CrossMark

Comparative analysis of sequences of mitochondrial genomes of wild abortive male sterile (WA-CMS) and male fertile lines of rice, development of functional markers for WA-CMS trait and their use in assessment of genetic purity of seeds of WA-CMS lines

K. Pranathi · B. C. Viraktamath · C. N. Neeraja · S. M. Balachandran ·
A. S. Hariprasad · P. Koteswara Rao · S. R. Kulkarni · P. Senguttuvel ·
S. K. Hajira · C. H. Balachiranjeevi · S. Bhaskar Naik · V. Abhilash ·
M. Anila · H. K. Mahadevaswamy · G. Rekha · M. S. Madhav ·
P. Revathi · G. Harika · T. Dilip · B. Kemparaju · R. M. Sundaram

Abstract WA-CMS system based rice hybrids are widely adopted in many rice growing countries, including India. Even though it is well known that the trait is controlled by mitochondria, the genes underpinning the trait remain enigmatic. In the present study, a complete genome-wide comparative sequence analysis was performed using draft mitochondrial genomes of WA-CMS and male fertile lines in a step-wise manner, progressively covering 5–10 kb every time through BLASTN tool. The sequence polymorphisms identified in different mitochondrial regions were targeted to develop two different sets of dominant PCR-based markers, one consisting of six markers targeting WA-CMS mitochondria, the other set consisting of five markers targeting male fertile mitochondria in addition to development of a set of eight co-dominant PCR-based markers targeting both the genomes. When a set of candidate genes/ORFs reported earlier to be associated with WA-CMS trait in rice were analyzed through RT-PCR of RNA isolated from immature rice florets, it was observed that the chimeric ORF, *WA352* is expressed only in WA-CMS line and hybrid (i.e. genotypes containing sterile mitochondria), indicating it's candidacy for the WA-CMS trait. Targeting the functional nucleotide polymorphism between WA-CMS and maintainer mitochondria with respect to *WA352*, two dominant markers, one targeting sterile and another targeting fertile mitochondria were developed. In addition, a robust, co-dominant functional marker targeting the candidate gene was also developed and validated for its utility in identification of genetic impurities in seed lots of WA-CMS lines.

K. Pranathi and R. M. Sundaram have contributed equally to this study.

K. Pranathi · B. C. Viraktamath · C. N. Neeraja ·
S. M. Balachandran · A. S. Hariprasad ·
P. Koteswara Rao · S. R. Kulkarni · P. Senguttuvel ·
S. K. Hajira · C. H. Balachiranjeevi · S. Bhaskar Naik ·
V. Abhilash · M. Anila · H. K. Mahadevaswamy ·
G. Rekha · M. S. Madhav · P. Revathi ·
G. Harika · T. Dilip · B. Kemparaju ·
R. M. Sundaram (✉)
Crop Improvement Section, Indian Institute of Rice Research, Rajendra nagar, Hyderabad 500030, India
e-mail: rms_28@rediffmail.com

Springer

Introduction

Rice is the staple food crop for more than half of world's population and is crucial to India's food security. The UN/FAO forecasts that global food production needs to be increased by over 40 % by 2030 and 70 % by 2050 and India needs to produce ∼125 million tons by 2030 to feed its ever increasing population of ∼150 billion. Hybrid rice is one of the feasible options for increasing rice yield and productivity, as hybrids yield ∼15–20 % more as compared to inbred varieties in rice (Virmani 1996). Production of hybrid seeds in self pollinating crop species like rice requires the use of male-sterile plants. Cytoplasmic male sterility (CMS) is most commonly employed in developing such hybrids.

Cytoplasmic male sterility (CMS) is a widespread phenomenon observed in >150 flowering plant species (Laser and Lersten 1972). CMS is a maternally inherited trait and is often associated with unusual open reading frames (ORFs), located in mitochondrial genomes, and in many instances, male fertility can be restored specifically by nuclear-genome encoded, fertility restorer (Rf) genes (Schnable and Wise 1998). Thus, CMS/Rf systems are ideal models for studying the genetic interaction and cooperative function of mitochondrial and nuclear genomes in plants. CMS/Rf systems have long been exploited for hybrid breeding to enhance the productivity. In rice (Oryza sativa), several CMS/Rf systems defined by the different CMS cytoplasm with distinct genetic features have been identified. These include CMS-BT (Boro II), CMS-WA (wild abortive) and CMS-HL (Honglian) (Shinjyo 1969; Lin and Yuan 1980; Rao 1988). These systems have been widely used for hybrid rice breeding in China and other Asian countries as hybrid rice crops that often produce higher yields than inbred varieties (Li and Yuan 2000).

Most of commercially released hybrids in rice are based on wild abortive (WA) cytoplasm, which forms the core of the three-line system of hybrid seed production. This system involves a WA-CMS (or A line) containing sterility inducing mitochondria, the cognate isonuclear line of the A line, called maintainer (or B line) possessing fertile mitochondria and a genetically divergent restorer line which possess nuclear Rf genes. As WA-CMS and maintainer lines are isonuclear and cannot be distinguished till flowering, the admixture of these lines during hybrid seed production is expected and in hybrid seed production, even slight impurity in WA-CMS lines impact the purity of hybrid seeds to a significant extent (Yashitola et al. 2004). Molecular markers can be helpful to distinguish the two lines during parental and hybrid seed production (Yashitola et al. 2002). However, the molecular markers that have been developed to distinguish WA-CMS lines from maintainers, target regions in the mitochondrial genome, which are not specific for WA-CMS trait and hence are efficient only to a certain extent (Yashitola et al. 2002, 2004; Rajendran et al. 2007; Rajendrakumar et al. 2007) in distinguishing impurities in seed-lots of WA-CMS lines. It will be ideal if functional markers specific for WA-CMS trait are deployed for this purpose, so that genetic impurities can be accurately identified and quantified in WA-CMS seed lots. Recent studies have reported on potential candidate genes for WA-CMS trait, viz., orfB (Das et al. 2010), orf126 (Bentolila and Stefanov 2012) and WA352 (Luo et al. 2013) and sequences of mitochondrial genomes of WA-CMS lines and maintainer lines are available in public domain for comparison (Bentolila and Stefanov 2012). However, none of these studies reported development and deployment of functional markers, which could unambiguously distinguish WA-CMS lines from their maintainers. The present study was therefore undertaken to analyze sequence polymorphisms in the mitochondria of WA-CMS and maintainer lines, develop a set of dominant and co-dominant markers specific for WA-CMS (i.e. sterile) mitochondria and maintainer (i.e. fertile) mitochondria, validate candidate genes reported for WA-CMS trait through expression analysis and develop functional marker(s) targeting the validated candidate gene for use in analysis of genetic impurities in seed lots of WA-CMS lines.

