Indian Journal of Biotechnology
Vol. 19, October 2020, pp 299-310

Identification and characterization of polymorphic genic SSR markers between cultivated (*Oryza sativa*) and Indian wild rice (*Oryza nivara*)

Umakanta Ngangkham[1,5]*, Jawahar Lal Katara[1], Shanmugavadivel P S[2], Manoj Kumar Yadav[1], Suresh Yadav[3], Nongthombam Devachandra[4], Sanghamitra Samantaray[1], Lotan Kumar Bose[1]

[1]ICAR-National Rice Research Institute, Cuttack-753006, Odisha, India

[2]ICAR-Indian Institute of Pulses Research (ICAR-IIPR), Kanpur, Uttar Pradesh 208024, India

[3]ICAR-Indan Agricultural Research Institute, New Delhi 110012, India

[4]College of Horticulture and Forestry, CAU, Pashighat-791102, Arunachal Pradesh

Received 3 August 2020; revised & accepted 12 September 2020

In the present study, we developed a set of 100 BC_2F_4 mapping lines derived from the cross between the *Oryza nivara* (AC100476) and high yielding *indica* rice, Lalat. Out of 410 RM markers used for polymorphism survey between the parental lines, we identified around 113 (28.9%) polymorphic rice microsatellite (RM) markers between the parental lines that were uniformly distributed among the 12 chromosomes except for few gaps on chromosome 1, 7 and 10. On the basis of motif length, the trinucleotide repeats-motif, (TTA)*n* was the longest with a maximum motif length of 177 nucleotides. Among the repeat motifs, di-nucleotide repeat-motifs was the most abundant (68.14%) with the motifs (AC/GT)*n* were the most abundant accounting 56.32% of the total dinucleotide repeat-motifs. Out of the 113 polymorphic RM markers, 56 (49.55%) were found to be genic markers and broadly distributed into three groups such as molecular, biological and cellular categories with 38%, 32% and 30% frequencies, respectively. The present finding would be useful for the identification and mapping of drought related traits and development of drought tolerant rice cultivars in rice breeding program.

Keywords: Indian wild rice, *Oryza nivara*, high yielding indica rice, *Oryza sativa*, gene introgression, mapping population

Introduction

Rice (*Oryza sativa* L.) is the main food crop providing food and nutrients for more than one-half of the world population[1]. Considering the diminishing arable land in the world along with present world population growth rate, improvement of rice productivity is the utmost goal of the rice breeding program. Due to climate change and several biotic and abiotic stresses, the challenges for increasing the rice production and productivity are becoming difficulty. In order to overcome these constraints, we need to develop genetically superior rice varieties having multiple resistances to biotic and abiotic stresses by adopting modern molecular tools and approaches[2]. During domestication and continuous selection of rice by humankind, genetic variability and important traits for biotic and abiotic stresses have been dwindling in the present rice cultivars. This is one of the main reasons of rice cultivars for becoming highly vulnerable to the biotic and abiotic stresses. However, wild rice and its relatives constitute an invaluable gene pool in terms of resistance/tolerance to biotic and abiotic stresses, which can be exploited for the development of sustainable and resilient varieties against the adverse climatic changes and biotic and abiotic stresses[3-4]. Development and deployment of rice varieties introgressed with resistance/tolerance genes for biotic and abiotic stresses is the most cost-effective and environment-friendly approach[5]. Therefore, the enormous rice genetic diversity available in the wild rice will be one of the best options for the foundation of the genetic improvement of the rice cultivars by introgressing them and unraveling the new genes and traits. It is also required to mine out genes for several agronomical traits including resistance to biotic and abiotic stresses and quality and productivity of wild rice. Unlike to perennial wild rice, Indian wild rice (*O. nivara*) is immediate wild progenitors of Asian cultivated rice *O. sativa* and annual with agregarious habit, photoperiod insensitive, synchronous flowering of tillers and bold seeds. Occurrences of wide genetic diversity in the *O. nivara* accessions from different parts of India and abroad have been reported[6-10].

*Author for correspondence:
ukbiotech@gmail.com
Present address: ICAR Research Complex for NEH Region, Umiam-793103, Meghalaya, India

Several attempts have been made for development of superior rice cultivars by introgressing genes/QTL for important traits and biotic and abiotic stress from *O. nivara* have been successfully practiced through conventional breeding methods and marker-assisted selection[11]. Considering the facts and importance, we are reporting the development of BC_2F_4 mapping lines derived from the cross between the *O. nivara* (AC100476) and high yielding *indica* rice, Lalat for future mapping and characterization of agronomically important genes in rice. Further, we identified around 113 polymorphic rice microsatellite markers between the parental lines which might be useful for identification and characterization of the gene of interest.

Materials and Methods

Plant Material and Phenotypic Evaluation

In the present study, an accession of *O. nivara* (AC100476) having drought tolerance during the vegetative stage was taken and crossed with an improved *indica* rice cultivar, Lalat as a recurrent parent to develop advanced mapping population. A total of 100 BC_2F_4 mapping population as an advanced mapping population were developed through two times back-crossing with Lalat and then by selfing up to F_4 generations. These 100 inbred lines (ILs) along with parental lines such as Lalat and *O. nivara* (hereafter referred as *nivara*) were grown in the rice research field of ICAR- National Rice Research Institute (NRRI), Cuttack, Odisha, India following standard agronomic practices. The rice grains were sown in the nursery and transplanted after 21 days of sowing in the main field of the ICAR-NRRI, Cuttack, India.

Genomic DNA Isolation, SSR Markers and PCR Amplification

The genomic DNA of Lalat and *nivara* were isolated from the young seedling following the modified the cetyltrimethyl ammonium bromide (CTAB) protocol. Here, young leaf of Lalat and *nivara* rice plants was collected and grounded in liquid nitrogen. This is followed by incubation in CTAB extraction buffer (800 µl) at 65°C for duration of 50 – 60 min with mixing 3 - 4 times. The extract mixture was mixed with an equal volume of phenol: chloroform: isoamyl alcohol solution (25:24:1) and centrifuged at 12,000 rpm for 10 min at room temperature. The aqueous phase was separated and again mixed with an equal volume of chloroform: isoamyl alcohol (24:1) solution and centrifuged at 12,000 rpm for 10 min at room temperature. The final aqueous phase was transferred to a fresh centrifuge tube and genomic DNA was precipitated overnight with equal volume of chilled isopropanol. The precipitated DNA pellet was washed with 70% ethanol by centrifuging at 8,000 rpm for 7 min at room temperature. The DNA pellet was then air-dried and dissolved in 100 µl of TE buffer (1X). The genomic DNA was diluted at the concentration of 20 - 30 ng/µl for PCR amplification by quantifying them through the agarose eletrophorosis method.

