J Appl Genet 48(2), 2007, pp. 107–113

Original article

IRAP and REMAP assessments of genetic similarity in rice

Castelo J. S. Branco[1], Eduardo A. Vieira[2], Gaspar Malone[1], Mauricio M. Kopp[1], Emilia Malone[1], Albina Bernardes[1], Claudete C. Mistura[1], Fernando I.F. Carvalho[1], Costa A. Oliveira[1]

[1]Eliseu Maciel School of Agronomy, Federal University of Pelotas, Pelotas, RS-Brazil
[2]Embrapa Cerrados, Brazil

Abstract. Rice is a model genome for cereal research, providing important information about genome structure and evolution. Retrotransposons are common components of grass genomes, showing activity at transcription, translation and integration levels. Their abundance and ability to transpose make them good potential markers. In this study, we used 2 multilocus PCR-based techniques that detect retrotransposon integration events in the genome: IRAP (inter-retrotransposon amplified polymorphism) and REMAP (retrotransposon-microsatellite amplified polymorphism). Markers derived from *Tos17*, a copia-like endogenous retrotransposon of rice, were used to identify genetic similarity among 51 rice cultivars (*Oryza sativa* L.). Genetic similarity analysis was performed by means of the Dice coefficient, and dendrograms were developed by using the average linkage distance method. A cophenetic correlation coefficient was also calculated. The clustering techniques revealed a good adjustment between matrices, with correlation coefficients of 0.74 and 0.80, or lower (0.21) but still significant, between IRAP and REMAP-based techniques. Consistent clusters were found for Japanese genotypes, while a subgroup clustered the irrigated Brazilian genotypes.

Keywords: *Oryza sativa* L., retrotransposons, *Tos 17.*

Introduction

Retroelements have been detected in most grasses, accounting for as much as 80% of the genome (Flavell et al. 1977; SanMiguel et al. 1996; IRGSP 2005). Because they transpose using a copy-and-paste mechanism of reverse transcription, retroelements can amplify to high copy numbers and may be the major contributors to genome size and genetic variability. In angiosperms, the most abundant are long terminal repeat (LTR) retrotransposons. Moreover, their ubiquitous presence in plant genomes is characterized by polymorphic insertion patterns within pools of many species (Vitte and Panaud 2005).

Many molecular strategies have been incorporated into breeding programs, with the goal of obtaining superior genotypes, such as genetic engineering (Ye et al. 2000), marker-assisted selection (Frary et al. 2000), induced mutations (Pandini et al. 1997), and genomic methods (McCouch 2001). Due to the abundance of retroelements and their ability to create new copies, their use as potential markers has been proposed (Amar and Hirochika 2001; Hirochika 2001). Three main types of markers have been developed: S-SAP (sequence-specific amplification polymorphism), IRAP (inter-retrotransposon amplified polymorphism) and REMAP (retrotransposon-microsatellite amplified polymorphism) markers (Waugh et al. 1997; Kalendar et al. 1999; Price et al. 2003). REMAP differs from IRAP in the sense that IRAP primers are combined with locus-specific simple sequence repeat (SSR) primers to identify polymorphic products of the amplification of a segment between the transposon and a SSR.

The aim of the present study was to assess the genetic variability of a collection of rice genotypes in order to evaluate the potential application of IRAP and REMAP techniques in rice breeding programs.

Received: June 27, 2006. Accepted: January 2, 2007.
Correspondence: C.A. de Oliveira, Bennetzen lab, Department of Genetics, University of Georgia, Athens, GA 30602, USA; e-mail: acostol@uga.edu

Materials and methods

Plant material and DNA quantification

A collection of 51 rice genotypes originating from Japan, Brazil and the Philippines, grown in irrigated (lowland) or upland systems were used, as described in Malone et al. (2006). Seeds were obtained from the germplasm bank belonging to the Plant Genomic Center, Eliseu Maciel School of Agronomy, Federal University of Pelotas (CGF/FAEM/UFPel). DNA was extracted from leaf samples, as described previously (Malone et al. 2006). DNA quantification was performed by comparing with a Low DNA Mass Ladder (Invitrogen, Life Technologies) on 0.8% agarose gel after ethidium bromide staining. DNAs were diluted to obtain a uniform concentration of 20 ng μL^{-1}.

