**水稻中microRNA介导的基因沉默相关SNP的研究**

**Investigation of SNPs involved in rice miRNA-mediated gene silencing**

**背景**

单核苷酸多态性(single nucleotide polymorphisms, SNPs), 就是群体的DNA序列发生的单一碱基变化 [[[1]](#endnote-1)]，因为拥有在大多数物种中数量多、分布广泛并且能够被用于高通量和超高通量自动测序等特点而迅速成为植物分子遗传学非常普及的分子标记 [[[2]](#endnote-2)-[[3]](#endnote-3)]。SNP也给分子标记辅助育种、数量性状位点分析和基因组关联分析提供了强有力的工具 [3-[[4]](#endnote-4)]。二代测序技术的应用(next-generation sequencing, NGS)通过重测序在不少物种中帮助鉴定出大量SNP，包括人[[[5]](#endnote-5)],，水稻[[[6]](#endnote-6)-[[7]](#endnote-7)[[8]](#endnote-8)]，玉米 [[[9]](#endnote-9)-[[10]](#endnote-10)]，大豆 [[[11]](#endnote-11)]，拟南芥[[[12]](#endnote-12)]等。全基因组分析发现SNP在不同的基因组区间的分布并不均匀， 在保守性比较高的区段比如编码区和调控因子等功能区段则会有较少的SNP [[[13]](#endnote-13)-[[14]](#endnote-14)]。

近些年有几个研究运用重测序进行SNP的鉴定和分析，除了研究蛋白编码基因中的SNP之外，也有开展一些针对水稻[[[15]](#endnote-15), [[16]](#endnote-16)]和拟南芥[[[17]](#endnote-17)]的miRNA (microRNA)相关的SNP研究。这些研究主要关注的是SNP对miRNA结构稳定性和靶基因的改变以及miRNA进化的影响。

miRNA是一类源自内源位点的非编码小RNA。这些miRNA基因会转录成能够自我配对并且形成“发夹”结构的的初级转录本pri-miRNA (primary RNA) [[[18]](#endnote-18)]，然后pri-miRNA会被DCL1剪切从而释放出前体RNA (pre-miRNA) [[[19]](#endnote-19)]。接下来，pre-miRNA会产生一个大约21 nt的二聚体RNA，而该二聚体会分离，其中 miRNA链将装入AGO蛋白形成RNA诱导沉默复合体 (RNA-induced silencing complex, RISC)，另一条miRNA\*链则会降解 [[[20]](#endnote-20)]。其中装入AGO蛋白的单链被称为成熟miRNA，是通过互补性以引导RISC识别到靶基因的miRNA结合位点。miRNA是植物生长发育过程中非常重要的调节物而且往往会调控那些本身就是调节物的靶基因，例如转录因子。已经有研究报导在miRNA介导基因沉默中相关的SNP会引起农艺性状变异。在水稻中，osa-miR156调控的OsSPL14的结合位点中出现一个SNP干扰了miRNA的调控，进而导致了一个分孽数减少、抗倒伏能力增强并且产量提升的“理想”植株 [[[21]](#endnote-21)]。而在大麦中，SNP干扰了miR172和其靶基因HvAP2的相互作用从而引进了大麦花序中穗状花序密度的变异 [[[22]](#endnote-22)]。

在植物中，miRNA主要以剪切转录本的方式抑制其具有高度互补性的靶基因 [[[23]](#endnote-23)]，而这种高度互补性则称为如今很多预测miRNA靶基因的生信软件的基础，其中使用比较广泛的是网页版的 psRNATarget (Plant Small RNA Target Analysis Server) [[[24]](#endnote-24)]. 除了生物信息方法，近年来，研究人员开发了一些方法验证真正的靶基因，比如过表达miRNA或者抗miRNA的靶基因，5’-RACE和降解组测序等 [[[25]](#endnote-25)]。植物miRNA对靶基因识别有很复杂的机制，所以生物信息的预测方式可能会产生并不能被相关miRNA作用的基因 [25]，因此如何筛选掉其中的假阳性结果仍有待研究。

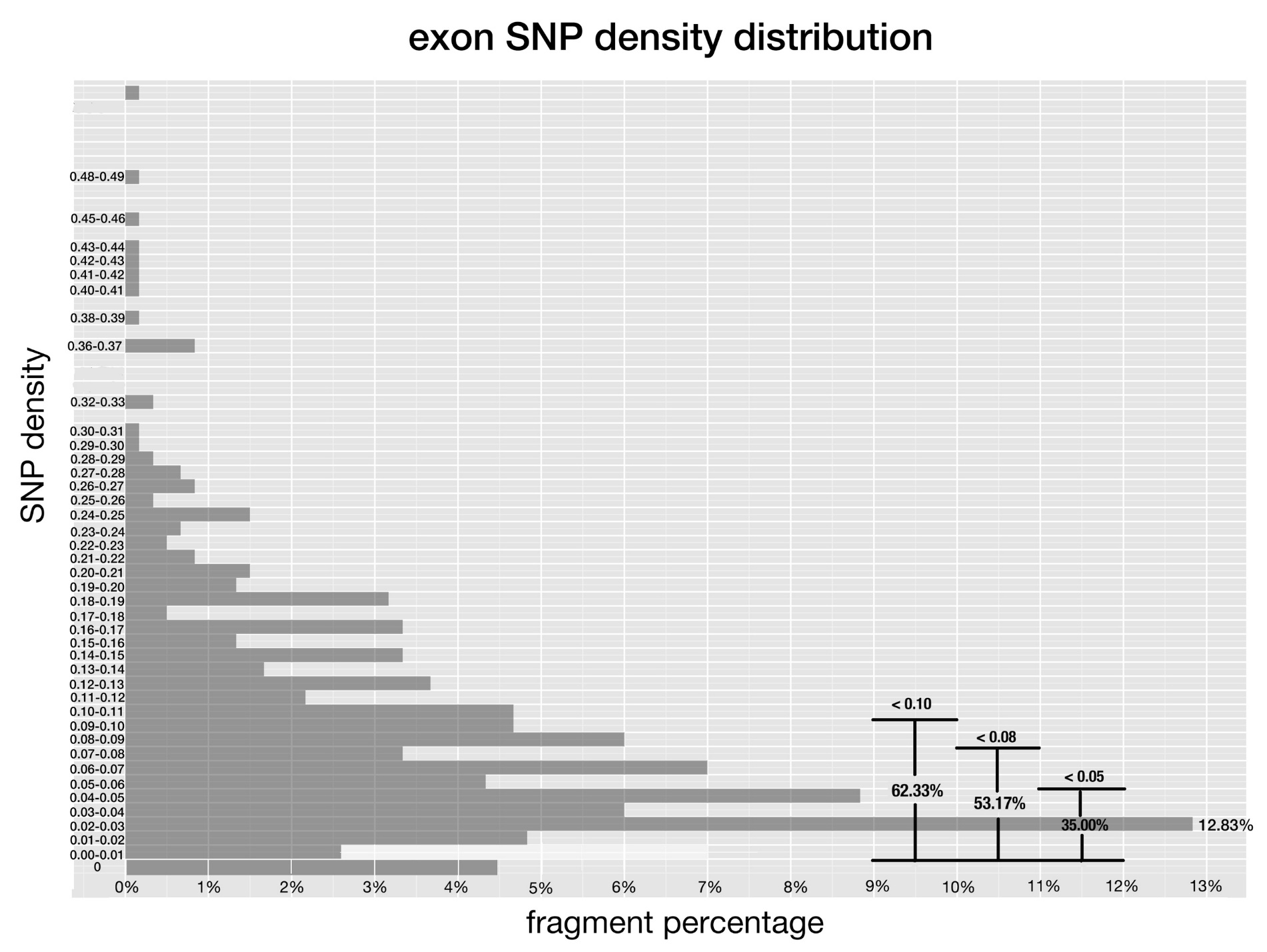
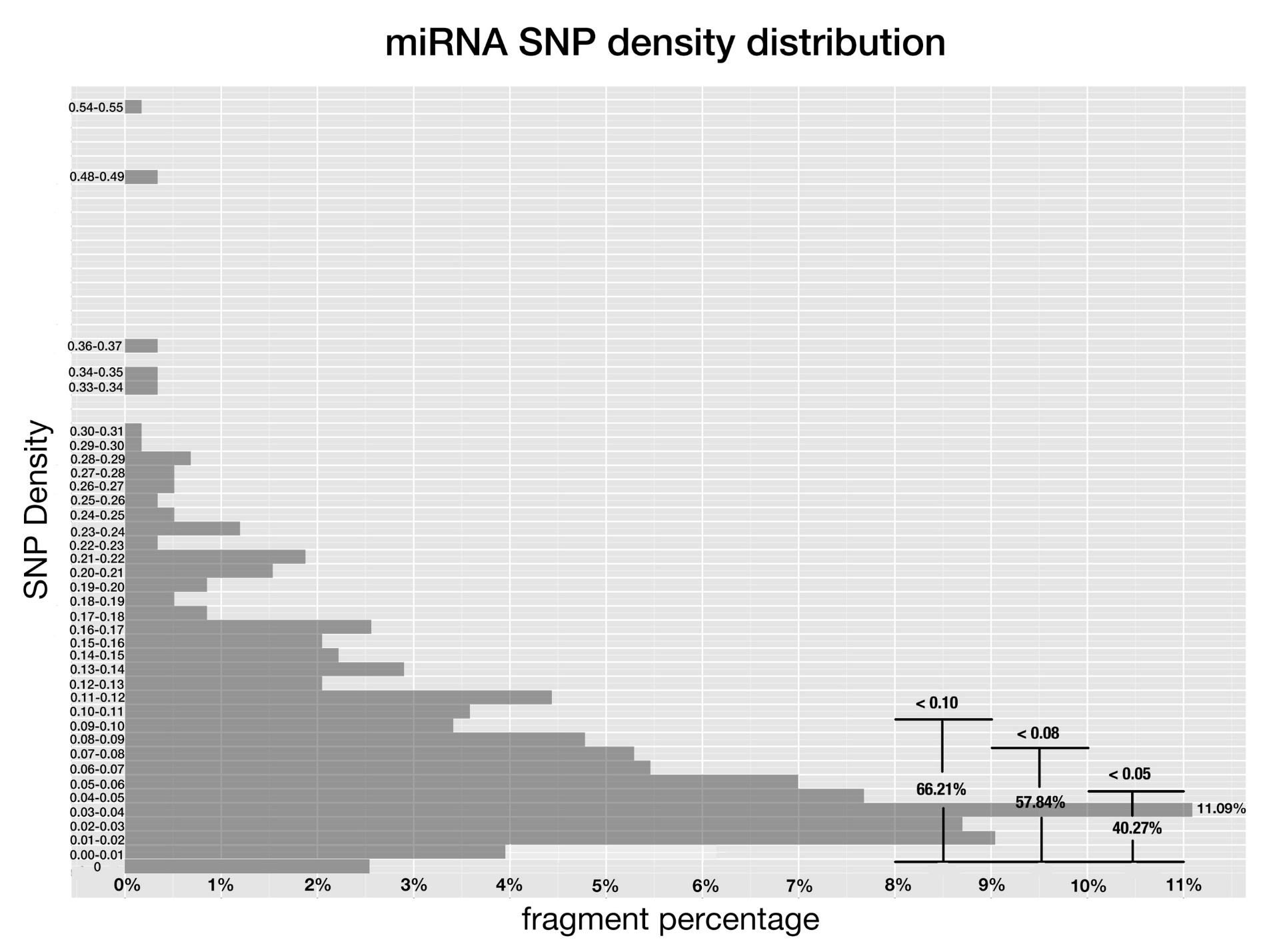
最近，3K水稻基因组项目对3 000多种水稻品种进行测序并且产生了上百万基因组片段 [[[26]](#endnote-26)]。同时通过比对这些序列鉴定出来数量极其众多的SNP [8]，其中包括一些稀有的三个甚至四个等位基因的SNP。如此丰富的SNP资源正好为本研究的全基因组miRNA介导调节相关SNP的研究提供助力。因为SNP能够反应不同水稻品系基因组的变异，所以SNP可以用以研究水稻品系间miRNA介导的调节过程变异和该变异对水稻品系表型的影响。基于3K水稻项目的SNP，我们研究了在miRBase[[[27]](#endnote-27)]收集的水稻miRNA和预测的靶基因上的SNP分布以及SNP对水稻中miRNA和靶基因互作的可能影响。进一步对miRNA和靶基因互作变异与相应表型差异的关系进行了分析。