Materials and methods

Plant material

Fourteen WA-CMS and their maintainer lines of rice along with popular restorer lines used in hybrid rice breeding in India, five public-bred rice hybrids and three Indian rice varieties (listed in Table 1) were used in the present study for validation of the developed markers. To demonstrate the utility of functional co-dominant marker identified for WA-CMS trait, a seed lot of the WA-CMS line APMS6A that has been mixed

manually with known percentage of contaminants (i.e. seeds) of the maintainer line, APMS6B (as suggested in Rajendrakumar et al. 2007) was utilized. All the seed materials were collected from Hybrid Rice section of Indian Institute of Rice Research (IIRR), Hyderabad, India.

Analysis of mitochondrial genomes/genes for polymorphic regions

Two recently sequenced mitochondrial genomes, available in public domain were downloaded from GenBank/EMBL database. One of it is WA-CMS mitochondrial genome of the F_1 resulting from the cross IR6888A (WA-CMS) X IR62161R (restorer), which is male sterile (EMBL: JF281154) and the other is mitochondrial genome of maintainer line (IR6888B), which is a male fertile isonuclear maintainer line (EMBL: JF281153) (Bentolila and Stefanov 2012). In addition to these, the mitochondrial genomes of a *japonica* cultivar Nipponbare (BA000029), *indica* cultivar 93-11 (DQ167399) and *Oryza rufipogon* (AP011076), which are all male fertile were also downloaded from Genbank (http://www.ncbi.nlm.nih.gov/GenBank/index.html) and used for validation of sequence polymorphisms identified between WA-CMS and maintainer genomes. Comparative sequence analysis between the mitochondrial genomes of WA-CMS line (401 kb) and its maintainer line (637 kb) was done, by taking a short sequence of successive 5–10 kb from each genome at every alignment step and were aligned against each other using BLASTN tool (Altschul et al. 1990) with e-value cutoffs at 1e-5. This analysis resulted in identification of polymorphism between the two sequence and also fertile and sterile genome specific mitochondrial regions, which were then targeted for designing specific dominant and co-dominant markers for distinguishing sterile and fertile genomes (Supplementary Table 1a).

The sequence of the reported candidate gene for WA-CMS trait, *WA352* was also downloaded from NCBI using the accession code JX131325 (Luo et al. 2013). The downloaded sequence of 17,460 bp is a chimeric gene that has sequences of *rpl5*, pseudo *rps14* followed by *WA352* gene (which constitutes *orf284*, *orf224*, *orf288* and an unknown region between *orf284* and *orf224*) and atp6 (F_0 subunit).

The genomic co-ordinates (start sites and end sites) of *WA352* gene (only CDS-1059 bp) were located in both WA-CMS and maintainer genomes using Bioedit tool version 7.0.9 (Hall 2007). Three start sites of *WA352* gene were found in both the genomes and one end site was found only in WA-CMS genome and no end site was found in maintainer genomes. *WA352* gene sequence polymorphisms were analyzed through alignment of *WA352* region from WA-CMS and *WA352* region from maintainer genome using CLUSTALW tool. The identified polymorphisms in *WA352* gene (see alignments in Supplementary Figure 1) were targeted for the development of two functional dominant and co-dominant markers (Supplementary Table 1b).

DNA isolation and PCR amplification

Total rice genomic DNA was isolated by the protocol of Kochert et al. (1989) from the leaves of 18–20 days old greenhouse grown rice plants. Mitochondrial DNA was isolated from 18–20 days old etiolated seedlings as per the procedure of Mulligan et al. (1988). PCR was performed in 25 µl reaction volumes containing $1\times$ PCR buffer [10 mM Tris.HCI (pH 8.3), 50 mM KCI, 1.5 mM $MgCL_2$, 0.01 % (v/v) gelatin] 50–100 ng of template DNA, 5 pmol of each primer, 200 µM (each) deoxyribonucleotides, and 1 unit of Taq polymerase (Bangalore Genei India Ltd.). Standard PCR cycling conditions were followed as recommended by Rajendrakumar et al. (2007) for the eight co-dominant markers, five dominant markers specific for fertile mitochondrial, RMS-3-WA352 functional marker and similar PCR cycling conditions with slight alterations were adopted for six dominant markers specific for WA-CMS genome and two functional markers RMS-CMS-WA352, RMS-MNT-WA352. The altered PCR conditions for the six dominant markers included an initial denaturation step at 94 °C for 5 min, followed by 30 cycles of 94 °C for 1 min, 60 °C for 1 min and 72 °C for 3 min and a final extension at 72 °C for 10 min, whereas the PCR conditions for dominant functional markers, included an initial denaturation step at 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, 64 °C for 30 s and 72 °C for 1 min, and a final extension at