IRAP reaction

The amplification reaction was performed according to the protocol described by Kalendar et al. (1999). Four LTR primers from *Tos17*, a copia-like endogenous retrotransposon of rice, were obtained according to their described sequences (Hirochika et al. 1996). The combination of 4 oligos, referred to as LTR-1, LTR-2, LTR-3 and LTR-4, enabled the use of 10 distinct primer pairs. IRAP amplifications were performed in a final volume of 25 µL containing 50 ng DNA, 1 × PCR buffer (20 mM Tris-HCl, pH 8.4, 50 mM KCl, 4 mM $MgCl_2$), 0.01% gelatin (w/v), 0.01% triton X-100 (v/v), 300 nM dNTPs, 1U Platinum® Taq DNA polymerase (Invitrogen), and 25 pmol of each LTR primer. Amplification was performed in a PTC-100™ thermocycler (MJ Research) in 0.2-mL microtubes. The amplification program consisted of initial denaturation at 94°C for 5 min, followed by 30 cycles composed of 94°C for 60 s, 50°C for 90 s, and 72°C for 120 s for denaturation, annealing, and extension, respectively. After amplification, a final extension step was performed at 72°C for 8 min. The amplification product was separated in polyacrylamide gel and silver-stained, as described by Briard et al. (2000).

REMAP reaction

Four primers synthesized from *Tos17* LTR sequences (Hirochika et al. 1996) were combined with rice SSR primers (*RM 55, RM 81, RM 202, RM 205, RM 207, RM 210*, and *RM 219*) performing 16 LTR-SSR primer combinations. REMAP amplifications were performed in a final volume of 25 µL, containing 50 ng DNA, 1 × PCR buffer (20 mM Tris-HCl, pH 8.4, 50 mM KCl, 4 mM $MgCl_2$), 0.01% gelatin, 0.01 triton X-100, 300 nM dNTPs, 1U Platinum® Taq DNA polymerase (Invitrogen); 25 pmol of LTR primer, and 50 pmol of SSR primer. The amplification was performed on a PTC-100™ thermocycler (MJ research) in 0.2-mL microtubes. The amplification program consisted of an initial denaturation cycle at 94°C for 120 s, followed by 30 cycles at 94°C for 30 s, 52°C for 120 s, and 72°C for 120 s for denaturation, annealing, and extension, respectively. A final extension step was performed at 72°C for 8 min. The amplification product was separated in polyacrylamide gel and silver-stained (Briard et al. 2000).

Data analysis

Amplification products were scored independently as 1 and 0 for presence and absence of bands, respectively, and the obtained binary data were used for the analyses. The genetic similarity between individual pairs of genotypes was analyzed by using the NTSYS pc 2.1 software (Rohlf 2000). The average similarity for all genotype pairs was used as a cutoff value for defining the clusters. For the estimation of genetic similarity, the Dice coefficient was used (Dice 1945) and basing on the 3 generated similarity matrices (MS_{IRAP}, MS_{REMAP} and $MS_{IRAP+REMAP}$), 3 dendrograms were obtained through clustering analysis by the UPGMA (unweighted pair group method with arithmetic means). To verify the adjustment between similarity matrices and respective dendrogram-derived matrices, the cophenetic correlation coefficient (r) was estimated according to Sokal and Rohlf (1962). To estimate the degree of correlation among the obtained similarity matrices (MS_{IRAP}, MS_{REMAP} and $MS_{IRAP+REMAP}$), a Mantel's test of matrix comparison with 1000 permutations was performed (Mantel 1967) by using the NTSYS pc 2.1 software (Rohlf 2000). The statistical stability of the clusters was estimated by a bootstrap analysis with 1000 replications, with the Winboot software (Yap and Nelson 1996).