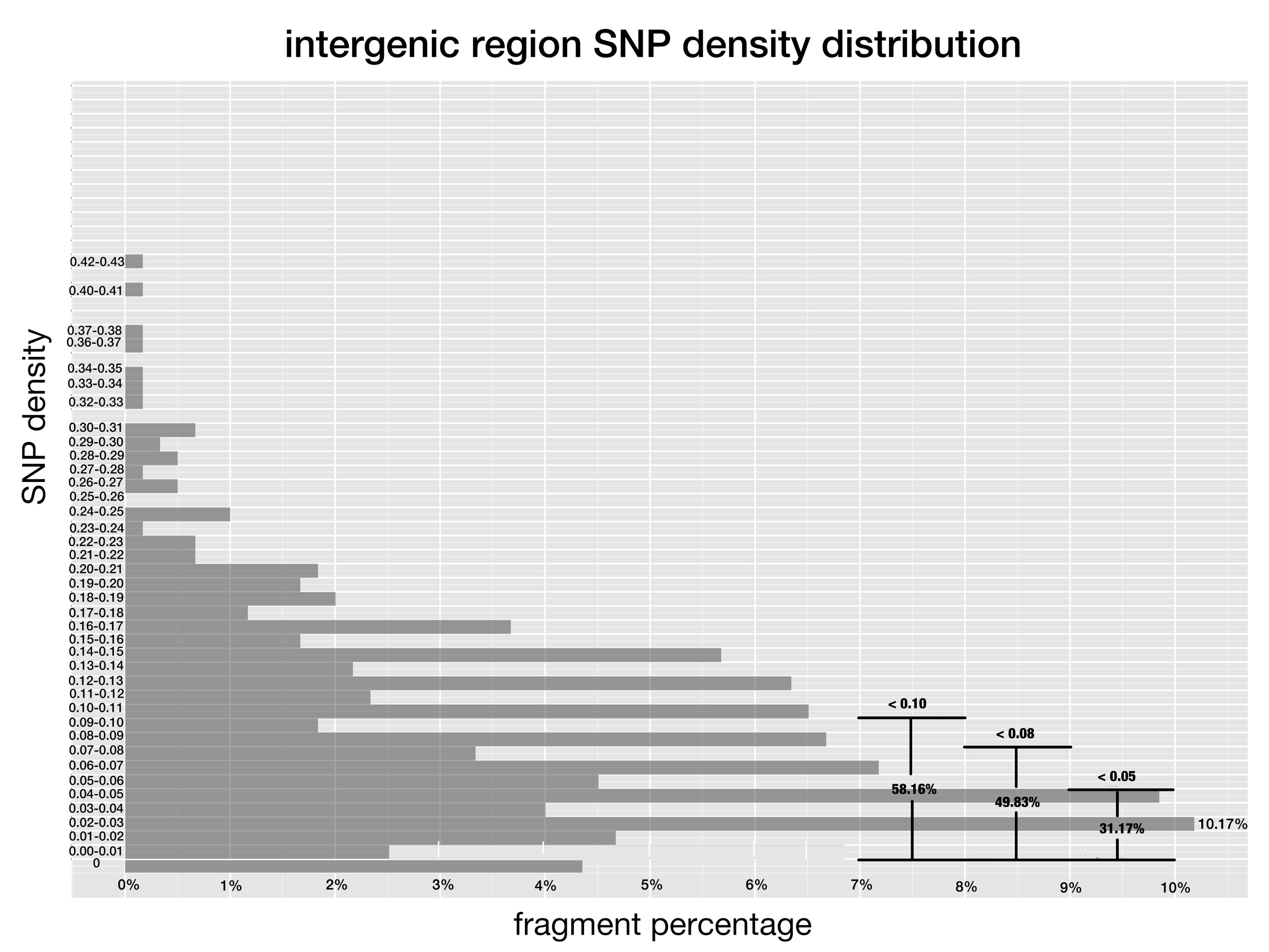
**结果与分析**

**基于3K水稻基因组项目，对全部水稻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 [[[28]](#endnote-28)]，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 [[[29]](#endnote-29)]。为了研究这个问题，我们从Rice SNP-Seek Database (3K水稻基因组项目， *snp-seek.irri.org/*) [8]收集了所有miRBase.org[27]中所有pre-miRNA、随机选择的基因间隔区 (intergenic region)片段和外显子 (exon)片段上的SNP，并且计算了其上的SNP密度，结果以图片展示出来。

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 (Fig. 1)。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，峰值为12.83% (Fig. 2)。然而，基因间隔区中则没有展现出类似的趋势 (Fig. 3)。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 (Fig. 1, 2, 3)。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。





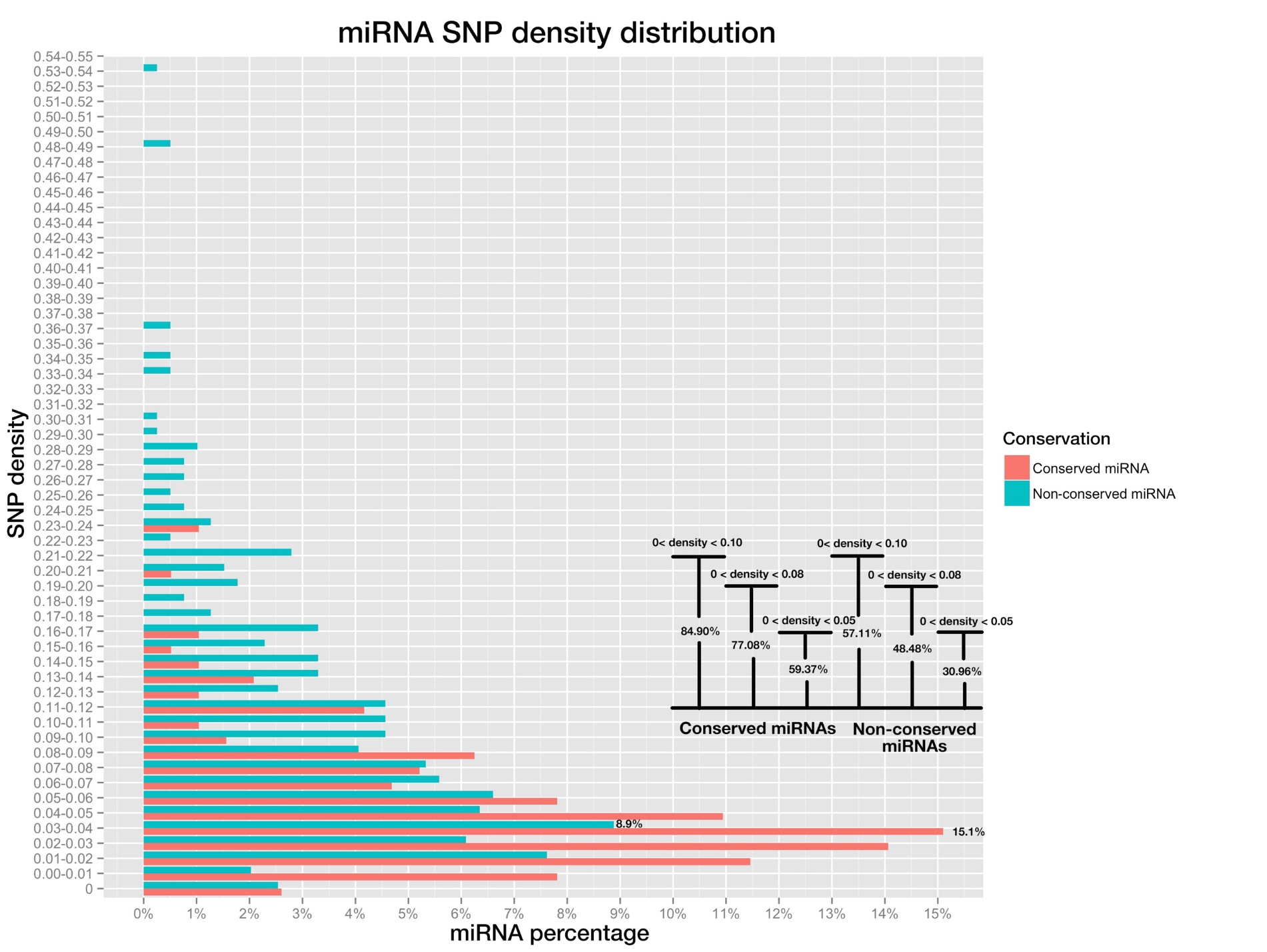
**图1-3 水稻pre-miRNA、外显子和基因间隔区的SNP密度**

*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 1-3. SNP density of pre-miRNAs, exon regions and intergenic regions in rice.**

*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 [20, [[30]](#endnote-30), [[31]](#endnote-31)]，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 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**

