**Background**

Single nucleotide polymorphisms (SNPs), generally defined as only a single base change in a DNA sequence among individuals [[[1]](#endnote-1)], quickly became one of the most popular molecular markers in plant molecular genetics with the advantage of being abundant, ubiquitous in most organisms and highly amenable to high- and ultra-high-throughput automation [[[2]](#endnote-2)-[[3]](#endnote-3)]. It also provided a powerful tool for marker-assisted breeding and quantitative trait locus (QTL) analysis and genome association analysis [3-[[4]](#endnote-4)]. The application of next-generation sequencing (NGS) technology has facilitated the identification of massive number of SNPs in various organisms through resequencing, including human [[[5]](#endnote-5)], rice [[[6]](#endnote-6)-[[7]](#endnote-7)[[8]](#endnote-8)], maize [[[9]](#endnote-9)-[[10]](#endnote-10)], soybean [[[11]](#endnote-11)], Arabidopsis [[[12]](#endnote-12)], etc. Genome-wide analysis revealed that SNPs were distributed unevenly on different genomic regions and fewer SNPs fell in some regions of higher conservation which are functional regions like coding sequences (CDSs) and regulatory elements [[[13]](#endnote-13)-[[14]](#endnote-14)]. SNPs found in plant genomes may cause serious phenotype changes or apparent trait variations. For example, while studying the loss of seed-shattering habit in rice domestication, an SNP was found in *qSH1* gene responsible for this important phenotype change event.

Recent years have witnessed numerous studies identifying and analyzing SNPs through resequencing with NGS, and in addition to great effort put on studying SNPs in protein-coding genes, several investigations of miRNA-related SNPs were performed in Arabidopsis [[[15]](#endnote-15)] as well as rice [[[16]](#endnote-16)-[[17]](#endnote-17)], in which researchers focused on the changes SNPs may bring to the miRNA structure stability and target alteration as well as miRNA evolution.

MiRNAs are small endogenous non-coding RNAs originated from endogenous loci. The miRNA gene loci are transcribed into self-complementary primary RNAs (pri-miRNAs) that can form hairpin structures [[[18]](#endnote-18)] in nucleus and then pri-miRNAs are excised to liberate precursor miRNAs (pre-miRNAs) by DCL1 [[[19]](#endnote-19)]. Next, a duplex of about 21nt RNAs with 3’ overhangs, is further produced from hairpin precursor, and after methylation on both 3’ nucleotides, the duplex would then be cleaved and loaded into AGO protein to form RNA-induced silencing complex (RISC) with a miRNA\* strand degraded [[[20]](#endnote-20)]. The remaining strand is the so-called mature miRNA and is thought to guide RISC to target RNAs through complementarity to the miRNA binding site. MiRNAs are key regulators in process of plant growth and development and often target genes that are themselves regulators such as transcription factors. Studies have reported SNPs involved in the miRNA-mediated gene silencing caused distinct changes to agronomic traits. In rice, it was reported that one point mutation in the osa-miR156 binding site of OsSPL14 perturbed the outcome of osa-miR156-mediated silencing, thus resulted in an ideal plant with reduced tiller number, increasing lodging resistance and enhanced grain yield [[[21]](#endnote-21)]. While in barley, SNPs perturbed the interaction between miR172 and its target gene HvAP2 and brought variations to the spike density of barley inflorescence [[[22]](#endnote-22)].

In plants, miRNAs repress the mRNAs of their targets with high complementarity mainly through transcript cleavage [[[23]](#endnote-23)], and this high complementarity requirement formed the basis of many bioinformatic software for miRNA target prediction, one of which is the web-based Plant Small RNA Target Analysis Server (psRNATarget) [[[24]](#endnote-24)]. Besides in silico method, recent years, there are several methods developed to verify the true miRNA:target relationship, such as overexpression of miRNA or miRNA-resistant target, RNA ligase-mediated 5’-RACE, degradome sequencing, and etc. [[[25]](#endnote-25)] Owing to the complexity of plant miRNA target recognition, bioinformatic methods may product miRNA targets that are not subjected to functionally relevant miRNA regulation [25], so how to filter the false-positives remain a headache.

Recent 3K rice genome project sequenced more than 3,000 rice cultivars and millions of genomic reads have been produced [[[26]](#endnote-26)]. Huge number of SNPs were identified by aligning the sequence reads [8], including some rare tri- and tetra-allelic SNPs which represent the precious genotypes of the minority population of cultivars. The abundance of SNPs provides good opportunity for genome-wide identification and analysis into SNPs in miRNA-mediated silencing processes. Since SNPs can reflect the variations of genomes of different rice cultivars, and in turn help us uncover the variations in miRNA-mediated regulation between rice cultivars and possible effects of SNPs involved in this process on the phenotypes of rice cultivars. Based on the SNP data derived from 3K rice genome project, we studied the SNP distribution on both rice miRNAs deposited in miRBase [[[27]](#endnote-27)] as well as identified miRNA targets and the effects SNPs brought to the miRNA:target interactions in rice. The relationship between variations of miRNA:target interactions and that of phenotypes was further analyzed.

