

华中农业大学

HUAZHONG AGRICULTURAL UNIVERSITY

博士学位论文

Ph D DISSERTATION

水稻光敏感雄性核不育基因 *pms1* 的克隆与功能分析

MAP-BASED CLONING AND FUNCTIONAL ANALYSIS OF
PHOTOPERIOD-SENSITIVE GENIC MALE STERILITY
GENE (*pms1*) IN RICE

研究生: 范优荣

CANDIDATE: FAN YOURONG

导师: 张启发 教授

SUPERVISOR: PROFESSOR ZHANG QIFA

专业: 生物化学与分子生物学

MAJOR: BIOCHEMISTRY AND MOLECULAR
BIOLOGY

研究方向: 水稻功能基因组学

FIELD: RICE FUNCTIONAL GENOMICS

中国 武汉

WUHAN, CHINA

二〇一六年十二月

DECEMBER, 2016

分类号

密级

华中农业大学博士学位论文

水稻光敏感雄性核不育基因 *pms1* 的克隆与功能分析

**Map-based Cloning and Functional Analysis of
Photoperiod-sensitive Genic Male Sterility Gene (*pms1*) in Rice**

博士研究生：范优荣

学 号：060250037

指导教师：张启发 教授

指导教师：张启发 教授

王石平 教授

熊立仲 教授

林拥军 教授

专业：生物化学与分子生物学 研究方向：水稻分子生物学

获得学位名称：理学博士 获得学位时间：2016年12月

华中农业大学生命科学与技术学院

二〇一六年十二月

华中农业大学学位论文独创性声明及使用授权书

学位论文 是否保密	否	如需保密, 解密时间	年 月 日
--------------	---	------------	-------

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知, 除了文中特别加以标注和致谢的地方外, 论文中不包含其他人已经发表或撰写过的研究成果, 也不包含为获得华中农业大学或其他教育机构的学位或证书而使用过的材料, 指导教师对此进行了审定。与我一同工作的同志对本研究所做的任何贡献均已在论文中做了明确的说明, 并表示了谢意。

研究生签名: 范优荣

时间: 2016年 12月 14日

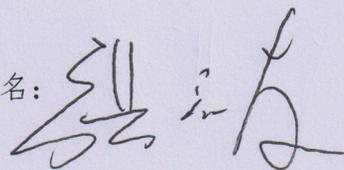
学位论文使用授权书

本人完全了解华中农业大学关于保存、使用学位论文的规定, 即学生必须按照学校要求提交学位论文的印刷本和电子版本; 学校有权保留提交论文的印刷版和电子版, 并提供目录检索和阅览服务, 可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。本人同意华中农业大学可以用不同方式在不同媒体上发表、传播学位论文的全部或部分内容, 为存在馆际合作关系的兄弟高校用户提供文献传递和交换服务, 同时本人保留在其他媒体发表论文的权力。

注: 保密学位论文(即涉及技术秘密、商业秘密或申请专利等潜在需要提交保密的论文)在解密后适用于本授权书。

学位论文作者签名: 范优荣

导师签名:



签名日期: 2016年 12月 14日

签名日期: 2016年 12月 14日

目录

摘 要	i
ABSTRACT	iv
缩略语表	viii
1 前言	1
1.1 研究问题的由来	1
1.2 植物雄性不育研究进展	2
1.2.1 细胞质雄性不育	4
1.2.2 细胞核雄性不育	7
1.2.3 水稻光温敏雄性核不育研究进展	9
1.2.3.1 光温敏雄性核不育水稻的发现及其光温特性	9
1.2.3.2 水稻光敏雄性核不育基因的定位和克隆	11
1.2.3.3 水稻温敏雄性核不育基因的定位和克隆	15
1.3 Long non-coding RNA 研究进展	17
1.3.1 Long non-coding RNA 的定义, 分类及功能	17
1.3.2 植物 lncRNA 研究进展	20
1.4 植物 phasiRNA 研究进展	22
1.4.1 tasiRNA 的发现和形成过程	23
1.4.2 phasiRNA 的形成和特异表达	25
1.5 研究的目的是和意义	28
2 材料和方法	28
2.1 水稻材料及田间种植	28
2.2 定位群体的构建	30
2.3 分子标记的开发	31
2.4 转化载体及菌株	31
2.5 遗传转化载体构建及转化	32
2.5.1 互补载体构建	32
2.5.2 抑制载体构建	36
2.5.3 超表达载体构建	37

2.5.4 Target mimicry 载体构建	38
2.5.5 农杆菌介导的遗传转化	39
2.6 水稻育性考察	39
2.7 水稻总 DNA 抽提	40
2.8 水稻材料基因型检测与转基因植株 Southern blot 拷贝数分析	41
2.9 水稻总 RNA 抽提	42
2.10 RACE 与 5'RLM-RACE	42
2.11 基因表达量分析与 small RNA 表达量分析	44
2.11.1 RT-PCR 与 qPCR	44
2.11.2 Stem-loop RT-PCR 与 page-northern blot	46
2.11.3 RPA (Ribonuclease Protection Assay)	47
2.12 Small RNA-seq 及 PARE 文库构建	47
2.13 DNA 甲基化分析	48
3 结果与分析	49
3.1 前人定位结果的验证	49
3.2 <i>pms1</i> 显隐性关系的确定	52
3.3 <i>pms1</i> 的精细定位	55
3.3.1 分子标记的获得	55
3.3.2 图位克隆 <i>pms1</i>	57
3.4 <i>pms1</i> 候选基因的分离和验证	60
3.4.1 互补转化结果	61
3.4.2 RACE 确定 <i>pms1</i> 转录本	64
3.4.3 抑制 <i>PMSIT</i> 表达的转化结果	65
3.4.4 超量表达 <i>PMSIT</i> 的转化结果	68
3.5 <i>PMSIT</i> 表达谱分析	70
3.6 <i>PMSIT</i> 是一个 long non-coding RNA	75
3.7 确定 <i>pms1</i> 的功能性序列差异	78
3.7.1 比较测序确定 SNP S2 是功能性序列差异	78
3.7.2 65 bp 插入对光敏感雄性不育的影响	79

3.8 <i>PMSIT</i> 是 PHAS 基因	85
3.8.1 <i>PMSIT</i> 是 miR2118 的靶基因.....	85
3.8.1.1 实验验证 <i>PMSIT</i> 被 miR2118 剪切.....	85
3.8.1.2 miR2118 家族表达量分析.....	88
3.8.1.3 miR2118 介导的剪切对光敏感雄性不育的重要性.....	93
3.8.1.4 Target mimicry	94
3.8.2 <i>PMSIT</i> 产生 phasiRNA	97
3.8.2.1 <i>PMSIT</i> 区段产生大量小 RNA.....	97
3.8.2.2 比较分析农垦 58S 和 NIL(MH)中 <i>PMSIT</i> -phasiRNA 的表达.....	100
3.8.2.3 转基因材料中 <i>PMSIT</i> -phasiRNA 分析	102
3.8.2.4 <i>PMSIT</i> -phasiRNA 的验证	104
3.9 SNP S2 与 <i>PMSIT</i> -phasiRNA.....	105
3.9.1 SNP S2 对 miR2118 介导的剪切效率的影响	106
3.9.2 <i>PMSIT</i> -phasiRNA 可能的靶基因	108
3.10 <i>PMSIT</i> 区间 DNA 甲基化分析.....	117
4 讨论	119
4.1 <i>pms1</i> 的作用机理研究.....	119
4.2 未来研究方向	122
4.3 应用前景	124
4.4 Small peptide 与 lncRNA.....	124
4.5 phasiRNA 与雄性不育	125
4.6 <i>pms1</i> 与 <i>pms3</i> 间关系的讨论	126
参考文献	128
附录 I 实验所用引物信息	149
附录 II 在读期间研究成果.....	155
致 谢	156

摘 要

水稻是全球一半以上人口的主粮，如何提高水稻产量一直是各国科学家致力于解决的关乎国计民生的重大课题。50年前袁隆平先生提出了杂交稻的构想，并随后将之应用于生产实践中，为我国粮食安全做出了巨大的贡献。中国杂交稻的生产经历了从三系稻到两系稻的发展，两系稻占我国杂交稻种植面积的比例越来越高。相对于三系不育系，两系不育系恢复系广、配组更为自由，大大简化了杂交过程，应用潜力巨大。两系不育系是指光温敏雄性不育系，其育性受到温度和光照长度的影响：光敏不育系在长日照下不育、短日照下可育；温敏不育系在高温下不育、低温下可育。粳稻品种农垦 58S 是第一个被发现也是研究和应用最广泛的光敏雄性不育系，以农垦 58S 为基因源培育出了大批优良的光温敏雄性不育系。为了更好地利用两系不育系，有必要分离控制光温敏雄性不育性的基因并阐明其分子作用机理。*pms1* 是农垦 58S 中控制光敏雄性核不育的主效基因，我们构建了农垦 58S 与明恢 63 的高世代回交群体以图位克隆 *pms1* 基因并对其进行功能分析，得到的主要研究结果如下：

1. 与以往的认识不同，通过对农垦 58S 与明恢 63 回交 F₂ 群体的结实率和基因型分析表明，光敏感雄性核不育基因 *pms1* 是一个不完全显性基因。在长日照下 BC₅F₂ 群体中纯合的农垦 58S *pms1* 位点表现为完全不育；纯合的明恢 63 *pms1* 位点表现为完全可育；而杂合基因型的育性呈现出从不育到可育连续分布的趋势，并且绝大部分植株育性低于 60%。明恢 63 纯合、杂合及农垦 58S 纯合这三种基因型单株数的比例符合 1:2:1 的单基因分离比。
2. 在确定了 *pms1* 显隐性的基础上对 6841 个 BC₅F₂ 群体单株用两端分子标记 *pj23* 和 *Fssr* 进行基因型分析，共得到 88 个重组单株。再利用新开发的 11 个分子标记进一步缩小定位区间。由于杂合基因型单株与农垦 58S 纯合基因型单株间育性的重叠，为了定位的可靠性，仅利用杂合基因型和明恢 63 纯合基因型发生交换的重组单株进行精细定位。最终将 *pms1* 定位于分子标记 P4 和 P6 之间 4.0 kb 的范围内，左右各 1 个重组单株，分子标记 P5 与 *pms1* 共分离。
3. 将农垦 58S 中包含有该 4.0 kb 的基因序列互补转化到近等基因系 NIL(MH)中，能降低受体长日照下的育性，短日照下无影响；反之，将对应来自于明恢 63 中

的序列转化到农垦 58S 中没有表型的改变，说明该区域内确实包含有 *pms1* 的候选基因。

4. 采用 RACE 的方法在定位区间内分离到一条全长 cDNA，将此转录本命名为 *PMSIT*。在长日照下抑制农垦 58S 中 *PMSIT* 的表达能够恢复育性，NIL(MH)中抑制 *PMSIT* 的表达对育性无影响；在 NIL(MH)中超量表达 *PMSIT* 也能导致育性下降。这些结果都表明 *PMSIT* 就是控制光敏雄性不育的基因 *pms1*。
5. 对 *PMSIT* 表达谱分析发现 *PMSIT* 在与花粉发育相关的组织中表达量较高，在营养生长器官中表达量很低，并且在二次枝梗原基分化期、雌雄蕊形成期和花粉母细胞形成期的幼穗中优势表达，这三个时期也正是光敏感雄性不育的敏感时期。qPCR 结果表明在这三个时期中长日照下农垦 58S 中 *PMSIT* 的表达量低于 NIL(MH)及短日照下的农垦 58S 和 NIL(MH)。
6. *PMSIT* 在农垦 58S 和明恢 63 中的长度分别为 1,453 bp 和 1,388 bp，两者相差 65 bp。在 *PMSIT* 的 5' 端还存在两个 SNP 位点 S1 和 S2，SNP S2 和分子标记 P5 与 *pms1* 共分离。在多个水稻品种中对这几个序列差异进行比较测序表明 SNP S2 是造成光敏雄性不育的功能性突变。
7. *PMSIT* 没有 intron，位于基因间区，没有基因注释。预测有三个短的 ORF，将农垦 58S 这三个预测 ORF 的起始密码子 ATG 后面插入碱基“G”，分别互补转化 NIL(MH)，仍然能够正常发挥功能，降低 NIL(MH)的育性，说明 *PMSIT* 不编码蛋白，是一个长链非编码 RNA。
8. 5' RLM-RACE 和降解组测序数据证实了 *PMSIT* 能够被 miR2118 剪切。对长、短日照下农垦 58S 和 NIL(MH) 雌雄蕊形成期、花粉母细胞形成期和减数分裂期幼穗进行 Small RNA-seq 分析发现 *PMSIT* 能够从 miR2118 剪切位点开始形成 18 对 21 nt phasiRNA。这些 *PMSIT*-phasiRNA 的表达量在花粉母细胞形成期和减数分裂期幼穗中高于雌雄蕊形成期，其中 4s phasiRNA 和 6s phasiRNA 的量远高于其他 phasiRNA。比较分析花粉母细胞形成期幼穗中的 *PMSIT*-phasiRNA，在长日照下的农垦 58S 中表达量高于其他三种材料。同时发现 *PMSIT*-phasiRNA 的表达量与 *PMSIT* 转录本的表达量呈负相关。此外还分析了长日照花粉母细胞形成期幼穗中 *PMSIT*-phasiRNA 在转基因植株中的表达量。在超量表达农垦 58S *PMSIT* 转基因植株中，phasiRNA 的表达量在阳性植株中高于阴性植株；与此相

吻合的是, 农垦 58S *PMSIT* 抑制表达的转基因阳性单株中 *phasiRNA* 的表达量低于阴性单株。这些结果表明 *PMSIT*-*phasiRNA* 与光敏雄性不育相关。转基因结果进一步证实失去了完整 *miR2118* 识别位点的农垦 58S *PMSIT* 不能发挥正常功能来降低 NIL(MH)长日照下的育性, 表明 *phasiRNA* 参与了光敏雄性不育的调控。*PMSIT* 是到目前为止鉴定到的第一个具有生物学功能的 *PHAS* 基因, 证明了这类小 RNA 对植物生长发育的重要性, 同时也加深了我们对于光敏感雄性核不育机理的理解。

关键词: 光敏感雄性核不育; *pms1*; 长链非编码 RNA; *miR2118*; *phasiRNA*

ABSTRACT

Rice is the staple food for more than half of the world population, and how to improve the rice yield has been the major strategic issues related to people's livelihood, national economic development and social stability, waiting to be solved by scientists. Fifty years ago, Longping Yuan proposed the theory and ideas of three-line hybrid rice, then applied it to practical production, making enormous contribution to our country food security. The hybrid rice production in China developed from three-line hybrid to two-line hybrid, and planting proportion of the latter in hybrid rice has been growing. Comparing with CMS lines, the fertility of two-line sterility lines could be recovered by more restorer lines with more freedom, which greatly simplifying the hybrid process and having great potential for application. The two-line sterility lines are classified into photoperiod-sensitive genic male sterility (PSMS) lines and thermo-sensitive genic male sterility (TGMS) lines, whose fertility are influenced by temperature and day-length. PSMS lines are male-sterile under long-day conditions and convert to male-fertile under short-day conditions, while TGMS are male-sterile under high temperatures and male-fertile under low temperatures. Nongken 58S (*japonica*) is the first found and also the most extensive researched and widely used PSMS line, whose gene source is transferred to other cultivars to breed lots of excellent two-line sterility lines. It is necessary to isolate the PSMS and TGMS genes and elucidate their molecular mechanism for better utilization. *pms1* is the major locus controlling PSMS in Nongken 58S, and the advanced backcross population between Nongken 58S (58S) and Minghui 63 (MH63) was constructed to map the *pms1* gene and be further investigate the gene function. The major results are as follows:

1. Combined analysis of Genotyping and the spikelet fertility of the advanced backcross F₂ population indicated that *pms1* was semi-dominant rather than completely recessive as previously assumed. Plants homozygous for the 58S allele were highly sterile and plants homozygous for the MH63 allele were fertile under long-day conditions; spikelet fertility of the heterozygotes varied from completely sterile to

- fully fertile, most of them were less than 60%. The 1:2:1 ratio of homozygous for the 58S allele to heterozygotes to homozygous for the MH63 allele was in accordance with Mendelian separation ratio of single gene.
2. For fine mapping *pms1*, genotyping 6841 individuals from the BC₃F₂ mapping population derived from a cross between 58S and NIL(MH) identified 88 recombinants between two flanking markers Fssr and pj23, covering a 138 kb genomic region in MH63. More molecular markers were developed surrounding Fssr and Rssr (now P5) based on the previous work. Because of extensive phenotypic overlap between heterozygotes and 58S homozygotes, only the recombinants between MH63 homozygotes and heterozygotes were used for further mapping. Genotyping the recombinants resolved the *pms1* locus to a 4.0 kb region flanked by markers P4 and P6 with one recombinant at each side, with the SSR marker P5 co-segregating with *pms1*.
 3. The genomic sequence covering this 4.0 kb from 58S was complementary transformed into the near isogenic line NIL(MH) and could reduce the spikelet fertility under long-day conditions, while no effect under short-day conditions. On the contrast, the corresponding fragment from MH63 could not change the spikelet fertility of 58S. These results confirmed that the transformed genomic fragment indeed contained the *pms1* locus.
 4. A full length cDNA named *PMSIT* by us was isolated within this mapping region using 5' RACE and 3' RACE. The suppression expression of *PMSIT* in 58S increased the spikelet fertility of transgenic plants under long-day conditions, while the spikelet fertility of transgenic plants with suppression expression of *PMSIT* in NIL(MH) did not change regardless of the day length. Moreover, overexpression of full-length *PMSIT* from 58S driven by the maize ubiquitin promoter in NIL(MH) greatly reduced fertility under long days. The results above confirmed that the *PMSIT* corresponding the *pms1* locus conferred the PSMS in rice.
 5. The expression pattern of *PMSIT* showed that *PMSIT* was expressed at low levels and in a tissue-specific manner. In general its expression levels in both 58S and

NIL(MH) were low in all tissues assayed and especially so in green tissues. It was expressed preferentially in young panicles and florets, with an expression peak at secondary branch primordium differentiation stage, pistil/stamen primordium differentiation stage and pollen mother cell formation stage. This expression pattern corresponded well with previous results that the most critical period for PSMS occurred. Moreover, the transcript abundance was lower in 58S long-day conditions than 58S under short-day conditions and also than NIL(MH) under both long-day and short-day conditions at these three stages.

6. This *PMSIT* transcript from 58S was 1,453 nt that was 65 nt longer than the one from MH63 (1,388 bp). The two SNPs S1 and S2 were both located at the 5' terminus of *PMSIT*, while SNP S2 and marker P5 both cosegregated with *pms1*. We compared sequences of 15 different rice lines at these sites, the sequencing results and recombination event narrowed the potential causal polymorphism to SNP S2.
7. There was no intron in *PMSIT* transcript locating in the intergenic region without annotation. Three small ORFs were predicted from *PMSIT*. The guanine residue was inserted immediately downstream the ATG start codons of putative 58S ORFs to disrupt the reading frames. These three vectors were transformed into NIL(MH) independently, could function properly to reduce the spikelet fertility of NIL(MH) under long-day conditions, demonstrating that *PMSIT* encoded a long non-coding RNA.
8. *PMSIT* was predicted to be targeted by miR2118, and we validated the cleavage site in *PMSIT* in both 58S and NIL(MH) by modified 5' RLM-RACE. The signature sequence of the highest abundance of Parallel Analysis of RNA ends (PARE) reads also corresponded to the cleavage site. Sequencing of small RNA libraries constructed using young panicle tissues from pistil/stamen primordium differentiation stage and pollen mother cell formation stage and meiosis stages of 58S and NIL(MH) grown under long-day and short-day conditions showed that 18 pairs of 21 nt phasiRNA could be generated from the cleavage site triggered by miR2118. Comparative analysis revealed that the quantities of 21-nt *PMSIT*-phasiRNAs were higher in pollen

mother cell formation stage and meiosis stages than in pistil/stamen primordium differentiation stage, and most phasiRNAs were generated from the sense strand corresponding to the 4s, 6s phasiRNAs. The phasiRNA reads in 58S under long days at pollen mother cell formation stage were the highest followed by 58S under short days, while the reads in NIL(MH) were much lower. We noticed a trend of negative relation between the abundance of phasiRNAs and the *PMSIT* transcript. We also quantified small RNAs from pollen mother cell formation stage in the transgenic plants, which showed that the 21-nt phasiRNA reads of *PMSIT* were much higher in the 58S *PMSIT* overexpression transgenic-positive plants than in negative ones under long-day conditions. Conversely, their levels were lower in 58S *PMSIT* suppression expression transgenic positive than in negative plants. These data suggested that *PMSIT*-phasiRNA associated with PSMS. Without the complete target sequence of miR2118, the transformed 58S *PMSIT* fragment did not reduce spikelet fertility of NIL(MH), indicating the important roles of phasiRNA for causing male sterility by *pms1*. *PMSIT* was the first identified *PHAS* gene with biological function, implying the importance of this kind of small RNA for plant growth and development, and also deepening our understanding about the mechanism of PSMS in rice.

Keywords: photoperiod-sensitive genetic male sterility; *pms1*; long non-coding RNA; miR2118; phasiRNA

缩略语表

缩略词	英文全称	中文全称
Abbreviation	English full name	Chinese meaning
CMS	Cytoplasmic Male Sterility	细胞质雄性不育
GUS	β -Glucuronidase	葡萄糖苷酸酶
LncRNA	Long non-coding RNA	长链非编码 RNA
InDel	Insertion/Deletion	插入/缺失
NIL	Near Isogenic Line	近等基因系
ORF	Open Reading Frame	开放阅读框
PAGE	Polyacrylamide Gel Electrophoresis	聚丙烯酰胺凝胶电泳
PARE	Parallel Analysis of RNA ends	RNA 末端并行分析
PCR	Polymerase chain reaction	聚合酶链式反应
phasiRNA	phased small interfering RNA	相位小干扰 RNA
PSMS	Photoperiod-sensitive genic Male Sterility	光敏感雄性核不育
qPCR	quantitative Polymerase Chain Reaction	定量 PCR
RACE	Rapid Amplification of cDNA Ends	快速扩增 cDNA 末端
RNAi	RNA interference	RNA 干涉
RT-PCR	Reverse Transcription PCR	反转录 PCR
SNP	Single Nucleotide Polymorphism	单核苷酸多态性
SSR	Simple Sequence Repeat	简单序列重复
tasiRNA	<i>trans</i> -acting small interfering RNAs	反式作用小干扰 RNA

1 前言

1.1 研究问题的由来

“仓廩实而知礼节，衣食足而知荣辱”，粮食问题永远是一个社会和国家首先需要考虑的至关重要的问题，事关国家和民族生存大计。只有粮食得到了充足的保障，社会才能得到繁荣的发展，人民才能安居乐业。但在人类几千年的文明发展史中，饥饿始终如影随形，在 societal 科技发展达到高度文明的今天，依然无法消弭。我国人口基数大，在已有人均耕地面积还不到世界平均水平一半的情况下，加大的城市化进程、环境的恶化、农业结构调整等原因导致土地可耕种面积逐年减少，多个县市低于联合国确定的人均耕地 0.8 亩的警戒线（<http://www.fao.org>;
<http://www.mlr.gov.cn/>）。我国面临的粮食安全问题形势严峻，如何提高粮食产量，是全社会人民需要共同承担的长期而艰巨的任务。

水稻 (*Oryza sativa* L.) 是包括我国在内的全球近一半以上人口的主粮，近年来随着生活水平的提高，居民饮食结构有了极大的调整和改善，依然改变不了其作为主粮的地位。水稻雌雄同花，属于自花授粉作物，以自交结实为主。千百年来水稻的种植都是自交收种，依靠自然突变和人工选择获得对当地自然和栽培环境具有良好适宜性的优良水稻品种，以提高水稻单株产量。发生在上世纪 50 年代初的第一次绿色革命就是利用矮秆基因培育出矮秆、耐肥、抗倒伏的高产水稻，使得世界粮食总产量有了大幅提升(Khush 2001)。但同时水稻也能接受种内其他单株的花粉，获得的杂交种存在不同程度的杂种优势，Jones (1926)首次提出水稻中具有杂种优势。育种学家们最关心的是杂种优势在产量上的体现和利用，而水稻自花授粉的特性及其特殊的花结构——花药包被于颖壳中，去雄困难，导致在生产中无法进行大规模的杂交以获得杂交种，进而达到增产的目的。50 年前袁隆平提出了通过选育雄性不育系、保持系和恢复系的三系配套法充分利用水稻杂种优势提高水稻产量的设想(袁隆平 1966)，并在随后几年将之应用于生产中，使得水稻产量在绿色革命的基础上进一步得到了大幅提升。1973 年石明松发现的光敏感雄性核不育株农垦 58S 拥有长日照不育、短日照下可育的特性(石明松 1985)，在不同的光温条件下能分作不育系和保持系，将三系法杂交育种简化为两系法。相对于三系配套，两系不育系配组自由、

恢复系广、杂交制种过程简化, 在杂交稻生产上占据了越来越重要的地位(斯华敏等 2011)。

相较于两系不育系在杂交稻生产中的广泛应用, 对控制其不育性的光温敏雄性不育基因的研究进展缓慢。得益于水稻基因组测序工作的完成(Goff et al 2002; Yu et al 2002)、水稻基因芯片的开发(Schena et al 1995)、大型突变体库的构建(Jeon et al 2000; Wu et al 2003; Piffanelli et al 2007)以及第二代测序技术的发展(Shendure and Ji 2008; Wang et al 2009), 使得光温敏雄性不育基因的研究取得了突破性进展, 几个基因相继被克隆, 并对其功能进行了解析, 他们的作用方式各不相同。目前生产中广泛使用的两系不育系中除了这些基因, 还存在更多未被克隆的控制光温敏雄性不育性状的基因, 对这些基因的分离和功能研究有助于加深对光温敏雄性不育调控机制的理解, 使之更好地应用于生产中, 对于指导选育和创造优良、新型、稳定的两系不育系具有重要的价值和意义。

1.2 植物雄性不育研究进展

植物雄性不育是指植物雄性生殖器官异常不能产生具有正常功能的雄配子——花粉, 但雌性生殖器官正常。这种性状分为可遗传与非遗传两类。可遗传的雄性不育由基因控制, 能一代代传递保存下去; 而非遗传的雄性不育主要由各种环境胁迫或者化学试剂诱导造成, 比如高温、冷害、干旱、化学杀雄剂处理等, 在这些不利于植物生长的环境下植物雄性生殖器官发育异常, 造成雄性不育, 但这种不育特性在后代中并不能遗传和保持, 只要恢复到正常生长条件下产生雄配子即变为可育。

植物雄性不育现象在高等植物中普遍存在, 早在 1763 年 Kölreuter 就观察到了在某些植物物种和种间杂交中存在花粉败育的现象, 后来一些学者陆续在其他植物中也发现了雄性不育。1988 年出版的《高等植物雄性不育》专著中总结出开花植物 43 个科、162 个属、320 个种中存在着雄性不育现象(Kaul 1988), 并且随着雄性不育在大量具有重要经济价值植物中比如玉米、水稻、小麦、高粱等作物及茄科、豆科、十字花科等双子叶植物中的利用, 有更多植物的雄性不育现象被发现和关注。

雄性不育对于植物个体来说并不是一个好的性状, 因为自身基因不能完整地传递下去, 但对整个群体的适应性和进化而言有着积极的推动作用。植株雄性不育就意味着必须接受外来花粉并正常受精才能获得后代, 这种携带有双亲不同遗传信息

的杂交后代在形态和适应性方面相对于其双亲具有极大的优势。从这方面来说雄性不育又不失为一种良好的遗传机制，能够拓宽高等植物的基因库和细胞质基础，增强对环境的遗传适应性，扩大植物的生存范围和生态环境。而且基于植物雄性不育进行杂交获得的杂种优势对于植株产量和实用品质相关的性状具有很大的经济价值。

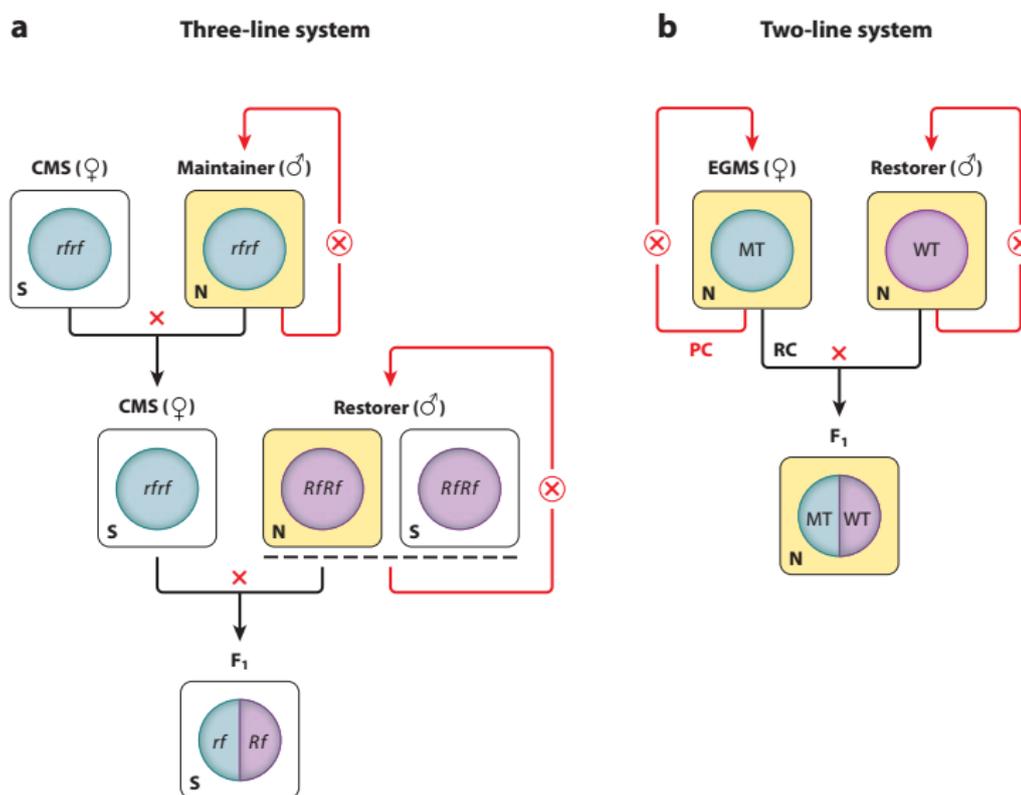


图 1 雄性不育系在杂交育种中的应用(Chen and Liu 2014)。

Figure 1 The application of CMS and GMS for hybrid seed production in a three-line and two-line system (Chen and Liu 2014).

利用杂种优势的前提是能得到不育性状稳定的雄性不育植株。这种可遗传的、稳定的雄性不育根据其控制的基因所在的细胞器和作用方式分为细胞核雄性不育 (Genic male sterility, GMS) 和细胞质雄性不育 (Cytoplasmic male sterility, CMS)。细胞核雄性不育由细胞核内的显性或隐性核不育基因控制，与细胞质遗传没有关系；细胞质雄性不育由细胞质和细胞核基因相互作用控制。以这两类雄性不育为基础在生产中分别发展出两系法和三系法育种以获得具有杂种优势的杂交种子。三系法包括：细胞质雄性不育系、保持系 (Maintainer) 和恢复系 (Restorer)。以细胞质雄性不育系为母本，与恢复系杂交用以生产杂交种子；不育系的繁殖依靠与保持系杂交

维持后代的不育性；恢复系和保持系均能自交繁殖（图 1）。两系法依赖一种环境敏感型的细胞核不育系（environment-sensitive GMS, EGMS），该不育系的育性在一定环境条件下能够恢复，可自交结实，因此不育系与保持系合二为一，将三系减少为两系；EGMS 在不育条件下作为母本，同样可与恢复系杂交产生杂交种（图 1）。

1.2.1 细胞质雄性不育

植物的细胞质雄性不育表现为母性遗传，早期研究认为细胞质雄性不育是由异常的细胞质基因组与隐性的核基因相互作用共同导致的，在正常的细胞质内或者显性的核基因存在的情况下育性得到恢复，CMS 是因为核内育性基因在不利的细胞质环境中不能正常表达，细胞核与细胞质基因组竞争传递自身遗传物质的结果(Jones 1950; Bailey 2002; Budar et al 2003)。细胞质内含有复制信息的有线粒体 DNA 和叶绿体 DNA，植物的叶绿体基因组较小，仅 120-160 kb，并且相对稳定，在进化上比较保守(Palmer 1985)。到目前为止还没有明确的证据表明叶绿体基因组与细胞质雄性不育相关(Frankel et al 1979; Liu and Li 1983; Chen et al 1993)。而线粒体基因组具有不均一性，存在分子间和分子内重排现象，这种线粒体 DNA 的重排、插入和缺失造成基因功能的缺失或失活，导致细胞质雄性不育(Schnable and Wise 1998; Hanson and Bentolila 2004; Chase 2007)。已知的植物线粒体基因仅 60 多个(Kubo and Newton 2008)，但蛋白质组分析表明线粒体中存在超过 1000 个的蛋白质，这些蛋白质由核基因组编码，在细胞核中合成，通过线粒体定位信号运输到线粒体中起作用(Kruft et al 2001)。而细胞质雄性不育确实能够被核基因组上的核恢复基因(*Restorer of fertility*, *Rf*) 发挥校正功能而得到恢复。因此，对细胞质雄性不育功能的解析涉及到细胞质雄性不育基因（CMS 基因）和 *Rf* 基因的分离、作用机制及两者之间的相互作用。

自然界存在很多带有 CMS 基因的植株，但在核基因 *Rf* 基因的存在下，这些植株依然是可育的，为了研究 CMS 基因，必须将含有 CMS 细胞质的植株作为母本与携带隐性 *rf* 基因的父本杂交，获得细胞质雄性不育单株用以研究 CMS 基因。通过比较分析 *Rf* 基因和 *rf* 基因存在的 CMS 细胞质中线粒体基因组在转录组和蛋白质组水平上基因序列或者表达量的差异来寻找 CMS 候选基因，为了验证这些基因的真实性，受限于线粒体转化技术的局限，通常在候选基因 5'端融合一段线粒体的定位信号，先转化到细胞核中，再通过定位信号将候选基因运送到线粒体中发挥作用。

到目前为止, 在 13 个作物中共分离到 28 个 CMS 基因: 包括玉米中的 *urf13*、*orf355* 和 *atp6-C*, *urf13* 是第一个被鉴定到的线粒体 CMS 基因(Dewey et al 1987; Dewey et al 1991; Zabala et al 1997); 水稻中的 *orf79*、*orfH79*、*L-orf79*、*orf307*、*WA352*、*orf352* 和 *orf113* (Wang et al 2006; Fujii et al 2007; Itabashi et al 2009; Hu et al 2012; Igarashi et al 2013; Luo et al 2013; Okazaki et al 2013; Wang et al 2013); 高粱中的 *orf107* (Tang et al 1996); 小麦中的 *orf256* (Song and Hedgcoth 1994); 甜菜中的 *preSatp6*、*orf129* 和 *cox2* (Ducos et al 2001; Yamamoto et al 2005; Yamamoto et al 2008); 油菜中的 *orf138*、*orf222*、*orf224*、*orf220*、*orf288* 和 *orf263* (Singh and Brown 1991; Landgren et al 1996; L'Homme et al 1997; Zhang et al 2003; Uyttewaal et al 2008; Jing et al 2012); 大豆中的 *orf239* (Abad et al 1995); 萝卜中的 *orf125* 和 *orf463* (Iwabuchi et al 1999; Park et al 2013); 胡萝卜中的 *orfB* (Nakajima et al 2001); 辣椒中的 *orf456* 和 *orf507* (Kim et al 2007; Gulyas et al 2010); 向日葵中的 *orf522* (Hans Köhler et al 1991)。这些基因大部分起源于线粒体的 *cox1*, *atp8* 和 *atp6* 基因, 由它们的序列插入、缺失或重排而来。在已阐明功能的 CMS 基因中, 大多编码膜蛋白和 ATP8-like membrane protein, 涉及到线粒体电子传递链途径(Chen and Liu 2014)。

每个 CMS 基因都有其对应的核恢复基因 *Rf*, 某些 CMS 基因存在着几个 *Rf* 基因; 也有些植物细胞质雄性不育系中仅找到了 *Rf* 基因, CMS 基因还是未知。以上描述的 CMS 基因对应的 *Rf* 基因位点鉴定到很多个, 其中仅仅只有 8 个被克隆。包括玉米 *urf13* 对应的 *Rf2* (Cui et al 1996; Liu et al 2001a); 水稻 *orf79* 对应的 *Rf1a* 和 *Rf1b* (Komori et al 2004; Wang et al 2006)、*orfH79* 对应的 *Rf5* (Hu et al 2012)、*L-orf79* 对应的 *Rf2* (Itabashi et al 2011)、*orf307* 对应的 *Rf17* (Fujii and Toriyama 2009); 高粱中的 *Rf1* (Klein et al 2005); 油菜 *orf138* 对应的 *Rfo* (Brown et al 2003; Uyttewaal et al 2008); 萝卜 *orf125* 对应的 *Rfk1* (*orf687*) (Koizuka et al 2003); 甜菜 *preSatp6* 对应的 *Rf1* (Matsuhira et al 2012)。此外, 在矮牵牛中还克隆到 *Rf* 基因 *Rf-PPR592* (Bentolila et al 2002)。水稻包台型细胞质雄性不育的恢复基因 *Rf-1* 最开始以为是一个基因 (Komori et al 2004), 后来发现实际是由两个临近的基因 *Rf1a* 和 *Rf1b* 组成(Wang et al 2006), 并且水稻红莲型细胞质雄性不育中的 *Rf5* 与 *Rf1a* 为同一个基因, 都编码 PPR791 蛋白(Hu et al 2012)。油菜中的 *Rfo* 基因是通过以萝卜和拟南芥的基因组共线性克隆得到的, 与萝卜中的 *Rfk1* 也为同一个基因, 都能恢复 Ogura 型 CMS 油菜的

育性(Uyttewaal et al 2008)。高粱中的 *Rf1* 调控的 CMS 基因未知。玉米中的 *Rf2* 是第一个被克隆的细胞质雄性不育恢复基因，编码线粒体乙醛脱氢酶(Liu et al 2001a)。这些 *Rf* 基因大部分都编码 PPR 蛋白：*Rf1a* 和 *Rf1b* 分别含有 18 个和 11 个 PPR motif，*Rf-PPR592*、*Rfo* 的 PPR-B 蛋白和高粱中的 *Rf1* 则分别有 14、17 和 14 个 PPR motif。

PPR (Pentatricopeptide repeat) 蛋白是指在序列中存在多个以 35 个氨基酸为单元进行串联重复排列的蛋白质，这样的 35 个氨基酸单元被称为 PPR motif。PPR 蛋白在植物基因组中大量存在，在动物和酵母中比较少见(Small and Peeters 2000; Lurin et al 2004)。PPR 蛋白与 RNA 编辑有关，PPR 蛋白的 N 端含有线粒体或者叶绿体的靶向序列，使其进入线粒体或者叶绿体中参与转录后 mRNA 的加工，譬如编辑、剪接、剪切、降解和翻译等过程(Kotera et al 2005; Schmitz-Linneweber and Small 2008; Barkan and Small 2014)。包含 PPR motif 的 *Rf* 基因主要是在核内翻译，运送到线粒体后通过剪切或者降解 CMS 基因的转录本来抑制 CMS 基因的表达，从而恢复育性(Wang et al 2006)。这些 *Rf* 基因附近存在成簇分布的 PPR 蛋白(Fujii and Toriyama 2008)，但仅其中一个或者两个能够对 CMS 基因发挥作用，说明这些 PPR 蛋白与靶基因之间存在特异的识别位点。目前 PPR 蛋白的结构已经被解析出来(Yin et al 2013)，证实了 PPR motif 中不同位置的氨基酸组合对应识别不同的 RNA 碱基(Shen et al 2015)，还揭示出四种密码氨基酸和其对应的四个 RNA 碱基之间特异识别机制是由氢键介导的(Shen et al 2016)。这些研究结果为预测 PPR 蛋白的靶标序列提供了参考依据，有利于今后确定候选 *Rf* 基因中到底是哪一个参与了 CMS 基因的认识和加工。

对于 *Rf* 基因和 CMS 基因相互作用调控 CMS 的分子机制，普遍认为是细胞核基因组和线粒体基因组之间顺行调控 (Anterograde regulation) 和逆行调控 (Retrograde regulation) 共同作用的结果(Fujii and Toriyama 2008; Woodson and Chory 2008; Fujii and Toriyama 2009; Chen and Liu 2014)。顺行调控是指细胞核基因对线粒体功能的影响：从细胞核基因组编码的线粒体定位基因 (mitochondrial-sorting gene, MSR) 产物，包括 RF 蛋白、组织特异性调控因子 (tissue-specific regulatory factors, TSRFs) 和线粒体电子传递链复合物 (mtETC) 亚基，被运送到线粒体内从基因组 (G)、mRNA (R)、蛋白质 (P) 或者代谢 (M) 水平上对 CMS 基因进行调控。而逆行调控是指从线粒体中产生的一些逆行信号会从线粒体传递到细胞核，造

成绒毡层或者小孢子的细胞程序性死亡，导致配子体不育。这些逆行信号是由与线粒体电子传递链复合物亚基功能相关的细胞色素 *c* 和过量 ROS (reactive oxygen species) 释放的(Ji et al 2013; Wang et al 2013)。

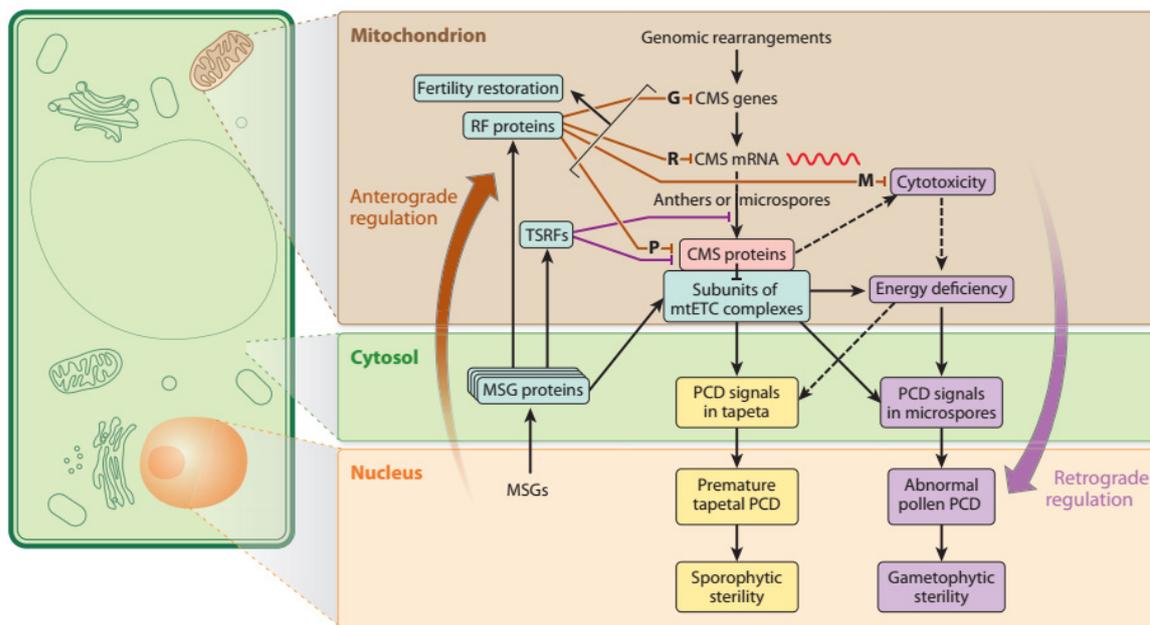


图 2 细胞质雄性不育的线粒体-核基因互作模型(Chen and Liu 2014)。

Figure 1 The general model of mitochondrial-nuclear gene interactions in CMS system (Chen and Liu 2014).

1. 2. 2 细胞核雄性不育

相较于细胞质雄性不育，细胞核雄性不育仅仅涉及到核基因，在遗传上相对较为简单，对其研究更为深入细致和全面。被子植物的花粉包括两种类型：2 细胞花粉和 3 细胞花粉，他们经历的花粉发育过程相似，都包括小孢子发生和雄配子发生两大过程：从花药原基中分化出的花粉母细胞经过减数分裂形成小孢子四分体，四分体解散后游离出来的小孢子再经过两次有丝分裂形成雄配子体，然后进入花粉成熟阶段(Smyth et al 1990; Ma 2005; Zhang and Wilson 2009; Zhang and Yang 2014)。只是在 3 细胞花粉形成过程中的生殖细胞的第二次有丝分裂发生在花粉成熟之前，而 2 细胞花粉在花粉成熟之前生殖细胞只发生了一次有丝分裂，形成 2 个细胞，要等花粉进入花粉管中才完成第二次有丝分裂(杨弘远 2005)。作为模式植物的水稻和拟南芥都属于 3 细胞花粉，因此对其相关方面的研究比较深入和广泛，对 2 细胞花粉

植物诸如烟草等的研究不是很多(Rutley and Twell 2015)。在花粉发育的任何一个过程中都有一系列的基因参与, 这些相关基因的突变、失活或者异常表达都会使发育受阻, 对花粉育性产生影响, 造成细胞核雄性不育(Goldberg et al 1993; Scott et al 2004; Ge et al 2010)。

对水稻全生育期各个组织表达基因进行分析发现, 与其他组织中高丰度和低丰度表达基因均匀紧密分布在组织平均表达量周围的特征不同, 花药中存在大量基因明显高于或低于平均表达水平, 说明有大量的基因在花药中被诱导上调表达或者被抑制下调表达(Wang et al 2010b)。选取拟南芥单核小孢子、双核花粉粒、三核未成熟花粉粒和成熟花粉粒进行 Affymetrix 芯片分析, 有 13,977 个基因至少在这四个组织其中之一表达, 占基因总数的 61.9%, 表达非常丰富; 与其他数据库比较后发现 有 1,355 个约 9.7% 的转录本是在花粉中特异表达, 主要是一些细胞周期基因、转录因子和翻译因子(Honys and Twell 2004)。利用显微切割技术特异性地对水稻花粉发育过程中减数分裂期到三核花粉期的 5 个时期的小孢子细胞和绒毡层细胞进行转录组分析显示, 共检测到 28,141 个基因在花药中表达, 其中大约 12.3% 的基因是属于花粉特异性表达。并且与拟南芥结果一致, 这些花粉特异表达基因具有发育时期特异性, 在二核花粉和三核花粉中最多, 主要涉及到转录、翻译、细胞周期、脂类代谢和次生代谢过程(Hobo et al 2008)。随着新一代基因表达检测技术 RNA-seq 的出现, 对各种植物花粉的转录组分析越来越细致, 有助于更好地寻找花粉发育相关基因(Rutley and Twell 2015)。

基于突变体库的反向遗传学是目前研究花粉发育相关基因最为快速和有效的手段, 在水稻和拟南芥中鉴定到了几十个与细胞核雄性不育相关的基因, 以调控减数分裂、绒毡层发育和花粉壁形成的基因研究最为透彻(Nguyen et al 2016)。这些基因之间相互作用或者独立, 形成了一个庞大的调控网络(Ge et al 2010; Chang et al 2011; Zhang et al 2011; Guo and Liu 2012; Luo et al 2014; Shi et al 2015), 并且许多调控路径在水稻和拟南芥之间是保守的(Wilson and Zhang 2009; Zhang et al 2011; Gómez et al 2015; Shi et al 2015), 这种进化上的保守也为研究其他植物的花粉发育提供了很好的借鉴。

虽然关于细胞核雄性不育的机理研究比较清楚, 对基础分子生物学的发展做出了重要的贡献, 但是在实际生产应用中由于这些细胞核雄性不育基因都是隐性遗传,

很难找到保持系在后代中保持其不育性，限制了应用。后来，一些环境敏感型细胞核雄性不育植株的发现，才使得两系法育种成为可能(Chen and Liu 2014)。环境敏感型细胞核雄性不育系 EGMS 的不育性在一定环境条件下能够被恢复，受到日照长度和温度的影响。虽然在其他植物中也有关于 EGMS 的报道，但以水稻中对光温敏感性不育系的研究和利用最为深入和广泛，而我国在这方面的研究开展最早，一直处于世界领先水平。

1.2.3 水稻光温敏雄性核不育研究进展

1.2.3.1 光温敏雄性核不育水稻的发现及其光温特性

自从袁隆平于 1966 年发文提出寻找雄性不育水稻利用杂种优势的设想(袁隆平 1966)，在我国水稻育种界掀起了寻找雄性不育株的热潮，开展了轰轰烈烈的杂交稻研究。那时候找到的雄性不育株都属于细胞质雄性不育，应用于三系杂交稻育种中。在此背景下，1973 年石明松在湖北沔阳沙湖原种场种植的晚粳品种农垦 58 的大田中发现了一株自然突变的雄性不育株农垦 58S，并且按照选育细胞质雄性不育的传统方法与其他水稻品种进行各种杂交以期寻找到保持系。但在这一过程中发现农垦 58S 的育性极不稳定，早播种时不育率和不育度都比较高，而晚播种的育性则比较高，从不育株上割菹再生的分蘖上也能收获比较多的种子。因此在 1980 年和 1981 年对农垦 58S 进行分批播种实验，结果表明单株的结实率和光照长度的相关性达到了极显著水平，进一步的遮光实验排除了温度对育性的影响，说明日照长度对农垦 58S 的育性起着关键性的作用：长日照下不育，短日照下可育。而且用了 500 多个材料与之进行测交，都不能恢复长日照下农垦 58S 的育性，排除了农垦 58S 是细胞质雄性不育株的可能性；正反交的 F_2 中可育株与不育株的比例符合孟德尔一对隐性基因控制的 3:1 的分离比。这些实验结果在 1985 总结成文，提出了其在杂交稻育种中的应用价值(石明松 1985)，一时引起了轰动。接下来的几年中，各省以农垦 58S 为供体培育出了一大批适应当地气候条件的光敏雄性不育系，并且配置的杂交组合表现出比对照明显的杂种优势，表明光敏雄性不育系巨大的应用前景：根据其育性转换特性在可育期内不需要依靠保持系，自身就可繁种；不育期内用作不育系进行杂交制种。这种光敏雄性不育系能够一系两用，大大简化三系杂交稻的育种过程。

此后, 育种学家们又陆续发现了育性受温度影响的温敏雄性不育系: 5460S, 安农 S-1 和衡农 S-1, 这三个早籼品种的育性表现为高温下不育、低温下可育, 此特性同样可应用于两系杂交稻的生产育种(杨仁崔等 1989; 周广洽等 1990; 蒋佐升等 1992; 邓华凤等 1999)。在 1986 年福建农学院从细胞质雄性不育恢复系 5460 中发现了雄性不育株, 而 5460 来自于 IR 54 的辐射后代。最开始认为该雄性不育株属于光敏雄性不育, 温度对其育性有一定的影响, 因此命名为 5460ps(杨仁崔等 1989)。利用人工气候箱设置不同光温条件发现在高温下无论光照长度如何, 大部分植株表现出完全不育或者高度不育, 低温下完全可育, 认为 5460ps 应为温敏雄性不育, 更名为 5460S(孙宗修等 1989)。湖南杂交水稻研究中心于 1987 年从超 40B/H285//6209-3 F₃ 群体中发现了雄性不育株后又经过两代后育成了不育系安农 S-1(邓华凤等 1999), 经历了与 5460S 相似的鉴定过程后, 最终确定其为温敏雄性不育系(周广洽等 1990), 并且成为了我国选育新型温敏不育系的主要基因源之一。衡农 S-1 则是 1986 年由衡阳市农业科学研究所从长芒野生稻/R0183//测 64 的 F₂ 群体中的不育株 87N123S 选育而来, 具有温敏特性(蒋佐升等 1992)。

目前国内用于生产的光温敏雄性不育系主要来自于农垦 58S 和安农 S-1 的衍生系(斯华敏等 2011)。而国家水稻数据中心 (<http://www.ricedata.cn/variety/>) 统计数据显示截至目前为止通过省级以上审定的两系不育系共有 99 个。这些光温敏不育系的培育保证了我国杂交稻生产的种质资源和可持续性。

研究表明光敏不育系农垦 58S 育性转换的光周期敏感时期为二次枝梗原基分化期、雌雄蕊原基分化期和花粉母细胞形成期, 并且后两个时期对育性的诱导作用强于二次枝梗原基分化期, 还存在一定的累积效应(元生朝等 1988)。接受光周期的敏感植株部位为最上两片叶及叶鞘, 尤以心叶为重(张平博等 2010)。同时, 农垦 58S 育性的转换还存在两个温度阈值, 即不育和可育的临界温度。在长日照下过低、短日照下过高的温度都会部分恢复育性; 而不利于水稻正常生长的极端高温和极端低温会降低育性(刘宜柏等 1991)。因此, 只有在一定温度范围内、育性转换的敏感时期确保敏感植株部位接受长度大于 14 h、光强高于 50 lux 的日照才能保证农垦 58S 的不育, 日照长度低于 13.5 h 则育性得到恢复(张自国等 1987)。此外, 在短日条件下进行暗期光中断可诱导不育(张自国等 1987), 并且不同波长的光质对育性转换诱导效果也不同: 蓝光和红光具有白光一样的效应, 绿光则没有; 红光的效应还能被

远红光逆转(李合生 and 卢世峰 1987; 陈克成 and 肖翊华 1991)。温敏不育系安农S-1的临界温度为26℃,高于此温度时植株表现为雄性不育,低于此温度则为可育(曾汉来等 1996)。其对温度敏感的时期为花粉母细胞形成期和减数分裂期(陈良碧 and 周广洽 1994)。

而由这些原始光温敏不育系品种为供体得到的一系列籼粳衍生系的育性转换的临界光长和临界温度却不尽相同,并且与原始供体相差较大,说明遗传背景对育性转换的条件存在一定的影响。同时,在光敏感的温度范围内温度对光照长度在育性影响上具有一定的补偿效应,光敏不育系在较高温度下育性转换的临界光长缩短,较低温度下则需要延长临界光长才能保证不育,所以很多光敏不育系尤其是籼型不育系的育性转换普遍受到光照和温度的双重调控;但光周期对温敏不育系的育性没有明显影响。

1.2.3.2 水稻光敏雄性核不育基因的定位和克隆

早期主要采用图位克隆的方法定位光敏雄性不育基因,由于所用群体材料的遗传背景不同导致对同一位点的定位结果也有所偏差。到目前为止,水稻中已被定位和克隆的光敏不育基因并不是很多,比如*pms1*和*pms3*等,此外利用反向遗传学的方法也找到了其他光敏不育基因,比如*CSA*。这些基因所在染色体等相关信息详见表1。

将农垦58S与其他粳稻和籼稻进行正反交和测交,表明农垦58S的不育性是由核内隐性基因控制,与农垦58之间存在一对主效基因的差异(石明松 and 邓景扬 1986),与其他品种间则存在1-2对主效基因的差异(梅明华 and 李泽炳 1995),并且光敏核不育水稻不育性的稳定性和育性的恢复性还受到其他QTL的影响(He et al 1999)。

对这些位点的定位主要是利用RFLP、AFLP、SSR等分子标记结合极端不育-隐性群的BSA分析法进行遗传连锁图谱的构建。Zhang等(1994)率先在32001S×明恢63的F₂分离群体中检测到两个光敏雄性不育位点*pms1*和*pms2*,并且位于第7染色体上的*pms1*基因的效应约为第3染色体上*pms2*的2-3倍,作为研究的重点在后来继续进行精细定位,最终将该基因定位于两个SSR分子标记Fssr和Rssr之间85 kb的范围内(Liu et al 2001b)。但在此研究过程中发现农垦58S和农垦58之间在*pms1*区域内存在多态性的标记与这两者之间的育性不存在相关性,说明农垦58S和农垦58在*pms1*位点上没有

育性的分离, 农垦58S中的光敏雄性不育基因另有所在(王风平等 1997)。从农垦58S与1514和轮回422这两个组合中鉴定到了除*pms1*外的另一个位于第12染色体上的光敏雄性不育位点*pms3*, 并最终证实*pms3*是导致农垦58突变为农垦58S的位点(Mei et al 1999a; Mei et al 1999b)。围绕*pms3*进行多次逐步定位, 将之缩小到了28.4 kb的范围内(Lu et al 2005)。此外, 浙江大学利用培矮64S与9311组合同样在第7染色体上定位到了光敏不育位点*pms1(t)*, 有意思的是, *pms1*位点的共分离标记RG477正好位于*pms1(t)*候选基因LOC_Os07g12130的第二个intron上, 但也不能说明这两个位点是一样的(Zhou et al 2011), 尽管农垦58S和培矮64S的不育主基因等位(邓晓建等 1997), 可能存在两相邻基因控制同一性状的可能性。

绵9S是四川绵阳农科所选育的籼型光敏不育系, 其不育基因来源与农垦58S无关(王志等 1999), 并且不育性也是由单个隐性核不育基因控制, 位于第4染色体上, 命名为*pms4*(Huang et al 2008)。表1归纳了这些基因的相关信息。

直到2012年华中农业大学和华南农业大学才相继报道了对农垦58S中控制的光敏雄性不育基因*pms3*的成功克隆和功能分析。研究结果表明他们分别利用不同的光温敏核不育系——农垦58S和培矮64S定位的不育位点*pms3*和*p/tms12-1*实为同一个基因。培矮64S虽然是由农垦58S转育而来, 两者之间存在等位点的核不育基因(邓晓建等 1997), 但更多地表现为温敏雄性不育: 高温下不育, 低温下可育。而*p/tms12-1*的候选基因片段在长日高温条件下都能恢复农垦58S和培矮64S的育性。由于两种不育系对光温反应的不同, 因此两个课题组对*pms3*功能的解析也略有不同。

前期研究已经表明是*pms3*基因的突变导致了农垦58变为光敏不育的农垦58S(Mei et al 1999a), Ding等发现这种突变是由*pms3*转录本上单个碱基造成的, 农垦58中碱基G变为了农垦58S中的碱基C, 引起*pms3*编码的1, 236 bp长的long non-coding RNA *LDMAR*的RNA二级结构发生了改变, 并造成其启动子区域DNA甲基化中CG甲基化程度的升高, 抑制了基因在长日照下幼穗中的表达量, 导致雄性不育(Ding et al 2012a)。此后进一步深入研究发现是由*LDMAR*启动子区长度为21-nt的小RNA Psi-*LDMAR*介导了该区间的DNA甲基化, 属于典型的RNA介导的DNA甲基化(RNA-directed DNA methylation, RdDM)(Ding et al 2012b)。Zhou等也证实了从培矮64S中鉴定到的*p/tms12-1*位点上的同一个碱基突变造成了光温敏雄性不育, 该碱基突变恰好位于21 nt的小RNA *osa-smR5864m*的第11位上, *osa-smR5864m*和野生型

*osa-smR5864w*都在幼穗中优先表达，但两者之间的表达差异并不受光照长度和温度的影响，推测其作用的发挥并不是通过小RNA的表达量来体现，可能是通过对小RNA结合的下游靶基因进行调控：*osa-smR5864w*会抑制其下游靶基因的表达，而*osa-smR5864m*则不存在这种抑制作用，使得光温敏雄性不育得以体现，因此*p/tms12-1*单碱基的突变是*osa-smR5864m*的丧失功能突变，使其在粳稻和籼稻背景中分别表现出光敏雄性不育和温敏雄性不育(Zhou et al 2012)。此前有研究报道将水稻低分子量GTP结合蛋白编码基因*osRACD*转化农垦58S能使其育性得到一定程度的恢复，抑制*osRACD*的表达会降低农垦58的育性，说明其表达参与了长日照下农垦58S育性的转换(叶建荣等 2004)，但还没有证据表明*osRACD*与*pms3*之间存在互作，是否作为*pms3*的下游靶基因参与光敏雄性不育还有待进一步的研究证实。

基于上述*pms3*基因定位的例子，同样是利用农垦58S和培矮64S分别定位到的光敏雄性不育基因*pms1*和光温敏雄性不育基因*pms1(t)*的候选基因区间存在部分重叠的现象，推测这两个位点极有可能也是由同一个基因控制。

*csa*突变体(*carbon starved anther*)最初是单纯作为一个雄性不育材料进行研究，后来发现其育性受光照长度的影响，在长日照下可育，短日照下不育，具有典型的短光敏雄性不育系特性(Zhang et al 2012)。相对于野生型，*csa*突变体叶片和茎中糖的含量升高，而花器官中糖和淀粉的积累降低，尤其是在花药发育后期花药中的碳水化合物含量降低，这种糖分配的不均最终导致了雄性不育(Zhang et al 2010)，但这种糖分配缺陷在长日照下能得以恢复，使得育性恢复(Zhang et al 2012)。无论是在籼稻还是粳稻中，*CSA*基因的突变都能表现出短光敏雄性不育，不受遗传背景的影响，具有极大的应用价值和潜力。应用CRISPR/Cas9技术对粳稻品种9522、交优5B和空育131中的*CSA*基因进行编辑，分别获得了无转化载体的反光敏雄性不育系9522^{*csa*}、*JY5B*^{*csa*}和反光温敏雄性不育系*KY131*^{*csa*-4}，可应用于水稻生产育种(Li et al 2016)。此外，研究表明油菜素内酯能促进*CSA*基因的表达；抑制油菜素内酯信号因子*OsBZR1*的表达会降低花粉育性，*OsBZR1*通过结合到*CSA*基因的启动子上调控*CSA*的表达(Zhu et al 2015b)。不过没有证据表明油菜素内酯是否会影响*CSA*基因参与的光敏雄性不育。

表1 水稻中已定位的光温敏雄性不育基因相关信息。

Table 1 The information of mapped photoperiod-sensitive and thermo-sensitive genic male sterility genes in rice.

