**Background**

A single nucleotide polymorphism (SNP), is defined as a single nucleotide variation in a peculiar genomic sequence among individuals within a population of a certain species [[[1]](#endnote-1)]. SNPs have quickly become the most popular molecular markers in plant molecular genetics for their abundance and ubiquity in most species. They are highly amenable to high- and ultra-high-throughput automation [[[2]](#endnote-2)-[[3]](#endnote-3)], being widely utilized in marker-assisted breeding, quantitative trait locus (QTL) analysis and genome association analysis [3-[[4]](#endnote-4)].

Recently, next-generation sequencing (NGS) technology has greatly facilitated the identification of massive number of SNPs in various organisms, 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)] and Arabidopsis [[[12]](#endnote-12)]. Genome-wide analyses have revealed that SNPs are distributed unevenly. Generally, there are fewer SNPs in regions with higher conservation, most of which are functional regions like protein coding sequences (CDSs) and regulatory elements [[[13]](#endnote-13)-[[14]](#endnote-14)]. SNPs found in plant genomes can cause significant changes to agronomic traits. For example, while studying the loss of seed-shattering habit during rice domestication can be attributed an SNP was found in *qSH1* (quantitative trait locus of seed shattering on chromosome 1) gene responsible for this important phenotype change event.

MiRNAs are small regulatory RNAs originated from the endogenous loci in plants and animals. The miRNA gene is transcribed into self-complementary primary RNA (pri-miRNA) and afterwards excised to be miRNA precursor (pre-miRNA). The pre-miRNA is processed again by DCL to produce a 21nt long duplex. One strand of the duplex, the so-called mature miRNA is incorporated into the RNA Induced Silencing Complex (RISC), guiding it to target mRNA through sequence complementarity and mediating the subsequent silencing mainly through transcript cleavage.

It is widely accepted that plants miRNAs are high complementary to their targets, and this formed the basis of many bioinformatic software for target prediction [[[15]](#endnote-15)]. However, the biological significance of many predicted targets is not guaranteed. 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. [[[16]](#endnote-16)] Owing to the complexity of plant miRNA target recognition, bioinformatic methods may produce miRNA targets that are not subjected to functionally relevant miRNA regulation [16], so how to filter the false-positives remain a headache.

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 causing distinct changes to agronomic traits. For example, one point mutation in the osa-miR156 binding site of OsSPL14 (SOUAMOSA PROMOTER BINDING PROTEIN-LIKE 14) perturbed the outcome of osa-miR156-mediated silencing, thus resulted in reduced tiller number, increasing lodging resistance and enhanced grain yield [[[17]](#endnote-17)]. While in barley, SNPs perturbed the interaction between miR172 and its target gene HvAP2 and brought variations to the spike density of barley inflorescence [[[18]](#endnote-18)].

Great effort has been put on studying SNPs in protein-coding genes, while several investigations of miRNA-related SNPs were performed in Arabidopsis [[[19]](#endnote-19)] as well as rice [[[20]](#endnote-20)-[[21]](#endnote-21)], where researchers focused on the changes SNPs may bring to the miRNA structure stability and target alteration as well as miRNA evolution and they found SNPs within pre-miRNA stems changed the energy of folding into secondary structure ,which may potential affect the mature miRNA production, and the putative domestication-related miRNAs were found to have lower SNP density than others which in suggests of natural or artificial selection. While how SNPs affect the interaction between plant miRNAs and their cognate targets remained unsearched and furthermore, change the agronomic traits involved in miRNA-mediated regulation, remained unstudied.

Recently, the 3K rice genome project sequenced more than 3,000 rice cultivars and obtained millions of genomic reads [[[22]](#endnote-22)]. A large number of SNPs were identified by aligning the sequence reads [8], including some rare tri- and tetra-allelic SNPs. The abundance of SNPs provides a good opportunity for genome-wide identification and analysis of SNPs involved in miRNA-mediated silencing. Since SNPs can reflect genomic variations different rice cultivars, by analyzing them in-depth, variations of miRNA-mediated regulation and their possible phenotypic effects may be uncovered. Here, we studied SNP distributions on selected rice miRNAs their cognate miRNA targets in over 3000 rice cultivars, in the hope to reveal their impacts on miRNA:target interactions. Furthermore, the potential relationship between variations of miRNA:target interactions their possible phenotypic effects was analyzed.

Logic and contents of introduction:

1. What is SNP?

2. Types of SNPs and SNP frequencies in different genomic regions.

3. Why is studying SNP important? What are the applications of SNPs? Describe specific examples related to your research. How NGS facilitates SNP studies？

