Molecular fingerprinting of the elite, fine-grain type rice cultivar, Samba Mahsuri (BPT 5204) and assessment of genetic purity in seed-lots of the variety using SSR markers

K. Aruna Kumari[1]*, Ch. V. Durgarani[1], S. Vanisree[1], S. Sivaramakrishnan[1], S.M. Balachandran[2] and R.M. Sundaram[2]

Abstract- Samba Mahsuri is an elite, fine-grain type rice variety which is grown extensively in India. In order to fingerprint Samba Mahsuri and to differentiate it from 26 other closely related fine-grain type varieties, we utilized a total of 113 SSR markers, which included 87 hyper variable SSRs and 26 $(GATA)_n$ motif specific SSRs distributed uniformly across the rice genome. Twenty-six SSRs were found to be polymorphic among the rice varieties analyzed, amplifying a total of 74 alleles with an average of 2.84 alleles per marker. Cluster analysis based on Jaccard's similarity coefficient using UPGMA grouped the 27 rice varieties into five clusters. Samba Mahsuri and another elite fine-grain type variety, Sona Mahsuri (BPT 3291) were observed to be grouped in the same cluster. A twin-marker combination JGT11-16.3 + RM19426 could unambiguously differentiate Samba Mahsuri from other fine-grain type varieties through multiplex-PCR. To utilize these SSR markers specific for Samba Mahsuri effectively for detection of impurities in Samba Mahsuri, a rapid and cost effective two-dimensional bulked DNA sampling strategy was designed and validated.

Index Terms- Genetic purity, Samba Mahsuri, SSRs, DUS, and Multiplex PCR

I. INTRODUCTION

Rice (*Oryza sativa* L.) is the most important food crop in the world and feeds more than half of the global population. India is the second largest producer of rice in the world with an annual production of 99.37 million tonnes (2008-2009) from a rice growing area of 44.5 million hectares with an average productivity of 2014 kg ha^{-1}. To feed the ever-increasing population of India, it becomes necessary to increase the production of rice by releasing high yielding varieties. Since 1965, India has released more than 800 improved rice varieties. About 25% area under cultivation of rice in the country is being occupied by different varieties developed by Acharya N G Ranga Agricultural University (ANGRAU), Hyderabad, India and contributing to $1/3^{rd}$ of country's rice production. Three out of the five best varieties in the country *viz*., Samba Mahsuri (BPT 5204), Swarna (MTU 7029) and Vijetha (MTU 1001) were released by ANGRAU. The variety, Samba Mahsuri is extremely popular amongst rice farmers and consumers because of its high yield, medium-slender, fine-grain type and excellent cooking and eating quality. This variety is being targeted for improvement of multiple traits like biotic and abiotic stress tolerance/resistance. The original variety along with its improved versions needs to be protected in terms of plant breeder's right in the present era of intellectual property protection. There is an imminent need to identify genotype specific markers for cultivar identity and trueness of elite varieties like Samba Mahsuri so that such molecular markers can be utilized to unambiguously identify the variety and purity of seed-lots of the variety. During the seed production chain, the variety can get contaminated with other medium-slender varieties (some of which are cultivated extensively in India) and such impurities may go on accumulating unnoticed, finally leading to deterioration of the genetic quality of the variety. It has been estimated that 1% impurity in seed lot may decrease the potential yield of varieties and hybrids by about 100 kg/ha (Mao et al., 1996). In a country like India, where contract farming is practiced at many places for seed production, with the active participation of private sector (Mishra et al., 2003), monitoring genetic purity at each stage of seed production becomes necessary. Assessment of seed purity is one of the most important quality control aspects in seed production. Traditionally, it has been the practice to carry out a grow-out test (GOT), based on morphological traits, for assessment of purity of seeds. GOT is time consuming (takes one full growing season for completion), space demanding and often does not allow the unequivocal identification of genotypes. Among the PCR based DNA markers, microsatellites or simple sequence repeats (SSRs) are well known for high level of polymorphism content and versatility and are preferred due to their reproducibility and amenability for automation (McCouch et al., 1997; McCouch et al., 2002; Nagaoka T and Ogihara Y., 1997; Rafalski et al., 1996: Salimah et al., 1995). More than twenty thousands of SSR markers have been developed in rice so far and their chromosomal location and polymorphism levels have been determined (McCouch et al., 2002). SSRs have been extensively used in rice for molecular fingerprinting and assessment of genetic purity of hybrids and parental lines (Nandakumar et al., 2004; Sundaram et al., 2008; Yashitola et al., 2002). Based on the above, we designed the present study with the following objectives (i) to identify genotype specific marker(s) for Samba Mahsuri in order to differentiate it from other popular fine-grain varieties and (ii) to design an SSR marker based assay to determine the genetic purity in seed-lots of Samba Mahsuri using the genotype specific markers.

II. MATERIALS AND METHODS

1) Rice genotypes

A total of twenty seven genotypes including Samba Mahsuri, which were developed and released by ANGRAU were analyzed for the identification of genotypic specific molecular finger prints SSR marker (Table 1). [Published at end of paper]

2) DNA isolation and polymerase chain reaction (PCR) analysis to identification of genotype specific SSR markers

For varietal fingerprinting, total genomic DNA was extracted from freshly germinated young seedlings (Dellaporta et al., 1983). A total of 87 hyper-variable SSR primer pairs (selected from http://www.gramene.org) and 26 $(GATA)_n$ motif specific SSR primer pairs (Rajendrakumar et al., 2009) distributed across the 12 rice chromosomes were used for PCR amplification (listed in supplementary Table 1). DNA samples (50 ng) were amplified in 10µl reaction volumes containing 1X PCR buffer [10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl2, 0.01% (v/v) gelatin] (Bangalore Genei, India), 0.2 mM of each dNTPs (Bangalore Genei, India), 10 pmol of each primer and 1 U of Taq polymerase (Bangalore Genei, India). PCR was carried out in a Thermal cycler (Perkin–Elmer, USA). A PCR profile consisting of was 5 min initial denaturation at 94°C, 35 cycles of 1 min denaturation at 94°C, 1 min annealing at 55°C, 2 min extension at 72°C and 7 min at 72°C final extension was followed. The amplified products were resolved on 4% Metaphor agarose gels (Lonza, USA), stained with ethidium bromide and visualized under UV in a gel documentation system (Alpha Innotech, USA). The sizes of the amplified fragments were estimated with the help of Alphaease software utility of the gel documentation system using 50 and 100 bp DNA ladders (MBI Fermentas, Lithuania) as the molecular size standards. The polymorphism information content (PIC) for each SSR marker was calculated according to the formula (Botstein et al. , 1980). PIC=1-ΣPi2 - ΣΣPi2 Pj2 where 'i' is the total number of alleles detected for SSR marker and 'Pi' is the frequency of the i^{th} allele in the set of 27 genotypes investigated and j = i+1. If a certain allele with respect to a particular SSR marker was observed uniquely in just only one genotype under study and absent in all the other rice genotypes, it was considered to be specific for that genotype and such SSR markers were considered as a genotype specific SSR markers.