类型	位点	染色体	不育亲本	候选区间	功能	参考文献
光敏	<i>pms1</i>	7	32001S	85 kb	—	(Zhang et al 1994; Liu et al 2001b)
光温敏	<i>pms1(t)</i>	7	培矮 64S	101.1 kb	—	(Zhou et al 2011)
光敏	<i>pms2</i>	3	32001S	17.6 cM	—	(Zhang et al 1994)
光敏	<i>pms3</i>	12	农垦 58S	LDMAR	long non-coding RNA	(Lu et al 2005; Ding et al 2012a)
光敏	<i>pms4</i>	4	绵 9S	6.5 cM	—	(Huang et al 2008)
光温敏	<i>p/tms12-1</i>	12	培矮 64S	osa-smR5864m	small RNA	(Zhou et al 2012)
短光敏	<i>CSA</i>	1	<i>csa</i> 突变体	LOC_Os01g16810	MYB 转录因子	(Zhang et al 2010; Zhang et al 2012)
短光敏	<i>rpms1</i>	8	宜 D1S	998 kb	—	(Peng et al 2008)
短光敏	<i>rpms2</i>	9	宜 D1S	68 kb	—	(Peng et al 2008)
光温敏	<i>ptgms2-1</i>	2	广占 63S	50.4 kb	—	(Xu et al 2011)
温敏	<i>tms1</i>	8	5460S	6.7 cM	—	(Wang et al 1995)
温敏	<i>tms2</i>	7	Norin PL12	1.7 cM	—	(Yamaguchi et al 1997)
温敏	<i>tms3(t)</i>	6	IR32364TGMS	2.4 cM	—	(Subudhi et al 1997)
温敏	<i>tms4(t)</i>	2	TGMS-VN1	3.3 cM	—	(Dong et al 2000)
温敏	<i>tms5</i>	2	安农 S-1, 株 1S	LOC_Os02g12290	RNase Z	(Wang et al 2003; Jiang et al 2006; Zhou et al 2014)
温敏	<i>tmsX</i>	2	籼 S	183 kb	—	(Peng et al 2010)
温敏	<i>tms6</i>	5	Sokcho-MS	2.0 cM	—	(Lee et al 2005)
温敏	<i>tms9</i>	2	株 1S	107.2 kb	—	(Sheng et al 2013)
温敏	<i>tms9-1</i>	9	衡农 S-1	162 kb	—	(Qi et al 2014)
温敏	<i>TGMS</i>	9	SA2	11.5 cM	—	(Reddy et al 2000)
温敏	<i>Ugp1</i>	9	Ugp1 共抑制	LOC_Os09g38030	UDP 葡萄糖焦磷酸化	(Chen et al 2007b)
反温敏	<i>tms6(t)</i>	10	G20S	1455 kb	—	(Liu et al 2010b)
反温敏	<i>rtms1</i>	10	J207S	7.6 cM	—	(Jia et al 2001)

1.2.3.3 水稻温敏雄性核不育基因的定位和克隆

相较于光敏雄性不育基因来源单一、绝大部分都是由农垦58S转育而来的情况，温敏雄性不育材料较为丰富，不育基因来源广泛，并且不育系之间的温敏雄性不育位点并不完全等位(陈顺辉等 1996)。尽管光敏不育系农垦58S与温敏不育系培矮64S不育性都由同一个基因控制，农垦58S与安农S-1及他们的衍生系都不存在等位的不育基因(向阳等 2002)。因此定位到的温敏雄性不育位点远多于光敏雄性不育位点，分布在多条染色体上(表1)。遗传研究表明安农S-1和衡农S-1的不育性都是由一对隐性基因控制(邓启云等 2001)，并且安农S-1与其衍生的不育系之间具有等位点的温敏不育基因(向阳等 2002)。相对来说，安农S-1的遗传行为比农垦58S要简单，并且利用其基因源培育出了一大批不育系，在生产中应用广泛，因此对安农S-1中温敏雄性不育基因的定位和克隆相关研究也最多。最早是在大连召开的第一届全国植物基因组大会摘要中由中科院遗传所提及了他们对安农S-1不育基因连锁图谱的构建和定位，将其定位于第2染色体上，命名为*tms5*，后来利用安农S-1和南京11的RIL F₈群体将*tms5*定位到标记C365-1和G221-7之间，遗传距离分别为1.04和2.08cM(Wang et al 2003)。随后华南农业大学用了三个群体进一步将该基因范围缩小至181 kb，这个定位结果与之前遗传所的定位位置非常相近，但并不重叠(Jiang et al 2006)。

后来两个单位展开了合作，最终完成了*tms5*的克隆。他们利用了两个群体(不育系分别为株1S和安农S-1的衍生系香125S)分别将温敏不育基因定位到了第2染色体上的同一个位置上。候选基因*LOC_Os02g12290*能够恢复高温下安农S-1和株1S的育性，说明就是*TMS5*基因(Zhou et al 2014)。*TMS5*编码一个保守短版RNA酶Z，命名为RNase Z^{S1}。单个碱基突变造成了安农S-1和株1S中RNase Z^{S1}编码蛋白质的翻译提前终止，但是RNase Z^{S1}本身并不受温度的影响，其mRNA和蛋白质水平在高、低温下无差别。与其他RNase Z功能相似，RNase Z^{S1}也具有核酸内切酶活性，能对tRNA的3'端进行加工，但这种加工在温敏不育材料和正常品种之间没有区别，最后通过RNA-seq的方法找到了RNase Z^{S1}加工的基因*Ubl40*(泛素核糖体L40蛋白)。RNase Z^{S1}能够将3个*Ubl40*的mRNA加工成多个片段，在温敏不育系中由于RNase Z^{S1}功能的缺失，不能正常加工*Ubl40*，使其过度积累引起了雄性不育；而常规品种中的*Ubl40* mRNA能被RNase Z^{S1}正常降解，不会造成积累，育性正常(Zhou et al 2014)。

之前扬州大学从广占 63S 中定位的温敏不育位点 *ptgms2-1* 也是将 *LOC_Os02g12290* 作为候选基因, 还检测到了同样的单碱基突变, 说明 *ptgms2-1* 实际上就是 *tms5*, 遗憾的是当时没有进行遗传转化验证(Xu et al 2011)。而 *RNase Z^{Sl}* 基因也正好位于 *tmsX* 的定位区间内(Peng et al 2010), 并且毗邻 *tms9* 的候选基因区段(Sheng et al 2013), 说明这些位点可能都由同一个基因控制。有意思的是用于定位 *tms5* 的温敏不育系安农 S-1 是在超 40B/H285//6209-3 F₅ 群体中发现的不育株(邓华凤等 1999), 株 1S 是在抗罗早//4342/02428 的 F₂ 群体中发现的(杨远柱等 2000), 广占 63S 是从质源为农垦 58S 的广亲和性不育系 N422S 与广占 63 杂交 F₂ 中选育而来(杨振玉等 2002), 粳 S 是在粳黄占中发现的自然突变株(Peng et al 2010), 并且这些不育系的不育性都由一对隐性主效基因控制, 但实际上他们的不育基因来源完全不同, 都是在不同群体中独立发现的自然不育株。广占 63S 中的 *RNase Z^{Sl}* 基因序列与常规品种 1587 之间除了已报道的功能性单碱基突变, 还存在 2 个 SNP(Xu et al 2011)。未来分析这些不育系中 *RNase Z^{Sl}* 基因其他位置处的序列变异, 可能有助于加深对 *tms5* 作用机理的理解。

同样通过反向遗传学的方法鉴定到抑制水稻 UDP 葡萄糖焦磷酸化酶同源基因 *Ugp1* 的表达会引起花粉细胞发育过程中正常的胍胍质积累过程被干扰, 导致雄性不育。*Ugp1* 的 RNAi 植株表现为稳定的雄性不育, 而其共抑制植株的育性受温度影响: 高温下不育, 低温下可育(Chen et al 2007b)。研究表明 *Ugp1* 的 mRNA 选择性剪切受温度调控: 在这些共抑制植株中由于转化所用的载体超量表达, 使得体内还存在大量内含子未被剪切的 mRNA, 在低温下 *Ugp1* 共抑制植株的共抑制过程也是正常的, 但在小花中, 与高温条件下相比, 这些内含子未被剪切的 mRNA 能进一步产生更多被正确剪切的转录本, 翻译出足够的 UGPase 蛋白质, 胍胍质积累正常, 育性不受影响(Chen et al 2007b)。基因芯片分析高温下 *Ugp1* 共抑制植株发展的温敏雄性不育系 TGMS-Co27 减数分裂时期幼穗中基因的表达, 发现大部分被抑制, 包括一些与花粉发育相关的重要基因(Pan et al 2014)。

在获得了光温敏雄性不育基因序列的基础上, 针对 *LDMAR* 和 *RNase Z^{Sl}* 的功能性碱基突变设计 CAPS 标记分析了 90 个光温敏雄性不育系中所携带的这两个基因的情况, 发现以农垦 58S 为唯一不育基因源的光温敏不育系中也有部分携带 *RNase Z^{Sl}*, 或者均不含这两个基因; 由安农 S-1 和株 1S 衍生的不育系则都含有 *RNase Z^{Sl}*; 其他不育基因与农垦 58S 及安农 S-1 和株 1S 无关的两系不育系中也会携带其中之一个

基因,但衡农 S-1 中都不携带这两个基因。而且携带 *RNase Z^{SI}* 的不育系在目前两系杂交稻生产中占据主导地位(张华丽等 2015)。该结果也从侧面说明了两系不育系中参与光温敏雄性不育的基因的复杂性,远远不止目前克隆的几个基因;而且不育性的体现还包含了光敏不育和温敏不育基因之间的互动,但已报道的几个光温敏雄性不育基因作用机理各不相同,他们之间的分子调控路径是否存在交叉点也是未来研究的一个方向。

1.3 Long non-coding RNA 研究进展

1.3.1 Long non-coding RNA 的定义,分类及功能

早期研究认为大部分从 RNA 聚合酶 II 复合体合成得到的转录本会被翻译成蛋白质,随着全基因组高通量测序和芯片技术的发展,发现大量转录本并没有 ORF (Open Reading Frame)或者翻译起始密码子序列,被称之为非编码 RNA(non-coding RNA, ncRNA)。ncRNA 在基因组中广泛存在,以人类为例,超过 70%的基因组序列能够被转录,但仅有 2%编码蛋白质(Bertone et al 2004; Birney et al 2007; ENCODE Project Consortium 2012)。根据核苷酸序列长度和结构将这些 ncRNA 分类(图 3),包括看家 ncRNA 和调控 ncRNA。看家 ncRNA 包括核糖体 RNA、转运 RNA、核仁小 RNA (small nucleolar RNA, snoRNA)、小核 RNA (small nuclear RNA, snRNA)和端粒 RNA (telomere RNA),看家 ncRNA 在细胞中组成型表达,而调控 ncRNA 中研究比较深入的是 piRNA (piwi-interacting RNA)、siRNA (short interfering RNA)和 miRNA (microRNA)这三类长度小于 50 nt 的短链 ncRNA,它们在进化上比较保守,通过碱基互补配对在转录、RNA 加工、RNA 稳定性、翻译及染色质结构等水平上对基因和基因组进行调控(Eddy 2001; Klattenhoff and Theurkauf 2008; Chen 2009)。长链 ncRNA (long non-coding RNA, lncRNA)是一类长度大于 200 nt 的非编码 RNA,最早是在 2002 年对小鼠全长 cDNA 进行大规模测序时发现了这类长链不编码蛋白质的 RNA,其转录和加工过程与编码蛋白质的 RNA 相同,大部分会经过 5'端加帽子和 3'端 poly(A)尾巴加工,也包含有 exon 和 intron(Okazaki et al 2002)。但 lncRNA 在进化上的保守程度很低,并且表达丰度普遍不高,存在很强的组织特异性,一度被认为是转录的“噪音”,直到 2006 年对 lncRNA 分析表明它们也是有功能的

(Pang et al 2006)。随后陆续在各个物种中发现了多个参与各种生物学路径调控的 lncRNA，对这些 lncRNA 功能的阐释及其与互作蛋白之间的相互作用说明了其在发育上的重要性，并且涉及到癌症与一些其他人类疾病(Wapinski and Chang 2011; Tano and Akimitsu 2012; Batista and Chang 2013)，目前已经成为生物学研究的一个热门领域。

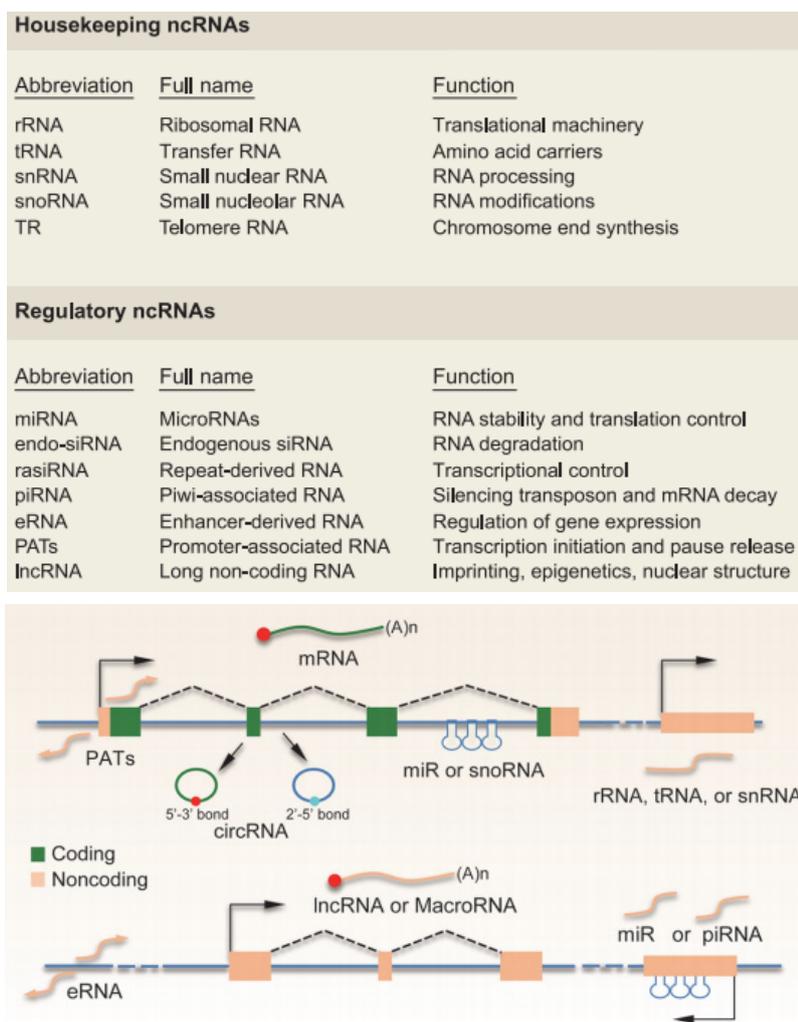


图3 非编码RNA的分类和形成(Fu 2014)。

Figure 3 The production of distinct families of non-coding RNAs(Fu 2014).

LncRNA 根据其与邻近蛋白质编码基因之间的相对位置可以简单地分为 5 类(图 4A): sense, 由正义链转录而来, 与编码基因转录本的一个或多个 exon 重叠; antisense, 从反义链转录而来, 也与编码基因转录本的一个或多个 exon 反向重叠; bidirectional, 转录位点位于编码基因上游, 与邻近编码基因的转录起始位点很近, 但方向相反, 没有重叠; intronic, 来自于编码基因的 intron; intergenic, 位于两个编码基因的基因间区, 是一个独立的转录单元。按照转录本的长度、与已知功能 DNA

element 之间的关系、与重复单元之间的关系、参与的生化路径、序列和结构保守性、亚细胞结构和生物学功能可进一步将其细分(St. Laurent et al 2015)。

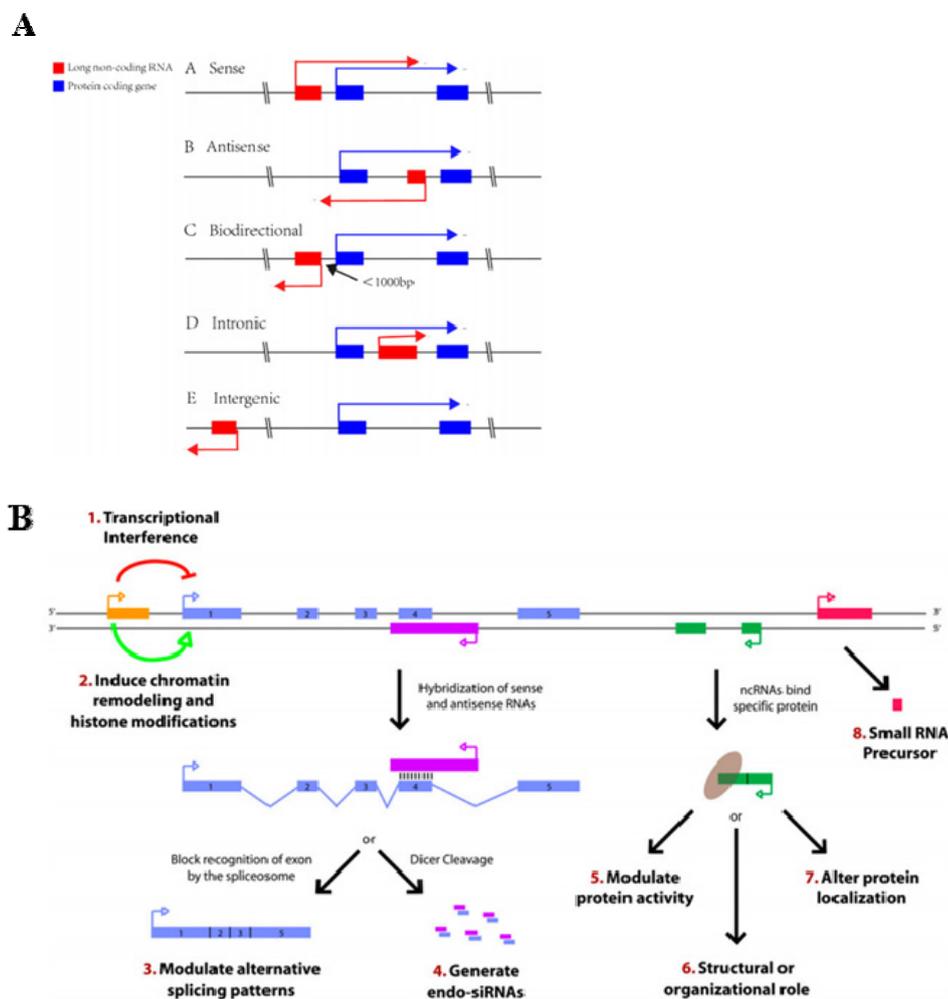


图 4 长链非编码 RNA 的分类和作用方式。

A, 长链非编码 RNA 的分类(Zhang et al 2014a); B, 长链非编码 RNA 的作用方式(Wilusz et al 2009)。

Figure 4 The categories and function styles of lncRNAs.

A, The categories of lncRNAs; B, The regulatory paradigms for how lncRNAs function.

LncRNA 通过染色质、转录调控及转录后 mRNA 加工等方面调控基因表达参与各种细胞学功能(Mercer et al 2009; Wilusz et al 2009; Fatica and Bozzoni 2014; Angrand et al 2015)。其作用方式可以分为 8 类(Wilusz et al 2009) (图 4B): 1, 转录干涉, 通过抑制 RNA 聚合酶 II 复合体的募集来抑制下游邻近基因的表达; 2, 诱导表达, 通过诱导染色质重塑和组蛋白修饰诱导下游基因的表达; 3, 调控选择性剪接, 来自于反义链上的 lncRNA 与正义链上的转录本的 exon 存在重叠, 这种杂交干扰了剪接复合体对 exon 的识别, 得到了不同的选择性剪接产物; 4, 产生内源 siRNA,

方式 3 的这种杂交有时会募集到 Dicer 复合体, 被剪切形成内源 siRNA; lncRNA 能够结合特异性的蛋白以调控蛋白质活性(方式 5), 或者作为一个结构单元介导 RNA-蛋白质复合体的形成(方式 6), 或者改变蛋白质的亚细胞定位(方式 7); 8, 作为前提被加工产生小 RNA, 如 miRNA、piRNA 等。

虽然 lncRNA 的作用方式多种多样, 但其作用机理归纳起来可以分为 4 种(Wang and Chang 2011)(图 5): 信号(signal)、诱饵(decoy)、向导(guide)和脚手架(scaffold)。LncRNA 的表达量可以作为一个信号, 反映结合于其启动子上的转录因子(图 5 左上彩色椭圆形)或其他信号路径在时间和空间上的基因调控, 由于 lncRNA 表达的时空特异性, 有时能作为时空表达和发育时期的信号标记; 作为诱饵, lncRNA 可竞争性与转录因子及其他蛋白质结合, 使它们远离染色质或者细胞核亚区, 比如 lncRNA 能够与某些 miRNA 的靶基因竞争性地结合 miRNA, 从而调控 miRNA 的功能; lncRNA 作为向导能够招募染色质修饰酶以改变染色质的结构, 以便以顺式作用方式近距离或者反式方式远距离调控目标基因的表达, 这种调控方式最终是通过 DNA 的表观修饰完成; lncRNAs 还能像一个脚手架一样能将多个蛋白质聚集起来形成核蛋白复合体, 作用于染色质影响组蛋白修饰, 另一方面, lncRNA 的脚手架作用也有助于形成和稳定核结构或某些信号复合物。

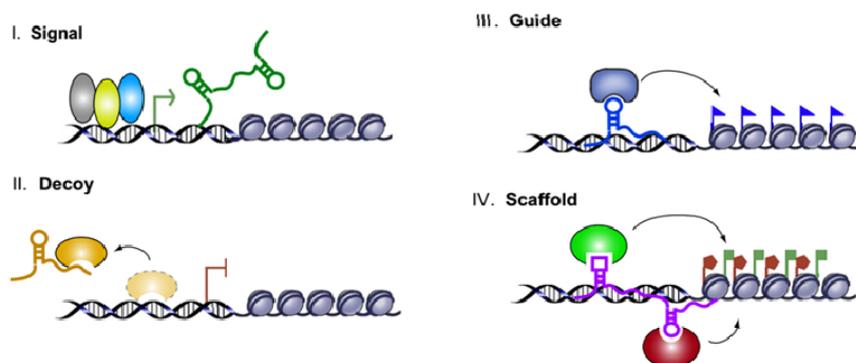


图 5 长链非编码 RNA 的作用机理(Wang and Chang 2011)。

Figure 5 Schematic diagram of lncRNA mechanism(Wang and Chang 2011).

1.3.2 植物 lncRNA 研究进展

相较于动物中对 lncRNA 单个基因及其互作蛋白等作用机理的深入研究, 植物中仅对少数几个 lncRNA 的功能有所认识, 大部分都还停留在全基因组 lncRNA 的发掘上。利用 RNA-seq 测序技术在水稻(Zhang et al 2014b; Wang et al 2015a)、玉米(Li

et al 2014b; Fan et al 2015; Wang et al 2015a; Lv et al 2016)、拟南芥(Liu et al 2012; Di et al 2014; Wang et al 2014a; Wang et al 2016; Yuan et al 2016)、小麦(Xin et al 2011)、棉花(Wang et al 2015b)、番茄(Zhu et al 2015a)、烟草(Xie and Fan 2016)、苜蓿(Wang et al 2015c)、向日葵(Flórez-Zapata et al 2016)、黄瓜(Hao et al 2015)、杨树(Chen et al 2016)中都检测到了 lncRNA 的存在,涉及到多个组织、各种农艺性状和环境响应,表明了 lncRNA 存在的普遍性和重要性(Zhang and Chen 2013; Liu et al 2015)。并且利用这些数据建立了植物 lncRNA 的数据库(Jin et al 2013; Paytuví Gallart et al 2015; Szcześniak et al 2015; Yi et al 2015)。

植物中第一个发现并深入研究的 lncRNA 基因是拟南芥中与春化诱导开花相关的 *COOLAIR* (cold induced long antisense intragenic RNA) (Swiezewski et al 2009)。*COOLAIR* 转录自整个开花基因 *FLOWERING LOCUS C* (*FLC*) 全长 7 kb 的反义链,并且存在不同的剪接模式。*FLC* 是自主途径 (autonomous pathway) 和春化中控制开花的关键基因,植物通过抑制 *FLC* 的表达来促进开花。在持续低温下 *COOLAIR* 诱导表达,通过多梳蛋白复合物介导的表观修饰沉默 *FLC* 的表达,促进开花 (Swiezewski et al 2009)。在自主途径中 *COOLAIR* 的各种选择性剪接模式导致组蛋白去甲基化修饰发生改变,影响 *FLC* 的表达量(Liu et al 2010a; Li et al 2015)。而且在 *COOLAIR* 的启动子和第一个 exon 上存在 R-loop 结构,稳定了 AtNDX 蛋白与 *COOLAIR* 的结合,抑制 *COOLAIR* 的转录,进而调控 *FLC* 的表达(Sun et al 2013)。*COOLAIR* 和 *FLC* 之间还存在负反馈调节机制,CDKC;2 (cyclin-dependent kinase C) 能影响 RNA 聚合酶 II 的磷酸化状态,失去功能后会抑制 *COOLAIR* 的转录,干扰 *COOLAIR* 对 *FLC* 的抑制作用,间接上调 *FLC* 的表达量;而这种干扰又反过来促进 *COOLAIR* 的表达(Wang et al 2014b)。在 *FLC* 正义链的第一个 intron 上还发现了一个新的 lncRNA *COLDAIR* (cold assisted intronic noncoding RNA),长 1.1 kb,只有一个 exon。虽然也是 RNA 聚合酶 II 转录而来,但与 *COOLAIR* 不同,*COLDAIR* 只有 5' 帽子结构,没有 poly(A)尾巴(Heo and Sung 2011)。*COLDAIR* 能够与 PRC2 (Polycomb Repression Complex 2) 直接互作,在春化过程中将 PRC2 招募到 *FLC* 上,PRC2 通过介导组蛋白 H3K27me3 甲基化抑制 *FLC* 的表达,促进开花。这种调控机制类似于图 5 中的脚手架作用,也是动物中 lncRNA 调控基因表达的重要方式,表明该机制在真核生物进化上的保守(Ietswaart et al 2012; Kim and Sung 2012; Chekanova 2015)。

与这两个 lncRNA 功能类似, 水稻中克隆得到的第一个 lncRNA 基因 *LDMAR* (long-day specific male-fertility associated RNA) 也是通过表观修饰发挥作用(Ding et al 2012a)。 *LDMAR* 全长 1, 236 bp, 没有 intron, 产生于基因间区, 控制光敏雄性不育性状。野生型中 *LDMAR* 的一个单碱基突变引起 RNA 二级结构的改变, 并造成 *LDMAR* 启动子区 DNA 甲基化水平的升高, 抑制了基因在长日照下幼穗中的表达, 导致雄性不育。这种表观修饰改变是由启动子区的一个小 RNA 介导的 RdDM(Ding et al 2012b)。而且 *LDMAR* 本身还能够产生 21 nt 的小 RNA osa-smR5864w, 恰好位于其第 11 位上的碱基突变可能解除了小 RNA 对下游靶基因的抑制作用, 恢复育性(Zhou et al 2012)。是否有其他蛋白质参与了 *LDMAR* 的表观修饰还未为可知, 需要更深入的研究发掘。

植物 lncRNA 也能作为诱饵与 miRNA 竞争性结合, 调控 miRNA 的功能。拟南芥中的 lncRNA *IPSI* 基因 (Induced by Phosphate Starvation 1) 含有一段 23 nt 长的序列, 可以与 miR-399 互补配对, 但在 miRNA 的第 10 到第 11 位上形成了 3 nt 的 loop, 使得 miR-399 不能对 *IPSI* 剪切, 这种竞争性结合导致了 miR-399 的靶基因 *PHO2* 的 mRNA 不能被正常降解, 表达量升高, 改变了茎中磷的含量。目前广泛应用于 miRNA 功能研究的 target mimicry 技术就是基于此现象(Franco-Zorrilla et al 2007)。大规模测序发现植物体内存在大量这类具有 eTM (endogenous target mimic) 能力的 lncRNA(Wu et al 2013; Ye et al 2014)。

1.4 植物 phasiRNA 研究进展

在植物中还还存在一类特异的 lncRNA——*PHAS* (phased siRNA producing loci) 基因, 它们作为小 RNA 的前体, 能产生 21 nt 或 24 nt 长的 phased small interfering RNA (phasiRNA)。顾名思义, phasiRNA 就是一簇小 RNA 以 21 nt 或者 24 nt 为相位, 在基因组上依次首尾相接连续排列。这种小 RNA 按照相位依次排列的现象在动物特有的 piRNA 中也有报道, 但 phasiRNA 和 piRNA 的形成过程和作用方式差别很大, 相位排列只是一种现象而已, 不能因此简单地将两者等而视之(Han et al 2015; Mohn et al 2015)。

1.4.1 tasiRNA 的发现和形成过程

phasiRNA 的概念其实引申自 tasiRNA (*trans-acting small interfering RNAs*), 最早在拟南芥中鉴定到 tasiRNA 时, 发现与之前报道的内源 siRNA 不同, 其产生需要 miRNA 形成过程中的一系列蛋白, 而且 tasiRNA 会抑制其起源的转录本的表达量, 并且通过反式作用方式调控目标基因的表达(Peragine et al 2004; Vazquez et al 2004), 后来研究发现 tasiRNA 的形成确实是由 miRNA 介导剪切产生的(Allen et al 2005; Yoshikawa et al 2005)。tasiRNA 来自于 TAS (*trans-acting siRNA producing loci*) 位点, TAS 位点转录成非编码的转录本, 包含有一个或两个 miRNA 的识别位点, 从其中一个 miRNA 介导的剪切位点开始在 RDR6 (RNA-dependent RNA polymerase 6) 的作用下变成 dsRNA, 再被 DCL 酶剪切成一串相连的以 21 nt 为相位排列的 siRNA(Allen et al 2005; Gascioli et al 2005; Xie et al 2005; Yoshikawa et al 2005; Chen et al 2007a; Howell et al 2007), 这些以二级 siRNA 作用的 tasiRNA 一方面能够像其他 siRNA 一样作用于下游靶基因, 还能进一步结合到其他转录本上, 启动下一轮 tasiRNA 的形成, 造成 siRNA 多米诺骨牌样的级联产生(Chen et al 2007a)。

到目前为止在拟南芥中共鉴定到 8 个 TAS 基因, 属于 4 个家族 (*TAS1*, *TAS2*, *TAS3* 和 *TAS4*)。miR173 介导 *TAS1a/1b/1c* 和 *TAS2* 产生 tasiRNA, 这些 tasiRNA 的靶基因编码 PPR 家族蛋白及一些未知功能的蛋白(Allen et al 2005; Yoshikawa et al 2005)。miR390 介导 *TAS3a/3b/3c* 产生的 tasiRNA 也被称之为 tasiARF, 因为它们的靶基因来自于 *ARF* (ausin response factor) 家族, 包括 *ARF3/ETTIN* 和 *ARF4* 等(Williams et al 2005), 主要参与调控叶片形态、叶片背腹极性、发育时期转换等过程, 并且 tasiARF 在植物的进化中非常保守(Allen et al 2005; Axtell et al 2006; Talmor-Neiman et al 2006; Montgomery et al 2008; Wang et al 2010a; Rajeswaran et al 2012)。*TAS4* 产生的 tasiRNA 是由 miR828 介导, 靶定到 3 个 MYB 转录因子上(Rajagopalan et al 2006)。

tasiRNA 的形成可分为两条途径, *TAS1*、*TAS2* 和 *TAS4* 与 *TAS3* 分属于两类(图 6)。miR828 和 miR173 都是 22 nt 的 miRNA, 在 *TAS1*、*TAS2* 和 *TAS4* 上都只有一个靶位点, 位于 5' 端, 被 AGO1 (argonaute 1) 识别, 在 SGS3 (suppressor of gene silencing 3) 和 RDR6 的作用下合成 dsRNA, 再被 DCL4 (DICER-like 4) 和 DRB4 (DICER RNA binding factor 4) 加工成 21 nt 的 tasiRNA(Allen et al 2005; Yoshikawa et

al 2013)。而在另一条路径上的 TAS3 转录本上有两个 21 nt 的 miR390 的靶位点，AGO7 都能识别这两位点，但只有 3'端的靶位点能够被 RISC(RNA-induced silencing complex) 结合并剪切，从而招募 SGS3 和 RDR6 合成 dsRNA(Axtell et al 2006; Rajeswaran et al 2012; Axtell 2013; Fei et al 2013)。最新研究表明 AGO1 的剪切活性对于 tasiRNA 形成 phase siRNA 非常重要，拟南芥的 AGO1 失去了剪切活性后，虽然也能从 *TAS* 位点上产生依赖于 RDR6/SGS3 的大量 siRNA，但这些 tasiRNA 并不是按照相位排列的 RDR6(Arribas-Hernández et al 2016)。

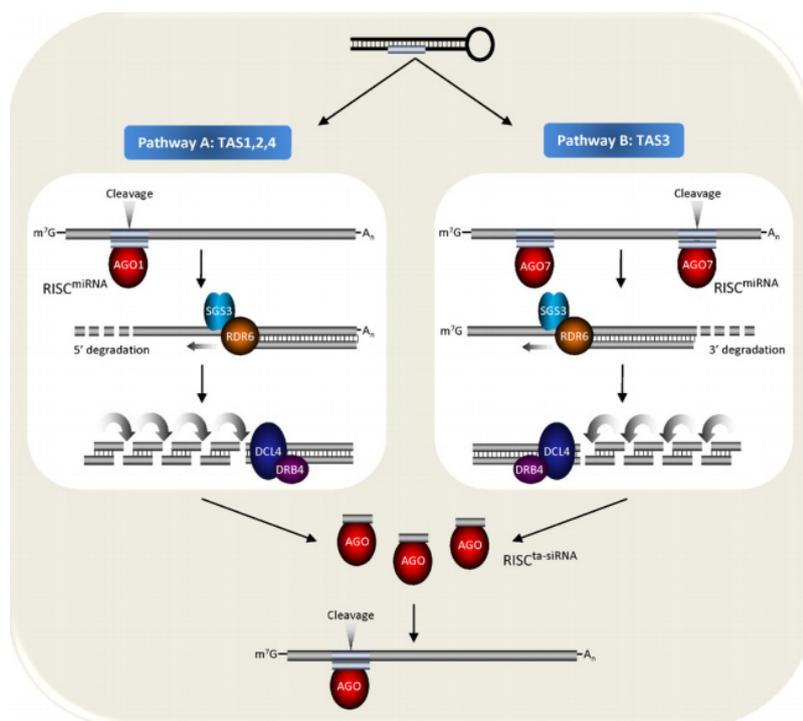


图 6 形成 tasiRNA 的两条途径(Allen and Howell 2010)。

Figure 6 TAS pathway in plants(Allen and Howell 2010).

目前还没有直接对这些 *TAS* 基因进行研究的报道，它们都是通过 tasiRNA 的靶基因发挥作用。除了拟南芥，在水稻(Abe et al 2010)、玉米(Petsch et al 2015)、大豆(Hu et al 2013)、烟草(Tang et al 2012)、苜蓿(Zhou et al 2013)、葡萄(Rock 2013)、苹果(Xia et al 2012)、桃(Zhu et al 2012)、苔藓(Talmor-Neiman et al 2006)、云杉(Källman et al 2013)中都有报道，但仅 tasiARF 有明确的功能，而且功能保守，其余 tasiRNA 的靶基因功能都还未知。针对烟草、马铃薯和番茄的 tasiRNA 建立了专门的网站，有助于分析和鉴定 tasiRNA 相关的 miRNA 及靶基因(Li et al 2012)。

1.4.2 phasiRNA 的形成和特异表达

在研究植物二级 siRNA 的过程中发现大量 miRNA 介导产生的二级 siRNA 找不到能被实验验证的靶位点，也不能确定是以顺式还是反式的方式作用于其他基因，将这类 siRNA 称之为 phasiRNA，从这个角度认为 phasiRNA 包含了以反式方式作用的 tasiRNA 和以顺式方式作用的 casiRNA(图 7A)，因此用 phasiRNA 指代由 miRNA 介导从转录本剪切位点开始产生的按照一定相位排列的二级 siRNA 具有更广泛的意义(Zhai et al 2011; Axtell 2013; Fei et al 2013)。phasiRNA 的形成包含 one-hit 和 two-hit 两条路径(图 7B-C)：one-hit 是指 *PHAS* 基因上仅包含 1 个 miRNA 的识别和剪切位点，该 miRNA 通常是 22 nt，还未有 21 nt miRNA 参与该途径的报道；two-hit 是指 *PHAS* 基因上包含 miRNA 的 2 个识别位点，该 miRNA 一般是 21 nt 或者 22 nt，这两个识别位点有可能都参与 phasiRNA 的剪切形成，或者仅其中之一剪切(Zhai et al 2011; Fei et al 2013; Xia et al 2013)。在苜蓿中超量表达 miRNA 会造成相应的 phasiRNA 表达量的提高(Fei et al 2015)。

在单子叶植物水稻和玉米的幼穗中发现 phasiRNA 主要是由 22 nt 的 miR2118 和 miR2275 介导形成的，得到的 phasiRNA 的长度分别为 21 nt 和 24 nt(Johnson et al 2009; Arikiti et al 2013; Zhai et al 2015)。这类幼穗特异表达的 phasiRNA 形成过程与 tasiRNA 非常相似，但很多参与其中的基因还未被实验证实(图 7D)。水稻中 miR2118 和 miR2275 的合成需要 OsDCL1，而它们介导形成的 21 nt 和 24 nt 的 phasiRNA 则分别由 OsDCL4 和 OsDCL3b 负责加工(Song et al 2012a)。在该过程的上游——将剪切后转录本转换为 dsRNA 的过程中同样需要 OsRDR6 的作用。在 *osrdr6-1* 突变体中，无论是 tasiRNA 还是 21 nt 和 24 nt 的 phasiRNA 的量相对于野生型中来说都极大地减少，表明 OsRDR6 对 phasiRNA 形成的重要性(Song et al 2012b)。在拟南芥中 24 nt siRNA 通常介导靶位点的 RdDM(Chapman and Carrington 2007; Law and Jacobsen 2010)，但对水稻这些 *PHAS* 位点进行 DNA 甲基化分析表明，24 nt phasiRNA 与 RdDM 过程无关，可能影响下游靶基因的 DNA 甲基化(Song et al 2012b)(图 7D)。

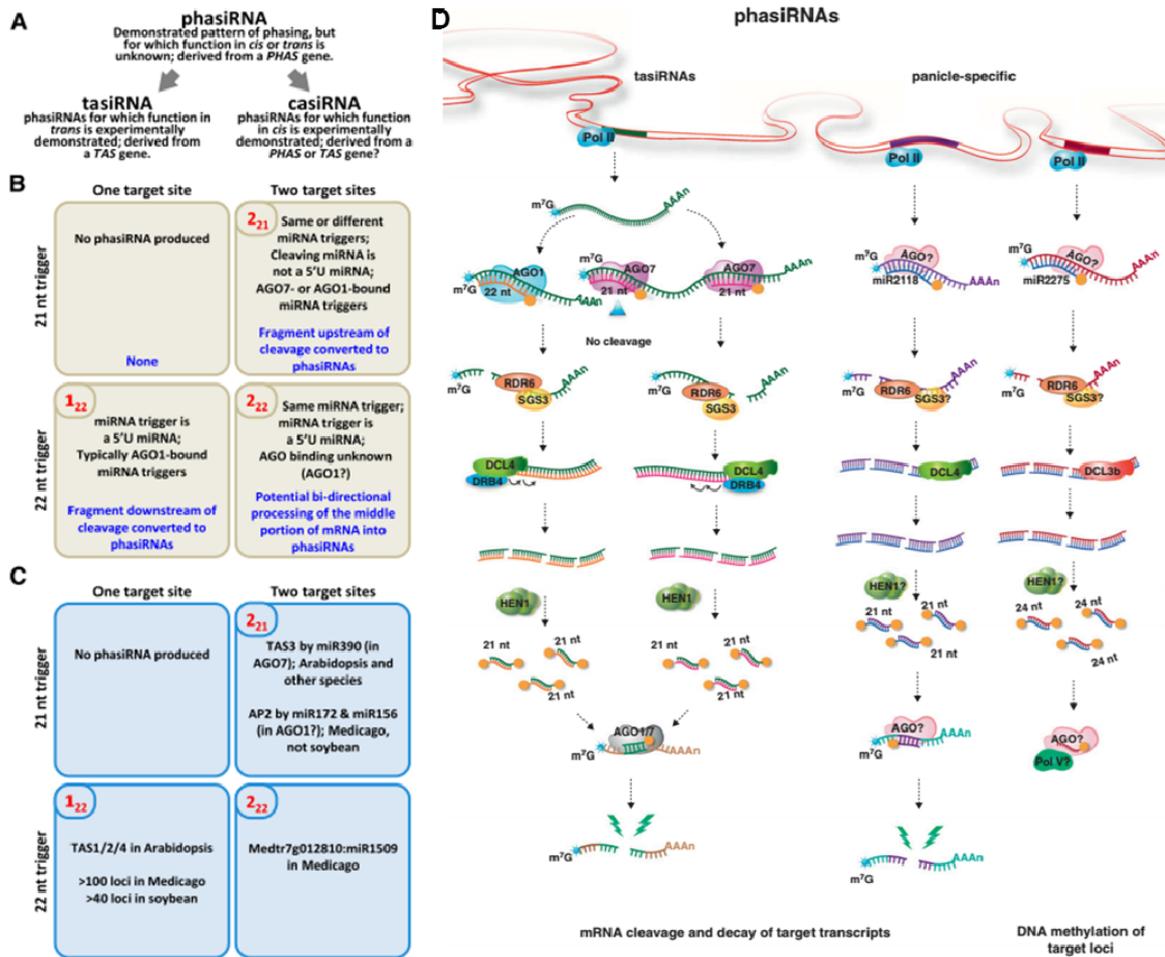


图 7 phasiRNA 的分类和形成过程（来自于 Zhai et al 2011; Arikita et al 2013）

A, phasiRNA 的定义; B, phasiRNA 的分类原则; C, 与 B 中各种类型 phasiRNA 相对应的实例; D, phasiRNA 形成过程和作用方式示意图。

Figure 7 Definition of phasiRNA classes and the schematic overview of their biogenesis (from Zhai et al 2011; Arikita et al 2013).

A, Definition of phased small RNA classes in plants. B, Rules and observations of phasiRNAs in plants. C, Examples of phasiRNAs matching the observations described in B. D, A schematic overview of phasiRNA biogenesis pathways and their modes of action.

研究发现禾本科作物玉米和水稻中 phasiRNA 的表达具有组织特异性，主要在幼穗和花药中表达(Johnson et al 2009; Song et al 2012a; Song et al 2012b; Zhai et al 2015)，并且在二穗短柄草的幼穗中也检测到 phasiRNA 的表达(The International Brachypodium Initiative 2010)，表明 phasiRNA 在植物花发育过程中的可能具有重要的作用，并且存在功能上的保守性。有意思的是水稻 *osrdr6* 突变体表现为花发育缺

陷,并且这种缺陷与温度相关,仅在长日高温下出现该表型,短日低温下则恢复(Song et al 2012b)。而 *OsDCL4* 突变后水稻小花形态也发育异常,很难收到种子(Liu et al 2007)。这些参与 phasiRNA 形成的重要基因也都与花发育相关,进一步证实了这类 siRNA 的重要性。此外,番茄中 22 nt 的 miR4376 能够结合到 Ca^{2+} -ATPase 基因 *ACA10* 上,介导产生 21 nt 的 phasiRNA,超表达 miR4376 后会导致花形态的改变和产量的降低(Wang et al 2011)。草莓中新鉴定到的 22 nt miRFBX7 能介导 6 个 *F-box* 基因产生 phasiRNA,这些 phasiRNA 在花和果实中的表达量比较高,还能进一步靶定其他的 *F-box* 基因(Xia et al 2015)。

另一方面,phasiRNA 还参与了植物抗病过程。在苜蓿、大豆和马铃薯中,phasiRNA 集中在花药和根瘤中表达量,与禾本科作物不同,很多 phasiRNA 来自于编码基因,其中 41%都属于 *NB-LRR* 基因(Zhai et al 2011; Arikrit et al 2014)。植物 *NB-LRR* 基因广为人知的功能是在微生物识别过程中激活植物的抗病反应(Eitas and Dangl 2010),植物基因组中存在的大量 *NB-LRR* 基因组成了识别病原菌的初级免疫系统(Meyers et al 2005)。这些能够介导 *NB-LRR* 基因被剪切的 22 nt miRNA 和来自于 *NB-LRR* 基因的 phasiRNA 都可能抑制 *NB-LRR* 基因的表达,从而限制根瘤菌进入根瘤过程中的抗病反应,而且 *NB-LRR* 基因产生如此大量的 phasiRNA,对于维持植物的长期抗病效应具有进化上的重要意义(Shivaprasad et al 2012; Fei et al 2013)。烟草中病毒或者细菌的侵染还会导致来自于 *NB-LRR* 的 phasiRNA 表达量下降(Shivaprasad et al 2012)。

到目前为止发现的 phasiRNA 主要涉及到植物抗病和生殖发育过程,但这些 phasiRNA 是否能对下游靶基因进行剪切,还没有 PARE 数据或者其他实验证实。因此,尽管在多个物种中得到了大量 phasiRNA,没有任何一个 phasiRNA 的靶基因被找到,而且对于 *PHAS* 基因本身功能的研究也一直处于空白。*PHAS* 基因是否仅仅只是作为产生 phasiRNA 的中间媒介,还是也能像其他 lncRNA 或者作为编码蛋白质发挥作用? phasiRNA 是不是也和 tasiRNA 一样能够触发下游级联 siRNA 的产生? 以顺式反式作用的 casiRNA 是否在植物体内真实存在? 对于 phasiRNA 研究领域存在着太多未知的地方等待我们的探索。

1.5 研究的目的是和意义

从 1985 年第一篇关于光敏雄性核不育水稻的研究报道开始,至今已过去 30 多年,我国育种学家们成功完成了两系杂交稻从理论到应用实践再到大面积种植的伟大历程,为保障我国粮食安全做出了巨大的贡献。两系杂交稻在杂交稻种植面积中的占比越来越大,大有超过三系杂交稻的趋势。但在实际的育种过程中,还存在很多问题,尤其是对于光照和温度等环境因素对光温敏雄性不育系育性的影响还不能精确控制。因此,有必要从分子水平上对控制光温敏雄性不育基因的机理进行深入研究,用于指导选育稳定、优良的光温敏雄性不育系以利于杂交稻生产。

在本研究开展之时,尚未有关于光温敏雄性不育基因的成功克隆。*pms1* 是控制光敏雄性核不育的一个主效位点,本室早期对该位点的遗传进行了详细分析和精细定位,本研究的目的是在此基础上分离克隆 *pms1* 基因,并对其功能进行分析,阐明光敏感雄性不育的作用机理。在研究进行的过程中先后有光敏雄性不育基因 *pms3* 和温敏雄性不育基因 *tms5* 的报道,*pms1* 是否与这些基因之间存在共同的分子机制?我们发现 *pms1* 与 *pms3* 相似,都编码一个长链非编码 RNA,并且能够产生 siRNA,但 *pms1* 还存在其独特的调控机制,这些丰富了我们对于光温敏雄性不育基因作用机理的认识。*pms1* 作为植物中首次报道、具有调控重要农艺性状的、有功能的 *PHAS* 基因,打破了我们以往对 *PHAS* 基因的简单认识,表明了 *phasRNA* 的重要性,为研究植物雄性不育提供了新的思路和方向。

2 材料和方法

2.1 水稻材料及田间种植

本研究主要用到的水稻材料是农垦 58S 和近等基因系 NIL(MH)。农垦 58S 是光敏核不育雄性不育系,属于粳稻品种农垦 58 的天然突变株,由石明松 1973 年在湖北沔阳沙湖的一季晚粳农垦 58 的大田中发现(石明松 1985)。近等基因系 NIL(MH)来自于以籼稻品种明恢 63 为亲本,农垦 58S 为轮回母本经多次回交和分子标记筛选获得的 BC₈F₂ 后代,其在 *pms1* 位点含有与明恢 63 相同的纯合基因型,其余背景与农垦 58S 一致。NIL(MH)与农垦 58S 的株型和生育期相似,但育性不随光周期长短

而改变，其与农垦 58S 在武汉长、短日照下的成熟期整株形态、小穗育性及花粉育性如图 8 所示。这两个亲本在研究中用作遗传转化的受体，及其他分子生物学分析。

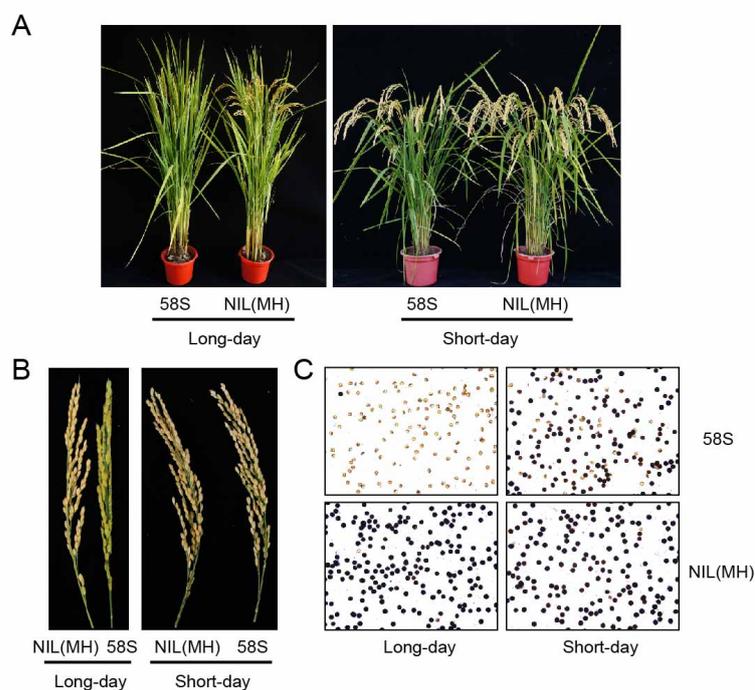


图 8 两个亲本在长、短日照下的表型。

农垦 58S 和 NIL(MH)在长短日照下的单株表型 (A)，小穗育型 (B) 和花粉育性 (C)。

Figure 8 The performance of Nongken58S and NIL(MH) under long-day and short-day conditions.

The whole plant (A), spikelet fertility (B) and pollen fertility (C) of Nongken 58S and NIL(MH) under long-day and short-day conditions.

根据已有文献报道(卢开阳等 1990)及本室多年种植经验，确定光敏感雄性不育材料的播种时间，分为两批，分别视为长日照和短日照条件。在武汉自然条件下，第一批于 4 月下旬至 5 月中旬播种，7 月下旬开始进入水稻幼穗分化阶段，8 月中旬开始到 9 月 3 日前抽穗，保证光敏感雄性不育材料对光长敏感的几个水稻幼穗分化时期都处于日照长度大于 14 h 的条件。6 月下旬进行第二批播种，8 月下旬开始幼穗分化，此时武汉地区日照时间逐渐低于 13.5 h，短于光敏雄性不育水稻的临界短日日长 13.75 h，9 月 20 日前开始抽穗。晚于 9 月 20 日抽穗的材料，往往因为低温，会对花粉育性和小穗育性造成一定的影响，不用于育性的考察和相关研究。

根据需要，部分材料于 12 月上旬在海南陵水播种以加代繁殖，3 月中下旬抽穗，光敏感雄性不育材料表现出部分可育。此地种植的材料仅收种或用作杂交的亲本，

不用于表型考察和性状研究。少量转基因 T₀ 代单株因为出苗时间的限制，恰好不处于水稻的正常生长时间，暂时种植于温室，待外界环境气温适宜后再移入大田。

2.2 定位群体的构建

本研究所用到的定位群体的原始亲本为农垦 58S 和明恢 63，是在本室於金生的工作基础上(於金生 2006)，进一步进行多次回交获得，具体构建示意图参见图 9。在前期工作中，於金生采用分子标记对 BC₁F₁ 进行了背景选择，去掉了另一个光敏感核不育基因 *pms3* 和籼粳杂种胚囊不育及广亲和位点 *S5* 的影响，并且与农垦 58S 回交获得了 BC₂F₁。我们将 BC₂F₃ 中分子标记显示 *pms1* 区段为明恢 63 纯合基因型的单株与农垦 58S 进行连续回交，得到的 BC₃F₁ 单株自交后获得了 BC₃F₂ 大群体，用于 *pms1* 的精细定位。同时，将 BC₃F₁ 继续回交，直至获得 BC₈F₁ 的单株，自交得到 BC₈F₂，进行分子标记分析，选取 *pms1* 区段为明恢 63 纯合基因型的植株，将其命名为 NIL(MH)，用于后续基因功能分析等。

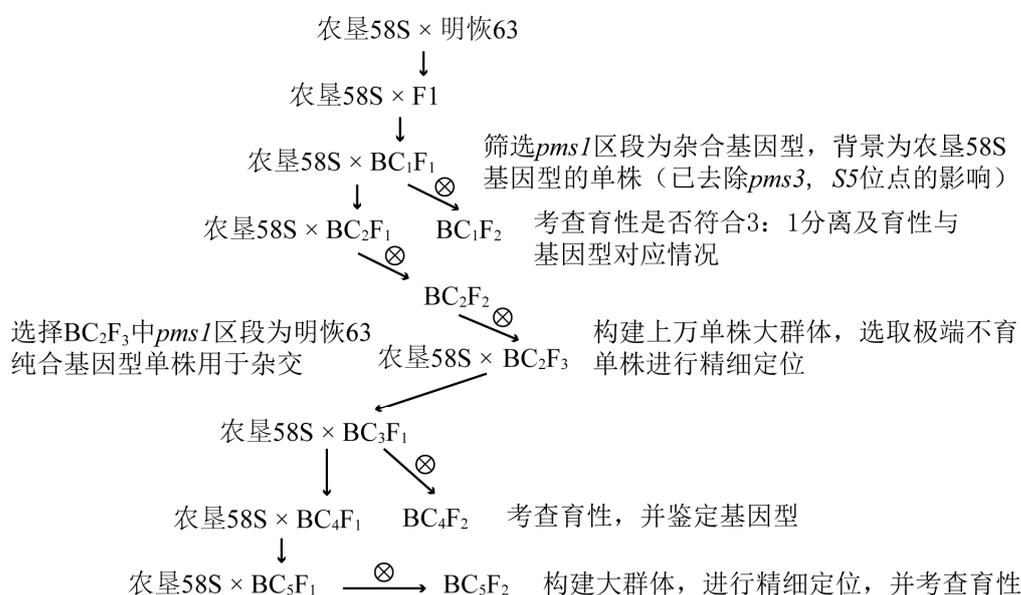


图 9 定位群体构建示意图。

Figure 9 A diagram of mapping population construction.

2.3 分子标记的开发

本研究主要针对明恢 63 与农垦 58S 的基因组序列差异进行分子标记的开发和利用。一方面利用 Gramene 数据库 (<http://www.gramene.org/>) 提供的 SSR 标记信息设计引物进行筛选。另一方面, 通过对农垦 58S 基因组 *pms1* 候选区段进行测序, 获得的序列与明恢 63 BAC 2109 (GenBank: DQ989628.1) 序列进行比对, 根据序列差异开发分子标记。优先选择两者间存在 10 bp 以上缺失的区段, 设计引物, 开发成 InDel 标记。针对单碱基差异, 将来自于两个基因组上包括单碱基及其左右两边各 30 bp 左右的序列在 dCAPS Finder 2.0 (<http://helix.wustl.edu/dcaps/dcaps.html>) 上进行分析, 系统会自动分析出该 SNP 位点能否开发成 CAPS 或者 dCAPS 标记, 同时提供备选引物信息。

2.4 转化载体及菌株

本研究构建的载体主要用于水稻愈伤组织转化。载体构建过程中使用的大肠杆菌菌株为 *E.coli* DH10B, 愈伤组织转化使用的农杆菌为 EHA105 菌株。电击转化所用到的大肠杆菌和农杆菌的感受态细胞制备方法参照本室《水稻生物学实验指南》(仅用于内部交流, 未公开发表) 进行, 热激转化用到的大肠杆菌感受态细胞来自于北京全式金生物技术有限公司相关试剂盒中提供的 Trans1-T1 Phage Resistant Chemically Competent Cell。

用于水稻愈伤组织转化的基本载体包括以下几种: 互补转化载体为双元载体 pCAMBIA1301, 来自澳大利亚 CAMBIA (Center for the Application of Molecular Biology to International Agriculture) 实验室; 超量表达载体 pCAMBIA1301S (pC1301S) 和 pU1301, 都是在载体 pCAMBIA1301 的基础上改造得到, 分别由 CaMV 35S 和玉米 Ubiquitin 启动子驱动表达, 其中 pC1301S 是周莹改造而成(Zhou et al 2009), pU1301 由邱德运改造获得(Qiu et al 2007); RNAi 抑制载体 pDS1301, 经本实验室袁斌改造得到(Yuan et al 2007), 基本骨架还是 pCAMBIA1301, 其携带有能表达 dsRNA 的构件来自于 pMCG161 载体(McGinnis et al 2005)。

2.5 遗传转化载体构建及转化

2.5.1 互补载体构建

本研究所用到的互补载体包括：58S-C，MH-C，58S-C-dP6，58S-ORF1+G，58S-ORF2+G 和 58S-ORF3+G。58S-C 和 MH-C 的外源片段分别来自于农垦 58S 和明恢 63 基因组相同的区段，经过 *Pst*I 和 *Sac*I 酶切后连到 pCAMBIA1301 上，但两者的构建方法略有不同，分别需要对农垦 58S 基因组进行 PCR 扩增和酶切明恢 63 BAC 克隆 2109（图 10）。

明恢 63 BAC 克隆 2109 从本室构建的明恢 63 BAC 文库中挑选得到(Peng et al 1998)，根据日本晴和 93-11 基因组序列设计引物进行测序(山燕 2000; 於金生 2006)，共获得长度为 107,663 bp 的序列（GenBank: DQ989628.1）。BAC 2109 经过 *Pst*I 和 *Sac*I 完全酶切后经过 1.0% TAE 琼脂糖凝胶电泳，将长度为 5,682 bp 的片段进行挖胶回收（DNA Gel Extraction Kit, Cat. No. k0513, Fermentas）（理论酶切片段长度参见图 10，实际酶切产物胶图见图 11），再与预先酶切好的 pCAMBIA1301 连接（T4 DNA Ligase, Cat. No. M0202T, NEB）。挑选酶切验证无误的克隆（图 10），再经过引物 pucF 和 pucR 进行末端测序验证后，命名为 MH-C，用于遗传转化。

构建互补载体 58S-C 时，由于没有农垦 58S 的 BAC 文库，只能根据农垦 58S 的基因组序列进行长片段的 PCR 扩增。PCR 反应体系如下：以农垦 58S 的大样 DNA 为模板 2 μ l，10 \times LA PCR buffer II (Mg^{2+} Free) 2 μ l，10 mM dNTPs 3.2 μ l，25 mM $MgCl_2$ 2 μ l，10 mM 引物（F/R）各 0.8 μ l，TaKaRa LA Taq（5 U/ μ l）0.8 μ l，加灭菌 ddH₂O 至 20 μ l。引物对为 58S-C-F 和 58S-C-R，见附录 I。采用两步法扩增，程序：94 $^{\circ}$ C 预变性 1 min；30 个 cycles；每个 cycle 94 $^{\circ}$ C 变性 20 sec，68 $^{\circ}$ C 复性延伸 6 min；最后 72 $^{\circ}$ C 延伸 7 min，降至室温。得到的 PCR 产物取 8 μ l 电泳检测，目标条带大小为 5,954 bp。余下产物经氯仿抽提后，加入预冷的 2 倍体积 95%乙醇沉淀后溶于灭菌 ddH₂O，再用 *Pst*I 和 *Sac*I 酶切，回收片段为 5,695 bp（图 11），再连接到 pCAMBIA1301 载体中。得到的阳性克隆通过测序与农垦 58S 序列比对无突变后，用于后续遗传转化。

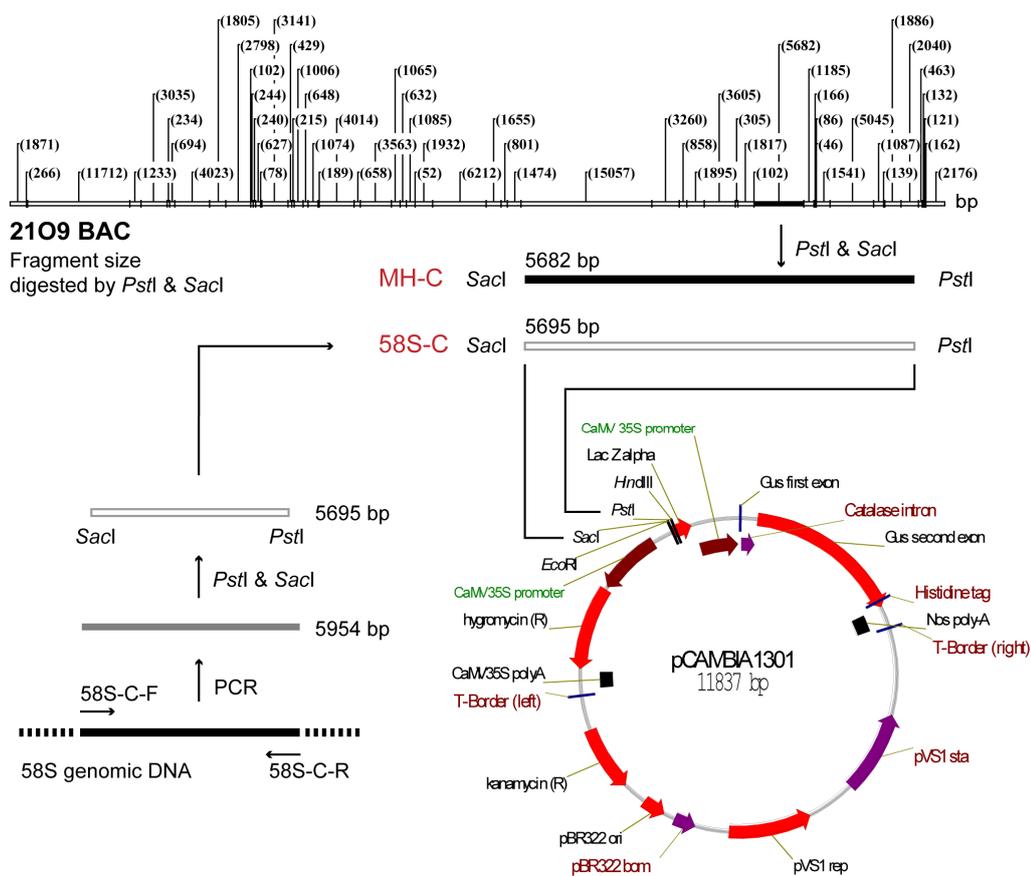


图 10 载体 58S-C 和 MH-C 构建示意图。

Figure 10 A procedure chart of 58S-C and MH-C vector constructions.

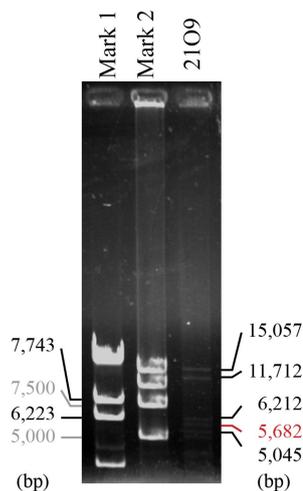


图 11 明恢 63 BAC 2109 经过 *PstI* 和 *SacI* 酶切后的胶图。

Mark 1 为 λ -*EcoT14 I* digest DNA Marker; Mark 2 为 DL15,000 DNA Marker; 红色标注为目标回收片段。

Figure 11 Gel result of Minghui 63 BAC 2109 digested with *PstI* and *SacI*.

Mark 1 is the λ -*EcoT14 I* digest DNA Marker, and Mark 2 is DL15,000 DNA Marker, the band used for extraction is marked in red.

58S-C-dP6 是经 58S-C 和 MH-C 改造而来, 构建流程图如图 12: 先用 *SpeI* 和 *SacI* 完全酶切 MH-C 和 58S-C 载体, 分别回收大小为 2,064 bp 和大约 15 kb 的片段, 再将这两段连接起来, 得到的 58S-C-dP6 载体与 58S-C 相比, 仅仅缺少了 P6 处的 65 bp 片段插入。

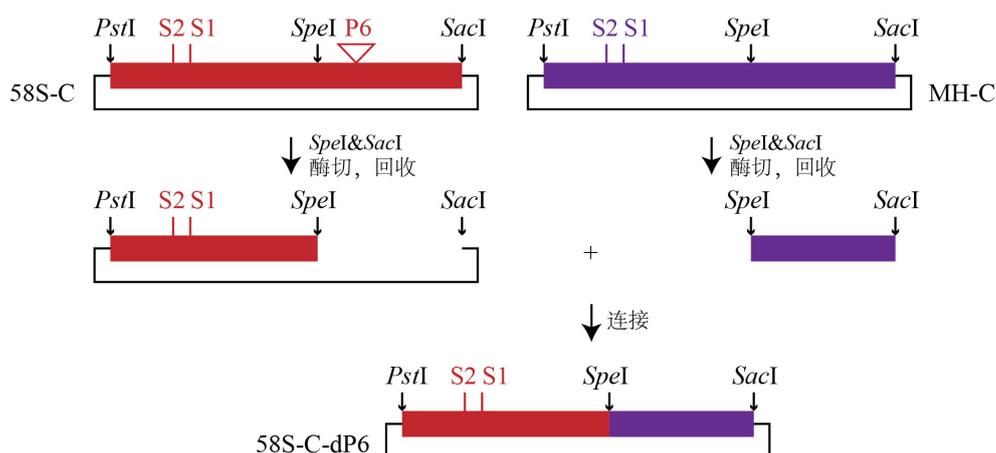


图 12 载体 58S-C-dP6 构建示意图。

Figure 12 A procedure chart of 58S-C-dP6 construction.

