665	**1.000**	0.887	0.309
Tos14519	0.291	0.366	0.364	0.366	0.354	0.401	0.342	0.372	0.421	0.369	0.354	0.379	0.662	0.711	**1.000**	0.354
Nipponbare	0.234	0.220	0.225	0.219	0.222	0.485	0.554	0.515	0.667	0.547	0.515	0.564	0.329	0.356	0.367	**1.000**

Table 4: The proportion of variation explained by the seven Principal Components (PC) from principal component analyses performed using 174 SSR and 83 EST polymorphic markers

	Proportion of explained variance (%)		
Component	SSR	EST	Combined
PC1	42.0	39.5	40.8
PC2	20.7	25.4	22.5
PC3	15.1	11.1	13.9
PC4	9.7	7.4	9.0
PC5	4.7	6.6	5.5
PC6	3.8	3.4	3.5
PC7	3.2	2.7	2.9
Total	99.2	96.1	98.2

Table 5: Correlations between cophenetic matrices (above diagonal) and similarity matrices (below diagonal) derived from EST, SSR and combined data of the two marker types. Cophenetic correlation coefficients for the UPGMA phenograms shown in Figure 1 and Figure 2 are given in bold on the leading diagonal

	EST	SSR	Combined
EST	**0.98**	0.96	0.98
SSR	0.96	**0.99**	1.00
Combined	0.98	0.99	**0.99**

of sufficient genetic difference between them to serve as parents for breeding programs. For varieties that belong to the same cluster (within the *glaberrima*, *indica* or NERICA and *japonica* group), other important agronomic traits should be considered in selecting parental genotypes for breeding, as has been reported by various other workers (Bhatt, 1970; Ariyo, 1987).

The four NERICAs were derived using CG 14 as a donor parent and WAB 56-104 as recurrent parent and have the same pedigree (WAB 56-104/CG 14). NERICA's similarity with CG 14 and WAB 56-104 varied from 26.8 to 46.8 and 80.4 to 92.7%, respectively, which is in agreement with the expectation. For both EST and SSR markers, the level of genetic similarity among the four NERICAs was the lowest between NERICA 1 and NERICA 6 (76.7% for SSRs and 86.4% for ESTs), indicating the presence of more genetic variation between these two varieties. Both varieties have also been reported to be different in yield, plant height, grain length, grain width, grain apex color, amylose content and aroma (WARDA unpublished).

The pattern of clustering of the 16 varieties into 4 groups was in agreement with the taxonomy of the species: *O. longistaminata*, *O. sativa* subsp. *indica*, *O. sativa* subsp. *japonica* and *O. glaberrima* formed separate groups. The NERICAs were clustered together with the *japonica* than the *glaberrima* group, which confirms the high proportion of genomic contribution by their recurrent than donor parent. However, the high proportion of similarity between CG 14 and NERICAs for ESTs (up to 41.8%) is an indicative of the introgressed *glaberrima* genome, which might have accounted for the good agronomic performance in NERICAs compared to their recurrent parent.

Correspondence between SSR and EST markers : In order to quantify the extent of difference between the phenograms derived from SSR and EST, cophenetic matrix was constructed for each marker type and compared using the Mantel matrix correspondence test. This test demonstrated the presence of greater correspondence ($r = 0.96$) between the two phenograms (Fig. 4; Table 5). The correspondence between phenograms derived from either SSR or EST and the combined data set of the two marker types was also very high. Rohlf (1993) stated that if one matrix is a cophenetic value and the other matrix upon which the clustering was based, the cophenetic correlation to be computed from the two matrices could be used as a measure of goodness of fit for a cluster analysis. The degree of fit can be interpreted subjectively as follows: $r > 0.9$ very good fit; $0.8 < r < 0.9$ good fit, $0.7 < r < 0.8$ poor fit and $r < 0.7$ very poor fit. According to this interpretation, both markers showed very good fit ($r = 0.98$ for EST and $r = 0.99$ for SSR). The reasons for the

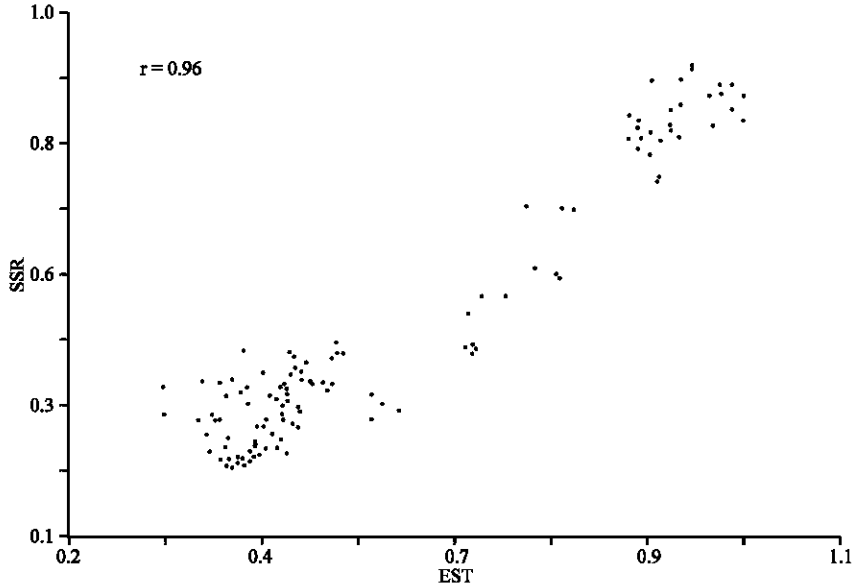

Fig. 4: Correlations between cophenetic matrices derived from 174 SSR and 83 EST markers

high correspondence between SSR and EST include codominant inheritance, high reproducibility and reliability and high genome coverage. The data, therefore, reflects the hyper-variability and higher resolving power of both marker types.