Initially an attempt was made to identify a single SSR marker-which could unambiguously differentiate Samba Mahsuri from the rest of the genotypes. Since no single marker could differentiate Samba Mahsuri from other varieties, combinations of two SSR markers which could be pre-PCR multiplexed were selected based on the amplification data of the SSR markers in order to identify genotype specific marker combinations. Only those markers that did not show significant primer sequence matches and those with amplicon sizes well separated from each other were selected for pre-PCR multiplexing.

3) Genetic diversity analysis based on SSR markers

To determine the genetic diversity among the 27 rice varieties, the binary data matrix generated from SSR marker data was subjected to cluster analysis. The excel file containing the binary data was imported into NT Edit in NTSYS-pc 2.02J (Rohlf, 1993). The 0/1 matrix was used to calculate similarity as Jaccard' coefficient using SIMQUAL subroutine in SIMILARITY routine. The resultant similarity matrix was employed to construct dendrogram using Sequential Agglomerative Hierarchical Non-overlapping (SAHN) based Unweighted Pair Group Method with Arithmetic means (UPGMA) to infer genetic relationships.

4) Purity assessment in the seed production chain of Samba Mahsuri variety through morphological characters and genotype-specific SSR markers

A total of 13 seed samples including nucleus seed from Rice Research Unit (RRU), Bapatla, Andhra Pradesh, India; breeder seed from three sources (RRU, Bapatla, Seed Research and Technology Centre, Rajendranagar, Hyderabad, India and Agricultural Research Station, Kampasagar, Nalgonda, Andhra Pradesh, India), foundation seed from four sources (Seed Production Farm, J.M.. Puram, Guntur district, Andhra Pradesh, India Farm Section, ANGRAU, Hyderabad, India, Andhra Pradesh State Seed Development Corporation, Hyderabad, India and National Seed Corporation, Hyderabad, India; certified seed from two sources (Andhra Pradesh State Seed Development Corporation, Hyderabad and National Seed Corporation, Hyderabad, India) and three farmer produced samples collected from three different villages in Karimnagar district, Andhra Pradesh, India were used to determine the genetic purity of Samba Mahsuri based on both morphological and molecular descriptors. The experimental material was planted in randomized block design in two replications. Treatments were randomized in each replication and one seedling per hill was transplanted in six rows. Twenty plants were planted in each row and the spacing between rows was 20 cm and spacing between plants was 15 cm. Thus, a total of 120 plants were maintained in each seed lot (six rows and 20 plants in each row). When the plants were at panicle initiation stage, leaf material was collected from each of the plant for DNA isolation. For reducing the sampling size, a two-dimensional DNA sampling strategy suggested (Nas et al ., 2002) and modified (Sundaram et al ., 2008) was adopted for bulking of leaf samples. Leaf samples were collected from each row and column were separately bulked. Thus a total of 26 DNA bulks representing 6 rows and 20 columns from a total 120 plants were used for isolation of total genomic DNA in genogrinder (OPC diagnostics, LLC, USA). Thus a total of 26 DNA bulks representing 6 rows and 20 columns were prepared and these were amplified through multiplex PCR using marker combination of RM 19426 and JGT 11-16.3, which can differentiate Samba Mahsuri from other genotypes with amplicon sizes of 272 bp and 139 bp, respectively. The rows and columns, which showed amplification pattern different from that of Samba Mahsuri, were considered as impure rows or columns. Plants located on hills where the impure row(s) and column(s) intersected were considered as suspected admixture(s). The plants were grown till maturity and the genotype as deduced from marker profiles was verified with the phenotype as deduced from DUS test based on 33 characters (Shobarani et al., 2004). Of the 33 morphological characters studied, 23 were qualitative in nature and ten were quantitative in nature. The details of 33 morphological characters and their stage

of observation are shown in Table 2. [Published at end of paper] Impurities in the seed-lot were estimated based on visual examination of qualitative traits and by measuring the quantitative traits.

III. Results and Discussion

DNA fingerprinting methods based on polymerase chain reaction have become methods of choice for germplasm characterization, diversity analysis and seed purity assays (Sundaram et al., 2008). A variety of DNA markers are now available for fingerprinting cultivars and for marker assisted selection. Even though the utility of RAPD fingerprints for establishing genotype identity is known in case of rice hybrids (Qian et al., 1996; Wang et al., 1994), SSR markers are considered more reliable because of their ability to produce high-fidelity profiles as a result of their co-dominant nature and chromosome specificity (Nanda Kumar et al ., 2004). In the present study, DNA from 27 rice genotypes including Samba Mahsuri and other popular fine-grain type rice varieties developed by ANGRAU, Hyderabad, India were subjected to SSR marker analysis for identification of genotype specific markers/marker combinations for Samba Mahsuri. We have identified genotype specific SSR markers for Samba Mahsuri which can distinguish this variety from other medium-grain type slender grain types and demonstrated the utility of these markers in the assessment of purity in seed-lots of Samba Mahsuri and compared their efficiency with morphological markers. The results obtained in the study are presented and discussed below.

1) SSR marker polymorphism among the rice genotypes

Of the 113 SSR markers utilized in the present study, 26 displayed polymorphism among the 27 rice genotypes analyzed and amplified a total of 74 alleles with an average of 2.84 alleles per marker. The number of alleles detected by the polymorphic SSR markers varied from 2 to 5. These include five alleles with one marker (RM16649), four alleles each with 2 markers (RM12469 and JGT07-22.8), three alleles each with 15 markers (RM14250, RM26329, RM15630, RM6745, RM400, RM 10936, RM16153, RM16592, RM23237, RM14140, JGT01-16.2, JGT03-29.2, JGT 12-20.2, JGT11-16.3, JGT04-11.5) and two alleles each with 8 markers (RM24542, RM27840, RM19426, RM11229, JGT03-36.1, JGT02-9.8, JGT03-0.01 and JGT10-4.3).