58S-ORF1+G, 58S-ORF2+G 和 58S-ORF3+G 也是以 58S-C 为骨架进行的改造。与 58S-C 相比, 这三个载体分别在预测的三个 ORF: ORF1, ORF2 和 ORF3 的起始密码子 ATG 的后面加入一个“G”, 通过在引物中引入突变获得。

58S-ORF1+G: 根据农垦 58S 的基因组序列, 设计可扩增 ORF1 的正向引物 ORF1G-muF (5'-CAAATCACAATAAAGATATATG^GGATCAAACCTTAATAAT-3') (下划线标明的字母为起始密码子“ATG”, 后面插入的一个“G”用斜体字母表示), 先与引物 ORF1G-R 扩增农垦 58S 的基因组序列, 得到 501 bp 长的片段; 同时用 ORF1G-muR 与 ORF1G-F 扩增获得 685 bp 长的片段。ORF1G-muF 与 ORF1G-muR 是完全反向互补的两条序列, 因此可以 501 bp 和 685 bp 的 PCR 产物为模板, 用引物 ORF1G-F 与 ORF1G-R 进行重叠延伸 PCR, 引物位置参见图 13 中 A 图。PCR 产物经过挖胶回收 1,144 bp, 连接到 T-vector 上面 (pGEM[®]-T Vector System, Cat. No. A3600, Promega), 通过酶切验证, 阳性克隆再用 *ApaI* 和 *SpeI* 双酶切, 回收 775 bp 的酶切片段, 与同样经过 *ApaI* 和 *SpeI* 双酶切的 58S-C 载体进行连接, 经过酶切和测序验证, 阳性克隆质粒命名为 58S-ORF1+G。测序验证的结果见图 14 的左图, 在 ATG 后面确实插入了一个碱基“G”。

58S-ORF2+G: 与 58S-ORF1+G 载体构建方法略有不同的是, *SpeI* 酶切位点正好位于 ORF2 的 ATG 后面, 相距仅 48 bp, 因此直接设计包含 *SpeI* 酶切位点和多一个“G”的长引物 ORF2G-muR (5'-AACTAGTATATAGTATATAGAGAAGCCCAA-GTACCGAACTTTTTCTGTTCGCCAT-3'), 与 ORF2G-F 进行 PCR 扩增 1,571 bp 长的片段。回收后同样经过 TA 克隆, 阳性质粒经 *ApaI* 和 *SpeI* 双酶切后与 58S-C 载体进行连接, 阳性克隆 58S-ORF2+G 的测序验证结果如图 14 的中图。引物位置及酶切位点见图 13B。

58S-ORF3+G: 构建方法与 58S-ORF1+G 载体的构建方法类似。设计了用于重叠延伸 PCR 的包含插入碱基“G”的正向引物 ORF3G-muF (5'-CCCATTGCCATCTCCGCCTAATGGAACGCCTTGCATGTTGG -3'), 以及与之先与之反向互补的反向引物 ORF3G-muR。将引物 ORF3G-muF 和 ORF3G-R 扩增农垦 58S 的基因组序列得到的 1,954 bp 片段, 与引物 ORF3G-muR 与 ORF3G-F 扩增获得的 349 bp 片段进行重叠延伸 PCR, 产物经过挖胶回收 2,261 bp 片段。该片段连到 pGEM-T Vector 后, 用 *SpeI* 和 *SacI* 进行双酶切, 酶切得到的 2,064 bp 片段再与经过同样双酶切的 58S-C 载体连接, 酶切、测序验证得到的阳性克隆 58S-ORF3+G 的测序验证结果如图 14 的右图。引物位置及酶切位点参见图 13 中 C 图。

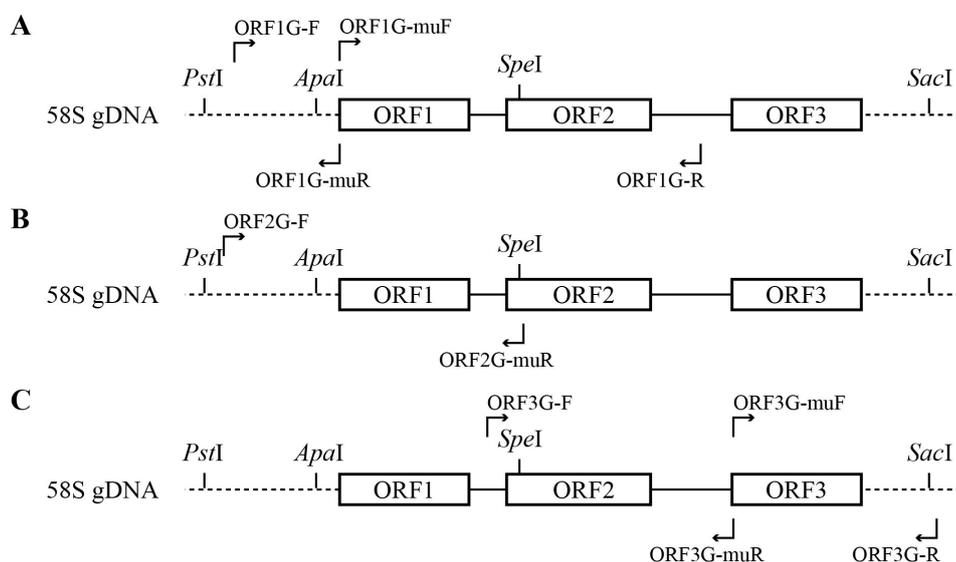


图 13 载体 58S-ORF1+G, 58S-ORF2+G 和 58S-ORF3+G 构建过程中使用的引物示意图。

Figure 13 The location of primers for 58S-ORF1+G, 58S-ORF2+G and 58S-ORF3+G vector construction.



图 14 载体 58S-ORF1+G, 58S-ORF2+G 和 58S-ORF3+G 阳性克隆测序结果。

Figure 14 The sequence results of 58S-ORF1+G, 58S-ORF2+G and 58S-ORF3+G positive clones.

2.5.2 抑制载体构建

本研究用到的抑制载体包括两种类型：双链 RNA 抑制以及人工 microRNA (artificial microRNA, amiRNA) 抑制，分别命名为 dsi 和 ami。两种类型载体的构建过程如下：

1, dsi 载体的构建。构建 dsi 时，先对要抑制表达的基因 *PMS1T* 进行序列分析，为避免脱靶，选择了 5'端特异性高，且在明恢 63 和农垦 58S 基因组序列无差异的 321 bp 区段设计引物，该区间位于 SNP S2 和 SNP S1 之间。引物 dsi-F 和 dsi-R 分别带有 *KpnI* 和 *SpeI*, *BamHI* 和 *SacI* 两个酶切位点。以农垦 58S 花粉母细胞形成期幼穗 RNA 反转录的 cDNA 为模板，使用这对引物进行 PCR 扩增，回收的 PCR 产物先经过 *KpnI* 与 *BamHI* 双酶切，再与经过同样双酶切的双链 RNA 抑制载体 pDS1301 连接，阳性克隆经过测序验证无误后得到第一链载体 dsi-sense。然后对 dsi-sense 使用 *SpeI* 和 *SacI* 进行双酶切，再与同样经过了 *SpeI* 和 *SacI* 双酶切的 dsi-F 与 dsi-R 扩增 PCR 产物进行连接，完成第二链的连接。测序验证后，得到的终载体命名为 dsi。其构建流程图参见图 15。

2, ami 载体的构建。在 WMD2 网站上 (Web MicroRNA designer, <http://wmd2.weigelworld.org/cgi-bin/mirnatools.pl>) 对可能抑制 *PMS1T* 功能的 amiRNA 进行预测，选取了其中最合适的两条 amiRNA 分别设计了 4 条引物。根据 WMD2 网站上提供的详细载体构建方法，以中间载体 pNW55 为模板，利用通用引物 G-4368

和 G-4369 以及 4 条引物，进行三段 PCR 产物的重叠延伸 PCR，得到的目标片段连接到 pGEM-T Vector 后，经过酶切和测序验证后，将阳性克隆经过 *KpnI* 和 *BamHI* 双酶切后的片段连接到经过相同酶切的超表达载体 pU1301 上，得到的终载体分别命名为 ami1, ami2。ami1, ami2 的靶序列分别位于 *PMSIT* 的 3' 和 5' 端，使用到的引物序列详见附录 I。

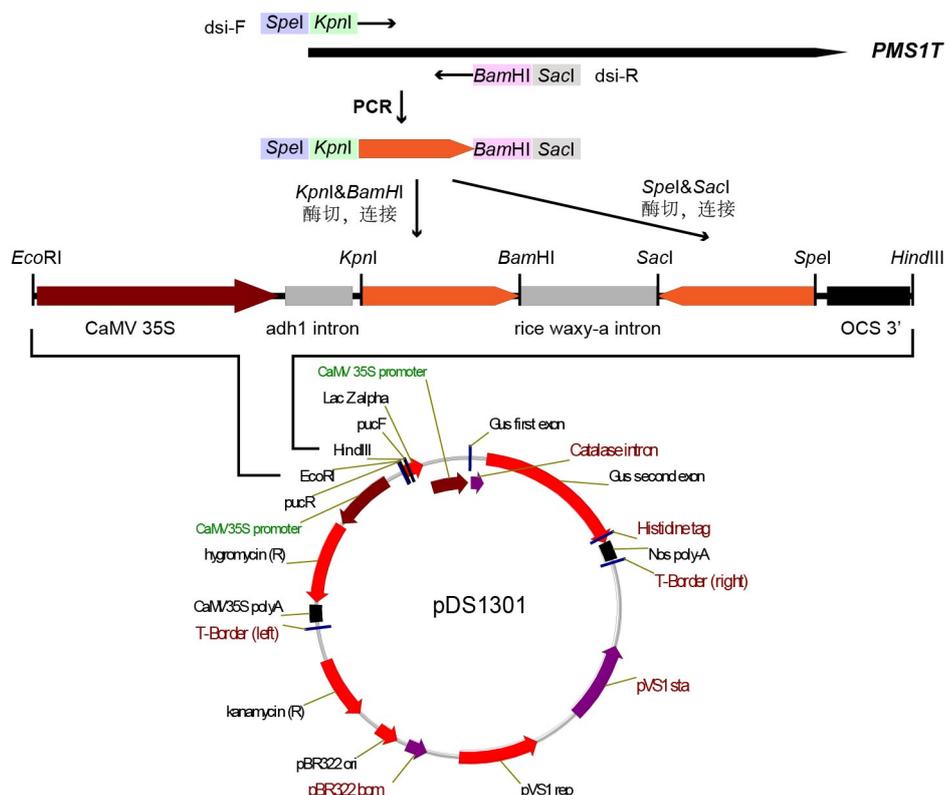


图 15 载体 dsi 构建示意图。

Figure 15 A procedure chart of dsi vector construction.

2.5.3 超表达载体构建

本研究所用到的超表达载体包括：35S:S, 35S:S-CS, 35S:M-CS, Ubi:S, Ubi:M, Ubi:CS+P6, Ubi:P6 和 Ubi:P6+polyA。35S:S, Ubi:S 和 Ubi:M 的外源片段都是通过引物 OX-1F 与 OX-1R 分别扩增农垦 58S 和明恢 63 基因组 DNA，得到包含 *PMSIT* 全长 cDNA 的序列，然后经过 *KpnI* 和 *BamHI* 双酶切，分别与经过同样酶切处理的 pC1301S 和 pU1301 空载体进行连接，获得的阳性克隆经过酶切、测序验证后，用于遗传转化。采用相同的构建方法，分别用引物 OX-1F 与 OX-3R, OX-3F 与 OX-3R, OX-3F 与 OX-1R 扩增农垦 58S 的基因组 DNA，经过 *KpnI* 和 *BamHI* 双酶切连接到

pU1301 空载体上, 得到 Ubi:CS+P6, Ubi:P6 和 Ubi:P6+polyA。35S:S-CS 和 35S:M-CS 的外源片段由引物 OX-2F 与 OX-2R 分别扩增农垦 58S 和明恢 63 基因组 DNA 得到, 再通过 *SacI* 和 *PstI* 双酶切后连到 pU1301S 载体上。这些超表达载体所包含的外源片段与 *PMS1T* 的相对位置如图 16 所示。

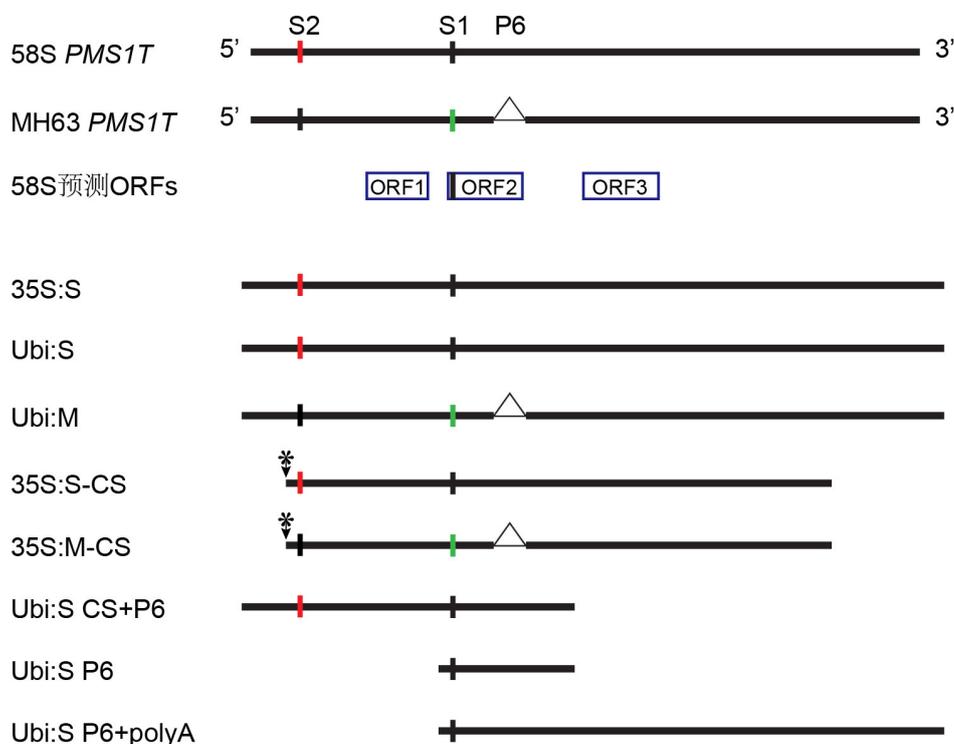


图 16 超表达载体外源片段示意图。

三角符号表明恢 63 在 P6 处缺失的 65 bp; 短竖线代表 SNP S2 和 S1, 红、黑和绿色分别为碱基 “T”、“G” 和 “A”; 带星号的箭头表示 miR2118 剪切位置。

Figure 16 The location of foreigner fragments in overexpression constructs.

Triangle marks the 65-bp deletion at P6 marker in MH63. Short vertical lines represent the SNP: red, T; black, G; green, A. Arrow with asterisk indicates the cleavage site of miR2118.

2.5.4 Target mimicry 载体构建

为了降低 miR2118d 和 miR2118o 的表达, 构建了 target mimicry 载体 MIM2118d 和 MIM2118o。其主要方法是通过重叠延伸 PCR 的方法将来自于拟南芥的非编码 RNA *IPSI* 基因中的 microRNA 模拟靶位点分别替换成 miR2118d 和 miR2118o 的模拟靶位点(Franco-Zorrilla et al 2007), 该模拟靶位点是在 miR2118d 和 miR2118o 的靶

基因的第 10 和 11 位碱基之间分别插入了额外的 3 bp,使其与 miRNA 不能完全配对,形成 3 bp 的 loop 结构。替换后的序列再通过 *KpnI* 和 *BamHI* 双酶切,连接到 pU1301 超表达载体上,经过酶切和测序验证,得到的阳性克隆命名为 MIM2118d 和 MIM2118o。

2.5.5 农杆菌介导的遗传转化

本研究中构建的遗传转化载体都是通过农杆菌介导的遗传转化方法转化水稻成熟胚诱导的愈伤组织。用到的水稻种子主要来自于农垦 58S 和 NIL(MH),都属于粳稻,转化方法、程序及培养基配方主要参照 Hiei 等(1994)和本室 Lin and Zhang(2005)建立的转化体系。

将新收获的成熟的水稻种子去壳后,去掉畸形、变色的种子,用 0.15%的升汞浸泡消毒 30 min 后,用灭菌的单蒸水反复涮洗,去掉残留的升汞,置于诱导培养基上,于 26℃暗培养 30-40 d 以诱导出愈伤组织。可挑选致密紧致、嫩黄色的愈伤组织在继代培养基上扩大培养一次,26℃暗培养两周后,用携带有目标载体的农杆菌菌株 EHA105 的菌液侵染该愈伤组织。经过 19℃共培养 2-3 d 以便完成农杆菌的侵染过程,然后用灭菌单蒸水多次洗涤去掉愈伤组织表面残留的农杆菌菌液,再置于添加了潮霉素和羧苄青霉素或者 G-418 的筛选培养基上,经过多次筛选获得抗性愈伤。抗性愈伤在分化培养基上 26℃光照培养 40 d 左右分化出苗,来自于同一颗愈伤组织的分化苗中挑选出一棵在生根培养基上生根,待长出强壮的根茎后,移栽到田间或者温室生长。这些转基因植株作为独立的转化单株用于后续基因型鉴定和表型观察。

2.6 水稻育性考察

本研究主要关注于水稻的育性变化,通过考察单株小穗育性和花粉育性来反映单株育性。所有用于育性考察的单株均于正季种植在武汉华中农业大学实验田(参见前文“田间种植”),温室或者海南的材料不用于育性考察。

单株小穗育性考察方法:水稻成熟期后选取单株的 3 个主穗,考察总的实粒数和空壳数,单株结实率即小穗育性为实粒数占实粒数和空壳数之和的百分比。在展

示家系中单株的结实率时，为了更客观、直观，主要采用 Weissgerber et al (2015)提供的散点作图法。

单株花粉育性观察：在水稻齐穗期于单株的 3-5 个主穗上分别选取 3-5 朵即将开花的颖花（雄蕊已到达颖壳的三分之二处），放置在 70%的酒精中，于 4℃短时间保存待用。观察花粉育性时，从中挑选 5-6 朵颖花，用镊子取出花药置于载玻片上，滴加 1-2 滴 1% 碘-碘化钾溶液，混合花药并捣碎，盖上盖玻片，在 Leica DM4000B 显微镜下观察，并拍照记录。

2.7 水稻总 DNA 抽提

根据实验目的不同，本研究采用两种方法抽提水稻总 DNA：大样法和小样法。两者基本原理相同，都是采用 CTAB 法抽提 DNA (Murray and Thompson 1980)。但大样法步骤繁琐，能获得高质量的总 DNA，一般用来进行 Southern blot 或者作为长片段 DNA 扩增的模板；小样法抽提水稻总 DNA 简单方便，虽然得到的 DNA 完整度不如大样法，但也能满足单株基因型检测、小片段 DNA 扩增等实验要求。

本研究使用的大样法抽提水稻总 DNA 在以往实验方法上经过改良，只需要少量水稻叶片。具体步骤如下：将 2-3 g 水稻幼嫩新鲜叶片用液氮研磨成粉末后放入 1.5 ml 离心管中，加入 500 μ l 已预热至沸腾的 1.5 \times CTAB，振荡器上混匀，65℃水浴 30 min，每隔 10 min 颠倒混匀几次后，加入等体积 24:1（氯仿:异戊醇），缓慢轻摇 20 min 后，12,000 r/min 离心 10 min。吸取上清 500 μ l，加入 1/10 体积的 56℃预热好的 10% CTAB 以及等体积的 24:1（氯仿:异戊醇），缓慢轻摇 15 min 后，12,000 r/min 离心 15 min。吸取上清 400 μ l，用已剪口的枪头加入 2 倍体积的 1% CTAB，轻轻旋转离心管，可看到 DNA 絮状成团。10,000 r/min 离心 5 min，倒掉上清，将离心管倒扣在吸水纸上，沥干 DNA 沉淀。加入 5 μ l RNase 和 300 μ l 1 M NaCl，56℃水浴过夜溶解 DNA。再加入 1 ml 预冷的 95%乙醇，轻轻旋转离心管，得到 DNA 絮状沉淀后 12,000 r/min 离心 3 min，倒掉上清，用 75%乙醇洗沉淀，超净台上吹干沉淀至透明后，加入 20 μ l TE 溶解，保存于-20℃。

2.8 水稻材料基因型检测与转基因植株 Southern blot 拷贝数分析

水稻材料基因型的检测主要采用 PCR 的方法，一般通用的反应体系为 15 μl ，包括：DNA 模板 2 μl ，10 \times buffer (Mg²⁺ plus) 1.5 μl ，10 mM dNTPs 0.3 μl ，10 μM 引物 (F/R) 各 0.2 μl ，Taq 聚合酶 (5 U/ μl) 0.2 μl ，加灭菌 ddH₂O 补至 15 μl 。PCR 扩增程序为三步法：94 $^{\circ}\text{C}$ 预变性 3 min；30-40 个 cycles，每个 cycle 94 $^{\circ}\text{C}$ 变性 30 sec，55-60 $^{\circ}\text{C}$ 复性 30 sec，72 $^{\circ}\text{C}$ 延伸（延伸时间根据 PCR 产物大小而定，一般 1 min/kb）；最后 72 $^{\circ}\text{C}$ 延伸 7 min，降至室温，于 4 $^{\circ}\text{C}$ 保存 PCR 产物。PCR 产物加入 10 \times loading buffer 后，进行适当浓度（1.0%-3.0%）的琼脂糖凝胶电泳，或者 6% 的变性 PAGE 胶电泳。

SSR 标记分析方法参见喻辉辉博士论文(喻辉辉 2011)。CAPS 和 dCAPS 标记检测过程中的酶切反应体系一般为 10 μl ，包括：PCR 产物 5 μl ，10 \times 酶切 buffer 1 μl ，限制性内切酶 3 Unit，加灭菌 ddH₂O 补至 10 μl ，37 $^{\circ}\text{C}$ 培养箱中酶切 2 小时或过夜。酶切产物加入 10 \times loading buffer 终止反应后，于 2.0% 的琼脂糖凝胶电泳检测。

对转基因植株进行阳性检测时，一般根据转化片段和转基因受体选择基因特异标记进行鉴定，互补转化和超表达转化单株鉴定最常用的是 InDel 分子标记 P6：相对于农垦 58S 的 *PMSIT*，明恢 63 中缺失了 65 bp；如果两者在 P6 处不存在差异，就针对载体上的报告基因 *GUS* 采用组织染色(Jefferson et al 1987)或者用引物对 GUS1.6F/GUS1.6F 进行 PCR 扩增的方法检测转基因阴/阳性。dsi 和 ami 转基因单株分别用引物对 PMCGF/PMCGR, G-11491/G-11494 进行 PCR 阳性检测。Target mimicry 载体 MIM2118d 和 MIM2118o 的转基因 T₀ 代单株采用 MIM-F/MIM-R 进行 PCR 阳性检测；T₁ 代单株则需要对 PCR 产物进行测序，根据额外插入的 3 bp 序列的测序结果判断转基因阳性单株的基因型是纯合还是杂合。

Southern blot 用来检测转基因植株的拷贝数，DNA 的抽提参照前面的改良的大样法完成，DNA 酶切、电泳、搭盐桥转膜及放射性同位素标记探针、杂交和洗膜等具体实验流程参照於金生博士论文(於金生 2006)进行。本研究中对总 DNA 进行酶切的体系优化为：总 DNA 4 μg ，10 \times M buffer 2 μl ，限制性内切酶 *Hind*III (Takara, 15 U/ μl) 1.2 μl ，spermidine (40 mmol/L) 1 μl ，加 ddH₂O 补至 20 μl ，37 $^{\circ}\text{C}$ 酶切过夜。用于放射性同位素标记的探针模板来自于 *GUS* 基因，采用 GUS1.6F/GUS1.6F 扩增

pCAMBIA1301 质粒得到的 1.57 kb 长的片段，经过挖胶回收、测浓度后保存于-20℃备用。杂交后的膜经过洗膜后，放置于磷屏（Fujifilm, BSA Cassette2 2025）中，压一天后，于磷屏扫描仪（Fujifilm, FLA-5100）上进行图片扫描和保存。

2.9 水稻总 RNA 抽提

根据不同的实验目的，本研究分别采用 TransZol（北京全式金，Cat. No. ET101）和 TRIzol 试剂盒（Invitrogen, Cat. No. 15596-026）抽提水稻组织总 RNA，两者步骤基本相同，但采用 Invitrogen 公司试剂盒得到的 RNA 质量更高一些。

某些实验对总 RNA 的 OD_{260/230} 比值有要求，必须大于 2.0，可通过重复一次氯仿抽提和 75%乙醇多次洗沉淀的方法降低其他化学物质的污染，提高样品的纯度。

2.10 RACE 与 5' RLM-RACE

RACE（Rapid Amplification of cDNA Ends）主要用来确定基因转录的 5'和 3'末端。本研究主要使用 SMARTTM RACE cDNA Amplification Kit（Clontech, Cat. No. 634914）分离 *PMSIT* 基因的全长 cDNA。分别以 1 μg 总 RNA 为模板，采用试剂盒中提供的 5'-CDS primer A，SMART II A oligo 和 3'-CDS primer A，根据说明书提供的方法进行 first-strand cDNA synthesis，得到 5'-RACE-Ready cDNA 和 3'-RACE-Ready cDNA。以这些 cDNA 为模板分别进行两轮的巢式 PCR，即 Rapid Amplification of cDNA Ends。5'-RACE PCR 和 3'-RACE PCR 使用共同的 UPM（Universal Primer）和 NUP（Nested Universal Primer A）引物，GSP（gene specific primer）和 NGSP（nested gene specific primer）引物根据基因特异序列设计。第一轮 PCR 的反应体系为 50 μl，组分如下：5'-RACE-Ready cDNA 或者 3'-RACE-Ready cDNA 2.5 μl，10 × UPM 5 μl，10 μM GSP 1 μl，10 mM dNTPs 1 μl，10 × LA buffer 5 μl，LA Taq（TaKaRa 5 U/μl）0.5 μl，ddH₂O 35 μl。PCR 反应程序为：94℃ 预变性 3 min；32 cycles，每个 cycle 94℃ 30 sec，55℃ 30 sec，72℃ 1 min；72℃ 7 min，降至 25℃，于 4℃ 保存 PCR 产物。PCR 产物稀释 50 倍后作为第二轮 PCR 的模板。第二轮 PCR 反应体系如下：模板 2 μl，10 × NUP 1 μl，10 μM NGSP 1 μl，10 mM dNTPs 1 μl，10 × LA buffer 5 μl，LA Taq（TaKaRa 5 U/μl）0.5 μl，ddH₂O 补至终体系 50 μl。PCR

反应程序与第一轮相同。PCR 产物取 5 μ l 用 2.0%琼脂糖凝胶电泳，根据电泳结果，如果是单条带，可直接将 PCR 产物经过酶消化后进行测序；多条带需要将对应的条带挖胶回收，连接到 T-vector 上后再挑选阳性克隆测序。PCR 产物酶消化体系如下：PCR 产物 5 μ l, 10 \times PCR buffer 0.3 μ l, ExoI (NEB 20 U/ μ l) 0.25 μ l, SAP (TaKaRa 1 U/ μ l) 0.26 μ l, ddH₂O 补至 8 μ l。过程为：37 $^{\circ}$ C 孵育 1 hr 后，80 $^{\circ}$ C 反应 20 min 终止反应。

本研究用到的 5' RLM-RACE (RNA Ligase Mediated Rapid Amplification of cDNA Ends) 是经过修改后的 5' RLM-RACE。根据 Llave et al (2002)报道的方法，按照 FirstChoice[®] RLM RACE Kit (Ambion, Cat. No. AM1700)的说明书进行 5' RLM-RACE 时，省略掉通过一系列酶的预处理以在 mRNA 5'帽子上添加 adapter 的过程，而是直接在 mRNA 暴露出来的 5'磷酸基团上添加 adapter，再进行反转录。利用该方法获得的 cDNA 5'末端序列可用来验证 microRNA 介导的对靶基因进行剪切的位点。5' RACE Adapter ligation 和反转录的过程参照试剂盒说明书完成，5' RLM-RACE PCR 也是两轮的巢式 PCR。第一轮 PCR 的反应体系如下：反转录产物 1 μ l, 10 \times HiFi buffer 5 μ l, 10 mM dNTPs 5 μ l, 5' RACE Outer primer 2 μ l, 10 μ M 5' RACE gene-specific outer primer 2 μ l, HiFi Taq (全式金, 1.25 U/ μ l) 0.3 μ l, ddH₂O 补至 50 μ l。PCR 反应程序采用 Touchdown 的方法，具体为：94 $^{\circ}$ C 预变性 3 min；第一步循环为 20 cycles, 起始 cycle 94 $^{\circ}$ C 30 sec, 60 $^{\circ}$ C 30 sec, 72 $^{\circ}$ C 30 sec, 之后每经过一个循环，退火温度下降 0.5 $^{\circ}$ C；第二步循环为 15 cycles, 起始 cycle 94 $^{\circ}$ C 30 sec, 50 $^{\circ}$ C 30 sec, 72 $^{\circ}$ C 30 sec, 之后每经过一个循环，退火温度下降 0.5 $^{\circ}$ C；72 $^{\circ}$ C 7 min, 降至 25 $^{\circ}$ C, 4 $^{\circ}$ C 保存 PCR 产物。PCR 产物取 1 μ l 作为第二轮 PCR 的模板。第二轮 PCR 反应体系如下：模板 1 μ l, 10 \times HiFi buffer 5 μ l, 10 mM dNTPs 5 μ l, 5' RACE Inner primer 1 μ l, 10 μ M 5' RACE gene-specific inner primer 1 μ l, HiFi Taq (全式金, 1.25 U/ μ l) 0.3 μ l, ddH₂O 补至 50 μ l。PCR 反应程序与第一轮相同。得到的 PCR 产物同样经过挖胶回收和测序后与基因序列比对，确定是否在预测位点发生剪切。

2.11 基因表达量分析与 small RNA 表达量分析

2.11.1 RT-PCR 与 qPCR

本研究中采用 RT-PCR 半定量与 qPCR 相对定量方法对基因进行表达量分析。水稻总 RNA 经过 Nanodrop 2000 测定 OD260/280 比值, 确认无降解后, 用于反转录。反转录主要步骤如下:

1, 去除基因组 DNA 污染。在 1.5 ml RNase-free 离心管中加入总 RNA 5 μg , 10 \times DNase I Reaction Buffer 1 μl , DNase I (Invitrogen, Cat. No. 18068-015, 1 U/ μl) 1 μl , 再加入 DEPC-treated ddH₂O 至终体系为 10 μl 。室温孵育 15 min, 加入 25 mM EDTA 1 μl , 65 $^{\circ}\text{C}$ 水浴 10 min, 冰上放置 5 min 以上。

2, First-strand cDNA synthesis。在上述反应物中加入 1 μl Oligo(dT)₁₅ (500 ng/ μl , Promega), 1 μl 10 mM dNTP, 混匀后 65 $^{\circ}\text{C}$ 反应 5 min, 迅速置冰上至少 2 min。再加入以下组分: 5 \times First Strand Buffer 4 μl , 0.1 M DTT 2 μl , SuperScriptTM III Reverse Transcriptase (Invitrogen, Cat. No. 18080-044, 200 U/ μl) 1 μl 。混匀后 50 $^{\circ}\text{C}$ 反应 60-90 min, 70 $^{\circ}\text{C}$ 水浴 15 min 灭活。最后再加入 80 μl ddH₂O, 取 1 μl 做 PCR 检测反转录的效率, 其余的保存于 -20 $^{\circ}\text{C}$ 待用。

使用一对跨 intron 的扩增看家基因 *actin 1* (LOC_Os03g50885) 的引物对 actin-inF/actin-inR 检测 RNA 样品的反转录效率。actin-inF/actin-inR 扩增基因组 DNA 和 cDNA 的长度分别是 459 bp 和 210 bp, 在 2.0% 琼脂糖凝胶上能够清晰地区分两条带。所有反转录得到的 cDNA, 先使用该引物对扩增后, 确信无基因组 DNA 污染, 并且能够得到明亮的 210 bp 条带, 说明反转录成功后, 再用于下一步的 RT-PCR 或者 qPCR。

RT-PCR 和 qPCR 的内参基因均为水稻 *ubiquitin* 基因 (*UBQ*, LOC_Os03g13170), 使用引物对 UBQ-F/UBQ-R 扩增。针对 *PMSIT* 基因设计了多对引物进行 qPCR, 最终根据引物的标准曲线和溶解曲线, 确定了最优引物对 RT-F/RT-R。

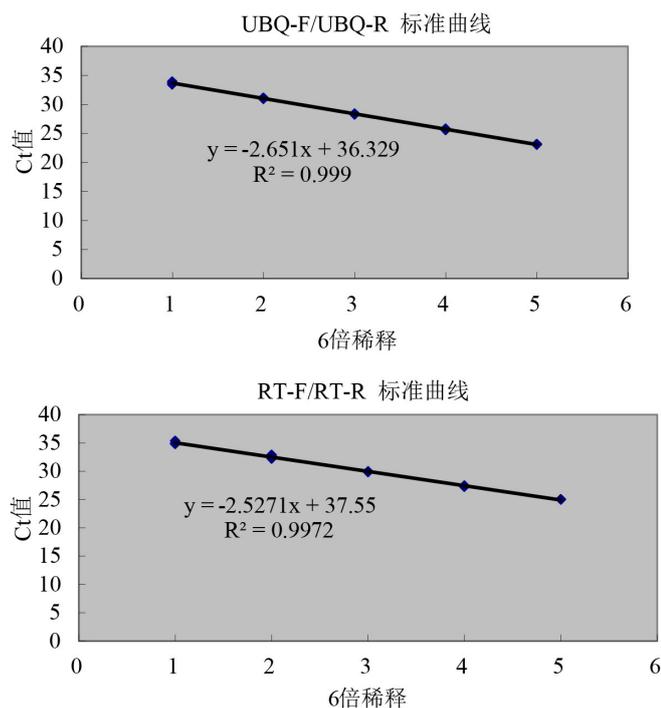


图 17 引物标准曲线图。

上图是引物对 UBQ-F/UBQ-R 的标准曲线，下图是 RT-F/RT-R 的标准曲线。

Figure 17 Standard curve of qPCR primer.

The upper image is the standard curve of primer UBQ-F/UBQ-R, and the lower image is that of RT-F/RT-R.

在进行 qPCR 之前，对于没有使用过的引物，要求做一次相对定量的标准曲线，以计算出引物的扩增效率。理论上引物的扩增效率是 100%，但实际上在 90%-110% 之间都是可以接受的，同时要求标准曲线的 $R^2 > 0.980$ 。本研究中，由于 *PMS1T* 基因表达量较低，我们对起始 cDNA 进行 6 倍的浓度梯度稀释，每个稀释样品重复三次。其中 UBQ-F/UBQ-R 和 RT-F/RT-R 的标准曲线图如图 17，计算得到的引物扩增效率分别为 97% 和 103%，可信度 R^2 分别为 0.999 和 0.9972。

qPCR 的反应采用 Roche 的相关试剂 (Cat. No. 04913850001) 完成，体系如下：cDNA 1-2 μ l, FastStart Universal SYBR Green Master (ROX) 5 μ l, 30 μ M forward primer 0.3 μ l, 30 μ M reverse primer 0.3 μ l, 加 ddH₂O 至终体系为 10 μ l, 在 Applied Biosystems ViiA 7 定量 PCR 仪上完成反应，程序为：95 $^{\circ}$ C 10 min; 45 个 cycles, 每个 cycle 95 $^{\circ}$ C 15 sec, 60 $^{\circ}$ C 30 sec; 另加上溶解曲线程序。反应完成后，收集每个样品的 Ct 值，根据文献报道的相对定量方法(Livak and Schmittgen 2001)计算基因的相对表达量。

2. 11. 2 Stem-loop RT-PCR 与 page-northern blot

分析 small RNA 的表达量时，主要采用 stem-loop RT-PCR 和 page-northern blot 进行相对定量和定性分析。

Page-northern blot 与 northern blot 原理相似，但分辨率更高，适用于小分子量 RNA 的杂交，是目前在体内检测 small RNA 的有效和最可信的方法。本研究使用的 page-northern blot protocol 主要参照威益军实验室，包括小分子量 RNA 的分离和纯化、跑胶、转膜、预杂交与杂交等步骤，但摒弃了传统的同位素标记的方法，改用生物素标记的探针进行杂交，并且针对文献提供的方法(Huang et al 2014)进行了改进。后期显影采用 Chemiluminescent Nucleic Acid Detection Module Kit (Thermo Scientific, Cat. No. 89880) 完成，blot 的曝光和图像采集在 Bio-Rad 化学发光成像系统 ChemiDOC XRS+ 中进行，条带亮度的相对定量使用 Image Lab 软件分析。

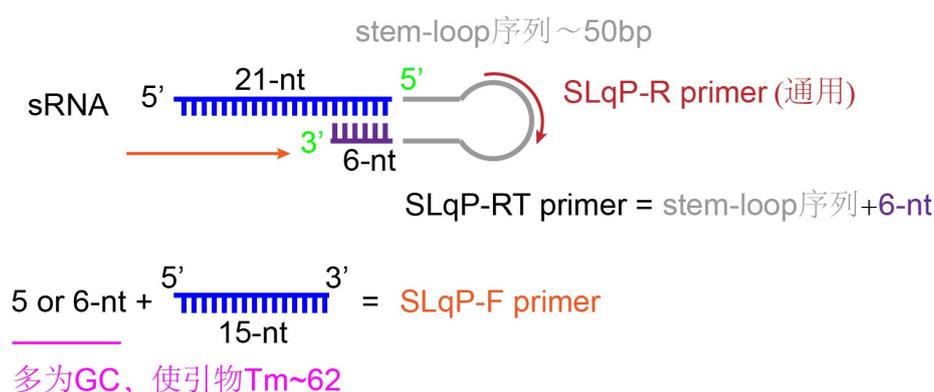


图 18 Stem-loop RT-PCR 引物设计方法 (以 21-nt small RNA 为例)。

Figure 18 Design method for stem-loop RT-PCR primers (take 21-nt small RNA as an example).

Stem-loop RT-PCR 可定量检测 small RNA 的表达量，灵敏度和便捷度都高于 page-northern blot，但因为需要经过体外 PCR 扩增的过程，可信度不如 page-northern blot。Stem-loop RT-PCR 的引物设计有别于普通 PCR，主要原理见图 18。以 21-nt 的 small RNA 为例，其中 SLqP-RT primer 用于反转录，为一段通用茎环序列 (5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGAC-3') 再加上 6-nt 与 small RNA 3' 末端互补的序列，总长度为 50 bp。SLqP-F 和 SLqP-R primer 用于 qPCR 反应，SLqP-F 引物的是由 small RNA 的前 15 个碱基再在 5' 端加上 5-6 个 G/C，

使整条引物最终的 T_m 值在 62 度左右。SLqP-R primer 是根据通用茎环序列设计的一段序列,所有 small RNA 均可以通用,序列为 5'-CCAGTGCAGGGTCCGAGGT-3'。Stem-loop RT-PCR 所使用的内参基因是 U6, 用作 qPCR 反应的引物 U6-R 同时也用来进行反转录。本研究中使用的一步法 Stem-loop RT-PCR 主要参照已有文献报道 (Chen et al 2005; Varkonyi-Gasic et al 2007; Shen et al 2010) 进行, 包括反转录和 qPCR 两大步骤。反转录的过程与普通 RT-PCR 的反转录方法相似, 但是用 small RNA 特异的引物代替 Oligo(dT)₁₅, 相应的第一链 cDNA 合成条件也进行了优化。

2.11.3 RPA (Ribonuclease Protection Assay)

RPA, 核糖核酸酶保护实验, 与 Northern blot 相同, 都是用来对体内 mRNA 进行检测和定量。相较于 Northern blot, 其分辨率更高, 可用于检测低丰度的 mRNA, 并且对 RNA 的完整度要求不高。其主要原理如下: 通过同位素或者非同位素的方法合成与目标 RNA 互补的一段序列, 这段标记好的探针与模板 RNA 在适宜的条件下进行杂交, 杂交后的混合物用核糖核酸酶处理, 以去除未杂交上的探针。同时, 在核糖核酸酶处理过程中, 因为探针与目标 RNA 的序列互补, 这段互补的序列会被保护, 不能被核糖核酸酶降解, 因此, 在聚丙烯酰胺变性胶上通过 PAGE (polyacrylamide gel electrophoresis, 聚丙烯酰胺凝胶电泳), 能够将目标 RNA 的全长序列与被保护的序列区分开。本研究的 RPA 实验按照 RPA IIITM Kit (Ambion, Cat. No. AM1415) 提供的说明书完成, 其中探针的标记使用生物素标记, 利用 MAXIscript[®] SP6/T7 Transcription Kit (Ambion, Cat. No. AM1320) 和 10 mM Bio-16-UTP (Invitrogen, Cat. No. AM8452) 进行体外转录实现。

2.12 Small RNA-seq 及 PARE 文库构建

Small RNA-seq 的建库在本室完成, 根据 illumina 的 TruSeq[®] Small RNA Sample Preparation Guide 完成, 相关试剂进行了减半使用。为了确保 RNA 的完整性, 除了用 Nanodrop 2000 测定 OD260/280 比值外, 还采用 Bio-Rad 的 The Experion 全自动电泳系统检测 RNA 的质量, 用于 small RNA-seq 建库的 RNA 要求其 RQI 值不低于 8.0。同时, 为了保证所有建库样品的总 RNA 量一致, 都为 3 μ g, 还进一步使用 Qubit[®]

2.0 荧光计 (Invitrogen) 对 RNA 进行精准定量。建库完成后, 利用本室测序平台的 Hi-seq 2000 (Illumina) 高通量测序仪进行测序和数据的收集。

PARE (Parallel Analysis of RNA Ends, RNA 末端平行分析法) 是基于降解组 (degradome) 进行的测序。文库的构建和数据分析委托美国特拉华大学 Blake 实验室根据已发表文献(Zhai et al 2014)完成。

2.13 DNA 甲基化分析

本研究采用 EpiTect Bisulfite Kit (QIAGEN, Cat. No. 59104) 进行 DNA 的甲基化修饰分析, 主要原理是: 真核生物中 DNA 的甲基化仅发生在胞嘧啶, DNA 变性后进行重亚硫酸钠处理, 会将未发生甲基化的胞嘧啶脱氨基变成尿嘧啶, 而发生了甲基化的胞嘧啶不会改变, 通过比较处理后和未处理的 DNA 序列, 就能找到发生了甲基化的位点。实验步骤参照试剂盒提供的 EpiTect[®] Bisulfite handbook 完成。小样法抽提的 DNA 就能满足实验需求, 但事先需要经过纯化, 即用等体积 24:1 (氯仿:异戊醇) 抽提一次后, 再加入 2 倍体积 95%乙醇和 1/10 体积的 NaAc 沉淀 DNA。

经过重亚硫酸钠处理后的 DNA, 还需要通过 PCR 扩增, 对目标 DNA 区段进行测序。PCR 扩增所使用的引物使用 MethPrimer 在线设计 (<http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi>), 一般采用多对引物重叠覆盖基因组目标区段, 引物序列详见附录 I。PCR 的反应体系如下: DNA 模板 1 μ l, 10 \times buffer (Mg²⁺ plus) 2 μ l, 10 mM dNTPs 0.4 μ l, 10 μ M 引物 (F/R) 各 0.4 μ l, rTaq (Takara, 5 U/ μ l) 0.2 μ l, 加灭菌 ddH₂O 补至 20 μ l。PCR 反应程序采用 Touchdown, 具体为: 94 $^{\circ}$ C 预变性 3 min; 第一步循环为 20 cycles, 起始 cycle 94 $^{\circ}$ C 30 sec, 60 $^{\circ}$ C 40 sec, 72 $^{\circ}$ C 2 min, 之后每经过一个循环, 褪火温度下降 0.5 $^{\circ}$ C; 第二步循环也是 20 cycles, 起始 cycle 94 $^{\circ}$ C 30 sec, 50 $^{\circ}$ C 40 sec, 72 $^{\circ}$ C 2 min, 之后每经过一个循环, 褪火温度下降 0.5 $^{\circ}$ C; 72 $^{\circ}$ C 7 min, 25 $^{\circ}$ C 10 min 结束反应。PCR 产物通过挖胶回收, 并且连接到 pGEM-T Vector 后, 用 M13F/M13R 引物对克隆进行菌落 PCR 以挑选阳性克隆。一般挑选 30 个左右的阳性独立克隆, 抽提质粒后用 T7 或者 SP6 引物进行测序, 获得的序列与原始序列一起在专门的网站 Kismeth 上进行分析 (<http://katahdin.mssm.edu/kismeth>), 计算序列上每个胞嘧啶发生甲基化的概率。

3 结果与分析

3.1 前人定位结果的验证

本室对水稻光敏感雄性核不育基因 *pms1* 的精细定位结果最早见于 2001 年(Liu et al 2001b), 刘南(2000)利用籼型不育组合 32001S/明恢 63 将 *pms1* 定位于两个分子标记 Fssr 和 Rssr 之间 85 kb 范围内。山燕(2000)和李亚芳(2002)构建了覆盖该区间的 6 个互补载体, 并将其转化到农垦 58S 中。於金生(2006)考察了这些转基因植株及其后代的表型, 进行了共分离检测, 结果发现这 6 个亚克隆片段都不能恢复长日照下农垦 58S 的育型。于是, 重新开始改用农垦 58S/明恢 63 组合对 *pms1* 进行定位, 最终将 *pms1* 定位于分子标记 Rssr 与 N6B5 之间 21 kb 范围内, 并且构建了覆盖此区间的来自于明恢 63 基因组外源片段的 5 个互补载体: y10、ulp1、y12、y8 和 J033, 同时还针对 ulp1 片段上的候选基因 ulp1 protease-like protein 和 j033 上的转录本构建了 RNAi 抑制载体: u1, u2 和 j1, j2。前人定位区间及载体示意图如图 19 所示。

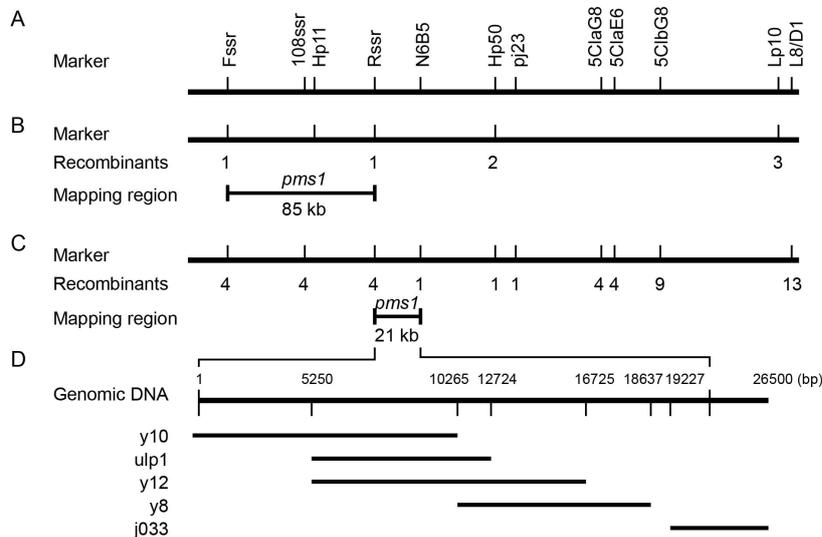


图 19 前人定位 *pms1* 结果。

A, 分子标记位置; B, 刘南(2000)对 *pms1* 进行精细定位的结果; C, 於金生(2006)对 *pms1* 进行精细定位的结果; D, 於金生(2006)构建的覆盖 21 kb 定位区间的 5 个互补载体示意图。

Figure 19 Previous results of *pms1* map-based cloning.

A, Position of molecular markers; B, Map-based cloning of *pms1* from Liu (2000); C, Map-based cloning of *pms1* from Yu (2000); D, Five overlapping complementary constructs constructed by Yu (2000) covering the 21 kb region.

本研究起始工作是先对前人的定位结果进行验证，基于前期对 *pms1* 遗传结果的认识，认为农垦 58S 中的 *pms1* 为隐性核不育基因，因此将 5 个互补载体分别转化到农垦 58S 愈伤组织中，分别得到转基因 T₀ 代植株：y10 150 株、ulp1 181 株、y12 98 株、y8 187 株和 j033 146 株。RNAi 抑制载体转化农垦 58S/明恢 63 组合 BC₂F₄ 中明恢 63 纯合基因型植株种子愈伤组织，得到转基因 T₀ 代植株：u1 46 株，u2 85 株和 j1 30 株，j2 61 株。将这些转基因植株于武汉大田长日照下种植，对这些单株进行了基因型的考察，并且检测了互补转化植株的拷贝数（图 20）。结果发现 RNAi 抑制转基因阳性单株的育性并没有下降，而部分互补转化单株的育性有所恢复，考虑到 T₀ 代转基因植株的育性可能会受到组培的影响，为了减少这种影响，从 T₀ 代单株中选择部分单株种植 T₁ 代进行共分离检测。y10、ulp1、y12、y8 和 j033 种植的 T₁ 代家系数目分别为 29、60、24、44 和 35，每家系 30 株。在武汉长日照下，所有这些种植的转基因互补家系中，有且仅有 y10 的 3 个 T₁ 代转基因家系中出现了育性恢复的单株。对这 3 个家系的 T₁ 代单株进行拷贝数检测和育性考察。

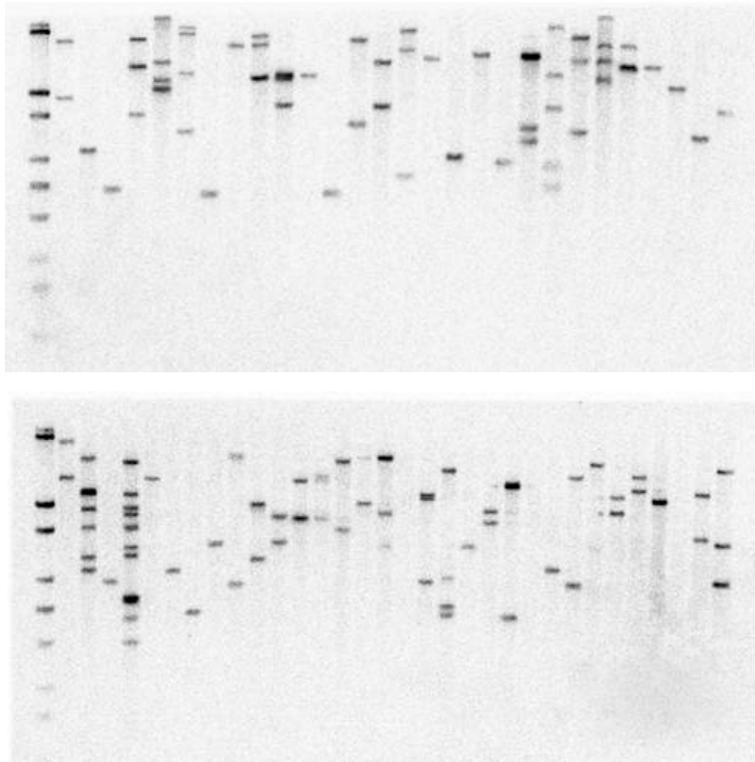


图 20 转基因 T₀ 代单株拷贝数检测。

最左边的孔为 λ -EcoT14 I digest DNA Marker

Figure 20 Southern blot of transgenic T₀ plants.

The left line is λ -EcoT14 I digest DNA Marker

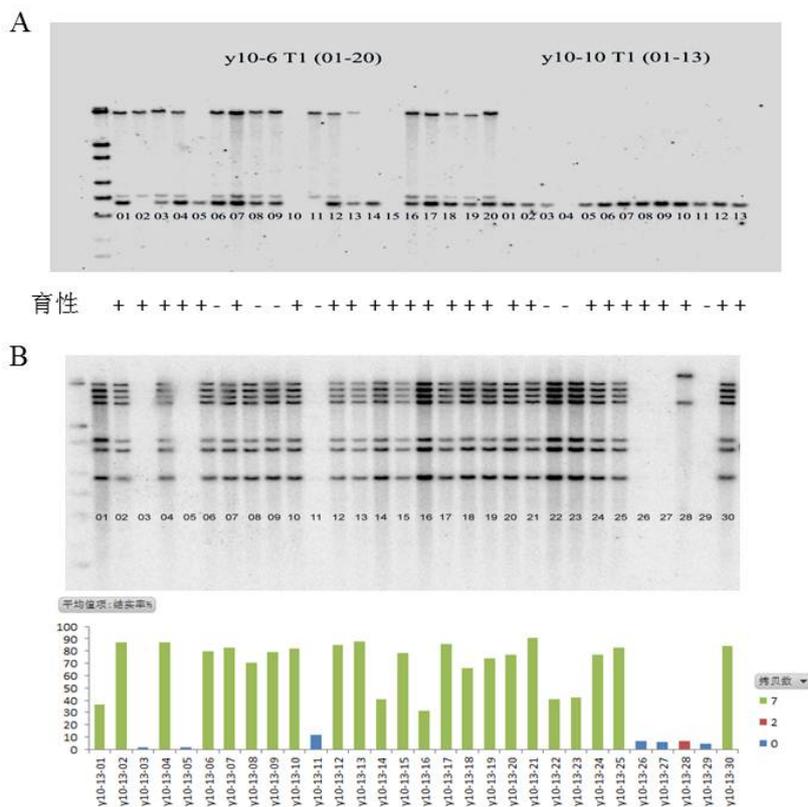


图 21 y10 转基因 T₁ 代家系拷贝数检测和育性分析。

A, y10-6 和 y10-10 两个 T₁ 代家系单株的拷贝数检测, 下排的“+”和“-”分别表示单株育性为可育和不育; B, 转基因 T₁ 代家系 y10-13 的单株拷贝数检测和结实率考察, 上排为拷贝数检测, 下排为对应单株的结实率。拷贝数检测结果中最左边的条带为 λ -EcoT14 I digest DNA Marker。

Figure 21 Southern blot and fertility analysis of y10 transgenic T₁ family.

A, Southern blot of T₁ family y10-6 and y10-10, the “+” and “-” at the bottom indicates “fertile” and “sterile”, respectively; B, Southern blot and spikelet fertility of y10-13 transgenic T1 family plants, the upper image is the southern blot result, and the spikelet fertility of individual plants is showed below. The left line in each blot is λ -EcoT14 I digest DNA Marker.

虽然有的单株育性恢复, 但在单拷贝家系 y10-10 中, 拷贝数与育性不共分离; 同样的, 在家系 y10-6 的 T₁ 代单株中, 3 个拷贝发生了分离, 但是没有任何一个拷贝与育性共分离 (图 21A)。在 y10-13 T₁ 代家系中, 除了第 28 株为 2 个拷贝外, 其余单株均为相同的 7 拷贝或者 0 拷贝, 而且第 28 株的带型明显与其他单株不一样 (图 21B), 说明是一个异常单株, 可能由串粉造成, 分析时对其不予考虑。分析剩下的 29 株, 发现拷贝数与育性共分离, 表现为: 0 拷贝单株为不育, 7 拷贝单株的育性得到不同程度的恢复; 而且相对于其他单株, 1, 14, 16, 22, 23 这 5 个单株的条

带亮度明显高些, 考虑到进行 southern blot 的样品的 DNA 量是一致的, 条带亮些的单株表明是转基因纯合, 其余的单株为转基因杂合, 那么, 杂合单株的育性明显高于纯合单株。这与我们的预期不太一致: 理论上, 在互补转化中, 如果转化片段含有 *pms1* 的候选基因, 纯合单株的育性应该高于杂合单株。而且其余的 y10 转基因 T₁ 代家系中阳性单株的育性并没有恢复, 有单株育性恢复的 2 个家系的基因型与育性也不共分离。y10-13 的育性恢复可能是由转基因造成的一个偶然事件, 比如与 T-DNA 插入位点相关, 或者多拷贝造成的共抑制等。

综合以上对之前 *pms1* 定位结果的验证, 并没有确切的证据证实得到了 *pms1* 的候选区段, 因此有必要重新考虑定位的思路。

3.2 *pms1* 显隐性关系的确定

於金生(2006)观察到长日照下, 农垦 58S/明恢 63 组合 BC₂F₄ 和 BC₂F₅ 的有些家系中不育单株数与可育单株数的比例由 BC₂F₃ 中的 1:3 变成了 3:1, 与之前报道的光敏感核不育基因 *pms1* 是隐性基因(Zhang et al 1994), F₂ 后代中不育单株数与可育单株数的分离比应该符合 1:3 的孟德尔单基因分离比的结论不相符合。另一方面, 之前对 *pms1* 进行精细定位所用的群体是 F₂ 群体或者低世代回交群体 BC₂F₃, 背景不纯, 对定位结果会产生极大的影响。因此, 本研究以农垦 58S/明恢 63 组合 BC₂F₃ 为基础, 继续与农垦 58S 进行连续回交, 以期获得高世代回交群体, 尽量排除背景的影响, 重新对 *pms1* 进行精细定位。

经过两轮回交后, 2006 年夏天获得 BC₄F₂ 群体, 长日照下种植了 24 个家系共 575 个单株。用分子标记 Fssr 检测这些单株的基因型, 得到明恢 63 纯合基因型(M)、农垦 58S/明恢 63 杂合基因型(H)和农垦 58S 纯合基因型(S)单株数分别为 145、280 和 140, 符合孟德尔单基因分离比 1:2:1。对这些单株长日照下的结实率进行分析(图 22), 发现 S 基因型单株的结实率都低于 10%, M 基因型单株的结实率都高于 40%, 而 H 杂合基因型单株的结实率都低于 40%。假设以结实率 40%作为不育和可育的判断标准, 那么, 整个群体中不育单株数与可育单株数的比例确实为 3:1, 与於金生(2006)的观察相符合。如果农垦 58S 中的 *pms1* 是隐性基因, BC₄F₂ 群体中 H 杂合基因型单株的结实率表现应该与 M 基因型单株相似, 但结果却是与 S 基因型单株更接近, 说明农垦 58S 中 *pms1* 可能是显性, 而不是隐性。

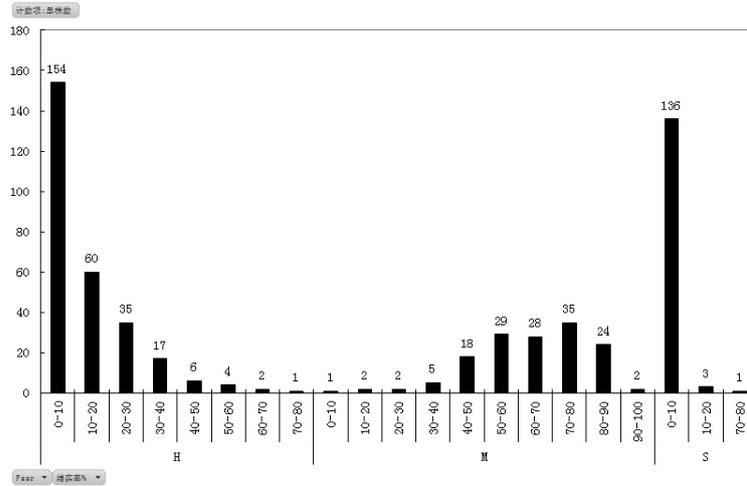


图 22 长日照下 BC₄F₂ 群体结实率。

Figure 22 The spikelet fertility of BC₄F₂ population under long-day condition.

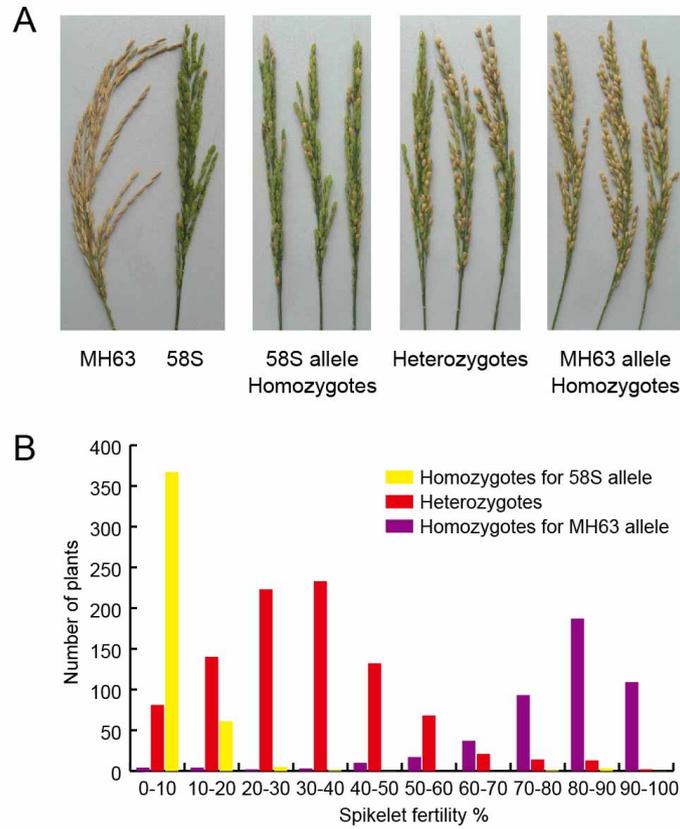


图 23 长日照下 BC₅F₂ 群体育性。

A, 群体中各种基因型单株的小穗; B, 群体中各种基因型单株育性分布图。

Figure 23 The spikelet fertility of BC₅F₂ population under long-day condition.

A, The spikelet fertility of plants in the BC₅F₂ population; B, The frequency distribution of the spikelet fertility of plants.

为了进一步验证该结论，我们又考察了 BC_3F_2 群体在长日照下的基因型与结实率（图 23）。H 杂合基因型单株的育性同样偏低，结实率的分布区间更倾向于 S 基因型单株。考虑到群体构建使用的亲本为明恢 63 和农垦 58S，属于籼粳交群体，杂合基因型单株育性偏低可能受到籼粳不育的影响，为了排除这种可能性，还需要考察短日照下群体的表型。因此我们考察了 BC_8F_2 群体在长日照下 287 个单株和短日照下 295 个单株的基因型与结实率（图 24）。

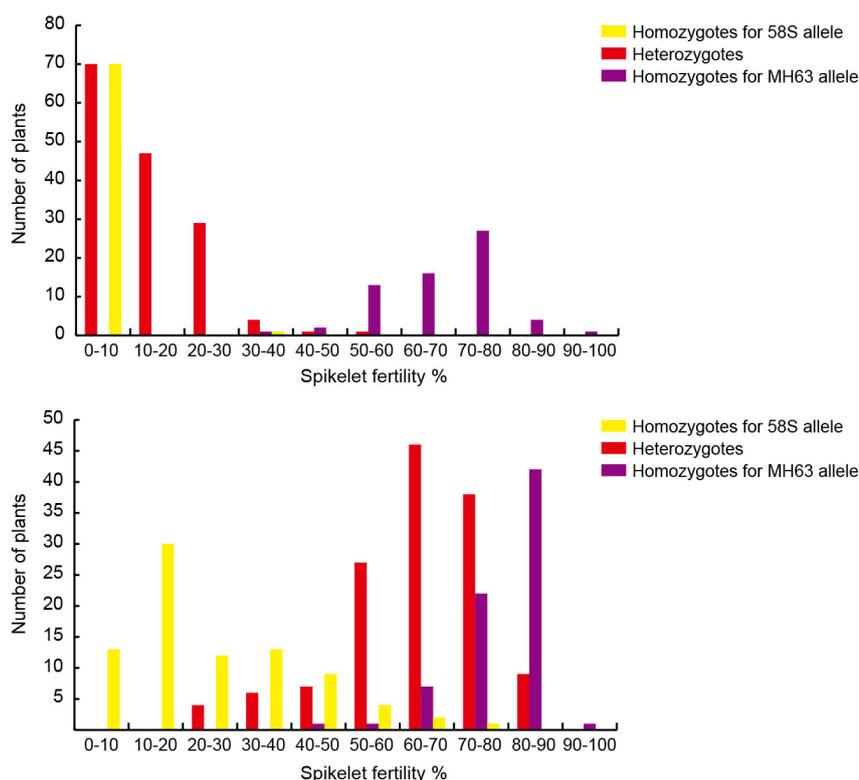


图 24 BC_8F_2 群体在长日照和短日照下的育性分布图。

上图，长日照；下图，短日照。

Figure 24 The frequency distribution of the spikelet fertility of BC_8F_2 population under long-day and short-day conditions.

