

Photoperiod- and thermo-sensitive genic male sterility in rice are caused by a point mutation in a novel noncoding RNA that produces a small RNA

Hai Zhou^{1,*}, Qinjian Liu^{1,*}, Jing Li^{1,**}, Dagang Jiang^{1,**}, Lingyan Zhou², Ping Wu¹, Sen Lu¹, Feng Li¹, Liya Zhu², Zhenlan Liu¹, Letian Chen¹, Yao-Guang Liu¹, Chuxiong Zhuang¹

¹State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources; College of Life Sciences, South China Agricultural University, Guangzhou, Guangdong 510642, China; ²Zhongkai University of Agriculture and Engineering, Guangzhou, Guangdong 510225, China

Photoperiod- and thermo-sensitive genic male sterility (PGMS and TGMS) are the core components for hybrid breeding in crops. Hybrid rice based on the two-line system using PGMS and TGMS lines has been successfully developed and applied widely in agriculture. However, the molecular mechanism underlying the control of PGMS and TGMS remains obscure. In this study, we mapped and cloned a major locus, *p/tms12-1* (photo- or thermo-sensitive genic male sterility locus on chromosome 12), which confers PGMS in the *japonica* rice line Nongken 58S (NK58S) and TGMS in the *indica* rice line Peiai 64S (PA64S, derived from NK58S). A 2.4-kb DNA fragment containing the wild-type allele *P/TMS12-1* was able to restore the pollen fertility of NK58S and PA64S plants in genetic complementation. *P/TMS12-1* encodes a unique noncoding RNA, which produces a 21-nucleotide small RNA that we named *osa-smR5864w*. A substitution of C-to-G in *p/tms12-1*, the only polymorphism relative to *P/TMS12-1*, is present in the mutant small RNA, namely *osa-smR5864m*. Furthermore, overexpression of a 375-bp sequence of *P/TMS12-1* in transgenic NK58S and PA64S plants also produced *osa-smR5864w* and restored pollen fertility. The small RNA was expressed preferentially in young panicles, but its expression was not markedly affected by different day lengths or temperatures. Our results reveal that the point mutation in *p/tms12-1*, which probably leads to a loss-of-function for *osa-smR5864m*, constitutes a common cause for PGMS and TGMS in the *japonica* and *indica* lines, respectively. Our findings thus suggest that this noncoding small RNA gene is an important regulator of male development controlled by cross-talk between the genetic networks and environmental conditions.

Keywords: rice; PGMS; TGMS; noncoding RNA; small RNA

Cell Research (2012) 22:649-660. doi:10.1038/cr.2012.28; published online 21 February 2012

Introduction

Asian cultivated rice (*Oryza sativa* L.) is one of the world's most important crops, meeting the stable food demand of more than half of the global population. The development of hybrid rice is a major approach for in-

creasing the yield potential of rice, since hybrid rice varieties have about 20% or more yield advantage over improved inbred varieties [1]. Hybrid rice technologies are mainly based on the three-line and two-line systems. The three-line system uses cytoplasmic male sterility lines, maintainer lines and restorer lines [2, 3]. The two-line hybrid rice is based on the discovery and application of environmentally sensitive genic male sterile (EGMS) lines, which serve as both the male sterile lines and maintainer lines under different environmental conditions [4]. Thus, the two-line hybrid rice system is an important innovation for the better exploitation of hybrid vigor (heterosis). Compared with the three-line system, the advantages of the two-line system include a wider range

*These two authors contributed equally to this work.

**These two authors are the co-second authors.

Correspondence: Chuxiong Zhuang^a, Yao-Guang Liu^b

^aE-mail: zhuangcx@scau.edu.cn

^bE-mail: ygliu@scau.edu.cn

Received 7 December 2011; revised 8 January 2012; accepted 10 January 2012; published online 21 February 2012

of germplasm resources used as breeding parents, better grain quality and higher yields, and benefiting from simpler procedures for breeding and hybrid seed production [5]. In recent years, two-line hybrid rice has been applied for large-scale grain production in China [6].

Cell development in plants is controlled by cross-talk between certain genetic networks and environmental conditions such as photoperiod and temperature. Photoperiod-sensitive genic male sterile (PGMS) lines and thermo-sensitive genic male sterile (TGMS) lines are two major types of EGMS germplasm resources, and have been widely used for the breeding of two-line hybrid rice. Nongken 58S (NK58S) is the first spontaneous PGMS mutant found in 1973 from the *japonica* (*O. sativa* ssp. *japonica*) cultivar Nongken 58 (NK58) [7, 8]. Therefore, the discovery of NK58S has opened the door to the development and large-scale application of two-line hybrid rice in agriculture. NK58S retains complete male sterility when the day length (photoperiod) is longer than 13.75 h during the anther development; it converts to partial or complete male fertility when the day length is shorter than 13.5 h. A high temperature (~29 °C) slightly promotes complete male sterility under long-day conditions [9].

NK58S has been used as a donor parent for breeding of a large number of rice EGMS lines by introgressive backcrossing [10]. Peiai 64S (PA64S) is one such NK58S-derived line with *indica* (*O. sativa* ssp. *indica*) genetic background. With wide compatibility and good agronomic traits, PA64S has become the most widely used female parent for two-line hybrid rice breeding, including being used as the parent of the super-rice variety Liangyoupei9 [6, 11, 12]. However, the male sterility-fertility conversion of PA64S and other *indica* lines derived from NK58S is controlled mainly by temperature rather than by day length: PA64S exhibits male sterility at temperatures higher than 23.5 °C during the anther development, but it converts to male fertility when the temperature is ~21–23 °C [13, 14]. Long-day (14 h) conditions can suppress the degree of sterility-to-fertility conversion under low temperatures (21–23 °C), but short-day (12 h) conditions cannot restore male fertility under high temperatures [14].

To date, a number of loci that control PGMS or TGMS, including reverse PGMS or reverse TGMS, in different rice lines have been mapped to different chromosomes: photoperiod-sensitive genic male sterile genes *pms1* [15, 16], *pms2* [15], *pms3* [17–20]; reverse photoperiod-sensitive genic male sterile genes *rpms1* [21], *rpms2* [21]; thermo-sensitive genic male-sterile genes *tms1* [22], *tms2* [23, 24], *tms3* [25, 26], *tms4* [27], *tms5* [5, 28–31], *tms6* [32], *tms6(t)* [33]; photoperiod-thermo-

sensitive genic male sterile genes *ptgms2-1* [34], *pms1(t)* [35]; and reverse thermo-sensitive genic male-sterile genes *rtms1* [36]. Despite the wide application of PGMS and TGMS lines in hybrid rice breeding, the genes responsible for PGMS and TGMS in these lines have not been cloned, and the molecular mechanisms underlying these traits remain unclear.

MicroRNAs (miRNAs) and small interfering RNAs (siRNAs) are two major types of endogenous small, 20–30 nucleotide (nt) RNAs that are widely distributed in eukaryotes. They regulate a number of biological processes, including cell development, metabolic control and adaptive responses to biotic and abiotic stresses, by suppressing gene expression via acting either on RNA to guide cleavage or translational repression or on DNA to mediate chromatin remodeling [37–40]. miRNAs are generated from single-stranded precursor transcripts that self-fold into stem-loop structures. Mature miRNAs form complexes mainly with the Argonaute protein; they can then bind to target mRNAs, after which they cleave and degrade the mRNA strands [41–43]. In addition, miRNAs are involved in translational suppression [44, 45]. siRNAs are derived from long double-stranded RNAs by an RNase III named DCL (Dicer-like) in plants [46–48], where several types of siRNAs have been discovered: *trans*-acting siRNAs (ta-siRNAs), natural antisense transcript-derived siRNAs (nat-siRNAs) and repeat-associated siRNAs (ra-siRNAs) [49]. ta-siRNAs arise from defined genetic loci (TAS loci) through an miRNA-dependent biogenesis pathway, and negatively regulate gene expression by guiding the cleavage of target mRNAs. nat-siRNAs originate from pairs of natural *cis*-antisense transcripts and guide the cleavage of the constitutive transcripts. ra-siRNAs originate from genomic repetitive sequences and are associated with DNA remodeling.

