Indian J. Genet., 79(3) 536-544 (2019)
DOI: 10.31742/IJGPB.79.3.2

Identification, distribution and comparative analysis of microsatellites in the chloroplast genome of *Oryza* species

Kabita Tripathy[1], Balwant Singh, Gauri Misra[1] and Nagendra K. Singh[*]

National Institute for Plant Biotechnology, Pusa Campus, New Delhi 110 012; [1]Amity Institute of Biotechnology, Amity University, Noida 201 303, Uttar Pradesh

(Received: May 2019; Revised: July 2019; Accepted: August 2019)

Abstract

Chloroplast genome is important because of its maternally inherited, conserved within species and stable in structure. This allows elucidation of inter-specific comparison, identification and establishment of evolutionary relationship among different species. Different classes of DNA elements were present in chloroplast genome of *Oryza* species, among them chloroplast microsatellites were supposed to be highly variable. In this study, a total of 102 sequences of chloroplast genome of 23 species of *Oryza* genus belonging to ten different genome types were downloaded from NCBI nucleotide database. The chloroplast DNA length among different species varied from 134,401bp to 136,133bp. Microsatellite analysis revealed that total SSR number varied from 268 in *O. coarctata* to 281 in *O. granulata*. Comparative analysis of repeat type revealed that mononucleotide and trinucleotide were least present, while penta and hexa nucleotide motif were the most common type among all the *Oryza* species. Frequency analysis of the repeat revealed that T repeat among mono, AT among di, AAT among tri, AATA among tetra, AATTC among penta and AAAGAA among hexa nucleotide repeat was found to be common among most of the *Oryza* species. Total of five set of SSR markers flanking to penta and hexa nucleotide repeat types were developed which uniquely present among species, namely, *O. barthi, O. sativa* Japonica*, O. cocarctata, O. rhizomatis and O. ridley*. Phylogeny based on SSR markers indicate separate evolution of different SSR markers among individual species.

Key words: Chloroplast, SSR, mining, genome, *Oryza* species

Introduction

The *Oryza* genus comprises 24 different diploid and tetraploid species (2n = 24, 48) representing ten different genome constitutions, namely AA, BB, CC, EE, FF, GG, BBCC, CCDD, KKLL and HHJJ types. These are further grouped into *O. sativa* complex (AA), the *O. officinalis* complex (BB, CC, EE, BBCC, and CCDD), the *O. ridleyi* complex (HHJJ), the *O. granulata* complex (GG), and others (FF). The *Oryza sativa* complex (AA genome) has six wild rice and two cultivated species (*O. sativa* and *O. glaberrima*) with tremendous genetic variation within each species, e.g. *Oryza sativa* itself has six well known cultivar groups (Kim et al. 2016). In India, out of the six AA genome species, two species were present, *O. rufipogon* Griff and *O. nivara* (Sharma and Shastry 1965). Phylogenetic relationships among the *Oryza* species have been studied for a long time but cataloguing of individual into their respective species class are yet to be fully resolved. Being a staple food crop, the present scenario of rice cultivation needs development of high yielding varieties that have tolerant to different stresses. Wild rice species are naturally grown and adopted to extreme habitats. During the course of evolution they have acquired gene and governing traits that showed better tolerance to biotic and abiotic stresses (Mishra et al. 2016). Cataloguing them into their respective species group may boost rice breeding program after incorporating these wild relatives.

Genetic diversity and its utilization in the breeding program is the key factor for crop improvement. High genetic variation leads to spurious classification of the wild rice species. Therefore, DNA based marker may resolve this false notation. DNA markers are highly polymorphic, co-dominant inheritance, frequent distributions in the genome, high reproducibility, easy availability and neutral to the

*Corresponding author's e-mail: nksingh4@gmail.com; balwants27@gmail.com
Published by the Indian Society of Genetics & Plant Breeding, A-Block, F2, First Floor, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by www.isgpb.org; indianjournals.com

environmental conditions. These characteristic features also help to measure genetic relationship in crop plants (Joshi et al. 1999). Simple sequence repeats (SSRs) or microsatellites are short DNA stretches of single specific loci, are tandemly repeated and exhibit a high degree of polymorphism (Schlotterer 2000). Due to different number of repeat units it shows length variation which utilized as DNA marker (Morgante and Olivieri 1993). Further, SSRs are preferred as DNA markers due to their abundance, multi-allelic nature, highly variable, co-dominant inheritance, reproducibility and amenability to automation and high throughput genotyping (Powell et al. 1996). In rice genomic microsatellites have been identified and are being in use for various studies including population genetics and plant breeding (Mc couch et al. 2002, Singh et al. 2018). However, these markers were insufficient to differentiate individual *Oryza* species because of their complex evolutionary path (Miah et al. 2013). Chloroplast DNA (cpDNA) is important due to its maternal inheritance and physical stability in structure. In photosynthetic organisms, cpDNA size lies between 115 and 165 kbp (Jansen et al. 2005). It is considered to be an informative and valuable resource for phylogenetic analysis in plants at multiple taxonomic levels (Nadachowska-Brzyska et al. 2015).