Upper, long-day conditions; lower, short-day conditions.

BC_8F_2 群体各基因型单株在长日照下的育性分布与 BC_4F_2 和 BC_3F_2 群体相似。短日照下，整个群体的育性都得到了不同程度的恢复，尤其是 H 杂合基因型单株的育性大部分都高于 50%。排除了长日照下杂合基因型单株育性偏低是受籼粳不育影响的可能性，因为籼粳不育并不受日照长度的影响，而群体中 H 杂合基因型单株在短日照下是可育的。

通过对几个回交 F₂ 群体的分析, 尤其是高世代 BC₈F₂ 群体的表现, 充分证实了农垦 58S 中的 *pms1* 不是隐性, 而是显性, 并且这种光敏雄性不育特性表现为不完全显性。

3.3 *pms1* 的精细定位

在确定了 *pms1* 显隐性的基础上, 比较容易理解前人的遗传转化为什么看不到预期的表型, 山燕(2000), 李亚芳(2002)和於金生(2006)都是将隐性的 *pms1* 基因转化到拥有显性 *pms1* 的农垦 58S 中, 因此无法通过转基因试验验证定位结果的准确性。根据 F₂ 群体的育性表现, 我们重新开始对 *pms1* 进行精细定位。

3.3.1 分子标记的获得

在对 *pms1* 进行精细定位的过程中, Liu et al (2001b)和於金生(2006)开发了一些分子标记, 但这些标记的密度远远不够, 而且大部分都是 RFLP 标记, 不方便用于大量植株的基因型鉴定。综合分析之前的定位结果后(图 19), 确定在分子标记 Fssr 和 pj23 之间寻找更多更合适的分子标记用于精细定位。Fssr 和 pj23 在明恢 63 基因组上的物理距离为 138 kb, 对应于日本晴基因组 300 kb(Yu et al 2007)。

在 Fssr 和 pj23 之间已有可用 SSR 标记 Rssr (重命名为“P5”)和 108ssr (重命名为“P11”), 并且分布比较均匀, 为了在其中进一步加密标记, 主要采用以下几种方法开发新的标记: 1, 根据明恢 63 和日本晴基因组序列在 Gramene 网站上搜寻 Fssr 和 pj23 之间可能存在的 SSR 标记, 并设计引物扩增明恢 63 和农垦 58S 基因组序列进行标记多态性验证; 2, 将明恢 63 和日本晴基因组序列进行比对, 针对序列差异开发标记; 3, 利用明恢 63 BAC 21O9 和农垦 58 BAC 18P4 在该区间的序列差异开发标记; 4, 於金生(2006)对农垦 58S 在 Rssr 和 N6B5 之间的序列进行了测序, 将这些测序结果与明恢 63 序列比对, 设计 CAPS/dCAPS 标记; 5, 通过上述方法获得了一定分子标记的基础上, 根据已有标记分布情况, 在需要加密标记的区段参考明恢 63 基因组序列, 设计一系列引物对农垦 58S 进行测序, 进一步开发标记。最终在 Fssr 和 pj23 之间共得到 11 个分子标记, 具体情况见表 2, 引物序列参见附录 I, 标记的实际检测情况见图 25。

表 2 分子标记信息。

Table 2 Information of molecular markers.

Name	Type	Detection
pj23	SSR	PCR
P1	CAPS	PCR, <i>Mun</i> I
P2	CAPS	PCR, <i>Cfr</i> 10I
P3	CAPS	PCR, <i>Eco</i> 52I
P4	dCAPS	PCR, <i>Alu</i> I
P5	SSR	PCR
P6	InDel	PCR
P7	InDel	PCR
P8	CAPS	PCR, <i>Dra</i> I
P9	CAPS	PCR, <i>Eco</i> RV
P10	SSR	PCR
P11	SSR	PCR
Fssr	SSR	PCR

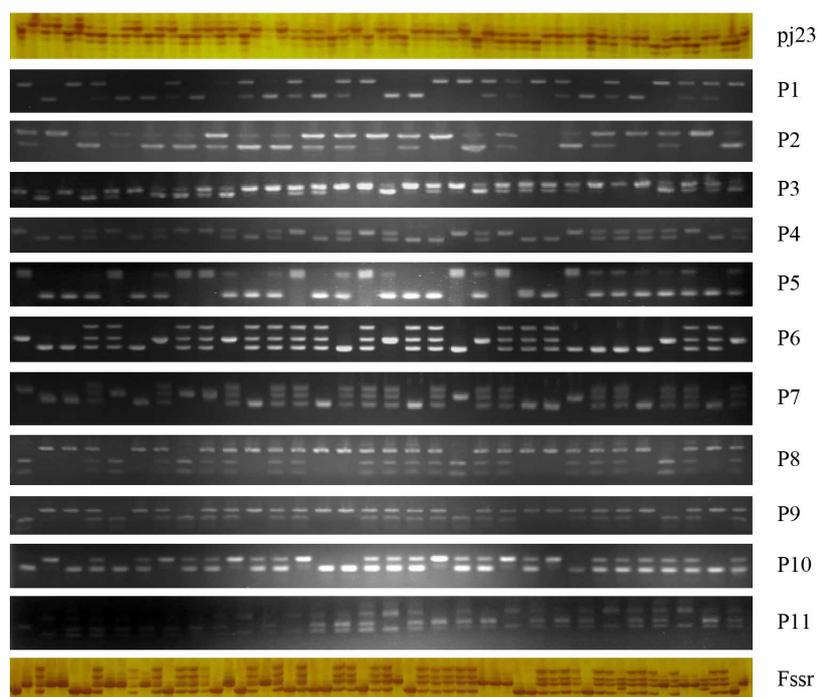


图 25 分子标记检测图。

Figure 25 The detection of molecular markers.

3.3.2 图位克隆 *pms1*

2007年夏天在武汉长日照下大田种植农垦 58S/明恢 63 组合 BC₅F₂ 群体用于图位克隆 *pms1*。在经典的图位克隆过程中,明确了基因的显隐性关系后,为了节省时间,通常采用极端隐性群法,选取群体中表现为极端隐性表型的单株上标记,鉴定重组单株(Zhang et al 1994; Liu et al 2001b; Lu et al 2005; Xue et al 2008)。但在对 *pms1* 进行图位克隆时,之前的结果表明 *pms1* 表现为不完全显性,从图 23A 中可以看到,杂合基因型单株的表型会与 S 纯合基因型单株的表型重叠,表现为极端不育性状的单株基因型可能是 H 或者 S。因此对 BC₅F₂ 群体中 6841 个单株全部用两端标记 pj23 和 Fssr 进行分析,共得到 88 个重组单株。

其中 H 和 S 之间的交换单株数为 43, H 和 M 之间的交换单株数为 45。考虑到发生了 H 和 S 交换的重组单株因为大部分 H 杂合基因型单株的育性也很低,与 S 纯合基因型单株的结实率差不多,无法根据该重组单株的结实率准确判断候选基因所在的方向,但是 H 和 M 基因型单株间的育性差别比较大,更容易判断单株发生交换的方向。因此,仅用 45 个在 H 和 M 之间发生交换的重组单株完成精细定位。先用 pj23 和 Fssr 之间剩下的 11 个分子标记进一步鉴定这 45 个单株的基因型,再通过重组单株的结实率判断 *pms1* 所在的区间。并且将这些重组单株收种,于次年夏天各种植 30 株进行了基因型鉴定和表型观察。最终将 *pms1* 限定在分子标记 P4 和 P6 之间 4.0 kb 的范围内,左右各 1 个重组单株,分子标记 P5 与 *pms1* 共分离(图 26)。

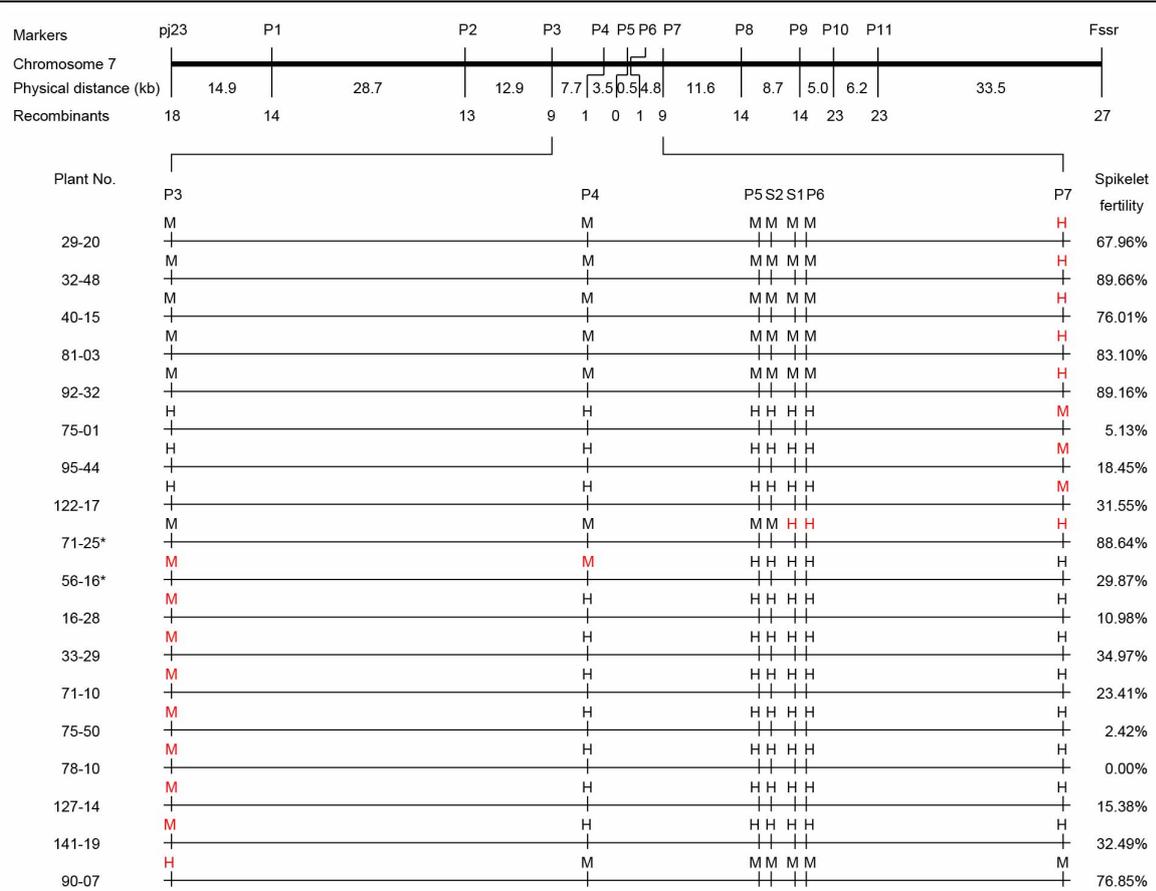


图 26 分子标记 pj23 和 Fssr 之间重组单株分析。

图中上部分显示各个分子标记处发生交换的重组单株数，物理距离基于明恢 63 基因组；下图显示标记 P3 和 P7 之间的 18 个重组单株的基因型和结实率。M，表示明恢 63 基因型纯合；H，表示明恢 63/农垦 58S 杂合基因型；发生交换的位点用红色显示。Plant No.是重组单株在 F₂ 群体中的编号，其单株结实率显示在右侧，带星号的交换单株编号为 P4 和 P6 处的重组单株。

Figure 26 The analysis of recombinants between molecular marker pj23 and Fssr.

The upper part presented the numbers of recombinants, the physical distance was the situation on Minghui 63 genome; and the genotypes and fertility of the recombinants between markers P3 and P7 was showing below. M, homozygous for MH63 genotype; H, heterozygote; red color indicates the recombination sites. Plant No., the identification number for recombinants in the population, and their spikelet fertility is shown on the right. *, the most interesting recombinants.

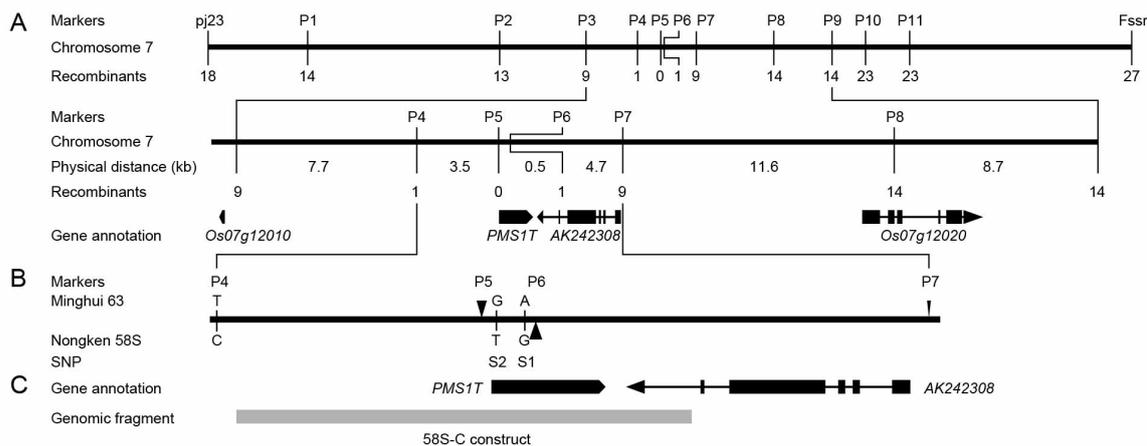


图 27 *pms1* 定位区间内基因注释和序列比较分析。

A, 定位区间附近的基因注释。中间的黑色方框代表 exon，黑线代表 intron，箭头示意基因的方向。B, 农垦 58S 和明恢 63 在分子标记 P4 和 P7 之间的序列比较。短竖线表示单碱基多态性，黑色三角表示多碱基插入。C, 互补载体 58S-C 覆盖的区间。

Figure 27 Gene annotation at the *pms1* mapping region and the comparative sequence analysis.

A, Gene annotation at the *pms1* mapping region. The black blocks in the middle represent exons, and the lines between them represent introns, arrows indicate the directions of genes. B, The sequence comparison between Nongken 58S and Minghui 63 at the region flanking by molecular marker P4 and P7. vertical lines represent the single nucleotide polymorphisms, and triangles for insertions. C, The fragment amplified from 58S for preparing the construct 58S-C.

对该区间内农垦 58S 测序后，与明恢 63 序列比较分析发现，除了 P4, P5 和 P6 处的序列差异外，两者之间有且仅有两个单碱基的差异，都位于 P5 和 P6 之间，将这两个 SNP 命名为 S2 和 S1 (图 27B)。在 SNP S2 位置，明恢 63 中的碱基 G (CTACTAGTGT) 突变为农垦 58S 中的碱基 T (CTACTATTGT)，结合附近序列分析，发现此突变位点正好落在限制性内切酶 *SpeI* 的识别位点上，突变后农垦 58S 中就丢失了 *SpeI* 的识别位点。利用此差异可以开发成 CAPS 标记：使用引物对 (F: 5'-GACTACATGGGCACCCCTTGAA -3', R: 5'-ACCTGGCATAGACCGATAGTTAC -3') 扩增长度为 390 bp，经过 *SpeI* 完全酶切后在明恢 63 中显示为 266 bp 和 124 bp 大小的两条带，在农垦 58S 中为单条带，如图 28 所示。但分析后发现 S1 处的 SNP 无法开发成分子标记。

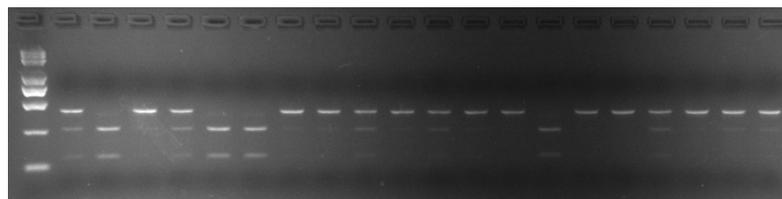


图 28 根据 S2 开发的 CAPS 标记。

Figure 28 The CAPS marker developing from S2 SNP.

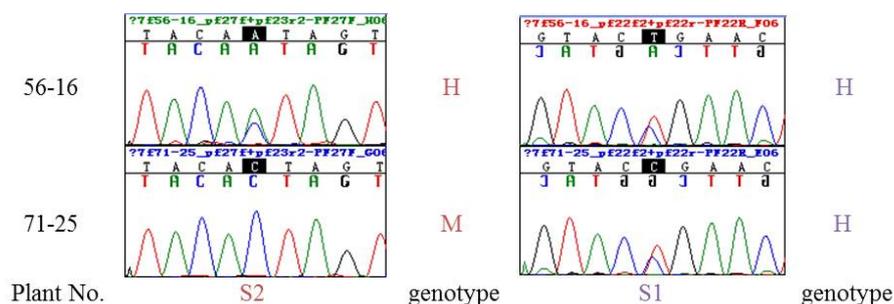


图 29 重组单株在 S2 和 S1 处的测序结果。

Figure 29 The sequencing result of recombinants at SNP S2 and S1.

由于两个 SNP 位点相距较近，对重组单株直接测序鉴定这两个位点处的基因型（图 29），发现 S2 与 *pms1* 共分离，S1 处还有 1 个重组单株（图 26）。根据此结果，可进一步缩小定位区间，将 *pms1* 定位于 P4 与 S1 之间 3.8 kb 内，左右各 1 个重组单株（图 26 和图 27A）。

3.4 *pms1* 候选基因的分离和验证

在 3.8 kb 的定位区间内，并没有发现有任何预测的基因（图 27B）。左右两侧距离最近的两个预测基因 *LOC_07g12010* 和 *LOC_07g12020* 离 P5 在明恢 63 基因组上分别为 7.7 kb 和 15.0 kb，如果是日本晴基因组，则距离更远。这两个基因都注释为 putative unclassified retrotransposon protein。此外，在 KOME 数据库 (<http://cdna01.dna.affrc.go.jp/cDNA/>) 中找到 1 条反义链上的全长 cDNA AK242308，位于 P6 的下游，其 3' 端距离 P6 有 1,139 bp。在无任何可供候选的预测基因的情况下，必须通过转基因的方法验证定位结果的可靠性，并分离到这个未被注释的 *pms1* 候选基因。

3.4.1 互补转化结果

根据已有遗传分析结果, 确定 *pms1* 是不完全显性, 而且不育是显性。据此构建了包含相同区段的农垦 58S 和明恢 63 基因组序列的互补载体 58S-C 和 MH-C, 其覆盖的区域见图 27C, 包含了从 P4 到 P6 及 AK242308 最后一个 exon 和最后一个 intron 的部分序列, 分别为 5,695 bp 和 5,682 bp。将 58S-C 和 MH-C 分别转化 NIL(MH) 和农垦 58S, 得到 79 和 47 株转基因 T₀ 代独立转化单株。由于组织培养过程对转基因植株育性的影响, 一般并不对 T₀ 代转基因植株进行共分离检测。仅完成转基因阳性鉴定后, 对阳性单株收种, 种植 T₁ 代进行共分离检测。采用基因内部分子标记 P6 检测转基因单株的阴/阳性, 每株选取 3 个穗子考察单株结实率。

在 58S-C 转化 NIL(MH) 的 T₁ 代转基因家系单株中, 基因型与表型完全共分离, 具体表现为: 长日照下, 相对于转基因阴性单株, 阳性单株株型、穗型都没有明显变化, 但结实率出现不同程度的降低; 而在短日照下, 转基因阴性和阳性单株之间的育性并没有明显的区别, 均表现为可育 (图 30A 和 B)。表 1 为其中两个 T₁ 代家系在长日照和短日照下的小穗育性统计结果, 经 *t*-test 检验, 长日照下阳性单株和阴性单株之间的结实率差异达到了极显著水平, 但在短日照下不显著。同时我们观察到同一家系中阳性单株的结实率虽然下降, 平均结实率在 40% 左右, 但各个单株育性降低的程度并不相同, 呈连续分布的趋势 (图 30C), 与之前得到的 F₂ 群体中 H 杂合基因型单株的结实率表现相似。而事实上, 通过转基因的方式经 T-DNA 插入外源片段后, 加上受体基因组原有的内源片段, 得到的就是杂合基因型。以上结果表明 58S-C 载体携带的外源片段能够显著降低受体 NIL(MH) 的育性, 并且这种效应只在长日照下发挥作用, 说明遗传转化片段包含有控制光敏感雄性不育的 *pms1* 基因, 而不仅仅只是一个控制雄性不育的基因。

分析 MH-C 转化农垦 58S 的 T₁ 代家系的育性 (图 30C)。长日照下, 转基因阳性单株与阴性单株之间的育性没有差别, 都为不育, 结实率之间不存在显著性差异; 短日照下, 所有转基因单株均为可育, 阴阳性单株之间的结实率也没有显著性差异 (表 3)。说明来自于明恢 63 的包含有 *pms1* 位点的基因组序列不能恢复长日照下农垦 58S 的育性。这也与我们的预期相符, 表明明恢 63 的 *pms1* 为隐性, 对农垦 58S 携带的显性 *pms1* 不会产生影响。

从 58S-C 和 MH-C 的互补转化结果可以确定转化的 5.7 kb 序列中包含有 *pms1* 的候选基因。

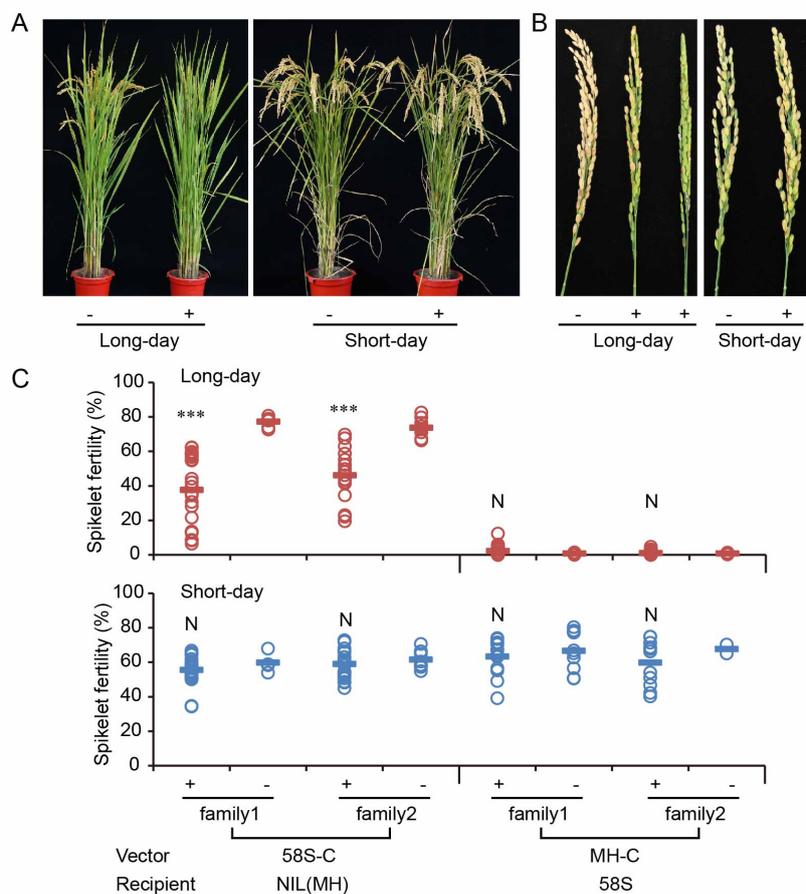


图 30 58S-C 和 MH-C 转基因 T₁ 代单株在长、短日照下的育性。

A, 58S-C 转基因 T₁ 代家系单株在长、短日照下的整株表型。B, 58S-C 转基因单株的小穗, +, 阳性单株; -, 阴性单株。C, 58S-C 和 MH-C 转基因 T₁ 代家系单株在长、短日照下的小穗育性。中间的短横线代表类群中结实率的平均值。用 *t*-test 检测每个家系中阳性单株与阴性单株之间的显著性, 其中***代表极显著 ($P < 0.001$), N 代表不显著 ($P > 0.05$)。

Figure 30 Fertility of 58S-C and MH-C T₁ transgenic plants under long-day and short-day conditions.

A, Whole plants of transgenic segregants from the T₁ family of 58S-C under long days and short days. B, Panicles of transgenic plants of 58S-C. +, positive plants; -, negative plants. C, Spikelet fertility of T₁ transgenic plants from 58S-C and MH-C under long-day and short-day conditions. The bars indicate the means. *** indicates statistically significant ($P < 0.001$) difference by *t*-test between the positive and negative plants in each family, N indicates non-significant difference ($P > 0.05$).

表 3 互补和 *dsi* 抑制转化来自于 2 个 T_0 单株的 T_1 家系在长短日照下的小穗育性。

Table 3 Spikelet fertility of T_1 families from two representative T_0 transgenic plants under long-day (LD) and short-day (SD) conditions.

Construct	Recipient	Genotype	LD Spikelet fertility (%)		SD Spikelet fertility (%)	
			Family 1	Family 2	Family 1	Family 2
58S-C	NIL(MH)	Positive	37.61±4.19 (23)	46.08±3.27 (19)	55.49±1.64 (24)	58.87±1.91 (20)
		Negative	77.22±1.08 (7)	73.73±1.78 (9)	59.69±2.94 (4)	61.47±1.61 (10)
		<i>P</i> value	0.0000	0.0000	0.3309	0.3855
MH-C	58S	Positive	2.16±0.56 (25)	1.04±0.24 (24)	63.26±2.70 (14)	59.73±3.98 (11)
		Negative	0.81±0.25 (5)	0.83±0.15 (6)	66.52±3.63 (10)	67.58±2.55 (2)
		<i>P</i> value	0.2992	0.6634	0.4691	0.4360
58S- <i>dsi</i>	58S	Positive	72.18±4.36 (13)	68.46±1.98 (18)	73.64±2.11 (9)	77.52±1.31 (16)
		Negative	24.35±0.57 (5)	30.31±2.08 (2)	71.37±3.90 (4)	73.91±1.16 (6)
		<i>P</i> value	0.0000	0.0000	0.5884	0.1272
MH- <i>dsi</i>	NIL(MH)	Positive	90.57±0.74 (19)	89.30±0.60 (15)	77.42±0.94 (18)	73.80±2.17 (16)
		Negative	91.68±1.17 (6)	90.32±0.79 (5)	76.17±1.66 (8)	73.11±4.48 (5)
		<i>P</i> value	0.4603	0.3775	0.4929	0.8829

所有数据展示为平均值±s.e.m。括号内的数字为被统计的单株数，*P* 值通过 *t*-test 比较阳性单株与阴性单株之间小穗育性得到。

All data presented are mean±s.e.m. Numbers in parentheses are numbers of plants analyzed. *P* value is obtained from the *t*-test of positive and negative plants.

3.4.2 RACE 确定 *pms1* 转录本

58S-C 和 MH-C 载体包含的外源片段中虽然包含了 AK242308 的最后一个 exon 和最后一个 intron 的部分序列, 但是 AK242308 的 3' 端距离分子标记 P6 还有 1,139 bp, P6 处还存在 1 个重组单株, 因此, 基本上可以排除 AK242308 作为 *pms1* 候选基因的可能性。剩下的区间属于基因间区, 表明 *pms1* 候选基因是一个未知的、从未被发现的基因。

在此区间内, 采用 prime walking 的方法, 从农垦 58S 和 NIL(MH) 的幼穗 cDNA 中扩增获得 *pms1* 候选基因的 EST 序列 (Expressed Sequence Tag)。在此基础上, 进一步采用试剂盒分别进行 5' RACE 和 3' RACE, 确定转录本的转录起始位点和终止位点。RACE 结果如图 31 所示, 将图中箭头标识的条带回收连接到 T-vector 上进行测序, 获得的序列去掉接头序列后与农垦 58S 和明恢 63 基因组序列进行比对, 得到 *pms1* 的全长 cDNA。

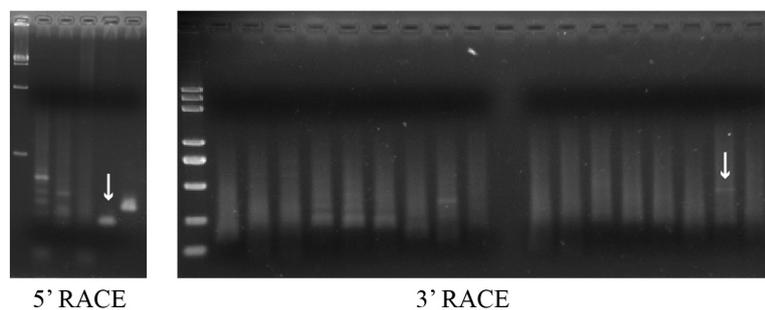


图 31 *pms1* 候选基因的 RACE 扩增。

最左边为 DL2,000 marker, 白色箭头指示用于测序的条带。

Figure 31 Rapid Amplification of cDNA Ends (RACE) of *pms1*.

The left line in each image was the DL2,000 marker, white arrows indicated the samples used for sequencing.

RACE 分析结果表明, 在定位区间内, *pms1* 的候选基因有且仅有 1 条转录本, 没有 intron, 仅包含 1 个 exon, 将该转录本命名为 *PMSIT* (图 27C)。*PMSIT* 在农垦 58S 和明恢 63 中的长度分别为 1,453 bp 和 1,388 bp, 两者相差 65 bp, 这 65 bp 在之前的图位克隆中已被开发为 InDel 分子标记 P6, 位于 *PMSIT* 的中部。两个 SNP 位点 S1 和 S2 都位于 *PMSIT* 的 5' 端。

3.4.3 抑制 *PMS1T* 表达的转化结果

为了进一步验证 *PMS1T* 的功能，构建了双链 RNA 抑制载体 *dsi* 以及 artificial microRNA 抑制载体 *ami1* 和 *ami2*（参见前文“材料与方法”）。将这 3 个载体分别转化农垦 58S 与 NIL(MH)，得到的转基因植株命名为 58S-*dsi*，MH-*dsi*，SAI1，NAI1，SAI2 和 NAI2。根据群体遗传分析及 *pms1* 互补转化结果，推测长日照下在农垦 58S 中抑制了显性 *pms1* 候选基因 *PMS1T* 的表达后，会降低农垦 58S 的育性；在 NIL(MH) 中，隐性 *PMS1T* 的表达量被降低后，不会对育性造成影响；短日照下，无论在农垦 58S 还是 NIL(MH) 中抑制 *PMS1T* 的表达都会表现为可育。

抑制载体 *dsi* 包含的 321 bp 片段恰好位于 SNP S2 和 SNP S1 之间，在明恢 63 与农垦 58S 基因组序列间并无差异。共获得了 58S-*dsi* 和 MH-*dsi* T₀ 代转基因植株 21 和 42 株，长日照下种植，用引物对 PMCGR/PMCGR 检测转基因阴/阳性，并对单株进行 *PMS1T* 表达量检测（图 32）。选取 *PMS1T* 表达量下降的 T₀ 代单株在长日照和短日照下分别种植 T₁ 代家系进行共分离检测。

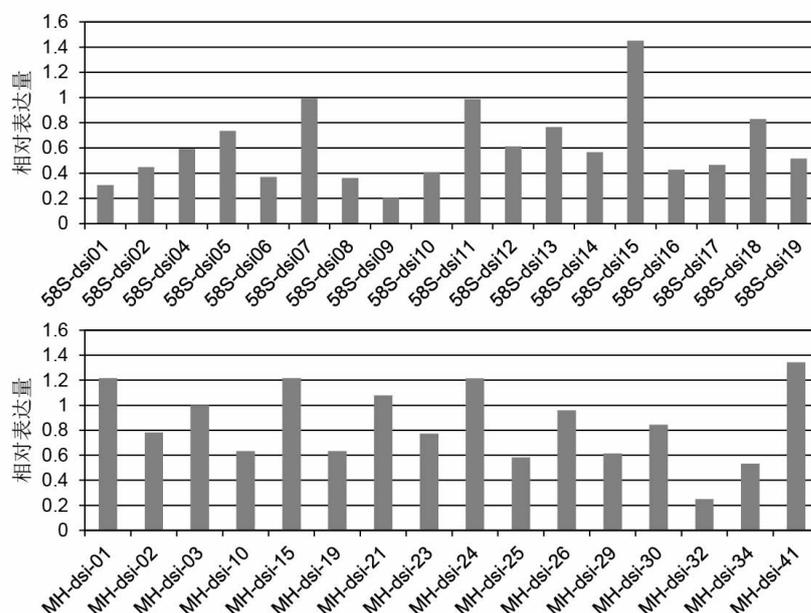


图 32 58S-*dsi* 和 MH-*dsi* 转基因 T₀ 代单株 *PMS1T* 相对表达量检测。

Figure 32 The relative expression level of *PMS1T* in T₀ transgenic plants of 58S-*dsi* and MH-*dsi*.

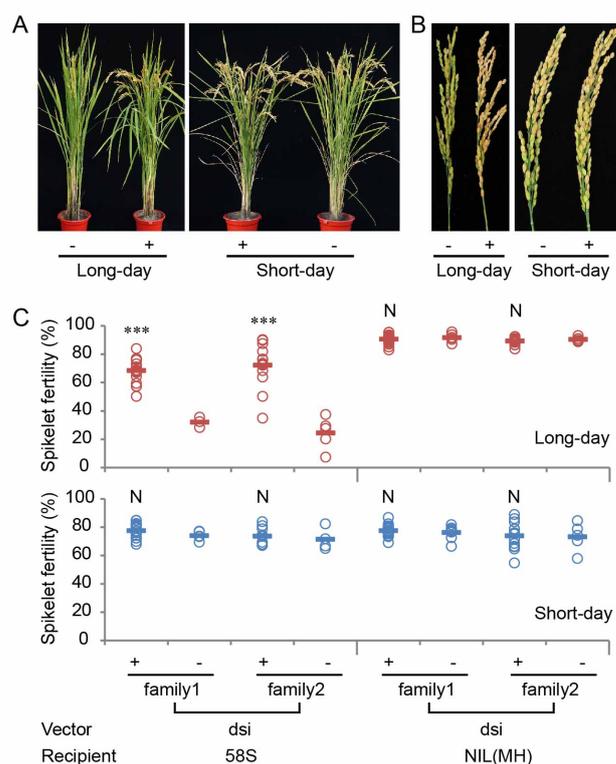


图 33 58S-dsi 和 MH-dsi 转基因 T₁ 代家系单株在长、短日照下的育性。

A, 58S-dsi 转基因 T₁ 代家系单株在长、短日照下的整株表型。B, 58S-dsi 转基因单株的小穗, +, 阳性单株; -, 阴性单株。C, 58S-dsi 和 MH-dsi 转基因 T₁ 代家系单株在长、短日照下的小穗育性。中间的短横线代表类群中结实率的平均值。用 *t*-test 检测每个家系中阳性单株与阴性单株之间的显著性, 其中***代表极显著 ($P < 0.001$), N 代表不显著 ($P > 0.05$)。

Figure 33 Fertility of 58S-dsi and MH-dsi T₁ transgenic plants under long-day and short-day conditions.

A, Whole plants of transgenic segregants from the T₁ family of 58S-dsi under long days and short days. B, Panicles of transgenic plants of 58S-dsi. +, positive plants; -, negative plants. C, Spikelet fertility of T₁ transgenic plants from 58S-dsi and MH-dsi under long-day and short-day conditions. The bars indicate the means. *** indicates statistically significant ($P < 0.001$) difference by *t*-test between the positive and negative plants in each family, N indicates non-significant difference ($P > 0.05$).

在这些 58S-dsi T₁ 代家系中, 长日照下, 单株之间的株型、株高、分蘖数等没有明显差异, 但在成熟期, 某些单株的小穗直立, 并不勾头 (图 33A), 考察基因型与结实率后, 发现在长日照下基因型与表型共分离: 转基因阴性单株小穗育性很低, 大部分为空壳, 转基因阳性单株的小穗育性得到极大的恢复 (图 33B 左图); 而在

短日照下，两者的小穗育性没有明显差异，均为可育。图 33C 进一步展示了其中 2 个家系的单株在长日照和短日照下的结实率，长日照下相对于阴性单株，阳性单株的结实率有不同程度的上升。统计测验也表明长日照下的差异达到了极显著水平(表 3)。在 MH-dsi T₁ 代转基因家系中，无论长、短日照，阴性单株和阳性单株之间的育性并没有差别，都为可育(表 3)。以上结果与我们的预期完全相符，表明 *PMSIT* 就是光敏感雄性核不育基因 *pms1*，也进一步证实了 58S 的 *PMSIT* 为显性。

另一方面，在构建双链 RNA 抑制载体 *dsi* 的同时，为了保险起见，我们还构建了 artificial microRNA 抑制载体 *ami1* 和 *ami2*，靶序列分别位于 *PMSIT* 的 3' 和 5' 端。共得到转基因 T₀ 植株数为：SAI1 19 株，NAI1 19 株，SAI2 18 株和 NAI2 9 株。同样也对这些转基因植株检测了 *PMSIT* 的表达量(图 34)，选取表达量下降的单株在长日照和短日照下分别种植 T₁ 代。

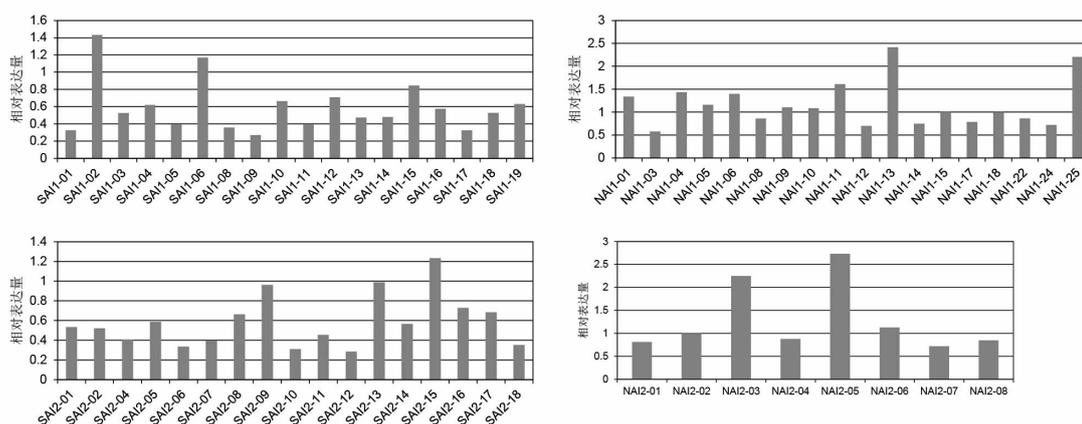


图 34 SAI1, SAI2, NAI1 和 NAI2 转基因 T₀ 代单株 *PMSIT* 相对表达量检测。

Figure 34 The relative expression level of *PMSIT* in T₀ transgenic plants of SAI1, SAI2, NAI1 and NAI2.

与 MH-dsi 相似，NAI1 和 NAI2 的转基因阳性单株和阴性单株在长、短日照下均为可育，结实率之间没有显著差异(表 4)。但同样是抑制农垦 58S 中 *PMSIT* 的表达，SAI1 和 SAI2 的转基因单株阴性和阳性在长日照下的表型一样，阳性单株的育性并不像 58S-dsi 那样会恢复到可育水平。说明在对 *PMSIT* 的抑制表达过程中，artificial microRNA 载体没有双链 RNA 载体的效率高，可能与两个载体发挥作用的方式不一样有关。*dsi* 载体的针对的是一段 321 bp 的长片段，该片段内可能包含有 *PMSIT* 发挥作用的重要元件，而 artificial microRNA 的靶序列仅 21 bp，*ami1* 和 *ami2* 的两个 21 bp 靶序列可能都对 *PMSIT* 的功能没有那么重要。

表 4 Artificial microRNA 抑制转化 T₁ 家系在长、短日照下的小穗育性。Table 4 Spikelet fertility of SAI1, SAI2, NAI1 and NAI2 T₁ families under long-day (LD) and short-day (SD) conditions.

Condition	Transgene	Genotype	Number of Plants	Spikelet fertility (%)
Long day	SAI1	Positive	18	12.48±4.03
		Negative	9	8.84±2.50
		<i>P</i> value		0.5508
	NAI1	Positive	19	57.68±2.07
		Negative	6	57.25±3.71
		<i>P</i> value		0.9209
	SAI2	Positive	20	26.53±3.83
		Negative	5	27.25±7.23
		<i>P</i> value		0.5508
	NAI2	Positive	25	72.30±1.70
		Negative	3	68.77±5.09
		<i>P</i> value		0.5034
Short day	SAI1	Positive	13	64.24±2.20
		Negative	7	67.98±1.72
		<i>P</i> value		0.2680
	NAI1	Positive	16	34.94±2.85
		Negative	9	27.63±5.81
		<i>P</i> value		0.2163
	SAI2	Positive	12	52.39±2.20
		Negative	1	56.12
		<i>P</i> value		0.6473
	NAI2	Positive	28	69.99±1.29
		Negative	2	73.98±0.61
		<i>P</i> value		0.4229

3.4.4 超量表达 *PMS1T* 的转化结果

之前对 *PMS1T* 的互补转化和抑制转化结果都与我们的预期相符,那么超量表达 *PMS1T* 实际又会是什么样的结果? 根据已有结果推测,在 NIL(MH)中超量表达显性的农垦 58S *PMS1T*, 应该会降低受体的育性, 并且降低的程度可能会高于互补转化的结果。超量表达载体 Ubi:S 包含农垦 58S 的全长 cDNA, 在玉米 *ubiquitin* 的启动子启动下进行组成型表达。Ubi:S 转化 NIL(MH)得到的 T₀ 代转基因单株完成表达量检测后 (图 35), 将超量表达 *PMS1T* 的单株继续种植 T₁ 代, 用于共分离分析。转

基因单株的基因型采用基因内部分子标记 P6 鉴定。长日照下考察单株的基因型与结实率，发现两者完全共分离：转基因阳性单株的育性非常低，转基因阴性单株完全可育，同一家系中阳性单株和阴性单株的结实率分为非常明显的两个类群（图 36A）。该结果表明在 NIL(MH)中超表达农垦 58S 的 *PMSIT* 基因能够显著降低 NIL(MH)的育性，符合之前的预期。综合之前的结果，在 NIL(MH)中无论是互补还是超表达来自于农垦 58S 的 *PMSIT* 基因均能够降低育性，但是仔细比较两种转化结果（图 30C 和图 36A），发现超表达的效应要大于自身启动子驱动表达的互补转化。在互补转化中，转基因阳性单株的结实率是不同程度的降低；在超表达转化中，转基因阳性单株的结实率却是普遍极低。

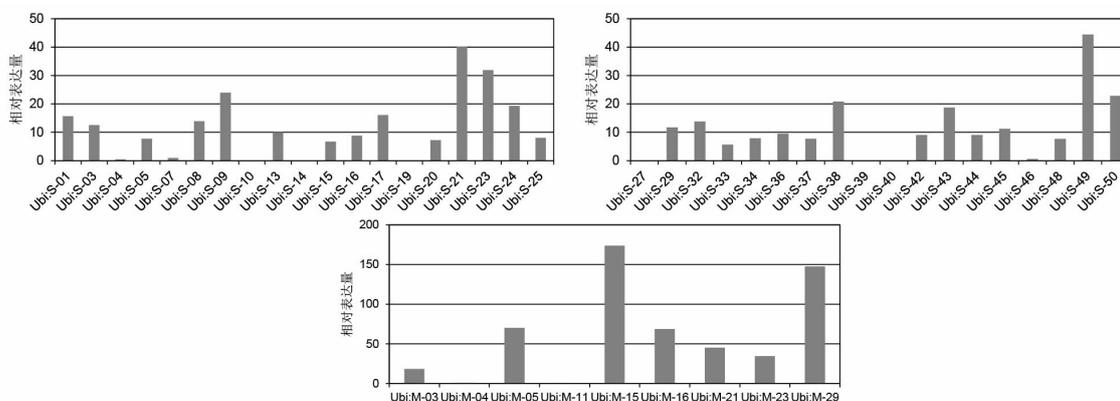


图 35 超量表达 *PMSIT* 转基因 T_0 代单株相对表达量检测。

上排：Ubi:S 转基因单株；下排：Ubi:M 转基因单株。

Figure 35 The relative expression level of *PMSIT* in T_0 transgenic plants of Ubi:S and Ubi:M.

The images in upper and bottom were the result of Ubi:S and Ubi:M transgenic plants, respectively.

虽然之前互补和抑制明恢 63 中的 *PMSIT* 都对育性没有影响，考虑到超表达是一种组成型表达，也许会有不一样的结果。Ubi:M 的外源片段来自于明恢 63 的 *PMSIT*，与 Ubi:S 都包括相同的全长 *PMSIT* 区段，转化受体为农垦 58S。同样选择 *PMSIT* 确实超量表达的单株（图 35）进行 T_1 代共分离分析。从图 36B 的 3 个家系共分离分析结果中明显看到不同基因型的单株结实率集中在不同的育性区间，长日照下 Ubi:M 转基因阳性单株的育性得到了不同程度的恢复，且基因型与表型共分离。比较图 36A 和 B，发现 Ubi:M 转基因阳性单株的育性恢复并不是完全的，结实率分布范围较广，不像 Ubi:S 中超量表达的作用那么彻底——阳性单株的结实率大部分为“0”，且范围集中，说明超量表达明恢 63 的 *PMSIT* 对育性有一定的恢复作用。

Ubi:M 的转化结果是我们预先没有设想到的，但是现在还并下结论说该结果与之前的遗传分析以及互补转化和抑制转化结果相矛盾。一方面，超量表达是组成性的表达，改变了基因原有的时空表达模式；另一方面，这种转化结果预计与 *PMS1T* 的作用机理相关，只有在阐明了 *PMS1T* 作用方式的情况下才能得到合理解释。

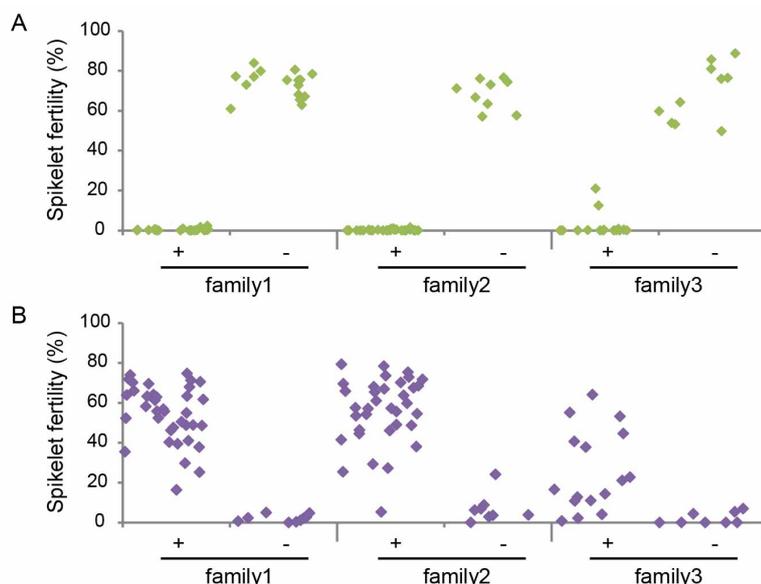


图 36 超量表达 *PMS1T* 转基因 T_1 代各 3 个家系长日照下共分离分析。

A, Ubi:S 转基因 T_1 家系; B, Ubi:M 转基因 T_1 家系。

Figure 36 The co-segregation analysis of *PMS1T* over-expression in transgenic plants in three T_1 families under long-day conditions.

A, T_1 families of Ubi:S; B, T_1 families of Ubi:M.

3.5 *PMS1T* 表达谱分析

PMS1T 位于基因间区，没有对应的基因注释，在 Affymetrix 芯片中不存在与之相应的探针，因此无法利用本室的水稻全生育期表达谱芯片数据(Wang et al 2010b)。在武汉大田长日照和短日照下分别种植农垦 58S 和 NIL(MH)，对各个时期的不同组织取材，用于 *PMS1T* 的表达谱分析。

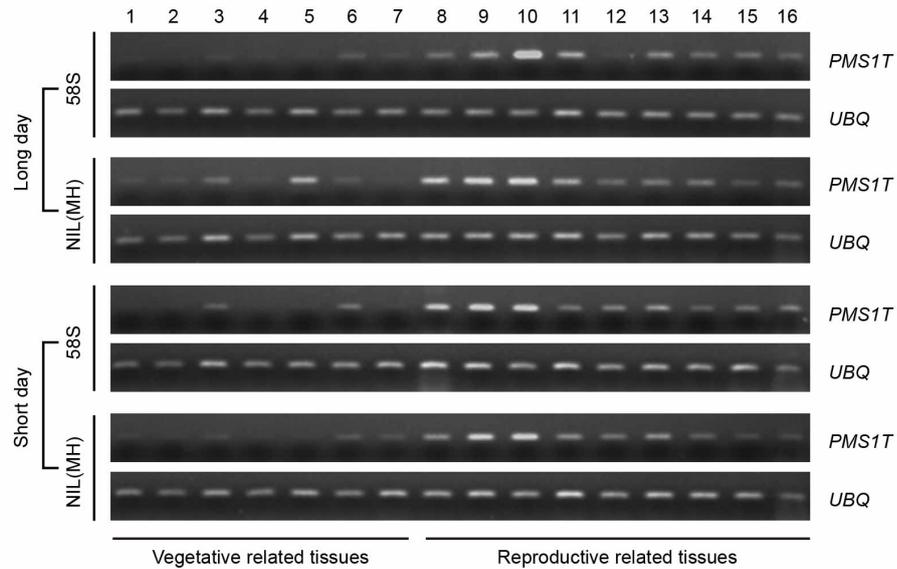


图 37 *PMS1T* 基因在长、短日照下农垦 58S 和 NIL(MH) 全生育期不同组织中的表达量分析。

每组图片中上排为 *PMS1T* 扩增结果，PCR 循环数为 35；下排为内参基因 *Ubiquitin* (*LOC_Os03g13170*) 扩增 25 个循环的结果。编号 1-16 分别代表水稻不同发育时期的不同组织：1，分蘖期叶片；2，雌雄蕊形成期叶片；3，雌雄蕊形成期倒二节间；4，雌雄蕊形成期叶片叶鞘；5，剑叶；6，小孢子发育晚期倒二节间；7，剑叶叶鞘；8，二次枝梗分化期幼穗 (<0.5 cm, P1)；9，雌雄蕊形成期幼穗 (0.5-1.0 cm, P2)；10，花粉母细胞时期形成期幼穗 (1.0-2.0 cm, P3)；11，小孢子减数分裂期小花 (3.0-4.0 mm, P4)；12，花粉母细胞减数分裂四分体时期小花 (4.0-4.5 mm, P5)；13，小孢子形成期小花 (4.5-5.0 mm, P6)；14，小孢子晚期小花 (5.0-6.0 mm, P7)；15，二细胞花粉形成期小花 (6.0-7.0 mm, P8)；16，花粉成熟期花药 (P9)。

Figure 37 Expression profiles of *PMS1T* in different tissues from 58S and NIL(MH) under long-day and short-day conditions.

In each panel, the upper row is the result from *PMS1T*, with 35 cycles of PCR amplification, and the lower row is the reference gene *Ubiquitin* (*LOC_Os03g13170*) with 25 cycles. Tissues: 1, leaf at tillering stage; 2, leaf at young panicle stage (0.5-1.0 cm); 3, penultimate internode at young panicle stage (0.5-1.0 cm); 4, leaf sheath at young panicle stage (0.5-1.0 cm); 5, flag leaf; 6, penultimate internode at P7 stage (see below description); 7, sheath of flag leaf; 8, young panicle at secondary branch primordium differentiation stage (less than 0.5 cm, P1); 9, young panicle at pistil/stamen primordium differentiation stage (0.5-1.0 cm, P2); 10, young panicle at pollen mother cell formation stage (1.0-2.0 cm, P3); 11, floret at microsporocyte meiosis stage (3.0-4.0 mm in length, P4); 12, floret at tetrad stage of pollen mother cell meiosis (4.0-4.5 mm, P5); 13, floret at microspore stage (4.5-5.0 mm, P6); 14, floret at late microspore stage (5.0-6.0 mm, P7); 15, floret at bicellular pollen stage (6.0-7.0 mm, P8); 16, stamen at mature pollen stage (P9).

从图 37 中可以看到，在所有检测的 16 个组织中，*PMSIT* 的表达量都非常低，内参基因 *Ubiquitin* 在 RT-PCR 过程中 25 个扩增循环就能看到明亮、清晰的条带，但是 *PMSIT* 的扩增循环数即使提高到 35 个循环，仍然有许多组织的样品只能看到微弱的条带。在此情况下，比较 *PMSIT* 在农垦 58S 和 NIL(MH) 的表达谱差异，发现两者中基因的表达整体趋势相同：在与营养相关的组织中（叶片，叶鞘，节间）表达量非常低，与生殖相关的组织中（幼穗，小花，雄蕊）表达量相对较高；其中，在二次枝梗分化期（P1 期）幼穗，雌雄蕊形成期（P2 期）幼穗和花粉母细胞形成期（P3 期）幼穗（图 37 中编号为 8, 9, 10）这三个组织中的表达量最高。仔细分析最高表达量的 P1, P2 和 P3 期，看到 *PMSIT* 在长日照下农垦 58S 中的表达量低于 NIL(MH)，及短日照下的农垦 58S 和 NIL(MH)。为了证实这一点，进而采用 qPCR 对幼穗发育早期的几个组织进行详细的表达量比较分析。

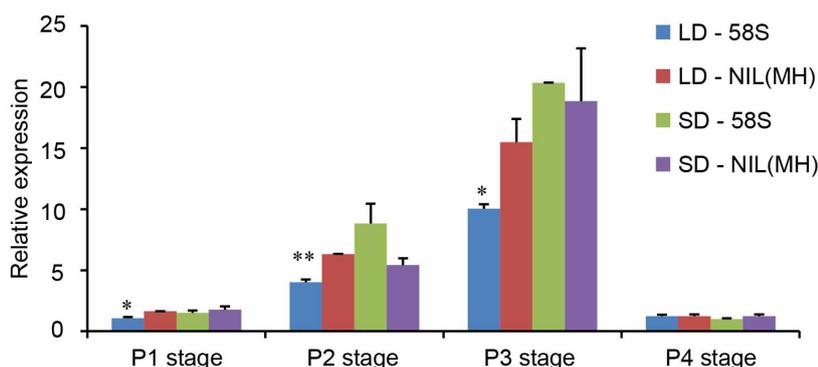


图 38 *PMSIT* 基因在长、短日照下农垦 58S 和 NIL(MH) 幼穗发育早期组织中的相对定量表达量分析。

P1 期到 P4 期组织的描述参见图*的说明，相对表达量数值为平均值 \pm SEM (n=3)。t-tests 检验 LD-58S 和 LD-NIL(MH)之间的显著性差异，*和**分别代表 $P<0.05$ 和 $P<0.01$ 。

Figure 38 Relative abundance of *PMSIT* in 58S and NIL(MH) under long-day and short-day conditions at the early stages of panicle development.

The descriptions about different stages are as in the legend of Figure*. Data are means \pm SEM (n=3). *, ** indicate statistically significant difference between LD-58S and LD-NIL(MH) by t-tests at $P<0.05$ and $P<0.01$.

qPCR 结果表明在幼穗发育早期的四个组织中，*PMSIT* 的表达量在 P2 期和 P3 期幼穗中相对较高，在 P3 期最高。之前研究表明，光敏感雄性不育的敏感时期为二次枝梗原基分化期、雌雄蕊原基形成期和花粉母细胞形成期，亦即 P1 期、P2 期

和 P3 期(Yuan et al 1988)。与图 37 RT-PCR 结果一致, *PMSIT* 的表达量在这三个时期的幼穗中确实表现为长日照下的农垦 58S 要低于长日照下的 NIL(MH), 除此之外, P2 期和 P3 期长、短日照下的农垦 58S 和 NIL(MH)的 *PMSIT* 的变化趋势相似。

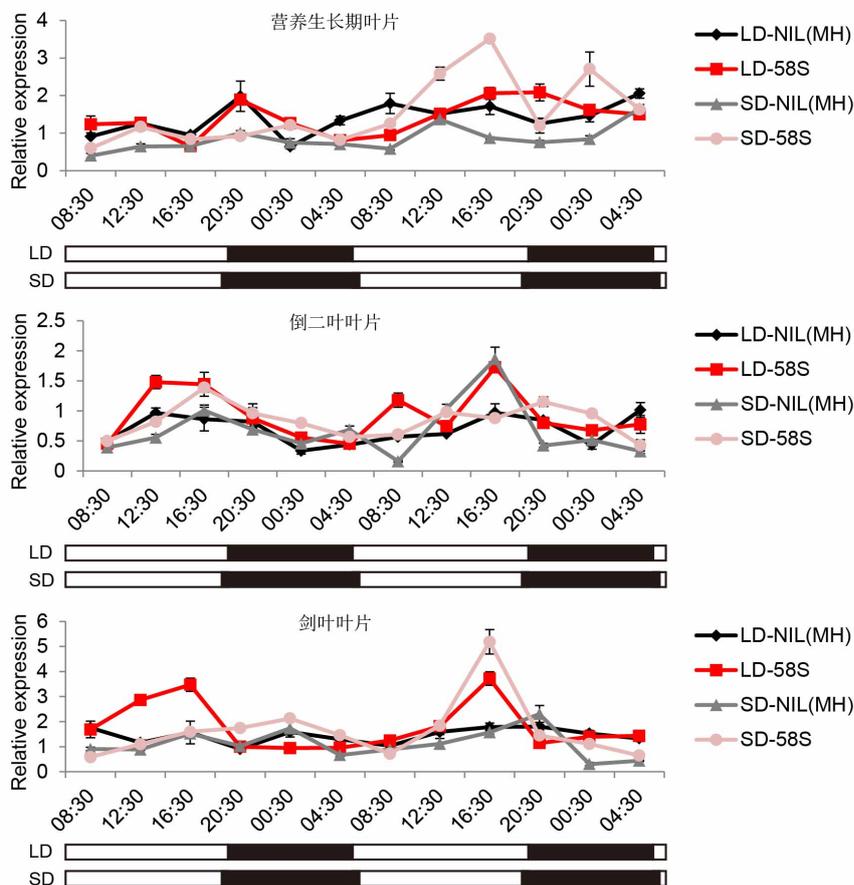


图 39 *PMSIT* 基因在长、短日照农垦 58S 和 NIL(MH) 各种叶片中的光周期和昼夜节律表达分析。

纵坐标为取样的各时间点; 下方白色矩形代表光照期, 黑色矩形代表黑暗期; 表达量数字为平均值 \pm SEM ($n=3$)。

Figure 39 The diurnal expression analysis of *PMSIT* in leaves of 58S and NIL(MH) under long-day and short-day conditions.

The y-axis indicated the time points for sample collecting. The photoperiod of long-day and short-day conditions showed at bottom, and the white and black rectangle represented the light and dark period, respectively. Data are means \pm SEM ($n=3$).

长日照和短日照是两种不同的光周期环境, 在这两种环境下与光周期开花相关基因的表达量会呈现出光周期变化和昼夜节律性。为了检测 *PMSIT* 的表达是否具有光周期和节律性, 以直接感受光照的组织——叶片为材料分不同时间点连续取材进行表达量检测。在 48 小时的光周期时间段内, 每隔 4 小时取样一次。选取的叶片

为水稻营养生长和生殖生长中有代表性的营养生长期叶片，剑叶叶片以及光敏感雄性不育的敏感时期 P2 期对应的倒二叶叶片。利用 qPCR 分析这些样品中 *PMSIT* 的表达量，结果如图 39 所示。无论是在营养生长期叶片，倒二叶叶片，还是剑叶叶片中，*PMSIT* 的表达量在农垦 58S 和 NIL(MH)长、短日照之间都没有表现出明显的差异及规律性：长日照和短日照下的表达量在任何一个亲本材料中没有差异，不表现出光周期性；而任一亲本的长日照或者短日照下两个昼夜之间的表达量不具备相似性，也看不到规律的昼夜节律性。

本室张平博对长日照下农垦 58S 遮光实验发现，在育性转换敏感期，倒二叶叶鞘对农垦 58S 的育性恢复有一定的作用(张平博等 2010)，暗示叶鞘在光周期长度感受方面可能发挥一些影响。因此，我们也对倒二叶叶鞘及剑叶叶鞘进行了 48 小时取样以检测 *PMSIT* 的表达量，方法跟之前叶片节律性分析相同。图 40 结果显示，倒二叶叶鞘和剑叶叶鞘中，*PMSIT* 的表达量在农垦 58S 和 NIL(MH)长、短日照之间同样没有表现出明显的差异及规律性，也不具有光周期效应或者节律性。

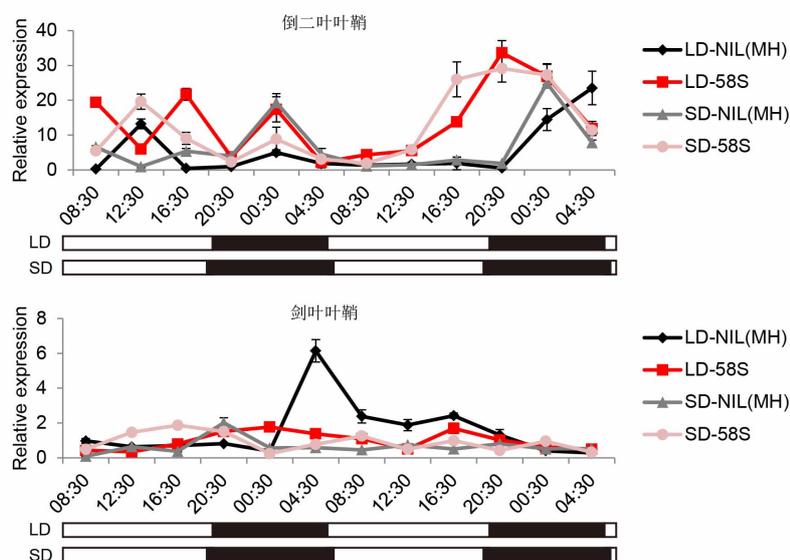


图 40 *PMSIT* 基因在长、短日照农垦 58S 和 NIL(MH) 各种叶片中的光周期和昼夜节律表达分析。

Figure 40 The diurnal expression analysis of *PMSIT* in leaf sheath of 58S and NIL(MH) under long-day and short-day conditions.