In this study, we report the identification and cloning of the major locus, *p/tms12-1*, that confers TGMS in PA64S and PGMS in NK58S. We found that the locus encodes a noncoding RNA precursor that produces a 21-nt small RNA. A point mutation is present in the small RNA of *p/tms12-1*. Transformation of NK58S and PA64S plants with the allele *P/TMS12-1* produced the wild-type small RNA and suppressed both PGMS and TGMS under long-day and high-temperature conditions. Our results reveal that this point mutation in *p/tms12-1* confers the PGMS and TGMS traits. We discuss possible mechanisms by which this mutation in the small RNA might cause the PGMS and TGMS traits in the *japonica* and *indica* lines, respectively.

Results

Identification of a major locus for TGMS in PA64S

We observed that PA64S was completely male sterile when grown under sterility-inducing high temperatures (25-30 °C), but became male fertile when it was subjected to lower temperatures (21-23 °C) during the microspore mother cell (MMC) to the meiosis stages [50], corresponding to the anther developmental stages 6-8a defined recently by Zhang *et al.* [51] (Figure 2A-2D). This feature of sterility-fertility conversion is similar to previous observations of this line [13, 14, 52]. To estimate the number of loci that control the TGMS trait in PA64S, an F₂ population was produced from a cross between PA64S and Peiai 64 (PA64) (the recurrent backcrossing parent used for breeding of PA64S). Genetic analysis of this F₂ population grown under the sterility-inducing temperatures (higher than 25 °C) showed that its segregation into 880 fertile plants and 105 sterile plants deviated from the ratio of 3:1 expected for a single locus. This suggested that TGMS in PA64S is determined by one major recessive locus and might be affected by other minor gene(s).

We then performed linkage analysis with these 105 sterile plants, using 256 simple sequence repeat markers that were distributed across the 12 chromosomes. One marker (RM2992) on chromosome 12 was found to be linked to the TGMS phenotype. Use of additional markers around RM2992 for primary mapping enabled us to define a 937-kilobase (kb) region on chromosome 12, which was flanked by two markers, PA107 and PA243 (Figure 1A). We temperately designated the recessive locus *tms12-1(t)* (thermo-sensitive genic male sterility locus on chromosome 12) in PA64S.

Fine mapping of the TGMS locus and detection of a point mutation within the allele

To fine-map *tms12-1(t)*, 690 completely sterile plants were selected from the F₂ population and more molecular markers (Supplementary information, Table S1) were developed from the initially mapped region. Finally, *tms12-1(t)* was mapped within a region of 5.8 kb flanked by the markers PA301 and PAIDL2. This region is carried in the BAC clone AL731757 (Figure 1A and 1B). A single nu-

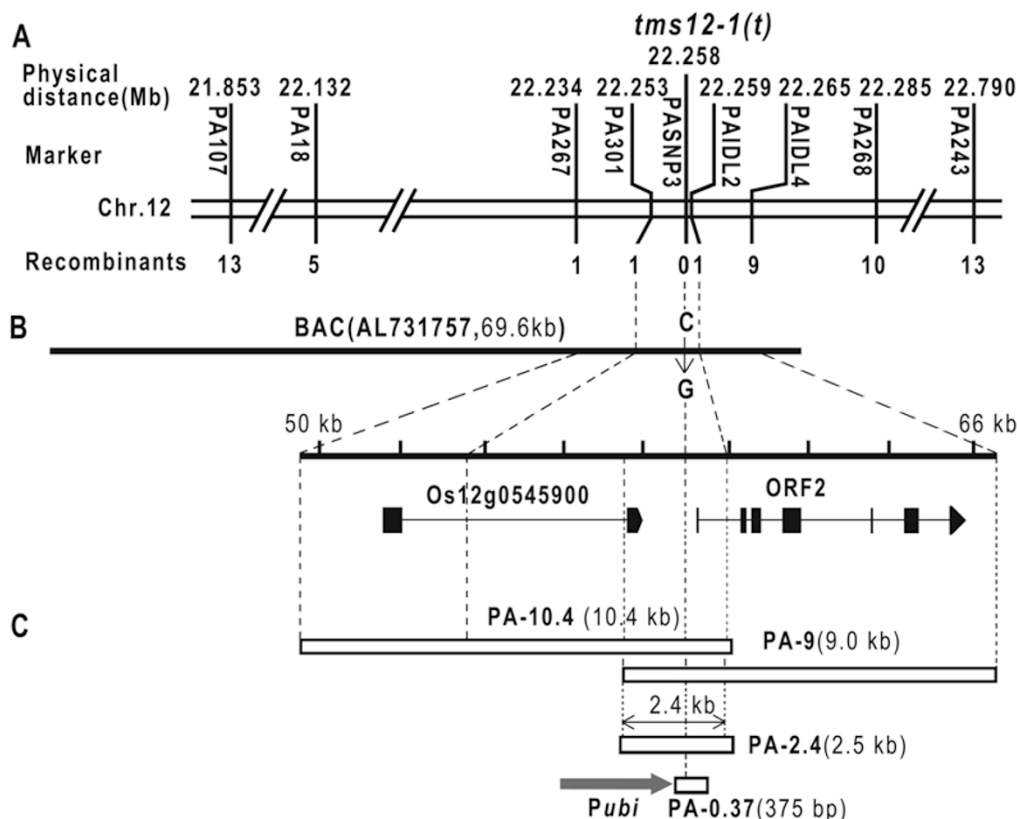


Figure 1 Molecular mapping of *tms12-1(t)*. **(A)** Fine mapping of *tms12-1(t)* to a 5.8-kb region on chromosome 12. **(B)** Predicted gene/ORF composition of the mapped region in a BAC sequence, in which the only mutation within the mapped regions is indicated. **(C)** Fragments used for genetic complementation tests, of which PA-0.37 was linked to the promoter of the maize ubiquitin gene (*Pubi*).

cleotide polymorphism (SNP) marker, PASNP3, cosegregated with *tms12-1(t)* (Figure 1A). This 5.8-kb sequence contained partial sequences of a gene Os12g0545900 (AK111270) of unknown function and of a predicted open reading frame (ORF2) (Figure 1B).

We performed sequence analysis of this mapped region in four sterile lines, NK58S, 7001S (a NK58S-derived *japonica* PGMS line), PA64S and GD-8S (another NK58S-derived *indica* TGMS line), as well as three normal fertile lines, PA64, NK58 and Zhonghua11 (a *japonica* fertile line). Only one SNP of the marker PASNP3 was identified, with a C residue in all the fertile lines and a G residue in all the PGMS and TGMS lines. The genomes of an *indica* cultivar, 93-11 (fertile line) and an African cultivated rice (*O. glaberrima* Steud), searched on the GenBank, also possess the C residue in this position. The G in the PGMS and TGMS lines was therefore derived from a point mutation in the recessive allele. A BLAST similarity search showed that the mapped sequence is unique in the rice genome.