Chloroplast genome has been utilized for phylogeny analysis and species differentiation using different marker systems such as SNPs and InDels (Song et al. 2018). In addition, some recent studies on chloroplast genome for development of markers for identification of closely related species have been published that utilized InDel and SNP markers for species differentiation among *Panax ginseng* subspecies (Kim et al. 2015), and InDel tandem repeat copy number variation markers for *Fagopyrum tataicum* and *Fagopyrum esculentum* (Cho et al. 2015). The chloroplast SSR (cpSSR) may significantly influence rice breeding program because of their wide application such as monitoring of seed-mediated gene flow, risk assessment of transferring transgenic into wild relative, parentage detection in hybrids and somatic hybrids, population genetic bottlenecks in natural populations (Bastia et al. 2001; Ryan et al. 2006; Atienza et al. 2007). The non-random evolution of genetic variations (Kumar et al. 2018) among intraspecific cpSSR are suited for phylogeographical population structure, therefore, they may be useful for rice domestication study as well (Powell et al. 1996b). Keeping this in view the study was focused on microsatellites (Bangar et al. 2018) mining for finding inter and intra specific variation among different *Oryza* species.

Materials and methods

Sequence retrieval and cpSSR mining

In order to analyze microsatellites present over different *Oryza* species, whole chloroplast genomes of 102 accessions of 23 different *Oryza* species belonging to AA, BB, BBCC, CC, CCDD, EE, FF, GG, HHJJ and KKLL genome types were downloaded from nucleotide database of NCBI (Table 1). Downloaded sequences were aligned using MAFFT version 7. For calculating microsatellite/SSR, chloroplast sequences of all 102 accessions of *Oryza* species were employed in MISA perl script (MIcroSAtellite, http://pgrc.ipkgatersleben.de/misa/) (Thiel et al. 2003). One to six nucleotide repeats were analyzed with different threshold values. For mononucleotide repeats the value was set as $\geq$12 bases, for dinucleotides (five), trinucleotides (four), tetranucleotides (three), pentanucleotide (two) and hexanucleotides (two).

Primer designing and In-silico PCR analysis

Based on the uniqueness of SSR markers a total of 5 species-specific primers were designed from 200bp of flanking sequences of the identified SSR motifs using Primer3 software (Untergasser et al. 2012, Koressaar et al. 2007). Primer designed for unique set cpSSRs to individual Oryza species and in-silico PCR was perform by SPCR (Cao et al 2005). Similarity of the primers sequences with the nuclear genome was check by NCBI BLAST and PCR product size was validated with primerBLAST tool. The primer sequences were not found any significant similarity with the nuclear genome.

Annotation of cpSSR

The identified cpSSRs (mono, di, tri, tetra, penta and hexa) were annotated in Excel 2013 by comparing cpSSR position and gene features of chloroplast genome. OrganellarGenomeDraw (*OGDRAW*) was used for representing annotation of microsatellites (Lohse et al. 2007).

Results and discussion

Chloroplast genome of 102 accessions belonging to 23 different *Oryza* species were downloaded from NCBI database. Among the available genome in the database, *O. sativa* Indica and *O. sativa* Japonica have

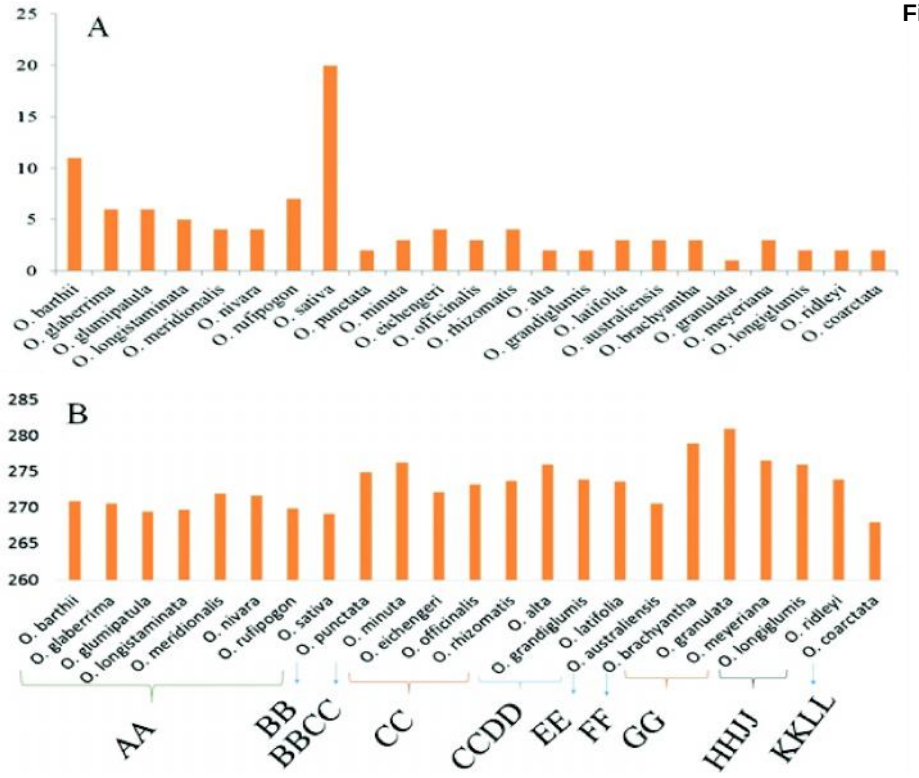

Fig. 1. Analysis of microsatellites (SSR) among 102 chloroplast genome sequence of 23 *Oryza* species belongs to 10 genome type A) Number of sequences downloaded from each species of Oryza B) Total number of Microsatellites present among individual *Oryza* species

highest ten number of available chloroplast genome each representing ten different cultivars, while *O. granulata* has only one genome (Fig. 1A). Highest number of chloroplast genome was available for the AA genome species. Genome size of chloroplast sequences varies from 1,34,401bp to 1,36,133bp. This study covers cp genome of all the 10