以上结果表明，无论是在育性转换敏感期的倒二叶还是剑叶的叶片和叶鞘中，*PMSIT* 的表达量都不具有光周期性和昼夜节律性，说明在光敏感雄性不育水稻农垦 58S 中，对光周期的感应和传递并不是通过 *PMSIT* 转录本的表达量来反映的，可能是通过 *PMSIT* 的其他作用形式来体现。

3.6 *PMS1T* 是一个 long non-coding RNA

应用 NCBI ORF Finder 对农垦 58S 和明恢 63 的 *PMS1T* 进行 ORF 分析, 发现可能存在三个短的 ORF, 其序列及对应位置见图 41。

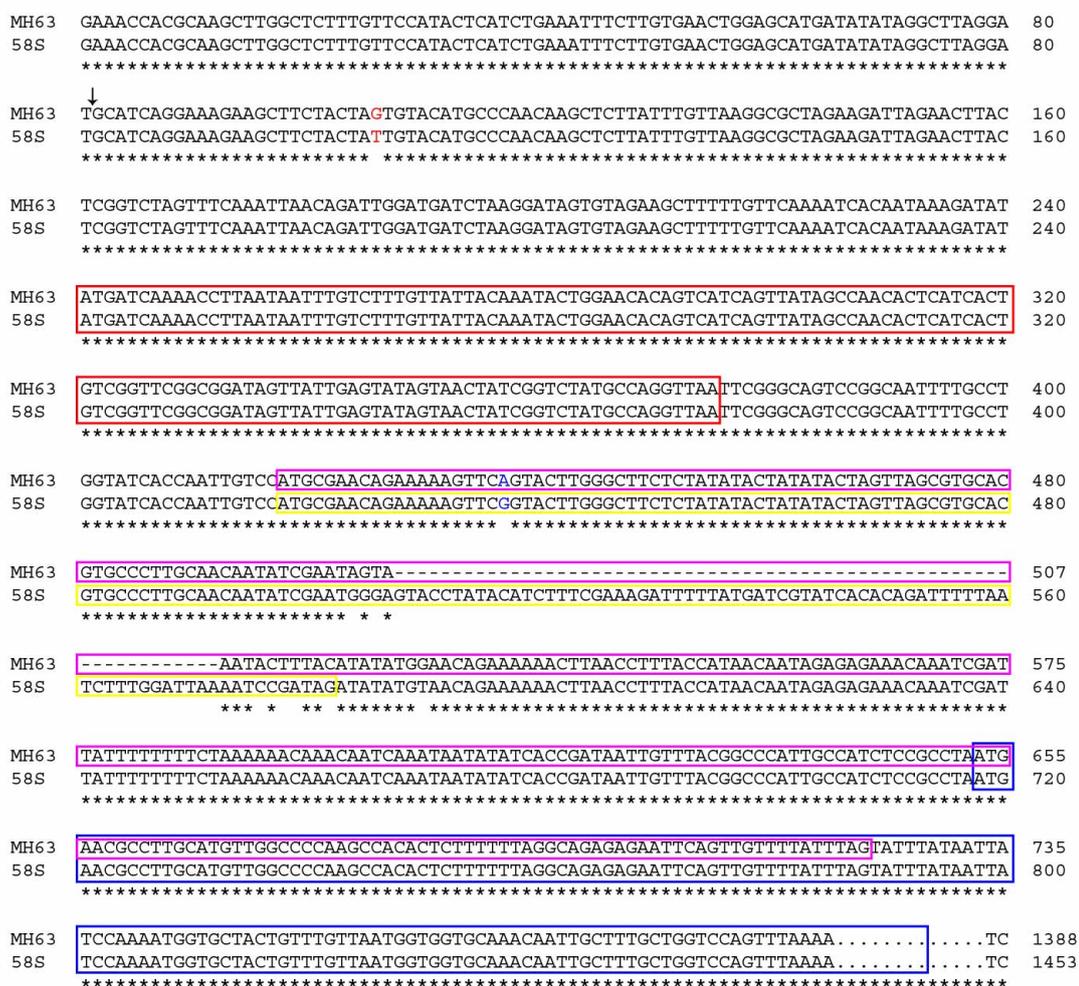


图 41 *PMS1T* 在农垦 58S 和明恢 63 中的序列比较及预测 ORFs 的示意图。

SNP S2 和 S1 分别用红色和蓝色显示, 竖箭头代表 miR2118 的剪切位点, 带颜色方框内为预测 ORF, 其中红色为 ORF1; 粉色和黄色为 ORF2M 和 ORF2S; 蓝色为 ORF3。图中 *PMS1T* 800bp 之后的序列在明恢 63 与农垦 58S 之间完全一样, 用省略号表示忽略。

Figure 41 Comparison of the genomic sequences of MH63 and 58S that encode *PMS1T*.

SNP S2 is in red and SNP S1 in blue. Vertical arrow indicates the cleavage site of miR2118. Predicted ORFs are boxed with colored lines: ORF1 is in red box, ORF2 in MH63 is in pink box, ORF2 in 58S is in yellow box, and ORF3 is in blue box overlapping with ORF2 in MH63. Sequences after 800 bp in MH63 are omitted.

ORF1 和 ORF3 在农垦 58S 和明恢 63 中的序列完全一样, 分别编码 44 和 51 个氨基酸。ORF2 在明恢 63 中 ORF2M 的编码氨基酸长度为 101, 但在农垦 58S 中由于 65 bp 序列的插入造成读码框移码和提前终止, 仅编码 54 个氨基酸, 将其命名为 ORF2S。明恢 63 中的 ORF2M 与 ORF3 有部分重叠。

对这几个 ORF 编码的蛋白质序列进行 BLASTP 搜索和分析, 均未发现有同源性较高的蛋白。此外, 多次尝试对这几个 ORF 进行原核表达和体外翻译, 但都没有得到结果, 说明这些 ORF 可能并不是真实存在, *PMSIT* 可能不编码蛋白质。为了验证此推论, 我们将 58S-C 载体进行改造, 分别在预测的三个 ORF——ORF1, ORF2S 和 ORF3 的起始密码子“ATG”后面插入碱基“G”, 以打乱 ORF 的读码框, 完全改变原有蛋白质的序列, 而相应 mRNA 的序列改变甚微。

将构建好的三个载体 58S-ORF1+G, 58S-ORF2+G 和 58S-ORF3+G 转化 NIL(MH), 分别得到 T₀代独立转化植株 60, 53 和 54 株。使用分子标记 P6 进行阳性检测, 观察长日照下这些单株的花粉碘染表型, 发现这三个转化的转基因 T₀阳性单株相对于阴性单株, 绝大部分花粉形态皱缩, 不能被染上色(图 42)。继续种植 T₁代进行共分离分析, 这三种转化材料在长日照下基因型与表型都表现为共分离, 阳性单株的小穗育性比阴性单株低(图 42), 转基因家系的结实率考察结果详见图 43。可看到, 同一家系中各转基因阳性单株的结实率下降程度也是不同的, 这种现象与未经改造的 58S-C 载体的转化结果相似。因此, 可得出结论: 改变这些 ORF 并不会影响农垦 58S 的 *PMSIT* 功能的正常行使。表明 *PMSIT* 是以 RNA 而不是蛋白质的方式发挥作用, *PMSIT* 是一个长链非编码 RNA。并且根据其相对于基因的位置, 应属于 long intergenic non-coding RNA (lincRNA)。

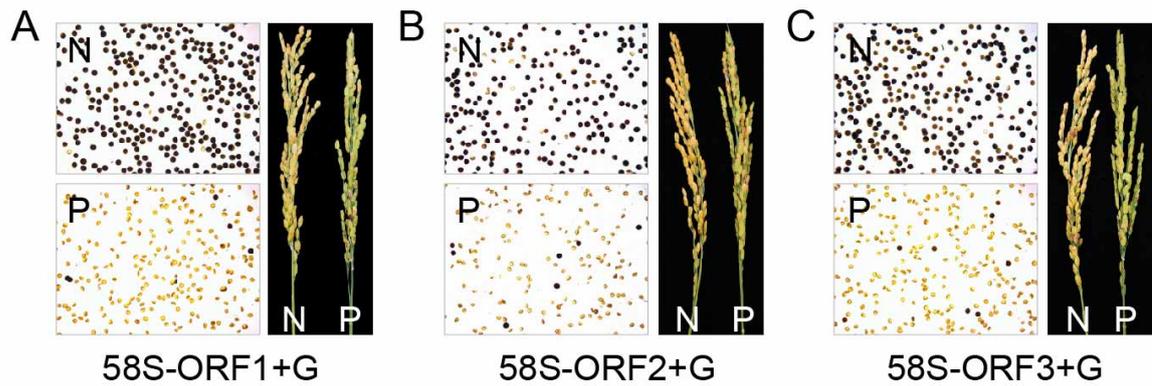


图 42 58S-ORF1+G, 58S-ORF2+G 和 58S-ORF3+G 转基因单株表型。

(A), (B) 和 (C) 中转基因 T_0 植株的花粉碘染结果显示在左侧, 右侧为 T_1 代家系中植株的成熟小穗, 其中 N 代表转基因阴性, P 代表转基因阳性。

Figure 42 The performance of 58S-ORF1+G, 58S-ORF2+G and 58S-ORF3+G transgenic plants.

The left images of (A), (B) and (C) were the results from pollen staining of transgenic T_0 plants, and the panicles of T_1 transgenic plants were showing in the right. N, negative; P, positive.

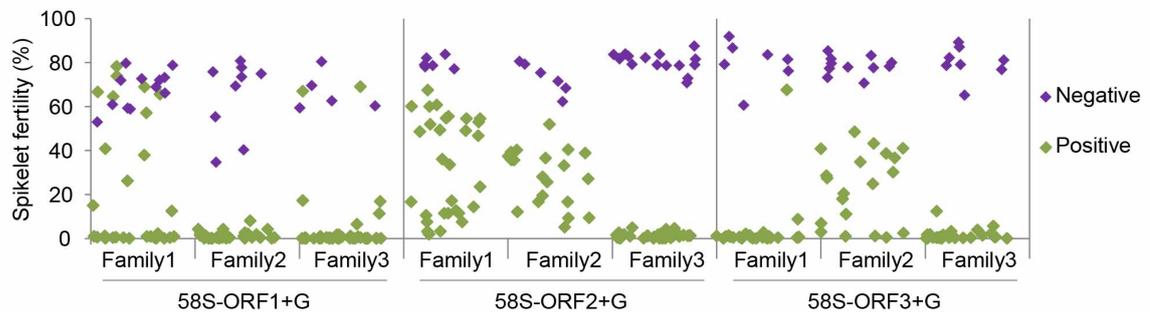


图 43 58S-ORF1+G, 58S-ORF2+G 和 58S-ORF3+G 各三个转基因 T_1 家系共分离分析。

Figure 43 The co-segregation analysis of 58S-ORF1+G, 58S-ORF2+G and 58S-ORF3+G transgenic plants of three T_1 families.

此时, 回过头来看之前得到的 *PMSIT* 表达谱分析结果, 发现 *PMSIT* 的表达特征也与大多数 lncRNA 一样: 表达量低, 并且具有组织特异性。

3.7 确定 *pms1* 的功能性序列差异

3.7.1 比较测序确定 SNP S2 是功能性序列差异

在确定了 *PMSIT* 就是 *pms1* 的基础上, 需要进一步明确农垦 58S 和明恢 63 的 *PMSIT* 序列差异中的哪一个导致最终的光敏感雄性不育表型差别, 通过对各种不同遗传背景的水稻品种进行比较测序, 根据序列差异与表型之间的相关性能推断出最终的功能性差异。从图 27B 和 C 中看到农垦 58S 和明恢 63 在 *PMSIT* 及其临近区段仅存在四处序列差异: SSR 标记 P5 位于 *PMSIT* 的启动子区; SNP S2、SNP S1 及 P6 都处于 *PMSIT* 的 5' 端。针对这些序列差异, 选取了 15 个品种进行比较测序, 其中包括 8 个由农垦 58S 转育而来的光敏雄性不育系, 结果见表 5。其中日本晴的序列是从公共数据库中下载得到, 明恢 63 的序列为已有的 BAC 21O9 测序结果, 其余品种的序列信息通过 PCR 测序获得。

本室早年遗传学分析已经确定轮回 422 和 1514 在 *pms1* 位点均为可育基因型 (Mei et al 1999b); 农垦 58 在 *pms1* 位点的基因型与农垦 58S 相同 (王风平等 1997)。在定位群体中, 位于 *PMSIT* 启动子区的 SSR 标记 P5 与 *pms1* 是共分离的, 从表 5 中看到, 这 15 个品种在 P5 处的 AT 重复数都不相同。明恢 63 中的重复数最多, 高达 40 个; 1514 和日本晴都包含 8 个重复; 农垦 58 和所有的 9 个光敏不育系的 AT 重复数都相同; 而轮回 422 和 9311 的重复数也相同, 说明 AT 的重复数在籼粳之间没有差别。比较携带 *pms1* 可育位点的品种明恢 63、轮回 422 和 1514 的 AT 重复数, 发现其相对于携带 *pms1* 不育位点的品种的 AT 重复数有多有少, 表明 AT 的重复数与 *pms1* 位点是可育还是不育没有相关性。尽管 P5 是个共分离标记, 并且在启动子区上, 但也仅仅为一处序列差异, 不太可能与 *PMSIT* 功能的发挥相关。SNP S2、SNP S1 及 P6 处的序列差异都在光敏不育系和非光敏不育系 (农垦 58 除外) 之间明显地分为两类, 但 SNP S2 和 S1 之间发生交换的一个重组单株排除了 SNP S1 和 P6 作为功能性差异的可能性。因此, 从比较测序的结果, 结合重组单株分析, 确定了 SNP S2 为 *PMSIT* 的功能性差异位点, 在 *pms1* 不育位点中为 “T”, 可育位点中为 “G”。

表 5 15 个品种在 *pms1* 区间的比较测序结果。Table 5 Comparative sequencing results of 15 rice lines in the *pms1* region.

Variety ^{a)}	P5	S2	S1	P6 ^{b)}	Sub species
Nongken 58S	(AT)14	T	G	65 bp	<i>japonica</i>
Nongken 58	(AT)14	T	G	65 bp	<i>japonica</i>
1514	(AT)8	G	A	—	<i>japonica</i>
Nipponbare	(AT)8	G	A	—	<i>japonica</i>
Lunhui 422	(AT)9	G	A	—	<i>japonica</i>
Minghui 63	(AT)40	G	A	—	<i>indica</i>
9311	(AT)9	G	A	—	<i>indica</i>
Peiai 64S	(AT)14	T	G	65 bp	<i>indica</i>
32001S	(AT)14	T	G	65 bp	<i>indica</i>
HN5S	(AT)14	T	G	65 bp	<i>indica</i>
95076S	(AT)14	T	G	65 bp	<i>japonica</i>
N5088S	(AT)14	T	G	65 bp	<i>japonica</i>
7001S	(AT)14	T	G	65 bp	<i>japonica</i>
29130S	(AT)14	T	G	65 bp	<i>japonica</i>
31301S	(AT)14	T	G	65 bp	<i>japonica</i>

a) 品种名称后面的后缀“S”代表光温敏感雄性不育系； b) 短横线表示在 P6 处缺失了 65 bp。

a) Suffix letter “S” indicates PSMS line; b) Dash represents the deletion of 65 bp at P6.

3.7.2 65 bp 插入对光敏感雄性不育的影响

之前的重组单株数据已排除 P6 处的 65 bp 插入作为 *PMSIT* 的功能性差异位点，但该处差异明显，此段多出来的序列到底来自于何处？虽然不能决定 *PMSIT* 的光敏感不育性，但是否会对 *PMSIT* 的功能产生一定的影响？因此，有必要对这段序列进行深入分析。

Organism	Blast Name	Score	Number of Hits	Description
Oryza	monocots		101	
. Oryza sativa	monocots		1	
. . Oryza sativa Indica Group	monocots	1568	16	Oryza sativa Indica Group hits
. . Oryza sativa Japonica Group	monocots	1546	81	Oryza sativa Japonica Group hits
. Oryza rufipogon	monocots	152	3	Oryza rufipogon hits
. Oryza sativa	monocots	150	1	Oryza sativa hits

图 44 58S *PMSIT* 的 discontinuous MEGABLAST 和 BLASTN 搜索结果系谱报告。Figure 44 The lineage report of 58S *PMSIT* hits according discontinuous MEGABLAST or BLASTN.

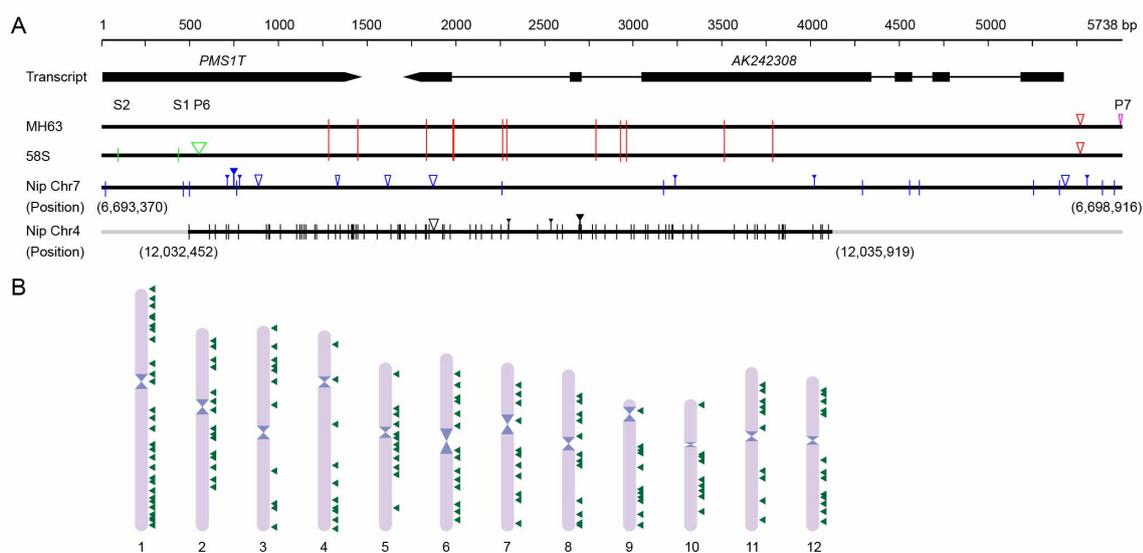


图 45 *PMSIT* 基因组区段序列特征。

A, *PMSIT* 基因组序列比较结果。灰色横线代表该区间日本晴第四染色体与 *PMSIT* 附近区段完全没有相似性；括号中的数字代表用于分析的日本晴序列起始和终止位点所在的染色体位置；序列差异用以下颜色和标记表示：短竖线，单核苷酸差异；三角符号，多个碱基插入；带三角形的竖线，单个或两个碱基插入；黑色，日本晴第四染色体特有的序列；蓝色，日本晴第七染色体特有的序列；红色，农垦 58S 和明恢 63 共同特有的序列；绿色，农垦 58S 特有；粉色，明恢 63 特有。B, 65 bp 序列在整个基因组上可匹配到的位点示意图，该图根据 Gramene 网站输出的 HSP 分布图修改得到。

Figure 45 The sequence features of the *PMSIT* genomic region.

A, Genomic sequence alignment of *PMSIT*. The nonhomologous sequences outside this region were marked in grey lines. The start and end points of Nipponbare (Nip) genomic fragment from *PMSIT* to the P7 marker are indicated in the parentheses. The sequence differences are indicated in various colors and signs: vertical short line, single nucleotide difference; triangle, large segment insertion; vertical short line with triangle, single or double nucleotide insertion; black, specific for Nipponbare chromosome 4; blue, specific for Nipponbare chromosome 7; red, specific for 58S and MH63; green, specific for 58S; pink, specific for MH63. B, Genome-wide distribution of matches from the BLASTN results of the 65 bp insertion sequence found in 58S, using a cut-off of 10 in E-value. The diagram was modified from the HSP distribution image output generated from the search result of the Gramene website (<http://www.gramene.org/>).

利用 NCBI 的 discontinuous MEGABLAST 和 BLASTN 分析农垦 58S 的 *PMSIT* 全长序列，得到相同的结果，仅在稻属 (*Oryza*) 中找到与其相似的序列 (图 44)，*O. rufipogon* 中匹配到的几条相似序列恰好位于 P6 处的插入序列区段。同时，发现来自于水稻日本晴基因组第四染色体上的一段序列与 *PMSIT* 高度相似，将之与农垦

58S 和明恢 63 及日本晴 *PMSIT* 基因组区段进行了比较分析。日本晴第四染色体上 12,032,452 到 12,035,919 之间的序列与第七染色体上 *PMSIT* 3'端及其下游序列之间仅存在 SNP 和少数碱基插入/缺失的差异 (图 45A)。而 65 bp 序列在整个基因组上能找到几百个高度相似区段, 并且均匀分布在各条染色体上 (图 45B)。大部分位于基因间区和启动子, 少数位于基因上, 其中包含一个预测编码转座子蛋白的 LOC_Os12g06040, 离 *PMSIT* 距离最近的两个基因也都注释为反转座子蛋白。因此, 我们推测农垦 58S 中插入的 65 bp 可能由转座子或者反转座子插入造成。

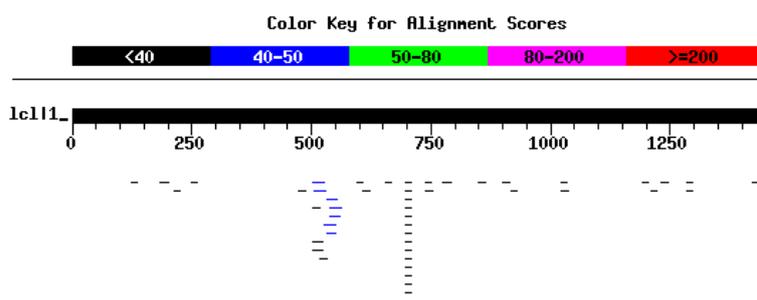


图 46 农垦 58S *PMSIT* 在 CSRDB 的 BLAST 结果。

Figure 46 The result of BLAST at CSRDB database using 58S *PMSIT* as query.

对另一个光敏感雄性不育基因 *pms3* 的研究表明, 该基因也是一个 lncRNA, 能产生 small RNA (Ding et al 2012b; Zhou et al 2012)。为了分析 *PMSIT* 是否也会产生 small RNA, 在水稻玉米 small RNA 数据库 CSRDB (<http://sundarlab.ucdavis.edu/smrnas/>) 中对农垦 58S 的 *PMSIT* 进行 BLAST 分析。有多条 small RNA 能够匹配到 *PMSIT* 上, 其中比值最高的几条 small RNA 恰好都位于 65 bp 内 (图 46)。同时在本室已有的 58S 雌雄蕊原基形成期幼穗 small RNA 数据库中也找到多条 small RNA 匹配到该区间。继续在 CSRDB 网站上分析此前 BLAST 得到的几条 small RNA 的靶基因, 发现它们拥有共同的靶基因 *LOC_Os03g14590*, 基因注释为 calcium-binding EF hand family protein。利用本室水稻全生育期表达谱芯片数据库分析该基因的表达, 发现其在各时期的叶片和叶鞘中的表达量最高。叶片和叶鞘是水稻直接感受光照的部位, 那么, *LOC_Os03g14590* 在叶片中的表达是否具有光周期或者节律性? qPCR 结果 (引物对为 CB-F/CB-R) 显示在农垦 58S 长、短日照 48 小时的两个昼夜循环中, *LOC_Os03g14590* 的表达量表现出光周期性: 在营养生长期叶片中, 长日照下的表达丰度低于短日照 (图 47A); 但在倒二叶叶片中情况恰好相反。而且, 属于育性转换敏感时期的倒二叶叶片中的

表达还呈现出昼夜节律性，在重复的两天内各时间点的表达趋势相同，光照条件下的表达量逐渐上升，到黑暗条件下开始下降（图 47B）。另外，分析 *LOC_Os03g14590* 在农垦 58S 幼穗发育早期和雄蕊中的表达量，在这些组织中，长日照下的表达丰度也都低于短日照下（图 47C）。虽然没有证据表明 *LOC_Os03g14590* 就是 65 bp 内可能产生的 small RNA 的靶基因，但从上面的结果推测，此 65 bp 是否与农垦 58S 的光周期相关？如果去掉该段序列，对农垦 58S 的长日照下的育性是否有影响？因此，我们针对该区间构建了一系列的载体进行遗传转化分析。

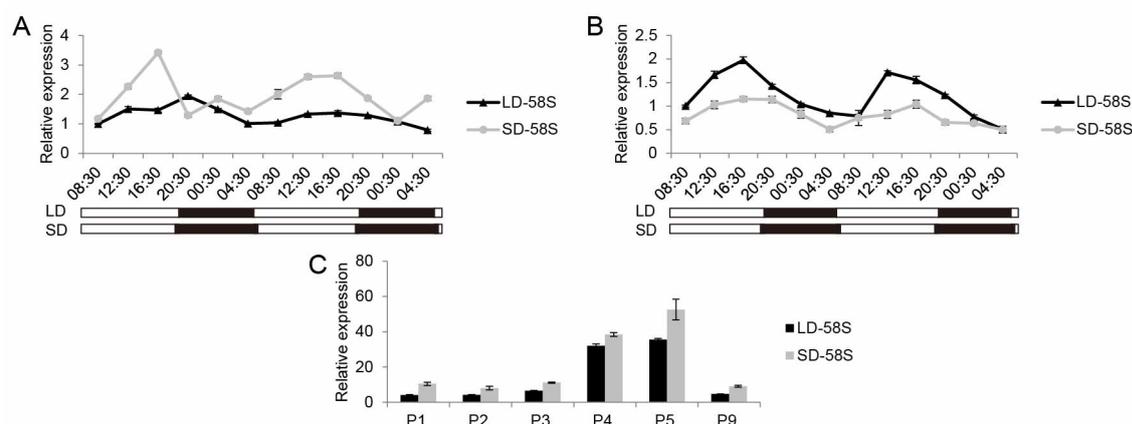


图 47 *LOC_Os03g14590* 基因在农垦 58S 长、短日照下的表达量分析。

A, 在营养生长期叶片 48 小时内的表达量分析；B, 在倒二叶叶片 48 小时内的表达量分析；C, 幼穗发育各时期的表达量分析。

Figure 47 The expression level of *LOC_Os03g14590* at 58S under long-day and short-day conditions.

A, leaf at tillering stage; B, leaf at young panicle stage (P2 stage); C, young panicle development stages.

首先，通过对互补载体 58S-C 进行改造，将其含有 65 bp 的区间替换为明恢 63 对应的区段，具体步骤参见“材料和方法”，得到终载体 58S-C-dP6。58S-C-dP6 与 58S-C 包含的外源片段相比，仅缺少了 P6 处的 65 bp（图 12）。将该载体转化 NIL(MH)，得到转基因 T_0 单株 33 株，分析考察 T_1 代转基因家系的结实率。在武汉长日照下，各 T_1 代独立转化家系中阳性单株的结实率相比阴性单株有不同程度的下降，不同家系中下降的程度会有所不同（图 48）。这种共分离结果与载体 58S-C 的转化结果相同（图 30A 和 C），说明农垦 58S *PMSIT* 去掉了 65 bp 后，在长日照下仍能行使正常功能，降低 NIL(MH) 的育性。同时为了验证 65 bp 对光周期的响应，分析 58S-C-dP6 在短日照下的表型。为了尽量减少温度对育性的影响，选择在海南

高温短日照下进行考种,以便与武汉高温长日照下的结果进行比较。图 48 结果显示,在海南高温短日条件下, T_1 转基因家系中,虽然阴性单株的结实率都不是很高,但显然阴性和阳性单株之间的结实率并没有明显差异,与 58S-C 在短日照下的表现结果,表明去掉了 65 bp 后也不影响农垦 58S *PMSIT* 在短日照下的功能。从以上结果推论,认为农垦 58S *PMSIT* 中 65 bp 的插入片段与光敏感雄性不育水稻农垦 58S 的光周期敏感性可能无关,

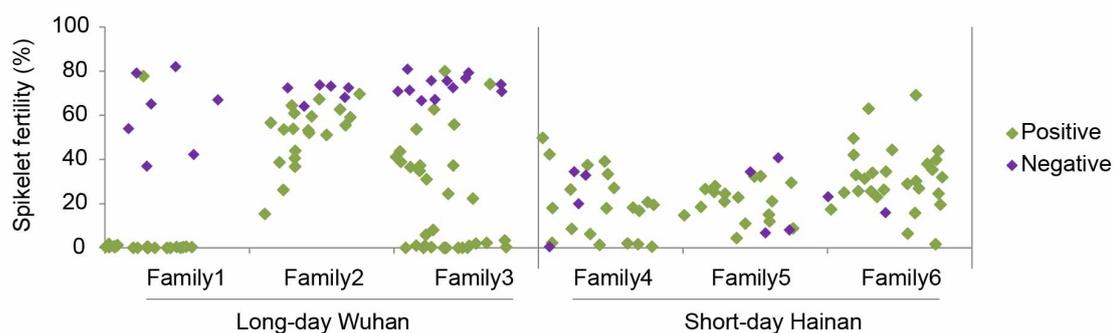


图 48 58S-C-dP6 转基因 T_1 代家系在武汉长日照和海南短日照下的结实率。

Figure 48 The spikelet fertility analysis of 58S-C-dP6 transgenic plants in T_1 families under long-day conditions at Wuhan and short-day conditions at Hainan.

此外,为了验证 65 bp 处可能产生的 small RNA 是否对育性有影响,针对该区间构建了三个超表达载体 Ubi:S CS+P6, Ubi:S P6 和 Ubi:S P6+polyA。Ubi:S P6 的外源片段包括农垦 58S *PMSIT* 的 65 bp 及其周围序列共 288 bp; Ubi:S CS+P6 和 Ubi:S P6+polyA 包含的片段是从该 288 bp 分别向上游延伸至 *PMSIT* 的转录起始位点和向下游延伸至 *PMSIT* 的转录终止位点(图 16)。同样将这三个超表达载体分别转化 NIL(MH), 得到 Ubi:S CS+P6, Ubi:S P6 和 Ubi:S P6+polyA 转基因 T_0 单株数为 80, 50 和 40 株。对这些单株进行 *PMSIT* 的表达量检测和花粉育性观察,发现长日照下 Ubi:S CS+P6 转基因阳性单株的花粉育性低于阴性单株,而 Ubi:S P6 和 Ubi:S P6+polyA 转基因阳性单株的花粉育性跟阴性单株的表现一样,高度可育(图 49)。检测这些单株中外源片段的表达量,将超量表达的单株(图 50)继续种植 T_1 代。

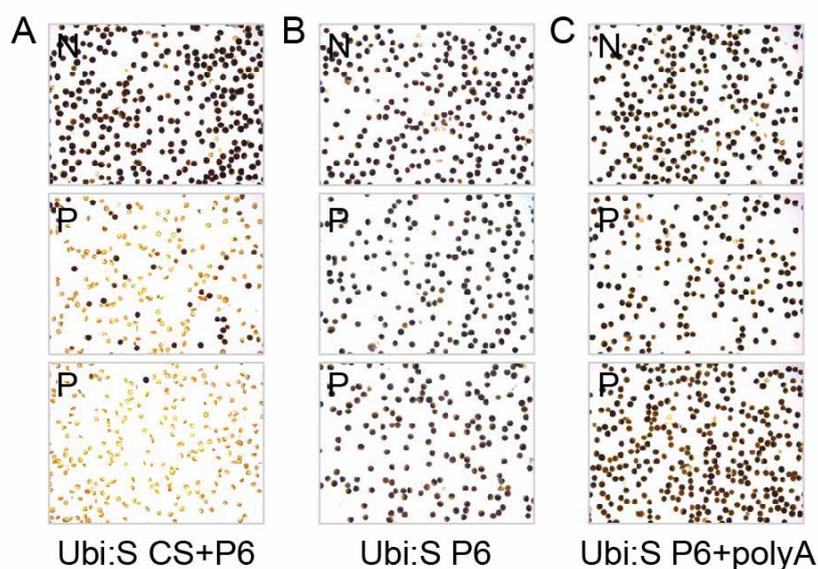


图 49 Ubi:S CS+P6, Ubi:S P6 和 Ubi:S P6+polyA 转基因 T₀ 植株长日照下花粉碘染。

Ubi:S CS+P6 (A), Ubi:S P6 (B) 和 Ubi:S P6+polyA (C) 中 N 代表转基因阴性单株, P 代表转基因阳性单株。

Figure 49 The pollen fertility of Ubi:S CS+P6, Ubi:S P6 and Ubi:S P6+polyA transgenic T₀ plants under long-day conditions.

N, negative transgenic T₀ plants of Ubi:S CS+P6 (A), Ubi:S P6 (B) and Ubi:S P6+polyA (C); P, positive ones.

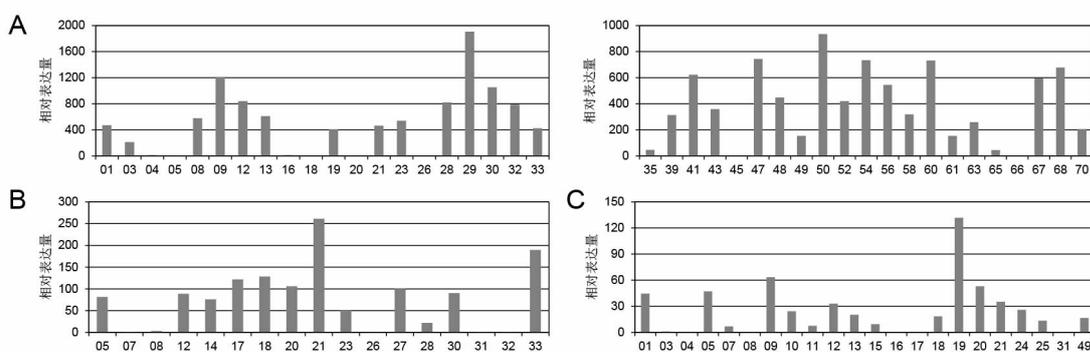


图 50 Ubi:S CS+P6, Ubi:S P6 和 Ubi:S P6+polyA 转基因 T₀ 代单株相对表达量检测。

A, Ubi:S CS+P6; B, Ubi:S P6; C, Ubi:S P6+polyA。

Figure 50 The relative expression level of *PMS1T* in T₀ transgenic plants of Ubi:S CS+P6, Ubi:S P6 and Ubi:S P6+polyA.

武汉大田长日照条件下, 仅超表达 65 bp 及其附近序列的 Ubi:S P6 的 T₁ 代家系中, 转基因阳性单株与阴性单株的结实率差不多, 两者之间没有明显区别, 都表现为可育 (图 51); 同样地, Ubi:S P6+polyA T₁ 家系中阳性单株也是完全可育的。该

结果表明农垦 58S 中插入的 65 bp 可能对育性没有直接影响。另一方面, Ubi:S P6 的转化片段中也包含了完整的 ORF2S, Ubi:S P6+polyA 中则包含了 ORF2S 和 ORF3, 该转化结果也进一步表明预测的 ORF2 和 ORF3 没有功能。

但在 Ubi:S CS+P6 转基因家系中, 阳性植株的结实率都低于阴性单株; 而且在各个家系中, 阴性单株完全可育, 阳性单株育性极低 (图 51), 与 Ubi:S 转化表现出相同的效应 (图 36)。相对于 Ubi:S 包含完整的 *PMS1T* 转录本, Ubi:S CS+P6 仅含有从 *PMS1T* 转录本转录起始位点开始的 696 bp 序列, 说明该段序列包含有与光敏感雄性不育相关的重要元件, 且 *PMS1T* 发挥正常的功能并不需要完整的转录本。

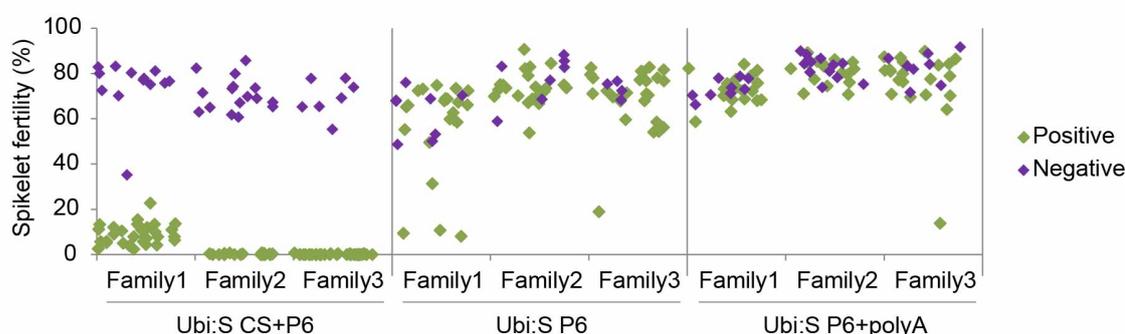


图 51 Ubi:S CS+P6, Ubi:S P6 和 Ubi:S P6+polyA 各三个转基因 T_1 家系长日照下共分离分析。

Figure 51 The co-segregation analysis of Ubi:S CS+P6, Ubi:S P6 and Ubi:S P6+polyA transgenic plants of three T_1 families under long-day conditions.

3.8 *PMS1T* 是 PHAS 基因

3.8.1 *PMS1T* 是 miR2118 的靶基因

3.8.1.1 实验验证 *PMS1T* 被 miR2118 剪切

研究表明许多 lncRNA 会被 microRNA 识别, 两者相互作用, 在调控基因表达上发挥重要的功能(Yoon et al., 2012; Leucci et al., 2013; Liu et al., 2013; Yoon et al., 2013)。而在水稻中, 通过大量数据分析发现很多在发育的小穗中优势表达的 lncRNA 是 microRNA2118 (miR2118) 或者 miR2275 潜在的靶基因(Johnson et al., 2009; Song et al., 2012)。因此, 利用植物 small RNA target 分析数据库 psRNATarget (<http://plantgrn.noble.org/psRNATarget/>) 对农垦 58S 和明恢 63 的 *PMS1T* 全长序列

进行分析,发现在 *PMSIT* 的 5' 端存在共同的 miR2118 识别位点,可能是 miR2118d 或者 miR2118o 的靶基因(图 52)。

miRNA Acc.	Target Acc.	Expectation (E)	Target Accessibility (UPE)	Alignment	Target Description	Inhibition	Multiplicity
osa-miR2118d	Your Sequence	3.0	20.41	miRNA 22 AMDCGIAACCCUCCGAGUCUW 1 Target 70 UAGGCUUAGGAUCCAUCAAGAA 91		Translation	1
osa-miR2118o	Your Sequence	3.0	20.41	miRNA 22 AMDCGIAACCCUCCGAGUCUW 1 Target 70 UAGGCUUAGGAUCCAUCAAGAA 91		Translation	1

图 52 psRNATarget 数据库预测的 *PMSIT* 可能的 miRNA 结合位点。

搜索时采用默认参数, Expectation 为 3.0, UPE 为 25.0。

Figure 52 The predicted small RNA target sites of *PMSIT* analyzed by psRNATarget server.

All the parameter was the default with the Expectation 3.0 and UPE 25.0.

为了验证 *PMSIT* 是否能被 miR2118 识别并且介导剪切, 分别以长日照下农垦 58S 和 NIL(MH) P3 期的幼穗 RNA 为模板进行 modified 5' RLM-RACE。根据预测结果, 设计了基因特异引物 RACE5-3 和 RACE5-4 完成两轮巢式 PCR, 结果如图 53 所示。将第二轮 PCR 产物连接到 T-vector 上后, 进行测序, 与 *PMSIT* 序列进行比较, 分析剪切发生的位点并计算各位点的剪切概率。

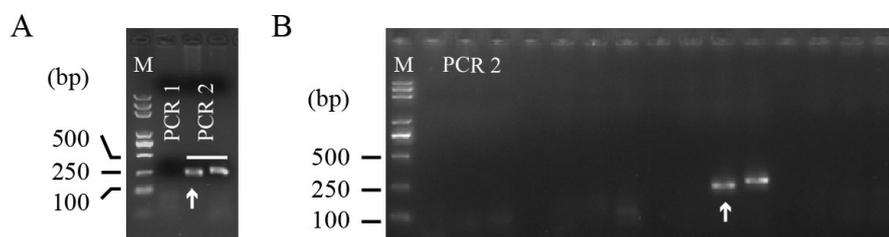


图 53 *PMSIT* 5' RLM-RACE PCR 扩增结果。

最左边为 *Trans2K Plus* DNA marker; 白色箭头指示用于测序的条带, 所用的扩增模板分别为长日照下农垦 58S P3 期幼穗 (A) 和 NIL(MH) B 为 P3 期幼穗 (B); PCR1 和 PCR2 分别为巢式 PCR 扩增的第一轮和第二轮 PCR 结果。

Figure 53 The nested PCR amplification of *PMSIT* according modified 5' RLM-RACE.

The left line in each image was the *Trans2K Plus* DNA marker; white arrows indicated the bands used for sequencing, the total RNA from young panicles of Nongken58S (A) and NIL(MH) (B) under long days were processed for nested PCR amplification. PCR1, outer 5' RLM-RACE PCR; PCR2, inner 5' RLM-RACE PCR.

分析 5' RLM-RACE 测序结果发现恰好在预测的 miR2118 结合位点上检测到 *PMSIT* 被剪切, 并且所有的剪切都发生在 miR2118 结合位点的第 10 到 11 位碱基之间(图 54), 证实了农垦 58S 和 NIL(MH)的 *PMSIT* 确实是 miR2118 的靶基因。但

还不能明确是该家族中的哪些成员参与了 miRNA 介导的剪切, 因为所有的 miR2118 序列相似性程度非常高, 但 miR2118d 和 miR2118o 的 Expectation 值最低, 可信度最高。

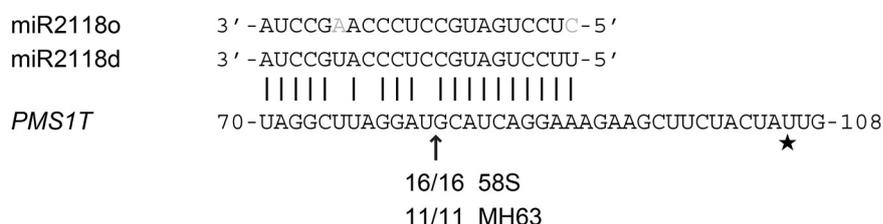


图 54 5' RLM-RACE 证实 miR2118 介导 *PMS1T* 被剪切。

箭头指示剪切发生的位点, 下面的数字表明在农垦 58S 和 NIL(MH)中剪切的频次; miR2118o 与 miR2118d 有差异的序列显示为灰色; 星号标注了 SNP S2 的位置。

Figure 54 Validation by 5' RLM-RACE that *PMS1T* is targeted by miR2118.

The arrowhead indicates the cleavage site and sequence frequencies are shown below. The differences between miR2118o and miR2118d were marked in grey. Star marks the position of S2 SNP.

同时, 我们还采用 PARE 的方法分析验证 *PMS1T* 序列被剪切的情况。PARE 是基于降解组进行的测序, 相对于 5' RLM-RACE 仅分析某个特异位点可能发生的剪切, PARE 是在全基因组水平上分析序列的剪切。分别以长、短日照下农垦 58S 和 NIL(MH)的 P3 期幼穗总 RNA 为模板构建 PARE 文库, PARE 文库信息详见表 6。将所有的 signature 序列的 reads 数采用 TP30M 进行平衡化后与基因组序列比对, 分析这些 signature 在 *PMS1T* 区间的 reads。结果发现信号值最高的位点就是 *PMS1T* 预测的 miR2118 介导的剪切位点, 并且在该位点上长日照下农垦 58S 中的信号值高于短日照及长、短日照下的 NIL(MH) (图 55)。

表 6 PARE 文库信息。

Table 6 The information of PARE libraries.

Condition	Material	Total Sequences	Genome Matched Reads	Distinct Genome Matched Reads	t/r/sn/sno RNA Matched Reads	Norm. Base Value	Max. Len. of Reads	Min. Len. of Reads
Long-day	NIL(MH)	34,120,199	30,548,637	5,850,129	397,756	30M	21	18
	58S	42,420,515	38,673,590	6,853,163	410,218	30M	21	18
Short-day	NIL(MH)	35,826,997	32,282,105	5,640,352	664,820	30M	21	18
	58S	35,575,232	32,607,464	6,417,295	344,834	30M	21	18

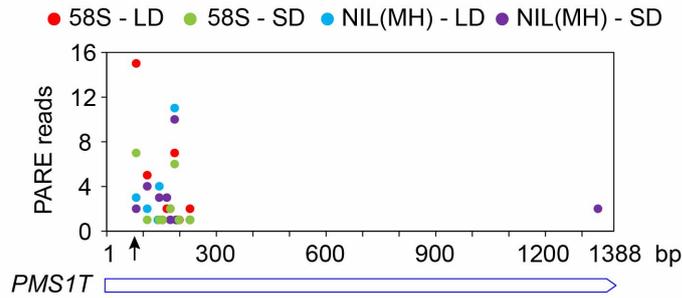


图 55 长、短日照下农垦 58S 和 NIL(MH) P3 期幼穗 PARE 文库中 *PMS1T* 被剪切的结果。

箭头表明了 miR2118 介导的剪切位点；所有的 reads 数均采用 TP30M 进行平衡化。

Figure 55 PARE results of *PMS1T* in 58S and NIL(MH) at P3 stage under long-day and short-day conditions.

Arrow indicates the cleavage sites directed by miR2118. The values of reads in each library are normalized to transcripts per 30 million (TP30M). LD, long-day conditions; SD, short-day conditions.

以上两方面的实验数据均证实了 *PMS1T* 是 miR2118 的靶基因，并且能够被 miR2118 介导剪切。

3.8.1.2 miR2118 家族表达量分析

miRBase (<http://www.mirbase.org/>) 数据库提供的信息显示水稻的 miR2118 是一个大家族，共有 18 个成员，编号从 miR2118a 到 miR2118r，大部分在第 4 染色体上，形成一个大的 cluster；miR2118p、q 和 r 则位于第 11 染色体上（图 56A）。所有的 miR2118 成熟序列长度均为 22 nt，其中某些 miR2118 成员的成熟序列完全相同（图 56B）(Johnson et al 2009; Song et al 2012a)。miR2118 成员的表达具有组织特异性，仅在未成熟的幼穗中检测到有丰富的表达(Johnson et al 2009; Song et al 2012b)，而 *PMS1T* 也是在幼穗发育早期中拥有相对较高的表达丰度。

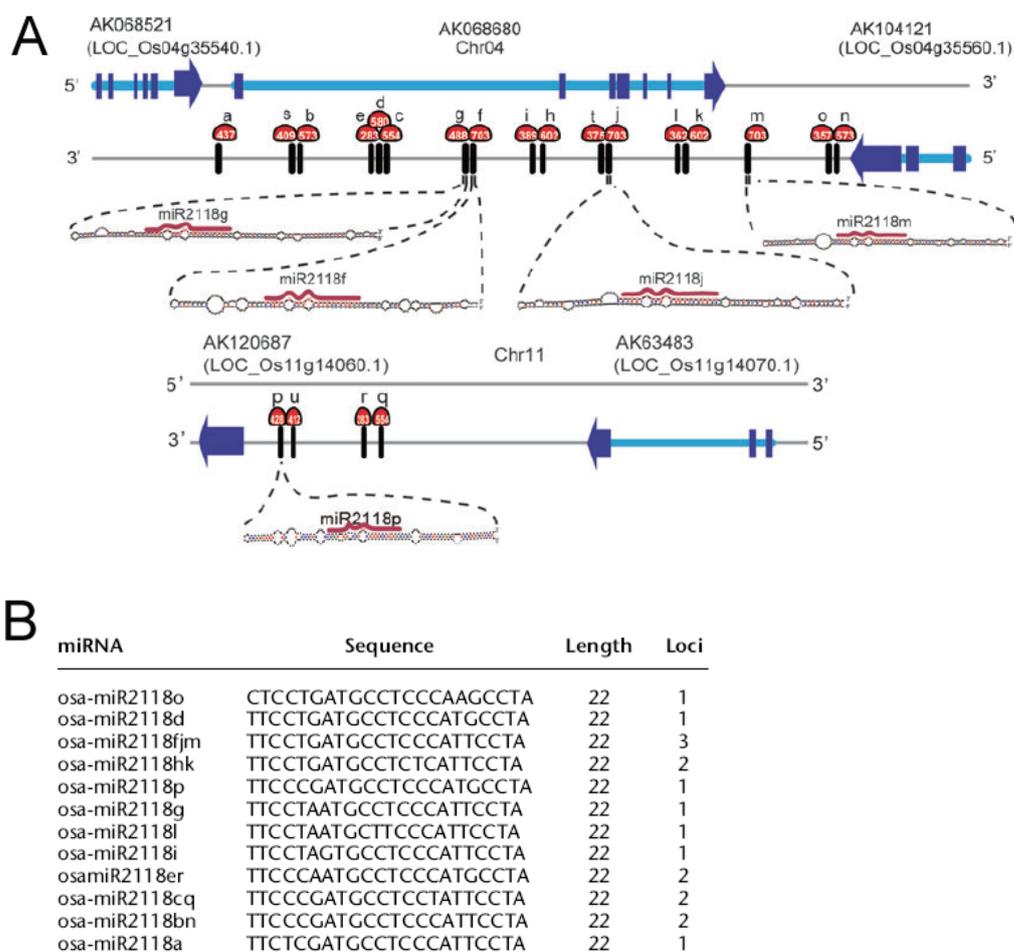


图 56 miR2118 家族序列及其在基因组上的位置。

A, miR2118 家族成员在 93-11 基因组上的位置, 来源于 Song et al 2012a; B, miR2118 家族成员序列信息, 来源于 Johnson et al 2009。

Figure 56 The sequences and locations of miR2118 family members.

A, The locations of miR2118 in the 93-11 genome, from Song et al 2012a; B, The sequences of miR2118, from Johnson et al 2009.

为了分析在幼穗发育早期过程中 small RNA 的表达情况, 我们构建了长、短日照下农垦 58S 和 NIL(MH)幼穗发育早期 P2 期、P3 期和 P4 期的幼穗或小花的 small RNA 文库, 每份样品有两次生物学重复, 文库测序信息详见表 7, 得到的 small RNA reads 数是经过 TP10M 进行平衡化后的结果。此外, 表中“Genome Matched Reads”和“Distinct Genome Matched Reads”的数值并不包含“t/r/sn/snoRNA Matched Reads”中的数值。在这三个时期的样品中检测到各种长度的 small RNA 的表达, 分析其中已知的 miRNA 的表达情况发现, 22 nt 的 miRNA 中仅有 miR2118、miR2275 和

miR1859 家族成员有表达。其中 miR2118 家族中所有成员都能检测到较高的表达丰度，具体表达情况如图 57 所示。

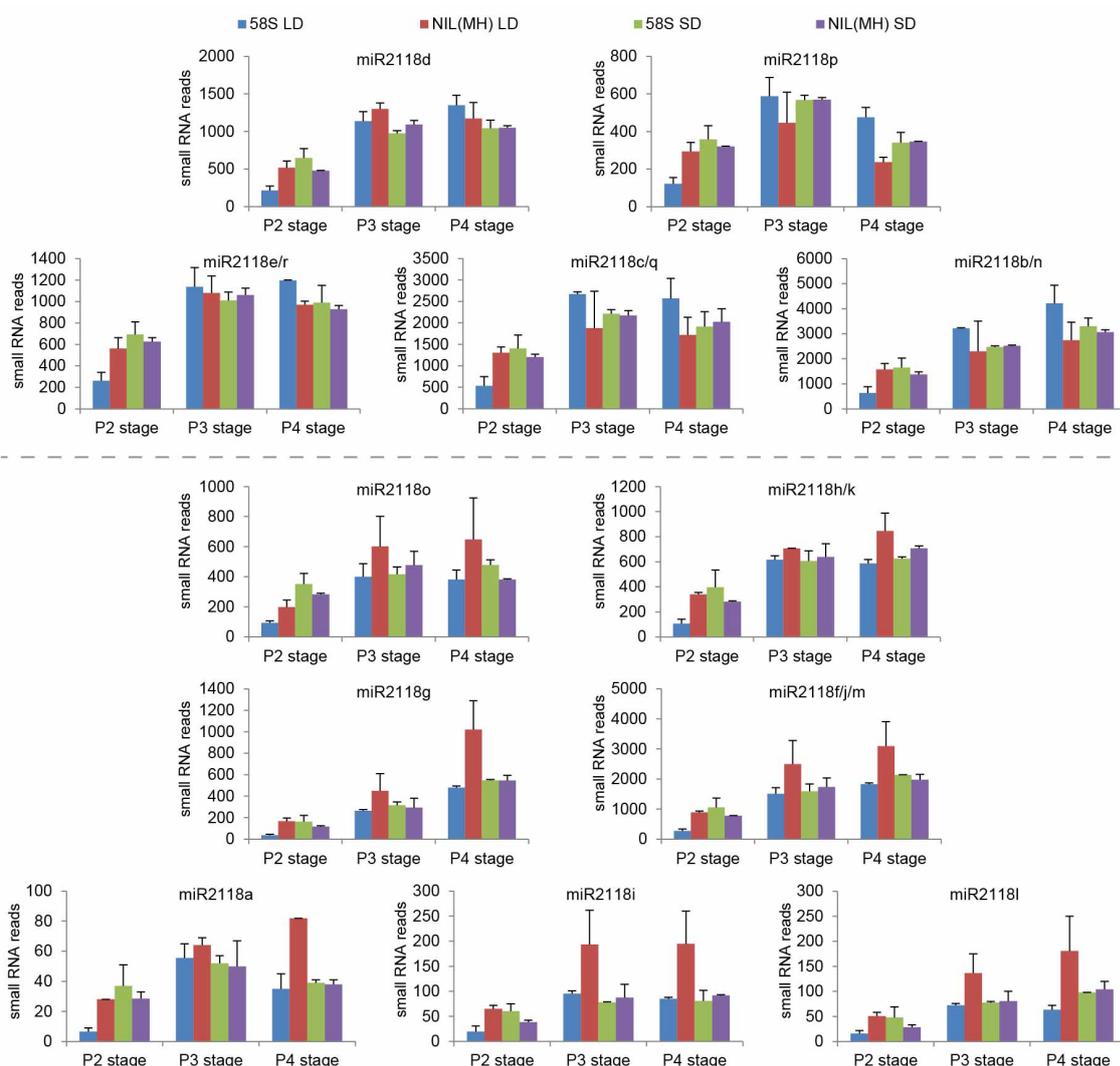


图 57 miR2118 家族在幼穗发育早期的表达量分析。

数值为两次生物学重复的平均值，误差线为 s.e.m。LD，长日照；SD，短日照。

Figure 57 Abundance of miR2118 in panicles at three stages of young panicle development.

Values are means of two biological replicates, and error bars represent the s.e.m. LD, long-day conditions; SD, short-day conditions.

18 个不同前体的 miR2118 被加工成 12 条成熟 miRNA，这些成熟 miRNA 虽然表达丰度有很大差别，但也呈现出相似的规律性：从整体上看，他们在 P3 期和 P4 期的表达量都明显高于 P2 期；P3 期和 P4 期拥有相似的表达模式，并且表达丰度差不多；所有的 miR2118 成员在长、短日照下在农垦 58S 和 NIL(MH) P2 期的表达趋势相同，表现为长日照下农垦 58S 中的表达丰度低于其他三种情况。此外，根据 P3

期和 P4 期中的表达模式差异可以将这些 miRNA 明显地分为两类：图 57 上部分显示的 5 个成熟体 miR2118 在 P3 期和 P4 期的长日照下农垦 58S 中的表达量最高；而图 57 下半部分的 7 个成熟体 miR2118 显示在短日照的农垦 58S 中表达量最高。miR2118d 和 miR2118o 可分别划分到这两个类群中。

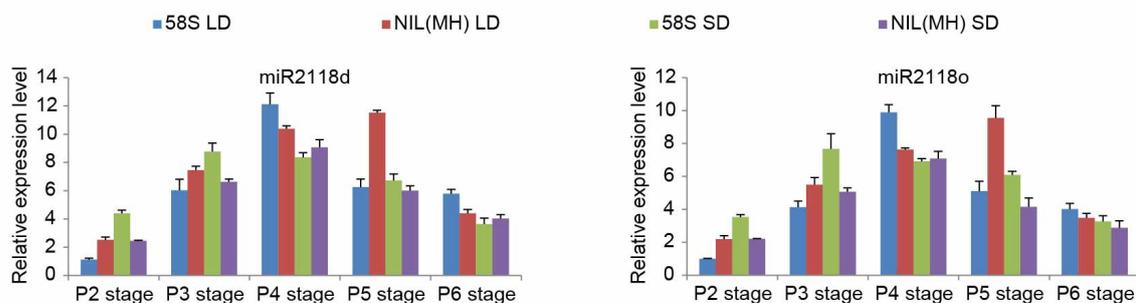


图 58 miR2118d 和 miR2118o 在长、短日照下农垦 58S 和 NIL(MH) 中幼穗发育各时期的相对表达量分析。

Figure 58 The relative expression level of miR2118d and miR2118o in panicles of 58S and NIL(MH) at long-day and short-day conditions.

我们进一步采用 stem-loop RT-PCR 的方法验证、分析 miR2118d 和 miR2118o 在幼穗发育各时期中的表达模式。stem-loop RT-PCR 反转录过程中用到的引物分别为 miR2118d-RT 和 miR2118o-RT, qPCR 时分别用引物 miR2118-qR 与 miR2118d-qF 和 miR2118o-qF 进行扩增。结果显示, 在幼穗发育的 P2 期到 P6 期, miR2118d 和 miR2118o 在长、短日照下农垦 58S 和 NIL(MH) 中的相对表达量趋势完全一样。从 P2 期开始表达量逐渐升高, 到 P4 期达到最高, 然后逐渐开始降低 (图 58)。P2 期和 P3 期中 miR2118d 和 miR2118o 在长日照农垦 58S 中的表达量低于 NIL(MH), 这与 small RNA-seq 的分析结果相一致。但 P4 期中 miR2118o 相对表达量的结果与 small RNA-seq 结果有些差别, 可能是针对 miR2118 这种家族成员成熟序列之间仅有 1-2 个碱基差异的情况下, stem-loop RT-PCR 的分辨率不够高, 不能特异地区分各条 microRNA 的相对表达量。

表 7 Small RNA 文库信息。

Table 7 The information of small RNA libraries.

Condition	Material	Stage	Repeat	Total Sequences	Genome Matched Reads	Distinct Genome Matched Reads	t/r/sn/snoRNA Matched Reads	Norm. Base Value	Max. Len. of Reads	Min. Len. of Reads
Long-day	58S	P2 stage	1	16,799,138	13,828,073	5,283,775	489,006	10M	34	18
			2	9,846,199	7,874,822	3,742,835	709,399	10M	34	18
		P3 stage	1	7,442,130	6,108,173	2,471,454	260,722	10M	34	18
			2	11,472,561	8,925,938	3,615,463	1,156,333	10M	34	18
		P4 stage	1	6,299,626	5,102,487	1,622,550	351,615	10M	34	18
			2	10,223,883	7,519,179	2,419,641	1,548,539	10M	34	18
Short-day	58S	P2 stage	1	8,055,854	6,584,870	2,925,738	320,222	10M	34	18
			2	18,566,211	14,567,930	5,584,821	1,800,317	10M	34	18
		P3 stage	1	7,744,473	6,437,977	2,579,416	214,773	10M	34	18
			2	14,180,440	11,744,920	4,390,679	726,283	10M	34	18
		P4 stage	1	9,176,074	7,336,665	2,388,519	562,714	10M	34	18
			2	15,378,568	12,179,632	3,726,780	1,466,345	10M	34	18
Long-day	NIL(MH)	P2 stage	1	12,429,931	10,295,016	4,424,623	349,892	10M	34	18
			2	10,256,625	7,447,357	3,206,714	1,251,021	10M	34	18
		P3 stage	1	29,509,417	24,349,067	6,541,673	972,279	10M	34	18
			2	7,850,158	5,939,909	2,561,898	878,977	10M	34	18
		P4 stage	1	8,386,684	6,847,925	2,010,762	270,493	10M	34	18
			2	10,022,686	7,598,623	2,598,967	1,205,930	10M	34	18
Short-day	NIL(MH)	P2 stage	1	8,597,907	6,802,183	3,014,875	508,312	10M	34	18
			2	10,704,936	8,578,192	3,931,715	694,281	10M	34	18
		P3 stage	1	6,415,895	5,204,476	2,226,559	251,562	10M	34	18
			2	11,674,329	9,135,441	3,734,960	937,276	10M	34	18
		P4 stage	1	7,818,855	6,479,341	2,098,046	177,344	10M	34	18
			2	11,982,739	9,196,113	2,861,791	1,271,148	10M	34	18

3.8.1.3 miR2118 介导的剪切对光敏感雄性不育的重要性

既然 *PMSIT* 转录本能够被 miR2118 识别, 并且介导剪切, 那么他们之间的这种作用方式对光敏感雄性不育又有什么影响? 我们构建了 35S:S-CS 超表达载体, 其外源片段为不完整的农垦 58S *PMSIT*, 缺少了 5'端, 仅从 miR2118 介导的剪切位点开始, 因此不含有完整的 miR2118 识别序列, 只包含 miR2118 识别位点的第 11 位到第 22 位 (图 16)。因为 35S:S-CS 载体是在 35S 启动子的驱动下进行的超表达, 为了与 Ubi:S 超表达农垦 58S *PMSIT* 全长的转化结果相比较, 另外将 Ubi:S 的外源片段连接到 pS1301 载体上, 得到 35S:S 载体以作对照。将 35S:S-CS 和 35S:S 分别转化 NIL(MH), 得到转基因 T₀ 代植株 27 株和 55 株。明恢 63 与农垦 58S 的 *PMSIT* 有相同的 miR2118 识别位点, 我们也将明恢 63 中与 35S:S-CS 外源片段相对应的序列同样进行扩增, 连接到 pS1301 上得到 35S:M-CS 载体, 转化农垦 58S 愈伤组织, 共得到 T₀ 代转基因植株 14 株。分析 35S:S-CS、35S:M-CS 和 35S:S T₀ 转基因植株中外源片段的表达情况 (图 59), 选取超量表达的植株种植 T₁ 代, 进行基因型与表型共分离分析。

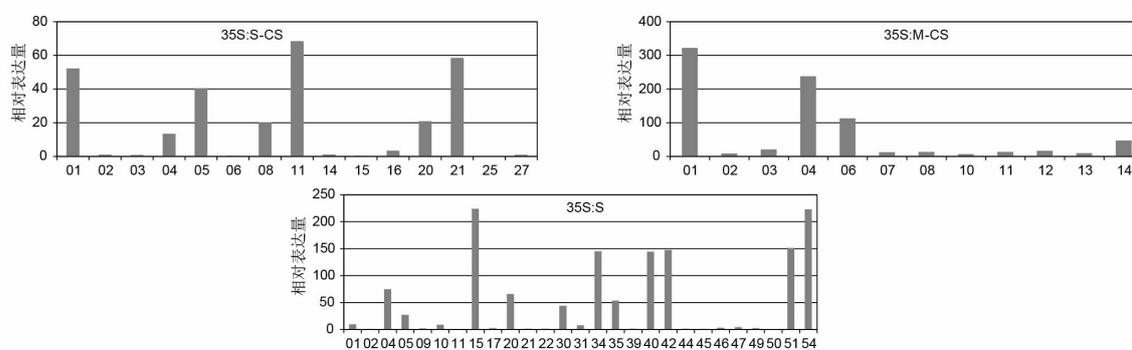


图 59 35S:S-CS, 35S:M-CS 和 35S:S 转基因 T₀ 代单株中 *PMSIT* 相对表达量检测。

Figure 59 The relative expression level of *PMSIT* in T₀ transgenic plants of 35S:S-CS, 35S:M-CS and 35S:S.

虽然在不同的超表达启动子驱动下表达相同的序列, 35S:S 与 Ubi:S 的转化结果相同 (图 60, 图 36), 长日照下转基因 T₁ 家系中阳性单株的育性低于阴性单株, 虽然育性降低的程度没有 Ubi:S 那么明显, 但转基因的效应是一致的。在此情况下, 继续分析 35S:S-CS 的转基因表型, T₁ 家系中阳性单株和阴性单株的结实率没有明显差别, 都是可育的 (图 60), 说明转化的外源片段并不能发挥 *PMSIT* 的正常功能,

不能降低 NIL(MH)的育性。同样地, 在 35S:M-CS T_1 家系中阴、阳性单株的结实率都很低 (图 60), 阳性单株并不像 Ubi:M 的转化结果那样能够恢复农垦 58S 的育性 (图 36)。与 35S:S, Ubi:S 和 Ubi:M 相比, 35S:S-CS 和 35S:M-CS 的外源片段都缺少完整的 miR2118 识别位点, 不能像前三者一样改变受体的育性, 说明该识别位点对 *PMSIT* 行使正常的功能起着至关重要的作用。

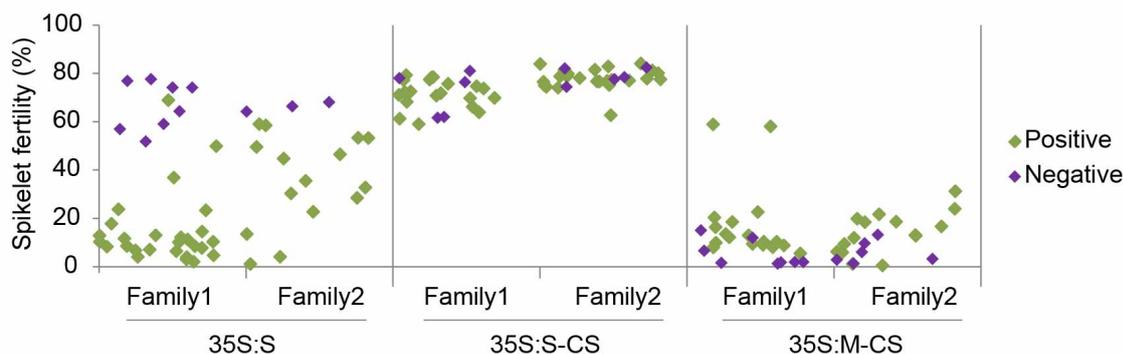


图 60 长日照下 35S:S, 35S:S-CS 和 35S:M-CS 转基因 T_1 家系共分离分析。

Figure 60 The co-segregation analysis of 35S:S, 35S:S-CS and 35S:M-CS transgenic plants of T_1 families under long-day conditions.