A 2.4-kb sequence of the wild-type allele restores male fertility to NK58S and PA64S

As the sequence of *tms12-1* in PA64S was identical to that of NK58S, we speculated that TGMS in PA64S and PGMS in NK58S are caused by this SNP. To test this hypothesis, we carried out genetic complementation tests. A 10.4-kb fragment (designated PA-10.4) and

a 9-kb fragment (PA-9), which contained the mapped region and overlapped each other by 2.4-kb including the SNP site (Figure 1C), were amplified from PA64. The fragments were cloned into a plant transformation binary vector, pCAMBIA1380, and the constructs were used to transform NK58S. Overall, 52 and 38 independent T_0 transgenic plants were obtained with PA-10.4 and PA-9, respectively. We observed that the pollen fertility of the transgenic plants (T_0 , T_1 and T_2) was restored under long-day (14 h) conditions by both PA-10.4 and PA-9 (Figure 2E and 2F), suggesting that the locus is located within this 2.4-kb region encompassing the SNP. We then prepared a construct that contained this 2.4-kb fragment (PA-2.4) and transformed it into PA64S and NK58S. Overall, 10 and 29 independent T_0 transgenic plants were obtained from PA64S and NK58S, respectively. NK58S- and PA64S-transgenic plants (T_0 and T_1) carrying this fragment became male fertile under long-day (14 h) and high-temperature (25-30 °C) conditions (Figure 2G and 2H). The results demonstrate that this 2.4-kb sequence from the wild-type allele *TMS12-1(t)* is able to suppress PGMS and TGMS, while the C-to-G mutation in this locus confers PGMS in the *japonica* line NK58S and TGMS in the *indica* line PA64S. Therefore, we renamed this locus *p/tms12-1* (photo- or thermo-sensitive genic male sterility locus on chromosome 12) and *P/TMS12-1* for the mutant and wild-type alleles, respectively.

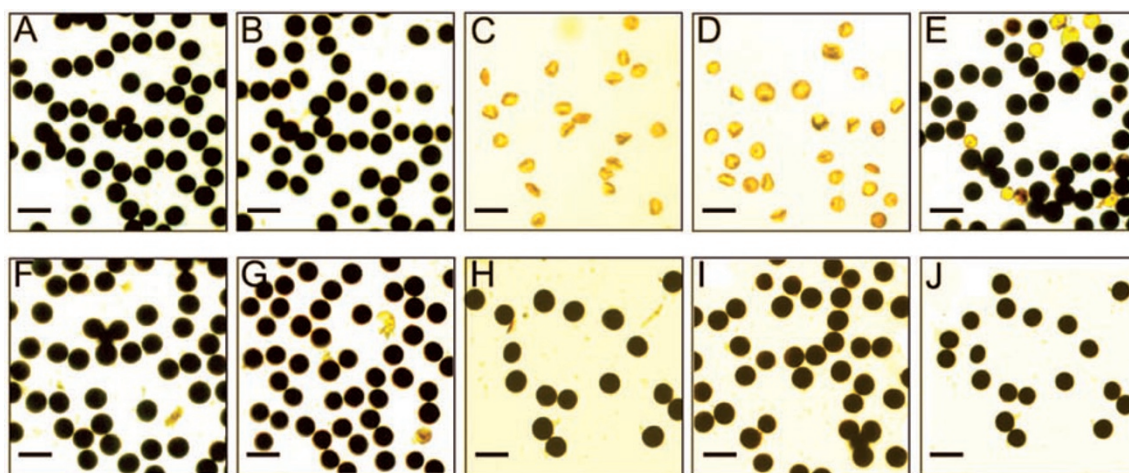


Figure 2 Pollen fertility of transgenic NK58S and PA64S plants is restored by *TMS12-1(t)*. (A, B) Fertile pollen of NK58S and PA64S plants grown under short-day (12 h) and low-temperature (22-23 °C) conditions during the MMC-meiosis stages. (C, D) Sterile pollen of NK58S and PA64S plants grown under sterility-inducing conditions (14-h day length, 25-30 °C). (E, F) Fertility-restored pollen of PA-10.4- and PA-9-transgenic T_1 NK58S plants, respectively. (G, H) Fertility-restored pollen of PA-2.4-transgenic T_1 plants of NK58S and PA64S, respectively. (I, J) Fertility-restored pollen of transgenic T_1 NK58S (I) and T_1 PA64S (J) plants overexpressing a 375-bp sequence (PA-0.37). All transgenic plants were grown under sterility-inducing conditions (14-h day length, 25-30 °C). Scale bars: 50 μ m.

Identification of a small RNA in *P/TMS12-1* and *p/tms12-1*

Since this 2.4-kb region excluded Os12g0545900 and ORF2 as candidate genes, we speculated that this locus might encode small RNA(s). By whole-genome small RNA sequencing, we detected a small RNA of 21 nt (CAUUGUUUGUCUACCAUCCAU; 21 reads in a total of 27 274 860 reads) from *P/TMS12-1* of PA64 and one (CAUUGUUUGUGUACCAUCCAU; 10 reads in a total of 27 957 497 reads) from *p/tms12-1* of PA64S. The C-to-G SNP between *P/TMS12-1* and *p/tms12-1* (underlined above) is present at position 11 of these small RNAs. The presence of the same small RNAs in NK58 and NK58S was confirmed by a specially designed 3' and 5' RACE assays (Figure 3A). Later we searched the small RNA database CSRDB (<http://sundarlab.ucdavis.edu/smrnas>) and also found that this small RNA (osa-smRNA615912) expressed in wild-type rice.

Overexpression of *P/TMS12-1* generates the same small RNA and restores male fertility

To test if the overexpression of a short sequence from *P/TMS12-1* could generate the small RNA and whether it was functional in suppressing the PGMS and TGMS phenotypes, we cloned a 375-bp fragment of *P/TMS12-1* (Figure 3C) containing the SNP site into a binary vector, in which the fragment was located downstream of the maize *ubiquitin* promoter (Figure 1C). This construct was transformed into NK58S and PA64S. In total, 78 and 31 NK58S- and PA64S-transgenic plants were obtained, respectively. Similar to the results of complementation tests with the native sequence constructs, the pollen fertility of NK58S- and PA64S-transgenic plants harboring this recombinant construct was also restored under long-day and high-temperature conditions (Figure 2I and 2J). Furthermore, 3' and 5' RACE assays confirmed that the same wild-type small RNA was produced (Figure 3A). Since the transgenic wild-type small RNA was expressed at much higher levels than those of endogenous one in the lines (Figure 5C), all the obtained overexpressing transgenic plants were restored to fully male fertile.

Determination of the precursor sequence of the small RNA

To determine the precursor RNA sequence that produced the small RNA in the overexpressing transgenic plants, we performed 5' and 3' RACE assays. Sequencing the 5' and 3' RACE products (Figure 3B) identified a 136-nt RNA containing the small RNA sequence (Figure 3C). The 3' RACE product could be specifically amplified from the RNA sample polyadenylated by poly(A) polymerase, but not from that without the *in vitro* 3' polyadenylation reaction (Supplementary information,

Figure S1), suggesting that this 136-nt RNA was an *in vivo*-processed intermediate precursor that lacked the 3' poly(A) end. These results indicate that this locus is a noncoding RNA gene that yields the small RNA. We named the wild-type and mutant small RNAs osa-smR5864w and osa-smR5864m, respectively.

To assess whether the small RNAs might have been generated from miRNA precursor structures, we performed a prediction of the secondary structures of the RNA precursors of *P/TMS12-1* and *p/tms12-1* using the CONTRA fold tool (<http://contra.stanford.edu/contrafold/server.html>). The sequences were predicted to adopt conformations (Figure 4A and 4B) that are markedly different from those of typical pre-miRNAs [53].

The small RNA is expressed preferentially in young panicles with little effect of day length and temperature

The expression patterns of the small RNAs were analyzed using a small RNA qRT-PCR (quantitative real-time PCR) detection kit (GeneCopoeia). Two osa-smR5864w- and osa-smR5864m-specific primers (Supplementary information, Table S1) were used for qRT-PCR. The expression level of endogenous osa-smR5864w in young panicles (at the MMC-meiosis stages) was at least 10-fold higher than those in other vegetative tissues under 27-30 °C/14-h day length conditions (Figure 5A). Differential expression between osa-smR5864w and osa-smR5864m in young panicles was observed under 27-30 °C/14-h day length conditions: the level in PA64 was 3.1-fold higher than that in PA64S, while the level in NK58 was about 4.8-fold higher than that in NK58S (Figure 5B). High-level production of osa-smR5864w in the overexpressing transgenic plants was detected (Figure 5C): of the analyzed samples, the expression levels were 366- to 4 086-fold higher than those of osa-smR5864m in NK58S and PA64S.