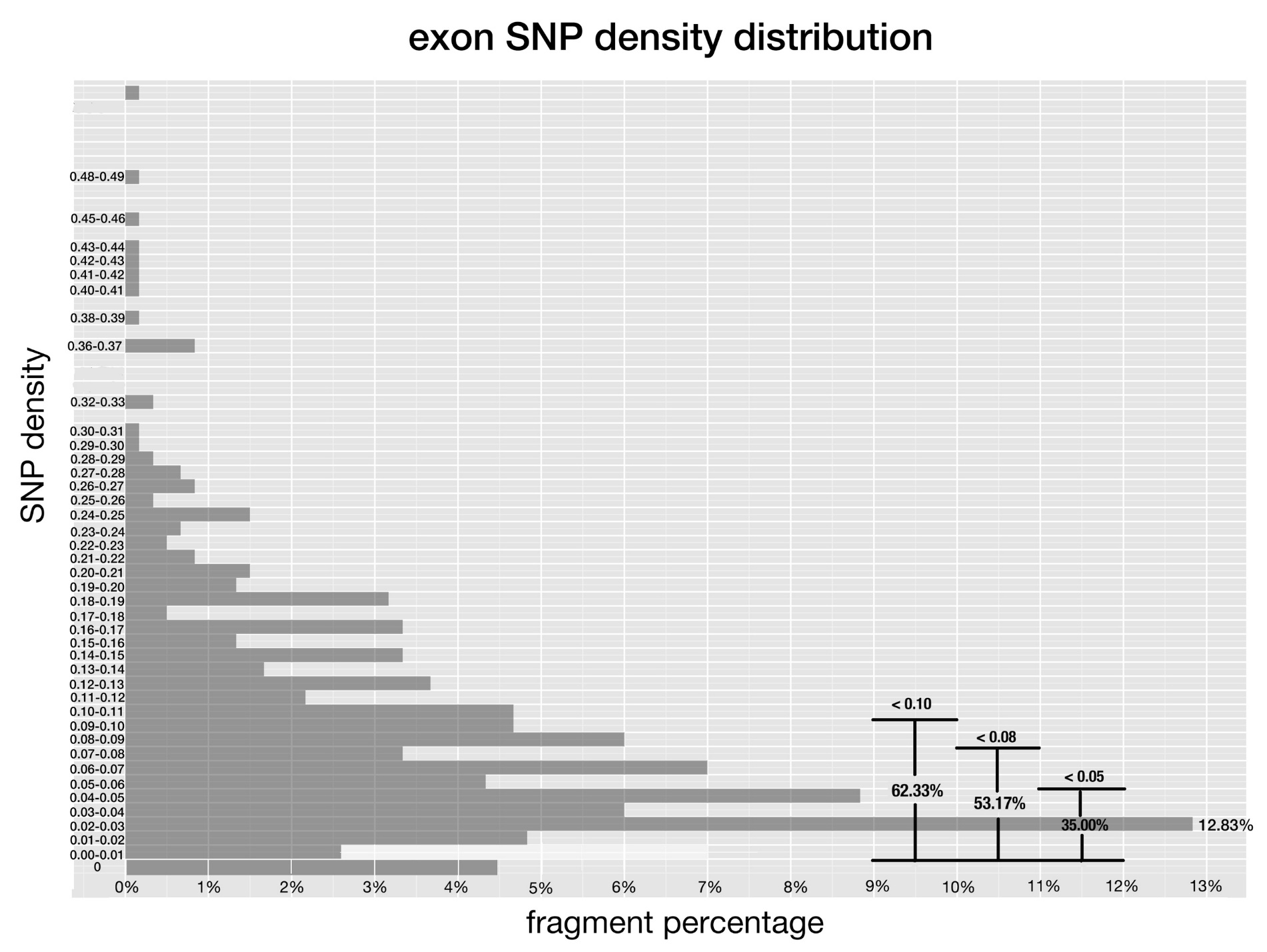
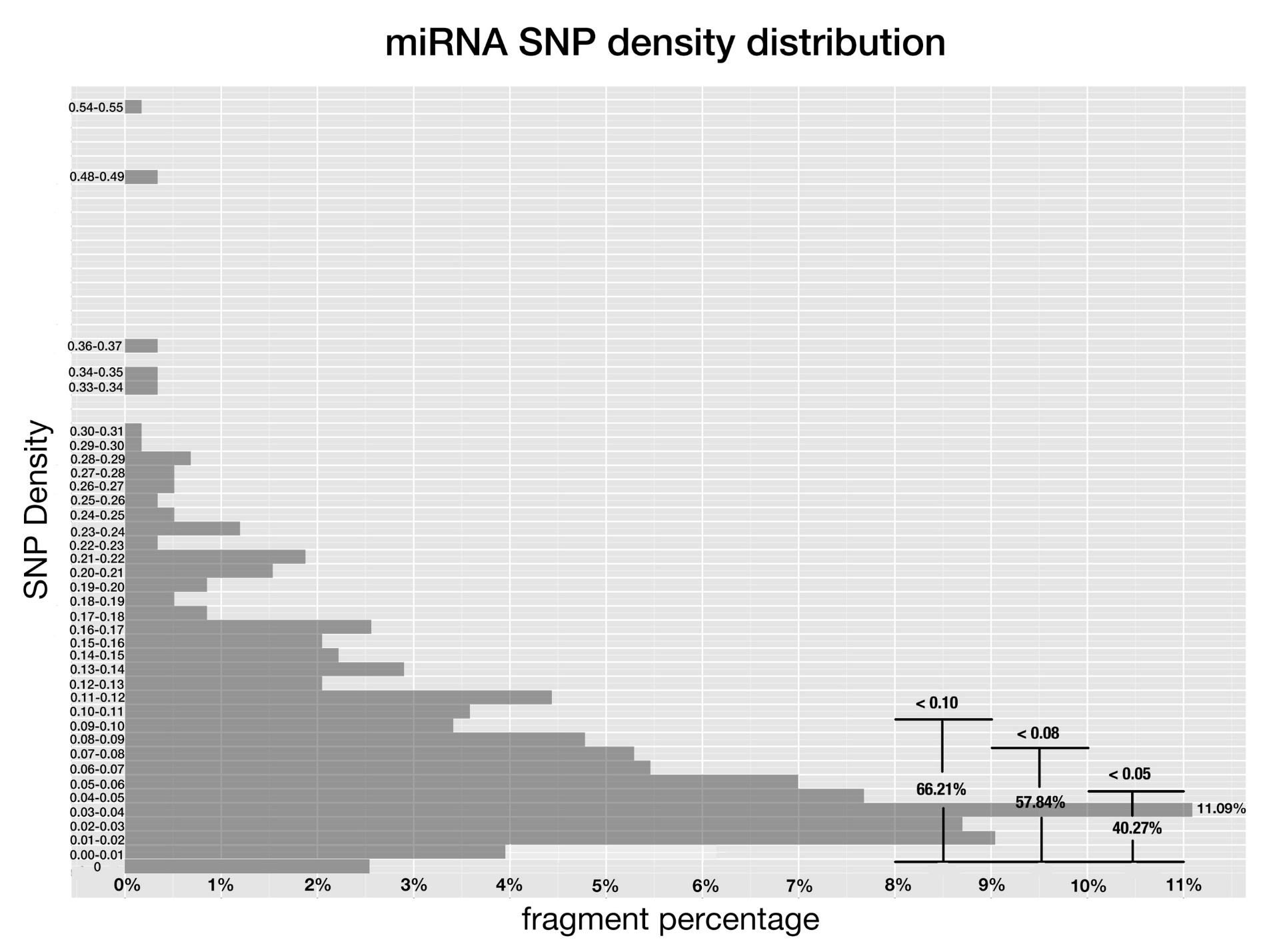
**Results**