A set of 410 rice microsatellites (RM) which are simple sequence repeats (SSR) covering uniformly the whole 12 chromosomes of rice were collected from the gramene marker database (https://archive.gramene.org/markers/) for identification of polymorphic RM markers between the two rice genotypes, *O. sativa* (Lalat) and *O. nivara*. The primers sequences of both the forward and reverse of RM markers was downloaded from the gramene web database. The PCR reaction mixture was prepared in a 10 µl reaction volume containing of 20 ng of genomic DNA, 10 pmol of both forward and reverse primers, 0.2 mM of dNTP mix, 1U of *Taq* polymerase (Bangalore Genie, India), 1X *Taq* buffer with 1.5 mM of $MgCl_2$. The PCR reaction program was set using a thermal cycler (Eppendorf, USA) with the following cycling steps; initial denaturation at 94°C for 5 min followed by 35 cycles of denaturation at 94°C for 45 sec, annealing at 55°C for 30 sec and polymerization at 72°C for 1 min and completed the reaction by a final extension at 72°C for 10 min. The amplified PCR products were detected in a 3.5% metaphor agarose gel (Lonza, Switzerland) stained with ethidium bromide. Gel pictures were visualized using the gel documentation system (Alpha Imager, USA). The PCR band showing different sizes between the parental lines, *indica* (Lalat) and *nivara* are considered as polymorphic RM markers. The distribution of polymorphic RM markers with the physical position along the 12 chromosomes was graphically presented using MapChart v2.32[12].

Identification of Genic and Non-Genic Polymorphic SSR Markers

The primer sequences of both the forward and reverse of RM markers were used for queries against the rice genome database using basic local alignment search tool (BLAST) online tool in the gramene web database. An RM marker is considered as a genic marker only when both the forward and reverse sequence of the RM marker overlaps with a gene or aligned within a gene sequence. The IDs of the gene

locus for GO (gene ontology) annotation were obtained from the rice genome annotation project; RGAP database (http://rice.plantbiology.msu.edu/annotation_pseudo_goslim.shtml) and used for GO analysis using WEGA 2.0[13] (http://wego.genomics.org.cn/). The genes showing polymorphism between the two rice genotypes were classified as three categories (biological process, molecular function and cellular component).

Results and Discussion

Compared to cultivars, wild rice and its relative are not much superior in yield and grain quality. However, many wild relatives of rice have been considered as hidden genetic resources for several agricultural importance traits including abiotic and biotic stresses. Due to the prevalence of very narrow genetic variability among the cultivated rice, they are highly vulnerable to several biotic and abiotic stresses. Therefore, several attempts have been practiced to broaden the genetic bases of presently cultivated popular rice by transferring useful genes for biotic and abiotic stress from wild relative rice via conventional breeding methods and marker-assisted selection[6,10-11]. Thus, it is required to introgress important gene/QTL from wild rice into popular rice varieties for sustainable and resilient rice breeding program. An accession of *Oryza nivara* (AC100476) previously known as drought tolerant at the vegetative stage was crossed with recurrent parent Lalat, a drought susceptible cultivar to develop advanced mapping population. A total of 100 BC_2F_4 was developed and grown in the field for the present studies.

Identification of Polymorphic RM Markers

Rice microsatellites are simple sequence repeats (SSRs) which are arranged in tandem repeats of one to six nucleotide long DNA motifs distributed throughout the rice genomes[14]. Allelic variation of RM markers is caused by variation in the number of repeat-motifs at a locus due to replication slippage and/or unequal crossing-over during meiosis and they are considered higher as compared to other PCR-based markers[15]. Besides, RM markers have been extensively utilized in rice molecular plant breeding because of the co-dominant nature of inheritance, hypervariability, highly reproducibility, multi-allelic nature, genome-wide distribution and amenable to high throughput genotyping. With the availability of complete genome sequence data of rice and advancement of whole genome sequencing technology such as next generation genome sequencing, RNA seq, etc. a huge number of microsatellite markers for rice are now available in public domains[16] (https://archive.gramene.org/markers/microsat/). Therefore, a total of 410 RM markers were selected uniformly distributed over the 12 chromosomes of rice and used for screening polymorphism survey

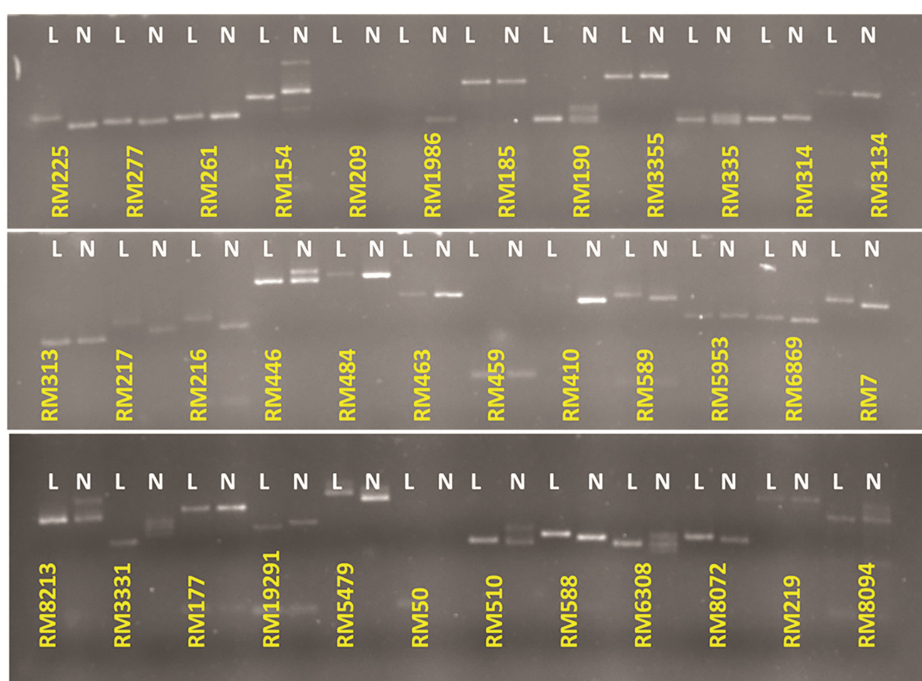


Fig. 1 — Identification of polymorphic RM markers between the Lalat (*Oryza sativa* subsp. *indica*) and *Oryza nivara* (AC100476). L and N denote DNA of Lalat (*Oryza sativa* subsp. *indica*) and *Oryza nivara*, respectively.

between two genotypes of rice, *O. nivara* (AC100476) having drought tolerance at the vegetative stage and *O. sativa* subsp. *indica*, Lalat. Out of 410 RM markers, 113 RM markers (28.9%) were found to be polymorphic (Fig. 1) which is comparatively higher than that has been reported between the parents of different mapping population in *indica* rice[17-18]. The detailed information of the polymorphic RM markers between the *O. sativa* subsp. *indica* (Lalat) and *O. nivara* was given in the Table 1. The results indicated that the genetic variability is more between the *indica* and *nivara* genome as compared to the genetic variability within the same *indica* species. These markers distributed uniformly throughout the twelve chromosomes except few gaps on chromosome 1, 7 and 10 (Fig. 2). On the basis of motif length, the