Table 1 List of genotypes used in the study

Line	Source
IR58025A	International Rice Research Institute (IRRI), Philippines
IR68897A	IRRI, Philippines
IR68888A	IRRI, Philippines
IR79156A	IRRI, Philippines
IR80555A	IRRI, Philippines
IR80561A	IRRI, Philippines
APMS 6A	Andhra Pradesh Rice Research Institute (APRRS), Maruteru, India
PUSA5A	Indian Agricultural Research Institute (IARI), New Delhi, India
DRR 4A	Indian Institute of Rice Research (DRR), Hyderabad, India
DRR 5A	Indian Institute of Rice Research (DRR), Hyderabad, India
DRR 6A	Indian Institute of Rice Research (DRR), Hyderabad, India
DRR 9A	Indian Institute of Rice Research (DRR), Hyderabad, India
DRR 10A	Indian Institute of Rice Research (DRR), Hyderabad, India
CRMS 32 A	Central Rice Research Institute (CRRI), Cuttack, India
IR58025B	IRRI, Philippines
IR68897B	IRRI, Philippines
IR68888B	IRRI, Philippines
IR79156B	IRRI, Philippines
IR80555B	IRRI, Philippines
IR80561B	IRRI, Philippines
APMS 6B	Andhra Pradesh Rice Research Institute (APRRS), Maruteru, India
PUSA5B	Indian Agricultural Research Institute (IARI), New Delhi, India
DRR 4B	Indian Institute of Rice Research (DRR), Hyderabad, India
DRR 5B	Indian Institute of Rice Research (DRR), Hyderabad, India
DRR 6B	Indian Institute of Rice Research (DRR), Hyderabad, India
DRR 9B	Indian Institute of Rice Research (DRR), Hyderabad, India
DRR 10B	Indian Institute of Rice Research (DRR), Hyderabad, India
CRMS 32 B	Central Rice Research Institute (CRRI), Cuttack, India
KMR3R	University of Agricultural Sciences (USA), V.C. Farm, Mandya, India
RPHR1005	Indian Institute of Rice Research (DRR), Hyderabad, India
DR-714-2R	Indian Institute of Rice Research (DRR), Hyderabad, India
IR40750R	IRRI, Philippines
BR 827-35	IRRI, Philippines
KRH2 (IR58025A X KMR3R)	University of Agricultural Sciences (USA), V.C. Farm, Mandya, India
DRRH2 (IR68897A X DR714-2R)	Indian Institute of Rice Research (DRR), Hyderabad, India
DRRH3 (APMS6A X RPHR1005)	Indian Institute of Rice Research (DRR), Hyderabad, India
TNAURH C04	Tamilnadu agriculture university (TNAU), Coimbatore, Tamilnadu
JRH8	JNKVV, Jabalpur
SAMBA MASURI	Acharya N.G. Ranga agriculture university (ANGRAU), Hyderabad, Telangana
IR64	IRRI, Philippines
SWARNA	Andhra Pradesh Rice Research Institute (APRRS), Maruteru, India

72 °C for 7 min. All amplified products were resolved in 2–3.5 % agarose gels (Lonza Inc., USA).

DNA Sequencing and analysis

The amplified PCR fragments of functional co-dominant marker RMS-3-WA352, from WA-CMS line, maintainer line, restorer line and hybrid were gel eluted and purified with Qiaquick Gel Extraction Kit (Qiagen, Hilden, Germany), cloned in TOPO-TA cloning kit (Invitrogen, Carlsbad, CA) and sequenced using an ABI Prism 3700 automated DNA sequencer (PerkinElmer, Wellesley, MA) as per the procedure suggested in Rajendrakumar et al. (2007). Similarly, the amplified PCR fragments of functional dominant marker RMS-CMS-WA352 from WA-CMS line, hybrid and amplified PCR fragments of RMS-MNT-WA352 from maintainer line and restorer line were also sequenced. Homology searches were performed by BLASTN algorithm (Altschul et al. 1990) through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/blast). These DNA sequences of WA-CMS, maintainer, restorer and hybrid of each marker were aligned using the software CLUSTALW (Higgins et al. 1994) to validate the indel polymorphisms those were identified through sequence analysis of WA352 genomic regions.

RNA isolation and reverse transcriptase PCR

Total RNA was isolated from panicles collected at pre-anthesis stage using Trizol Reagent (Invitrogen, Carlsbad, CA, USA) according to the protocol supplied by the manufacturer and RNA was incubated with RNase free DNase I (Fisher Bio Reagents) at 37 °C for 50 min to clear contaminating trace amounts of DNA, followed by enzyme inactivation by incubating at 75 °C for 5 min. The total RNA was reverse transcribed into first strand cDNA in a 20 μl volume with AMV reverse transcriptase kit (New England Bio labs). The samples were diluted with ultra pure water to 50 μl volume and 3 μl (∼1 ug RNA equivalent) was used in a 25 μl reaction with 1X PCR buffer [10 mM Tris HCI (pH 8.3), 50 mM KCI, 1.5 mM MgCL$_2$, 0.01 % (v/v) gelatin], 5 pmol of each primer, 200 μM (each) deoxyribonucleotides, and 1 unit of Taq polymerase (Bangalore Genei India Ltd.). Supplementary Table 2 contains list of RT primers specific for genes/ORFs reported to be specific for CMS trait in rice (viz., orfB, WA352 and orf126), which were used in expression analysis.

Results

Identification of mitochondrial regions polymorphic between WA-CMS and maintainer lines

A sequence similarity search was performed using BLASTN tool between sequences from WA-CMS and maintainer mitochondrial genomes to identify the polymorphic regions. These local alignments identified ten different polymorphic regions in the form of indels that ranged from 3 to 65 bp (Supplementary Table 3). In addition to indel polymorphisms, four genome-specific mitochondrial regions of sequence length varying from 5 to 11 kb were also identified (Supplementary Table 4). The sequence from 513,537 to 518,789 bp in the fertile mitochondrial genome was found to be unique, as no similar sequence was found in entire WA-CMS mitochondrial genome. Similarly, three different regions (12,917 to 19,714 bp; 97,932 to 104,799 bp and 391,154 to 396,194 bp) were found to be unique for WA-CMS mitochondrial genome, with no sequence similarity in other mitochondrial genomes.

Development of WA-CMS and maintainer mitochondrial genome specific markers and their validation

Polymorphic regions identified through local alignment were targeted for designing PCR primer pairs, which can differentiate WA-CMS lines and hybrids (possessing sterile mitochondria) from their maintainer lines and other lines possessing fertile mitochondria in dominant and co-dominant pattern. The results obtained are summarized in Tables 2 and 3 and discussed briefly below.

Sterile mitochondria specific markers

Based on the results of sequence analysis that has identified three different mitochondrial regions, which are unique to WA-CMS genome, a set of dominant markers were designed. Targeting each unique mitochondrial region, two dominant markers were designed, totaling to six WA-CMS specific dominant

markers. Each dominant marker amplifies product size of 2–3 kb with lines possessing only sterile mitochondria and null amplification with the lines possessing fertile mitochondria (Supplementary Figure 2a). Apart from these, a functional dominant marker has been designed from *WA-352* CDS region, which has been identified to be the candidate gene for WA-CMS trait. This functional dominant marker RMS-CMS-WA352 amplifies 2523 bp fragment from all sterile mitochondria possessing lines and null amplification with fertile mitochondria possessing lines (Fig. 1a). The sequenced PCR fragments of RMS-CMS-WA352 from WA-CMS line and hybrid were aligned to *WA352* genomic sequence to validate the identified polymorphic region (Supplementary Figure 3).

Fertile mitochondria specific markers

The mitochondrial sequences, which were analyzed to be unique to fertile mitochondrial genome, were targeted for development of five dominant markers.

