

H. Akagi · A. Nakamura · Y. Yokozeki-Misono ·
A. Inagaki · H. Takahashi · K. Mori · T. Fujimura

Positional cloning of the rice *Rf-1* gene, a restorer of BT-type cytoplasmic male sterility that encodes a mitochondria-targeting PPR protein

Received: 22 September 2003 / Accepted: 5 January 2004 / Published online: 14 February 2004
© Springer-Verlag 2004

Abstract The combination of cytoplasmic male sterility (CMS) in one parent and a restorer gene (*Rf*) to restore fertility in another are indispensable for the development of hybrid varieties. We have found a rice *Rf-1* gene that restores BT-type CMS by applying a positional cloning strategy. Using linkage analysis in combination with 6,104 BC₁F₃ progeny derived from a cross between two near-isogenic lines (NILs) differing only at the *Rf-1* locus, we delimited the *Rf-1* gene to a 22.4-kb region in the rice genome. Duplicate open reading frames (*Rf-1A* and *Rf-1B*) with a pentatricopeptide (PPR) motif were found in this region. Since several insertions and/or deletions were found in the regions corresponding to both the *Rf-1A* and

Rf-1B genes in the maintainer's allele, they may have lost their function. *Rf-1A* protein had a mitochondria-targeting signal, whereas *Rf-1B* did not. The *Rf-1B* gene encoded a shorter polypeptide that was determined by a premature stop codon. Based on the function of the *Rf-1* gene, its product is expected to target mitochondria and may process the transcript from an *atp6/orf79* region in the mitochondrial genome. Since the *Rf-1A* gene encodes a 791-amino acid protein with a signal targeting mitochondria and has 16 repeats of the PPR motif, we concluded that *Rf-1A* is the *Rf-1* gene. Nine duplications of *Rf-1A* homologs were found around the *Rf-1* locus in the Nipponbare genome. However, while some of them encoded proteins with the PPR motif, they do not restore BT-type CMS based on the lack of co-segregation with the restoration phenotype. These duplicates may have played diversified roles in RNA processing and/or recombination in mitochondria during the co-evolution of these genes and the mitochondrial genome.

Communicated by M. Mackill

Sequence data from this article have been deposited with the EMBL/Genbank Data Libraries under accession nos. AB112808, AB112809, AB112810, AB112811.

H. Akagi (✉) · H. Takahashi · K. Mori
Laboratory of Plant Breeding and Genetics,
Department of Biological Production,
Faculty of Bioresource Sciences,
Akita Prefectural University,
Kaidoubata-Nishi 241-7, Shimoshinjo-Nakano,
010-0195 Akita, Japan
e-mail: akagi@akita-pu.ac.jp
Tel.: +81-18-8721645
Fax: +81-18-8721678

A. Nakamura · Y. Yokozeki-Misono · A. Inagaki
Biochemical Technology Section, Life Science Laboratory,
Performance Materials R&D Center,
Mitsui Chemicals, Togo 1144, 297-0017 Mobara, Japan

T. Fujimura
Institute of Agricultural and Forest Engineering,
University of Tsukuba,
Tennoudai 1-1-1, Tsukuba, 305-8572 Ibaraki, Japan

Present address:

A. Inagaki, Department of Synthetic Chemistry
and Biological Chemistry, Graduate School of Engineering,
Kyoto University, Yoshida-Honmachi, Sakyo-ku,
606-8501 Kyoto, Japan

Introduction

Rice is one of the most important food crops in the world and is the main nutritional staple for approximately 40% of the world's population. Annual rice production on a worldwide basis must increase from its current level of 520 million tons per year to 810 million tons per year by 2025 (Hossain 1996) to support the rapid increase in the human population that has been predicted. Since rice hybrids show heterosis, which subsequently results in them having yields 15–30% higher than inbred varieties (Yuan 1994; Fujimura et al. 1996), these hybrids may offer a solution to this problem.

The combination of cytoplasmic male sterility (CMS) and a nuclear gene for restoration of fertility (*Rf*) are essential for breeding hybrid varieties and for hybrid seed production. CMS, which eliminates the possibility of self-pollination, is commercially used in the production of hybrid seeds for economically important plants (Newton 1988). The *Rf* gene, on the other hand, restores the self-

fertilization ability of a hybrid plant and is indispensable in crops in which seeds are harvested, such as hybrid rice. Despite its importance in agriculture, the molecular mechanisms of CMS and its recovery are still unclear.

CMS is common in higher plants and derives from the incompatibility between nuclear and cytoplasmic gene products, which results in a failure to produce mature pollen grains (Newton 1988; Levings and Brown 1989). The reorganization of mitochondrial genomes in CMS plants has been reported in numerous species, and in many cases it results in the formation of chimeric genes in these regions that are responsible for the CMS characteristic (Schuabel and Wise 1998). Transcripts produced from such chimeric genes may inhibit expression of the normal gene and impair the formation of fertile pollens (Levings and Vasil 1995). In a number of plant species, such malformed transcripts in CMS mitochondria are altered in the presence of the *Rf* genes, leading to the production of fertile pollen and self-pollinated seeds (Pruitt and Hanson 1991; Iwabuchi et al. 1993; Akagi et al. 1994; Tang et al. 1996, 1998; Wise et al. 1999). These facts have suggested that some *Rf* genes are involved in the processing of mitochondrial (mt) mRNA associated with CMS.

Proteins that contain a pentatricopeptide repeat (PPR) are scattered throughout genome of higher plants (Small and Peeters 2000). In the *Arabidopsis* genome, more than 450 genes encode members of the PPR family (Aubourg et al. 2000). The PPR motif is presumed to play a role in binding to macromolecules such as RNA, and it has been predicted that many *Arabidopsis* PPR proteins target organelles (Aubourg et al. 2000). Members of the PPR family are thought to be involved in controlling organelle gene expression by processing or editing CMS-associated transcripts (Small and Peeters 2000), and it is through this mechanism that some PPR proteins are thought to participate in the suppression of CMS. In fact, one PPR protein, the product of the *Rf-PPR592* gene, counteracted CMS in *Petunia* (Bentolila et al. 2002), and another PPR protein in radish also suppressed CMS in *Brassica napus* (Brown et al. 2003; Koizuka et al. 2003). However, it is unclear which PPR proteins suppress CMS, since many genes in a plant genome encode PPRs (Aubourg et al. 2000). Moreover, as several CMS/*Rf* systems have been identified in rice (Virmani and Shinjyo 1988), it is necessary to specify the *Rf* gene and its corresponding CMS.