Table 1 — Detail information of polymorphic RM markers between the Lalat (*O. sativa* subsp. *indica*) and *O. nivara*

Sl. No.	Markers	Forward & reverse sequences (5' to 3')	Position (Mb)	Motif repeat	GR/IGR
Chromosome 01					
1	RM495	AATCCAAGGTGCAGAGATGG CAACGATGACGAACACAACC	0.215	(CTG)7	Os01g0104200
2	RM1	GCGAAAACACAATGCAAAAA GCGTTGGTTGGACCTGAC	4.635	(GA)26	IGR
3	RM579	TCCGAGTGGTTATGCAAATG AATTGTGTCCAATGGGCTGT	8.451	(GA)25	IGR
4	RM582	TCTGTTGCCGATTTGTTCG AAATGGCTTACCTGCTGTCTC	9.190	(TC)20	IGR
5	RM580	GATGAACTCGAATTTGCATCC	9.605	(CTT)19	IGR
6	RM11865	CACTCCCATGTTTGGCTCC	36.264	(TTA)59	IGR
7	RM431	TCCTGCGAACTGAAGAGTTG AGAGCAAAACCCTGGTTCAC	38.893	(AG)16	Os01g0894600
8	RM12119	CCTCCTCCTCTTCTTCTAGCTTCC TCCACCACCACATCACTTTCG	40.451	(CT)32	Os01g0923950
9	RM529	CCCTCCCTTCTGTAAGCTCC GAAGAACAATGGGGTTCTGG	40.670	(CT)12	Os01g0927300
Chromosome 02					
1	RM154		1.083	(GA)12	Os02g0120800
2	RM12412	CTCACAGCAACATGTGAGGTACG GATCGATGGCCTTAGGTTTGC	1.731	(TG)11	IGR
3	RM423	AGCACCCATGCCTTATGTTG CCTTTTTCAGTAGCCCTCCC	3.836	(TTC)9	Os02g0170300
4	RM71	CTAGAGGCGAAAACGAGATG GGGTGGGCGAGGTAATAATG	8.760	(ATT)10T(ATT)4	Os02g0255066
5	RM13308	CAAAGGCGGATTCTCATTAGACG CAATCGACAGACACAGTTGTTCG	18.896	(TG)17	IGR
6	RM341	CAAGAAACCTCAATCCGAGC CTCCTCCCGATCCCAATC	19.341	(CTT)20	IGR
7	RM475	CCTCACGATTTTCCTCCAAC ACGGTGGGATTAGACTGTGC	23.604	(TATC)8	IGR
8	RM13679	AGATGACAAGGTGAGAGCACTGG TGGAGCCCAGAATTTCTAGATCG	26.102	(TC)34	IGR
9	RM13709	GTGGCTTGATTTCTGCAACTTCC GCTGGTACCTACCAAGTATCATTTCG	26.485	(AAT)21	IGR
10	RM6	GTCCCCTCCACCCAATTC TCGTCTACTGTTGGCTGCAC	29.579	(AG)16	IGR
11	RM166	GGTCCTGGGTCAATAATTGGGTTACC TTGCTGCATGATCCTAAACCGG	34.352	(T)12	Os02g0805200
12	RM208	TCTGCAAGCCTTGTCTGATG TAAGTCGATCATTGTGTGGACC	35.135	(CT)17	IGR
13	RM266	TAGTTTAACCAAGACTCTC GGTTGAACCCAAATCTGCA	35.431	(GA)19	Os02g0824800
14	RM535	ACTACATACACGGCCCTTGC CTACGTGGACACCGTCACAC	35.778	(AG)11	Os02g0832150

(*Contd.*)

Table 1 — Detail information of polymorphic RM markers between the Lalat (*O. sativa* subsp. *indica*) and *O. nivara* — (*Contd.)*

Sl. No.	Markers	Forward & reverse sequences (5' to 3')	Position (Mb)	Motif repeat	GR/IGR
Chromosome 03					
1	RM14239	CAAGTTCACCCGCCTTCTCG TTTCCATCATTAGCAGGCAGTAGC	0.106	(AATT)6	Os03g0101100
2	RM81B	GAGTGCTTGTGCAAGATCCA CTTCTTCACTCATGCAGTTC	1.945	(TCT)10	IGR
3	RM545	CAATGGCAGAGACCCAAAAG CTGGCATGTAACGACAGTGG	4.947	(GA)30	Os03g0195350
4	RM14602	GGCTTACTGGCTTCGATTTG CGTCTCCTTTGGTTAGTGCC	6.167	(CT)24	Os03g0217200
5	RM232	CCGGTATCCTTCGATATTGC CCGACTTTTCCTCCTGACG	9.755	(CT)24	Os03g0283900
6	RM7	TTCGCCATGAAGTCTCTCG CCTCCCATCATTTCGTTGTT	9.829	(GA)19	IGR
7	RM14981	GGCGAGCAGAAGTATAATCCAGAAGG CGCTTGTGGCTTACTGGCTTGG	13.886	(AG)48	IGR
8	RM15809	AAAGCTGCGACGAACACGAACG CGCCGCAGCAGAGAAGAGAAGG	29.045	(AG)17	IGR
9	RM16131	CAGCATTGCAGTCTTGCTTGC GGGTAGCAGTAGGTAGTCAGAGTTGG	34.271	(AG)16	IGR
10	RM514	AGATTGATCTCCCATTCCCC CACGAGCATATTACTAGTGG	35.281	(AC)12	Os03g0839400
Chromosome 04					
1	RM551	AGCCCAGACTAGCATGATTG GAAGGCGAGAAGGATCACAG	0.177	(AG)18	Os04g0102500
2	RM471	ACGCACAAGCAGATGATGAG GGGAGAAGACGAATGTTTGC	18.824	(GA)12	IGR
3	RM16838	AGAAATGGATCGGACTGAACATGC AGACACTCGGACGCACAAGC	18.996	(AG)12	IGR
4	RM19605	GAGCAAGATATGGTAGGTACTGC GGTAGCAATCCAATGTTAGTGG	19.528	(TTA)23	Os04g0398800
5	RM17063	ACGGAGACCGACCCAAGTAAGC CCACAGGTCAAGATGGAACAGC	22.120	(CT)24	Os04G0447100
6	RM17182	TGCAGCGTCTCATCATAAAGTCG GCTTAGTGCTGTGAACTGTGAAGACC	24.719	(CA)22	IGR
7	RM252	TTCGCTGACGTGATAGGTTG ATGACTTGATCCCGAGAACG	25.369	(CT)19	IGR
8	RM17256	CTCGAACCACAGCCACTTCACC CAATGTAGGCACCCAGTGATTCC	25.999	(AC)10	IGR
9	RM241	GAGCCAAATAAGATCGCTGA TGCAAGCAGCAGATTTAGTG	26.857	(CT)31	Os04g0540500
10	RM17322	TCTGCTAGCCTGCACACAGAAGG GCACAGGAGACCAAAGAATCAGG	27.808	(AG)20	Os04g0555600
11	RM17349	TAGCTGCTGGATGTTACCACTGC ACCATCTATCGCCACATCTCACC	28.256	(GA)15	IGR
12	RM17337	CCCTCCCGTAGACCTTGTACCC CCACAGCTAACCAATCCTTCTCC	28.275	(AAG)14	Os04g0564000
13	RM303	GCATGGCCAAATATTAAAGG GGTTGGAAATAGAAGTTCGGT	28.574	[AC(AT)2-10] 9(GT)7(ATGT)6	IGR
14	RM280	ACACGATCCACTTTGCGC TGTGTCTTGAGCAGCCAGG	34.989	(GA)16	IGR
Chromosome 05					
1	RM153	GCCTCGAGCATCATCATCAG ATCAACCTGCACTTGCCTGG	0.189	(GAA)9	Os05g0103300
2	RM413	GGCGATTCTTGGATGAAGAG TCCCCACCAATCTTGTCTTC	2.212	(AG)11	Os05g0138000