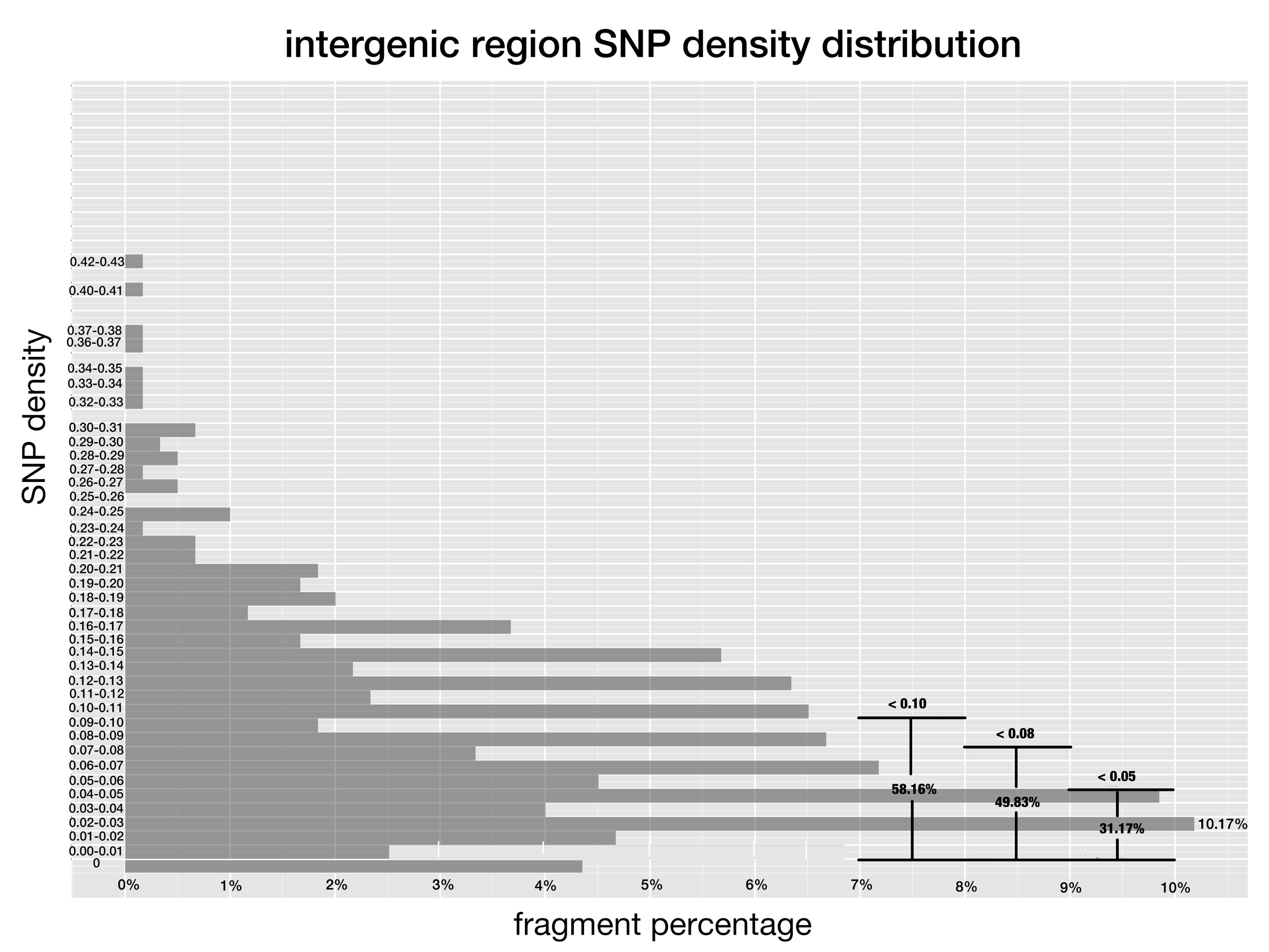
**Overall SNP densities of rice miRNAs from 3K Rice Genome Project**

SNP density can reflect the selection pressure imposed on given genomic regions. The higher the pressure is; the lower the SNP density will be [[[28]](#endnote-28)], and vice versa. Since miRNAs are functional genomic units and master regulators, they should have gone through different selection pressure compared with intergenic regions, and exons [[[29]](#endnote-29)]. To investigate this, SNPs were collected from Rice SNP-Seek Database (derived from 3,000 Rice Genome Project, *snp-seek.irri.org/*) [8], for all pre-miRNAs deposited in *miRBase.org* (miRBase release v20) [27] and randomly chosen intergenic regions and exons across the rice genome. Then SNP density distribution, defined as the division of SNP number by the length of the genetic region, of rice pre-miRNAs, intergenic regions and exons were plotted respectively and compared.

As expected, SNP density of pre-miRNAs and exons were found to be lower than that of intergenic regions. For pre-miRNAs, the percentage of genomics fragments kept

rising with the SNP density until it peaked at the range of 0.03-0.04. It then decreased gradually after 0.04 (Fig. 1). The same trend was observed for the overall SNP distribution in exons sampled in this study, except that the abundance of corresponding fragments peaked at the SNP density range of 0.02-0.03 with a percentage of 12.83 (Fig. 2). No such trend was seen for the intergenic regions analyzed (Fig. 3). Through comparing the percentage of fragments fell within the lower SNP density ranges of 0-0.10, 0-0.08 and 0-0.05 individually (Fig. 1, 2, 3), it is clear that most pre-miRNAs clustered at similar density as exons, where significantly less intergenic regions fell into the same ranges. This demonstrates that pre-miRNAs have gone through stricter evolutionary selections than both intergenic regions and exons, which is consistent with their established roles as master regulators in many genetic pathways.