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

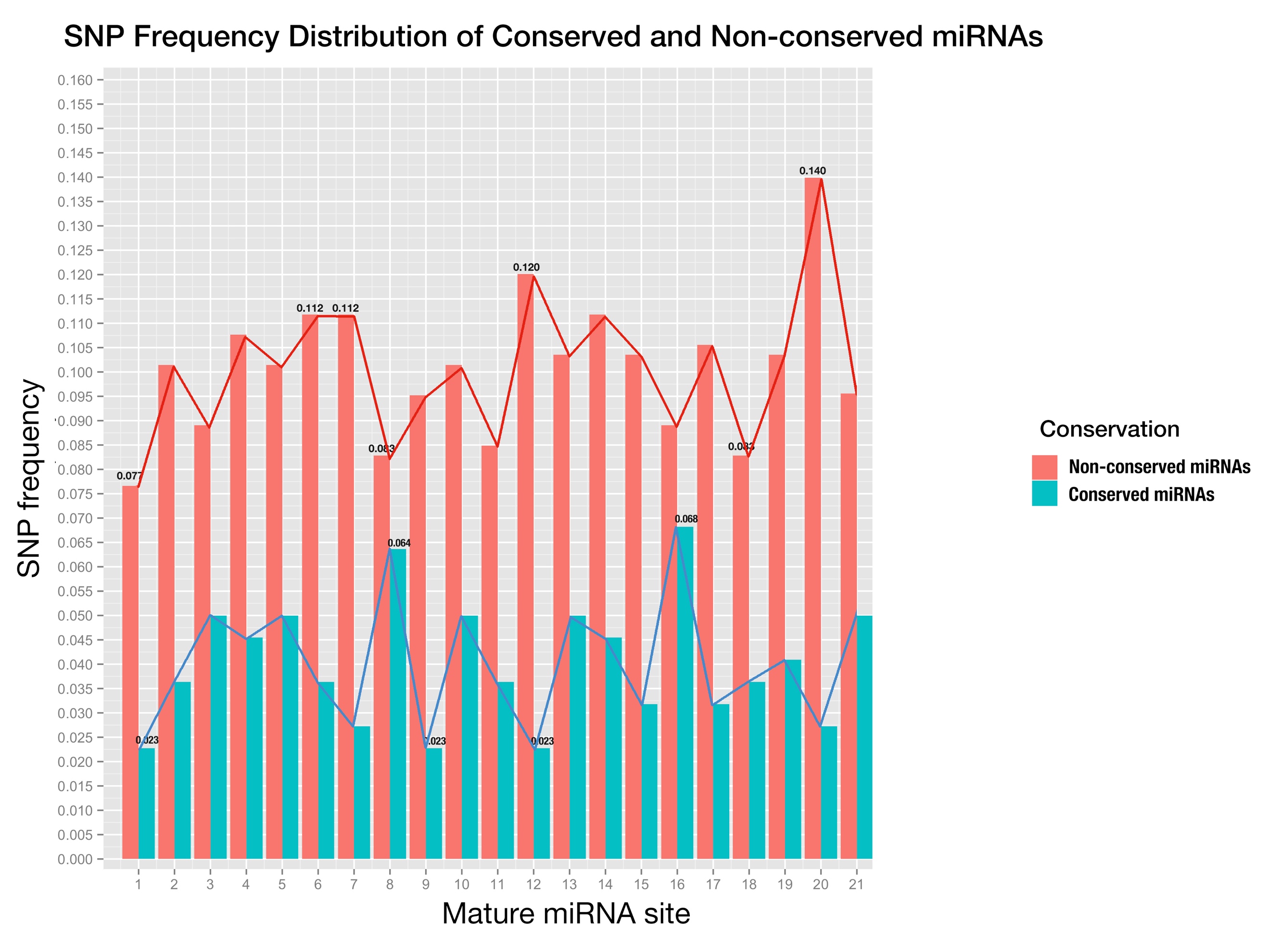
*Bar plot on the bottom right shows the percentage of miRNAs whose SNP densities are below 0.10, 0.08 and 0.05, for conserved and non-conserved miRNAs respectively.*

**保守miRNA和非保守miRNA的位点SNP频率分布是不同的**

植物miRNA对不同的靶标其调节效率有差异，而这差异归因于它们的互补模式，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 [30, [[32]](#endnote-32), [[33]](#endnote-33)]。这项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都处在相同的进化压力的话，两者在不同位点的SNP频率应该会有相似的秩 (ranking)。然而结果并非如此， 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 [[[34]](#endnote-34), [[35]](#endnote-35)]，此外其它的位点的排序都不一样。其中位点8和位点12的SNP频率在保守和非保守miRNA的秩是相反的，另外我们对两者的SNP频率进行了Pearson相关性测试，结果也显示之间没有显著的相关性 (r=-0.163, p-value=0.4473)。两者SNP频率秩的差异显示了在不同的位点之间所处进化压力的分布是不同的，进而暗示保守miRNA和非保守miRNA对靶基因的作用机理之间存在差异。

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 [[[36]](#endnote-36)-[[37]](#endnote-37)[[38]](#endnote-38)]，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，这个结果和葱经验推断而衍生出来的假设不符。



**图5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**

*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。*

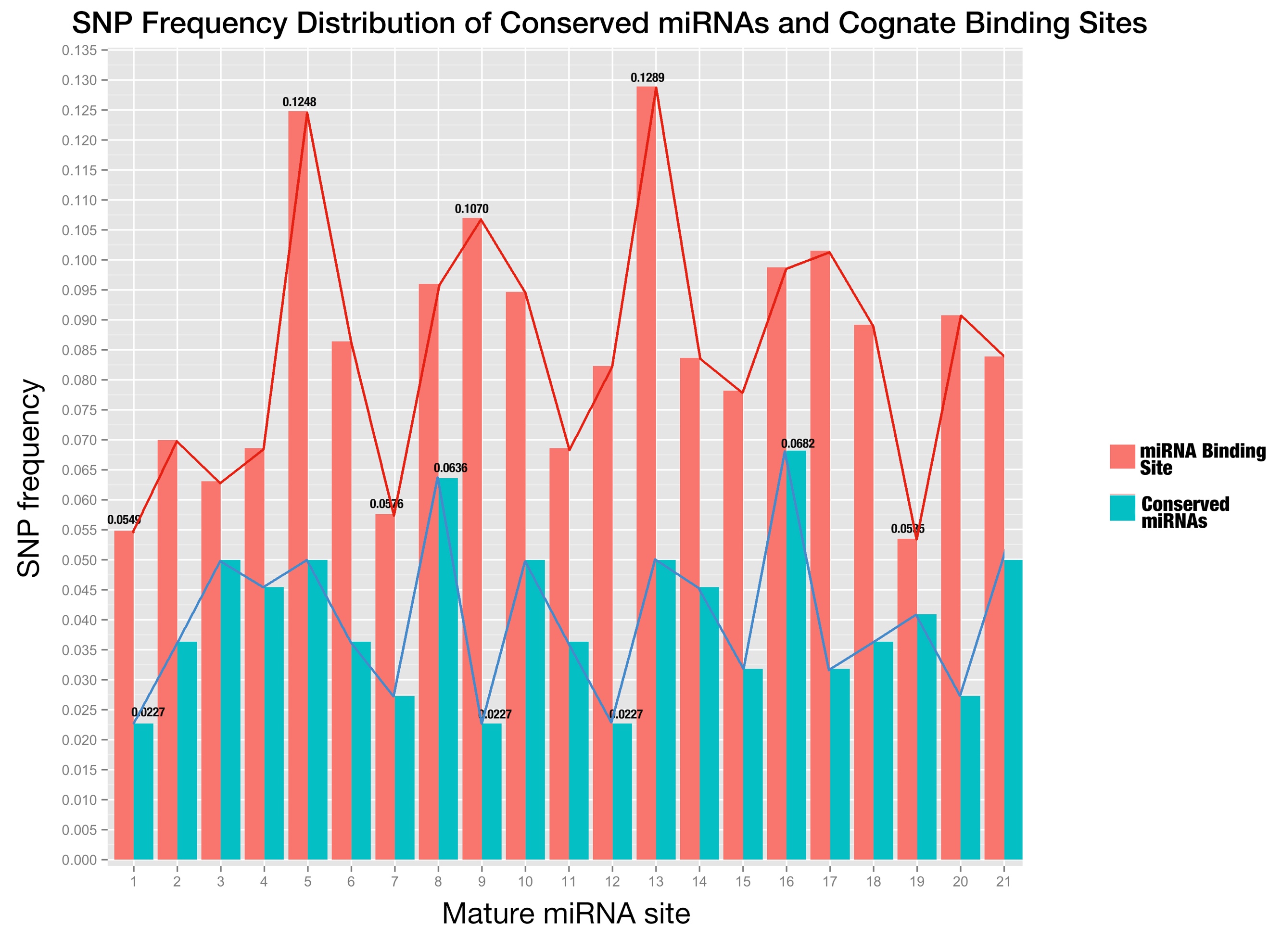
**Fig 5. Positional SNP distribution of conserved miRNAs (blue) and non-conserved miRNAs (red).**

*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和相应靶基因位点SNP频率的正相关性解释了两者之间的共同进化关系**

进一步使用3K水稻基因组的SNP数据来研究成熟miRNA和其结合位点相应的位置SNP频率的功能相关性，可能会揭示它们的功能重要性。

因为植物中，保守的miRNA比非保守的miRNA拥有更多可以鉴别的靶基因 [20]而且这些基因在功能上更加重要，所以我们只研究保守miRNA靶基因上的结合位点。已有研究报导过miRNA和其结合位点存在共同进化 [37, [[39]](#endnote-39)]。这种共同进化将会导致在成熟miRNA和相应的结合位点上各位置的SNP频率之间存在一定的相关性。为了检测两者之间是否有相关性，我们先用在线的miRNA预测工具 *psRNATarget*进行预测 [24]并结合已公开的基于全转录组的降解组实验所验证的miRNA靶基因 [[[40]](#endnote-40)]，总共找到了823个由保守miRNA作为靶标的基因。然后保守miRNA和它们结合位点的SNP频率分布展示在以下的条形图，图6。随后，我们进行了两者的Pearson相关性测试，其结果显示两者之间存在中等的正相关性 (r=0.5891, p-value=2.455e-3)，进而暗示了miRNA和相应靶基因上的结合位点之间存在共同进化。另外，比较之下，在miRNA结合位点上的每一个位置SNP频率都比相应的成熟miRNA的要高，这显示了在保守的成熟miRNA位点上的进化压力比在相应结合位点上的要大。



**图6成熟miRNA和其结合位点上每一个位置的SNP频率分布**

*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 6. SNP distribution of all sites along conserved mature miRNAs and their binding sites.**

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

**已验证的miRNA和靶基因之间表达量之间没有显著的关联性**

和动物miRNA不同的是，植物miRNA和靶基因之间有高度的互补性，这个特性也被用做生物信息预测靶基因的基础。但是如何筛选掉其中的不具有生物相关性的假阳性结果却是一个难题。在普遍的假设下，也就是互补性为基因沉默唯一决定因素以及植物中任何载入了miRNA的RISC都可以单独行动 [[[41]](#endnote-41), [[42]](#endnote-42)]，靶基因将会被相应的miRNA下调，进而导致植物miRNA和相应靶基因mRNA表达量之间呈现负相关，而且已经有实验支持了这个观点：在植物中过表达miR319时，5个编码TCP转录因子的基因都被下调 [[[43]](#endnote-43)]。为了验证这个假设，并且进一步应用到本研究中以筛选出具有功能相关性的靶基因，我们从RiceFREND数据库[[[44]](#endnote-44)]提取了水稻miRNA和基因的表达数据，并且选取了被降解组测序验证过的靶基因和miRNA作为实验组进行相关性分析 [36]。