To understand whether the expression of the small RNAs is regulated by photoperiod and/or temperature, we then examined the expression levels in young panicles of NK58, NK58S, PA64 and PA64S under different day length and temperature conditions. The results showed that, although the expression levels of the small RNAs in the same lines were different between long-day (14 h) and short-day (12 h) conditions or between high (27-30 °C) and low (21-23 °C) temperatures (Figure 5D), the differences were only 1.3- to 1.43-fold. The results suggest that expression of the small RNAs in the plants is not greatly affected by photoperiod or temperature conditions.

Candidate targets of osa-smR5864w

To identify candidate gene(s) targeted by the small

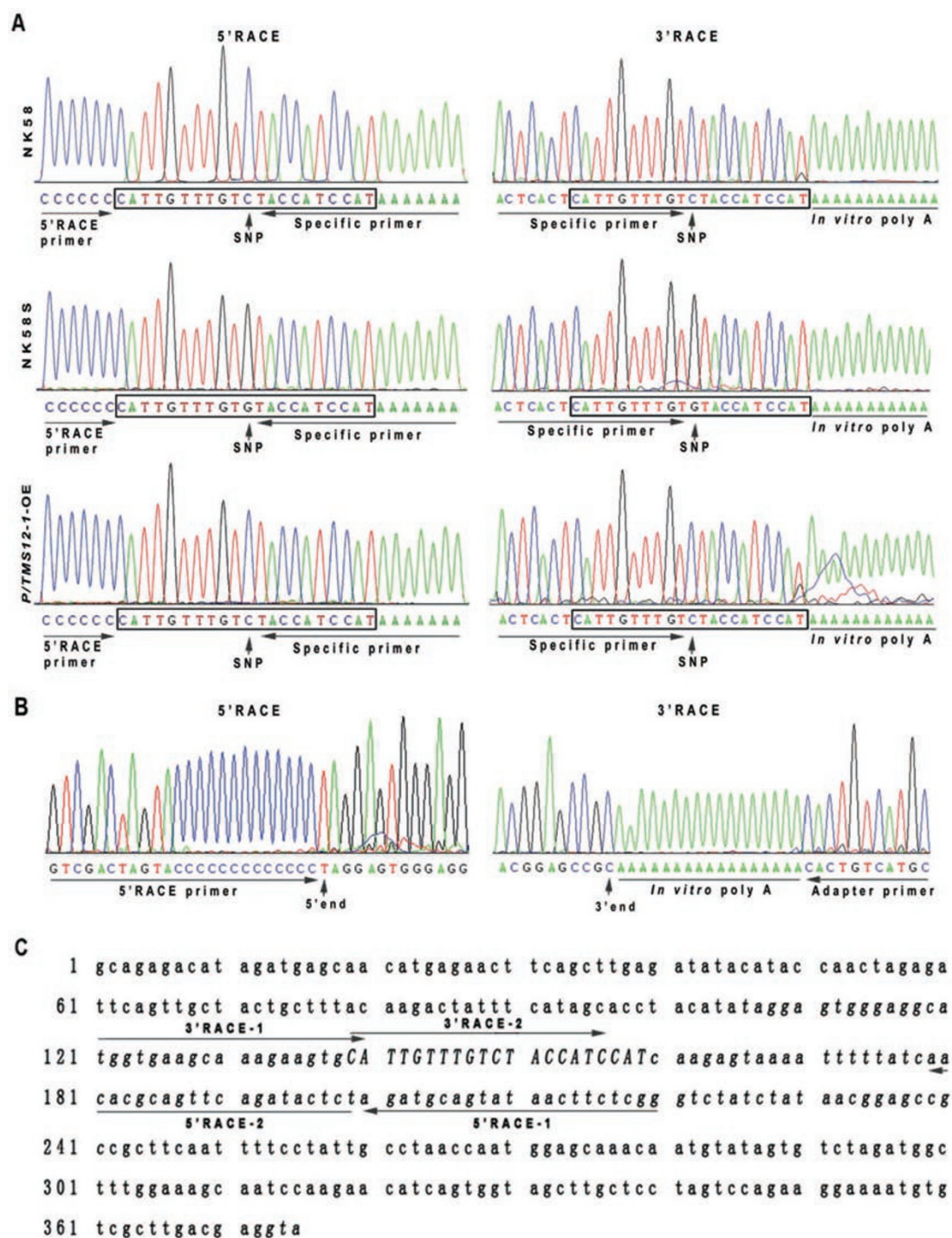


Figure 3 Identification of the small RNA and its precursor. **(A)** Confirmation of the small RNAs in NK58, NK58S, and overexpressing (*P/TMS12-1-OE*) plants by 3' RACE and 5' RACE assays and sequencing, by which the 5' and 3' termini of the small RNAs (shown by rectangles) were determined. The RNA samples were treated with poly(A) polymerase to polyadenylate the 3' end of small RNAs for reverse transcription with a poly dT-adapter primer. The 5' part (5'-ACTCACT-3') of the specific primer is an adapted arbitrary sequence. **(B)** Identification of the 5' and 3' termini of the precursor RNA in an overexpressing plant (P-2 shown in Figure 5C) by 5' and 3' RACE assays and sequencing. **(C)** The 375-bp sequence from *P/TMS12-1* used for overexpression in NK58S and PA64S, in which the transcribed sequence generating the processed intermediate RNA precursor (positions 106-241) is shown in italic and that generating the small RNA is in capital letters. Specific primers used for semi-nested PCR of the 5' or 3' RACE product are shown.

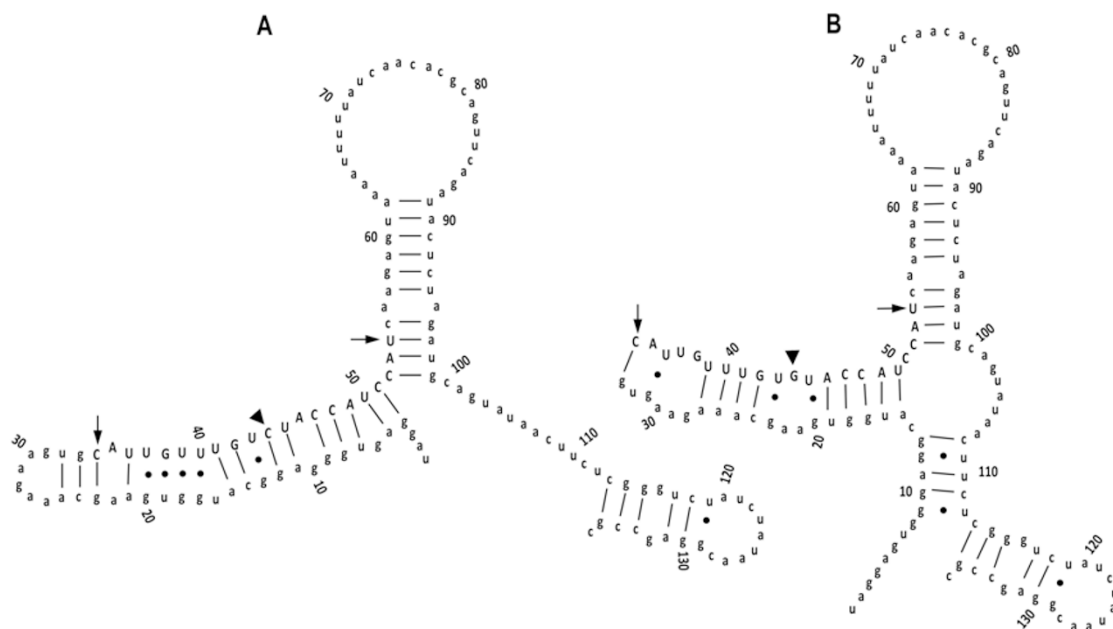


Figure 4 Prediction of the secondary structures of the small RNA precursors. **(A)** The secondary structure of the precursor from *P/TMS12-1*. **(B)** The secondary structure of the precursor from *p/tms12-1*. The small RNAs are located at positions 34–54 indicated by arrows and the SNP sites are shown by arrowheads. In secondary RNA structures, U residue can pair with G residue (shown with dots).