In rice, CMS occurs when the cytoplasm of *japonica* or *indica* rice is replaced with that of *indica* rice or wild rice (Katsuo and Mizushima 1958; Li and Zho 1986; Virmani and Shinjyo 1988). One CMS system, based on the cytoplasm of Chinsurah Boro II (*indica* rice) and the nucleus of Taichung 65 (*japonica* rice), is called ms-bo type or BT-type. BT-type CMS is restored by the nuclear restorer gene *Rf-1*, which was initially identified in Chinsurah Boro II and which has been shown to be due to a single locus in the nuclear genome (Shinjyo 1969). The *Rf-1* gene was mapped on chromosome 10 (Shinjyo 1975, 1984; Akagi et al. 1996). Since BT-type CMS is a type of

gametophytic CMS, only pollen with the *Rf-1* gene can develop normally into F₁ plants (Shinjyo 1969). In BT-type CMS mitochondria of rice, the 3' part of *atp6* was recombined to form a chimeric configuration (Iwabuchi et al. 1993; Akagi et al. 1994) that was joined to a novel open reading frame (ORF), *orf79* (Akagi et al. 1994). This di-cistronic mRNA was processed or edited by the action of the *Rf-1* gene, intact *atp6* mRNA was formed, and fertility was recovered (Iwabuchi et al. 1993; Akagi et al. 1994).

Kazama and Toriyama (2003) recently reported the candidate for the *Rf-1* gene by the transgenic strategy. They mapped the *Rf-1* locus within 1 cM. Using a corresponding region of Nipponbare genome sequence, they searched for ORFs encoding mitochondria targeting PPR proteins. Three candidate genes were cloned from Miliyang 23, and these were introduced into callus of an *rf-1/rf-1* CMS line. In a transgenic rice callus, one of the candidate genes (*PPR8-1*) encoded a 791-amino acid protein with 18 PPR repeats and was involved in processing of the transcript in mitochondrial *atp6*. However, these researchers did not confirm that *PPR8-1* restored fertility in regenerated plants.

A set of near-isogenic lines (NILs) that differ only at the locus of the *Rf-1* gene has been developed by repeated backcrossing to replace most of the nuclear genome of Chinsurah Boro II with that of Taichung 65 (Shinjyo 1975). We have also developed DNA markers that are tightly linked to the *Rf-1* gene using these NILs (Akagi et al. 1996; Ichikawa et al. 1997). In the investigation reported here, we attempted to identify the *Rf-1* gene with a positional cloning strategy using these NILs for precise linkage analysis. We found duplicate open reading frames (ORFs) (*Rf-1A*, *Rf-1B*) that encode a PPR protein at the *Rf-1* locus. Of these, *Rf-1B* encoded only a truncated protein; therefore we concluded that only *Rf-1A* could be *Rf-1*.

Materials and methods

Plant materials and linkage analysis

Two NILs, MTC-10A [(*cms-bo*)*rf-1/rf-1*] and MTC-10R [(*cms-bo*)*Rf-1/Rf-1*], were used (Akagi et al. 1994) throughout this work. The cytoplasm of MTC-10A was derived from Chinsurah Boro II, and its nuclear genome was derived from Taichung 65; thus MTC-10A is a BT-type CMS line. MTC-10R has the same cytoplasmic and nuclear genomes as MTC-10A except for the presence of *Rf-1*, which was derived from Chinsurah Boro II (Shinjyo 1975). The cytoplasm and the restoring gene of the NILs had been introduced into Taichung 65 from Chinsurah Boro II by recurrent backcrosses. These lines were originally referred to as BT-C and BT-A, respectively, by Shinjyo (1969). IR24, IR36 and MTC-18R, which can also restore the fertility of BT-type CMS, were also used. MTC-18R retained the *Rf-1* gene derived from IR8 (Ichikawa et al. 1997).

A population consisting of 300 plants of BC₁ progeny was used to analyze linkage of the *Rf-1* gene and DNA markers. These plants were derived from backcrossing the F₁ hybrid of MTC-10A and MTC-10R with MTC-10A. Since MTC-10A shows gametophytic CMS, pollen without the *Rf-1* gene is abortive and only pollen with the *Rf-1* gene can develop normally in F₁ plants between MTC-10A

and MTC-10R. Therefore, the genotype of the BC₁ plants between the MTC-10A and F₁ was *Rf-1/rf-1*. The recombination value was converted into map distances (centiMorgans) using the Kosambi function (Kosambi 1944).

Seven BC₁ plants with recombination around the *Rf-1* gene were selected and then used for fine mapping of the *Rf-1* gene. While the ovules of BC₁ plants with the BT-type cytoplasm can develop normally even without the *Rf-1* gene, their pollen will develop only when they carry the *Rf-1* gene. Therefore, BC₁F₂ progeny in these BC₁S were either heterozygous (*Rf-1/rf-1*) or homozygous (*Rf-1/Rf-1*) for the *Rf-1* gene. Twenty-two BC₁F₂ plants were selected with a heterozygous allele of the microsatellite marker (68923-7). We used 6,104 BC₁F₃ progeny derived from self-pollination of these BC₁F₂ plants to precisely locate the *Rf-1* gene.

Polymerase chain reaction

Total rice DNA was prepared from green leaves according to Lichtenstein and Draper (1985). Crude DNA was also extracted from the leaves of BC₁, BC₁F₂ and BC₁F₃ seedlings (Akagi et al. 1997). Each leaf was dried at 70°C for 2 h and then homogenized with a small glass bead in a 1.5-ml Eppendorf tube using a vortex mixer. Crude DNA was then extracted according to the method of Edwards et al. (1991) and subjected to PCR for linkage analysis.

Primer sets were designed using genomic sequences of rice cv. Nipponbare. A total of 112 primer sequences were selected from the region between 5,5071 and 7,7373 of the nucleotide sequence of the bacterial artificial chromosome (BAC) clone OSJNBa0017E08 (AC068923, gi: 17298629). Corresponding genomic regions in MTC-10A (recessive allele) and MTC-10R (dominant allele) were amplified by PCR using these respective primers. After DNA amplification, PCR products were purified using a PCR purification kit (QIAGEN, Valencia, Calif.) and then sequenced. DNA markers were developed on the basis of polymorphisms between sequences of two NILs (Table 1). PCR analyses were run in 20- μ l aliquots of a buffer consisting of 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.5 or 1 U of *TAKARA Taq* (TAKARA), 4 nmol dNTP, 10 ng of genomic DNA and 10 pmol of each set of primers listed in Table 1 in a Thermal Cycler 9600 (Perkin-Elmer, Foster City, Calif.). Thirty-five PCR cycles, each consisting of 10 s of denaturation at 94°C, 30 s of annealing at 55°C or 60°C and 2 min of polymerization at 72°C, were performed. PCR products for cleaved amplified polymorphic sequence (CAPs) markers were digested with a corresponding restriction endonuclease. The PCR products were electrophoresed on 1.8–3% MetaPhor Agarose gels (FMC, Rockland, Me.).