Each of these markers amplifies 800 bp–1 kb fragment with the lines possessing fertile mitochondria and null amplification in sterile mitochondria containing lines (Supplementary Figure 2b). In addition, a functional dominant marker RMS-MNT-WA352 specific for fertile mitochondrial genome was designed by targeting the promoter region of *WA352* gene. This functional marker can efficiently identify lines possessing fertile mitochondria (Fig. 1b). The sequenced PCR fragments of RMS-MNT-WA352 from maintainer and restorer lines were aligned to *WA352* genomic sequence to validate the identified polymorphic region (Supplementary Figure 4).

Co-dominant markers that can differentiate both sterile and fertile lines

A set of highly efficient and robust co-dominant markers, which can distinguish lines possessing sterile and fertile mitochondria in a co-dominant fashion were developed (Supplementary Table 1a). Out of 10 indel

Table 2 List of polymorphic markers with their amplification pattern in WA-CMS, Maintainer, Restorer lines, Hybrids

Serial no.	Primer name	Expected PCR amplicon size (bp)	PCR product size (bp) in			
			WA-CMS line	Maintainer line	Restorer line	Hybrid
1	RMS-CMS-1	3625	3625	No amplification	No amplification	3625
2	RMS-CMS-2	3480	3480	No amplification	No amplification	3480
3	RMS-CMS-3	2386	2386	No amplification	No amplification	2386
4	RMS-CMS-4	2577	2577	No amplification	No amplification	2577
5	RMS-CMS-5	3680	3680	No amplification	No amplification	3680
6	RMS-CMS-6	3190	3190	No amplification	No amplification	3190
7	RMS-MNT-1	1372	No amplification	1372	1372	No amplification
8	RMS-MNT-2	1152	No amplification	1152	1152	No amplification
9	RMS-MNT-3	1760	No amplification	1760	1760	No amplification
10	RMS-MNT-4	1775	No amplification	1775	1775	No amplification
11	RMS-MNT-5	820	No amplification	820	820	No amplification
12	RMS-DRR-1	259/239	259	239	239	259
13	RMS-DRR-2	205/184	205	205	205	205
14	RMS-DRR-3	176/113	176	176	176	176
15	RMS-DRR-4	167/159	167	159	159	167
16	RMS-DRR-5	216/207	216	207	207	216
17	RMS-DRR-6	211/192	211	192	192	211
18	RMS-DRR-7	246/182	246	182	182	246
19	RMS-DRR-8	199/190	199	190	190	199

S. no., indicates serial number in the above table

Table 3 List of *WA352* gene based functional markers and their amplification pattern

Serial no.	Primername	Expected PCRamplicon size (bp)	PCR product size (bp) IN				
			WA-CMS line	Maintainer line	Restorer line	Hybrid	NCBI accession no.
1	RMS-CMS-WA352	2523	2523	No amplification	No amplification	2523	Pr032754351
2	RMS-MNT-WA352	244	No amplification	244	244	No amplification	Pr032754352
3	RMS-3-WA352	227/247	247	227	227	247	KU064722–KU064725

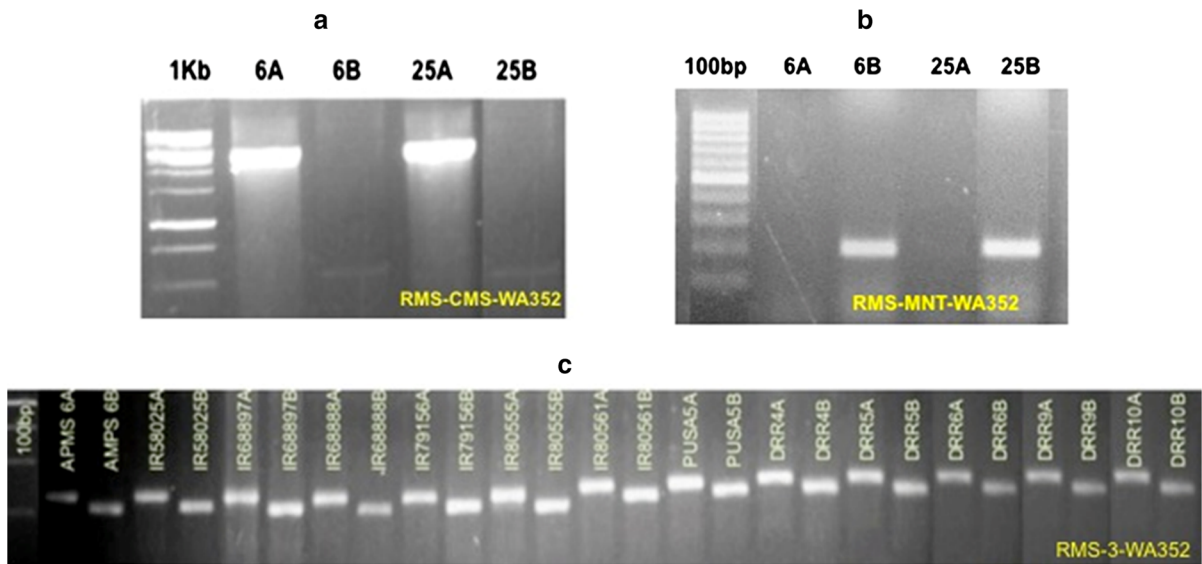S. no., indicates serial number in the above table

Fig. 1 Functional markers targeting *WA352* can distinguish WA-CMS line from their maintainer line in dominant (**a**, **b**) or co-dominant manner (**c**): targeting the candidate gene for WA-CMS trait (i.e. *WA352*), a dominant functional marker specific for sterile mitochondria, RMS-CMS-WA352 (**a**), a dominant functional marker specific for fertile mitochondria, RMS-MNT-WA352 (**b**) and (**c**) a co-dominant functional marker which can distinguish rice genotypes possessing sterile and fertile mitochondrial, RMS-3-WA352 were developed. The co-dominant marker could distinguish all known WA-CMS lines from their cognate isonuclear maintainer lines. 1–1 Kb ladder molecular weight marker, 100–100 bp ladder molecular weight marker, 6A-APMS6A, 25A-IR58025A, 6B-APMS6B, 25B-IR58025B

polymorphisms identified at various regions of mitochondrial genome, those with size difference of >9 bp were targeted to design eight co-dominant markers. Each of these co-dominant markers amplifies ~180–250 bp amplicon from sterile and fertile mitochondria possessing lines. The amplification pattern of co-dominant markers RMS-DRR-1, RMS-DRR-4, RMS-DRR-6, RMS-DRR-7 and RMS-DRR-8, which can be easily resolved in 1.5–3.5 % agarose gels is given in Supplementary Figure 2c. In addition, a highly efficient robust functional co-dominant marker RMS-3-WA352 was designed targeting a 20 bp in-del polymorphism within *WA352* genomic region. This marker amplifies 247 bp product from sterile mitochondria possessing lines and 227 bp product from fertile mitochondria possessing lines, thus distinguishing all WA-CMS lines from their isonuclear cognate maintainer lines and also restorer lines (Fig. 1c). The sequenced PCR fragments of RMS-3-WA352 from WA-CMS line, maintainer, restorer line and hybrid were aligned to *WA352* genomic sequence to validate the identified polymorphic region (Supplementary Figure 5).

