上海交通大学硕士学位论文

水稻中microRNA介导的基因沉默相关SNP的研究

|  |  |  |
| --- | --- | --- |
| 硕士研究生 | ： | 黄飘飘 |
| 学号 | ： | 1140809051 |
| 导 师 | ： | 张大兵教授、李俊彦博士 |
| 申请学位 | ： | 理学硕士 |
| 学科 | ： | 生物学 |
| 所 在 单 位 | ： | 生命科学技术学院 |
| 答 辩 日 期 | ： | 20 年 月 |
| 授予学位单位 | ： | 上海交通大学 |

Dissertation Submitted to Shanghai Jiao Tong University for the Degree of Master

INVESTIGATION OF SNPS INVOLVED IN RICE MIRNA-MEDIATED GENE SILENCING

|  |  |
| --- | --- |
| Candidate： | Piaopiao Huang |
| Student ID: | 1140809051 |
| Supervisor： | Prof. Dabing Zhang & Dr. Junyan Li |
| Academic Degree Applied for： | Master of Science |
| Speciality： | Biology |
| Affiliation： | School of Life Science and Biotechnology |
| Date of Defence： |  |
| Degree-Conferring-Institution： | Shanghai Jiao Tong University |

**上海交通大学**

**学位论文原创性声明**

本人郑重声明：所呈交的学位论文《水稻中microRNA介导的基因沉默相关SNP的研究》，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日

**上海交通大学**

**学位论文版权使用授权书**

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权上海交通大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

**保密**□，在 年解密后适用本授权书。

本学位论文属于

**不保密**□。

（请在以上方框内打“**√**”）

学位论文作者签名： 指导教师签名：

日期： 年 月 日 日期： 年 月 日

水稻中microRNA介导的基因沉默相关SNP的研究

摘 要

miRNA是重要的调节因子，在植物的生长发育和抗逆性中扮演不可替代的角色。和miRNA介导的基因沉默相关的单核苷酸多态性(SNP)可能导致非常严重的植物农艺性状改变。为了进一步了解在miRNA和其靶基因上的进化压力以及SNP如何通过影响miRNA和靶基因对的互补模式从而影响miRNA相关的性状，我们进行了针对miRNA介导的调节的全基因组SNP的研究，并且分析了这些SNP对miRNA和靶基因互补模式以及相关性状的影响。我们发现pre-miRNA上SNP的密度比基因间隔区和外显子区域都要低，这和miRNA是植物中主要调控因子的角色相符。对比成熟的保守miRNA和非保守miRNA发现，两者的SNP分布并不相同，暗示了两者各位点上的进化压力不同；而对比成熟的保守miRNA和其相应的靶基因结合位点，则发现它们之间SNP分布有相关性，这也支持了miRNA和相应靶基因结合位点的共同进化的观点。在本研究中，我们将单倍型分析拓展成联合互补模式分析从而可以应用在miRNA和其结合位点上，并且我们找到了两个靶基因的结合位点上携带有两个可能对miRNA介导的调节产生重要影响的SNP，但是并没有发现这些SNP给水稻带来显著的形状变化。本研究是全基因组分析SNP对miRNA和靶基因相互作用的全新尝试，可能加深甚至改变我们对miRNA介导的基因沉默机理的理解以及SNP给miRNA和靶基因互作带来的影响。

关键词：水稻，microRNA，单核苷酸多态性，互补性，性状改变

INVESTIGATION OF SNPS INVOLVED IN RICE MIRNA-MEDIATED GENE SILENCING

ABSTRACT

MiRNAs are key regulators and play inevitable role in plant growth, development and stress tolerance. Single nucleotide polymorphisms (SNPs) that are involved in miRNA-mediated gene silencing might cause serious changes to plant agronomic traits. To further understand the evolutionary pressure imposed on miRNAs and their targets as well as how SNPs could affect the complementarity of miRNA:target pairs and further bring changes to miRNA-involved phenotypes, we performed a genome-wide scan of SNPs involved in miRNA-mediated regulation, and analyzed their effects on miRNA:target complementarity and related phenotypes. We found that SNP density of pre-miRNAs was lower than both intergenic regions and exons, consistent with their established roles as master regulators in many genetic pathways. And comparison between conserved mature miRNAs and non-conserved mature miRNAs showed the SNP distributions were rather different, implying the differential selection pressure upon them; while comparison between conserved mature miRNAs and their binding sites showed similar SNP distribution, and this supported the co-evolution of miRNAs and their binding sites of cognate targets. In this study, we extended haplotype analysis into combined complementarity pattern analysis to apply on miRNA:target pairs, and found two target genes carrying SNPs which potentially may bring great changes to miRNA-mediated regulation, but we didn’t find obvious phenotypical changes for these SNPs. This study provided a new attempt of analyzing genome-wide SNPs on miRNA:target interactions, and might change and deepen our understand of the mechanism of miRNA-mediated gene silencing and the effects of SNPs on the miRNA:target interaction.

KEY WORDS: Rice, microRNA, SNP, complementarity, phenotypical change

目 录

↑

（黑体3号字，段前0.7厘米，段后0；目录题目与条目之间空两行）

[第一章 正文文字格式 1](#_Toc251590708)

[1.1 论文正文 1](#_Toc251590709)

[1.2 字数要求 1](#_Toc251590710)

[1.2.1 硕士论文字数要求 1](#_Toc251590711)

[1.2.2 博士论文字数要求 1](#_Toc251590712)

[1.3 论文的主要内容与章节安排 1](#_Toc251590713)

[第二章 图表、公式格式 2](#_Toc251590714)

[2.1 图表格式 2](#_Toc251590715)

[2.2 公式格式 4](#_Toc251590716)

[2.3 引用说明 4](#_Toc251590717)

[2.4 本章小结 4](#_Toc251590718)

[第三章 其他格式要求 5](#_Toc251590719)

[3.1 页码 5](#_Toc251590720)

[3.2 页眉 5](#_Toc251590721)

[3.3 目录 5](#_Toc251590722)

[3.4 正文的层次安排 5](#_Toc251590723)

[3.5 打印要求 6](#_Toc251590724)

[3.5.1 页面设置 6](#_Toc251590725)

[3.5.2 字体 6](#_Toc251590726)

[3.5.3 字号 6](#_Toc251590727)

[第四章 结束语 7](#_Toc251590728)

[4.1 主要工作与创新点 7](#_Toc251590729)

[4.2 后续研究工作 7](#_Toc251590730)

[参 考 文 献 8](#_Toc251590731)

[附录**1** 10](#_Toc251590732)

[致 谢 11](#_Toc251590733)

[攻读硕士学位期间已发表或录用的论文 12](#_Toc251590734)

图 录

↑

（黑体3号字，段前0.7厘米，段后0；目录题目与条目之间空两行）

[图2-1论文页面设置图 2](#_Toc251150896)

[图2-2内热源沿径向的分布 3](#_Toc251150897)

表 录

↑

（黑体3号字，段前0.7厘米，段后0；目录题目与条目之间空两行）

[表2-1高频感应加热的基本参数 3](#_Toc251151029)

[表3-1论文的层次代号与说明 6](#_Toc251151065)

# 正文文字格式

↑

（黑体3号字居中，段前0.7厘米，段后0.7厘米，单倍行距）

## 论文正文

↑

（黑体4号字顶格，段前0.7厘米，段后0.7厘米，单倍行距）

论文正文是主体，一般由标题、文字叙述、图、表格和公式等五个部分构成。写作形式可因科研项目的性质不同而变化，一般可包括理论分析、计算方法、实验装置和测试方法，经过整理加工的实验结果分析和讨论，与理论计算结果的比较以及本研究方法与已有研究方法的比较等。

论文内容一般应由十个主要部分组成，依次为：1、封面，2、中文摘要，3、英文摘要，4、目录，5、符号说明，6、论文正文，7、参考文献，8、附录，9、致谢，10、攻读学位期间发表的学术论文目录。

## 字数要求

### 硕士论文字数要求

↑

（黑体小4号字顶格，段前0.7厘米，段后0.7厘米，单倍行距）

硕士学位论文字数为4～5万字左右。

硕士专业学位论文一般为2～5万字左右；

### 博士论文字数要求

博士学位论文字数为8～10万字左右。

论文中一般不出现四级标题。

## 论文的主要内容与章节安排

本文主要分为……除了第一章，每章结束都应该有小结。

# 水稻miRNA信息整理和相应靶基因的预测和整理

## 水稻miRNA整理和相应信息收集

miRBase中收集和整理了目前所有已发表的miRNA的序列以及相关的注释，包括miRNA描述、所在基因家族、“茎环”结构 (stem-loop)、深度测序结果、基因组内容(genome context)等[[1]](#endnote-1)，其中也包含了所有已发表的水稻miRNA的序列和注释，而水稻miRNA是以MSU7.0作为参考基因组。 所以，我们首先从miRBase中将所有的水稻miRNA及其相关注释信息下载下来，主要是包括pre-miRNA序列、成熟产物 (mature miRNA)、基因组内容。在此总共收集得592个pre-miRNA和相应的713个成熟miRNA。

但是，其中有14个pre-miRNA以及相应的24个成熟miRNA的基因组坐标信息并没有提供，其中包括著名的调节MADS-box基因的osa-miR444家族，但由于随后的SNP分析需要miRNA的基因组坐标。所以，我们用本地BLAST，将MSU7.0的水稻基因组序列创建成本地的BLAST数据库，并且用这些pre-miRNA的序列作为信息进行比对。对于BLAST结果的筛选标准是所有的序列完全匹配，其中发现osa-miR444家族的所有pre-miRNA以及osa-miR810a在参照基因组上比对的匹配结果中基因组坐标是间断的，说明它们基因的转录本经过了剪切和拼接。另外在线的BLAST也得出相同的结果，更加有意思的是发现osa-miR444的基因组位置正处在MADS-box基因的反义链上，这正是之前有报道过的天然反义miRNA (nat-miRNA) [[2]](#endnote-2)。通过运用BLAST，最终我们得到了585个pre-miRNA和703个成熟miRNA的所需信息，可以供本实验进一步研究使用。

miRBase中成熟miRNA的基因组位置是其在pre-miRNA上的相对坐标，使用Python脚本将相对坐标结合pre-miRNA的基因组坐标而得到所有相应的成熟miRNA的基因组绝对坐标。