(*Contd.*)

Table 1 — Detail information of polymorphic RM markers between the Lalat (*O. sativa* subsp. *indica*) and *O. nivara* — (*Contd.)*

Sl. No.	Markers	Forward & reverse sequences (5' to 3')	Position (Mb)	Motif repeat	GR/IGR
3	RM249	GGCGTAAAGGTTTTGCATGT ATGATGCCATGAAGGTCAGC	10.776	(AG)5A2(AG)14	Os05g0268500
4	RM430	AAACAACGACGTCCCTGATC GTGCCTCCGTGGTTATGAAC	18.691	(GA)25	IGR
5	RM164	TCTTGCCCGTCACTGCAGATATCC GCAGCCCTAATGCTACAATTCTTC	19.259	(GA)16TT(GT)4	Os05g0395600
6	RM421	AGCTCAGGTGAAACATCCAC ATCCAGAATCCATTGACCCC	23.976	(AGA)6	Os05g0489000
7	RM480	GCTCAAGCATTCTGCAGTTG GCGCTTCTGCTTATTGGAAG	27.313	(AC)30	IGR
Chromosome 06					
1	RM19255	TTAAGCTAGGGAATCAGCGGTTAGC GGAGTTGCAGTGTGGTGTGTGG	0.534	(GAA)22	Os06g0109100
2	RM19291	CACTTGCACGTGTCCTCTGTACG GTGTTTCAGTTCACCTTGCATCG	1.215	(ATAC)7	IGR
3	RM589	ATCATGGTCGGTGGCTTAAC CAGGTTCCAACCAGACACTG	1.380	(GT)24	IGR
4	RM8072	GATCACTCAGGTCATCCATTC AATCAGAGAGGCTAAAGACAATAAT	1.408	(CATC)9	IGR
5	RM588	GTTGCTCTGCCTCACTCTTG AACGAGCCAACGAAGCAG	1.611	(TGC)9	Os06g0130100
6	RM510	AACCGGATTAGTTTCTCGCC TGAGGACGACGAGCAGATTC	2.831	(GA)15	Os06g0154900
7	RM204	GTGACTGACTTGGTCATAGGG GCTAGCCATGCTCTCGTACC	3.169	(CT)44	Os06g0162800
8	RM225	TGCCCATATGGTCTGGATG GAAAGTGGATCAGGAAGGC	3.416	(CT)18	Os06g0167500
9	RM584	AGAAAGTGGATCAGGAAGGC GATCCTGCAGGTAACCACAC	3.417	(CT)14	Os06g0167500
10	RM217	ATCGCAGCAATGCCTCGT GGGTGTGAACAAAGACAC	4.235	(CT)20	IGR
11	RM314	CTAGCAGGAACTCCTTTCAGG AACATTCCACACACACACGC	4.845	(GT)8(CG)3(GT)5	Os06g0195600
12	RM253	TCCTTCAAGAGTGCAAAACC GCATTGTCATGTCGAAGCC	5.425	(GA)25	Os06g0207000
13	RM30	GGTTAGGCATCGTCACGG TCACCTCACCACACGACACG	27.253	(AG)9A(GA)12	Os06g0661200
14	RM340	GGTAAATGGACAATCCTATGGC GACAAATATAAGGGCAGTGTGC	28.599	(CTT)8T3(CTT)14	Os06g0686050
Chromosome 07					
1	RM542	TGAATCAAGCCCCTCACTAC CTGCAACGAGTAAGGCAGAG	12.712	(CT)22	IGR
2	RM21421	CGCTCCTTTCTAGCTCCATCTCC AACTGCAACGAGTAAGGCAGAGG	12.713	(AG)22	IGR
3	RM248	TCCTTGTGAAATCTGGTCCC GTAGCCTAGCATGGTGCATG	29.339	(CT)25	Os07g0689800
4	RM22164	TGGATCTTCGATCTCTCACTCACC GCATATGCATGTTCCATGATCG	29.389	(TA)43	Os07g0691100
5	RM172	TGCAGCTGCGCCACAGCCATAG CAACCACGACACCGCCGTGTTG	29.561	(AGG)6	Os07g0694325
Chromosome 08					
1	RM337	GTAGGAAAGGAAGGGCAGAG CGATAGATAGCTAGATGTGGCC	0.152	(CTT)4-19-(CTT)8	Os08g0102700
2	RM152	GAAACCACCACACCTCACCG CCGTAGACCTTCTTGAAGTAG	0.682	(GGC)10	Os08g0112800

(*Contd.)*

Table 1 — Detail information of polymorphic RM markers between the Lalat (*O. sativa* subsp. *indica*) and *O. nivara* — (*Contd.*)