Analysis of RNA

Total RNA from various rice tissues was extracted using an RNeasy Plant Mini kit (QIAGEN) following the manufacturer's protocol. To remove contaminating DNA, total RNA was suspended in 100 μ l 0.8 mM MgCl₂, 5 mM DTT and 40 mM Tris-HCl, pH 7.5 and digested with RNase-free DNaseI (50 U/ml; TAKARA) for 30 min at 37°C. It was then extracted with phenol-chloroform and precipitated with EtOH. First-strand cDNA was synthesized from 1 μ g of total RNA with Oligo dT Adaptor Primer FB using a High Fidelity RNA PCR kit (TAKARA) according to the supplier's protocol. PCR was performed with primer pairs (see below) using *TAKARA Ex Taq* (TAKARA) or *TAKARA LA Taq* with GC I buffer (TAKARA). The primer pairs used were as follows: a-1F, ttgttacaggtgcacagacc; a-1R, gctctgatctacggcttgaa; a-2F, agcttgaccctgcataacg; a-2R, taagacaaactcgttgcgg; a-3F, caaccgatgagaatgcgta; a-3R, taagacaaactcgttgcgg; b-1F, acagtccaggaaccaca; b-1R, ctgttgactctggcatgcta; b-2F, gtaggtcacaacatttggcg; b-2R, aagctctc-gagactactgcac; b-3F, gtaggtcacaacatttggcg; b-3R, tgctgcacctgacg-tag; b-4F, gtaggtcacaacatttggcg; b-4R, gctgcacctgacgtaggc. For the β -tubulin gene, the primers used were: rice tubulin F (tggtcg-gattcgcccgctg) and rice tubulin R (ttacatgtcctcagcctct) (Nakagawa et al. 2000). Amplified DNA fragments were cloned into pT7Blue (Novagen) and sequenced. For 3'-RACE (rapid amplification of cDNA end), the first-strand cDNA was amplified by PCR with Adaptor Primer FB and a specific primer for each gene. Nested PCR was also performed with Adaptor Primer FB and another specific primer for each gene. Amplified DNA fragments were cloned into pT7Blue (Novagen) and then sequenced.

Sequence analysis

Genomic regions covering the *Rf-1* locus of MTC-10R and MTC-10A were amplified using primers based on nucleotide sequences obtained from Nipponbare. The nucleotide sequences of the amplicons were determined by a direct sequencing analysis. We analyzed the nucleotide sequences of mRNA using at least five cloned cDNA fragments and also by a direct sequencing analysis. Nucleotide sequences were determined using a DNA sequencing system (ABI 373S; Applied Biosystems, Foster City, Calif.). DNA sequences were analyzed using GENETYX-MAC (Software Development, Tokyo). Genomic sequences were also analyzed using gene prediction programs, GENESCAN and FGENESH.

Table 1 DNA markers around the *Rf-1* gene

Marker	Forward primer	Reverse primer	Detection
G2155	cagggataggattgtatggtg	gaggtaacatttgggagatgtaaag	CAPs; <i>DdeI</i>
G4003	tcgacgagtacctgaacgatcc	ttgattgtcctcaaagggatg	CAPs; <i>HindIII</i>
C1361R	gtcacttcagtgctgcttc	ggaggctaaggtcacgctgctc	R-Dominant
C1361A	tgctctgtgtccagcc	cacaccagatcttctaatac	A-Dominant
G291	tgcttgatgcaactattgacttg	agatgtgtaatggactttcta	CAPs; <i>PvuII</i>
79888-1	acaaggtggtgggctagatg	ctcgcgttctttttccac	A-Dominant
79888-2	gtggagaaaagaacgcgag	gatccgtgctcattccagat	CAPs; <i>MboI</i>
68923-1	agggttacttgacggacctg	tggttctgctgaacacatc	CAPs; <i>HaeIII</i>
68923-2	tcgttttgataaggatacgg	aacgtactctcctcagactg	Microsatellite
68923-3	cactggcgcacttcaataaa	tcataaacacgcatccactt	Direct sequence
68923-4	acatgcgccgaaatttcta	cagcatagccacatcttcta	CAPs; <i>XhoI</i>
68923-5	tgaccagatcacctctctc	tgctgcttaagtaccatgc	Direct sequence
68923-6	tgcatgtgtctgtctgtctg	cagccactcttaggggtgaa	CAPs; <i>MboI</i>
68923-7	ccaacaagtcccacaatag	ctcgtatagtaacttgagacg	Microsatellite
68923-8	tcctctctaatagactg	ggtgtcgtataccactgtca	Insertion/deletion
68923-9	tgtagtctgagctctcgaa	actgttgacactgtttctgg	CAPs; <i>MspI</i>
68923-10	aagttcaagctgcgttccac	gaggcagcaaggttgaatc	Direct sequence
68923-11	agctgagtcgcttctctg	catttccagctgccttcc	CAPs; <i>HinfI</i>
68923-12	gtgttctgttccctgcatt	gtgaaaacttctttagatg	R-Dominant

Results

Mapping of the *Rf-1* gene

To clarify the precise position of the *Rf-1* gene on chromosome 10, we selected four RFLP markers, G2155, G4003, G291 and C1361, from a high-density molecular genetic map (Kurata et al. 1994) and converted them into PCR-based markers (Table 1). When a plant lacked the allele from MTC-10R at one of these markers, it was recombined between the marker and the *Rf-1* gene. The calculated position and the distance between the *Rf-1* gene and each DNA marker on chromosome 10 are shown in Fig. 1. The *Rf-1* gene was located between C1361 and fL601, with map distances of 1.00 ± 0.58 cM and 1.50 ± 0.67 cM, respectively (Fig. 1). The distance between the *Rf-1* gene and OSRRf was almost the same that reported by Kurata et al. (1994) and Akagi et al. (1996).

By a BLAST homology search, the sequence of fL601 was located on the rice BAC clone of OSJNBa0017E08, and the sequences of C1361 and OSRRf were located on that of OSJNBa0078O01 (Fig. 1). These two BAC clones overlapped each other on chromosome 10 (Fig. 1). Thus, the *Rf-1* gene should lie between C1361 and fL601 on these BAC clones at a physical distance of 219 kb.

High-resolution mapping of the *Rf-1* gene

The locus of the *Rf-1* gene within this 219-kb fragment of chromosome 10 was clarified more precisely with 6,104 BC₁F₃ progeny. All of the 723 plants that were randomly selected from these 6,104 plants set seeds normally.

Fourteen DNA markers between fL601 and C1361 were newly developed based on the nucleotide sequences of the two BAC clones (Table 1, Fig. 2). The population of BC₁F₃ plants was analyzed using OSRRf, C1361 and 68923-11, and nine plants were found to be carrying a homozygous allele from MTC-10A (Fig. 2), thereby showing recombination around the *Rf-1* gene. The genotypes of these plants were characterized more precisely using 17 DNA markers between C1361 and fL601 (Fig. 2). Plants FM12-19 and FM24-26 were recombinant between 68923-6 and 68923-7, while FM0-626 was recombinant between 68923-6 and 68923-9 (Fig. 2). The *Rf-1* gene lay between two markers, 68923-6 and 68923-9, and the remaining parts (shown by open bars) were derived from MTC-10A (Fig. 2).