### miRNA根据保守性分类

水稻miRNA可以根据其在不同的植物物种中的保守性分成两类：一类是保守miRNA (conserved miRNA)，也就是在其它的物种中有相应的同源miRNA；另一类则是非保守miRNA (non-conserved miRNA)，通常也称为水稻特异miRNA (rice-specific miRNAs)。两者之间有一些明显的差别，首先就其定义而言，它们的保守性不同；其次，一般说来，多数的保守miRNA的表达量都比非保守miRNA的表达量高；最后，最关键的差别是，保守miRNA几乎都有可辨认的靶基因，而且这些靶基因往往也在植物物种中有保守性，甚至是一些调控因子或其它的调控基因，比如DCL1，ARGONAUTE和生长素受体TIR1，但是非保守miRNA却鲜有可辨认的靶基因[[3]](#endnote-3)。由此可以看出，保守miRNA在植物生长发育等过程中会比非保守miRNA扮演更重要的角色，并且两者之间因其进化过程和靶基因的结合模式不同所以作用机理也可能略有差异，所以在分析研究中将两者分别进行研究和比较非常有必要。

使用miRBase提供的miRNA家族分类标准对pre-miRNA进行保守性分类，分类标准是：该pre-miRNA所在的家族如果还有其它的植物物种的miRNA，就认定其为保守miRNA，否则就认为非保守miRNA。最后的分类结果是，191个pre-miRNA归类为保守miRNA，401个pre-miRNA归类为非保守miRNA；同时，220个成熟miRNA归类为保守成熟miRNA，另外493个归类为非保守成熟miRNA。

## miRNA靶基因预测和整理

鉴于保守miRNA拥有更多重要的靶基因，并且针对这些miRNA的研究更加充分，本实验只针对保守miRNA的靶基因进行研究。而现有的miRNA靶基因的预测方法主要是生物信息方法，比如在线植物小RNA靶基因预测服务器 (psRNATarget, http://plantgrn.noble.org/psRNATarget/)；而实验中鉴定靶基因的方法有miRNA过表达或抗miRNA靶基因 (miRNA-resistant target)，RNA连接酶催化的5’-RACE，降解组测序等[[4]](#endnote-4)。

本实验采用生物信息预测方法结合降解组测试的结果获得miRNA靶基因。使用生物预测在线软件psRNATarget，采用默认的参数进行预测[[5]](#endnote-5)。另外 Li et.al 2010对水稻进行了全转录组的miRNA靶基因鉴定[[6]](#endnote-6)，我们收集了其鉴定的结果。但是这些鉴定出来的靶基因缺乏更加具体的miRNA结合位点的信息，我们使用psRNATarget用这些靶基因反向搜索其对应的miRNA，从而获得相应的结合位点的信息。最终，结合两者的结果，我们得到了823个靶基因，相对应的有2113个miRNA靶基因对。

## 本章小节

本章详细叙述了如何获得所有的miRNA，以及研究所必需其在的水稻基因组的坐标信息，并且使用BLAST得到了一些miRNA所缺失的基因组坐标。进而，根据miRBase的miRNA家族分类，将miRNA分成保守和非保守的miRNA。最后，结合生物信息预测和降解组的数据，得到保守miRNA的靶基因823个，形成miRNA和靶基因的作用对总共2113对。

# 水稻miRNA靶基因的生物相关性筛选

所得到的823个靶基因中，46个靶基因是降解组测序所证实的，另外777个靶基因是基于植物miRNA和其靶基因具有高度互补性(complementarity)的特征而预测出来的。但是用这种生物信息方式得出的预测结果，其生物相关性(biological relevancy)并不能得到保证，也就是这些预测的靶基因并不能确定真的能够被该miRNA调控，并且能够造成植物生理上的影响。因为除了miRNA和靶基因结合位点的序列互补性之外，还有其它因素会影响到miRNA和靶基因的相互结合4：

1. 靶基因的可接近性，这个是由靶基因mRNA的二级结构决定，如果靶基因的结合位点部分需要很高的能量来释放从而可以和成熟miRNA结合的话，miRNA和靶基因的结合能力就相应下降[[7]](#endnote-7)；
2. miRNA和靶基因结合位点序列的整理互补性模式，即使当互补程度相同时，错配出现的地方不同，也会非常影响到miRNA对靶基因的调节[[8]](#endnote-8)；
3. RNA结合蛋白等。

除了上述列举的因素之外，即便是实验的方法也并不能完全保证所鉴定的靶基因的生物相关性，比如miRNA过表达，虽然可以通过过表达的转基因植株表现出的表型来判断潜在的靶基因，但这种人工的过量表达和原本miRNA在植物中的表达水平不同，在原本的植物组织中可能并不具有调节能力的miRNA因此却被鉴定出可以调节某靶基因。

鉴于此，我们需要对所预测出的靶基因生物相关性筛选。由于植物miRNA介导的基因沉默主要是通过靶基因转录子的剪切这种方式，在通常的假设下，也就是，互补性是基因沉默的唯一决定因子和任何给定的装载miRNA的RNA介导的沉默复合体(miRNA-loaded RNA-induced silencing complex, miRISC)都能够独立运作[[9]](#endnote-9),[[10]](#endnote-10)，靶基因mRNA的表达量会被相应的miRNA下调，因此植物miRNA的表达量和其靶基因的表达量就呈现负相关。并且也有实验表明过表达miR319时，五个编码TCP转录因子的基因下调，呈现负相关[[11]](#endnote-11)。因此，我们考虑以miRNA和靶基因的表达量负相关性来筛选具有生物相关性的靶基因。

从RiceFrend[[12]](#endnote-12)的水稻整个生长期间各种组织/器官的基因表达实验中得到表达量的数据，该表达量数据现存于EMBL-EBI数据库，编号为E-GEOD-21396。其中有48个样本以代表生长发育不同阶段的组织／器官，而其中包含了169个miRNA的探针以检测在样本中的miRNA表达量。

## miRNA的表达热图(heatmap)

在所得到的表达量数据中，每个miRNA都有相对应的六个探针的数据，分别是pre-miRNA在3’端、5’端和成熟miRNA序列各有两个探针，所以这些探针中3’端和5’端的探针数据看作是表示pre-miRNA的表达量，而成熟miRNA序列中的探针看作是成熟miRNA的表达量。分别对pre-miRNA的四个探针数据和成熟miRNA的两个探针数据取平均值，就得到相应的pre-miRNA和成熟miRNA的表达量。

利用Genesis软件[[13]](#endnote-13)分别绘制了pre-miRNA和成熟miRNA的表达热图。

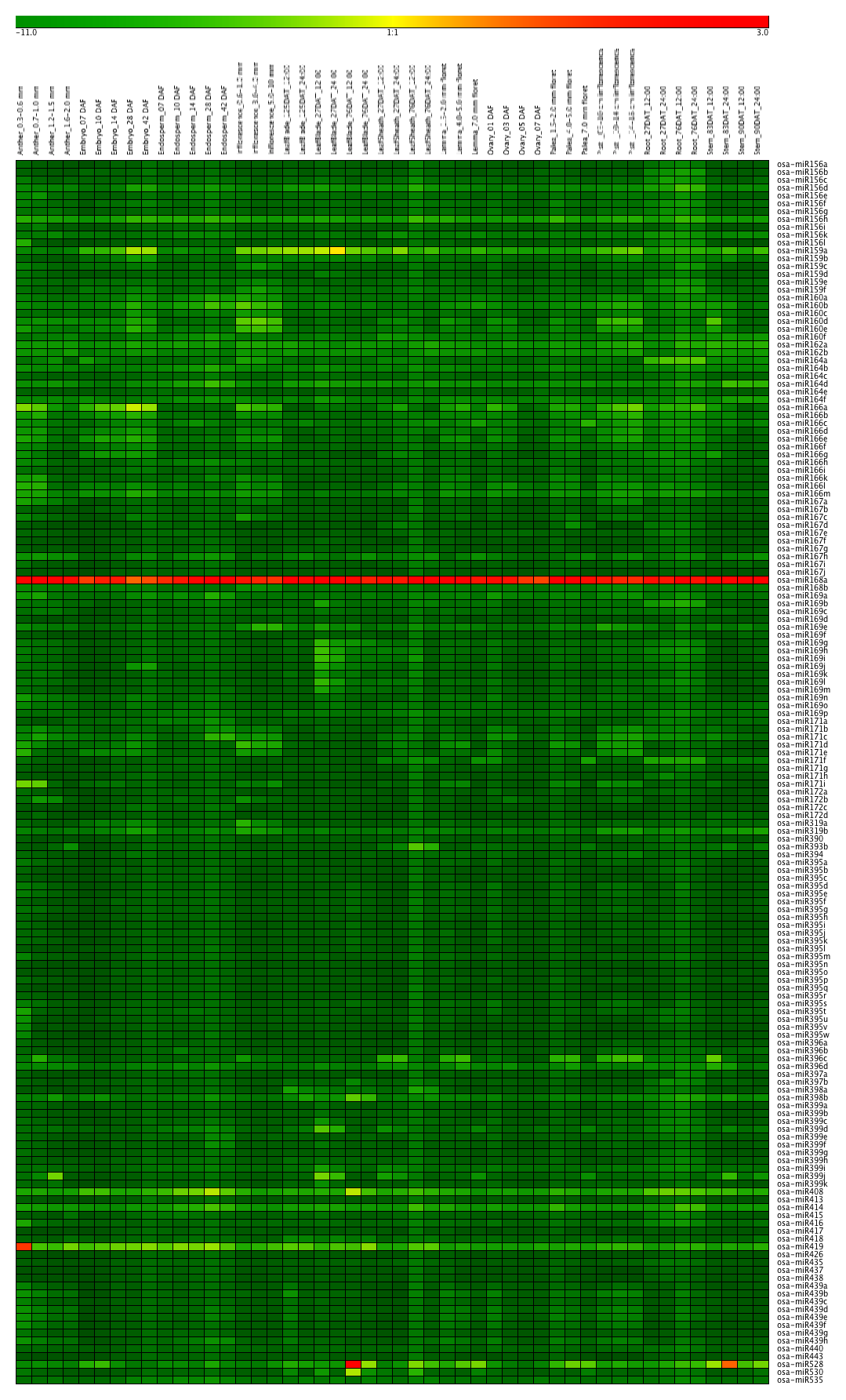


图3-1 pre-miRNA在水稻各个不同时期的组织／器官中的表达热图

Fig.3-1 Heatmap of expression levels of pre-miRNAs in various tissues/organs at various stages