Results and discussion

IRAP and REMAP analyses

Retrotransposons can potentially integrate in either orientation, enabling the finding of members of a retrotransposon family as head-to-head,

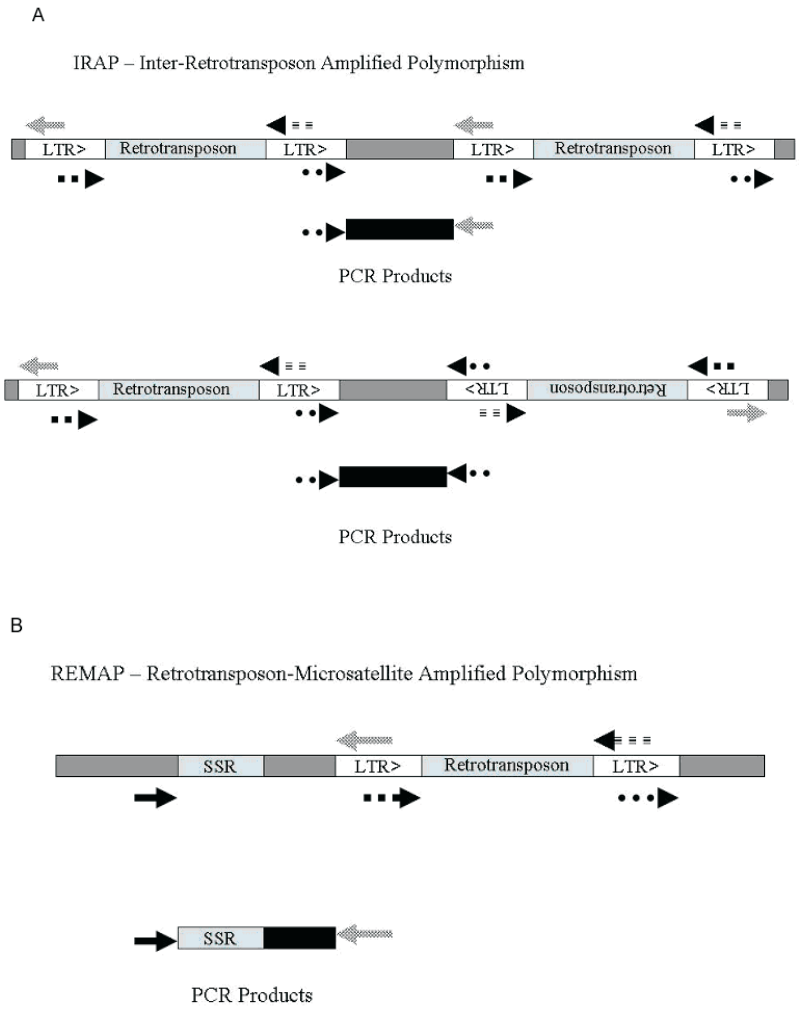

Figure 1. Diagram showing annealing of primers and potential products for 2 techniques: (A) IRAP = inter-retrotransposon amplified polymorphism; and (B) REMAP = retrotransposon-microsatellite amplified polymorphism

head-to-tail and tail-to-tail (Figure 1 A). To increase the probability of finding bands, one can combine primers from both 5' and 3' LTR ends or combine LTR primers with SSR primers (Figure 1B) to amplify intervening genomic DNA.

As expected, the IRAP analysis produced a high level of polymorphism (96% of bands appeared to be polymorphic). The 6 primer combinations used produced on average 9.16 bands; the highest and the smallest numbers of bands were obtained with the combinations LTR-2/LTR-3 (17 bands) and LTR-4/LTR-4 (5 bands), respectively (Table 1).

The average similarity of 0.74 and the cophenetic correlation coefficient of 0.90 (Figure 2) suggested a relatively high similarity and a good adjustment between the original and the dendrogram-derived matrix, respectively (Rohlf 1972). Cultivars Yonoashi and Matsuyama were the most similar genotypes (0.98), suggesting a close relationship in this subset of Japanese cultivars. The most dissimilar pair was K. Sim and IAC 5544 (0.17). Four distinct clusters were observed: I = Khao Sin; II = KhaoXiou Khay; III = Rusip; and IV = remaining genotypes. A subgroup of cluster IV was formed by irrigated Brazilian cultivars; this was consistent with the early AFLP analysis performed by our group (Malone et al. 2006). A novel finding is the apparent high similarity between the Brazilian irrigated cultivars and the upland Brazilian cultivars Jaguari and Birigui (Figure 2). These results are not in agreement with AFLP data from former studies (Malone et al. 2006).