3.8.1.4 Target mimicry

miR2118 的识别和介导的剪切对 *PMSIT* 调控的光敏感雄性不育非常重要, 那么, miR2118 本身在光敏感雄性不育过程中又扮演着什么样的角色? 之前 psRNATarget 预测结果认为在 miR2118 家族中, miR2118d 和 miR2118o 是最有可能识别 *PMSIT* 的 (图 52); 并且对 miR2118 进行 small RNA-seq 分析发现该家族成员的表达模式可分为两类, 而 miR2118d 和 miR2118o 恰好分属于两类 (图 57), 因此我们采用 target mimicry 的方法分别降低这两个 miR2118 的表达量, 构建得到 MIM2118d 和 MIM2118o 载体, 将这两个载体分别转化农垦 58S 和 NIL(MH), 获得转基因植株 MIMdS 43 株, MIMdN 42 株, MIMoS 26 株和 MIMoN 54 株。Stem-loop RT-PCR 结果显示大部分转基因植株中 miR2118d 或者 miR2118o 表达量显著下降 (图 61)。

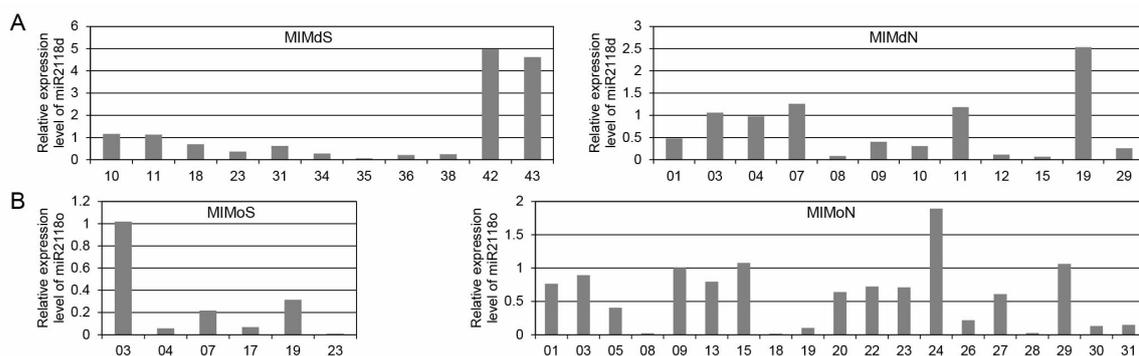


图 61 Target mimicry 转基因植株中 miR2118 的相对表达量检测。

A, MIMdS 和 MIMdN 中 miR2118d 的相对表达量; B, MIMoS 和 MIMoN 中 miR2118o 的相对表达量。

Figure 61 The relative expression level of miR2118d and miR2118o in transgenic plants of target mimicry vectors.

A, The relative expression level of miR2118d in transformants of MIMdS and MIMdN; B, The relative expression level of miR2118o in transformants of MIMoS and MIMoN.

进一步考察这些 miR2118 表达被抑制的单株下一代的基因型和表型。长日照下, 在农垦 58S 中单独抑制 miR2118d 或者 miR2118o 表达的转基因 T_1 家系中, 阳性单株的结实率相对于阴性单株并没有升高 (图 61A); 而在 NIL(MH) 中单独抑制 miR2118d 或者 miR2118o 的表达也不会降低单株的结实率 (图 61B)。说明在农垦 58S 和 NIL(MH) 中只抑制 miR2118d 或者 miR2118o 的表达并不会对育性产生影响。由于 miR2118d 和 miR2118o 都有可能识别 *PMSIT*, 或许只抑制其中一个小 RNA 的表达还不足以引起育性的改变, 因此将 MIM2118d 和 MIM2118o 转基因植株进行杂交以分析共同抑制这两个 miRNA 的表型。

分别分析了来自不同独立转基因 T_0 植株之间进行三个组合杂交得到的 F_1 杂交种的结实率, MIMdN 和 MIMoN 杂交种是完全可育的 (图 61C)。绝大部分 MIMdS 和 MIMoS 杂交种结实率极低, 但在每个组合中都有那么一两株的育性略微有所上升 (图 61C), 从每个组合中各挑选一个代表了育性恢复的不同程度的 F_1 杂种继续种植 F_2 , 以考察这种育性恢复的可靠性和真实性。长日照下, 田间观察发现来自 MIMdS 和 MIMoS 杂交组合 2 的 30 个 F_2 单株的育性没有差别, 都是完全不育; 而组合 1 和组合 3 的 F_2 群体中育性出现了分离, 进而对这些单株进行了基因型的鉴定和考种分析。无论是在组合 1 还是组合 3 的 F_2 中, 分离出的三种基因型 (MIM2118d 纯合; MIM2118d 和 MIM2118o 杂合; MIM2118o 纯合), 尤其是杂合基因型单株的

结实率都有高有低，基因型与表型不共分离（图 61D 和 E），说明 F₁ 中获得的育性恢复并不是由小 RNA 的抑制表达造成的。以上结果表明在农垦 58S 或者 NIL(MH) 中同时抑制 miR2118d 和 miR2118o 的表达都对育性没有影响。

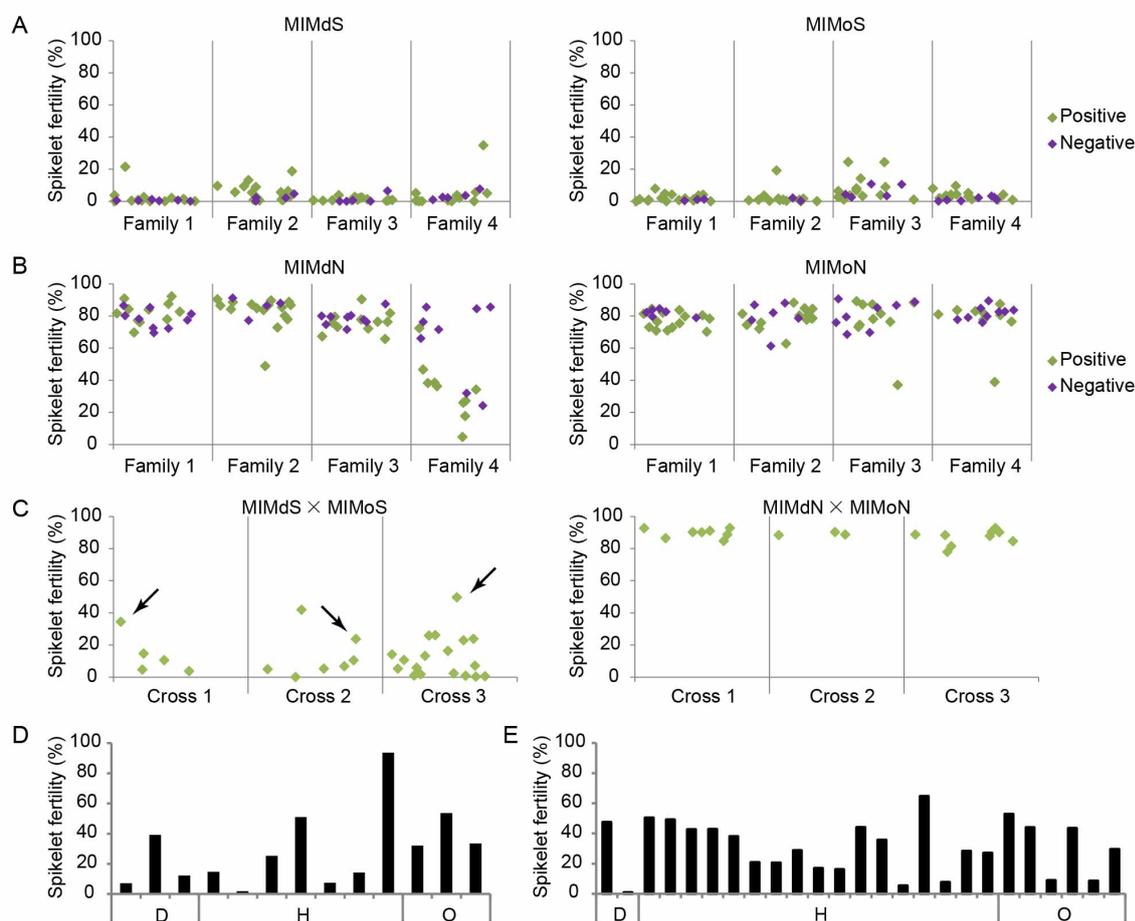


图 61 miR2118d 和 miR2118o target mimicry 转基因单株长日照下的表型。

A-B, 在农垦 58S (A) 和 NIL(MH) (B) 中抑制 miR2118d 和 miR2118o 表达的转基因 T₁ 家系结实率分析; C, MIMdS 和 MIMoS 及 MIMdN 和 MIMoN 杂交种 F₁ 的结实率分析; D-E, MIMdS 和 MIMoS 杂交种 F₂ 的结实率分析, 其中 D 图来自 Cross 1 单株; E 图来自 Cross 3 单株。“D”, MIM2118d 纯合; “H”, MIM2118d 和 MIM2118o 杂合; “O”, MIM2118o 纯合。

Figure 61 The spikelet fertility of transgenic plants from MIM2118d and MIM2118o under long-day condition.

A-B, The spikelet fertility of transgenic plants from T₁ families target mimicing miR2118d or miR2118o in Nongken 58S (A) and NIL(MH) (B); C, The spikelet fertility of hybrid F₁ from MIMdS and MIMoS or MIMdN and MIMoN; D-E, The spikelet fertility of hybrid F₂ from cross 1 (D) or cross 3 (D) of MIMdS and MIMoS. “D”, MIM2118d homozygotes; “H”, MIM2118d and MIM2118o heterozygotes; “O”, MIM2118o homozygotes.

因此, 尽管在 miR2118 家族中, miR2118d 和 miR2118o 是最有可能识别 *PMS1T* 并介导剪切的小 RNA, 但在农垦 58S 和 NIL(MH) 中无论是单独还是同时抑制这两个小 RNA 的表达, 对 *PMS1T* 造成的光敏感雄性不育都没有影响。我们仅分析了该家族中两个 miRNA 的作用, 这些小 RNA 之间极有可能存在功能冗余性, 仅抑制其中少数几个 miRNA 的表达, 并不能阻止家族中其他成员对 *PMS1T* 的识别和剪切, 因此还不能排除 miR2118 在光敏感雄性不育中的重要作用。

3.8.2 *PMS1T* 产生 phasiRNA

Johnson(2010)在对水稻未成熟的小穗中的小 RNA 进行分析时发现大量的 21-nt 小 RNA 会在染色体上成簇存在, 形成 21-mer cluster, 并且进一步聚集成 supercluster。其中高达 80.8% 的 21-mer cluster 中的小 RNA 会形成 phase, 即在基因组上以小 RNA 的长度为相位连续排列。而在 89% 的能形成 phase 小 RNA 的 21-mer cluster 前面 12 nt 的位置处都能找到一个共同的 motif, 在小 RNA 文库中共有 12 条 22-nt 的小 RNA 能靶定该 motif, 可能诱导在该 motif 第 12 位核苷酸后的剪切, 并形成一系列 phase 小 RNA。这 12 条小 RNA 来自于第四染色体的 20 kb 和第十一染色体的 3.0 kb 范围内, 该范围内的序列具有相同的特征, 均由 2 到 3 个拷贝的大约 170 nt 的反向重复序列间隔排列而成, 这也是 miRNA 前体的典型特征。因此认为这些 22-nt 的小 RNA 可能组成了一个新的 miRNA 家族, 称之为 miR2118。

在其后的研究中, 又在水稻、苜蓿、马铃薯、大豆、玉米中进一步发现和证实了 miR2118 能够诱导靶基因被剪切产生 21-nt 的 phasiRNA (Zhai et al 2011; Shivaprasad et al 2012; Song et al 2012a; Arikiti et al 2014; Zhai et al 2015)。因此, 我们也对 *PMS1T* 区段进行小 RNA 分析。

3.8.2.1 *PMS1T* 区段产生大量小 RNA

分析之前构建的长、短日照下农垦 58S 和 NIL(MH) P2、P3 和 P4 三个时期的幼穗 small RNA 文库(表 8), 发现在 *PMS1T* 区段存在大量的小 RNA, 从 18-nt 到 30-nt 不等, 但在各时期中 21-nt siRNA 占绝大多数(表 8), 并且在 P3 期中的表达丰度最高。

表 8 Small RNA 文库中 *PMSIT* 区段产生的 siRNA。Table 8 Abundance of small RNAs generated from the *PMSIT* region.

Length	Stage	58S		NIL(MH)	
		Long-day	Short-day	Long-day	Short-day
20 nt	P2	0	1	0	1
	P3	5	4	3	0
	P4	5	6	4	2
21 nt	P2	19	162	47	33
	P3	364	297	140	153
	P4	169	125	68	56
22 nt	P2	0	1	0	0
	P3	0	3	2	0
	P4	6	1	1	1
23 nt	P2	0	1	0	0
	P3	0	0	1	0
	P4	1	0	0	1
24 nt	P2	1	2	2	1
	P3	2	3	2	1
	P4	2	3	1	0

仔细分析这些 21-nt 小 RNA，发现他们都集中在 *PMSIT* 的第 82 位到 459 位碱基之间，从 miR2118 剪切位点开始到 SNP S1 附近（图 62A）。这些小 RNA 既来自于正义链，也来自于反义链；并且某些分别来自于两条链的小 RNA 之间能够互补配对，形成在两个 3'端各突出 2-nt 的 duplex（图 62A）。

进一步将这些小 RNA 排布在两条链上分别分析，发现他们恰好是从 miR2118 介导的剪切位点开始，以 21-nt 的间距进行排列，首尾相连，形成一系列的 phasiRNA。以长日照下农垦 58S P3 期的小 RNA 测序结果为例，从剪切位点开始，在正义链和反义链上各得到 18 个 phasiRNA（图 62B），并且两条链上匹配到同一位置上的 sense-phasiRNA 和 antisense-phasiRNA 之间有 2-nt 的错位，该特征表明这些 siRNA 的加工和形成是由 DCL4 和 RDR6 参与的(Rajeswaran and Pooggin 2012; Song et al 2012a)。除了这些 phasiRNA 之外，*PMSIT* 区间还存在少数的 21-nt siRNA 不是以 21-nt 的相位排列，称之为 non-phasiRNA（图 62B）。这 18 对 phasiRNA 虽然都来自于 *PMSIT* 转录本，但他们之间的表达丰度也存在明显差异。

由此可知，与已报道的 miR2118 功能相同，在水稻中 miR2118 识别 *PMSIT* 后，同样会介导 *PMSIT* 的剪切，并且从剪切位点开始形成 21-nt 的 phasiRNA。

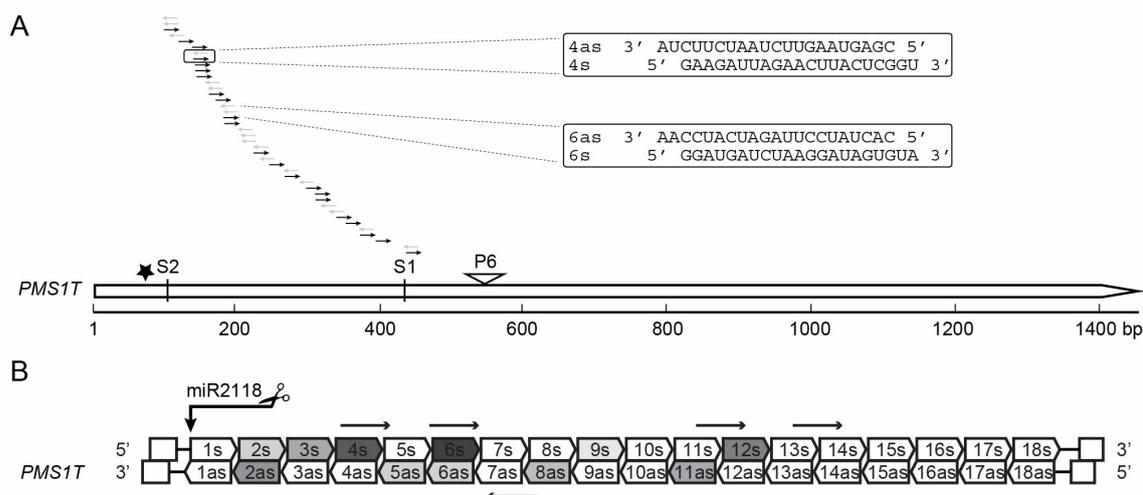


图 62 *PMS1T* 区段产生的 21-nt 小 RNA 的序列特征。

A, *PMS1T* 区段产生的 21-nt 小 RNA 的定位。数据来源于长、短日照下农垦 58S 和 NIL(MH) P2、P3 和 P4 期幼穗的小 RNA 文库测序结果。方框中示意来自于正义链和反义链的两条 21-nt 小 RNA 能互补配对形成 2-nt 3' 端突出的 duplex。星号代表 miR2118 的剪切位点。B, *PMS1T* 区段形成的 21-nt phasiRNAs 示意图, 以长日照下农垦 58S P3 期小 RNA 测序结果为例。垂直箭头示意 miR2118 介导的剪切位点, 带箭头的方框代表 phasiRNA; 带箭头的横线代表 non-phasiRNA, 方框的灰度表示 phasiRNA 的表达丰度。s, 来自于正义链的 phasiRNA; as, 来自于反义链的 phasiRNA。

Figure 62 The feature of 21-nt small RNA producing from *PMS1T*.

A, Locations of 21-nt small RNAs produced from *PMS1T*. The data are generated from small RNA libraries of young panicles at P2, P3 and P4 stages from 58S and NIL(MH) grown under long-day and short-day conditions. Examples of duplexes with 2-nt 3' overhang formed between 21-nt sense and antisense phased small RNAs are illustrated in the boxes. The star indicates the cleavage site of miR2118. B, Schematic diagram of the 21-nt phasiRNAs generated from the *PMS1T* transcript in 58S at P3 stage under long-day conditions. The vertical arrow indicates the cleavage site directed by miR2118. Boxes with arrow show small RNAs in phase while horizontal arrows indicate ones that are not in phase. Intensity of grey color displays the abundance of 21-nt phasiRNA. s, phasiRNAs from sense strand; as, phasiRNAs from antisense strand.

3.8.2.2 比较分析农垦 58S 和 NIL (MH) 中 *PMS1T*-phasiRNA 的表达

PMS1T 被 miR2118 剪切后能形成 phasiRNA, 那么这些 *PMS1T*-phasiRNA 是否与光敏感雄性不育相关? 我们进一步比较了长、短日照下农垦 58S 和 NIL(MH) 中 *PMS1T*-phasiRNA 的表达差异。比较分析表明, 从整体上来看 21-nt *PMS1T*-phasiRNA 的表达丰度在 P3 期和 P4 期要远高于 P2 期; P3 期和 P4 期的表达模式相似; sense-phasiRNA 的表达量要高于 antisense-phasiRNA, 且绝大部分的 phasiRNA 集中在 4s、6s 和 12s 这三条 phasiRNA 上 (图 63)。

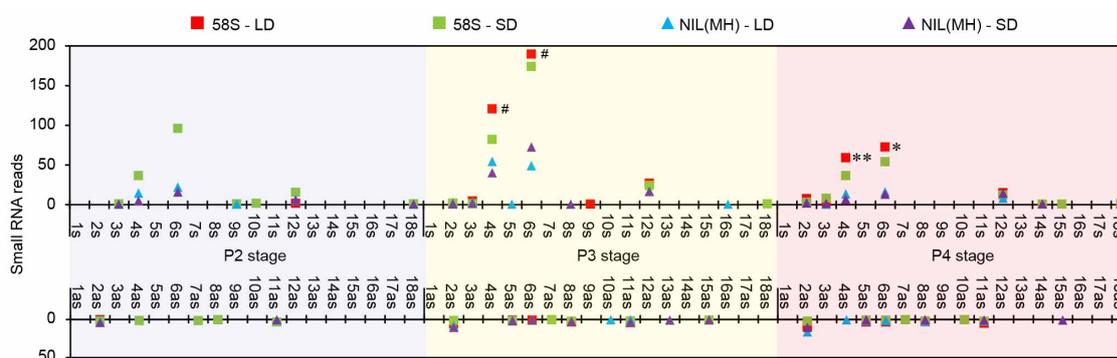


图 63 长、短日照下农垦 58S 和 NIL(MH) 中 P2, P3 和 P4 期 21-nt *PMS1T*-phasiRNA 的表达丰度比较分析。

图中小 RNA 的 reads 数为 TP10M 平衡化后的结果, 为两次生物学重复的平均值。对 58S-LD 和 NIL(MH)-LD 进行了 *t*-test 显著性检验, 显著程度表示为: ** $P < 0.01$, * $P < 0.05$, 及 # $P < 0.1$ 。

Figure 63 Abundance of 21-nt *PMS1T*-phasiRNAs from young panicles of 58S and NIL(MH) at P2, P3 and P4 stages under long-day and short-day conditions.

The small RNA reads are normalized to the whole library with transcripts per 10 million (TP10M) and presented as the mean from two biological replicates. Differences were detected between 58S-LD and NIL(MH)-LD by *t*-test at ** $P < 0.01$, * $P < 0.05$, or # $P < 0.1$.

具体分析长日照下农垦 58S 和 NIL(MH) 中 phasiRNA 的表达量, 发现 P3 期和 P4 期的幼穗中 phasiRNA 的表达丰度在长日照农垦 58S 中高于 NIL(MH) 中, 统计学上达到了显著水平 (图 63)。虽然在 P3 期两者之间统计测验的 *P* 值大于 0.05, 但也小于 0.1, 并且从数值上看差异明显, 因为仅两次生物学重复, 如果增加生物学重复数, 应该能得到更理想的结果。同时, 长日照下农垦 58S 中 phasiRNA 的表达量也高于短日照下农垦 58S 和 NIL(MH) (图 63)。该结果表明在不育材料中

PMSIT-*phasRNA* 的丰度比可育材料中高, 暗示 *PMSIT* 调控的光敏感雄性不育可能与 *phasRNA* 相关。

P2 期和 P3 期都是农垦 58S 育性转换的关键时期, 但 P2 期 21-nt *PMSIT*-*phasRNA* 的表达丰度并不高, 且不表现出与 P3 和 P4 期相同的表达趋势。之前对农垦 58S 进行组织切片和 TUNEL 分析的结果发现花粉的败育起始于 P3 期 (花粉母细胞形成期), 而相对提前的细胞程序性死亡是从 P4 期开始(Shi et al 2009; Ding et al 2012a)。说明在 *phasRNA* 参与的 *PMSIT* 调控光敏感雄性不育过程中, P2 期可能并没有 P3 期和 P4 期关键。

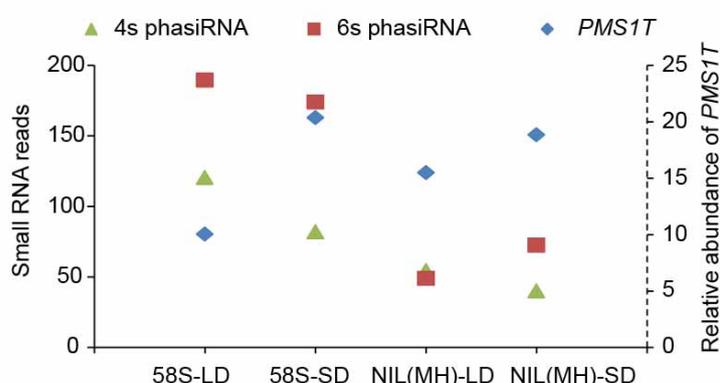


图 64 P3 期 *PMSIT*-*phasRNAs* 表达量和 *PMSIT* 相对表达量之间的相关分析。

小 RNA 表达丰度和 *PMSIT* 相对表达量的数据分别来自于图*和图*。

Figure 64 Comparison of the relative abundance of *PMSIT*-*phasRNAs* and *PMSIT* in panicles at P3 stage under long-day and short-day conditions.

The small RNA reads and the relative abundance of *PMSIT* are collected from the data in Figure * and Figure *.

之前对 *PMSIT* 进行表达量分析时表明在 P3 期时 *PMSIT* 在长日照下农垦 58S 中的表达量低于 NIL(MH), 及短日照下的农垦 58S 和 NIL(MH); 而现在分析得到 *PMSIT*-*phasRNA* 的表达量却在长日照农垦 58S 中最高。将 *PMSIT* 转录本相对表达量与 *PMSIT*-*phasRNA* 的表达量进行比较, 发现两者之间呈现负相关: 在长、短日照下 P3 期的农垦 58S 和 NIL(MH) 中, 当 4s *phasRNA* 和 6s *phasRNA* 的表达丰度高时, *PMSIT* 转录本的相对表达量就较低(图 64)。表明在细胞中有越多的 *phasRNA* 形成时, 作为 *PMSIT*-*phasRNA* 前体的 PHAS 基因——*PMSIT* 完整转录本保留得就越少, 因为更多的转录本需要被 miR2118 介导剪切以形成 *phasRNA*。

3.8.2.3 转基因材料中 *PMS1T*-*phas*iRNA 分析

为了进一步确证 *phas*iRNA 与 *PMS1T* 调控的光敏感雄性不育相关, 构建了转基因植株长日照下 P3 期幼穗的小 RNA 文库进行 *PMS1T*-*phas*iRNA 分析, 主要包括: 农垦 58S *PMS1T* 超量表达的 Ubi:S 转基因 T₁ 代阳性单株和阴性单株(简称为“OX”), 及农垦 58S *PMS1T* knock-down 表达的 58S:dsi 转基因 T₁ 代阳性单株和阴性单株(简称为“Dsi”)。每份转基因材料有两次生物学重复, 数据分析时采用 TP20M 进行平衡化, 表 9 展示了这些小 RNA 文库的相关信息。

表 9 转基因材料小 RNA 文库信息。

Table 9 The information of small RNA libraries constructed from transgenic plants.

Material	Repeat	Total Sequences	Genome Matched Reads ^{c)}	Distinct Genome Matched Reads ^{c)}	t/r/sn/snoRNA Matched Reads	Norm. Base Value	Max. Len. of Reads	Min. Len. of Reads
OX ^{a)}	1	21,112,310	15,937,293	5,900,039	1,755,444	20M	51	16
negative	2	17,722,618	13,697,068	5,158,162	1,306,628	20M	51	16
OX ^{a)}	1	24,179,355	18,200,240	5,993,498	2,267,349	20M	51	16
positive	2	23,512,057	17,982,634	6,113,594	1,993,204	20M	51	16
Dsi ^{b)}	1	23,101,075	17,081,599	6,189,342	1,900,938	20M	51	16
negative	2	18,979,485	14,239,463	5,243,912	1,411,383	20M	51	16
Dsi ^{b)}	1	17,293,190	12,982,001	4,970,789	1,115,132	20M	51	16
positive	2	19,603,763	14,327,817	5,627,864	1,545,708	20M	51	16

OX^{a)}, Ubi:S transgenic T₁ plants; Dsi^{b)}, 58S:dsi transgenic T₁ plants; ^{c)}, Does not include the data listed in the column “t/r/sn/snoRNA Matched Reads”.

选择文库中能匹配到 *PMS1T* 基因区段的长度为 20-nt 到 24-nt 的小 RNA 进行分析, 与农垦 58S 和 NIL(MH)中小 RNA 测序得到的结果相似, 21-nt 长度的小 RNA 具有优势表达(表 8 和表 10)。但在转基因阳性植株中, 无论是 OX 还是 Dsi, 相对于两个亲本的测序结果来说, 22-nt 和 24-nt 小 RNA 的表达量都显著上升; 同时, 在 21-nt 小 RNA 中 non-phased 小 RNA 所占的比例也非常高, 几乎是绝大部分。这两种情况在转基因阴性单株中却不存在, 说明这些现象并不是由于组织培养过程造成的, 可能与转基因 T-DNA 插入相关; 或者是在转化的过程中由于载体的超量表达诱导了其他长度的小 RNA 的产生, 因为超量表达载体 Ubi:S 和抑制表达载体 58S:dsi 的外源片段分别在 Ubiquitin 和 35S 启动子驱动下进行超量表达。

表 10 长日照下转基因植株 P3 时期幼穗 small RNA 文库中 *PMSIT* 区段产生的小 RNA。

Table 10 Abundance of small RNAs generated from the *PMSIT* region in transgenic plants at P3 stage under long-day conditions.

Length	OX positive	OX negative	Dsi positive	Dsi negative
20 nt	16	3	280	1
21 nt	9412(678 [*])	222(218 [*])	6195(189 [*])	287(281 [*])
22 nt	300	2	7658	8
23 nt	30	0	538	3
24 nt	81	4	2682	13

^{*}, the numbers of 21-nt small RNAs in phase.

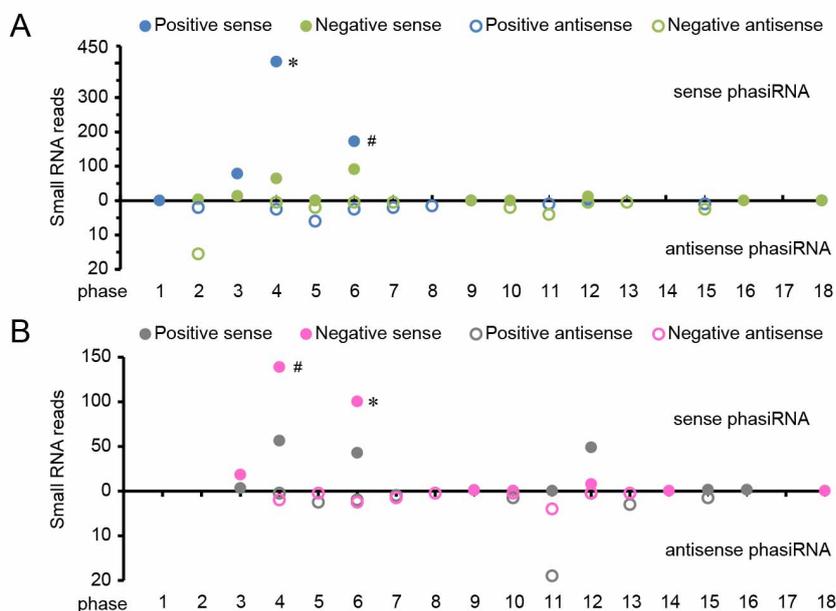


图 65 长日照下转基因 T_1 代植株 P3 期 21-nt *PMSIT*-phasiRNA 的表达丰度。

A, 来自于 Ubi:S 超表达转基因植株; B, 来自于 58S:dsi 抑制表达转基因植株。t-test 检验为转基因阳性单株与阴性单株之间的比较, 其中* $P < 0.05$ 及# $P < 0.1$ 。数据处理方法与图*相同, 但采用 TP20M 进行平衡化。

Figure 65 Abundance of 21-nt phasiRNAs from young panicles of transgenic T_1 plants under long-day conditions at P3 stage.

The data were collected from transgenic T_1 plants of Ubi:S (A) and 58S:dsi (B). Differences between transgenic positive and negative plants were detected by t-test at * $P < 0.05$ or # $P < 0.1$, respectively. The small RNA data collecting and handing were as in Figure *, but normalizing with TP20M.

分析 OX 和 Dsi 中 21-nt *PMSIT*-phasiRNA 的表达量, 发现其中 sense-phasiRNA 的表达量也比 antisense-phasiRNA 的表达量高, 且 4s phasiRNA 和 6s phasiRNA 的表

达丰度最高（图 65）。在 OX 转基因植株中阳性单株的 *PMSIT*-phasiRNA 表达量显著高于阴性单株（图 65A）；而在 Dsi 转基因植株中正好相反，阳性单株的 *PMSIT*-phasiRNA 表达量低于阴性单株（图 65B）。该结果说明 21-nt *PMSIT*-phasiRNA 的表达量与农垦 58S *PMSIT* 的表达量负相关，并且与光敏感雄性不育相关：在长日照下，超量表达农垦 58S *PMSIT*，转基因阳性单株中 21-nt *PMSIT*-phasiRNA 大量升高，单株表现出雄性不育；抑制农垦 58S 中 *PMSIT* 表达，转基因阳性单株中 *PMSIT*-phasiRNA 相应降低，单株育性得到恢复。

转基因结果与长、短日照下亲本小 RNA 测序结果相一致，都表明：miR2118 识别并介导 *PMSIT* 的剪切产生的 21-nt *PMSIT*-phasiRNA，在长日照下的农垦 58S 中累积，造成雄性不育。

3.8.2.4 *PMSIT*-phasiRNA 的验证

小 RNA 测序结果显示 *PMSIT* 能产生 21-nt phasiRNA，那么这些 phasiRNA 是否在体内真实存在？其表达模式如何？我们选择对其中表达丰度最高的 6s phasiRNA 进行 northern blot 分析。在实验过程中，我们摒弃了传统的采用放射性同位素的方法，改用非放射性的生物素（biotin）标记的探针定性分析体内小 RNA 的表达量。

在预实验中，我们合成了与 6s phasiRNA 序列相同的 21-nt 长度的 RNA（5'-GGAUGAUCUAAGGAUAGUGUA-3'）作为 marker 与来自于农垦 58S 和转基因植株的样品总 RNA 一起与探针杂交，合成的 DNA 探针序列其 5'端带有生物素标记（序列为 5'-TACACTATCCTTAGATCATCC-3'）。结果显示农垦 58S 和转基因植株样品杂交得到的条带与 6s phasiRNA 大小相同，且条带唯一。在此基础上，我们检测长、短日照下农垦 58S 和 NIL(MH)中 P3 期幼穗中 6s phasiRNA 的表达量，以 U6 作为内参进行半定量分析（5'端 biotin 修饰的 U6 探针序列为 5'-TGTATCGTTCCAATTTTATCGGATGT-3'），其结果与小 RNA 测序结果一致，6s phasiRNA 的表达丰度在长日照下农垦 58S 中最高（图 66A）；同样地，在 Ubi:S 转基因阳性植株中 6s phasiRNA 的表达量高于阴性单株，而在 58S-dsi 转基因植株中则相反（图 66B）。

Page-northern 的结果验证了 6s phasiRNA 在体内确实真实存在，也说明了小 RNA 测序的结果是可靠的：在体内 *PMS1T* 能产生 21-nt phasiRNA。但该结果反映的样品之间的表达差异并不是非常明显，可能与体内 6s phasiRNA 的表达量太低相关。事实上，在曝光显影的过程中，6s phasiRNA 探针杂交后的膜需要 10 min 以上才能看到模糊的条带，而经过探针剥离后用 U6 探针杂交的膜在同样的条件下仅需 1 sec 就能显出异常清晰的条带。

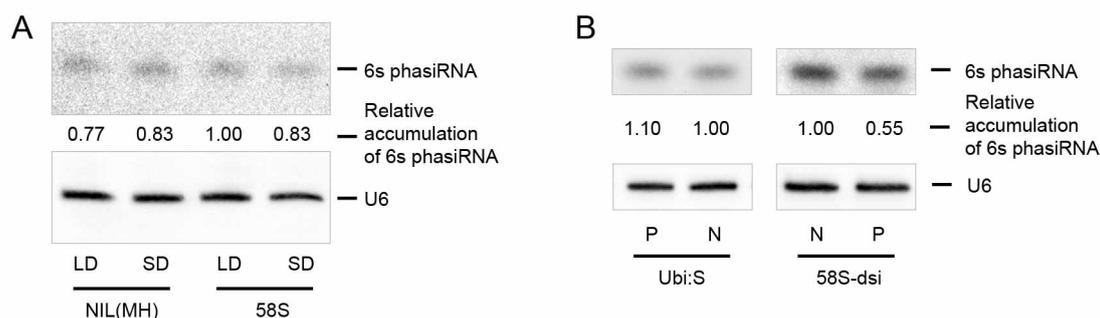


图 66 6s phasiRNA page-northern blot 分析。

A, 长、短日照下农垦 58S 和 NIL(MH); B, 长日照下转基因 T_1 植株分析。样品均来自于 P3 期幼穗，上样量为 10 μ g 总 RNA。膜上探针剥离后再用 U6 探针进行杂交。LD, 长日照; SD, 短日照。P, 转基因阳性单株; N, 转基因阴性单株。

Figure 66 The page-northern blot analysis of 6s phasiRNA.

A, Nongken 58S and NIL(MH) under long-day and short-day conditions; B, Transgenic plants from T_1 families under long-day conditions. 10 μ g small RNAs per sample from young panicles at P3 stage were loaded on the gel. The blots were stripped off and rehybridized with U6 probe. LD: long days, SD: short days. P, positive plants; N, negative plants.

3.9 SNP S2 与 *PMS1T*-phasiRNA

之前研究表明 SNP S2 是造成 *PMS1T* 功能性差异的序列多态性位点，而 SNP S2 恰好位于 2s phasiRNA 的第 4 个核苷酸和 2as phasiRNA 的第 16 个核苷酸上，并且在 miR2118 介导的剪切位点下游，相距仅 24 nt (图 54 和图 67)。文献报道 miRNA 识别位点附近的 SNP 能够影响 miRNA 对靶基因的表达调控(Mishra et al 2007; Chen et al 2011; Li et al 2014a)。因此，结合 SNP S2 与 phasiRNA 之间的位置关系，推测 SNP S2 可能通过两种途径参与调控 *PMS1T*-phasiRNA 影响的光敏感雄性不育：首先，因为 SNP S2 距离 miR2118 介导的剪切位点非常近，可能从空间位置上影响

miR2118 对 *PMSIT* 的剪切；其次，SNP S2 的多态性位点可能影响其所在的 2s phasiRNA 和 2as phasiRNA 对下游靶基因的调控。



图 67 SNP S2 和 SNP S1 在 *PMSIT*-phasiRNA 上的位置。

灰色字母突出显示了 SNP S2 和 SNP S1 的序列。

Figure 67 The location of SNP S2 and SNP S1 at 21-nt phasiRNAs.

Grey characters indicate the SNPs S2 and S1.

3.9.1 SNP S2 对 miR2118 介导的剪切效率的影响

我们利用软件分析了农垦 58S 和明恢 63 *PMSIT* 的 RNA 二级结构 (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>), 结果显示两者之间存在极大的差异 (图 68), 而这种差异主要是由 65 bp 的插入/缺失造成的。之前分析表明, 将农垦 58S *PMSIT* 中的 65 bp 插入去掉并不影响 *PMSIT* 对育性的影响, 分析该载体 58S-C-dP6 所含有的 *PMSIT* 序列的 RNA 二级结构, 与明恢 63 *PMSIT* 相比, 在序列上两者仅存在 SNP S2 和 SNP S1 这两个碱基的差异, 但在 RNA 二级结构上两者却有很大的不同。而这种不同主要体现在 SNP S2 所在的附近区域, 该区域位于整个结构的上部, miR2118 的识别位点也因此发生了极大的改变。相对来说, SNP S1 处于的整个结构的后半部分在两者间的差异微乎其微 (图 68)。由此推测由 SNP S2 造成的这种结构上的改变, 从空间结构位置上将影响 miR2118 与靶序列的结合, 及 miR2118 介导的剪切过程中 DCL4 蛋白的结合与接下来的剪切形成 phasiRNA 的加工过程, 使得长日照的农垦 58S 中能形成更多的 21-nt phasiRNA。

那么如何利用实验证明该推测的成立? 常规考虑是利用 Northern blot 比较体内被剪切后的 *PMSIT* 转录本的含量, 以确定 *PMSIT* 被剪切的效率。但是 miR2118 介导的剪切位点位于 *PMSIT* 转录本的第 81 个碱基上, 剪切后的 *PMSIT* 与完整的转录

本之间仅 81 bp 的差别，Northern blot 的琼脂糖凝胶电泳无法将两者区分开来。因此尝试采用以聚丙烯酰胺变性凝胶电泳为基础的 RPA 进行检测。

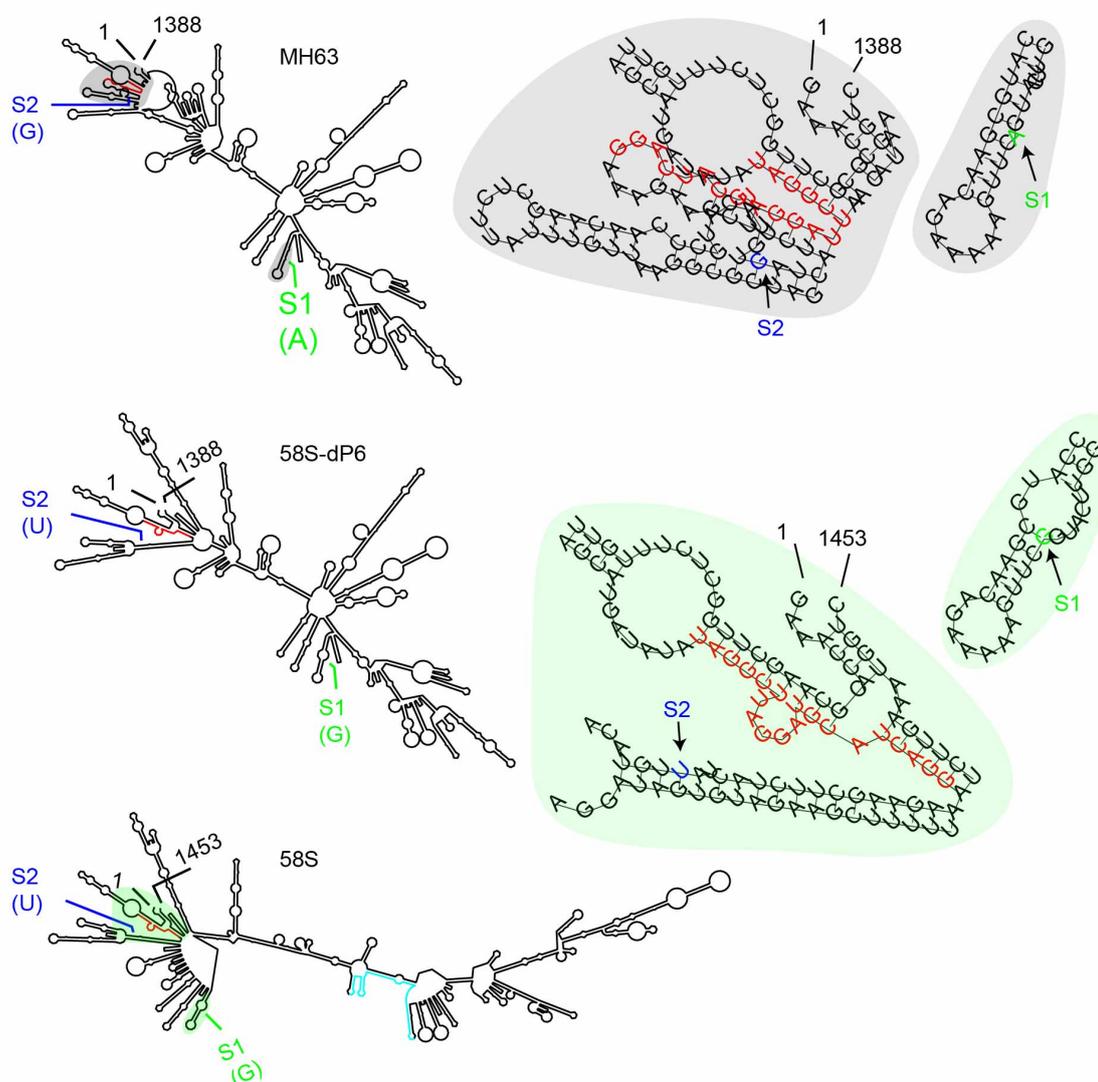


图 68 *PMSIT* RNA 二级结构预测结果。

MiR2118 的靶位点序列标为红色，65-bp 区域标为蓝绿色。SNP S2 和 SNP S1 分别以蓝色和绿色标记。明恢 63 和农垦 58S 中两个 SNP 附近区域的二级结构及其放大图分别以灰色和绿色阴影突出显示。

Figure 68 Prediction of RNA secondary structure of *PMSIT*.

The sequence of the miR2118 target is indicated in red, and the 65-bp region in cyan color. The positions and bases of SNP S2 and S1 are marked in blue and green colors respectively. Magnification of the structures of the regions nearby the SNPs from MH63 and 58S are presented under the grey and green background, respectively.

我们对 *PMSIT* 转录起始位点开始的 200 bp 序列进行扩增，引物序列见附录 I，并利用试剂盒进行生物素标记互补序列，然后与农垦 58S 和 NIL(MH)长、短日照下

P3 期幼穗的总 RNA 进行杂交。经过核糖核酸酶处理后，探针与目标 RNA 序列互补的区段将会被保护，不能被核糖核酸酶降解，经过聚丙烯酰胺胶电泳及转膜显影后，预期能够看到三条带：200 bp 未被剪切条带，miR2118 剪切后的 81 bp 和 119 bp 两条带。通过比较分析各样品之间 81 bp 或者 119 bp 条带在三条带中所占的比例就能推断 miR2118 的剪切效率高。但遗憾的是，经过多次实验，除了对照外，在样品中未看到任何条带，可能与 *PMS1T* 的表达丰度过低相关，已经超出了该实验能检测到的范围：一方面，*PMS1T* 被剪切后在体内是一个暂时的状态，会被 DCL4 蛋白继续剪切以便 phasiRNA 的形成；另一方面，未被剪切、得以保留的 *PMS1T* 完整转录本的量相对也不会太高。因此，到目前为止，虽然有线索指出 SNP S2 可能影响 miR2118 介导的对 *PMS1T* 的剪切和 phasiRNA 的形成，但还没有更好的实验手段进行证实。

3.9.2 *PMS1T*-phasiRNA 可能的靶基因

研究表明 tasiRNA 能够作为小 RNA 去结合靶基因，并介导靶基因进一步形成二级 tasiRNA，触发一系列小 RNA 形成的级联反应(Chen et al 2007a)。虽然到目前为止仅生物学信息预测，还没有确切的实验证据表明 phasiRNA 也能结合下游靶基因，并调控靶基因的表达，但不能排除这种可能性。因此 *PMS1T*-phasiRNA 可能通过对下游靶基因的调控在光敏感雄性不育中发挥作用，其可能的方式有两种：首先，位于 2s phasiRNA 和 2as phasiRNA 上的 SNP S2 使得农垦 58S 和明恢 63 的这两条 phasiRNA 序列有差别，这种差别将导致它们会识别不同的靶基因，这些靶基因可能参与雄性不育；其次，虽然其他 *PMS1T*-phasiRNA 在农垦 58S 和明恢 63 之间没有序列差异，但 SNP S2 导致了农垦 58S 和明恢 63 的 phasiRNA 表达量不同，使得它们对下游与雄性不育相关的靶基因的调控能力也不同。我们将分析这两种推论的可能性。

在 psRNATarget 数据库中预测分析农垦 58S 和明恢 63 的 2s phasiRNA 和 2as phasiRNA 的下游靶基因，设定 Maximum expectation value 为 3.0，其结果如表 11 所示。SNP S2 导致它们拥有不同的靶基因，同时相同的靶基因与 phasiRNA 作用的 expectation 值有差别（表 11）。这些基因绝大部分都没有被克隆或者进行过功能研究，其中已报道的基因 *LOC_Os05g44310* 和 *LOC_Os04g56580* 的已知功能也与育性

或者光周期无关。尽管如此，也不能排除这些预测的靶基因参与光敏感雄性不育过程，或者其被 phasiRNA 结合后进一步产生 siRNA，调控更下游基因表达的可能。

表 11 2s phasiRNA 和 2as phasiRNA 预测的靶基因。

Table 11 The putative targets of 2s phasiRNA and 2as phasiRNA.

phasiRNA	Putative target	E ^{a)}	Symbol	Annotation
2s-s ^{b)}	LOC_Os01g70020	2.0		DEK C terminal domain containing protein, expressed
2s-m ^{c)}	LOC_Os05g22600	2.5		expressed protein
	LOC_Os01g02010	3.0		expressed protein
	LOC_Os06g22060	3.0		pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha, putative, expressed
	LOC_Os03g63520	3.0		Der1-like family domain containing protein, expressed
	LOC_Os11g43760	3.0		lipase class 3 family protein, putative, expressed
	LOC_Os01g60590	3.0		transposon protein, putative, Pong sub-class, expressed
	2as-s ^{b)}	LOC_Os04g51040	2.0	
LOC_Os03g15430		2.0		Ser/Thr protein phosphatase family protein, putative, expressed
LOC_Os05g44310		2.5	<i>OsSec18</i>	vesicle-fusing ATPase, putative, expressed
LOC_Os09g30462		2.5		expressed protein
LOC_Os01g06920		3.0		resistance protein SIVe1 precursor, putative, expressed
2as-m ^{c)}	LOC_Os01g06920	2.0		resistance protein SIVe1 precursor, putative, expressed
	LOC_Os03g04260	2.5		glutathione S-transferase, putative, expressed
	LOC_Os05g34600	3.0		no apical meristem protein, putative, expressed
	LOC_Os04g56580	3.0	<i>IPK1</i>	inositol 1,3,4,5,6-pentakisphosphate 2-kinase, putative, expressed
	LOC_Os04g51040	3.0		OsWAK50 - OsWAK receptor-like protein kinase, expressed
	LOC_Os11g47180	3.0		receptor-like protein kinase 2 precursor, putative, expressed
	LOC_Os07g23710	3.0		cytochrome P450, putative, expressed
	LOC_Os07g44110	3.0		cytochrome P450 72A1, putative, expressed
LOC_Os07g23570	3.0		cytochrome P450 72A1, putative, expressed	

^{a)} Expectation value; ^{b)} from Nongken 58S; ^{c)} from Minghui 63.

为了验证这些预测的靶基因在体内是否真正会被 phasiRNA 结合并介导剪切，我们分析了之前得到的长、短日照下农垦 58S 和 NIL(MH) P3 期幼穗的 PARE 文库，发现在 2s-s phasiRNA 的靶基因 *LOC_Os01g70020* 预测的 siRNA 识别序列上能检测到发生断裂的信号值（图 69）。但我们也注意到在 *LOC_Os01g70020* 的整个转录本上多处发生了序列断裂，并且信号值还高于 2s-s phasiRNA 预测的剪切位点（图 69A）；而且在 NIL(MH)中也不是完全检测不到剪切位点的 PARE 信号（图 69B）。说明该剪切位点上检测到的信号可能是在 PARE 文库构建过程中产生的机械断裂或者降解，与小 RNA 介导的剪切无关。进而尝试用 5' RLM-RACE 验证该剪切位点在体内是否存在，由于该剪切位点恰好位于基因的 3'-UTR 上，限制了 RACE PCR 引

物的设计，经过多次实验，也未能检测到符合预期的条带。因此还无法证实 *LOC_Os01g70020* 是否是 2s-s phasiRNA 的靶基因。

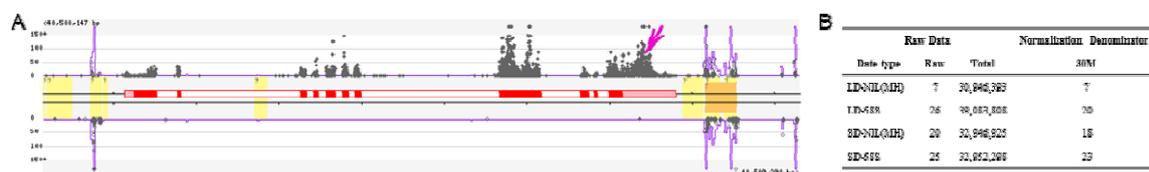


图 69 *LOC_Os01g70020* 的 PARE 结果。

A, 整个转录本的 PARE 结果，粉红色箭头示意预测的 2s-s phasiRNA 剪切位点，灰色点代表序列在 4 个 PARE 文库中检测到的 read 值总和。B, 2s-s phasiRNA 预测剪切位点的 PARE 数据。

Figure 69 The PARE data of *LOC_Os01g70020*.

A, The PARE data of *LOC_Os01g70020* transcript. The pink arrow indicated the putative 2s-s phasiRNA cleavage site, and the grey dots were the sum reads of signature from all four PARE libraries. B, The PARE data of putative cleavage site mediated by 2s-s phasiRNA on *LOC_Os01g70020*.

此外我们还对剩下的 2s phasiRNA 和 2as phasiRNA 靶基因在长日照下农垦 58S 和 NIL(MH) P3 期幼穗的 RNA 中进行 5' RLM-RACE 验证，除开某些靶基因不能设计合适的引物，这些靶基因 RACE PCR 都没有得到预期的条带。引物信息参见附录 I 表 2。因此，到目前为止，2s phasiRNA 和 2as phasiRNA 的靶基因都未能得到证实。

另一方面，分析农垦 58S 和明恢 63 之间没有序列差异的 *PMS1T*-phasiRNA 的靶基因，相关信息如表 12 所示，其中 3as, 14as 和 15as phasiRNA 在 Expectation 值为 3.0 的阈值下没有预测到靶基因。12s phasiRNA 的预测靶基因全为反转录转座子蛋白。某些预测靶基因在水稻中已经进行了功能研究，其中三个基因与育性相关：8as 的预测靶基因 *LOC_Os05g14170* (*restoration of fertility complex3, RFC3*) 编码一个线粒体定位的包含 DUF1620 的类 WD40 重复蛋白，也是水稻红莲型细胞质雄性不育育性恢复复合体的组分之一，抑制该基因的表达导致花粉育性降低 50%(Qin et al 2016); 10as 的预测靶基因 *LOC_Os05g03430* (*SUMO E3 Ligase, OsSIZ1*) 编码 SUMO E3 连接酶，调控水稻花药的开裂，突变体小穗育性低(Thangasamy et al 2011); 11as 的预测靶基因 *LOC_Os01g11040* (*TUTOUI/ early senescence1, TUT1/ES1*) 编码 cAMP 类受体蛋白的一个抑制因子，在穗发育中发挥重要作用，突变后导致抽穗延迟、小穗退化、花药和花粉发育异常，育性下降(Bai et al 2015)。分析这三个基因预测剪切位点的 PARE 信息，遗憾的是都没有检测到 PARE 信号。

表 12 *PMSIT*-phasiRNA 靶基因预测。

Table 12 The putative targets of *PMSIT*-phasiRNA.

phasiRNA	Putative target	Symbol	Annotation
1s	LOC_Os10g09870		HVA22, putative, expressed
	LOC_Os12g19549		telomerase reverse transcriptase, putative, expressed
	LOC_Os11g43770	<i>OsTERT</i>	Leucine Rich Repeat family protein, expressed
	LOC_Os04g49510	<i>OsCDPK7</i>	CAMK_CAMK_like.27-CAMK includes calcium/calmodulin dependent protein kinases, expressed
	LOC_Os02g54110		expressed protein
3s	LOC_Os11g30910	<i>STVII</i>	sulfotransferase domain containing protein, expressed
4s	LOC_Os09g38790		ZOS9-19-C2H2 zinc finger protein, expressed
5s	LOC_Os01g52851		expressed protein
	LOC_Os05g17830		retrotransposon protein, putative, unclassified, expressed
	LOC_Os01g62070	<i>OsMTP2</i>	cation efflux family protein, putative, expressed
	LOC_Os02g48740		fimbrin-like protein 2, putative, expressed
	LOC_Os02g36770		galactosyltransferase family protein, putative, expressed
6s	LOC_Os03g10540		OsFBX78-F-box domain containing protein, expressed
	LOC_Os06g08560		multidrug resistance-associated protein 11, putative, expressed
	LOC_Os10g40830		metalloendoproteinase 1 precursor, putative, expressed
	LOC_Os02g40400		expressed protein
	LOC_Os03g51900		DEAD-box ATP-dependent RNA helicase, putative, expressed
7s	LOC_Os09g20900		hypoxia-responsive family protein, putative, expressed
	LOC_Os09g30130		CSLE6-cellulose synthase-like family E, expressed
	LOC_Os07g46300		expressed protein
	LOC_Os03g19104		retrotransposon protein, putative, unclassified, expressed
	LOC_Os06g35470		L1P family of ribosomal proteins domain containing protein, expressed
	LOC_Os05g19250		retrotransposon protein, putative, Ty1-copia subclass, expressed
8s	LOC_Os08g25260		transposon protein, putative, unclassified, expressed
9s	LOC_Os12g01760		F-box/LRR domain containing protein, putative, expressed
	LOC_Os05g41820		expressed protein
	LOC_Os02g58160		expressed protein
	LOC_Os12g22590		retrotransposon protein, putative, Ty3-gypsy subclass, expressed
	LOC_Os02g28980	<i>rFKBP75</i>	peptidyl-prolyl isomerase, putative, expressed
10s	LOC_Os02g02470		oxidoreductase, short chain dehydrogenase/reductase family domain containing protein, expressed
	LOC_Os07g12730		DUF567 domain containing protein, putative, expressed
	LOC_Os01g69020		retrotransposon protein, putative, unclassified, expressed
	LOC_Os10g41400		peptide methionine sulfoxide reductase, putative, expressed
11s	LOC_Os09g15220		retrotransposon protein, putative, unclassified, expressed
	LOC_Os12g29360		retrotransposon protein, putative, unclassified, expressed
	LOC_Os12g28900		retrotransposon protein, putative, unclassified

11s	LOC_Os07g17480		retrotransposon protein, putative, unclassified
	LOC_Os10g22380		glutamyl-tRNA synthetase, cytoplasmic, putative, expressed
	LOC_Os03g19104		retrotransposon protein, putative, unclassified, expressed
	LOC_Os01g72200		expressed protein
	LOC_Os01g32250		retrotransposon protein, putative, unclassified, expressed
12s	LOC_Os11g12850		retrotransposon protein, putative, Ty3-gypsy subclass
	LOC_Os11g12874		retrotransposon protein, putative, Ty3-gypsy subclass, expressed
	LOC_Os11g37430		retrotransposon protein, putative, Ty3-gypsy subclass, expressed
	LOC_Os04g06160		retrotransposon protein, putative, Ty3-gypsy subclass, expressed
	LOC_Os09g30494		retrotransposon protein, putative, Ty3-gypsy subclass, expressed
	LOC_Os10g12630		retrotransposon protein, putative, Ty3-gypsy subclass, expressed
	LOC_Os01g28960		retrotransposon protein, putative, Ty3-gypsy subclass, expressed
	LOC_Os11g27410		retrotransposon protein, putative, Ty3-gypsy subclass, expressed
	LOC_Os05g23110		retrotransposon protein, putative, Ty3-gypsy subclass, expressed
	LOC_Os09g16340		retrotransposon protein, putative, Ty3-gypsy subclass
	LOC_Os10g24620		retrotransposon protein, putative, Ty3-gypsy subclass, expressed
	LOC_Os10g24460		retrotransposon protein, putative, Ty3-gypsy subclass, expressed
13s	LOC_Os01g25340		expressed protein
	LOC_Os07g41570		expressed protein
	LOC_Os08g30340		PAS2, putative, expressed
	LOC_Os08g33150		MYB family transcription factor, putative, expressed
	LOC_Os08g39070	<i>CSN6</i>	COP9 signalosome complex subunit 6a, putative, expressed
14s	LOC_Os02g46350		GPI transamidase subunit PIG-U domain containing protein, expressed
	LOC_Os03g08270		ataxin-2 C-terminal region family protein, expressed
15s	LOC_Os03g53650		cysteine synthase, putative, expressed
16s	LOC_Os03g11440		protein transport protein Sec61 subunit alpha, putative, expressed
	LOC_Os11g37740		stripe rust resistance protein Yr10, putative, expressed
17s-s	LOC_Os01g19630		expressed protein
17s-m	LOC_Os09g15380		retrotransposon protein, putative, unclassified, expressed
	LOC_Os01g19630		expressed protein
	LOC_Os01g01307		translocon-associated protein beta domain containing protein, expressed
	LOC_Os01g19630		expressed protein
	LOC_Os01g01307		translocon-associated protein beta domain containing protein, expressed
	LOC_Os03g12230		calcosin related protein, putative, expressed
18s	LOC_Os12g44150		plasma membrane ATPase, putative, expressed
	LOC_Os01g56780		plus-3 domain containing protein, expressed
	LOC_Os02g37090		hydrolase, alpha/beta fold family domain containing protein, expressed
	LOC_Os01g01200		retrotransposon protein, putative, unclassified, expressed
	LOC_Os01g55300		DNA-directed polymerase, putative, expressed

水稻光敏感雄性核不育基因 *pms1* 的克隆与功能分析

1as	LOC_Os06g48770	<i>OsABC1-8</i>	OsSBeL1-Putative Serine Beta-Lactamase homologue, expressed
	LOC_Os01g42960		TTL1, putative, expressed
	LOC_Os05g44922		6-phosphofructokinase, putative, expressed
	LOC_Os02g05630	<i>OsPP18</i>	protein phosphatase 2C, putative, expressed
	LOC_Os06g08860		transporter family protein, putative, expressed
	LOC_Os09g16910		cysteine desulfurase 1, mitochondrial precursor, putative, expressed
3as	-	-	
4as	LOC_Os05g15180		expressed protein
	LOC_Os09g32250		DNA repair ATPase-related, putative, expressed
	LOC_Os09g08900		retrotransposon protein, putative, Ty3-gypsy subclass, expressed
5as	LOC_Os01g72230		stromal membrane-associated protein, putative, expressed
	LOC_Os12g33740		retrotransposon protein, putative, unclassified
	LOC_Os01g17250		BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative, expressed
6as	LOC_Os03g31410		expressed protein
	LOC_Os07g43380		zinc finger, C3HC4 type domain containing protein, expressed
	LOC_Os06g12530		CS domain containing protein, putative, expressed
	LOC_Os06g28550		nmrA-like family domain containing protein, expressed
	LOC_Os01g52540		B3 DNA binding domain containing protein, expressed
7as	LOC_Os03g38440		transposon protein, putative, unclassified, expressed
	LOC_Os03g30760		retrotransposon protein, putative, unclassified, expressed
	LOC_Os04g56850	<i>OsARF11</i>	auxin response factor, putative, expressed
	LOC_Os10g05980	<i>OsPRP2</i>	POEI13-Pollen Ole e I allergen and extensin family protein precursor, expressed
	LOC_Os07g28670		retrotransposon protein, putative, unclassified, expressed
	LOC_Os08g11970		retrotransposon protein, putative, Ty3-gypsy subclass, expressed
8as	LOC_Os06g20320		peptidyl-prolyl cis-trans isomerase, FKBP-type, putative, expressed
	LOC_Os01g07740		DEAD-box ATP-dependent RNA helicase 14, putative, expressed
	LOC_Os07g37950		expressed protein
	LOC_Os08g35000	<i>OsARID4</i>	ARID/BRIGHT DNA-binding domain containing protein, expressed
	LOC_Os05g14170	<i>RFC3</i>	expressed protein
9as	LOC_Os03g19600		retrotransposon protein, putative, unclassified, expressed
10as	LOC_Os12g41700		LSD1 zinc finger domain containing protein, expressed
	LOC_Os08g03560		chloroplast unusual positioning protein, putative, expressed
	LOC_Os03g27054		expressed protein
	LOC_Os05g12140		Leucine Rich Repeat family protein, expressed
	LOC_Os07g17280		Ser/Thr protein phosphatase family protein, putative, expressed
	LOC_Os05g03430	<i>OsSIZ1</i>	ATSIZ1/SIZ1, putative, expressed
	LOC_Os03g14600		transposon protein, putative, unclassified, expressed
	LOC_Os06g10740		transposon protein, putative, unclassified, expressed