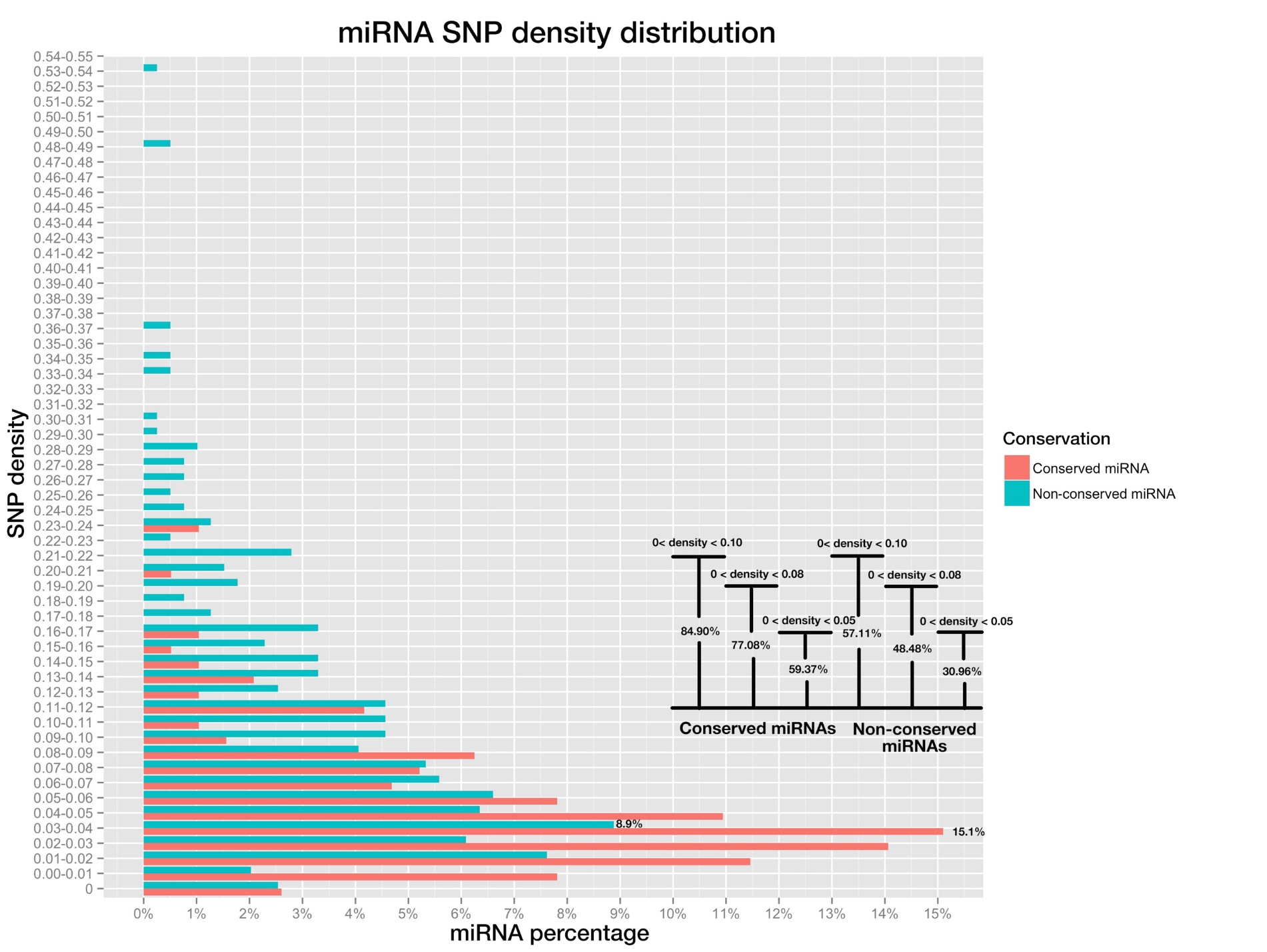




**Fig 1-3. SNP density of pre-miRNAs, exon regions and intergenic regions in rice.**

*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.*

Due to their evolutionary conservation [20, [[30]](#endnote-30), [[31]](#endnote-31)], conserved rice miRNAs should have lower SNP density than that of non-conserved ones. This is the case as shown in Figure 4, most conserved miRNAs clustered at lower SNP density ranges compared to non-conserved ones after comparing the percentage of fragments fell within the lower SNP density ranges of 0-0.10, 0-0.08 and 0-0.05, respectively.



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

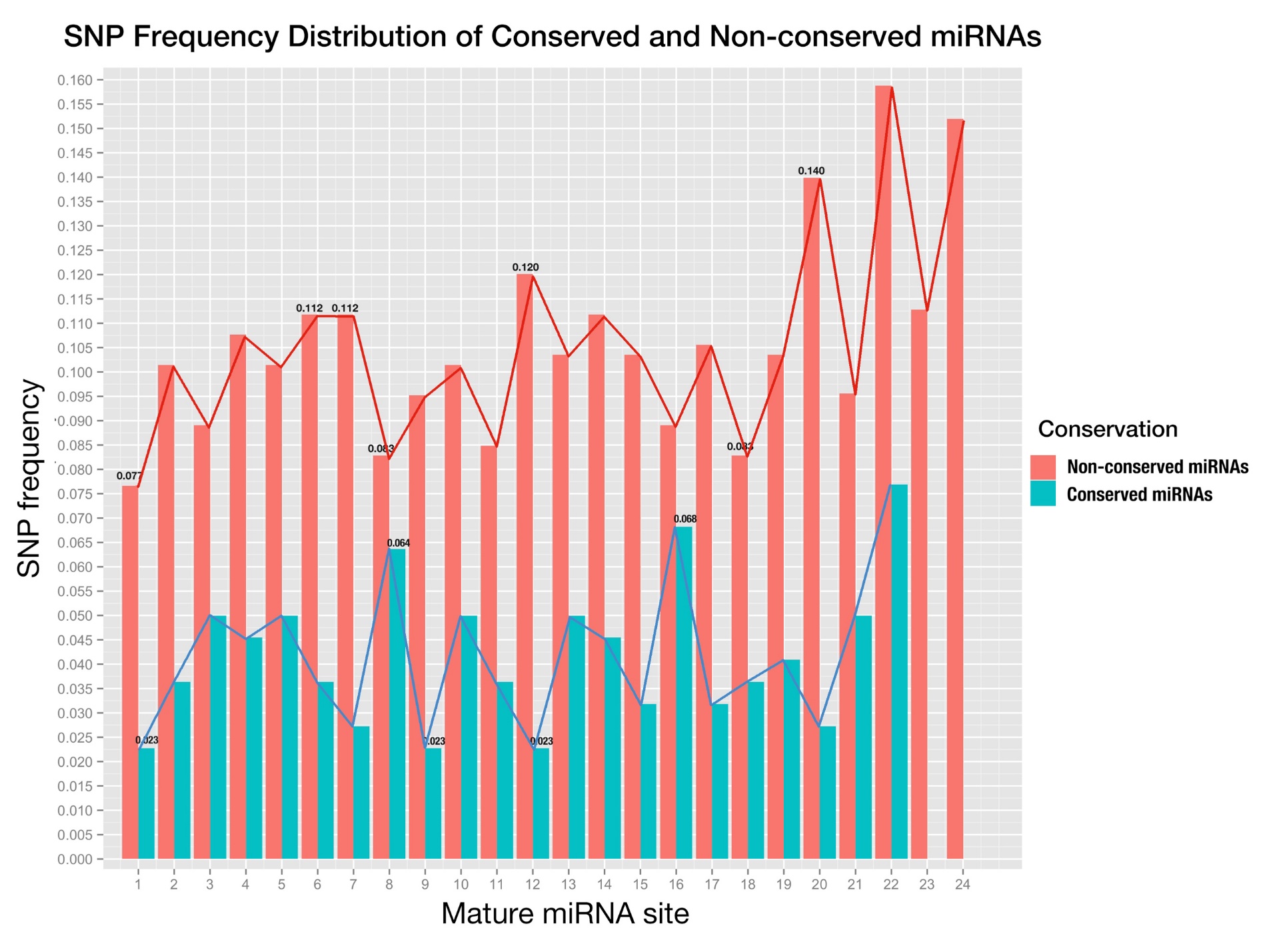
*Bar plot on the bottom right shows the percentage of miRNAs whose SNP densities are below 0.10, 0.08 and 0.05, for conserved and non-conserved miRNAs respectively.*

**Positional SNP distributions are differential in conserved and non-conserved miRNAs**

Plant miRNAs have variable efficacies toward different targets. This can be attributed to their complementarity patterns, which are determined by mature miRNAs and the cognate target sites co-jointly. Moreover, there is evidence suggesting that different positions along the 21nt mature miRNA functionally diversify in target recognition and silencing [[[32]](#endnote-32)], as mutations on certain positions cause complete abortion of silencing while others do not have obvious impact [30, [[33]](#endnote-33), [[34]](#endnote-34)]. This may also be reflected by the differential SNP frequency of each position, as lower frequency implies higher selection pressure. Therefore, collective analyses of SNP frequencies of individual positions along mature miRNAs may further reveal their differential functional importance in silencing.

Positional SNP frequencies of conserved and non-conserved mature miRNAs were calculated separately and compared. Since most miRNAs are 21-nt in length, this study focused only on position 1 to 21. As expected, all positions on conserved miRNAs had lower SNP frequencies than corresponding positions on non-conserved miRNAs. Despite so, it was expected that they would share similar rankings of SNP frequencies among different positions, given both are under the same selection pressure mechanistically. However, this doesn’t appear to be the case, except position one was the lowest SNP frequencies for both conserved and non-conserved miRNAs, which may be explained by its importance in the loading of miRNAs into the AGO proteins [[[35]](#endnote-35), [[36]](#endnote-36)]. Position 8 and 12 had inverted rankings among conserved and non-conserved mature miRNAs, and the Pearson correlation test of the SNP frequencies showed no significant correlation between them (r=-0.163, p-value=0.4473). And the differential ranking showed the differential selection pressure on each position, thus suggested of the differential function mechanisms of miRNAs toward targets between conserved and non-conserved miRNAs.