The sequenced amplicons from WA-CMS, maintainer, restorer lines and hybrid with functional co-dominant marker developed in the study were deposited in the public databases with accession numbers KU064722 to KU064725 and sequences of primers of functional dominant markers was deposited as Pr032754351 and Pr032754352 in NCBI.

Utility of functional co-dominant marker in detecting contaminants in parental/hybrid seed lots

The set of co-dominant markers and a functional co-dominant marker developed in the study have wider applications in identification of contaminants in WA-CMS parental and hybrid seed lots. An assay has been developed to demonstrate the utility of RMS-3-WA352 functional marker using predetermined mixture of 450 CMS seeds of APMS6A with its maintainer APMS6B (Rajendrakumar et al. 2007). This predetermined mixture of seeds was planted individually in research farm of IIRR in wet season of 2013. Each plant was coded and leaf material was collected from 20 day old seedling to isolate DNA. The DNA sample from each individual seedling was used to set up PCR reaction with RMS-3-WA352 functional marker. After resolving the amplified products in agarose gel the 18 contaminants were detected among 450 coded individual plants based on the differences in amplicon sizes (Fig. 2). These plants were allowed to grow till maturity to assess the pollen fertility and spikelet fertility in CMS seed lot. The seed set (i.e. pollen fertility) was observed in these 18 coded individual plants. A perfect correlation was observed with respect to identification of contaminants through RMS-3-WA352 marker analysis and spikelet fertility data collected from individual coded plants. Similar results were observed with individual coded plants in dry season of 2014, where impurities were rapidly and accurately detected among WA-CMS seed lots using single seed/seedling based functional marker assay.

Expression analysis of putative candidate genes associated with CMS trait

The expression of the genes/ORFs *orfB*, *orf126* and *WA352* that were earlier reported to be associated with CMS trait in rice were analyzed. *WA352* and *orf126* showed differential expression among WA-CMS lines, their maintainers, restorers and hybrids with expression only in the WA-CMS lines and hybrids possessing sterile mitochondria (Supplementary Figure 6).

Discussion

Rice hybrids based on WA-CMS system were first released in 1973 in China (Shaoqing et al. 2007). As on date, except in China, all the commercially released hybrids in the hybrid rice growing countries including India, Philippines, Indonesia, Vietnam are based on the WA-CMS system. Several attempts have been made to identify the molecular basis of CMS trait in rice (Yashitola et al. 2002, 2004; Rajendrakumar et al. 2007; Rajendran et al. 2007; Xie et al. 2014). Few putative candidate genes associated with the trait have been reported. These include *rps3–rpl16–nad3–rpl12* gene cluster (Yashitola et al. 2004), *orfB* (Das et al. 2010), *orf126* (Bentolila and Stefanov 2012) and *WA352* (Luo et al. 2013).

In the present study, the whole mitochondrial genome sequence of WA-CMS (A line) and maintainer (B line) rice genotypes was compared and several genomic regions polymorphic between the two genotypes have been identified. Targeting these polymorphisms, a set of dominant and co-dominant markers has been developed in this study. Through gene expression studies, we confirmed the candidacy of chimeric ORF *WA352* with respect to WA-CMS trait. Further, targeting the sequence polymorphism in *WA352*, we also developed a co-dominant, functional marker based on the sequence polymorphism and using the marker, we developed single seed/seedling based assay for distinguishing WA-CMS lines from maintainer lines and for identification of impurities in seed lots from WA-CMS lines (Fig. 2).

When this study was initiated, WA-CMS mitochondrial genome sequence of a F_1 derived from cross between WA-CMS line and Restorer (i.e. IR6888AX IR62162R), and the maintainer line (i.e. IR6888B) were made available in the public domain for the first time (Bentolila and Stefanov 2012), but the actual candidate gene underlying the trait was not yet confirmed. Hence, we ventured to analyze the relative sequence polymorphism between the two mitochondrial genomes in the present study using simple online tools, viz., CLUSTALW (Higgins et al. 1994) and
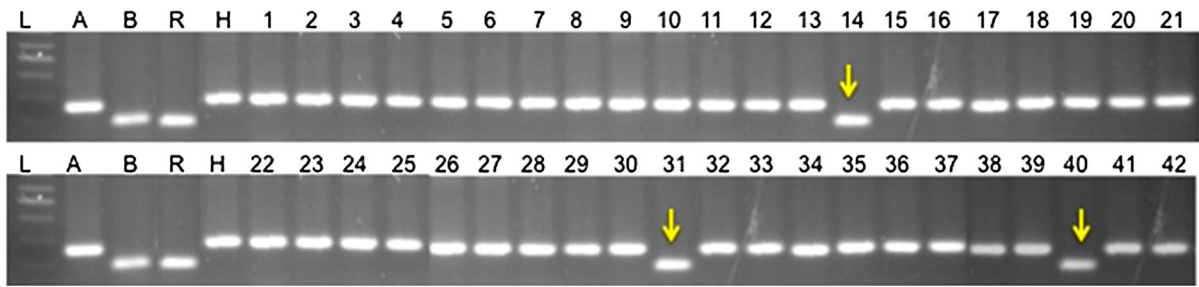
Fig. 2 Detection of impurities in seed-lots of a WA-CMS lines using the functional marker RMS-3-WA352: When a seed-lot of the popular WA-CMS line, APMS6A was analyzed with RMS-3-WA352, the functional, co-dominant marker targeting the candidate gene for WA-CMS trait (i.e. *WA352*), admixtures of the maintainer line seeds (i.e. APMS6B) could easily be identified (indicated by *arrows*). *L*—100 bp ladder molecular weight marker, A-APMS6A, B-APMS6B, R-RPHR1005 (restorer line), H-DRRH3 (hybrid), 1–42—samples collected from a seed-lot of APMS6A