The SSR marker JGT11-16.3, targeting a $(GATA)_n$ motif, clearly distinguished Samba Mahsuri from other medium slender varieties except MTU 2077, MTU 2067, Early samba, and WGL 14 by amplifying a 139 bp fragment (Figure 1), whereas another SSR marker RM19426 clearly distinguished Samba Mahsuri from the above mentioned four medium slender rice varieties by amplifying a 272 bp fragment (Figure 2). When these two markers were pre-PCR multiplexed, they could clearly distinguish Samba Mahsuri from all the other rice varieties and hence the fragments amplified by both these markers (i.e. in combination) could be considered as a molecular ID for Samba Mahsuri (Figure 3). This can be considered the first step towards DNA fingerprinting of elite, quality rice varieties like Samba Mahsuri and thus protect the IPR of the institutions developing such unique and well accepted varieties. Followed (Sundaram et al., 2008) a similar multiplexing approach for checking the purity of hybrid rice parental lines

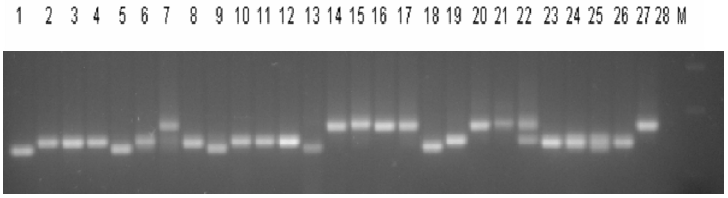

Figure-1 Amplification pattern of the marker JGT 11-16.3

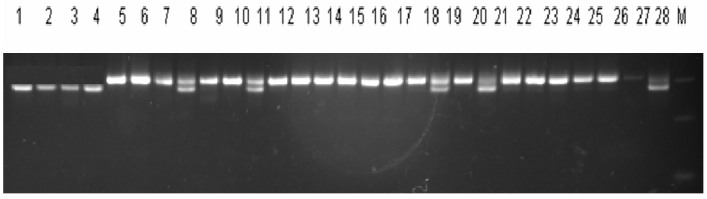

Figure-2 Amplification pattern of the marker RM 19426

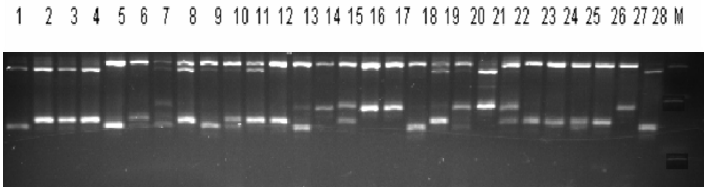

Figure-3 Amplification pattern generated through multiplex PCR of JGT 11-16.3 and RM 19426

2) Cluster analysis

A Cluster analysis carried out based on the similarity index data derived from the SSR markers grouped the 27 varieties into five major clusters (Figure 4).

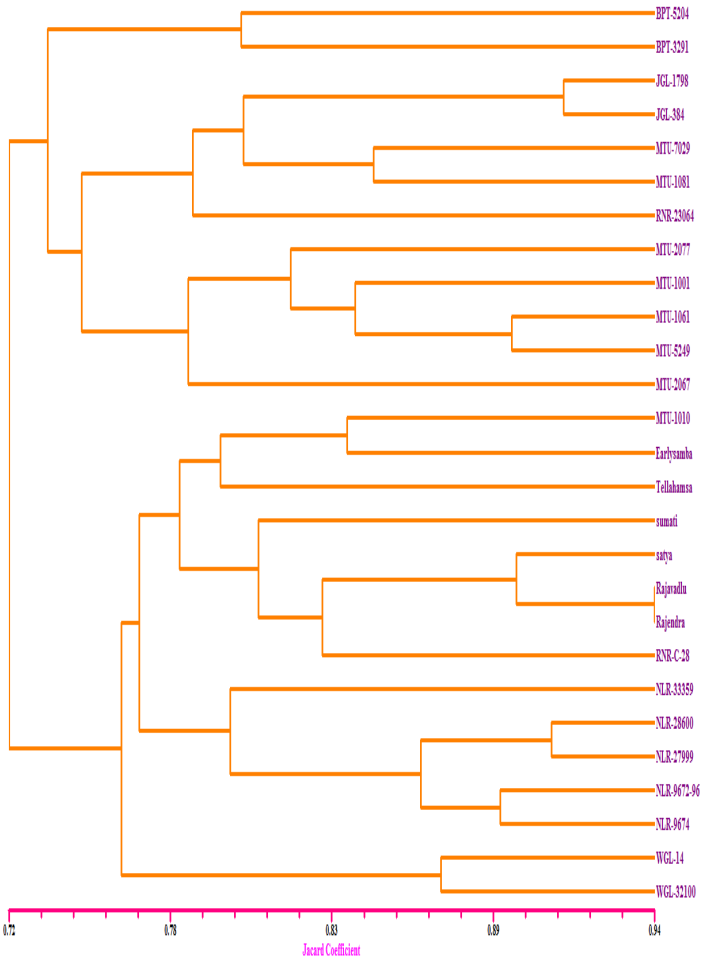

Figure -4 Dendrogram of 27 varieties by using 26 polymorphic microsatellite markers

The Jaccard's similarity coefficient ranged from 0.63 to 0.94 with an average similarity index of 0.74 indicating that most of the varieties used in the present study were having common ancestors in their pedigree. Samba Mahsuri shared maximum similarity with Sona Mahsuri; both are having Mahsuri as one of the parents in their pedigree. Cluster I comprised of two genotypes (Samba Mahsuri and Sona Mahsuri) sharing a similarity of 91%. Cluster II consisted of five genotypes viz., MTU 7029, MTU 1081, RNR 23064, JGL 384 and JGL 1798, which were further divided into two minor sub-clusters. Sub-cluster I with two genotypes (JGL 384 and JGL 1798) shared 83 per cent of genetic similarity, while sub-cluster II consisted of two genotypes (MTU 7029 and MTU 1081). RNR 23064 was observed to be divergent from MTU 7029, MTU 1081, JGL 384 and JGL 1798. The major cluster III comprised of five genotypes, in which the varieties MTU 1061, MTU 5249 shared a genetic similarity of 69%. Cluster IV was observed to be divided into two minor sub-clusters: sub-cluster I consisted of eight genotypes, among which Rajendra and Rajavadlu shared a maximum genetic similarity of 94%. All varieties developed at Nellore Agricultural Research Station were grouped under sub-cluster II in major cluster IV. Major cluster V consisted of two varieties (WGL 14 and WGL 32100), which displayed a genetic similarity of 81%. Samba Mahsuri shared maximum genetic similarity with BPT 3291 (91%), minimum similarity with JGL 1798 (67%) and the average genetic similarity with 26 varieties was about 74%. The genetic diversity among these varieties was observed to be very less. NLR 27999 (short bold) and JGL 1798 (medium slender) in cluster I and cluster IV, respectively shared a minimum similarity of 63%. However, phenotypically, grain character is different for these two varieties. Based on these results, it can be concluded that the allelic diversity revealed by 26 SSR markers is sufficient to distinguish the varieties based on their genetic constitution and grain characters. These results are on expected lines and could be explained based on a close pedigree and genetic relatedness of the varieties analyzed.