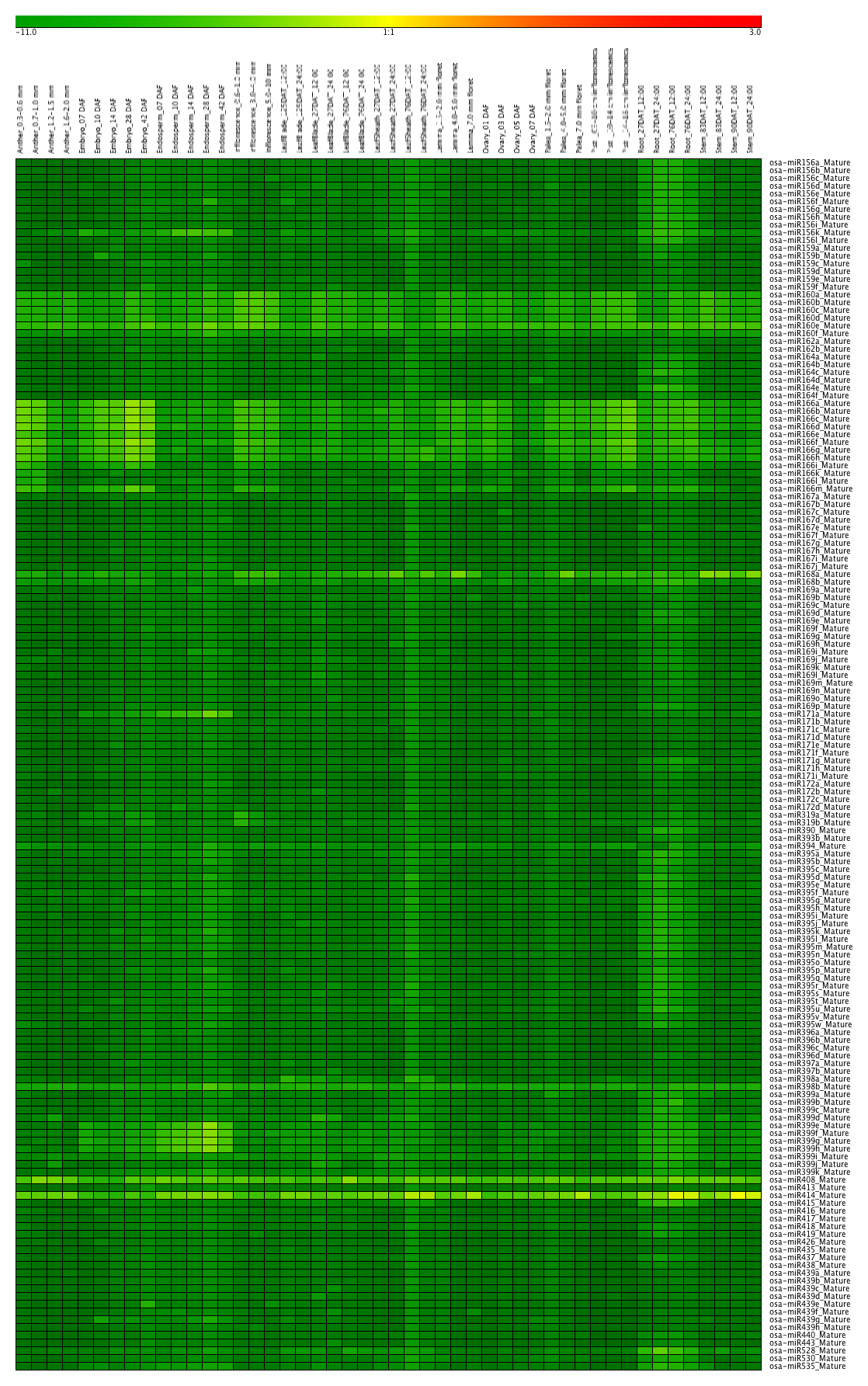


图3-2 成熟miRNA在水稻各个不同时期的组织／器官中的表达热图

Fig.3-2 Heatmap of expression levels of mature miRNAs in various tissues/organs at various stages

以上的图3-1和图3-2中，显示了同一个miRNA在不同阶段的组织和器官中的表达量存在差别，也说明miRNA在不同的组织和器官中的调节强度有差异。另外图3-1中，同一个家族的miRNA表达存在差异，也暗示了miRNA存在的亚功能化 (subfunctionalization)，但是图3-2中，同一个家族的miRNA之间的差异较小，因为同一个家族的成熟miRNA序列非常相似，所以其探针所得到的表达量也很相似。图3-1显示，osa-miR168a是所有检测的pre-miRNA中表达量最高的，图3-2则显示，osa-miR414是所有检测的成熟miRNA中表达量最高的。

## 降解组验证miRNA和靶基因作用对的表达量相关性检验

收集的靶基因所包含的46个靶基因是用降解组测序验证的，可以作为真正的靶基因。而根据传统的理论，在以上提到的两个前提下，它们和其相应的miRNA之间应该总是有表达量负相关的关联。上节的图3-2结果显示，两者同一个miRNA家族的成熟miRNA表达量几乎相同，源自于探针几乎相同，而真正有区别的是pre-miRNA的表达量，所以此处我们对pre-miRNA和靶基因的表达量进行皮尔森相关性检验(Pearson correlation coefficient test)。

因为降解组测序实验所采用的样本是3周大的水稻幼苗，所以在进行相关性检验时我们选择其中27天大的水稻幼苗样本，选择的组织时叶片和叶鞘。

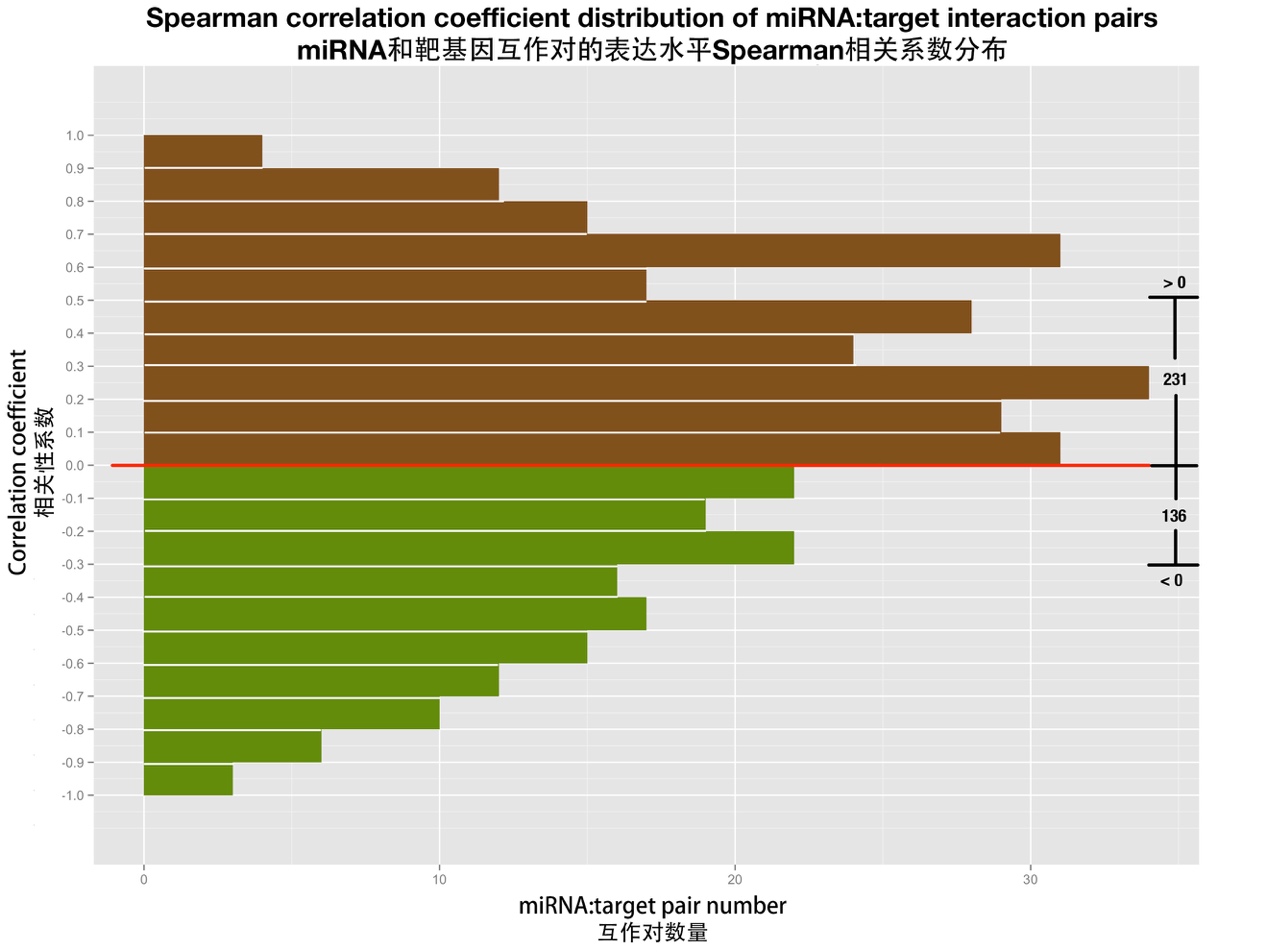


图3-3由降解组验证的miRNA:靶基因作用对之间Spearman相关性系数图

注：X坐标是落在特定相关系数区间的作用对数量，其中下方负相关的作用对的条形图用绿色标出，而上方正相关的作用对则用棕色标出。

Fig. 3-3 Heatmap of expression levels of mature miRNAs in various tissues/organs at various stages

Note: X-axis is the number of pairs that fall on the specified range of correlation coefficient; so bars with green color in the lower part denote the negatively correlated miRNA:target pairs, by contrast, bars with brown color in the upper part denote the positively correlated miRNA:target pairs.