RNA, we used CSRDB to search the rice genome and found 10 candidate target genes of *osa-smR5864w* (Table 1). The functions of these candidate genes are unknown. All of them possess imperfectly paired annealing sites for the small RNA. All but one (*LOC_Os08g05710.1*) have the nucleotide G complementary to the SNP nucleotide C of *osa-smR5864w*. Therefore, the mutation in *osa-smR5864m* reduces the degree of its matching to these nine candidate targets.

A number of anther development-related genes have been cloned in rice, including *Gibberellins MYB* gene (*GAMYB*) [54, 55], *Undeveloped Tapetum1* (*UDT1*) [56], *Tapetum Degeneration Retardation* gene (*TDR*) [57], *CYP704B2* [58], *OsC6* [59], *MADS3* [60], *Carbon Starved Anther* (*CSA*) [61], *APOPTOSIS* (*API5*) [62] and *MEIOSIS ARRESTED AT LEPTOTENE1* (*MEL1*) [63]. These genes were not recognized as the potential targets of *osa-smR5864w* as searched by the CSRDB tool. To understand whether these genes are regulated indirectly by *P/TMS12-1*, we carried out qRT-PCR analysis using specific primers (Supplementary information, Table S1) to investigate their expression profiles. The result showed that the expression levels of these genes were similar between NK58S and NK58 and between PA64S and PA64, although four genes, *GAMYB*, *TDR*, *MADS3* and *API5*, exhibited 1.3- to 1.5-fold difference between some pairs

of the lines (Supplementary information, Figure S2). The result suggests that the regulatory pathway of *P/TMS12-1* does not involve these genes.

Discussion

PGMS and TGMS are very useful germplasm resources for hybrid breeding in crops. In this study, we identified a major locus that controls TGMS in the *indica* rice line PA64S and mapped the locus to a 5.8-kb region on chromosome 12. Sequence analysis revealed a C-to-G mutation in this region of PA64S and its donor *japonica* parent line NK58S. Genetic complementation tests showed that a 2.4-kb sequence of the wild-type allele, *P/TMS12-1*, could restore male fertility of transgenic NK58S and PA64S plants. As the locus *pms3* conferring PGMS in NK58S has been mapped to the region of chromosome 12 [20] in which *p/tms12-1* identified in this study is located, we speculate that *p/tms12-1* and *pms3* may represent the same locus.

We identified a novel 21-nt small RNA from *P/TMS12-1* and *p/tms12-1*, and the only polymorphic site between *P/TMS12-1* and *p/tms12-1* is located in the center of this small RNA sequence. Furthermore, overexpression of a 375-bp sequence of *P/TMS12-1* in transgenic PA64S and NK58S produced the same small RNA and restored pol-

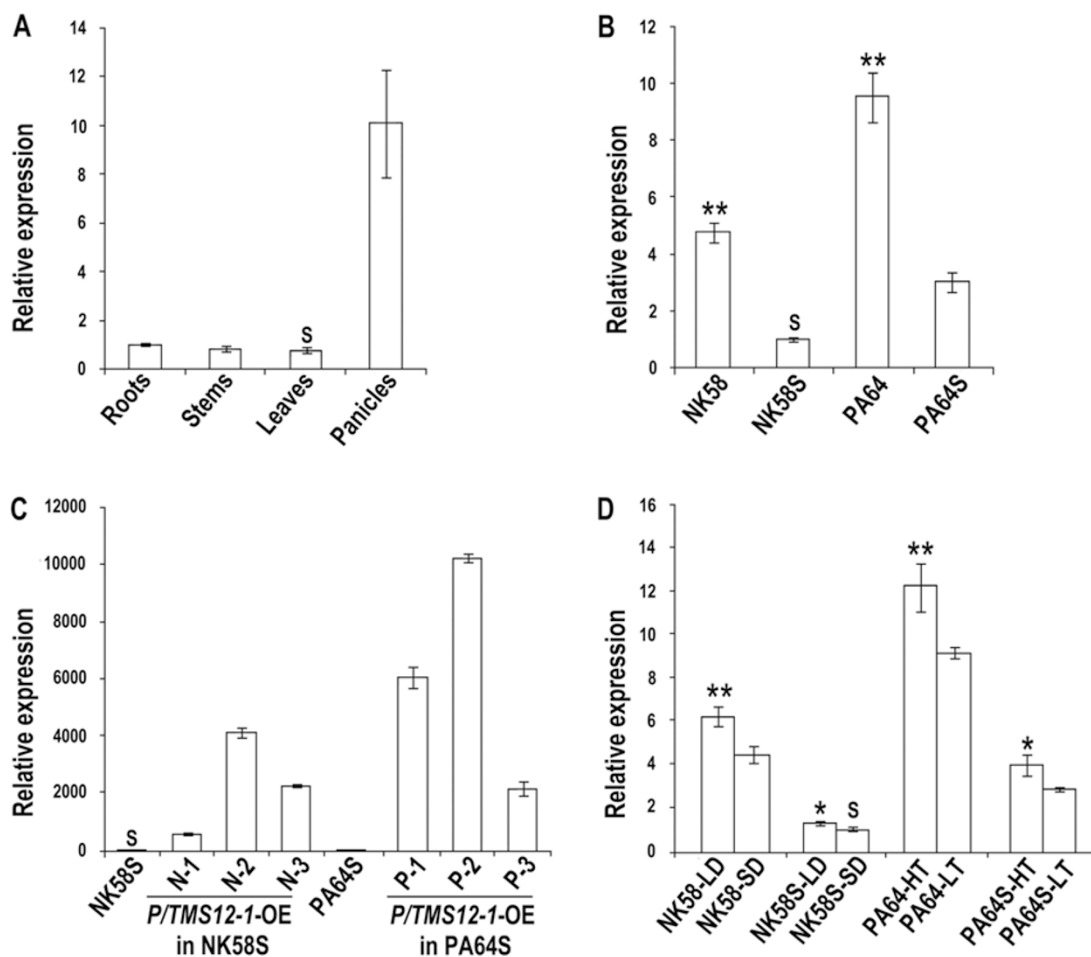


Figure 5 Expression patterns of *osa-smR5864w* and *osa-smR5864m*. **(A)** Expression patterns for *osa-smR5864w* in different organs of PA64. **(B)** Expression levels of *osa-smR5864w* (NK58 and PA64) and *osa-smR5864m* (NK58S and PA64S) in young panicles under sterility-inducing conditions. *t*-tests were performed between NK58 and NK58S and between PA64 and PA64S, with significant level of $P < 0.01$ (**). **(C)** High levels of *osa-smR5864w* were detected in overexpressing transgenic plants (three independent plants were assayed for each of NK58S-OE and PA64S-OE). **(D)** Expression levels of *osa-smR5864w* and *osa-smR5864m* in young panicles of NK58 and NK58S under long-day (14 h, LD) and short-day (12 h, SD) conditions, respectively; and of PA64 and PA64S under high- temperature (27–30 °C, HT) and low-temperature (22–23 °C, LT) conditions (* and ** indicate $P < 0.05$ and $P < 0.01$, respectively). All qRT-PCR assays used the 5S RNA as an internal control for normalization, and the samples marked with S were used as standards (relative value=1) against which expression in the other samples was compared.

len fertility under sterility-inducing conditions. Since the sequence of *P/TMS12-1* is unique in the rice genome and no gene displays appreciable similarity to the precursor RNA sequence except for those that match partially with *osa-smR5864w* (Table 1), it is unlikely that the precursor RNA itself acts directly on target gene(s). Therefore, the small RNA *osa-smR5864w* encoded by *P/TMS12-1* is the most likely player for suppression of the PGMS and TGMS traits; the point mutation in *p/tms12-1* may lead to a loss-of-function of *osa-smR5864m*, and thereby confer PGMS and TGMS in the *japonica* and *indica* genetic

backgrounds, respectively. Our findings suggest that this small RNA gene is an important regulator of genetic networks controlling the development of the male reproductive organ under certain environmental conditions.