3) Purity assessment of seed-lots of samba Mahsuri

An attempt was made to validate the utility of Samba Mahsuri specific fragments of size 139 bp and 272 bp, amplified by the genotype specific SSR markers, JGT11-16.3 and RM19426, respectively for monitoring purity of different classes of seeds of the elite fine-grain type rice variety. All the 13 seed samples of Samba Mahsuri were analyzed with the two SSR markers JGT11-16.3 and RM19426 through pre-PCR multiplexing strategy involving bulking of samples grown in 6 x 20 matrices and through analysis of morphological characters as explained in materials and methods. The percentage of purity and number of contaminants in different classes of seed by molecular markers and morphological markers are given in Table 3 [Published at end of paper]. Impure plants in the SRTC breeder seed of Samba Mahsuri are displayed in the matrix (Figure5).

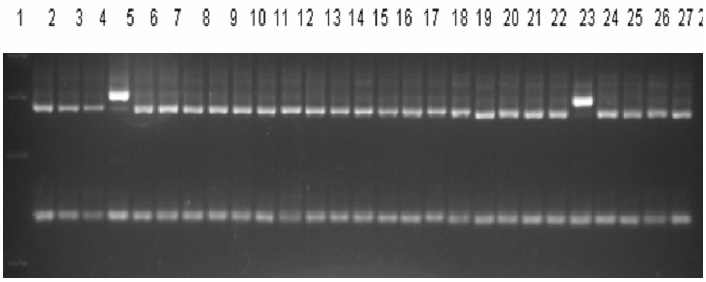

Fig -5: Amplification pattern of RM 19426 and JGT 11-16.3 in SRTC-Breeder seed (M values 1 to 27)

Based on the molecular marker analysis, the purity ranged from 83.1 % to 100 %. Nucleus seed showed 100% purity with out any contamination. The farmer seed sample II showed the least genetic purity (83.1%) probably due to the genetic admixtures resulting from the use of non-certified seed. As per the seed certification standards, nucleus seed should have 100% purity, whereas breeder seed, foundation seed and certified seed should have a minimum of 99, 98 and 97 % seed purity, respectively (Seed Management, 2003-IRRI). Nucleus and breeder seed showed genetic purity as per the seed certification standards, but the foundation and certified seed lots deviated significantly from the set standards. Since these are progenies of breeder seed, some contamination might have occurred probably due to cross pollination, mistakes in bagging and tagging and

lack of adequate care at the time of processing during seed production.

Of the 33 morphological characters analyzed, seven characters (days to 50% flowering, panicle number per plant, attitude of panicle branching, time of maturity and secondary branching, grain length and width) differentiated the impurities in foundation seed, certified seed and farmer seed. Among these, days to 50% flowering exhibited maximum variation (Table 3) [Published at end of paper]. Genetic purity detected by morphological markers ranged from 89 to 100%. The variant plants identified on the basis of morphological characters also showed variation on molecular basis. Similar observations were made in rice (Kalaichelvan, 2009). The percentage of contaminants detected based on SSR marker analysis was higher than those detected by conventional GOT assay (Keshavalu, 2006). Molecular markers detected some additional impurities, which were not detected through analysis of morphological characters. For example in SRTC breeder seed, one impure plant was detected based on SSR analysis, which was not detected by GOT analysis. This demonstrates the better discriminatory power and efficiency of SSR markers in genetic purity assessments and these markers could even accurately detect residual heterozygosity in the seed. Similar results have been reported by others (Nanda Kumar et al., 2004; Sundaram et al., 2008; Ye-Yun et al., 2005). Based on the results obtained from this study, it is clear that there is a need to critically assess the genetic purity of elite varieties of premium quality like Samba Mahsuri at each and every stage of seed multiplication and processing with the help of molecular markers so that the seeds cultivated by farmers are true-to-type and fetches premium price.

An important aspect of the present study is deployment of the strategy of pre-PCR multiplexing for assessment of genetic impurity in seed-lots of Samba Mahsuri. Pre-PCR Multiplexing is a cost saving strategy, wherein, analysis can be carried out simultaneously using two or more markers in a single PCR reaction tube, with negligible addition to the total cost of assay and with enhanced accuracy. Sometimes analysis using single markers (Nanda Kumar et al., 2004: Yashitola et al., 2002) may not help in accurate estimation of seed impurities in certain cases, and wherever possible, it is better to deploy more than one marker through multiplex PCR to get authentic results (Sundaram et al., 2008). Adopting the two-dimensional DNA sampling method is a cost-effective way of analyzing admixtures in seed-lots. Others (Nas et al., 2002; Sundaram et al., 2008) also demonstrated the utility of a similar two-dimensional DNA sampling strategy for assessing purity was demonstrated in the present study.

IV. CONCLUSION

Based on the results of the present study, it can be concluded that SSR markers can be efficiently used to generate locus specific allelic information which can serve as molecular IDs for elite rice varieties like Samba Mahsuri and such information can be used effectively in assessment of seed purity using these markers compared to conventional GOT, saving on time, resource and energy. The same strategy can be extended to other crops of economic importance, wherever SSR markers are available. The SSR marker information developed through this study is also helpful in the assessment of genetic contaminants in seed supply chain and under take necessary corrective action to supply the high quality seed to farmers.

Matrix showing impure plants in field in SRTC breeder seed

1	2	3	**4**	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21	22	23	**24**	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
41	**42**	**43**	***44***	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	**60**
61	62	63	**64**	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
81	82	83	**84**	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
101	102	103	**104**	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120

In the above gel, M - 100bp DNA ladder, lanes 1-20 represents bulked sample analysis column wise, lanes 21-26 represents row wise bulking and 27 represents the amplification pattern of the check variety BPT 5204. Black arrow represents impure row or column. Matrix represents 120 plants, bold colored numbers represents impure rows and columns and italic number represents impure plant.