BLASTN (Altschul et al. 1990) and identified five polymorphic regions (four of which are unique to WA-CMS genome and one specific to fertile mitochondrial genome) and a total of 10 in-del polymorphisms between the two types of mitochondrial genomes (Supplementary Tables 3, 4). This approach of comparative genome sequence analysis for identification of structural variations between normal and sterile lines, as carried out in the present study was very rapid, reliable and efficient in quick identification of genetic polymorphisms, in contrast to earlier methods of RFLP (Liu et al. 2007), RAPD (Cai et al. 1998), STS (Yashitola et al. 2002) and CAPS (Pranathi et al. 2014). The availability of whole rice mitochondrial genome in public domain has revolutionized the approach to identify polymorphic regions in mitochondria. Through this approach, Bentolila and Stefanov (2012) identified a putative ORF, *orf126* to be associated with the trait of WA-CMS in rice; however, the candidacy of the ORF was not demonstrated clearly. Due to the lack of *orf126* nucleotide or protein sequence in public domain, using the reported primers targeting *orf126* (Bentolila and Stefanov 2012), corresponding primer binding co-ordinates (340,284–340,609 bp) were identified in this study in WA-CMS mitochondrial genome using Bioedit tool version 7.0.9 (Hall 2007). Interestingly, it is observed that the reported RT-PCR primers specific for *orf126*, actually targets CDS of *WA352* gene only (Supplementary Figure 6).Luo et al. (2013), later identified another putative candidate gene, *WA352* associated with WA-CMS trait, cloned it and demonstrated its candidacy by complementation assay. Luo et al. (2013) also reported that the sequence of *orf126* is part of *WA352* and has two insertions that lead to premature stop codons. The nucleotide sequence of *orf126* might contain sequencing errors and may likely be identical to *WA352* gene as indicated by Okazaki et al. (2013). Even though both *orf126* and *WA352* have shown differential expression in this study (Supplementary Figure 6) and based on earlier reports (Luo et al. 2013; Okazaki et al. 2013) and based on points mentioned above, we conclude that *WA352* and not *orf126,* is the candidate for the trait of WA-CMS. Interestingly, another ORF, *orfB* (Das et al. 2010) does not show distinct expression pattern in WA-CMS lines and hybrids (Supplementary Figure 6).

There are a few reports on comparative mitochondrial sequence analysis for identifying chimeric ORFs and putative candidate genes associated with different CMS types in rice. Fujii et al. (2010) compared the mitochondrial genome sequences of LD-CMS and CW-CMS and identified unique chimeric genes specific to each CMS type and also demonstrated that the ORF, CW-*orf307* is specific to CW-CMS type. Igarashi et al. (2013) identified that *orf113* in RT98 cytoplasm derived from *O. rufipogon* is associated with the trait of CMS in RT cytoplasm through sequencing and comparing whole mitochondrial genomes of RT and non-RT mitochondria. Later, Okazaki et al. 2013 identified *orf352* candidate gene through whole genome sequencing and transcriptional analysis of RT102 type CMS derived from *O. rufipogon* and also reported that *WA352* and *orf352* are sequence variants with 99.5 % sequence similarity (five nucleotide differences) resulting in four amino acid substitutions. There are similar such reports in other crops on discovery of putative candidate genes

for CMS trait. For e.g. in *Raphanus sativus* L. Ogura type CMS trait was observed to be controlled by *orf138* (Tanaka et al. 2012). Similarly, in *Cajanus cajan*, sequencing of mitochondrial genomes of male sterile and fertile lines resulted in identification of CMS specific open reading frames (Tuteja et al. 2013). Later, the association of *nad7a* gene and *nad4L* gene with CMS trait was analyzed through comparative sequence analysis and expression analysis between sterile and fertile lines (Sinha et al. 2015). Shuangping et al. (2014), identified that *orf288* is associated with the trait of CMS in *hau* cytoplasm in *Brassica juncea* through comparative sequence analysis between *hau* cytoplasmic male sterile line and its isonuclear maintainer line. In the present study also, through comparative sequence analysis, we were able to identify several regions polymorphic between fertile and sterile mitochondria and were able to demonstrate the candidacy of *WA352* with respect to WA-CMS trait.

Maintenance of genetic purity of CMS seed lots is essential for realizing complete yield potential in hybrid rice breeding. The admixture of isonuclear maintainer line seeds with CMS is usually expected during various stages of seed production, including transplantation, harvesting and storage. It is difficult to differentiate these lines until reproductive stage, and generally seed-producers employ a morphological assay called Grow out test (GOT) to identify impurities in seed lots (Yan 2000). As GOT has many disadvantages (Yashitola et al. 2002), many groups including ours ventured to identify molecular markers, which can distinguish WA-CMS lines from their maintainer lines, so that impurities can be efficiently identified at seed/seedling stage. A few dominant (Yashitola et al. 2002; Rajendran et al. 2007) and a co-dominant marker (Rajendrakumar et al. 2007) were earlier identified through these studies and their utility in assessment of impurities in seed lots has been demonstrated. However, these markers did not essentially target the candidate genomic region specific for WA-CMS trait (i.e. *WA352*) and hence they have limited efficiency in accurate identification of impurities in seed-lots of WA-CMS lines.

In this study, we have developed eight co-dominant markers targeting in-del polymorphisms between WA-CMS and maintainer mitochondrial genomes (Supplementary Tables 1a, 3). Among the eight markers, RMS-DRR-1, RMS-DRR-4, RMS-DRR-6,