10as	LOC_Os10g12140		transposon protein, putative, unclassified, expressed
	LOC_Os02g02480		basic helix-loop-helix DND-binding domain containing protein, expressed
11as	LOC_Os07g22534		WD domain, G-beta repeat domain containing protein, expressed
	LOC_Os05g35530		expressed protein
	LOC_Os01g11040	<i>TUT1/ESI</i>	SCAR-like protein 2, putative, expressed
12as	LOC_Os05g37700		periplasmic beta-glucosidase precursor, putative, expressed
	LOC_Os11g32010		expressed protein
	LOC_Os04g19870		expressed protein
	LOC_Os10g41920		expressed protein
13as	LOC_Os01g63930		cytochrome P450, putative, expressed
14as	-		-
15as	-		-
16as	LOC_Os03g19840		cysteine-rich repeat secretory protein 55 precursor, putative, expressed
	LOC_Os01g08960		phosducin-like protein 3, putative, expressed
	LOC_Os01g59620	<i>SDG709</i>	histone-lysine N-methyltransferase, H3 lysine-9 specific SUVH1, putative, expressed
	LOC_Os11g03700	<i>SDG733</i>	histone-lysine N-methyltransferase, H3 lysine-9 specific SUVH1, putative, expressed
	LOC_Os01g02880		fructose-bisphosphate aldolase isozyme, putative, expressed
17as	LOC_Os11g15490		expressed protein
	LOC_Os01g66250		S-locus-like receptor protein kinase, putative, expressed
	LOC_Os02g57070		3-5 exoribonuclease CSL4, putative, expressed
	LOC_Os10g18220		GHMP kinases ATP-binding protein, putative, expressed
	LOC_Os03g09970		sulfate transporter, putative, expressed
18as-s	LOC_Os03g18550	<i>MIT</i>	mitochondrial carrier protein, putative, expressed
	LOC_Os02g36110	<i>CYP76M7</i>	cytochrome P450, putative, expressed
	LOC_Os03g23080		retrotransposon protein, putative, unclassified, expressed
18as-m	LOC_Os03g18550	<i>MIT</i>	mitochondrial carrier protein, putative, expressed
	LOC_Os02g36110	<i>CYP76M7</i>	cytochrome P450, putative, expressed
	LOC_Os03g23080		retrotransposon protein, putative, unclassified, expressed

在如此众多的靶基因中如何确定哪一个才是真正的靶基因？归根结底还是必须用 5' RLM-RACE 在体内进行验证。考虑到 *PMSIT*-phasiRNA 中 sense phasiRNA 的表达丰度普遍高于 antisense phasiRNA，可能更重要，于是对所有这些 sense phasiRNA 的预测靶基因进行 5' RLM-RACE 验证，附录 I 附表 2 中列出了所有能设计特异 RACE PCR 引物的预测靶基因及其引物序列。但最终没有任何一个预测靶基因得到证实。

至此我们并没有得到很多有用的信息，因此，我们将 psRNATarget 的 Maximum expectation value 提高到 4.0，共得到 1634 个可能的靶基因，其中 10s phasiRNA 的

预测靶基因 *LOC_Os02g02820* (*TAPETUM DEGENERATION RETARDATION, TDR*) 最值得关注。*TDR* 优先在水稻绒毡层表达, 对水稻绒毡层降解和花药发育是必需的, 突变后将导致雄性不育(Li et al 2006; Frouin et al 2014)。10s phasiRNA 预测结合到 *TDR* 转录本的第 2,028 bp 到 2,048 bp 位点上, 位于 3' UTR 上(图 70A, B)。预测的剪切位点在长日照的 NIL(MH)和短日照农垦 58S 的 P3 期幼穗 PARE 文库中能检测到信号(图 70B)。但在进行 5' RLM-RACE 验证时发现该剪切位点距离 *TDR* 转录本的末端仅 116 bp, 无法满足 5' RLM-RACE 引物设计要求距离 5'端至少 150 bp 的条件, 尝试设计了引物 TDR-R-RACE (5'-TAGCCACATTTTCAGTCATCAAGTC-3') 和 TDR-R-RACE-2 (5'-TCCAACAAAAGTGGAAAAACAC-3') 进行扩增, 预期片段大小分别为 53 bp 和 101 bp, 都未得到目标条带。

此外, 分析之前的农垦 58S 和 NIL(MH)的 small RNA 文库数据, 发现在 *TDR* 基因区段不会产生 siRNA (图 70C), 排除了 *TDR* 基因能形成次级 siRNA 并进一步 target 其他基因的可能性。

在水稻绒毡层发育过程中, *TDR* 及其互作基因构成了一个调控绒毡层降解的网络(图 70D), *TIP2* (*TDR INTERACTING PROTEIN 2*) 位于 *TDR* 的上游, 并且两者能形成异源二聚体, 结合到 *EAT1* (*ETERNAL TAPETUM 1*) 的启动子区, 协调调控 *EAT1* 的表达; 同时 *EAT1* 也能与 *TDR* 结合形成异源二聚体, 两者都能调控 *CPI* (*CYSTEINE PROTEASE 1*) 基因的表达(Lee et al 2004; Li et al 2006; Niu et al 2013; Fu et al 2014; Ko et al 2014)。细胞学研究表明农垦 58S 的花粉败育也是由绒毡层的推迟降解造成(Shi et al 2009; Ding et al 2012a), 因此分析长、短日照下农垦 58S 和 NIL(MH) P2 期到 P5 期的幼穗中 *TDR* 相关基因的表达谱, 以判断其是否也参与了光敏感雄性不育的花粉发育过程。结果表明 *TDR* 与其上游调控基因 *TIP2* 从 P2 期到 P5 期的整体表达量呈逐渐上升的趋势, 与 *PMSIT* 的表达趋势正好相反(图 38 和图 70E); 在 P3 期、P4 期和 P5 期, *TDR* 和 *CPI* 基因的表达量在长日照农垦 58S 中低于其他三种可育材料, 这种趋势与 P3 期 *PMSIT* 和 P4 期 *TIP2* 的表达趋势相似, 而与 P5 期 *EAT1* 的表达趋势相反。该结果表明 *TDR* 及其相关基因可能受 *PMSIT* 调控, 导致长日照农垦 58S 的绒毡层细胞降解异常及 PCD 过程提前, 造成雄性不育。

虽然到目前为止还没有确切的证据表明 *TDR* 是 *PMSIT*-phasiRNA 的靶基因, 但在以后的研究中应将此作为重点。

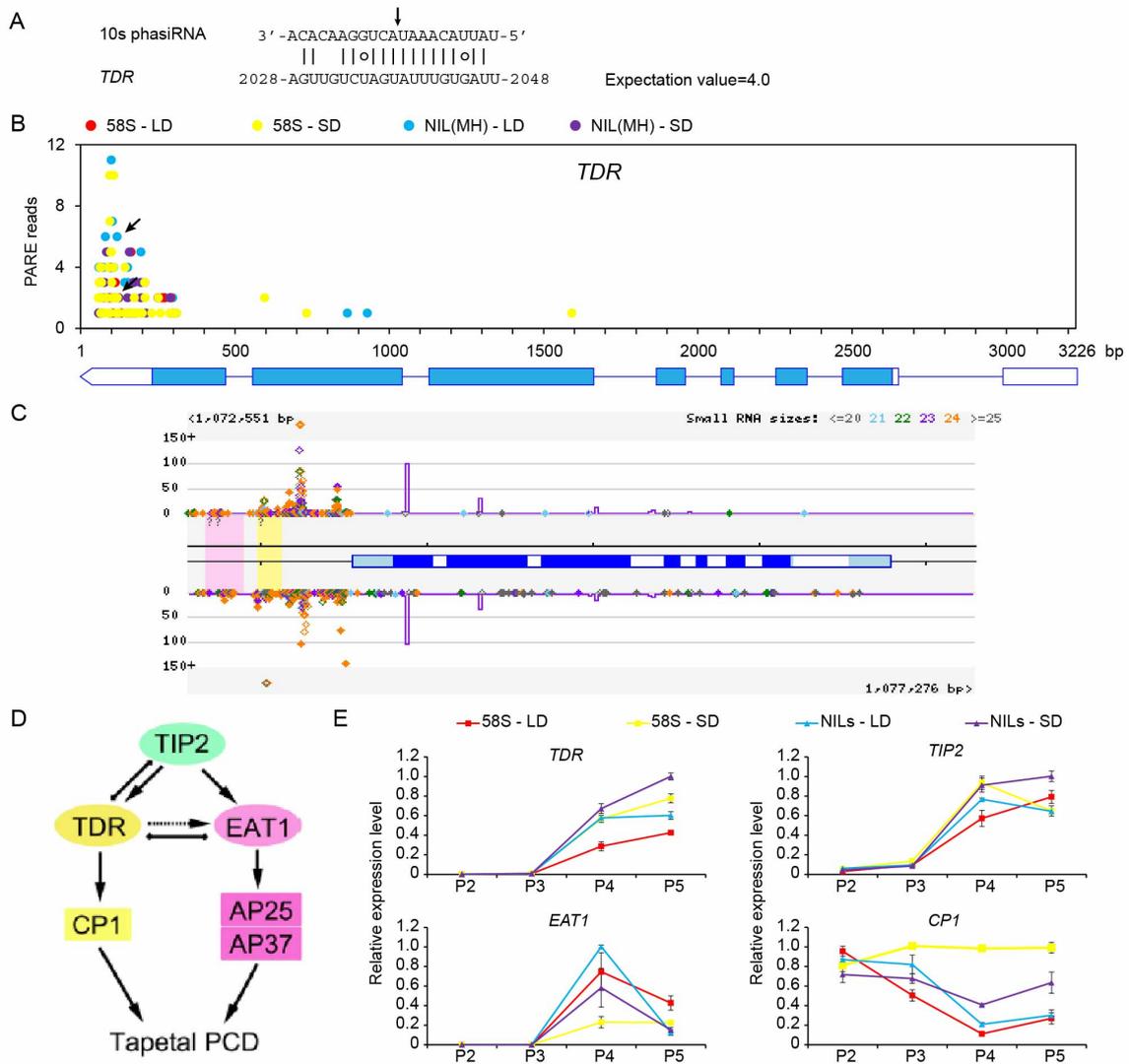


图 70 TDR 可能是 10s phasiRNA 的靶基因。

A, 预测的 10s phasiRNA 和 *TDR* 转录本之间的 miRNA:mRNA 匹配结果, 图中竖线和圆圈分别代表碱基完全配对和 G:U 配对。B, *TDR* 转录本在长、短日照下农垦 58S 和 NIL(MH) P3 期幼穗中的 PARE 结果, 箭头指示了 10s phasiRNA 可能的剪切位点处的信号, 图下方对应着 *TDR* 的基因结构图。C, 长、短日照下农垦 58S 和 NIL(MH) P2, P3 和 P4 期幼穗中 *TDR* 基因区段的小 RNA 测序结果。基因结构图和方向与 (B) 中相同。D, *TIP2*, *TDR* 和 *EAT1* 共同调控绒毡层细胞的 PCD 过程, 此图来自于 Fu et al 2014。E, *TDR* 及其相关基因在长、短日照下农垦 58S 和 NIL(MH) 中的相对表达量分析。

Figure 70 TDR could be the potential target of 10s phasiRNA.

A, Putative miRNA:mRNA alignment between 10s phasiRNA and *TDR*. Vertical lines and circles represent matched and G:U base pairs, respectively. B, PARE data of *TDR* transcript from libraries of Nongken 58S and NIL(MH) panicles at P3 stage under long-day and short-day conditions. Arrows indicate the cleavage sites of miRNA or siRNA.

The diagrams of gene model are shown below. C, The small RNA data of TDR region from libraries of Nongken 58S and NIL(MH) panicles at P2, P3 and P4 stages under long-day and short-day conditions. The diagrams of gene model and the direction of transcript was as same as in (B). D, The interaction between *TIP2*, *TDR* and *EAT1* in promoting tapetal PCD. The picture was collected from Fu et al 2014. E, The expression pattern of *TDR* and related genes in Nonken 58S and NIL(MH) under long-day and short-day conditions.

3. 10 *PMSIT* 区间 DNA 甲基化分析

水稻中早先克隆的另一个光敏感雄性核不育基因 *pms3* 同样编码一个 lncRNA LDMAR, 农垦 58S 和农垦 58 序列间的单碱基差异引起长日照下农垦 58S 中 LDMAR 启动子区 DNA 甲基化水平升高, 造成表达量下降, 导致雄性不育(Ding et al 2012a)。因此我们用重亚硫酸盐测序的方法也对 *PMSIT* 启动子和转录本区间的 DNA 甲基化修饰情况进行了分析, 引物序列详见附录 I, 选取的材料包括长日照下的 NIL(MH) 和长、短日照下农垦 58S P3 期的幼穗及该发育时期生长的最后一片展开叶。

在叶片中, *PMSIT* 启动子区的 DNA 甲基化集中在距离转录本起始位点 300 bp 到 700 bp 处, 该区段内 DNA 甲基化程度非常高, 以 CG 和 CHG 甲基化为主, 但在长日照下的 NIL(MH) 和长、短日照下农垦 58S 中并没有明显区别, 仅在 P5 标记的上游处的单个碱基上存在 CG 甲基化水平的差异; 而 *PMSIT* 转录本区间的 DNA 甲基化从 SNP S1 位点下游开始, 集中在农垦 58S 插入的 65 bp 处, 除开这段插入的序列, 检测的 3 个叶片材料在 *PMSIT* 转录本上的 DNA 甲基化水平也没有明显差异(图 71A)。以上两方面说明在叶片中 *PMSIT* 启动子和转录本上的 DNA 甲基化水平不受光照长度影响, 并且在农垦 58S 和 NIL(MH) 中也没有区别。在此基础上, 进一步分析幼穗中以上两个重点区间的 DNA 甲基化修饰情况, 得到与叶片中类似的结果(图 71B)。表明 *PMSIT* 的甲基化修饰也不存在组织差异, 并且与 *PMSIT* 调控的光敏感雄性不育无关。

但是我们也注意到两个比较有意思的地方: 首先, *PMSIT* 转录本的 5'端 DNA 甲基化水平很低, 从 SNP S1 的下游开始 CG 甲基化程度急剧升高, 而 SNP S1 恰好位于 17s phasiRNA 和最后一个 phasiRNA——18as phasiRNA 上, 推测 SNP S1 的下游的 DNA 高度甲基化可能造成 miR2118 介导的 *PMSIT*-phasiRNA 产生的终止, 导致仅形成 18 对 phasiRNA, 而不是一直延续下去。其次, 农垦 58S 插入的 65 bp 序

列的 DNA 甲基化程度非常高。研究表明植物基因组中转座子通常会发生高度的 DNA 甲基化(Zhang et al 2006), 并且在哺乳动物中发现转座子通常存在于成熟的 lncRNA 序列中, 可能与 lncRNA 的进化及组织特异性表达相关(Kelley and Rinn 2012; Kapusta et al 2013; Fatica and Bozzoni 2014), 之前根据此 65 bp 在基因组中拷贝数非常高等特征推测此 65 bp 可能由转座子或者反转座子插入造成, 此处结果更进一步说明该段序列可能与转座子插入相关, 并且可能对 lncRNA *PMSIT* 的进化或其他方面起着重要作用。

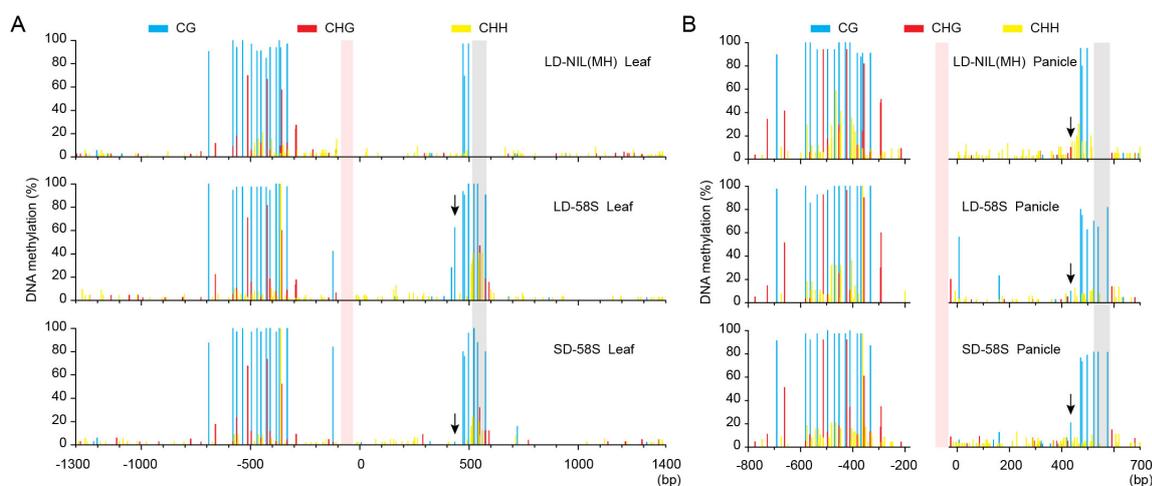


图 71 长、短日照下不同组织中 *PMSIT* 基因区段的 DNA 甲基化水平。

A, 农垦 58S 和 NIL(MH)叶片中 *PMSIT* 基因区段 DNA 甲基化水平, 包含 1,300 bp 启动子和 1,400 bp 转录本区间。B, 农垦 58S 和 NIL(MH)幼穗中 *PMSIT* 基因区段 DNA 甲基化水平, 包含 800 bp 启动子和 700 bp 转录本区间。垂直箭头指示了 SNP S1 所在的位置。红色和灰色方框分别对应明恢 63 在分子标记 P5 处插入的 52 bp 区间和农垦 58S 在分子标记 P6 处插入的 65 bp 区间。

Figure 71 DNA methylation levels of *PMSIT* region from different tissues under long-day and short-day conditions.

A, DNA methylation levels of *PMSIT* region including 1,300 bp putative promoter and 1,400 bp *PMSIT* transcript from leaves of Nongken 58S and NIL(MH). B, DNA methylation levels of *PMSIT* region including 800 bp putative promoter and 700 bp *PMSIT* transcript from young panicles of Nongken 58S and NIL(MH). The vertical arrows pointed the position of SNP S1. Red and grey blank covered the area of 52 bp insertion at Minghui 63 in P5 marker and 65 bp insertion at Nongken 58S in P6 marker, respectively.

4 讨论

4.1 *pms1* 的作用机理研究

基于本研究的实验结果，我们提出 *pms1* 调控水稻光敏感雄性不育的分子模型（图 72）：光敏感雄性核不育 *pms1* 位点转录出长链非编码 RNA *PMSIT*，22 nt 的 miR2118 会结合到 *PMSIT* 的 5' 端，并介导剪切，触发从剪切位点开始形成 21 nt 的 phasiRNA，距离剪切位点 24 bp 的碱基在明恢 63 和农垦 58S 中存在从 G 到 T 的突变（图 72 中分别用黑点和红点示意），突变后造成 *PMSIT* RNA 二级结构改变，导致长日照下 miR2118 对农垦 58S *PMSIT* 的剪切效率大大增加，能够形成更多的 phasiRNA，这些 phasiRNA 靶定到其他靶基因，对靶基因转录本进行剪切或者抑制蛋白翻译从而抑制靶基因的表达，引起雄性不育。而在短日照下，虽然 SNP 存在，但并不影响 miR2118 介导的剪切效率，都处于较低的水平，与长日照下的 NIL(MH) 中一样，不造成体内 *PMSIT*-phasiRNA 的积累，因此育性不受影响。

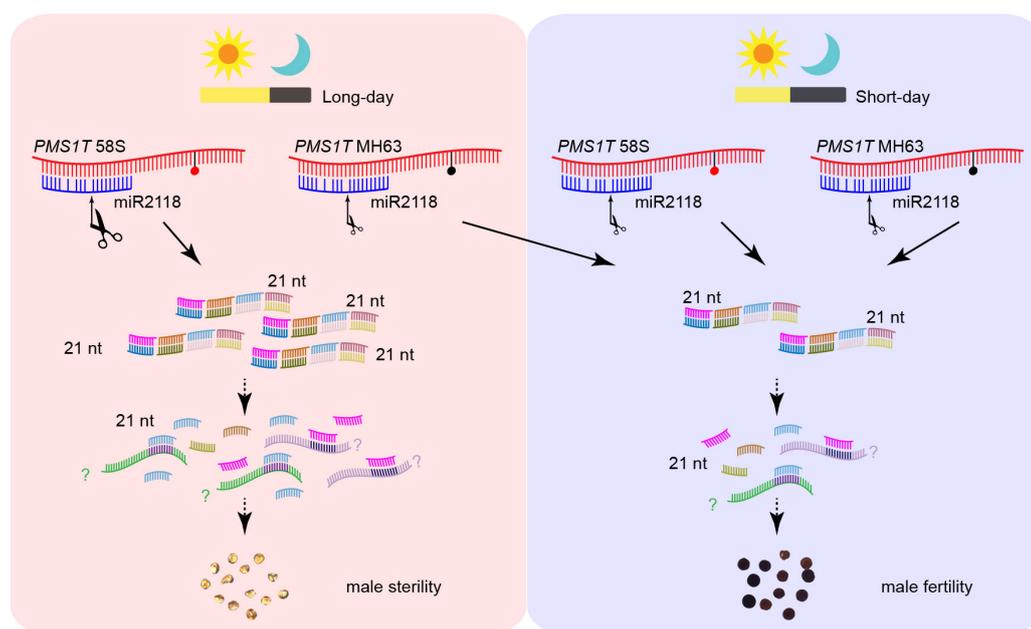


图 72 *pms1* 调控光敏感雄性核不育的作用机理。

图中剪刀的大小示意了 miR2118 介导的剪切效率的高低。

Figure 72 A hypothetical mechanistic model for *pms1*-regulated photoperiod-sensitive male sterility in rice.

The size of the scissors at the cleavage site of miR2118 indicates the cleavage efficiency mediated by miR2118.

在我们的研究过程中，将农垦 58S 的 *PMSIT* 互补转化 NIL(MH) 时，长日照下会降低 NIL(MH) 的育性，而将供体和受体反过来进行转化的 MH-C 转基因植株的育性并没有改变 (图 30)；与该结果一致，抑制农垦 58S 和 NIL(MH) 中 *PMSIT* 的表达，分别会降低和不影响育性 (图 33)。这些结果一方面证实了 *PMSIT* 就是 *pms1* 的候选基因，另一方面也与群体遗传结果相印证，表明农垦 58S 中的 *pms1* 是一个不完全显性位点。后来在 NIL(MH) 中超量表达农垦 58S 的 *PMSIT* 确实能够极大地降低育性，但在农垦 58S 中超量表达明恢 63 的 *PMSIT* 居然也能够恢复长日照下的育性 (图 36)，似乎与之前的结果有些矛盾。我们注意到在不包含完整 miR2118 结合位点的超表达植株 35S:S-CS 和 35S:M-CS 中对育性都没有任何影响 (图 60)。结合后来得到的 *pms1* 通过 phasiRNA 发挥作用的分子机理，可以很好地解释 Ubi:M 的转化结果：农垦 58S 和明恢 63 的 *PMSIT* 会竞争结合 miR2118，导致 phasiRNA 的丰度改变。在 Ubi:M 转基因植株中，超量表达的明恢 63 *PMSIT* 会与内源的农垦 58S *PMSIT* 竞争结合 miR2118，但又不能有效地产生 phasiRNA，因此 Ubi:M 中 *PMSIT*-phasiRNA 不会累积，育性得到恢复；相反，在 35S:M-CS 中由于超量表达的外源明恢 63 *PMSIT* 不包含完整的 miR2118 结合位点，不能与内源完整的农垦 58S *PMSIT* 竞争结合 miR2118，也就不能恢复农垦 58S 的育性。

PMSIT 被剪切后从剪切位点开始，理论上一共能形成 18 对 phasiRNA，这些 phasiRNA 并不是每一条都在体内能够检测到表达，并且表达水平各不一样，其中以 4s phasiRNA 和 6s phasiRNA 的表达丰度最高 (图 63)。载体 58S-ORF1+G，58S-ORF2+G 和 58S-ORF3+G 是在预测的 *PMSIT* 的 3 个 ORF 的起始密码子 ATG 后面插入碱基 “G” 以打乱正常的蛋白质读码框，同时也会造成 mRNA 序列的改变，进而干扰 phasiRNA 的形成。miR2118 的剪切位点位于 *PMSIT* 的第 81 个碱基上，而 ORF1 的预测翻译起始位点位于第 241 个碱基上 (图 41)，恰好是 8s phasiRNA 的第 13 位，因此 58S-ORF1+G 中插入的碱基 “G” 位于 8s phasiRNA 的第 16 位。同样地，58S-ORF2+G 中插入的碱基 “G” 位于 17s phasiRNA 的第 4 位，而 58S-ORF3+G 中插入的碱基 “G” 已经超出了 *PMSIT*-phasiRNA 的形成范围。所以在这些载体中插入的碱基并不会影响 miRNA2118 识别 *PMSIT* 和介导 phasiRNA 的形成，但会导致从 8s phasiRNA 开始相位的移位，而在所有 *PMSIT*-phasiRNA 中占大多数的 4s phasiRNA 和 6s phasiRNA 的产生在这三个载体中不会有改变。该结果

也间接地将与 *PMSIT* 光敏感雄性不育相关的 sense phasiRNA 的范围缩小到前面的 7 个 phasiRNA 内。

虽然我们建立了以 phasiRNA 为基础的 *pms1* 调控光敏感雄性不育的作用机理，但还有许多地方值得探讨。从上游上来看，*PMSIT* 是如何对长、短日照响应？其中一种猜测是 SNP2 会影响 *PMSIT* 在不同光照长度下的 RNA 二级结构的改变，可能存在两种方式：一是光照长度从物理上改变 RNA 的二级结构，已知温度和一些化学物质会改变 RNA 的折叠，但并没有关于光照时间或光质影响 RNA 二级结构的报道，目前我们也没有很好的方法来验证此推论；另外，可能 *PMSIT* 的 RNA 在折叠的过程中需要其他蛋白或者 RNA 的帮助，光照长度会影响这些蛋白和 RNA 的功能，或者影响 SNP2 可能参与介导 *PMSIT* 与这些蛋白或 RNA 之间的结合。

另一方面，miRNA2118 介导了 *PMSIT*-phasiRNA 的形成，miR2118 是否与光敏感雄性不育相关？我们用 target mimicry 的方法抑制了农垦 58S 和 NIL(MH) 中 miR2118d 和 miR2118o 的表达，没有看到任何表型的改变。在苜蓿中超量表达 miR2118a/b/c 会造成大量 phasiRNA 表达量的上升(Fei et al 2015)。番茄 22 nt 的 miR4376 能够结合到钙离子 ATP 酶基因 *ACA10* 上，产生 21 nt 的 phasiRNA。将 miR4376 超表达后 *ACA10* mRNA 的表达量下降，会改变花形态，同时降低成熟果实的产量。奇怪的是将 *ACA10* 上 miR4376 的识别位点上的第 9 个碱基突变后同样进行超表达，也能表现出与 miR4376 相同的表型，而不是相反，推测可能过高或者过低的 *ACA10* 表达量会造成相似的表型(Wang et al 2011)。那么，如果 target mimicry miR2118d 和 miR2118o 后因为 miR2118 家族中基因功能的冗余导致无法看到表型改变，超量表达这些 miR2118 后按理会降低体内 *PMSIT* 的表达量，因此，在农垦 58S 中进行 miR2118 的超量表达可能导致与 58S-dsi 相似的表型，即育性恢复。同时，可突变农垦 58S *PMSIT* 上 miR2118 结合位点再进行互补或者超表达 NIL(MH) 看是否能降低受体的育性。值得注意的是，与以往认为影响 miRNA 与靶基因的结合和剪切的碱基位于第 9 到 11 位不同(Mallory et al 2004; Liu et al 2014)，miRNA 的 3' 末端几个碱基对 phasiRNA 的产生更为重要(Fei et al 2015)。因此，对这几个碱基对应的位点进行突变可能更重要。

另外，参与 phasiRNA 形成过程中的各种蛋白在 *PMSIT*-phasiRNA 形成过程中扮演什么样的角色？在农垦 58S 中突变 DCL4、RDR6、SGS3 是否能够恢复长日照

下的育性？到目前为止对于 *phasiRNA* 形成过程中需要哪些基因参与还不是研究得很透彻，农垦 58S 可能作为一个重要的材料能够用于鉴定这些基因。

但最让我们关注的是 *PMSIT-phasiRNA* 的靶基因是什么。虽然我们也花了大量的时间和精力去验证预测靶基因是否能被 *PMSIT-phasiRNA* 结合，遗憾的是都没有得到证实。其中最有可能的是 10s *phasiRNA* 的靶基因 *TDR*，有 PARE 数据的支持(图 70B)，并且 *TDR* 的表达量在长日照下农垦 58S 中的表达量比 NIL(MH)中低(图 70E)，在没有 5' RLM-RACE 结果的证实下还不能断言 *TDR* 就是 *PMSIT-phasiRNA* 的靶基因。*TDR* 突变后会造成雄性不育，可以预见在 NIL(MH)中突变 *TDR* 也会降低育性，那么在农垦 58S 中超量表达 *TDR*，以弥补 *PMSIT-phasiRNA* 对 *TDR* 的抑制作用，能否恢复长日照下的育性？如果答案是肯定的，那么也能从遗传上证实 *TDR* 确实位于 *pms1* 的下游。此外，利用 RNA-seq 寻找 *PMSIT-phasiRNA* 的靶基因也不失为一个好办法。根据我们对 *pms1* 作用机理的研究，应该重点分析长日照下 P3 期幼穗中在农垦 58S 中相对 NIL(MH)下降表达的差异基因，这些极有可能是 *PMSIT-phasiRNA* 的靶基因，然后分析差异基因上是否存在 *phasiRNA* 的靶位点，再通过实验验证及后续遗传转化验证。

4.2 未来研究方向

基因调控网络是多方面的，*pms1* 可能不止通过 *PMSIT-phasiRNA* 发挥功能这一条途径，还可能存在其他的作用方式，为了更深入地研究 *pms1* 的作用机理，未来还可以从以下几个方面开展研究：

其一是对 *PMSIT* 本身进行分析。共分离标记 P5 位于 *PMSIT* 的启动子区，除此之外农垦 58S 和明恢 63 的启动子区没有其他序列差异，DNA 甲基化水平也不存在明显差异（图 71）。P5 处的差异可能造成启动子顺式元件的差异，导致其他转录因子与 *PMSIT* 结合发生改变；另外就是改变基因的时空表达。之前我们也尝试过构建农垦 58S 和明恢 63 的启动子连报告基因的载体，试图用报告基因追踪 *PMSIT* 的时空表达，并比较两者是否存在差异。由于 P5 在农垦 58S 和明恢 63 中分别有 9 个和 40 个 AT 重复，在进行 PCR 扩增时极容易出现打滑或者扩增不完全的情况，无法得到无突变的完整启动子片段，尤其是明恢 63 的。如今高效构建载体的方法层出不穷，有助于我们完成这项工作，从整个植株的水平上比较 *PMSIT* 的时空表达。

与此相类似的,也可以采用原位杂交的方法分析 *PMSIT* 的表达模式。我们也在这方面进行过尝试,由于 *PMSIT* 的表达量非常低,即便是在高表达的幼穗中也无法获得清晰的结果,细化到小孢子发育过程中的各层细胞的杂交结果非常模糊,与杂交背景之间的对比度不高。如果尝试优化原位杂交的各种条件,有可能得到较好的结果。

植物感受光照长度的组织是叶片,水稻中控制光周期开花的基因许多都在叶片中高表达(Xue et al 2008),而 *PMSIT* 主要在幼穗中表达,在叶片中表达量极低,几乎不表达,是否 *PMSIT* 在植物体内存在从叶片到幼穗的运输?为了验证这一点,可以利用 *PMSIT* 自身启动子再加上基因序列融合 GFP 报告基因,然后转化水稻植株,分析在长、短日照下 *PMSIT* 基因在各组织内的动态表达情况。

研究表明 lncRNA 在细胞核与细胞质中都有表达,大部分在细胞核中表达,与染色质相关(Zhang et al 2014a; Chekanova 2015)。将 *PMSIT* 序列与报告基因融合,进行亚细胞定位,有助于我们认识 *PMSIT* 发挥功能的场所。

其二是对 *PMSIT* 互作蛋白的分析。在动物细胞中筛选 lncRNA 的互作蛋白有一些比较成熟的方法,成功分离到了许多与 lncRNA 直接互作的蛋白,并得到实验证实(McHugh et al 2015; Yang et al 2015)。但在植物中还未有报道,一方面是因为植物中已知功能的 lncRNA 基因没几个,另一方面植物细胞由于有细胞壁,无法像动物细胞一样操作,这些方法无法直接应用。在 lncRNA 的作用机理中,lncRNA 作为“向导”、“诱饵”和“脚手架”都需要与其他蛋白质结合发挥作用(图 5)(Wang and Chang 2011; Yang et al 2015),因此分离互作的蛋白质对于理解 lncRNA 的分子机理具有重要意义(Goff and Rinn 2015; Chi 2016)。

其三是对 *PMSIT*-phasiRNA 的研究,包括确定 18 对 phasiRNA 中到底是哪几个起作用及他们的功能。从已有实验结果推测,4s phasiRNA 和 6s phasiRNA 可能对光敏感雄性不育更为重要。可以通过将 58S:C 和 Ubi:S 载体中包含这两个 phasiRNA 的前体序列分别缺失和共同缺失后再进行转化,以观察表型是否变化。同时,为了减少内源 *PMSIT* 基因的影响,还可以考虑用 CRISPR 的方法直接对农垦 58S 中的 *PMSIT* 进行敲除整个 4s phasiRNA 和 6s phasiRNA,这项技术在水稻中已得到广泛应用(Ma et al 2016)。然后再对这些转基因材料进行 RNA-seq 分析,更有针对性和目的地寻找其下游靶基因。针对 phasiRNA 还可以进行原位杂交分析,对玉米花药

特异表达的 *phasiRNA* 原位杂交分析表明, *miR2118* 在花药的表皮中表达, 而 *21-PHAS* 基因在小孢子中表达, *21 nt phasiRNA* 在除表皮层外的其他花药细胞中都有表达; 介导产生 *24 nt phasiRNA* 的 *miR2275* 和 *24-PHAS* 基因及 *24 nt phasiRNA* 都在绒毡层和小孢子母细胞中优势表达(Zhai et al 2015)。因此, 从细胞学上比较分析 *PMSIT*-*phasiRNA* 在长、短日照下的表达部位和表达量差异是可行的。

4.3 应用前景

作为一个不完全显性基因, 雄性不育是显性的特征限制了 *pms1* 在生产中的应用, 但是只要 *pms3* 和 *pms1* 有一个位点是可育的, 植株就表现为可育, 说明 *pms1* 与 *pms3* 间存在着某种联系, 也解释了之前光敏雄性不育的复杂性。利用培矮 64S 与 9311 组合定位到的光敏不育位点 *pms1(t)* 与之前报道的精细定位的 *pms1* 的位置非常接近, 结合两个课题组分别利用培矮 64S 和农垦 58S 定位都定位到相同的 *pms3* 基因的结果, 推测 *pms1(t)* 的候选基因极有可能就是 *pms1*, 而 *pms1(t)* 实际离 *PMSIT* 的距离超过了 70 kb, 由于对基因显隐性关系的认识不同, 极有可能导致出现这种定位的偏差。为了验证这一点, 可以在培矮 64S 中对 *PMSIT* 进行抑制表达, 看能否恢复其育性。这也提示我们在克隆其他光温敏雄性不育基因的过程中有必要对基因的显隐性关系进行严格分析, 低世代群体的表型由于受到背景基因的影响太大, 可能造成误导。

将农垦 58S 的 *PMSIT* 转到 NIL(MH) 中会得到长日照下不育、短日照可育的材料, 为创造新型光敏雄性不育系提供了新的思路。结合 *phasiRNA* 参与的光敏雄性不育过程, 如果将农垦 58S 中的 *PMSIT* 超量表达, 得到更多的 *phasiRNA*, 能否克服温度的影响, 获得育性不受温度波动影响的光敏材料? 同时也表明了 *lncRNA* 和 *phasiRNA* 在控制重要农艺性状上的重要性。

4.4 Small peptide 与 *lncRNA*

lncRNA 的定义就是长链非编码 RNA, 但数据分析表明动植物的许多 *lncRNA* 也具有编码短肽的能力(Andrews and Rothnagel 2014; Bazzini et al 2014; Ruiz-Orera et al 2014)。苜蓿和拟南芥的 *miRNA* 前体 *pri-miR171b* 和 *pri-miR165a* 分别产生 20 aa

的短肽 miPEP171b 和 18 aa 的 miPEP165a, 这些短肽能促进其成熟 miRNA 的富集, 增强对靶基因的下降调控, 被称为 miPEP (miRNA-encoded peptide)。除此外还发现了另外 5 个具有类似激活功能的 miPEP, 说明这种短肽在植物中广泛存在。对苜蓿施加合成的 miPEP171b, 相比对照短肽能明显减少侧根的形成, 同样对拟南芥施加 miPEP165a 会增加主根的根长, 表明 miPEP 在农业生产上具有利用价值 (Lauressergues et al 2015)。后来在大豆中也找到了一个编码 16 个氨基酸的 miPEP, miPEP172c。在水中加入 miPEP172c 灌溉大豆后能提高根瘤的数目, 而对根的形成无影响, 这些根瘤都能有效固氮 (Couzigou et al 2016)。在小鼠和人类的骨骼肌特异表达 lncRNA 中找到一段保守的 ORF, 编码 46 个氨基酸的短肽, 命名为 MLN (myoregulin), MLN 通过抑制 SERCA (sarcoplasmic reticulum Ca^{2+} -ATPase) 的活性来调控骨骼肌中钙离子的浓度。SERCA 是哺乳动物中调控横纹肌表现及反馈心血管和骨骼肌病理信号的中心调节子 (Anderson et al 2015)。通过同样的方法在小鼠中找到另一个来自于 lncRNA 的编码 34 个氨基酸的短肽 DWORF (dwarf open reading frame), 与 MLN 功能相反, DWORF 能增强 SERCA 的活性。这些实例都证实了 lncRNA 能作为 RNA 和短肽参与基因的表达调控 (de Andres-Pablo et al 2016)。

那么 lncRNA *PMSIT* 是否也具有编码短肽的能力? 之前我们是对农垦 58S *PMSIT* 预测的 3 个最长的 ORF 进行读码框干扰后仍能正常发挥功能, 因此认为 *PMSIT* 不编码蛋白质。但是 *PMSIT* 上可能存在能编码更短的短肽的 ORF, 要验证这一点, 除了花费大量的时间对 *PMSIT* 可能编码的短肽进行一一验证外, 还可以通过分离总蛋白中的低分子量蛋白进行测序分析, 再与预测短肽序列比对。根据 Ubi:S CS+P6 的转化结果, 可以将目标区段缩小至 *PMSIT* 的前 696 bp。

4.5 phasiRNA 与雄性不育

植物中 phasiRNA 在幼穗中优势表达, 与雄性不育密切相关。水稻 *MEL1* (*MEIOSIS ARRESTED AT LEPTOTENE1*) 基因编码 Agronaute 蛋白, 是拟南芥 AGO5 的同源蛋白, 能够结合到从基因组上超过 700 个 lincRNA 产生的 21 nt phasiRNA 上, 这些 lincRNA 包含有一个或多个 miR2118 的识别位点 (Komiya et al 2014)。MEL1 是小孢子发育减数分裂期同源染色体联会所必须的, 在精细胞中特异表达, 突变后会造造成雄性不育 (Nonomura et al 2007), 说明 21 nt phasiRNA 对水稻雄性不育发育的重

要性。对玉米花粉发育各时期 *phasiRNA* 分析表明 21 nt *phasiRNA* 在减数分裂前期表达量最高，而 24 nt *phasiRNA* 的表达峰值出现在减数分裂完成后(Zhai et al 2015)。*PMSIT-phasiRNA* 调控光敏雄性核不育的例子进一步证实了 21 nt *phasiRNA* 在植物雄性不育中发挥重要功能。

在动物细胞中特有的小 RNA piRNA 非常特异地在精细胞中表达，是精细胞发生所必须的，piRNA 要与属于 AGRONAUTE 亚家族的 PIWI 蛋白结合才能发挥作用。禾本科 *phasiRNA* 与哺乳动物 piRNA 存在很多相似的地方（图 73），可能在植物与动物中是平行进化的(Zhai et al 2015)。近来研究更是表明 piRNA 同样也能按照相位排列，也是 phased 小 RNA(Han et al 2015; Mohn et al 2015)。

Characteristic	Grass (maize) <i>phasiRNA</i> *		Mammalian piRNA [†]	
	Anther		Testis	
Organ	Anther		Testis	
Developmental stage	Premeiotic	Meiotic	Prepachytene	Pachytene
Peak timing	0.4-mm anther	2.0-mm anther	12.5 dpp	17.5 dpp
Size	21 nt	24 nt	26–27 nt	29–30 nt
AGO partner	AGO5c [‡]	AGO18b [‡]	MILI, MIWI2	MIWI
Master regulator	miR2118	miR2275	Unknown	A-myb
Number of loci	463	176	~900	~100
Abundance	Very high			
Distribution	Present on all chromosomes; most loci are clustered			
Precursor	Noncoding, Pol II transcripts			
Single copy?	Nearly all			
Repeat associated?	Few			
Impacts on fertility	Misregulated in sterile mutants; biogenesis defects can cause sterility			
Targeting	The majority lack complementarity to TEs or other loci			
Function(s)	Unknown		Mainly unknown; a subset silence TEs	

图 73 *phasiRNA* 与 piRNA 之间的异同(Zhai et al 2015)。

Figure 73 Parallels between grass *phasiRNAs* and mammalian piRNAs(Zhai et al 2015).

4.6 *pms1* 与 *pms3* 间关系的讨论

pms1 和 *pms3* 都调控水稻光敏感雄性核不育，这两个基因存在很多相似的地方：都编码 lncRNA，在幼穗中特异表达，功能性突变由 SNP 造成，都能产生 small RNA。也有很多不同：*pms1* 控制的雄性不育是不完全显性，而在 *pms3* 中是隐性；长日照下需要足量的 *LDMAR* 转录本才能保持可育，而 *PMSIT-phasiRNAs* 的累积会造成育性下降；*pms3* 的分子机制主要是 RdDM，而 *pms1* 涉及到 *phasiRNA*。在农垦 58S 中必须这两个基因都是不育位点才能表现为光敏雄性不育，任何一个位点为可育，植株都是可育的，表明这两个基因在调控光敏雄性不育上可能不是直接的上下游关系，而是属于两条不同的途径，但他们的下游基因存在交叉。

为了阐明这两个基因之间的关系，我们拟从以下几个方面展开研究：

首先从细胞学水平上分析两个基因的时空表达，主要采用原位杂交的方法，一方面是研究两个 lncRNA 在花粉发育过程中的各个时期的各层细胞中的表达异同，另一方面是分析两个基因产生的小 RNA 的时空表达。以长日照下的农垦 58S 为主，分析两个基因在这一个材料中的表达模式是否存在差异，再结合各自的可育亲本共同比较两个基因在 mRNA 和小 RNA 水平上是否存在促进或抑制表达的情况。

其次是分析这两个 lncRNA 的互作蛋白，可能存在某个蛋白能够同时结合到这两个 lncRNA 上。

另外是将两个基因进行杂交分析。以往的研究中将农垦 58S 与 1514 杂交分析了两个位点对育性的影响(Li et al 2001)，现在有了两个基因的近等基因系及转基因材料，将这些材料进行杂交，能最大程度地降低其他背景基因的影响。获得了杂交材料后可对这些材料进行 DNA 甲基化分析、RNA-seq 和 small RNA-seq 测序。DNA 甲基化结果能够用来分析 *pms1* 是否能影响 *pms3* 的甲基化程度；而 small RNA-seq 结果可分析 *pms3* 是否对 *PMSIT*-phasiRNA 产生影响；杂交材料的 RNA-seq 结果有助于寻找 *pms1* 和 *pms3* 产生的小 RNA 的共同下游靶基因。

参考文献

1. 陈良碧, 周广洽. 温光条件对水稻安农 S-1, 衡农 S-1 的育性及生理的影响. 植物学报, 1994, 36: 119-123
2. 曾汉来, 张自国, 张端品, 元生朝. 光(温)敏核不育水稻发育与育性转换光温反应特性鉴定. 湖北农业科学, 1996, 33-37
3. 陈克成, 肖翊华. 不同波长的光进行黑夜一中断对 HPGMR 育性的影响. 武汉大学学报(理学版), 1991, 127-128
4. 陈顺辉, 卢浩然, 杨聚宝. 籼型温敏核雄性不育基因的等位性研究. 中国农业科学, 1996, 29: 27-33
5. 邓华凤, 舒福北, 袁定阳. 安农 S-1 的研究及其利用概况. 杂交水稻, 1999, 14: 1-3
6. 邓启云, 盛孝邦, 段美娟, 李新奇. 籼型光温敏核不育水稻雄性不育性遗传研究. 杂交水稻, 2001, 16: 47-51
7. 邓晓建, 李仁端, 周开达, 杨仁崔. 水稻光温敏核不育系不育基因的等位性研究. 西南农业学报, 1997, 10: 1-6
8. 蒋佐升, 徐庆国, 董延瑜. 籼型两用不育系衡农 S-1 育性转换及育性遗传的研究. 作物研究, 1992, 6: 12-14
9. 李合生, 卢世峰. 湖北光敏感核不育水稻育性转换与光敏色素相关性的初步研究. 华中农业大学学报, 1987, 6: 397-398
10. 李亚芳. 光敏核不育水稻基因 *pms1* BAC 克隆微物理图谱的构建及相应片段农杆菌介导的遗传转化. [硕士学位论文]. 武汉: 华中农业大学图书馆, 2002
11. 刘南. 光敏核不育水稻基因 *pms1* 物理图谱的构建及候选基因的确定. [硕士学位论文]. 武汉: 华中农业大学图书馆, 2000
12. 刘宜柏, 贺浩华, 饶治祥, 孙义伟, 黄英金. 光温条件对水稻两用核不育系育性的作用机理研究. 江西农业大学学报, 1991, 13: 1-7
13. 卢开阳, 元生朝, 张自国. 光照长度对不同类型光敏核不育水稻生育期的影响. 华中农业大学学报, 1990, 112-115
14. 梅明华, 李泽炳. 水稻光(温)敏核不育系与核质互作不育系的遗传关系剖析. 遗传, 1995, 17: 199-204
15. 山燕. 水稻光敏核不育基因 *pms1* 目标 BAC 克隆的序列测定、分析及候选基因的确定. [硕士学位论文]. 武汉: 华中农业大学图书馆, 2000
16. 石明松. 对光照长度敏感的隐性雄性不育水稻的发现与初步研究. 中国农业科学, 1985, 44-48
17. 石明松, 邓景扬. 湖北光感核不育水稻的发现、鉴定及其利用途径. 遗传学报, 1986, 107-112+165
18. 斯华敏, 刘文真, 付亚萍, 孙宗修, 胡国成. 我国两系杂交水稻发展的现状和建议. 中国水稻科学, 2011, 544-552
19. 孙宗修, 熊振民, 闵绍楷, 斯华敏. 温度敏感型雄性不育水稻的鉴定. 中国水稻科学, 1989, 3: 49-55
20. 王风平, 梅明华, 徐才国, 张启发. 光敏核不育水稻农垦 58S 与正常品种“农垦 58”

- 在 *pms1* 区段无育性基因分离. 植物学报, 1997, 39: 922-925
21. 王志, 吴凡, 唐益文, 李乾安, 赵德明, 黄湘. 水稻新资源粳型两用核不育系绵 9S 的选育. 西南农业学报, 1999, 12: 11-14
 22. 向阳, 李必湖, 吴厚雄, 陈良碧. 光敏、温敏核不育水稻核不育基因等位性及基因对数的研究. 种子, 2002, 37-39
 23. 杨弘远. 水稻生殖生物学. 浙江: 浙江大学出版社, 2005.
 24. 杨仁崔, 李维明, 王乃元, 梁康迳, 陈青华. 粳稻光敏核不育种质 5460 ps 的发现和初步研究. 中国水稻科学, 1989, 3: 47-48
 25. 杨远柱, 唐平徕, 杨文才, 刘爱民, 陈运泉, 凌文彬, 石天宝. 水稻广亲和温敏不育系株 1S 的选育及应用. 杂交水稻, 2000, 15: 6-7
 26. 杨振玉, 张国良, 张从合, 陈金节, 王合勤, 张金家, 严志. 中粳型优质光温敏核不育系广占 63S 的选育. 杂交水稻, 2002, 17: 4-6
 27. 叶建荣, 黄美娟, 赵淑慧, 吴乃虎. *osRACD* 基因表达与光敏核不育水稻光周期育性转换的相关性. 自然科学进展, 2004, 14: 166-172
 28. 於金生. 水稻光敏核不育基因 *pms1* 的精细定位以及候选基因区段序列分析. [博士学位论文]. 武汉: 华中农业大学图书馆, 2006
 29. 喻辉辉. 水稻杂种育性基因 *f5* 的克隆及高通量分子标记技术的开发与应用. [博士学位论文]. 武汉: 华中农业大学图书馆, 2011
 30. 元生朝, 张自国, 许传桢. 光照诱导湖北光敏感核不育水稻育性转变的敏感期及其发育阶段的探讨. 作物学报, 1988, 14: 7-13
 31. 袁隆平. 水稻的雄性不孕性. 科学通报, 1966, 11: 185-188
 32. 张华丽, 陈晓阳, 黄建中, 鄂志国, 龚俊义, 舒庆尧. 中国两系杂交水稻光温敏核不育基因的鉴定与演化分析. 中国农业科学, 2015, 48: 1-9
 33. 张平博, 丁寄花, 张启发. 光敏核不育水稻对日照长度处理反应的叶片及部位. 分子植物育种, 2010, 08: 641-646
 34. 张自国, 元生朝, 许传桢. 光周期条件对湖北光敏感核不育水稻育性转变的影响. 中国水稻科学, 1987, 1: 137-143
 35. 周广洽, 李训贞, 谭周鎡, 陈良碧. 水稻粳型不育系安农 s_{-1} 不育基因表达条件和育性转换规律的研究. 湖南师范大学自然科学学报, 1990, 13: 365-372
 36. Abad AR, Mehrstens BJ, Mackenzie SA. Specific expression in reproductive tissues and fate of a mitochondrial sterility-associated protein in cytoplasmic male-sterile bean. *Plant Cell*, 1995, 7: 271-285
 37. Abe M, Yoshikawa T, Nosaka M, Sakakibara H, Sato Y, Nagato Y, Itoh J-i. *WAVYLEAF1*, an ortholog of Arabidopsis *HEN1*, regulates shoot development by maintaining microRNA and trans-acting small interfering RNA accumulation in rice. *Plant Physiol*, 2010, 154: 1335-1346
 38. Allen E, Howell MD. miRNAs in the biogenesis of *trans*-acting siRNAs in higher plants. *Semin Cell Dev Biol*, 2010, 21: 798-804
 39. Allen E, Xie Z, Gustafson AM, Carrington JC. MicroRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell*, 2005, 121: 207-221
 40. Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R, Olson Eric N. A micropeptide encoded

- by a putative long noncoding RNA regulates muscle performance. *Cell*, 2015, 160: 595-606
41. Andrews SJ, Rothnagel JA. Emerging evidence for functional peptides encoded by short open reading frames. *Nat Rev Genet*, 2014, 15: 193-204
 42. Angrand PO, Vennin C, Le Bourhis X, Adriaenssens E. The role of long non-coding RNAs in genome formatting and expression. *Front Genet*, 2015, 6: 165
 43. Arikiti S, Xia R, Kakrana A, Huang K, Zhai J, Yan Z, Valdés-López O, Prince S, Musket TA, Nguyen HT, Stacey G, Meyers BC. An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. *Plant Cell*, 2014, 26: 4584-4601
 44. Arikiti S, Zhai J, Meyers BC. Biogenesis and function of rice small RNAs from non-coding RNA precursors. *Curr Opin Plant Biol*, 2013, 16: 170-179
 45. Arribas-Hernández L, Marchais A, Poulsen C, Haase B, Hauptmann J, Benes V, Meister G, Brodersen P. The slicer activity of ARGONAUTE1 is required specifically for the phasing, not production, of *trans*-acting short interfering RNAs in Arabidopsis. *Plant Cell*, 2016, 28: 1563-1580
 46. Axtell MJ. Classification and comparison of small RNAs from plants. *Annu Rev Plant Biol*, 2013, 64: 137-159
 47. Axtell MJ, Jan C, Rajagopalan R, Bartel DP. A two-hit trigger for siRNA biogenesis in plants. *Cell*, 2006, 127: 565-577
 48. Bai J, Zhu X, Wang Q, Zhang J, Chen H, Dong G, Zhu L, Zheng H, Xie Q, Nian J, Chen F, Fu Y, Qian Q, Zuo J. Rice *TUTOUI* encodes a suppressor of cAMP receptor-like protein that is important for actin organization and panicle development. *Plant Physiol*, 2015, 169: 1179-1191
 49. Bailey MF. A cost of restoration of male fertility in a gynodioecious species, *Lobelia siphilitica*. *Evolution*, 2002, 56: 2178-2186
 50. Barkan A, Small I. Pentatricopeptide repeat proteins in plants. *Annu Rev Plant Biol*, 2014, 65: 415-442
 51. Batista Pedro J, Chang Howard Y. Long noncoding RNAs: cellular address codes in development and disease. *Cell*, 2013, 152: 1298-1307
 52. Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC, Giraldez AJ. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J*, 2014, 33: 981-993
 53. Bentolila S, Alfonso AA, Hanson MR. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc Natl Acad Sci USA*, 2002, 99: 10887-10892
 54. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, Gerstein M, Snyder M. Global identification of human transcribed sequences with genome tiling arrays. *Science*, 2004, 306: 2242-2246
 55. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P et al.

- Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 2007, 447: 799-816
56. Brown GG, Formanová N, Jin H, Wargachuk R, Dendy C, Patil P, Laforest M, Zhang J, Cheung WY, Landry BS. The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant J*, 2003, 35: 262-272
 57. Budar F, Touzet P, De Paepe R. The nucleo-mitochondrial conflict in cytoplasmic male sterilities revisited. *Genetica*, 2003, 117: 3-16
 58. Chang F, Wang Y, Wang S, Ma H. Molecular control of microsporogenesis in *Arabidopsis*. *Curr Opin Plant Biol*, 2011, 14: 66-73
 59. Chapman EJ, Carrington JC. Specialization and evolution of endogenous small RNA pathways. *Nat Rev Genet*, 2007, 8: 884-896
 60. Chase CD. Cytoplasmic male sterility: a window to the world of plant mitochondrial–nuclear interactions. *Trends Genet*, 2007, 23: 81-90
 61. Chekanova JA. Long non-coding RNAs and their functions in plants. *Curr Opin Plant Biol*, 2015, 27: 207-216
 62. Chen A, Yu K, Fan L, Li J, Yang C, Huang AJ, Shao Z. Germline genetic variants disturbing the *Let-7/LIN28* double-negative feedback loop alter breast cancer susceptibility. *Plos Genet*, 2011, 7: e1002259
 63. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, Lao KQ, Livak KJ, Guegler KJ. Real-time quantification of microRNAs by stem–loop RT–PCR. *Nucleic Acids Res*, 2005, 33: e179
 64. Chen H, Li Y, Wu S. Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in *Arabidopsis*. *Proc Natl Acad Sci USA*, 2007a, 104: 3318-3323
 65. Chen L, Liu Y-G. Male sterility and fertility restoration in crops. *Annu Rev Plant Biol*, 2014, 65: 579-606
 66. Chen M, Wang C, Bao H, Chen H, Wang Y. Genome-wide identification and characterization of novel lncRNAs in *Populus* under nitrogen deficiency. *Mol Genet Genomics*, 2016, 291: 1663-1680
 67. Chen R, Zhao X, Shao Z, Wei Z, Wang Y, Zhu L, Zhao J, Sun M, He R, He G. Rice UDP-glucose pyrophosphorylase1 is essential for pollen callose deposition and its cosuppression results in a new type of thermosensitive genic male sterility. *Plant Cell*, 2007b, 19: 847-861
 68. Chen X. Small RNAs and their roles in plant development. *Annu Rev Cell Dev Bi*, 2009, 25: 21-44
 69. Chen Z, Muthukrishnan S, Liang GH, Schertz KF, Hart GE. A chloroplast DNA deletion located in RNA polymerase gene *rpoC2* in CMS lines of sorghum. *Mol Gen Genet*, 1993, 236: 251-259
 70. Chi KR. Finding function in mystery transcripts. *Nature*, 2016, 529: 423-425
 71. Couzigou J-M, André O, Guillotin B, Alexandre M, Combiér J-P. Use of microRNA-encoded peptide miPEP172c to stimulate nodulation in soybean. *New Phytol*, 2016, 211: 379-381

72. Cui X, Wise RP, Schnable PS. The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. *Science*, 1996, 272: 1334-1336
73. de Andres-Pablo A, Morillon A, Wery M. LncRNAs, lost in translation or licence to regulate? *Curr Genet*, 2016, 1-5
74. Dewey RE, Timothy DH, Levings CS. A mitochondrial protein associated with cytoplasmic male sterility in the T cytoplasm of maize. *Proc Natl Acad Sci USA*, 1987, 84: 5374-5378
75. Dewey RE, Timothy DH, Levings CS. Chimeric mitochondrial genes expressed in the C male-sterile cytoplasm of maize. *Curr Genet*, 1991, 20: 475-482
76. Di C, Yuan J, Wu Y, Li J, Lin H, Hu L, Zhang T, Qi Y, Gerstein MB, Guo Y, Lu ZJ. Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant J*, 2014, 80: 848-861
77. Ding J, Lu Q, Ouyang Y, Mao H, Zhang P, Yao J, Xu C, Li X, Xiao J, Zhang Q. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc Natl Acad Sci USA*, 2012a, 109: 2654-2659
78. Ding J, Shen J, Mao H, Xie W, Li X, Zhang Q. RNA-directed DNA methylation is involved in regulating photoperiod-sensitive male sterility in rice. *Mol Plant*, 2012b, 5: 1210-1216
79. Dong VN, Subudhi KP, Luong NP, Quang DV, Quy DT, Zheng GH, Wang B, Nguyen TH. Molecular mapping of a rice gene conditioning thermosensitive genic male sterility using AFLP, RFLP and SSR techniques. *Theor Appl Genet*, 2000, 100: 727-734
80. Ducos E, Touzet P, Boutry M. The male sterile G cytoplasm of wild beet displays modified mitochondrial respiratory complexes. *Plant J*, 2001, 26: 171-180
81. Eddy SR. Non-coding RNA genes and the modern RNA world. *Nat Rev Genet*, 2001, 2: 919-929
82. Eitas TK, Dangl JL. NB-LRR proteins: pairs, pieces, perception, partners, and pathways. *Curr Opin Plant Biol*, 2010, 13: 472-477
83. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 2012, 489: 57-74
84. Fan C, Hao Z, Yan J, Li G. Genome-wide identification and functional analysis of lincRNAs acting as miRNA targets or decoys in maize. *BMC Genomics*, 2015, 16: 793
85. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet*, 2014, 15: 7-21
86. Fei Q, Li P, Teng C, Meyers BC. Secondary siRNAs from *Medicago NB-LRRs* modulated via miRNA-target interactions and their abundances. *Plant J*, 2015, 83: 451-465
87. Fei Q, Xia R, Meyers BC. Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell*, 2013, 25: 2400-2415
88. Flórez-Zapata NMV, Reyes-Valdés MH, Martínez O. Long non-coding RNAs are major contributors to transcriptome changes in sunflower meiocytes with different recombination rates. *BMC Genomics*, 2016, 17: 490
89. Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, Garcia JA, Paz-Ares J. Target mimicry provides a new mechanism for

- regulation of microRNA activity. *Nat Genet*, 2007, 39: 1033-1037
90. Frankel R, Scowcroft WR, Whitfeld PR. Chloroplast DNA variation in isonuclear male-sterile lines of *Nicotiana*. *Mol Gen Genet*, 1979, 169: 129-135
 91. Frouin J, Filloux D, Taillebois J, Grenier C, Montes F, de Lamotte F, Verdeil J-L, Courtois B, Ahmadi N. Positional cloning of the rice male sterility gene *ms-IR36*, widely used in the inter-crossing phase of recurrent selection schemes. *Mol Breeding*, 2014, 33: 555-567
 92. Fu X. Non-coding RNA: a new frontier in regulatory biology. *Natl Sci Rev*, 2014, 1: 190-204
 93. Fu Z, Yu J, Cheng X, Zong X, Xu J, Chen M, Li Z, Zhang D, Liang W. The rice basic helix-loop-helix transcription factor TDR INTERACTING PROTEIN2 is a central switch in early anther development. *Plant Cell*, 2014, 26: 1512-1524
 94. Fujii S, Komatsu S, Toriyama K. Retrograde regulation of nuclear gene expression in CW-CMS of rice. *Plant Mol Biol*, 2007, 63: 405-417
 95. Fujii S, Toriyama K. Genome barriers between nuclei and mitochondria exemplified by cytoplasmic male sterility. *Plant Cell Physiol*, 2008, 49: 1484-1494
 96. Fujii S, Toriyama K. Suppressed expression of *RETROGRADE-REGULATED MALE STERILITY* restores pollen fertility in cytoplasmic male sterile rice plants. *Proc Natl Acad Sci USA*, 2009, 106: 9513-9518
 97. Gómez JF, Talle B, Wilson ZA. Anther and pollen development: A conserved developmental pathway. *J Integr Plant Biol*, 2015, 57: 876-891
 98. Gascioli V, Mallory AC, Bartel DP, Vaucheret H. Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing *trans*-acting siRNAs. *Curr Biol*, 2005, 15: 1494-1500
 99. Ge X, Chang F, Ma H. Signaling and transcriptional control of reproductive development in *Arabidopsis*. *Curr Biol*, 2010, 20: R988-R997
 100. Goff LA, Rinn JL. Linking RNA biology to lncRNAs. *Genome Res*, 2015, 25: 1456-1465
 101. Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, 2002, 296: 92-100
 102. Goldberg RB, Beals TP, Sanders PM. Anther development: basic principles and practical applications. *Plant Cell*, 1993, 5: 1217-1229
 103. Gulyas G, Shin Y, Kim H, Lee J-S, Hirata Y. Altered transcript reveals an *orf507* sterility-related gene in chili pepper (*Capsicum annuum* L.). *Plant Mol Biol Rep*, 2010, 28: 605-612
 104. Guo J-X, Liu Y-G. Molecular control of male reproductive development and pollen fertility in rice. *J Integr Plant Biol*, 2012, no-no
 105. Han BW, Wang W, Li C, Weng Z, Zamore PD. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science*, 2015, 348: 817-821
 106. Hans Köhler R, Horn R, Lössl A, Zetsche K. Cytoplasmic male sterility in sunflower is correlated with the co-transcription of a new open reading frame with the *atpA* gene. *Mol*

- Gen Genet*, 1991, 227: 369-376
107. Hanson MR, Bentolila S. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell*, 2004, 16: S154-S169
 108. Hao Z, Fan C, Cheng T, Su Y, Wei Q, Li G. Genome-wide identification, characterization and evolutionary analysis of long intergenic noncoding RNAs in cucumber. *PLoS ONE*, 2015, 10: e0121800
 109. He QY, Yang J, Xu GC, Zhang ZG, Zhang Q. Genetic bases of instability of male sterility and fertility reversibility in photoperiod-sensitive genic male-sterile rice. *Theor Appl Genet*, 1999, 99: 683-693
 110. Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*, 2011, 331: 76-79
 111. Hiei Y, Ohta S, Komari T, Kumashiro T. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J*, 1994, 6: 271-282
 112. Hobo T, Suwabe K, Aya K, Suzuki G, Yano K, Ishimizu T, Fujita M, Kikuchi S, Hamada K, Miyano M, Fujioka T, Kaneko F, Kazama T, Mizuta Y, Takahashi H, Shiono K, Nakazono M, Tsutsumi N, Nagamura Y, Kurata N et al. Various spatiotemporal expression profiles of anther-expressed genes in rice. *Plant Cell Physiol*, 2008, 49: 1417-1428
 113. Honys D, Twell D. Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biol*, 2004, 5: 1-13
 114. Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, Givan SA, Kasschau KD, Carrington JC. Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell*, 2007, 19: 926-942
 115. Hu J, Wang K, Huang W, Liu G, Gao Y, Wang J, Huang Q, Ji Y, Qin X, Wan L, Zhu R, Li S, Yang D, Zhu Y. The rice pentatricopeptide repeat protein RF5 restores fertility in Hong-Lian cytoplasmic male-sterile lines via a complex with the glycine-rich protein GRP162. *Plant Cell*, 2012, 24: 109-122
 116. Hu Z, Jiang Q, Ni Z, Chen R, Xu S, Zhang H. Analyses of a *Glycine max* degradome library identify microRNA targets and microRNAs that trigger secondary siRNA biogenesis. *J Integr Plant Biol*, 2013, 55: 160-176
 117. Huang Q, Mao Z, Li S, Hu J, Zhu Y. A non-radioactive method for small RNA detection by northern blotting. *Rice*, 2014, 7: 1-7
 118. Huang T, Wang Z, Hu Y, Shi S, Peng T, Chu X, Shi J, Xiang Z, Liu D. Genetic analysis and primary mapping of *pms4*, a photoperiod-sensitive genic male sterility gene in rice (*oryza sativa*). *Rice Science*, 2008, 15: 153-156
 119. Ietswaart R, Wu Z, Dean C. Flowering time control: another window to the connection between antisense RNA and chromatin. *Trends Genet*, 2012, 28: 445-453
 120. Igarashi K, Kazama T, Motomura K, Toriyama K. Whole genomic sequencing of RT98 mitochondria derived from *Oryza rufipogon* and northern blot analysis to uncover a cytoplasmic male sterility-associated gene. *Plant Cell Physiol*, 2013, 54: 237-243

121. Itabashi E, Iwata N, Fujii S, Kazama T, Toriyama K. The fertility restorer gene, *Rf2*, for Lead Rice-type cytoplasmic male sterility of rice encodes a mitochondrial glycine-rich protein. *Plant J*, 2011, 65: 359-367
122. Itabashi E, Kazama T, Toriyama K. Characterization of cytoplasmic male sterility of rice with Lead Rice cytoplasm in comparison with that with Chinsurah Boro II cytoplasm. *Plant Cell Rep*, 2009, 28: 233-239
123. Iwabuchi M, Koizuka N, Fujimoto H, Sakai T, Imamura J. Identification and expression of the kosenra radish (*Raphanus sativus* cv. Kosenra) homologue of the ogura radish CMS-associated gene, *orf138*. *Plant Mol Biol*, 1999, 39: 183-188
124. Jefferson RA, Kavanagh TA, Bevan MW. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J*, 1987, 6: 3901-3907
125. Jeon J-S, Lee S, Jung K-H, Jun S-H, Jeong D-H, Lee J, Kim C, Jang S, Lee S, Yang K, Nam J, An K, Han M-J, Sung R-J, Choi H-S, Yu J-H, Choi J-H, Cho S-Y, Cha S-S, Kim S-I et al. T-DNA insertional mutagenesis for functional genomics in rice. *Plant J*, 2000, 22: 561-570
126. Ji J, Huang W, Yin C, Gong Z. Mitochondrial cytochrome c oxidase and F1Fo-ATPase dysfunction in peppers (*Capsicum annuum* L.) with cytoplasmic male sterility and its association with *orf507* and *Ψatp6-2* genes. *Int J Mol Sci*, 2013, 14: 1050-1068
127. Jia HJ, Zhang SD, Li YC, Qu PX, Wang WS, Chamarek V, Nguyen TH, Wang B. Molecular mapping of the reverse thermo-sensitive genic male-sterile gene (*rtms1*) in rice. *Theor Appl Genet*, 2001, 103: 607-612
128. Jiang D, Lu S, Zhou H, Wu X, Zhuang C, Liu Y, Mei M. Mapping of the rice (*Oryza sativa* L.) thermo-sensitive genic male sterile gene *tms5* with EST and SSR markers. *Chinese Sci Bull*, 2006, 51: 417-420
129. Jin J, Liu J, Wang H, Wong L, Chua N-H. PLncDB: plant long non-coding RNA database. *Bioinformatics*, 2013, 29: 1068-1071
130. Jing B, Heng S, Tong D, Wan Z, Fu T, Tu J, Ma C, Yi B, Wen J, Shen J. A male sterility-associated cytotoxic protein ORF288 in *Brassica juncea* causes aborted pollen development. *J Exp Bot*, 2012, 63: 1285-1295
131. Johnson C, Kasprzewska A, Tennessen K, Fernandes J, Nan G-L, Walbot V, Sundaresan V, Vance V, Bowman LH. Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. *Genome Res*, 2009, 19: 1429-1440
132. Jones DF. The interrelation of plasmagenes and chromogenes in pollen production in maize. *Genetics*, 1950, 35: 507-512
133. Jones JW. Hybrid vigor in rice. *Agron J*, 1926, 18: 423-428
134. Källman T, Chen J, Gyllenstrand N, Lagercrantz U. A significant fraction of 21-nucleotide small RNA originates from phased degradation of resistance genes in several perennial species. *Plant Physiol*, 2013, 162: 741-754
135. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet*, 2013, 9: e1003470
136. Kaul MLH. Male sterility in higher plants. Berlin: Springer-Verlag, 1988.

137. Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol*, 2012, 13: R107
138. Khush GS. Green revolution: the way forward. *Nat Rev Genet*, 2001, 2: 815-822
139. Kim DH, Kang JG, Kim B-D. Isolation and characterization of the cytoplasmic male sterility-associated *orf456* gene of chili pepper (*Capsicum annuum* L.). *Plant Mol Biol*, 2007, 63: 519-532
140. Kim E-D, Sung S. Long noncoding RNA: unveiling hidden layer of gene regulatory networks. *Trends Plant Sci*, 2012, 17: 16-21
141. Klattenhoff C, Theurkauf W. Biogenesis and germline functions of piRNAs. *Development*, 2008, 135: 3-9
142. Klein RR, Klein PE, Mullet JE, Minx P, Rooney WL, Schertz KF. Fertility restorer locus *Rf1* of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. *Theor Appl Genet*, 2005, 111: 994-1012
143. Ko S, Li M, Sun-Ben Ku M, Ho Y, Lin Y, Chuang M, Hsing H, Lien Y, Yang H, Chang H, Chan M. The bHLH142 transcription factor coordinates with TDR1 to modulate the expression of *EAT1* and regulate pollen development in rice. *Plant Cell*, 2014, 26: 2486-2504
144. Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, Sakai T, Kawasaki S, Imamura J. Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosena radish. *Plant J*, 2003, 34: 407-415
145. Komiya R, Ohyanagi H, Niihama M, Watanabe T, Nakano M, Kurata N, Nonomura K-I. Rice germline-specific Argonaute MEL1 protein binds to phasiRNAs generated from more than 700 lincRNAs. *Plant J*, 2014, 78: 385-397
146. Komori T, Ohta S, Murai N, Takakura Y, Kuraya Y, Suzuki S, Hiei Y, Imaseki H, Nitta N. Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). *Plant J*, 2004, 37: 315-325
147. Kotera E, Tasaka M, Shikanai T. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature*, 2005, 433: 326-330
148. Kruft V, Eubel H, Jansch L, Werhahn W, Braun H-P. Proteomic approach to identify novel mitochondrial proteins in Arabidopsis. *Plant Physiol*, 2001, 127: 1694-1710
149. Kubo T, Newton KJ. Angiosperm mitochondrial genomes and mutations. *Mitochondrion*, 2008, 8: 5-14
150. L'Homme Y, Stahl JR, Li X-Q, Hameed A, Brown GG. *Brassica nap* cytoplasmic male sterility is associated with expression of a mtDNA region containing a chimeric gene similar to the *pol* CMS-associated *orf224* gene. *Curr Genet*, 1997, 31: 325-335
151. Landgren M, Zetterstrand M, Sundberg E, Glimelius K. Alloplasmic male-sterile *Brassica* lines containing *B. tournefortii* mitochondria express an ORF 3' of the *atp6* gene and a 32 kDa protein. *Plant Mol Biol*, 1996, 32: 879-890
152. Laressergues D, Couzigou J-M, Clemente HS, Martinez Y, Dunand C, Becard G, Combier J-P. Primary transcripts of microRNAs encode regulatory peptides. *Nature*, 2015,

- 520: 90-93
153. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*, 2010, 11: 204-220
 154. 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
 155. Lee S, Jung K-H, An G, Chung Y-Y. Isolation and characterization of a rice cysteine protease gene, *OsCPI*, using T-DNA gene-trap system. *Plant Mol Biol*, 2004, 54: 755-765
 156. Li F, Orban R, Baker B. SoMART: a web server for plant miRNA, tasiRNA and target gene analysis. *Plant J*, 2012, 70: 891-901
 157. Li J, Reichel M, Millar AA. Determinants beyond both complementarity and cleavage govern microR159 efficacy in *Arabidopsis*. *Plos Genet*, 2014a, 10: e1004232
 158. Li L, Eichten S, Shimizu R, Petsch K, Yeh C-T, Wu W, Chetoor A, Givan S, Cole R, Fowler J, Evans M, Scanlon M, Yu J, Schnable P, Timmermans M, Springer N, Muehlbauer G. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol*, 2014b, 15: R40
 159. Li N, Zhang D, Liu H, Yin C, Li X, Liang W, Yuan Z, Xu B, Chu H, Wang J, Wen T, Huang H, Luo D, Ma H, Zhang D. The rice *Tapetum Degeneration Retardation* gene is required for tapetum degradation and anther development. *Plant Cell*, 2006, 18: 2999-3014
 160. Li P, Tao Z, Dean C. Phenotypic evolution through variation in splicing of the noncoding RNA *COOLAIR*. *Genes Dev*, 2015,
 161. Li Q, Zhang D, Chen M, Liang W, Wei J, Qi Y, Yuan Z. Development of *japonica* photo-sensitive genic male sterile rice lines by editing *Carbon Starved Anther* using CRISPR/Cas9. *J Genet Genomics*, 2016, 43: 415-419
 162. Li X, Lu Q, Wang F, Xu C, Zhang Q. 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
 163. Lin Y, Zhang Q. Optimising the tissue culture conditions for high efficiency transformation of indica rice. *Plant Cell Rep*, 2005, 23: 540-547
 164. Liu B, Chen Z, Song X, Liu C, Cui X, Zhao X, Fang J, Xu W, Zhang H, Wang X, Chu C, Deng X, Xue Y, Cao X. *Oryza sativa Dicer-like4* reveals a key role for small interfering RNA silencing in plant development. *Plant Cell*, 2007, 19: 2705-2718
 165. Liu F, Cui X, Horner HT, Weiner H, Schnable PS. Mitochondrial aldehyde dehydrogenase activity is required for male fertility in maize. *Plant Cell*, 2001a, 13: 1063-1078
 166. Liu F, Marquardt S, Lister C, Swiezewski S, Dean C. Targeted 3' processing of antisense transcripts triggers *Arabidopsis FLC* chromatin silencing. *Science*, 2010a, 327: 94-97
 167. Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua N-H. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in

- Arabidopsis*. *Plant Cell*, 2012, 24: 4333-4345
168. Liu N, Shan Y, Wang F, Xu C, Peng K, Li X, Zhang Q. Identification of an 85-kb DNA fragment containing *pms1*, a locus for photoperiod-sensitive genic male sterility in rice. *Mol Genet Genomics*, 2001b, 266: 271-275
169. Liu Q, Wang F, Axtell MJ. Analysis of complementarity requirements for plant microRNA targeting using a *Nicotiana benthamiana* quantitative transient assay. *Plant Cell*, 2014, 26: 741-753
170. Liu X, Hao L, Li D, Zhu L, Hu S. Long non-coding RNAs and their biological roles in plants. *Genomics Proteomics Bioinformatics*, 2015, 13: 137-147
171. 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*, 2010b, 53: 119-124
172. Liu Y, Li J. Chloroplast DNA and cytoplasmic male sterility. *J Genet Genomics*, 1983, 10: 114-122
173. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods*, 2001, 25: 402-408
174. Llave C, Xie Z, Kasschau KD, Carrington JC. Cleavage of scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science*, 2002, 297: 2053-2056
175. Lu Q, Li X, Guo D, Xu C, 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
176. Luo D, Xu H, Liu Z, Guo J, Li H, Chen L, Fang C, Zhang Q, Bai M, Yao N, Wu H, Wu H, Ji C, Zheng H, Chen Y, Ye S, Li X, Zhao X, Li R, Liu Y-G. A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. *Nat Genet*, 2013, 45: 573-577
177. Luo Q, Li Y, Shen Y, Cheng Z. Ten years of gene discovery for meiotic event control in rice. *J Genet Genomics*, 2014, 41: 125-137
178. Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, Bruyère C, Caboche M, Debast C, Gualberto J, Hoffmann B, Lecharny A, Le Ret M, Martin-Magniette M-L, Mireau H, Peeters N, Renou J-P, Szurek B, Taconnat L, Small I. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, 2004, 16: 2089-2103
179. Lv Y, Liang Z, Ge M, Qi W, Zhang T, Lin F, Peng Z, Zhao H. Genome-wide identification and functional prediction of nitrogen-responsive intergenic and intronic long non-coding RNAs in maize (*Zea mays* L.). *BMC Genomics*, 2016, 17: 350
180. Ma H. Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Annu Rev Plant Biol*, 2005, 56: 393-434
181. Ma X, Zhu Q, Chen Y, Liu Y-G. CRISPR/Cas9 platforms for genome editing in plants: developments and applications. *Mol Plant*, 2016, 9: 961-974
182. Mallory AC, Reinhart BJ, Jones - Rhoades MW, Tang G, Zamore PD, Barton MK, Bartel DP. MicroRNA control of *PHABULOSA* in leaf development: importance of pairing to the microRNA 5' region. *EMBO J*, 2004, 23: 3356-3364
183. Matsuhira H, Kagami H, Kurata M, Kitazaki K, Matsunaga M, Hamaguchi Y, Hagihara E,

- Ueda M, Harada M, Muramatsu A, Yui-Kurino R, Taguchi K, Tamagake H, Mikami T, Kubo T. Unusual and typical features of a novel *Restorer-of-fertility* gene of sugar beet (*Beta vulgaris* L.). *Genetics*, 2012, 192: 1347-1358
184. McGinnis K, Chandler V, Cone K, Kaeppler H, Kaeppler S, Kerschen A, Pikaard C, Richards E, Sidorenko L, Smith T, Springer N, Wulan T. Transgene-induced RNA interference as a tool for plant functional genomics. *Methods Enzymol*, 2005, 392: 1-24
185. McHugh CA, Chen C-K, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A, Sweredoski MJ, Shishkin AA, Su J, Lander ES, Hess S, Plath K, Guttman M. The *Xist* lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*, 2015, 521: 232-236
186. 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*, 1999a, 42: 316-322
187. Mei M, Dai X, 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.*, 1999b, 39: 1711-1715
188. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet*, 2009, 10: 155-159
189. Meyers BC, Kaushik S, Nandety RS. Evolving disease resistance genes. *Curr Opin Plant Biol*, 2005, 8: 129-134
190. Mishra PJ, Humeniuk R, Mishra PJ, Longo-Sorbello GSA, Banerjee D, Bertino JR. A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance. *Proc Natl Acad Sci USA*, 2007, 104: 13513-13518
191. Mohn F, Handler D, Brennecke J. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science*, 2015, 348: 812-817
192. Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, Alexander AL, Chapman EJ, Fahlgren N, Allen E, Carrington JC. Specificity of ARGONAUTE7-miR390 interaction and dual functionality in *TAS3* trans-acting siRNA formation. *Cell*, 2008, 133: 128-141
193. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res*, 1980, 8: 4321-4325
194. Nakajima Y, Yamamoto T, Muranaka T, Oeda K. A novel orfB-related gene of carrot mitochondrial genomes that is associated with homeotic cytoplasmic male sterility (CMS). *Plant Mol Biol*, 2001, 46: 99-107
195. Nguyen TD, Moon S, Nguyen VNT, Gho Y, Chandran AKN, Soh M-S, Song JT, An G, Oh SA, Park SK, Jung K-H. Genome-wide identification and analysis of rice genes preferentially expressed in pollen at an early developmental stage. *Plant Mol Biol*, 2016, 1-18
196. Niu N, Liang W, Yang X, Jin W, Wilson ZA, Hu J, Zhang D. EAT1 promotes tapetal cell death by regulating aspartic proteases during male reproductive development in rice. *Nat Commun*, 2013, 4: 1445
197. Nonomura K-I, Morohoshi A, Nakano M, Eiguchi M, Miyao A, Hirochika H, Kurata N. 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
198. Okazaki M, Kazama T, Murata H, Motomura K, Toriyama K. Whole mitochondrial genome sequencing and transcriptional analysis to uncover an RT102-type cytoplasmic male sterility-associated candidate gene derived from *Oryza rufipogon*. *Plant Cell Physiol*, 2013, 54: 1560-1568
199. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H, Yamanaka I, Kiyosawa H, Yagi K, Tomaru Y, Hasegawa Y, Nogami A, Schönbach C, Gojobori T, Baldarelli R, Hill DP et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, 2002, 420: 563-573
200. Palmer JD. Comparative organization of chloroplast genomes. *Annu Rev Genet*, 1985, 19: 325-354
201. Pan Y, Li Q, Wang Z, Wang Y, Ma R, Zhu L, He G, Chen R. Genes associated with thermosensitive genic male sterility in rice identified by comparative expression profiling. *BMC Genomics*, 2014, 15: 1-17
202. Pang KC, Frith MC, Mattick JS. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet*, 2006, 22: 1-5
203. Park JY, Lee Y-P, Lee J, Choi B-S, Kim S, Yang T-J. Complete mitochondrial genome sequence and identification of a candidate gene responsible for cytoplasmic male sterility in radish (*Raphanus sativus* L.) containing DCGMS cytoplasm. *Theor Appl Genet*, 2013, 126: 1763-1774
204. Paytuví Gallart A, Hermoso Pulido A, Anzar Martínez de Lagrán I, Sanseverino W, Aiese Cigliano R. GREENC: a Wiki-based database of plant lncRNAs. *Nucleic Acids Res*, 2015, 44: D1161-D1166
205. Peng H, Zhang Z, Wu B, Chen X, Zhang G, Zhang Z, Wan B, Lu Y. 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
206. Peng HF, Chen XH, Lu YP, Peng YF, Wan BH, Chen ND, Wu B, Xin SP, Zhang GQ. 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
207. Peng K, Zhang H, Zhang Q. BAC library constructed to the rice cultivar “Minghui 63” for cloning gene of agronomic importance. *Acta Botanica Sinica* 1998, 40: 1108-1114
208. Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS. *SGS3* and *SGS2/SDE1/RDR6* are required for juvenile development and the production of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev*, 2004, 18: 2368-2379
209. Petsch K, Manzotti PS, Tam OH, Meeley R, Hammell M, Consonni G, Timmermans MCP. Novel DICER-LIKE1 siRNAs bypass the requirement for DICER-LIKE4 in maize development. *Plant Cell*, 2015, 27: 2163-2177
210. Piffanelli P, Droc G, Mieulet D, Lanau N, Bès M, Bourgeois E, Rouvière C, Gavory F, Cruaud C, Ghesquière A, Guiderdoni E. Large-scale characterization of Tos17 insertion sites in a rice T-DNA mutant library. *Plant Mol Biol*, 2007, 65: 587-601

211. Qi Y, Liu Q, Zhang L, Mao B, Yan D, Jin Q, He Z. Fine mapping and candidate gene analysis of the novel thermo-sensitive genic male sterility *tms9-1* gene in rice. *Theor Appl Genet*, 2014, 127: 1173-1182
212. Qin X, Huang Q, Xiao H, Zhang Q, Ni C, Xu Y, Liu G, Yang D, Zhu Y, Hu J. The rice DUF1620-containing and WD40-like repeat protein is required for the assembly of the restoration of fertility complex. *New Phytol*, 2016, 210: 934-945
213. Qiu D, Xiao J, Ding X, Xiong M, Cai M, Cao Y, Li X, Xu C, Wang S. OsWRKY13 mediates rice disease resistance by regulating defense-related genes in salicylate- and jasmonate-dependent signaling. *Mol Plant Microbe In*, 2007, 20: 492-499
214. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev*, 2006, 20: 3407-3425
215. Rajeswaran R, Aregger M, Zvereva AS, Borah BK, Gubaeva EG, Pooggin MM. Sequencing of RDR6-dependent double-stranded RNAs reveals novel features of plant siRNA biogenesis. *Nucleic Acids Res*, 2012,
216. Rajeswaran R, Pooggin MM. RDR6-mediated synthesis of complementary RNA is terminated by miRNA stably bound to template RNA. *Nucleic Acids Res*, 2012, 40: 594-599
217. Reddy KOU, Siddiq AE, Sarma PN, Ali J, Hussain JA, Nimmakayala P, Ramasamy P, Pammi S, Reddy SA. Genetic analysis of temperature-sensitive male sterility in rice. *Theor Appl Genet*, 2000, 100: 794-801
218. Rock CD. Trans-acting small interfering RNA4: key to nutraceutical synthesis in grape development? *Trends Plant Sci*, 2013, 18: 601-610
219. Ruiz-Orera J, Messeguer X, Subirana JA, Alba MM. Long non-coding RNAs as a source of new peptides. *Elife*, 2014, 3: e03523
220. Rutley N, Twell D. A decade of pollen transcriptomics. *Plant Reprod*, 2015, 28: 73-89
221. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 1995, 270: 467-470
222. Schmitz-Linneweber C, Small I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci*, 2008, 13: 663-670
223. Schnable PS, Wise RP. The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends Plant Sci*, 1998, 3: 175-180
224. Scott RJ, Spielman M, Dickinson HG. Stamen structure and function. *Plant Cell*, 2004, 16: S46-S60
225. Shen C, Wang X, Liu Y, Li Q, Yang Z, Yan N, Zou T, Yin P. Specific RNA recognition by designer pentatricopeptide repeat protein. *Mol Plant*, 2015, 8: 667-670
226. Shen C, Zhang D, Guan Z, Liu Y, Yang Z, Yang Y, Wang X, Wang Q, Zhang Q, Fan S, Zou T, Yin P. Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. *Nat Commun*, 2016, 7: 11285
227. Shen J, Xie K, Xiong L. Global expression profiling of rice microRNAs by one-tube stem-loop reverse transcription quantitative PCR revealed important roles of microRNAs in abiotic stress responses. *Mol Genet Genomics*, 2010, 284: 477-488
228. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotech*, 2008, 26: 1135-1145

229. Sheng Z, Wei X, Shao G, Chen M, Song J, Tang S, Luo J, Hu Y, Hu P, Chen L. Genetic analysis and fine mapping of *tms9*, a novel thermosensitive genic male-sterile gene in rice (*Oryza sativa* L.). *Plant Breeding*, 2013, 132: 159-164
230. Shi J, Cui M, Yang L, Kim Y-J, Zhang D. Genetic and biochemical mechanisms of pollen wall development. *Trends Plant Sci*, 2015, 20: 741-753
231. Shi Y, Zhao S, Yao J. Premature tapetum degeneration: a major cause of abortive pollen development in photoperiod sensitive genic male sterility in rice. *J Integr Plant Biol*, 2009, 51: 774-781
232. Shivaprasad PV, Chen HM, Patel K, Bond DM, Santos BA, Baulcombe DC. A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant Cell*, 2012, 24: 859-874
233. Singh M, Brown GG. Suppression of cytoplasmic male sterility by nuclear genes alters expression of a novel mitochondrial gene region. *Plant Cell*, 1991, 3: 1349-1362
234. Small ID, Peeters N. The PPR motif – a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci*, 2000, 25: 45-47
235. Smyth DR, Bowman JL, Meyerowitz EM. Early flower development in *Arabidopsis*. *Plant Cell*, 1990, 2: 755-767
236. Song J, Hedgcoth C. A chimeric gene (*orf256*) is expressed as protein only in cytoplasmic male-sterile lines of wheat. *Plant Mol Biol*, 1994, 26: 535-539
237. Song X, Li P, Zhai J, Zhou M, Ma L, Liu B, Jeong D-H, Nakano M, Cao S, Liu C, Chu C, Wang X-J, Green PJ, Meyers BC, Cao X. Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. *Plant J*, 2012a, 69: 462-474
238. Song X, Wang D, Ma L, Chen Z, Li P, Cui X, Liu C, Cao S, Chu C, Tao Y, Cao X. Rice RNA-dependent RNA polymerase 6 acts in small RNA biogenesis and spikelet development. *Plant J*, 2012b, 71: 378-389
239. St. Laurent G, Wahlestedt C, Kapranov P. The Landscape of long noncoding RNA classification. *Trends Genet*, 2015, 1-13
240. 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
241. Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C. R-loop stabilization represses antisense transcription at the *Arabidopsis FLC* locus. *Science*, 2013, 340: 619-621
242. Swiezewski S, Liu F, Magusin A, Dean C. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature*, 2009, 462: 799-802
243. Szcześniak MW, Rosikiewicz W, Makałowska I. CANTATAdb: a collection of plant long non-coding RNAs. *Plant Cell Physiol*, 2015, 57: e8
244. Talmor-Neiman M, Stav R, Klipcan L, Buxdorf K, Baulcombe DC, Arazi T. Identification of *trans*-acting siRNAs in moss and an RNA-dependent RNA polymerase required for their biogenesis. *Plant J*, 2006, 48: 511-521
245. Tang HV, Pring DR, Shaw LC, Salazar RA, Muza FR, Yan B, Schertz KF. Transcript processing internal to a mitochondrial open reading frame is correlated with fertility

- restoration in male-sterile sorghum. *Plant J*, 1996, 10: 123-133
246. Tang S, Wang Y, Li Z, Gui Y, Xiao B, Xie J, Zhu Q-H, Fan L. Identification of wounding and topping responsive small RNAs in tobacco (*Nicotiana tabacum*). *BMC Plant Biol*, 2012, 12: 1-16
247. Tano K, Akimitsu N. Long non-coding RNAs in cancer progression. *Front Genet*, 2012, 3: 219
248. Thangasamy S, Guo C-L, Chuang M-H, Lai M-H, Chen J, Jauh G-Y. Rice SIZ1, a SUMO E3 ligase, controls spikelet fertility through regulation of anther dehiscence. *New Phytol*, 2011, 189: 869-882
249. The International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, 2010, 463: 763-768
250. Uyttewaal M, Arnal N, Quadrado M, Martin-Canadell A, Vrielynck N, Hiard S, Gherbi H, Bendahmane A, Budar F, Mireau H. Characterization of *Raphanus sativus* pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. *Plant Cell*, 2008, 20: 3331-3345
251. Varkonyi-Gasic E, Wu R, Wood M, Walton E, Hellens R. Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods*, 2007, 3: 12
252. Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gascioli V, Mallory AC, Hilbert J-L, Bartel DP, Crété P. Endogenous *trans*-acting sirnas regulate the accumulation of *Arabidopsis* mRNAs. *Mol Cell*, 2004, 16: 69-79
253. Wang B, Xu WW, Wang JZ, Wu W, Zheng HG, Yang ZY, Ray JD, Nguyen HT. 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
254. Wang H, Chung PJ, Liu J, Jang I-C, Kean M, Xu J, Chua N-H. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res*, 2014a, 24: 444-453
255. Wang H, Niu Q-W, Wu H-W, Liu J, Ye J, Yu N, Chua N-H. Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lncRNAs associated with agriculture traits. *Plant J*, 2015a, 84: 404-416
256. Wang J, Gao X, Li L, Shi X, Zhang J, Shi Z. Overexpression of Osta-siR2141 caused abnormal polarity establishment and retarded growth in rice. *J Exp Bot*, 2010a, 61: 1885-1895
257. Wang K, Gao F, Ji Y, Liu Y, Dan Z, Yang P, Zhu Y, Li S. ORFH79 impairs mitochondrial function via interaction with a subunit of electron transport chain complex III in Honglian cytoplasmic male sterile rice. *New Phytol*, 2013, 198: 408-418
258. Wang Kevin C, Chang Howard Y. Molecular mechanisms of long noncoding RNAs. *Mol Cell*, 2011, 43: 904-914
259. Wang L, Xie W, Chen Y, Tang W, Yang J, Ye R, Liu L, Lin Y, Xu C, Xiao J, Zhang Q. A dynamic gene expression atlas covering the entire life cycle of rice. *Plant J*, 2010b, 61: 752-766
260. Wang M, Yuan D, Tu L, Gao W, He Y, Hu H, Wang P, Liu N, Lindsey K, Zhang X. Long

- noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp.). *New Phytol*, 2015b, 207: 1181-1197
261. Wang T, Liu M, Zhao M, Chen R, Zhang W. Identification and characterization of long non-coding RNAs involved in osmotic and salt stress in *Medicago truncatula* using genome-wide high-throughput sequencing. *BMC Plant Biol*, 2015c, 15: 131
262. Wang Y, Itaya A, Zhong X, Wu Y, Zhang J, van der Knaap E, Olmstead R, Qi Y, Ding B. Function and evolution of a microRNA that regulates a Ca²⁺-ATPase and triggers the formation of phased small interfering RNAs in tomato reproductive growth. *Plant Cell*, 2011, 23: 3185-3203
263. Wang YG, Xing QH, Deng QY, Liang FS, Yuan LP, Weng ML, Wang B. Fine mapping of the rice thermo-sensitive genic male-sterile gene *tms5*. *Theor Appl Genet*, 2003, 107: 917-921
264. Wang Z-W, Wu Z, Raitskin O, Sun Q, Dean C. Antisense-mediated FLC transcriptional repression requires the P-TEFb transcription elongation factor. *Proc Natl Acad Sci USA*, 2014b, 111: 7468-7473
265. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 2009, 10: 57-63
266. Wang Z, Schwacke R, Kunze R. DNA damage-induced transcription of transposable elements and long non-coding RNAs in *Arabidopsis* is rare and ATM-dependent. *Mol Plant*, 2016, 9: 1142-1155
267. Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, Su D, Chen Y, Guo J, Luo D, Long Y, Zhong Y, Liu Y-G. 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
268. Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol*, 2011, 21: 354-361
269. Weissgerber TL, Milic NM, Winham SJ, Garovic VD. Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS Biol*, 2015, 13: e1002128
270. Williams L, Carles CC, Osmont KS, Fletcher JC. A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the *Arabidopsis* *ARF2*, *ARF3*, and *ARF4* genes. *Proc Natl Acad Sci USA*, 2005, 102: 9703-9708
271. Wilson ZA, Zhang D-B. From *Arabidopsis* to rice: pathways in pollen development. *J Exp Bot*, 2009, 60: 1479-1492
272. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev*, 2009, 23: 1494-1504
273. Woodson JD, Chory J. Coordination of gene expression between organellar and nuclear genomes. *Nat Rev Genet*, 2008, 9: 383-395
274. Wu C, Li X, Yuan W, Chen G, Kilian A, Li J, Xu C, Li X, Zhou D-X, Wang S, Zhang Q. Development of enhancer trap lines for functional analysis of the rice genome. *Plant J*, 2003, 35: 418-427
275. Wu H, Wang Z, Wang M, Wang X. Wide-spread long non-coding RNAs (lncRNAs) as endogenous target mimics (eTMs) for microRNAs in plants. *Plant Physiol*, 2013, 161:

- 1875-1884
276. Xia R, Meyers BC, Liu Z, Beers EP, Ye S, Liu Z. MicroRNA superfamilies descended from miR390 and their roles in secondary small interfering RNA biogenesis in eudicots. *Plant Cell*, 2013, 25: 1555-1572
277. Xia R, Ye S, Liu Z, Meyers B, Liu Z. Novel and recently evolved miRNA clusters regulate expansive F-box gene networks through phasiRNAs in wild diploid strawberry. *Plant Physiol*, 2015, 169: 594-610
278. Xia R, Zhu H, An Y-q, Beers EP, Liu Z. Apple miRNAs and tasiRNAs with novel regulatory networks. *Genome Biol*, 2012, 13: R47
279. Xie J, Fan L. Nicotine biosynthesis is regulated by two more layers: small and long non-protein-coding RNAs. *Plant Signal Behav*, 2016, 11: e1184811
280. Xie Z, Allen E, Wilken A, Carrington JC. DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*, 2005, 102: 12984-12989
281. Xin M, Wang Y, Yao Y, Song N, Hu Z, Qin D, Xie C, Peng H, Ni Z, Sun Q. Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biol*, 2011, 11: 61
282. Xu J, Wang B, Wu Y, Du P, Wang J, Wang M, Yi C, Gu M, Liang G. 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
283. Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, Zhou H, Yu S, Xu C, Li X, Zhang Q. Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nat Genet*, 2008, 40: 761-767
284. Yamaguchi Y, Ikeda R, Hirasawa H, Minami M, Ujihara A. Linkage analysis of thermosensitive genic male sterility gene, *tms-2* in rice (*Oryza sativa* L.). *Breeding Sci*, 1997, 47: 371-373
285. Yamamoto MP, Kubo T, Mikami T. The 5' -leader sequence of sugar beet mitochondrial *atp6* encodes a novel polypeptide that is characteristic of Owen cytoplasmic male sterility. *Mol Genet Genomics*, 2005, 273: 342-349
286. Yamamoto MP, Shinada H, Onodera Y, Komaki C, Mikami T, Kubo T. A male sterility-associated mitochondrial protein in wild beets causes pollen disruption in transgenic plants. *Plant J*, 2008, 54: 1027-1036
287. Yang Y, Wen L, Zhu H. Unveiling the hidden function of long non-coding RNA by identifying its major partner-protein. *Cell Biosci*, 2015, 5: 59
288. Ye C, Xu H, Shen E, Liu Y, Wang Y, Shen Y, Qiu J, Zhu Q, Fan L. Genome-wide identification of non-coding RNAs interacted with microRNAs in soybean. *Front Plant Sci*, 2014, 5:
289. Yi X, Zhang Z, Ling Y, Xu W, Su Z. PNRD: a plant non-coding RNA database. *Nucleic Acids Res*, 2015, 43: D982-D989
290. Yin P, Li Q, Yan C, Liu Y, Liu J, Yu F, Wang Z, Long J, He J, Wang H-W, Wang J, Zhu J-K, Shi Y, Yan N. Structural basis for the modular recognition of single-stranded RNA

- by PPR proteins. *Nature*, 2013, 504: 168-171
291. Yoshikawa M, Iki T, Tsutsui Y, Miyashita K, Poethig RS, Habu Y, Ishikawa M. 3' fragment of miR173-programmed RISC-cleaved RNA is protected from degradation in a complex with RISC and SGS3. *Proc Natl Acad Sci USA*, 2013, 110: 4117-4122
292. Yoshikawa M, Peragine A, Park MY, Poethig RS. A pathway for the biogenesis of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev*, 2005, 19: 2164-2175
293. Yu J, Fan Y, Liu N, Shan Y, Li X, Zhang Q. Rapid genome evolution in *Pms1* region of rice revealed by comparative sequence analysis. *Chinese Sci Bull*, 2007, 52: 912-921
294. Yu J, Hu S, Wang J, Wong GK-S, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, 2002, 296: 79-92
295. Yuan B, Shen X, Li X, Xu C, Wang S. Mitogen-activated protein kinase OsMPK6 negatively regulates rice disease resistance to bacterial pathogens. *Planta*, 2007, 226: 953-960
296. Yuan J, Zhang Y, Dong J, Sun Y, Lim BL, Liu D, Lu ZJ. Systematic characterization of novel lncRNAs responding to phosphate starvation in *Arabidopsis thaliana*. *BMC Genomics*, 2016, 17: 1-16
297. Yuan S, Zhang Z, Xu C. Studies on the critical stage of fertility change induced by light and its phase development in HPGMR. *Acta Agron Sin*, 1988, 14: 7-13
298. Zabala G, Gabay-Laughnan S, Laughnan JR. The nuclear gene *Rf3* affects the expression of the mitochondrial chimeric sequence R implicated in S-type male sterility in maize. *Genetics*, 1997, 147: 847-860
299. Zhai J, Arikait S, Simon SA, Kingham BF, Meyers BC. Rapid construction of parallel analysis of RNA end (PARE) libraries for Illumina sequencing. *Methods*, 2014, 67: 84-90
300. Zhai J, Jeong D-H, De Paoli E, Park S, Rosen BD, Li Y, González AJ, Yan Z, Kitto SL, Grusak MA, Jackson SA, Stacey G, Cook DR, Green PJ, Sherrier DJ, Meyers BC. MicroRNAs as master regulators of the plant *NB-LRR* defense gene family via the production of phased, *trans*-acting siRNAs. *Genes Dev*, 2011, 25: 2540-2553
301. Zhai J, Zhang H, Arikait S, Huang K, Nan G-L, Walbot V, Meyers BC. Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proc Natl Acad Sci USA*, 2015, 112: 3146-3151
302. Zhang D, Luo X, Zhu L. Cytological analysis and genetic control of rice anther development. *J Genet Genomics*, 2011, 38: 379-390
303. Zhang D, Wilson ZA. Stamen specification and anther development in rice. *Chinese Sci Bull*, 2009, 54: 2342-2353
304. Zhang D, Yang L. Specification of tapetum and microsporocyte cells within the anther. *Curr Opin Plant Biol*, 2014, 17: 49-55
305. Zhang H, Liang W, Yang X, Luo X, Jiang N, Ma H, Zhang D. *Carbon starved anther* encodes a MYB domain protein that regulates sugar partitioning required for rice pollen development. *Plant Cell*, 2010, 22: 672-689
306. Zhang H, Xu C, He Y, Zong J, Yang X, Si H, Sun Z, Hu J, Liang W, Zhang D. Mutation in *CSA* creates a new photoperiod-sensitive genic male sterile line applicable for hybrid

- rice seed production. *Proc Natl Acad Sci USA*, 2012, 110: 76-81
307. Zhang K, Shi Z-M, Chang Y-N, Hu Z-M, Qi H-X, Hong W. The ways of action of long non-coding RNAs in cytoplasm and nucleus. *Gene*, 2014a, 547: 1-9
308. Zhang M-F, Chen L-P, Wang B-L, Yang J-H, Chen Z-J, Hirata Y. Characterization of *atpA* and *orf220* genes distinctively present in a cytoplasmic male-sterile line of tuber mustard. *J Horti Sci Biotech*, 2003, 78: 837-841
309. 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
310. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker Joseph R. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell*, 2006, 126: 1189-1201
311. Zhang Y, Chen Y. Long noncoding RNAs: new regulators in plant development. *Biochem Bioph Res Co*, 2013, 436: 111-114
312. Zhang Y, Liao J, Li Z, Yu Y, Zhang J, Li Q, Qu L, Shu W, Chen Y. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol*, 2014b, 15: 512
313. Zhou C, Han L, Fu C, Wen J, Cheng X, Nakashima J, Ma J, Tang Y, Tan Y, Tadege M, Mysore KS, Xia G, Wang Z-Y. The *trans*-acting short interfering RNA3 pathway and NO APICAL MERISTEM antagonistically regulate leaf margin development and lateral organ separation, as revealed by analysis of an *argonaute7/lobed leaflet1* mutant in *Medicago truncatula*. *Plant Cell*, 2013,
314. Zhou H, Liu Q, Li J, Jiang D, Zhou L, Wu P, Lu S, Li F, Zhu L, Liu Z, Chen L, Liu Y-G, Zhuang C. 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. *Cell Res*, 2012, 22: 649-660
315. Zhou H, Zhou M, Yang Y, Li J, Zhu L, Jiang D, Dong J, Liu Q, Gu L, Zhou L, Feng M, Qin P, Hu X, Song C, Shi J, Song X, Ni E, Wu X, Deng Q, Liu Z et al. RNase Z^{S1} processes *Ubl40* mRNAs and controls thermosensitive genic male sterility in rice. *Nat Commun*, 2014, 5: 4884-4892
316. Zhou Y, Cai H, Xiao J, Li X, Zhang Q, Lian X. Over-expression of aspartate aminotransferase genes in rice resulted in altered nitrogen metabolism and increased amino acid content in seeds. *Theor Appl Genet*, 2009, 118: 1381-1390
317. Zhou Y, Zhang X, Xue Q. Fine mapping and candidate gene prediction of the photoperiod and thermo-sensitive genic male sterile gene *pms1(t)* in rice. *J Zhejiang Univ-Sc B*, 2011, 12: 436-447
318. Zhu B, Yang Y, Li R, Fu D, Wen L, Luo Y, Zhu H. RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. *J Exp Bot*, 2015a, 66: 4483-4495
319. Zhu H, Xia R, Zhao B, An Y-q, Dardick CD, Callahan AM, Liu Z. Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs. *BMC Plant Biol*, 2012, 12: 149

320. Zhu X, Liang W, Cui X, Chen M, Yin C, Luo Z, Zhu J, Lucas WJ, Wang Z, Zhang D. Brassinosteroids promote development of rice pollen grains and seeds by triggering expression of Carbon Starved Anther, a MYB domain protein. *Plant J*, 2015b, 82: 570-581

附录 I 实验所用引物信息

附表 1 本研究所用引物信息。

Supplementary Table 1 The information of primers used in this study.

Primer name	Sequence (from 5' to 3')	Purpose	
58S-C-F	ATGAAGGACCGAGAAGAAGC	Complementary vector construction	
58S-C-R	GTTCTCAGATGATATGGAAGTGTG		
ORF1G-F	AGGTGTCGGTTCTAATGGTAGAAC	RNAi vector construction	
ORF1G-R	CATTAGGCGGAGATGGCAAT		
ORF1G-muF	CAAATCACAATAAAGATATATGGATCAAACCT TAATAAT		
ORF1G-muR	ATTATTAAGGTTTTGATCCATATATCTTTATTGTGA TTTTG		
ORF2G-F	CCATTTATAGGCTACTCCTTTCC		
ORF2G-muR	AACTAGTATATAGTATATAGAGAAGCCCAAGTAC CGAACTTTTTCTGTTCGCCAT		
ORF3G-F	CAATTTTGCCTGGTATCACCAA		
ORF3G-R	CATTAGGCGGAGATGGCAAT		
ORF3G-muF	CCCATTGCCATCTCCGCCTAATGGAACGCCTTGC ATGTTGG		
ORF3G-muR	CCAACATGCAAGGCGTTCCATTAGGCGGAGATG GCAATGGG		
dsi-F	TCGACTAGTTGGCAGGGTACCTGTACATGCCCA ACAAGCTCT		RNAi vector construction
dsi-R	GGCGAGCTCGGATCCTGTTCGCATGGACAATTG GTG		
ami1-1	AGTCCGATTTTAAACTGGACGAGCAGGAGATTC AGTTTGA	Artificial miRNA vector construction	
ami1-2	TGCTCGTCCAGTTTAAAATCGGACTGCTGCTGCT ACAGCC		
ami1-3	CTCTCGTGCAGATTTAAAATCGGATTCCTGCTGCT AGGCTG		
ami1-4	AATCCGATTTTAACTGACGAGAGAGAGGCAA AAGTGAA		
ami2-1	AGTAATTTGAAACTAGACCGAGCCAGGAGATTC AGTTTGA		
ami2-2	TGGCTCGGTCTAGTTTCAAATTACTGCTGCTGCT ACAGCC		

ami2-3	CTGCTCGCTCTTGTTCAAATTATTCCTGCTGCT AGGCTG	
ami2-4	AATAATTTGAAACAAGAGCGAGCAGAGAGGCA AAAGTGAA	
OX-1F	TATGGTACCGACTACATGGGCACCCCTTGAA	Overexpression vector construction
OX-1R	TATGGATCCCGTGATTTCAGCAGGTGGAGTTAA	
OX-2F	TCAGAGCTCGCATCAGGAAAGAAGCTTCTACTA	
OX-2R	TGACCTGCAGTACTTCAGTACCATCAGCGTGAC	
OX-3F	TATGGTACCAATTGTCCATGCGAACAGAAAA	
OX-3R	TATGGATCCCGTAAACAATTATCGGTGATA	
MIM-F	ATTGGTACCTGGCCATCCCCTAGCTAGGT	Target mimicry vector construction
MIM-R	ATTGGATCCCGGAAGCAAATTTACATGCACT	
MIMd-F	TTTAGGCATGGGATAACGCATCAGGAAAGCTTC GGTCCCCTCGGAATCA	
MIMd-R	CTTTCCTGATGCGTTATCCCATGCCTAAATTTCTA GAGGGAGATAAACA	
MIMo-F	TTTAGGCTTGGGATAACGCATCAGGAGAGCTTC GGTCCCCTCGGAATCA	
MIMo-R	CTCTCCTGATGCGTTATCCCAAGCCTAAATTTCT AGAGGGAGATAAACA	
P6-F2	AATTGTCCATGCGAACAGAAAA	Transgenic plants genotyping
P6-R2	CCGTAAACAATTATCGGTGATA	
GUS1.6F	CCAGGCAGTTTTTAACGATCAGTTCGC	
GUS1.6R	GAGTGAAGATCCCTTTCTTGTTACCG	
PMCGF	GGCTACCAAACCTTAAACAA	
PMCGR	CTGAGCTACACATGCTCAGGTT	
G-11491	TCGGTACCCAGCAGCAGCCACAGCAAA	
G-11494	TCGGATCCGCTGCTGATGCTGATGCCAT	
RACE5-1	CTGATGACTGTGTTCCAGTATTTG	5' RACE
RACE5-2	ACGCAAGCTTGGCTCTTTGTT	
RACE5-3	ACCTGGCATAGACCGATAGTTAC	5' RLM-RACE
RACE5-4	CTGATGACTGTGTTCCAGTATTTG	
RACE3-1	CAATTTTGCTGGTATCACCAA	3' RACE
RACE3-2	CAACATCATTAGGTTGCTGTGAT	
actin-inF	TGCTATGTACGTCGCCATCCAG	RT-PCR and qPCR
actin-inR	AATGAGTAACCACGCTCCGTCA	
UBQ-F	AACCAGCTGAGGCCCAAGA	
UBQ-R	ACGATTGATTTAACCAGTCCATGA	
RT-F	AAAGTTCGGTACTTGGGCTTCTCT	

水稻光敏感雄性核不育基因 *pms1* 的克隆与功能分析

RT-R	ACTCCATTCGATATTGTTGCAAGGGC	
CB-F	GGACCTGAGGGACTTTGTTGAC	
CB-R	TTGAGAAGATGGCAGCAGCTT	
U6-F	TACAGATAAGATTAGCATGGCCCC	Stem-loop RT-PCR
U6-R	GGACCATTCTCGATTTGTACGTG	
miR2118d-RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCA CTGGATACGACTAGGCA	
miR2118o-RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCA CTGGATACGACTAGGCT	
miR2118d-qF	GCCCTTCCTGATGCCTCCCA	
miR2118o-qF	CGCCCTCCTGATGCCTCCCA	
miR2118-qR	CCAGTGCAGGGTCCGAGGT	
pj23-F	AGAGAGAACTCTGCCAATATC	Molecular marker
pj23-R	AAAGTAACCCCTGCATAGACT	
P1-F	TCCTGTGGTCCAATCAAACG	
P1-R	CAGATCGTCTGATGATGTTGC	
P2-F	GCAATGGCTGTTGTCCGAAT	
P2-R	CTATGACGTTTGCTTTGTGCC	
P3-F	CTCCTTCGCAGACTCTACAC	
P3-R	ATGTCAATCTCTGATATCGGC	
P4-F	GCGAGCACATAAGTCCTCTA	
P4-R	ATAATTGTTTAAAGGTGTCAG	
P5-F	AGGCGCAGTAAAAACACCTG	
P5-R	CTTGCGTGGTTTCAAGGG	
P6-F	CATTAGGCGGAGATGGCAAT	
P6-R	ATAGCCAACACTCATCACTGTCTG	
P7-F	AAGGAGATGCAGAGAGGGTTTAC	
P7-R	AGGAAGGATAGGTGTTTGATGAAC	
P8-F	TCAGATGACTTAGCAACACAAGC	
P8-R	CTTGCTTATCCCAAATGTCGT	
P9-F	AGGAAACGATAGGCGAACAAC	
P9-R	ACAAGGCTTCCAACTTCCAA	
P10-F	CACGACCTTACTTATGAGCATG	
P10-R	AAAAGAACACAGCAGGGGAG	
P11-F	GTCCCTCATCTTTAGGCTGTG	
P11-R	GCCTAAAGAGGAGGGACGAA	
Fssr-F	TTAGTGCGTGTGGCATAGATG	
Fssr-R	GGCCTTTTTGTGAATTGGG	
RPA-F	GAAACCACGCAAGCTTGG	RPA
RPA-R	TCCTTAGATCATCCAATCTGTTAAT	
BS1-1F	TTTTGTTTTATTTTTTGAAAA	Bisulfite sequencing
BS1-1R	AACTRCACCATCAAATTCARAT ^{a)}	

BS1-2F	TGAGGTATTTTATTTTTYTGAATGG ^{b)}
BS1-2R	ACARCTRCCAACTRTCATTCAACA ^{a)}
BS1-3F	ATYTGGAAAGTTTTTTTTTATGATG ^{b)}
BS1-3R	ACCAAAAAAAAAATAAATAACCCAA
BS1-4F	TTYAGGATATATTTTTATTGGGTT ^{b)}
BS1-4R	TAAAACCCAATAAATATCARTTTC ^{a)}
BS1-5F	GATATTAATGATTTTGATTATTGGTG
BS1-5R	CARTTCACAAAAAATTCARATAA ^{a)}
BS1-6F	ATGATTATATGGGTATTTTTTGAA
BS1-6R	TCATCCAATCTRRTAATTTAAAACTAA ^{a)}
BS1-7F	TTTTGTGAAYTGGAGTATGATATA ^{b)}
BS1-7R	AAACAATTAATAATACCARACAAAA ^{a)}
BS1-8F	ATATATGATTAATAATTTAATAATTTGTTT
BS1-8R	TAATATATAAAAAAACCCAAATACC
BS1-9F	AATTTTGTYTGGTATTATTAATTGTT ^{b)}
BS1-9R	TAACAAACARTAACACCATTTTAA ^{a)}
BS1-10F	GATAATTGTTTAYGGTTTATTGTT ^{b)}
BS1-10R	AAATTTATCCCTRTCACCACTAAC ^{a)}
BS1-11F	TTAATGGTGGTGTAATAAATTGTTT
BS1-11R	TATTAACAACCATCCAATTTAACA
BS1-12F	AGTGTTTATGTAAATGGTTATTTTTT
BS1-12R	TCAACTRAAAAATAACAAAAATTTACT ^{a)}
BS1-13F	AAGTTGTGAGTTAAAAGGAAGAATAY ^{b)}
BS1-13R	ACTTAACCATTTCAAAATTTTCAT

^{a)} R 为 A 和 G; ^{b)} Y 为 C 和 T。

附表 2 *PMS1T*-phasiRNA 预测靶基因 5' RLM-RACE 所用引物信息。

Supplementary Table 2 The information of primers used in the 5' RLM-RACE.

phasiRNA	Putative target	Sequence (from 5' to 3')	purpose
2s-s	LOC_Os01g70020	ATTGCCGTTTGGGCTGTG	Inner primer
		AGCACAGAGGCTGGGGCA	Outer primer
2s-m	LOC_Os06g22060	CGCTTAACGCTGCCTTCA	Inner primer
		CAGAGAGATACGGATAACCAGGATA	Outer primer
	LOC_Os03g63520	AAATACACGGGAACATTGCTT	Inner primer
		AGATGTCAATCGTCGCAAAA	Outer primer
LOC_Os11g43760	CCCTATTACCTCGGTGTGAAG	Inner primer	
	GCCTCGGATGTGCCAGAT	Outer primer	
2as-s	LOC_Os04g51040	CCAACAATACGGTCATCGCT	Inner primer
		TCCTAAGACGATGCTCCCAT	Outer primer
	LOC_Os05g44310	GGTTCATTTCCGTTTCAGCAG	Inner primer
2as-m	LOC_Os01g06920	TGCCTCAACACCATCTATCTTT	Outer primer
		AGGGATGACAAATCAGCGA	Inner primer
	LOC_Os01g06920	GGTAAAGATCCAGACAAACCAC	Outer primer
		AGGGATGACAAATCAGCGA	Inner primer
	LOC_Os04g51040	GGTAAAGATCCAGACAAACCAC	Outer primer
		CCAACAATACGGTCATCGCT	Inner primer
	LOC_Os07g23710	TCCTAAGACGATGCTCCCAT	Outer primer
		CTCGGTCCGCTGGAGAAG	Inner primer
	LOC_Os07g44110	GGATCATAGGGAGCCCGTAC	Outer primer
		CGAAATTCTGCCCGATGC	Inner primer
LOC_Os07g23570	GTACAACAAAGCGGTGCTAAA	Outer primer	
	CCTAAACTCATCCGCATCCT	Inner primer	
1s	LOC_Os02g54110	CGAAGTGGCAAACCTTTATAGAA	Outer primer
		GGCATTCAAGCCTGTAGCG	Inner primer
	LOC_Os10g09870	CCTTCCTCCCGTCATTTTC	Outer primer
		GCAAACCTGTGTTGAGCG	Inner primer
	LOC_Os11g43770	AACTCAGGTATTAGCAGGCAAA	Outer primer
		GGAAGAAAAGATGGACGAGATT	Inner primer
	LOC_Os12g19549	GCAGCAAAGGACGAAAAGG	Outer primer
		AACAGTCGCACCTTGAATGA	Inner primer
	LOC_Os04g49510	TGCTACCAGGTATCCTAAAGAAA	Outer primer
		GCCACTACAGCGAGCGGA	Inner primer
3s	LOC_Os11g30910	GCAAGTCCATTTCCAAGCG	Outer primer
		CATAGTTGTAGTGTCCGATTTGC	Inner primer
4s	LOC_Os09g38790	CCTTTCTCCCCTTTACAGAGT	Outer primer
		TGACGCACAACTAAAGGGA	Inner primer
6s	LOC_Os02g40400	CGCTAATGGGCTCAGAAGAA	Outer primer
		GAGTTGCCCCACATAGC	Inner primer
	LOC_Os03g10540	TGTATCGGTCAAAATGTGCC	Outer primer
		CCACGCTCTGGAGGTTGT	Inner primer
	LOC_Os03g51900	GCTTTCTTGGCTTCGGTGA	Outer primer
		CGTAAGGGTTCAGAATAAAAGGG	Inner primer
	LOC_Os06g08560	AGACTTATTGTCATCCGTTGCT	Outer primer
		GCCTTGGCAAATCGTGAGA	Inner primer
	LOC_Os10g40830	AGAGTTTATGTTCCCTACCAGC	Outer primer
		CCGAATGAGCACAGGACC	Inner primer

10s	LOC_Os01g69020	GGTTCACCAATCCGCTT	Outer primer
		ACCGTCTGATACCACCAACC	Inner primer
		GCGCGAACAGGCAGAACA	Outer primer
LOC_Os02g02470	LOC_Os02g02470	CAATACGTGTAGGGGTGCCATACGA	Inner primer
		CTCTCCTGGGATTGTTTCATACA	Outer primer
		CTCAAATGAACCGTAAGCACA	Inner primer
11s	LOC_Os01g32250	AGCAAGTCTGCGATTTCAAG	Outer primer
		TGGTCGAAGTTGTAATCTCCG	Inner primer
		CAAACATCGTGAGTCGTTACC	Outer primer
LOC_Os01g72200	LOC_Os01g72200	TGATGAATGGCTACTCCGAAG	Inner primer
		GCCAGACATAGGATTGGTAGTG	Outer primer
		ACTGTAATGTCGCTTATGTGTCA	Inner primer
LOC_Os10g22380	LOC_Os10g22380	GAGGAAAGGCTATTACAGGTGC	Outer primer
		CCACGAGTATGAGCAGGGT	Inner primer
		CATGGCGATAATCCACGAT	Outer primer
LOC_Os12g29360	LOC_Os12g29360	CAACAGTATCCGAACCCGA	Inner primer
		GCCGTTACCTTTGACCTTG	Outer primer
		CGCAGGAAGACCCATTA ACT	Inner primer
12s	LOC_Os01g28960	CCACTGGCTGGACCCCTG	Outer primer
		AAGACCCTTTAGCTGAGCGT	Inner primer
		AGCGACTCCTGGAACCTG	Outer primer
LOC_Os09g16340	LOC_Os09g16340	AAGACCCTTTAACTGAGCGTT	Inner primer
		TTCGCTGTTGCTGCCCTC	Outer primer
		GCGGGCGGA ACTAGAGGT	Inner primer
LOC_Os10g12630	LOC_Os10g12630	CGACTGACCCAAACAAGCA	Outer primer
		GTGTGGACCATCCCAAACA	Inner primer
		CCAGGCTGACGCAGGACC	Outer primer
LOC_Os10g24460	LOC_Os10g24460	GGTGTAGACCATCCCAAACA	Inner primer
		CCAGGCTGACGCAGGACC	Outer primer
		GTCTGCTCTTACGGGGC	Inner primer
LOC_Os11g12850	LOC_Os11g12850	CAAGCAACCTGGTCCCTCT	Outer primer
		AACAAGCAACCTGGTCCCT	Inner primer
		GGCAAGCGTCGGATTCTC	Outer primer
LOC_Os11g12874	LOC_Os11g12874	GGA ACTTGAGGTGCTGCG	Inner primer
		AGACCATCCCAAACAAGCA	Outer primer
		GATGCTGGGGCGAAGGAG	Inner primer
LOC_Os11g37430	LOC_Os11g37430	CCGCTGCCTGACTGTGAG	Outer primer
		GCATAGAACCTGTTGAATTATCCA	Inner primer
		ATTATCAGGTCCAGGCATCTC	Outer primer
14s	LOC_Os03g08270	GCTGATTCTGAGGTCTCCCAT	Inner primer
		GCAGGGAGACATAACACCAGAT	Outer primer
		ACAACAGTGGCAGGACGG	Inner primer
16s	LOC_Os03g11440	TCTCCCCCTCTACCATTTGA	Outer primer
		CGGGCGATGATGTCGTATG	Inner primer
		TCGAGAACGAGAATACCCTTC	Outer primer
18s	LOC_Os01g01200	GGCACAGGGTTTGTGGAA	Inner primer
		TTTTGAAATGCCCTCTGT	Outer primer
		CATCTCTGAAGACCCTCGCA	Inner primer
LOC_Os01g56780	LOC_Os01g56780	AAAAAAGGGTCACGGCGA	Outer primer
		ATGGATGGATTGCGCCCGC	Inner primer
		CTGCTGGGAGTAACTCTGCC	Outer primer

附录 II 在读期间研究成果

一、攻读博士期间发表论文

1. **Fan Y**, Yang J, Mathioni SM, Yu J, Shen J, Yang X, Wang L, Zhang Q, Cai Z, Xu C, Li X, Xiao J, Meyers BC, Zhang Q. *PMSIT*, producing phased small interfering RNAs, regulates photoperiod-sensitive male sterility in rice. *Proc Natl Acad Sci USA*, (Accepted)
2. **范优荣**, 曹晓风, 张启发. 光温敏雄性不育水稻的研究进展. 科学通报 (Accepted)
3. 肖景华, 吴昌银, 袁猛, 王妮丽, **范优荣**, 杨猛, 欧阳亦聃, 阮一骏, 张启发. 中国水稻功能基因组研究进展与展望. 科学通报, 2015, 60:1711-1722
4. **Fan Y**, Zhang Q. Understanding a key gene for thermosensitive genic male sterility in rice. *Sci China Life Sci*, 2014, 57:1241-1242
5. Yu J, **Fan Y**, Liu N, Shan Y, Li X, Zhang Q. Rapid genome evolution in *Pms1* region of rice revealed by comparative sequence analysis. *Chinese Sci Bull*, 2007, 52: 912-921

二、参加会议摘要

1. Fan Y, Yang J, Mathioni SM, Yu J, Xu C, Li X, Xiao J, Meyers BC, Zhang Q. *PMSIT*, a PHAS gene producing phased small interfering RNAs, regulates photoperiod-sensitive male sterility in rice. Abstract, The 14th International Symposium on Rice Functional Genomics, 2016, Montpellier, France
2. Fan Y, Yang J, Mathioni SM, Yu J, Shen J, Yang X, Wang L, Zhang Q, Cai Z, Xu C, Li X, Xiao J, Meyers BC, Zhang Q. The map-based cloning and functional analysis of photoperiod-sensitive genic male sterility gene (*pms1*) in rice. Abstract, The 17th Plant Genomic in China, 2016, Fuzhou, China

三、申请专利

张启发, **范优荣**. 水稻光敏感核不育基因 *pms1* 的分离克隆及应用. 中国专利. 专利申请号: 201610347215.

致 谢

很多人都问过我，作为女生花了这么多年的时间去拿一个博士学位到底值不值得，有没有后悔过？确实，从亲朋好友、老师和同学眼里一直的好学生变为始终没能按时毕业的“超龄留级生”，心里会有很大的压力，但如果再让我重新选择一次，依然会毫不犹豫地选择在华农最好的实验室里跟随最好的导师走上科研的道路。犹记得初入实验室时张老师指着师兄对我说“他是研究‘是什么’，你的任务是研究‘为什么’”，多年以后弄清楚了“是什么”，但“为什么”愧不能交给张老师一份满意的答卷。在这个纷繁复杂的年代，处处充满着诱惑和各种隐规则的社会里，张老师极尽所能地为我们提供了一片极为难得的不受外界干扰、公平公正、积极向上的科研净土。感谢导师这么多年来对我工作的理解和包容，即便是在得到与前人研究结果完全不同的情况下依然坚定地支持我的想法，鼓励我的每一次探索和尝试，很难想象世界上会有几个导师能够容忍自己的学生读这么多年的博士，还保障学生的生活，为学生取得的每一个结果而高兴。他以他亲身的经历和深厚的人生阅历为我们讲述生活和学习的各种道理，传授“激扬梦想、追求卓越”的科学理念，支持和鼓励我们出去开会、交流学习，为我们提供充分的学习和科研成长的空间。在论文写作和投稿的过程中更是进一步深刻体会和感受到了导师严谨、细致、一丝不苟的科研作风。这些都值得我一辈子去追求和学习。

非常感谢美国丹佛斯植物研究中心 Blake Meyers 教授在我们研究最困难的时候为我们打开了一扇窗，并且在 phasiRNA 分析和 PARE 文库构建和分析方面给予了不遗余力的帮助。是他让我切实体会到了“科研无国界”的深刻内涵，也明白了闭门造车是永远也出不了好的研究成果，交流和思想的碰撞才是科研发展的原动力。

在作物遗传改良重点实验室待了很多年，见证了身边很多师兄师姐从学生到老师的蜕变，他们为我们树立了良好的榜样，其中赵毓、胡红红、欧阳亦聃、陈炯炯等师姐更是为我们女生之楷模。感谢引领我走上科研之路的吴昌银老师、袁文雅和於金生师兄，吴昌银老师是我来实验室除张老师外第一个接触到的科研人，至今还记得初次见面时他给我画 T-DNA 突变体库构建原理时认真的模样和那细致的讲解。感谢王磊博士多年来与我一起并肩战斗，不断给我提供好的建议，无私分享各种科研资源。感谢同一级的已经毕业的同学们，他们都做出了相当好的工作，发表了高

档次的文章，也是激励我坚持下来的原因之一。感谢邵林博士让我知道了外面的世界有多么的精彩，科研界的人生活是多么的丰富，而不仅仅只是科研论文上一个冷冰冰的名字。感谢沈建强在实验上给予我的很多帮助和建议，难忘那段文章投稿修改的虐心过程中我俩相互鼓励加油的岁月。感谢赵晓波和杨雪菲在实验过程中的大力协助。尤其需要感谢课题组这么多年来一起携手走过的兄弟姐妹们，在此不一一列举，感谢大家组成我们这个团结、和谐、积极向上的科研团体，跟你们一起工作、学习、生活感觉非常愉快。

还要感谢那些为实验室正常运转付出了辛勤劳动的老师。感谢徐才国老师多年来在田间的辛苦管理，没有您耐心细致的管理，我们的科研材料不可能正常生长、收获。也要感谢继任的张庆路老师对我们的继续支持。感谢李香花老师在试剂订购上为我们提供的便利和对整个实验室仪器设备正常运行的保障。感谢肖景华老师在各种管理事务上的尽心尽力。还要感谢刘亚萍、朱君、严蒙、周中文、甘道清、刘祖刚、方国芹在实验室各种后勤事务上提供的各项帮助。

最后特别感谢多年来默默守候在我身边的家人，你们是我坚持下去的最大支柱！

本研究是在作物遗传改良国家重点实验室和国家植物基因研究中心（武汉）完成，得到了国家自然科学基金、国家科技部重大专项和梅琳达·盖茨基金的资助。

范优荣

2016年9月 狮子山