4. What is miRNA? Why miRNA is important?

5. Why studying SNPs in miRNAs are important? What are the previous

6. What are you trying to study? Why is this important?

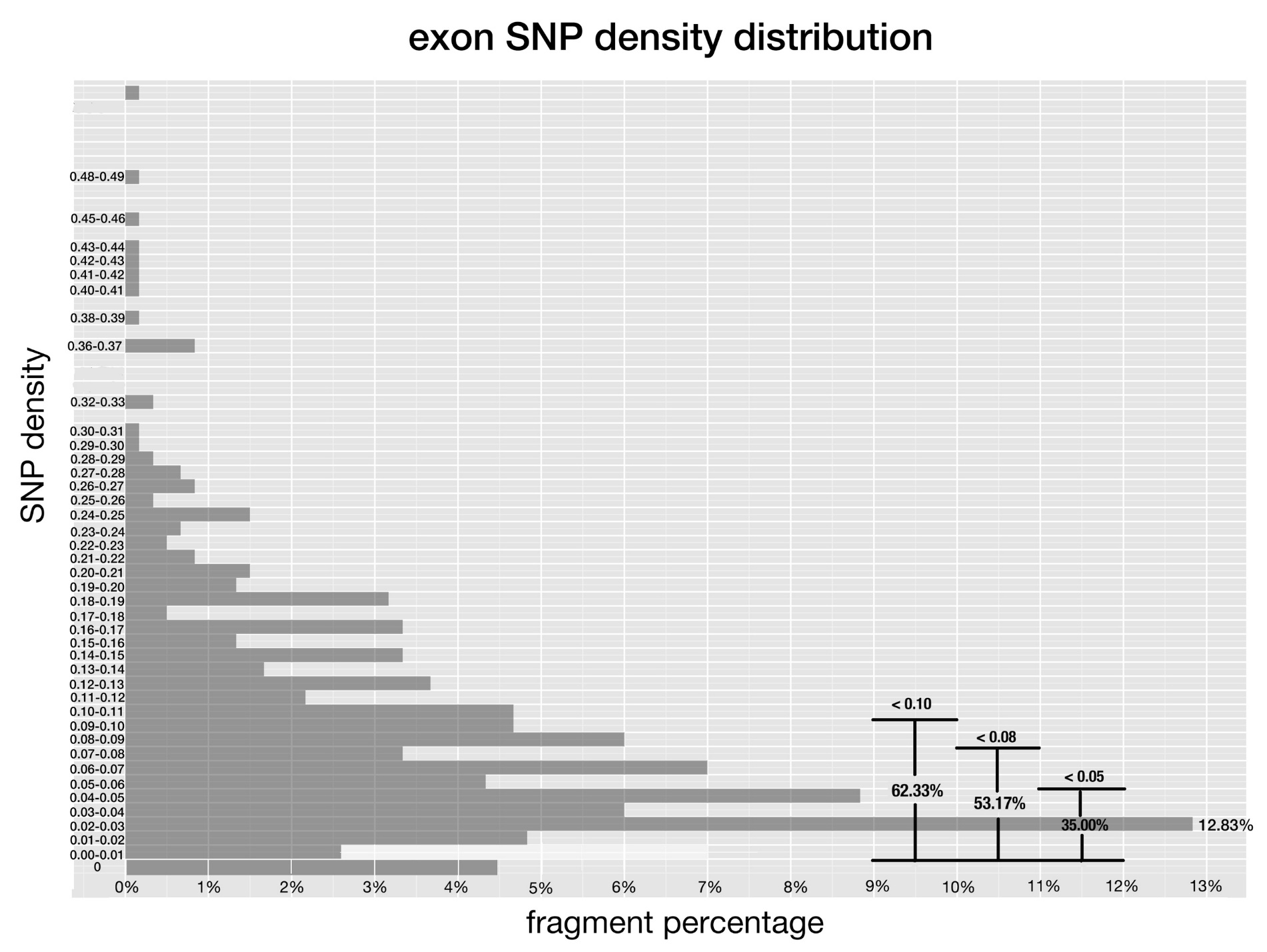
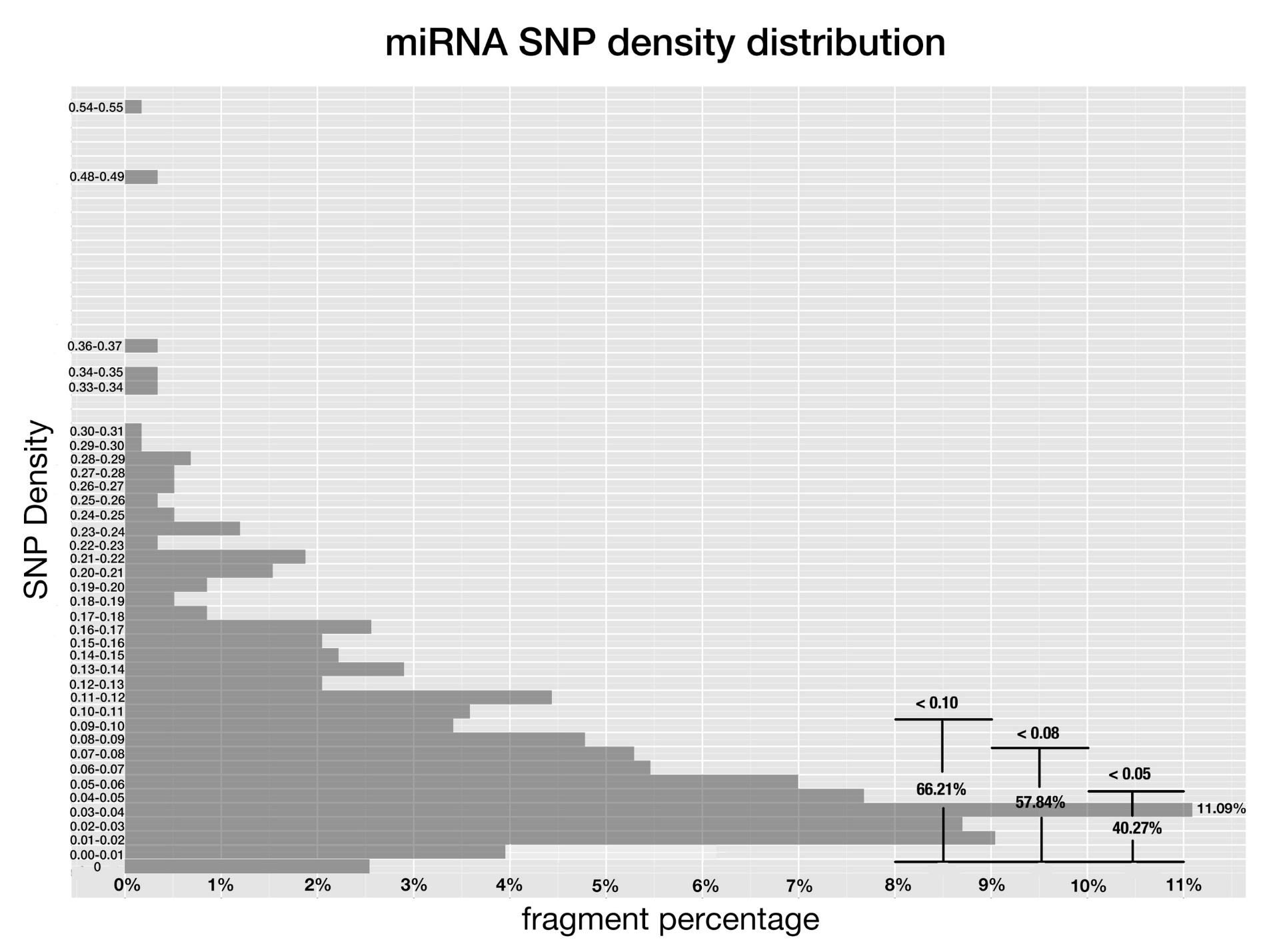
I can’t do much editing if your logic is not sound and your writing is not good. I am not supposed to write it for you. And I don’t want to either.