Sequence analysis of the *Rf-1* gene locus

This region between 68923-6 and 68923-9 is about 22.7 kbp long in the Nipponbare genome. Because Nipponbare is a non-restorer for BT-type CMS and presumably carries the recessive gene, *rf-1*, the nucleotide sequence of the corresponding region of MTC-10R was determined. In this MTC-10R sequence, we searched

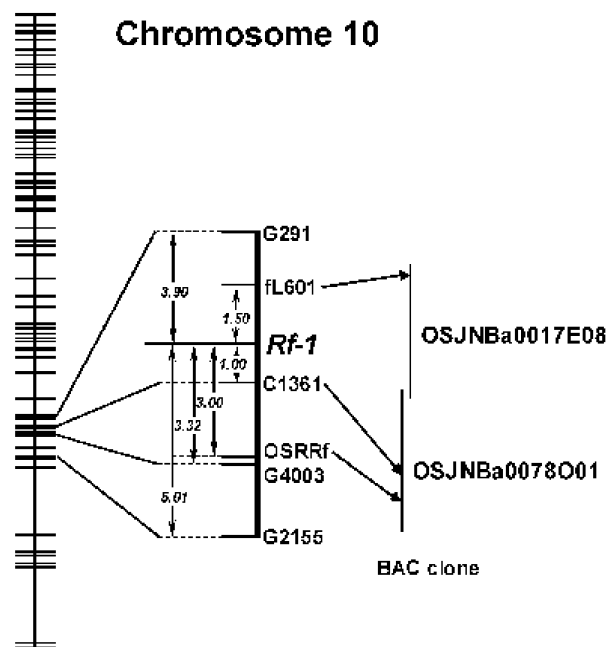


Fig. 1 Genetic mapping of the *Rf-1* gene on chromosome 10 of rice. Linkage orders between the *Rf-1* gene and six DNA markers are shown to the right of the vertical bar; genetic distances (in centiMorgans) are shown to the left of the vertical bar. The positions of four DNA markers designed from RFLP markers are indicated on the high-density RFLP map of chromosome 10 from the Rice Genome Research Program (Kurata et al. 1994). The relative positions of the sequences homologous to fL601, C1361 and OSRRf are indicated on the BAC contig

coding regions using both gene prediction programs and reverse transcription (RT)-PCR for detecting mRNA.

Two ORFs (*Rf-1A* and *Rf-1B*) without introns were found in this 22.4-kb region of MTC-10R (Fig. 3). cDNAs corresponding to these ORFs were cloned by RT-PCR and expanded to their 3' ends by 3'-RACE (Fig. 3). An additional mRNA, transcribed from the region downstream of the *Rf-1A* gene containing several introns, was also detected by RT-PCR. The mRNA contained several stop codons and may be translated into a truncated and nonfunctional protein. Thus, only *Rf-1A* and *Rf-1B* were considered candidates for the *Rf-1* gene.

In the counterpart of both genes on the MTC-10A genome, we found 1-bp and 574-bp deletions in the region corresponding to the *Rf-1A* gene, and 1-bp and 21-bp deletions in the region corresponding to *Rf-1B*. An insertion of a 970-bp fragment homologous to the long terminal region (LTR) was also found in the *Rf-1B* gene (Fig. 3). These genes on the MTC-10A genome may have lost their function.

Rf-1 gene encodes PPR protein

The nucleotide sequences of the *Rf-1A* and *Rf-1B* genes are highly conserved. The *Rf-1A* gene encodes a 791-amino acid protein and the *Rf-1B* gene encodes a 332-

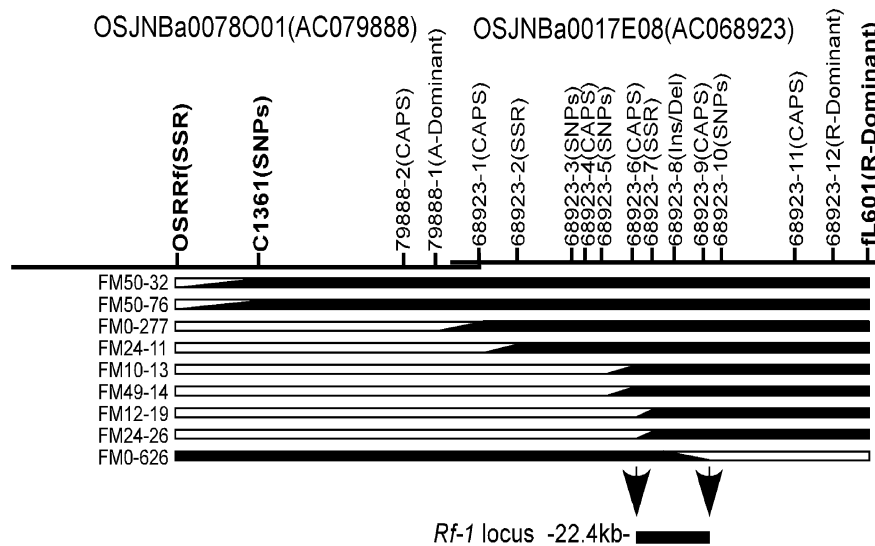


Fig. 2 Delimitation of the *Rf-1* locus and positions of DNA markers around the *Rf-1* locus. The positions of 17 DNA markers designed from the sequences of BAC clones (OSJNBa0078O01 and OSJNBa0017E08) and the genotype of nine plants in which recombination occurred between OSRRf and fL601 are indicated. Except for fL601 and 79888-1, DNA markers are co-dominant, allowing the discrimination of *Rf-1/Rf-1* plants from *Rf-1/rf-1* ones.

White regions are chromosomal segments that are homozygous for the recessive allele from MTC-10A, and black regions are ones that are heterozygous for the allele from MTC-10A and MTC-10R (*Rf-1/rf-1*). Recombination occurred in the region of transition from black to white. The *Rf-1* locus was delimited to the approximately 22.7-kb-long region between 68923-6 and 68923-9 in the Nipponbare genome