从图3-3可以看出，367个miRNA和靶基因对中，只有136个呈现负相关，而原定的假设是真正的靶基因和其miRNA在相互作用的组织中总是呈现负相关，所得的结果并不支持原定的假设。不但如此，超过一半(197/367)的作用对之间是弱相关(-0.4~0.4)，这暗示在这些miRNA和靶基因对的表达量之间可能根本就不存在直接的相关性。

Wen et al.,2016也发现了类似的现象，在他们的研究中正相关的miRNA和靶基因作用对也是比负相关的要多。而这种现象可能是由更加复杂的作用机理比如负反馈环路 (negative feedback loops, FBLs) 和不合理正反馈环路 (incoherent feedforward loops, FFLs) 导致[[14]](#endnote-14)。

以上的结果也再次强调了植物miRNA介导的基因沉默还有很多位置的复杂因素未被研究清楚。同时也说明，表达量负相关的方法不能用以筛选具有生物相关性的miRNA和靶基因相互作用对。

## miRNA家族内不同miRNA和靶基因表达量相关性比较

水稻中，很多保守miRNA家族都有不止一个成员，成员之间的成熟miRNA的序列有很高的相似性，同时它们会有重叠度非常高的靶基因集合，而且通常都是来自于同一个基因家族。比如osa-miR156家族有12个成员，它们的靶基因则都是集中在SBP基因家族。对比osa-miR156家族中miRNA和SBP中靶基因的表达量相关性（仍然采用皮尔森相关性系数）。

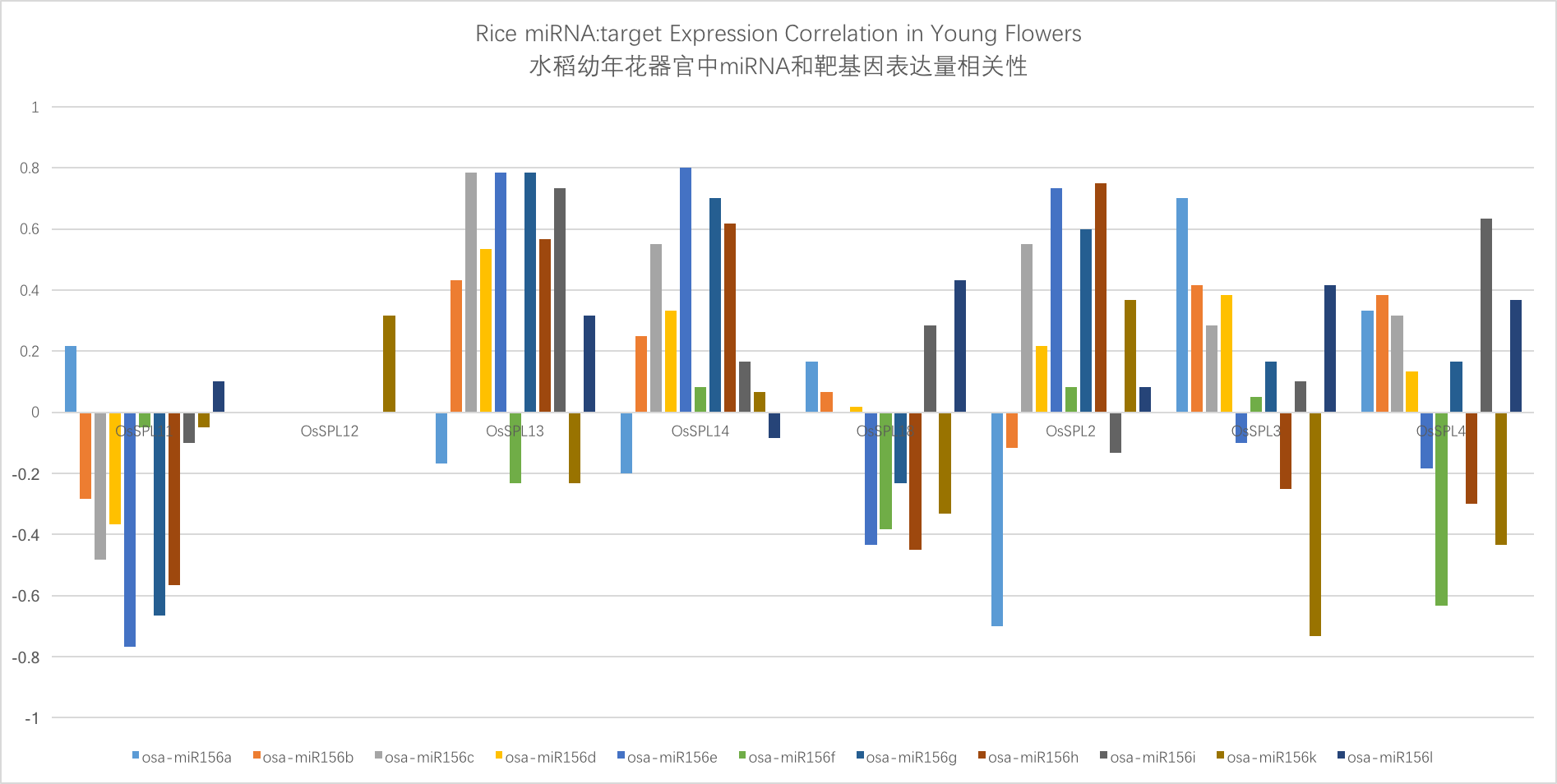


图3-4 osa-miR156家族和SBP基因家族之间表达量的皮尔森相关性系数\*

Fig. 3-4 Pearson correlation coefficient of expression levels of osa-miR156 family and its target gene family SBP

\*幼年花器官包括：0.6-1.0mm花序、3.0-4.0mm花序和5.0-10mm花序

对比miR156家族和同一个靶基因的相关性系数，可以看出相关性系数之间相差很大，这再次暗示了miRNA家族中有亚功能化(subfunctionalization)。

## 本章小节

本章尝试用miRNA和靶基因的表达量水平相关性来筛选出其中具有生物相关性的miRNA和靶基因作用对。但是在对经降解组验证的miRNA和靶基因对进行表达量水平相关性检验时，我们却发现它们之间并没有绝对的负相关性。因此，我们也不能用这种方法对miRNA和靶基因的生物相关性进行筛选，同时也对普遍所认为的miRNA和真正的靶基因之间的表达量具有负相关的观点提出质疑。

另外，对比miRNA家族的表达量和它们与相同靶基因之间的表达量相关性系数，可以发现miRNA家族成员的调节作用并不总是一样，反而呈现亚功能化。

# miRNA及其靶基因上结合位点的SNP获得和分析

## 本地SNP数据库的建立

最近的3K水稻基因组测序项目产生了23M SNP，如此大量的数据可以给水稻中的SNP研究提供很多帮助，而本研究中的所有的SNP都是来源于此(SNP-Seek database, http://snp-seek.irri.org/)。为了方便研究使用，我们将在线的SNP数据下载到本地，并且使用MySQL建立起本地的SNP数据库。

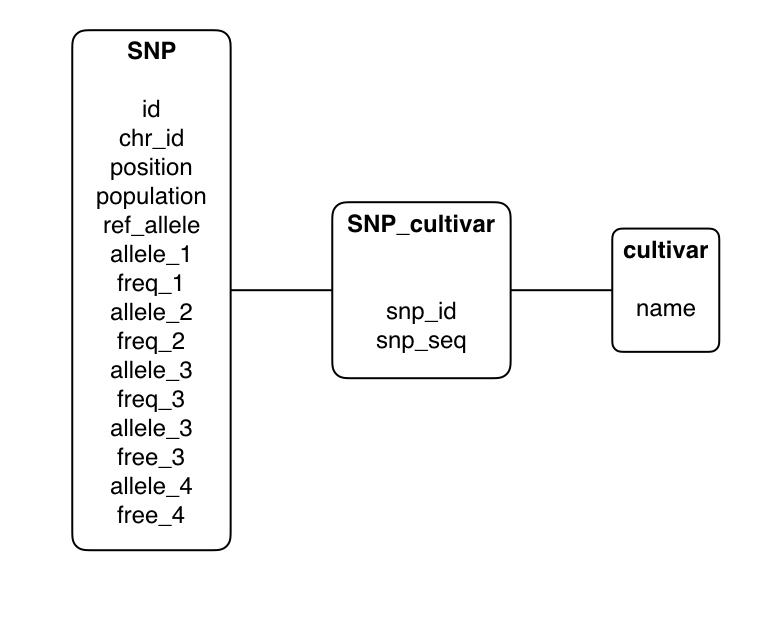


图4-1本地SNP数据库模式图

Fig.4- Basic schema of local SNP database

使用本地Python脚本，搜索在上述的pre-miRNA、成熟miRNA和靶基因上miRNA结合位点的SNP，我们得到大量的SNP数据。

表4-1 miRNA和靶基因上的SNP数量统计

Table 4-1 Summary of SNPs on miRNAs and targets

|  |  |  |
| --- | --- | --- |
| **Fragment type 基因组片段类型** | **Number 总数** | **Number of SNPs SNP的数量** |
| pre-miRNAs | 585 | 7193 |
| mature miRNAs 成熟miRNA | 793 | 1270 |
| conserved miRNAs 保守miRNA | 220 | 188 |
| target gene (miRNA binding site) 靶基因上miRNA结合位点 | 823 | 1169 |
| flanking region 结合位点侧翼 (~100bp上下游) | 823 | 9217 |

### SNP数据库之间的比较

RiceVarMap (Rice Variation Map, <http://ricevarmap.ncpgr.cn)>[[15]](#endnote-15)是在3K水稻基因测序项目之前的水稻基因组突变数据库，其中包括了从1479个水稻品系测序数据而得的6,551,358个SNP以及1,214,627个INDEL (insertions/deletions)。本实验初期也曾使用RiceVarMap进行分析。由于该数据库所采用的水稻参照基因组是Nipponbare, MSU 6.1，在SNP搜索中就需要进行参照基因组转换 (genome lift)。我们成功转换了560个pre-miRNA和665个成熟miRNA的基因组坐标，并且使用RiceVarMap进行SNP搜索。在pre-miRNA上总共得到1724个SNP，而其中323个SNP落在成熟miRNA上。

对比RiceVarMap和SNP-seek database的SNP结果，发现两者之间1670个SNP是共有的，而只有54个SNP是RiceVarMap所特有的，两者的重合率达到96.9%，所以可以看出SNP-seek database的SNP是具有很高的可信度。

另外，在本实验研究过程中，SNP-seek database经历过版本更新，水稻品系总数量从3,000增加到3,024，同时所得到的SNP总数量也增加了。旧版本中pre-miRNA上SNP数量是4617，成熟miRNA上SNP数量则是798。