Pairing at position 10 and 11 is thought to be critical for plant miRNA functioning through cleavage [[[37]](#endnote-37)-[[38]](#endnote-38)[[39]](#endnote-39)], which adds another level of restriction on the evolution of both positions and this may result in the lower SNP densities than other miRNA positions. But the SNP frequencies of both positions were not found among the lowest SNP frequency positions, not consistent with the deduced hypothesis from empirical claims.



**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;*

**Positive correlations between positional SNP frequencies of conserved miRNAs and cognate targets reveal co-evolving constrain**

To further utilize the 3K data to investigate the functional correlations between SNP frequencies along mature miRNAs and the corresponding positions on its binding sites, may reveal their functional importance.

Since conserved miRNAs have more identifiable targets than non-conserved miRNAs in plants [20] and are functionally more important, we only studied the binding sites of genes targeted by conserved miRNAs here. Co-evolution of miRNA and its binding site at target genes had been reported by several studies before [37, [[40]](#endnote-40)]. Then the SNP frequencies of positions along mature miRNA would have positive correlations with those of their cognate binding sites if they have co-evolved during selection history. With the help of online miRNA target prediction tool, *psRNATarget* [24] and transcriptome-wide degradome validation of rice miRNA targets [[[41]](#endnote-41)], a total number of 823 genes were found being targeted by conserved miRNAs. Then distribution of SNP frequencies of positions along both conserved miRNAs and their binding sites was shown side-by-side in a bar-plot (Fig. 6). And then, a Pearson correlation test was performed on the SNP frequencies of mature miRNAs and their binding sites. The result of the test (r=0.5891, p-value=2.455e-3) showed that there was moderate positive correlation between them, and furthermore, suggested the co-evolution of miRNA and binding site of its cognate target. In addition, compared with conserved mature miRNAs, SNP frequency of each position of miRNA binding site was higher, and this suggested that selection pressure upon conserved mature miRNAs is higher than that upon binding site on cognate targets.



**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);*

**No obvious correlation was found for the expression of confirmed miRNA:target pairs**

In contrast to animal miRNAs, plant miRNAs target genes with high degree of complementarity, which allows the confident prediction with bioinformatics tools. But how to filter those false positive predictions remains a headache. Under the general assumptions that complementarity is the sole determinant of silencing and that any given plant miRNA-loaded RNA-induced silencing complex (miRISC) is able to act independently [[[42]](#endnote-42), [[43]](#endnote-43)], target mRNAs are downregulated by corresponding miRNAs and the expression level of plant miRNA is negatively correlated with that of cognate target mRNAs, which was also supported by experiment in which five mRNAs encoding TCP transcription factors are downregulated in plants overexpressing miR319 [[[44]](#endnote-44)]. To verify this hypothesis, expression data of rice miRNAs and genes was extracted from RiceFREND database [[[45]](#endnote-45)], and correlation test was performed on miRNA:target pairs validated by degradome with 3-week old rice seedling samples [36].

Only 136 out of 367 miRNA:target pairs were examined to be negatively correlated on their expression level (Fig. 7) and this strongly objects the hypothesis that there is negative expression correlation between confirmed targets and their miRNAs. Furthermore, more than half (197 out of 367) pairs fell in the range of weak correlation (-0.4~0.4), which there were even no direct correlation between these pairs.



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

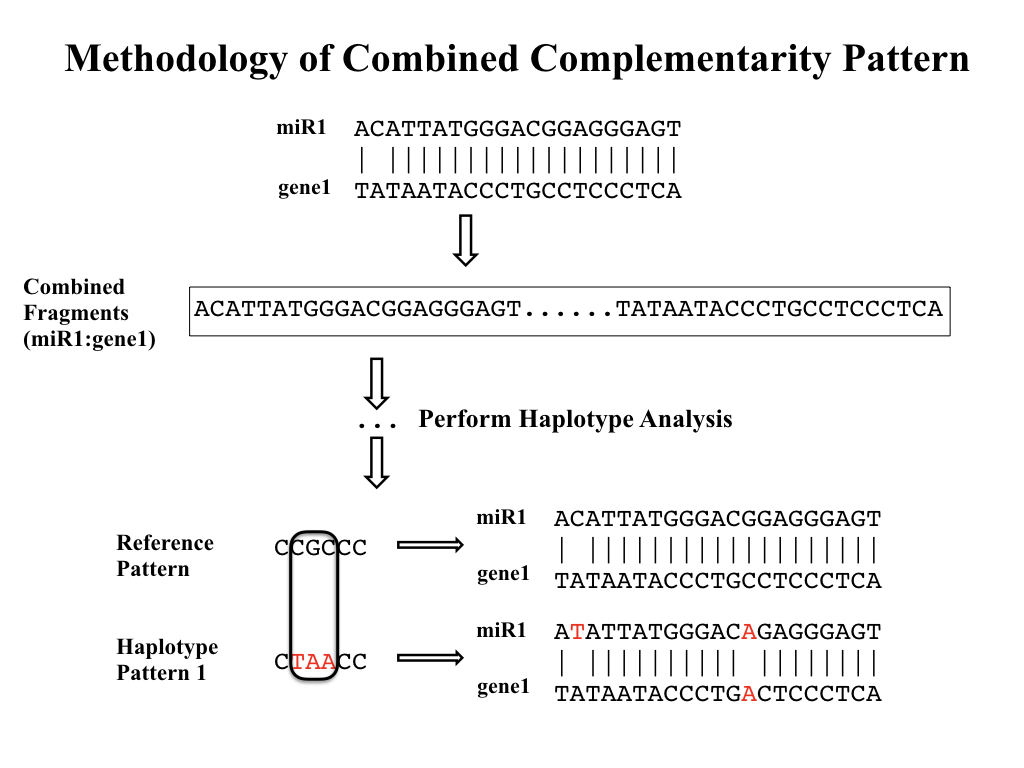
*X-axis is the number of pairs that fall on the specified range of correlation coefficient; so bars with green color in the lower part denote the negatively correlated miRNA:target pairs, by contrast, bars with brown color in the upper part denote the positively correlated miRNA:target pairs.*

**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.

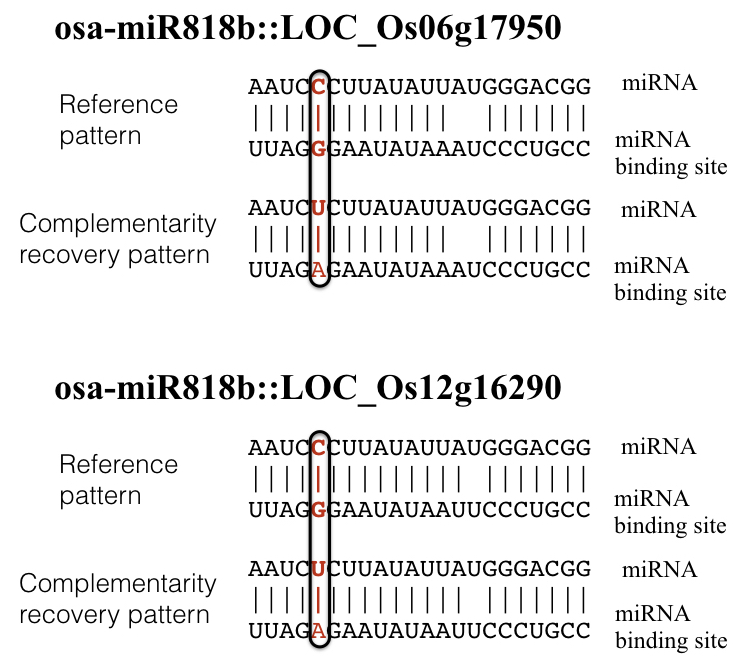


**Fig 8. Workflow of Combined Complementarity Pattern Analysis (CCPA)**

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 [[[46]](#endnote-46)]. 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 [[[47]](#endnote-47)], but changes of expression of osa-miR818 was detected in japonica rice infected with rice ragged stunt disease [[[48]](#endnote-48)] and two target genes with unknown functions were examined for osa-miR818 by L. Y. Li et al. [[[49]](#endnote-49)]. 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**