The REMAP analysis was performed with 16 primer combinations, generating a total of 101 bands, of which 97% were polymorphic (Table 1). The primer combinations that amplified the highest and lowest number of bands were LTR-3/*RM55-R* (11 bands) and LTR-2/ *RM55-F*

Table 1. Number of bands obtained with each primer combination, considering 51 evaluated rice genotypes (CGF/FAEM/UFPel, 2006)

Primer combination	Total number of bands
LTR-2 and LTR-2	6
LTR-2 and LTR-3	17
LTR-2 and LTR-4	6
LTR-2 and *RM202-F*	6
LTR-2 and *RM207-R*	6
LTR-2 and *RM55-F*	1
LTR-2 and *RM55-R*	1
LTR-2 and *RM81-R*	8
LTR-3 and *LTR-3*	13
LTR-3 and *LTR-4*	8
LTR-3 and *RM202-R*	6
LTR-3 and *RM205-R*	7
LTR-3 and *RM207-R*	8
LTR-3 and *RM55-R*	11
LTR-3 and *RM81-R*	10
LTR-4 and *LTR-4*	5
LTR-4 and *RM205-R*	9
LTR-4 and *RM207-R*	2
LTR-4 and *RM210-R*	5
LTR-4 and *RM219-R*	7
LTR-4 and *RM55-R*	10
LTR-4 and *RM81-R*	4

(1 band) and LTR-2/ *RM55-R* (1 band), respectively (Table 1). The results obtained with the REMAP technique (Figure 3) indicate an average similarity of 0.51 and a cophenetic correlation of 0.83. The highest similarity was found between El Paso L-144 and BRS-Taim, which are not related according to pedigree data (data not shown). The most dissimilar genotypes were Khao Xiou Khay and Bacaba, showing a similarity of 0.15. Six clusters were formed: I = Khao Xiou Khay; II = Mun1; III = AUS 8 and Zebu Branco; and 2 clusters (IV and V) formed by the remaining genotypes. The clustering pattern suggests a high similarity between Brazilian irrigated cultivars, in agreement with previous results from Malone et al. (2006), using the AFLP technique. This may indicate that *Tos17* did not move after these genotypes had been selected from a common ancestor genotype, or maybe the pedigree information does not go back enough to describe the similar origin of these genotypes. The cultivars Patinai 6, E-Nawn and Ketebei demonstrated high similarity, grouping with a fairly high bootstrap value