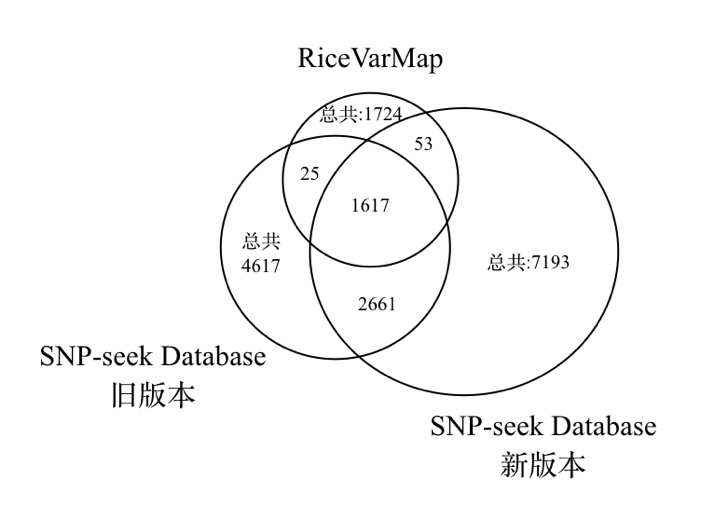


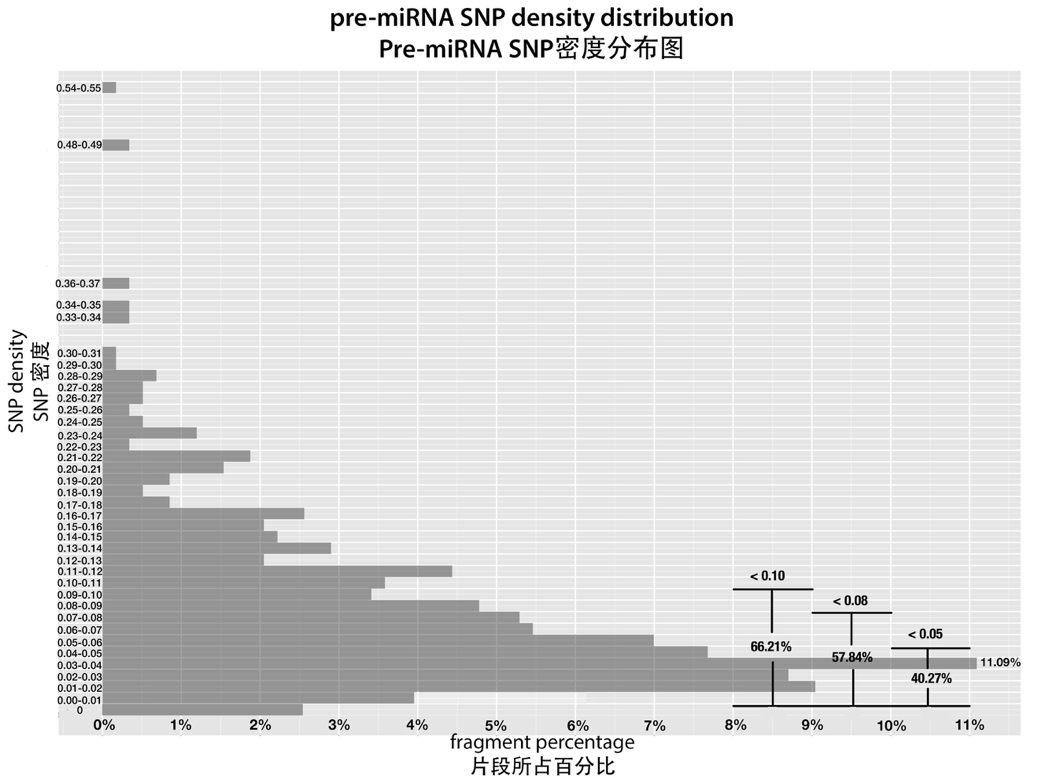
图4-2 RiceVarMap和SNP-seek database新旧版本中pre-miRNA上SNP数量

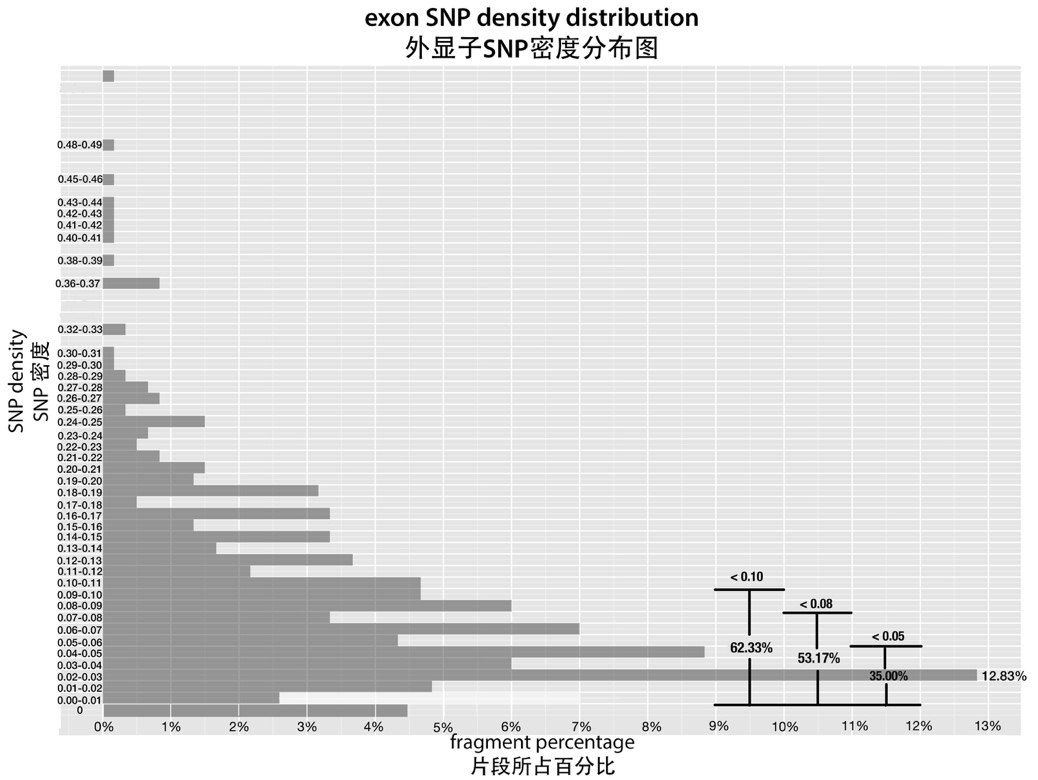
Fig.4-2 Number of SNPs on pre-miRNAs in RiceVarMap, and in both versions of SNP-seek database

## 水稻全部miRNA的SNP密度分析

SNP密度，也就是SNP数量除以相应基因组区段长度，能够反应该基因组区间的选择压力。其选择压力越大，则其上的SNP密度越低[[16]](#endnote-16)，反之亦然。因为miRNA是功能基因组元件并且是主要调节物，相比于基因间隔区和外显子，它们会经历不同的选择压力[[17]](#endnote-17)。为了研究这个问题，我们在水稻基因组中随机选择了长度为150碱基对长度的基因间隔区 (intergenic region)片段和外显子 (exon)片段作为对照，总共分别选择了600个片段，并且在本地的SNP数据库中对这些片段进行SNP搜索。

正如所预期的，pre-miRNA和外显子上的SNP密度比基因间隔区的更低。对于pre-miRNA，在SNP密度达到0.03-0.04范围之前，基因组片段的百分比随着SNP密度的增加而增加，而在超过0.04之后逐渐下降 (图4-2)。在本研究所取样的外显子也显示了类似的趋势，只是相应的基因组片段百分比最高值是在SNP密度为0.02-0.03区间达到 (图4-2)。然而，基因间隔区中则没有展现出类似的趋势 (图4-2)。通过分别比较三图中落在0-0.10、 0-0.08 和 0-0.05区间片段的百分比，可以很明显看出落在相应区间的pre-miRNA比例多余外显子，而外显子的比例则多余基因间隔区。这显示了相较于外显子和基因间隔区，pre-miRNA经过了更加严格的进化选择，而这结果也和miRNA是很多调节通路的主要调节因子的角色一致。





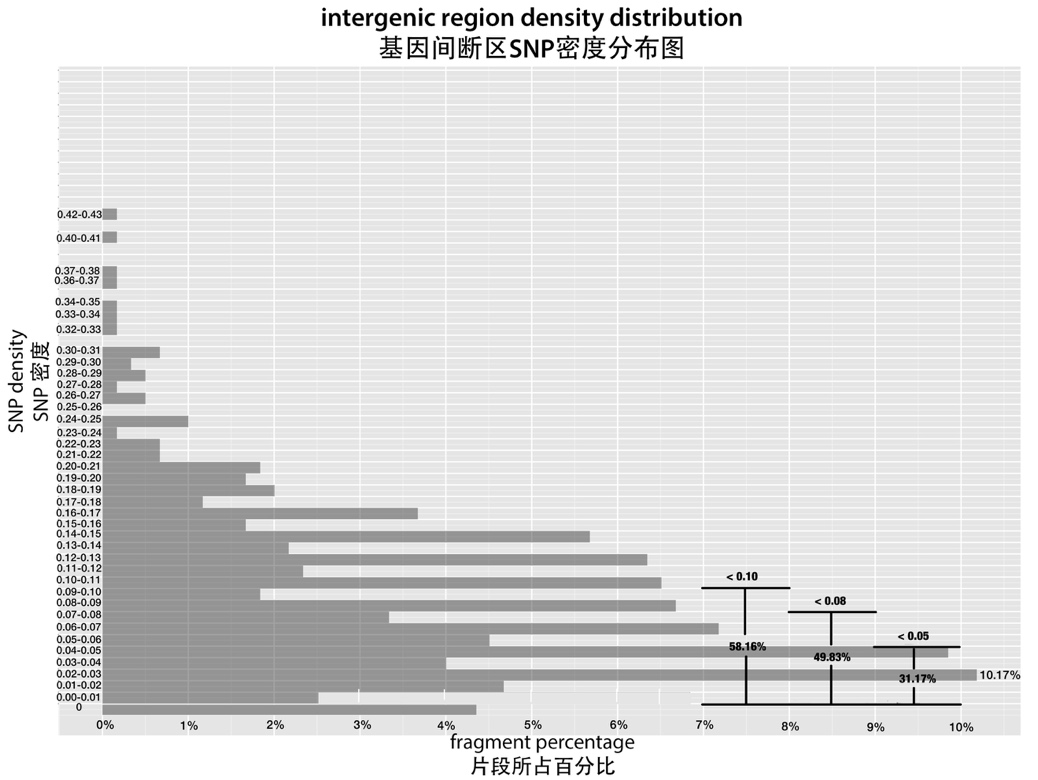


图4-2水稻pre-miRNA、外显子和基因间隔区的SNP密度

注: 其中，SNP密度是SNP的数量除以相应基因组区段的长度。而X坐标则是SNP密度落在相应区间的基因片段百分比

Fig. 4-2 SNP density of pre-miRNAs, exon regions and intergenic regions in rice.