Table 1. Average number of microsatellites repeats present among ten genome type of 23 Oryza species

Genome Type	species	P1	P2	P3	P4	P5	P6	c	c*
AA	*O. barthii*	1	3	2	8	130	56	65	5
AA	*O. glaberrima*	1	3	2	8	130	56	66	5
AA	*O. glumipatula*	1	3	1	8	128	56	68	4
AA	*O. longistaminata*	1	3	1	8	129	55	68	4
AA	*O. meridionalis*	0	3	2	8	134	54	68	4
AA	*O. nivara*	1	3	1	8	130	56	68	5
AA	*O. rufipogon*	0	3	2	8	130	55	67	5
AA	*O. sativa*	0	3	2	7	131	54	66	5
BB	*O. punctata*	2	3	2	8	129	58	66	7
BBCC	*O. minuta*	1	4	1	7	128	59	69	7
CC	*O. eichengeri*	2	3	0	8	128	60	67	5
CC	*O. officinalis*	2	3	0	8	129	60	67	4
CC	*O. rhizomatis*	2	3	0	8	130	59	67	5
CCDD	*O. alta*	4	3	0	8	127	60	69	5
CCDD	*O. grandiglumis*	3	3	1	8	125	59	70	5
CCDD	*O. latifolia*	3	3	0	7	127	59	69	5
EE	*O. australiensis*	2	4	2	8	122	59	68	6
FF	*O. brachyantha*	0	2	3	8	134	70	54	8
GG	*O. granulata*	1	4	2	10	132	59	68	5
GG	*O. meyeriana*	0	4	2	9	130	59	67	5
HHJJ	*O. longiglumis*	0	3	2	6	124	64	69	8
HHJJ	*O. ridleyi*	0	3	2	6	122	63	70	3
KKLL	*O. coarctata*	0	3	2	7	133	54	63	6

P1 = Mononucleotide; P2 = Di nucleotide; P3 = Tri nucleotide; P4 = Tetra nucleotide; P5 = Penta nucleotide; 6 = Hexa nucleotide repeats and c and C* = Compound repeat microsatellites

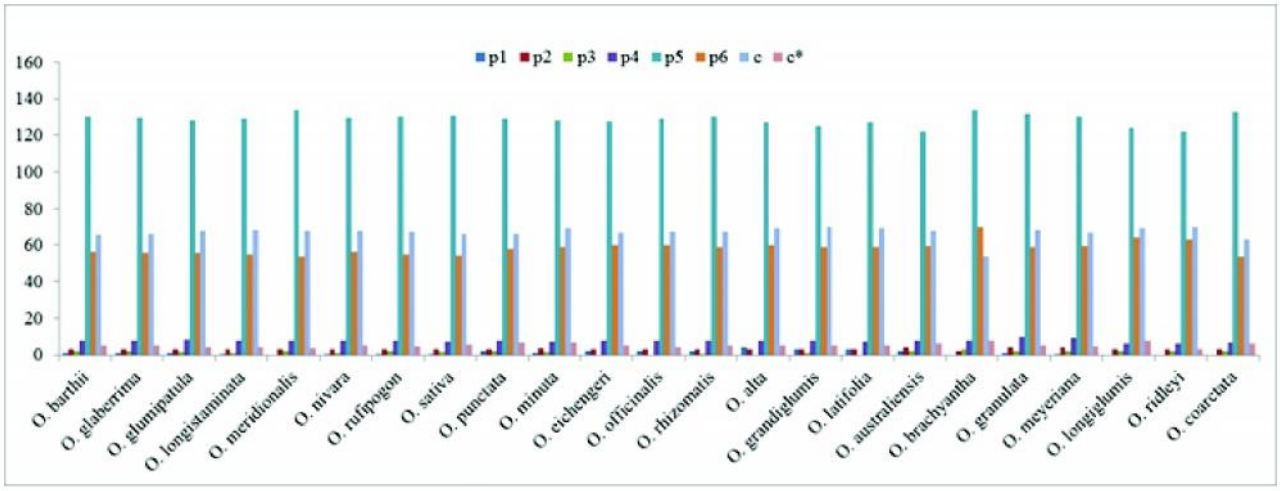

Fig. 2. Comparative analysis of average number of SSRs types present among Oryza species. Here: P1, mononucleotide; P2, di nucleotide; P3 tri nucleotide; P4 tetra nucleotide; P5 Penta nucleotide; 6, Hexa nucleotide repeat and c and C* compound repeatmicrosatellites

genome types first of its kind.