RMS-DRR-7and RMS-DRR-8 showed clear, unambiguous amplification pattern in WA-CMS and maintainer lines in a co-dominant fashion. However, in further analysis, it was observed that these markers could not distinguish all WA-CMS, maintainer lines and other lines (data not shown) and hence, as indicated earlier, it can be concluded that all regions polymorphic between WA-CMS and maintainer mitochondria may not be amenable for development of markers to distinguish A and B lines and it will be desirable to target polymorphisms in the candidate gene(s) implicated with WA-CMS trait. In order to implicate a candidate gene for WA-CMS trait, we carried out gene expression analysis of putative candidate genes earlier reported to be associated with the trait of WA-CMS (Das et al. 2010; Bentolila and Stefanov 2012; Luo et al. 2013) and confirmed that *WA352* is associated with the trait (Supplementary Figure 6). Even though, Das et al. (2010) indicated that *orfB* is the candidate ORF for WA-CMS trait in the CMS line of APMS6B in an earlier study, the results obtained in our study clearly demonstrates that *WA352* is the candidate and not *orfB*. Das et al. (2010) reported expression of 1.1 kb fragment using specific probe/RT primer pair (Mtg-1 and Corf) specific for *orfB* only in the WA-CMS line, APMS6A and no amplification in fertile lines. To our surprise, when the same probe was used in this study, we observed expression of a ~ 0.7 kb in all the samples, viz., WA-CMS lines, maintainer line, restorer line and hybrid (Supplementary Figure 6), instead of the reported 1.1 kb fragment, which should have been amplified only in the WA-CMS line. Interestingly, we also found presence of binding sites of the reported RT primer pair (Mtg-1 and Corf) in WA-CMS (7413–8195 bp) and male fertile (400,185–409,303 bp) mitochondrial genome. Thus, based on these observations, we conclude that *orfB* is not candidate for WA-CMS trait.

The chimeric ORF, *WA352* was observed to be of size 17,460 bp in WA-CMS genome (Luo et al. 2013), and genomic co-ordinates of *WA352* gene identified in maintainer genome was found to be longer compared to the gene size in WA-CMS genome (i.e. 18,176 bp), with two identical copies in the genome. Interestingly, the size of maintainer mitochondria is longer (637 kb) as compared to that of WA-CMS mitochondria (401 kb). The CDS of *WA352* gene was 1059 bp long in WA-CMS genome, with three transcription start sites distributed across the genome and with only one

transcription end site (Luo et al. 2013). Two of the transcripts were shorter and were from the inter-genic region, while the third was longer and was observed to co-transcribe with *rpl5*. By comparing *WA352* gene sequence from WA-CMS genome and maintainer genome, we identified a discontinuous indel polymorphism of length 20 bp (5, 5 and 9 bp) in *WA352* genomic region (Supplementary Figure 1) and validated the indel polymorphism through amplicon sequencing (Table 3 and Supplementary Figure 5). Targeting this indel polymorphism, we designed a functional co-dominant marker RMS-3-WA325, which can clearly distinguish all the WA-CMS lines from their maintainer lines (Fig. 1c). Interestingly, we identified the primer binding sites for RMS-3-WA352 marker in RT102A mitochondrial genome (which is another CMS mitochondria different from WA-CMS type; Okazaki et al. 2013; DDBJ accession: AP012528) and observed complete similarity in the primer binding sites targeted by the marker RMS-3-WA352 in both RT102A and WA-CMS mitochondrial genomes (data not shown). This indicates *orf352* discovered by Okazaki et al. (2013) and *WA352* identified by Luo et al. (2013) are sequence variants.

Using the functional co-dominant marker RMS-CMS-WA352, an assay based on single seed/seedling has been developed in the present study to identify contaminants in WA-CMS seed lots using the co-dominant marker and observed a perfect match between marker genotype and the plant phenotype as assessed using GOT (Fig. 2). Similar to our study, Suzuki et al. (2013) reported development of a functional marker based on in-del polymorphism identified in *atp6* gene to distinguish CMS-D8 type and its isonuclear line in Cotton. Recently, Sinha et al. (2015), reported development of a functional marker targeting a 10 bp indel polymorphism in *nad-7a* gene to distinguish A$_4$ cytoplasmic male sterile lines and their maintainer lines of pigeon pea. In order to analyze the efficiency of the functional co-dominant marker in identification of WA-CMS mitochondria containing rice genotypes, we amplified a germplasm pool of 83 wild rice lines and the marker amplified WA-CMS specific fragment in many Indian accessions of *O. rufipogon* (similar to the report of Luo et al. 2013), and few Indian accessions of *O. nivara* (Supplementary Table 5). Interestingly, all the wild rice accessions possessing sterile mitochondria were also observed to possess a restoring allele in the major

fertility restoration gene, *Rf4* when analyzed with the gene-specific marker DRCG Rf4-14 (Balaji et al. 2012), indicating that the trait of WA-CMS has possibly originated independently in China and India and possibly several other countries (Singh et al. 2015), and not exclusively in China as thought earlier.

In conclusion, through the present study, we have undertaken a comparative sequence analysis of mitochondrial genomes of WA-CMS and maintainer lines, identified regions polymorphic between the two genomes, developed a set of dominant and co-dominant markers capable of distinguishing the two lines, established the candidacy of the novel ORF, *WA352,* earlier discovered by Luo et al. (2013) with respect to the trait of WA-CMS through gene expression analysis and developed novel, functional markers targeting the *WA352*, which can be used for assessment of impurities in seed lots of WA-CMS lines.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Balaji S, Vemireddy LR, Srikanth B, Dharika N, Sambasiva Rao KRS, Hemanth Kishore V, Sundaram RM, Viraktamath BC, Subhakara Rao I, Ramesha MS, Neeraja CN (2012) Fine mapping of Rf3 and Rf4 fertility restorer loci of WA-CMS of rice (*Oryza sativa* L.) and validation of the developed marker system for identification of restorer lines. Euphytica 187:421–435

Bentolila S, Stefanov S (2012) A reevaluation of rice mitochondrial evolution based on the complete sequence of male-fertile and male-sterile mitochondrial genomes. Plant Physiol 158:996–1017

Cai CL, Yang Z, Zhu YG (1998) Analysis of rice mitochondrial DNA of sporophyte male sterility lines by RAPD. Yi Chuan Xue Bao 25:525–530

Das S, Sen S, Chakrabort A, Chakraborti P, Maiti MK, Basu A, Basu D, Sen KS (2010) An unedited 1.1 kb mitochondrial *orfB* gene transcript in the wild abortive cytoplasmic male sterility (WA-CMS) system of *Oryza sativa* L. subsp. Indica. BMC Plant Biol 10:39

Fujii S, Kazama T, Yamada M, Toriyamaws K (2010) Discovery of global genomic re-organization based on comparison of two newly sequenced rice mitochondrial genomes with cytoplasmic male sterility-related genes. BMC Genom 11:209

Hall T (2007) BioEdit version 7.0.9. Carlsbad, CA: Computer Program and documentation, Ibis Biosciences. http://www.mbio.ncsu.edu/BioEdit/bioedit.html

Higgins D, Thompson J, Gibson Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Igarashi K, Kazama T, Motomura K, Toriyama K (2013) Whole genomic sequencing of RT98 mitochondria derived from *Oryza rufipogon* and northern blot analysis to uncover a cytoplasmic male sterility-associated gene. Plant Cell Physiol 54:237–243