Table-1: Details of twenty seven rice varieties analyzed in the present study

S No	Variety	Year of Release	Parentage	Grain type
1	Samba Mahsuri (BPT 5204)	1986	GEB 24 / T(N)1 // Mahsuri	MS
2	Sona Mahsuri (BPT 3291)	1982	Sona / Mahsuri	MS
3	Jagtial Sannalu (JGL 1798)	2002	BPT 5204 / Kavya	MS
4	Polasa Prabha (JGL 384)	2002	Kavya / BPT 5204	MS
5	Krishnaveni (MTU 2077)	1989	Sowbhagya /ARC 5984	MS
6	Vijetha (MTU 1001)	1995	MTU 5249 / MTU 7014	MS
7	Cottondora Sannalu (MTU 1010)	2000	Krishnaveni / IR 64	LS
8	Swarna (MTU 7029)	1982	Vasishtha / Mahsuri	MS
9	Chaitanya (MTU 2067)	1988	Sowbhagya /ARC 5984	MS
10	Indra (MTU 1061)	2007	PLA1100/MTU1010	MS
11	MTU 1081	Completed minikit trails	Ajaya/BPT 5204	MS
12	Vajram (MTU 5249)	1986	MTU 4569 /ARC 6650	LS
13	Early Samba (RNR M-7)	2000	Mutant of BPT 5204	MS
14	Satya (RNR 1446)	1987	Tellahamsa / Rasi	LS
15	Sumati(RNR 18833)	2002	Chandan / Pak Basmati	LS
16	Rajavadlu(RNR 99377)	1993	Rajendra / IR30	LS
17	Rajendra (RNR 12329)	1976	IJ52 / TN1	MS
18	Warangal samba (WGL 14)	2005	BPT 5204 /ARC 5984 // BPT 3291	MS
19	Warangal Sannalu (WGL 32100)	2007	Divya / BPT5204	MS
20	RNR –C-28	Completed minikit trails	IR 64 / IET 9994	LS
21	Taramati (RNR 23064)		BPT 5204 / Tellahamsa	MS
22	Shravani (NLR-33359)	2000	Selection from IR 50	LS

23	Simhapuri (NLR-28600)	1991	RP 5-32 / Bulk H9	SB
24	Tikkana (NLR-27999)	1988	RP 31-49-2 / BCP 2	SB
25	Pinakini (NLR-9672-96)	1987	Bulk H 9 / Millekkunning	MS
26	Kottamolagolukulu(NLR-9674)	1982	Bulk HG 9 / Millekkunning	SB
27	Tellahamsa (C10754)	1971	HR 12 / T(N)1	MS

Table-2: Details of the 33 DUS characters considered in the present study, characters which are shown variations are present in italics

S. No.	Character	Stage of observation
1	Basal leaf: Sheath colour	Booting
2	Leaf: Anthocyanin colouration	Booting
3	Leaf: Pubescence of blade surface	Booting
4	Leaf: Auricles	Booting
5	Leaf: Colour of ligule	Booting
6	Leaf: Length of blade	Booting
7	Leaf: Width of blade	Booting
8	Culm: Attitude	Booting
9	*Time of Heading (50% of plants with panicles)*	*Half of inflorescence emerged*
10	Flag leaf: Attitude of blade (Early observation)	Beginning of anthesis
11	Spikelet: Density of pubescence of lemma	Beginning of anthesis
12	Lemma: Anthocyanin colouration of keel	Anthesis half way
13	Lemma: Anthocyanin colouration of area below apex	Anthesis half way
14	Spikelet: Colour of stigma	Anthesis half way
15	Stem: Thickness	Milk development stage
16	Stem: Length (excluding panicle)	Milk development stage
17	Stem: Anthocyanin Colouration of nodes	Milk development stage
18	Stem: Intensity of anthocyanin colouration of nodes	Milk development stage
19	Stem: Anthocyanin colouration of internodes	Milk development stage
20	Panicle: Length of Main axis	Milk development stage
21	Flag leaf: Attitude of blade (Late observation)	Ripening
22	Panicle: Curvature of main axis	Ripening
23	*Panicle: Number per plant*	*Ripening*
24	Spikelet: Colour of tip of lemma	Ripening
25	Panicle: Awns	Ripening
26	Panicle: Length of longest awn	Ripening

27	Panicle: Presence of secondary branching	Ripening
28	*Panicle: Secondary branching*	*Ripening*
29	*Panicle: Attitude of branches*	*Ripening*
30	Panicle: Exertion	Ripening
31	*Time maturity (days)*	*Ripening*
32	*Grain: Length*	*caryopsis hard*
33	*Grain: Width*	*caryopsis hard*

Table-3: Comparison of percent purity on the basis of molecular and morphological data

S. No.	Class of seed	% of Morphological purity					% of Molecular purity		
		No of contaminants	% purity	Impure plant No	Variation in morphological character		No of contaminants	% Purity	Impure plant No.
1	RRU BPT-Nucleus seed	0	100	-	---		0	100	-
2	RRU BPT-Breeder seed	0	100	-	---		2	100	-
3	Kampasagar-Breeder seed	0	100	-	Days to 50% flowering, Time of maturity		4	96.6	24, 37, 84, 97
4	SRTC-Breeder seed	0	100	-	---		1	99.1	44
5	APSSDC-Foundation seed	0	100	-	---		3	97.5	87, 90, 120
6	NSC-Foundation seed	1	99.1	16	Days to 50% flowering, Time of maturity		4	96.6	9, 10, 16, 20
7	ANGRAU-Foundation seed	2	98.3	23, 53	Days to 50% flowering ,Time of maturity		4	96.6	23, 29, 53, 79
8	Jangamaheswarapuram-Foundation seed	2	98.3	89, 111	Days to 50% flowering, Time of maturity and panicle number per plant		5	95.8	1, 89, 111, 116, 119
9	APSSDC-Certified seed	2	98.3	3, 101	Days to 50% flowering, Time of maturity and panicle number per plant		5	95.8	3, 46, 85, 101, 120
10	NSC-Certified seed	3	97.5	71, 82, 97	Days to 50% flowering, Time of maturity and panicle number per plant		12	90	62, 65, 67, 71, 72, 77, 82, 85, 87, 91, 92, 97
11	Farmer sample-1	6	94.1	71, 93, 100, 109, 114, 116	Presence of secondary branching, days to 50% flowering, attitude of panicle branching		12	90	2, 22, 46, 71, 84, 93, 94, 100, 106, 109, 114, 116
12	Farmer sample-2	13	89.1	7, 11, 18, 25, 27, 31, 38, 47, 51, 58, 107, 111, 118	Grain length and width, days to 50%flowering, panicle number per plant		20	83.1	5, 6, 7, 11, 18, 25, 26, 27, 31, 38, 45, 46, 47, 51, 58, 105, 106, 107, 111, 118
13	Farmer sample-3	3	97.5	23, 92, 108	Days to 50% flowering, attitude of panicle branching.		3	97.5	23, 92, 108

Supplementary table 1- List of SSR markers used for identification of genotype specific marker for BPT 5204