Note: SNP density is the division of SNP number by the length of the genetic region. And x-axis corresponds to the percentage of fragments that have SNP density at the given range.

由于进化上的保守性不同3,[[18]](#endnote-18),[[19]](#endnote-19)，水稻中保守的miRNA上的SNP密度应该会比非保守的miRNA上的更低。在图4-3中，通过分别比较落在0-0.10, 0-0.08 和0-0.05区间的比例，可以发现相较于非保守的miRNA，大部分的保守miRNA都聚集在比较低的SNP密度区间中。结果和预期的一样。

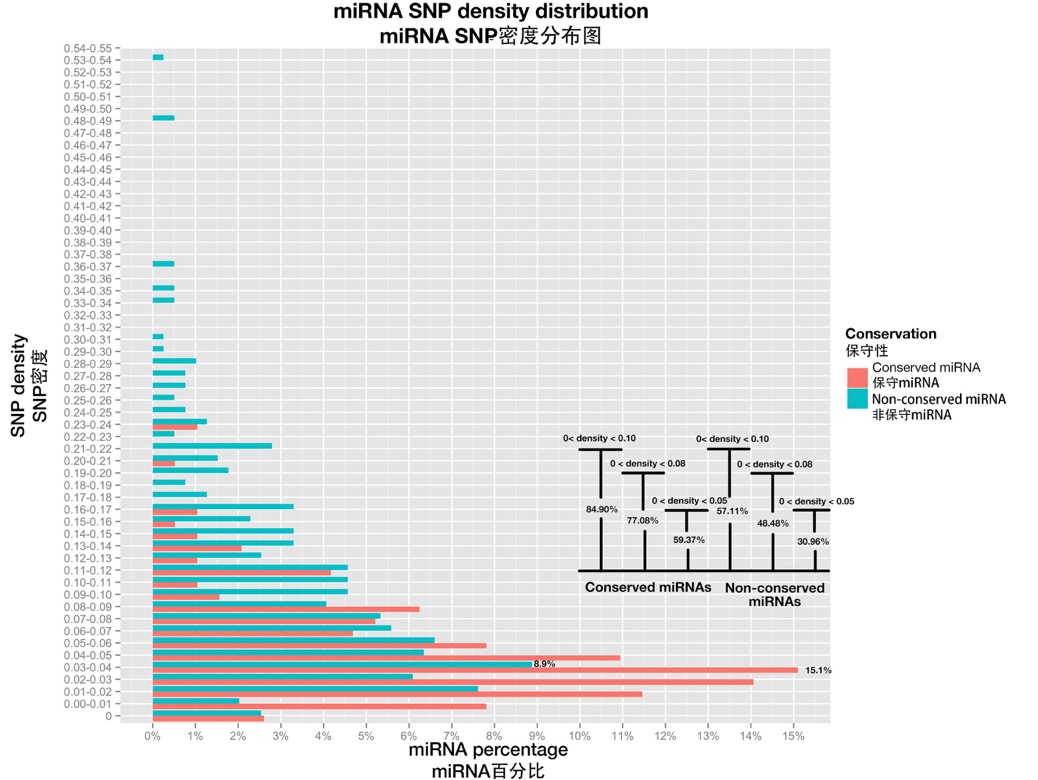


图4-3 pre-miRNA的SNP密度分布图，其中保守miRNA以红色标明，非保守miRNA则以蓝色标明

注: 右下方的条形图是SNP密度分别低于0-0.10, 0-0.08 和0-0.05的miRNA比例。

Fig. 4-3 miRNA SNP density distribution of pre-miRNAs, both conserved miRNAs (in red color) and non-conserved miRNAs (blue color).

Note: Bar plot on the bottom right shows the percentage of miRNAs whose SNP densities are below 0.10, 0.08 and 0.05, for conserved and non-conserved miRNAs respectively.

## 成熟的保守miRNA和非保守miRNA上各位点SNP频率

植物miRNA对不同的靶标基因其调节效率有差异，而这差异归因于它们的互补模式的不同，当然这互补模式是由成熟miRNA和所对应的靶标位点共同决定的。此外，已有证据指明成熟miRNA的21个碱基位点在基因识别和沉默中所起的作用是有差别的8，因为在一些位点上的突变会导致基因沉默调节实效而其它位点上的则不会有明显的影响 17,[[20]](#endnote-20),[[21]](#endnote-21)。这种差异也可以用每个位点的SNP频率的差异性来反应出来（本研究中某位点的SNP频率定义为在该位点出现SNP的miRNA占总的miRNA数量的比例）。更低的SNP频率暗示此位点经受了更大的选择压力。因此，对于成熟miRNA每一个位点进行全体的SNP频率分析可能进一步揭示它们在沉默中所起到的差异性作用。

因为大多数的miRNA都是21 nt的长度，所以本研究主要关注1-21位点上的分析。通过分别计算和比较保守miRNA和非保守miRNA每个位点的SNP频率，我们发现，结果如同预期：保守miRNA上的每一个位点的SNP频率都比非保守的要低。尽管如此，在理论上假如保守miRNA和非保守miRNA都处在相似的进化压力的话，两者在不同位点的SNP频率应该会有相似的秩 (ranking)，并且这相似的秩会导致保守miRNA和非保守miRNA各位点的SNP频率具有线性或者近似线性的关系，这线性关系可以用皮尔森相关性进行检验。然而结果并非和预期相同。位点20是非保守miRNA上SNP频率最高的位点，但在保守miRNA中却是SNP频率第四低的位点，位点12在保守miRNA上是SNP频率最低的位点，但在非保守miRNA上却是频率第二高的位点。同时，皮尔森相关性检验发现两者的各位点SNP频率之间并没有显著的相关性(r=-0.163, p-value=0.4473)。唯一的例外是，两者都在位点一有最低的SNP频率，而这个可以由位点一在miRNA载入AGO蛋白起到重要作用解释[[22]](#endnote-22),[[23]](#endnote-23)。总结来说，保守miRNA和非保守miRNA上位点SNP频率的不同的秩揭示了两者在各位点经历了不同的选择压力，暗示它们可能使用不同的沉默组份 (silencing component) 来调节靶基因。

通常认为在位点10和11的配对对于植物miRNA的mRNA剪切有非常重要的作用[[24]](#endnote-24),[[25]](#endnote-25),[[26]](#endnote-26)，而这个也对这两个位点在进化上加了一层的限制，并且可能导致它们的SNP频率比其它的位点都要低。然而，我们的结果显示这两个位点在保守miRNA和非保守miRNA中都不是SNP频率最低的位点，这和经验结论并不相符，进而对中心位点（10位和11位）完美配对对植物miRNA介导的沉默必要性的观点提出质疑。

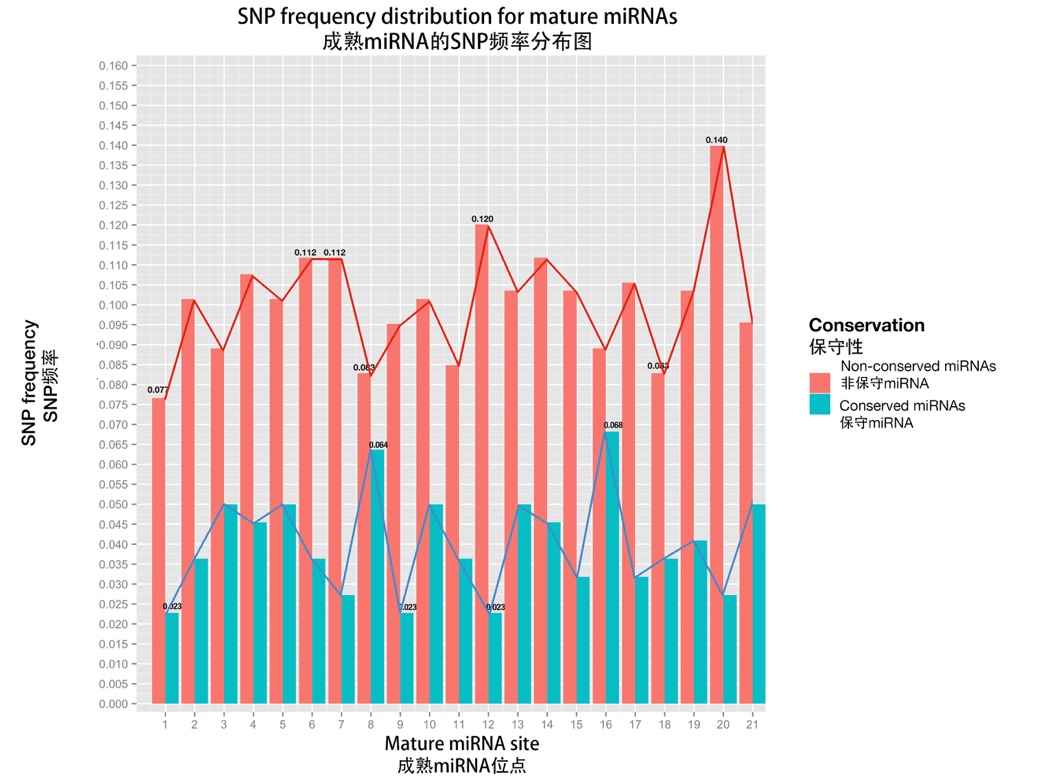


图4-4成熟miRNA每个位点的SNP频率，其中保守miRNA是蓝色而非保守的则是红色

注: X坐标是以成熟miRNA 5’到3’的顺序排列，而Y坐标则是SNP频率。

Fig. 4-4 Positional SNP distribution of conserved miRNAs (blue) and non-conserved miRNAs (red).