In order to find microsatellite variation MISA perl script was run with defined SSR units from 1 to 6 among 102 chloroplast genome sequences. Number of SSR varied from 268 in *O. coarctata* to 281 *O. granulata.* The two cultivated rice species have 269 SSR in *O. sativa* and 271 in *O. glaberriema* (Fig. 2B). However, more than 400 SSRs were identified among different species (Asaf et al. 2017). Comparative analysis of the different repeats revealed that trinucleotide repeat was absent in *O. alta*, *O. eichengeri*, *O. latifolia*, *O. officinalis* and Mononucleotide repeats were absent in *O. coarctata*, *O. brachyantha*, *O. longiglumis*, *O. meridionalis*, *O. ridley*, *O. rufipogon* and *O. sativa* Japonica (Fig. 3). Interestingly, mononucleotide was present in *O. nivara* and *O. sativa* Indica. Further analysis of repeats revealed that composite SSR and five nucleotide repeat were most variable among 10 of the 23 species while tetra nucleotide repeat was the most conserved. Penta nucleotide repeat was highest in number while mononucleotides were lowest (Fig. 3). Penta nucleotide repeat ranged from 128 to 135 in *O. rufipogon* and highest was in *O. rufipogon* JN005833.1. The mononucleotide repeats may be used for differentiation of *O. rufipogon* from *O. nivara* and *O. sativa* Indica from *O. sativa* Japonica (Table 1). Previously, it has been identified that mono and dinucleotide were the most common in cp genome however, the present study identified penta and hexa nucleotides which are the most common and that this variation must be due to selection of size of nucleotide in a repeat (Asaf et al. 2017). In this study we have selected length of minimum 12 bp size for a repeat

Table 2. Primers sequences flanking to penta and hexa nucleotide SSR motifs specific to five *Oryza* species

Species	SSR type	Forward primer	Reverse primer	start	end	product	Location
O. barthi	p5	TTCCGCTCCTTTTC TATCCA	TGATGTAGGGAAAA GCTGGTT	3354	3599	246	intergenic matk and rps16
O. sativa	p5	CGCTATGGATGGGG ATTATG	CTAACAGCGGTCG ATTGAAA	57386	57629	244	intergenic psaI and ycf4
O. cocarctata	p5	CCCTACTACAGG CCAAGCAG	GCTATGCATGGTT CCTTGGT	304	513	210	psbA
O. rhizomatis	p6	CGAATTAAGAGCC TTAGGTCGAT	GAAGTGGGTCGGT CTCAAAA	44671	44818	148	intergenic trnS-GGA and rps4
O. ridley	p6	GGCCCCCATACCTT ATATCTG	CTATTGCCGCAAA TCAATCC	103989	104238	250	ndhF

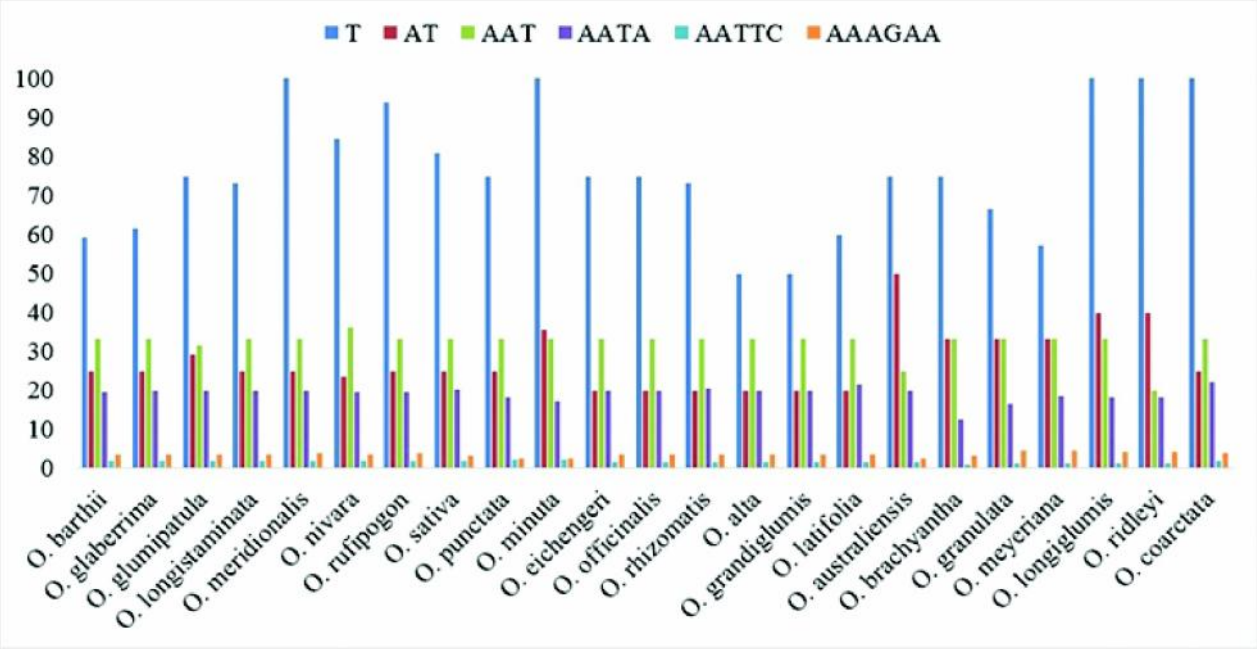

Fig. 3. Major nucleotide repeats having highest percent frequency among individual repeat class present in different *Oryza* species