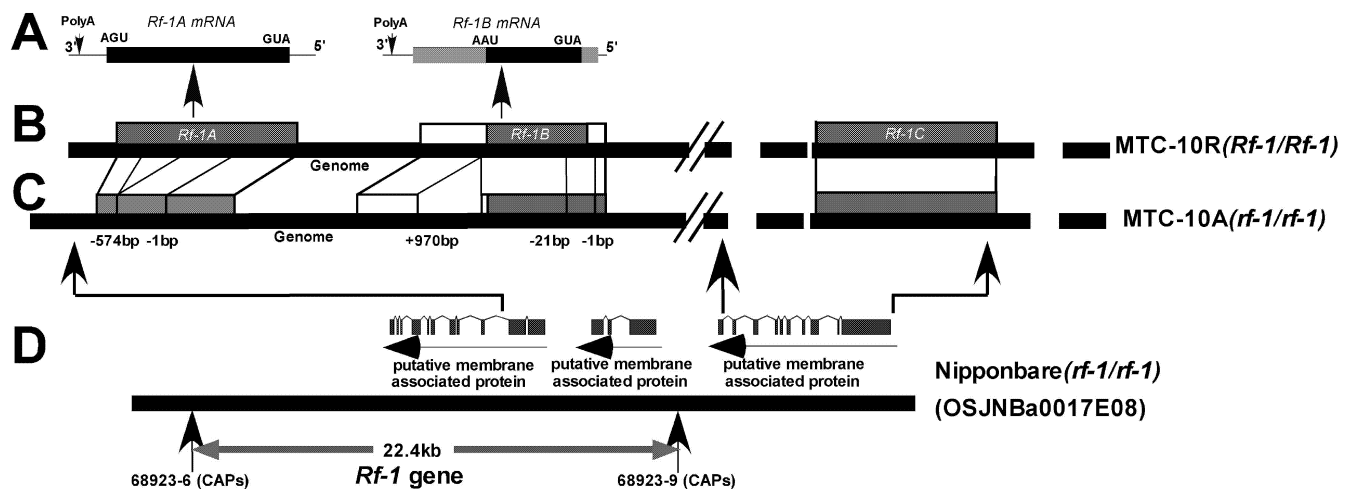


Fig. 3A–D Genomic organization of the *Rf-1* locus. **A** The structural features and direction of transcripts corresponding to the *Rf-1A* gene and the *Rf-1B* gene of MTC-10R. Arrows indicate the positions of poly(A) sites. **B** The genomic structure of the *Rf-1* locus in MTC-10R containing two duplicate open reading frames (*Rf-1A* and *Rf-1B*). **C** The genomic structure of the recessive allele (MTC-10A). The relative positions and lengths of deletions (–) and an insertion (+) were found in MTC-10A. Another homolog

(*Rf-1C*) is located upstream from the *Rf-1B* gene in both MTC-10R (**B**) and MTC-10A (**C**). **D** The structures and relative positions of three putative genes around the *Rf-1* locus in the Nipponbare genome. The positions of the boundary DNA markers are indicated by arrows. The putative genes in the Nipponbare genome corresponding to *Rf-1A*, *Rf-1B* and *Rf-1C* are registered in the database as OSJNBa001717E08.19, OSJNBa001717E08.19 and OSJNBa001717E08.19, respectively

amino acid protein (Fig. 4). The deduced amino acid sequence of Rf-1A showed 86.7% identity to that of Rf-1B (Fig. 4). NCBI Conserved Domain Search identified the 16 sets of a PPR motif in the *Rf-1A* gene (Fig. 4). The PPR-motives (PPR-1 to PPR10) were identified with E-values above 0.01, and the remaining motives (PPR-11 to PPR16) were identified with E-values in the range of 0.01

to 1. The *Rf-1B* gene also had nine sets of the PPR motif (Fig. 4). The protein encoded by the *Rf-1A* gene contained a putative mitochondrial transit peptide and may locate in mitochondria, as predicted by the method reported by Emanuelsson et al. (2000). Rf-1B may lack this function, in that a transit peptide was not predicted from its putative amino acid sequence even though its nucleotide sequence

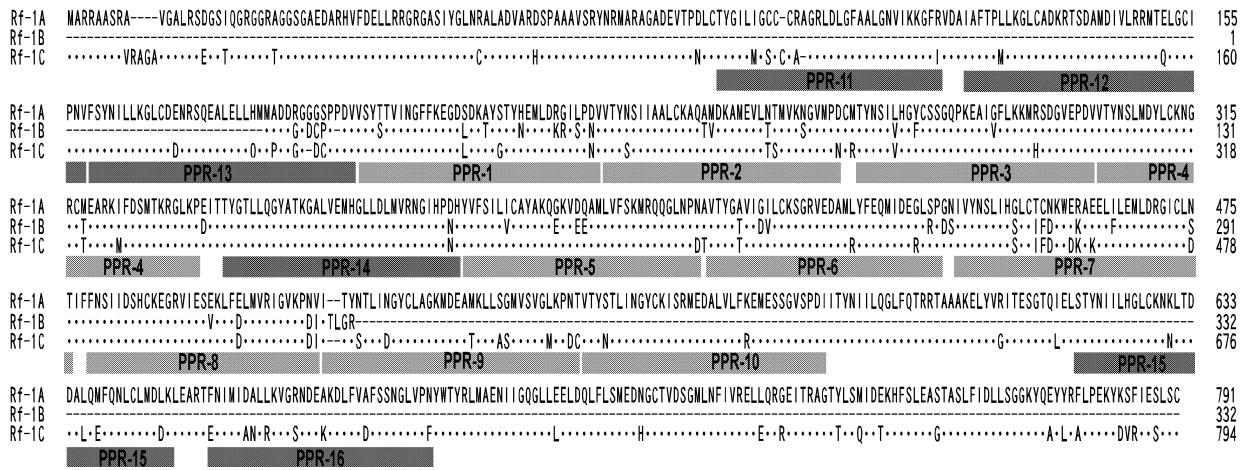


Fig. 4 Comparison of deduced amino acid sequences between Rf-1A, Rf-1B and Rf-1C of MTC-10R. Dots represent amino acids identical to those in Rf-1A. Sixteen repeats of the PPR motif, which

consists of 35 amino acids are indicated *under* the three amino acid sequences

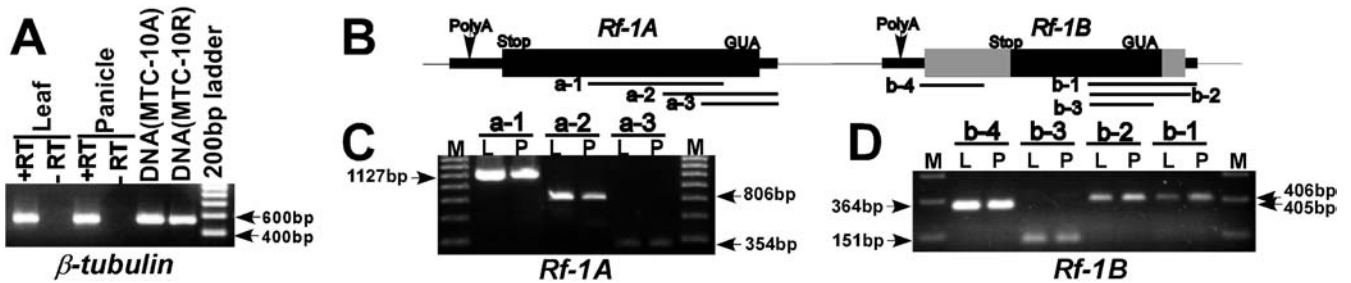


Fig. 5A–D Expression of the *Rf-1A* and *Rf-1B* genes. RNA of leaves (L) and developing panicles (P) of MTC-10R (*Rf-1/Rf-1*) was subjected to RT-PCR with primers specific for the β -tubulin gene (A), *Rf-1A* gene (C) and *Rf-1B* gene (D). –RT, +RT With and

without reverse transcriptase in the reaction, respectively. **B** The regions amplified by RT-PCR are indicated *under* the structures of *Rf-1A* and *Rf-1B* mRNAs. M Marker. Sizes of amplicons are also indicated

is highly homologous to that of the *Rf-1A* gene. Furthermore, the *Rf-1B* gene in MTC-10R has another premature stop codon (TAA) (Figure 3) encoding a truncated and nonfunctional protein.