如图7所示，367个miRNA:靶基因作用对中，只有136个在表达水平上存在负相关，这个和假设不一致。进一步可以从图中看出，过半 (197/367)的作用对落在弱相关区间(-0.4~0.4)，这表明这些作用对之间并没有很直接而强烈的相关性。



**图7由降解组验证的miRNA:靶基因作用对之间Spearman相关性系数图**

*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 7. The spearman correlation coefficient of degradome validated miRNA:target relationships;**

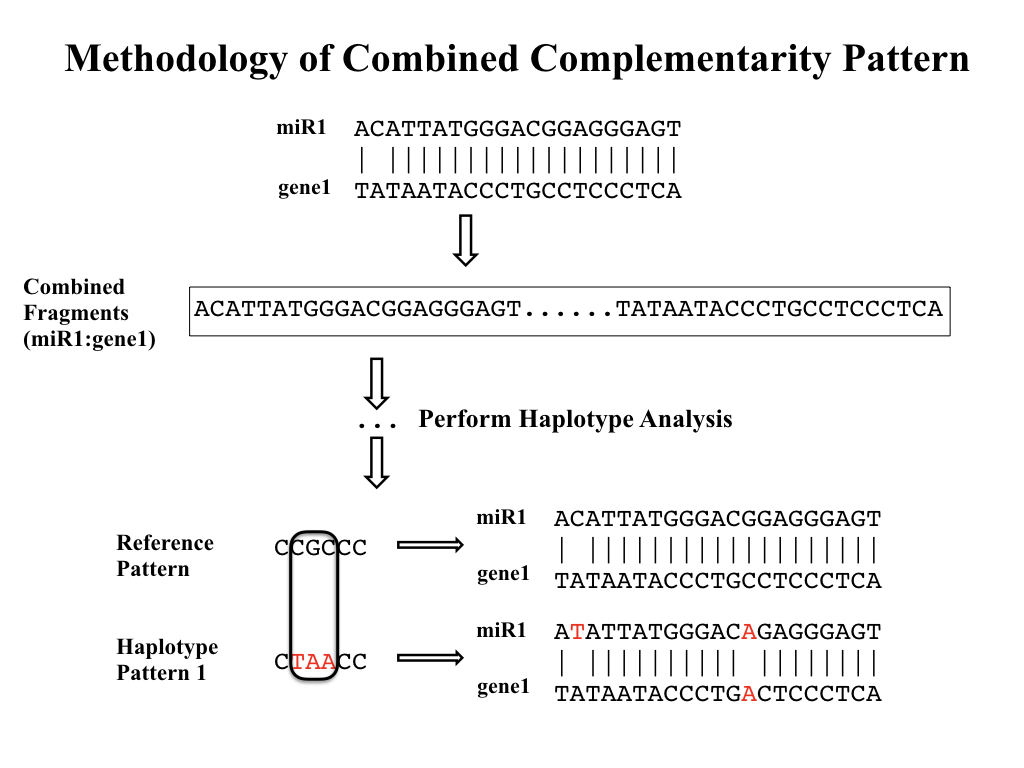
*X-axis is the number of pairs that fall on the specified range of correlation coefficient; so bars with green color in the lower part denote the negatively correlated miRNA:target pairs, by contrast, bars with brown color in the upper part denote the positively correlated miRNA:target pairs.*

**Combined Complementarity Pattern Analysis (CCPA) discovered miR818 family had positions recovered their pairing with two SNPs on both miRNA and binding site**

Haplotype analysis is frequently used in population genetic analysis, and rice cultivars can be grouped together into different sets according to the haplotype patterns when studying the function of a specific genomic region. MiRNA functions as a regulator through interaction with its target gene, to be specific, the binding site of cognate target gene. So, I extended the haplotype analysis into Combined Complementarity Pattern Analysis (CCPA) to work for both sequences on mature miRNA and miRNA binding site that are interacting while silencing.

SNP was adopted as the sole biological marker in CCPA. All SNPs were obtained from the database querying with the genomic region of both the mature miRNA and cognate miRNA binding on question, and then sorted in the ascending order of their id, which composed of 11 digits starting with number 1, followed by 2 digits representing rice genome chromosome id and then zero-filling 8 digits representing genomic position of the SNP. Each rice cultivar was denoted with a haplotype pattern which is a sequence of alleles at the sorted SNP positions, and one allele for one SNP position. After that, all 3,024 rice cultivars were divided into different sets according to their haplotype patterns.

Then, each haplotype pattern was mapped back to the sequences of mature miRNA and miRNA binding site to form the specified genotype of a rice cultivar set concerning the miRNA:target pair in question. Both sequences would then be aligned in the reverse order to form the complementarity pattern (and that’s why this method is called combined complementarity pattern analysis), for the orientation of mature miRNA is opposite to that of binding site of its target while binding, and the complementarity pattern would be compared with that of reference genome which is called the reference complementarity pattern to unfold the changes that SNPs do to the miRNA:target complementarity.

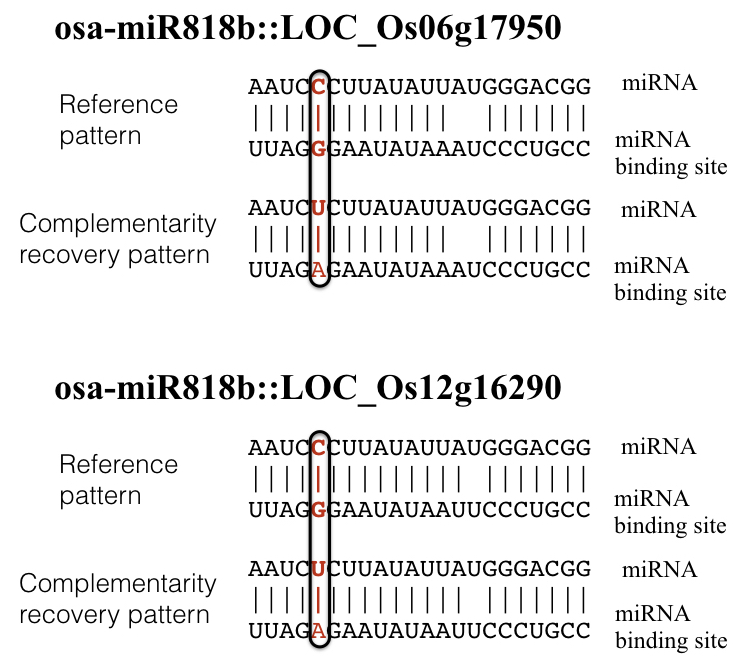


**Fig 8. Workflow of Combined Complementarity Pattern Analysis (CCPA)**

Generally, positions in complementarity pattern of miRNA:target pair could undergo four types of changes when SNPs take place in either mature miRNA or miRNA binding site (or both of them). The complementarity of the positions could switch from pairing to mismatch, from mismatch to pairing, from pairing to pairing (remain pairing) or from mismatch to mismatch (remain mismatch). Applying CCPA to all conserved miRNAs in rice as well as their target genes, a special type of complementarity change was found that positions remained pairing after two SNPs were introduced to both sequences at the given position of the complementarity pattern, and in this study, we called it complementarity recovery phenomenon. And this type of change took place in osa-miR818a-e, osa-miR1436, osa-miR1439, osa-miR1442, osa-miR1862b, osa-miR444a/b/d and their cognate targets. Osa-miR444 family was previously reported belonging to a unique class of miRNAs called natural antisense miRNAs that derive from the natural cis-antisense transcript pairs, and can generate mature miRNAs that perfectly match their targets [[[45]](#endnote-45)]. The recovery of complementarity can be explained by the antisense property because a single SNP introduced to the rice genome would cause mutations on both miRNA and its target but they still remain perfectly matched. But for other miRNAs, two SNPs were needed to keep the position still pairing. More interestingly, all other miRNAs belong to the same miRNA family of related hairpin sequences, osa-miR818, according to *miRBase.org*.

The osa-miR818 family was poorly studied till now [[[46]](#endnote-46)], but changes of expression of osa-miR818 was detected in japonica rice infected with rice ragged stunt disease [[[47]](#endnote-47)] and two target genes with unknown functions were examined for osa-miR818 by L. Y. Li et al. [[[48]](#endnote-48)]. And the SNP analysis in this study of all conserved miRNA:target interaction pairs also found 290 out of 313 interaction pairs carrying more than 5 SNPs, fell in the osa-miR818 family. The big number of SNPs of interaction pairs of this family implies that the selection pressure on the miRNA as well as binding sites is relatively small and would allow the random mutation on both sequences and furthermore produce the complementarity recovery phenomenon.

But among the complementarity recovery interaction pairs, there were two of them carrying only four SNPs and through mathematical calculation, the possibility of complementarity recovery phenomenon was about 3.1% with only 4 SNPs. Complementarity recovery in both cases happened at position 5 at both sequences with double mutations from CG to UA (Fig. 9). Closer examining of both cases found there were other rice cultivars with only single mutation from CG to CA and from CG to UG. And a possible explanation of this phenomenon would be that during the evolution of rice miRNA and its target, for some rice cultivars, constraint of complementarity became looser and allowed the mismatch at position 5. So, the single mutation could happen at this position, but afterward, the constraint arose again and forced the miRNA:target interaction pair to become pairing again at this position.