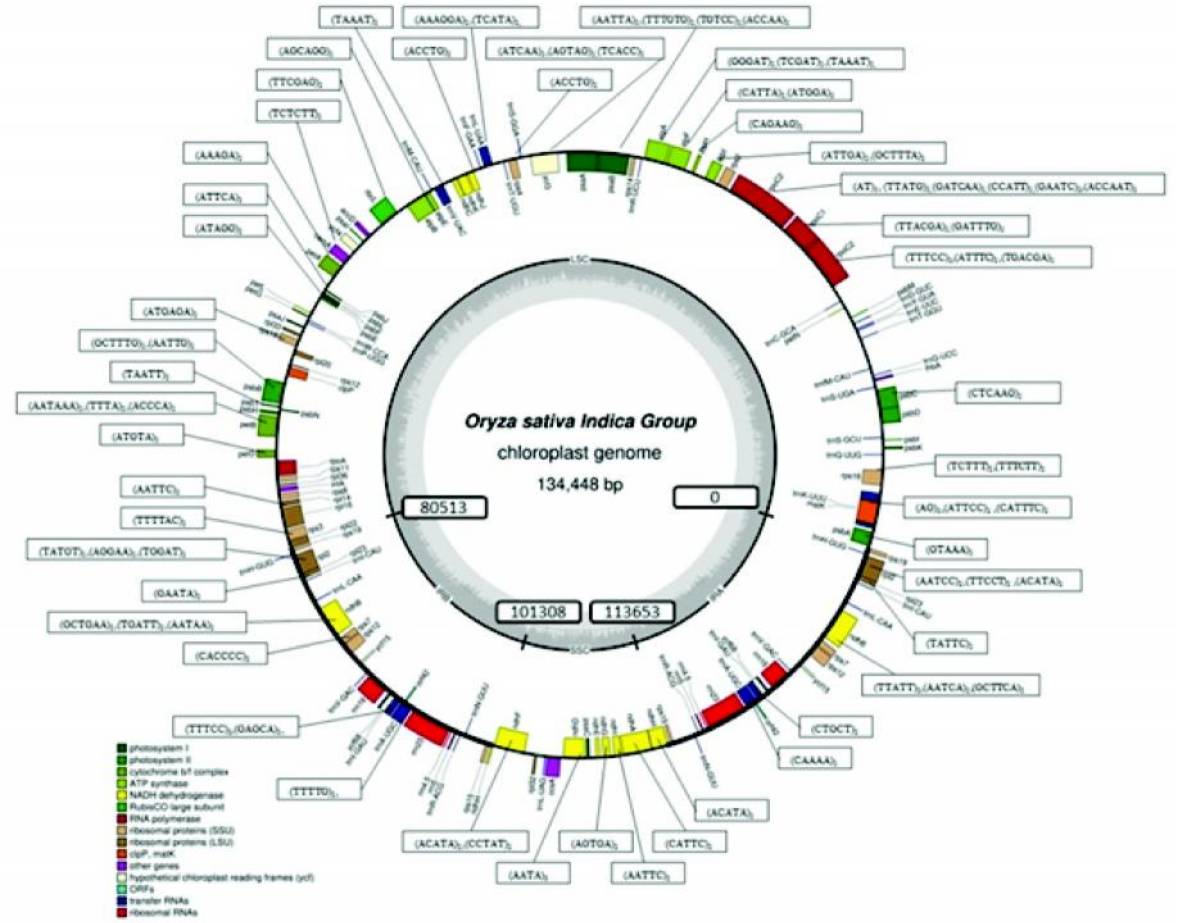

Fig. 4. Annotation of different SSR repeats present in coding region of the chloroplast genome

Table 3. SSR repeats present in coding region of different genes of the chloroplast genome

S. No.	SSR	SSR position	Gene name	Gene position	SSR type	SSR size
1	(GTAAA)2	925-934	psbA	82-1143	p5	10
2	(CATTTC)2	2296-2307	trnK-UUU	1363-3937	p6	12
3	(AG)5	3223-3232	trnK-UUU	1363-3937	p2	10
4	(ATTCC)2	3735-3744	matK	1668-3296	p5	10
5	(TTTCTT)2	4500-4511	rps16	4486-5560	p6	12
6	(TCTTT)2	4906-4915	rps16	4486-5560	p5	10
7	(CTCAAG)2	10773-10784	psbC	9845-11266	p6	12
8	(ATTTC)2	19308-19317	rpoC2	19167-22394	p5	10
9	(TGACGA)2	20069-20080	rpoC2	19167-22394	p6	12
10	(TTTCC)2	20745-20754	rpoC2	19167-22394	p5	10
11	(TTACGA)2	22903-22914	rpoC1	22432-24480	p6	12
12	(GATTTG)2	23044-23055	rpoC1	22432-24480	p6	12
13	(AT)5	25431-25440	rpoC2	24680-29221	p2	10
14	(TTATG)2	25592-25601	rpoC2	24680-29221	p5	10
15	(GATCAA)2	26174-26185	rpoC2	24680-29221	p6	12
16	(CCATT)2	26415-26424	rpoC2	24680-29221	p5	10
17	(ACCAAT)2	27761-27772	rpoC2	24680-29221	p6	12
18	(GAATC)2	27938-27947	rpoC2	24680-29221	p5	10
19	(GCTTTA)2	30868-30879	atpI	30454-31197	p6	12
20	(ATTGA)2	31193-31202	atpI	30454-31197	p5	10
21	(CAGAAG)2	32119-32130	atpH	31992-32237	p6	12
22	(CATTA)2	33521-33530	atpF	32693-34072	p5	10
23	(ATGGA)2	34055-34064	atpF	32693-34072	p5	10
24	(GGGAT)2	34269-34278	atpA	34163-35686	p5	10
25	(TCGAT)2	34623-34632	atpA	34163-35686	p5	10
26	(TAAAT)2	35071-35080	atpA	34163-35686	p5	10
27	(AATTA)2	36853-36862	psaB	36721-38925	p5	10
28	(TTTGTG)2	38127-38138	psaB	36721-38925	p6	12
29	(TGTCC)2	38348-38357	psaB	36721-38925	p5	10
30	(ACCAA)2	38854-38863	psaB	36721-38925	p5	10
31	(ATTCAA)2	41804-41815	ycf3	41802-43779	p6	12
32	(AGTAG)2	43399-43408	ycf3	41802-43779	p5	10
33	(TCACC)2	43529-43538	ycf3	41802-43779	p5	10
34	(ACCTG)2	45111-45120	rps4	44750-45355	p5	10
35	(AAAGGA)2	46652-46663	trnL-UAA	46482-47107	p6	12
36	(TCATA)2	46817-46826	trnL-UAA	46482-47107	p5	10
37	(CATATA)2	48256-48267	ndhJ	47918-48397	p6	12
38	(TAAAT)2	50576-50585	trnV-UAC	50292-50963	p5	10
39	(AGCAGG)2	52217-52228	atpB	51740-53236	p6	12
40	(TTCGAG)2	55432-55443	rbcL	54034-55467	p6	12
41	(TCTCTT)2	58678-58689	cemA	58611-59303	p6	12
42	(AAAGA)2	59917-59926	petA	59533-60495	p5	10
43	(ATTCA)2	61825-61834	psbL	61750-61866	p5	10
44	(ATAGG)2	61983-61992	psbF	61889-62008	p5	10
45	(ATGAGA)2	65892-65903	rps18	65573-66064	p6	12
46	(GCTTTG)2	69179-69190	psbB	68731-70257	p6	12
47	(AATTG)2	69522-69531	psbB	68731-70257	p5	10