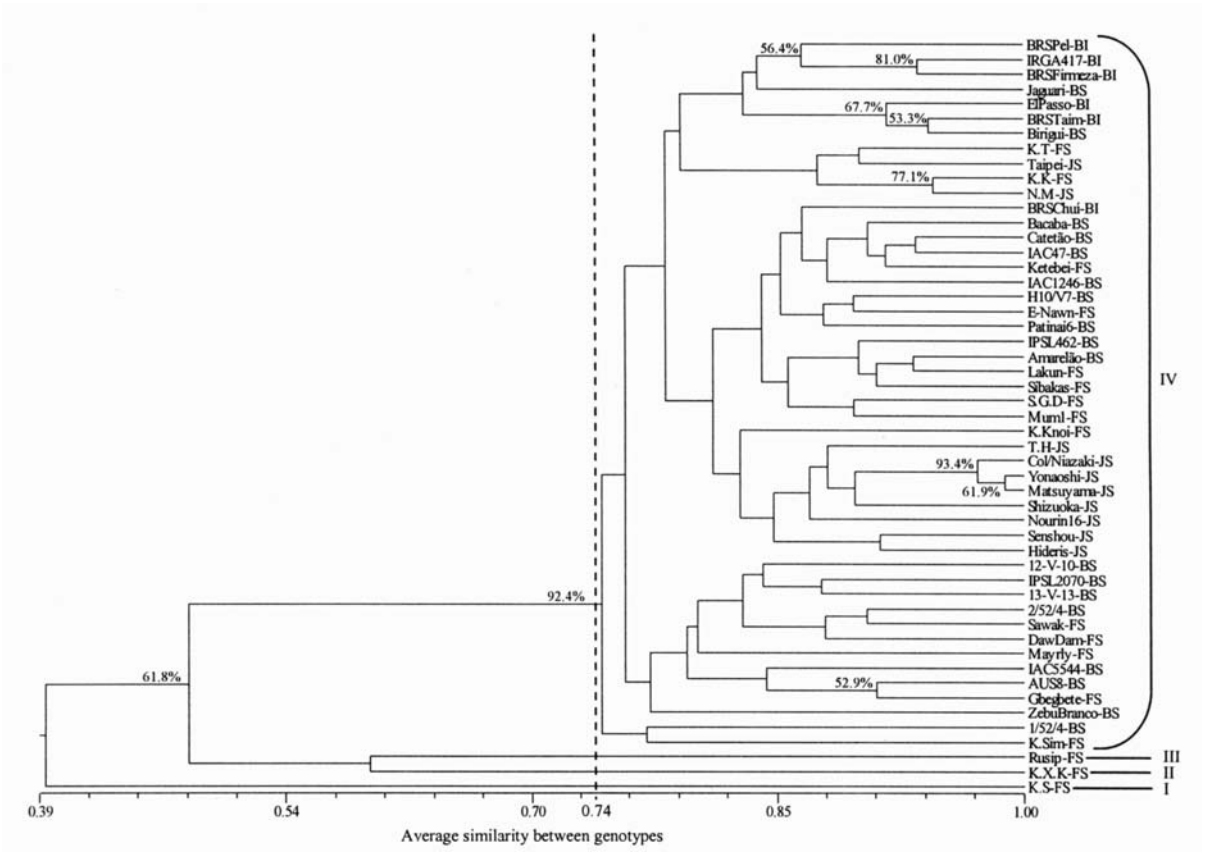

Figure 2. Dendrogram of 51 rice cultivars based on IRAP marker analysis using the Dice similarity index (Dice, 1945) and UPGMA clustering method. The percentage values for groups represent 1000 bootstrap cycles. Cophenetic correlation coefficient for the matrix is 0.90 (r). BS = Brazilian upland, BI = Brazilian irrigated, FS = Philippine upland, JS = Japanese upland (CGF/FAEM/UFPel, 2006).