Sl. No.	Markers	Forward & reverse sequences (5' to 3')	Position (Mb)	Motif repeat	GR/IGR
3	RM310	CCAAAACATTTAAAATATCATG GCTTGTTGGTCATTACCATTC	5.115	(GT)19	Os08g0187500
4	RM72	CCGGCGATAAAACAATGAG GCATCGGTCCTAACTAAGGG	6.763	(TAT)5C(ATT)15	IGR
5	RM223	GAGTGAGCTTGGGCTGAAAC GAAGGCAAGTCTTGGCACTG	20.650	(CT)25	IGR
6	RM210	TCACATTCGGTGGCATTG CGAGGATGGTTGTTCACTTG	22.471	(CT)23	IGR
7	RM256	GACAGGGAGTGATTGAAGGC GTTGATTTCGCCAAGGGC	24.270	(CT)21	IGR
8	RM281	ACCAAGCATCCAGTGACCAG GTTCTTCATACAGTCCACATG	27.895	(GA)21	Os08g0557500
9	RM264	GTTGCGTCCTACTGCTACTTC GATCCGTGTCGATGATTAGC	27.926	(GA)27	Os08g0557900
Chromosome 09					
1	RM566	ACCCAACTACGATCAGCTCG CTCCAGGAACACGCTCTTTC	14.704	(AG)15	IGR
2	RM434	GCCTCATCCCTCTAACCCTC CAAGAAAGATCAGTGCGTGG	15.662	(TC)12	IGR
3	RM410	GCTCAACGTTTCGTTCCTG GAAGATGCGTAAAGTGAACGG	17.642	(TA)13	Os09g0465400
4	RM257	CAGTTCCGAGCAAGAGTACTC GGATCGGACGTGGCATATG	17.719	(CT)24	IGR
5	RM24559	AGTTGAGTGGCAAACCACAGAGC CTTCACTTGGGTTTGGGTGATGG	18.450	(ATA)32	Os09g0177300
6	RM24717	CCTCACTCCCGTACAGTTGAACC TAAGGCCATTCCGTTGATGTGG	21.196	(AAT)28	Os09g0538450
Chromosome 10					
1	RM216	GCATGGCCGATGGTAAAG TGTATAAAACCACACGGCCA	5.352	(CT)18	Os10g0177300
Chromosome 11					
1	RM286	GGCTTCATCTTTGGCGAC CCGGATTCACGAGATAAACTC	0.383	(GA)16	IGR
2	RM26213	GCCACAGGAGACAGCAAGAACC CGATCCAATTCCAGCCTAGATAGC	4.750	(TA)46	IGR
3	RM26269	GGAGGTAGGGAAATCAGGTGAGG GTGCACGTGACCATAAACACTCC	6.187	(TG)11	IGR
4	RM26393	GGAGGTGATCAACAACAGATAAGC ACCTCCAGTCTTGCTACTGTGC	8.833	(AG)21	IGR
5	RM26546	AGCGCATAGGCCTTTCCATTAGC CGTGTTCGTGGGTCTTGTTATGC	11.984	(AAT)19	Os11g0311550
6	RM287	TTCCCTGTTAAGAGAGAAATC GTGTATTTGGTGAAAGCAAC	16.767	(GA)21	IGR
7	RM209	ATATGAGTTGCTGTCGTGCG CAACTTGCATCCTCCCCTCC	17.808	(CT)18	IGR
8	RM21	ACAGTATTCCGTAGGCACGG GCTCCATGAGGGTGGTAGAG	19.639	(GA)18	IGR
9	RM206	CCCATGCGTTTAACTATTCT CGTTCCATCGATCCGTATGG	22.014	(CT)21	IGR
10	RM27235	CTGGCCAACGTCTCCAGTAGC AACCCTATCTAACCGCGAGTGACC	26.143	(TG)12	IGR
11	RM224	ATCGATCGATCTTCACGAGG TGCTATAAAAGGCATTCGGG	27.67	(AAG)8(AG)13	Os11g0684000
Chromosome 12					
1	RM27487	CCAAGCACCACATTTGGTTTCC AACCTTGCTCAGCAGGACAGC	1.612	(AG)22	Os12g0133201

(*Contd.*)

Table 1 — Detail information of polymorphic RM markers between the Lalat (*O. sativa* subsp. *indica*) and *O. nivara* — (*Contd.*)

Sl. No.	Markers	Forward & reverse sequences (5' to 3')	Position (Mb)	Motif repeat	GR/IGR
2	RM453	CGCATCTCTCTCCCTTATCG CTCTCCTCCTCGTTGTCGTC	2.690	(TC)10	IGR
3	RM27697	TGAATCCACACTCGCAGATCG AAATCAGCTCGGAGGGAACAGC	4.920	(AG)20	IGR
4	RM28034	CTCGGAAGCAGACAGCGAAGG GAGAGACTGGATCGGGTTCAAGC	13.668	(GGT)12	Os12g0428200
5	RM519	AGAGAGCCCCTAAATTTCCG AGGTACGCTCACCTGTGGAC	19.903	(AAG)8	Os12g0514600
6	RM28367	CGTATCTCCACCTCCCGAGAAGC GCCAAATCTCACGGATCGAAGC	21.173	(AG)21	IGR
7	RM6869	GAGCTCCTTGTAGTGACCCG ATCAGCCTCGCCAGCTTC	22.219	(TGG)8	Os12g0549800
8	RM3331	CCTCCTCCATGAGCTAATGC AGGAGGAGCGGATTTCTCTC	23.460	(CT)15	Os12g0570600
9	RM5479	AACTCCTGATGCCTCCTAAG TCCATAGAAACAATTTGTGC	24.378	(TC)21	IGR
10	RM270	GGCCGTTGGTTCTAAAATC TGCGCAGTATCATCGGCGAG	25.002	(GA)13	IGR
11	RM28637	TCACATGTCATACGGCTACAGACC CACAATCACAGTGTGTGCAAAGG	25.066	(CA)12	IGR
12	RM235	AGAAGCTAGGGCTAACGAAC TCACCTGGTCAGCCTCTTTC	26.107	(CT)24	Os12g0616200
13	RM17	TGCCCTGTTATTTTCTTCTCTC GGTGATCCTTTCCCATTTCA	26.988	(GA)12	IGR