comparative to different size for different repeat types used in earlier studies.

Frequency of different nucleotide repeats

Among the mononucleotide repeats, T repeat frequency was found most common among different *Oryza* species which ranged from 50% in *O. alta* and *O. grandiglumis* to 100 per cent (Supplementary Table S1). While analyzing repeats it was observed that T mononucleotides repeat was most frequent. It has been also reported that most cpSSR were (A/T) repeats that were used for identification of inter and intra specific variation among different rice species (Provan et al. 1997) Out of 5 dinucleotide repeat (AG, AT, GA, TA, and TC) identified over different chloroplast genome, TA repeat was preset over species having FF, GG and HHJJ genome types. Among dinucleotide repeats, AT was the major microsatellite repeat ranged from 20 to 50 per cent. High content AT repeats were also reported in chloroplast genome of Poaceae family members (Melotto-Passarin et al. 2017). Only four type of trinucleotides repeats (AAT, TAT, TCT and TTC) were identified among the *Oryza* species. Among them, AAT was the major trinucleotide repeat and it ranged from 20 to 33 per cent. Out of 12 tetra nucleotide repeat, ATAG

48	(TAATT)2	70527-70536	psbT	70413-70529	p5	10
49	(AATAAA)2	71377-71388	petB	71164-72622	p6	12
50	(TTTA)3	71633-71644	petB	71164-72622	p4	12
51	(ACCCA)2	71840-71849	petB	71164-72622	p5	10
52	(ATGTA)2	73708-73717	petD	73520-74044	p5	10
53	(AATTC)2	77214-77223	rpl14	77176-77547	p5	10
54	(TTTTAC)2	79434-79445	rps3	79266-79985	p6	12
55	(TATGT)2	81493-81502	rpl2	81097-82581	p5	10
56	(AGGAA)2	81680-81689	rpl2	81097-82581	p5	10
57	(TGGAT)2	82504-82513	rpl2	81097-82581	p5	10
58	(GAATA)2	82625-82634	rpl23	82527-82880	p5	10
59	(GCTGAA)2	86005-86016	ndhB	85311-87555	p6	12
60	(TGATT)2	86372-86381	ndhB	85311-87555	p5	10
61	(AATAA)2	87239-87248	ndhB	85311-87555	p5	10
62	(CACCCC)2	87923-87934	rps7	87855-88325	p6	12
63	(TTTCC)2	93575-93584	trnI-GAU	93011-94029	p5	10
64	(GAGCA)2	93733-93742	trnI-GAU	93011-94029	p5	10
65	(TTTTG)2	94794-94803	trnA-UGC	94094-94977	p5	10
66	(ACATA)2	101381-101390	ndhF	101350-103554	p5	10
67	(CCTAT)2	101999-102008	ndhF	101350-103554	p5	10
68	(AATA)3	106521-106532	ndhD	106319-107821	p4	12
69	(AGTGA)2	109374-109383	ndhG	109148-109678	p5	10
70	(AATTC)2	110055-110064	ndhI	109921-110463	p5	10
71	(ATAAA)2	110879-110888	ndhA	110558-112633	p5	10
72	(CATTC)2	112904-112913	ndhH	112635-113816	p5	10
73	(CAAAA)2	120159-120168	trnA-UGC	119985-120868	p5	10
74	(CTGCT)2	121219-121228	trnI-GAU	120933-121951	p5	10
75	(TTATT)2	127714-127723	ndhB	127407-129651	p5	10
76	(AATCA)2	128581-128590	ndhB	127407-129651	p5	10
77	(GCTTCA)2	128944-128955	ndhB	127407-129651	p6	12
78	(TATTC)2	132328-132337	rpl23	132082-132435	p5	10
79	(AATCC)2	132448-132457	rpl2	132381-133865	p5	10
80	(TTCCT)2	133273-133282	rpl2	132381-133865	p5	10
81	(ACATA)2	133460-133469	rpl2	132381-133865	p5	10

specific markers can be generated by exploring these penta and hexa nucleotide repeats. Five such repeats (three penta and two hexa) have been examined and primers were developed for PCR based selection of the species (Table 2). The amplicon size was validated through *in silico* PCR analysis.