S. No	Marker	Chromosome	Forward primer sequence	Reverse primer sequence
1	RM 10936	1	ACGGTTTGGAAGTGTTCGTAGG	TGGTACTGCATAATCTCAGCATCG
2	RM 11229	1	TGACAGAAACAAAGCGGAAGG	TCCAAACCGCTATTCTTGTAGC
3	RM 3917	1	CGGACGACTCCGACAACACG	CGAACGAACGAGGACGAACG
4	RM 17600	1	CAGCATCAAGGGTCGCTACACG	CGATGGCGTGGGATAATTACGG
5	RM 11278	1	ACTTCTTGTAGCACTGCACCTTCG	CCTCGGCAACTGCTTCAAGG
6	RM 10649	1	GACCAATATGGAGTGGTCTCTTGG	ACGCTCCGCCTGCTATTTAGG
7	RM 8126	1	TGGGCCTCTTTGTTTCATACTCC	TCCTCATCTCTCTCCGTGTCTCC
8	RM 11865	1	CAACTCATTCGGCTTCTCTTTCC	GCAATTGCATGAGTTGGAATGG
9	RM 6942	2	CGAACTCCCAATTACAGATCACG	TCATCCAAGACTTAGGTGGAGAAGC
10	RM 13155	2	TACCAACAGGGAGTTGTCTCTCG	AGCGACGGTGTAAAGAATAAGTCG
11	RM 6933	2	AATGCCTAGCACTCATCCTTGC	AGGCACCCTACGATGAAATAGTGG
12	RM 12469	2	ACTCCATCGAACCCTGTTAGAGC	GTCCATGTTTGCTTACGTGTTTGC
13	RM 13154	2	GGTACTTAGCGTGCAACTTTAACC	TAGGTAACTAGACGAAGCGATAGAGG
14	RM 13812	2	AACACTAACCGGGACTAAAGATCG	CTGAGATAACGACTACTGCTACATCG
15	RM 6938	2	CCGATTAGCGATTGATATGGAGTAGG	AGTGCACAGCCATGGAATTATGC
16	RM 14140	2	CCTCCCTCTCCAAACACATTGC	TCATCAGCAGACAGAATGTTGACC
17	RM 15679	3	TAGATGTATGAGTCGGAATGGAGTCG	CAGACGCAGTGTGTGTATGAAGTTCC
18	RM 15630	3	AACTCGAAGGATCTCGCCCAACC	ACCCACCTCCTCACGCTGTACG
19	RM 14250	3	GATTACTGCCGATTCGATAGC	AAATGGGACATGTTCTCTCG
20	RM 14931	3	GCTTCACGCACATATTGCCTTCC	ATCCTCCCACTCCCAAATTATCTCC
21	RM 15331	3	GGTTCGGTGCTTTCTCTTTCAGC	AGCGATGCCGTCCTTACACC
22	RM 5928	3	CTTACCTTCTGAAATGGAGGTAGC	CAAGGTGAAAGACGAAGAATGC
23	RM 15831	3	AACACAATTCACCGGTCCTTAGC	ACGTGGGTGATCGTGTCTGC
24	RM 15727	3	ATCTGTGCGGACCACCATGC	CGCTACGAGCAGCGTACTTAAAGC
25	RM 15466	3	CGGAGTGATCATCAGAGTCG	GTAGAATCGCATGTAGAGTCTTAGGG
26	RM 16153	3	TGGTTGTGGTATAGCACGGTAAGC	TGACCCAAGGAGATACTAGGTTGC

27	RM 15288	3	GAGAGGCTCCTACTGCGGTTGC	CCTTTCCATCTCCTTGTTTCTCTGC
28	RM 6759	3	CCCGAGTCTTCATAGAGATATTCC	ATCCCTAGCTAGCCTTCCTTCC
29	RM 14582	3	ATAACCAAGATCTCGGCCCATCC	CCCATGTCACCGACAACTATCTAGC
30	RM 14320	3	CACCTGTAAATTAGGACACTGG	CAGTGTACTTTGAACTGCCTAGC
31	RM 14860	3	GGAAGGTGATTTCATCCGGTAGC	TGGCATGTTTAATGCTGGTTCG
32	RM 16606	4	TGCACTTCTTTAGAGTAGGAGGAAGC	CATGCATGTGTCCAAAGATTCG
33	RM 16592	4	CTTAGCACGGACACTCATATTTGG	CACAATACGTTTGATGGCTTGC
34	RM 16801	4	CGTTCAAGGAGCTTGTGTTGATCC	GGACCGATTTAAGTGAACGTTGATGG
35	RM 17127	4	CTCAATGTTTCCCACAGTTACCG	TGTGTTATGTGTGCGTGATGAGC
36	RM 16652	4	TGACATTAGTTGTGGCAGATCC	CCTAGAATCTCATCTGTCTTCTGG
37	RM 16649	4	CTCCCTTCATGCGTAAGCTCTCC	GCAAACAGGATCCTCCACAAAGG
38	RM 16343	4	AGAACCAGCAGTTTCTTTGC	TTCTATCCCTACAGTCTTGACAGC
39	RM 16913	4	GTGTACGTGTTGGCTCTCTGTACG	GATGTTGCTTGTGCTGCAACC
40	RM 18222	5	TGATTCCTCTATATGCAGCCTTGG	TATCGTGGTTTCATCGTGTGTGC
41	RM 5592	5	CTCGCTCTTACAACTTTCAAGC	CACTTACCTCCACTTCTCAACC
42	RM 18857	5	ACCGGTCCTTAGCTTCTGAGC	ACCAACCGGGACTAAAGATCG
43	RM 18888	5	GGTGCAAATGCTTGTAGTCCTATGC	GCACCACCAACCTAACAGTGAGC
44	RM 18065	5	CGATGGTGAGTGGTGATTCATGC	ATCATCCGCGCATTAGCATTTCC
45	RM 20060	6	CACACATGAGTGGTTAGGTAAGATGC	CAGTGACAAGAGCGAAATGATCC
46	RM 20458	6	GTGTGGTGGTCTTGGGATGTTGG	TATAGCCCAGCAAGGGTGGTACG
47	RM 20196	6	GTGGACCCACTACACAACATGG	TGTTCTACACTACGCCATTAGGC
48	RM 19697	6	AACAACCTGAGAACACCTCTTGG	GGACAAACACATGGTGATCTGC
49	RM 20096	6	CGGTAAGCCATAAATAGATCCCAAGG	TTTGAACAGCGACACGGTTTCC
50	RM 19682	6	CCCACTGATTGACCAAGAGC	TTTGCTAAAGGAGCACTACGC
51	RM 19426	6	CGGTGTCTTCTTTAAACAGC	ATGGATAAGCGGTATGTTCC
52	RM 400	6	TTACACCAGGCTACCCAAACTCG	TTGCTGAGTTCCCTCGTCTATCC
53	RM 19660	6	TTTGTCCCTGCCGTACTTGC	AGCCACGTTGGGTGAAATTAGC
54	RM 20818	7	AGATGCAGATAGATGCATGTCACG	ACCGATCATCCACGATCCTACG