**Fig 9. Complementarity recovery patterns**

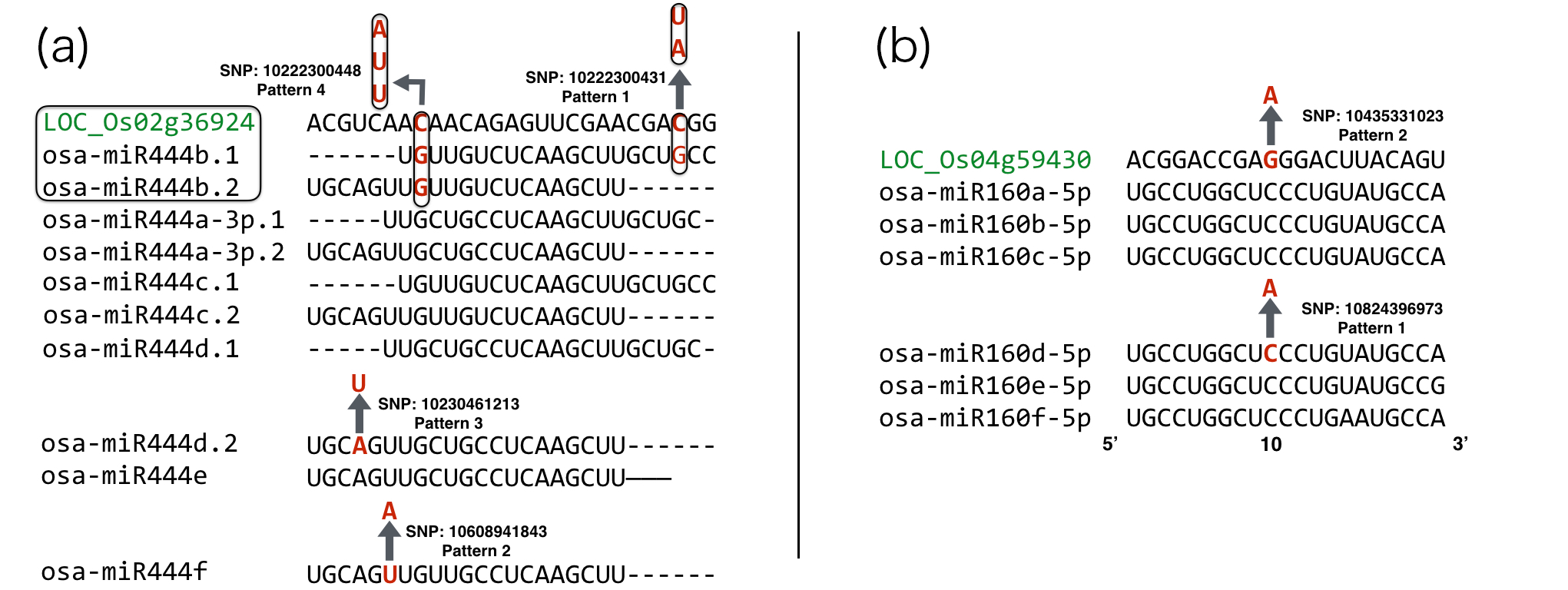
*In the complementarity pattern, a vertical bar would be placed denoting match, a blank denotes mismatch.*

**Changes to conserved miRNA:target complementarity brought by SNPs didn’t lead to distinct phenotypical changes**

Previous studies have found variations of miRNA-mediated regulation caused by SNP can have huge impact on agronomic phenotypes [21-22]. Considering the functional redundancy of plant miRNAs that members of miRNA families share similar sequences and are often predicted to target the same or overlapping sets of genes [[[49]](#endnote-49)-[[50]](#endnote-50)], we mainly focus on SNPs on miRNA binding sites of target genes rather than SNPs on mature miRNAs, whose influence on the outcome of miRNA regulation might be buffered by functional redundancy. Finally, 7 target genes of conserved miRNAs were found to carry SNPs on them. Apart from the complementarity pattern of miRNA:target, target accessibility to the miRNAs has also been identified as an important factor that are involved in target recognition [[[51]](#endnote-51)]. In order to evaluate the potential influences these SNPs could bring to the miRNA regulation, we mapped the SNPs to the miRNA binding site positions in the positional order of 5’-3’ on mature miRNAs, as well as the changes of total free energy of binding after the mutations caused by SNPs (Table 1, Fig 10).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 1. Summary of target genes carrying SNPs on the binding site** | | | | | |  |
| **Gene locus** | **SNP id** | **Predicted targeting miRNA family** | **Position on miRNA binding site** | **Average changes of free energy of binding (kcal/mol)** | **Gene name** | **Effect on proteins** |
| LOC\_Os12g41860 | 11225924993 | osa-miR166 | 1 | 2.116 | OSHB3 | Synonymous mutation |
| LOC\_Os12g41680 | 11225805945 | osa-miR164 | 8 | 1.902 | ONAC60 | Synonymous mutation |
| LOC\_Os05g25960 | 10515090268 | osa-miR164 | 5 | 4.782 |  | Missense, Phe128Leu |
| LOC\_Os04g59430 | 10435331023 | osa-miR160 | 10 | 2.58 | OsARF13 | Missense, Gly403Glu |
| LOC\_Os04g24190 | 10413845263 | osa-miR196 | 2 | 1.294 | OsGRF11 | Missense, Phe277Val |
| LOC\_Os02g49840 | 10230461213 | osa-miR444 | / | 1.16 | OsMADS57 | Missense, Leu99Gln |
| 10230461236 | -0.004 | Missense, Lys107Gln |
| LOC\_Os02g36924 | 10222300431 | osa-miR444 | / | -0.032 | OsMADS27 | Missense, Ala93Val |
| 10222300448 | 6.575 | Missense, Gln99Lys |

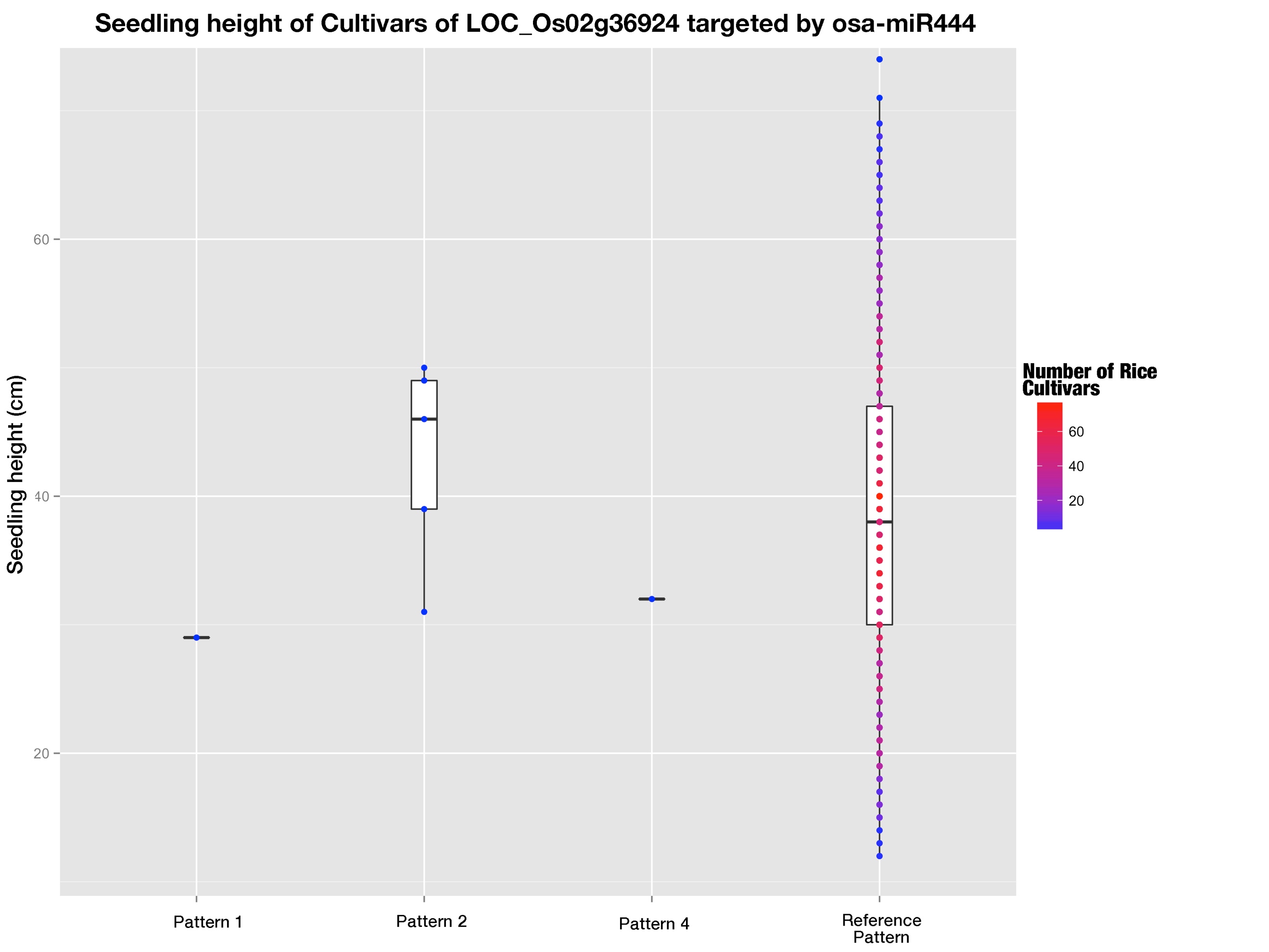
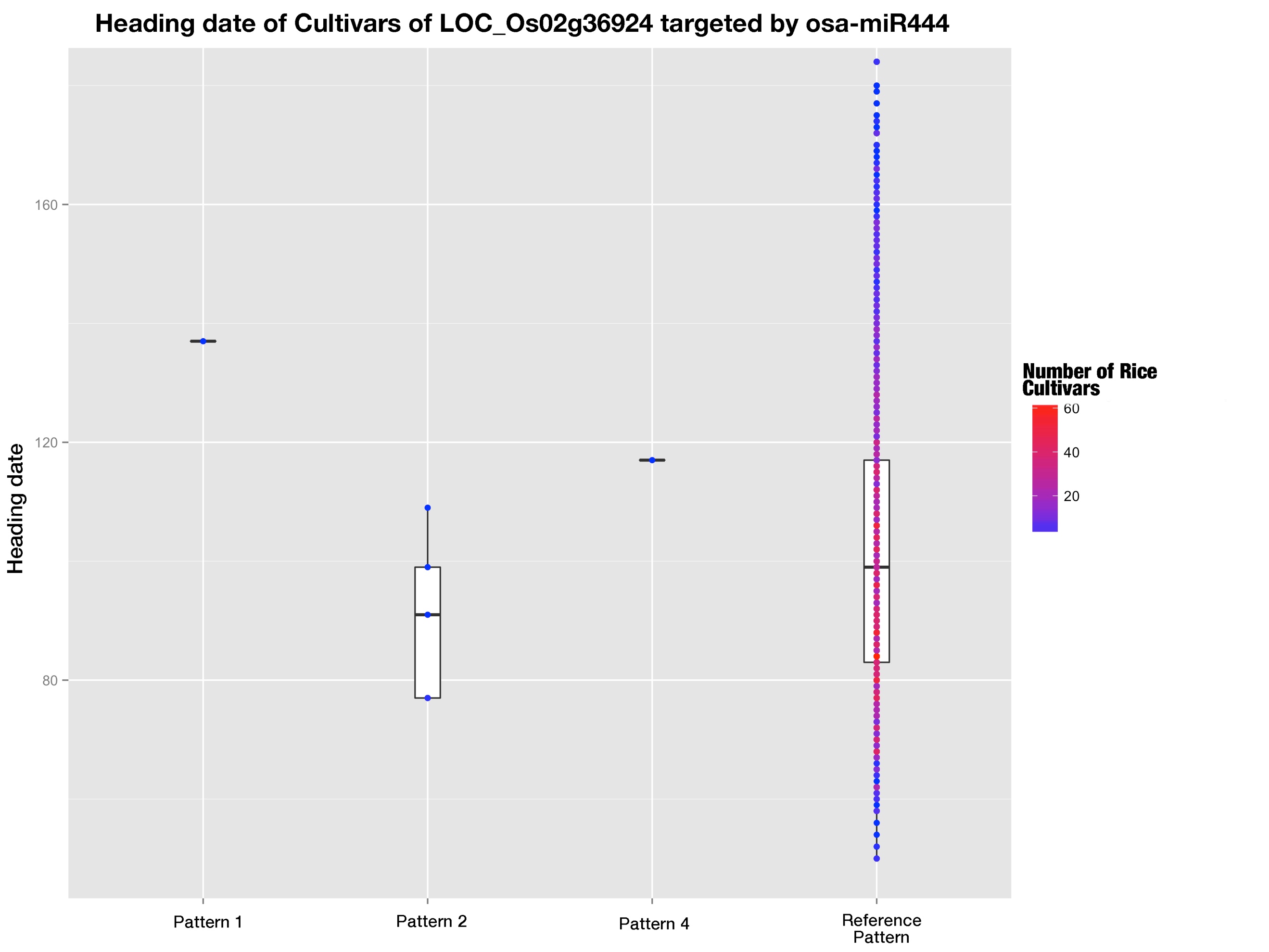
*Position on miRNA binding site is in the order of 5’ to 3’ on mature miRNA, and total free energy of binding was calculated using RNAup program in Vienna Package [[[52]](#endnote-52)].*