GR: Genic region, IGR: Intergenic region

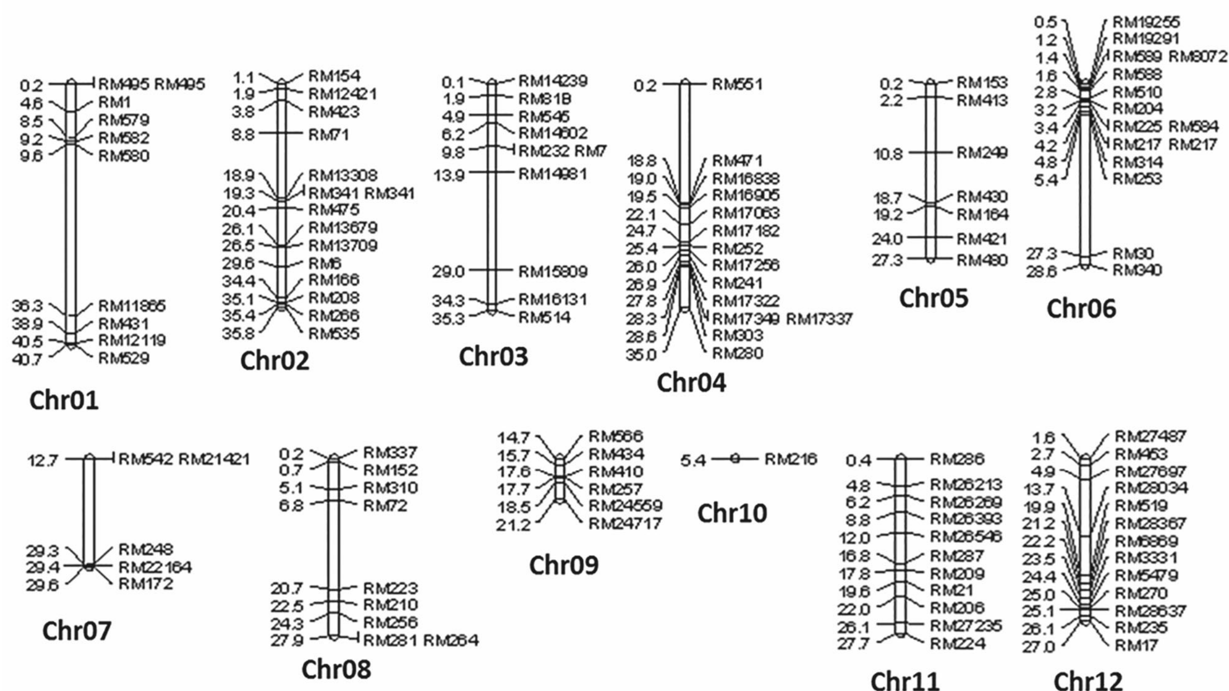


Fig. 2 — Distribution of polymorphic RM markers between parents, Lalat (*Oryza sativa* subsp. *Indica)* and *Oryza nivara* (AC100476) on 12 chromosomes of rice. The numerical figure on the left side of the chromosomes is the physical position of markers in Mb unit.

trinucleotide repeats-motif, (TTA)*n* was the longest with a maximum motif length of 177 nucleotides followed by dinucleotide repeat-motif, (AG)*n* with 96 nucleotides and tetra nucleotide repeat-motif, (CATC)*n* with 36 nucleotides length. Among the repeat motifs, di-nucleotide repeat-motifs were the most abundant

(68.14%) followed by tri-nucleotide repeat-motifs(18.58%), tetra-nucleotide repeat-motifs (3.53%) and mono-nucleotide repeat-motifs (0.88%). There was followed by the compound repeat-motifs accounting 8.84% among the polymorphic genic RM markers (Fig. 3). Higher frequency of dinucleotide repeat-motifs were also observed in several earlier studies in grasses[19-20]. Among the dinucleotide repeat-motifs, the motifs (AC/GT)*n* were the most abundant accounting 56.32% of the total dinucleotide repeat-motifs. This is followed by a repeat-motifs (GA/TC)*n* (27.58%), (TG/CA)*n* (6.89%), (AC/GT)*n* (5.74%) and (TA/AT)*n* (3.44%). In the case of trinucleotide repeat-motifs, the most abundant was (CTT/AAG)*n* (19.04%) followed by (TTC/GAA)*n*, (AAT/TTA)*n*, (TCT/AGA)*n*, (CTG/CAG)*n*, (TGC/GCA)*n*, (AGG/CTT)*n*, (GGC/GCC)*n*, (ATA/TAT)*n*, (GGT/ACC)*n* and (TGG/CCA)*n*. In case of tetra nucleotide repeat-motifs, there were four repeat-motifs, (TATC/GATA)*n*, (AATT/TTAA)*n*, (ATAC/GTAT)*n* and (CATC/GATG)*n* with one each number (Fig. 3).

Identification of Genic-Polymorphic SSR Markers

Simple sequence repeat (SSR) markers can be developed from either intergenic regions of the genome or exclusively genic regions which includes transcribing DNA segments and regulatory sequences. In comparison to the inter-genic SSR markers, genic-SSR markers are reported to be highly promising as there is a high chance of finding a marker trait association. Therefore, several genic-SSR markers have been developed in many important crops for genetic mapping of agronomic traits, regulation of gene expression by repeat-motifs, association mapping[21-27]. Among 113 polymorphic RM markers, 56 (49.55%) were found to be genic markers (Table 2). The highest number of genic markers was recorded in chromosome 6 (10), followed by chromosomes 2, 4 and 12 (6 each), chromosomes 3, 5, and 8 (5 each), chromosome 1 (4),

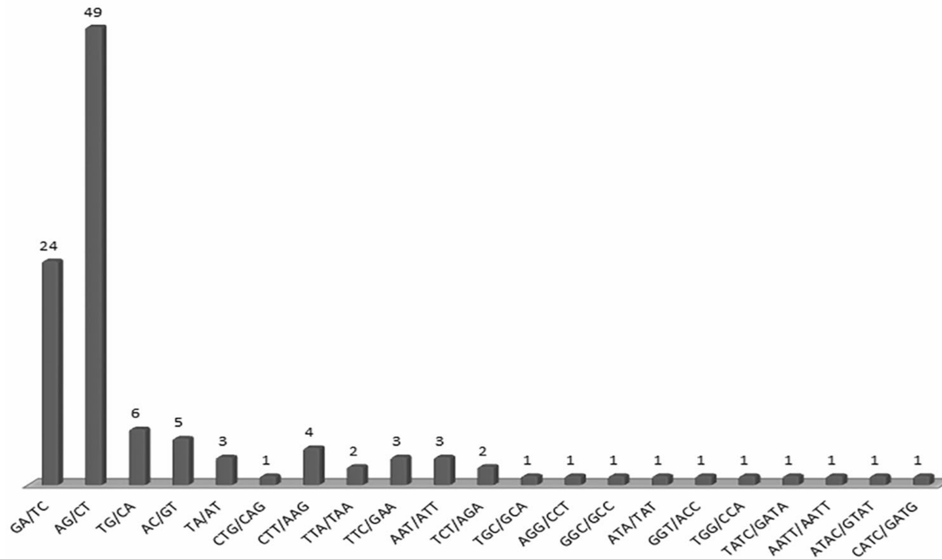


Fig. 3 — Frequency and distribution of different repeat-motifs (di-, tri- and tetra-nucleotides) of polymorphic RM markers between parents, Lalat (*Oryza sativa* subsp. *indica*) and *Oryza nivara* (AC100476).