55	RM 6697	7	TATTCCCGGGAGATCCAACAGC	AAGATCCAGTCGATTTGGTTCAGG
56	RM 21069	7	ATCTAGTACCGGATGTGACACG	AGACAGAGGCATGACAGAAAGG
57	RM 5720	7	GACTCGTCACTGACACTGATACG	CTTGTTAGGAAGAGCATTCTGC
58	RM 21258	7	TATCATTCCGGTCCAAAGTGTCG	TCCGGTCCAAAGTCTCATTTGC
59	RM 21564	7	CGACGGAAACACAATTCATAGG	ACCAACCGGGACTAAAGATCG
60	RM 21813	7	NA	NA
61	RM 23237	8	TAAAGCATTTGGACGGTGGATGG	GAGGTGGGTGTGACCCTTGG
62	RM 22622	8	TAGGCCGTTCTGACGTAATACCC	CAGTGATGGTGATGCGATTTAGC
63	RM 22585	8	CACCGATTATTGTCGTATGG	AGTGAGGAAGGGAAGAATACG
64	RM 5933	8	AGCGATTCAGAACGAATCAACG	TGCCAAAGCTACACAAATCTGACC
65	RM 6699	8	TGGTATCACCATTCACCAACACC	TCCTTACGCTACCACGACTTCC
66	RM 24542	9	ATCCACAAGAGCACCGATGAGG	TGACCTGGTAGTGGTGAGTGTGC
67	RM 24038	9	GTCCGGGCAACAATAACAAAGC	CACCAACCGGGACTAAAGATGG
68	RM 24780	9	GACTAGCCAGCCAAGGTTTGAGC	TGCTGCATGTGTGTATGTGACTACG
69	RM 25368	10	TATAGTTAAGGGAGCCACGCAAGC	CCACCTCGTAAGAACATGGAGAACC
70	RM25328	10	CATTTGATCTGTCCGGCTCTAGG	ATTGCTAGTGGGTGATGATGTGG
71	RM 6745	10	GAAGCCTCCTTGATGTTGACTGG	CAACGATTATGCGTCCTTAGATGC
72	RM 26329	11	TAACCGGGACTAAAGATAGAGC	CTACGTCGAAATCGTAACTAGC
73	RM 26086	11	CAGCGCTTTGTACATCTCACTCC	GCATCCTCTGCACCTATTCTTGG
74	RM 25970	11	TGTGAATCTGACTTGCCAATCG	AAGGCTGTTATGTGTCTTGCAACC
75	RM 27311	11	TTACCAACCGGGACTAAAGATCG	CAATTCATAACGTCGGTCCTTCC
76	RM 27034	11	AGGCCCTCGCGTGTACATACC	ATCCGACCCACGGTAATCTGAGG
77	RM 27101	11	CGGGCACCATGGTATCTAATGC	ATGCTCGGTCCACAGAGAATAGG
78	RM 28157	12	GCTTAATTTCTGACAGACCAGTGC	GATCTAAACACAGCCTTCCTTGG
79	RM 27840	12	TTTGCGTGCTAGGGAGATTAGC	CATTATGTACTTACTCCCTCCCTTCC
80	RM 27406	12	TGGTAGGTGTGCAATAGAAGTAGG	AATGCATGCAAACACAGTGG
81	RM 3103	12	CTGGAGTGGAGAAGAGAGAACAGG	TCTCCGCTCGGTTTCATCTAGG
82	RM 28024	12	CCGGGTGTTTGACCTAGTTTCC	CACGATGGTCGTCTTGTTACTGC

83	RM 27864	12	NA	NA
84	RM 28464	12	CTATATCCGCACCAATGGAAACC	TAGGAATTGCGACGAGTTTGTGG
85	RM 27861	12	TGCTAGCATGCTCATCCACTTGC	AATTGCCTCTGTAATACCGCCCTTCC
86	RM 27905	12	AACACAATTCACCGGTCCTTAGC	ATCGATCTTTAGTCCCGGATTGG
87	RM 27882	12	NA	NA
88	JGT 01-16.2	1	TGTTGGCTGAATGATTGATGGATAGA	GGACCTATGTGGCACGAATTTGAA
89	JGT 01-23.1	1	CGGCGTAGGAGTATTTGTAGGAAGTT	GCGGCAATGATGATGGTGGA
90	JGT 01-29.9	1	TCGATACTTTCGGAAATGGGAAGG	GCTAGCTCTAACATCCAAGCTTCC
91	JGT 02 - 35.2	2	CGGGGCGAATCGAGGTCAG	TCGCGGTGGATTAGATGCCTC
92	JGT 02-9.8	2	GAGCGTTTTAAGGAAGCCACATG	CACCGAGTCCCAAGCTGATGA
93	JGT 03-0.01	3	CCGATGCACCAACACCCTCAC	GACGCATGGTGAGTTTGGAGTGAT
94	JGT 03 - 36.1	3	GCGGCAACACGACCAGCTT	CATCCTGAGTTTTGAGAAGCACCATA
95	JGT 03-29.2	3	CCTGTAACCTGAACGGAGAGAGTA	TGAAAAAAAAAAGTCTCCCAATGTC
96	JGT 04 - 11.5	4	AGGGCTCGTAGAAAGTGTCCAATTAG	CCGCTGGGTATTGCCTAACTG
97	JGT 04-28.5	4	TTGAATAGATCGCACCGTTGAC	AAAAGGGCAGCTACGTCTGAGC
98	JGT 04 - 34.9	4	ACGAACTCTAGTCCCTCCGCTC	ACGACATTATCCTAACACCACG
99	JGT 05 -05	5	TGTAGGAAGACATGTGGCAGCT	TACAACTAGATGCAATGGCACC
100	JGT 05-7.8	5	ACGGTGAGATTAGGTATTGCCA	AGTTCGGACAAGGGGGAGTACT
101	JGT 06-18.1	6	GAAGAACGTGGTTTGCATCGT	TGTACCGTCGGCGAAGAACGCACC
102	JGT 07 22.7	7	AAGGGGTTATTGATTTAGTCCATTT	AGGGGTTGAACTTTGATGCTAATATG
103	JGT 07-22.8	7	TGGCGATCTAGGAGCGTCTGT	TGTAAACATTTCAAAAGGGCACTAA
104	JGT 08-19.5	8	TTCTGAAAAAGCTCTGACCAAGC	ACTAGCTACATGCTGCAGTGCAT
105	JGT 08-10.2	8	GTACTAAACAGATGGAGCAGTACG	GTGGTACTGCTAATTAGGGGGTA
106	JGT 08-16.7	8	GAAAGAATTTGTTCCCACGACCTC	GCTACAATTGCATGCTGGATGTCG
107	JGT 09-4.8	9	CGGGATACTAACAGCAAGCAAAT	TCTCGCACCACTAACAGGGAACA

108	JGT 10-0.3	10	TGGCGATTTAGGAGCATTTGTAGAA	GCCAACGACCACTCTCACCCTAC
109	JGT 10-4.3	10	TGGCGACTTAGGAGCGTTTGTAG	GCGGCACTTCCCCAAACAA
110	JGT 11-16.3	11	GGCGGCGTATTAGCGTTGTA	AGGTTCTAGCCCATGTTAAATCTTCT
111	JGT 12-20.2	12	CTTACATGGATATGCACATCGAGC	CAAGGGAGCACACGAACAACAG
112	JGT 12-7.7	12	GAGCGTTTTGAGGAAGTTTAATGGAC	GGCACCGTTTGATTTAGATTATTACC
113	JGT 12-25.9	12	GGCGGCATAGGAGTGTTTGTAG	CGCGGTGGTCGGGATGAG

ACKNOWLEDGMENT

We thank very much to Directorate of Rice Research, Acharya, N. G. Ranga, Agricultural University, Hyderabad 500 030, India.