Annotations of nucleotide repeats

Structurally chloroplast genomes have four regions namely small single-copy region (SSC), a large single-copy region (LSC) and a pair of inverted repeats (IRs) i.e., IRA and IRB. The SSR was annotated which revealed that mono and trinucleotides were present in noncoding regions while di, tetra, penta and hexa were distributed in both coding and noncoding regions. About 705 of the SSRs were found in noncoding region. Genome of *O. sativa* Indica (JN 861109.1) was drawn by using Organellar Genome Draw (*OGDRAW*) and SSRs

was identified in *O. australiensis* (EE) and only 5.2% of the tetra nucleotides in *O. glumepaetula* (AA). Major tetra nucleotide repeat was AATA motif and ranged from 12 to 22 per cent. Penta and hexa nucleotide repeats was the most common microsatellites among chloroplast genomes of *Oryza* speices. Among pentanucleotide repeats, AATTC motif and hexanucleotide AAAGAA motif were the major repeats. Pentanucleotide repeat AACCC and ACCAT were present in *O. brahcyantha*, ACACA and GGGAA in *O. coarctata*, AGAAT and ATGTT were present in *O. minuta* and *O. puntacta*. ATTGG was identified in *O. rhyzomatis*, whereas GAGAG was present in all the species of AA genome and *O. coarctata*. Species were placed in respective position (Table 3). Out of the six dinucleotides, AG and AT were present in coding region of trn-UUU and rpoC2. Out of 12 tetra nucleotides repeat two repeats TTTA, AATA found to be present in coding region of petB and ndhD gene (Fig. 4). Maximum six repeats were present in a copy of rpoC2 gene. Out of six, one was dinucleotide (AT) motif, three penta and two hexnucleotide motifs were present (Table 3). Only four genes namely, trnK-UUU, rpoC2, petB and ndhD respectively have dinucleotide (AG)5 motifs, tetranucleotide (TTTA)3 and (AATA) motifs. Other annotated genes have penta or hexanucleotide motifs.

Annotation of cpSSR revealed that 12 (*O. punctata*) to 26 (*O. alta*) per cent of the cpSSRs were present within the exonic region of the genes. This indicate intergenic cpSSRs were more abundant among oryza species. This has been also reported in the entire Poaceae family (Melotto-Passarin et al. 2011). Coding region of the genes namely rpoC2, rpl2, psaB, petB, matK, ndhB, atpF and atpA have multiple types of cpSSR motifs. However, some of the repeat motifs were specifically absent from individual species type. A repeat type was absent form rpoC2 gene of *O. sativa*, rpl2 gene of *O. rufipogon*, *O. barthii* and *O. glabrema*; psaB genes of *Oryza sativa* species complex and ndhB genes of *O. nivara*.

Microsatellite analysis revealed that maximum 281 SSR were found in *O. granulata*. Mononucleotide and trinucleotide motifs were least present andpenta and hexa nucleotide motif were the most common type among all the *Oryza* species for this selection criteria. For the first time all ten *Oryza* genome types were included in this study for comparative analysis of chloroplast SSRs. Repeats analysis showed presence of species specific SSRs among different *Oryza* species. Five set of unique cpSSR primers were designed and validated by *in silico* PCR analysis among five *Oryza* species. These species specific cpSSR markers may be used for identification of seed/ pollen mediated gene flow, risk assessment of transferring transgenic into wild relative and phyllogeographic origin of cultivated rice. Generated data set will be stored in Indian wild rice database (iwrdb) for comprehensive information (Tripathy et al. 2018).

Authors' contribution

Conceptualization of research (NKS, BS, GM); Designing of the experiments (KT, BS); Contribution of experimental materials (KT); Execution of field/lab experiments and data collection (KT, BS); Analysis of data and interpretation (KT, BS); Preparation of the manuscript (KT, BS).

Declaration

The authors declare no conflict of interest.

Acknowledgements

We are thankful to Indian Council of Agricultural Research (ICAR) for financial support in the form of 'ICAR-National Professor, B. P. Pal Chair' project.

References

Asaf S., Waqas M., Khan A. L., Khan M. A., Kang S. M., Imran Q. M., Shahzad R., Bilal S., Yun B. W. and Lee I. J. 2017. The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species. Front. Plant Sci., **8**: p.304.

Atienza S. G., Martin A. C., Ramirez M. C., Martin A. and Ballesteros J. 2007. Effects of *Hordeum chilense* cytoplasm on agronomic traits in common wheat. Plant Breed.,**126**(1): 5-8.

Bangar P., Chaudhury A., Umdale S., Kumari R., Tiwari B., Kumar S., Gaikwad A. B. and Bhat K. V. 2018. Detection and characterization of polymorphic simple sequence repeats markers for the analysis of genetic diversity in Indian mungbean [*Vigna radiata* (L.) Wilczek]. Indian J. Genet., **78**(1): 111-117. doi: 10.5958/0975.6906.2018.00013.5.

Bastia T., Scotti N. and Cardi T. 2001. Organelle DNA analysis of Solanum and Brassica somatic hybrids by PCR with 'universal primers'. Theor. Appl. Genet., **102**(8): 1265-1272.