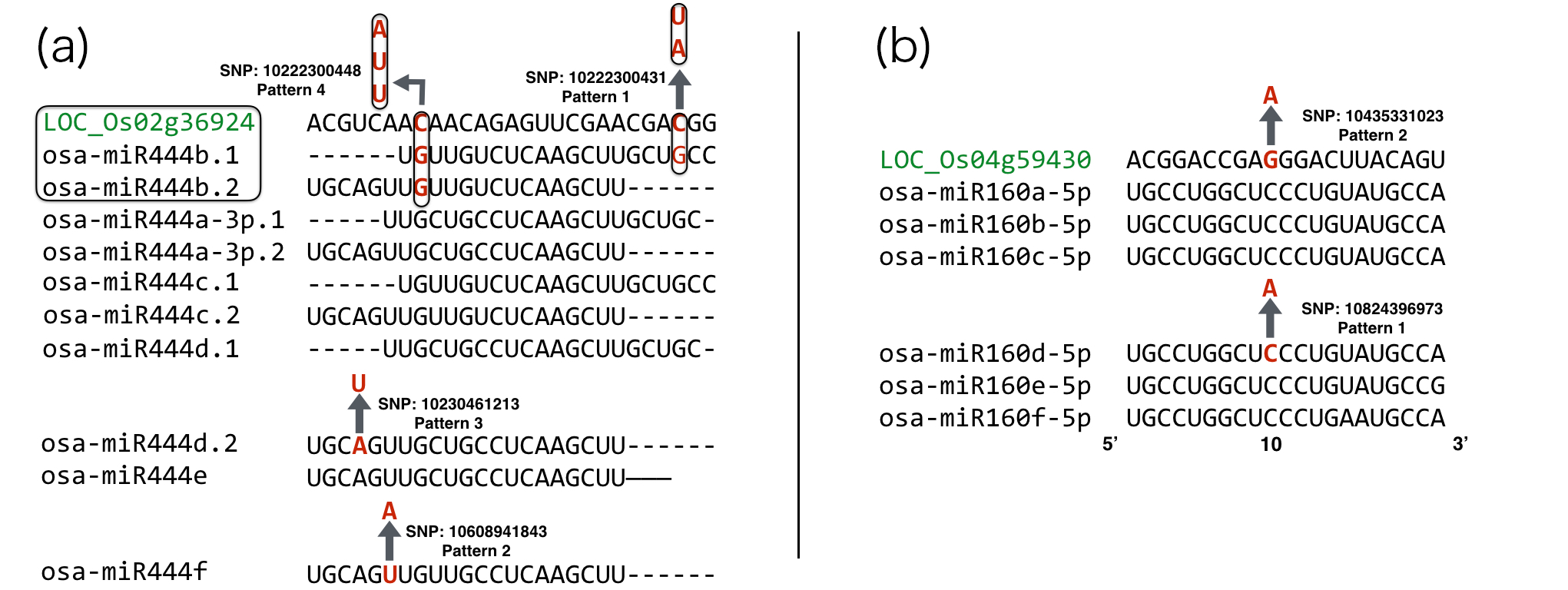
*In the complementarity pattern, a vertical bar would be placed denoting match, a blank denotes mismatch.*

**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 [[[50]](#endnote-50)-[[51]](#endnote-51)], 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 [[[52]](#endnote-52)]. 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 |