REFERENCES

[1] Botstein D, White R L, Skolnick M and Davis R W (1980). Construction of genetic linkage map in man using restriction fragment length polymorphism (RFLP). American Journal of Human Genetics. 32: 314-331.

[2] Dellaporta SL, Wood J, Hicks JB (1983). A plant DNA minipreparation. Plant Mol Biol Rep. 1:19–21

[3] Kalaichelvan C (2009). Studies on identification of rice (Oryza sativa L.) cultivars using morphological and molecular markers. M.Sc thesis, Acarya N G Ranga Agricultural University, Rajendranagar, Hyderabad.

[4] Keshavulu K (2006). Identification of microsatellite markers for distinguishing elite fine grain rice varieties and hybrids and their utilization in seed purity assessments. Ph.D. thesis, Tamil Nadu Agricultural University, Coimbatore.

[5] Mao CX, Virmani SS, Kumar I (1996). Technological innovations to lower the costs of hybrid rice seed production. In: Virmani SS et al (eds) Advances in hybrid rice technology. Proceedings of Third International Symposium on Hybrid Rice, Directorate of Rice Research, Hyderabad, India

[6] McCouch S R, Chen X, Panaud O, Temnykh S, Xu Y, Cho Y G, Huang N, Ishii T and Blair M (1997). Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Molecular Biology. 35: 89-99.

[7] McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono Z, Yano M, Fjellstrom R, Declerck G, Schneider D, Cartinhour S, D.Ware D, Stein L (2002). Development and Mapping of 2240 New SSR Markers for Rice (*Oryza sativa* L.). DNA Res 9:199–207

[8] Mishra B, Viraktamath BC, Ilyas Ahmed M, Ramesha MS, Vijayakumar CHM (2003). Hybrid rice research and development in India. In: Virmani SS, Mao CX, Hardy B (eds) Hybrid rice for food security, poverty alleviation, and environmental protection. Proceedings of the 4th International Symposium on Hybrid Rice, 14–17 May 2002, Hanoi, Vietnam. Los Baños (Philippines), International Rice Research Institute, pp 265–283

[9] Nagaoka T and Ogihara Y (1997). Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. Theoretical and Applied Genetics. 94: 597-602.

[10] Nandakumar N, Singh AK, Sharma RK, Mohapatra T, Prabhu KV, Zaman FU (2004). Molecular Wngerprinting of hybrids and assessment of genetic purity of hybrid seeds in rice using microsatellite markers. Euphytica. 136:257–264

[11] Nas TM, Toledo RS, Li Z, Virmani SS (2002). Two-dimensional DNA sampling: a new strategy for assessing hybridity using molecular markers. Paper presented at the 4th International Symposium on Hybrid Rice, International Rice Research Institute, Hanoi, 14–17 May 2002

[12] Qian Q, Chen H, Sun Z X and Zhu L H (1996). The study on determining true and false hybrid rice II You 63 Using RAPD molecular markers. Chinese Journal of Rice Science. 10: 241-242.

[13] Rafalski J A, Vogel J M, Morgante M, Powell W, Andre C and Tingey S V (1996). Generating and using DNA markers in plants. In : Birren B and Lai E (eds) Analysis of non-mammalian genomes – A practical guide. Academic Press, New York, pp 75-134.

[14] Rajendrakumar P, Akshaya Kumar Biswal, Kannabiran Sakthivel, Maganti Sheshu Madhav, Chirravuri Neeraja Sena M. Balachandran, Kommoju Srinivasarao Podishetty Natarajkumar Yadla Hari Kalidindi Sujatha Raman M (2009). Targeting (GATA)n repeat motifs in rice. Euphytica. DOI 10.1007/s10681-009-9964-x.

[15] Rohlf EJ (1993). Numerical taxonomy and multivariate analysis system, version 1.80- Exeter Software, Setauket, NY.

[16] Salimah S S, De Oliveira A C and Godwin I D (1995). Assesment of genomic origin and genetic diversity in the genus Eleusine with DNA markers. Genomics. 38: 757-763

[17] Shobarani N, Subba Rao L V, Viraktamath B C and Mishra B (2004). National Guidelines for the conduct of tests for Distinctness, Uniformity and Stability. Technical Bulletin No. 8, Directorate of Rice Research, Hyderabad.

[18] Sundaram R M, Naveenkumar B, Biradar S K, Balachandran S M, Mishra B, Ilyas Ahmed M, Viraktamath B C, Ramesha M S and Sarma N P (2008). Identification of informative SSR markers capable of distinguishing hybrid rice parental lines and their utilization in seed purity assessment. Euphytica, 163: 215-224.

[19] Wang G S, Csatigline S, Zhang J, Fu R, Ma J S, Wenbin L, Youngrusuh and Sala F (1994). Hybrid rice (Oryza sativa L.): Identification and parentage determination by RAPD fingerprinting. Plant Cell Reports. 14: 112-115.

[20] Yashitola J, Thirumurugan T, Sundaram R M, Naseerullah M K, Ramesh M S, Sarma N P and Sonti R V (2002). Assessment of purity of rice hybrids using microsatellite and STS markers. Crop Science. 42: 1369-1373.

[21] Ye-Yun X, Zhan Z, Yi-Ping X and Long-Ping Y (2005). Identification and purity test of super hybrid rice with SSR molecular markers. Rice Science. 12 (1) 7-12.

[1] Department of Agricultural Biotechnology, College of Agriculture, Rajendranagar, Acharya. N G Ranga Agricultural University, Rajendranagar, Hyderabad 500 030, India.

[2] Crop Improvement Section, Directorate of Rice Research, Rajendrangar, Hyderabad 500 030, India.

***Correspondence Author:** K. Aruna Kumari, e-mail: arunaagbsc@gmail.com