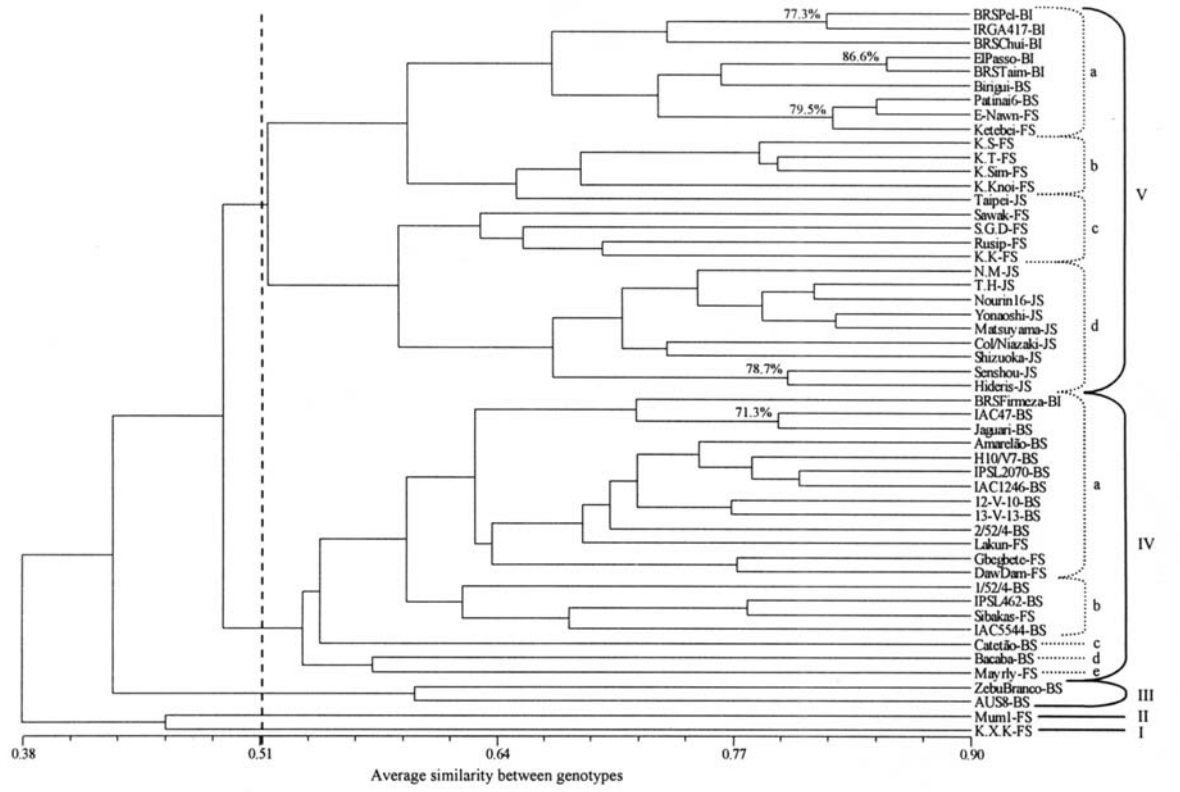

Figure 3. Dendrogram of 51 rice cultivars based on REMAP marker analysis using the Dice similarity index (Dice, 1945) and UPGMA clustering method. The percentage values for groups represent 1000 bootstrap cycles. Cophenetic correlation coefficient for the matrix is 0.70 (r). BS = Brazilian upland, BI = Brazilian irrigated, FS = Philippine upland, JS = Japanese upland (CGF/FAEM/UFPel, 2006).

(79.5%). The cultivar Birigui was present in the cluster formed by Brazilian irrigated cultivars. It was observed, in accordance with IRAP results, that the upland Japanese genotypes did not form clusters with any other genotypes.

Combined analysis

From the combined data analyses, we built a dendrogram to evaluate the power of both techniques when accumulated (Figure 4). It has been observed that different clustering techniques can be applied to the same data sets and that the most frequent structure is accepted as the most adequate (Oliveira et al. 1996). This attitude avoids the misleading conclusions based on artifacts, since each technique can impose a determined data structure. By observing the dendrogram based on the combined analysis of IRAP and REMAP and using the average similarity (0.66) as a cutoff line, 8 isolated groups were distinguished: I = Khao Sin; II = KhaoXiou Khay; III = Rusip; IV = Mum 1; and 4 groups containing 15 (V), 3 (VI), 18 (VII) and 9 (VIII) genotypes. This also confirmed a cluster containing the irrigated Brazilian cultivars and Birigui, agreeing with former analyses. The combined analysis reveals a subgroup formed by cultivars Patinai 6, E-Nawn and Ketebei in cluster VIII. Besides, the clustering of Japanese cultivars in group V reinforces the idea that the Japanese cultivars are more homogeneous and have a narrow genetic basis. Although the Japanese genotypes used in this study are not currently being used, this result corroborates recent findings that stated the narrow basis of Japanese cultivars, compared to Indica types (Garry et al. 2005). The most similar genotypes were Yonaoshi and Matsuyama (0.90), agreeing with IRAP results. The genotypes showing the lowest similarity were K. Sim and Mayrly (0.26).