Endogenous small RNAs can be generated by the miRNA and siRNA pathways. Most endogenous siRNA families are generated from double-stranded RNAs derived from repeated sequences or transposable elements. As the *P/TMS12-1* sequence is unique in the rice genome, *osa-smR5864w* (or *osa-smR5864m*) does not belong to the repeated sequence-based siRNA family.

Table 1 Target gene candidates and the targeting sites of osa-smR5864w (the SNP site is underlined)

Transcript ID	Alignment	Description
LOC_Os11g19350.1	34-AAGUUGGAUGGGAGACAAACCAUAGUC-60 UACCUACCAUCUGUUUGUUAC	Putative retrotransposon protein, unclassified
LOC_Os01g36870.1	1206-CAUUUGGAUGCUGGGCGAACAAUGCUG-1232 : : : UACCUACCAUCUGUUUGUUAC	Putative phox domain-containing protein
LOC_Os11g40640.1	1695-GUGUAUGGAUGGUGGACAUCGAUGGGG-1721 : : UACCUACCAUCUGUUUGUUAC	Putative transposon protein, mutatorsub-class
LOC_Os07g36544.1	1677-GAAUUGGAUGGAGGGCAGACAGUGGCA-1703 : : : : UACCUACCAUCUGUUUGUUAC	D-mannose binding lectin family protein
LOC_Os12g27680.1	261-CUCAUGGGAGGUAGACGACGUGGAC-287 : : : UACCUACCAUCUGUUUGUUAC	Hypothetical protein
LOC_Os12g17420.1	1186-GAAUUGGGUGGUGGACUAAGCAUGGUG -1212 : : UACCUACCAUCUGUUUGUUAC	Disease resistance protein RPM1 homolog
LOC_Os01g40120.1	609-GCUAUGGAUGGUGGACAAACGCGCGCU-635 : : UACCUACCAUCUGUUUGUUAC	Unknown protein
LOC_Os07g33110.1	1862-AUGGAUGGUGACGACGGAUAAAAGGU-1888 : : UACCUACCAUCUGUUUGUUAC	Putative calcium-dependent protein kinase, isoform 2
LOC_Os06g16790.1	2499-AAGCUGGAUGGUGGAUAAGCAGACCGG-2525 : : : : : UACCUACCAUCUGUUUGUUAC	NB-ARC domain containing protein
LOC_Os08g05710.1	1846-GACAAUGGAAGUGGAACAAGCAAUGAG-1872 : : UACCUACCAUCUGUUUGUUAC	ABC transporter family protein

On the other hand, miRNAs are generated from RNA precursors that form fold-back stem-loop structures. Our analysis showed that the RNA precursors of *P/TMS12-1* and *p/tms12-1* are predicted to form secondary structures (Figure 4A and 4B). However, their conformations are obviously distinct from the pre-miRNA structures of conventional miRNA precursors [53]. Thus, the small RNAs are unlikely to be generated by the miRNA pathway. However, the high-level production of osa-smR5864w in the overexpressing transgenic plants indicates that the small RNA was generated from overexpressed precursor RNA. This result suggests that the small RNAs are not likely to be produced from *cis* sense-antisense genes of the *P/TMS12-1* and *p/tms12-1* alleles by the nat-siRNA pathway [53]. It is possible that the precursor RNAs of

P/TMS12-1 and *p/tms12-1* are converted into double-stranded RNA structures by the RNA-dependent RNA polymerase pathway [64], after which they are processed into mature ta-siRNAs by 21-nt phased small RNA biogenesis pathway as defined in rice [46].

The major function of small RNAs is to repress the expression of target genes [53]. We searched the CSRDB database and found a number of candidates target genes for osa-smR5864w (Table 1), whose functions are unknown. However, all these candidate target genes have imperfectly paired sites to anneal with osa-smR5864w, with a maximum match of 19/21 nt; the point mutation in osa-smR5864m further reduces the degree of matching. Therefore, it is possible that the decreased match of osa-smR5864m to its target(s), together with its reduced

expression level (Figure 5B), results in the loss of its ability for repressing the expression of the target gene(s) and thereby inducing PGMS and TGMS.

Our results showed that *P/TMS12-1* (*osa-smR5864w*) was expressed preferentially in young panicles, which is consistent with its proposed function in the control of male fertility. In addition, the levels of *osa-smR5864w* and *osa-smR5864m* were not substantially affected by different day lengths in NK58S or temperatures in PA64S (Figure 5D). Our findings in this study therefore raise some intriguing questions, such as how the same mutant allele *p/tms12-1* causes PGMS in the *japonica* line (NK58S) but TGMS in the derived *indica* lines represented by PA64S, and how different day lengths or temperatures control the sterility-fertility conversion in these lines of different subspecies. A plausible explanation is that the expression of the same or different target gene(s) of the small RNA may be upregulated by long day in *japonica* lines but by high temperature in *indica* lines, probably due to their different genetic backgrounds. The increased expression of the target gene(s) may impair the anther development. In wild-type rice plants, *osa-smR5864w* may downregulate the expression of the target gene(s) in the *japonica* and *indica* lines under long-day or high-temperature conditions to suppress PGMS and TGMS, whereas the mutant *osa-smR5864m* in the PGMS and TGMS lines fails to suppress the expression of the target gene(s), thus PGMS or TGMS occurs. Further study of the biogenesis of the small RNA and the functions of its target gene(s) will elucidate the mechanistic connection between the actions of the *P/TMS12-1* and *p/tms12-1* alleles and the male sterility-fertility conversion in the PGMS and TGMS lines, and will thus promote a more efficient exploitation of heterosis in rice.

Materials and Methods

Plant materials

The rice lines used in this study were PA64, NK58, NK58S, 7001S (a NK58S-derived *japonica* PGMS line), PA64S and GD-8S (NK58S-derived *indica* TGMS lines) and Zhonghua11 (*japonica* line). An F₂ population was generated by a cross between PA64S and PA64. The F₂ plants were grown in summer of 2006 in Guangzhou for genetic analysis and screening of TMGS individuals. Parents and transgenic plants requiring artificial conditions were reared in phytotrons. Pollen fertility was assayed by staining with potassium iodide (1% I₂-KI).

Molecular cloning of *p/tms12-1*

Primary and fine mapping of *p/tms12-1* were carried out using TGMS plants obtained from the F₂ population. New polymorphic markers (Supplementary information, Table S1) for fine mapping were developed based on sequence differences between *indica* variety 93-11 and *japonica* variety Nipponbare available in NCBI

(<http://www.ncbi.nlm.nih.gov>) and by sequence analysis among the materials described above.

Preparation of binary constructs and rice transformation

To prepare the constructs for complementation tests, a 10.4-kb (PA-10.4) and a 9-kb (PA-9) sequences of *P/TMS12-1* were amplified using primers PA-10.4 F/R and PA-9 F/R (Supplementary information, Table S1) and cloned into the *EcoR* I/*Spe* I sites (for PA-10.4) or *EcoR* I/*Pst* I sites (for PA-9) of the pCAMBIA1380 binary vector. To yield the construct PA-2.4, a 2.5-kb sequence of *P/TGMS12-1* was amplified using primers PA-2.4 F/R (Supplementary information, Table S1) and cloned into the *Spe* I/*Bam*HI sites of pCAMBIA1380. To prepare the overexpression construct of *P/TMS12-1*, a 375-bp fragment (PA-0.37) of *P/TMS12-1* was amplified and inserted into the *Hind* III/*Spe* I sites downstream of the promoter of maize *ubiquitin (Pubi)* in a pCAMBIA-based vector. These constructs were transformed into the *Agrobacterium tumefaciens* strain EHA105 and were used for rice transformation by a previously described method [65].