The rice lines IR24, IR36 and MTC-18R can also restore BT-type CMS. However, the proteins encoded by the *Rf-1B* gene in their respective genomes were predicted to have a function of mitochondrial targeting and 14 sets of a PPR motif (PPR-11 and PPR-12 missing; data not shown). A frame-shift mutation produced by a single nucleotide extension in MTC-10R (data not shown) and a 26-bp deletion that generated new stop codons resulted in a loss of function of Rf-1B in MTC-10R.

A gene (*Rf-1C*) homologous to the *Rf-1A* gene was found upstream of the *Rf-1B* gene (Fig. 3). This gene also encoded a PPR protein with 88.6% homologous to the Rf-1A protein (Fig. 4). The deduced amino acid sequence of the *Rf-1C* gene of MTC-10R was 99.9% and 99.5% identical to those of MTC-10A and Nipponbare, respectively. While the *Rf-1A* gene and the *Rf-1C* gene encoded highly homologous PPR proteins, the *Rf-1C* gene is beyond the cross-over exhibited by the recombinant FM0-626 (Fig. 2) and cannot be the *Rf-1* gene.

Expression profiles

The expression of both the *Rf-1A* gene (Fig. 5B, C) and *Rf-1B* gene (Fig. 5B, D) was detected in panicles at the booting stage with the development of pollen and in green leaves of MTC-10R by RT-PCR. These DNA fragments were only amplified in the presence of reverse transcriptase, similar to the results for β -tubulin (Fig. 5A). We confirmed that these DNA fragments had been specifically amplified from the mRNA of each gene by a direct sequencing analysis (data not shown).

Duplicate *Rf-1* gene homologs in the rice genome

In addition to the *Rf-1A*, *Rf-1B* and *Rf-1C* genes, six copies of genes homologous to the *Rf-1A* gene were found on chromosome 10 in the Nipponbare genome (Fig. 6). Seven of them encoded only short proteins. The remaining two genes (*Rf-68950-2* and *Rf-1C*) encoded proteins as long as Rf-1A of MTC-10R and may be functional (Fig. 6). The *Rf-1C* gene encoded a protein containing a

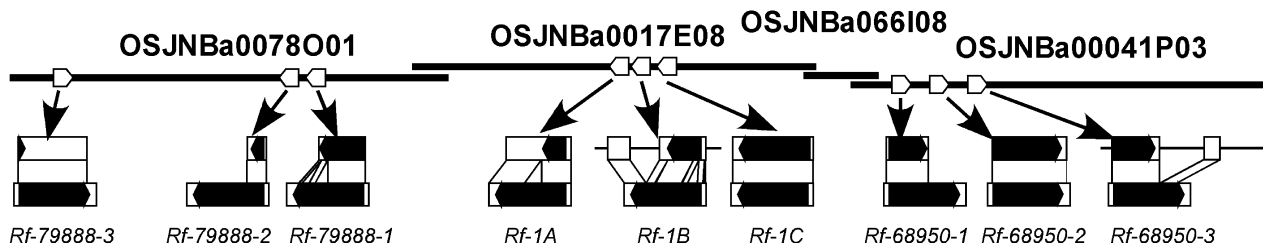


Fig. 6 BAC contig covering the *Rf-1* locus in the Nipponbare genome showing the genomic structure of the regions homologous to the *Rf-1A* gene in the Nipponbare genome. *Broad open arrows* represent the positions and direction of nine *Rf-1A* gene homologs. The structures of these homologues are indicated *under* the contig and are compared with the *Rf-1A* gene. *Broad black arrows*

represent regions that can be translated into a protein. *Rf-79888-3*, *Rf-79888-2*, *Rf-79888-3*, *Rf-68950-1*, *Rf-68950-2* and *Rf-68950-3* partly correspond to the predicted genes, OSJNBa0078O01.29, OSJNBa0078O01.20, OSJNBa0078O01.15, OSJNBa0041P03.13, OSJNBa0041P03.12 and OSJNBa0041P03.10, respectively

mitochondria-targeting peptide and 16 repeats of the PPR motif (Fig. 4).

Discussion

Positional cloning of the *Rf-1* gene

During the course of the investigation reported here, the *Rf-1* locus was delimited in a genomic region of 22.4 kb by linkage analysis using 6,104 BC₁F₃ progeny (Fig. 2). Two ORFs (*Rf-1A* and *Rf-1B*) encoded proteins containing a PPR motif (Fig. 3) that are transcribed in different parts of the plant with the MTC-10R genome (Figs. 3, 5). Furthermore, with respect to the recessive allele (*rf-1*), the genes corresponding to both the *Rf-1A* and *Rf-1B* genes possess frame-shift mutations and are predicted to be nonfunctional (Fig. 3).

Rf-1A protein with a signal peptide was predicted to localize in mitochondria, whereas *Rf-1B* of MTC-10R lacked such a signal peptide (Fig. 4). In addition, the *Rf-1B* gene of MTC-10R has not only a frame-shift mutation but also a 26-bp deletion that generates new stop codons within the gene; therefore, the *Rf-1B* gene product lacked functionality (data not show). In comparison with the *Rf-1B* gene in MTC-10R, in three other restorer lines, the gene corresponding to the *Rf-1B* gene encoded a protein with a mitochondria-targeting signal and 16 repeats of a PPR motif. The segregates that carried only the *Rf-1A* gene as a functional gene in the *Rf-1* locus had fertile seeds. This supports the conclusion that the *Rf-1A* gene is the *Rf-1* gene.

The *Rf-1* gene encodes a PPR protein

Proteins with a PPR motif constitute a large family in higher plants (Aubourg et al. 2000; Small and Peeters 2000). The PPR motif consists of a 35-amino acid repeat and has been predicted to be a macromolecule-binding motif (Small and Peeters 2000). It has been suggested that PPR proteins containing tandem repeats of a PPR motif are sequence-specific RNA- or DNA-binding proteins

(Lahmy et al. 2000; Small and Peeters 2000). In the *Arabidopsis* genome, about two-thirds of these proteins are predicted to be targeted to organelles (Small and Peeters 2000). These findings suggest the possibility that PPR proteins are involved in RNA metabolism in organelles. In many plant species, *Rf* genes alter transcription profiles in CMS mitochondria. In fact, one PPR protein, the product of the *Rf-PPR592* gene, can restore CMS in *Petunia* (Bentolila et al. 2002).