The combined analysis also displayed the cultivars Rusip, KhaoXiou Khay and Khao Sim as unique very distant genotypes, as in the case of the IRAP technique. The estimated correlations between the 3 generated similarity matrices (MS_{REMAP}, MS_{IRAP}, and $MS_{IRAP+REMAP}$) evidenced a high and significant correlation of $MS_{IRAP+REMAP}$ with MS_{REMAP} ($r = 0.80$) and MS_{IRAP} ($r = 0.74$). However, the similarity matrices estimated by the techniques individually revealed a low but significant correlation ($r = 0.21$, Table 2). These results suggest that, similarly to what is found in barley (Kalendar et al. 1999), the addition of SSR primers enables the amplification of DNA regions that

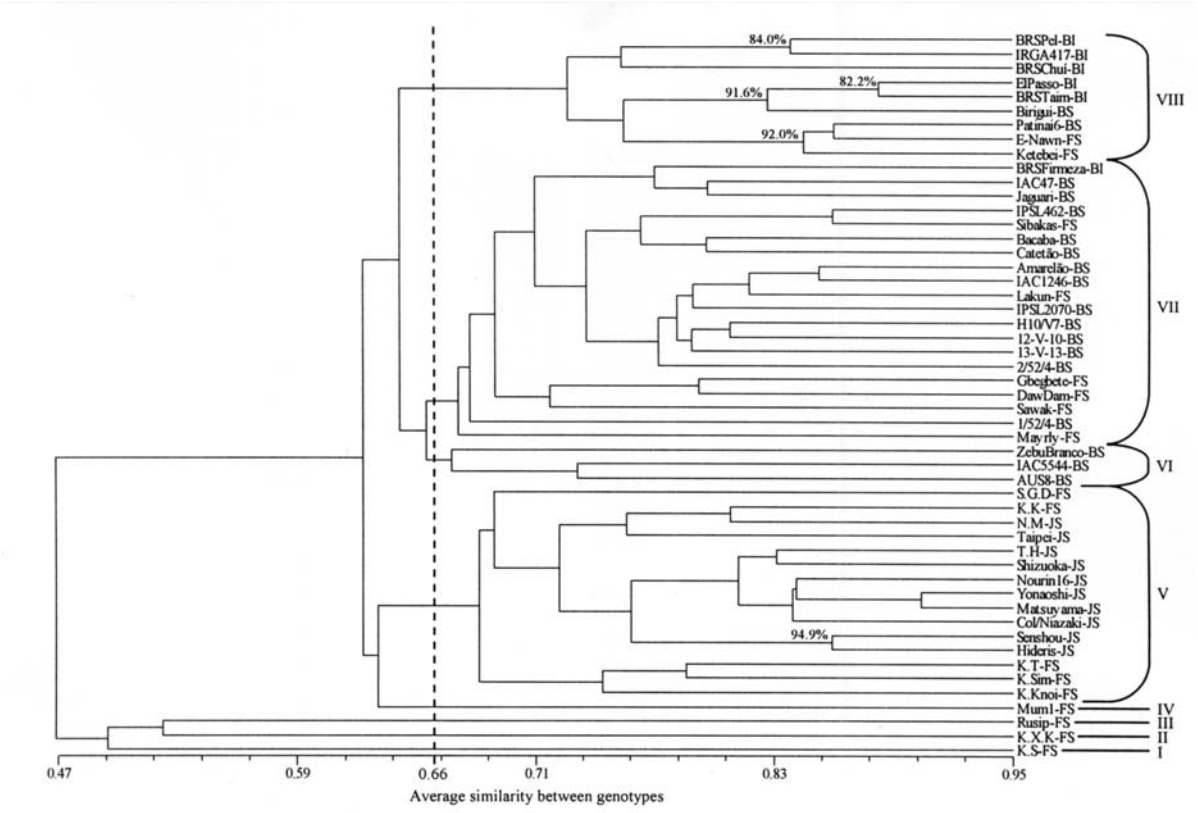

Figure 4. Dendrogram of 51 rice cultivars obtained from the combined marker analysis IRAP and REMAP, using

could not be covered by PCR methods using IRAP.

The use of retrotransposon-based markers can be a valuable tool for rice breeders, as it is for barley (Kalendar et al. 1999). The IRAP and REMAP techniques can be used separately or combined for a more complete genome survey. The ubiquitous presence of LTR retrotransposons in plant genomes suggests that the use of these techniques (either isolated or in combination) would allow breeders to obtain markers close to virtually any important agronomical trait and that the hypervariable nature of these repeat elements should make them excellent sources of polymorphic markers. However, one should point out that it is advisable to use different elements to screen the complete genome or at least elements that do not insert preferentially. When comparing the results obtained with IRAP/REMAP, they proved to be as reliable molecular markers as AFLPs, but they also bring additional information, showing a great potential use in genome assessments for fingerprinting, mapping and diversity studies.