ACKNOWLEDGEMENTS

We would like to thank the JIRCAS for providing the generous financial support for this work.

REFERENCES

Akkaya, M.S., A.A. Bhagwat and P.B. Cregan, 1992. Length polymorphism of simple sequence repeat DNA in soybean. Genetics, 132: 1131-1139.

Ariyo, O.J., 1987. Multivariate analysis and the choice of parents for hybridization in Okra (*Abelmoschus esculentus* (L.) Moench). Theor. Applied Genet., 74: 361-363.

Becker, J. and M. Heun, 1995. Barley microsatellites: Allele variation and mapping. Plant Mol. Biol., 27: 835-845.

Bhatt, G.M., 1970. Multivariate analysis approach to selection of parents for hybridization aiming at yield improvement in self-pollinated crops. Aust. J. Agric. Res., 21: 1-7.

Botstein, D., R.L. White, M. Skolnick and R.W. Davis, 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet., 32: 314-331.

Chen X., S. Temnykh, Y. Xu, Y.G. Cho and S.R. McCouch, 1997. Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). Theor. Applied Genet., 100: 563-567.

Gupta, P.K., R.K. Varshney, P.C. Sharma and B. Ramesh, 1999. Molecular markers and their application in wheat breeding: A review. Plant Breeding, 118: 369-390.

Havukkala, I.J., 1996. Cereal genome analysis using rice as a model. Curr. Opin. Genet. Dev., 6: 711-714.

Izawa, T. and K. Shimamoto, 1996. Becoming a model plant: The importance of rice to plant science. Trends Plant Sci., 1: 95-99.

Jermstad, K.D., D.L. Bassoni, C.S. Kinlaw and B.D. Neale, 1998. Partial DNA sequencing of Douglas-Fir cDNAs used for RFLP mapping. Theor. Applied Genet., 97: 771-776.

Jones, M.P., M. Dingkuhn, G.K. Aluko and M. Semon, 1997. Interspecific *O. sativa* L. x *O. glaberrima* Steud: Progenies in upland rice improvement. Euphytica, 92: 237-246.

Jongeneel, C.V., 2000. Searching the Expressed Sequence Tag (EST) databases: Panning for genes. Briefings in Bioinformatics, 1: 76-92.

Litt, M. and J.A. Luty, 1989. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet., 44: 397-401.

Mantel, N.A., 1967. The detection of disease clustering and a generalized regression approach. Cancer Res., 27: 209-220.

Murray, M.G. and W.F. Thompson, 1980. Rapid isolation of high molecular weight plant DNA. Nucl. Acids Res., 8: 4321-4325.

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 microsatellites and RFLP markers. Genome, 40: 370-378.

Peeters, J.P. and J.A. Martinelli, 1989. Hierarchical cluster analysis as a tool to manage variation in germplasm collections. Theor. Applied Genet., 78: 42-48.

Peleman, J.D. and J. R. van der Voort, 2003. Breeding by design. Trends Plant Sci., 8: 330-334.

Rohlf, F.J., 1993. NTSYS-pc, Numerical Taxonomy and Multivariate Analysis System. Applied Biosystematics Inc., New York.

Rohlf, F.J., 1998. NTSYS-pc, Numerical Taxonomy and Multivariate Analysis System. Version 2.0, Exeter Software, New York.

Schlötterer, C., B. Amos and D. Tauzt, 1991. Conservation of polymorphic simple sequence loci in cetacean species. Nature, 354: 63-65.

Semon, M., R. Nielsen, M.P. Jones and S.R. McCouch, 2005. The population structure of African cultivated rice *Oryza glaberrima* (Steud.): Evidence for elevated levels of linkage disequilibrium caused by admixture with *O. sativa* and ecological adaptation. Genetics, 169: 1639-1647.

Souza, E. and M.E. Sorrells, 1991. Relationships among North American oat germplasms: II. Cluster analysis using quantitative Characters. Crop Science, 31: 605-612.

Talbert, L.E., N.K. Blake, P.W. Chee, T.K. Blake and G.M. Magyar, 1994. Evaluation of sequence-tagged-site-facilitated PCR products as molecular markers in wheat. Theor. Applied Genet., 87: 789 794.

Tautz, D., 1989. Hypervariability of the simple sequences as a general source of polymorphic DNA markers. Nucl. Acids. Res., 17: 6463-6471.

Temnykh, S, G. DeClerck, A. Lukashova, L. Lipovich, S. Carthinhour and S.R. McCouch, 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations and genetic marker potential. Genome Res., 11: 1441-1452.

Thormann, C.E., M.E. Ferreira, L.E.A. Camargo, J.G. Tivang and T.C. Osborn, 1994. Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. Theor. Applied Genet., 88: 973-980.

Vos, P., R. Hogers, M. Reijans, T. van de Lee, M. Hornes, A. Friters, J. Pot, J. Peleman, M. Kupier and M. Zabeau, 1995. AFLP: a new technique for DNA fingerprinting. Nucl. Acids Res., 23: 4407-4414.

Williams, J.G.K., A.R.K. Kubelik, J.L. Livak, J.A. Rafalski and S.V. Tingey, 1990. DNA polymorphisms amplified by random primers are useful as genetic markers. Nucl. Acids Res., 18: 6531-6535.

Wu, J., T. Maehara, T. Shimokawa, S. Yamamoto, C. Harada, Y. Takazaki, N. Ono, Y. Mukai, K. Koike, J. Yazaki, F. Fujii, A. Shomura, T. Ando, I. Kono, K. Waki, K. Yamamoto, M. Yano, T. Matsumoto and T. Sasaki, 2002. A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell, 14: 525-535.