Table 2 — Chromosome wise frequencies of repeat motifs of polymorphic RM markers between the Lalat (*Oryza sativa* subsp. *indica*) and *Oryza nivara*

Repeat motif	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Chr 6	Chr 7	Chr 8	Chr 9	Chr 10	Chr 11	Chr 12	Total
PM	9	14	10	14	7	14	5	9	6	1	11	13	113
Mono-	0	1	0	0	0	0	0	0	0	0	0	0	1
Di-	6	8	8	11	3	7	4	6	4	1	9	10	77
Tri-	3	3	1	2	2	2	1	1	2	0	1	3	21
Tetra-	0	1	1	0	0	2	0	0	0	0	0	0	4
Compound	0	1	0	1	2	3	0	2	0	0	1	0	10
GR	4	6	5	6	5	10	3	5	3	1	2	6	56

PM: Polymorphic markers; GR: Genic markers

chromosomes 7, and 9 (3 each), chromosome 11 (2) and chromosome 10 (1). In the present study, the polymorphism level of genic markers was higher as compared to other genic markers in different crops. Among the repeat motifs, di-nucleotide repeat-motifs were the most abundant (77.68%) followed by tri-nucleotide repeat-motifs (19%), tetra-nucleotide repeat-motifs (3) and mono-nucleotide repeat-motifs (1%). There was followed by the compound repeat-motifs accounting 9% among the polymorphic genic RM markers (Fig. 4).

In order to understand the various functions of the genes from which genic-RM markers were developed, we differentiated the 56 genes into three different categories such as the biological process, molecular

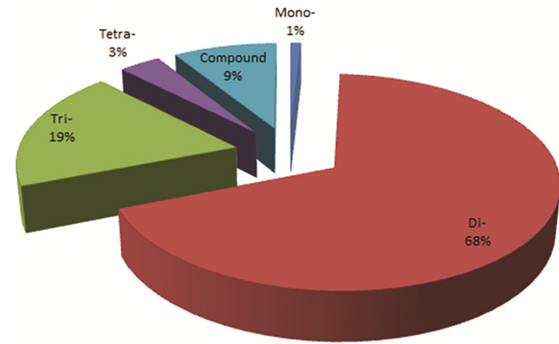
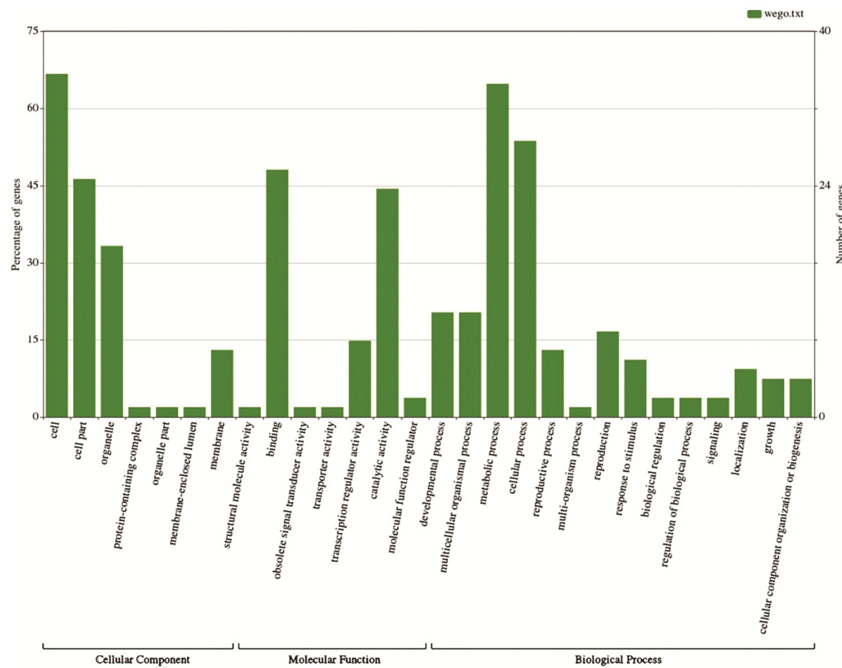


Fig. 4 — Distribution of different repeat-motifs of polymorphic genic RM markers between parents, Lalat (*Oryza sativa* subsp. *Indica*) and *Oryza nivara* (AC100476). Dinucleotide was the most abundant repeat motif (77. 68%), which was followed by trinucleotide motifs (21.19%).



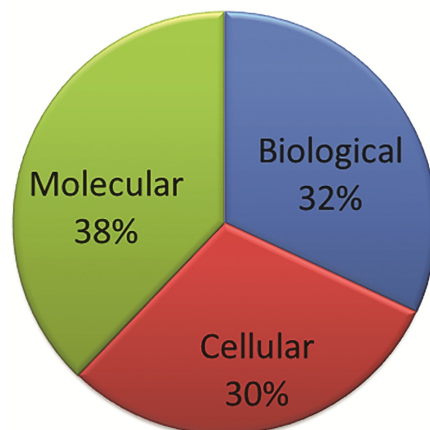


Fig. 5 — Depiction of the histogram of gene ontology (GO) classification of polymorphic genic RM markers. The genes were classified as three categories (biological process, molecular function and cellular component).

Table 3 — Classification of polymorphic genic-SSR markers based on gene ontology

Sl. No.	Cellular component	Molecular function	Biological process
1	RM529	RM495	RM495
2	RM154	RM431	RM529
3	RM423	RM529	RM154
4	RM166	RM154	RM423
5	RM14239	RM423	RM166
6	RM232	RM166	RM14239
7	RM551	RM535	RM14602
8	RM17063	RM14239	RM232
9	RM17337	RM14602	RM17063
10	RM153	RM232	RM17337
11	RM249	RM17063	RM249
12	RM164	RM17322	RM164
13	RM588	RM17337	RM421
14	RM204	RM249	RM588
15	RM225	RM164	RM510
16	RM253	RM421	RM204
17	RM248	RM588	RM225
18	RM152	RM510	RM314
19	RM310	RM204	RM253
20	RM281	RM225	RM248
21	RM264	RM314	RM310
22	RM410	RM253	RM281
23	RM24559	RM248	RM264
24	RM235	RM310	RM410
25		RM281	RM24559
26		RM264	RM224
27		RM410	RM28034
28		RM24559	RM235
29		RM224	
30		RM28034	
31		RM6869	
32		RM235	

function and cellular component (Fig. 5). These 56 genic-SSR markers localized on the 56 different genes were grouped into three categories based on gene ontology (Table 3). There are 32 SSR markers in the category of molecular function followed by 28 SSR markers in biological process and 24 SSR markers in cellular component.

All the 56 genes have been broadly distributed into three groups such as the molecular, biological and cellular categories with 38%, 32% and 30% frequencies respectively. The cellular component was further subdivided into the cell, cell part, organelle, protein-containing complex, organelle part, membrane-enclosed lumen, membrane and structural molecular activity. Among them, genes for cell, cell part and organelle were the highest abundant with 66.7%, 46.3% and 33.3%, respectively. Similarly, genes in molecular function have been categories as binding, obsolete signal transducer, transporter, transcription regulator, catalytic, molecular function regulator and development process. Among them, genes for binding were the highest abundant accounting 26 (48.1%) followed by catalytic activity (38.09%) and transcription regulator (14.4%). However, the genes in the biological process have been found to be multicelular organismal, metabolic, cellular, reproductive, multi-organism process, reproduction, response to a stimulus, biological regulation, regulation of the biological process, signaling, localization, growth and cellular component organization or biogenesis. In biological process, the metabolic process was the most abundant with 35 genes (64.8%) followed by a cellular (53.7%), developmental (20.4%), multicellular organismal process (20.4%) and reproduction (16.7%). Therefore, these genic-RM markers identified in the present study were found to be involved in the modulation of genes expression at the transcriptional and post-transcriptional level of the molecular, biological and cellular function which might be potential targets of selection in rice genetic improvement program.