Kochert G, Tanksley SD, Price JP (1989) RFLP training course laboratory manual. Rockefeller Program on Rice Biotechnology. Cornell University, Ithaca, pp 5–6

Laser KD, Lersten NR (1972) Anatomy and cytology of microsporogenesis in cytoplasmic male sterile angiosperms. Bot Rev 38:425–454

Li JM, Yuan LP (2000) Hybrid rice, genetics, breeding, and seed production. In: Janick J (ed) Plant breeding reviews, vol 17. Wiley, New York, pp 15–158

Lin SC, Yuan LP (1980) Hybrid rice breeding in China. Innovative approaches to rice breeding. International Rice Research Institute, Manila, pp 35–51

Liu ZL, Xu H, Guo JX, Liu YG (2007) Structural and expressional variations of the mitochondrial genome conferring the wild abortive type of cytoplasmic male sterility in rice. J Integr Plant Biol 49:908–914

Luo DP, Xu H, Liu ZL, Guo JX, Li HY, Chen LT, Fang C, Zhang QY, Bai M, Yao N, Wu H, Wu H, Ji CH, Zheng HQ, Chen YL, Ye S, Li XY, Zhao XC, Li RQ, Liu YG (2013) A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. Nat Genet 45:573–577

Mulligan RM, Dement E, Andre C, Walbot V (1988) Isolation of RNA and DNA from rice seedling mitochondria. Rice Genet. Newsl. 5:151–154

Okazaki M, Kazama T, Murata H, Motomura K, Toriyama K (2013) Whole mitochondrial genome sequencing and transcriptional analysis to uncover an RT 102-type cytoplasmic male sterility-associated candidate gene derived from *Oryza rufipogon*. Plant Cell Physiol 54(9):1560–1568

Pranathi K, Viraktamath BC, Hajira S, Balachiranjeevi CH, Koteshwar Rao P, Bhaskar S, Abhilashkumar V, Rekha G, Anila M, Mahadevaswamy HK, Harika G, Swapnilkulkarni BVN, Kousik M, Balachandran SM, Madhav MS, Neeraja CN, Hari Prasad AS, Sundaram RM (2014) Analysis of WA-CMS and maintainer lines using mitochondrial genome specific molecular markers. Progress Res (Conf Spl) 9:1159–1163

Rajendrakumar P, Biswal AK, Balachandran SM, Ramesha MS, Viraktamath BC, Sundaram RM (2007) A mitochondrial repeat specific marker for distinguishing wild abortive type cytoplasmic male sterile rice lines from their cognate isogenic maintainer lines. Crop Sci 47:207–211

Rajendran N, Ramkumar G, Singh Sukhpal, Kadirvel P (2007) Development of a DNA marker for distinguishing CMS lines from fertile lines in rice (*Oryza sativa* L.). Euphytica 156:129–139

Rao YS (1988) Cyto histology of cytoplasmic male sterile lines in hybrid rice. In: Smith WH, Bostian LR, Cervantes EP (eds) Hybrid rice. International Rice Research Institute, Manila, pp 115–128

Schnable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. Trends Plant Sci 3:175–180

Shaoqing L, Daichang Y, Yingguo Z (2007) Characterization and use of male sterility in hybrid rice breeding. J Integr Plant Biol 49:791–804

Shinjyo C (1969) Cytoplasmic genetic male sterility in cultivated rice, *Oryza sativa* L. II. The inheritance of male sterility. Jpn J Genet 44:149–156

Shuangping H, Chao W, Bing J, Zhengjie W, Jing W, Bin Y, Chaozhi M, Jinxing T, Tingdong F, Jinxiong S (2014) Comparative analysis of mitochondrial genomes between the *hau* cytoplasmic male sterility (CMS) line and its iso-nuclear maintainer line in *Brassica juncea* to reveal the origin of the CMS-associated gene *orf288*. BMC Genom 15:322

Singh N, Jayaswal PK, Panda K, Mandal P, Kumar V, Singh B, Mishra S, Singh Y, Singh R, Rai V, Gupta A, Sharma TR, Singh NK (2015) Single-copy gene based 50 K SNP chip for genetic studies and molecular breeding in rice. Sci Rep 5:11600. doi:10.1038/srep11600 **(published online)**

Sinha P, Saxena KB, Rachit KS, Vikas KS, Suryanarayana V, Sameer Kumar CV, Katta Mohan AVS, Aamir WK, Rajeev KV (2015) Association of *nad7*a gene with cytoplasmic male sterility in pigeonpea. Plant Genome 8(2):1–12

Suzuki H, Yu J, Wang F, Zhang J (2013) Identification of mitochondrial DNA sequence variation and development of single nucleotide polymorphic markers for CMS-D8 in cotton. Theor Appl Genet 126:1521–1529

Tanaka Y, Tsuda M, Yasumoto K, Yamagishi H, Terachi T (2012) A complete mitochondrial genome sequence of Ogura-type male-sterile cytoplasm and its comparative analysis with that of normal cytoplasm in radish (*Raphanus sativus* L.). BMC Genom 13:352

Tuteja R, Saxena RK, Davila J, Shah T, Chen W, Xiao YL, Fan G, Saxena KB, Alverson AJ, Spilliane C, Town C, Varshney RK (2013) Cytoplasmic male sterility-associated chimeric open reading frames identified by mitochondrial genome sequencing of four Cajanus genotypes. DNA Res 20:485–495

Virmani SS (1996) Hybrid rice. Adv Agron 57:377–462

Xie H, Wang J, Qian M, Li N, Zhu Y, Li S (2014) Mitotype-specific sequences related to cytoplasmic male sterility in Oryza species. Mol Breeding 33:803–811

Yan W (2000) Crop heterosis and herbicide. US Patent No. 6,066,779

Yashitola J, Thirumurugan T, Sundaram RM, Naseerullah MK, Ramesha MS, Sarma NP, Sonti RV (2002) Assessment of purity of rice hybrids using microsatellite and STS markers. Crop Sci 42:1369–1373

Yashitola J, Sundaram RM, Biradar SK, Thirumurugan T, Vishnupriya MR, Rajeshwari R, Viraktamath BC, Sarma NP, Sonti RV (2004) A sequence specific PCR marker for distinguishing rice lines on the basis of wild abortive cytoplasm from their cognate maintainer lines. Crop Sci 44:920–924