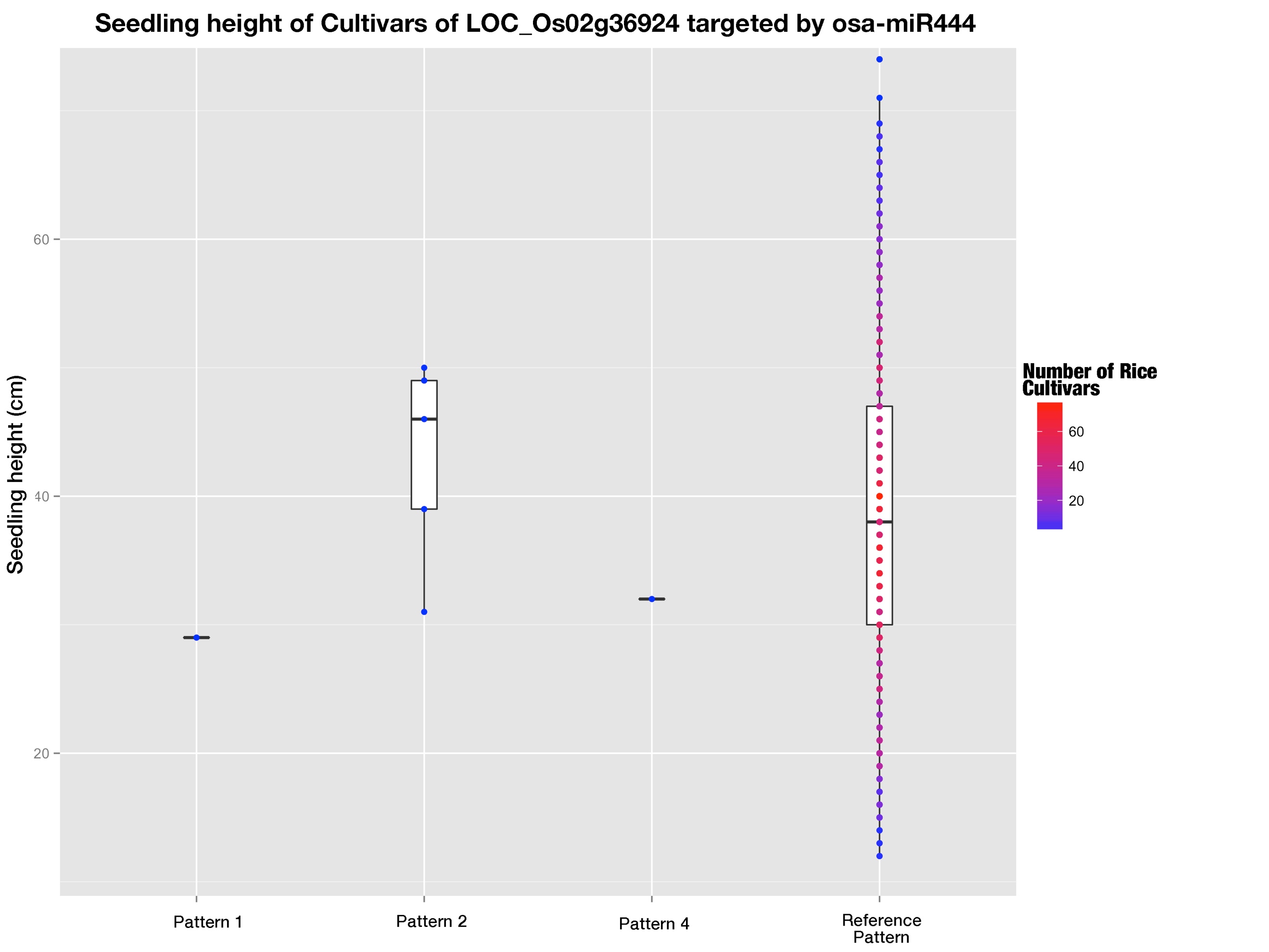
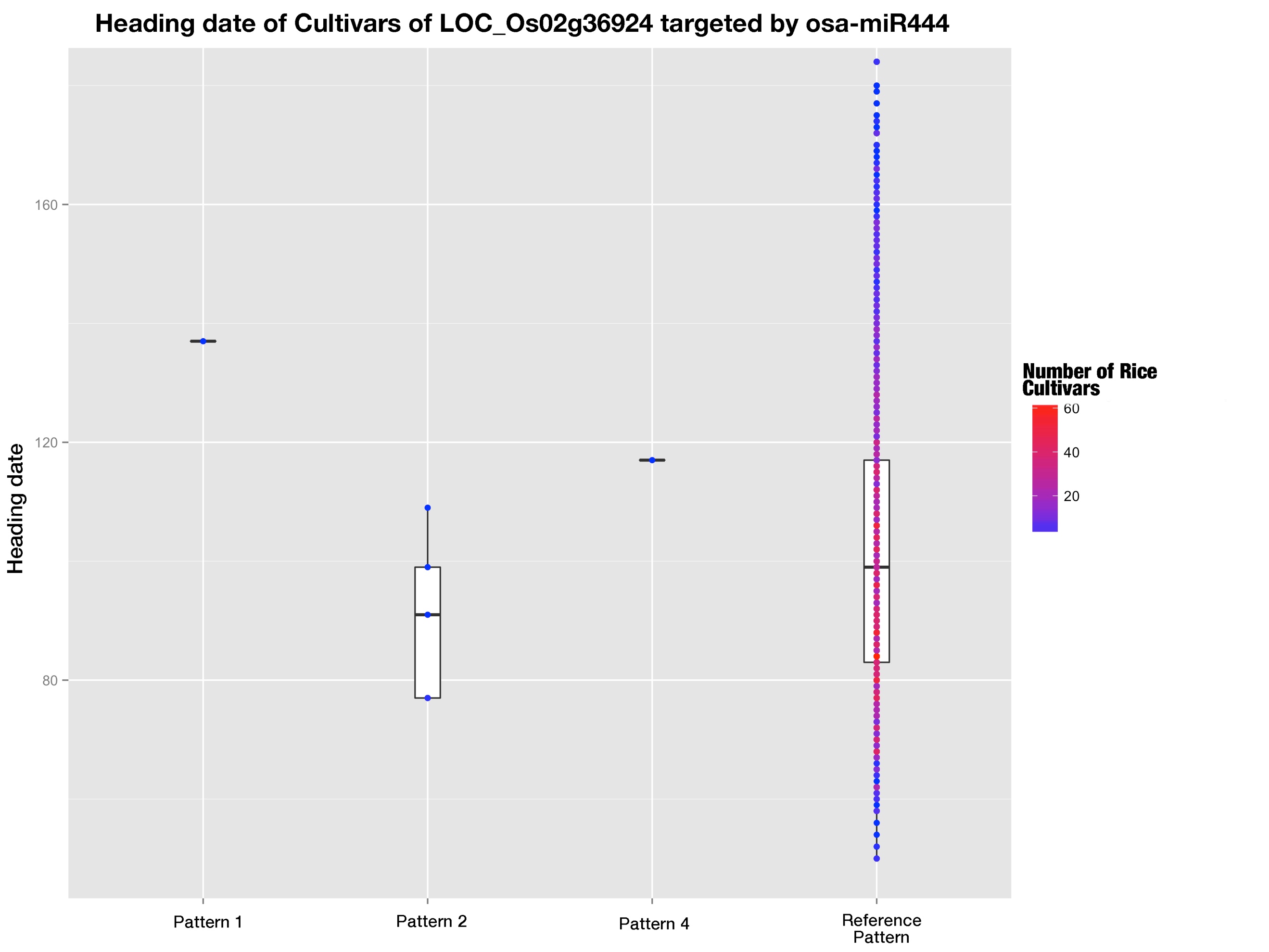
*Position on miRNA binding site is in the order of 5’ to 3’ on mature miRNA, and total free energy of binding was calculated using RNAup program in Vienna Package [[[53]](#endnote-53)].*