Table 2. Correlations between the genetic similarity matrices estimated from IRAP (MS_{IRAP}), REMAP (MS_{REMAP}) and from the combined analysis ($MS_{IRAP+REMAP}$) of 51 evaluated rice genotypes (CGF/FAEM/UFPel, 2006)

	MSREMAP	MSIRAP+REMAP
MSIRAP	0.21*	0.74*
MSREMAP	–	0.80*

*Significant at P 0.05 (Mantel's test with 1000 permutations)

Acknowledgments. We thank CNPq, FAPERGS and FAO/IAEA for providing supporting grants and fellowships. Also, we are grateful to Dr. Clementine Vitte, Bennetzen lab, for important comments and suggestions.

REFERENCES

Amar K, Hirochika H, 2001. Applications of retrotransposons as genetic tools in plant biology. Trends in Plant Sci 6:127–133.

Briard M, Clera VLE, Gnzebelus D, Senalik D, Simon PW, 2000. Modified protocols for rapid carrot genomic DNA extraction and AFLP™ analysis using silver stain on radioisotopes. Plant Mol Biol Rep 18: 235–241.

Dice LR, 1945. Measures of the amount of ecological association between species. Ecology 26: 297–307.

Flavell RB, Rimpau J, Smith DB, 1977. Repeated sequence DNA relationships in four cereal genomes. Chromosoma 63: 205–222.

Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, et al. 2000. *fw 2,2*: A quantitative trait locus key to the evolution of tomato fruit size. Science 289: 85–88.

Garry AJ, Tai TH, Coburn J, Kresovich S, McCouch S, 2005. Genetic structure and diversity in *Oryza sativa* L. Genetics 169: 1631–1638.

Hirochika H, 2001. Contribution of the *Tos17* retrotransposon to rice functional genomics. Curr Op Plant Biol 4: 118–122.

Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M, 1996. Retrotransposon of rice involved in mutation by culture. Proc Natl Acad Sci 93: 7783–7788.

[IRGSP] International Rice Genome Sequencing Project, 2005. The map-based sequence of the rice genome. Nature 436: 793–800.

Kalendar R, Grob T, Regina M, Suoniemi A, Schulman I, 1999. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 98: 704–711.

Malone G, Zimmer PD, Mattos LAT, Kopp MM, Carvalho FIF, de Oliveira AC, 2006. Assessment of the genetic variability among rice cultivars revealed by AFLP. Revista Brasileira de Agrociencia 12: 21–25.

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

McCouch SR, 2001. Genomics and synteny. Plant Physiol 125: 152–155.

Oliveira AC de, Richter T, Bennetzen, JL, 1996. Regional and racial specificities in sorghum germplasm assessed with DNA markers. Genome 39: 579–587.

Pandini F, Carvalho FIF, Barbosa Neto JF, 1997. Plant height reduction in populations of triticale (× *Triticosecale* Wittmack) by induced mutations and artificial crosses. Braz J Genet 20: 483–488.

Price Z, Schulman AH, Mayes S, 2003. Development of new marker methods – an example from oil palm. Plant Genetics Res 1: 103–113.

Rohlf FJ, 1972. An empirical comparison of three ordination techniques in numerical taxonomy. Systematic Zool 21: 271–280.

Rohlf FJ, 2000. NTSYS-pc: numerical taxonomy and multivariate analysis system, version 2.1. Exeter Software, New York.

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, et al. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274: 765–768.

Sokal RR, Rohlf FJ, 1962. The comparison of dendrograms by objective methods. Taxon 11:30–40.

Vitte C, Panaud O, 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. Cytogenet Genome Res 110: 91–107.

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT, Powell W, 1997. Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet 253: 687–694.

Yap IV, Nelson RJ, 1996. Winboot: a program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms. Manila: IRRI.

Ye X, Al-Babili S, Kloti A, Zhang J, Lucca P, Potrykus I, 2000. Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287: 303–305.