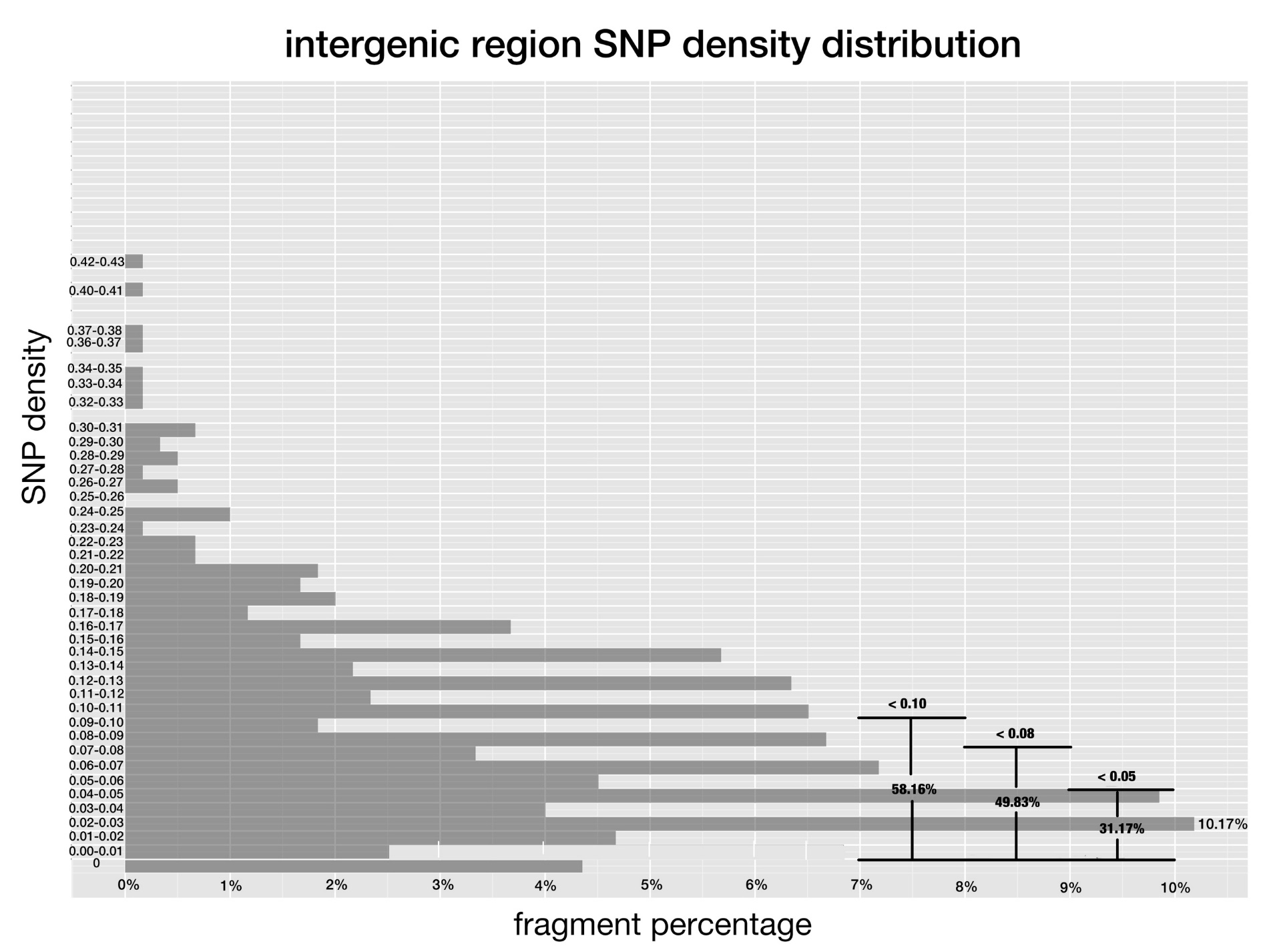
**Results**

**The overall distribution of SNP density in rice miRNAs**

SNP density can reflect the selection pressure imposed on given genomic regions. The higher the pressure; the lower the SNP density [[[23]](#endnote-23)], 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 [[[24]](#endnote-24)]. 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) [[[25]](#endnote-25)]. Randomly chosen fragments of 150 bp in length of intergenic regions and exons across the rice genome were included as controls. SNP densities, defined as the division of the total SNP number of a genetic region by its length were plotted respectively and compared.

As expected, the SNP density of pre-miRNAs and exons were found to be lower than that of the intergenic regions. For pre-miRNAs, the percentage of genomic fragments kept rising with the SNP density until it peaked at the range of 0.03-0.04 (Fig. 1). It then decreased gradually after 0.04. The same trend was observed for the overall SNP distribution in exons sampled, except that the abundance of corresponding fragments peaked at the SNP density range of 0.02-0.03 (Fig. 2). No such trend was seen for the intergenic regions analyzed (Fig. 3). Through comparing the percentage of genomic 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 was clear that more pre-miRNAs clustered at similar density than exons, where significantly much less intergenic regions fell into the same ranges. HIGHLIGHT YOUR SIGNIFICANT FINDINGS: how did you get to the conclusion that the SNP densities of pre-miRNAs are lower than exons??? 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.

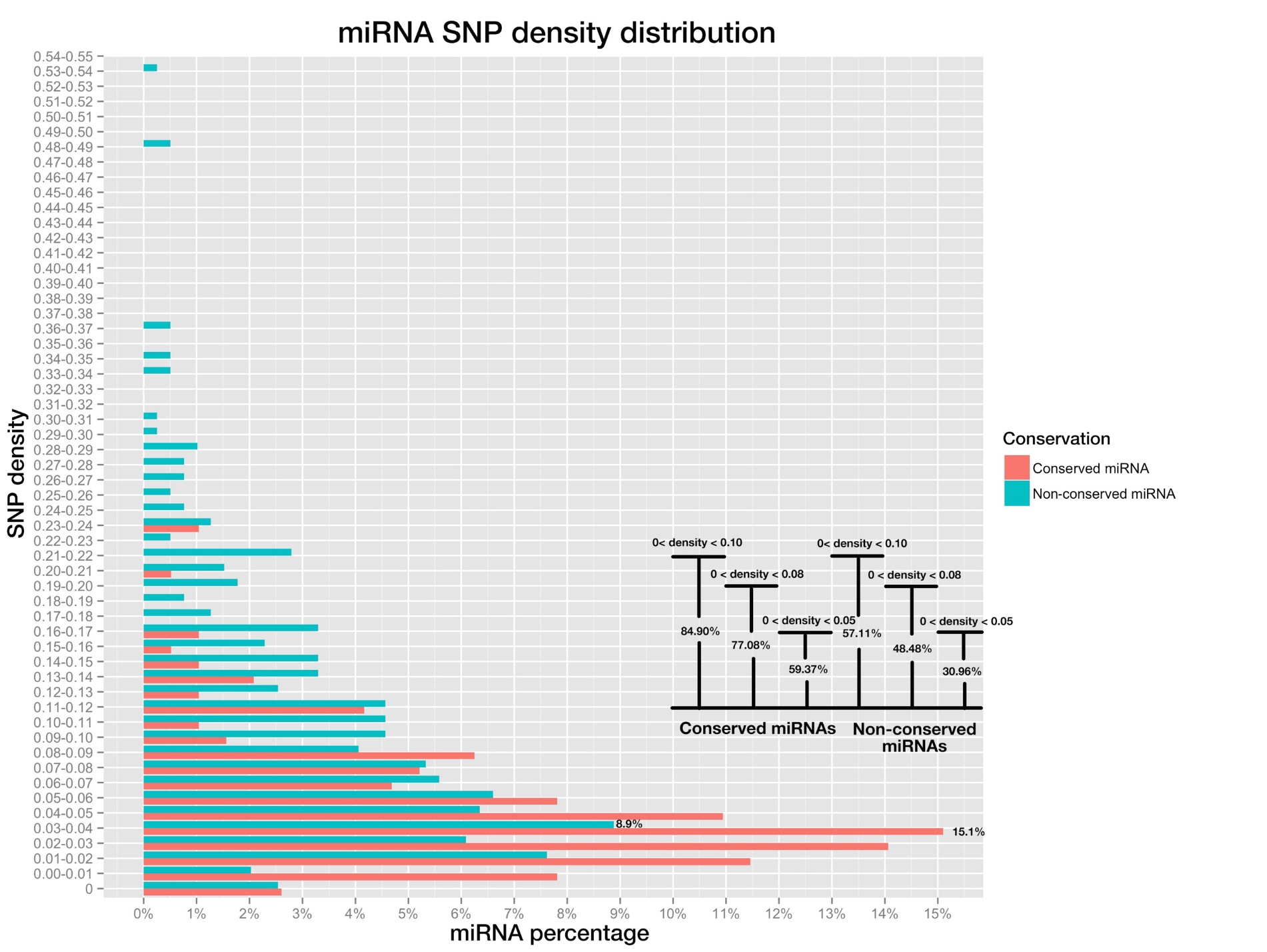




**Fig 1-3. SNP density of pre-miRNAs, exon regions and intergenic regions in rice.**

*SNP density is the division of total numbers of SNP on a given genomic fragment by its length in base pair. And x-axis corresponds to the percentage of fragments that have SNP density at the given range.* YOU HAVE TO DESCRIBE YOUR FINDINGS HERE!

Due to their evolutionary conservation [[[26]](#endnote-26), [[27]](#endnote-27), [[28]](#endnote-28)], conserved rice miRNAs should have lower SNP density than that of non-conserved ones. This was the case as shown in Figure 4. More conserved miRNAs clustered at lower SNP density ranges compared to non-conserved ones, comparing the percentage of fragments fell within the lower SNP density ranges of 0-0.10, 0-0.08 and 0-0.05, respectively. DECRIBE YOUR FINDINGS!!!!! WHAT ARE THE DIFFERENCES BETWEEN CONSERVED AND NON-CONSERVED MIRNAS IN TERMS OF SNP DENSITIES?