The *Rf-1A* gene identified here encodes a 791-amino acid protein (*Rf-1A*) consisting of a signal peptide that targeted mitochondria and 16 repeats of a PPR motif. Since the *Rf-1* gene is expected to play a role in processing the transcript from an *atp6/orf79* region in mitochondria (Iwabuchi et al. 1993; Akagi et al. 1994), there may be no discrepancy between the expected function of the *Rf-1* gene and the deduced function of the *Rf-1A* gene in processing *atp6/orf79* transcript and suppressing BT-type CMS.

In *Petunia*, the restorer gene *Rf-PPR592* is expressed only in floral buds (Bentolila et al. 2002). In the case of rice, the *Rf-1A* gene, which restores BT-type CMS, is expressed not only in an organ that contains immature pollen but also in a vegetative organ (Fig. 5). Furthermore, the *atp6/orf79* transcript was also found to be processed in the presence of the *Rf-1* gene in callus (Akagi et al. 1994). The production of intact *atp6* mRNA would not be essential to the somatic growth of plants but would be essential for the maturation of pollen. Considerable energy is required for the maturation of pollen, and this energy can be generated by an intact form of *atp6* mRNA and/or a sufficient amount of *atp6*; if not, aborted pollen results.

Co-evolution of *Rf* genes and mitochondrial genome

In the Nipponbare genome, nine regions of sequences homologous to the *Rf-1A* gene were found adjacent to the *Rf-1* locus (Fig. 6). Their highly conserved sequences suggest that they might have evolved by replication from their ancestor during rice evolution. Some of these sequences could encode functional PPR proteins even

though the *Rf-1A* gene and the *Rf-1B* gene have lost their function in the Nipponbare genome. In *Brassica*, two kinds of cytoplasm, *nap* and *pol*, cause CMS. A nuclear restorer gene called *Rfp* is able to modify transcripts of the *orf224/atp6* region in *pol* cytoplasm. On the other hand, *nap* CMS is restored by another nuclear gene, *Rfn* (*Mmt*), which is involved in modifying three kinds of transcripts from different mitochondrial regions, including *nap* CMS-associated *orf222*. Based on the map positions of *Rfn* and *Rfp*, these restorer genes are located on the same nuclear locus (Singh et al. 1996; Li et al. 1998). In *Sorghum* A3 CMS, genes related to processing of the *orf107* and *urf209* transcripts were found to cosegregate, indicating that they may be a single nuclear gene or tightly linked genes (Tang et al. 1998). Two duplicate genes, *Rf-PPR592* and *Rf-PPR591*, were located at the *Rf* locus in *Petunia*. While *Rf-PPR592* was able to restore *petunia* CMS, the function of *Rf-PPR591* was unclear (Bentolila et al. 2002). Koizuka et al. (2003) suggested that only a four amino acid substitution in the PPR of the *Rf* gene of CMS Kosenia radish changed the gene function. In rice, at least four copies of PPR genes exist around the *Rf-1* locus, and these vary somewhat with respect to their amino acid sequence. The *Rf-1A* gene and *Rf-1B* gene showed a loss of function in MTC-10A (Fig. 3). The *Rf-1B* gene lost its function in MTC-10R but not in other restorer lines, such as IR24, IR36 and MTC-18R. Such diverse and duplicate PPR genes may account for specialized functions in processing specific transcripts in rice.

Relationship between duplicate *Rf-1* gene homologs and CMS in rice

Within rice, there are several CMS/*Rf* systems (Virmani and Shinjyo 1988). The cytoplasm derived from wild rice, called *wild abortive*, causes WA-type CMS in a sporophytic manner and is widely used for the *indica* subspecies. Two fertility restorer genes, *Rf3* and *Rf4*, are required for the production of viable pollen in WA-type CMS. *Rf3* and *Rf4* have been mapped to chromosomes 1 and 10, respectively (Yao et al. 1997; Zhang et al. 1997). The latter is located adjacent to the *Rf-1* gene (Tan et al. 1998; Jing et al. 2001). Since MTC-10R (referred to as T65R in the literature) does not restore WA-type CMS (Teng and Shen 1994), the *Rf-1A* gene is not involved in the restoration of WA-type CMS. Although the *Rf-1B* gene, located at the delimited *Rf-1* locus, is nonfunctional in MTC-10R, it did encode a functional protein in other restorer lines, such as IR24, IR36 and MTC-18R (Fig. 5). Unlike MTC-10R, IR24 and MTC-18R can also restore WA-type CMS. Since the structure of the functional *Rf-1B* protein is almost the same as that of *Rf-1A* protein, *Rf-1B* might target mtRNAs other than the *atp6/orf79* transcript and restore WA-type CMS; thus, it may be *Rf4*.

Recently, Kazama and Toriyama (2003) reported that a gene (*PPR8-1*) encoding a PPR protein is a candidate for the *Rf-1* gene based on results obtained using a transgenic

strategy. They demonstrated that this gene is involved in the processing of the transcript in mitochondrial *atp6* in a transgenic rice callus. However, they did not confirm that *PPR8-1* restored fertility in regenerated plants. In our study, we were able to show that *Rf-1A* is the *Rf-1* gene by applying a precise positional cloning strategy. In our analysis, only segregates that carried a limited region around the *Rf-1A* gene had fertile seeds, and we presumed that *Rf-1A* restores fertility through processing of the *atp6/orf79* transcript. The nucleotide sequences of *Rf-1A* and *PPR8-1* were identical even in their flanking regions, indicating these are the same gene. Putting these two results together, it can be concluded that the product of the *Rf-1A* gene processes the transcript from the *atp6/orf79* region, subsequently restoring fertility as a result, and that the *Rf-1* gene itself is present in diversified forms among restorers for BT-type CMS.

Acknowledgements The authors thank Dr. Shinjyo for kindly providing the seeds of the near-isogenic lines of rice. This research was supported in part by a Grant-in-Aid for Scientific Research C (No. 12660008) from the Ministry of Education, Science and Culture, Japan.