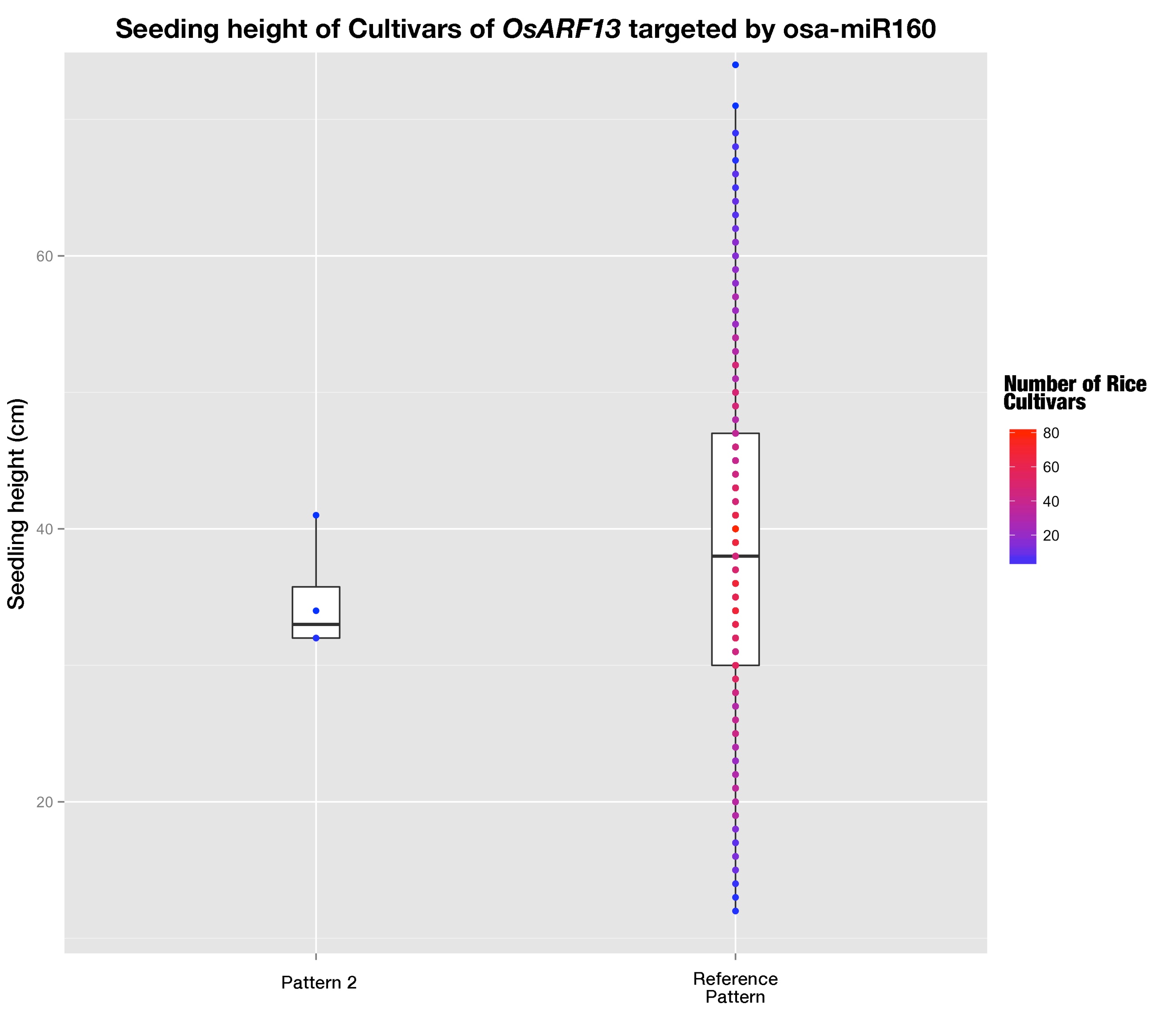
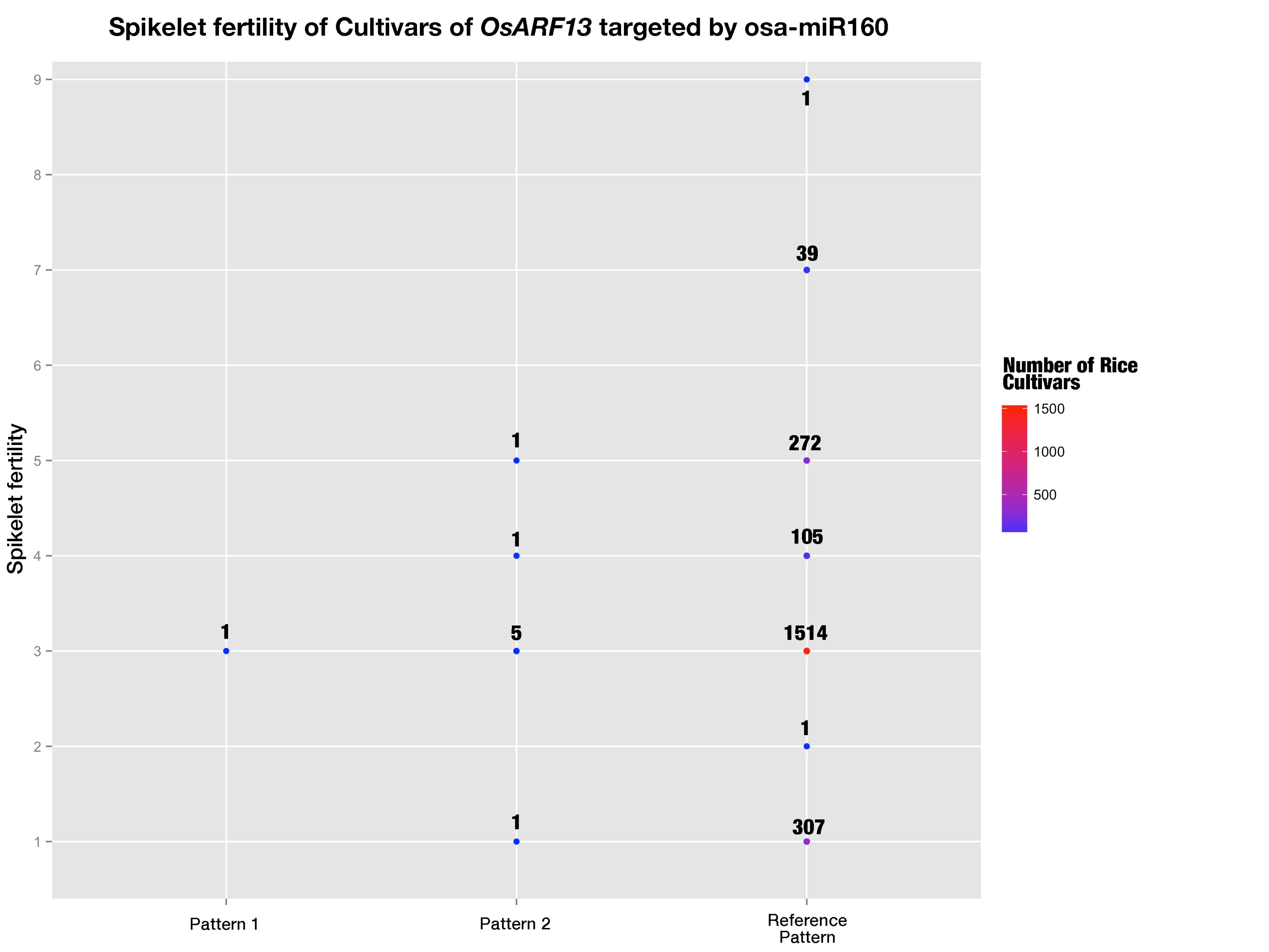


**Fig 10. Complementarity pattern of 2 target genes with their targeting miRNA family**

*Gene loci are marked in green color; each arrow stands for a mutation caused by SNP and is denoted as a separate haplotype pattern. Enclosed frame of LOC\_Os02g36924 and osa-miR444b.1/b.2 means they are on the same genomic region except for the opposite strand and SNPs in this region would cause mutations on both mature miRNAs and miRNA binding site simultaneously.*

By utilizing CCPA on the combined analysis of miRNAs and their corresponding miRNA binding sites of target genes, rice cultivars were obtained and classified to different haplotype patterns. Auxin response factors (ARF) in rice were reported to be involved with fertility, height and grain yield [[[53]](#endnote-53)] and rice MADS-box genes were reported to be involved with heading date and plant height [[[54]](#endnote-54)]. With the aid of the phenotypical data provided by Rice SNP-seek Database, the quantitative phenotypes of rice cultivars were plotted according to their haplotype pattern (Fig 11-14).





**Fig 11-14. Phenotypes of rice cultivars belonging to different haplotype patterns**

The color gradient stands for the number of rice cultivars at that dot.

SNP 10435331023 on miRNA binding site of OsARF13 caused a mismatch at position 10, which was claimed to be the important cleavage site and a mismatch at this position would cause fatal damages to silencing efficacies of miRNAs [37-3839]. And SNP 10222300448 on miRNA binding site of OsMADS27 increased the total free energy of binding by 6.575 kcal/mol, and would predictably brought down the silencing efficacy of osa-miR444 family to a large extend. But the predicted changes of outcomes of miRNA regulation didn’t lead to distinct phenotypical change (Fig 11-14).