RT-PCR, RACE and qRT-PCR

Total RNAs from roots, stems, leaves and young panicles (at MMC-meiosis stages) were prepared using TRIzol (Invitrogen, Carlsbad, CA, USA) in accordance with the manufacturer's recommendations. cDNA was synthesized from 1 µg of DNaseI-treated total RNAs with a Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV-RT) kit (Takara Biotechnology, Dalian, China). An aliquot (1 µl) of the cDNA was subjected to PCR amplification using gene-specific primers (Supplementary information, Table S1). For 5' and 3' RACE reactions to determine the termini of the precursor RNA, the cDNAs were synthesized using the All-in-One miRNA qRT-PCR Detection Kit (GeneCopoeia, Rockville, MD, USA), with which 3' polyadenylation and reverse transcription (using a polydT-adaptor primer) proceed in the same reaction. For the control 3' RACE reaction without *in vitro* 3' polyadenylation, the poly(A) polymerase was not added. For 5' RACE, the first strand of the cDNA was tailed at the 3' end by terminal deoxynucleotidyl transferase (Takara Biotechnology) and dGTP. Expression levels of the small RNAs were analyzed by qRT-PCR using the All-in-One miRNA qRT-PCR Detection Kit. The standard curve was prepared from series dilutions of template cDNA. Expression levels of the small RNAs were calculated after normalization to the expression levels of 5S ribosomal RNA as an internal control.

Small RNA sequencing

Whole-genome small RNA sequencing was performed by Beijing Genomics Institute using high-throughput pyrosequencing technology developed by Illumina. RNA samples were quantified and equalized so that equivalent amounts of RNA from mutant and wild-type plants were analyzed. Small RNAs were first separated from total RNAs from young panicles of PA64 and PA64S by size fractionation and recovery from a 15% acrylamide/urea gel. The isolated small RNAs were then ligated to 5' and 3' adaptors using T4 RNA ligase (Promega, Madison, WI, USA). The adaptor-ligated small RNAs were subsequently transcribed into cDNAs by SuperScript II Reverse Transcriptase (Invitrogen). The cDNA samples were loaded onto an Illumina 1G Genome Analyzer for sequencing. The sequencing data was analyzed by Science Cor-

poration of Gene. The target candidates of the small RNA were search using the CSRDB webpage (<http://sundarlab.ucdavis.edu/smrnas>).

Acknowledgments

We thank Dr Ian Smith (Nara Institute of Science and Technology, Japan) for critical reading and editing of the manuscript. This work was supported by the National Basic Research Programs of China (2007CB108705 and 2011CB100203) and Genetically Modified Organisms Breeding Major Projects (2011ZX08001-004).

References

- Cheng SH, Zhuang JY, Fan YY, Du JH, Cao LY. Progress in research and development on hybrid rice: a super-domesticated in China. *Ann Bot* 2007; **100**:959-966.
- Wang Z, Zou Y, Li X, et al. Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell* 2006; **18**:676-687.
- Lin SC, Yuan LP. Hybrid rice breeding in China. In: Innovative approaches to rice breeding. Manila, Philippines: International Rice Research Institute, 1980: 35-52.
- Yuan LP. Purification and production of foundation seed of rice PGMS and TGMS lines. *Hybrid Rice* 1994; **6**:1-3.
- Yang Q, Liang C, Zhuang W, et al. Characterization and identification of the candidate gene of rice thermo-sensitive genic male sterile gene *tms5* by mapping. *Planta* 2007; **225**:321-330.
- Si HM, Liu WZ, Fu YP, Sun ZX, Hu GC. Current situation and suggestions for development of two-line hybrid rice in china. *Chin J Rice Sci* 2011; **25**:544-552.
- Shi MS. The discovery and study of the photosensitive recessive male-sterile rice (*Oryza sativa* L. subsp. *japonica*). *Sci Agric Sin* 1985; **2**:44-48.
- Shi MS, Deng JY. The discovery, determination and utilization of the Hubei photosensitive genic male-sterile rice (*Oryza sativa* subsp. *japonica*). *Acta Agric Sin* 1986; **13**:107-112.
- He HH, Yuan SC. Analyses of plant character in Hubei Photoperiodsensitive Genic Male-sterile Rice (HPGMR) under different light and temperature conditions. *Hybrid Rice* 1989; **5**:42-44.
- Yang SH, Cheng BY, Shen WF, Xia JH. Progress of application and breeding on two-line hybrid rice in China. *Hybrid Rice* 2009; **24**:5-9.
- Luo XH, Xia ST, Luo S, Lu ZC, Zhang ZG, Wang GY. Breeding of dual purpose genic male sterile lines and their application in two-line super rice. *Hybrid Rice* 2010; **S1**:59-63.
- Luo XH, Qiu ZZ, Li RH. Pei'ai 64S, a dual purpose sterile line whose sterility is induced by low critical temperature. *Hybrid Rice* 1992; **1**:27-29.
- Lu CG, Zou JS, Hu N, Yao KM. Plant temperature for sterile alteration of a temperature-sensitive genic male sterile rice, Pei'ai64S. *Sci Agric Sin* 2007; **6**:1283-1290.
- Xu ML, Zhou GQ, Chen LB. Response of fertility of Pei'ai 64S to temperature and photoperiod in rice. *Acta Agric Sin* 1999; **25**:772-776.
- Zhang Q, Shen BZ, Dai XK, Mei MH, Saghai Maroof MA, Li ZB. Using bulked extremes and recessive class to map genes for photoperiod-sensitive genic male sterility in rice. *Proc Natl Acad Sci USA* 1994; **91**:8675-8679.
- Liu N, Shan Y, Wang FP, et al. Identification of an 85-kb DNA fragment containing *pms1*, a locus for photoperiod-sensitive genic male sterility in rice. *Mol Genet Genomics* 2001; **266**:271-275.
- Mei M, Chen L, Zhang Z, Li Z, Xu C, Zhang Q. *Pms3* is the locus causing the original photoperiod-sensitive male sterility mutation of 'Nongken 58S'. *Sci China C Life Sci* 1999; **42**:316-322.
- Mei MH, Dai XK, Xu C, Zhang Q. Mapping and genetic analysis of the genes for photoperiod-sensitive genic male sterility in rice using the original mutant Nongken 58S. *Crop Sci* 1999; **39**:1711-1715.
- Li XH, Lu Q, Wang FL, Xu CG, Zhang QF. Separation of the two-locus inheritance of photoperiod sensitive genic male sterility in rice and precise mapping the *pms3* locus. *Euphytica* 2001; **119**:343-348.
- Lu Q, Li XH, Guo D, Xu CG, Zhang Q. Localization of *pms3*, a gene for photoperiod-sensitive genic male sterility, to a 28.4-kb DNA fragment. *Mol Genet Genomics* 2005; **273**:507-511.
- Peng HF, Zhang ZF, Wu B, et al. Molecular mapping of two reverse photoperiod-sensitive genic male sterility genes (*rpms1* and *rpms2*) in rice (*Oryza sativa* L.). *Theor Appl Genet* 2008; **118**:77-83.
- Wang B, Xu WW, Wang JZ, et al. Tagging and mapping the thermo-sensitive genic male-sterile gene in rice (*Oryza sativa* L.) with molecular markers. *Theor Appl Genet* 1995; **91**:1111-1114.
- Yamagushi Y, Ikeda R, Hirasawa H, Minami M, Ujihara P. Linkage analysis of the thermo-sensitive genic male sterility gene *tms2* in rice (*Oryza sativa* L.). *Breed Sci* 1997; **47**:371-377.
- Lopez MT, Toojinda T, Vanavichit A, Tragoonrun S. Microsatellite markers flanking the *tms2* gene facilitated tropical TGMS rice line development. *Crop Sci* 2003; **43**:2267-2271.
- Subudhi PK, Borkakati RP, Virmani SS, Huang N. Molecular mapping of a thermosensitive genetic male sterility gene in rice using bulked segregant analysis. *Genome* 1997; **40**:188-194.
- Lang NT, Subudhi PK, Virmani SS, et al. Development of PCR-based markers for thermosensitive genetic male sterility gene *tms3(t)* in rice (*Oryza Sativa* L.). *Hereditas* 1999; **131**:121-127.
- Dong NV, Subudhi PK, Luong PN, et al. Molecular mapping of a rice gene conditioning thermosensitive genic male sterility using AFLP, RFLP and SSR techniques. *Theor Appl Genet* 2000; **100**:727-734.
- Wang YG, Xing QH, Deng QY, et al. Fine mapping of the rice thermo-sensitive genic male-sterile gene *tms5*. *Theor Appl Genet* 2003; **107**:917-921.
- Nas TM, Sanchez DL, Diaz GQ, Mendioro MS, Virmani SS. Pyramiding of thermosensitive genetic male sterility (TGMS) genes and identification of a candidate *tms5* gene in rice. *Euphytica* 2005; **145**:67-75.
- Jiang D, Lu S, Zhou H, et al. Mapping of the rice (*Oryza*