References

- Akagi H, Sakamoto M, Shinjyo C, Shimada H, Fujimura T (1994) A unique sequence located downstream from the rice mitochondrial *atp6* may cause male sterility. *Curr Genet* 25:52–58
- Akagi H, Yokozeki Y, Inagaki A, Nakamura A, Fujimura T (1996) A codominant DNA marker closely linked to the rice nuclear restorer gene, *Rf-1*, identified with inter-SSR fingerprinting. *Genome* 39:1205–1209
- Akagi H, Yokozeki Y, Inagaki A, Fujimura T (1997) Highly polymorphic microsatellites of rice consist of AT repeats, and classification of closely related cultivars with these microsatellite loci. *Theor Appl Genet* 94:61–67
- Aubourg S, Boudet N, Kreis M, Lecharny A (2000) In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. *Plant Mol Biol* 42:603–613
- Bentolila S, Alfonso AA, Hanson MR (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc Natl Acad Sci USA* 99:10887–10892
- Brown GG, Formanová N, Jin H, Wargachuk R, Dendy C, Patil P, Laforest M, Zhang J, Cheung WY, Landry BS (2003) The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant J* 35:262–272
- Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genome DNA for PCR analysis. *Nucleic Acids Res* 19:1349
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* 300:1005–1016
- Fujimura T, Akagi H, Oka M, Nakamura A, Sawada R (1996) Establishment of a rice protoplast culture and application of an asymmetric protoplast fusion technique to hybrid rice breeding. *Plant Tissue Cult Lett* 13:243–247
- Hossain M (1996) Economic prosperity in Asia: implications for rice research. In: Khush GS (ed) *Rice genetics III*. IRRI, Manila, Philippines, pp 3–16
- Ichikawa N, Kishimoto N, Inagaki A, Nakamura A, Koshino Y, Yokozeki Y, Oka M, Samoto S, Akagi H, Higo K, Shinjyo C, Fujimura T, Shimada H (1997) A rapid PCR-aided selection of a rice line containing the *Rf-1* gene, which is involved in

- restoration of the cytoplasmic male sterility. *Mol Breed* 3:195–202
- Iwabuchi M, Kyojuka J, Shimamoto K (1993) Processing followed by complete editing of an altered mitochondrial *atp6* RNA restores fertility of cytoplasmic male sterile rice. *EMBO J* 12:1437–1446
- Jing R, Li X, Yi P, Zhu Y (2001) Mapping fertility-restoring genes of rice WA cytoplasmic male sterility using SSLP markers. *Bot Bull Acad Sin* 42:167–171
- Katsuo K, Mizushima U (1958) Studies on the cytoplasmic difference among rice varieties, *Oryza sativa* L. *Jpn J Breed* 8:1–5
- Kazama T, Toriyama K (2003) A pentatricopeptide repeat-containing gene that promotes the processing of aberrant *atp6* RNA of cytoplasmic male-sterile rice. *FEBS Lett* 544:99–102
- Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, Sakai T, Kawasaki S, Imamura J (2003) Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosenia radish. *Plant J* 34:407–415
- Kosambi DD (1944) The estimation of map distances from recombination values. *Ann Eugen* 12:172–175
- Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin S-Y, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Krihara T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang Z-X, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y (1994) A 300-kilobase interval genetic map of rice including 883 expressed sequences. *Nat Genet* 8:365–372
- Lahmy S, Barnech F, Derancourt J, Filipowicz W, Delseny M, Echeverria M (2000) A chloroplastic RNA-binding protein is a new member of the PPR family. *FEBS Lett* 480:255–260
- Levings CSIII, Brown GG (1989) Molecular biology of plant mitochondria. *Cell* 56:171–179
- Levings CS III, Vasil IK (1995) The molecular biology of plant mitochondria. Kluwer, Dordrecht
- Li X-Q, Jean M, Landry BS, Brown GG (1998) Restorer genes for different forms of *Brassica* cytoplasmic male sterility map to a single nuclear locus that modifies transcripts of several mitochondrial genes. *Proc Natl Acad Sci USA* 95:10032–1 0037
- Li Z, Zho Y (1986) Rice male-sterile cytoplasm and fertility-restoration. In: IRRI (ed) Hybrid rice. IRRI, Manila, Philippines, pp 85–102
- Lichtenstein CP, Draper J (1985) Genetic engineering of plants. In: Glover DM (ed) DNA cloning, vol 3. Practical approach. IRL Press, Oxford, pp 67–119
- Nakagawa Y, Machida C, Machida Y, Toriyama K (2000) Frequency and pattern of transposition of the maize transposable element *Ds* in transgenic rice plants. *Plant Cell Physiol* 41:733–742
- Newton KJ (1988) Plant mitochondrial genomes: Organization, expression and variation. *Annu Rev Plant Physiol Plant Mol Biol* 39:503–532
- Pruitt KD, Hanson MR (1991) Transcription of the *Petunia* mitochondrial CMS-associated *Pcf* locus in male sterile and fertility-restored lines. *Mol Gen Genet* 227:348–355
- Schuable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends Plant Sci* 3:175–180
- Shinjo C (1969) Cytoplasmic genetic male sterility in cultivated rice, *Oryza sativa* L. II. The inheritance of male sterility. *Jpn J Genet* 44:149–156
- Shinjo C (1975) Genetical studies of cytoplasmic male sterility and fertility restoration in rice, *Oryza sativa* L. *Sci Bull Coll Agric Univ Ryukyus* 22:1–51
- Shinjo C (1984) Cytoplasmic male sterility and fertile restoration in rice having genome A. Biology of rice. Japan Science Societies, Elsevier Press, Tokyo
- Singh M, Hamel N, Menassa R, Li X-Q, Young B, Jean M, Landry BS, Brown GG (1996) Nuclear genes associated with a single *Brassica* CMS restorer locus influence transcripts of three different mitochondrial gene regions. *Genetics* 143:505–516
- Small ID, Peeters N (2000) The PPR motif—a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* 25:46–47
- Tan KL, Vanavichit A, Amornsilpa S, Tragoonrung S (1998) Genetic analysis of rice CMS-WA fertility restoration based on QTL mapping. *Theor Appl Genet* 96:994–999
- Tang HV, Pring DR, Shaw LC, Salazar RA, Muza FR, Yan B, Schertz KF (1996) Transcript processing internal to a mitochondrial open reading frame is correlated with fertility restoration in male-sterile sorghum. *Plant J* 10:123–133
- Tang HV, Chang R, Pring DR (1998) Cosegregation of single genes associated with fertility restoration and transcript processing of sorghum mitochondrial *orf107* and *urf209*. *Genetics* 150:383–391
- Teng L-S, Shen Z-T (1994) Inheritance of fertility restoration for cytoplasmic male sterility in rice. *Rice Genet Newsl* 11:95–97
- Virmani SS, Shinjo C (1988) Current status of analysis and symbols for male-sterile cytoplasm and fertility-restoring genes. *Rice Genet Newsl* 5:9–15
- Wise RP, Gobelman-Werner K, Pei D, Dill CL, Schnable PS (1999) Mitochondrial transcript processing and restoration of male fertility in T-cytoplasm maize. *J Hered* 90:380–385
- Yao FY, Xu CG, Yu SB, Li JX, Gac YJ, Li XH, Zhang Q (1997) Mapping and genetic analysis of two fertility restorer loci in the wild-abortive cytoplasmic male sterility system of rice (*Oryza sativa* L.). *Euphytica* 98:183–187
- Yuan LP (1994) Increasing yield potential in rice by exploitation of heterosis. In: Virmani SS (ed) Hybrid rice technology. New developments and future prospects. IRRI, Manila, Philippines, pp 1–6
- Zhang Q, Bharaj TS, Virmani SS, Huang H (1997) Mapping of the *Rf-3* nuclear fertility restoring gene for WA cytoplasmic male sterility in rice using RAPD and RFLP markers. *Theor Appl Genet* 94:27–33