Cho K. S., Yun B. K., Yoon Y. H., Hong S. Y., Mekapogu M., Kim K. H. and Yang T. J. 2015. Complete chloroplast genome sequence of tartary buckwheat (*Fagopyrum tataricum*) and comparative analysis with common buckwheat (*F. esculentum*). PLoS One, **10**(5): p.e0125332.

Cao Y., Wang L., Xu K., Kou C., Zhang Y., Wei G., He J., Wang Y. and Zhao L. 2005. Information theory-based algorithm for *in silico* prediction of PCR products with whole genomic sequences as templates. BMC bioinformatics, **6**(1): 190.

Joshi S. P., Ranjekar P. K. and Gupta V. S. 1999. Molecular markers in plant genome analysis. Current Science, **77**(2): 230-240.

Katoh K., Misawa K., Kuma K. I. and Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res., **30**(14): 3059-3066.

Kim H., Jung J., Singh N., Greenberg A., Doyle J. J., Tyagi W., Chung J. W., Kimball J., Hamilton R. S. and McCouch S. R. 2016. Population dynamics among six major groups of the *Oryzarufipogon* species complex, wild relative of cultivated Asian rice. Rice, **9**(1): p.56.

Kim K., Lee S. C., Lee J., Lee H. O., Joh H. J., Kim N. H., Park H. S. and Yang T. J. 2015. Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within *Panax ginseng* species. PloS One, **10**(6): p.e0117159.

Kumar P., Kumari R. and Sharma V. K. 2018. Choice of microsatellite markers for isolation of fertility restorers of wild abortive (WA) type cytoplasmic male sterility in rice. Indian J. Genet., **78**: 202-210. doi: 10.5958/ 0975-6906.2-18.00026.3.

Lohse M., Drechsel O. and Bock R. 2007. Organellar Genome DRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr. Genet., **52**(5-6): 267-274.

McCouch S. R., Teytelman L., Xu Y., Lobos K. B., Clare K., Walton M., Fu B., Maghirang R., Li Z., Xing Y. and Zhang Q. 2002. Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res., **9**(6): 199-207.

Melotto-Passarin D. M., Tambarussi E. V., Dressano K., De Martin V. F., Carrer H. 2011. Characterization of chloroplast DNA microsatellites from *Saccharum* spp. and related species. Genet. Mol. Res., **12**(3): 2024-33.

Miah G., Rafii M., Ismail M., Puteh A., Rahim H., Islam K. and Latif M. 2013. A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. Int. J. Mol. Sci., **14**(11): 22499-22528.

Mishra S., Singh B., Misra P., Rai V. and Singh N. K. 2016. Haplotype distribution and association of candidate genes with salt tolerance in Indian wild rice germplasm. Plant Cell Reports, **35**(11): 2295-2308.

Morgante M. and Olivieri A. M. 1993. PCR amplified microsatellites as markers in plant genetics. The Pl. J., **3**(1): 175-182.

Nadachowska-Brzyska K., Li C., Smeds L., Zhang G. and Ellegren H. 2015. Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. Curr. Biology, **25**(10): 1375-1380.

Park I., Kim W. J., Yang S., Yeo S. M., Li H. and Moon B. C. 2017. The complete chloroplast genome sequence of *Aconitum coreanum* and *Aconitum carmichaelii* and comparative analysis with other *Aconitum* species. PloS One, **12**(9): p.e0184257.

Park I., Yang S., Choi G., Kim W. and Moon B. 2017. The complete chloroplast genome sequences of *Aconitum pseudolaeve* and *Aconitum longecassi-datum* and development of molecular markers for distinguishing species in the Aconitum subgenus lycoctonum. Molecules, **22**(11): p. 2012.

Provan J., Corbett G., McNicol J. W. and Powell W. 1997. Chloroplast DNA variability in wild and cultivated rice (*Oryza* spp.) revealed by polymorphic chloroplast simple sequence repeats. Genome, **40**: 104-110.

Provan J., Powell W. and Hollingsworth P. M. 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecology Evol., **16**(3): 142-147.

Powell W., Machray G. C. and Provan J. 1996. Polymorphism revealed by simple sequence repeats. Trends Pl. Sci., **1**(7): 215-222.

Ryan E., Mullins E., Burke J., Downes M. and Meade C. 2006. Tracing field hybridization in Ryegrass species using microsatellite and morphological markers. Environmental Biosafety Res., **5**(2): 111-117.

Schlötterer C. 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma, **109**(6): 365-371.

Sharma S. D. and Shastry S. V. 1965. Taxonomic studies in genus *Oryza* L.: A modified classification. Indian J. Genet., **25**(2): 173.

Singh B., Singh N., Mishra S., Tripathi K., Singh B. P., Rai V., Singh A. K. and Singh N. K. 2018. Morphological and molecular data reveal three distinct populations of Indian wild rice *Oryza rufipogon* Griff. species complex. Front. Pl. Sci., **9**: p.123.

Song Y., Chen Y., Lv J., Xu J., Zhu S., Li M. and Chen N. 2017. Development of chloroplast genomic resources for *Oryza* species discrimination. Front. Pl. Sci., **8**: p.1854.

Thiel T., Michalek W., Varshney R. and Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet., **106**(3): 411-422.

Tripathy K., Singh B., Singh N., Rai V., Misra G. and Singh N. K. 2018. A database of wild rice germplasm of *Oryza rufipogon* species complex from different agro-climatic zones of India. Database 1. http://nksingh.nationalprof.in:8080/iwrdb/index.jsp.