**Discussion**

Single nucleotide polymorphisms are good indicators of evolutionary selection for different genomic regions and have already been employed to study natural selections on human miRNAs [28, 29]. In these studies, SNPs on functional regions such as pre-miRNAs especially seed-regions as well as miRNA binding site were very rare and much less than other conserved sequence motifs in 3’ UTR [28, 29]. And similar phenomena were also observed in this study that we found SNP density in pre-miRNAs were fewer than that in intergenic regions as well as exons, implying miRNAs are subjected to stricter evolutionary pressure than intergenic regions and exons. This is consistent with the role of miRNAs as master regulators in plants. There are big differences between conserved miRNAs and non-conserved miRNAs, the so-called rice specific miRNAs or newly arising miRNAs. Generally, conserved miRNAs are conserved across different species and have identifiable binding sites on targets which are also conserved among species and they tend to target genes encoding transcriptional factors (TFs), while non-conserved miRNAs are commonly taken as new-comers with few if any identifiable binding sites in targets both by bioinformatics and degradome sequencing [20]. Our comparative analysis of SNPs fallen unto conserved miRNAs and non-conserved ones had produced similar results, SNP density of conserved miRNA precursors was significantly less than that of non-conserved ones, showing that the more important regulation role of conserved miRNAs would impose greater selection pressure on non-conserved miRNAs. Previous study conducted by Liu Q et al. found when less conserved miRNAs were excluded, pre-miRNAs accumulated much fewer SNPs, and this also confirmed the stricter purifying pressure on conserved miRNAs. Considering the different evolutionary processes of conserved and non-conserved miRNAs [30, 31] and the fact that by the common target prediction methods few identifiable targets are found for non-conserved miRNAs, the functioning mechanisms of miRNA towards miRNA binding sites of targets may be different. In our study, the comparison of positional SNP frequency distributions between conserved mature miRNAs and non-conserved mature miRNAs showed distinct rankings of SNP frequencies along the positions, which indicated different selection pressure distribution among the positions for conserved and non-conserved miRNAs. It was reported that there was co-evolution of miRNAs and their cognate target genes [37, 40], and here in our study, correlation tests of positional SNP frequencies of conserved miRNAs and that of cognate targets showed moderate positive correlation with statistical significance, which provided further evidence for the coevolution of miRNAs and their cognate targets. And SNP frequency of miRNA was lower than that of cognate binding site at every position in this study. Several factors may be accountable for this result. Firstly, plant miRNA regulates multiple targets with high complementarity, adding more constraints to mature miRNAs and thus cause the SNP frequency to be lower. Secondly, a large part of the targets was obtained from the bioinformatic methods and false positive target genes might be included; these false positive target genes would not subject to miRNA regulation and the predicted binding sites on them might not be exposed to selection pressure from miRNA-mediated regulation.

In plants, miRNAs serve as master regulator through high complementarity towards binding site of targets and previous researches revealed the different importance of different positions on target recognition and cleavage [32-3334]. 5’ terminal nucleotide, which is position 1 on mature miRNA, determines which Argonaut protein to load for miRNA [35], and this layer of constraint on position 1 was reflected in this study as the lowest SNP frequency position both for conserved and non-conserved miRNAs. But unexpectedly, position 10 and 11, conventionally regarded as cleavage sites and were required to be perfect pairing to binding sites [37-3839], were not even among the lowest SNP frequency positions (Fig. 5). And this implied selection constraints imposed other sites with lower SNP frequencies are stronger than the cleavage site constraint imposed on these two positions. Liu Q et al. had similar finding for position 1 and 10, except that position 11 was reported to have the lowest SNP frequency and the conflict might be caused by the separation of conserved and non-conserved miRNAs. The separation of conserved miRNAs and non-conserved miRNAs would specify the general trend for the two class of miRNAs which have been reported to have different functioning mechanisms.

In order to search for biologically relevant target genes for miRNAs, generally ways such as 5’-RACE or degradome sequencing would be adopted. But for the abundant outcome of bioinformatic miRNA target prediction programs, expression correlation of miRNA and mRNA of cognate target genes was thought to be a feasible way to search for those biologically relevant ones. In this study, degradome validated miRNA:target pairs were found not to be fully negatively correlated and in contrast, more interaction pairs were positively correlated than negative correlated pairs. Ming Wen et al. had found similar phenomenon, in whose study positively correlated interaction pairs prevailed [[[55]](#endnote-55)]. And this may be caused by more complex mechanisms such as negative feedback loops (FBLs) and incoherent feedforward loops (FFLs) mentioned by Ming Wen et al. Also, the results showed that it’s not practical to use expression correlation for target screening.

No studies before had tried to adopt haplotype analysis to study the actual mutations caused by SNPs of miRNA-mediated regulations that happen to rice cultivars. In this study, haplotype analysis was extended to be combined complementarity pattern analysis (CCPA) and could help to study the polymorphisms of interactions between a family of miRNAs and their common target gene among different rice cultivars. And the afterward found complementarity recovery phenomenon of osa-miR818 family implied that less conserved miRNAs subject to low selection pressure and allow more flexible changes to the miRNA:target complementarity patterns.

Several studies showed the SNPs involved in miRNA-mediated regulation would cause apparent changes to plant phenotypes [21, 22]. While here in this study, genome-wide analysis of SNPs involved in miRNA-mediated regulations found 7 target genes carrying SNPs on their miRNA binding sites and two of them were promising in bringing big effects to miRNA’s regulation. One of them was located in cleavage site position 10 and the other brought up the free energy of binding by 6.575 kcal/mol. But unexpectedly, the comparison of relevant phenotypes of the SNP mutated rice cultivars and those of the reference pattern rice cultivars didn’t show distinct changes. And this implied that these plant phenotypes are controlled by more than one gene so that predicted changes in a single gene expression would not be reflected in the phenotype changes.

**Conclusion**

Based on the recently identified millions of SNPs from 3K Rice Genome Project, a genome-wide investigation of SNPs in rice miRNAs as well as their cognate binding sites of target genes was carried out. We found that pre-miRNAs tend to accumulate less SNPs compared with exons and intergenic regions, which suggests of stricter selection pressure imposed by the role of miRNAs as master regulatory units. The rankings of SNP frequency along the mature miRNAs differed between conserved miRNAs and non-conserved ones, showing the different recognizing and functioning mechanisms of miRNA towards target genes between them; while the positive correlation between SNP frequencies of conserved miRNAs and their cognate binding sites may be suggestive of the co-evolution of miRNAs and their target genes. The SNP found within binding sites of target genes at the critical cleavage position 10 and the other SNP that would increase the free energy of binding were potential to influence the miRNA regulation, but the indistinct phenotypical changes may be due to the multigene controlling of plant phenotypes. These findings are important for better understand and further investigation how SNPs would affect the miRNA-mediated regulation and further the miRNA-regulated plant phenotypes.

**Methods**

**Sequence data**

The rice miRNA data including sequence data and genomic location of both precursor miRNAs as well as mature miRNAs were obtained from miRBase database (release 21, in June 2014). A small fraction of pre-miRNAs whose genomic locations were not provided were used as query to search against the MSU7 rice genomic sequence using BLASTN with E value cutoff 10-10 and only those miRNAs that could be exactly mapped to reference genome were recorded, in which osa-miR1882bl whose precursor was mapped to a sequence on MSU7 with only one mismatch, was also recorded considering that it could be perfectly mapped to indica genome(ASM165v1). Totally, 585 pre-miRNAs along with 703 mature were recorded for further use. SNPs were downloaded from SNP-Seek Database (<http://snp-seek.irri.org/)> and then loaded to local MySQL database. After that, genomic coordination of miRNAs was used as query against the local SNP database and we’ve got 7193 SNPs fallen on pre-miRNAs and 1270 SNPs on mature miRNAs.

**MiRNA target identification**

Because in this study, we focused on analyzing targets of conserved miRNAs. We classified the miRNAs according to their conservation aided by miRNA family classification downloaded from miRBase (miFAM.dat file). Those miRNAs who had at least one member from other plant species at the same miRNA family were classified to be conserved miRNAs. PsRNATarget web server was employed with default prediction parameter using sequences of mature miRNAs to predict the target genes. In addition, some targets were collected from Liu Q. et al paper, and these sequences were used as target transcript candidates to predict which miRNA could target it in psRNATarget web server to obtain its miRNA binding sites. In turn, 823 target genes were recorded with their binding sites’ genomic coordination. Then, they were queried against local SNP database and 1169 SNPs were found fallen within the miRNA binding sites of these target genes.

**Identification and analysis of SNPs involved in miRNA-mediated regulation**

To compare the SNP density of pre-miRNAs and that of exons as well as intergenic regions, we randomly selected 600 sequence fragments with the length of 150nt from exon regions and intergenic regions across all rice genomes using in-house Python script, respectively. The SNP density was calculated as SNP numbers per base and they were plotted using R package “ggplot”. Then, SNP frequency that is assessed as the division of number of miRNAs that has SNP at the given position by total miRNA number, was calculated for each position along mature miRNAs for conserved miRNAs and non-conserved miRNAs as well as binding sites of cognate target genes. And SNP frequency distribution was plotted using R package “ggplot” as well.

**Expression correlation analysis**

The expression data of both miRNAs along with their degradome validated target genes, was downloaded from EMBL-EBI database with accession number E-GEOD-21396 (packed data derived from RiceFREND). The expression level of both pre-miRNAs and their cognate target genes were used to do Pearson correlation test with 27 day-after-transplanting seedlings as samples.

**Combined complementarity pattern analysis**

Detailed description of CCPA was stated in the result part. CCPA was applied to all conserved miRNAs and we focused on target genes carrying SNPs on the binding sites, finally, 7 target gene were found with SNPs on the binding site. The phenotype data were downloaded from SNP-Seek database. The cultivars belonging to these non-reference patterns were extracted from the local MySQL database. Then phenotypes of different rice cultivars belonging to these combined haplotype patterns were compared.