**Fig 4. SNP density distribution on pre-miRNAs of conserved miRNAs and non-conserved miRNAs (blue color).**

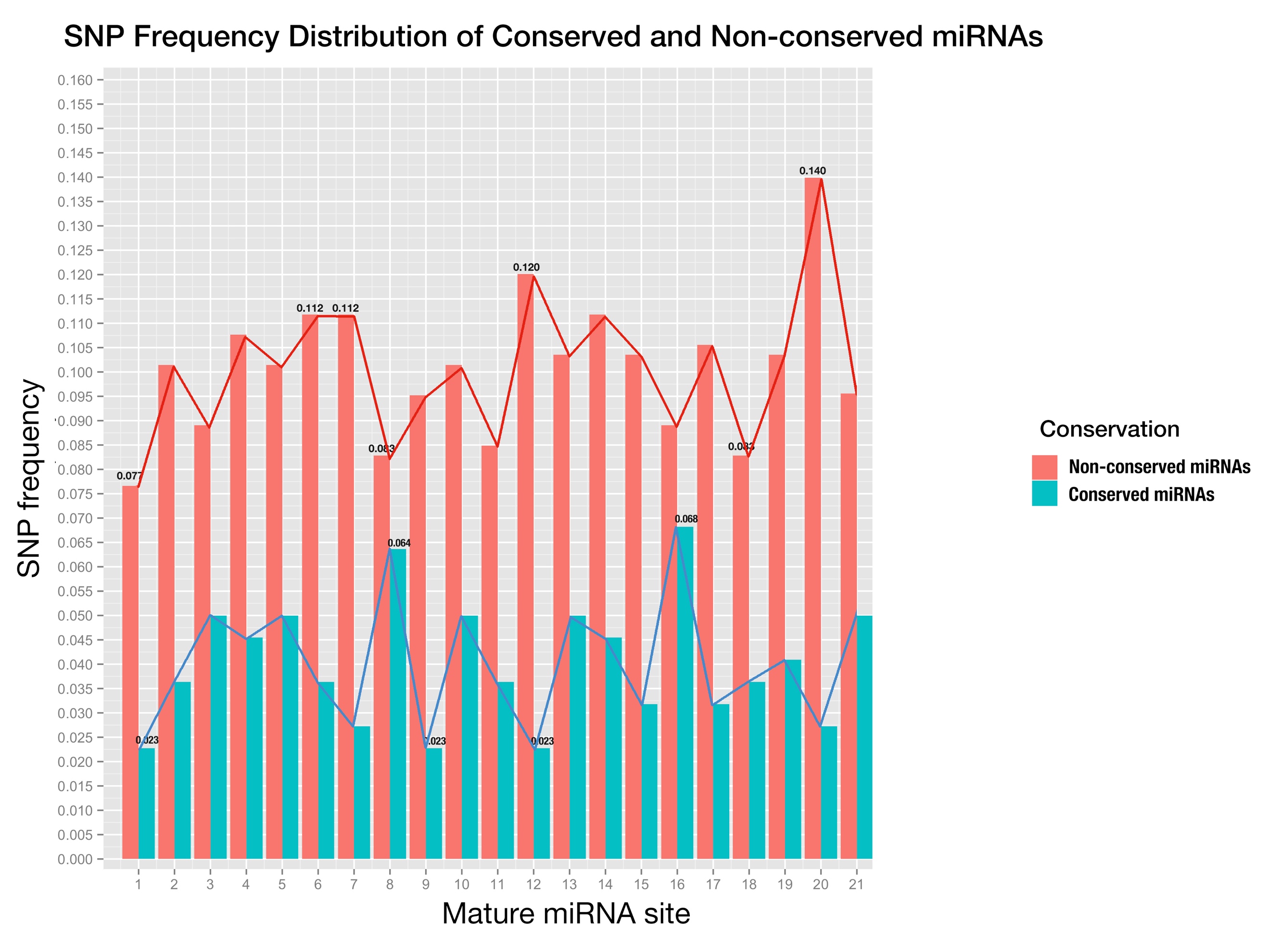
*Bar plot on the bottom right shows the percentage of miRNAs whose SNP densities are below 0.10, 0.08 and 0.05, for conserved and non-conserved miRNAs respectively.*

**Positional SNP distributions are differential between 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 [[[29]](#endnote-29)], as mutations on certain positions can cause complete abortion of silencing while others do not have obvious impact [27, [[30]](#endnote-30), [[31]](#endnote-31)]. 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 only focused 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 supposed to operate through the same silencing machinery, therefore are likely to go through similar mechanistic selection pressure. Moreover, the expected similar rankings of them would produce a linear or near-linear relationship between SNP frequencies of conserved mature miRNAs and those of non-conserved ones’ position-by-position and the linear relationship could be tested by Pearson correlation test. However, this doesn’t appear to be the case. Position 20 ranked the highest SNP frequency position at non-conserved miRNAs, but ranked 4th lowest SNP frequency position at conserved miRNAs; and as for position 8, it ranked 2nd lowest SNP frequency position at non-conserved miRNAs, while for conserved miRNAs, it was the 2nd highest SNP frequency position. Moreover, position 12 ranked 2nd highest SNP frequency position at non-conserved miRNAs, but was the lowest SNP frequency position at conserved miRNAs. Besides, the Pearson correlation test of the SNP frequencies showed no significant correlation between them (r=-0.163, p-value=0.4473). The only exception was position one which had the lowest SNP frequencies for both conserved and non-conserved miRNAs. This may be explained by its known importance in the loading of miRNAs into the AGO proteins [[[32]](#endnote-32), [[33]](#endnote-33)]. To conclude, the differential ranking of SNP frequencies for positions at conserved and non-conserved miRNAs showed conserved and non-conserved miRNAs have gone through differential selection pressure on each position, suggesting they may utilize distinct silencing components to regulate targets.

Perfect pairing to targets at the central position 10 and 11 where transcript cleavage happens is thought to be crucial for the target recognition and subsequent silencing of plant miRNAs [[[34]](#endnote-34)-[[35]](#endnote-35)[[36]](#endnote-36)]. This should add another level of restrict­­­ion on the evolution of those two positions compared to the rest positions, and is likely to result in lower SNP densities. However, the SNP frequencies of both position were not among the lowest positional miRNA SNP frequencies. This is inconsistent with the empirical claims and raised the question whether perfect central match is indispensable for miRNA-mediated silencing in plants again.



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

*Axis X shows the mature miRNA positions from 5’ to 3’end, and axis y is SNP frequency, which is calculated by dividing the number of SNPs at this site by the total number of miRNAs.*

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

Next, we tried to analyze the positional SNP frequencies of miRNA binding site on target genes. We only chose targets of conserved miRNAs, as they are better known and are functionally more important than non-conserved ones. Using online miRNA target prediction tool, *psRNATarget* [24] and degradome validated targets of rice miRNA targets found by Li YF et al. [[[37]](#endnote-37)], a total number of 823 genes were obtained.Since most predicted targets for miRNAs cannot be validated experimentally, we attempted to screen selected targets by correlating their expression with that of corresponding miRNAs, as it is widely accepted that a negative correlation should be observed for miRNA:target interaction pair.

Expression data of conserved rice miRNAs and target genes was obtained from RiceFREND database [[[38]](#endnote-38)]. Firstly, correlation test was performed on selected miRNA:target pairs in 3-week old rice seedling samples, the interactions of which have been validated by degradome previously [37]. Surprisingly, only 136 out of 367 miRNA:target pairs examined exhibited negative correlations (Fig. 6) , which opposes the assumption that there are always negative correlations between *bona fide* targets and cognate miRNAs in tissue they interact. Furthermore, more than half (197 out of 367) pairs fell in the range of weak correlation (-0.4~0.4), which suggests there were no direct correlations between the expression of those pairs. This highlights the unknown complexity of plant miRNA-mediated gene silencing.



**Fig 6. 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.*

Since the negative correlation test failed on *bona fide* miRNA:target interaction pairs, all 823 targets obtained from previous methods were kept for further analyses in this study.