Acknowledgements

The authors wish to thank Director, ICAR-National Rice Research Institute, Odisha, India for providing all the facilities required in the present studies.

References

1 Ngangkham U, Parida S K, De S, Kumar A R, Singh A K *et al*, Genic markers for wild abortive (WA) based male sterility and its fertility restoration in rice, *Mol Breed*, 26 (2010) 275-292

2 Ngangkham U, Nath M, Dokku P, Mithra S V, Ramamurthy S *et al*, An EMS induced new sequence variant, TEMS5032 in the coding region of SRS3 gene leads to shorter grain length in rice (*Oryza sativa* L.), *J Appl Genet*, 59 (4) (2018) 377-389

3 Brondani R P V, Zucchi M I, Brondani C, Rangel P H N, Borba TCDO *et al*, Genetic structure of wild rice *Oryza glumaepatula* populations in three Brazilian biomes using microsatellite markers, *Genetica*, 125 (2005)115-123

4 Jena K K & Khush G S, Introgression of genes from *O. officinalis* Wall. ex Watt to cultivated rice *O. sativa* L, *Theor Appl Genet*, 80 (1990) 737-745

5 Yadav M K, Aravindan S, Ngangkham U, Shubudhi H N, Bag M K *et al*, Use of molecular markers in identification and characterization of resistance to rice blast in India, *Plos One*, 12 (4) (2017) e0176236.

6 Swamy B P M, Kaladhar K, Rani N S, Prasad G S V, Viraktamath B C *et al*, QTL analysis for grain quality traits in 2 BC_2F_2 populations derived from crosses between *Oryza sativa* cv Swarna and 2 accessions of *O. nivara*, *J Hered*, 103 (3) (2012) 442-452

7 Joshi S P, Gupta V S, Aggarwal R K, Ranjekar P K & Brar D S, Genetic diversity and phylogenetic relationship as revealed

by inter simple sequence repeat (ISSR) polymorphism in the genus *Oryza*, *Theor Appl Genet*, 100 (2000) 311-1320.
8 Sarla N, Bobba S & Siddiq E A, ISSR and SSR markers based on AG and GA repeats delineate geographically diverse *Oryza nivara* accessions and reveal rare alleles, *Curr Sci,* 84 (2003) 683-690.
9 Juneja S, Das A, Joshi S V, Sharma S, Yogesh V Y *et al*, *Oryza nivara* (Sharma *et* Shastry) the progenitor of *O. sativa* (L.) subspecies *indica* harbours rich genetic diversity as measured by SSR markers, *Curr Sci*, 91 (2006) 1079-1085.
10 Swamy BPM, Kaladhar K, Ramesha M S, Viraktamath B C & Sarla N, Molecular mapping of QTLs for yield and related traits in *Oryza sativa* cv Swarna × *O. nivara* (IRGC81848) backcross population. *Rice Sci*, 18 (2011) 178-186
11 Addanki K R, Balakrishnan D, Yadavalli V R, Surapaneni M, Mesapogu S *et al*, Swarna × *Oryza nivara* introgression lines: A resource for seedling vigour traits in rice, *Plant Genet Resour*, 17 (1) (2018) 12-23
12 Voorrips R E, Map Chart: Software for the graphical presentation of linkage maps and QTLs, *The J Hered*, 93 (1) (2002) 77-78
13 Ye J, Zhang Y, Cui H, Liu J, Wu Y *et al*,. WEGO 2.0: A web tool for analyzing and plotting GO annotations, update, *Nucleic Acids Res*, 46 (W1) (2018) W71-W75
14 Parida S K, Kalia S K, Kaul S, Dalal V, Hemaprabha G *et al*, Informative genomic microsatellite markers for efficient genotyping applications in sugarcane, *Theor Appl Genet,* 1:118 (2) (2009) 327-38.
15 Goldstein D B & Schlotterer C, Microsatellites: Evolution and applications. Oxford University Press, Oxford (1999).
16 McCouch S R, Teytelman L, Xu Y, Lobos K B, Clareet K *et al*, Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.), DNA *Res*, 9 (2002) 199-207
17 Yoon D B, Kang K H, Kim H J, Ju H G, Kwon S J *et al*, Mapping quantitative trait loci for yield components and morphological traits in an advanced backcross population between *Oryza grandiglumis* and the *O. sativa japonica* cultivar Hwaseongbyeo, *Theor. Appl. Genet*, 112 (2006) 1052-1062.
18 Govindaraj P, Arumugasamy S & Maheswaran M, Bulked segregants analysis to detect main effect QTL associated with grain quality parameters in Basmati 370/ASD 16 cross in rice (*Oryza sativa* L.) using SSR markers. *Euphytica*, 144 (2005) 61-68
19 Powell W, Morgante M, Andre C, Hanafey M, Vogel J *et al*, The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis, *Mol Breed*, 2 (1996) 225-238
20 Chen X, Temnykh S, Xu Y, Cho Y G & McCouch S R, Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.), *Theor Appl Genet*, 95 (1997) 553-567
21 Singh A K, Chaurasia S, Kumar S, Singh R, Kumari J *et al*, Identification, analysis and development of salt responsive candidate gene based SSR markers in wheat, *BMC Plant Biol*, 18 (2018) 249
22 Asadi A A & Monfared S R, Characterization of EST-SSR markers in durum wheat EST library and functional analysis of SSR-containing EST fragments, *Mol Gen Genomics*, 289 (2014) 625-40.
23 Yu J K, Rota M L, Kantety R & Sorrells M, EST derived SSR markers for comparative mapping in wheat and rice. *Mol Gen Genomics*, 271 (2004) 742-51.
24 Varshney R K, Thiel T, Stein N, Langridge P & Graner A, *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol Biol Lett,* 7 (2A) (2002) 537-46.
25 Becher R, EST derived microsatellites as a rich source of molecular markers for oats. *Plant Breed*, 126 (2007) 274-8
26 Dutta S, Kumawat G, Singh B P, Gupta D K, Singh S *et al*, Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biol*, 11 (1) (2011) 17.
27 Mahato A K, Sharma N, Singh A, Srivastav M, Singh S K *et al*, Leaf transcriptome sequencing for identifying genic-SSR markers and SNP heterozygosity in crossbred mango variety 'Amrapali'(*Mangifera indica* L.), *PLoS One*, 11 (10) (2016) e0164325.