- sativa* L.) thermo-sensitive genic male sterile gene *tms5* with EST and SSR markers. *Chin Sci Bull* 2006; **51**:417-420.
- 31 Peng HF, Chen XH, Lu YP, *et al.* Fine mapping of a gene for non-pollen type thermosensitive genic male sterility in rice (*Oryza sativa* L.). *Theor Appl Genet* 2010; **120**:1013-1020.
- 32 Lee DS, Chen LJ, Suh HS. Genetic characterization and fine mapping of a novel thermo-sensitive genic male-sterile gene *tms6* in rice (*Oryza sativa* L.). *Theor Appl Genet* 2005; **111**:1271-1277.
- 33 Liu X, Li X, Zhang X, Wang S. Genetic analysis and mapping of a thermosensitive genic male sterility gene, *tms6(t)*, in rice (*Oryza sativa* L.). *Genome* 2010; **53**:119-124.
- 34 Xu J, Wang B, Wu Y, *et al.* Fine mapping and candidate gene analysis of *ptgms2-1*, the photoperiod-thermo-sensitive genic male sterile gene in rice (*Oryza sativa* L.). *Theor Appl Genet* 2011; **122**:365-372.
- 35 Zhou YF, Zhang XY, Xue QZ. Fine mapping and candidate gene prediction of the photoperiod and thermo-sensitive genic male sterile gene *pms1(t)* in rice. *J Zhejiang Univ Sci B* 2011; **12**:436-447.
- 36 Jia JH, Zhang DS, Li CY, *et al.* Molecular mapping of the reverse thermo-sensitive genic male-sterile gene (*rtms1*) in rice. *Theor Appl Genet* 2001; **103**:607-612.
- 37 Xie Z, Qi X. Diverse small RNA-directed silencing pathways in plants. *Bio Bio Acta* 2008; **1779**:720-724.
- 38 Ramachandran V, Chen X. Small RNA metabolism in *Arabidopsis*. *Tren Plant Sci* 2008; **13**:368-374.
- 39 Chen X. Small RNAs and their roles in plant development. *Annu Rev Cell Dev Biol* 2009; **25**:21-44.
- 40 Ruiz-Ferrer V, Voinnet O. Roles of plant small RNAs in biotic stress responses. *Annu Rev Plant Biol* 2009; **60**:485-510.
- 41 Olivier V. Origin, biogenesis, and activity of plant microRNAs. *Cell* 2009; **136**:669-687.
- 42 Wu L, Zhang Q, Zhou H, Ni F, Wu X, Qi Y. Rice microRNA effector complexes and targets. *Plant Cell* 2009; **21**:3421-3435.
- 43 Ji L, Liu X, Yan J, *et al.* ARGONAUTE10 and ARGONAUTE1 regulate the termination of floral stem cells through two microRNAs in *Arabidopsis*. *PLoS Genet* 2011; **7**:e1001358. doi:10.1371/journal.pgen.1001358
- 44 Chen X. A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science* 2004; **303**:2022-2025.
- 45 Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, *et al.* Widespread translational inhibition by plant miRNAs and siRNAs. *Science* 2008; **320**:1185-1190.
- 46 Song X, Li P, Zhai J, *et al.* Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. *Plant J* 2012; **69**:462-474.
- 47 Liu B, Li P, Li X, *et al.* Loss of function of OsDCL1 affects microRNA accumulation and causes developmental defects in rice. *Plant Physiol* 2005; **139**:296-305
- 48 Liu B, Chen Z, Song X, *et al.* *Oryza sativa* Dicer-like4 reveals a key role for small interfering RNA silencing in plant development. *Plant Cell* 2007; **19**:2705-2018.
- 49 Zhu QH, Spriggs A, Matthew L, *et al.* A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. *Genome Res* 2008; **18**:1456-1465.
- 50 Feng JH, Lu YG, Liu XD, Xu XB. Pollen development and its stages in rice (*Oryza sativa* L.). *Chinese J Rice Sci* 2001; **15**:21-28.
- 51 Zhang D, Luo X, Zhu L. Cytological analysis and genetic control of rice anther development. *J Genet Genomics* 2011; **38**:379-390.
- 52 Yu YH, Si HM, Hu GC, Fu YP, Sun ZX. Analysis on fertility transformation reaction to photo-thermo treatments of Pei-ai64S from different sources. *Sci Agric Sin* 2006; **39**:1064-1068.
- 53 Vaucheret H. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev* 2006; **20**:759-771.
- 54 Liu Z, Bao W, Liang W, Yin J, Zhang D. Identification of *gamyb-4* and analysis of the regulatory role of *GAMYB* in rice anther development. *J Integ Plant Biol* 2010; **52**:670-678.
- 55 Kaneko M, Inukai Y, Ueguchi-Tanaka M, *et al.* Loss-of-function mutations of the rice *GAMYB* gene impair α -Amylase expression in aleurone and flower development. *Plant Cell* 2004; **16**:33-44.
- 56 Jung KH, Han MJ, Lee YS, *et al.* Rice *Undeveloped Tapetum1* is a major regulator of early tapetum development. *Plant Cell* 2005; **17**:2705-2722.
- 57 Li N, Zhang DS, Liu HS, *et al.* The rice tapetum degeneration retardation gene is required for tapetum degradation and anther development. *Plant Cell* 2006; **18**:2999-3014.
- 58 Li H, Pinot F, Sauveplane V, *et al.* Cytochrome P450 family member CYP704B2 catalyzes the ω -hydroxylation of fatty acids and is required for anther cutin biosynthesis and pollen exine formation in rice. *Plant Cell* 2010; **22**:173-190.
- 59 Zhang D, Liang W, Yin C, Zong J, Gu F, Zhang D. OsC6, encoding a lipid transfer protein, is required for postmeiotic anther development in rice. *Plant Physiol* 2010; **154**:149-162.
- 60 Hu L, Liang W, Yin C, *et al.* Rice MADS3 regulates ROS homeostasis during late anther development. *Plant Cell* 2011; **23**:515-533.
- 61 Zhang H, Liang W, Yang X, *et al.* Carbon starved anther encodes a MYB domain protein that regulates sugar partitioning required for rice pollen development. *Plant Cell* 2010; **22**:672-689.
- 62 Li X, Gao X, Wei Y, *et al.* Rice APOPTOSIS INHIBITOR5 coupled with two DEAD-Box adenosine 5'-triphosphate-dependent RNA helicases regulates tapetum degeneration. *Plant Cell* 2011; **23**:1416-1434.
- 63 Nonomuraa K-I, Morohoshia A, Nakanoa M, *et al.* A germ cell-specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. *Plant Cell* 2007; **19**:2583-2594.
- 64 Dalmay T, Hamilton A, Rudd S, Angell S, Baulcombe DC. An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* 2000; **101**:543-553.
- 65 Li J, Jiang D, Zhou H, *et al.* Expression of RNA-interference/antisense transgenes by the cognate promoters of target genes is a better gene-silencing strategy to study gene functions in rice. *PLoS ONE* 2011; **6**:e17444. doi:10.1371/annotation/6818176e-4f00-4229-a133-a83d0c3b11da

(Supplementary information is linked to the online version of the paper on the Cell Research website.)