The distribution of positional SNP frequencies along the miRNA binding sites of these genes and their cognate miRNAs were shown side-by-side in Fig. 7. And it was found that all positions of mature miRNA had lower SNP frequencies than those of miRNA binding site. Co-evolution of miRNA and its target binding site had been reported by several studies before [34, [[39]](#endnote-39)]. Thus, the selection histories of them would be reciprocally affected by each other, and moreover, the silencing function of miRNA is determined by the sequences of mature miRNAs and their cognate binding sites co-jointly, it would be properly assumed that the selection pressures imposed on complementary positions of mature miRNAs and binding site are similar. The similarity of selection pressures could be reflected on positive correlation of SNP frequencies of all position on both of them. Position 1 on both sequences has the lowest SNP frequency, and position 13, which has the highest SNP frequency on miRNA binding site, ranked 3rd highest SNP frequency on mature miRNA. And a Pearson correlation test was then performed on the SNP frequencies of conserved mature miRNAs and their binding sites on cognate targets. The result of the test (r=0.5891, p-value=2.455e-3) showed that there was significant positive correlation at moderate level between them. These results supported the assumption concerning relationship between SNP frequencies of mature miRNA and miRNA binding sites.



**Fig 7. 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);*

**Combined Complementarity Pattern Analysis (CCPA) discovered complementarity recovery phenomenon in miR818 family**

Haplotype (haploid genotype), referred as an individual collection of specific mutations within a given genetic segment, is frequently employed in disease association studies and is a powerful discriminator between case and control in these studies. Thus, haplotype helps to divide a population of organisms into groups with different genotypes while studying a specific genetic segment. Hu Zhao et al. [[[40]](#endnote-41)] showed a good example of utilizing haplotype to study rice hybrid sterility and found the rice cultivars that can overcome the sterility of the hybrid were grouped into a single haplotype group.

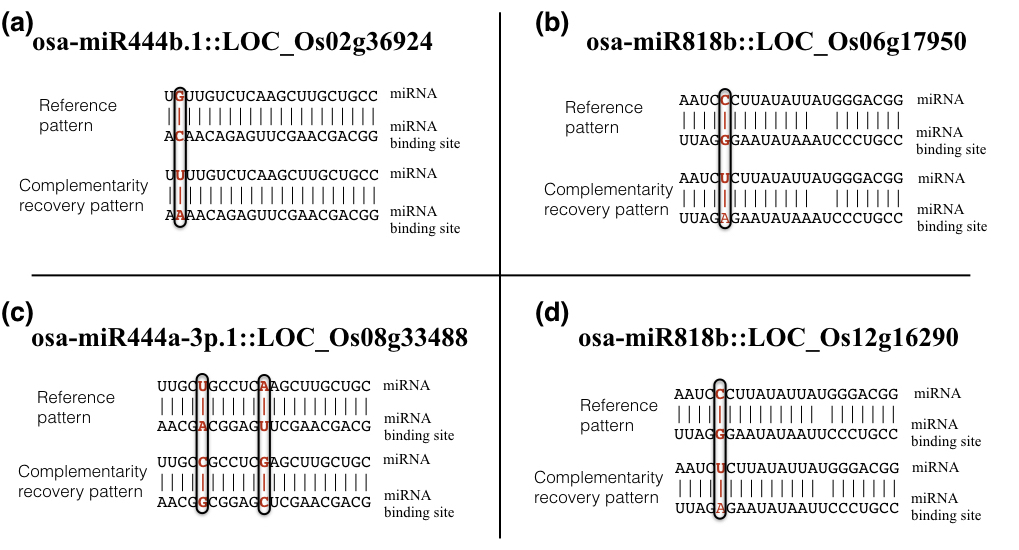
The general procedures of haplotype analysis, in which SNPs are adopted as molecular markers, are firstly gathering up SNPs for a specific genetic segment, secondly dividing the whole population into groups with each group sharing the same alleles for all SNPs gathered. Finally, the features of different groups were compared to see whether SNPs have any links to these features.

Here in this study, the haplotype analysis was extended into Combined Complementarity Pattern Analysis (CCPA) to work for both sequences on mature miRNA and miRNA binding site that are interacting while silencing. Firstly, all SNPs found on both mature miRNA and cognate miRNA binding site, were put together in ascending order according to their id. Then groups, which is defined as a group of alleles at the SNP loci. After that, each haplotype pattern was mapped back to mature miRNA and cognate miRNA binding site, to form a specified genotype for each rice cultivar group. The mutated sequences of mature miRNA and its cognate miRNA binding site were then aligned to form a complementarity pattern. And this pattern was compared with that of reference pattern, of which the rice cultivars have the same genotype as the reference genome concerning these SNPs (Fig. S1).

Generally, complementarity of reference miRNA:target pair can undergo four types of changes when there are SNPs fall within the mature miRNA and/or the miRNA binding site: class 1, Paired to mismatched; class 2, mismatched to paired; class 3, paired to paired; and class 4, (c)mismatched to mismatched. Class 3 and 4 would not change the miRNA complementarity pattern and might bring lesser impact to miRNA-mediated regulation than class 1 and 2. Class 1 mutation would bring in new mismatches to the complementarity pattern and would often weaken the miRNA’s regulation, while class 2 in the reverse direction would increase the complementarity and usually strengthen the miRNA’s regulation. After applying CCPA to all 823 targets with their regulating miRNAs, 74.62% of all mutated sites on complementarity pattern were class 1 mutation, 16.67% were class 2 mutation, 0.68% were class 3 mutation and 8.03% were class 4 mutation. Class 4 mutation, with a very small percentage, required 2 SNPs introduced to the complementarity pattern at the same position on both sequences of mature miRNA and cognate binding site, so was very rare and here we refer it as complementarity recovery phenomenon.

Complementarity recovery phenomenon were observed for osa-miR444 family (Fig. 9 a, c). This is not surprising, as it belongs 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 [[[41]](#endnote-42)]. The retaining of the complementarity between mature miRNA and targets can be explained by their antisense property.

Interestingly, for miRNAs osa-miR818a-e, osa-miR1436, osa-miR1439, osa-miR1442, osa-miR1862b, SNPs were detected for both miRNA and target at matching positions simultaneously, thus complementarity was maintained. According to *miRBase.org*, all these miRNAs including osa-miR818a-e, osa-miR1436, osa-miR1439, osa-miR1442 and osa-miR1862b, belong to the same miRNA family of related hairpin sequences, osa-miR818. Furthermore, among all these pairs where complementarity was maintained, two of them carrying four SNPs, where the possibility to maintain complementarity was as low as 9.3e-5 considering the SNP density of overall rice genome as 0.055 per base, for 3K Rice Genome Project found 23M SNPs in total. The maintenance of complementarity in both cases happened at position 5 at both sequences with double mutations from CG to UA (Fig. 9 b, d). 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 8. Complementarity recovery patterns**