Note: X-axis is the sites in mature miRNA from 5’end - 3’end, and y-axis is SNP frequency which is calculated as number of SNPs at this site divided by number of miRNAs;

## 保守miRNA和其靶基因保守位点上SNP频率

下一步，我们尝试分析靶基因上miRNA结合位点上的SNP频率。由于miRNA和靶基因表达量相关性检验发现，它们之间并没有存在必然的负相关，我们就不能使用这种负相关对靶基因进行生物相关性筛选。为此，所有收集而得的823个保守miRNA靶基因都进行统计和研究。

已有研究报导过miRNA和其结合位点存在共同进化23,[[27]](#endnote-27)。两者之间在自然选择的过程中互相影响，不但如此，两者的序列对miRNA的调节作用效率都起到重要的影响。所以，我们推测成熟miRNA喝结合位点上各位点的选择压力相似。而这种选择压力的相似性可以表现在两者各位点SNP频率的相关性上。图4-5显示成熟miRNA和结合位点之间SNP频率分布之间的确有相似性，位点1在miRNA结合位点上是SNP频率第二低的位点，同时在成熟miRNA上是频率最低的位点；而位点13在miRNA结合位点上是频率最高的位点，在成熟miRNA上则是频率第三高的位点。通过皮尔森相关性检验，也发现两者之间有显著的正相关 (r=0.5891, p-value=2.455e-3) 。所得的结果也支持了关于成熟miRNA和miRNA结合位点之间有共同进化关系的假设。

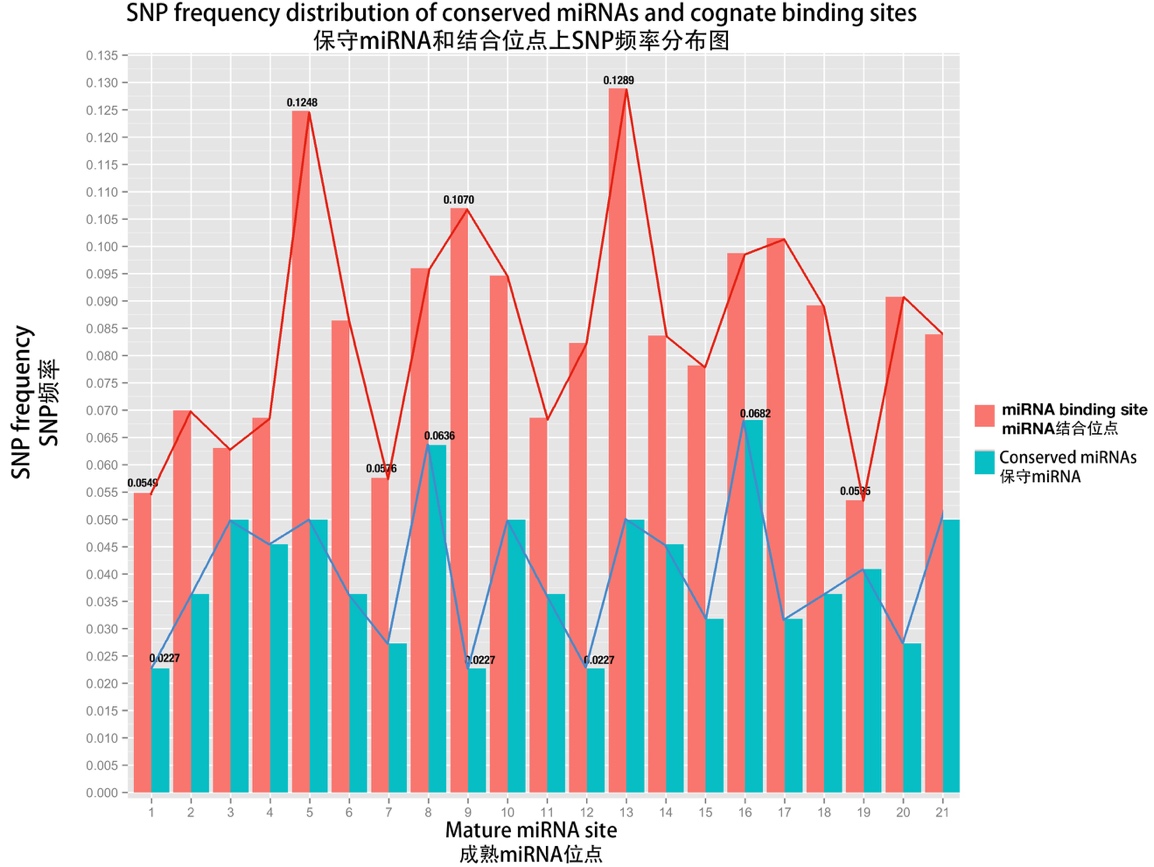


图4-5成熟miRNA和其结合位点上每一个位置的SNP频率分布

注: 结合位点上位置的排列顺序和成熟miRNA的顺序相同，都是按照成熟miRNA 5’到3’的顺序排列

Fig. 4-5 The sites of miRNA binding site are placed in the same order as mature miRNAs (from 5’end to 3’end in the miRNA);

## 本章小节

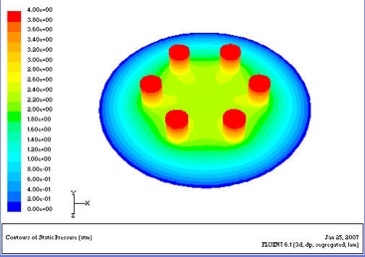
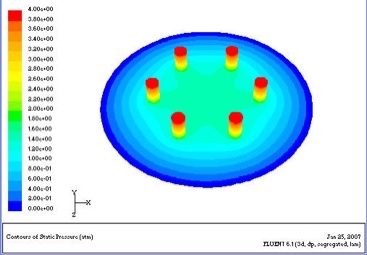
# 联合互补模式分析

# 单倍体表型分析和对比



图2-2内热源沿径向的分布

Fig.2-2 Energy distribution along radial



b)**=2.5mm时轴承的压力分布云图

b) Pressure contour of bearing when

**=2.5mm

a)**=1.5mm时轴承的压力分布云图

a) Pressure contour of bearing when

**=1.5mm

图2-3图中包含子图的格式范例

Fig.2-3 Example of …

（五号，单倍行距，此处空一行）

表2-1高频感应加热的基本参数

Table 2-1 The parameters of ...（段后0.5倍行距）

|  |  |  |  |
| --- | --- | --- | --- |
| 感应频率  （KHz） | 感应发生器功率  (%×80Kw) | 工件移动速度  (mm/min) | 感应圈与零件间隙  (mm) |
| 250 | 88 | 5900 | 1.65 |

表名中不允许使用标点符号，表名后不加标点。表题置于表上，硕士学位论文可以用中、英文两种文字居中排写，中文在上，也可以只用中文，表名与正文之间空一行。数字空缺的格内加横线“－”（占2个数字宽度）。

表内文字或数字上、下或左、右相同时，采用通栏处理方式（合并单元格），不允许用“〃”、“同上”之类的写法。

表内文字说明，起行空一格、转行顶格、句末不加标点。

如某个表需要转页接排，在随后的各页上应重复表的编号。编号后跟表题（可省略）和“（续）”，如所“表1-1 xxxx（续）” 或“表1-1（续）”，续表均应重复表头和关于单位的陈述。编号后加“（续表）”，表题可省略。续表应重复表头。

表格不加左、右边线。表的编排建议采用国际通行的三线表（三线表，以其形式简洁、功能分明、阅读方便而在科技论文中被推荐使用。三线表通常只有3条线，即顶线、底线和栏目线，没有竖线）。表中文字用宋体5号字。表中单元格的间距合适，紧促美观。

## 公式格式

 （2-1）

公式的序号右端对齐。公式后应注明编号，按章顺序编排。公式中字符大小合适，基本字符为5号字体，不宜较大。

公式应另起一行居中排，公式较长时最好在等号“＝”处转行，如难实现，则可在＋、－、×、÷运算符号处转行，转行时运算符号仅书写于转行式前，不重复书写。

公式中第一次出现的物理量代号应给予注释，注释的转行应与破折号“——”后第一个字对齐。破折号占二个字，注释物理量需用公式表示时，公式后不应出现公式序号，如（3-1）。公式下面的“式中”空两个字起排，单独占一行。公式中所要解释的符号按先左后右，先上后下顺序分行空两个字排，再用破折号与释文连接，回行时与上一行释文对齐。上下行的破折号对齐

式中└┘└┘———试样断裂前的最大扭矩()；

———试样断裂时的单位长度上的相对扭转角，

公式中应注意分数线的长短（主、副分数线严格区分），长分数线与等号对齐，如：



## 引用说明

正文中引用文献的标示应置于所引内容最后一个字的右上角，所引文献编号用阿拉伯数字置于方括号“[ ]”中，用小4号字体的上角标，引用单篇文献时如“二次铣削[1]”；引用两篇文献时如“原位生成的TiB主要有针状或晶须状[21, 22]”；引用多篇文献时如“蠕变断裂以沿晶断裂为主[5-7]”。当提及的参考文献为文中直接说明时，则用小4号字与正文排齐，如“由文献[8,10-13]可知”。

不得将引用文献标示置于各级标题处。

## 本章小结

本章介绍了……

# 其他格式要求

## 页码

页码在版芯下边线之下隔行居中放置；摘要、目录、物理量名称及符号表等文前部分的页码用罗马数字单独编排，正文以后的页码用阿拉伯数字编排。

## 页眉

页眉居中为“上海交通大学硕士学位论文”。

字体采用小五号。

边框采用双横线，粗线在上，细线在下，粗细为2.25磅。

中英文封面没有页眉，也没有边框。

## 目录

目录、图录、表录二字采用黑体3号字，段前0.7厘米，段后0厘米。“目录”与所列条目之间空两行。

目录应包括论文中全部章、节、条三级标题及其页号，含：

正文章节题目（要求编到第3级标题，即×.×.×。一级标题顶格书写，二级标题缩进一格，三级标题缩进两格。）

参考文献

附录

致谢

攻读硕士学位期间发表或录用的论文

目录中各章题序及标题用黑体小4号字。其余用宋体小4号字，段前段后均为0。所有标题1.25倍行距。

图录与表录得条目均使用宋体小4号字，段前段后均为0，1.25倍行距。

## 正文的层次安排

正文层次的编排建议用表3-1所示格式。

表3-1 论文的层次代号与说明

|  |  |  |
| --- | --- | --- |
| 层次名称 | 示 例 | 说 明 |
| 章 | 第一章└┘□□……□ | 章序及章名居中排 |
| 节 | 1.1└┘□□……□ | 题序顶格书写,与标题间空一格（注意标题格式是否已经默认加入一个空格）,阐述内容另起一段 |
| 条 | 1.1.1└┘□□……□ |
| ↑　　　　　　　　　　　　　　　　　　　　↑  　　　版心左边线　　　　　　　　　　　　　 　　版心右边线 | | |