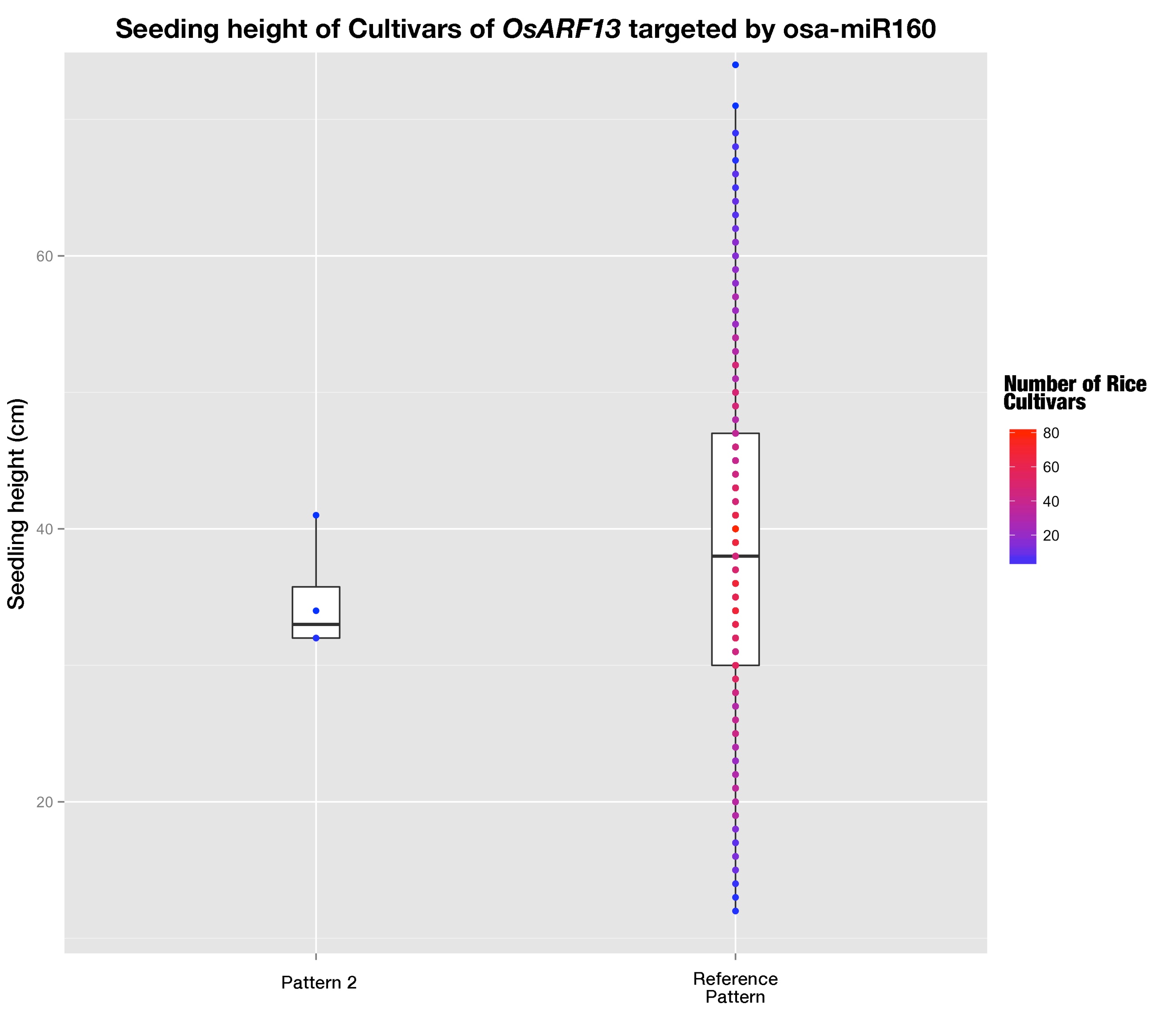
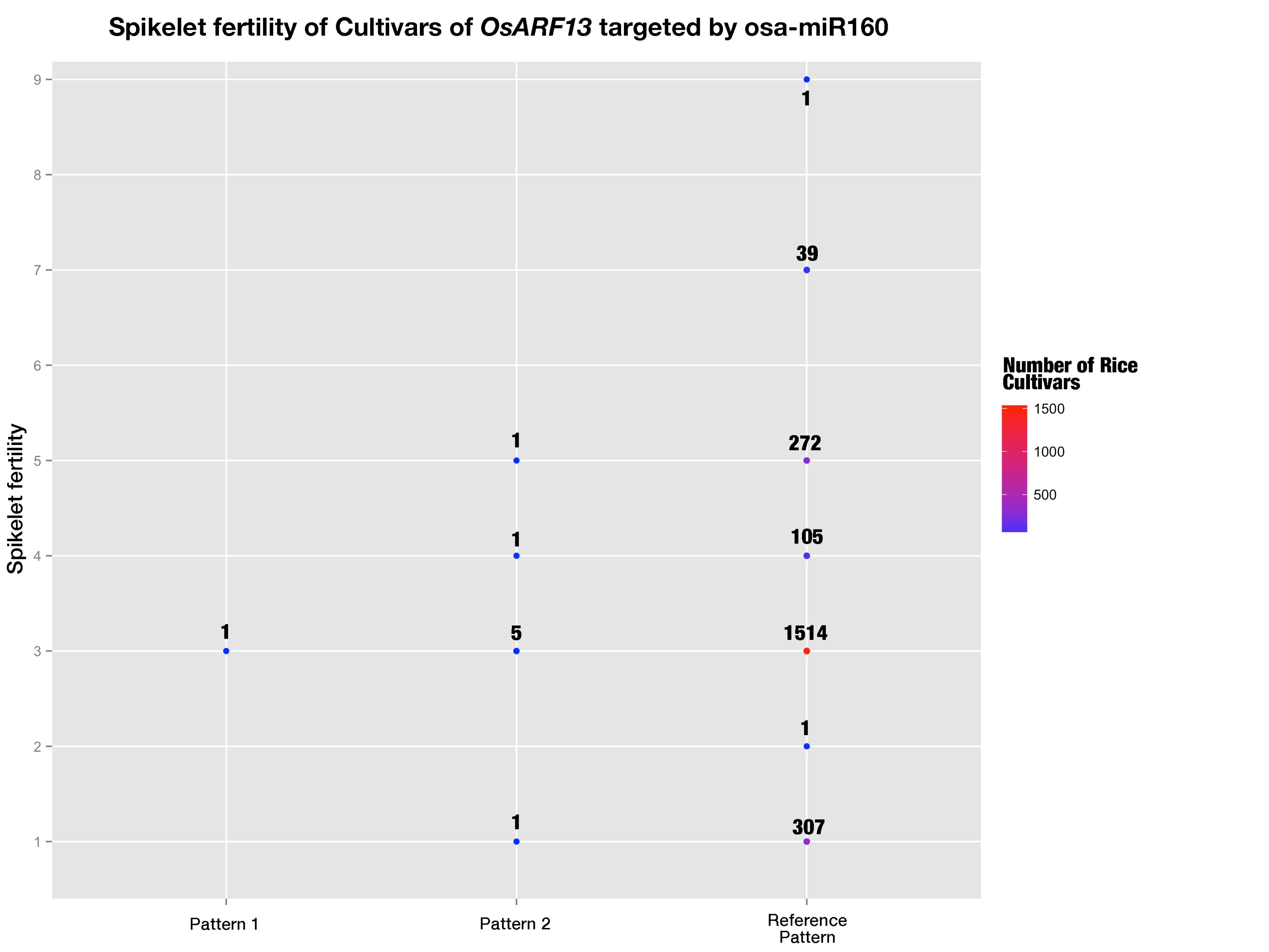


**Fig 10. Complementarity pattern of 2 target genes with their targeting miRNA family**

*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 [[[54]](#endnote-54)] and rice MADS-box genes were reported to be involved with heading date and plant height [[[55]](#endnote-55)]. 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 [[[56]](#endnote-56)]. 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. Ehrenreich IM, Purugganan MD, 2008, Sequence variation of microRNAs and their binding sites in Arabidopsis. (Plant Physiol) 146:1974-1982. [↑](#endnote-ref-15)
16. 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-16)
17. 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-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. Liu Q, Wang F, Axtell M.J., 2014, Analysis of complementarity requirements for plant microRNA targeting using a Nicotiana benthamiana quantitative transient assay. (Plant Cell) 26: 741-753 [↑](#endnote-ref-32)
33. 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-33)
34. 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-34)
35. 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-35)
36. Mallory A, Vaucheret H, 2010, Form, function, and regulation of ARGONAUTE proteins. (Plant Cell) 22: 3879–3889 [↑](#endnote-ref-36)
37. 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-37)
38. 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-38)
39. 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-39)
40. 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-40)
41. 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-41)
42. Iwakawa, H. and Tomari, Y. 2013, Molecular insights into microRNA-mediated translational repression in plants. (Mol. Cell) 52:591-601 [↑](#endnote-ref-42)
43. Tang, G. et al. 2003, A biochemical framework for RNA silencing in plants. (Genes Dev.) 17:49-63 [↑](#endnote-ref-43)
44. 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-44)
45. 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-45)
46. 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-46)
47. 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-47)
48. 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-48)
49. 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-49)
50. 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-50)
51. 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-51)
52. 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-52)
53. 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-53)
54. 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-54)
55. 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-55)
56. 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-56)