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

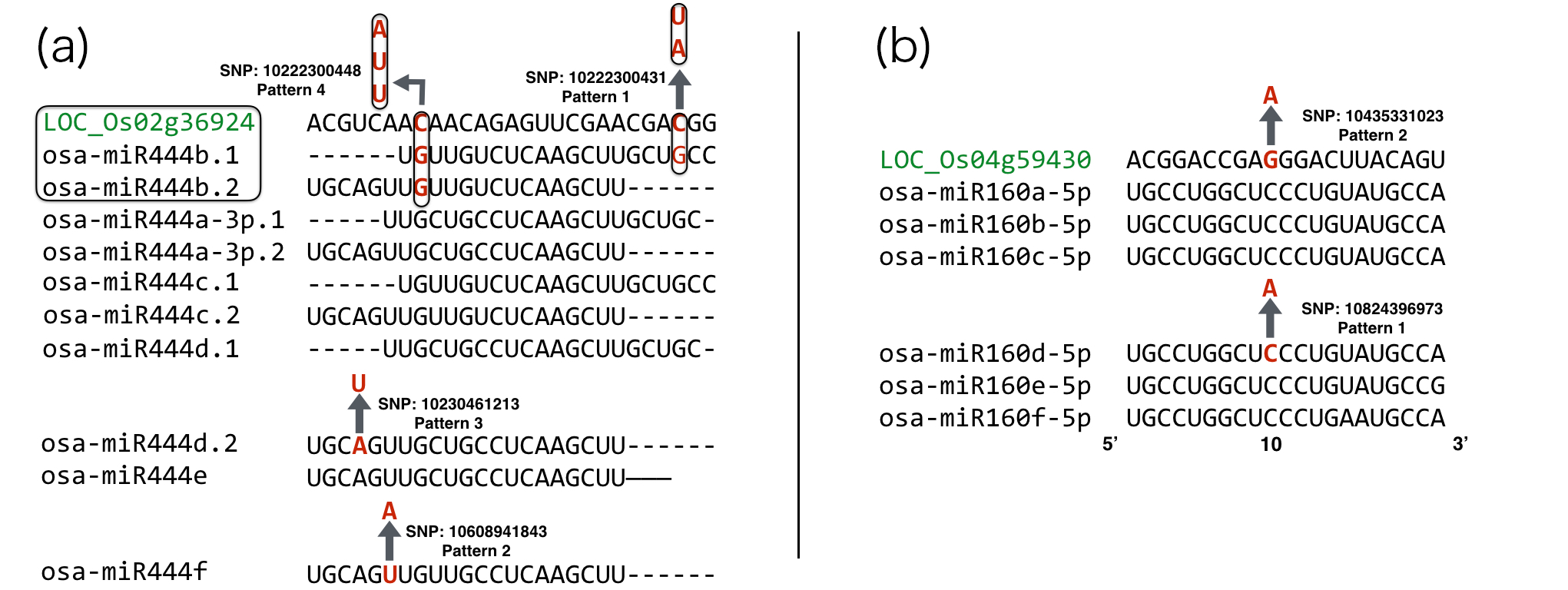
**No clear link can be found between complementary changes of known miRNA:target pairs and distinct phenotypical changes**

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 [[[42]](#endnote-46)-[[43]](#endnote-47)], 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.

Here, we examined almost all well-studied conserved miRNAs (detailed list were given at method part) and seven target genes of the miRNAs families examined were found to carry SNPs on their miRNA binding site. Apart from complementarity, accessibility of target mRNA to cognate miRNAs has also been identified as an important factor that for target recognition [[[44]](#endnote-48)], it is also considered. 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 [[[45]](#endnote-49)]. WHAT ARE THE PROPOSED/KNOWN FUNCTION OF THESE GENES?*

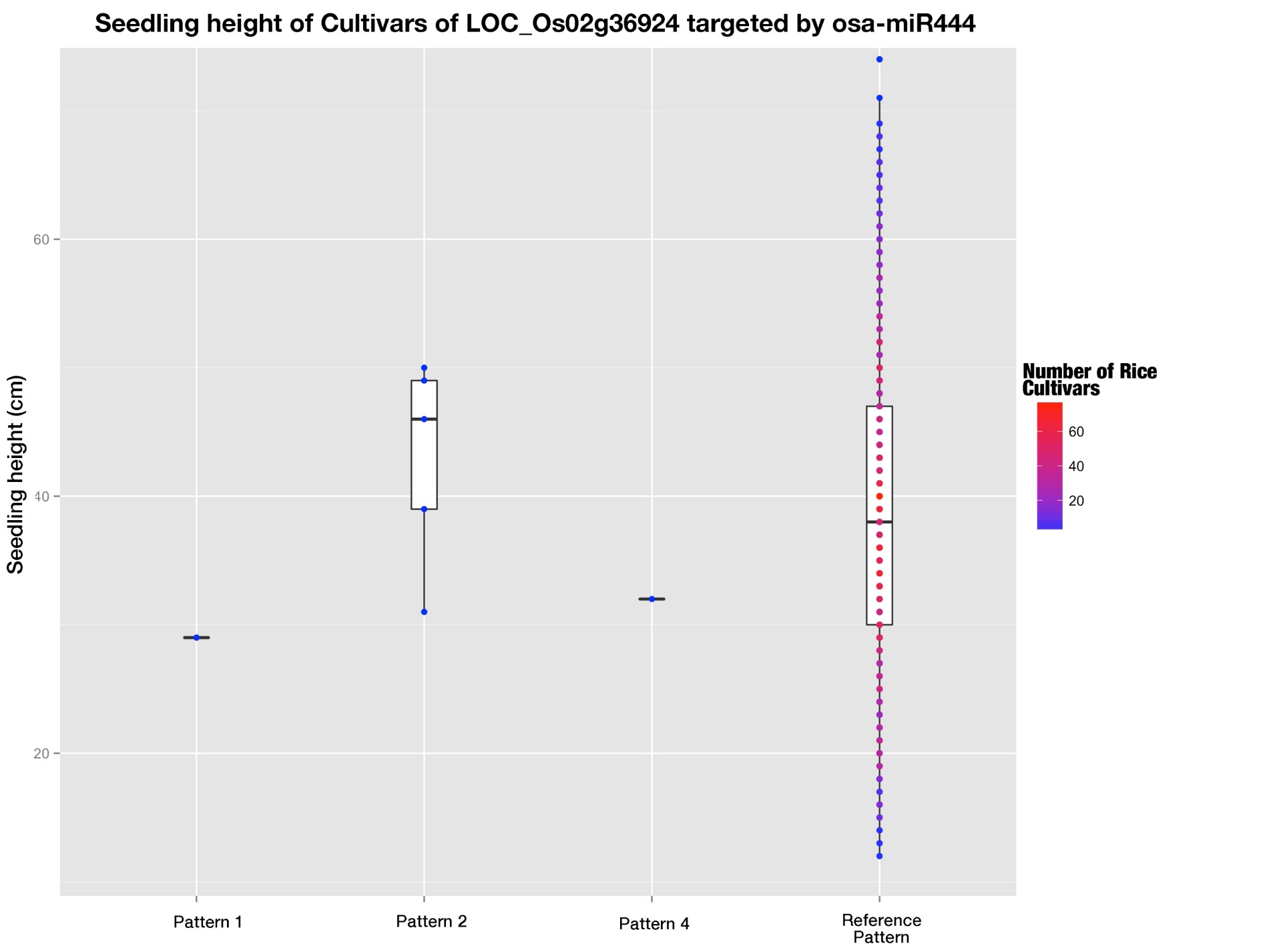
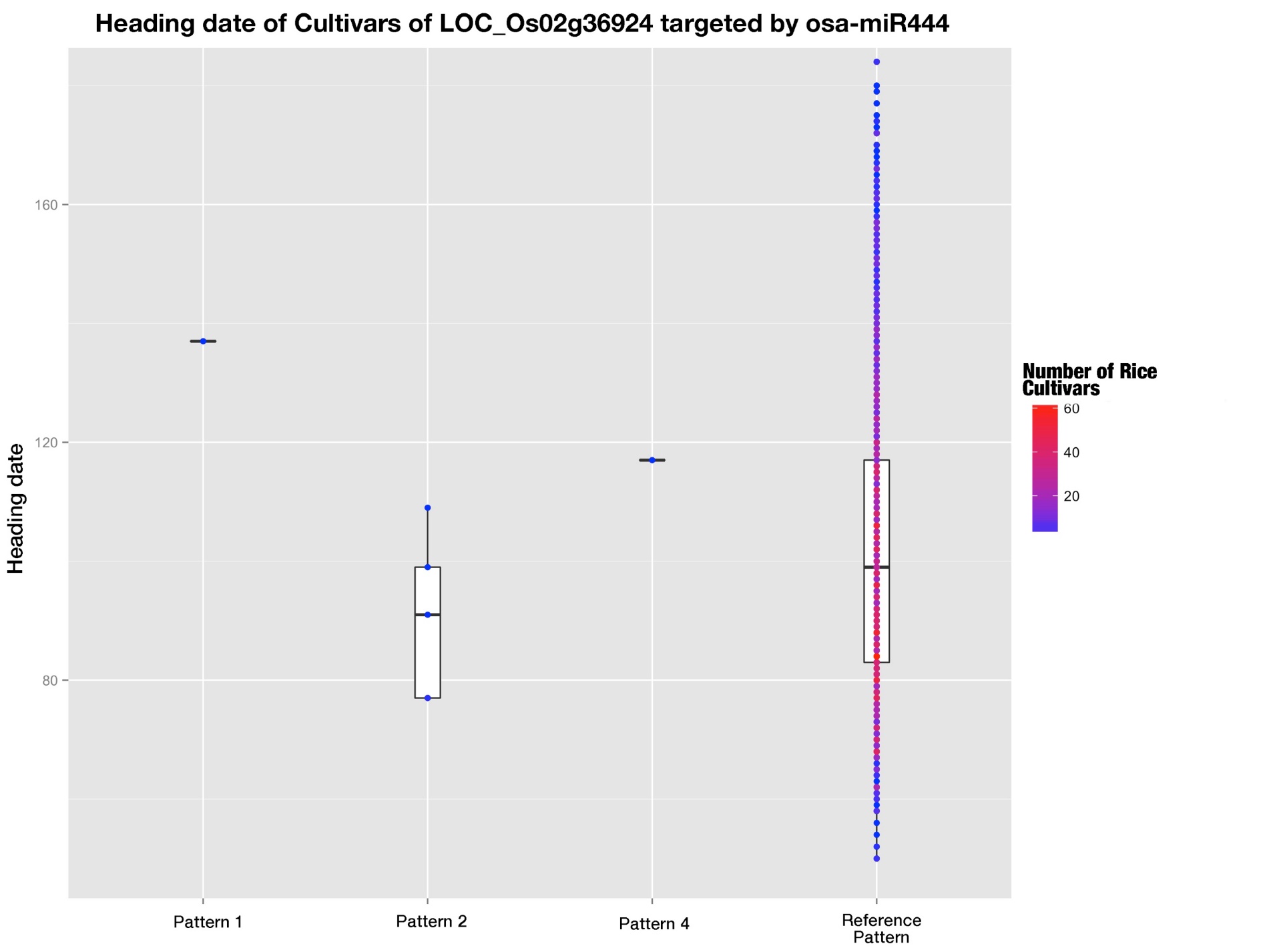