各层次题序及标题不得置于页面的最后一行（孤行）。

## 打印要求

### 页面设置

页边距：上3.5厘米，下4厘米，左右均为2.8厘米，装订线靠左0.5厘米位置。页眉：2.5厘米。页脚：3厘米。

无网格。

### 字体

论文中的中文字体除了图表题注为楷体\_GB312外，其余全用宋体；英文字体则要求为Times New Roman。如果英文与数字夹杂出现在黑体中文中，则将英文与数字作为Times New Roman字体再加粗处理。

### 字号

1、目录题目（目录、图录、表录）——黑体三号，居中，段前段后0.7cm，单倍行距；

2、章标题（第x章）——黑体三号，居中，段前段后0.7cm，单倍行距；

3、节标题（x.x）——黑体四号，左对齐，段前段后0.7cm，单倍行距；

4、条标题（x.x.x）——黑体小四号，左对齐，段前段后0.3cm，单倍行距；

5、正文——宋体小四号，，首行缩进2字符，1.25倍行距；

6、正文后的题目（参考文献、致谢、攻读xx期间发表的论文）——黑体三号，居中，段前0.7cm，段后0，单倍行距。

# 结束语

## 主要工作与创新点

本文主要……

## 后续研究工作

更深入的研究……

参 考 文 献

↑

（黑体3号字居中，段前0.7厘米，段后0厘米，单倍行距，与参考文献内容之间空两行）

[1] 杨瑞林, 李力军. 新型低合金高强韧性耐磨钢的研究. 钢铁. 1999（7）: 41~45.

[2] Schinstock, D.E., Cuttino, J.F. Real time kinematic solutions of a non-contacting, three dimensional metrology frame[J]. Precision Engineering. 2000, 24(1): 70-76.

[3] 温诗铸. 摩擦学原理. 北京: 清华大学出版社. 1990: 296-300.

[4] 贾名字. 工程硕士论文撰写规范[硕士论文].上海: 上海交通大学. 2000.

↑

（参考文献内容小四号宋体，1.25倍行距，[标号]与作者姓名之间空一格，换行内容与作者姓名的第一个字母对齐。）

要求：

1、所有被引用文献均要列入参考文献中，必须按顺序标注，但同一篇文章只用一个序号。

2、博士学位论文的参考文献数一般不少于100篇，硕士学位论文的参考文献一般不少于40篇，其中外文文献一般不少于总数的1/2。参考文献中近五年的文献数一般应不少于总数的1/3，并应有近两年的参考文献。专业硕士学位论文的参考文献一般不少于20篇，其中外文文献一般不少于总数的1/2。参考文献中近五年的文献数一般应不少于总数的1/3，并应有近两年的参考文献。

3、教材、产品说明书、未公开发表的研究报告（著名的内部报告如PB、AD报告及著名大公司的企业技术报告等除外）等通常不宜作为参考文献引用。

4、引用网上参考文献时，应注明该文献的准确网页地址，网上参考文献和各类标准不包含在上述规定的文献数量之内。

5、本人在攻读本学位期间发表的论文不应列入参考文献。

6、序号应按文献在论文中的被引用顺序编排。换行时与作者名第一个字对齐。若同一文献中有多处被引用，则要写出相应引用页码，各起止页码间空一格，排列按引用顺序，不按页码顺序。

7、示例：

①期刊：[序号] 作者，题名，刊名，出版年份，卷号（期号），起止页码

②专著：[序号] 作者，书名，版本（第1版不标注），出版地，出版者，出版年，起止页码

③论文集：[序号] 作者，题名，见（英文用In），主编，论文集名，出版地，出版年，起止页码

④学位论文：[序号] 作者，题名，［学位论文］（英文用［Dissertation］），保存地点，保存单位，年份

⑤专利：[序号] 专利申请者，题名，国别，专利文献种类，专利号，出版日期

⑥技术标准：[序号] 起草责任者，标准代号，标准顺序号－发布年，标准名称，出版地，出版者，出版年度

附录1

致 谢

↑

（黑体3号字居中，段前0.7厘米，段后0厘米，单倍行距，与致谢内容之间空两行）

本文需要感谢……致谢内容：宋体，小四号（“论文正文”样式）。

攻读硕士学位期间已发表或录用的论文

↑

（黑体3号字居中，段前0.7厘米，段后0厘米，单倍行距，与内容之间空两行）

[1] 张三，李四. 已经发表一篇学术论文. XXXXXXX学报 （已录用）

（采用“参考文献内容”样式）

1. Kozomara A, Griffiths-Jones S., miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014, 42:D68-D73 [↑](#endnote-ref-1)
2. Lu, C., et al. Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). Proc. Natl. Acad. Sci. USA 2008, 105: 4951–4956. [↑](#endnote-ref-2)
3. Jones-Rhoades M.W., Conservation and divergence in plant microRNAs. Plant Mol Biol. 2011, 80:3–16 [↑](#endnote-ref-3)
4. Li, J.Y. et al. 2014, The functional scope of plant microRNA-mediated silencing. Trends Plant Sci. 2014, 19:785-756. [↑](#endnote-ref-4)
5. Dai X., Zhao P.X., psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. 2011, 39:W155-W159 [↑](#endnote-ref-5)
6. Li Y.F., Zheng Y., Addo-Quaye C., et al., Transcriptome-wide identification of microRNA targets in rice. Plant J. 2010, 62:742-759 [↑](#endnote-ref-6)
7. Kertesz M., Iovino N., Unnerstall U., et al., The role of site accessibility in microRNA target recognition. Nat. Genet., 2007, 39:1278-1284. [↑](#endnote-ref-7)
8. Liu Q, Wang F, Axtell M.J., Analysis of complementarity requirements for plant microRNA targeting using a Nicotiana benthamiana quantitative transient assay. Plant Cell, 2014, 26: 741-753 [↑](#endnote-ref-8)
9. Iwakawa H., Tomari Y., Molecular insights into microRNA-mediated translational repression in plants. Mol. Cell, 2013, 52:591-601 [↑](#endnote-ref-9)
10. Tang G., Reinhart B.J., Bartel D.P., et al., A biochemical framework for RNA silencing in plants. Genes Dev. 2003, 17:49-63 [↑](#endnote-ref-10)
11. Palatnik J.F., Allen E., Wu X., Schommer C., et al., Control of leaf morphogenesis by microRNAs. Nature, 2003 425:257–63 [↑](#endnote-ref-11)
12. Sato Y., Namiki N., Takehisa H., et al., RiceFREND: a platform for retrieving coexpressed gene networks in rice. Nucleic Acids Research, 2013, 41:D1214-D1221 [↑](#endnote-ref-12)
13. Sturn A, Quackenbush J, Trajanoski Z. Genesis: Cluster analysis of microarray data., Bioinformatics. 2002,18(1):207-8. [↑](#endnote-ref-13)
14. Wen M., Xie M.N., He L., Wang Y.S., Shi S.H., Tang T., 2016, Expression Variations of miRNAs and mRNAs in Rice Oryza sativa. Genome Biology and Evolution 8:3529-3544 [↑](#endnote-ref-14)
15. Zhao H., Yao W., Ouyang Y., et al., RiceVarMap: a comprehensive database of rice genomic variations. Nucleic Acids Research, 2014, 43:D1018–D1022. [↑](#endnote-ref-15)
16. Chen K., Rajewsky N., Natural selection on human miRNA binding sites inferred from SNP data. Nature Genet., 2006, 38:1452–1456 [↑](#endnote-ref-16)
17. Saunders M. A., Liang H., Li, W. H., Human polymorphism at microRNAs and microRNA target sites. Proc. Natl Acad. Sci. USA, 2007 104, 3300–3305 [↑](#endnote-ref-17)
18. Fahlgren N., Jogdeo S., Kasschau K.D., et al., MicroRNA gene evolution in Arabidopsis lyrata and Arabidopsis thaliana. Plant Cell, 2010, 224:1074–1089 [↑](#endnote-ref-18)
19. Rajagopalan R., Vaucheret H., Trejo J., Bartel D.P., A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes Dev., 2006, 2024:3407–3425 [↑](#endnote-ref-19)
20. Mallory A.C., Reinhart B.J., Jones-Rhoades M.W., et al., MicroRNA control of PHABULOSA in leaf development: Importance of pairing to the microRNA 59 region. EMBO J., 2004, 23: 3356-3364. [↑](#endnote-ref-20)
21. Parizotto E.A., Dunoyer P., Rahm N., et al., In vivo investigation of the transcription, processing, endonucleolytic activity, and functional relevance of the spatial distribution of a plant miRNA. Genes Dev., 2004, 18: 2237-2242 [↑](#endnote-ref-21)
22. Mi S.J., Cai T., Hu Y.G., et al., Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5’ terminal nucleotide. Cell, 2008, 133: 116-127 [↑](#endnote-ref-22)
23. Mallory A., Vaucheret H., Form, function, and regulation of ARGONAUTE proteins. Plant Cell, 2010, 22: 3879–3889 [↑](#endnote-ref-23)
24. Schwab R., Palatnik J.F., Riester M., et al., Specific effects of microRNAs on the plant transcriptome. Dev. Cell, 2005, 8: 517–527 [↑](#endnote-ref-24)
25. Franco-Zorrilla J.M., Valli A., Todesco M., et al., Target mimicry provides a new mechanism for regulation of microRNA activity. Nat. Genet., 2007, 39:1033-1037 [↑](#endnote-ref-25)
26. Todesco M., Rubio-Somoza I., Paz-Ares J., et al., A collection of target mimics for comprehensive analysis of microRNA function in Arabidopsis thaliana. PLoS Genet., 2010, 6: e1001031 [↑](#endnote-ref-26)
27. Arikit S., Zhai J., Meyers B.C., Biogenesis and function of rice small RNAs from non-coding RNA precursors. Curr. Opin. Plant Biol., 2013, 162:170–179. [↑](#endnote-ref-27)