**References**

1. Arai-Kichise Y, Shiwa Y, Nagasaki H, Ebana K, Yoshikawa H, Yano M, et al. (2011), Discovery of genome-wide DNA polymorphisms in a landrace cultivar of Japonica rice by whole-genome sequencing. (Plant Cell Physiol), 52:274-282. [↑](#endnote-ref-1)
2. J. Mammadov, R. Aggarwal, R. Buyyarapu, S. Kumpatla, 2012, SNP markers and their impact on plant breeding (Int. J. Plant Genom.), 2012:728398 [↑](#endnote-ref-2)
3. Arai-Kichise Y, Shiwa Y, Nagasaki H, Ebana K, Yoshikawa H, Yano M, et al. (2011), Discovery of genome-wide DNA polymorphisms in a landrace cultivar of Japonica rice by whole-genome sequencing. (Plant Cell Physiol), 52:274-282. [↑](#endnote-ref-3)
4. Jena KK, Mackill DJ, 2008, Molecular markers and their use in marker-assisted selection in rice. (Crop Sci), 48:1266–1276 [↑](#endnote-ref-4)
5. Lee SH, van der Werf JHJ, Hayes BJ, Goddard ME, Visscher PM, 2008, Predicting unobserved phenotypes for complex traits from whole-genome SNP data. (PLoS Genet) 4:e1000231 [↑](#endnote-ref-5)
6. Huang X, et al. 2012, A map of rice genome variation reveals the origin of cultivated rice. (Nature) 490:497–501 [↑](#endnote-ref-6)
7. Xu X, et al. 2012, Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. (Nat Biotechnol) 30:105–111 [↑](#endnote-ref-7)
8. Alexandrov, N. et al. 2015, SNP-Seek database of SNPs derived from 3000 rice genomes. (Nucleic Acids Res.) 43:1023–1027. [↑](#endnote-ref-8)
9. Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, et al. 2012, Maize HapMap2 identifies extant variation from a genome in flux. (Nat. Genet.) 44:803–7 [↑](#endnote-ref-9)
10. Lai J, Li R, Xu X, Jin W, Xu M, et al. 2010, Genome-wide patterns of genetic variation among elite maize inbred lines. (Nat. Genet.) 42:1027–30 [↑](#endnote-ref-10)
11. Lam HM, Xu X, Liu X, Chen W, Yang G, et al. 2010, Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. (Nat. Genet.) 42:1053–59 [↑](#endnote-ref-11)
12. Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, et al. 2010, Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. (Nature) 465:627–631 [↑](#endnote-ref-12)
13. Castle JC, 2011, SNPs occur in regions with less genomic sequence conservation. (PLoS ONE) 6:e20660 [↑](#endnote-ref-13)
14. Yamamoto, T., Nagasaki, H., Yonemaru, J., Ebana, K., Nakajima, M., Shibaya, T. et al. 2010, Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. (BMC Genomics) 11:267. [↑](#endnote-ref-14)
15. Liu Q, Wang H, Zhu L, Hu H, Sun Y, 2013, Genome-wide identification and analysis of miRNA-related single nucleotide polymorphisms (SNPs) in rice. (Rice) 6:10 [↑](#endnote-ref-15)
16. Liu Q, Wang H, Hu H, Zhang H, 2015, Genome-wide identification and evolutionary analysis of positively selected miRNA genes in domesticated rice. (Mol Genet Genomics) 290(2):593–602 [↑](#endnote-ref-16)
17. Ehrenreich IM, Purugganan MD, 2008, Sequence variation of microRNAs and their binding sites in Arabidopsis. (Plant Physiol) 146:1974-1982. [↑](#endnote-ref-17)
18. Meyers BC, Axtell MJ, Bartel B et al. 2008, Criteria for annotation of plant MicroRNAs. (Plant Cell) 20(12):3186–3190 [↑](#endnote-ref-18)
19. Sun G, 2012, MicroRNAs and their diverse functions in plants. (Plant Mol Biol) 80:17–36 [↑](#endnote-ref-19)
20. Jones-Rhoades MW, 2011, Conservation and divergence in plant microRNAs. (Plant Mol Biol) 80:3–16 [↑](#endnote-ref-20)
21. Jiao, Y., et al. 2010, Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. (Nat Genet) 42(6): 541-544. [↑](#endnote-ref-21)
22. Houston K, McKim SM, Comadran J, et al. 2013, Variation in the interaction between alleles of HvAPETALA2 and microRNA172 determines the density of grains on the barley inflorescence. (Proceedings of the National Academy of Sciences), USA 110:16675–16680. [↑](#endnote-ref-22)
23. Li, S. et al. 2013, MicroRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in Arabidopsis. Cell 153, 562–574 [↑](#endnote-ref-23)
24. Dai, X. and Zhao, P.X., 2011, psRNATarget: a plant small RNA target analysis server. (Nucleic Acids Res.) 39:W155-W159 [↑](#endnote-ref-24)
25. Li, J. et al. 2014, The functional scope of plant microRNA-mediated silencing. (Trends Plant Sci.) 19:785-756. [↑](#endnote-ref-25)
26. 3K R.G.P. 2014, The 3,000 rice genomes project. (Gigascience), 3:7. [↑](#endnote-ref-26)
27. Kozomara A, Griffiths-Jones S., 2014, miRBase: annotating high confidence microRNAs using deep sequencing data. (Nucleic Acids Res.) 42:D68-D73 [↑](#endnote-ref-27)
28. Chen, K. & Rajewsky, N., 2006, Natural selection on human miRNA binding sites inferred from SNP data. (Nature Genet.) 38:1452–1456 [↑](#endnote-ref-28)
29. Saunders, M. A., Liang, H. & Li, W. H., 2007, Human polymorphism at microRNAs and microRNA target sites. (Proc. Natl Acad. Sci. USA) 104, 3300–3305 [↑](#endnote-ref-29)
30. Fahlgren N, Jogdeo S, Kasschau KD et al, 2010, MicroRNA gene evolution in Arabidopsis lyrata and Arabidopsis thaliana. (Plant Cell) 22(4):1074–1089 [↑](#endnote-ref-30)
31. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP, 2006, A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. (Genes Dev) 20(24):3407–3425 [↑](#endnote-ref-31)
32. Mallory, A.C., Reinhart, B.J., Jones-Rhoades, M.W., Tang, G., Zamore, P.D., Barton, M.K., and Bartel, D.P. 2004, MicroRNA control of PHABULOSA in leaf development: Importance of pairing to the microRNA 59 region. (EMBO J.) 23: 3356-3364. [↑](#endnote-ref-32)
33. Parizotto, E.A., Dunoyer, P., Rahm, N., Himber, C., and Voinnet, O.

    2004, In vivo investigation of the transcription, processing, endonucleolytic activity, and functional relevance of the spatial distribution of a plant miRNA. (Genes Dev.) 18: 2237-2242. [↑](#endnote-ref-33)
34. Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H, Long C, et al., 2008, Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5′ terminal nucleotide. (Cell) 133: 116–127 [↑](#endnote-ref-34)
35. Mallory A, Vaucheret H, 2010, Form, function, and regulation of ARGONAUTE proteins. (Plant Cell) 22: 3879–3889 [↑](#endnote-ref-35)
36. Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M., and Weigel, D. 2005, Specific effects of microRNAs on the plant transcriptome. (Dev. Cell) 8: 517–527 [↑](#endnote-ref-36)
37. Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., García, J.A., and Paz- Ares, J. 2007, Target mimicry provides a new mechanism for regulation of microRNA activity. (Nat. Genet.) 39: 1033-1037 [↑](#endnote-ref-37)
38. Todesco, M., Rubio-Somoza, I., Paz-Ares, J., and Weigel, D. 2010, A collection of target mimics for comprehensive analysis of microRNA function in Arabidopsis thaliana. (PLoS Genet) 6: e1001031 [↑](#endnote-ref-38)
39. Arikit S, Zhai J, Meyers BC, 2013, Biogenesis and function of rice small RNAs from non-coding RNA precursors. (Curr Opin Plant Biol) 16(2):170–179. [↑](#endnote-ref-39)
40. Li YF, Zheng Y, Addo-Quaye C, Zhang L, Saini A, Jagadeeswaran G, Axtell MJ, Zhang W, Sunkar R., 2010, Transcriptome-wide identification of microRNA targets in rice. (Plant J.) 62:742-759 [↑](#endnote-ref-40)
41. Iwakawa, H. and Tomari, Y. 2013, Molecular insights into microRNA-mediated translational repression in plants. (Mol. Cell) 52:591-601 [↑](#endnote-ref-41)
42. Tang, G. et al. 2003, A biochemical framework for RNA silencing in plants. (Genes Dev.) 17:49-63 [↑](#endnote-ref-42)
43. Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, et al. 2003, Control of leaf morphogenesis by microRNAs. (Nature) 425:257–63 [↑](#endnote-ref-43)
44. Sato Y, Namiki N, Takehisa H, Kamatsuki K, Minami H, Ikawa H, Ohyanagi H, Sugimoto K, Itoh J, Antonio B, Nagamura Y 2013, RiceFREND: a platform for retrieving coexpressed gene networks in rice. (Nucleic Acids Research) 41:D1214-D1221. [↑](#endnote-ref-44)
45. Lu, C., et al. 2008, Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). (Proc. Natl. Acad. Sci. USA) 105: 4951–4956. [↑](#endnote-ref-45)
46. Luo YC, Zhou H, Li Y, Chen JY, Yang JH, Chen YQ, Qu LH, 2006, Rice embryogenic calli express a unique set of microRNAs, suggesting regulatory roles of microRNAs in plant post-embryogenic development, (FEBS Lett.) 580:5111-5116 [↑](#endnote-ref-46)
47. Y Zhang, X Chen, et al. 2016, miRNA: A Novel Link Between Rice Ragged Stunt Virus and Oryza sativa, (Indian Journal of Microbiology), 56:219-224 [↑](#endnote-ref-47)
48. L.Y. Li, C. Yang, Y. He, R.Q. Fang, Z.H. Tian, J.X. Li, 2014, Expression patterns of microRNAs in different organs and developmental stages of a superhybrid rice LYP9 and its parental lines, (Plant Biol.), 16:878-887 [↑](#endnote-ref-48)
49. Sieber et al., 2007 P. Sieber, F. Wellmer, J. Gheyselinck, J.L. Riechmann, E.M. Meyerowitz, 2007, Redundancy and specialization among plant microRNAs: role of the MIR164 family in developmental robustness (Development), 134 (2007), pp. 1051–1060 [↑](#endnote-ref-49)
50. R.S. Allen, et al. 2007, Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family (Proc. Natl. Acad. Sci. U. S. A.), 104 pp. 16371–16376 [↑](#endnote-ref-50)
51. Kertesz,M., Iovino,N., Unnerstall,U., Gaul,U. and Segal,E. 2007, The role of site accessibility in microRNA target recognition. (Nat. Genet.), 39:1278-1284. [↑](#endnote-ref-51)
52. Muckstein,U., Tafer,H., Hackermuller,J., Bernhart,S.H., Stadler,P.F. and Hofacker,I.L. 2006, Thermodynamics of RNA-RNA binding. (Bioinformatics), 22, 1177–1182. [↑](#endnote-ref-52)
53. D. Wang, K. Pei, Y. Fu, Z. Sun, S. Li, H. Liu, K. Tang, B. Han, Y. Tao, 2007,

    Genome-wide analysis of the auxin response factor (ARF) gene family in rice (Oryza sativa) (Gene), 394 pp. 13-24 [↑](#endnote-ref-53)
54. Jeon, J., Lee, S., Jung, K.H., Yang, W.S., Yi, G.H., Oh, B.G., and An, G., 2000, Production of transgenic rice plants showing reduced heading date and plant height by ectopic expression of rice MADS-box genes. (Mol. Breed.) 6:581-592. [↑](#endnote-ref-54)
55. M Wen, M Xie et al. 2016, Expression Variations of miRNAs and mRNAs in Rice (Oryza sativa). (Genome Biology and Evolution) 8:3529-3544 [↑](#endnote-ref-55)