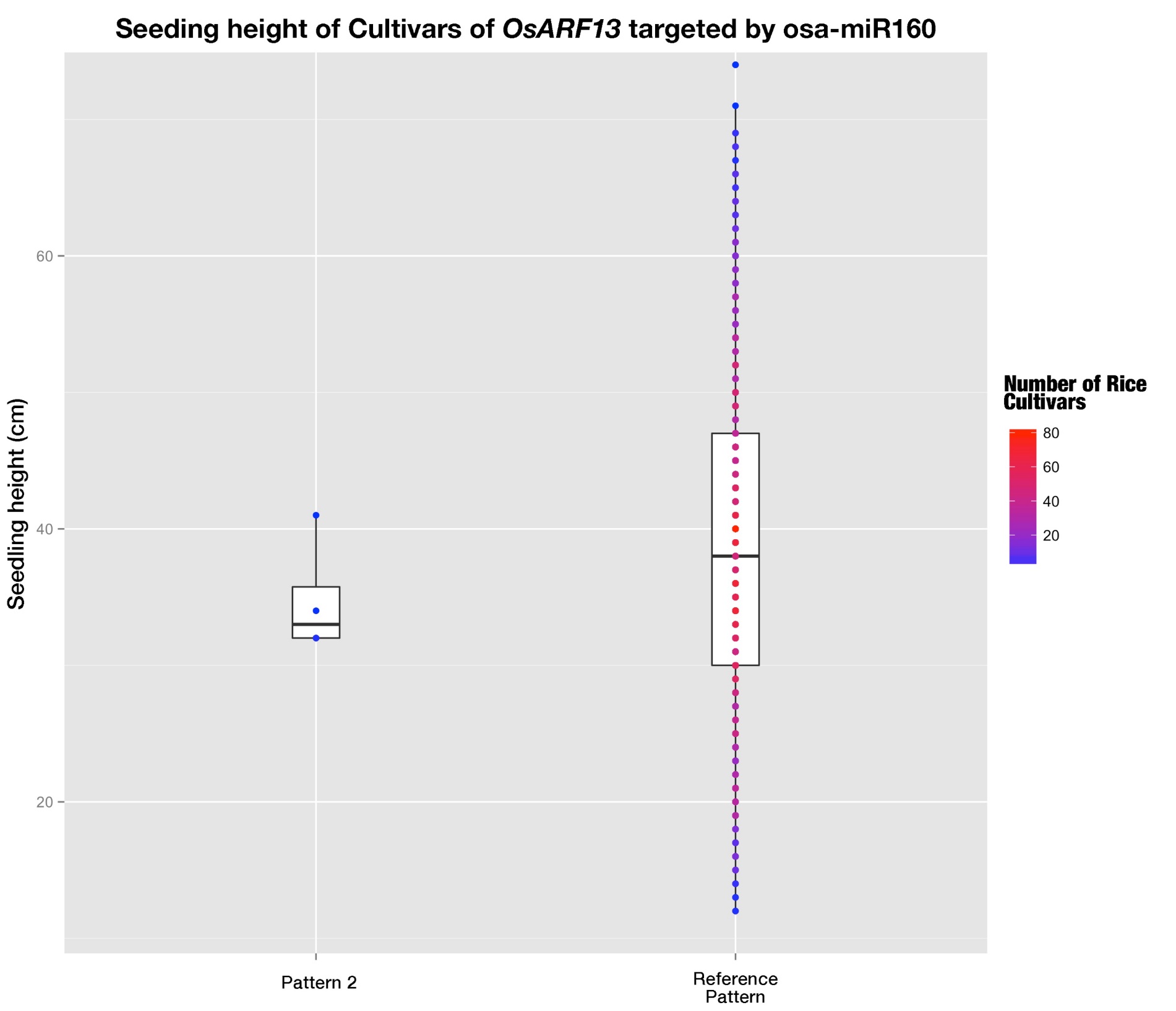
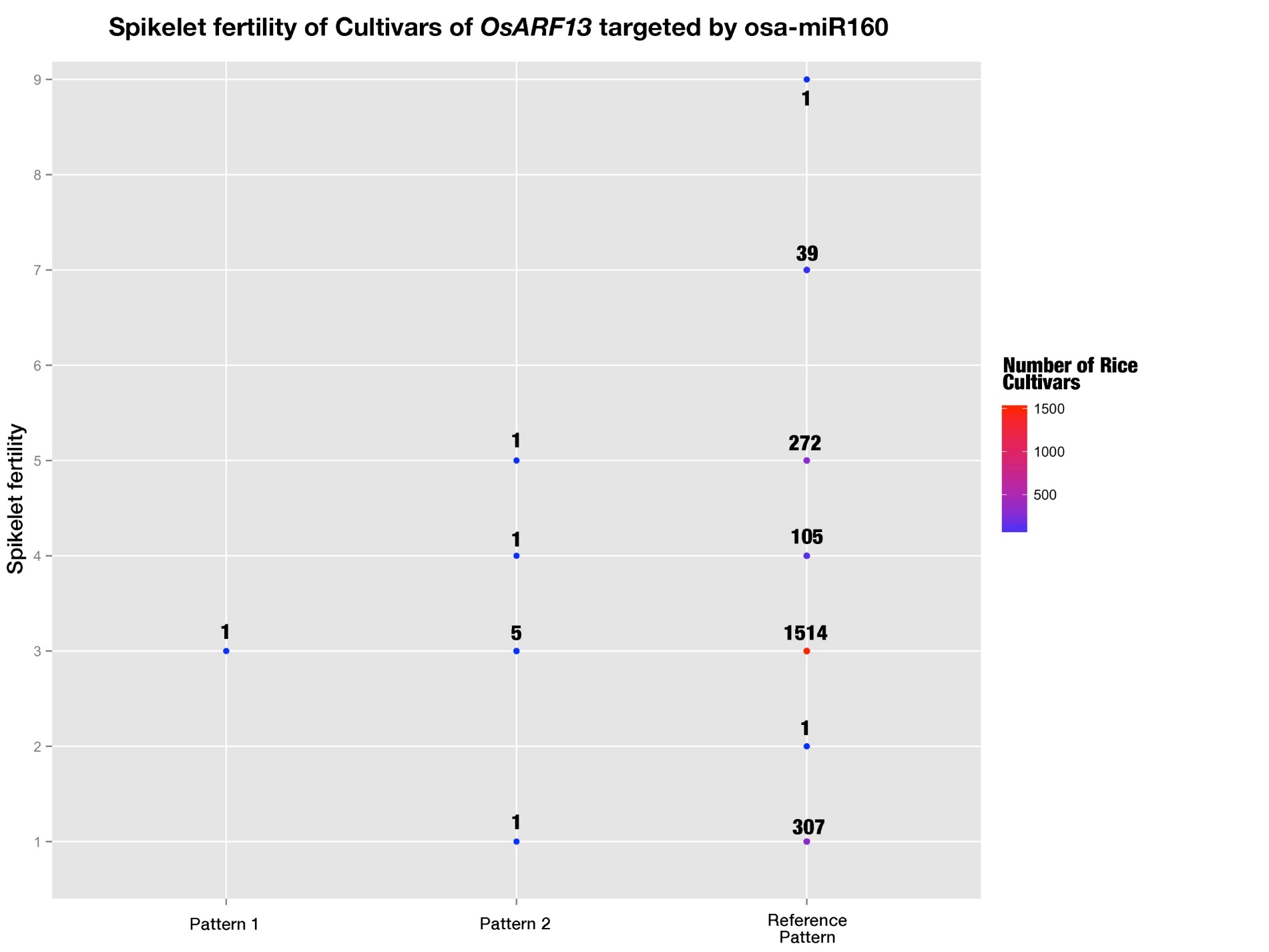
**Fig 9. Complementarity pattern of 2 target genes with their targeting miRNA family**

*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.*

SNP 10435331023 on miRNA binding site of OsARF13 caused a mismatch at position 10, which was at the cleavage site and the SNP was likely to negatively impact its silencing outcome by miRNAs [34-3536]. 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. So, we further looked into the phenotypical changes concerning the above two genes.

Auxin response factors (ARF) in rice were reported to be involved with fertility, height and grain yield [[[46]](#endnote-50)] and rice MADS-box genes were reported to be involved with heading date and plant height [[[47]](#endnote-51)]. 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 10-13).





**Fig 10-13. Phenotypes of rice cultivars belonging to different haplotype patterns**

*The color gradient stands for the number of rice cultivars at that dot. In Fig 10, the numbers of rice cultivars are 1 for pattern 1, 5 for pattern 2, 1 for pattern 4, and 2198 for reference pattern; and in Fig 11, the numbers of rice cultivars are 1 for pattern 1, 5 for pattern 2, 1 for pattern 4, and 1808 for reference pattern. While in Fig 14, the numbers of rice cultivars are 4 for pattern 2 and 1863 for reference pattern.*

As shown in fig. 10-13, though these two genes were found to carry important SNPs which may potentially bring big impact to plant phenotypes, the phenotypes of rice cultivars carrying mutations caused by these SNPs were not different from those of reference patterns. And the predicted changes of outcomes of miRNA regulation didn’t lead to distinct phenotypical change (Fig 11-14). What is your conclusion?

**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 coevolution 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 were obtained through 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 challenges the empirical claim that perfect matches are required for the cleavage sites . 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 filter 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 [[[48]](#endnote-52)]. 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 the possible coevolution history of miRNA and its binding site of cognate target genes, that they reciprocally affect evolution of each other.

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. In addition, in this study, the lack of phenotypical data of some other important agronomic traits reported to be controlled by regulation of miRNAs in question, might also be the cause of no distinct phenotypic changes found.

**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. Thus, it would be more recommended to start the research with changed phenotypes and then deeper into genetic level. 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 selected samples. The selected samples were leaf blade and leaf sheath of 27 day-after-transplanting rice seedlings at both daytime and nighttime.

**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.

In this study, the miRNA families examined to see whether there are SNPs on binding sites of their targets were osa-miR156, 159, 160, 164, 166, 167, 169, 171, 172, 390, 395, 396, 399, 444.

**Supplementary information:**

**FigS1**

**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. Dai, X. and Zhao, P.X., 2011, psRNATarget: a plant small RNA target analysis server. (Nucleic Acids Res.) 39:W155-W159 [↑](#endnote-ref-15)
16. Li, J. et al. 2014, The functional scope of plant microRNA-mediated silencing. (Trends Plant Sci.) 19:785-756. [↑](#endnote-ref-16)
17. Jiao, Y., et al. 2010, Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. (Nat Genet) 42(6): 541-544. [↑](#endnote-ref-17)
18. 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-18)
19. Ehrenreich IM, Purugganan MD, 2008, Sequence variation of microRNAs and their binding sites in Arabidopsis. (Plant Physiol) 146:1974-1982. [↑](#endnote-ref-19)
20. 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-20)
21. 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-21)
22. 3K R.G.P. 2014, The 3,000 rice genomes project. (Gigascience), 3:7. [↑](#endnote-ref-22)
23. Chen, K. & Rajewsky, N., 2006, Natural selection on human miRNA binding sites inferred from SNP data. (Nature Genet.) 38:1452–1456 [↑](#endnote-ref-23)
24. 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-24)
25. Kozomara A, Griffiths-Jones S., 2014, miRBase: annotating high confidence microRNAs using deep sequencing data. (Nucleic Acids Res.) 42:D68-D73 [↑](#endnote-ref-25)
26. Jones-Rhoades MW, 2011, Conservation and divergence in plant microRNAs. (Plant Mol Biol) 80:3–16 [↑](#endnote-ref-26)
27. 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-27)
28. 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-28)
29. 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-29)
30. 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-30)
31. 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-31)
32. 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-32)
33. Mallory A, Vaucheret H, 2010, Form, function, and regulation of ARGONAUTE proteins. (Plant Cell) 22: 3879–3889 [↑](#endnote-ref-33)
34. 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-34)
35. 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-35)
36. 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-36)
37. 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-37)
38. 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-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. Zhao H, Yao W, Ouyang Y, Yang W, Wang G, Lian X, Xing Y, Chen L, Xie W, 2014. RiceVarMap: a comprehensive database of rice genomic variations. (Nucleic Acids Research) 43:D1018–D1022. [↑](#endnote-ref-41)
41. 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-42)
42. 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-46)
43. 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-47)
44. 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-48)
45. 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-49)
46. 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-50)
47. 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-51)
48. 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-52)